Nonparametric Estimation of a Polarization Measure[*]

Gordon Anderson[†]
University of Toronto

Oliver Linton[‡]
The London School of Economics

Yoon-Jae Whang[§]
Seoul National University

The Suntory Centre
Suntory and Toyota International Centres for
Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel: 020 7955 6679

Discussion paper
No. EM/2009/534
June 2009

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

[‡] Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: o.linton@lse.ac.uk. Thanks to the ESRC and ERC for financial support. This paper was partly written while I was a Universidad Carlos III de Madrid-Banco Santander Chair of Excellence, and I thank them for financial support. I thank seminar participants at CEMFI for helpful comments.

[§] Department of Economics, Seoul National University, Seoul 151-742, Korea. Email: whang@snu.ac.kr.

Abstract

This paper develops methodology for nonparametric estimation of a polarization measure due to Anderson (2004) and Anderson, Ge, and Leo (2006) based on kernel estimation techniques. We give the asymptotic distribution theory of our estimator, which in some cases 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.

Key words: Kernel Estimation; Inequality; Overlap coefficient; Poissonization

JEL Classification Number: C12, C13, C14.

1   Introduction

Polarization is the process whereby a social or political group is divided into two opposing sub-groups with fewer and fewer members of the group remaining neutral or holding an intermediate position. It is the subject of some interest in economics as it is both cause and consequence of much economic behavior. There have been several proposed univariate polarization indices which focus on an arbitrary number of groups (Esteban and Ray (1994), Esteban, Gradin and Ray (1998), Zhang and Kanbur (2001) and Duclos, Esteban and Ray (2004)) and a similar number of bi-polarization measures that focus on just two groups (Alesina and Spolaore (1997), Foster and Wolfson (1992), Wolfson (1994) and Wang and Tsui (2000)).[1] All consider polarization within a distribution to be a tendency toward multiple separating modes, in essence addressing the potential for an observed distribution to be a mixture of several (or just two) unobserved sub-distributions. Gigliarano and Mosler (2008) develop a family of multivariate bi-polarization measures based upon estimates of between and within group multivariate variation and relative group size which require observations on the two sub-distributions. It is worth noting that, within applications in economics, the measure is readily adapted to general tests of distributional differences in examining issues of convergence, independence and mobility for example. The measure presented here falls into this category (though it can be readily extended to many groups) in that it reflects the polarization of the distributions of two or more identifiable groups.

Duclos, Esteban and Ray (2004) evaluate polarization measures on the basis of the extent to which they satisfy axioms formed around a notional univariate density $f(x)$ that is a mixture of kernels $f(x, a_i)$ that are symmetric uni-modal on a compact support of $[a_i, a_{i+2}]$ with $E(x) = \mu_i = a_{i+1}$. The kernels are subject to slides (location shifts) that preserve the shape of the kernels and squeezes that are location preserving shrinkages of the kernel to their respective locations. Potential indices are evaluated in the context of such changes in terms of whether they satisfy axioms that require symmetric squeezes and outward slides to increase the polarization measure. To the extent that the kernels remain overlapped this will be the case for the index proposed here. It is also required that common population scaling of the kernels will preserve the ordering which will also be the case here. They argue that valid measures lie in the class $P_\alpha(f) = \int \int f(x)^{1+\alpha} f(y) |y - x| dy dx$, where $\alpha \in [0.25, 1]$. They propose an estimator of $P_\alpha(f)$ and is establish its asymptotic properties.

We will focus on measurement of polarization between two well-defined groups. One technique for assessing polarization between two groups is to evaluate how much they have in common according to some objective outcome variable, such a measure corresponds to non-alienation and its negative corresponds to a degree of alienation. Anderson (2004) and Anderson, Ge, and Leo (2006) proposes an overlap measure as an index of convergence and a function of its negative as a measure of alienation.

---

[1] An excellent summary of the properties of the univariate indices is to be found in Esteban and Ray (2007).

The extent to which two densities f, g overlap is given by

$$\theta = \int_{-\infty}^{\infty} \min\{f(x), g(x)\}dx. \tag{1}$$

It is a number between zero and one with zero corresponding to no overlap and one to the perfect matching of the two distributions. It follows that $1 - \theta$ is a measure of the extent to which the distributions do not match or are alienated. 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. The definition can easily be extended to the multivariate x case, and indeed we treat this case below. This quantity has received a lot of attention in medical statistics, Mizuno, Yamaguchi, Fukushima, Matsuyama, and Ohashi (2005), and Ecology, where it is known as the overlap coefficient or the coefficient of community, see for example Ricklefs and Lau (1980). To provide a sense of the magnitude of θ we looked at the male-female height distributions from the National Health and Nutrition Survey of 1999 for the US age group 20-29. For these data, θ is approximately 0.40.

Previous work, Anderson, Ge and Leo (2006), has shown how to estimate θ and conduct inference about it 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. Also, there is no consensus on appropriate parametric models for income distributions for example, 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. The properties of the estimated θ can be nonstandard depending on the contact set $\{x : f(x) = g(x) > 0\}$. 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 Giné et al. (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, 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.

2   Estimation

We assume that there are population random variables X, Y, where X has density f and Y has density g. Note 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. Note also that $\theta = 1 - \frac{1}{2}\int_{-\infty}^{\infty} |f(x) - g(x)| \, dx$, which shows that $1 - \theta$ defines a pseudometric on the space of densities. An alternative representation of θ is as an expectation

$$\theta = E\left[\min\left\{1, \ell_{g,f}(X)\right\}\right] = E\left[\min\left\{1, \ell_{f,g}(Y)\right\}\right], \tag{2}$$

where $\ell_{g,f}(x) = g(x)/f(x)$ is the likelihood ratio, which can be convenient for computing estimators, see below.

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. In this case 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 $\{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ of size n on the population. In some cases one might have different sizes n, m for the two samples, but we shall leave this discussion till later. We propose to estimate θ by

$$\widehat{\theta} = \int_C \min\{f_n(x), g_n(x)\}dx, \tag{3}$$

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

where $C \subseteq \mathbb{R}^d$ is the union of the supports or some subset of interest, while K is a multivariate kernel and $K_b(.) = K(./b)/b^d$ and b is a bandwidth sequence. For simplicity we suppose that the same bandwidth is 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. In practice one has to compute a multivariate integral in (3) and a simple approach is to just replace $\widehat{\theta}$ 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 on (2).[2]

---

[2]An alternative estimator here, based on the transformation idea, is

$$\widehat{\theta} = \int_0^1 \min\{1, g_n^X(u)\}du,$$

3 Asymptotic Properties

We next discuss the asymptotic behavior of $\widehat{\theta}$ as $n \to \infty$. 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.[3] Schmid and Schmidt (2006) have recently established consistency of $\widehat{\theta}$ in the univariate compactly supported special case.

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

$$C_{f,g} = \left\{ x \in \mathbb{R}^d : f(x) = g(x) > 0 \right\}, \tag{4}$$

$$C_f = \{ x : f(x) < g(x) \}, \tag{5}$$
$$C_g = \{ x : f(x) > g(x) \}. \tag{6}$$

Let $\lambda = (\lambda_1, \ldots, \lambda_d)^\top$ denote a vector of nonnegative integer constants. For such vector, we define $|\lambda| = \sum_{i=1}^d \lambda_i$ and, for any function $h(x) : \mathbb{R}^d \to \mathbb{R}$, $D^\lambda h(x) = \partial^{|\lambda|} / (\partial x_1^{\lambda_1} \cdots \partial x_d^{\lambda_d})(h(x))$, where $x = (x_1, \ldots, x_d)^\top$ 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) (i) 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, where s is an integer that satisfies $s > d$. (ii) The kernel satisfies $\rho(t) = 1 - c \, ||t||^\alpha + o(||t||^\alpha)$ as $||t|| \to 0$ for some positive constants c and α.

(A2) The densities f and g are strictly positive on $\mathbb{R}^d$, bounded and absolutely continuous with respect to Lebesgue measure and s-times continuously differentiable with uniformly bounded derivatives.

(A3) The bandwidth satisfies: (i) $nb^{2s} \to 0$, (ii) $nb^{2d} \to \infty$, (iii) $nb^d / (\log n) \to \infty$, (iv) $nb^{d(1+\gamma)/\gamma} / (\log n)^{(1+2\gamma)/\gamma} \to \infty$, where γ is a positive constant that satisfies A5 when $\mu(C_{f,g}) > 0$ and $\gamma = \infty$ otherwise.

(A4) $\{X_i : i \geq 1\}$ and $\{Y_i : i \geq 1\}$ are i.i.d. and independent from each other with support $\mathbb{R}^d$.

where $g_n^X(x)$ is the density estimate based on transformed data $F_n^X(Y_i)$, $i = 1, \ldots, n$, where $F_n^X(\cdot) = n^{-1} \sum_{i=1}^n 1(X_i \leq \cdot)$ is the empirical process of X. It turns out this has identical asymptotic distribution to our estimator. Schmid and Schmidt (2006) find not much difference between the estimators in simulation experiments.

[3] This implicitly rules out the case $\theta = 0$.

(A5) (i) 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 $\varepsilon \to 0$ for some positive constant γ. (ii) $\int_{C_{f,g}} f^{1/2}(x)dx < \infty$.

Remarks. By the triangle inequality we have

$$2\left|\widehat{\theta} - \theta\right| \leq \int |f_n(x) - f(x)|\, dx + \int |g_n(x) - g(x)|\, dx, \tag{7}$$

so that under weaker conditions than A3, specifically just $b \to 0$ and $nb^d \to \infty$, we have $\widehat{\theta} - \theta = O_p(b^s) + O_p(n^{-1/2}b^{-d/2})$. The assumption A3 is needed for the asymptotic normality. To fulfil A3 we require that $s > d$ and $s > d(1 + \gamma)/2\gamma$; in the univariate case with $\gamma > 1$ 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. If $s \leq d$, the smoothing bias term dominates and prevents the distribution theory below applying. In A4 we assume that the variables are mutually independent and the sample is i.i.d. This can be relaxed under further conditions. 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 $h = f - g$ 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 h is bounded away from zero outside of $C_{f,g}$, then γ can be set to be ∞. Assumption A5 is used to derive the asymptotic distribution of $\sqrt{n}(\widehat{\theta} - \theta)$ and to get a consistent estimator of the centering term a_n in Theorem 2 below. The requirement in A5(ii) that $\int_{C_{f,g}} f^{1/2}(x)dx = E\left[f(X)^{-1/2}1(f(X) = g(X))\right] < \infty$ 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 A5(ii) is implied by the condition that $E\left[||X||^{1+\eta}1(f(X) = g(X))\right] < \infty$ for some $\eta > 0$.

