Kernel-weightd test statistics under general distributions Preliminary and incomplete

Sid Kankanala
Yale University

Victoria Zinde-Walsh
McGill University and CIREQ

Abstract

Local kernel-weighted test statistics are widely used and have been applied in a variety of settings including inference on propensity score, panel models and non-stationary regression. The limit properties of the statistics have been developed under absolutely continuous distributions of the conditioning variables. This paper studies robustness of the properties of the kernel-weighted statistics to contamination of the conditioning distribution by singular components. It establishes the (slower) rate of convergence to limit Gaussian under the null and demonstrates sensitivity of the rate to degree of singularity. Limit power properties of the statistics are shown to be affected by the contamination as well. Finite sample examination of the kernel-weighted statistics and the competing functionals-based Bierens statistics shows that contamination results in a steep loss of power for both types of tests. The loss of power is more pronounced for the kernel-based tests reflecting the theoretical result that for local tests the rates are sensitive to contamination while for functionals based tests the rates remain the same.

1 Introduction

Kernel-weighted statistics are widely used for testing hypotheses about functions, such as density, conditional distribution, regression function, derivatives of the functions. Denote the class of functions that satisfy the null hypothesis, $V(H_0)$ and a function from that class by g^0. Let the class of unrestricted measurable functions under consideration be denoted by $V(H)$, with $V(H_0) \subset V(H)$ and a function from that class by g. The statistic is based on the difference between the estimated under the null function, $\hat{g}^0$, and the true $\bar{g} \in V(H)$. The true $\bar{g}$ can be replaced by the function non-parametrically estimated without restriction, $\hat{g}$, or by the observed measurements with error u, as in the case of test for a regression function when the true function is replaced with observed $y = \bar{g}(.) + u$.

The null hypothesis for $\bar{g} \in V(H)$ is

$$H_0 : \bar{g} = g^0 \in V(H_0), \text{ a.s.}$$

the alternative

$$H_1 : \bar{g} \notin V(H_0).$$

We focus here on non-parametric tests of a parametric conditional mean, where $V(H_0)$ consists of a single point. Moreover, if the function g^0 is parametrically estimated with a sample of size n, as $n \to \infty$ the estimated function in most metrics is in the $n^{-1/2}$ range from the true function, thus with non-parametric local estimators that under standard conditioning distributions are farther from the true function, the asymptotic distribution of the statistic is not affected by treating the estimated function as the true function in $V(H_0)$.

The difference $u = \bar{g} - g^0$ is zero a.s. under H_0. For any weighting matrix W_n and vector $\bar{u} = (u_1, ..., u_n)'$ the form $\bar{u}'W_n\bar{u} = 0$ under the null. The test statistic uses the $n \times 1$ vector $\hat{u}$ of estimated $\hat{u}_i = \widehat{\bar{g}}(X_i) - \hat{g}^0(X_i)$.

The typical form of the basic statistic is

$$\hat{\tau}_n\left(X, g\left(u\right)\right) = \frac{I_n}{\sqrt{\sigma_n^2}} = \frac{\hat{u}' W_n \hat{u}}{\sqrt{\hat{\sigma}^2\left(\hat{u}' W_n \hat{u}\right)}} \tag{1}$$

with $\hat{\sigma}^2\left(\hat{u}' W_n \hat{u}\right)$ representing the estimated variance of, $\hat{u}' W_n \hat{u}$. The kernel weighting matrix is based on a multivariate kernel function $K\left(.\right)$ and a bandwidth h_n. It employs as W_n the matrix with i, j entries $K\left(\frac{x_i - x_j}{h_n}\right)$ for $i \neq j$ and with 0 for $i = j$.

We shall focus on statistics where the specific form of the numerator is

$$I_n = \frac{1}{n(n-1)} \Sigma_{i=1}^n \Sigma_{j \neq i}^n \hat{H}_n\left(\hat{z}_i, \hat{z}_j\right); \tag{2}$$

$$\hat{H}_n\left(\hat{z}_i, \hat{z}_j\right) = \hat{u}_i \hat{u}_j K\left(\frac{X_i - X_j}{h}\right) \tag{3}$$

the squared denominator is

$$\hat{\sigma}_n^2 = \frac{2}{n(n-1)} \Sigma_{i=1}^n \Sigma_{j \neq i}^n \hat{u}_i^2 \hat{u}_j^2 K\left(\frac{X_i - X_j}{h}\right)^2. \tag{4}$$

The kernel-weighted test statistic can be examined as a ratio of second-order U-statistics (Hoeffding, 1948), whose limit properties rely on convergence rates for the moments. Hall (1984) provided the relation between the moments for U-statistics that correspond to the kernel-weighted structure in (1). While for many $U-$statistics asymptotic normality is established for any distribution of the argument as long as the variance has the parametric rate, the derivation of the asymptotic normality results with kernel function $K\left(\frac{X_i - X_j}{h_n}\right)$ and $h_n \to 0$ in the U statistic relied on assumptions of absolute continuity of the distribution $F_X\left(x\right)$ of the $X's$.

In testing for a parametric form of a conditional mean kernel-based tests have been widely used in the traditional regression context as in Hong and White (1995), Zheng (1996). Those types of statistics also are employed in various extensions, such as propensity score (Shaikh et al, 2009, Sant'ana and Song, 2019), panel data models (Lin et al, 2014) and non-stationary regression (Kasparis et al, 2015, Phillips and Wang, 2012), to name a few. Extensions of the tests to non-i.i.d. context required extensions to the results on the limit distributions of the U-statistics, such as e.g. those in Fan and Li (1996, 1999). All the available analyses assumed absolutely continuous distributions for X.

Here, we focus on the properties of the straightforward test of the parametric regression function in the i.i.d. setting under the general distribution of the conditioning variables. Our derivations have implications for the more complicated settings in which these tests are employed. We refer frequently to the paper by Zheng (1996) as Z-96. The statistic there is defined as

$$V_n = \frac{n h^{q/2} \frac{1}{n(n-1)h^q} \Sigma_{i=1}^n \Sigma_{j \neq i = 1}^n \hat{u}_i \hat{u}_j K\left(\frac{x_i - x_j}{h}\right)}{\left\{\frac{2}{n(n-1)h^q} \Sigma_{i=1}^n \Sigma_{j \neq i = 1}^n \hat{u}_i^2 \hat{u}_j^2 K\left(\frac{x_i - x_j}{h}\right)^2\right\}^{1/2}} \tag{5}$$

which is related to $\hat{\tau}_n$ here via the scaling: $V_n = n h^{q/2} \hat{\tau}_n$.

We investigate the impact on the asymptotic distribution of the statistic $(1, 2, 4)$ of relaxing the assumptions on the distribution F_X. To simplify the exposition we assume that the joint distribution of $z_i = (u_i, X_i)$ is i.i.d. with $E\left(u|X\right) = 0$.

The general distribution for conditioning variables X, $F_X\left(x\right)$, has the well-known Lebesgue decomposition: it can be represented as a mixture of a discrete distribution, $F_X^d\left(x\right)$, and a distribution given by a continuous function, $F_X^c\left(x\right)$, which is in turn, a mixture of a absolutely continuous distribution, $F_X^{a.c.}\left(x\right)$, and a singular distribution, $F_X^s\left(x\right)$. The singular distribution is given by a continuous distribution function but there is no density function that integrates to it. We examine the general distribution given by the mixture

$$\begin{aligned} F_X(x) &= \rho_d F_X^d(x) + \rho_s F_X^s(x) + \rho_{a.c.} F_X^{a.c.}(x); \\ \rho_d + \rho_s + \rho_{a.c.} &= 1; \rho_d \geq 0; \rho_s \geq 0; \rho_{a.c.} \geq 0. \end{aligned} \tag{6}$$

The first contribution of this paper is the derivation of the limit behavior of the U statistics related to the numerator and denominator of the statistic $(1,2,4)$:

$$U_n = \frac{1}{n(n-1)} \Sigma_{i=1}^n \Sigma_{j \neq i}^n H_n(z_i, z_j); \tag{7}$$

$$H_n(z_i, z_j) = u_i u_j K\left(\frac{X_i - X_j}{h}\right); \tag{8}$$

$$S_n^2 = 2\frac{1}{n(n-1)} \Sigma_{i=1}^n \Sigma_{j \neq i}^n \tilde{H}_n(z_i, z_j); \tilde{H}_n(z_i, z_j) = u_i^2 u_j^2 K\left(\frac{X_i - X_j}{h}\right)^2 \tag{9}$$

for classes of distributions with singular components. Note that we do not scale the kernel function by $\frac{1}{h^q}$ here because it will not facilitate the derivations when the distribution is not absolutely continuous. We show that $\frac{nU_n}{\sqrt{S_n^2}}$ converges to $N(0,1)$ for a wide class of distributions for X.

The joint distribution of $z_i = (u_i, X_i)$ differs from the distribution of $\hat{z}_i = (\hat{u}_i, X_i)$. Under the absolutely continuous distribution $F_X(x) = F_X^{a.c.}(x)$, the terms coming from the difference $\hat{u} - u$ do not affect the asymptotic distribution. These terms may result in a finite sample bias, the impact of which could be substantial, but goes away asymptotically (Ellison and Ellison, 2000). Most of the literature does not take account of that bias in asymptotic derivations. Our **second contribution** is to show that with a general distribution for X there may be asymptotic bias.

Apart from the asymptotic distribution the bootstrap distribution is of interest as many tests are implemented with bootstrap. To avoid mass points in the bootstrap distribution typically wild bootstrap that smooths the empirical distribution is used in testing when absolute continuity of the distribution is assumed. However, if the true underlying distribution is not absolutely continuous, the wild bootstrap distribution may smooth away the singularities and mass points and be a poor approximation to the true distribution of the statistic. Our **third contribution** is a finite sample evaluation of the wild bootstrap procedure that illustrates a dramatic loss of accuracy under the null and loss of power when the distribution of X does not conform well with the usual assumptions. Monte Carlo experiments in the literature all utilize well behaved (normal and uniform) distributions; we demonstrate significant power loss for the case of a mixture of three normals where large values of the derivative of the density may undermine performance.

We also consider the main alternative, what Fan and Li (2000) call Bierens-type tests (Bierens, 1982 and subsequent work) to kernel-based tests of conditional mean and compare the two types of test in simulation. We refer to the paper by Fan and Li (2000) as FL-2000. The test statistic first introduced by Bierens (1982) in the i.i.d. case and later extended to stationary time series uses OLS residuals $\hat{u}_j$ and weights $w(x,\tau)$ (e.g. Fourier transform weights $w(x,\tau) = \exp\left(\mathbf{i}\boldsymbol{\tau}'\right)$ ($\mathbf{i}^2 = -1$), or some transformed version); the Integrated Conditional Moment (ICM) statistic is

$$\hat{T}_n = \int \left| \frac{1}{\sqrt{n}} \Sigma_{j=1}^n \hat{u}_j w(x_j, \tau) \right|^2 d\tau. \tag{10}$$

Distributional results in Bierens, Ploberger (1997) permitted to establish power against sequences of $n^{-\frac{1}{2}}$ drifting alternatives. The limit null distribution is a mixture of independent χ_1^2; simulations or bootstrap are required to control the test size. No assumptions on the marginal distribution, F_X, of X are made for the Bierens-type tests. The relation between "local" kernel-weighted and Bierens type functionals based tests was examined theoretically in FL-2000. A comparison of the performance of the two types of tests in application to propensity score is examined in Sant'ana and Song, 2019. **The fourth contribution** of this paper is a finite-sample comparison of the performance of the two types of tests in a non-standard set-up for the distribution F_X. We show that both tests exhibit a dramatic decline in performance but the functional-based Bierens-type test is more stable and more powerful. An additional point that is made in the literature is that despite lack of power against $O\left(n^{-1/2}\right)$ local alternatives the kernel-based tests provide better than $O\left(n^{-1/2}\right)$ power against "spiked" alternatives (e.g., FL-2000). We note that this result may not hold in the case of a general distribution for X.

