

Nonparametric Estimation and Inference about the Overlap of Two Distributions*

Gordon Anderson[†]

University of Toronto

Oliver Linton[‡]

Cambridge University

Yoon-Jae Whang[§]

Seoul National University

Abstract

This paper develops methodology for nonparametric estimation of a measure of the overlap of two distributions based on kernel estimation techniques. This quantity has been proposed as a measure of economic polarization between two groups, Anderson (2004) and Anderson, Ge, and Leo (2006). In ecology it has been used to measure the overlap of species. We give the asymptotic distribution theory of our estimator, which in some cases of practical relevance is nonstandard due to a boundary value problem. We also propose a method for conducting inference based on estimation of unknown quantities in the limiting distribution and show that our method yields consistent inference in all cases we consider. We investigate the finite sample properties of our methods by simulation methods. We give an application to the study of polarization within China in recent years using household survey data from two provinces taken in 1987 and 2001. We find a big increase in polarization between 1987 and 2001 according to monetary outcomes but less change in terms of living space.

Some key words: Kernel Estimation; Inequality; Overlap coefficient; Poissonization; Total Variation

JEL Classification Number: C12, C13, C14, D63

*Thanks to David Mason for helpful comments and providing us with copies of his work. GAUSS programs that carry out the computations in this paper are available from the second author upon request. We thank Sorawoot Srisuma for helpful research assistance.

[†]Gordon Anderson, Department of Economics, University of Toronto, 150 St. George Street, N303, Canada. Email address: anderson@chass.utoronto.ca

[‡]Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD. Email: obl20@cam.ac.uk. Thanks to the ERC for financial support.

[§]Department of Economics, Seoul National University, Seoul 151-746, Korea. Email: whang@snu.ac.kr. This work was supported by the Korea Research Foundation Grant funded by the Korean Government (KRF-2009-327-B00094).

1 Introduction

This paper considers the question of estimation and inference about the scalar parameter

$$\theta = \int_{-\infty}^{\infty} \min\{f(y), g(y)\} dy, \quad (1)$$

where f, g are two densities. This is a number between zero and one with zero corresponding to the distributions having supports with no intersection and one to the perfect matching of the two distributions, i.e., $f(y) = g(y)$ almost everywhere. It follows that θ is a measure of the extent to which the distributions overlap. This quantity was first introduced by Weitzman (1970) in a comparison of income distributions by race. Note that θ is a unit free measure, invariant to a common smooth monotonic transformation. This quantity has received a lot of attention in Medical statistics, Mizuno et. al. (2005), and Ecology, where it is known as the overlap coefficient or the coefficient of community, see for example Ricklefs and Lau (1980). In Economics, this measure has recently been proposed as a measure of the polarization between two well-defined groups, defined by race, occupation, gender, or location of dwelling, see Anderson (2004) and Anderson, Ge, and Leo (2010). In our application we will consider this latter usage, and many of our comments will be addressed to this literature, although our work has something to add to other application areas.

Previous work, Anderson, Ge and Leo (2010), has shown how to estimate θ and conduct inference about this quantity when f, g are parametric, albeit in the very special setting where effectively there are a finite number of cells and the frequency of each cell can be estimated at square root of sample size accuracy. The discretized setting can be expected to lose information in general and raises the spectre of test inconsistency through an injudicious choice of partitions (see Anderson Ge and Leo (2010)). Also, there is no consensus on appropriate parametric models for income distributions for example, see Kleiber and Kotz (2003) for a discussion, and the issue of misspecification bias suggests a nonparametric approach where this can be done effectively. We propose a nonparametric estimator of θ using kernel density estimates of f, g plugged into the population functional. Although these estimates and regular functionals of them are well understood, the population parameter θ is a nonsmooth functional of f, g and so standard methods based on Taylor series expansion cannot be applied to treat the estimator. Schmid and Schmidt (2006) have recently established strong consistency of the same estimator in the univariate compactly supported special case but they do give any distribution theory (they also provide simulation evidence on its performance under various scenarios). We consider a multivariate setting with unbounded support. We find that the properties of the estimated θ can be nonstandard depending on the contact set $\{y : f(y) = g(y) > 0\}$. This set can be empty, it can contain a countable number of isolated points, or it can be a union of intervals. In the first case, the asymptotics are trivial because this implies that one density always lies strictly below the other, and is not very interesting. The second case yields standard normal type asymptotics as in between the contact points one density estimate always prevails. The third case is a sort of ‘boundary value’ case. It is of interest because it corresponds to the case where the

distributions overlap perfectly over some range. This is an important case because one hypothesis of interest is that the two distributions are identical (or identical over a range) as one might believe in some applications. In that case there are binding inequality restrictions, which may be expected to induce non-normal asymptotics. We show the distribution theory for this latter case using some Poissonization techniques due to Beirlant and Mason (1995) and Giné, Mason, and Zaitsev (2003). It turns out that the limiting distribution is normal after a bias correction. In practice, we do not know which of these three cases arises and so our inference method should be robust to these different possibilities. In addition, it can be that the two densities while not identical are close to each other over a range of values, the so-called drifting distribution case, Andrews and Guggenberger (2009), and this would induce a distortion in the usual asymptotic approximation. We develop an analytical approach to inference and show that it yields consistent inference whatever the nature of the contact set. We are also interested in testing various hypotheses about θ such as it is equal to one, which corresponds to the two densities being equal, or that it is constant over time, and provide the theory to carry out these tests.

We investigate the finite sample properties of our methods by simulation methods, and show that the finite sample bias can be large and that our bias correction method improves matters. We also investigate the choice of smoothing and tuning parameters that our procedures require. We give an application to the study of polarization within China in recent years, which has been an important policy issue, using novel household survey data.

Finally, we remark that the theory we develop here can be applied to a range of other estimation and testing problems and we give a discussion of this below.

2 Estimation

We first discuss some properties of the parameter of interest and the connection with other concepts. Suppose the population random variables X and Y have joint density h and marginal densities f and g , respectively. Then

$$\theta = 1 - \frac{1}{2} \int_{-\infty}^{\infty} |f(x) - g(x)| dx = 1 - \frac{1}{2} TV, \quad (2)$$

where TV denotes total variation. This shows that $1 - \theta$ defines an L_1 pseudometric on the space of densities. Note also that θ is invariant to monotonic transformations of X, Y , that is, if $\tau_X = \tau(X)$ and $\tau_Y = \tau(Y)$ for a strictly increasing differentiable transformation τ , and τ_X and τ_Y have densities f_τ and g_τ , then $\theta = \int_{-\infty}^{\infty} \min\{f_\tau(t), g_\tau(t)\} dt$ by standard application of the law of transformation. This is a big advantage of this measure compared with L_2 or other type metrics, as this is the only member of the L_p class that possesses this property, Devroye and Györfi (1985). The representation (2) shows that θ can be used as a measure of the equality of the two densities; a special case of our results can be used to conduct a test of this hypothesis, which constitutes a two-sample version of the L_1 testing procedure considered in Giné, Mason, and Zaitsev (2003).

An alternative representation of θ is as an expectation

$$\theta = E [\min \{1, \ell_{g,f}(X)\}] = E [\min \{1, \ell_{f,g}(Y)\}], \quad (3)$$

where $\ell_{g,f}(x) = g(x)/f(x)$ is the likelihood ratio, which can be convenient for computing estimators, see below, and see Schmid and Schmidt (2006) where more discussion of (2) and (3) is given.

We will assume a multivariate setting where X, Y are d -dimensional vectors. In this case we shall assume that the integral is over all of the variables. It is also possible to consider integrating with respect to a subset of variables or to consider conditional densities, but we shall leave that for future work.

We suppose that there is a sample $\{(X_1, Y_1), \dots, (X_n, Y_n)\}$ of size n on the population. In some cases one might have different sizes n, m for the two samples; we will comment on this case later (see Remark 5 to Theorem 1), but for now treat the central theoretical issue by considering the same sample size. We propose to estimate θ by

$$\hat{\theta} = \int_C \min\{f_n(x), g_n(x)\} dx = \int_C \min\left\{1, \frac{g_n(x)}{f_n(x)}\right\} f_n(x) dx, \quad (4)$$

$$f_n(x) = \frac{1}{n} \sum_{i=1}^n K_b(x - X_i) \quad ; \quad g_n(x) = \frac{1}{n} \sum_{i=1}^n K_b(x - Y_i),$$

where $C \subseteq \mathbb{R}^d$ is the union of the supports or some subset of interest, while K is a multivariate kernel and $K_b(\cdot) = K(\cdot/b)/b^d$ and b is a bandwidth sequence. For simplicity we suppose that the same bandwidth and kernel are used in both estimations and at each point x . When $K \geq 0$, $f_n(x), g_n(x) \geq 0$. When X, Y have unbounded support, $\int f_n(x) dx = \int g_n(x) dx = 1$. There is an issue about boundary effects in the case where the support is compact and the densities are positive on the boundary. In that case one might use some boundary correction method, see Chen (1999). In practice one has to compute a multivariate integral in (4) and a simple approach is to just replace $\hat{\theta}$ by a sample average over a set of grid points on the support. Alternatively, one can take the sample average over the observations of the empirical version of (3), $\hat{\theta}_E = n^{-1} \sum_{i=1}^n \min\{1, g_n(X_i)/f_n(X_i)\} = \int_C \min\{1, g_n(x)/f_n(x)\} dF_n(x)$, where F_n is the empirical distribution of X_i . Note that this defines a different estimator from (4), since the integrating measure in (4) is the smoothed empirical. However, the differences in terms of the asymptotic properties are of smaller order and can be ignored when conducting inference. Schmid and Schmidt (2006) consider these and a number of other estimators and find not much difference between the estimators in simulation experiments.

3 Asymptotic Properties

We next discuss the asymptotic behavior of $\hat{\theta}$ as $n \rightarrow \infty$. We treat the case where X, Y have unbounded support $\mathbb{R}^d$ as this is more challenging and perhaps of more interest for applications

(a common parametric specification here is the log normal distribution, for which the logarithm of outcome has unbounded support).¹ Schmid and Schmidt (2006) have recently established consistency of $\hat{\theta}$ in the univariate compactly supported special case. We remark that as usual our assumptions are sufficient for the result but may not all be necessary.

We use the following notation. Define the contact set and its complements:

$$C_{f,g} = \{x \in \mathbb{R}^d : f(x) = g(x) > 0\}, \quad (5)$$

$$C_f = \{x : f(x) < g(x)\}, \quad (6)$$

$$C_g = \{x : f(x) > g(x)\}. \quad (7)$$

Let $\lambda = (\lambda_1, \dots, \lambda_d)^\top$ denote a vector of nonnegative integer constants. For such vector, we define $|\lambda| = \sum_{i=1}^d \lambda_i$ and, for any function $h(x) : \mathbb{R}^d \rightarrow \mathbb{R}$, $D^\lambda h(x) = \partial^{|\lambda|} / (\partial x_1^{\lambda_1} \dots \partial x_d^{\lambda_d})(h(x))$, where $x = (x_1, \dots, x_d)^\top$ and $x^\lambda = \prod_{j=1}^d x_j^{\lambda_j}$. For a Borel measurable set $A \subset \mathbb{R}^d$, we define $\mu(A)$ to be the Lebesgue measure of A and

$$\mu_f(A) = \int_A f^{1/2}(x) dx.$$

Let

$$\|K\|_2^2 = \int_{\mathbb{R}^d} K^2(u) du \text{ and } \rho(t) = \int_{\mathbb{R}^d} K(u) K(u+t) du / \|K\|_2^2.$$

ASSUMPTIONS

(A1) K is a s -th order kernel function having support in the closed ball of radius $1/2$ centered at zero, symmetric around zero, integrates to 1, and s -times continuously differentiable on the interior of its support, where s is an integer that satisfies $s > d$.

(A2) (i) The densities h , f and g are strictly positive, bounded and absolutely continuous with respect to Lebesgue measure and s -times continuously differentiable with uniformly bounded derivatives. (ii) $\int_{C_{f,g}} f^{1/2}(x) dx < \infty$. (iii) For all λ with $0 \leq |\lambda| \leq s$, $\int |D^\lambda f(x)| dx, \int |D^\lambda g(x)| dx < \infty$.

(A3) The bandwidth satisfies: (i) $nb^{2s} \rightarrow 0$, (ii) $nb^{2d} \rightarrow \infty$ and (iii) $nb^d / (\log n) \rightarrow \infty$.

(A4) $\{(X_i, Y_i) : i \geq 1\}$ are i.i.d. with support $\mathbb{R}^d \times \mathbb{R}^d$.

Assumption A1 requires the kernel function to have a compact support. This assumption facilitates the use of the Poissonization technique in our proofs, see Appendix A.1 for an intuitive explanation. The requirement in A2(ii) that

$\int_{C_{f,g}} f^{1/2}(x) dx = E[f(X)^{-1/2} 1(f(X) = g(X))] < \infty$ rules out the case where both f, g are the same Cauchy density since $\int_{C_{f,g}} f^{1/2}(x) dx = \infty$ in this case; condition A2(ii) is implied by the condition that $E[\|X\|^{1+\eta} 1(f(X) = g(X))] < \infty$ for some $\eta > 0$. Pareto-like tails are quite common in applications, for example income distributions, see Champernowne and Cowell (1998), but a Cauchy density

¹This implicitly rules out the case $\theta = 0$. This case corresponds to the supports of X, Y not overlapping. One can test this hypothesis by different methods.

would be quite extreme from the point of view of most applications. Assumption A3 is needed for the asymptotic normality. To fulfil A3 we require that $s > d$; in the univariate case, it suffices to have twice differentiable densities and bandwidth in the range $n^{-1/4}$ to $n^{-1/2}$, i.e., undersmoothing relative to what would be the optimal bandwidth for estimation of the densities themselves but not too much undersmoothing, this is a common requirement in semiparametric problems. In A4 we assume that the sample is i.i.d. but we allow the random vectors X and Y to be mutually dependent, which may be important in some applications, see Linton, Maasoumi and Whang (2005).

By the triangle inequality we can bound $|\hat{\theta} - \theta|$ by $\int |f_n(x) - f(x)| dx + \int |g_n(x) - g(x)| dx$, so that under weaker conditions than A3, specifically just $b \rightarrow 0$ and $nb^d \rightarrow \infty$, we have $\hat{\theta} - \theta = O_p(b^s) + O_p(n^{-1/2}b^{-d/2})$, which yields consistency. Our results below establish the further result of asymptotic normality.

Define:

$$p_0 = \Pr(X \in C_{f,g}) = E[1(f(X) = g(X))] = \Pr(Y \in C_{f,g}) = E[1(f(Y) = g(Y))] \quad (8)$$

$$p_f = \Pr(X \in C_f) = E[1(f(X) < g(X))], \quad p_g = \Pr(Y \in C_g) = E[1(f(Y) > g(Y))] \quad (9)$$

$$p_{f,g} = \Pr(X \in C_f, Y \in C_g) = E[1(f(X) < g(X)) 1(f(Y) > g(Y))] \quad (10)$$

$$\sigma_0^2 = \|K\|_2^2 \int_{T_0} \text{cov}\left(\min\{Z_1, Z_2\}, \min\left\{\rho(t)Z_1 + \sqrt{1-\rho(t)^2}Z_3, \rho(t)Z_2 + \sqrt{1-\rho(t)^2}Z_4\right\}\right) dt$$

$$\sigma_1^2 = p_f(1-p_f) + p_g(1-p_g) + 2(p_{f,g} - p_f p_g),$$

$$a_n = b^{-d/2} \|K\|_2 \int_{C_{f,g}} f^{1/2}(x) dx \cdot E \min\{Z_1, Z_2\}, \quad v = p_0 \sigma_0^2 + \sigma_1^2,$$

where Z_1, Z_2, Z_3 , and Z_4 are independent standard normal random variables and $T_0 = \{t \in \mathbb{R}^d : \|t\| \leq 1\}$.

THEOREM 1. *Suppose that Assumptions A1-A4 hold. Then, we have:*

$$\sqrt{n}(\hat{\theta} - \theta) - a_n \implies N(0, v).$$

Remarks.

1. The bias term a_n depends on the integral of the square root of either density over the contact set, and this is non zero whenever this set has some measure. In fact, $E \min\{Z_1, Z_2\} = -0.56$ and so $a_n \leq 0$, so that the estimator is downward biased. The bias can be arbitrarily large depending on the magnitude of $\int_{C_{f,g}} f^{1/2}(x) dx$. We show below how to compute a feasible bias corrected estimator that achieves root-n consistency, but to do that we will require additional conditions.

2. The limiting variance depends on the magnitudes of the sets $C_{f,g}$, C_f , and C_g under the relevant probability measures along with constants that just depend on the kernel chosen. It is not known what is the optimal kernel here, but we suspect that the uniform kernel is optimal due to its minimum variance property. We have calculated σ_0^2 for various kernels in the univariate case and present the results below:

*** Table 1 here ***

In the special case that the contact set is of zero measure, $p_0 = 0$ and $a_n = 0$ so that $\sqrt{n}(\hat{\theta} - \theta) \Rightarrow N(0, \sigma_1^2)$. This asymptotic variance is actually the semiparametric efficiency bound for the case where the sets C_f and C_g are known, so that $\hat{\theta}$ is fully efficient in this case.

3. Suppose that the two densities are not identical but close to each other. Specifically, suppose that for all x , $f(x) = g(x) + n^{-\alpha}\delta(x)$, for $\alpha \geq 1/2$, where of necessity $\int \delta(x)dx = 0$. Then $\theta_n = \int \min\{f(x), g(x)\}dx = 1 + n^{-\alpha} \int 1(\delta(x) < 0)\delta(x)dx < 1$. In this case, the contact set $C_{f,g}$ may be empty. First, consider the case $\alpha > 1/2$. Under some conditions, we have $\sqrt{n}(\hat{\theta} - \theta_n) - a_n \Rightarrow N(0, v)$, so exactly the same limiting distribution result as in Theorem 1. In this case, note that even though the contact set may be empty, the same asymptotic bias shows up. The second situation corresponds to larger deviations $\alpha = 1/2$. Then $\sqrt{n}(\hat{\theta} - \theta_n) - a_n \Rightarrow N(c_0, v)$, where $c_0 = -\int_{\mathbb{R}^d} 1(\delta(x) < 0)\delta(x)dx > 0$. In this case, the limiting distribution is shifted according to the precise form of δ .²

4. The proof of Theorem 1 uses the decomposition of the estimation error into three stochastic terms plus a remainder term:

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta) &= \sqrt{n} \int_{C_f} \{f_n(x) - Ef_n(x)\}dx + \sqrt{n} \int_{C_g} \{g_n(x) - Eg_n(x)\}dx \\ &\quad + \sqrt{n} \int_{C_{f,g}} \min\{f_n(x) - Ef_n(x), g_n(x) - Eg_n(x)\}dx + R_n, \end{aligned} \quad (11)$$

where $R_n = O_p(\sqrt{nb^s}) = o_p(1)$. The first two terms are more or less standard in the semiparametric literature as integrals of semiparametric estimators over some domain. The final term is what causes the main issue, at least when $C_{f,g}$ has positive measure. This term is similar in spirit to what is obtained in other boundary estimation problems, Andrews (1999). For example, consider the problem of estimating $\beta = \min\{\mu_X, \mu_Y\}$, where $\mu_X = EX$ and $\mu_Y = EY$. When $\mu_X = \mu_Y$, the usual estimator $\hat{\beta} = \min\{\bar{X}, \bar{Y}\}$ satisfies $\sqrt{n}(\hat{\beta} - \beta) = \min\{\sqrt{n}(\bar{X} - \mu_X), \sqrt{n}(\bar{Y} - \mu_Y)\} \Rightarrow \min\{Z_X, Z_Y\}$, where $[\sqrt{n}(\bar{X} - \mu_X), \sqrt{n}(\bar{Y} - \mu_Y)] \Rightarrow [Z_X, Z_Y] = Z$ and Z is bivariate normal with zero mean. In this case, the limiting distribution of $\hat{\beta}$ has a negative mean and is non-normal. In the case of $\hat{\theta}$ there is a negative bias term but after subtracting that off one has asymptotic normality. The intuitive reason is that our estimator involves averages of approximately independent random variables. The formal justification though is more complex because the behavior of the stochastic process $\nu_n(x) = [f_n(x) - Ef_n(x), g_n(x) - Eg_n(x)]$ for $x \in C_{f,g}$ is not standard. If $f_n(x), g_n(x)$ were c.d.f.'s we could apply the functional central limit and continuous mapping theorems to obtain the limiting distribution, but this is not available here even at the slower rate of the pointwise convergence of $\nu_n(x)$ because of a lack of tightness. If $\nu_n(x)$ and $\nu_n(x')$ for $x \neq x'$ were independent we could instead argue that $\int_{C_{f,g}} \min\{f_n(x) - Ef_n(x), g_n(x) - Eg_n(x)\}dx$ is like a sum of independent

²One can alternatively represent this limiting result as $\sqrt{n}(\hat{\theta} - 1) - a_n \Rightarrow N(0, v)$, which shows that centering at one gives the same asymptotic distribution as in Theorem 1.

random variables and apply a central limit theorem after recentering. Although $\nu_n(x)$ and $\nu_n(x')$ are asymptotically independent for $x \neq x'$ they are not exactly so and in any case the integral requires we treat also the case where $x - x' = tb$ for $||t|| \leq 1$, and for such sequences $\nu_n(x)$ and $\nu_n(x')$ can be highly dependent. In Appendix B we provide further discussion of this point. The argument to exploit asymptotic independence and establish normality is based on the so-called Poissonization, which was originally used by Kac (1949). The idea behind Poissonization is that the behavior of a fixed population problem should be close to that of the same problem under a Poisson model having the fixed population problem size as its mean. The additional randomness introduced by Poissonization allows for application of techniques that exploit the independence of the increments and the behavior of moments. The Poissonization technique generally requires weaker assumptions on the underlying densities and the parameters of estimates than the strong Gaussian approximation, see Rosenblatt (1975, p.7) and Berline (1999, p.91) for more discussions on its advantages. This technique has also been used in a number of places including combinatorial mathematics and analysis of algorithms. It can be useful in a number of econometrics contexts, as we discuss further below.