Define:

$$
\begin{aligned}
p_0 &= \Pr(X \in C_{f,g}) = E\left[1\left(f(X) = g(X)\right)\right] = \Pr(Y \in C_{f,g}) = E\left[1\left(f(Y) = g(Y)\right)\right] \\
p_f &= \Pr(X \in C_f) = E\left[1\left(f(X) < g(X)\right)\right], \ p_g = \Pr(Y \in C_g) = E\left[1\left(f(Y) > g(Y)\right)\right] \\
\sigma_1^2 &= p_f(1 - p_f) + p_g(1 - p_g), \quad v = p_0\sigma_0^2 + \sigma_1^2.
\end{aligned}
$$

THEOREM 1. *Suppose that Assumptions A1-A5 hold. Then, we have:*

$$\sqrt{n}\left(\widehat{\theta} - \theta\right) - a_n \Longrightarrow N(0, v),$$

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

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

Remarks.

1. The statistic $\widehat{\theta}$ is consistent provided only $b \to 0$ and $nb^d \to \infty$ as follows from the above discussion. Our result shows under stronger conditions that the statistic is asymptotically normal after subtracting a bias term that is of order $n^{-1/2}b^{-d/2}$. The bias term 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. We have $v \leq ||K||_2^2 \mu(T_0) + 1/2$ in general, and in the scalar case with the uniform kernel we have $v \leq 5/2$. 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:

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

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

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

$$\sqrt{n}(\widehat{\theta} - \theta) = \sqrt{n}\int_{C_f}\{f_n(x) - Ef_n(x)\}dx + \sqrt{n}\int_{C_g}\{g_n(x) - Eg_n(x)\}dx$$
$$+\sqrt{n}\int_{C_{f,g}}\min\{f_n(x) - Ef_n(x), g_n(x) - Eg_n(x)\}dx + R_n,$$

where $R_n = O_p(\sqrt{n}b^s) = o_p(1)$. 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

$\widehat{\beta} = \min\{\overline{X}, \overline{Y}\}$ satisfies $\sqrt{n}(\widehat{\beta} - \beta) = \min\{\sqrt{n}(\overline{X} - \mu_X), \sqrt{n}(\overline{Y} - \mu_Y)\} \implies \min\{Z_X, Z_Y\}$, where $[\sqrt{n}(\overline{X} - \mu_X), \sqrt{n}(\overline{Y} - \mu_Y)] \implies [Z_X, Z_Y] = Z$ and Z is bivariate normal with zero mean. In this case, the limiting distribution of $\widehat{\beta}$ has a negative mean and is non-normal. In the case of $\widehat{\theta}$ 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 $\nu_n(x) = (f_n(x) - Ef_n(x), g_n(x) - Eg_n(x))$, $x \in C_{f,g}$ 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 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 give a further discussion about this. 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. This technique has been used in a number of places including combinatorical mathematics and analysis of algorithms.

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, 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

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

(A7) The tuning parameter satisfies $c_n \to 0$, $nb^{2d}c_n \to \infty$ and $nb^d c_n^{2+2\gamma} \to 0$, where γ is a positive constant that satisfies A5 when $\mu(C_{f,g}) > 0$ and $\gamma = \infty$ otherwise.

The condition A6 is needed in the case where $C_{f,g} = \mathbb{R}^d$ as it is used to bound the estimation error of $\int_{C_{f,g}} f^{1/2}(x)dx = E[f^{-1/2}(X)1\left(f(X) = g(X)\right)]$, which can be badly affected by heavy tails. It imposes a further restriction on the bandwidth, and so for small values of p one needs a lot of smoothness in f, g to compensate. If X, Y are Gaussian 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.

Define the bias corrected estimator and asymptotic variance estimator:

$$\widehat{\theta}^{bc} = \widehat{\theta} - \hat{a}_n/n^{1/2}, \quad \widehat{v} = \hat{p}_0\sigma_0^2 + \hat{\sigma}_1^2$$

$$\hat{a}_n = -0.56\frac{||K||_2}{2b^{d/2}}\left(\int_{\hat{C}_{f,g}} f_n^{1/2}(x)dx + \int_{\hat{C}_{f,g}} g_n^{1/2}(x)dx\right)$$

$$\hat{C}_f = \left\{x \in \mathbb{R}^d : f_n(x) - g_n(x) < -c_n, f_n(x) > 0, g_n(x) > 0\right\}$$

$$\hat{C}_g = \left\{x \in \mathbb{R}^d : f_n(x) - g_n(x) > c_n, f_n(x) > 0, g_n(x) > 0\right\}$$

$$\hat{C}_{f,g} = \left\{x \in \mathbb{R}^d : |f_n(x) - g_n(x)| \leq c_n, f_n(x) > 0, g_n(x) > 0\right\}$$

$$\hat{p}_0 = \frac{1}{2}\left(\int_{\hat{C}_{f,g}} f_n(x)dx + \int_{\hat{C}_{f,g}} g_n(x)dx\right)$$

$$\hat{p}_f = \int_{\hat{C}_f} f_n(x)dx, \ \hat{p}_g = \int_{\hat{C}_g} g_n(x)dx$$

$$\hat{\sigma}_1^2 = \hat{p}_f(1 - \hat{p}_f) + \hat{p}_g(1 - \hat{p}_g).$$

Then, we have the following result:

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

$$\sqrt{n}\left(\widehat{\theta}^{bc} - \theta\right) \Rightarrow N(0, v) \tag{8}$$

$$\widehat{v} \xrightarrow{p} v. \tag{9}$$

Remarks.

1. Note that the bias corrected estimator falls outside of $[0, 1]$ with positive probability, which motivates the construction of a winsorized version $\widehat{\theta}_w^{bc} = \widehat{\theta}^{bc}1(\widehat{\theta}^{bc} \in [0, 1]) + 1(\widehat{\theta}^{bc} > 1)$. When $\theta \in [0, 1)$, $\sqrt{n}(\widehat{\theta}^{bc} - \widehat{\theta}_w^{bc}) = o_p(1)$. However, if $\theta = 1$, the limiting distribution is not normal, specifically $\sqrt{n}(\widehat{\theta}_w^{bc} - \theta)/\sqrt{\widehat{v}} \Rightarrow \min\{Z, 0\}$, where $Z \sim N(0, 1)$.

2. This theorem can be used to construct consistent confidence intervals for θ. In order to make the interval respect the fact that the parameter space here is $[0, 1]$ one can make the interval for $H(\widehat{\theta}_w^{bc})$ and back transform, where $H : [0, 1] \to \mathbb{R}$ is strictly increasing and continuously differentiable, for example the logit transform. Specifically, let z_α be the αth quantile from a standard normal distribution, then the two-sided interval

$$\mathcal{C}_\alpha = H^{-1}[H(\widehat{\theta}_w^{bc}) - z_{\alpha/2}\sqrt{h^2(\widehat{\theta})\widehat{v}/n}, H(\widehat{\theta}_w^{bc}) + z_{\alpha/2}\sqrt{h^2(\widehat{\theta})\widehat{v}/n}]$$

has asymptotic coverage $1 - \alpha$ and lies inside $[0, 1]$ with probability one. Here, $h(\theta) = \partial H(\theta)/\partial\theta$.

3. 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.

4. If one strongly believes that $C_{f,g}$ is of measure zero, then one can conduct inference using the uncorrected estimator $\widehat{\theta}$ and the variance estimator $\widetilde{v} = \widetilde{p}_f(1 - \widetilde{p}_f) + \widetilde{p}_g(1 - \widetilde{p}_g)$, where

$$\widetilde{p}_f = \int f_n(x)1(f_n(x) < g_n(x))dx$$

$$\widetilde{g}_f = \int g_n(x)1(f_n(x) \geq g_n(x))dx,$$

that is, the tuning parameter $c_n = 0$. In this case, $\widetilde{v} \xrightarrow{p} v = \sigma_1^2$. However, if it turned out that $C_{f,g}$ has positive measure then $\widehat{\theta}$ is biased and the standard errors are inconsistent. Specifically, it can be shown that $\widetilde{p}_f \xrightarrow{p} p_f + p_0/2$ and $\widetilde{p}_g \xrightarrow{p} p_g + p_0/2$ using a similar argument to that used in Theorem 2.

5. If one strongly believes that $C_{f,g}$ is of measure zero and that there is a single crossing point x_0 such that $f(x) < g(x)$ for all $x < x_0$ and $f(x) > g(x)$ for all $x > x_0$ (scalar case), then one can estimate the crossing point by finding the minimum of $|f_n(x) - g_n(x)|$ over x. Under some conditions such an estimator $\widehat{x}$ is consistent and even asymptotically normal, and satisfies $\sqrt{nb}(\widehat{x} - x_0) \Rightarrow N(0, \omega)$ with $\omega = 2||K||_2^2 f(x_0)/(f'(x_0) - g'(x_0))^2$, see Eddy (1980).

6. 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 $\theta \to 1$, i.e., when the contact set has positive measure.

4   A Simulation Study

Here we look at the small-sample performance of $\widehat{\theta}$ and $\widehat{\theta}^{bc}$. 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 $C_{f,g} = [0, 0.5]$. 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 $\rho(t) = (1 + t)1(-1 \leq t \leq 0) + (1 - t)1(0 \leq t \leq 1)$. 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 $b_s = 1.84sn^{-1/5}$, where s is the sample standard deviation, or the smaller value $b_s^{3/2}$. In construction of the bias corrected estimator $\widehat{\theta}$ we choose the tuning parameter c_n to be either the bandwidth b, the smaller value $b^{3/2}$ or the larger value $b_s^{2/3}$.

The supports of interested 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 $[-0.5, 1]$.

In Figure 1 we show the result of a typical sample. Although in the population the two densities are equal throughout the interval $[0, 0.5]$, this happens with probability zero in sample. In this example there are seven crossings of the two density estimates.
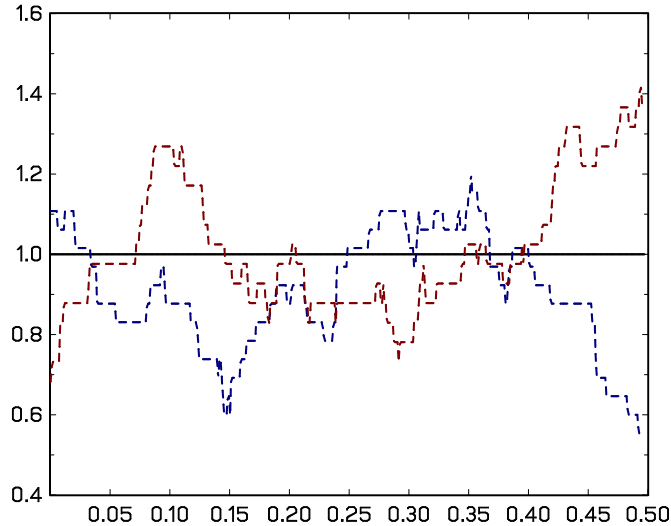

Figure 1. Shows two estimates of f, g reported over the intersection of their supports for case $n = 100$.