Finally, our simulation experiment evaluates the properties of the test with conditioning distribution that differs from of the cases most extensively considered in the simulation experiments in the literature (normality or uniformity). The simulation demonstrates that even under standard assumptions the power of the test is dramatically reduced under e.g. a mixture of normals distribution with narrow peaks; these results are similar to the ones in (Kotlyarova et al, 2016). Our **fifth contribution** is an evaluation of the impact of the deviation of the conditioning distribution from the ones routinely used in evaluation of the performance of the test.

The next section 2 provides the asymptotic results for the U-statistics and the test statistic; section 3 evaluates the asymptotic bias for kernel-based tests. Section 4 provides a discussion of the implications of the general distribution for conditioning variables. Section 5 provides simulations evidence. The proofs are in the Appendix.

2 Asymptotics for kernel-function based second order degenerate U statistics

This part develops the asymptotics for the U-statistics in (7) and (9).

Assumption 1. *The vector* $(u_i, x_i')'$ *for* $i = 1, ..., n$ *is i.i.d. in* $R \times \mathbb{R}^q$.

Assumption 2.(a) $E(u_i|X_i) = 0$; *(b) the functions* $\mu_2(x) = \sigma^2(x) = E\left(u_i^2|x\right)$ *and* $\mu_4(x) = E\left(u_i^4|x\right)$ *are continuous and bounded:* $0 < b \leq \mu_l(x) \leq B < \infty$.

Assumption 1 could be generalized but is made here to facilitate the focus on the distribution of conditioning variables. Assumptions 1 and 2 (a) imply that the statistic U_n in (7) is degenerate. The conditional moments assumptions 2(b) are common, although often the lower bound assumption is not explicitly made; of course, it applies under homoscedasticity.

Assumption 3. (a) *For any* $x, y \in R^q$ *with* $x = \left(x^1, ..., x^q\right)', y = \left(y^1, ..., y^q\right)'$ *the function* $K(.)$ *is a product kernel* $K\left(\frac{x-y}{h}\right) = \Pi_{l=1}^q k\left(\frac{x^l - y^l}{h}\right)$; *we assume that the bandwidth is the same for all the* q *components; (b) The kernel function,* $k(v)$, *is a bounded symmetric second order non-negative kernel monotone over* $v > 0$ *with support in* $[-1, 1]$; *(c) the bandwidth satisfies* $h \to 0, h^q n \to \infty$ as $n \to \infty$.

Assumptions 3 (a,b) assume a commonly used kernel function, e.g. Epanechnikov kernel. Here the product kernel does not involve division by h^q as is common; this does not affect the statistic in (1) and makes highlighting the rate of convergence to the limit process more straightforward. 3 (c) gives the usual restriction on the speed at which the bandwidth goes to zero. When the bandwidth is fixed the U-statistic kernel does not depend on n and the standard results for asymptotic normality hold. Assumption 3 could be generalized to admit a wider class of kernel functions as well as different bandwidths for different components, but here we simplify to highlight the impact of the distribution F_X.

Let F_X be the distribution function associated to X. For continuous functions $\mu_2(t), \mu_4(t)$ and Borel set $A \subseteq \mathbb{R}^q$ we define

$$\Omega_l(A) = \int_A \mu_l(t) dF_X(t), l = 2, 4.$$

This defines a Borel measure $d\Omega_l$ on $\mathbb{R}^q$ with $d\Omega_l(t) = \mu_l(t) dF_X(t)$, so that any such measure is absolutely continuous with respect to the measure associated with the distribution F_X. Consider $x \in R^q, w_1, w_2 \in R_+^q$, then define the set A by $A(x - w_1, x + w_2) = \{y \in \mathbb{R}^q : x^i - w_1^i \leq y^i \leq x^i + w_2^i, \ \forall i\}$. Denote by $\Omega_l(x - w_1, x + w_2)$ the value of $\Omega_l(A(x - w_1, x + w_2))$.

We shall use the following additional assumption about the kernel function.

Assumption 4. *The kernel function, k, is twice continuously differentiable.*

The reason for this assumption of kernel differentiability is that in the event that density does not exist one can use integration by parts to find the local behavior of the moments of the U statistics. Examination of limit properties of kernel statistics when density may not exists employs the theory of generalized functions (see, e.g. Gelfand and Shilov, 1964 a,b) and was given in Zinde-Walsh, 2008, 2017.

For a suitably differentiable function, $g(x)$ with $x = \left(x^1, ..., x^q\right) \in \mathbb{R}^q$ denote by $\partial_x g(x)$ the derivative $\frac{\partial^q g}{\partial x^1 ... \partial x^q}(x)$. Note that by our assumptions for the product kernel function, $K\left(\frac{x_1 - x_2}{h}\right)$, the derivative $\partial_{x_.} K(.)$ exists for any h.

Next, we derive the moments for powers and cross-products of $H_n(.,.)$, that determine the asymptotic behavior of the $U-$statistic U_n in (7). Denote by ι the vector of ones. Denote by $G_n(Z_1, Z_2)$ the conditional expectation $E\left(H\left(Z_1, Z\right) H\left(Z_2, Z\right) | Z_1, Z_2\right)$.

Lemma 1 *Under Assumptions* $1-4$*, we have that :*

$$\mathbb{E}\big(H_n(Z_1, Z_2)\big) = 0 \tag{11}$$

$$\mathbb{E}\big(|H_n(Z_1, Z_2)|^2\big) = \mathbb{E}\bigg(\mu_2(X)(-1)^q \int_{\mathbb{R}^q} \Omega_2(X - h\iota, X + hv)\partial_v K^2(v) dv\bigg) \tag{12}$$

$$\mathbb{E}\big(|H_n(Z_1, Z_2)|^4\big) = \mathbb{E}\bigg(\mu_4(X)(-1)^q \int_{\mathbb{R}^q} \Omega_4(X - h\iota, X + hv)\partial_v K^4(v) dv\bigg) \tag{13}$$

$$\begin{aligned} &\mathbb{E}\big(|G_n(Z_1, Z_2)|^2\big) \\ = \; &\mathbb{E}\bigg((-1)^q \mu_2(X) \int_{\mathbb{R}^q_+} \Omega_2(X, X + hv)\partial_u \Big[\int_{\mathbb{R}^q} \Omega_2(X - h\iota, X + hv)\partial_v \big[K(v)K(v+u)\big] dv\Big]^2 du\bigg) \end{aligned} \tag{14}$$

If X has a density f_X with respect to the Lebesgue measure, then :

$$h^{-q}\mathbb{E}\big(|H_n(Z_1, Z_2)|^2 \xrightarrow[h \to 0^+]{} \mathbb{E}\bigg((-1)^q \mu_2^2(X) f_X(X) \int_{\mathbb{R}^q} \big[\partial_v K^2(v)\big] \prod_{i=1}^q v_i dv\bigg) \tag{15}$$

If the distribution of X has a non trivial singular component under its Lebesgue-Radon-Nikodym decomposition, then :

$$h^{-q}\mathbb{E}\big(|H_n(Z_1, Z_2)|^2 \xrightarrow[h \to 0^+]{} +\infty \tag{16}$$

This Lemma provides the standard result for limit of $\mathbb{E}\big(\big|\frac{1}{h^q} H_n(Z_1, Z_2)\big|^2\big)$ when density exists, but demonstrates that when there is singularity at the standard rate $\mathbb{E}\big(\big|\frac{1}{h^q} H_n(Z_1, Z_2)\big|^2\big)$ diverges. The next Lemma provides bounds on the moments that will help determine the limit behaviour of the U-statistics in (7) and (9). For a univariate function $\psi(x)$ and $v = \left(v^1, ..., v^q\right) \in R^q$ define $D(v)$: $D(v) = \Pi_{i=1}^q \psi\left(v^i\right)$. Consider the functions

$$\begin{aligned} \psi_1(x) &= -\left(k^2(x)\right)', \\ \psi_2(x) &= \max\left\{\left|\left[k(x)k(x+u)\right]'\right|, \left|\left[k(-x)k(-x+u)\right]'\right|\right\}, \\ \psi_3(x) &= \max\left\{\left|\left[k(x)k(x+u)'\right]'\right|, \left|\left[k(-x)k(-x+u)'\right]'\right|\right\}, \psi_4(x) = \left|\left(k^4(x)\right)'\right|. \end{aligned}$$

The corresponding product functions are denoted $D_1(v)$, $D_2(v)$, $D_3(v)$ and $D_4(v)$.

Note that $D_1(v) = \partial_v K^2(v)$ and the proof of Lemma 1 implies that

$$\mathbb{E}\big(|H_n(Z_1, Z_2)|^2\big) = \mathbb{E}\bigg(\mu_2(X) \int_{\mathbb{R}^q_+} \Omega_2(X - hv, X + hv) D_1(v) dv\bigg).$$

We now develop some bounds on the moments. Denote $\int_{X-a\iota}^{X+a\iota} dF_X$ by $F_X(X + a\iota, X - a\iota)$.

Lemma 2 *Under Assumptions* $1-4$*, for any* $\varepsilon : 1 > \varepsilon > 0$ *we have that*

(i)

$$\begin{aligned} b^2 M_1 \mathbb{E}\Big([F_X(X - h\varepsilon\iota, X + h\varepsilon\iota)] \Big) &\leq \mathbb{E}\big(|H_n(Z_1, Z_2)|^2 \big) \\ &\leq B^2 M_1 \mathbb{E}\Big([F_X(X - h\iota, X + h\iota)] \Big) \end{aligned}$$

where $M_1 = \int D_1(v) I\left(v : v^i > \varepsilon, i = 1, ..., q\right) dv$*;*

(ii)

$$\mathbb{E}\big(|H_n(Z_1, Z_2)|^4 \big) \leq 2^q B^2 M_4 \mathbb{E}\Big([F_X(X - h\iota, X + h\iota)] \Big)$$

where $M_4 = \int D_4(v) dv$*;*

(iii)

$$\begin{aligned} & \mathbb{E}\big(|G_n(Z_1, Z_2)|^2 \big) \\ \leq\ & 2^{2q} B^4 M_{2,3} \mathbb{E}\Big([F_X(X - h\iota, X + h\iota)]^3 \Big) \end{aligned}$$

where $M_{2,3} = \int_{\mathbb{R}_+^q} \left[\int_{\mathbb{R}_+^q} D_2(v_1) \, dv_1 \right] D_3(v_2) \, dv_2.$

The class of distributions in the literature where the behavior of the statistic was analysed so far includes only absolutely continuous distributions with bounded density functions. We now propose an expanded class of distributions where we can establish asymptotic normality of the statistic.