5. In the application we consider below, and in other applications, there are different sample sizes n and m that apply to the estimation of f and g . The distribution theory for this case is only a trivial modification of the theory presented above. In particular, suppose that the two samples are mutually independent and that $m/n \rightarrow \tau \in (0, \infty)$. Then the asymptotic distribution is as in Theorem 1 with

$$\begin{aligned} a_n &= b^{-d/2} ||K||_2 \int_{C_{f,g}} f^{1/2}(x) dx \cdot E \min \{Z_1, Z_2/\tau\} \\ \sigma_1^2 &= p_f(1 - p_f) + p_g(1 - p_g)/\tau \\ \sigma_0^2(\tau) &= ||K||_2^2 \int_{T_0} \text{cov} \left(\min \{Z_1, Z_2/\tau\}, \min \left\{ \rho(t)Z_1 + \sqrt{1 - \rho(t)^2}Z_3, \rho(t)Z_2/\tau + \sqrt{1 - \rho(t)^2}Z_4/\tau \right\} \right) dt, \end{aligned}$$

where $E \min \{Z_1, Z_2/\tau\} = \frac{\sqrt{1+1/\tau}}{2} E \min \{Z_1, Z_2\} = -\frac{\sqrt{1+1/\tau}}{2} \sqrt{\frac{2}{\pi}}$ and $\sigma_0^2(\tau) = \sigma_0^2(1)(1 + 1/\tau)/2$.

6. By the relationship (2), the result of Theorem 1 implies that the estimator $\widehat{TV}$ of the L_1 (or total variation) distance $TV = \int_{-\infty}^{\infty} |f(x) - g(x)| dx$ between two densities f, g satisfies

$$\sqrt{n} \left(\widehat{TV} - TV \right) + 2a_n \implies N(0, 4v),$$

where

$$\widehat{TV} = \int_{-\infty}^{\infty} |f_n(x) - g_n(x)| dx.$$

We next discuss how to conduct consistent inference on the parameter θ using the theory presented in Theorem 1. For inference we must estimate consistently the quantities p_0 , p_f , $p_{f,g}$ and p_g , and estimate $\int_{C_{f,g}} f^{1/2}(x) dx$ consistently at a better rate than $b^{d/2}$. We require some additional conditions:

ASSUMPTIONS

(A5) Whenever $\mu(C_{f,g}) > 0$, the densities f and g satisfy $h^*(\varepsilon) := \mu_f(\{x : 0 < |f(x) - g(x)| \leq \varepsilon\}) = O(\varepsilon^\gamma)$ as $\varepsilon \rightarrow 0$ for some positive constant γ .

(A6) (i) $E[\|X\|^p 1(f(X) = g(X))] < \infty$ for some $p > 2$ such that $n^{(p-2)/2p} b^d \rightarrow \infty$. (ii) For all λ with $0 \leq |\lambda| \leq s$, $\int_{C_{f,g}} |D^\lambda f(x)|/f^{1/2}(x) dx < \infty$.

(A7) There is a sequence $c_n \rightarrow 0$ such that $nb^{2d}c_n^2/\log n \rightarrow \infty$ and $nb^d c_n^{2+2\gamma} \rightarrow 0$, where γ is the positive constant in A5 when $\mu(C_{f,g}) > 0$, and $\gamma = \infty$ otherwise.

Assumption A5 controls the behavior of the density functions near the boundary of the contact set $C_{f,g}$. It has to do with the sharpness in the decrease of $\phi = f - g$ to zero, see Härdle, Park and Tsybakov (1995), Hall (1982), and Cuevas and Fraiman (1997) for related concepts. It is like a tail thickness condition except that it only applies in the vicinity of $C_{f,g}$. If ϕ is bounded away from zero outside of $C_{f,g}$, then γ can be set to be ∞ . For small values of γ , this imposes a binding restriction on bandwidth and smoothness rather like dimensionality of the covariates. Assumption A5 is used to get a consistent estimator of the centering term a_n in Theorem 2 below. The condition A6(i) is needed in the case where $C_{f,g} = \mathbb{R}^d$ as it is used to bound the estimation error of $\int_{C_{f,g}} f^{1/2}(x) dx = E[f^{-1/2}(X) 1(f(X) = g(X))]$, which can be badly affected by heavy tails. It imposes a further restriction on the bandwidth: for small values of p one must take a large bandwidth to make $n^{(p-2)/2p} b^d \rightarrow \infty$ but then one needs s to be very large to ensure that $nb^{2s} \rightarrow 0$. If X, Y are Gaussian ($p = \infty$), then only an additional logarithmic constraint is imposed on the bandwidth. Condition A7 implicitly imposes a stronger restriction on the bandwidth than A3. Generally there is both an upper and lower bound on the tuning parameter; in the case that $\gamma = \infty$ there is only a lower bound on the tuning parameter, e.g. when $d = 1$, $s = 2$, and $b \propto n^{-1/3}$, c_n should be of larger order than $n^{-1/6}$.

Define the bias corrected estimator and asymptotic variance estimator:

$$\begin{aligned}\hat{\theta}^{bc} &= \hat{\theta} - \hat{a}_n/n^{1/2}, \quad \hat{v} = \hat{p}_0\sigma_0^2 + \hat{\sigma}_1^2 \\ \hat{a}_n &= -0.56 \frac{\|K\|_2}{2b^{d/2}} \left(\int_{\hat{C}_{f,g}} f_n^{1/2}(x) dx + \int_{\hat{C}_{f,g}} g_n^{1/2}(x) dx \right) \\ \hat{C}_f &= \{x \in \mathbb{R}^d : f_n(x) - g_n(x) < -c_n, f_n(x) > 0, g_n(x) > 0\} \\ \hat{C}_g &= \{x \in \mathbb{R}^d : f_n(x) - g_n(x) > c_n, f_n(x) > 0, g_n(x) > 0\} \\ \hat{C}_{f,g} &= \{x \in \mathbb{R}^d : |f_n(x) - g_n(x)| \leq c_n, f_n(x) > 0, g_n(x) > 0\} \\ \hat{p}_0 &= \frac{1}{2n} \sum_{i=1}^n \{1(X_i \in \hat{C}_{f,g}) + 1(Y_i \in \hat{C}_{f,g})\} \\ \hat{p}_f &= \frac{1}{n} \sum_{i=1}^n 1(X_i \in \hat{C}_f), \quad \hat{p}_g = \frac{1}{n} \sum_{i=1}^n 1(Y_i \in \hat{C}_g), \quad \hat{p}_{f,g} = \frac{1}{n} \sum_{i=1}^n 1(X_i \in \hat{C}_f, Y_i \in \hat{C}_g) \\ \hat{\sigma}_1^2 &= \hat{p}_f(1 - \hat{p}_f) + \hat{p}_g(1 - \hat{p}_g) + 2(\hat{p}_{f,g} - \hat{p}_f\hat{p}_g).\end{aligned}$$

Then, we have the following result:

THEOREM 2. *Suppose that Assumptions A1-A7 hold. Then, we have:*

$$\sqrt{n} \left(\hat{\theta}^{bc} - \theta \right) \Rightarrow N(0, v) \quad (12)$$

$$\hat{v} \xrightarrow{p} v. \quad (13)$$

This theorem can be used to construct hypothesis tests about and consistent confidence intervals for θ . One leading hypothesis of interest here is the test of perfect overlap, i.e., equality of distributions, which corresponds to $H_0 : \theta = 1$ versus $H_A : \theta < 1$. Note that under this null hypothesis $\sigma_1^2 = 0$ and so the asymptotic variance comes purely from the boundary case. In this case, we reject whenever $\hat{\theta}^{bc} < 1 - z_\alpha \hat{p}_0^{1/2} \sigma_0 / \sqrt{n}$, where z_α is the normal critical value of level α .

The approach we have advocated has been to make the inference robust to the possibility that the densities are equal or close to each other over a range of values. However, in some cases one may be willing to assert that this is not the case. If one strongly believes that $C_{f,g}$ is of measure zero, then one can conduct inference using the uncorrected estimator $\hat{\theta}$ and the variance estimator $\tilde{v} = \tilde{p}_f(1 - \tilde{p}_f) + \tilde{p}_g(1 - \tilde{p}_g) + 2(\tilde{p}_{f,g} - \tilde{p}_f\tilde{p}_g)$, where

$$\begin{aligned} \tilde{p}_f &= \frac{1}{n} \sum_{i=1}^n 1(X_i \in \tilde{C}_f) \quad ; \quad \tilde{g}_f = \frac{1}{n} \sum_{i=1}^n 1(Y_i \in \tilde{C}_g) \\ \tilde{p}_{f,g} &= \frac{1}{n} \sum_{i=1}^n 1(X_i \in \tilde{C}_f, Y_i \in \tilde{C}_g) \\ \tilde{C}_f &= \{x : f_n(x) < g_n(x)\} \quad ; \quad \tilde{C}_g = \{x : f_n(x) \geq g_n(x)\}, \end{aligned}$$

which corresponds to the definitions above Theorem 2 setting the tuning parameter $c_n = 0$. In this case, $\tilde{v} \xrightarrow{p} v = \sigma_1^2$.³ The bootstrap is an alternative method for providing confidence intervals. In the special case where the contact set has zero measure, standard bootstrap resampling algorithms can be applied to conduct inference. However, as reported in Clemons and Bradley (2000) the standard bootstrap confidence intervals start performing badly when $\theta \rightarrow 1$, i.e., when the contact set has positive measure. In order to make the bootstrap work in the general case one must make modifications to the resampling algorithm that involve an additional tuning parameter just as in our asymptotic approach.

Note that the bandwidth parameter b and the tuning parameter c_n are asymptotically negligible, and only affect higher order properties, which are hard to analyze theoretically. We investigate the choice of these parameters in the simulation study below.

Our theory is for the estimation of θ from a single sample. In the application below we compute the overlap measure for two different regions at two different points in time. In some cases one may

³However, if it turned out that $C_{f,g}$ has positive measure then $\hat{\theta}$ is biased and the standard errors are inconsistent. Specifically, it can be shown that $\tilde{p}_f \xrightarrow{p} p_f + p_0/2$ and $\tilde{p}_g \xrightarrow{p} p_g + p_0/2$ using a similar argument to that used in Theorem 2.

have a panel data setting, so that we observe (X_{it}, Y_{it}) , $i = 1, \dots, n$ and $t = 1, \dots, T$, where (X_{it}, Y_{it}) and (X_{is}, Y_{is}) are not mutually independent. We obtain $\hat{\theta}_t$, $t = 1, \dots, T$ from each period. Then the $\hat{\theta}_t$ and $\hat{\theta}_s$ are asymptotically correlated. However, the correlation can easily be accounted for, especially since it only affects the non-boundary terms.

4 A Simulation Study

Here we look at the small-sample performance of $\hat{\theta}$ and $\hat{\theta}^{bc}$. Anderson and Ge (2004) have investigated the performance of an estimator of θ in the case where there are either one or two crossings. We consider the more interesting case where the contact set has positive measure. The design is $X_i \sim U[-0.5, 0.5]$, and $Y_i \sim U[0, 1]$, where $\{X_i\}$ and $\{Y_i\}$ are independent, so that $\theta = 0.5$ and $C_{f,g} = [0, 0.5]$.⁴ We consider samples sizes $n = 100, 200, 400, 800$, and 1600 and take a thousand replications of each experiment. The estimator is computed using the uniform kernel, i.e., $K(u) = 1(|u| \leq 0.5)$ for which $\|K\|_2^2 = 1$ and $\rho(t) = (1+t)1(-1 \leq t \leq 0) + (1-t)1(0 \leq t \leq 1)$, which satisfies A1 with $\alpha = 1$ and $s = 2$. In this case $p_f = p_g = 0$ and $p_0 = 0.5$. It follows that $a_n = -0.28b^{-1/2}$, while $v = p_0\sigma_0^2 + \sigma_1^2 = 0.3067$. The bandwidth takes two values, either the Silverman's rule of thumb value, in this case $b_s = 1.84\hat{\sigma}n^{-1/5}$, where $\hat{\sigma}$ is the sample standard deviation, or the smaller value $b_s^{3/2}$. In construction of the bias corrected estimator $\hat{\theta}$ we choose the tuning parameter c_n to be either the bandwidth b , the smaller value $b^{3/2}$ or the larger value $b^{2/3}$.

The supports of interest are estimated from the sample, specifically the common support set in this case is estimated by the interval $[\max\{\min_{1 \leq i \leq n} X_i, \min_{1 \leq i \leq n} Y_i\}, \min\{\max_{1 \leq i \leq n} X_i, \max_{1 \leq i \leq n} Y_i\}]$. The integrals are computed based on a grid of five hundred equally spaced points in $[-0.5, 1]$.

We report our results in Table 2. We give the bias, the median bias (mbias), the standard deviation (std), and the interquartile range divided by 1.349 (iqr) for the two estimators for the various combinations of samples sizes, bandwidths, and tuning parameters. The results can be summarized as follows:

1. The bias is quite large compared to the standard deviation
2. The performance measures improve with sample size at a rate roughly predicted by the theory (as can be confirmed by least squares regression of $\ln(-\text{bias})$ on a constant and $\ln n$)
3. The bias corrected estimator has a smaller bias and larger standard deviation and a smaller mean squared error at the best choice of bandwidth and tuning parameters.
4. The best performance for $\hat{\theta}$ is when bandwidth is b_s although there is not a lot of difference for the larger sample sizes

⁴Note that this doesn't precisely fit in the framework we have chosen to present our main results since the random variables have compact support. However, the theory we have presented also applies in the compactly supported case after a boundary adjustment. By using a uniform kernel on a uniform density the boundary effect is mitigated.

5. The best performance in terms of standard deviation for $\hat{\theta}^{bc}$ is when bandwidth is $b_s^{3/2}$, although for the smaller samples sizes bias is best at b_s . The best value of the tuning parameter for bias is the larger one $b^{2/3}$, whereas for variance $b^{3/2}$ is better.

Finally, we look at the quality of the distributional approximation. In Figure 1 we show the qq plot for standardized $\hat{\theta}$ in the case $n = 800$ and bandwidth is b_s . The approximation seems quite good, with most discrepancy in the left tail.

Figure 1 here

In Table 3 we report the results of a multivariate simulation with local to unity θ . Specifically, we took $X_k \sim N(0, 1)$ and $Y_k \sim N(1/\sqrt{n}, 1)$ with X, Y being vectors of dimensions one through five with independent components. In the univariate case, $\theta_n = 2\Phi(-1/2\sqrt{n})$, Reiser and Faraggi (1999). We implemented as above with the best combinations of bandwidth/tuning parameter uncovered in Table 2, with the same uniform kernel. We present the results for $\hat{\theta} - 1$, which should be exactly as in Theorem 1. The bias correction method seems to produce much better bias with a small cost in terms of increased variability. Note that the curse of dimensionality makes performance deteriorate with dimension, so for five dimensions the performance is terrible, but the bias correction does seem to improve performance substantially according to mean squared error. In the working paper version of this paper we reported results for the case where $Y_k \sim N(0, 1)$. The difference between these two cases is not substantial, which shows that the local to unity asymptotics are quite good in this case.

5 Application

Much ink has been spilled on how the economic reforms in China benefited cities on the eastern seaboard relative to those in the interior. Evidence on per capita urban incomes suggests greater advances for seaboard provinces than for inland provinces. Partly the result of regional comparative advantage, it also reflected weak government regional equalization policy, imperfect capital markets, and initial preferential policies on FDI and exports and from the growth of tax revenues as their development proceeded for the seaboard provinces (Anderson and Ge (2004), Gustafsson, Li, and Sicular (2007)). Urbanization also took place differentially on the seaboard and inland with cities growing more rapidly both in size and number in the seaboard provinces than in the interior (Anderson and Ge (2006, 2008)). The question arises as to whether the consequences of the reforms have translated into an improvement in the relative wellbeing of individuals in seaboard as compared to interior provinces. To investigate this, samples of urban households in two Chinese provinces, Guangdong - an eastern seaboard province and Shaanxi - a province in the interior (see the map of China below),

***Figure 2 here ***

taken in 1987 and 2001 are employed.⁵ We emphasize here that these data are not panel in structure, so it seems reasonable to suppose that the observations are independent across the time frame, across the cross section, and across the provinces.

One approach to the relative wellbeing issue is to examine whether or not household wellbeing in central and seaboard provinces has polarized. Esteban and Ray (1994) and Duclos, Esteban and Ray (2004) (see also Wang and Tsui (2000)) posited a collection of axioms whose consequences should be reflected in a Polarization measure. The axioms are founded upon a so-called Identification-Alienation nexus wherein notions of polarization are fostered jointly by an agent's sense of increasing within-group identity and between-group distance or alienation. When one distribution stochastically dominates the other it can be argued that such measures also reflect a sense of relative ill-being of the impoverished group and when there is a multiplicity of indicators, measures of "Distributional overlap" appear to perform quite well Anderson (2008).⁶

Indicators employed to reflect household wellbeing are total expenditures per household member (Exp p.c.) and household living area per household member (Area p.c.). Table 4 presents summary statistics for the samples, some observations are appropriate. Both provinces have advanced in terms of their consumption expenditures and living space per person so that overall wellbeing may be considered to have advanced in both provinces. The gap between expenditures, which reflects the alienation component of polarization and favors Guangdong, widened and the gap between living space (again favoring Guangdong) remained unchanged so that polarization may well have increased in terms of the alienation component. Movements in the dispersion of these components have less clear implications for the identification part of polarization. In Guangdong dispersion of living space per person diminished whereas in Shaanxi it increased, with respect to dispersion of expenditures they increased in both provinces but much more so in Shaanxi than in Guangdong to the extent that Shaanxi overtook Guangdong in its expenditure per person dispersion over the period. This suggests that little can be said about polarization by piecemeal analysis of its components.

*** Table 4 here ***

We first show the univariate density plots, which were calculated with Gaussian kernel and Silverman's rule of thumb bandwidth. These confirm the general trends identified in the sample statistics. Note that empirically there is only one crossing for the expenditure data but the housing variable has several crossing points.

*** Figure 3 here ***

⁵These data were obtained from the National Bureau of Statistics as part of the project on Income Inequality during China's Transition organized by Dwayne Benjamin, Loren Brandt, John Giles and Sangui Wang.

⁶Using a multivariate Kolmogorov-Smirnov criterion the hypothesis that the Guangdong joint distribution first order stochastically dominates the Shaanxi joint distribution could not be rejected in both years whereas the hypothesis that Shaanxi dominates Guangdong could (details from the authors on request)

We next compute the univariate and multivariate polarization measures. Let $\hat{\theta}_e$, $\hat{\theta}_h$, and $\hat{\theta}_{eh}$ denote respectively the measure computed on the univariate expenditure data series, the univariate housing series, and the bivariate data. We computed these quantities with a uniform kernel and bandwidth either equal to the Silverman's rule of thumb bandwidth b_s or $b_s^{3/2}$. We also computed the bias corrected estimators denoted with superscript bc using tuning parameter $b^{2/3}$. We compute both our standard errors and the standard errors that assume that the contact set is of zero measure, these are denoted by sc . The data are from repeated cross-sections so we assume the samples are independent. In this dataset there are different sample sizes n and m that apply to the estimation of f and g . For the bivariate product uniform kernel $\sigma_0^2(1) = 0.5835$. We computed the bias correction and standard errors using these modifications and computed standard normal test statistics for the polarization hypothesis (decreased overlap) based upon 1987 and 2001 being independent samples. The results are shown in Table 5. The results show a substantial reduction in the value of the overlap measure for the joint distribution and also the univariate measure for expenditure. There is a slight decrease also in the overlap of the housing variable, but this is not statistically significant. The level of the overlap is quite high in general and the bias correction increases it quite substantially. The estimators are relatively insensitive to the choice of bandwidth. The standard errors are quite small and there is not much difference between the full standard errors and the standard errors that impose zero measure on the contact set. Evidently there has been a significant polarization (reduction in overlap) between the provincial joint distributions of consumption expenditures and living space reflecting deterioration in the wellbeing of households in Shaanxi relative to those in Guangdong.