We report our results in Table 1. 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(-bias) on a constant and $\ln n$)

3. The bias corrected estimator has a smaller bias and larger standard deviation

4. The best performance for $\widehat{\theta}$ is when bandwidth is b_s although there is not a lot of difference for the larger sample sizes

5. The best performance in terms of standard deviation for $\widehat{\theta}^{bc}$ 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 2 we show the qq plot for standardized $\widehat{\theta}$ in the case $n = 800$ and bandwidth is b_s. The approximation seems quite good, with most discrepancy in the left tail.
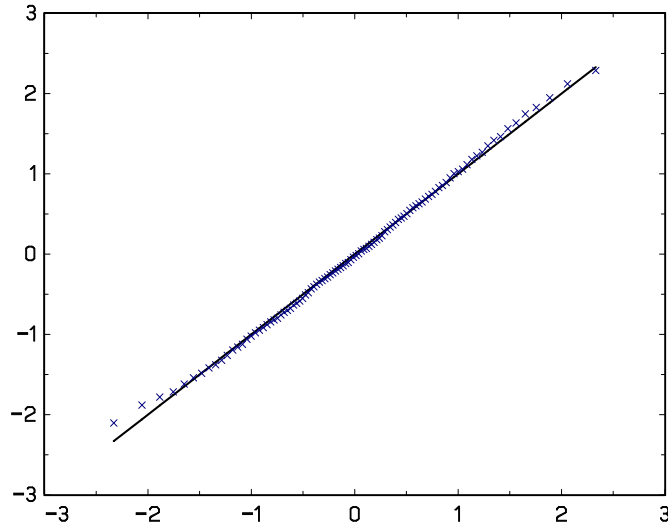
10

Figure 2. QQ-plot of sample data versus standard normal

In Table 2 we report the results of a multivariate simulation. The design was standard normal densities in dimensions 1,2,3,4 and 5. In this case $\theta = 1$ and $C_{f,g} = \mathbb{R}^d$. We implemented as above with the best combinations of bandwidth/tuning parameter uncovered in Table 1. The results are even more dramatic in this case. The bias correction method seems to produce much better bias with a small cost in terms of increased variability.

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), taken in 1987 and 2001 are employed.[4]

---

[4]These 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.

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).[5]

Indicators employed to reflect household wellbeing are total expenditures per household member and household living area per household member. Table 3 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.

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.

---

[5]Using 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)
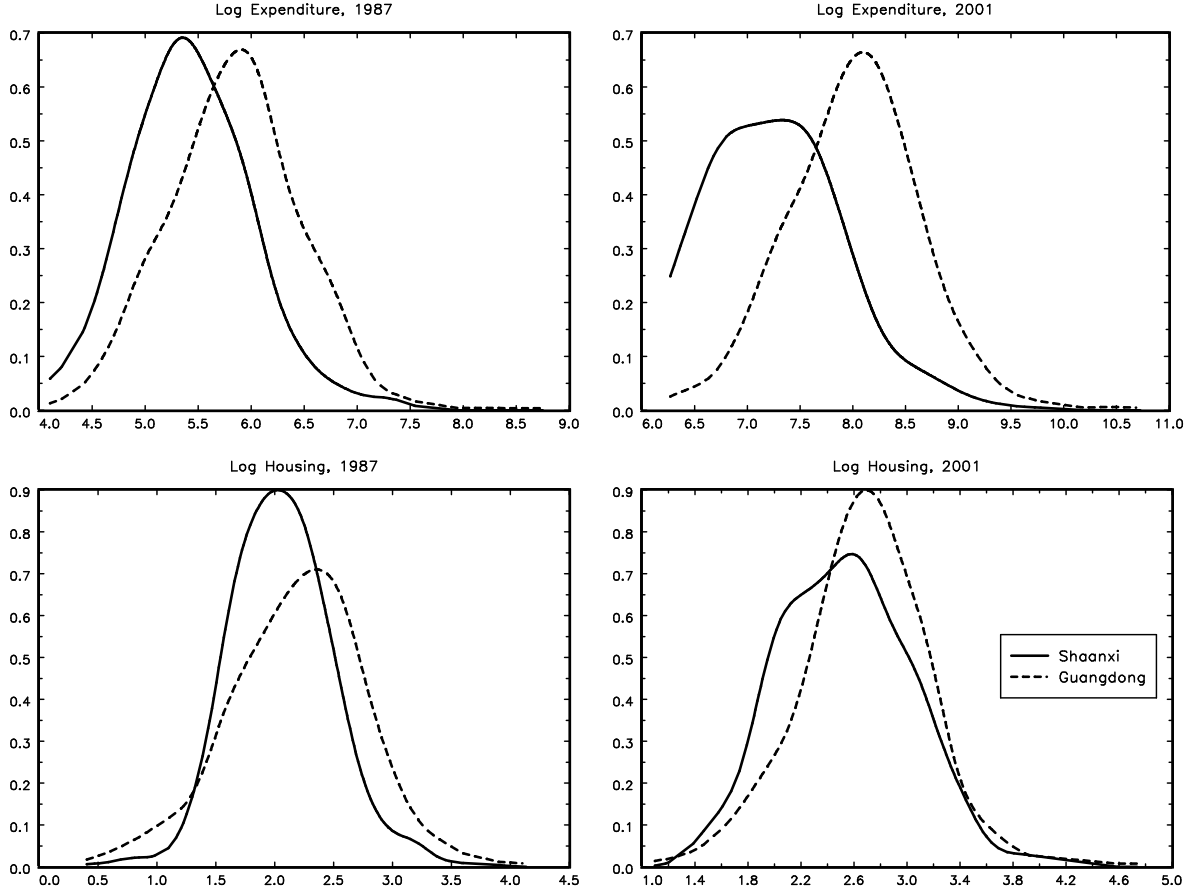
Figure 3

We next compute the univariate and multivariate polarization measures. Let $\widehat{\theta}_e, \widehat{\theta}_h$, and $\widehat{\theta}_{eh}$ 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. In this dataset 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 $m/n \rightarrow \tau \in (0, \infty)$, then the asymptotic distribution is as in Theorem 1 with

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

13

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 $\sigma_0^2(\tau) = \sigma_0^2(1)(1 + 1/\tau)/2$. For the bivariate product uniform kernel $\sigma_0^2(1) = 0.5835$. We computed the bias correction and standard errors using these modifications. The results are shown in Table 4. 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 statistical 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

Note that an alternative to the overlap measure could be obtained by computing the Duclos Esteban and Ray (2004) polarization measure 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 for the two years and record the results below.

	$\alpha = 0.25$	$\alpha = 0.5$	$\alpha = 0.75$	$\alpha = 1.0$
1987	0.27170	0.348900	0.459800	0.625500
se	0.00031	0.000410	0.000562	0.000784
2001	0.24020	0.331100	0.469700	0.680100
se	0.00024	0.000354	0.00052	0.000784

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

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 $\widehat{\theta}$ based on the asymptotic representation $\sqrt{n}(\widehat{\theta} - \theta) = n^{-1/2}\sum_{i=1}^n 1(X_i \in C_f) + n^{-1/2}\sum_{i=1}^n 1(Y_i \in C_g) + o_p(1)$, assuming some restriction on the strength of the dependence.

Our theory extends to the case of k densities $f_1, \ldots, f_k$ in an obvious way. In that case, one might also be interested in measuring quantities related to a partial overlap. Specifically, suppose

that $f_{i_1}(x) \leq \ldots \leq f_{i_k}(x)$, then $\min\{f_1(x), \ldots, f_k(x)\} = f_{i_1}(x)$ and $\max\{f_1(x), \ldots, f_k(x)\} = f_{i_k}(x)$. Then, $f_{i_r}(x)$, for $1 < r < k$ represents a situation where r of the densities overlap.[6]

We remark that the functional θ is related to some recent work of Kitagawa (2009) who has discussed the density envelope $\delta(f, g) = \int \max\{f(x), g(x)\}dx$, which arises as a quantity of interest from consideration of a partially identified model. Related to this quantity is an alternative overlap measure given by

$$\vartheta = \frac{\int_{-\infty}^{\infty} \min\{f(x), g(x)\}dx}{\int_{-\infty}^{\infty} \max\{f(x), g(x)\}dx}, \tag{10}$$

which has similar properties to θ. This quantity has the advantage of being sensible for cases where f, g are hazard functions or spectral densities that do not themselves integrate to one; in that case θ can take any non-negative value, whereas ϑ lies between zero and one. The distribution theory for the analogue estimator of ϑ follows by similar arguments to ones we have given above.

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 $\sqrt{n}(\widehat{\theta} - \theta)$ by A_n, given by (12) below, which decomposes the estimation error into three different terms, defined over the disjoint sets C_f, C_g and C_{fg}, respectively.

2. Get the asymptotic distribution of $A_n^P(\overline{B})$, 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 original sequence $\{(X_i, Y_i) : i \geq 1\}$ and the integral is taken over a subset $\overline{B}$ of the union of the supports of X and Y.

3. D*e-Poissonize* $A_n^P(\overline{B})$ to derive the asymptotic distribution of A_n and hence $\sqrt{n}(\widehat{\theta} - \theta)$.

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

---

[6]This is of interest in a number of biomedical applications. See for example http://collateral.knowitall.com/collateral/95391-OverlapDensityHeatMap.pdf

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 (2007) 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 $\sum_{i=0}^{N} 1(X_i \in A_k)\delta_{X_i}$, $k = 1, 2, ...$, are independent. 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 $A_n^P(\overline{B})$. Also, since a Poisson process is infinite divisible, we can write $\sum_{i=0}^{N} X_i \overset{d}{=} \sum_{i=0}^{n} Z_i$, where $\{Z_i : i \geq 1\}$ are i.i.d with $Z \overset{d}{=} \sum_{i=0}^{\eta_1} X_i$ 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 $A_n^P(\overline{B})$, using standard machineries including 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 $A_n^P(\overline{B})$ 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 $S_n = n^{-1/2} \sum_{i=1}^{N} \{1(X_i \in B) - \Pr(X \in B))\}$. The asymptotic distribution of S_n is given by $N(0, p_B)$, which is different from that of Λ_n. However, letting $U_n = n^{-1/2} \sum_{i=1}^{N} \{1(X_i \in C) - \Pr(X \in C))\}$ and $V_n = n^{-1/2} \sum_{i=1}^{N} \{1(X_i \in \mathbb{R}\backslash C) - \Pr(X \in \mathbb{R}\backslash C))\}$, where $B \subset C \subset \mathbb{R}$ is a Borel set, and applying the Poissonization lemma, we get that the conditional distribution of S_n given $N = n$ has the same asymptotic distribution as Λ_n.