The Assumption below defines a wide class $\mathcal{D}$ of continuous distribution functions F_X .

Assumption 5. *$\mathcal{D}$ is a class of continuous distribution functions for $X \in R^{q\in}$ such that any F_X in $\mathcal{D}$ satisfies the following moment conditions as $h \to 0$*

(i) $E\left[F_X(X - h\iota, X + h\iota)\right] \to 0;$

(ii) *for some $\varepsilon : 0 < \varepsilon < 1$ and $h_0 > 0$ for any $h > h_0$* there is $\bar{B} < \infty$ for which

$$\frac{E\left[F_X(X - h\iota, X + h\iota)\right]}{E\left[F_X(X - (h\varepsilon)\,\iota, X + (h\varepsilon)\,\iota)\right]} < \bar{B}. \tag{17}$$

(iii) as $h \to 0$

$$\frac{E\left[F_X(X - h\iota, X + h\iota)^3\right]}{\left(E\left[F_X(X - h\iota, X + h\iota)\right]\right)^2} \to 0. \tag{18}$$

Any absolutely continuous distribution $F_X^{a.c.}$ with a bounded uniformly continuous density function, f_X, satisfies the Assumption. Indeed, then at every X as $h \to 0$

$$|F(X - hv, X + hv)| = f_X(X)\,(vh)^q\,[1 + \xi]\,;\ \xi \to 0.$$

The conditions (i-iii) of the Assumption hold then and $EF_X\left[(X - h\iota, X + h\iota)\right]$ goes to zero at the rate h^q.

More generally, (i) may hold, but at a different rate from that of the case of absolutely continuous F_X. A sufficient condition for (i) boundedness of the support of X; then the property holds by uniform continuity. However, (i) can also hold for many distributions with unbounded support.

For (ii) a sufficient condition is that the F_X distribution be doubling on support of F_X, possibly all R^q (see, e.g. Vol'berg and Konyagin, 1988, Theorem 2), since then

$$F_X(X - h\iota, X + h\iota) \leq C\varepsilon^{-q} F_X(X - h\varepsilon\iota, X + h\varepsilon\iota). \tag{19}$$

By monotonicity of F_X

$$F_X(X - h\varepsilon\iota, X + h\varepsilon\iota) \leq F_X(X - h\iota, X + h\iota)$$

and thus by dominated expectation (ii) is established. However, even when the doubling property may not hold, the bounds that depend on the expectations could still apply.

Since F_X is a continuous distribution function $E\left(F_X(X - h\iota, X + h\iota)^3\right) < E\left[F_X(X - h\iota, X + h\iota)\right]$ and goes to zero under Assumption 5(i), but (iii) requires further that it go to zero faster than $\left(EF_X\left[(X - h\iota, X + h\iota)\right]\right)^2$.

There is a rich class of distributions beyond the absolutely continuous ones that satisfies Assumption 5 and belongs to $\mathcal{D}$. This is a class of mixtures of absolutely continuous and singular distributions in (6) with $\rho_c + \rho_s = 1$, that satisfy for some $0 < s < 1$ the bounds

$$\phi_L(X) \leq F_X(X - h\iota, X + h\iota)/h^{sq} \leq \phi_H(X) < B < \infty, \tag{20}$$

with $\phi_L(X) > 0$ a.e. F_X^s, implying that $EF_X(X - h\iota, X + h\iota) > \rho_s E_{F_X^s}(\phi_L(X))\, h^{sq} > 0$. This could hold if, e.g. the singular part is absolutely continuous on a lower-dimensional subspace (e.g. with $s = \frac{q-1}{q}$) or if F_X is a fractal distribution, e.g. Cantor-type distribution or a multivarite extension on R^q for some $s > 0$, then as $h \to 0$ the bounds (20) hold.

For a distribution satifying (20) then the rate in (i) is of order $O(h^{sq})$ and the Assumption holds. Part (ii) holds with $\bar{B} = \varepsilon^{-s}$ as long as $E\frac{\phi_H(X)}{\phi_L(X)}$ is bounded; for example, for the univariate Cantor-type distribution or its multivariate extensions when the distribution is uniform on the support set we have $\frac{\phi_H(X)}{\phi_L(X)} = 1$. The rate for the ratio in (iii) is $O(h^{2sq})$ thus (iii) also holds. Suppose that the distribution is a mixture of absolutely continuous with possibly many singular components each satisfying (20) with some s_t, for t in a set T, if then $s_{\min} = \min_{t \in T} s_t$ with $s_{\min} > 0$, then (20) holds on replacing s with $s_{\min}$; Assumption 5 holds.

In the theorem below asymptotic normality for U_n and convergence to expectation of S_n is established for any distribution function $F_X \in \mathcal{D}$.

Theorem 3 . *For $F_X \in \mathcal{D}$ under Assumptions 1-5 as $n \to \infty$*

$$\frac{nU_n}{\sqrt{2E\left(H_n^2\right)}} \underset{d}{\rightarrow} N(0,1);$$
$$\left|S_n^2 - E\left(H_n^2\right)\right| = o_p\left(E\left(H_n^2\right)\right).$$

This result shows that the U-statistic U_n converges to the Gaussian at a rate $O_p\left(n\left(E\left(H_n^2\right)\right)^{-1/2}\right)$, so that if there is e.g. a component with $s_{\min}$ $0 < s_{\min} \leq 1$, for $s = s_{\min}$ the rate of convergence is $O_p\left(nh^{-sq/2}\right)$. Applying this to a statistic that is scaled by h^{-q}, $\tilde{U}_n = \frac{1}{h^q}U_n$, we get the rate $h^{-2q}E\left(H_n^2\right)$ in the correspondingly scaled denominator, and thus the rate $O_p\left(nh^{(2-s)q/2}\right)$ for $\tilde{U}_n$. For $s = 1$ the usual rate $nh^{q/2}$ (e.g. Z-96, Lemma 3.3) applies, but when the distribution is not absolutely continuous the rate is slower.

Corollary 4 . *Under the Assumptions of the Theorem the statistic*

$$\frac{nU_n}{\sqrt{2S_n^2}} \underset{d}{\rightarrow} N(0,1).$$

Next, we consider the statistic I_n in (2) that differs from U_n in that $\hat{u}_.$ replaces $u_.$ in the expression for $H_n(z_i, z_j)$ and $\hat{\sigma}_n^2$ that similarly differs from S_n^2. The difference comes from the fact that the estimated residuals replace the model errors to form U statistics in which the $\hat{z}_i$ are no longer i.i.d.. In Lemmas 3.3(b-e) Z-96 shows that under the null the extra contribution from the discrepancy between u_i and $\hat{u}_i$ in a parametric regression model that satisfies the usual regularity conditions for the parametric model (Assumptions 2-4 of that paper) is not important enough to change the asymptotic distribution under the null. Indeed, with those assumptions for an M-estimator such as nonlinear least squares the residual vector, $\hat{u}$ is obtained by a (possibly random) projection of vector u so that $I_n = U_n\left(1 + O_p\left(n^{-1}\right)\right)$ and similarly $\hat{\sigma}_n^2 = S_n^2\left(1 + O_p\left(n^{-1}\right)\right)$.

Lemma 5 *Under the assumptions of the Theorem and with* $I_n = U_n\left(1 + O_p\left(n^{-1}\right)\right)$ *and* $\hat{\sigma}_n^2 = S_n^2\left(1 + O_p\left(n^{-1}\right)\right)$ *asymptotic normality holds:*

$$\frac{nI_n}{\sqrt{2\hat{\sigma}_n^2}} \underset{d}{\rightarrow} N(0,1).$$

Under the assumptions the asymptotic distribution of the statistic is unbiased. The next section will show that bias in the limit distribution of the test statistic may arise when F_X has mass points.

3 Bias and asymptotic bias

In finite sample generally $E\left(\hat{H}_n(\hat{z}_i, \hat{z}_j)\right)$ may not be zero. Ellison and Ellison (2000) show that in the case of testing a linear regression model with homoscedastic errors the finite sample expectation conditional on X of the numerator, I_n, converges to $-\sigma^2\left(Tr\left(P_X W_n\right)\right)$, where σ^2 is the error variance, P_X is the projection onto the space of the regressors (including a constant), W_n is the kernel weights matrix. They propose a euristic correction term, $\frac{(rank(X)+1)}{\sqrt{2\hat{\sigma}_n^2}}$, for the statistic $\frac{I_n}{\sqrt{2\hat{\sigma}_n^2}}$ that makes use of this quantity.

Generally, using the regression residuals, $\hat{u} = M_X u$ with $M_X = I - P_X$ we have (omitting the subsripts)

$$\hat{u}'W\hat{u} - u'Wu = -tr\left(uu'PW\right) - tr\left(WPuu'\right) + tr\left(uu'PWP\right).$$

Under assumptions 1 and 2 $E_{|X}\left(uu'\right)$ is a diagonal matrix, $\Lambda = \Lambda_X$, with positive diagonal entries between b and B. It follows that conditionally on X,

$$E_{|X}\left(\hat{u}'W\hat{u} - u'Wu\right) = -tr\left[PW\Lambda + WP\Lambda - PWP\Lambda\right];$$

thus bias is

$$-trE\left[PW\Lambda + WP\Lambda - PWP\Lambda\right].$$

Under homoscedasticity with $\Lambda = \sigma^2 I$, this reduces to $-\sigma^2 trE\left(\left(X'X\right)^{-1}X'WX\right)$.

When the distribution is a mixture of a continuous and discrete distribution the bias can be computed by adding the expectation over the support of the discrete part to the expectation for the continuous distribution. If the continuous part of the mixture satisfies our assumptions the corresponding part of the expectation converges to zero and thus there is no possible contribution to limit bias. To explore the asymptotic bias we examine the discrete component of the distribution, $\rho_d F_X^d$; it is supported on at most a countable number of mass points. Denoting by E_{F^d} the expectation for the discrete component of the distribution we get that the limit bias is given by

$$-\rho_d tr\left(E_{F^d}\left[PW\Lambda + WP\Lambda - PWP\Lambda\right]\right).$$

For $\Lambda = \sigma^2 I$ with a finite number of points of support, $X_k^d \in R^q$, $k = 1, \ldots n_d$, for F_X^d, the limit bias equals

$$-\rho_d K(0)\sigma^2 E_{F^d}\Sigma_{k=1}^{n_d}\Sigma_{i\neq j} I\left(X_i = X_j = X_k^d\right) trX_i\left(X'X\right)^{-1}X_j', \tag{21}$$

since as $h \rightarrow 0$ once h becomes greater than the minimal distance between points in support of the discrete part for any $X_{k_1}^d \neq X_{k_2}^d$ the value of the weight $K\left(\frac{X_{k_1}^d - X_{k_2}^d}{h}\right)$ becomes 0. As long as $K(0) > 0$ the limit bias is negative.

Thus when the distribution function F_X is discontinuous there is generally a bias in the numerator of the statistic and since the denominator for a distribution with mass points converges in probability to a bounded non-zero limit, the bias is present in the limit distribution of the statistic.