We perform the formal test of the hypotheses

$$H_0 : \theta_j = 1 \text{ vs. } H_A : \theta_j < 1, \quad j = 1987, 2001$$

$$H_0 : \theta_{1987} = \theta_{2001} \text{ vs. } H_A : \theta_{1987} \neq \theta_{2001}.$$

These hypotheses are rejected at the 1% level except for the housing case (full results are available from the authors upon request). The second hypothesis is rejected if

$$\left| \hat{\theta}_{1987}^{bc} - \hat{\theta}_{2001}^{bc} \right| \geq z_{\alpha/2} SE,$$

where $SE = \hat{v}_{1987}/\sqrt{n} + \hat{v}_{2001}/\sqrt{m}$. This rejection region is valid under the reasonable assumption of independence across time and household.

Note that an alternative to the overlap measure could be obtained by computing the Duclos Esteban and Ray (2004) polarization measure, specifically, $P_\alpha(f) = \int \int f(x)^{1+\alpha} f(y) |y - x| dy dx$, where $\alpha \in [0.25, 1]$, generalized to the multivariate case and based on the pooled distribution. This is a somewhat more general index of the multiplicity and diversity of modes and requires specifying a polarization sensitivity parameter α which should lay between 0.25 and 1. We computed this measure separately for the consumption and housing variables and for the joint distribution for the two years and record the results for the joint variables below. Note that, unlike the overlap statistic,

no significant change in the polarization index is detected in any of the cases reflecting the fact that the index is trying to identify polarization from changes in the mixture distribution. Intuitively this is because this approach is faced with the task of detecting changes in an unspecified number of bumps in a surface which is the result of pooling two distributions whereas the overlap is simply charged with detecting whether or not they are separating.

*** Table 6 here ***

Note the index is sensitive to the choice of their polarization sensitivity parameter α : at low levels of sensitivity the index actually diminishes over time whereas at high levels it increases.

6 Extensions

We have developed the asymptotic properties of the specific polarization measure; but note that our techniques are applicable for a more general class of problems. A more general quantity of interest is

$$\delta = \int \varphi(f(y), g(y)) dy, \quad (14)$$

where f, g are nonparametric objects like densities or regression functions and φ is a known function that has a kink at one or more points. Some examples include: $\varphi(s, t) = \min(s, t)$, $\varphi(s, t) = \max(s, t)$, $\varphi(s, t) = |s - t|^\alpha$, $\alpha > 0$, and $\varphi(s, t) = (\alpha - 1(s - t < 0))(s - t)$, $\alpha \in (0, 1)$. In these examples, the function is not differentiable at the point $s = t$. We discuss in detail one example. Kitagawa (2010) considers a partially identified version of the treatment effects model, Imbens and Rubin (1997). Specifically, Y is an outcome measure of interest that is not perfectly observable. Of interest is the density $f_{Y|X}(y|x)$, where X are observed covariates, but we only observe $(Y \cdot D, D, X, Z)$, where D is an indicator of missingness, and Z is a instrument. One can observe or estimate $f_{Y \cdot D|Z, X}(y, D = 1|Z = z, X = x)$. Under the exclusion assumption that Y is independent of Z given X one can show that the identified region includes only those densities $f_{Y|X}$ that are larger than the envelope $\sup_z f_{Y \cdot D|Z, X}(y, D = 1|Z = z, X = x)$. In the case where Z is binary, which is quite common in practice, the identification region is determined by the envelope $\max\{f(y|x), g(y|x)\}$, where $f(y|x) = f_{Y \cdot D|Z, X}(y, D = 1|Z = 0, X = x)$ and $g(y|x) = f_{Y \cdot D|Z, X}(y, D = 1|Z = 1, X = x)$. If the integrated density envelope $\delta(f, g)(x) = \int \max\{f(y|x), g(y|x)\} dy$ is larger than one for any x , the identification region is empty. Kitagawa (2010) develops a test of the hypothesis that $\delta = 1$ versus the general alternative. In this case, it can be for some values of x that $f(y|x) \simeq g(y|x)$ because the instrument is "weak", i.e., does not have a strong influence on the conditional outcome density. Our theory readily extends to cover this case.

Our theory extends to the case of k densities $f_1, \dots, f_k$ in an obvious way. In that case, one might also be interested in measuring quantities related to a partial overlap. Specifically, suppose that $f_{i_1}(x) \leq \dots \leq f_{i_k}(x)$, then $\min\{f_1(x), \dots, f_k(x)\} = f_{i_1}(x)$ and $\max\{f_1(x), \dots, f_k(x)\} = f_{i_k}(x)$.

Then, $f_{i_r}(x)$, for $1 < r < k$ represents a situation where r of the densities overlap.⁷ In this context the full overlap measure has been used as an index of inequality between k populations see Yalonetzky (2009). Kitagawa (2010) considered in general the case with k densities corresponding to k different values of the discrete instrument.

Finally, there are some applications where the data come from a time series and one would like to allow for dependence in the observations. For example, we might like to compare two or more forecast densities. In this case the theory becomes more complicated and it is not clear that the Poissonization method can be applied. However, in the special case that the contact set has zero measure, one can derive the limiting distribution for $\hat{\theta}$ based on the asymptotic representation $\sqrt{n}(\hat{\theta} - \theta) = n^{-1/2} \sum_{i=1}^n 1(X_i \in C_f) + n^{-1/2} \sum_{i=1}^n 1(Y_i \in C_g) + o_p(1)$, assuming some restriction on the strength of the dependence.

7 Conclusion

We have shown how to conduct robust inference for a polarization measure based upon the overlap of two potentially multivariate distributions. The theory accommodates the possibility that the contact set between two distributions is non empty and deals with estimation biases in the general case. An analytic approach to inference demonstrates that it is consistent whatever the nature of the contact set and most importantly when the distributions are very close. We have also provided for the case where two overlap estimates being compared are based upon non-independent samples (for example in the context of a panel data set). A simulation study examines the case where the contact set is non- empty and considers the uncorrected and bias corrected estimates of the overlap measure. It highlights the fact that the biases are quite large and that the bias correction comes at a cost of increased variability of the estimator. The asymptotic behaviour is much as the theory predicts and some evidence as to advisable tuning parameter choice is also presented.

The application, a multidimensional examination of urban household wellbeing in an internal (Shaanxi) and a coastal (Guangdong) province in China over the period of the economic reforms, demonstrates that the two groups of agents have bi-polarized significantly. Application of a multidimensional version of the most popular polarization index to the pooled sample of households did not yield the same result in that no significant polarization was detected. This is not surprising since the latter is a very general index which seeks changes in an unspecified number of points of modality and will have less power than the overlap statistic which provides a very specific test of bipolarization.

Finally (using the results herein presented) the overlap measure has already been used to examine symmetry properties of a function $f(x)$ centered on 0 so that $f(x) = f(-x)$ under the null (see Samawi et. al. (2010)). The theory we have developed is useful in a number of other theoretical and

⁷This is of interest in a number of biomedical applications. See for example <http://collateral.knowitall.com/collateral/95391-OverlapDensityHeatMap.pdf>

applied problems. The overlap measure can be employed to compare the coincidence or otherwise of $f(x)$ and $g(x)$ where these are more general non-parametric objects and the comparison function exhibits some sort of kink. Hence tests of separability of the f function and statistical independence of the elements of x are also facilitated. The extension to many functions $f_i(x)$, $i = 1, \dots, k$, is also straightforward (in this context the overlap measure has been used as an index of inequality between k populations see Yalonetzky (2009)).

A Appendix

A.1 Informal Discussion of the Proof Technique

Although the estimators and confidence intervals are easy to use in practice, the asymptotic theory to prove Theorem 1 involves several lengthy steps. Since establishing these steps require techniques that are not commonly used in econometrics, we now give a brief informal description of our proof techniques. Specifically, our proof of Theorem 1 consists of the following three steps:

1. The asymptotic approximation of $\sqrt{n}(\hat{\theta} - \theta)$ by A_n , given by the non-negligible terms in (11), which decomposes the estimation error into three different terms, defined over the disjoint sets C_f , C_g and $C_{f,g}$, respectively.
2. Get the asymptotic distribution of $A_n^P(\bar{B})$, a *Poissonized* version A_n , where the sample size n is replaced by a Poisson random variable N with mean n that is independent of the sequence $\{(X_i, Y_i) : i \geq 1\}$ and the integral is taken over a subset $\bar{B}$ of the union of the supports of X and Y .
3. De-*Poissonize* $A_n^P(\bar{B})$ to derive the asymptotic distribution of A_n and hence $\sqrt{n}(\hat{\theta} - \theta)$.

In step 1, we make the bias of kernel densities asymptotically negligible by using the smoothness assumptions on true densities and properties of kernel functions, which allows us to write A_n as a functional of the centered statistics $f_n(x) - Ef_n(x)$ and $g_n(x) - Eg_n(x)$. Also, the decomposition into three terms is related to the recent result in the moment inequality literature that, under inequality restrictions, the asymptotic behavior of statistics of interest often depend only on binding restrictions, see, e.g. Chernozhukov, Hong, and Tamer (2007), Andrews and Guggenberger (2009) and Linton, Maasoumi and Whang (2005).

In step 2, Poissonization of the statistic A_n gives a lot of convenience in our asymptotic analysis. In particular, it is well known that if N is a Poisson random variable independent of the i.i.d sequence $\{X_i : i \geq 1\}$ and $\{A_k : k \geq 1\}$ are disjoint measurable sets, then the processes $\sum_{i=0}^N 1(X_i \in A_k) \delta_{X_i}$, $k = 1, 2, \dots$, are independent, where δ denotes the Dirac delta. This implies, for example, that, since the kernel function K is assumed to have a compact support, the Poissonized kernel densities $f_N(x)$

and $f_N(y)$ are independent if the distance between x and y is greater than a certain threshold. This facilitates computation of asymptotic variance of $A_n^P(\bar{B})$. Also, since a Poisson process is infinitely divisible, we can write $\sum_{i=0}^N X_i \stackrel{d}{=} \sum_{i=0}^n Z_i$, where $\{Z_i : i \geq 1\}$ are i.i.d with $Z \stackrel{d}{=} \sum_{i=0}^{\eta_1} X_i$ and η_1 is a Poisson random variable with mean 1 and independent of $\{X_i : i \geq 1\}$. The fact is used repeatedly in our proofs to derive the asymptotic distribution of $A_n^P(\bar{B})$, using standard machineries including the CLT and Berry Esseen theorem for i.i.d. random variables.

In step 3, we need to de-Poissonize the result because the asymptotic behavior of the Poissonized variable $A_n^P(\bar{B})$ is generally different from A_n . For this purpose, we use the de-Poissonization lemma of Beirlant and Mason (1995, Theorem 2.1, see also Lemma A.2 below). To illustrate the Lemma in a simple context, consider a statistic $\Lambda_n = n^{-1/2} \sum_{i=1}^n [1(X_i \in B) - \Pr(X \in B)]$, where $B \subset \mathbb{R}$ is a Borel set. By a CLT, we know that $\Lambda_n \Rightarrow N(0, p_B(1 - p_B))$, where $p_B = \Pr(X \in B)$. Now, consider a Poissonized statistic $S_n = n^{-1/2} \sum_{i=1}^N 1(X_i \in B) - n^{1/2} \Pr(X \in B)$. The asymptotic distribution of S_n is given by $N(0, p_B)$, which is different from that of Λ_n . However, letting $U_n = n^{-1/2} \sum_{i=1}^N 1(X_i \in C) - n^{1/2} \Pr(X \in C)$ and $V_n = n^{-1/2} \sum_{i=1}^N 1(X_i \in \mathbb{R} \setminus C) - n^{1/2} \Pr(X \in \mathbb{R} \setminus C)$, where $B \subset C \subset \mathbb{R}$ is a Borel set, and applying the Poissonization lemma, we see that the conditional distribution of S_n given $N = n$ coincides with the distribution of Λ_n asymptotically.

Although the above steps closely follow those of Giné et. al. (2003), we need to extend their results to the general multi-dimensional variates $d \geq 1$, multiple kernel densities, and norms different from the L_1 - norm. Such extensions, to our best knowledge, are not available in the literature and are not trivial.

A.2 Proof of the Main Theorems

Under our conditions, we have

$$\sup_{x \in \mathbb{R}^d} |f_n(x) - f(x)| = O(b^s) + O\left(\sqrt{\frac{\log n}{nb^d}}\right) \quad a.s., \quad (15)$$

by Giné and Guillou (2002, Theorem 1) and standard treatment of the bias term, and likewise for $g_n(x) - g(x)$. We use this result below.

Let

$$\begin{aligned} A_n &= \sqrt{n} \int_{C_{f,g}} \min \{f_n(x) - Ef_n(x), g_n(x) - Eg_n(x)\} dx \\ &\quad + \sqrt{n} \int_{C_f} [f_n(x) - Ef_n(x)] dx + \sqrt{n} \int_{C_g} [g_n(x) - Eg_n(x)] dx \\ &=: A_{1n} + A_{2n} + A_{3n} \end{aligned} \quad (16)$$

We will show that the asymptotic distribution of A_n is normal when suitably standardized.

Theorem A1. *Under Assumptions (A1)-(A4), we have*

$$\frac{A_n - a_n}{\sqrt{p_0 \sigma_0^2 + \sigma_1^2}} \Rightarrow N(0, 1).$$

The proof of Theorem A1 will be given later. Given Theorem A1, we can establish Theorem 1.

Proof of Theorem 1. We will show below that

$$\sqrt{n}(\widehat{\theta} - \theta) = A_n + o_p(1). \quad (17)$$

Then, this result and Theorem A1 yield the desired result of Theorem 1. To show (17), write

$$\begin{aligned} \sqrt{n}(\widehat{\theta} - \theta) &= \int_{\mathbb{R}^d} \sqrt{n} [\min\{f_n(x), g_n(x)\} - \min\{f(x), g(x)\}] dx \\ &= \int_{C_{f,g}} \sqrt{n} \min\{f_n(x) - f(x), g_n(x) - g(x)\} dx \\ &\quad + \int_{C_f} \sqrt{n} \min\{f_n(x) - f(x), g_n(x) - f(x)\} dx \\ &\quad + \int_{C_g} \sqrt{n} \min\{f_n(x) - g(x), g_n(x) - g(x)\} dx \\ &=: \Lambda_{1n} + \Lambda_{2n} + \Lambda_{3n}. \end{aligned} \quad (18)$$

Consider Λ_{1n} first. Write

$$\begin{aligned} \Lambda_{1n} &= \sqrt{n} \int_{C_{f,g}} \min\{f_n(x) - Ef_n(x), g_n(x) - Eg_n(x)\} dx \\ &\quad + \int_{C_{f,g}} \sqrt{n} [\min\{f_n(x) - f(x), g_n(x) - g(x)\} - \min\{f_n(x) - Ef_n(x), g_n(x) - Eg_n(x)\}] dx \\ &=: A_{1n} + \Lambda_{12n}. \end{aligned} \quad (19)$$

We have

$$\begin{aligned} |\Lambda_{12n}| &\leq 2\sqrt{n} \left(\int_{C_{f,g}} \{|Ef_n(x) - f(x)| + |Eg_n(x) - g(x)|\} dx \right) \\ &\leq 2\sqrt{n}b^s \left(\int_{C_{f,g}} \int_{\mathbb{R}^d} \sum_{|\lambda|=s} \frac{1}{s!} \left| D^\lambda f(x - \tilde{b}u) \right| |u^\lambda K(u)| du dx \right) \\ &\quad + 2\sqrt{n}b^s \left(\int_{C_{f,g}} \int_{\mathbb{R}^d} \sum_{|\lambda|=s} \frac{1}{s!} \left| D^\lambda g(x - \tilde{b}u) \right| |u^\lambda K(u)| du dx \right) \\ &= O(n^{1/2}b^s) \rightarrow 0, \end{aligned} \quad (20)$$

where the first inequality uses the elementary result $|\min\{a+c, b+d\} - \min\{a, b\}| \leq 2(|c| + |d|)$, the second inequality holds by a two term Taylor expansion with $0 < \tilde{b} < b$ and Assumption A1, the last equality holds by Assumptions A1 and A2, and the convergence to zero follows from Assumption A3.

We next consider Λ_{2n} . By (15), we have: given $\eta > 0$, there exists $\delta > 0$ such that

$$\Pr \left(\left(\frac{nb^d}{\log n} \right)^{1/2} \sup_{x \in \mathbb{R}^d} |f_n(x) - Ef_n(x)| > \delta/2 \right) < \eta. \quad (21)$$

Also, for each $x \in C_f$, there exists n_0 such that for $n \geq n_0$, $(nb^d/\log n)^{1/2} [g(x) - f(x)] > \delta$. Let $\varepsilon > 0$ be a constant. Then, for $n \geq n_0$,

$$\begin{aligned} & \Pr \left(\int_{C_f} \sqrt{n} \max\{[f_n(x) - Ef_n(x)] - [g_n(x) - Eg_n(x)] - [g(x) - f(x)], 0\} dx > \varepsilon \right) \\ & \leq \Pr \left(\left(\frac{nb^d}{\log n} \right)^{1/2} \left\{ \sup_{x \in \mathbb{R}^d} |f_n(x) - Ef_n(x)| + \sup_{x \in \mathbb{R}^d} |g_n(x) - Eg_n(x)| \right\} > \delta \right) < \eta. \end{aligned} \quad (22)$$

Therefore,

$$\begin{aligned} |\Lambda_{2n} - A_{2n}| & \leq \int_{C_f} \sqrt{n} \max\{[f_n(x) - Ef_n(x)] - [g_n(x) - Eg_n(x)] - [g(x) - f(x)], 0\} dx + O(n^{1/2}b^s) \\ & = o_p(1) + O(n^{1/2}b^s) = o_p(1), \end{aligned}$$

where the first inequality follows from an argument similar to the one to establish (20) and triangle inequality and the first equality holds by (22) since η is arbitrary. Likewise,

$$\Lambda_{3n} = A_{3n} + o_p(1).$$

Now, this establishes (17), as desired. ■

We prove Theorem A1 using the Poissonization argument of Giné et. al. (2003). To do this, we need to extend some of the results of Giné et. al. (2003) to the general multi-dimensional case $d \geq 1$ with multiple kernel densities. Also, we need to consider a functional $\varphi_0 : (f, g) \mapsto \int \min\{f(x), g(x)\} dx$, which is different from the L_1 -functional $\varphi_1 : f \mapsto \int |f(x)| dx$. We first introduce some concepts used throughout the proofs. Let N be a Poisson random variable with mean n , defined on the same probability space as the sequence $\{(X_i, Y_i) : i \geq 1\}$, and independent of this sequence. Define

$$f_N(x) = \frac{1}{n} \sum_{i=1}^N K_b(x - X_i), \quad g_N(x) = \frac{1}{n} \sum_{i=1}^N K_b(x - Y_i)$$

where $b = b(n)$ and where the empty sum is defined to be zero. Notice that

$$Ef_N(x) = Ef_n(x) = EK_b(x - X) \quad (23)$$

$$k_{f,n}(x) = n \text{var}(f_N(x)) = EK_b^2(x - X) \quad (24)$$

$$n \text{var}(f_n(x)) = EK_b^2(x - X) - \{EK_b(x - X)\}^2. \quad (25)$$

Similar results hold for $g_N(x)$ and $g_n(x)$.

Let $C_{f,g}$, C_f and C_g denote the sets defined in (5)-(7) with the corresponding measures p_0 , p_f and p_g defined in (8)-(10). For a constant $M > 0$, let $\mathcal{B}(M) \subset \mathbb{R}^d$ denote a Borel set with nonempty interior with Lebesgue measure $\mu(\mathcal{B}(M)) = M^d$. For $v > 0$, define $\mathcal{B}(M, v)$ to be the v -contraction

of $\mathcal{B}(M)$, i.e., $\mathcal{B}(M, v) = \{x \in \mathcal{B}(M) : \rho(x, \mathbb{R}^d \setminus \mathcal{B}(M)) \geq v\}$, where $\rho(x, B) = \inf\{\|x - y\| : y \in B\}$. Let ϕ denote f or g . Take $\varepsilon \in (0, p_0)$ and $\varepsilon_\phi \in (0, p_\phi)$ to be arbitrary constants. (If any of p_0 , p_f and p_g is zero, then the corresponding sets defined subsequently have to be empty.) Choose $M, M_\phi, v, v_\phi > 0$ and Borel sets B_0, B_ϕ such that

$$\mathcal{B}(M) \subset C_{f,g}, \quad \mathcal{B}(M_\phi) \subset C_\phi, \quad (26)$$

$$B_0 \subset \mathcal{B}(M, v), \quad B_\phi \subset \mathcal{B}(M_\phi, v_\phi) \quad (27)$$

$$\int \int_{\mathbb{R}^{2d} \setminus \mathcal{T}(M)} f(x)g(y)dx dy = : \alpha > 0 \quad (28)$$

$$\int_{B_0} f(x)dx = \int_{B_0} g(x)dx > p_0 - \varepsilon, \quad \int_{B_\phi} \phi(x)dx > p_\phi - \varepsilon_\phi, \quad (29)$$

and f and ϕ are bounded away from 0 on B_0 and B_ϕ , respectively, where

$$\mathcal{T}(M) = (\mathcal{B}(M_f) \times \mathbb{R}^d) \cup (\mathcal{B}(M) \times \mathcal{B}(M)) \cup (\mathbb{R}^d \times \mathcal{B}(M_g)) \subset \mathbb{R}^{2d}. \quad (30)$$

Such $M, M_\phi, v, v_\phi, B_\phi$, and B_0 exist by continuity of f and g .