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

$$\sup_{x \in \mathbb{R}^d} |f_n(x) - f(x)| = O(b^s) + O\left(\sqrt{\frac{\log n}{nb^d}}\right) \quad a.s., \tag{11}$$

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

$$
\begin{aligned}
A_n &= \sqrt{n} \int_{C_{f,g}} \min \{f_n(x) - Ef_n(x), g_n(x) - Eg_n(x)\} \, dx \\
&\quad + \sqrt{n} \int_{C_f} [f_n(x) - Ef_n(x)] \, dx + \sqrt{n} \int_{C_g} [g_n(x) - Eg_n(x)] \, dx \\
&=: A_{1n} + A_{2n} + A_{3n}
\end{aligned}
\tag{12}
$$

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

Theorem A1. *Under Assumptions (A1)-(A5), 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

$$
\sqrt{n} \left(\widehat{\theta} - \theta \right) = A_n + o_p(1).
\tag{13}
$$

Then, this result and Theorem A1 yield the desired result of Theorem 1. To show (13), write

$$
\begin{aligned}
\sqrt{n} \left(\widehat{\theta} - \theta \right) &= \int_{\mathbb{R}^d} \sqrt{n} \left[\min\{f_n(x), g_n(x)\} - \min\{f(x), g(x)\} \right] dx \\
&= \int_{C_{f,g}} \sqrt{n} \min\{f_n(x) - f(x), g_n(x) - g(x)\} dx \\
&\quad + \int_{C_f} \sqrt{n} \min\{f_n(x) - f(x), g_n(x) - f(x)\} dx \\
&\quad + \int_{C_g} \sqrt{n} \min\{f_n(x) - g(x), g_n(x) - g(x)\} dx \\
&=: \Lambda_{1n} + \Lambda_{2n} + \Lambda_{3n}.
\end{aligned}
\tag{14}
$$

Consider Λ_{1n} first. Write

$$
\begin{aligned}
\Lambda_{1n} &= \sqrt{n} \int_{C_{f,g}} \min \{f_n(x) - Ef_n(x), g_n(x) - Eg_n(x)\} \, dx \\
&\quad + \int_{C_{f,g}} \sqrt{n} \left[\min\{f_n(x) - f(x), g_n(x) - g(x)\} - \min \{f_n(x) - Ef_n(x), g_n(x) - Eg_n(x)\} \right] dx \\
&=: A_{1n} + \Lambda_{12n}.
\end{aligned}
\tag{15}
$$

We have

$$
\begin{aligned}
|\Lambda_{12n}| &\leq 2\sqrt{n}\left(\int_{C_{f,g}}\{|Ef_n(x)-f(x)|\,dx+|Eg_n(x)-g(x)|\}\,dx\right)\\
&= 4\sqrt{n}\left(\int_{C_{f,g}}|Ef_n(x)-f(x)|\,dx\right)\\
&\leq 4\sqrt{n}b^s\left(\int_{C_{f,g}}\int_{\mathbb{R}^d}\sum_{|\lambda|=s}\frac{1}{s!}\left|D^\lambda f(x-\widetilde{b}u)\right|\left|u^\lambda K(u)\right|\,du\,dx\right)\\
&= O(n^{1/2}b^s)\to 0,
\end{aligned}
\tag{16}
$$

where the first inequality uses the elementary result $|\min\{a+c,b+d\}-\min\{a,b\}|\leq 2\,(|c|+|d|)$, the first equality follows from the definition of $C_{f,g}$, the second inequality holds by a two term Taylor expansion with $0<\widetilde{b}<b$ 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}. We have

$$
\begin{aligned}
\Lambda_{2n} &= \int_{C_f}\sqrt{n}\min\{[f_n(x)-Ef_n(x)]\,,[g_n(x)-Eg_n(x)]+[g(x)-f(x)]\}dx+O(n^{1/2}b^s)\\
&= \int_{C_f}\sqrt{n}\,[f_n(x)-Ef_n(x)]\,dx+O(n^{1/2}b^{2s})+o_p(1)\\
&= A_{2n}+o_p(1),
\end{aligned}
\tag{17}
$$

where the first equality follows from an argument similar to the one to establish (16) and second equality holds by the following argument: Let $\varepsilon>0$ be a constant and write

$$
\begin{aligned}
\Delta_{2n} &= :\left|\int_{C_f}\sqrt{n}\min\{[f_n(x)-Ef_n(x)]\,,[g_n(x)-Eg_n(x)]+[g(x)-f(x)]\}dx\right.\\
&\qquad\left.-\int_{C_f}\sqrt{n}\,[f_n(x)-Ef_n(x)]\,dx\right|\\
&= \int_{C_f}\sqrt{n}\max\{[f_n(x)-Ef_n(x)]-[g_n(x)-Eg_n(x)]-[g(x)-f(x)]\,,0\}dx\\
&= \int_{C_{f,1}(\varepsilon)}\sqrt{n}\max\{[f_n(x)-Ef_n(x)]-[g_n(x)-Eg_n(x)]-[g(x)-f(x)]\,,0\}dx\\
&\quad+\int_{C_{f,2}(\varepsilon)}\sqrt{n}\max\{[f_n(x)-Ef_n(x)]-[g_n(x)-Eg_n(x)]-[g(x)-f(x)]\,,0\}dx\\
&= :\Delta_{21n}+\Delta_{22n},
\end{aligned}
$$

where

$$
\begin{aligned}
C_{f,1}(\varepsilon) &= \{x\in\mathbb{R}^d:0<g(x)-f(x)\leq\varepsilon\},\\
C_{f,2}(\varepsilon) &= \{x\in\mathbb{R}^d:g(x)-f(x)>\varepsilon\}.
\end{aligned}
$$

18

Take $\varepsilon = \delta(nb^d)^{-1/2}(\log n)$ for $\delta > 0$. Then, we have

$$
\begin{aligned}
|\Delta_{21n}| & \leq \int_{C_{f,1}(\varepsilon)} \sqrt{n} \max\{[f_n(x) - Ef_n(x)] - [g_n(x) - Eg_n(x)], 0\}dx \\
& \leq \sqrt{n}\left(\sup_{x\in\mathbb{R}^d}|f_n(x) - Ef_n(x)| + \sup_{x\in\mathbb{R}^d}|g_n(x) - Eg_n(x)|\right)\mu\left(C_{f,1}(\varepsilon)\right) \\
& \leq O_p\left(b^{-d/2}(\log n)^{1/2}\right)O\left((nb^d)^{-\gamma/2}(\log n)^\gamma\right) = o_p(1),
\end{aligned}
\tag{18}
$$

where the last inequality holds by the uniform consistency result (11) and Assumption A5, and the convergence to zero holds by Assumption A3. Also, for each $\eta > 0$, we have

$$
\begin{aligned}
\Pr\left(|\Delta_{22n}| > \eta\right) & \leq \Pr\left(\sup_{x\in\mathbb{R}^d}|f_n(x) - Ef_n(x)| + \sup_{x\in\mathbb{R}^d}|g_n(x) - Eg_n(x)| > \varepsilon\right) \\
& = \Pr\left(\left(\frac{nb^d}{\log n}\right)^{1/2}\left\{\sup_{x\in\mathbb{R}^d}|f_n(x) - Ef_n(x)| + \sup_{x\in\mathbb{R}^d}|g_n(x) - Eg_n(x)|\right\} > \delta(\log n)^{1/2}\right) \\
& \to 0.
\end{aligned}
\tag{19}
$$

Therefore, (17) follows from (18) and (19). Also, similarly to (17), we have

$$
\Lambda_{3n} = A_{3n} + o_p(1).
\tag{20}
$$

Now, (15), (16), (17) and (20) establish (13), as desired. $\blacksquare$

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 norms different from L_1- norm. 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

$$
\begin{aligned}
Ef_N(x) & = Ef_n(x) = EK_b(x - X) \tag{21} \\
k_{f,n}(x) & = n\mathrm{var}(f_N(x)) = EK_b^2(x - X) \tag{22} \\
n\mathrm{var}(f_n(x)) & = EK_b^2(x - X) - \left\{EK_b^2(x - X)\right\}^2. \tag{23}
\end{aligned}
$$

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 (4)-(6). Suppose p_0, p_f and p_g are strictly positive. (When any of p_0, p_f or p_g is zero, we can trivially take $\mathcal{B}(M)$, $\mathcal{B}(M_\phi)$, $\mathcal{B}(M, v)$, $\mathcal{B}(M_\phi, v_\phi)$, B_0 or $B_{\phi 0}$

19

for $\phi = f, g$ defined below to be an empty set in our subsequent discussion.) 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 $v_j > 0$, define $\mathcal{B}(M, v)$ to be the v-contraction of $\mathcal{B}(M)$, i.e., $\mathcal{B}(M, v) = \{x \in \mathcal{B}(M) : \rho(x, \mathbb{R}^d \backslash \mathcal{B}(M)) \geq v\}$, 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. Choose $M, M_\phi, v, v_\phi > 0$ and Borel sets B_0, B_ϕ such that

$$\mathcal{B}(M) \subset C_{f,g}, \ \mathcal{B}(M_\phi) \subset C_\phi, \tag{24}$$

$$B_0 \subset \mathcal{B}(M, v), \ B_\phi \subset \mathcal{B}(M_\phi, v_\phi) \tag{25}$$

$$\int\int_{\mathbb{R}^{2d} \backslash \mathcal{T}(M)} f(x)g(y)dxdy = : \alpha > 0 \tag{26}$$

$$\int_{B_0} f(x)dx = \int_{B_0} g(x)dx > p_0 - \varepsilon, \ \int_{B_\phi} \phi(x)dx > p_\phi - \varepsilon_\phi, \tag{27}$$

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

$$\mathcal{T}(M) = \big(\mathcal{B}(M_f) \times \mathbb{R}^d\big) \cup \big(\mathcal{B}(M) \times \mathcal{B}(M)\big) \cup \big(\mathbb{R}^d \times \mathcal{B}(M_g)\big) \subset \mathbb{R}^{2d}. \tag{28}$$

Such $M, M_\phi, v, v_\phi, B_\phi$, and $B_{\phi 0}$ exist by continuity of f and g.

Let $\overline{B} = B_0 \cup B_f \cup B_g$. Define a Poissonization version of A_n (minus its expectation, restricted to the Borel set $\overline{B}$) to be

$$A_n^P(\overline{B}) = A_{1n}^P(B_0) + A_{2n}^P(B_f) + A_{3n}^P(B_g), \tag{29}$$

where

$$A_{1n}^P(B_0) = \int_{B_0} \sqrt{n} \min\{f_N(x) - Ef_n(x), g_N(x) - Eg_n(x)\} dx \tag{30}$$

$$- \int_{B_0} \sqrt{n}E \min\{f_N(x) - Ef_n(x), g_N(x) - Eg_n(x)\} dx$$

$$A_{2n}^P(B_f) = \sqrt{n} \int_{B_f} [f_N(x) - Ef_n(x)] dx \tag{31}$$

$$A_{3n}^P(B_g) = \sqrt{n} \int_{B_g} [g_N(x) - Eg_n(x)] dx. \tag{32}$$

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

$$\sigma_n^2(\overline{B}) = \text{var}\left(A_n^P(\overline{B})\right). \tag{33}$$

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

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

Let $Z = (Z_1, \ldots, Z_4)^\top$ be multivariate normal with mean vector 0 and variance-covariance matrix

$$\Sigma = EZZ^\top = EWW^\top = \begin{pmatrix} 1 & 0 & \rho_1 & 0 \\ 0 & 1 & 0 & \rho_2 \\ \rho_1 & 0 & 1 & 0 \\ 0 & \rho_2 & 0 & 1 \end{pmatrix}.$$

Then, there exist universal positive constants A_1 and A_2 such that

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

and, whenever $\rho_1^2 < 1$ and $\rho_2^2 < 1$,

$$\left| E \left[\frac{1}{n} \min \left\{ \sum_{i=1}^{n} W_{1i}, \sum_{i=1}^{n} W_{2i} \right\} \min \left\{ \sum_{i=1}^{n} W_{3i}, \sum_{i=1}^{n} W_{4i} \right\} \right] - E \left[\min \{Z_1, Z_2\} \min \{Z_3, Z_4\} \right] \right|$$

$$\leq \frac{1}{(1 - \rho_1^2)^{3/2}} \frac{1}{(1 - \rho_2^2)^{3/2}} \frac{A_2}{\sqrt{n}} \left(E |W_1|^3 + E |W_2|^3 + E |W_3|^3 + E |W_4|^3 \right).$$

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

$$\Sigma = EZZ^\top = EWW^\top = \begin{pmatrix} 1 & 0 & \rho_1 \\ 0 & 1 & \rho_2 \\ \rho_1 & \rho_2 & 1 \end{pmatrix}.$$

Then, whenever $\rho_1^2 + \rho_2^2 < 1$,

$$\left| E \left[\min \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{1i}, \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{2i} \right\} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{3i} \right] - E \left[\min \{Z_1, Z_2\} Z_3 \right] \right|$$

$$\leq \frac{1}{(1 - \rho_1^2 - \rho_2^2)^{3/2}} \frac{A_3}{\sqrt{n}} \left(E |W_1|^3 + E |W_2|^3 + E |W_3|^3 \right).$$

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*

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

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 γ 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

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

where Z_1 denotes the standard normal random variable.