4 Power of the kernel-based test

To facilitate comparison to the results in the literature we consider in this section the scaled statistic $\tilde{I}_n = \frac{1}{h^q} I_n$ for the i.i.d. case. The results currently available apply in models with absolutely continuous F_X (typically with some further smoothness assumptions on the density). For this class of distributions the test was shown to have power against a sequence of alternative models that can be represented as

$$Y_i = g_0(X_i, \beta_0) + \gamma_n \delta(X_i) + u_i,$$

that approach the null at the rate no faster than $\gamma_n = O\left(\left(nh^{q/2}\right)^{-1/2}\right)$ (see, e.g. Z96). As remarked in FL-2000 there is thus power against a Pitman sequence with rate $\gamma_n = O\left(n^{-1/2}\right)$ for the case of a fixed bandwidth h, but with $h \to 0$ there is no power against such a sequence. For F_X with singular parts, e.g. for $F_X \in \mathcal{D}(s)$ the sequence γ_n, correspondingly, by Theorem ? can approach the null no faster than $O\left(\left(nh^{(2-s)q/2}\right)^{-1/2}\right)$ for the test to have any power. This rate is worse than $\left(nh^{q/2}\right)^{-1/2}$ thus there is no power against $n^{-1/2}$ alternatives.

It is known that local tests may have extra power against localized alternatives. The local "singular" alternative models represented as

$$Y_i = g_0(X_i, \beta_0) + \gamma_n \delta_n(X_i) + u_i, \tag{22}$$

with γ_n a deterministic sequence and $\alpha_n = \int \delta_n^2(x)\,dx \to 0$ approach the null (given by $g_0(X_i, \beta_0)$) at the rate $O(\alpha_n \gamma_n)$. Under some conditions that further specify the class of "singular" alternatives it is shown in FL-2000 that for the kernel test with $h \to 0$ there is a class of alternatives (characterized by conditions on the pairs (α_n, γ_n)) where the kernel-based test has power. Moreover the power is non-trivial for alternatives with $\gamma_n \alpha_n$ approaching zero at a rate arbitrarily close to $n^{-2/3}$.

We examine whether the result is robust against the case of a more general continuous distribution function that is not non-absolutely continuous. In particular, we show that with the absolutely continuous part satisfying all the usual assumptions even a small contamination with a singular distribution can lead to a drastic reduction in power. We make the following assumption.

Assumption 6. (a) The distribution $F_X \in \mathcal{D}(s)$ with $0 < \rho_{a.c.} < 1$, $0 < s < 1$ and the support of the singular part, F_X^s does not include $(-1, 1)$.

(b) The alternative is represented as (22) with γ_n a deterministic sequence and $\alpha_n = \int \delta_n^2(x)\,dx \to \alpha_n \to 0$; the support of δ_n is in ζ_n-neighborhood of 0 with $|\zeta_n| < 1$.

(c) Assumptions (C1)-(C4) of FL-2000 apply with $f(.)$ representing the density of the absolutely continuous component, $F_X^{a.c.}$.

The Assumption 6(a) requires F_X to have a non-trivial singular part that is supported outside of a neighborhood of zero. As a specific example for (a) consider a product measure with the absolutely continuous components with positive density on $(-2, 2)$ and a fractal singular component, $F_{X^i}^s$, e.g. given by a Cantor distribution on $[1, 2]$ with $s = \frac{\ln 2}{\ln 3}$.

The form of the alternative given in 6(b) is the same as in FL-2000, with the only difference that here we restrict the class of alternatives by requiring the support of $\delta_n(.)$ to be a neighborhood of zero; this serves to simplify the derivations. A specific representative of this class is the function suggested by FL-2000, $\delta_n(x) = I\left(x \in [0, \zeta_n]^q\right)$, or the one used in the simulations in that paper. By Assumption 6(a) the support of $\delta_n(x)$ does not intersect with the support of the singular part of F_X.

Assumption 6(c) here implies that the assumptions (C1)-(C4) of FL-2000 apply to the absolutely continuous component. If in contravention to Assumption 6(a) we had $\rho_{a.c.} = 1$, then under 6(b-c) the result of Theorem 3.1. in FL-2000 would hold and better than $n^{-1/2}$ power would be possible for the statistic $\tilde{I}_n$ over the class of alternatives that satisfy

$$\lim_{n\to\infty} \left(nh^{q/2}\gamma_n^2 \alpha_n\right) > 0; \tag{23}$$

$$\gamma_n \alpha_n = o\left(n^{-1/2}\right). \tag{24}$$

Theorem 3.1 in FL-2000 asserts that when the class of alternatives is such that (23) holds then

$$nh^{q/2}\tilde{I}_n \to_d N\left(C_1 C_2, \sigma_0^2\right),$$

with C_1, C_2, σ_0^2 positive constants and there is power against $O(\alpha_n \gamma_n)$ alternatives. The discussion in that paper (p. 1029) highlights several choices of γ_n, α_n that provide power for rates better than $n^{-1/2}$, (24). For example, selecting a subclass with e.g. $\alpha_n = c_\alpha n^{-1/2}$ and $\gamma_n = c_\gamma n^{-1/4} h^{-q/4}$ for non-zero bounded sequences c_α, c_γ leads to the rate $\alpha_n \gamma_n = o\left(n^{-1/2}\right)$; selecting $\alpha_n = c_\alpha n^{-1+\varepsilon}$ for small $\varepsilon > 0$ and $\gamma_n = c_\gamma n^{1/4}$ provides $\alpha_n \gamma_n = O\left(n^{-3/4+\varepsilon}\right)$, close to $n^{-3/4}$.

Next, we develop the limit distribution for the statistic $\tilde{I}_n$ given the singular part for F_X defined in Assumption 6(a). For this distribution by Lemma 3? the bounds on $E\left(h^{-2q} H_n(Z_1, Z_2)^2\right)$ provide that this expectation is $h^{-(2-s)q}\sigma^2$, where σ^2 is such that

$$b^2 M_1 E\left(\phi_L(X)\right) \le \sigma^2 \le B^2 M_1 E\left(\phi_H(X)\right). \tag{25}$$

Theorem 6. *Under Assumption 1-6 provided that*

$$\gamma_n^2 \alpha_n n h^{(2-s)q/2} = \tilde{C}_{1n}(s) \to \tilde{C}_1(s)$$

the statistic $\tilde{I}_n$ *converges as*

$$nh^{(2-s)q/2}\tilde{I}_n \to_d N\left(\tilde{C}(s), \sigma^2\right), \tag{26}$$

where $\tilde{C}(s) = \tilde{C}_1(s) C_2 > 0$ *if*

$$\lim_{n\to\infty}\left(\gamma_n^2 \alpha_n n h^{(2-s)q/2}\right) > 0. \tag{27}$$

We see that here the condition (27) differs from (23), and may not necessarily hold for class of alternatives with $(23-24)$. Indeed, for the same statistic with the same values for α_n, γ_n we have that $\tilde{C}_{1n}(s) = C_{1n} h^{(1-s)q/2}$; if $C_1 < \infty$, then $\tilde{C}_1(s) = 0$. Thus when F_X is contaminated with a singular distribution there is no power for the class of "singular" alternatives that provided better than $n^{-1/2}$ power in the absolutely continuous class of distributions. With the knowledge of the specific structure of the singular components of the distribution, it would have been possible to indicate a calss of alternatives where the test would have power with rate better than $n^{-1/2}$, but such power results will not be robust to slight contaminations of the distribution.

To conclude, the test against local alternatives may have power when the alternatives approach the null at a rate $O\left(\left(nh^{(2-s)q/2}\right)^{-1/2}\right)$ when $F_X \in \mathcal{D}(s)$ but not for sequences of $O\left(n^{-1/2}\right)$ alternatives. Thus over this class of distributions the local kernel-based test (with $h \to 0$) does not exhibit a power advantage over Bierens test.

5 Simulations

5.1 The mixture of normal distribution for X and size properties.

Here we consider the homoscedastic linear model under H_0

$$Y = 1 + X_1 + X_2 + e \tag{28}$$

typically examined in the literature. Instead of the usually considered distributions of the conditioning variables such as uniform or normal we use a mixture of normals with large values of density derivative: X_1, X_2 are independent observations from the normal mixture model NM:

$$F_X = 0.5N(0,1) + 0.3N(0.8, 0.01^2) + 0.2N(1.2, 0.01^2).$$

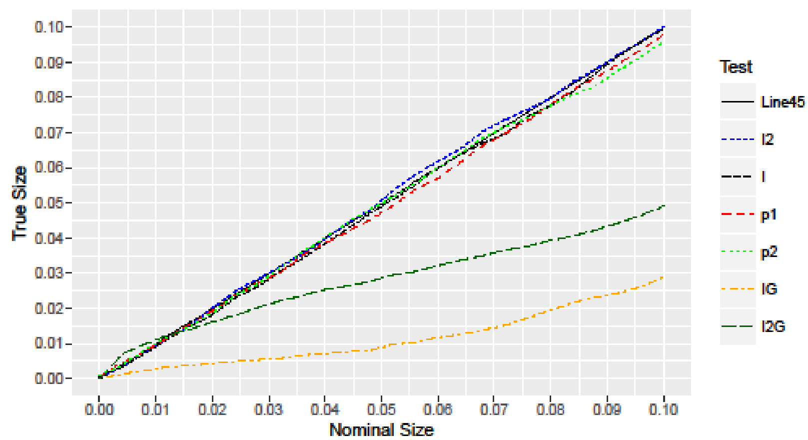


Figure (1) Figure 3. Size discrepancy.

We mainly compare our numerical results to those of FL-2000 where the X_{1i}, X_{2i} were drawn as $X_{1i} = V_i + V_{ii}$; $X_{2i} = V_i + V_{i2}$ with V_i and V_{i1}, V_{i2} independent draws from a uniform $[-\pi, \pi]$. The joint distribution that we consider is simpler in that it eliminates dependence between the two regressors, on the other hand, we focus on the irregular nature of the marginal.

Figure 3 provides evidence about test size discrepancy under the null in (28) where X_1, X_2 are NM, $u \sim N(0,1), n = 2000$. Here and below $I, I2$ denote wild bootstrapped kernel test for $h = \widehat{sd}(x)\, n^{-a}$, with $\alpha_{I,I2} = 1/6, 1/3$, correspondingly. We denote $p1, p2$ wild bootstrapped Bierens tests. $IG, I2G$ asymptotic test for kernel stats. 1499 bootstraps in each simulation.

Figure 3 demonstrates that asymptotic critical values for the most part are undersized, with the worse distortion for a larger bandwidth indicating that even at this bigger sample size (2000 as opposed to 500 in Figure 2) finite sample bias remains a significant problem for the asymptotic approximation. Bootstrap statistics all perform much better with close to the true size under the null.

5.2 Power properties with normal mixture covariates.

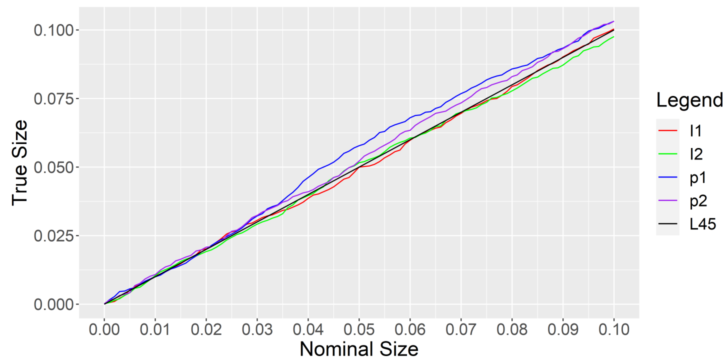
In this section we provide simulation evidence that confirms the comments in the previous section about the asymptotic performance of the two tests, kernel-based and Bierens type. We consider some of the examples investigated e.g. by FL-2000. The difference here is that the two regressors, X_1, X_2, are independently distributed as NM, with densities that exhibit very high derivatives relaive to all the simulations for hernel statistics considered in the literature. We use the same models as in FL-2000; $q = 2$.

$$H_0 : Y = 1 + X_1 + X_2 + e \quad \text{DGP1}$$

$$\begin{aligned} Y &= 1 + X_1 + X_2 + 0.15 X_1 X_2 + e \quad \text{DGP2} \\ &\quad \text{(fixed alternative)} \end{aligned}$$

$$\begin{aligned} Y &= 1 + X_1 + X_2 + \frac{1}{\sqrt{n}}(X_1^2 + X_2^2) + e \quad \text{DGP3} \\ &\quad \text{(local alternative)} \end{aligned}$$

$$\begin{aligned} Y &= 1 + X_1 + X_2 + 10 n^{-19/60} \sin(X_1 n^{1/10}) \sin(X_2 n^{1/10}) 1\{|X_1| \wedge |X_2| \le n^{-1/10}\} + e \quad \text{DGP4} \\ &\quad \text{(singular local alternative)} \end{aligned}$$



$$
\begin{aligned}
Y &= 1 + X_1 + X_2 + 0.5\sin(X_1)\sin(X_2) + e \quad \text{DGP5} \\
&\quad \text{(high frequency alternative)}
\end{aligned}
$$

Here as is standard $e \sim N(0,1)$, independent of X_1 and X_2.

Our F_X is **normal mixture** NM for each X_1, X_2; independent:

$$F_X = 0.5N(0,1) + 0.3N(0.8, 0.01^2) + 0.2N(1.2, 0.01^2)$$

The distribution exhibits high values for the derivative of density.

Following FL-2000, let I_1 and I_2 represent kernel statistics with a bandwidth choice of $h_i = 0.5n^{-1/6}\hat{sd}(X_i)$ and $h_i = n^{-1/6}\hat{sd}(X_i)$, respectively. For the Bierens statistics, we use

$$
\begin{aligned}
p_i &= \frac{1}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\int \hat{u}_j \hat{u}_k \exp(\zeta'(Z_j + Z_k))d\mu_i(\zeta) \qquad i = 1, 2 \\
Z_j &= \arctan\left(\frac{X_j - \bar{X}}{\hat{sd}(X)}\right)
\end{aligned}
$$

where μ_1 is a product $U[-1,1]$ distribution and μ_2 is a product $N(0,1)$ distribution. We consider a sample size of $n = 200$.

Wild bootstrap test.

Size.

As the figure indicates, the wild bootstrap controls size remarkably well for all the statistics, even at a relatively low sample size of $n = 200$.

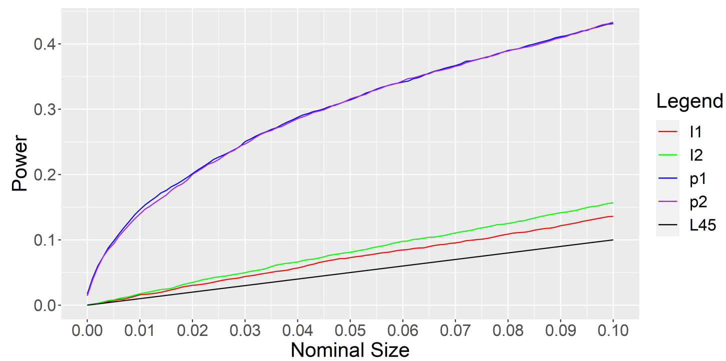
Power against DGP2 (fixed alternative) for NM and $n = 500$. Fan&Li: with $x's$ given by sums of uniforms all the tests had similar power. Here the Bierens-type tests clearly dominate in terms of power.

Power against DGP3 (local alternative), NM, $n = 500$. Fan&Li: kernel tests had trivial power that declined with n; power of the $p1, p2$ tests increased. Here same for I,I2 tests; Bierens test has some power, that for p1 declined slightly, for p2 unchanged.

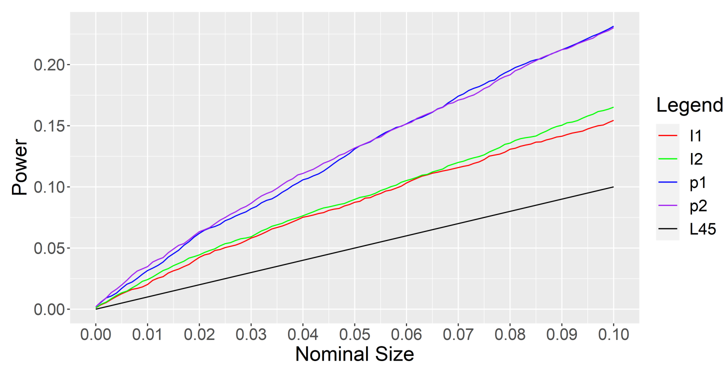
In Fan and Li the observation was that the kernel test I had only trivial power that declined with n, while the power of the $p1, p2$ tests increased. Here the I,I2 tests do have trivial power and their power declines as n increases, and the Bierens test has some discriminatory power that does not change much as sample size increases.

Power curves for DGP4 (spiked alternative). $x_1, x_2 \sim NM$; $n = 500$; $I, I2$ wild bootstrapped kernel test ($h = \widehat{sd}\,(x)\, n^{-a}$, $\alpha_{I,I2} = 1/6, 1/3$), $p1, p2$ wild bootstrapped Bierens tests.

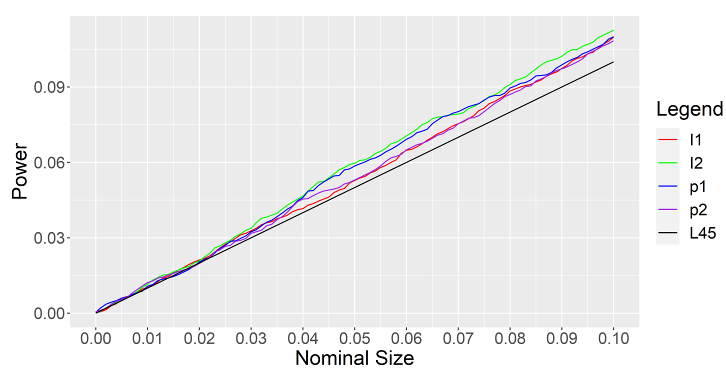
Power against DGP 5



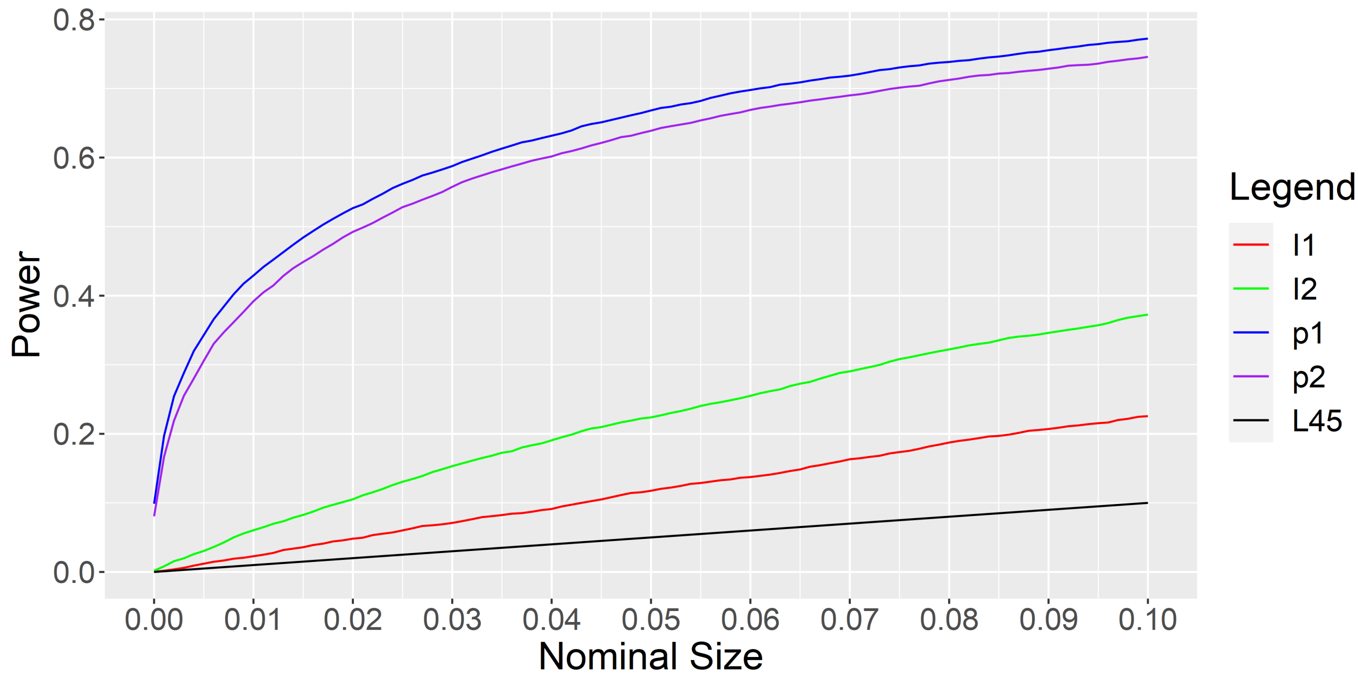
Power
Nominal Size
Legend
l1
l2
p1
p2
L45



Power
Nominal Size
Legend
l1
l2
p1
p2
L45



Power
Nominal Size
Legend
l1
l2
p1
p2
L45



6 Appendix A. Proofs.

Lemma 6 *Axilliary Lemma. Let $f : \mathbb{R}^q \to \mathbb{R}$ be bounded and $g : \mathbb{R}^q \to \mathbb{R}$ be a function with ∂g a bounded function with support contained in $\prod_{i=1}^{q}[l_i, u_i]$ Then :*

$$\int_{\mathbb{R}^q} f(x)g(x)dF_X(x) = (-1)^q \int_{\mathbb{R}^q} \Omega_f(t,l)\partial g(t)dt \tag{29}$$

where $\Omega_f(t,l)$ is the function

$$\Omega_f(t,l) = \int_{\mathbb{R}^q} f(x) \prod_{i=1}^{q} 1_{l_i \le x_i \le t_i} dF_X(x) \tag{30}$$

In particular, for fixed $z, y \in \mathbb{R}^q$ and $h > 0$ this implies that :

$$\int_{\mathbb{R}^q} f(x)K^2\left(\frac{x-z}{h}\right) dF_X(x) = (-1)^q \int_{\mathbb{R}^q} \Omega_f(z - h\iota, z + hv)\partial_v K^2(v)dv \tag{31}$$

$$\int_{\mathbb{R}^q} f(x)K\left(\frac{x-z}{h}\right) K\left(\frac{x-y}{h}\right) dF_X(x) = (-1)^q \int_{\mathbb{R}^q} \Omega_f(z - h\iota, z + hv)\partial_v \left[K(v)K\left(\frac{z-y}{h} + v\right)\right] dv \tag{32}$$

Proof.

Since f is bounded, the function $A \to \int_A f(x)dF_X(x)$ represents a finite Borel measure on $\mathbb{R}^q$ which we will refer to as Ω_f. With this notation, it follows that :

$$\int_{\mathbb{R}^q} f(x)g(x)dF_X(x) = \int_{\mathbb{R}^q} g(x)d\Omega_f(x)$$

Given the support of g, this can be expressed as :

$$\int_{\mathbb{R}^q} g(x) \prod_{i=1}^{q} 1_{l_i \leq x_i \leq u_i} d\Omega_f(x)$$

For any $x \in \prod_{i=1}^{q}[l_i, u_i]$, we have by the fundamental theorem of calculus

$$g(x) = (-1)^q \int_{x_n}^{u_n} \dots \int_{x_1}^{u_1} \partial g(t) dt_1 \dots dt_n$$

and so it follows that :

$$\int_{\mathbb{R}^q} g(x) d\Omega_f(x) = (-1)^q \int_{\mathbb{R}^q} \int_{\mathbb{R}^q} \partial g(t) \prod_{i=1}^{q} 1_{l_i \leq x_i \leq u_i} 1_{x_i \leq t_i \leq u_i} dt d\Omega_f(x)$$

Fubini's Theorem now implies that :

$$\int_{\mathbb{R}^q} g(x) d\Omega_f(x) = (-1)^q \int_{\mathbb{R}^q} \Omega_f(t, l) \partial_t g(t) dt \tag{33}$$

where $\Omega_f(t, l)$ is the function

$$\Omega_f(t, l) = \int_{\mathbb{R}^q} \prod_{i=1}^{q} 1_{l_i \leq x_i \leq u_i} 1_{x_i \leq t_i \leq u_i} d\Omega_f(x) = \Omega_f\big(\{x \in \mathbb{R}^q : l_i \leq x_i \leq t_i \ \ \forall i\}\big)$$

Substitution of $g(x) = K^2((x - z)h^{-1})$, which by Assumption 3 has support contained in $\prod_{i=1}^{q}[z_i - h, z_i + h]$, into (33) and using (30) gives :

$$\int_{\mathbb{R}^q} f(x) K^2\Big(\frac{x - z}{h}\Big) dF_X(x) = (-1)^q \int_{\mathbb{R}^q} \Omega_f(z - h\iota, t) \partial_t K^2\Big(\frac{t - z}{h}\Big) dt$$

The change of variables $t_i \to z_i + hv_i$ now yields (**??**):

$$\int_{\mathbb{R}^q} f(x) K^2\Big(\frac{x - z}{h}\Big) dF_X(x) = (-1)^q \int_{\mathbb{R}^q} \Omega_f(z - h\iota, X + hv) \partial_v K^2(v) dv$$

Likewise, substitution of $g(x) = K\big((x-z)h^{-1}\big) K\big((x-y)h^{-1}\big)$ which has support contained in $\prod_{i=1}^{q}[z_i - h\iota, z_i + h\iota]$ yields :

$$\int_{\mathbb{R}^q} f(x) K\left(\frac{x - z}{h}\right) K\left(\frac{x - y}{h}\right) dF_X(x) = (-1)^q \int_{\mathbb{R}^q} \Omega_f(z - h\iota, t) \partial_t \Big[K\Big(\frac{t - z}{h}\Big) K\Big(\frac{t - y}{h}\Big)\Big] dt$$

and the desired result (32) follows from the change of variables $t_i \to z_i + hv_i$. ■

Proof of Lemma 1. Let Z_1, Z_2, Z_3 be independent copies of (X, u) and define :

$$H_n(Z_1, Z_2) = u_1 u_2 K\Big(\frac{X_1 - X_2}{h}\Big) \qquad G_n(Z_1, Z_2) = \mathbb{E}\Big[H_n(Z_3, Z_1) H_n(Z_3, Z_2) \Big| Z_1, Z_2\Big].$$

From Assumptions 1-3 (11) immediately follows. We will now derive the expression for $\mathbb{E}\big(|G_n(Z_1, Z_2)|^2\big)$. The definition of the conditional expectation and independence implies that :

$$\mathbb{E}\big(|G_n(Z_1, Z_2)|^2\big) = \mathbb{E}\Big(\mu_2(X_1)\mu_2(X_2)\Big[\int_{\mathbb{R}^q} \mu_2(x) K\Big(\frac{x - X_1}{h}\Big) K\Big(\frac{x - X_2}{h}\Big) dF_X(x)\Big]^2\Big)$$

where $\mu_2(X) = \mathbb{E}[u^2|X]$. By (32) of Lemma 6 applied to the expression in squared brackets, this evaluates to :

$$\mathbb{E}\big(\,|G_n(Z_1,Z_2)|^2\,\big) = \mathbb{E}\bigg(\mu_2(X_1)\mu_2(X_2)\bigg[\int_{\mathbb{R}^q}\Omega_2(X_1-h\iota,X_1-hv)\partial_v\bigg[K(v)K\bigg(\frac{X_1-X_2}{h}+v\bigg)\bigg]dv\bigg]^2\bigg)$$

Using (29) of Lemma 6 with the function $f(x_2)=\mu_2(x_2)$ and

$$g(x_2) = \bigg[\int_{\mathbb{R}^q}\Omega_2(X_1-h\iota,X_1-hv)\partial_v\bigg[K(v)K\bigg(\frac{X_1-x_2}{h}+v\bigg)\bigg]dv\bigg]^2$$

which has support contained in $\prod_{i=1}^q[X_1^i,X_1^i+2h]$ implies that

$$\mathbb{E}\big(\,|G_n(Z_1,Z_2)|^2\,\big) = \mathbb{E}\bigg((-1)^q\mu_2(X_1)\int_{\mathbb{R}^q}\Omega_2(t,X_1)\partial_t\bigg[\int_{\mathbb{R}^q}\Omega_2(X_1-h\iota,X_1-hv)\partial_v\bigg[K(v)K\bigg(\frac{X_1-t}{h}+v\bigg)\bigg]dv\bigg]^2dt\bigg)$$

The desired result (14) now follows from the change of variables $u=(t-X_1)h^{-1}$

$$\mathbb{E}\big(\,|G_n(Z_1,Z_2)|^2\,\big) = \mathbb{E}\bigg((-1)^q\mu_2(X)\int_{\mathbb{R}^q_+}\Omega_2(X,X+hv)\partial_u\bigg[\int_{\mathbb{R}^q}\Omega_2(X-h\iota,X-hv)\partial_v\big[K(v)K(v+u)\big]dv\bigg]^2du\bigg)$$

A similar argument shows that :

$$\begin{aligned}\mathbb{E}\big(\,|H_n(Z_1,Z_2)|^2\,\big) = \mathbb{E}\bigg(\mu_2(X_1)\mu_2(X_2)K^2\bigg(\frac{X_1-X_2}{h}\bigg)\bigg) &= \mathbb{E}\bigg(\mu_2(X)(-1)^q\int_{\mathbb{R}^q}\Omega_2(X-h\iota,X-hv)\partial_vK^2(v)dv\bigg)\\ &= \mathbb{E}\bigg(\mu_2(X)(-1)^q\int_{[-1,1]^q}\Omega_2(X-h\iota,X-hv)\partial_vK^2(v)dv\bigg)\end{aligned}$$

to provide (12) and for (13) we get

$$\mathbb{E}\big(\,|H_n(Z_1,Z_2)|^4\,\big) = \mathbb{E}\bigg(\mu_4(X)(-1)^q\int_{\mathbb{R}^q}\Omega_4(X-h\iota,X-hv)\partial_vK^4(v)dv\bigg).$$

When X has a density f_X with respect to the Lebesgue measure :

$$\Omega_2(x+hv,x-h) = \int_{x_q-h}^{x_q+hv_q}\cdots\int_{x_1-h}^{x_1+hv_1}\mu_2(y_1,\ldots,y_q)f_X(y_1,\ldots,y_q)dy$$

where x_i is the i^{th} coordinate of $x\in\mathbb{R}^q$. For any fixed $v\in[-1,1]^q$, Theorem 7.10 in Rudin (2006) and the fact that $\mu_2(.)$ is bounded from above implies that :

$$h^{-q}\Omega_2(x+hv,x-h)\xrightarrow[h\to0^+]{}\mu_2(x)f_X(x)\prod_{i=1}^q(1+v_i)$$

at almost every x (with respect to the Lebesgue measure on $\mathbb{R}^q$). Since the set of of points for which this is not true is also a null set under F_X (when X has a density), an application of the dominated convergence theorem yields :

$$h^{-q}\mathbb{E}\big(\,|H_n(Z_1,Z_2)|^2\xrightarrow[h\to0^+]{}\mathbb{E}\bigg((-1)^q\mu_2^2(X)f_X(X)\int_{\mathbb{R}^q}\big[\partial_vK^2(v)\big]\prod_{i=1}^q(v_i+1)dv\bigg)$$

$$=\quad h^{-q}\mathbb{E}\big(\,|H_n(Z_1,Z_2)|^2\xrightarrow[h\to0^+]{}\mathbb{E}\bigg((-1)^q\mu_2^2(X)f_X(X)\int_{\mathbb{R}^q}\big[\partial_vK^2(v)\big]\prod_{i=1}^q v_idv\bigg)$$

where the last equality follows from the fact that under Assumption 3(a-b) $\frac{d}{dv_i}k^2(v_i) = -\frac{d}{dv_i}k^2(-v_i)$, thus $\int_{\mathbb{R}^q}\left[\partial_v K^2(v)\right]dv = 0$.

Before considering the case where X does not admit a density, we first do some preliminary calculations. Define (v_1, v_{-1}) to be the partitioned vector $(v_1, \ldots, v_n)$ with $v_{-1} = (v_2, ..., v_q)$. For any fixed choice of $v_{-1} \in \mathbb{R}^{q-1}$, we have that :

$$\begin{aligned}
&\int_{[-1,1]} \Omega_2(X - h\iota, X + h(v_1, v_{-1}))\partial_{v_1} k^2(v_1)dv_1 \\
&= \int_0^1 \Omega_2(X - h\iota, X + h(v_1, v_{-1}))\partial_{v_1} k^2(v_1)dv_1 + \int_{-1}^0 \Omega_2(X - h\iota, X + h(v_1, v_{-1}))\partial_{v_1} k^2(v_1)dv_1 \\
&= \int_0^1 \Omega_2(X - h\iota, X + h(v_1, v_{-1}))\partial_{v_1} k^2(v_1)dv_1 + \int_0^1 \Omega_2(X - h\iota, X + h(-v_1, v_{-1}))\partial_{v_1} k^2(-v_1)dv_1 \\
&= \int_0^1 \Omega_2(X - h\iota, X + h(v_1, v_{-1}))\partial_{v_1} k^2(v_1)dv_1 - \int_0^1 \Omega_2(X - h\iota, X + h(-v_1, v_{-1}))\partial_{v_1} k^2(v_1)dv_1 \\
&= \int_0^1 \Big[\Omega_2(X - h\iota, X + h(v_1, v_{-1})) - \Omega_2(X - h\iota, X + h(-v_1, v_{-1}))\Big]\partial_{v_1} k^2(v_1)dv_1 \\
&= \int_0^1 \Big[\Omega_2(X + h(-v_1, v_{-1}), X + h(v_1, v_{-1}))\Big]\partial_{v_1} k^2(v_1)dv_1
\end{aligned}$$

where the second to last equality follows from the fact that $\frac{d}{dv_1}k^2(-v_1) = -\frac{d}{dv_1}k^2(v_1)$ and the last from the fact that the sets defined by $\{X - h\iota, X + h(v_1, v_{-1})\}$ and $\{X - h\iota, X + h(-v_1, v_{-1}))\}$ intersect on $\{X + h(-v_1, v_{-1}), X + h(v_1, v_{-1})\}$ and the linearity of the measure. Iterating this procedure from v_1 to v_q shows that :

$$\int_{[-1,1]^q} \Omega_2(X - h\iota, X - hv)\partial_v K^2(v)dv = \int_{[0,1]^q} \Big(\Omega_2\big(X + hv, X - hv\big)\Big)\partial_v K^2(v)dv$$

Since $\partial_{v_i} k^2(v_i) = -\partial_{v_i} k^2(-v_i)$ and $\partial_{v_i} k^2(-v_i) \geq 0$ on $[0,1]$, it follows that we can express $\mathbb{E}\big(|H_n(Z_1, Z_2)|^2\big)$ through a non-negative integrand via :

$$\mathbb{E}\big(|H_n(Z_1, Z_2)|^2\big) = \mathbb{E}\Big(\mu_2(X)\int_{[0,1]^q}\Big(\Omega_2\big(X + hv, X - hv\big)\Big)\partial_v K^2(-v)dv\Big) \tag{34}$$

Let $F_X = F_X^a + F_X^s$ refer to the Lebesgue Radon-Nikodym decomposition of F_X where F_X^a and F_X^s are respectively, absolutely continuous and singular with respect to the Lebesgue measure on $\mathbb{R}^q$. Suppose F_X is not absolutely continuous with respect to the Lebesgue measure so that F_X^s is a non trivial measure. Let $\mathbb{E}_{F_X^s}$ represent integration with respect to F_X^s so that by using the same steps as given above we get that :

$$\begin{aligned}
\mathbb{E}\big(|H_n(Z_1, Z_2)|^2\big) &= \mathbb{E}\Big(\mu_2(X_1)\mu_2(X_2)K^2\Big(\frac{X_1 - X_2}{h}\Big)\Big) \\
&\geq \mathbb{E}_{F_X^s}\Big(\mu_2(X_1)\mu_2(X_2)K^2\Big(\frac{X_1 - X_2}{h}\Big)\Big) \\
&= \mathbb{E}_{F_X^s}\Big(\mu_2(X)\int_{[0,1]^q}\Big(\Omega_2^s\big(X - hv, X + hv\big)\Big)\partial_v K^2(-v)dv\Big)
\end{aligned}$$

where Ω_2^s is defined the same way as Ω_2 but with the measure $d\Omega_2^s = \mu_2(.)dF_X^s$ (which is also singular to the Lebesgue measure since $\mu_2(.)$ is bounded away from 0). For any fixed $v \in (0,1)^q$, Theorem 7.15 in Rudin (2006) implies that :

$$h^{-q}\Omega_2^s(x - h\iota, x + h\epsilon v) \xrightarrow[h \to 0^+]{} +\infty$$

for almost every x (with respect to the measure $d\Omega_2^s$). Since $\mu_2(.)$ is bounded away from 0, a null set under $d\Omega_2^s$ is also a null set under dF_X^s and so Fatou's Lemma implies that :

$$\begin{aligned}\liminf_{h\to 0^+} h^{-q}\mathbb{E}\big(\,|H_n(Z_1,Z_2)|^2\,\big) &\geq \liminf_{h\to 0^+} h^{-q}\mathbb{E}_{F_X^s}\left(\mu_2(X)\int_{[0,1]^q}\left(\Omega_2^s\big(X-hv,X+hv\big)\right)\partial_v K^2(-v)dv\right)\\ &\geq \mathbb{E}_{F_X^s}\left(\mu_2(X)\int_{[0,1]^q}\left(\liminf_{h\to 0^+} h^{-q}\Omega_2^s\big(X-hv,X+hv\big)\right)\partial_v K^2(-v)dv\right)\\ &= +\infty\end{aligned}$$

■

Proof of Lemma 2.3.

By (34) of Lemma 2.2, using monotonicity of $\Omega_2(X-hv,X+hv)$, the fact that $D_1(v) = (-1)^q\,\partial_v K^2(v)\geq 0$ for positive v and Assumption 2(b) we get a lower bound

$$\begin{aligned}&\mathbb{E}\big(\,|H_n(Z_1,Z_2)|^2\,\big)\\ =\;&\mathbb{E}\left(\mu_2(X)(-1)^q\int_{\mathbb{R}_+^q}\Omega_2(X-hv,X+hv)\partial_v K^2(v)dv\right)\\ =\;&\mathbb{E}\left(\mu_2(X)\int_{\mathbb{R}_+^q}\Omega_2(X-hv,X+hv)D_1(v)dv\right)\\ \geq\;&\mathbb{E}\left(\mu_2(X)\int_{\mathbb{R}_+^q}\Omega_2(X-hv,X+hv)D_1(v)I\left(v:v^i>\varepsilon,i=1,...,q\right)dv\right)\\ \geq\;&M_1\mathbb{E}\bigg(\mu_2(X)\left[\Omega_2(X-(h\varepsilon)\,\iota,X+(h\varepsilon)\,\iota)\right]\bigg)\\ \geq\;&b^2M_1\mathbb{E}\bigg(\left[F_X(X-(h\varepsilon)\,\iota,X+(h\varepsilon)\,\iota)\right]\bigg),\end{aligned}$$

where $M_1=\int D_1(v)I\left(v:v^i>\varepsilon,i=1,...,q\right)dv$, and the last inequality arises from the definition of $\Omega_2\,(.,.)$ and the lower bound on $\mu_2\,(.)$.

Un upper bound on $\left|\mathbb{E}\big(H_n(Z_1,Z_2)^4\big)\right|$ is obtained by considering $D_4\,(v)$ and bounding the integral over $[-1,1]^q$ by 2^q integrals over $[0,1]^q$.

$$\begin{aligned}&\mathbb{E}\big(\,|H_n(Z_1,Z_2)|^4\,\big)=\mathbb{E}\left(\mu_4(X)\int_{\mathbb{R}^q}\Omega_4(X-h\iota,X-hv)\left|\partial_v K^4(v)\right|dv\right).\\ \leq\;&2^q\mathbb{E}\left(\mu_4(X)\int_{\mathbb{R}_+^q}\Omega_4(X+h,X-h)D_4\,(v)\,dv\right)\\ \leq\;&2^qM_4\mathbb{E}\bigg(\mu_4(X)\left[\Omega_4(X+h,X-h)\right]\bigg)\\ \leq\;&2^qB^2M_4\mathbb{E}\bigg(\left[F_X(X+h,X-h)\right]\bigg).\end{aligned}$$

with $M_4=\int_{\mathbb{R}_+^q}D_4(v)dv$. The upper bound on $\left|\mathbb{E}\big(H_n(Z_1,Z_2)^2\big)\right|$ is similar.

In the expression for $\mathbb{E}\big(\,|G_n(Z_1,Z_2)|^2\,\big)$ consider

$$\begin{aligned}&\partial_u\left[\int_{\mathbb{R}^q}\Omega_2(X-h\iota,X-hv)\partial_v\big[K(v)K(v+u)\big]dv\right]^2\\ =\;&2\left[\int_{\mathbb{R}^q}\Omega_2(X-h\iota,X-hv)\partial_v\big[K(v)K(v+u)\big]dv\right]\\ &\times\left[\int_{\mathbb{R}^q}\Omega_2(X-h\iota,X-hv)\partial_v\big[K(v)\partial_u K(v+u)\big]dv\right].\end{aligned}$$

Bound on the first factor:

$$\begin{aligned} & \left| \int_{\mathbb{R}^q} \Omega_2(X-h\iota, X-hv)\partial_v \left[K(v)K(v+u)\right] dv \right| \\ \leq\ & 2^q B \int_{\mathbb{R}^q_+} D_2\left(v_1\right) dv_1 \mathbb{E}\Big(\left[F_X(X+h, X-h)\right] \Big); \end{aligned}$$

on the second:

$$\begin{aligned} & \left| \int_{\mathbb{R}^q} \Omega_2(X-h\iota, X-hv)\partial_v \left[K(v)\partial_u K(v+u)\right] dv \right| \\ \leq\ & 2^q B \int_{\mathbb{R}^q_+} D_3\left(v_1\right) dv_1 \mathbb{E}\Big(\left[F_X(X+h, X-h)\right] \Big). \end{aligned}$$

Combining and completing with the bound

$$\mu_2\left(X\right) \Omega_2(X, X+hv) \leq B^2 F_X(X-h\iota, X+h\iota)$$

we obtain

$$\begin{aligned} & \mathbb{E}\big(\left|G_n(Z_1, Z_2)\right|^2 \big) \\ \leq\ & 2^{2q} B^4 M_{2,3} \mathbb{E}\Big(\left[F_X(X-h\iota, X+h\iota)\right]^3 \Big) \end{aligned}$$

where $M_{2,3} = \int_{\mathbb{R}^q_+} \left[\int_{\mathbb{R}^q_+} D_2\left(v_1\right) dv_1 \right] D_3\left(v_2\right) dv_2$.

Proof of Theorem 2.3.

By Theorem 1 of Hall (1984) a sufficient condition for asymptotic normality requires that the moments satisfy the conditions:

$$E\left(H_n\left(z_1, z_2\right) | z_1\right) = 0 \text{ a.s.}; \tag{35}$$

$$E\left(H_n\left(z_1, z_2\right)\right)^2 < \infty \text{ for all } n; \tag{36}$$

$$\frac{E\left[E\left(G_n^2\left(z_1, z_2\right)\right)\right] + \frac{1}{n}E\left(H_n\left(z_1, z_2\right)\right)^4}{\left\{E\left(H_n\left(z_1, z_2\right)\right)^2\right\}^2} \rightarrow 0 \text{ as } n \rightarrow \infty, \tag{37}$$

We next verify that under our assumptions that conditions of that theorem hold when $EH_n^2\left(z_i, z_j\right)$ goes to zero as $h \rightarrow 0$. Condition (35) is (11) of Lemma 1. As well, (36) holds by Assumption 5 (i). All that is needed is to show as $n \rightarrow \infty$

$$(a) \frac{E\left[E\left(G_n^2\left(z_1, z_2\right)\right)\right]}{\left\{E\left(H_n\left(z_1, z_2\right)\right)^2\right\}^2} \rightarrow 0, \tag{38}$$

$$(b) \frac{\frac{1}{n}E\left(H_n\left(z_1, z_2\right)\right)^4}{\left\{E\left(H_n\left(z_1, z_2\right)\right)^2\right\}^2} \rightarrow 0. \tag{39}$$

In (38) by Lemma 2 (iii) the numerator is bounded from above by $2^{2q}B^4 M_{2,3} E\left(\left|F\left(X-h\iota, X+h\iota\right)\right|^3\right)$. By Lemma 2(i) the denominator has a lower bound $b^4 M_1^2 \left\{E\left(\left|F\left(X-\varepsilon h\iota, X+\varepsilon h\iota\right)\right|\right)\right\}^2$. By Assumption 5 (ii) $E\left(\left|F\left(X-\varepsilon h\iota, X+\varepsilon h\iota\right)\right|\right) \geq \frac{1}{B}E\left(\left|F\left(X-h\iota, X+h\iota\right)\right|\right)$. Then the ratio in (38) is bounded from above by

$$\frac{2^{2q}B^4M_{2,3}E\left(\left|F\left(X-h\iota,X+h\iota\right)\right|^3\right)}{\left(b^4M_1^2/\bar{B}\right)\left\{E\left(\left|F\left(X-h\iota,X+h\iota\right)\right|\right)\right\}^2}.$$

By Assumption 5(iii) then this goes to zero and (38) holds.

For (39) similarly using Lemma 2 and Assumption 5(ii) we bound the ratio from above with

$$\frac{\frac{1}{n}2^qB^2M_4E\left(\left|F\left(X-h\iota,X+h\iota\right)\right|\right)}{\left(b^4M_1^2/\bar{B}\right)\left\{E\left(\left|F\left(X-h\iota,X+h\iota\right)\right|\right)\right\}^2}=\frac{2^qB^2M_4}{\left(b^4M_1^2/\bar{B}\right)nE\left(\left|F\left(X-h\iota,X+h\iota\right)\right|\right)}.$$

By $(15, 16)$ of Lemma 1 regardless of whether F_X is absolutely continuous or has singular components we have that $h^{-q}E\left(H_n\left(z_1,z_2\right)\right)^2>c>0$. implying that $h^{-q}E\left(\left|F\left(X-h\iota,X+h\iota\right)\right|\right)>\xi>0$. Then

$$nE\left(\left|F\left(X-h\iota,X+h\iota\right)\right|\right)=nh^q(h^{-q}E\left(\left|F\left(X-h\iota,X+h\iota\right)\right|\right))>nh^q\xi.$$

By Assumption 3 (c) $nh^q\to\infty$. Thus (39) follows.

Proof of proposition 6.

Similarly to the proof of Theorem 3.1 in FL-2000 we focus on the orders in probability of the terms in

$$\begin{aligned}\tilde{I}_n &= \frac{1}{n\left(n-1\right)h^q}\Sigma_i\Sigma_{j\neq i}\hat{u}_i\hat{u}_jK\left(\frac{X_i-X_j}{h}\right)\\ &= A_1+A_2+A_3+2\left(A_4+A_5+A_6\right),\end{aligned}$$

(i) We obtain similarly to FL-2000 that

$$A_1=\frac{1}{n\left(n-1\right)}\Sigma_i\Sigma_{j\neq i}\left(\hat{g}-g_o\right)_i\left(\hat{g}-g_o\right)_jK\left(\frac{X_i-X_j}{h}\right)=O_p\left(n^{-1}\right),$$

then

$$nh^{(2-s)q/2}A_1=O_p\left(h^{(2-s)q/2}\right)=o_p\left(1\right).$$

(ii) For

$$A_2=\frac{\gamma_n^2}{n\left(n-1\right)h_n^q}\Sigma_i\Sigma_{j\neq i}\delta_n\left(X_i\right)\delta_n\left(X_j\right)K\left(\frac{X_i-X_j}{h}\right)$$

noting that the computation of moments is restricted to the neighborhood of zero we get similarly to FL-2000 but with normalization suitable to the presence of singularity with s

$$\begin{aligned}E\left(nh^{(2-s)q/2}A_2\right) &= nh^{(2-s)q/2}\gamma_n^2\alpha_n\left(C_2+o(1)\right);\\ var\left(nh^{(2-s)q/2}A_2\right) &= O(nh^{(2-s)q}\gamma^4\alpha)\\ &= O\left(\gamma^2\alpha nh^{(2-s)q/2}\right)^2\left(\alpha^{-1}h^q\right)\left(nh^q\right)^{-1}=o\left(1\right).\end{aligned}$$

Then $nh^{(2-s)q/2}A_2\to_p\tilde{C}\left(s\right)=C_2\tilde{C}_1$ provided that

$$\gamma^2\alpha nh^{(2-s)q/2}=\tilde{C}_{1n}\to\tilde{C}_1. \tag{40}$$

(iii) For the term

$$A_3=\frac{1}{n\left(n-1\right)h_n^q}\Sigma_i\Sigma_{j\neq i}K_i\left(\frac{X_i-X_j}{h}\right)\hat{u}_i\hat{u}_j$$

from Theorem 3 and Lemma 5 of this paper it follows that

$$nh^{(2-s)q/2}A_3 \to_d N\left(0,\sigma^2\right)$$

with $\sigma^2 > 0$ that satisfies (25).

(iv) The term

$$A_4 = \left(\hat{\beta} - \beta_0\right)' \left\{ \frac{\gamma_n}{(n(n-1)h^d} \Sigma_i \Sigma_{j \neq i} \partial g_0 (X_i, \beta_*) \delta_n (X_j) K \left(\frac{X_i - X_j}{h} \right) \right\}$$

is supported in a neighborhood of zero and provides as in FL-2000

$$A_4 = O_p\left(n^{-1/2}\gamma_n \alpha_n\right),$$

thus

$$nh^{(2-s)q/2}A_4 = O_p\left(\gamma_n \alpha_n n^{1/2} h^{(2-s)q/2}\right) = o_p(1).$$

(v) For the term

$$A_5 = \frac{1}{n(n-1)h_n^q} \Sigma_i \Sigma_{j \neq i} K_i \left(\frac{X_i - X_j}{h} \right) \left[g_0 (X_i, \beta_0) - \bar{g}(X_i) \right] u_j$$

computation of moments differs from those in the proof of FL-2000, as it has to take into account the singular part of the distribution. Moments are computed similarly to the derivations in the proof of Lemma 1 here on considering $\Omega_{\Delta g} = \int \Delta g (X, \beta_.) dF_X$. These computations yield

$$A_5 = O_p\left(n^{-1}h^{(s-1)q/2}\right),$$

then

$$nh^{(2-s)q/2}A_5 = O_p\left(h^{(2-s)q/2}h^{(s-1)q/2}\right) = O_p\left(h^{q/2}\right) = o_p(1).$$

(vi) Since the term A_6 is defined on the neighborhood of 0 only, the result in FL-2000 holds and provides

$$A_6 = O_p\left(n^{-1/2}\alpha^{1/2}\gamma\right);$$

where under the condition (40)

$$\begin{aligned} nh^{(2-s)q/2}A_6 &= O_p\left(n^{1/2}h^{(2-s)q/2}\alpha^{1/2}\gamma\right) \\ &= O_p\left(h^{(2-s)q/4}\right) = o_p(1). \end{aligned}$$

On combining these results we obtain (26).

■

References

[1] Alberti, G. & Marchese, 2016, On the differentiability of Lipschitz functions with respect to measures in the Euclidean space, A. Geom. Funct. Anal. 26, 1–66.

[2] Bierens, H. 1982, Consistent model specification tests, Journal of Econometrics, 20, 105-134.

[3] Bierens, H. and Ploberger, W., 1997, Asymptotic theory of integrated conditional moment tests, Journal of Econometrics, 65, 1129-1151.

[4] Bierens, H., 2017, Econometric Model Specification: consistent model specification tests and semi-nonparametric modeling and inference, World Scientific.

[5] Calonico, S., M. D. Cattaneo and M. H. Farrell, 2018, On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference, Journal of the American Statistical Association, 767-7791.

[6] Ellison, G. and S.F. Ellison, 2000, A simple framework for nonparametric specification testing, Journal of Econometrics, 96, 1-28.

[7] J.C. Escanciano a, S.C. Goh, 2014, Specification analysis of linear quantile models, Journal of Econometrics 178, 495–507.

[8] Fan, Y. and Q. Li, 2000,Consistent Model Specification Tests: Nonparametric Versus Bierens' Tests, Econometric Theory, 16, 1016-1041.

[9] Gao, J., M.King, Z. Lu and D. Tjostheem, 2009, Specification testing in nonlinear and nonstationary time series autoregression, The Annals of Statistics, 37, 3893-3928.

[10] Gelfand, I.M. & G.E. Shilov (1964a) Generalized Functions, Volume 1, Properties and Operations, Academic Press.

[11] Gelfand, I.M. & G.E. Shilov (1964b) Generalized Functions, Volume 2, Spaces of Test Functions and Generalized Functions. Academic Press.

[12] Hall, P., 1984, Central limit theorem for integrated square error of multivariate nonparametric density estimators, Journal of Multivariate Analysis 14, 1-16.

[13] Hoeffding, W. (1948) A class of statistics with asymptotically normal distributions. Annals of Statistics, 19:293–325.

[14] Hong, Y. and H. White, 1995, Consistent specification testing via nonparametric series regression, Econometrica, 63, 1133-1159.

[15] Kasparis I., Andreoua, E. and P. C.B. Phillips, 2015, Nonparametric predictive regression, Journal of Econometrics 185, 468–494

[16] Kotlyarova, Yulia, Marcia Schafgans and Victoria Zinde-Walsh, 2016, Exploration of Smoothness: Bias and Efficiency of Nonparametric Kernel Estimators in Carter Hill, Gloria Gosalez-Riviera and Tae-Hwe Lee (eds.) Advances in Econometrics, Vol. 36, Essays in Honor of Aman Ullah, 561-589.

[17] Li, Q. and J. Racine, 2007, Nonparametric Econometrics: Theory and Practice, Princeton University Press.

[18] Li, Q. and S. Wang , 1998, A simple consistent bootstrap test for a parametric regression function, Journal of Econometrics, 87, 145-165.

[19] Lin Z., Qi Li, and Yiguo Sun, 2015, A consistent nonparametric test of parametric regression functional form in fixed effects panel data models,Journal of Econometrics 178, 167–179.

[20] Pagan, A and A. Ullah, 1999, Nonparametric Econometrics, New York, Cambridge University Press.

[21] Rudin, Walter. 2006, Real and complex analysis, McGraw-Hill

[22] Shaikh, A., M.Simonsen,E.Vytlacil and N.Yildiz, 2009, A specification Test for Propensity Score Using its Distribution Conditional on Participation, Journal of Econometrics, 151, 33-46.

[23] Sant'anna, P.H.C. and X. Song. 2019, Specification tests for the propensity score, Journal of Econometrics, 210, 379-404.

[24] A L Vol'berg and S V Konyagin 1988, On measures with the doubling condition, Math. USSR Izv.30 629-638.

[25] Zheng J.X.,1996, A consistent test of functional form via nonparametric estimation techniques, Journal of Econometrics, 75, 263-289.

[26] Zinde-Walsh, V. 2017, Kernel Estimation when Density May not Exist – A Corrigendum, Econometric Theory, 33, 1259–1263.

[27] Zinde-Walsh, V. 2008, Kernel estimation when density may not exist, Econometric Theory, 24, pp. 696-725.