Let $\bar{B} = B_0 \cup B_f \cup B_g$. In $\bar{B}$, the sets are separated with no common boundary and the suitable measures of the corresponding sets are strictly positive. The construction of $\bar{B}$ is necessary to apply the de-Poissonization argument (Lemma A2). Define a Poissonization version $A_n^P(\bar{B})$ of A_n based on the expression (16):

$$A_n^P(\bar{B}) = A_{1n}^P(B_0) + A_{2n}^P(B_f) + A_{3n}^P(B_g), \quad (31)$$

where

$$\begin{aligned} A_{1n}^P(B_0) &= \sqrt{n} \int_{B_0} \min \{f_N(x) - Ef_n(x), g_N(x) - Eg_n(x)\} dx \\ &\quad - \sqrt{n} \int_{B_0} E \min \{f_N(x) - Ef_n(x), g_N(x) - Eg_n(x)\} dx \end{aligned} \quad (32)$$

$$A_{2n}^P(B_f) = \sqrt{n} \int_{B_f} [f_N(x) - Ef_n(x)] dx \quad (33)$$

$$A_{3n}^P(B_g) = \sqrt{n} \int_{B_g} [g_N(x) - Eg_n(x)] dx. \quad (34)$$

Also, define the variance of the Poissonization version $A_n^P(\bar{B})$ to be

$$\sigma_n^2(\bar{B}) = \text{var} (A_n^P(\bar{B})). \quad (35)$$

To investigate the asymptotic distribution of $A_n^P(\bar{B})$, we will need the following lemma, which is related to the classical Berry-Esseen theorem.

Lemma A1. (a) Let $\{W_i = (W_{1i}, \dots, W_{4i})^\top : i \geq 1\}$ be a sequence of i.i.d. random vectors in $\mathbb{R}^4$ such that each component has mean 0, variance 1, finite absolute moments of third order, and

$$EWW^\top = \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{pmatrix}.$$

Let $Z = (Z_1, \dots, Z_4)^\top$ be multivariate normal with mean vector 0 and positive definite variance-covariance matrix $EZZ^\top = \Sigma$. Then, there exist universal positive constants A_1 and A_2 such that

$$\left| E \min \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n W_{1i}, \frac{1}{\sqrt{n}} \sum_{i=1}^n W_{2i} \right\} - E \min \{Z_1, Z_2\} \right| \leq \frac{A_1}{|\det(\Sigma_{11})|^{-3/2}} \frac{1}{\sqrt{n}} (E|W_1|^3 + E|W_2|^3)$$

and

$$\begin{aligned} & \left| E \left[\frac{1}{n} \min \left\{ \sum_{i=1}^n W_{1i}, \sum_{i=1}^n W_{2i} \right\} \min \left\{ \sum_{i=1}^n W_{3i}, \sum_{i=1}^n W_{4i} \right\} \right] - E [\min \{Z_1, Z_2\} \min \{Z_3, Z_4\}] \right| \\ & \leq \frac{A_2}{|\det(\Sigma)|^{-3/2}} \frac{1}{\sqrt{n}} (E|W_1|^3 + E|W_2|^3 + E|W_3|^3 + E|W_4|^3). \end{aligned}$$

(b) Let $\{W_i = (W_{1i}, \dots, W_{3i})^\top : i \geq 1\}$ be a sequence of i.i.d. random vectors in $\mathbb{R}^3$ such that each component has mean 0, variance 1, finite absolute moments of third order, and $EW_i W_i^\top = \Sigma$. Let $Z = (Z_1, Z_2, Z_3)^\top$ be multivariate normal with mean vector 0 and positive definite variance-covariance matrix $EZZ^\top = \Sigma$. Then,

$$\begin{aligned} & \left| E \left[\min \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n W_{1i}, \frac{1}{\sqrt{n}} \sum_{i=1}^n W_{2i} \right\} \frac{1}{\sqrt{n}} \sum_{i=1}^n W_{3i} \right] - E [\min \{Z_1, Z_2\} Z_3] \right| \\ & \leq \frac{A_3}{|\det(\Sigma)|^{-3/2}} \frac{1}{\sqrt{n}} (E|W_1|^3 + E|W_2|^3 + E|W_3|^3). \end{aligned}$$

Proof of Lemma A1. The results of Lemma A1 follow directly from Sweeting (1980, Theorem1). ■

We also need the following basic result of Beirlant and Mason (1995, Theorem 2.1), which is needed to "de-Poissonize" our asymptotic results on the Poissonized random variables.

Lemma A2. Let $N_{1,n}$ and $N_{2,n}$ be independent Poisson random variables with $N_{1,n}$ being Poisson($n(1 - \alpha)$) and $N_{2,n}$ being Poisson($n\alpha$), where $\alpha \in (0, 1/2)$. Denote $N_n = N_{1,n} + N_{2,n}$ and set

$$U_n = \frac{N_{1,n} - n(1 - \alpha)}{\sqrt{n}} \text{ and } V_n = \frac{N_{2,n} - n\alpha}{\sqrt{n}}.$$

Let $\{S_{N_n} : n \geq 1\}$ be a sequence of random variables such that (i) for each $n \geq 1$, the random vector (S_{N_n}, U_n) is independent of V_n , (ii) for some $\sigma^2 > 0$ and ξ such that $(1 - \alpha)\sigma^2 - \xi^2 > 0$,

$$(S_{N_n}, U_n)^\top \Rightarrow N(0, \Sigma),$$

where

$$\Sigma = \begin{pmatrix} \sigma^2 & \xi \\ \xi & 1 - \alpha \end{pmatrix}.$$

Then, for all x , we have

$$\Pr(S_{N_n} \leq x \mid N_n = n) \rightarrow \Pr\left(\sqrt{\sigma^2 - \xi^2} Z_1 \leq x\right),$$

where Z_1 denotes the standard normal random variable.

The proofs of Lemmas A3-A8 are in Appendix A.3. The following lemma derives the asymptotic variance of $A_n^P(\bar{B})$.

Lemma A3. *Whenever Assumptions (A1)-(A4) hold and B_0 , B_f , and B_g satisfy (26)-(29), we have*

$$\lim_{n \rightarrow \infty} \sigma_n^2(\bar{B}) = p_{0,B} \sigma_0^2 + \bar{\sigma}_{1,B}^2, \quad (36)$$

where $p_{0,B} = \Pr(X \in B_0) = \Pr(Y \in B_0)$, $\bar{\sigma}_{1,B}^2 = p_{f,B} + p_{g,B} + 2p_{f,g,B}$, $p_{f,B} = \Pr(X \in B_f)$, $p_{g,B} = \Pr(Y \in B_g)$, $p_{f,g,B} = \Pr(X \in B_f, Y \in B_g)$ and σ_0^2 is defined above Theorem 1.

Define

$$\begin{aligned} U_n &= \frac{1}{\sqrt{n}} \left\{ \sum_{j=1}^N 1((X_j, Y_j) \in \mathcal{T}(M)) - n \Pr((X, Y) \in \mathcal{T}(M)) \right\} \\ V_n &= \frac{1}{\sqrt{n}} \left\{ \sum_{j=1}^N 1((X_j, Y_j) \in (\mathbb{R}^{2d} \setminus \mathcal{T}(M))) - n \Pr((X, Y) \in (\mathbb{R}^{2d} \setminus \mathcal{T}(M))) \right\}, \end{aligned}$$

where $\mathcal{T}(M)$ is defined in (30). We next establish the following convergence in distribution result.

Lemma A4. *Under Assumptions (A1)-(A4), we have*

$$(A_n^P(\bar{B}), U_n)^\top \Rightarrow N(0, \Sigma),$$

where

$$\Sigma = \begin{pmatrix} p_{0,B} \sigma_0^2 + \bar{\sigma}_{1,B}^2 & p_{f,B} + p_{g,B} \\ p_{f,B} + p_{g,B} & 1 - \alpha \end{pmatrix}$$

and α is defined in (28).

The following theorem gives the asymptotic bias formula.

Lemma A5. *Under Assumptions (A1)-(A4), we have*

$$\begin{aligned} \text{(a)} \quad \lim_{n \rightarrow \infty} \int_{B_0} \left[\sqrt{n} E \min \{f_n(x) - E f_n(x), g_n(x) - E g_n(x)\} dx - E \min \{Z_1, Z_2\} k_{f,n}^{1/2}(x) \right] dx &= 0 \\ \text{(b)} \quad \lim_{n \rightarrow \infty} \int_{B_0} \left[\sqrt{n} E \min \{f_n(x) - E f_n(x), g_n(x) - E g_n(x)\} dx - E \min \{Z_1, Z_2\} k_{f,n}^{1/2}(x) \right] dx &= 0, \end{aligned}$$

for $k_{f,n}$ defined in (24), where Z_1 and Z_2 are standard normal random variables.

Define

$$\begin{aligned} A_n(\bar{B}) &= \sqrt{n} \int_{B_0} [\min \{f_n(x) - E f_n(x), g_n(x) - E g_n(x)\} \\ &\quad - E \min \{f_n(x) - E f_n(x), g_n(x) - E g_n(x)\}] dx \\ &\quad + \sqrt{n} \int_{B_f} [f_n(x) - E f_n(x)] dx + \sqrt{n} \int_{B_g} [g_n(x) - E g_n(x)] dx. \end{aligned} \quad (37)$$

Using the de-Poissonization lemma (Lemma A2), we can show that the asymptotic distribution of $A_n(\overline{B})$ is normal.

Lemma A6. *Under Assumptions (A1)-(A4), we have*

$$A_n(\overline{B}) \Rightarrow \sqrt{p_{0,B}\sigma_0^2 + \sigma_{1,B}^2} Z_1,$$

where $\sigma_{1,B}^2 = p_{f,B}(1 - p_{f,B}) + p_{g,B}(1 - p_{g,B}) + 2(p_{f,g,B} - p_{f,B}p_{g,B})$ and Z_1 stands for the standard normal random variable.

The following two lemmas are useful to investigate the behavior of the difference between the statistics $A_n(\overline{B})$ and A_n .

Lemma A7. *Let $\{X_i = (X_{1i}^\top, X_{2i}^\top)^\top \in \mathbb{R}^d \times \mathbb{R}^d : i = 1, \dots, n\}$ be i.i.d random vectors with $E\|X\| < \infty$. Let $h_j : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ be a real function such that $Eh_j(X_j, x) = 0$ for all $x \in \mathbb{R}^d$ for $j = 1, 2$. Let*

$$T_n = \int_{\mathcal{B}} \min \left\{ \sum_{k=1}^n h_1(X_{1k}, x), \sum_{k=1}^n h_2(X_{2k}, x) \right\} dx,$$

where $\mathcal{B} \subset \mathbb{R}^d$ is a Borel set. Then, for any convex function $g : \mathbb{R} \rightarrow \mathbb{R}$, we have

$$Eg(T_n - ET_n) \leq Eg \left(4 \sum_{j=1}^2 \sum_{k=1}^n \varepsilon_k \int_{\mathcal{B}} |h_j(X_{jk}, x)| dx \right),$$

where $\{\varepsilon_i : i = 1, \dots, n\}$ are i.i.d random variables with $\Pr(\varepsilon = 1) = \Pr(\varepsilon = -1) = 1/2$, independent of $\{X_i : i = 1, \dots, n\}$.

Lemma A8. *Suppose that Assumptions (A1)-(A4) hold. Then, for any Borel subset $\mathcal{B}$ of $\mathbb{R}^d$, we have*

$$\overline{\lim}_{n \rightarrow \infty} E \left(\sqrt{n} \int_{\mathcal{B}} \{h_n(x) - Eh_n(x)\} dx \right)^2 \leq D \left(\sup_u |K(u)| \right)^2 \int_{\mathcal{B}} [f(x) + g(x)] dx,$$

for some generic constant $D > 0$, where

$$h_n(x) = \min\{f_n(x) - Ef_n(x), g_n(x) - Eg_n(x)\}.$$

Now, we are now ready to prove Theorem A1.

Proof of Theorem A1. By Lemma 6.1 of Giné et. al.(2003), there exists increasing sequences of Borel sets $\{B_{0k} \subset C_{f,g} : k \geq 1\}$, $\{B_{fk} \subset C_f : k \geq 1\}$, and $\{B_{gk} \subset C_g : k \geq 1\}$, each with finite Lebesgue measure, such that

$$\lim_{k \rightarrow \infty} \int_{C_{f,g} \setminus B_{0k}} f(x) dx = \lim_{k \rightarrow \infty} \int_{C_{f,g} \setminus B_{0k}} g(x) dx = 0 \quad (38)$$

$$\lim_{k \rightarrow \infty} \int_{C_f \setminus B_{fk}} f(x) dx = 0; \quad \lim_{k \rightarrow \infty} \int_{C_g \setminus B_{gk}} g(x) dx = 0. \quad (39)$$

Let $\bar{B}_k = B_{0k} \cup B_{fk} \cup B_{gk}$ for $k \geq 1$. Notice that for each $k \geq 1$, by Lemma A6, we have

$$A_n(\bar{B}_k) \Rightarrow \sqrt{p_{0,B_k}\sigma_0^2 + \sigma_{1,B_k}^2} Z_1 \text{ as } n \rightarrow \infty. \quad (40)$$

By (38) and (39),

$$\sqrt{p_{0,B_k}\sigma_0^2 + \sigma_{1,B_k}^2} Z_1 \Rightarrow \sqrt{p_0\sigma_0^2 + \sigma_1^2} Z_1 \text{ as } k \rightarrow \infty. \quad (41)$$

Also, by Lemma A8, we have

$$\overline{\lim}_{n \rightarrow \infty} E \left(\sqrt{n} \int_{C_{f,g} \setminus B_{0k}} \{h_n(x) - E h_n(x)\} dx \right)^2 \leq D \left(\sup_u |K(u)| \right)^2 \int_{C_{f,g} \setminus B_{0k}} [f(x) + g(x)] dx, \quad (42)$$

where

$$h_n(x) = \min\{f_n(x) - E f_n(x), g_n(x) - E g_n(x)\}.$$

Similarly, we have

$$\overline{\lim}_{n \rightarrow \infty} E \left(\sqrt{n} \int_{C_f \setminus B_{fk}} \{f_n(x) - E f_n(x)\} dx \right)^2 \leq D \left(\sup_u |K(u)| \right)^2 \int_{C_f \setminus B_{fk}} f(x) dx \quad (43)$$

$$\overline{\lim}_{n \rightarrow \infty} E \left(\sqrt{n} \int_{C_g \setminus B_{gk}} \{g_n(x) - E g_n(x)\} dx \right)^2 \leq D \left(\sup_u |K(u)| \right)^2 \int_{C_g \setminus B_{gk}} g(x) dx. \quad (44)$$

The right hand sides of (42), (43), and (44) converge to zero as $k \rightarrow \infty$ by (38), (39). Therefore, we have

$$\lim_{k \rightarrow \infty} \overline{\lim}_{n \rightarrow \infty} \Pr(|A_n(\bar{B}_k) - (A_n - E A_n)| > \varepsilon) = 0 \quad \forall \varepsilon > 0. \quad (45)$$

Now, by (40), (41) and (45) and Theorem 4.2 of Billingsley (1968), we have, as $n \rightarrow \infty$,

$$\begin{aligned} & A_n - E A_n \\ &= \sqrt{n} \int_{C_{f,g}} [\min\{f_n(x) - E f_n(x), g_n(x) - E g_n(x)\} \\ &\quad - E \min\{f_n(x) - E f_n(x), g_n(x) - E g_n(x)\}] dx \\ &\quad + \sqrt{n} \int_{C_f} [f_n(x) - E f_n(x)] dx + \sqrt{n} \int_{C_g} [g_n(x) - E g_n(x)] dx \\ &\Rightarrow \sqrt{p_0\sigma_0^2 + \sigma_1^2} Z_1. \end{aligned} \quad (46)$$

Now, the proof of Theorem A is complete since, similarly to Lemma A5, we have

$$\lim_{n \rightarrow \infty} |E A_n - a_n| = 0.$$

■

Proof of Theorem 2. To establish (12) we must show consistency of the bias correction. By the triangle inequality, we have

$$\begin{aligned} \left| \int_{\hat{C}_{f,g}} f_n^{1/2}(x) dx - \int_{C_{f,g}} f^{1/2}(x) dx \right| &\leq \int_{\hat{C}_{f,g} \Delta C_{f,g}} f^{1/2}(x) dx + \int_{\hat{C}_{f,g}} |f_n^{1/2}(x) - f^{1/2}(x)| dx \\ &=: D_{1n} + D_{2n}, \end{aligned}$$

where Δ denotes the symmetric difference. For notational simplicity, let

$$h_n(x) = f_n(x) - g_n(x) \text{ and } h(x) = f(x) - g(x).$$

Define

$$\begin{aligned} \tilde{C}_{f,g} &= \{x : |h(x)| \leq 2c_n\} \\ E_n &= \{x : |h_n(x) - h(x)| \geq c_n\}. \end{aligned}$$

We first establish $D_{1n} = o_p(b^{d/2})$ using an argument similar to Cuevas and Fraiman (1997, Theorem 1). Using $C_{f,g} \subset \tilde{C}_{f,g}$, $\mu_f(\hat{C}_{f,g} \cap \tilde{C}_{f,g}^c \cap E_n^c) = 0$ and $\mu_f(\hat{C}_{f,g}^c \cap C_{f,g} \cap E_n^c) = 0$, we have

$$\begin{aligned} D_{1n} &= \mu_f(\hat{C}_{f,g} \Delta C_{f,g}) = \mu_f(\hat{C}_{f,g} \cap C_{f,g}^c) + \mu_f(\hat{C}_{f,g}^c \cap C_{f,g}) \\ &\leq \mu_f(\hat{C}_{f,g} \cap \tilde{C}_{f,g}^c) + \mu_f(\tilde{C}_{f,g} \cap C_{f,g}^c) + \mu_f(\hat{C}_{f,g}^c \cap C_{f,g}) \\ &= \mu_f(\hat{C}_{f,g} \cap \tilde{C}_{f,g}^c \cap E_n) + \mu_f(\tilde{C}_{f,g} \cap C_{f,g}^c) + \mu_f(\hat{C}_{f,g}^c \cap C_{f,g} \cap E_n) \\ &\leq 2\mu_f(E_n) + \varsigma_n, \end{aligned} \tag{47}$$

where superscript c stands for complement and $\varsigma_n = 2h^*(2c_n)$ by Assumption A5. Also, by the rates of convergence result of the L_1 -errors of kernel densities (see, e.g., Holmström and Klemelä (1992)), we have that

$$\int |h_n(x) - h(x)| f^{1/2}(x) dx \leq O_p(b^s + (nb^d)^{-1/2}). \tag{48}$$

Let $\rho_n = \min\{b^{-s}, (nb^d)^{1/2}\}$. Then, for any $\varepsilon > 0$,

$$\begin{aligned} \Pr(b^{-d/2} D_{1n} > \varepsilon) &\leq \Pr(2\mu_f(E_n) + \varsigma_n > \varepsilon b^{d/2}) \\ &\leq \Pr\left(\frac{1}{c_n} \int |h_n(x) - h(x)| f^{1/2}(x) dx > \frac{\varepsilon b^{d/2} - \varsigma_n}{2}\right) \\ &\leq \Pr\left(\rho_n \int |h_n(x) - h(x)| f^{1/2}(x) dx > \frac{\varepsilon \rho_n c_n b^{d/2}}{2}\right) + o(1) \\ &\rightarrow 0, \end{aligned} \tag{49}$$

where the first inequality holds by (47), the second inequality holds by the inequality $1(E_n) \leq |h_n(x) - h(x)|/c_n$, the third inequality follows from Assumptions A5 and A7 which implies $\rho_n c_n \varsigma_n \rightarrow 0$, and the last convergence to zero holds by (48) and $\rho_n c_n b^{d/2}/(\log n)^{1/2} \rightarrow \infty$ using Assumption A7. This now establishes that $D_{1n} = o_p(b^{d/2})$.

We next consider D_{2n} . First note that with probability one

$$\hat{\theta} = \int \min\{f_n(x), g_n(x)\} dx = \int_{C_n} \min\{f_n(x), g_n(x)\} dx,$$

where $C_n = [l_{1n}, u_{1n}] \times \cdots \times [l_{dn}, u_{dn}]$ with

$$\begin{aligned} l_{jn} &= \max\left\{\min_{1 \leq i \leq n} X_{ji}, \min_{1 \leq i \leq n} Y_{ji}\right\} - b/2 \\ u_{jn} &= \min\left\{\max_{1 \leq i \leq n} X_{ji}, \max_{1 \leq i \leq n} Y_{ji}\right\} + b/2. \end{aligned}$$

This holds since the kernel has compact support with radius $1/2$. It follows that we can restrict any integration to sets intersected with C_n .