The following lemma derives the asymptotic variance of $A_n^P(\overline{B})$.

Lemma A3. *Whenever Assumptions (A1)-(A4) hold and B_0, B_f, and B_g satisfy (24)-(27), we have*

$$\lim_{n\to\infty} \sigma_n^2(\overline{B}) = p_{0,B}\sigma_0^2 + \bar{\sigma}_{1,B}^2, \tag{34}$$

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

Define

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

where $\mathcal{T}(M)$ is defined in (28). We next establish the following convergence in distribution result.

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

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

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 (26).

The following theorem gives the asymptotic bias formula.

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

(a) $\displaystyle \lim_{n\to\infty} \int_{B_0} \left[\sqrt{n}E \min\left\{f_N(x) - Ef_n(x), g_N(x) - Eg_n(x)\right\} dx - E\min\left\{Z_1, Z_2\right\} k_{f,n}^{1/2}(x)\right] dx = 0$

(b) $\displaystyle \lim_{n\to\infty} \int_{B_0} \left[\sqrt{n}E \min\left\{f_n(x) - Ef_n(x), g_n(x) - Eg_n(x)\right\} dx - E\min\left\{Z_1, Z_2\right\} k_{f,n}^{1/2}(x)\right] dx = 0,$

where Z_1 and Z_2 are standard normal random variables.

Define

$$
\begin{aligned}
A_n(\overline{B}) \;=\; & \sqrt{n} \int_{B_0} \left[\min\left\{ f_n(x) - E f_n(x), g_n(x) - E g_n(x) \right\} \right. \qquad\qquad (35) \\
& \left. - E \min\left\{ f_n(x) - E f_n(x), g_n(x) - E g_n(x) \right\} \right] dx \\
& + \sqrt{n} \int_{B_f} \left[f_n(x) - E f_n(x) \right] dx + \sqrt{n} \int_{B_g} \left[g_n(x) - E g_n(x) \right] dx.
\end{aligned}
$$

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 $\sigma_{1,B}^2 = p_{f,B}(1 - p_{f,B}) + p_{g,B}(1 - p_{g,B})$ 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 $\{X_i = (X_{1i}^\top, X_{2i}^\top)^\top \in \mathbb{R}^d \times \mathbb{R}^d : i = 1, \ldots, n\}$ be i.i.d random vectors with $E\|X\| < \infty$. Let $h_j : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be a real function such that $E h_j(X_j, x) = 0$ 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 $g : \mathbb{R} \to \mathbb{R}$, we have

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

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

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

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

for some generic constant $D > 0$, where

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

Now, we are now ready to prove Theorem A.

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

$$\lim_{k \to \infty} \int\limits_{C_{f,g} \setminus B_{0k}} f(x)dx = \lim_{k \to \infty} \int\limits_{C_{f,g} \setminus B_{0k}} g(x)dx = 0 \tag{36}$$

$$\lim_{k \to \infty} \int\limits_{C_f \setminus B_{fk}} f(x)dx = 0; \quad \lim_{k \to \infty} \int\limits_{C_g \setminus B_{gk}} g(x)dx = 0. \tag{37}$$

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

$$A_n(\overline{B}_k) \Rightarrow \sqrt{p_{0,B_k}\sigma_0^2 + \sigma_{1,B_k}^2}\, Z_1 \text{ as } n \to \infty. \tag{38}$$

By (36) and (37),

$$\sqrt{p_{0,B_k}\sigma_0^2 + \sigma_{1,B_k}^2}\, Z_1 \Rightarrow \sqrt{p_0\sigma_0^2 + \sigma_1^2}\, Z_1 \text{ as } k \to \infty. \tag{39}$$

Also, by Lemma A8, we have

$$\overline{\lim_{n \to \infty}} E \left(\sqrt{n} \int\limits_{C_{f,g} \setminus B_{0k}} \{h_n(x) - Eh_n(x)\}\, dx \right)^2 \leq D \left(\sup_u |K(u)| \right)^2 \int\limits_{C_{f,g} \setminus B_{0k}} [f(x) + g(x)]\, dx, \tag{40}$$

where

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

Similarly, we have

$$\overline{\lim_{n \to \infty}} E \left(\sqrt{n} \int\limits_{C_f \setminus B_{fk}} \{f_n(x) - Ef_n(x)\}\, dx \right)^2 \leq D \left(\sup_u |K(u)| \right)^2 \int\limits_{C_f \setminus B_{fk}} f(x)dx \tag{41}$$

$$\overline{\lim_{n \to \infty}} E \left(\sqrt{n} \int\limits_{C_g \setminus B_{gk}} \{g_n(x) - Eg_n(x)\}\, dx \right)^2 \leq D \left(\sup_u |K(u)| \right)^2 \int\limits_{C_g \setminus B_{gk}} g(x)dx. \tag{42}$$

Therefore, (40), (41), and (42) imply

$$\lim_{k \to \infty} \overline{\lim_{n \to \infty}} \Pr \left(\left| A_n(\overline{B}_k) - (A_n - EA_n) \right| > \varepsilon \right) = 0 \; \forall \varepsilon > 0. \tag{43}$$

Now, by (38), (39) and (43) and Theorem 4.2 of Billingsley (1968), we have, as $n \to \infty$,

$$
\begin{aligned}
& A_n - EA_n \\
= \; & \sqrt{n} \int_{C_{f,g}} [\min\{f_n(x) - Ef_n(x), g_n(x) - Eg_n(x)\} \\
& - E\min\{f_n(x) - Ef_n(x), g_n(x) - Eg_n(x)\}]\, dx \\
& + \sqrt{n} \int_{C_f} [f_n(x) - Ef_n(x)]\, dx + \sqrt{n} \int_{C_g} [g_n(x) - Eg_n(x)]\, dx \\
\Rightarrow \; & \sqrt{p_0\sigma_0^2 + \sigma_1^2}\, Z_1.
\end{aligned}
\tag{44}
$$

24

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

$$\lim_{n\to\infty} |EA_n - a_n| = 0.$$

∎

Proof of Theorem 2. To establish (8) we must show consistency of the bias correction. By the triangle inequality, we have

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

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

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

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

$$\begin{aligned}
D_{1n} &= \mu_f(\hat{C}_{f,g}\Delta C_{f,g}) = \mu_f(\hat{C}_{f,g} \cap C_{f,g}^c) + \mu_f(\hat{C}_{f,g}^c \cap C_{f,g}) \\
&\leq \mu_f(\hat{C}_{f,g} \cap \tilde{C}_{f,g}^c) + \mu_f(\tilde{C}_{f,g} \cap C_{f,g}^c) + \mu_f(\hat{C}_{f,g}^c \cap C_{f,g}) \\
&= \mu_f(\hat{C}_{f,g} \cap \tilde{C}_{f,g}^c \cap E_n) + \mu_f(\tilde{C}_{f,g} \cap C_{f,g}^c) + \mu_f(\hat{C}_{f,g}^c \cap C_{f,g} \cap E_n) \\
&\leq 2\mu_f(E_n) + \varsigma_n,
\end{aligned} \tag{45}$$

where $\varsigma_n = 2h(2c_n)$. 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

$$\int |h_n(x) - h(x)| \, f^{1/2}(x)dx \leq O_p\left(b^s + (nb^d)^{-1/2}\right). \tag{46}$$

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

$$\begin{aligned}
\Pr\left(b^{-d/2}D_{1n} > \varepsilon\right) &\leq \Pr\left(2\mu_f(E_n) + \varsigma_n > \varepsilon b^{d/2}\right) \\
&\leq \Pr\left(\frac{1}{c_n} \int |h_n(x) - h(x)| \, f^{1/2}(x)dx > \frac{\varepsilon b^{d/2} - \varsigma_n}{2}\right) \\
&\leq \Pr\left(\rho_n \int |h_n(x) - h(x)| \, f^{1/2}(x)dx > \frac{\varepsilon \rho_n c_n b^{d/2}}{2}\right) + o(1) \\
&\to 0,
\end{aligned} \tag{47}$$

where the first inequality holds by (45), the second inequality holds by the inequality $1(E_n) \leq |h_n(x) - h(x)|/c_n$, the third inequality follows from Assumptions A3 and A5 which implies $\rho_n c_n \varsigma_n \to 0$, and the last convergence to zero holds by (46) and $\rho_n c_n b^{d/2} \to \infty$ using Assumption A3. This now establishes that $D_{1n} = o_p(b^{d/2})$.

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

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

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

$$l_{jn} = \max\{\min_{1 \leq i \leq n} X_{ji}, \min_{1 \leq i \leq n} Y_{ji}\} - b/2$$
$$u_{jn} = \min\{\max_{1 \leq i \leq n} X_{ji}, \max_{1 \leq i \leq n} Y_{ji}\}] + b/2.$$

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

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

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

$$\left(\frac{nb^d}{||K||_2^2}\right)^{1/2} E\left[\frac{|f_n(x) - f(x)|}{f^{1/2}(x)}\right] \leq \left(\frac{nb^d}{||K||_2^2}\right)^{1/2} E^{1/2}\left[|f_n(x) - f(x)|^2\right] \frac{1}{f^{1/2}(x)}$$

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

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

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

$$\leq 2 + \left[\left(\frac{nb^d}{||K||_2^2}\right)^{1/2} b^s \sum_{|\lambda|=s} \frac{1}{s!} \frac{|D^\lambda f(x)|}{f^{1/2}(x)} \int u^\lambda K(u)du\right] + o(1),$$

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,

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

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

26

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

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

provided $E|X_{ji}|^p < \infty$. Therefore, we take $\delta_n = n^{1/p}L(n)$, where $L(n) \to \infty$ 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 (8). The consistency of the standard error follows by similar arguments. ∎

A.3   Proofs of Lemmas

Proof of Lemma A1. The results of Lemma A1 follow directly from Bhattacharya (1975, Theorem). ∎

Proof of Lemma A2. See Beirlant and Mason (1995, Proof of Theorem 2.1, pp.5-6). ∎

Proof of Lemma A3. To establish (34), 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

$$\mathrm{cov}\left(A_{1n}^P(B_0), A_{2n}^P(B_f)\right) = \mathrm{cov}\left(A_{1n}^P(B_0), A_{3n}^P(B_g)\right) = 0. \tag{48}$$

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

$$\begin{aligned}
\mathrm{var}\left(A_{2n}^P(B_f)\right) &= E\left(\int_{B_f} K_b(x - X)\,dx\right)^2 \to p_{f,B} \\
\mathrm{var}\left(A_{3n}^P(B_g)\right) &= E\left(\int_{B_g} K_b(x - Y)\,dx\right)^2 \to p_{g,B} \\
\mathrm{cov}\left(A_{2n}^P(B_f), A_{3n}^P(B_g)\right) &= E\left(\int_{B_f} K_b(x - X)\,dx\right)\left(\int_{B_g} K_b(x - Y)\,dx\right) \to p_{f,B}\,p_{g,B}\,.
\end{aligned} \tag{49}$$

Therefore, by (48) and (49), the proof of Lemma A3 is complete if we show

$$\lim_{n \to \infty} \mathrm{var}\left(A_{1n}^P(B_0)\right) = p_{0,B}\sigma_0^2. \tag{50}$$

27

To show (50), note that

$$\text{var} \left(A_{1n}^{P}(B_0) \right)$$
$$= n \int_{B_0} \int_{B_0} \text{cov} \left(\min \left\{ f_N(x) - E f_n(x), g_N(x) - E g_n(x) \right\}, \min \left\{ f_N(y) - E f_n(y), g_N(y) - E g_n(y) \right\} \right) dxdy$$
$$= n \int_{B_0} \int_{B_0} 1 \left(\|x - y\| \leq b \right)$$
$$\times \text{cov} \left(\min \left\{ f_N(x) - E f_n(x), g_N(x) - E g_n(x) \right\}, \min \left\{ f_N(y) - E f_n(y), g_N(y) - E g_n(y) \right\} \right) dxdy.$$

Let

$$T_{f,N}(x) = \frac{\sqrt{n} \left\{ f_N(x) - E f_n(x) \right\}}{\sqrt{k_{f,n}(x)}}, \tag{51}$$

$$T_{g,N}(x) = \frac{\sqrt{n} \left\{ g_N(x) - E g_n(x) \right\}}{\sqrt{k_{g,n}(x)}}, \tag{52}$$

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

$$\sup_{x \in B_0} \left| \sqrt{k_{f,n}(x)} - b^{-d/2} \|K\|_2 \sqrt{f(x)} \right| = O(b^{d/2}) \tag{53}$$

$$\sup_{x \in B_0} \left| \sqrt{k_{g,n}(x)} - b^{-d/2} \|K\|_2 \sqrt{g(x)} \right| = O(b^{d/2}) \tag{54}$$

$$\int_{B_0} \int_{B_0} 1 \left(\|x - y\| \leq b \right) dxdy = O(b^d) \tag{55}$$

$$\sup_{x,y \in \mathbb{R}^d} \left| \text{cov} \left(\min\{T_{f,N}(x), T_{g,N}(x)\}, \min \left\{ T_{f,N}(y), T_{g,N}(y) \right\} \right) \right| = O(1), \tag{56}$$

where (53) and (54) holds by two term Taylor expansions and (56) follows from Cauchy-Schwarz inequality and the elementary result $|\min\{a, b\}| \leq |a| + |b|$. Therefore, from (53) - (56), we have that

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

where

$$\bar{\sigma}_{n,0}^2 = \int_{B_0} \int_{B_0} 1 \left(\|x - y\| \leq b \right) \text{cov} \left(\min\{T_{f,N}(x), T_{g,N}(x)\}, \min \left\{ T_{f,N}(y), T_{g,N}(y) \right\} \right)$$
$$\times b^{-d} \|K\|_2^2 \sqrt{f(x) f(y)} dxdy. \tag{57}$$

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

$$(Z_{1n}(x), Z_{2n}(x), Z_{3n}(y), Z_{4n}(y))$$
$$\stackrel{d}{=} \left(Z_1, Z_2, \rho_{f,n}^*(x, y) Z_1 + \sqrt{1 - \left(\rho_{f,n}^*(x, y) \right)^2} Z_3, \rho_{g,n}^*(x, y) Z_2 + \sqrt{1 - \left(\rho_{g,n}^*(x, y) \right)^2} Z_4 \right),$$

28

where Z_1, Z_2, Z_3, and Z_4 are independent standard normal random variables and

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

Let

$$
\begin{aligned}
\overline{\tau}_{n,0}^2 &= \int_{B_0}\int_{B_0} 1\left(\|x - y\| \leq b\right) \mathrm{cov}\left(\min\{Z_{1n}(x), Z_{2n}(x)\}, \min\{Z_{3n}(y), Z_{4n}(y)\}\right) \\
&\quad \times b^{-d}\|K\|_2^2\sqrt{f(x)f(y)}dxdy.
\end{aligned}
\tag{58}
$$

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

$$
\begin{aligned}
\overline{\tau}_{n,0}^2 &= \int_{B_0}\int_{T_0} 1(x \in B_0)1\left(x + tb \in B_0\right) \mathrm{cov}\left(\min\{Z_{1n}(x), Z_{2n}(x)\}, \min\{Z_{3n}(x+tb), Z_{4n}(x+tb)\}\right) \\
&\quad \times \|K\|_2^2\sqrt{f(x)f(x+tb)}dxdt,
\end{aligned}
\tag{59}
$$

where $T_0 = \{t \in \mathbb{R}^d : \|t\| \leq 1\}$. Observe that, for almost every $x \in B_0$, we have

$$
\begin{aligned}
\rho_{f,n}^*(x, x+tb) &= \frac{b^{-d}E\left[K\left((x-X)/b\right)K\left((x-X)/b + t\right)\right]}{\sqrt{b^{-2d}EK^2\left((x-X)/b\right)EK^2\left((x-X)/b+t\right)}} \\
&\to \frac{\int_{\mathbb{R}^d}K(u)K(u+t)du}{\|K\|_2^2} = \rho(t).
\end{aligned}
\tag{60}
$$

Similarly, we have $\rho_{g,n}^*(x, x+tb) \to \rho(t)$ as $n \to \infty$ for almost all $x \in B_0$. (In fact, under our assumptions, the convergence (60) holds uniformly over $(x, t) \in B_0 \times T_0$.) Therefore, by the bounded convergence theorem, we have

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

Now, the desired result (34) holds if we establish

$$
\overline{\tau}_{n,0}^2 - \overline{\sigma}_{n,0}^2 \to 0.
\tag{61}
$$

Set

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

Notice that

$$
\int_{B_0}\int_{T_0} G_n(x, t)dxdt \leq \|K\|_2^2\mu(T_0 \times B_0)\sup_{x\in B_0}|f(x)| =: \overline{\beta} < \infty.
\tag{62}
$$

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

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

$$\overline{\tau}_{n,0}^2(\varepsilon_n) = \int_{B_0}\int_{T_{0,n,}} 1(x \in B_0)1(x + tb \in B_0)\operatorname{cov}\left(\min\{Z_{1n}(x), Z_{2n}(x)\}, \min\{Z_{3n}(x+tb), Z_{4n}(x+tb)\}\right)$$
$$\times\|K\|_2^2\sqrt{f(x)f(x+tb)}dxdt.$$

To show (61), we first establish
$$\overline{\tau}_{n,0}^2(\varepsilon_n) - \overline{\sigma}_{n,0}^2(\varepsilon_n) \to 0. \tag{63}$$

We have

$$\overline{\tau}_{n,0}^2(\varepsilon_n) - \overline{\sigma}_{n,0}^2(\varepsilon_n) = \left|\int_{B_0}\int_{T_{0,n}}\left[\operatorname{cov}\left(\min\{Z_{1n}(x), Z_{2n}(x)\}, \min\{Z_{3n}(x+tb), Z_{4n}(x+tb)\}\right)\right.\right.$$
$$\left.\left. -\operatorname{cov}\left(\min\{T_{f,N}(x), T_{g,N}(x)\}, \min\{T_{f,N}(x+tb), T_{g,N}(x+tb)\}\right)\right]G_n(x,t)dxdt\right|$$
$$\leq \int_{B_0}\int_{T_{0,n}}|E\min\{Z_{1n}(x), Z_{2n}(x)\}E\min\{Z_{3n}(x+tb), Z_{4n}(x+tb)\}$$
$$-E\min\{T_{f,N}(x), T_{g,N}(x)\}E\min\{T_{f,N}(x+tb), T_{g,N}(x+tb)\}|\,G_n(x,t)dxdt$$
$$+\int_{B_0}\int_{T_{0,n}}|E\min\{Z_{1n}(x), Z_{2n}(x)\}\min\{Z_{3n}(x+tb), Z_{4n}(x+tb)\}$$
$$-E\min\{T_{f,N}(x), T_{g,N}(x)\}\min\{T_{f,N}(x+tb), T_{g,N}(x+tb)\}|\,G_n(x,t)dxdt$$
$$=\ :\Delta_{1n} + \Delta_{2n}. \tag{64}$$

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

$$Q_{f,n}(x) = \left[\sum_{j \leq \eta_1} K\left(\frac{x - X_j}{b}\right) - EK\left(\frac{x - X}{b}\right)\right]\Big/\sqrt{EK^2\left(\frac{x - X}{b}\right)} \tag{65}$$

$$Q_{g,n}(x) = \left[\sum_{j \leq \eta_1} K\left(\frac{x - Y_j}{b}\right) - EK\left(\frac{x - Y}{b}\right)\right]\Big/\sqrt{EK^2\left(\frac{x - Y}{b}\right)}. \tag{66}$$

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

$$\sup_{x \in B_0} E\left|Q_{\phi,n}(x)\right|^3 = O(b^{-d/2})\ \text{for}\ \phi = f\ \text{and}\ g. \tag{67}$$

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

$$T_{f,N}(x) = \frac{\sqrt{n}\{f_N(x) - Ef_n(x)\}}{\sqrt{b^{-2d}EK^2\left((x - X)/b\right)}} \stackrel{d}{=} \frac{\sum_{i=1}^n Q_{f,n}^{(i)}(x)}{\sqrt{n}} \tag{68}$$

$$T_{g,N}(x) = \frac{\sqrt{n}\{g_N(x) - Eg_n(x)\}}{\sqrt{b^{-2d}EK^2\left((x - Y)/b\right)}} \stackrel{d}{=} \frac{\sum_{i=1}^n Q_{g,n}^{(i)}(x)}{\sqrt{n}}. \tag{69}$$

Therefore, by Lemma A1(a), (67), (68), and (69), we have

$$\sup_{x \in B_0}|E\min\{T_{f,N}(x), T_{g,N}(x)\} - E\min\{Z_{1n}(x), Z_{2n}(x)\}| \leq O\left(\frac{1}{\sqrt{nb^d}}\right). \tag{70}$$

The results (70) and (62) imply that $\Delta_{1n} = o(1)$. We next consider Δ_{2n}. We have

$$
\begin{aligned}
\Delta_{2n} \leq \quad & \sup_{(x,t) \in B_0 \times T_{0,n}} |E \min\{Z_{1n}(x), Z_{2n}(x)\} \min\{Z_{3n}(x + tb), Z_{4n}(x + tb)\} \\
& -E \min\{T_{f,N}(x), T_{g,N}(x)\} \min\{T_{f,N}(x + tb), T_{g,N}(x + tb)\}| \cdot \overline{\beta} \\
\leq \quad & O\left(\left(\frac{\varepsilon_n}{b}\right)^{-3\alpha}\right) \cdot O\left(\frac{1}{\sqrt{nb^d}}\right),
\end{aligned} \tag{71}
$$

where the first inequality uses (62) and the second inequality holds by Lemma A1(a), (67), Assumption A1(ii), and the fact that $\lim_{n\to\infty} \rho^*_{f,n}(x, x + tb) = \lim_{n\to\infty} \rho^*_{g,n}(x, x + tb) = \rho(t)$ a.s. uniformly over $(x, t) \in B_0 \times T_0$. Since ε_n is arbitrary, we can choose $\varepsilon_n = c_0 \cdot b (\log n)^{-1/(6\alpha)}$ for some constant c_0, where $\alpha > 0$ satisfies Assumption A1(ii). This choice of ε_n makes the right hand side of (71) $o(1)$, using Assumption A3. Therefore, we have $\Delta_{2n} = o(1)$, and hence (63).