Using the identity $x - y = (x^{1/2} - y^{1/2})(y^{1/2} + x^{1/2})$, we have

$$\int_{\hat{C}_{f,g} \cap C_n} |f_n^{1/2}(x) - f^{1/2}(x)| dx = \int_{\hat{C}_{f,g} \cap C_n} \frac{|f_n(x) - f(x)|}{f^{1/2}(x) + f_n^{1/2}(x)} dx \leq \int_{\hat{C}_{f,g} \cap C_n} \frac{|f_n(x) - f(x)|}{f^{1/2}(x)} dx$$

because $f_n^{1/2}(x) \geq 0$ for $x \in \hat{C}_{f,g} \cap C_n$. Then note that by the Cauchy-Schwarz and triangle inequalities

$$\begin{aligned} \left(\frac{nb^d}{\|K\|_2^2}\right)^{1/2} E \left[\frac{|f_n(x) - f(x)|}{f^{1/2}(x)} \right] &\leq \left(\frac{nb^d}{\|K\|_2^2}\right)^{1/2} E^{1/2} [|f_n(x) - f(x)|^2] \frac{1}{f^{1/2}(x)} \\ &\leq \left[\left(\frac{nb^d}{\|K\|_2^2}\right) \frac{|Ef_n(x) - f(x)|^2}{f(x)} + \left(\frac{nb^d}{\|K\|_2^2}\right) \frac{\text{var}(f_n(x))}{f(x)} \right]^{1/2}. \end{aligned}$$

Then, we use the inequality $(a + b)^{1/2} \leq 1 + a^{1/2} + b^{1/2}$ for all positive a, b , to obtain that

$$\begin{aligned} \left(\frac{nb^d}{\|K\|_2^2}\right)^{1/2} E \left[\frac{|f_n(x) - f(x)|}{f^{1/2}(x)} \right] &\leq 1 + \left[\left(\frac{nb^d}{\|K\|_2^2}\right)^{1/2} \frac{|Ef_n(x) - f(x)|}{f^{1/2}(x)} \right] + \left[\left(\frac{nb^d}{\|K\|_2^2}\right) \frac{\text{var}(f_n(x))}{f(x)} \right]^{1/2} \\ &\leq 2 + \left[\left(\frac{nb^d}{\|K\|_2^2}\right)^{1/2} b^s \sum_{|\lambda|=s} \frac{1}{s!} \frac{|D^\lambda f(x)|}{f^{1/2}(x)} \int u^\lambda K(u) du \right] + o(1), \end{aligned}$$

where the second inequality follows by standard kernel arguments and is uniform in $x \in \mathbb{R}^d$. The bias term is of smaller order under our conditions given the absolute integrability of $D^\lambda f(x)/f^{1/2}(x)$. Therefore,

$$\int_{\hat{C}_{f,g} \cap C_n} |f_n^{1/2}(x) - f^{1/2}(x)| dx \leq 2\mu(C_n) \left(\frac{\|K\|_2^2}{nb^d}\right)^{1/2} (1 + o(1)),$$

$$b^{-d/2} D_{2n} = O_p(n^{-1/2} b^{-d} r_n),$$

where $r_n = \prod_{j=1}^d (u_{jn} - l_{jn})$. For Gaussian X, Y , $r_n = O_p((\log n)^{d/2})$ so that this term is small under our conditions. More generally we can bound r_n under our moment condition. Specifically, by the Bonferroni and Markov inequalities

$$\Pr \left[\max_{1 \leq i \leq n} |X_{ji}| > \delta_n \right] \leq \sum_{i=1}^n \Pr [|X_{ji}| > \delta_n] \leq n \frac{E|X_{ji}|^p}{\delta_n^p},$$

provided $E|X_{ji}|^p < \infty$. Therefore, we take $\delta_n = n^{1/p}L(n)$, where $L(n) \rightarrow \infty$ is a slowly varying function. We then show that we can condition on the event that $\{r_n \leq n^{1/p}L(n)\}$, which has probability tending to one.

This completes the proof of (12). The consistency of the standard error follows by similar arguments. ■

A.3 Proofs of Lemmas

Proof of Lemma A3. To establish (36), we first notice that, for $\|x - y\| > b$, the random variables $(f_N(x) - Ef_n(x), g_N(x) - Eg_n(x))$ and $(f_N(y) - Ef_n(y), g_N(y) - Eg_n(y))$ are independent because they are functions of independent increments of Poisson processes and the kernel K vanishes outside of the closed ball of radius $1/2$. This implies that

$$\text{cov}(A_{1n}^P(B_0), A_{2n}^P(B_f)) = \text{cov}(A_{1n}^P(B_0), A_{3n}^P(B_g)) = 0. \quad (50)$$

On the other hand, by standard arguments for kernel densities, we have as $n \rightarrow \infty$

$$\begin{aligned} \text{var}(A_{2n}^P(B_f)) &= E \left(\int_{B_f} K_b(x - X) dx \right)^2 \rightarrow p_{f,B} \\ \text{var}(A_{3n}^P(B_g)) &= E \left(\int_{B_g} K_b(x - Y) dx \right)^2 \rightarrow p_{g,B} \\ \text{cov}(A_{2n}^P(B_f), A_{3n}^P(B_g)) &= E \left(\int_{B_f} K_b(x - X) dx \right) \left(\int_{B_g} K_b(x - Y) dx \right) \rightarrow p_{f,g,B}. \end{aligned} \quad (51)$$

Therefore, by (50) and (51), the proof of Lemma A3 is complete if we show

$$\lim_{n \rightarrow \infty} \text{var}(A_{1n}^P(B_0)) = p_{0,B} \sigma_0^2. \quad (52)$$

To show (52), note that

$$\begin{aligned} &\text{var}(A_{1n}^P(B_0)) \\ &= n \int_{B_0} \int_{B_0} \text{cov}(\min\{f_N(x) - Ef_n(x), g_N(x) - Eg_n(x)\}, \min\{f_N(y) - Ef_n(y), g_N(y) - Eg_n(y)\}) dx dy \\ &= n \int_{B_0} \int_{B_0} 1(\|x - y\| \leq b) \\ &\quad \times \text{cov}(\min\{f_N(x) - Ef_n(x), g_N(x) - Eg_n(x)\}, \min\{f_N(y) - Ef_n(y), g_N(y) - Eg_n(y)\}) dx dy. \end{aligned}$$

Let

$$T_{f,N}(x) = \frac{\sqrt{n}\{f_N(x) - Ef_n(x)\}}{\sqrt{k_{f,n}(x)}}, \quad (53)$$

$$T_{g,N}(x) = \frac{\sqrt{n}\{g_N(x) - Eg_n(x)\}}{\sqrt{k_{g,n}(x)}}, \quad (54)$$

where $k_{f,n}(x) = \text{nvar}(f_N(x))$ and $k_{g,n}(x) = \text{nvar}(g_N(x))$. By standard arguments, we have that, with $\mu(B_0) < \infty$,

$$\sup_{x \in B_0} \left| \sqrt{k_{f,n}(x)} - b^{-d/2} \|K\|_2 \sqrt{f(x)} \right| = O(b^{d/2}) \quad (55)$$

$$\sup_{x \in B_0} \left| \sqrt{k_{g,n}(x)} - b^{-d/2} \|K\|_2 \sqrt{g(x)} \right| = O(b^{d/2}) \quad (56)$$

$$\int_{B_0} \int_{B_0} 1(\|x - y\| \leq b) dx dy = O(b^d) \quad (57)$$

$$\sup_{x, y \in \mathbb{R}^d} |\text{cov}(\min\{T_{f,N}(x), T_{g,N}(x)\}, \min\{T_{f,N}(y), T_{g,N}(y)\})| = O(1), \quad (58)$$

where (55) and (56) holds by two term Taylor expansions and (58) follows from Cauchy-Schwarz inequality and the elementary result $|\min\{a, b\}| \leq |a| + |b|$. Therefore, from (55) - (58), we have that

$$\text{var}(A_{1n}^P(B_0)) = \bar{\sigma}_{n,0}^2 + o(1),$$

where

$$\begin{aligned} \bar{\sigma}_{n,0}^2 &= \int_{B_0} \int_{B_0} 1(\|x - y\| \leq b) \text{cov}(\min\{T_{f,N}(x), T_{g,N}(x)\}, \min\{T_{f,N}(y), T_{g,N}(y)\}) \\ &\quad \times b^{-d} \|K\|_2^2 \sqrt{f(x)f(y)} dx dy. \end{aligned} \quad (59)$$

Now, let $\mathbb{Z}_n(x, y) = (Z_{1n}(x), Z_{2n}(x), Z_{3n}(y), Z_{4n}(y))^\top$ for $x, y \in \mathbb{R}^d$, be a mean zero multivariate Gaussian process such that for each $x, y \in \mathbb{R}^d$, $\mathbb{Z}_n(x, y)$ and $\mathbb{T}_n(x, y) = (T_{f,N}(x), T_{g,N}(x), T_{f,N}(y), T_{g,N}(y))^\top$ have the same covariance structure. That is,

$$\mathbb{Z}_n(x, y) \stackrel{d}{=} \Sigma_n^{-1/2}(x, y) Z$$

where $Z = (Z_1, Z_2, Z_3, Z_4)^\top$ is the standard normal ($N(0, I_4)$) random vector and

$$\Sigma_n(x, y) = E\mathbb{T}_n(x, y)\mathbb{T}_n(x, y)^\top. \quad (60)$$

Let

$$\begin{aligned} \bar{\tau}_{n,0}^2 &= \int_{B_0} \int_{B_0} 1(\|x - y\| \leq b) \text{cov}(\min\{Z_{1n}(x), Z_{2n}(x)\}, \min\{Z_{3n}(y), Z_{4n}(y)\}) \\ &\quad \times b^{-d} \|K\|_2^2 \sqrt{f(x)f(y)} dx dy. \end{aligned} \quad (61)$$

By a change of variables $y = x + tb$, we can write

$$\begin{aligned} \bar{\tau}_{n,0}^2 &= \int_{B_0} \int_{T_0} 1(x \in B_0) 1(x + tb \in B_0) \text{cov}(\min\{Z_{1n}(x), Z_{2n}(x)\}, \min\{Z_{3n}(x + tb), Z_{4n}(x + tb)\}) \\ &\quad \times \|K\|_2^2 \sqrt{f(x)f(x + tb)} dx dt, \end{aligned} \quad (62)$$

where $T_0 = \{t \in \mathbb{R}^d : \|t\| \leq 1\}$. Let

$$\begin{aligned}\rho_{f,n}^*(x, y) &= E [T_{f,N}(x)T_{f,N}(y)] \\ \rho_{g,n}^*(x, y) &= E [T_{g,N}(x)T_{g,N}(y)] \\ \rho_{f,g,n}^*(x, y) &= E [T_{f,N}(x)T_{g,N}(y)].\end{aligned}$$

Observe that, for almost every $x \in B_0$, we have

$$\begin{aligned}\rho_{f,n}^*(x, x+tb) &= \frac{b^{-d}E[K((x-X)/b)K((x-X)/b+t)]}{\sqrt{b^{-2d}EK^2((x-X)/b)EK^2((x-X)/b+t)}} \\ &\rightarrow \frac{\int_{\mathbb{R}^d} K(u)K(u+t)du}{\|K\|_2^2} = \rho(t).\end{aligned}\tag{63}$$

Similarly, we have $\rho_{g,n}^*(x, x+tb) \rightarrow \rho(t)$ as $n \rightarrow \infty$ for almost all $x \in B_0$. (In fact, under our assumptions, the convergence (63) holds uniformly over $(x, t) \in B_0 \times T_0$.) Similarly, for almost every $x, y \in B_0$, we have

$$\begin{aligned}\rho_{f,g,n}^*(x, y) &= \frac{b^{-d}E[K((x-X)/b)K((y-Y)/b)]}{\sqrt{b^{-2d}EK^2((x-X)/b)EK^2((y-Y)/b)}} \\ &= O(b^d) \rightarrow 0.\end{aligned}\tag{64}$$

Therefore, by (63), (64) and the bounded convergence theorem, we have

$$\lim_{n \rightarrow \infty} \bar{\tau}_{n,0}^2 = p_{0,B}\sigma_0^2.$$

Now, the desired result (36) holds if we establish

$$\bar{\tau}_{n,0}^2 - \bar{\sigma}_{n,0}^2 \rightarrow 0.\tag{65}$$

Set

$$G_n(x, t) = \|K\|_2^2 1(x \in B_0) 1(x+tb \in B_0) \sqrt{f(x)f(x+tb)}.$$

Notice that

$$\int_{B_0} \int_{T_0} G_n(x, t) dx dt \leq \|K\|_2^2 \mu(T_0 \times B_0) \sup_{x \in B_0} |f(x)| =: \bar{\beta} < \infty.\tag{66}$$

Let $\varepsilon_n \in (0, b]$ be a sequence such that $\varepsilon_n/b \rightarrow 0$. Letting $T_{0,n} = \{t \in \mathbb{R}^d : \varepsilon_n/b \leq \|t\| \leq 1\}$, define

$$\begin{aligned}\bar{\sigma}_{n,0}^2(\varepsilon_n) &= \int_{B_0} \int_{T_{0,n}} 1(x \in B_0) 1(x+tb \in B_0) \text{cov}(\min\{T_{f,N}(x), T_{g,N}(x)\}, \min\{T_{f,N}(x+tb), T_{g,N}(x+tb)\}) \\ &\quad \times \|K\|_2^2 \sqrt{f(x)f(x+tb)} dx dt,\end{aligned}$$

$$\begin{aligned}\bar{\tau}_{n,0}^2(\varepsilon_n) &= \int_{B_0} \int_{T_{0,n}} 1(x \in B_0) 1(x+tb \in B_0) \text{cov}(\min\{Z_{1n}(x), Z_{2n}(x)\}, \min\{Z_{3n}(x+tb), Z_{4n}(x+tb)\}) \\ &\quad \times \|K\|_2^2 \sqrt{f(x)f(x+tb)} dx dt.\end{aligned}$$

To show (65), we first establish

$$\bar{\tau}_{n,0}^2(\varepsilon_n) - \bar{\sigma}_{n,0}^2(\varepsilon_n) \rightarrow 0. \quad (67)$$

We have

$$\begin{aligned} \bar{\tau}_{n,0}^2(\varepsilon_n) - \bar{\sigma}_{n,0}^2(\varepsilon_n) &= \left| \int_{B_0} \int_{T_{0,n}} [\text{cov}(\min\{Z_{1n}(x), Z_{2n}(x)\}, \min\{Z_{3n}(x+tb), Z_{4n}(x+tb)\}) \right. \\ &\quad \left. - \text{cov}(\min\{T_{f,N}(x), T_{g,N}(x)\}, \min\{T_{f,N}(x+tb), T_{g,N}(x+tb)\})] G_n(x, t) dx dt \right| \\ &\leq \int_{B_0} \int_{T_{0,n}} |E \min\{Z_{1n}(x), Z_{2n}(x)\} E \min\{Z_{3n}(x+tb), Z_{4n}(x+tb)\} \\ &\quad - E \min\{T_{f,N}(x), T_{g,N}(x)\} E \min\{T_{f,N}(x+tb), T_{g,N}(x+tb)\}| G_n(x, t) dx dt \\ &\quad + \int_{B_0} \int_{T_{0,n}} |E \min\{Z_{1n}(x), Z_{2n}(x)\} \min\{Z_{3n}(x+tb), Z_{4n}(x+tb)\} \\ &\quad - E \min\{T_{f,N}(x), T_{g,N}(x)\} \min\{T_{f,N}(x+tb), T_{g,N}(x+tb)\}| G_n(x, t) dx dt \\ &=: \Delta_{1n} + \Delta_{2n}. \end{aligned} \quad (68)$$

First, consider Δ_{1n} . Let η_1 denote a Poisson random variable with mean 1 that are independent of $\{(X_i, Y_i) : i \geq 1\}$ and set

$$Q_{f,n}(x) = \left[\sum_{j \leq \eta_1} K\left(\frac{x - X_j}{b}\right) - EK\left(\frac{x - X}{b}\right) \right] / \sqrt{EK^2\left(\frac{x - X}{b}\right)} \quad (69)$$

$$Q_{g,n}(x) = \left[\sum_{j \leq \eta_1} K\left(\frac{x - Y_j}{b}\right) - EK\left(\frac{x - Y}{b}\right) \right] / \sqrt{EK^2\left(\frac{x - Y}{b}\right)}. \quad (70)$$

Notice that, with $f(x) = g(x) \geq \delta > 0$ for all $x \in B_0$, we have

$$\sup_{x \in B_0} E |Q_{\phi,n}(x)|^3 = O(b^{-d/2}) \text{ for } \phi = f \text{ and } g. \quad (71)$$

Let $Q_{\phi,n}^{(1)}(x), \dots, Q_{\phi,n}^{(n)}(x)$ be i.i.d $Q_{\phi,n}(x)$ for $\phi = f$ and g . Clearly, we have

$$T_{f,N}(x) = \frac{\sqrt{n} \{f_N(x) - Ef_n(x)\}}{\sqrt{b^{-2d} EK^2((x - X)/b)}} \stackrel{d}{=} \frac{\sum_{i=1}^n Q_{f,n}^{(i)}(x)}{\sqrt{n}} \quad (72)$$

$$T_{g,N}(x) = \frac{\sqrt{n} \{g_N(x) - Eg_n(x)\}}{\sqrt{b^{-2d} EK^2((x - Y)/b)}} \stackrel{d}{=} \frac{\sum_{i=1}^n Q_{g,n}^{(i)}(x)}{\sqrt{n}}. \quad (73)$$

Therefore, by Lemma A1(a), (71), (72), and (73), we have

$$\sup_{x \in B_0} |E \min\{T_{f,N}(x), T_{g,N}(x)\} - E \min\{Z_{1n}(x), Z_{2n}(x)\}| \leq O\left(\frac{1}{\sqrt{nb^d}}\right). \quad (74)$$

The results (74) and (66) imply that $\Delta_{1n} = o(1)$. We next consider Δ_{2n} . Under Assumption A1, we can expand $\rho(t)$ in a Taylor series to get

$$\rho(t) = 1 - C \|t\|^\alpha + o(\|t\|^\alpha) \text{ as } \|t\| \rightarrow 0 \quad (75)$$

for some constants $C \geq 0$ and $0 < \alpha \leq 2$, see Liero (1982, p.178). Therefore, we have

$$\begin{aligned} \Delta_{2n} &\leq \sup_{(x,t) \in B_0 \times T_{0,n}} |E \min\{Z_{1n}(x), Z_{2n}(x)\} \min\{Z_{3n}(x+tb), Z_{4n}(x+tb)\} \\ &\quad - E \min\{T_{f,N}(x), T_{g,N}(x)\} \min\{T_{f,N}(x+tb), T_{g,N}(x+tb)\}| \cdot \bar{\beta} \\ &\leq O\left(\left(\frac{\varepsilon_n}{b}\right)^{-3\alpha}\right) \cdot O\left(\frac{1}{\sqrt{nb^d}}\right), \end{aligned} \quad (76)$$

where the first inequality uses (66) and the second inequality holds by Lemma A1(a), (71), (75), and the fact that $\lim_{n \rightarrow \infty} \rho_{f,n}^*(x, x+tb) = \lim_{n \rightarrow \infty} \rho_{g,n}^*(x, x+tb) = \rho(t)$ a.s. uniformly over $(x, t) \in B_0 \times T_0$. Since ε_n is arbitrary, we can choose $\varepsilon_n = b(\log n)^{-1/(6\alpha)}$. This choice of ε_n makes the right hand side of (76) $o(1)$, using Assumption A3. Therefore, we have $\Delta_{2n} = o(1)$, and hence (67).

On the other hand, using (58) and argument similar to (74), we have

$$\begin{aligned} &|[\bar{\sigma}_{n,0}^2 - \bar{\tau}_{n,0}^2] - [\bar{\sigma}_{n,0}^2(\varepsilon_n) - \bar{\tau}_{n,0}^2(\varepsilon_n)]| \\ &\leq O\left(\frac{1}{\sqrt{nb^d}}\right) + O\left(\left(\frac{\varepsilon_n}{b}\right)^d\right) = o(1). \end{aligned}$$

This and (67) establish (65), and hence (52), as desired. ■

Proof of Lemma A4. Let

$$\begin{aligned} \Delta_n(x) &= \sqrt{n} [\min\{f_N(x) - Ef_n(x), g_N(x) - Eg_n(x)\} \\ &\quad - nE \min\{f_N(x) - Ef_n(x), g_N(x) - Eg_n(x)\}], \\ \Delta_{f,n}(x) &= \sqrt{n} \{f_N(x) - Ef_n(x)\}, \quad \Delta_{g,n}(x) = \sqrt{n} \{g_N(x) - Eg_n(x)\} \end{aligned}$$

We first construct partitions of $\mathcal{B}(M)$, $\mathcal{B}(M_f)$ and $\mathcal{B}(M_g)$. Consider the regular grid

$$G_{\mathbf{i}} = (x_{i_1}, x_{i_1+1}] \times \cdots \times (x_{i_d}, x_{i_d+1}],$$

where $\mathbf{i} = (i_1, \dots, i_d)$, $i_1, \dots, i_d \in \mathbb{Z}$ and $x_i = ib$ for $i \in \mathbb{Z}$. Define

$$\begin{aligned} R_{0,\mathbf{i}} &= G_{\mathbf{i}} \cap \mathcal{B}(M), R_{f,\mathbf{i}} = G_{\mathbf{i}} \cap \mathcal{B}(M_f), R_{g,\mathbf{i}} = G_{\mathbf{i}} \cap \mathcal{B}(M_g), \\ \mathcal{I}_n &= \{\mathbf{i} : R_{0,\mathbf{i}} \cup R_{f,\mathbf{i}} \cup R_{g,\mathbf{i}} \neq \emptyset\}. \end{aligned}$$

Then, we see that $\{R_{0,\mathbf{i}} : \mathbf{i} \in \mathcal{I}_n \subset \mathbb{Z}^d\}$, $\{R_{f,\mathbf{i}} : \mathbf{i} \in \mathcal{I}_n \subset \mathbb{Z}^d\}$ and $\{R_{g,\mathbf{i}} : \mathbf{i} \in \mathcal{I}_n \subset \mathbb{Z}^d\}$ are partitions of $\mathcal{B}(M)$, $\mathcal{B}(M_f)$ and $\mathcal{B}(M_g)$, respectively, with

$$\begin{aligned} \mu(R_{0,\mathbf{i}}) &\leq d_0 b^d, \quad \mu(R_{f,\mathbf{i}}) \leq d_1 b^d, \quad \mu(R_{g,\mathbf{i}}) \leq d_2 b^d \\ m_n &= : \mu(\mathcal{I}_n) \leq d_3 b^{-d} \end{aligned} \quad (77)$$

for some positive constants $d_0, \dots, d_3$, see, e.g., Mason and Polonik (2009) for a similar construction of partitions in a different context. Set

$$\begin{aligned}\alpha_{\mathbf{i},n} &= \int_{R_{0,\mathbf{i}}} 1(x \in B_0) \Delta_n(x) dx + \int_{R_{f,\mathbf{i}}} 1(x \in B_f) \Delta_{f,n}(x) dx + \int_{R_{g,\mathbf{i}}} 1(x \in B_g) \Delta_{g,n}(x) dx, \\ u_{\mathbf{i},n} &= \frac{1}{\sqrt{n}} \sum_{j=1}^N [1 \{ (X_j \in R_{f,\mathbf{i}}) \cup (X_j \in R_{0,\mathbf{i}}, Y_j \in R_{0,\mathbf{i}}) \cup (Y_j \in R_{g,\mathbf{i}}) \} \\ &\quad - n \Pr \{ (X \in R_{f,\mathbf{i}}) \cup (X \in R_{0,\mathbf{i}}, Y \in R_{0,\mathbf{i}}) \cup (Y \in R_{g,\mathbf{i}}) \}].\end{aligned}$$

Then, we have

$$A_n^P(\bar{B}) = \sum_{\mathbf{i} \in \mathcal{I}_n} \alpha_{\mathbf{i},n} \text{ and } U_n = \sum_{\mathbf{i} \in \mathcal{I}_n} u_{\mathbf{i},n}.$$

Notice that

$$\text{var}(A_n^P(\bar{B})) = \sigma_n^2(\bar{B}) \text{ and } \text{var}(U_n) = 1 - \alpha.$$

For arbitrary λ_1 and $\lambda_2 \in \mathbb{R}$, let

$$y_{\mathbf{i},n} = \lambda_1 \alpha_{\mathbf{i},n} + \lambda_2 u_{\mathbf{i},n}.$$

Notice that $\{y_{\mathbf{i},n} : \mathbf{i} \in \mathcal{I}_n\}$ is an array of mean zero one-dependent random fields. Below we will establish that

$$\begin{aligned}\text{var} \left(\sum_{\mathbf{i} \in \mathcal{I}_n} y_{\mathbf{i},n} \right) &= \text{var}(\lambda_1 A_n^P(\bar{B}) + \lambda_2 U_n) \\ &\rightarrow \lambda_1^2 (p_{0,B} \sigma_0^2 + \bar{\sigma}_{1,B}^2) + \lambda_2^2 (1 - \alpha) + 2\lambda_1 \lambda_2 (p_{f,B} + p_{g,B}),\end{aligned} \tag{78}$$

and

$$\sum_{\mathbf{i} \in \mathcal{I}_n} E |y_{\mathbf{i},n}|^r = o(1) \text{ for some } 2 < r < 3. \tag{79}$$

Then, the result of Lemma A4 follows from the CLT of Shergin (1990) and Cramér-Wold device.

We first establish (78). By Lemma A3, (78) holds if we have

$$\text{cov}(A_n^P(\bar{B}), U_n) \rightarrow p_{f,B} + p_{g,B}. \tag{80}$$

Recall that

$$A_n^P(\bar{B}) = A_{1n}^P(B_0) + A_{2n}^P(B_f) + A_{3n}^P(B_g).$$

Therefore, (80) holds if we have

$$\text{cov}(A_{2n}^P(B_f), U_n) \rightarrow p_{f,B}, \quad \text{cov}(A_{3n}^P(B_g), U_n) \rightarrow p_{g,B}, \tag{81}$$

$$\begin{aligned}\text{cov}(A_{1n}^P(B_0), U_n) &= \text{cov} \left(\sqrt{n} \int_{B_0} \min \{ f_N(x) - E f_n(x), g_N(x) - E g_n(x) \} dx, U_n \right) \\ &= O \left(\frac{1}{\sqrt{n b^{2d}}} \right)\end{aligned} \tag{82}$$

(81) holds by a standard argument. We show below (82). For any $x \in B_0$, for $T_{f,N}, T_{g,N}$ defined in (72), (73), we have

$$\left(T_{f,N}(x), T_{g,N}(x), \frac{U_n}{\sqrt{\Pr((X, Y) \in \mathcal{T}(M))}} \right) \stackrel{d}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(Q_{f,n}^{(i)}(x), Q_{g,n}^{(i)}(x), U^{(i)} \right), \quad (83)$$

where $(Q_{f,n}^{(i)}(x), Q_{g,n}^{(i)}(x), U^{(i)})$ for $i = 1, \dots, n$ are i.i.d $(Q_{f,n}(x), Q_{g,n}(x), U)$, with $(Q_{f,n}(x), Q_{g,n}(x))$ defined as in (69) and (70) and

$$U = \left[\sum_{j \leq \eta_1} 1((X_j, Y_j) \in \mathcal{T}(M)) - \Pr((X, Y) \in \mathcal{T}(M)) \right] / \sqrt{\Pr((X, Y) \in \mathcal{T}(M))}.$$

Notice that, for $\phi = f$ and g , we have

$$\sup_{x \in B_0} |\text{cov}(Q_{\phi,n}(x), U)| = O(b^{d/2}), \quad (84)$$

which in turn is less than or equal to ε for all sufficiently large n and any $0 < \varepsilon < 1/2$. This result and Lemma A1(b) imply that

$$\sup_{x \in B_0} |\text{cov}(\sqrt{n} \min\{f_N(x) - Ef_n(x), g_N(x) - Eg_n(x)\}, U_n)| \leq O\left(\frac{1}{\sqrt{nb^{2d}}}\right),$$

which, when combined with $\mu(B_0) < \infty$, yields (82) and hence (78), as desired.

We next establish (79). Notice that, with $2 < r < 3$, using Liapunov inequality and c_r -inequality, we have

$$\begin{aligned} & (E|\alpha_{\mathbf{i},n}|^r)^{3/r} \\ & \leq 9 \left(\int_{R_{0,\mathbf{i}}} \int_{R_{0,\mathbf{i}}} \int_{R_{0,\mathbf{i}}} 1_{B_0}(x) 1_{B_0}(y) 1_{B_0}(z) E|\Delta_n(x) \Delta_n(y) \Delta_n(z)| dx dy dz \right. \\ & \quad + \int_{R_{f,\mathbf{i}}} \int_{R_{f,\mathbf{i}}} \int_{R_{f,\mathbf{i}}} 1_{B_f}(x) 1_{B_f}(y) 1_{B_f}(z) E|\Delta_{f,n}(x) \Delta_{f,n}(y) \Delta_{f,n}(z)| dx dy dz \\ & \quad \left. + \int_{R_{g,\mathbf{i}}} \int_{R_{g,\mathbf{i}}} \int_{R_{g,\mathbf{i}}} 1_{B_g}(x) 1_{B_g}(y) 1_{B_g}(z) E|\Delta_{g,n}(x) \Delta_{g,n}(y) \Delta_{g,n}(z)| dx dy dz \right), \end{aligned} \quad (85)$$

where $1_B(x) = 1(x \in B)$. Also, by c_r -inequality again and the elementary result $|\min\{X, Y\}| \leq |X| + |Y|$, we have: for some constant $D > 0$,

$$E|\Delta_n(x)|^3 \leq Dn^{3/2} \{E|f_N(x) - Ef_n(x)|^3 + E|g_N(x) - Eg_n(x)|^3\}. \quad (86)$$

By the Rosenthal's inequality (see, e.g., Lemma 2.3 of Giné et al. (2003)), we have:

$$\sup_{x \in B_0} n^{3/2} E|f_N(x) - Ef_n(x)|^3 \leq O\left(\frac{1}{b^{3d/2}} + \frac{1}{n^{1/2}b^{2d}}\right). \quad (87)$$

A similar result holds for g_N . Now, Assumption A3, (85), (86), (87), the elementary result $E|XYZ| \leq E(|X| + |Y| + |Z|)^3$ and the fact that $\mu(R_{0,\mathbf{i}}) \leq d_1 b^d$ imply that the first term on the right hand side of (85) is bounded by a $O(b^{rd/2})$ term uniformly in $\mathbf{i} \in \mathcal{I}_n$. Similar results hold for the other terms on the right hand side of (85). Therefore, we have

$$E|\alpha_{\mathbf{i},n}|^r \leq O(b^{rd/2}) \text{ uniformly in } \mathbf{i} \in \mathcal{I}_n. \quad (88)$$

This implies that

$$\sum_{\mathbf{i} \in \mathcal{I}_n} E|\alpha_{\mathbf{i},n}|^r \leq O(m_n b^{rd/2}) = O(b^{(r-2)/d}) = o(1). \quad (89)$$

On the other hand, set

$$p_{\mathbf{i},n} = \Pr \{ (X \in R_{f,\mathbf{i}}) \cup (X \in R_{0,\mathbf{i}}, Y \in R_{0,\mathbf{i}}) \cup (Y \in R_{g,\mathbf{i}}) \}.$$

Then, by the Rosenthal's inequality, there exists a constant $D_1 > 0$ such that

$$\begin{aligned} \sum_{\mathbf{i} \in \mathcal{I}_n} E|u_{\mathbf{i},n}|^r &\leq D_1 n^{-r/2} \sum_{\mathbf{i} \in \mathcal{I}_n} ((np_{\mathbf{i},n})^{r/2} + np_{\mathbf{i},n}) \\ &\leq D_1 \max_{\mathbf{i} \in \mathcal{I}_n} ((p_{\mathbf{i},n})^{(r-2)/2} + n^{-1/2}) \rightarrow 0. \end{aligned} \quad (90)$$

Therefore, combining (89) and (90), we have (79). This completes the proof of Lemma A4. $\blacksquare$

Proof of Lemma A5. Consider part (a) first. Let $T_{f,N}(x)$ and $T_{g,N}(x)$ be defined as in (53) and (54), respectively. We have

$$\begin{aligned} &\left| \int_{B_0} \left[\sqrt{n} E \min \{f_N(x) - Ef_N(x), g_N(x) - Eg_N(x)\} dx - E \min \{Z_1, Z_2\} k_{f,n}^{1/2}(x) \right] dx \right| \\ &\leq \sup_{x \in B_0} |E \min \{T_{f,N}(x), T_{g,N}(x)\} - E \min \{Z_1, Z_2\}| \cdot \sup_{x \in B_0} k_{f,n}^{1/2}(x) \cdot \mu(B_0) \\ &= O(n^{-1/2} b^{-d/2}) O(b^{-d/2}) = O(n^{-1/2} b^{-d}) = o(1), \end{aligned}$$

by Lemma A1(a), (55), Assumption A3, and the fact $\mu(B_0) < \infty$. Similarly, we have

$$\left| \int_{B_0} \left[\sqrt{n} E \min \{f_n(x) - Ef_n(x), g_n(x) - Eg_n(x)\} dx - E \min \{Z_1, Z_2\} \sqrt{n \text{var}(f_n(x))} \right] dx \right| = O(n^{-1/2} b^{-d}).$$

Therefore, part (b) also holds since, using (24) and (25),

$$\sup_{x \in B_0} \left| k_{f,n}^{1/2}(x) - \sqrt{n \text{var}(f_n(x))} \right| = O(b^{d/2}) = o(1).$$

$\blacksquare$

Proof of Lemma A6. Consider $A_n^P(\overline{B})$ defined in (31). Conditional on $N = n$, we have

$$\begin{aligned} A_n(\overline{B}) &\stackrel{d}{=} \sqrt{n} \int_{B_0} [\min \{f_n(x) - Ef_n(x), g_n(x) - Eg_n(x)\} \\ &\quad - E \min \{f_N(x) - Ef_N(x), g_N(x) - Eg_N(x)\}] dx \\ &\quad + \sqrt{n} \int_{B_f} [f_n(x) - Ef_n(x)] dx + \sqrt{n} \int_{B_g} [g_n(x) - Eg_n(x)] dx \\ &=: A_n^C(\overline{B}). \end{aligned} \quad (91)$$

By Lemmas A2 and A4, we have

$$\begin{aligned} A_n^C(\overline{B}) &\Rightarrow [p_{0,B}\sigma_0^2 + \bar{\sigma}_{1,B}^2 - (p_{f,B} + p_{g,B})^2]^{1/2} Z_1 \\ &= \sqrt{p_{0,B}\sigma_0^2 + \sigma_{1,B}^2} Z_1. \end{aligned}$$

Now, the result of Lemma A6 holds since, as $n \rightarrow \infty$

$$\begin{aligned} &A_n^C(\overline{B}) - A_n(\overline{B}) \\ &= \int_{B_0} [\sqrt{n}E \min \{f_N(x) - Ef_n(x), g_N(x) - Eg_n(x)\} \\ &\quad - \sqrt{n}E \min \{f_n(x) - Ef_n(x), g_n(x) - Eg_n(x)\}] dx \rightarrow 0. \end{aligned}$$

■

Proof of Lemma A7. We can establish Lemma A7 by modifying the majorization inequality results of Pinelis (1994). Let $(X_1^*, \dots, X_n^*)$ be an independent copy of $(X_1, \dots, X_n)$. For $i = 1, \dots, n$, let E_i and E_i^* denote the conditional expectations given $(X_1, \dots, X_i)$ and $(X_1, \dots, X_{i-1}, X_i^*)$. Let

$$\xi_i = E_i T_n - E_{i-1} T_n, \quad (92)$$

$$\eta_i = E_i (T_n - T_{n,-i}), \quad (93)$$

where

$$T_{n,-i} = \int_{\mathcal{B}} \min \left\{ \sum_{k \neq i}^n h_1(X_{1k}, x), \sum_{k \neq i}^n h_2(X_{2k}, x) \right\} dx.$$

Then, we have

$$T_n - ET_n = \xi_1 + \dots + \xi_n, \quad (94)$$

$$\xi_i = \eta_i - E_{i-1} \eta_i, \quad (95)$$

$$|\eta_i| \leq 2 \sum_{j=1}^2 \int_{\mathcal{B}} |h_j(X_{ji}, x)| dx, \quad (96)$$

where (94) follows from (92), (95) holds by independence of X_i 's, and (96) follows from the elementary inequality $|\min\{a+b, c+d\} - \min\{a, c\}| \leq 2(|b| + |d|)$. Let

$$\eta_i^* = E_i^* (T_{n,i}^* - T_{n,-i}),$$

where

$$T_{n,i}^* = \int_{\mathcal{B}} \min \left\{ \sum_{k \neq i}^n h_1(X_{1k}, x) + h_1(X_{1i}^*, x), \sum_{k \neq i}^n h_2(X_{2k}, x) + h_2(X_{2i}^*, x) \right\} dx.$$

Notice that the random variables η_i and η_i^* are conditionally independent given $(X_1, \dots, X_{i-1})$, and the conditional distributions of η_i and η_i^* given $(X_1, \dots, X_{i-1})$ are equivalent. Therefore, for any

convex function $q : \mathbb{R} \rightarrow \mathbb{R}$, we have

$$\begin{aligned}
E_{i-1}q(\xi_i) &= E_{i-1}q(\eta_i - E_{i-1}\eta_i) \\
&\leq E_{i-1}q(\eta_i - E_{i-1}\eta_i - \eta_i^* - E_{i-1}\eta_i^*) \\
&\leq \frac{1}{2}E_{i-1}[q(2\eta_i) + q(-2\eta_i^*)] \\
&\leq \frac{1}{2}E_{i-1}\left[q\left(4\sum_{j=1}^2\int_{\mathcal{B}}|h_j(X_{ji}, x)|dx\right) + q\left(-4\sum_{j=1}^2\int_{\mathcal{B}}|h_j(X_{ji}, x)|dx\right)\right] \\
&= Eq\left(4\epsilon_i\sum_{j=1}^2\int_{\mathcal{B}}|h_j(X_{ji}, x)|dx\right),
\end{aligned}$$

where the first inequality follows from Berger (1991, Lemma 2.2), the second inequality holds by the convexity of q and the last inequality follows from the convexity of q and (96). Now, the result of Lemma A7 holds by (94) and Lemma 2.6 of Berger (1991). $\blacksquare$

Proof of Lemma A8. Let $R(x, r) = \prod_{i=1}^d [x_i - r, x_i + r]$ denote a closed rectangle in $\mathbb{R}^d$. We have

$$\begin{aligned}
&E\left(\sqrt{n}\int_{\mathcal{B}}\{h_n(x) - Eh_n(x)\}dx\right)^2 \\
&\leq D\left\{E\left(\frac{1}{b^d}\int_{\mathcal{B}}\left|K\left(\frac{x-X}{b}\right)\right|dx\right)^2 + E\left(\frac{1}{b^d}\int_{\mathcal{B}}\left|K\left(\frac{x-Y}{b}\right)\right|dx\right)^2\right\} \\
&\leq D\left(\sup_u |K(u)|\right)^2 \int_{\mathcal{B}} [f(x) + g(x)]dx \\
&\quad + D\left(\sup_u |K(u)|\right)^2 \int_{-\infty}^{\infty} |b^{-d}\Pr(X \in R(x, b/2)) - f(x)|dx + \int_{-\infty}^{\infty} |b^{-d}\Pr(Y \in R(x, b/2)) - g(x)|dx \\
&= D\left(\sup_u |K(u)|\right)^2 \int_{\mathcal{B}} [f(x) + g(x)]dx + o(1)
\end{aligned}$$

as $n \rightarrow \infty$, where the first inequality follows from Lemma A7. This results completes the proof of Lemma A8. $\blacksquare$

References

- [1] Anderson G.J. (2004) ‘‘Toward an Empirical Analysis of Polarization’’ *Journal of Econometrics* 122 1-26

- [2] Anderson G.J. and Y. Ge (2004) "Do Economic Reforms Accelerate Urban Growth: The Case of China" *Urban Studies* 11 2197-2210.
- [3] Anderson G.J. and Y. Ge (2005) "The Size Distribution of Chinese Cities" *Regional Science and Urban Economics* 35 756-776
- [4] Anderson G.J., Y. Ge and T.W. Leo (2010) "Distributional Overlap: Simple Multivariate Parametric and Non-Parametric Tests for Alienation, Convergence and General Distributional Difference Issues" *Econometric Reviews* 29, 1532-4168.
- [5] Andrews, D.W.K. (1995), "Nonparametric Kernel Estimation for Semiparametric Models." *Econometric Theory* 11, 560-596.
- [6] Andrews, D. W. K. (1999), Estimation When a Parameter Is on a Boundary. *Econometrica* 67, 1341-1383.
- [7] Andrews, D. W. K. and P. Guggenberger (2009), "Hybrid and Size-Corrected Subsampling Methods," *Econometrica* 77, 721-762.
- [8] Beirlant, J. and D. M. Mason (1995), "On the Asymptotic Normality of L_p -Norms of Empirical Functionals," *Math. Methods Statist.* 4, 1-19.
- [9] Berline A. (1999), "How to get central limit theorems for global errors of estimates," *Applications of Mathematics* 44, 81-96.
- [10] Berger, E. (1991), "Majorization, Exponential Inequalities and Almost Sure Behavior of Vector-valued Random Variables, *Annals of Probability* 19, 1206-1226.
- [11] Champernowne, D. G. and Frank Alan Cowell (1998). *Economic inequality and income distribution*. Cambridge University Press.
- [12] Chen, Song Xi (1999) Beta kernel estimators for density functions. *Computational Statistics & Data Analysis* 31, 2, 131-145
- [13] Chernozhukov, V., H. Hong, and E. Tamer (2007), "Estimation and confidence regions for parameter sets in econometric models. *Econometrica*, 75, 1243-1284.
- [14] Clemons, T.E. and E.L. Bradley (2000). "A nonparametric measure of the overlapping coefficient," *Computational Statistics & Data Analysis* 34, 51-61.
- [15] Cuevas A. and R. Fraiman (1997), "A plug-in approach to support estimation," *Annals of Statistics* 25, 2300-2312.

- [16] Devroye, L., and L. Györfi (1985). *Nonparametric Density Estimation: The L_1 view*. John Wiley, New York.
- [17] Duclos J-Y., J. Esteban and D. Ray (2004) "Polarization: Concepts, Measurement and Estimation" *Econometrica* 72 1737-1772.
- [18] Eddy, W.F. (1980). "Optimum kernel estimators of the mode," *The Annals of Statistics* 8, 870-882.
- [19] Esteban, J-M. And D. Ray (1994) "On the Measurement of Polarization" *Econometrica* 62, 819-851
- [20] Giné, E., D. M. Mason, and A. Y. Zaitsev (2003), "The L_1 -Norm Density Estimator Process," *Annals of Probability* 31, 719-768.
- [21] Giné, E., and A. Guillou (2002), "Rates of strong uniform consistency for multivariate kernel density estimators," *Annals of the institute of Henri Poincare* 6, 907-921.
- [22] Gustafsson, Bjorn, Li Shi and Terry Sicular (eds) (2007). *Inequality and Public Policy in China*, New York and Cambridge: Cambridge University Press
- [23] Hall, P. (1982), "On estimating the endpoint of a distribution," *Annals of Statistics* 10, 556-568.
- [24] Härdle, W., B. U. Park, and A. B. Tsybakov (1995), "Estimation of non-sharp support boundaries," *Journal of Multivariate Analysis* 55, 205-218.
- [25] Holmström and Klemelä (1992), "Asymptotic bounds for the expected L^1 error of a multivariate kernel density estimator," *Journal of Multivariate Analysis* 42, 245-266.
- [26] Horváth, L. (1991), "On L_p - norms of multivariate density estimators," *Annals of Statistics* 19, 1933-1949.
- [27] Imbens, G. W. and D. B. Rubin (1997), "Estimating outcome distributions for compliers in instrumental variables models," *Review of Economic Studies* 64, 555-574.
- [28] Kac, M. (1949). On deviations between theoretical and empirical distributions. *Proc. Natl. Acad. Sci.*, 35, 252-257.
- [29] Kitagawa, T. (2010). "Testing for Instrument Independence in the Selection Model," Working Paper, University College London.
- [30] Kleiber, C. and S. Kotz (2003) *Statistical Size Distributions in Economics and Actuarial Sciences*. Wiley Interscience.

- [31] Knight, John , Zhao Renwei and Li Shi (2001). ‘A spatial analysis of wages and incomes in urban China: divergent means, convergent inequality’, in Carl Riskin, Zhao Renwei and Li Shi (eds), *China’s Retreat from Equality: Income Distribution and Economic Transition*, Armonk, New York: M.E. Sharpe: 133-66.
- [32] Liero, H. (1982), "On the maximal deviation of the kernel regression function estimate," *Math. Operationsforsch. Statist., Ser. Statistics* 13, 171-182.
- [33] Linton, O., E. Maasoumi, and Y. -J. Whang (2005), "Consistent testing for stochastic dominance under general sampling schemes," *Review of Economic Studies*, 72, 735-765.
- [34] Mason, D. M. and W. Polonik (2009), "Asymptotic Normality of Plug-In Level Set Estimates," *Annals of Applied Probability* 19, 1108-1142.
- [35] Mizuno, S., T. Yamaguchi, A. Fukushima, Y. Matsuyama, and Y. Ohashi (2005): "Overlap coefficient for assessing the similarity of pharmacokinetic data between ethnically different populations", *Clinical Trials*, Vol. 2, No. 2, 174-181
- [36] Pinelis, I. F. (1994), "On a Majorization Inequality for Sums of Independent Random Variables," *Statistics and Probability Letters* 19, 97-99.
- [37] Prakasa Rao, B.L.S. (1983). *Nonparametric Functional Estimation*, Academic Press.
- [38] Reiser, B., and D. Faraggi (1999). Confidence intervals for the overlapping coefficient: the normal equal variance case. *Journal of the Royal Statistical Society, Series D.* 48, 413-418.
- [39] Ricklefs, R.E. and M. Lau (1980). "Bias and dispersion of overlap indices: results of some monte carlo simulations," *Ecology* 61, 1019-1024.
- [40] Rosenbaum, S. (1961), "Moments of truncated bivariate normal distribution," *Journal of the Royal Statistical Society: Series B* 23, 405-408.
- [41] Rosenblatt, M. (1975), "A quadratic measure of deviation of two-dimensional density estimates and a test of independence," *Annals of Statistics* 3, 1-14.
- [42] Samawi, H. M., Helu, A. and Vogel R. (2010). “A nonparametric test of symmetry based on the overlapping coefficient.” Forthcoming in *Journal of Applied Statistics*.
- [43] Schmid, F., and A. Schmidt (2006), "Nonparametric estimation of the coefficient of overlapping - theory and empirical application," *Computational Statistics and Data Analysis* 50, 1583-1596.
- [44] Shergin, V. V. (1990), "The Central Limit Theorem for Finitely Dependent Random Variables," In: *Proc. 5th Vilnius Conf. Probability and Mathematical Statistics*, Vol. II, Grigelionis, B., Prohorov, Y. V., Sazonov, V. V., and Statulevicius V. (Eds.) 424-431.

- [45] Sweeting, T.J. (1980), "Speeds of Convergence and Asymptotic Expansions in the Central Limit Theorem: A Treatment by Operators," *Annals of Probability* 8, 281-297.
- [46] Wand, M.P. and M.C. Jones (1995) *Kernel Smoothing*. Chapman and Hall, Boca Raton.
- [47] Weitzman, M (1970) "Measures of Overlap of Income Distributions of White and Negro Families in the U.S." Technical Paper 22 Bureau of the Census.
- [48] Yalonetsky G. (2010) "A Comparison between the Pearson – based dissimilarity index and the multiple – group overlap index" mimeo OPHI Oxford University.
- [49] Yitzhaki S., (1994) "Economic Distance and overlapping of distributions." *Journal of Econometrics* 61 147-159.
- [50] Zhang, X. and R. Kanbur (2001), "What difference do polarisation measures make? an application to China," *The Journal of Development Studies* 37, 85-98.

B Tables and Figures

Kernel	$K(u)$	σ_0^2	$\ K\ _2^2$
Uniform	$\mathbf{1} [u \leq 0.5]$	0.6135	1.000
Triangular	$(2 + 4u) \mathbf{1} [-0.5 \leq u \leq 0] + (2 - 4u) \mathbf{1} [0 < u \leq 0.5]$	0.6248	1.3334
Normal	$\frac{\phi(u)}{1-2\Phi(-0.5)} \mathbf{1} [u \leq 0.5]$	0.6167	1.0014
Epanechnikov	$6 \left(\frac{1}{4} - u^2\right) \mathbf{1} [u \leq 0.5]$	0.6175	1.1999
Biweight	$30 \left(\frac{1}{4} - u^2\right)^2 \mathbf{1} [u \leq 0.5]$	0.6169	1.4275

Table 1. Constants for selected kernels

n	b	c	$\hat{\theta}$				$\hat{\theta}^{bc}$			
			bias	mbias	std	iqr	bias	mbias	std	iqr
100	b_s	b	-0.09244	-0.09037	0.03757	0.03793	-0.06353	-0.06129	0.04442	0.06358
		$b^{3/2}$	-0.09244	-0.09037	0.03757	0.03793	-0.07839	-0.07568	0.04138	0.04959
		$b^{2/3}$	-0.09244	-0.09037	0.03757	0.03793	-0.04953	-0.04739	0.04462	0.07284
	$b_s^{3/2}$	b	-0.10204	-0.10212	0.03922	0.03938	-0.08702	-0.08765	0.04263	0.05168
		$b^{3/2}$	-0.10204	-0.10212	0.03922	0.03938	-0.09724	-0.09751	0.04032	0.04345
		$b^{2/3}$	-0.10204	-0.10212	0.03922	0.03938	-0.07043	-0.07082	0.04569	0.06578
200	b_s	b	-0.07618	-0.07574	0.02686	0.02591	-0.05260	-0.05216	0.03146	0.04526
		$b^{3/2}$	-0.07618	-0.07574	0.02686	0.02591	-0.06518	-0.06495	0.02959	0.03516
		$b^{2/3}$	-0.07618	-0.07574	0.02686	0.02591	-0.04116	-0.04005	0.03090	0.05381
	$b_s^{3/2}$	b	-0.08067	-0.08110	0.02711	0.02669	-0.06856	-0.06882	0.02930	0.03651
		$b^{3/2}$	-0.08067	-0.08110	0.02711	0.02669	-0.07722	-0.07766	0.02773	0.02935
		$b^{2/3}$	-0.08067	-0.08110	0.02711	0.02669	-0.05325	-0.05279	0.03149	0.04830
400	b_s	b	-0.06512	-0.06472	0.01951	0.01935	-0.04582	-0.04496	0.02278	0.03510
		$b^{3/2}$	-0.06512	-0.06472	0.01951	0.01935	-0.05649	-0.05643	0.02149	0.02639
		$b^{2/3}$	-0.06512	-0.06472	0.01951	0.01935	-0.03701	-0.03605	0.02176	0.04148
	$b_s^{3/2}$	b	-0.06614	-0.06508	0.02010	0.02144	-0.05632	-0.05548	0.02168	0.02926
		$b^{3/2}$	-0.06614	-0.06508	0.02010	0.02144	-0.06362	-0.06305	0.02051	0.02337
		$b^{2/3}$	-0.06614	-0.06508	0.02010	0.02144	-0.04267	-0.04219	0.02337	0.04041
800	b_s	b	-0.05417	-0.05449	0.01383	0.01395	-0.03821	-0.03821	0.01585	0.02683
		$b^{3/2}$	-0.05417	-0.05449	0.01383	0.01395	-0.04730	-0.04724	0.01517	0.01954
		$b^{2/3}$	-0.05417	-0.05449	0.01383	0.01395	-0.03194	-0.03194	0.01486	0.03097
	$b_s^{3/2}$	b	-0.05260	-0.05247	0.01407	0.01393	-0.04474	-0.04480	0.01515	0.01999
		$b^{3/2}$	-0.05260	-0.05247	0.01407	0.01393	-0.05078	-0.05055	0.01429	0.01522
		$b^{2/3}$	-0.05260	-0.05247	0.01407	0.01393	-0.03275	-0.03249	0.01648	0.02940
1600	b_s	b	-0.04594	-0.04558	0.01022	0.01032	-0.03266	-0.03188	0.01165	0.02083
		$b^{3/2}$	-0.04594	-0.04558	0.01022	0.01032	-0.04031	-0.03951	0.01134	0.01485
		$b^{2/3}$	-0.04594	-0.04558	0.01022	0.01032	-0.02871	-0.02837	0.01069	0.02338
	$b_s^{3/2}$	b	-0.04272	-0.04246	0.01050	0.01040	-0.03639	-0.03627	0.01129	0.01540
		$b^{3/2}$	-0.04272	-0.04246	0.01050	0.01040	-0.04137	-0.04108	0.01067	0.01147
		$b^{2/3}$	-0.04272	-0.04246	0.01050	0.01040	-0.02577	-0.02590	0.01233	0.02362

Table 2. Univariate Monte Carlo Simulations

d	n	$\hat{\theta}$				$\hat{\theta}^{bc}$			
		bias	mbias	std	iqr	bias	mbias	std	iqr
1	100	-0.18971	-0.18749	0.03707	0.03745	-0.00541	-0.00414	0.03810	0.03917
	200	-0.14374	-0.14205	0.02518	0.02542	-0.00097	0.00055	0.02564	0.02528
	400	-0.11026	-0.10883	0.01826	0.01857	-0.00046	0.00109	0.01847	0.01874
	800	-0.08466	-0.08429	0.01252	0.01213	-0.00062	-0.00018	0.01258	0.01212
	1600	-0.06422	-0.06349	0.00862	0.00845	0.00002	0.00078	0.00865	0.00859
2	100	-0.48512	-0.48419	0.03945	0.03999	-0.09285	-0.09210	0.04499	0.04654
	200	-0.40276	-0.40251	0.02857	0.02972	-0.06028	-0.06027	0.03164	0.03131
	400	-0.33145	-0.33172	0.01888	0.01882	-0.03869	-0.03921	0.02060	0.02051
	800	-0.27078	-0.27047	0.01227	0.01131	-0.02506	-0.02517	0.01311	0.01205
	1600	-0.21984	-0.21987	0.00825	0.00852	-0.01598	-0.01609	0.00871	0.00908
3	100	-0.80363	-0.80339	0.03575	0.03821	-0.28710	-0.28707	0.03837	0.04088
	200	-0.73850	-0.73884	0.02658	0.02625	-0.24065	-0.24136	0.02902	0.02811
	400	-0.66891	-0.66990	0.02078	0.02116	-0.19509	-0.19645	0.02327	0.02298
	800	-0.59617	-0.59578	0.01455	0.01495	-0.15201	-0.15169	0.01636	0.01656
	1600	-0.52514	-0.52502	0.00998	0.01011	-0.11562	-0.11563	0.01129	0.01104
4	100	-0.95036	-0.95056	0.02062	0.01914	-0.39984	-0.40115	0.02155	0.02176
	200	-0.92756	-0.92822	0.01754	0.01771	-0.38106	-0.38184	0.01840	0.01907
	400	-0.90124	-0.90164	0.01434	0.01381	-0.36023	-0.36021	0.01497	0.01533
	800	-0.86748	-0.86784	0.01174	0.01174	-0.33398	-0.33440	0.01261	0.01236
	1600	-0.82785	-0.82800	0.00876	0.00918	-0.30430	-0.30464	0.00959	0.00963
5	100	-0.98838	-0.99000	0.01052	0.01432	-0.43053	-0.43184	0.01090	0.01239
	200	-0.98403	-0.98500	0.00860	0.00741	-0.42692	-0.42744	0.00899	0.00863
	400	-0.97722	-0.97750	0.00711	0.00741	-0.42119	-0.42162	0.00737	0.00775
	800	-0.96915	-0.96916	0.00593	0.00608	-0.41462	-0.41469	0.00606	0.00602
	1600	-0.95855	-0.95866	0.00475	0.00463	-0.40594	-0.40602	0.00498	0.00504

Table 3. Multivariate Monte Carlo Simulations

		Mean	Std Deviation	Median	Maximum	Minimum
Guangdong (n=600) 2001	Ln(Exp p.c.)	8.0137	0.6277	8.0338	10.7260	6.2687
	Ln(Area p.c.)	2.6685	0.4726	2.6968	4.7361	1.0986
	Family size	3.3233	0.7723	3.0000	7.0000	1.0000
	Expend p.c.	11431	7866.2	9638.5	105886	2329.0
	Sq Meters p.c.	52.5800	29.7991	48.000	342.00	9.0000
Shaanxi (n=500) 2001	Ln(Exp p.c.)	7.2324	0.6728	7.2092	9.7712	5.5977
	Ln(Area p.c.)	2.5336	0.4902	2.5328	4.1997	1.3863
	Family size	3.1020	0.7725	3.0000	6.0000	1.0000
	Expend p.c.	4918.4	3407.9	3942.0	30806.0	1094.0
	Sq Meters p.c.	42.0840	21.0892	38.000	200.00	12.000
Guangdong (n=595) 1987	Ln(Exp p.c.)	5.8395	0.6147	5.8452	8.7497	4.1015
	Ln(Area p.c.)	2.1868	0.5702	2.2246	4.1352	0.4055
	Family size	3.8958	1.0836	4.0000	8.0000	2.0000
	Expend p.c.	1447.1	898.79	1243.7	12617.0	411.60
	Sq Meters p.c.	38.8588	21.6403	36.000	156.00	6.0000
Shaanxi (n=546) 1987	Ln(Exp p.c.)	5.4156	0.5866	5.3898	7.3705	3.5231
	Ln(Area p.c.)	2.0575	0.4346	2.0369	3.8067	0.4055
	Family size	3.7216	1.0507	4.0000	8.0000	2.0000
	Expend p.c.	883.15	434.22	779.16	4524.3	241.00
	Sq Meters p.c.	30.4469	13.0470	28.000	100.00	6.0000

Table 4. Sample Statistics

		$\hat{\theta}_e$	sesc_e	$\hat{\theta}_e^{bc}$	se_e	$\hat{\theta}_h$	sesc_h	$\hat{\theta}_h^{bc}$	se_h	$\hat{\theta}_{eh}$	sesc_{eh}
1987	b_s	0.70201	0.02803	0.77217	0.03282	0.78201	0.02948	0.85663	0.03282	0.63403	0.027
2001	b_s	0.52999	0.02694	0.59666	0.03366	0.76591	0.02833	0.83333	0.03323	0.50644	0.026
1987	$b_s^{3/2}$	0.69455	0.02861	0.77096	0.03432	0.77748	0.02947	0.84607	0.03456	0.59050	0.027
2001	$b_s^{3/2}$	0.52244	0.02752	0.58314	0.03487	0.82130	0.02709	0.90226	0.03518	0.47045	0.025

Table 5. Parameter estimates along with standard errors

	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1.0$
1987	0.27170	0.348900	0.459800	0.625500
se	0.03935	0.10460	0.26326	0.54095
2001	0.24020	0.331100	0.469700	0.680100
se	0.03168	0.085039	0.228473	0.50276

Table 6. Duclos, Esteban and Ray polarization measure from pooled sample

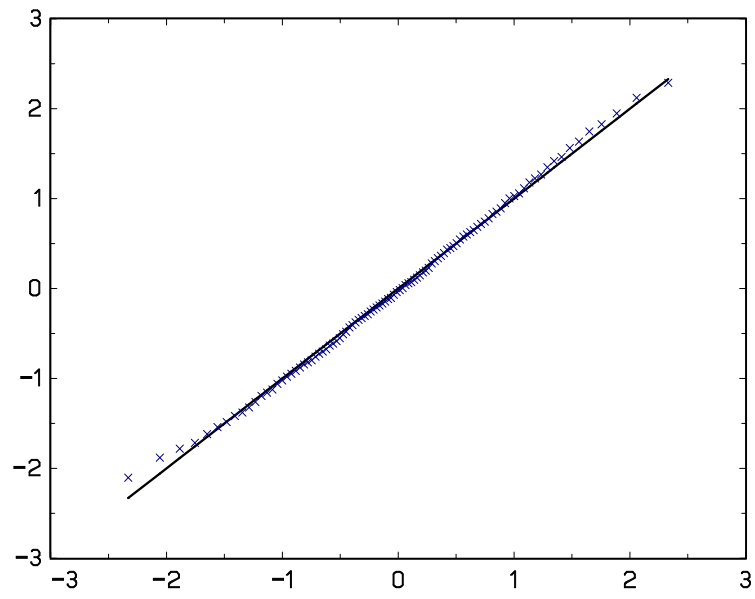


Figure 1. qq-plot of $\hat{\theta}$ versus normal distribution.



Figure 1:

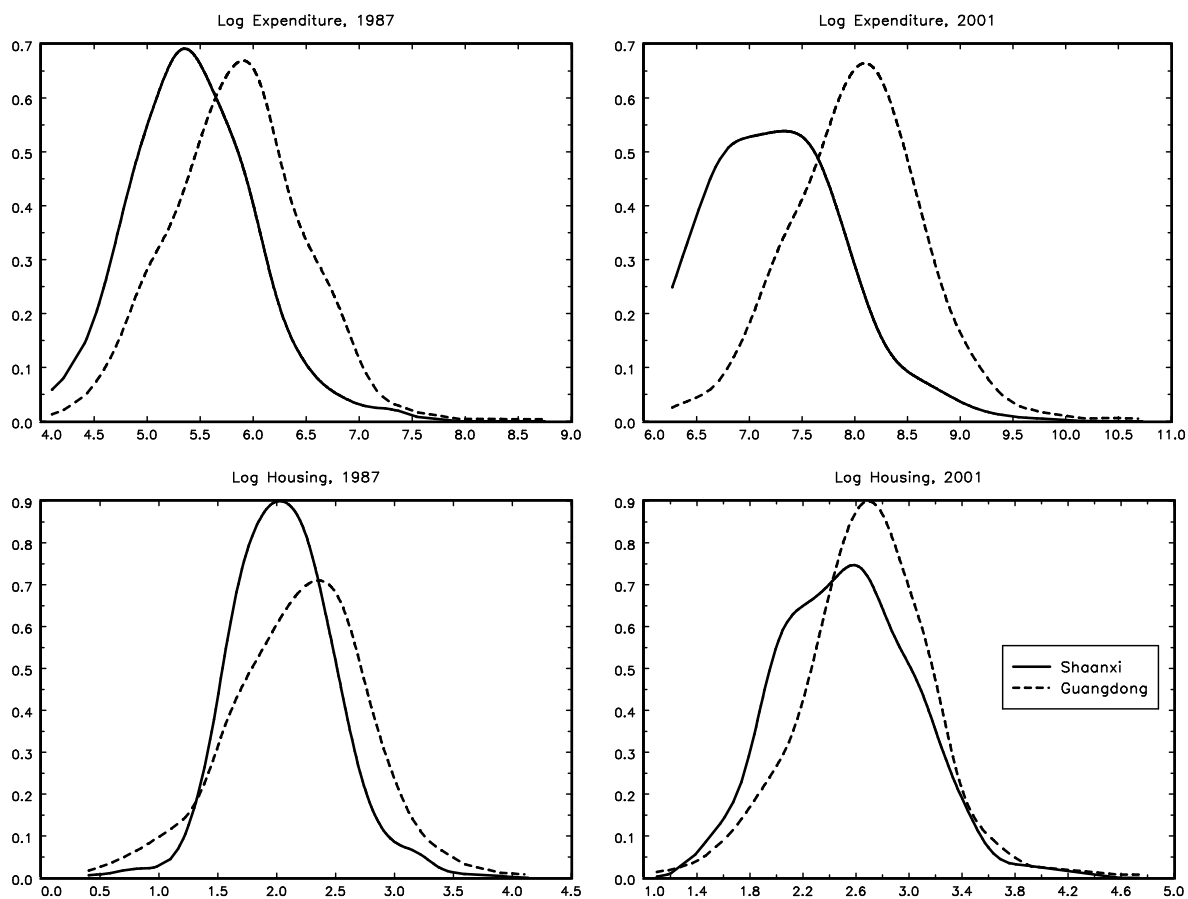


Figure 3. The kernel density estimates of expenditure and housing for the two provinces and the two years.

C Further Material

C.1 Some Intuition

We give below an alternative estimator/intuition about our estimator. Suppose that the common support is $[0, 1]$ and let $x^1 = b/2, x^2 = 3b/2, \dots, x^T = 1 - b/2$, where $T = 1/b$, and suppose the kernel is supported on $[-0.5, 0.5]$. Then let

$$\hat{\theta} = \frac{1}{T} \sum_{t=1}^T \min\{\hat{f}(x^t), \hat{g}(x^t)\}.$$

Note that although $\hat{f}(x^t)$ and $\hat{f}(x^s)$ are not strictly independent for $s \neq t$, they are approximately so. Specifically, $\text{cov}(\hat{f}(x^t), \hat{f}(x^s)) = O(n^{-2})$, since

$$\begin{aligned} E[\hat{f}(x^t)\hat{f}(x^s)] &= \frac{1}{n^2} \sum_{i=1}^n E[K_b(x^t - X_i) K_b(x^s - X_i)] + \frac{n(n-1)}{n^2} E[K_b(x^t - X)] E[K_b(x^s - X)] \\ &= \left(1 - \frac{1}{n^2}\right) E[\hat{f}(x^t)] E[\hat{f}(x^s)]. \end{aligned}$$

We shall suppose for heuristic reasons that $\min\{\hat{f}(x^t), \hat{g}(x^t)\}$ is strictly independent of $\min\{\hat{f}(x^s), \hat{g}(x^s)\}$, although technically we should still apply the Poissonization technique used in this paper to proceed. Suppose we can dispose of the smoothing bias terms as before and suppose for simplicity that $C_{f,g} = [0, 1]$, then

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta) &= \sqrt{n} \frac{1}{T} \sum_{t=1}^T \min\{\hat{f}(x^t) - E\hat{f}(x^t), \hat{g}(x^t) - E\hat{g}(x^t)\} + o_p(1) \\ &= \sum_{t=1}^T \xi_{nt} + o_p(1), \end{aligned}$$

where $\xi_{nt} = T^{-1}n^{1/2} \min\{\hat{f}(x^t) - E\hat{f}(x^t), \hat{g}(x^t) - E\hat{g}(x^t)\}$. We can apply a triangular array CLT here after subtracting off the mean of ξ_{nt} , but to make things simple let us make the further step of the normal approximation. Then

$$\begin{aligned} \sqrt{n}(\hat{\theta} - \theta) &= \frac{1}{\sqrt{nb}} \sqrt{n} \frac{1}{T} \sum_{t=1}^T \min\{f(x^t)^{1/2} \|K\|_2 Z_{ft}, g(x^t)^{1/2} \|K\|_2 Z_{gt}\} + o_p(1), \\ &= \|K\|_2 \sqrt{b} \sum_{t=1}^{1/b} f(x^t)^{1/2} \min\{Z_{ft}, Z_{gt}\} + o_p(1), \end{aligned}$$

where Z_{ft}, Z_{gt} are standard normal random variables mutually independent when X, Y are but otherwise correlated. We then have the following result

$$\begin{aligned} & \sqrt{n}(\hat{\theta} - \theta) - \|K\|_2 b^{-1/2} \int f(x)^{1/2} dx E \min\{Z_{ft}, Z_{gt}\} \\ &= \|K\|_2 \sqrt{b} \sum_{t=1}^{1/b} f(x^t)^{1/2} [\min\{Z_{ft}, Z_{gt}\} - E \min\{Z_{ft}, Z_{gt}\}] + o_p(1), \end{aligned}$$

which is asymptotically normal with mean zero and variance

$$\|K\|_2^2 \int f(x) dx \text{var} [\min\{Z_{ft}, Z_{gt}\}].$$

This should be compared with our σ_0^2 . One can show that $E[\min\{Z_1, Z_2\}^2] = 1$, so that

$$\text{var} [\min\{Z_{ft}, Z_{gt}\}] = 1 - 0.56^2 = 0.6874,$$

which is bigger than $\sigma_0^2 = 0.61$ for the uniform kernel. This says that our more complicated estimator that averages also over non-independent points delivers about 10% improvement in variance in this case. Note that the bias is the same for both estimators.

You can apply the same type of method in other semiparametric problems and our sense is that the optimal kernel is the uniform if you only take account of the first order effect on variance. Example is suppose

$$Y = \mu + \varepsilon_i,$$

then the sample mean is BLUE. We can compute the kernel estimator against some covariates X and average, thus

$$\hat{\theta} = \frac{1}{T} \sum_{t=1}^T \hat{m}(x^t),$$

where $\hat{m}(x^t)$ is a kernel estimator. Then when ε_i is independent of X_i and X_i is uniform on $[0, 1]$, the asymptotic variance of $\hat{\theta}$ is $\sigma^2 \|K\|_2^2 / n$, which is minimized by taking K to be uniform on $[-0.5, 0.5]$. In general the asymptotic variance of this method is worse than averaging over the sample points (which is equivalent to the sample mean when K is uniform), but always asymptotic variance is minimized by taking uniform kernel. In our case, even averaging over the sample points gives this nasty looking asymptotic variance but it still makes sense that the variance minimizing kernel is the uniform.

Finally, we compare the asymptotics of the nonparametric estimator with those of a natural parametric alternative. The parametric problem generally has different asymptotics. In particular, suppose that f, g are parametrically specified so that $f_\phi, \phi \in \Phi$, and $g_\lambda, \lambda \in \Lambda$. Let $\hat{\phi}$ and $\hat{\lambda}$ denote root- n consistent estimators of ϕ and λ , such that $[\sqrt{n}(\hat{\phi} - \phi), \sqrt{n}(\hat{\lambda} - \lambda)]^\top \implies U \sim N(0, \Omega)$, and let $\hat{\theta} = \int_C \min\{f_{\hat{\phi}}(x), g_{\hat{\lambda}}(x)\} dx$. Then we have

$$\sqrt{n}(\hat{\theta} - \theta) = -\frac{1}{2} \int_C \left| (\dot{s}_\phi(x) f_\phi(x), -\dot{s}_\lambda(x) g_\lambda(x)) U \right| dx + o_p(1),$$

where $\dot{s}_\phi(x) = \partial \log f_\phi(x) / \partial \phi$ and $\dot{s}_\lambda(x) = \partial \log g_\lambda(x) / \partial \lambda$ are the score functions. When the contact set has positive measure, the asymptotic distribution is non-normal with a negative mean.

C.2 Computation of kernel constants

This note is regarding computation of σ_0^2 in cases where sample sizes are not equal. We have to calculate

$$R(\rho) = \text{cov} \left(\min \{Z_1, Z_2\}, \min \left\{ \rho Z_1 + \sqrt{1 - \rho^2} Z_3, \rho Z_2 + \sqrt{1 - \rho^2} Z_4 \right\} \right)$$

for given ρ , where $Z_i \sim N(0, v_i)$ with $v_1 = v_3 = 1$ and $v_2 = v_4 = \omega$.

We have

$$\begin{aligned} 2 \min \{Z_1, Z_2\} &= U - |V| \\ 2 \min \left\{ \rho Z_1 + \sqrt{1 - \rho^2} Z_3, \rho Z_2 + \sqrt{1 - \rho^2} Z_4 \right\} &= W - |Y| \end{aligned}$$

where

$$\begin{aligned} U &= Z_1 + Z_2 \\ V &= Z_1 - Z_2 \\ W &= \rho Z_1 + \sqrt{1 - \rho^2} Z_3 + \rho Z_2 + \sqrt{1 - \rho^2} Z_4 \\ Y &= \rho Z_1 + \sqrt{1 - \rho^2} Z_3 - \rho Z_2 - \sqrt{1 - \rho^2} Z_4, \end{aligned}$$

and

$$\begin{aligned} \begin{pmatrix} U \\ V \\ W \\ Y \end{pmatrix} &= \sqrt{1 + \omega} \begin{pmatrix} U' \\ V' \\ W' \\ Y' \end{pmatrix} \\ \begin{pmatrix} U' \\ V' \\ W' \\ Y' \end{pmatrix} &\sim N \left(0, \begin{bmatrix} 1 & 0 & \rho & 0 \\ 0 & 1 & 0 & \rho \\ \rho & 0 & 1 & 0 \\ 0 & \rho & 0 & 1 \end{bmatrix} \right). \end{aligned}$$

Therefore,

$$E \min \{Z_1, Z_2\} = -\frac{1}{2} E|V| = -\frac{\sqrt{1 + \omega}}{2} E|V'| = -\frac{\sqrt{1 + \omega}}{2} \sqrt{\frac{2}{\pi}} = -\sqrt{\frac{1 + \omega}{2\pi}}.$$

We have

$$\begin{aligned}
R(\rho) &= \frac{1}{4} \text{cov}(U - |V|, W - |Y|) \\
&= \frac{1+\omega}{4} \text{cov}(U' - |V'|, W' - |Y'|) \\
&= \frac{1+\omega}{4} \text{cov}(U', W') - \frac{1+\omega}{4} \text{cov}(|V'|, W') - \frac{1+\omega}{4} \text{cov}(U', |Y'|) + \frac{1+\omega}{4} \text{cov}(|V'|, |Y'|) \\
&= \frac{1+\omega}{4} [\rho + \text{cov}(|V'|, |Y'|)],
\end{aligned}$$

because for zero mean normals $\text{cov}(|V'|, W') = \text{cov}(U', |Y'|) = 0$. Write $|V'| = V'1(V' > 0) - V'1(V' \leq 0)$ and $|Y'| = Y'1(Y' > 0) - Y'1(Y' \leq 0)$, so that

$$\begin{aligned}
\text{cov}(|V'|, |Y'|) &= E[|V'||Y'|] - E[|V'|]E[|Y'|] \\
&= E[V'Y'1(V' > 0)1(Y' > 0)] + E[V'Y'1(V' \leq 0)1(Y' \leq 0)] \\
&\quad - E[V'Y'1(V' > 0)1(Y' \leq 0)] - E[V'Y'1(V' \leq 0)1(Y' > 0)] - \frac{2}{\pi} \\
&= 2E[V'Y'1(V' > 0)1(Y' > 0)] - 2E[V'Y'1(V' \leq 0)1(Y' > 0)] - \frac{2}{\pi}
\end{aligned}$$

by symmetry. From Rosenbaum (1961, JRSS B) we have

$$\begin{aligned}
E[V'Y'1(V' > 0)1(Y' > 0)] &= F(\rho)\rho + \frac{1}{\sqrt{2\pi}} \sqrt{\frac{1-\rho^2}{2\pi}} \\
E[V'Y'1(V' \leq 0)1(Y' > 0)] &= E[V'Y'1(V' > -\infty)1(Y' > 0)] - E[V'Y'1(V' > 0)1(Y' > 0)] \\
&= \frac{1}{2}\rho - F(\rho)\rho - \frac{1}{\sqrt{2\pi}} \sqrt{\frac{1-\rho^2}{2\pi}} \\
&= \rho \left(\frac{1}{2} - F(\rho) \right) - \frac{1}{\sqrt{2\pi}} \sqrt{\frac{1-\rho^2}{2\pi}},
\end{aligned}$$

where $F(\rho) = \Pr[V' > 0, Y' > 0]$. Therefore,

$$\text{cov}(|V'|, |Y'|) = 2\rho \left(2F(\rho) - \frac{1}{2} \right) + \frac{4}{\sqrt{2\pi}} \sqrt{\frac{1-\rho^2}{2\pi}} - \frac{2}{\pi}.$$

In the special case that $\rho = 1$ we have $F(\rho) = 1/2$ so that

$$\text{cov}(|V'|, |Y'|) = E(|V'|^2) - \frac{2}{\pi} = 1 - \frac{2}{\pi}.$$

In the special case that $\rho = 0$,

$$\text{cov}(|V'|, |Y'|) = \frac{4}{\sqrt{2\pi}} \sqrt{\frac{1}{2\pi}} - \frac{2}{\pi} = 0$$

as expected.

In conclusion,

$$\begin{aligned}
R(\rho) &= \frac{1+\omega}{4} [\rho + \text{cov}(|V'|, |Y'|)] \\
&= \frac{1+\omega}{4} \left[4\rho F(\rho) + \frac{4}{\sqrt{2\pi}} \sqrt{\frac{1-\rho^2}{2\pi}} - \frac{2}{\pi} \right] \\
&= (1+\omega) \left[\rho F(\rho) + \frac{\sqrt{1-\rho^2}-1}{2\pi} \right].
\end{aligned}$$

The main thing is that this shows that

$$R(\rho; \omega) = \frac{1+\omega}{2} R(\rho; 1)$$

so that we can calculate this quantity once and for all and apply it in situations where sample sizes m and n are different.

C.3 Drifting DGP

Suppose that the two densities are not identical but close to each other. Specifically, suppose that

$$f(x) = g(x) + \frac{\delta(x)}{\sqrt{n}},$$

Then

$$\theta_n = 1 + \frac{1}{\sqrt{n}} \int \delta_f(x) dx,$$

where $\delta_f(x) = 1(x \in C_f)\delta(x)$ and $\delta_g(x) = 1(x \in C_g)\delta(x)$. We have

$$\begin{aligned}
\sqrt{n}(\hat{\theta} - \theta_n) &= \int \sqrt{n} [\min\{f_n(x), g_n(x)\} - \min\{f(x), g(x)\}] dx \\
&= \int_{C_{f,g}} \sqrt{n} \min\{f_n(x) - f(x), g_n(x) - g(x)\} dx \\
&\quad + \int_{C_f} \sqrt{n} \min\{f_n(x) - f(x), g_n(x) - g(x) - \frac{\delta(x)}{\sqrt{n}}\} dx \\
&\quad + \int_{C_g} \sqrt{n} \min\{f_n(x) - f(x) + \frac{\delta(x)}{\sqrt{n}}, g_n(x) - g(x)\} dx \\
&= \int \min\{\sqrt{n}[f_n(x) - f(x)] + \delta_g(x), \sqrt{n}[g_n(x) - g(x)] - \delta_f(x)\} dx.
\end{aligned}$$

Using an argument similar to the proof of Theorem 1 to deal with the bias of kernel estimators, we have

$$\begin{aligned}
\sqrt{n}(\hat{\theta} - \theta) &= \int \min\{\sqrt{n}[f_n(x) - Ef_n(x)] + \delta_g(x), \sqrt{n}[g_n(x) - Eg_n(x)] - \delta_f(x)\} dx + o(1) \\
&=: A_n + o(1), \text{ say.}
\end{aligned}$$

Define a Poissonized version of A_n (minus its expectation restricted to a Borel set $B \subset \mathbb{R}^d$) to be

$$\begin{aligned} A_n^P(B) &= \int_B \min \left\{ \sqrt{n} [f_N(x) - Ef_n(x)] + \delta_g(x), \sqrt{n} [g_N(x) - Eg_n(x)] - \delta_f(x) \right\} dx \\ &\quad - \int_B E \min \left\{ \sqrt{n} [f_N(x) - Ef_n(x)] + \delta_g(x), \sqrt{n} [g_N(x) - Eg_n(x)] - \delta_f(x) \right\} dx. \end{aligned}$$

Let

$$\begin{aligned} T_{f,N}(x) &= \frac{\sqrt{n} \{f_N(x) - Ef_n(x)\}}{\sqrt{k_{f,n}(x)}}, \\ T_{g,N}(x) &= \frac{\sqrt{n} \{g_N(x) - Eg_n(x)\}}{\sqrt{k_{g,n}(x)}}. \end{aligned}$$

Note that

$$\sup_{x \in B} \left| \sqrt{k_{f,n}(x)} - b^{-d/2} \|K\|_2 \sqrt{f(x)} \right| = O(b^{d/2}) \quad (97)$$

$$\sup_{x \in B} \left| \sqrt{k_{g,n}(x)} - b^{-d/2} \|K\|_2 \sqrt{g(x)} \right| = O(b^{d/2}), \quad (98)$$

so that

$$\begin{aligned} &\sup_{x \in B} \left| \sqrt{k_{f,n}(x)} - \sqrt{k_{g,n}(x)} \right| \\ &= \sup_{x \in B} \left| \sqrt{f(x)} - \sqrt{g(x)} \right| + O(b^{d/2}) \\ &= O(n^{-1/2}) + O(b^{d/2}) = o(1). \end{aligned} \quad (99)$$

Therefore, the bias term is given by

$$\begin{aligned} &\int_B E \min \left\{ T_{f,N}(x) \sqrt{k_{f,n}(x)} + \delta_g(x), T_{g,N}(x) \sqrt{k_{g,n}(x)} - \delta_f(x) \right\} dx \\ &= \int_B E \min \left\{ T_{f,N}(x) + \frac{\delta_g(x)}{\sqrt{k_{f,n}(x)}}, T_{g,N}(x) - \frac{\delta_f(x)}{\sqrt{k_{f,n}(x)}} \right\} \sqrt{k_{f,n}(x)} dx + o(1) \\ &= \int_B E \min \left\{ Z_1 + \frac{\delta_g(x)}{\sqrt{k_{f,n}(x)}}, Z_2 - \frac{\delta_f(x)}{\sqrt{k_{f,n}(x)}} \right\} \sqrt{k_{f,n}(x)} dx + o(1) \\ &= \int_B E \min \left\{ Z_1 + \frac{\delta_g(x) + \delta_f(x)}{\sqrt{k_{f,n}(x)}}, Z_2 \right\} \sqrt{k_{f,n}(x)} dx - \int_B \delta_f(x) dx + o(1), \end{aligned} \quad (100)$$

where the first equality holds by (99), the second equality holds by Lemma A1 and the last equality holds by rearranging terms. For each constant $\lambda \rightarrow 0$, some calculations yield

$$\begin{aligned} &E \min \{Z_1 + \lambda, Z_2\} \\ &= \lambda \left\{ 1 - \Phi \left(\frac{\lambda}{\sqrt{2}} \right) \right\} - E\phi(Z - \lambda) - E\phi(Z + \lambda) \\ &= E \min \{Z_1, Z_2\} + \frac{\lambda}{2} + O(\lambda^2). \end{aligned}$$

Therefore, using (97) and (98), we have: uniformly over $x \in B$,

$$\begin{aligned} & E \min \left\{ Z_1 + \frac{\delta_g(x) + \delta_f(x)}{\sqrt{k_{f,n}(x)}}, Z_2 \right\} \sqrt{k_{f,n}(x)} \\ &= E \min \{Z_1, Z_2\} \sqrt{k_{f,n}(x)} + \frac{\delta_g(x) + \delta_f(x)}{2} + O(b^{d/2}). \end{aligned}$$

Therefore, the bias term on the right hand side of (100) is given by

$$\begin{aligned} & b^{-d/2} \|K\|_2 \int_B f^{1/2}(x) dx \cdot E \min \{Z_1, Z_2\} + \frac{1}{2} \int_B [\delta_g(x) + \delta_f(x)] dx - \int_B \delta_f(x) dx + o(1) \\ &= a_n(B) + \frac{1}{2} \int_B [\delta_g(x) - \delta_f(x)] dx + o(1). \end{aligned}$$

By letting B increase to $\mathbb{R}^d$, we have

$$a_n(B) + \frac{1}{2} \int_B [\delta_g(x) - \delta_f(x)] dx \rightarrow a_n + \frac{1}{2} \int_{\mathbb{R}^d} [\delta_g(x) - \delta_f(x)] dx.$$

On the other hand, using an argument similar to the proof of Lemma A3, we can show that

$$\lim_{n \rightarrow \infty} \text{var} (A_n^P(B)) = p_{0,B} \sigma_0^2 + \bar{\sigma}_{1,B}^2$$

as before, which in turn converges to v as B increases to $\mathbb{R}^d$. Therefore, combining the above results, we have

$$\begin{aligned} \sqrt{n} (\hat{\theta} - \theta) - a_n &= [A_n - EA_n] + [EA_n - a_n] + o(1) \\ &\Rightarrow N \left(\frac{1}{2} \int_{\mathbb{R}^d} [\delta_g(x) - \delta_f(x)] dx, v \right), \end{aligned}$$

so that there is a location shift in the distribution. This implies that

$$\begin{aligned} & \sqrt{n} (\hat{\theta} - 1) - a_n \\ &= \sqrt{n} (\hat{\theta} - \theta) - a_n + \int \delta_f(x) dx \\ &\Rightarrow N(0, v). \end{aligned}$$