On the other hand, using (56) and argument similar to (70), we have

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

This and (63) establish (61), and hence (50), as desired. ∎

Proof of Lemma A4. Let

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

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 $\mathbf{i} = (i_1, \ldots, i_d), i_1, \ldots, i_d \in \mathbb{Z}$ 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

$$
\begin{aligned}
\mu(R_{0,\mathbf{i}}) & \leq d_0 b^d, \ \mu(R_{f,\mathbf{i}}) \leq d_1 b^d, \ \mu(R_{g,\mathbf{i}}) \leq d_2 b^d \tag{72} \\
m_n & = : \mu(\mathcal{I}_n) \leq d_3 b^{-d}
\end{aligned}
$$

for some positive constants $d_0, .., d_3$, see, e.g., Mason and Polonik (2008) 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

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

Notice that

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

For arbitrary λ_1, λ_2 and $\lambda_3 \in \mathbb{R}$, let

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

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

$$
\begin{aligned}
\text{var}\left(\sum_{\mathbf{i} \in \mathcal{I}_n} y_{\mathbf{i},n}\right) &= \text{var}(\lambda_1 A_n^P(\overline{B}) + \lambda_2 U_n) \tag{73} \\
&\to \lambda_1^2\left(p_{0,B}\sigma_0^2 + \bar{\sigma}_{1,B}^2\right) + \lambda_2^2(1 - \alpha) + 2\lambda_1\lambda_2\left(p_{f,B} + p_{g,B}\right),
\end{aligned}
$$

and

$$
\sum_{\mathbf{i} \in \mathcal{I}_n} E|y_{i,n}|^r = o(1) \text{ for some } 2 < r < 3. \tag{74}
$$

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

We first establish (73). By Lemma A3, (73) holds if we have

$$
\text{cov}\left(A_n^P(\overline{B}), U_n\right) \to p_{f,B} + p_{g,B}. \tag{75}
$$

Recall that

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

Therefore, (75) holds if we have

$$
\begin{aligned}
\text{cov}\left(A_{2n}^P(B_f), U_n\right) &\to p_{f,B}, \qquad \text{cov}\left(A_{3n}^P(B_f), U_n\right) \to p_{f,B}, \tag{76} \\
\text{cov}\left(A_{1n}^P(\overline{B}), U_n\right) &= \text{cov}\left(\sqrt{n}\int_{B_0} \min\{f_N(x) - Ef_n(x), g_N(x) - Eg_n(x)\}dx, U_n\right) \tag{77} \\
&= O\left(\frac{1}{\sqrt{n}b^{2d}}\right)
\end{aligned}
$$

(76) holds by a standard argument. We show below (77). For any $x \in B_0$, we have

$$\left(T_{f,N}(x), T_{g,N}(x), \frac{U_n}{\sqrt{\Pr\left((X,Y) \in \mathcal{T}(M)\right)}} \right) \overset{d}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(Q_{f,n}^{(i)}(x), Q_{g,n}^{(i)}(x), U^{(i)} \right), \tag{78}$$

where $\left(Q_{f,n}^{(i)}(x), Q_{g,n}^{(i)}(x), U^{(i)} \right)$ for $i = 1, \ldots, n$ 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 (65) and (66) and

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

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

$$\sup_{x \in B_0} |\text{cov}\left(Q_{\phi,n}(x), U\right)| = O(b^{d/2}), \tag{79}$$

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

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

which, when combined with $\mu(B_0) < \infty$, yields (77) and hence (73), as desired.

We next establish (74). Notice that, with $2 < r < 3$, using Liapunov inequality and c_r-inequality, we have

$$\alpha_{\mathbf{i},n} = \int_{R_{0,\mathbf{i}}} 1(x \in B_0)\Delta_n(x)dx + \int_{R_{f,\mathbf{i}}} 1(x \in B_f)\Delta_{f,n}(x)dx \int_{R_{g,\mathbf{i}}} 1(x \in B_g)\Delta_{g,n}(x)dx,$$

$$\begin{aligned}
&\left(E\left|\alpha_{\mathbf{i},n}\right|^r\right)^{3/r} \\
&\le 9\left(\int_{R_{0,\mathbf{i}}} \int_{R_{0,\mathbf{i}}} \int_{R_{0,\mathbf{i}}} 1_{B_0}(x)1_{B_0}(y)1_{B_0}(z)E\left|\Delta_n(x)\Delta_n(y)\Delta_n(z)\right| dxdydz \right. \tag{80}\\
&\quad + \int_{R_{f,\mathbf{i}}} \int_{R_{f,\mathbf{i}}} \int_{R_{f,\mathbf{i}}} 1_{B_f}(x)1_{B_f}(y)1_{B_f}(z)E\left|\Delta_{f,n}(x)\Delta_{f,n}(y)\Delta_{f,n}(z)\right| dxdydz \\
&\quad + \left. \int_{R_{g,\mathbf{i}}} \int_{R_{g,\mathbf{i}}} \int_{R_{g,\mathbf{i}}} 1_{B_g}(x)1_{B_g}(y)1_{B_g}(z)E\left|\Delta_{g,n}(x)\Delta_{g,n}(y)\Delta_{g,n}(z)\right| dxdydz \right),
\end{aligned}$$

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

$$E\left|\Delta_n(x)\right|^3 \le Dn^{3/2}\left\{E\left|f_N(x) - Ef_n(x)\right|^3 + E\left|g_N(x) - Eg_n(x)\right|^3\right\}. \tag{81}$$

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

$$\sup_{x \in B_0} n^{3/2}E\left|f_N(x) - Ef_n(x)\right|^3 \le O\left(\frac{1}{b^{3d/2}} + \frac{1}{n^{1/2}b^{2d}}\right). \tag{82}$$

33

A similar result holds for g_N. Now, Assumption A3, (80), (81), (82), 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 (80) 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 (80). Therefore, we have

$$E|\alpha_{\mathbf{i},n}|^r \leq O(b^{rd/2}) \text{ uniformly in } \mathbf{i} \in \mathcal{I}_n. \tag{83}$$

This implies that

$$\sum_{\mathbf{i} \in \mathcal{I}_n} E|\alpha_{\mathbf{i},n}|^r \leq O(m_n b^{rd/2}) = O(b^{(r-2)/d}) = o(1). \tag{84}$$

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

$$\sum_{\mathbf{i} \in \mathcal{I}_n} E|u_{\mathbf{i},n}|^r \leq D_1 n^{-r/2} \sum_{\mathbf{i} \in \mathcal{I}_n} \left((np_{\mathbf{i},n})^{r/2} + np_{\mathbf{i},n}\right) \tag{85}$$

$$\leq D_1 \max_{\mathbf{i} \in \mathcal{I}_n} \left((p_{\mathbf{i},n})^{(r-2)/2} + n^{-1/2}\right) \to 0.$$

Therefore, combining (84) and (85), we have (74). This completes the proof of Lemma A4. ∎

Proof of Lemma A5. Consider part (a) first. Let $T_{f,N}(x)$ and $T_{g,N}(x)$ be defined as in (51) and (52), respectively. We have

$$\left|\int_{B_0} \left[\sqrt{n}E\min\{f_N(x) - Ef_n(x), g_N(x) - Eg_n(x)\} dx - E\min\{Z_1, Z_2\} k_{f,n}^{1/2}(x)\right] dx\right|$$

$$\leq \sup_{x \in B_0} |E\min\{T_{f,N}(x), T_{g,N}(x)\} - E\min\{Z_1, Z_2\}| \cdot \sup_{x \in B_0} k_{f,n}^{1/2}(x) \cdot \mu(B_0)$$

$$= O\left(n^{-1/2}b^{-d/2}\right) O(b^{-d/2}) = O(n^{-1/2}b^{-d}) = o(1),$$

by Lemma A1(a), (53), Assumption A3, and the fact $\mu(B_0) < \infty$. Similarly, we have

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

Therefore, part (b) also holds since, using (22) and (23),

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

∎

Proof of Lemma A6. Consider $A_n^P(\overline{B})$ defined in (29). Conditional on $N = n$, we have

$$A_n(\overline{B}) \stackrel{d}{=} \sqrt{n} \int_{B_0} [\min\{f_n(x) - Ef_n(x), g_n(x) - Eg_n(x)\} \tag{86}$$

$$- E\min\{f_N(x) - Ef_n(x), g_N(x) - Eg_n(x)\}] dx$$

$$+ \sqrt{n} \int_{B_f} [f_n(x) - Ef_n(x)] dx + \sqrt{n} \int_{B_g} [g_n(x) - Eg_n(x)] dx$$

$$=: A_n^C(\overline{B}).$$

34

By Lemmas A2 and A4, we have

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

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

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

∎

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

$$
\begin{aligned}
\xi_i &= E_i T_n - E_{i-1} T_n, \tag{87} \\
\eta_i &= E_i (T_n - T_{n,-i}), \tag{88}
\end{aligned}
$$

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

$$
\begin{aligned}
T_n - E T_n &= \xi_1 + \cdots \xi_n, \tag{89} \\
\xi_i &= \eta_i - E_{i-1} \eta_i, \tag{90} \\
|\eta_i| &\leq 2 \sum_{j=1}^2 \int_{\mathcal{B}} |h_j(X_{ji}, x)| \, dx, \tag{91}
\end{aligned}
$$

where (89) follows from (87), (90) holds by independence of X_i 's, and (91) follows from the elementary inequality $|\min\{a + b, c + d\} - \min\{a, c\}| \leq 2 (|b| + |d|)$. Let

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

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 $(X_1, \ldots, X_{i-1})$, and the conditional distributions of η_i and η_i^* given $(X_1, \ldots, X_{i-1})$ are equivalent. Therefore, for any

35

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

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

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 (91). Now, the result of Lemma A7 holds by (89) and Lemma 2.6 of Berger (1991). ∎

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

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

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

B   Further Material

B.1   Some Intuition

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

kernel is supported on $[-0.5, 0.5]$. Then let

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

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

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

We shall suppose for heuristic reasons that $\min\{\widehat{f}(x^t), \widehat{g}(x^t)\}$ is strictly independent of $\min\{\widehat{f}(x^s), \widehat{g}(x^s)\}$, 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

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

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

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

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

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

which is asymptotically normal with mean zero and variance

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

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

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

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

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

where $\widehat{m}(x^t)$ 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 $\widehat{\theta}$ 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 $\widehat{\phi}$ and $\widehat{\lambda}$ denote root-n consistent estimators of ϕ and λ, such that $[\sqrt{n}(\widehat{\phi} - \phi), \sqrt{n}(\widehat{\lambda} - \lambda)]^\top \implies U \sim N(0, \Omega)$, and let $\widehat{\theta} = \int_C \min\{f_{\widehat{\phi}}(x), g_{\widehat{\lambda}}(x)\} dx$. Then we have

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

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.

B.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

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

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

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

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

$$\begin{pmatrix} U \\ V \\ W \\ Y \end{pmatrix} = \sqrt{1 + \omega} \begin{pmatrix} U' \\ V' \\ W' \\ Y' \end{pmatrix}$$

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

Therefore,

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

We have

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

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

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

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

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

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

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

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

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

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

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

as expected.

In conclusion,

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

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.

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., Ge Y. and Leo T.W. (2006) "Distributional Overlap: Simple Multivariate Parametric and Non-Parametric Tests for Alienation, Convergence and General Distributional Differences" mimeo Economics Department University of Toronto. Forthcoming Econometric Reviews (2009).

[5] Anderson G.J. (2008) "Polarization of the Poor: Multivariate Poverty Measurement Sans Frontiers"mimeo University of Toronto Department of Economics.

[6] Andrews, D.W.K. (1995), "Nonparametric Kernel Estimation for Semiparametric Models." *Econometric Theory* 11, 560-596.

[7] Andrews, D. W. K. (1999), Estimation When a Parameter Is on a Boundary. Econometrica 67, 1341-1383.

[8] Andrews, D. W. K. and P. Guggenberger (2007), "The limit of exact size and a problem with subsampling and with the m out of n bootstrap," Working paper.

[9] Beirlant, J. and D. M. Mason (1995), "On the Asymptotic Normality of L_p-Norms of Empirical Functionals," *Math. Methods Statist.* 4, 1-19.

[10] Berger, E. (1991), "Marjorization, Exponential Inequalities and Almost Sure Behavior of Vector-valued Random Variables, *Annals of Probability* 19, 1206-1226.

[11] Bhattacharya, R. N. (1975) "On Errors of Normal Approximation," *Annals of Probability* 3, 815-828.

[12] Chernozhukov, V., H. Hong, and E. Tamer (2007), "Estimation and confidence regions for parameter sets in econometric models. *Econometrica,* 75, 1243-1284.

[13] Clemons, T.E. and E.L. Bradley (2000). "A nonparametric measure of the overlapping coefficient," Computational Statistics & Data Analysis 34, 51-61.

[14] Cuevas A. and R. Fraiman (1997), "A plug-in approach to support estimation," *Annals of Statistics* 25, 2300-2312.

[15] Duclos J-Y., J. Esteban and D. Ray (2004) "Polarization: Concepts, Measurement and Estimation" Econometrica 72 1737-1772.

[16] Eddy, W.F. (1980). "Optimum kernel estimators of the mode," The Annals of Statistics 8, 870-882.

[17] Esteban, J-M. And D. Ray (1994) "On the Measurement of Polarization" Econometrica 62, 819-851

[18] Giné, E., D. M. Mason, and A. Y. Zaitsev (2003), "The L_1-Norm Density Estimator Process," Annals of Probability 31, 719-768.

[19] Giné, E., and A. Guillou (2002), "Rates if strong uniform consistency for multivariate kernel density estimators," Annals of the institute of Henri Poincare 6, 907-921.

[20] Gustafsson, Bjorn, Li Shi and Terry Sicular (eds) (2007). Inequality and Public Policy in China, New York and Cambridge: Cambridge University Press

[21] 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.

[22] Kac, M. (1949). On deviations between theoretical and empirical distributions. *Proc. Natl. Acad. Sci.,* 35, 252-257.

[23] Kitagawa, T. (2009). "Testing for Instrument Independence in the Selection Model," Working paper, Brown University.

[24] 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.

[25] 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.

[26] Mason, D. M. and W. Polonik (2008), "Asymptotic Normality of Plug-In Level Set Estimates," Unpublished manuscript.

[27] 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

[28] Pinelis, I. F. (1994), "On a Majorization Inequality for Sums of Independent Random Variables," *Statistics and Probability Letters* 19, 97-99.

[29] Prakasa Rao, B.L.S. (1983). Nonparametric Functional Estimation, Academic Press.

[30] Ricklefs, R.E. and M. Lau (1980). "Bias and dispersion of overlap indices: results of some monte carlo simulations," *Ecology* 61, 1019-1024.

[31] Schmid, F., and A. Schmidt (2006), "Nonparametric estimation of the coefficient of overlapping - theory and empirical application," Computational Statistics and Data Analaysis 50, 1583-1596.

[32] 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.

[33] Wand, M.P. and M.C. Jones (1995) Kernel Smoothing. Chapman and Hall, Boca Raton.

[34] 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.

[35] Yitzhaki S., (1994) "Economic Distance and overlapping of distributions." Journal of Econometrics 61 147-159

C   Tables

			$\widehat{\theta}$				$\widehat{\theta}^{bc}$			
n	b	c	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 1. b_s is ROT bandwidth for density estimation $1.84sn^{-1/5}$

		$\widehat{\theta}$				$\widehat{\theta}^{bc}$			
d	n	bias	mbias	std	iqr	bias	mbias	std	iqr
	100	-0.18417	-0.18217	0.03551	0.03707	0.00020	0.00274	0.03675	0.03764
	200	-0.14226	-0.14069	0.02442	0.02569	0.00043	0.00210	0.02496	0.02609
1	400	-0.10921	-0.10820	0.01757	0.01742	0.00056	0.00154	0.01780	0.01756
	800	-0.08346	-0.08307	0.01282	0.01288	0.00065	0.00115	0.01285	0.01303
	1600	-0.06300	-0.06262	0.00840	0.00863	0.00123	0.00163	0.00842	0.00856
	100	-0.48052	-0.47917	0.04104	0.04008	-0.08851	-0.08728	0.04618	0.04543
	200	-0.40087	-0.39981	0.02927	0.02890	-0.05859	-0.05755	0.03273	0.03221
2	400	-0.32999	-0.32966	0.01796	0.01820	-0.03750	-0.03711	0.01991	0.02000
	800	-0.27177	-0.27166	0.01293	0.01308	-0.02623	-0.02600	0.01396	0.01426
	1600	-0.21953	-0.21956	0.00880	0.00878	-0.01563	-0.01593	0.00928	0.00943
	100	-0.80170	-0.80233	0.03372	0.03302	-0.28484	-0.28504	0.03684	0.03630
	200	-0.73569	-0.73673	0.02678	0.02638	-0.23728	-0.23776	0.02956	0.02879
3	400	-0.66785	-0.66731	0.02031	0.02012	-0.19374	-0.19286	0.02277	0.02246
	800	-0.59575	-0.59547	0.01434	0.01416	-0.15148	-0.15177	0.01603	0.01564
	1600	-0.52451	-0.52463	0.01007	0.00996	-0.11515	-0.11554	0.01144	0.01163
	100	-0.94858	-0.95000	0.02033	0.02004	-0.39799	-0.39857	0.02090	0.02046
	200	-0.92844	-0.92957	0.01757	0.01695	-0.38182	-0.38278	0.01823	0.01868
4	400	-0.90074	-0.90068	0.01377	0.01366	-0.35991	-0.36023	0.01450	0.01377
	800	-0.86798	-0.86761	0.01165	0.01191	-0.33469	-0.33433	0.01238	0.01266
	1600	-0.82744	-0.82727	0.00859	0.00858	-0.30376	-0.30374	0.00932	0.00954
	100	-0.98818	-0.99000	0.00997	0.01451	-0.43027	-0.43088	0.01038	0.01239
	200	-0.98413	-0.98500	0.00858	0.00741	-0.42703	-0.42799	0.00882	0.00851
5	400	-0.97766	-0.97779	0.00713	0.00741	-0.42160	-0.42198	0.00737	0.00749
	800	-0.96921	-0.96910	0.00599	0.00613	-0.41462	-0.41453	0.00621	0.00634
	1600	-0.95828	-0.95827	0.00482	0.00497	-0.40574	-0.40571	0.00502	0.00502

Table 2.

		Mean	Std Deviation	Median	Maximum	Minimum
Guangdong	Ln(Exp p.c.)	8.0137	0.6277	8.0338	10.7260	6.2687
(n=600)	Ln(Area p.c.)	2.6685	0.4726	2.6968	4.7361	1.0986
2001	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	Ln(Exp p.c.)	7.2324	0.6728	7.2092	9.7712	5.5977
(n=500)	Ln(Area p.c.)	2.5336	0.4902	2.5328	4.1997	1.3863
2001	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	Ln(Exp p.c.)	5.8395	0.6147	5.8452	8.7497	4.1015
(n=595)	Ln(Area p.c.)	2.1868	0.5702	2.2246	4.1352	0.4055
1987	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	Ln(Exp p.c.)	5.4156	0.5866	5.3898	7.3705	3.5231
(n=546)	Ln(Area p.c.)	2.0575	0.4346	2.0369	3.8067	0.4055
1987	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 3

		$\widehat{\theta}_e$	sesc_e	$\widehat{\theta}_e^{bc}$	se_e	$\widehat{\theta}_h$	sesc_h	$\widehat{\theta}_h^{bc}$	se_h	$\widehat{\theta}_{eh}$	sesc_{eh}	$\widehat{\theta}_{eh}^{bc}$	se_{eh}
1987	b_s	0.70201	0.02803	0.77217	0.03282	0.78201	0.02948	0.85663	0.03282	0.63403	0.02782	0.79142	0.03214
2001	b_s	0.52999	0.02694	0.59666	0.03366	0.76591	0.02833	0.83333	0.03323	0.50644	0.02665	0.65010	0.03289
1987	$\mathrm{b}_s^{3/2}$	0.69455	0.02861	0.77096	0.03432	0.77748	0.02947	0.84607	0.03456	0.59050	0.02771	0.81563	0.03279
2001	$\mathrm{b}_s^{3/2}$	0.52244	0.02752	0.58314	0.03487	0.82130	0.02709	0.90226	0.03518	0.47045	0.02571	0.65866	0.03434

Table 4

Figure 1: