

Article

Testing Symmetry of Unknown Densities via Smoothing with the Generalized Gamma Kernels

Masayuki Hirukawa ^{1,*} and Mari Sakudo ^{2,3,4}

¹ Faculty of Economics, Setsunan University, 17-8 Ikeda Nakamachi, Neyagawa, Osaka 572-8508, Japan

² Research Institute of Capital Formation, Development Bank of Japan, 9-7, Otemachi 1-chome, Chiyoda-ku, Tokyo 100-8178, Japan; marisakudo@gmail.com

³ Waseda Institute of Political Economy, Waseda University, 6-1 Nishiwaseda 1-chome, Shinjuku-ku, Tokyo 169-8050, Japan

⁴ Japan Economic Research Institute, 2-1, Otemachi 2-chome, Chiyoda-ku, Tokyo 100-0004, Japan

* Correspondence: hirukawa@econ.setsunan.ac.jp; Tel.: +81-72-839-8095

Academic Editor: Kerry Patterson

Received: 29 December 2015; Accepted: 30 May 2016; Published: 17 June 2016

Abstract: This paper improves a kernel-smoothed test of symmetry through combining it with a new class of asymmetric kernels called the generalized gamma kernels. It is demonstrated that the improved test statistic has a normal limit under the null of symmetry and is consistent under the alternative. A test-oriented smoothing parameter selection method is also proposed to implement the test. Monte Carlo simulations indicate superior finite-sample performance of the test statistic. It is worth emphasizing that the performance is grounded on the first-order normal limit and a small number of observations, despite a nonparametric convergence rate and a sample-splitting procedure of the test.

Keywords: asymmetric kernel; degenerate U -statistic; generalized gamma kernels; nonparametric kernel testing; smoothing parameter selection; symmetry test; two-sample goodness-of-fit test

MSC: 62G10; 62G20

JEL: C12; C14

1. Introduction

Symmetry and conditional symmetry play a key role in numerous fields of economics and finance. Economists' focuses are often on asymmetry of price adjustments (Bacon [1]), innovations in asset markets (Campbell and Hentschel [2]) or policy shocks (Clarida and Gertler [3]). In addition, the mean-variance analysis in finance is consistent with investors' portfolio decision making if and only if asset returns are elliptically distributed (e.g., Chamberlain [4]; Owen and Rabinovitch [5]; Appendix B in Chapter 4 of Ingersoll [6]). Moreover, conditional symmetry in the distribution of the disturbance is often a key regularity condition for regression analysis. In particular, convergence properties of adaptive estimation and robust regression estimation are typically explored under this condition. For the former, Bickel [7] and Newey [8] demonstrate that conditional symmetry of the disturbance distribution in the contexts of linear regression and moment-condition models, respectively, suffices for adaptive estimators to attain their efficiency bounds. For the latter, Carroll and Welsh [9] warn invalidity in inference based on robust regression estimation when the regression disturbance is asymmetrically distributed. Indeed, symmetry of the disturbance distribution is often a key assumption for consistency of parameter estimators in certain versions of robust regression estimation (e.g., Lee [10,11]; Zinde-Walsh [12]; Bondell and Stefanski [13]). Based on their simulation studies,

Baldauf and Santos Silva [14] also argue that lack of conditional symmetry in the disturbance distribution may lead to inconsistency of parameter estimates via robust regression estimation.

In view of the importance in the existence of symmetry, a number of tests for symmetry and conditional symmetry have been proposed. The tests can be classified into kernel and non-kernel methods. Examples for the former include Fan and Gencay [15], Ahmad and Li [16], Zheng [17], Diks and Tong [18], and Fan and Ullah [19]. The latter falls into the tests based on: (i) sample moments (Randles *et al.* [20]; Godfrey and Orme [21]; Bai and Ng [22]; Premaratne and Bera [23]); (ii) regression percentile (Newey and Powell [24]); (iii) martingale transformation (Bai and Ng [25]); (iv) empirical processes (Delgado and Escanciano [26]; Chen and Tripathi [27]); and (v) Neyman's smooth test (Fang *et al.* [28]). Our focus is on the test by Fernandes, Mendes and Scaillet [29] (abbreviated as "FMS" hereafter). While this test can be viewed as the kernel-smoothed one, it has a unique feature. When a probability density function ("pdf") is symmetric about zero, its shapes on positive and negative sides must be mirror images each other. Then, after estimating pdfs on positive and negative sides separately using positive and absolute values of negative observations, respectively, FMS examine whether symmetry holds through gauging closeness between two density estimates. By this nature, we call the test the split-sample symmetry test ("SSST") hereafter. One of the features of the SSST is that it relies on asymmetric kernels with support on $[0, \infty)$ such as the gamma ("G") kernel by Chen [30]. Asymmetric kernel estimators are nonnegative and boundary bias-free, and achieve the optimal convergence rate (in the mean integrated squared error sense) within the class of nonnegative kernel estimators. It is also reported (e.g., p. 597 of Gospodinov and Hirukawa [31]; p. 651 of FMS) that asymmetric kernel-based estimation and inference possess nice finite-sample properties. The split-sample approach is expected to result in efficiency loss. However, it can attain the same convergence rate as the smoothed symmetry tests using symmetric kernels do. Furthermore, unlike these tests, the SSST does not require continuity of density derivatives at the origin.

The aim of this paper is to ameliorate the SSST further through combining it with the generalized gamma ("GG") kernels, a new class of asymmetric kernels with support on $[0, \infty)$ that have been proposed recently by Hirukawa and Sakudo [32]. Our particular focus is on two special cases of the GG kernels, namely, the modified gamma ("MG") and Nakagami- m ("NM") kernels. While superior finite-sample performance of the MG kernel has been reported in the literature, the NM kernel is also anticipated to have an advantage when applied to the SSST. It is known that finite-sample performance of a kernel density estimator depends on proximity in shape between the underlying density and the kernel chosen. As shown in Section 2, the NM kernel collapses to the half-normal pdf when smoothing is made at the origin, and the shape of the density is likely to be close to those on the positive side of single-peaked symmetric distributions. We also pay particular attention to the smoothing parameter selection. While existing articles on asymmetric kernel-smoothed tests (e.g., Fernandes and Grammig [33]; FMS) simply borrow the choice method based on optimality for density estimation, we tailor the idea of test-oriented smoothing parameter selection by Kulasekera and Wang [34,35] to the SSST.

The SSST with the GG kernels plugged in preserves all appealing properties documented in FMS. First, the SSST has a normal limit under the null of symmetry and it is also consistent under the alternative. Hence, unlike the tests by Delgado and Escanciano [26] and Chen and Tripathi [27], no simulated critical values are required. Second, Monte Carlo simulations indicate superior finite-sample performance of the SSST smoothed by the GG kernels. The performance is confirmed even when the entire sample size is 50, despite a nonparametric convergence rate and a sample-splitting procedure. Remarkably, the superior performance is based simply on first-order asymptotic results, and thus the assistance of bootstrapping appears to be unnecessary, unlike most of the smoothed tests employing fixed, symmetric kernels. This result complements previous findings on asymmetric kernel-smoothed tests by Fernandes and Grammig [33] and FMS.

The remainder of this paper is organized as follows. In Section 2 a brief review of a family of the GG kernels is provided. Section 3 proposes symmetry and conditional symmetry tests based

on the GG kernels. Their limiting null distributions and power properties are also explored. As an important practical problem, Section 4 discusses the smoothing parameter selection. Our particular focus is on the choice method for power optimality. Section 5 conducts Monte Carlo simulations to investigate finite-sample properties of the test statistics. Section 6 summarizes the main results of the paper. Proofs are provided in the Appendix.

This paper adopts the following notational conventions: $\Gamma(a) = \int_0^\infty y^{a-1} \exp(-y) dy$ ($a > 0$) is the gamma function; $\mathbf{1}\{\cdot\}$ signifies an indicator function; $[\cdot]$ denotes the integer part; $\|A\| = \{\text{tr}(A'A)\}^{1/2}$ is the Euclidian norm of matrix A ; and $c (> 0)$ denotes a generic constant, the quantity of which varies from statement to statement. The expression “ $X \stackrel{d}{=} Y$ ” reads “A random variable X obeys the distribution Y .” The expression “ $X_n \sim Y_n$ ” is used whenever $X_n/Y_n \rightarrow 1$ as $n \rightarrow \infty$. Lastly, in order to describe different asymptotic properties of an asymmetric kernel estimator across positions of the design point $x (> 0)$ relative to the smoothing parameter $b (> 0)$ that shrinks toward zero, we denote by “interior x ” and “boundary x ” a design point x that satisfies $x/b \rightarrow \infty$ and $x/b \rightarrow \kappa$ for some $0 < \kappa < \infty$ as $b \rightarrow 0$, respectively.

2. Family of the GG Kernels: A Brief Review

Before proceeding, we provide a concise review on a family of the GG kernels. The family constitutes a new class of asymmetric kernels, and it consists of a specific functional form and a set of common conditions, as in Definition 1 below. The name “GG kernels” comes from the fact that the pdf of a GG distribution by Stacy [36] is chosen as the functional form. A major advantage of the family is that for each asymmetric kernel generated from this class, asymptotic properties of the kernel estimators (e.g., density and regression estimators) can be delivered by manipulating the conditions directly, as with symmetric kernels.

Definition 1. (Hirukawa and Sakudo [32], Definition 1) Let $(\alpha, \beta, \gamma) = (\alpha_b(x), \beta_b(x), \gamma_b(x)) \in \mathbb{R}_+^3$ be a continuous function of the design point x and the smoothing parameter b . For such (α, β, γ) , consider the pdf of GG $(\alpha, \beta\Gamma(\alpha/\gamma)/\Gamma\{(\alpha+1)/\gamma\}, \gamma)$, i.e.,

$$K_{GG}(u; x, b) = \frac{\gamma u^{\alpha-1} \exp\left[-\left\{\frac{\beta\Gamma(\frac{\alpha}{\gamma})/\Gamma(\frac{\alpha+1}{\gamma})}{\Gamma(\frac{\alpha}{\gamma})}\right\}^\gamma u\right]}{\left\{\beta\Gamma\left(\frac{\alpha}{\gamma}\right)/\Gamma\left(\frac{\alpha+1}{\gamma}\right)\right\}^\alpha \Gamma\left(\frac{\alpha}{\gamma}\right)} \mathbf{1}\{u \geq 0\}. \tag{1}$$

This pdf is said to be a family of the GG kernels if it satisfies each of the following conditions:

Condition 1. $\beta = \begin{cases} x & \text{for } x \geq C_1 b \\ \varphi_b(x) & \text{for } x \in [0, C_1 b) \end{cases}$, where $0 < C_1 < \infty$ is some constant, the function $\varphi_b(x)$ satisfies $C_2 b \leq \varphi_b(x) \leq C_3 b$ for some constants $0 < C_2 \leq C_3 < \infty$, and the connection between x and $\varphi_b(x)$ at $x = C_1 b$ is smooth.

Condition 2. $\alpha, \gamma \geq 1$, and for $x \in [0, C_1 b)$, α satisfies $1 \leq \alpha \leq C_4$ for some constant $1 \leq C_4 < \infty$. Moreover, connections of α and γ at $x = C_1 b$, if any, are smooth.

Condition 3. $M_b(x) := \frac{\Gamma(\frac{\alpha}{\gamma})\Gamma(\frac{\alpha+2}{\gamma})}{\{\Gamma(\frac{\alpha+1}{\gamma})\}^2} = \begin{cases} 1 + (C_5/x)b + o(b) & \text{for } x \geq C_1 b \\ O(1) & \text{for } x \in [0, C_1 b) \end{cases}$, for some constant $0 < |C_5| < \infty$.

Condition 4. $H_b(x) := \frac{\Gamma(\frac{\alpha}{\gamma})\Gamma(\frac{2\alpha}{\gamma})}{2^{1/\gamma}\Gamma(\frac{\alpha+1}{\gamma})\Gamma(\frac{2\alpha-1}{\gamma})} = \begin{cases} 1 + o(1) & \text{for interior } x \\ O(1) & \text{for boundary } x \end{cases}$.

Condition 5. $A_{b,\nu}(x) := \left\{ \frac{\gamma \Gamma(\frac{\alpha+1}{\gamma})}{\beta} \right\}^{\nu-1} \frac{\Gamma\left\{\frac{\nu(\alpha-1)+1}{\gamma}\right\}}{\nu^{\frac{\nu(\alpha-1)+1}{\gamma}} \left\{\Gamma\left(\frac{\alpha}{\gamma}\right)\right\}^{2\nu-1}} \sim \begin{cases} V_I(\nu) (xb)^{\frac{1-\nu}{2}} & \text{for interior } x \\ V_B(\nu) b^{1-\nu} & \text{for boundary } x \end{cases}$,
 $\nu \in \mathbb{R}_+$, where constants $0 < V_I(\nu), V_B(\nu) < \infty$ depend only on ν .

The family embraces the following two special cases¹. Putting

$$(\alpha, \beta) = \begin{cases} \left(\frac{x}{b}, x\right) & \text{for } x \geq 2b \\ \left(\frac{1}{4}\left(\frac{x}{b}\right)^2 + 1, \frac{x^2}{4b} + b\right) & \text{for } x \in [0, 2b) \end{cases}$$

and $\gamma = 1$ in (1) generates the MG kernel

$$K_{MG}(u; x, b) = \frac{u^{\alpha-1} \exp\{-u/(\beta/\alpha)\}}{(\beta/\alpha)^\alpha \Gamma(\alpha)} \mathbf{1}\{u \geq 0\}.$$

It can be found that this is equivalent to the one proposed by Chen [30] by recognizing that $\alpha = \rho_b(x)$ on p. 473 of Chen [30] and $\beta/\alpha = b$. The same (α, β) and $\gamma = 2$ also yields the NM kernel

$$K_{NM}(u; x, b) = \frac{2u^{\alpha-1} \exp\left[-\left\{u / \left(\beta \Gamma\left(\frac{\alpha}{2}\right) / \Gamma\left(\frac{\alpha+1}{2}\right)\right)\right\}^2\right]}{\left\{\beta \Gamma\left(\frac{\alpha}{2}\right) / \Gamma\left(\frac{\alpha+1}{2}\right)\right\}^\alpha \Gamma\left(\frac{\alpha}{2}\right)} \mathbf{1}\{u \geq 0\}.$$

The GG kernels are designed to inherit all appealing properties that the MG kernel possesses. We conclude this section by referring to the properties. Two properties below are basic ones. First, by construction, the GG kernels are free of boundary bias and always generate nonnegative density estimates everywhere. Second, the shape of each GG kernel varies according to the position at which smoothing is made; in other words, the amount of smoothing changes in a locally adaptive manner. To illustrate this property, Figure 1 plots the shapes of the MG and NM kernels at four different design points ($x = 0.0, 0.5, 1.0, 2.0$) at which the smoothing is performed. For reference, the G kernel is also drawn in each panel. When smoothing is made at the origin (Panel (A)), the NM kernel collapses to a half-normal pdf, whereas others reduce to exponential pdfs. As the design point moves away from the boundary (Panels (B–D)), the shape of each kernel becomes flatter and closer to symmetry. We should stress that Figure 1 is drawn with the value of the smoothing parameter fixed at $b = 0.2$. Unlike variable bandwidth methods for fixed, symmetric kernels (e.g., Abramson [37]), adaptive smoothing of these kernels can be achieved by a single smoothing parameter, which makes them much more appealing in empirical work.

The remaining three properties are on density estimates using the GG kernels. Third, when best implemented, each GG density estimator attains Stone’s [38] optimal convergence rate in the mean integrated squared error within the class of nonnegative kernel density estimators. Fourth, the leading bias of each GG density estimator contains only the second-order derivative of the true density over the interior region, unlike many other asymmetric kernels including the G kernel. Fifth, the variance of the GG estimator tends to decrease as the design point moves away from the boundary. This property is particularly advantageous to estimating the distributions that have long tails with sparse data.

¹ Hirukawa and Sakudo [32] present the Weibull kernel as yet another special case. However, it is not confirmed that this kernel satisfies Lemma 1 below, and thus the kernel is not investigated throughout.

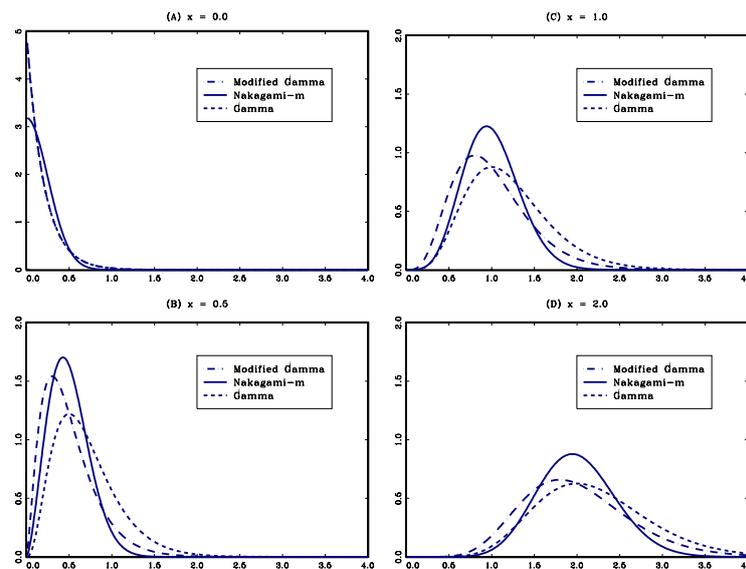


Figure 1. Shapes of the GG Kernels When $b = 0.2$.

3. Tests for Symmetry and Conditional Symmetry Smoothed by the GG Kernels

3.1. SSST as a Special Case of Two-Sample Goodness-of-Fit Tests

This section proposes to combine the SSST with the GG kernels, explores asymptotic properties of the test statistic, and finally expands the scope of the test to testing the null of conditional symmetry. The SSST can be characterized as a special case of two-sample tests for equality of two unknown densities investigated by Anderson, Hall and Titterton [39]. Suppose that we are interested in testing symmetry of the distribution of a random variable $U \in \mathbb{R}$. Without loss of generality, we hypothesize that the distribution is symmetric about zero. If U has a pdf, then under the null, its shapes on positive and negative sides of the entire real line \mathbb{R} must be mirror images each other. Let f and g be the pdfs to the right and left from the origin, respectively. Then, we would like to test the null hypothesis

$$H_0 : f(u) = g(u) \text{ for almost all } u \in \mathbb{R}_+$$

against the alternative

$$H_1 : f(u) \neq g(u) \text{ on a set of positive measure in } \mathbb{R}_+.$$

Accordingly, a natural test statistic should be built on the integrated squared error (“ISE”)

$$\begin{aligned} I &= \int_0^\infty \{f(u) - g(u)\}^2 du \\ &= \int_0^\infty \{f(u) - g(u)\} dF(u) - \int_0^\infty \{f(u) - g(u)\} dG(u), \end{aligned}$$

where F and G are cumulative distribution functions corresponding to f and g , respectively.

The name of the SSST comes from the way to construct a sample analog to I . A random sample of N observations $\{U_i\}_{i=1}^N$ is split into two sub-samples, namely, $\{X_i\}_{i=1}^{n_1} := \{U_i : U_i \geq 0\}_{i=1}^{n_1}$ and $\{Y_i\}_{i=1}^{n_2} := \{-U_i : U_i < 0\}_{i=1}^{n_2}$, where $N = n_1 + n_2$. Given the sub-samples, f and g can be estimated using a GG kernel with the smoothing parameter b as

$$\hat{f}(u) = \frac{1}{n_1} \sum_{i=1}^{n_1} K_{GG}(X_i; u, b) \text{ and } \hat{g}(u) = \frac{1}{n_2} \sum_{i=1}^{n_2} K_{GG}(Y_i; u, b),$$

respectively². Similarly, (F, G) is replaced with their empirical measures (F_{n_1}, G_{n_2}) . In addition, because $n_1 \sim n_2$ under H_0 , without loss of generality and for ease of exposition, we assume that N is even and that $n := n_1 = n_2 = N/2$. Using a short-handed notation $K_X(Y) = K_{GG}(Y; X, b)$ finally yields the sample analog to I as

$$\begin{aligned} I_n &= \frac{1}{n} \sum_{i=1}^n \left\{ \hat{f}(X_i) + \hat{g}(Y_i) - \hat{g}(X_i) - \hat{f}(Y_i) \right\} \\ &= \sum_{i=1}^n \frac{1}{n^2} \left\{ K_{X_i}(X_i) + K_{Y_i}(Y_i) - K_{Y_i}(X_i) - K_{X_i}(Y_i) \right\} \\ &\quad + \sum_{j=1}^n \sum_{i=1, i \neq j}^n \frac{1}{n^2} \left\{ K_{X_j}(X_i) + K_{Y_j}(Y_i) - K_{Y_j}(X_i) - K_{X_j}(Y_i) \right\} \\ &= I_{1n} + I_n. \end{aligned}$$

Although we could use \bar{I}_n itself as the test statistic, the probability limit of I_{1n} plays a role in a non-vanishing center term of the asymptotic null distribution. Because the term is likely to cause size distortions in finite samples, we focus only on I_n to construct the testing statistic. Now I_n can be rewritten as

$$I_n := \sum_{1 \leq i < j \leq n} \Phi_n(Z_i, Z_j) := \sum_{1 \leq i < j \leq n} \frac{1}{n^2} \left\{ \phi_n(Z_i, Z_j) + \phi_n(Z_j, Z_i) \right\},$$

where $Z_i := (X_i, Y_i)$ and $\phi_n(Z_i, Z_j) := K_{X_j}(X_i) + K_{Y_j}(Y_i) - K_{Y_j}(X_i) - K_{X_j}(Y_i)$. Observe that $\Phi_n(Z_i, Z_j)$ is symmetric between Z_i and Z_j and that $E \left\{ \Phi_n(Z_i, Z_j) \mid Z_i \right\} = 0$ almost surely under H_0 . It follows that I_n is a degenerate U -statistic, and thus we may apply a martingale central limit theorem (e.g., Theorem 1 of Hall [40]; Theorem 4.7.3 of Koroljuk and Borovskich [41]).

Before describing the asymptotic properties of I_n , we make two remarks. First, applying the idea of two-sample goodness-of-fit tests to the symmetry test is not new. Ahmad and Li [16] and Fan and Ullah [19] have also studied the symmetry test based on closeness of two density estimates measured by the ISE. They estimate densities using two samples, namely, the original entire sample $\{X_i\}_{i=1}^N := \{U_i\}_{i=1}^N$ and the one obtained by flipping the sign of each observation $\{Y_i\}_{i=1}^N := \{-U_i\}_{i=1}^N$ in our notations. Because each of X and Y has support on $(-\infty, \infty)$ by construction, a standard symmetric kernel is employed for density estimation unlike the SSST. Second, if X and Y are taken from two different distributions with support on $[0, \infty)$, then I_n can be viewed as a pure two-sample goodness-of-fit test. It can be immediately applied to the testing for equality of two unknown distributions of nonnegative economic and financial variables such as incomes, wages, short-term interest rates, and insurance claims.

To present the convergence properties of I_n , we make the following assumptions.

Assumption 1. Two random samples $\{X_i\}_{i=1}^{n_1}$ and $\{Y_i\}_{i=1}^{n_2}$ are drawn independently from univariate distributions that have pdfs f and g with support on $[0, \infty)$, respectively.

Assumption 2. f and g are twice continuously differentiable on $[0, \infty)$, and $E |Xf''(X)|^2, E |X^2f''(X)g''(X)|, E |Y^2f''(Y)g''(Y)|, E |Yg''(Y)|^2 < \infty$.

Assumption 3. The smoothing parameter $b (= b_n)$ satisfies $b + (nb)^{-1} \rightarrow 0$ as $n \rightarrow \infty$.

² It is possible to use different asymmetric kernels and/or different smoothing parameters to estimate f and g . For convenience, however, we choose to employ the same asymmetric kernel function and a single smoothing parameter.

Assumption 4. Let (X_1, X_2) and (Y_1, Y_2) be two independent copies of X and Y , respectively. Then, the followings hold:

- (a) $E \{K_{X_2}(X_1) K_{Y_2}(X_1)\} \sim E \{f(X) g(X)\}$; and $E \{K_{Y_2}(Y_1) K_{X_2}(Y_1)\} \sim E \{f(Y) g(Y)\}$.
- (b) $E \{K_{X_2}(X_1) K_{X_1}(X_2)\} \sim b^{-1/2} V_I(2) E \{X^{-1/2} f(X)\}$; $E \{K_{X_2}(Y_1) K_{Y_1}(X_2)\} \sim b^{-1/2} V_I(2) E \{X^{-1/2} g(X)\}$; $E \{K_{Y_2}(X_1) K_{X_1}(Y_2)\} \sim b^{-1/2} V_I(2) E \{Y^{-1/2} f(Y)\}$; and $E \{K_{Y_2}(Y_1) K_{Y_1}(Y_2)\} \sim b^{-1/2} V_I(2) E \{Y^{-1/2} g(Y)\}$, where $V_I(2)$ is a kernel-specific constant given in Condition 5 of Definition 1.

Assumptions 1–3 are standard in the literature of asymmetric kernel smoothing. On the other hand, Assumption 4 has a different flavor. Convergence results on I_n are built on several different moment approximations. While Definition 1 implies the statements in Lemma A2 in the Appendix, it is unclear whether the definition may even admit such approximations as in Assumption 4. The difficulty comes from the fact that unlike symmetric kernels, roles of design points and data points are nonexchangeable in asymmetric kernels. What makes the problem more complicated is that functional forms of $(\alpha, \beta, \gamma) = (\alpha_b(x), \beta_b(x), \gamma_b(x))$ in the GG kernels are not fully specified in Definition 1. Considering that not all GG kernels may admit the moment approximations (a) and (b), we choose to make an extra assumption. Note that the MG and NM kernels fulfill Assumption 4, as documented in the next lemma.

Lemma 1. *If Assumptions 1–3 hold, then each of the MG and NM kernels satisfies Assumption 4.*

The theorem below delivers the convergence properties of I_n and provides a consistent estimator of its asymptotic variance.

Theorem 1. *Suppose that Assumptions 1–4 and $n_1 = n_2 = n$ hold.*

- (i) Under H_0 , $nb^{1/4} I_n \xrightarrow{d} N(0, \sigma^2)$ as $n \rightarrow \infty$, where

$$\sigma^2 = 2V_I(2) E \left[X^{-1/2} \{f(X) + g(X)\} + Y^{-1/2} \{f(Y) + g(Y)\} \right],$$

which reduces to $\sigma^2 = 8V_I(2) E \{X^{-1/2} f(X)\}$ under H_0 , and $V_I(2)$ is a kernel-specific constant given in Condition 5 of Definition 1.

- (ii) A consistent estimator of σ^2 is given by

$$\hat{\sigma}^2 = 2V_I(2) \frac{1}{n} \sum_{i=1}^n \left[X_i^{-1/2} \{ \hat{f}(X_i) + \hat{g}(X_i) \} + Y_i^{-1/2} \{ \hat{f}(Y_i) + \hat{g}(Y_i) \} \right]. \tag{2}$$

We make a few remarks. First, it follows from Lemma 1 and Theorem 1 that the MG and NM kernels can be safely employed for the SSST, where values of $V_I(2)$ for these kernels are $(V_{I, MG}(2), V_{I, NM}(2)) = (1/(2\sqrt{\pi}), 1/\sqrt{2\pi})$. It also follows from Proposition 1 of FMS and Theorem 1 that limiting null distributions of $nb^{1/4} I_n$ using the G and MG kernels coincide, as expected. Second, while a similar form to the asymptotic variance σ^2 can be found in Proposition 1 of FMS, σ^2 takes a more general form. Accordingly, the variance estimator $\hat{\sigma}^2$ is consistent under both H_0 and H_1 . Third, it can be inferred from Theorem 1 that the test statistic becomes $T_n := nb^{1/4} I_n / \hat{\sigma}$. As a consequence, the SSST is a one-sided test that rejects H_0 in favor of H_1 if $T_n > z_\alpha$, where z_α is the upper α -percentile of $N(0, 1)$.

The next proposition refers to consistency of the SSST. Observe that the power approaches one for local alternatives with convergence rates no faster than $nb^{1/4}$, as well as for fixed alternatives.

Proposition 1. *If Assumptions 1–4 hold, then under H_1 , $\Pr(T_n > B_n) \rightarrow 1$ as $n \rightarrow \infty$ for any non-stochastic sequence B_n satisfying $B_n = o(nb^{1/4})$.*

3.2. SSST When Two Sub-Samples Have Unequal Sample Sizes

Convergence results in the previous section rely on the assumption that the sample sizes of two sub-samples $\{X_i\}_{i=1}^{n_1}$ and $\{Y_i\}_{i=1}^{n_2}$ are the same, i.e., so far $n_1 = n_2$ has been maintained. In reality, $n_1 \neq n_2$ is often the case, in particular, when the entire sample size $N = n_1 + n_2$ is odd or when H_1 is true.

Handling this case requires more tedious calculation. When $n_1 \neq n_2$, I_n can be rewritten as

$$\begin{aligned}
 I_{n_1, n_2} = & \sum_{j=1}^{n_1} \sum_{i=1, i \neq j}^{n_1} \frac{1}{n_1^2} K_{X_j}(X_i) + \sum_{j=1}^{n_2} \sum_{i=1, i \neq j}^{n_2} \frac{1}{n_2^2} K_{Y_j}(Y_i) \\
 & - \sum_{j=1}^{n_2} \sum_{i=1, i \neq j}^{n_1} \frac{1}{n_1 n_2} K_{Y_j}(X_i) - \sum_{j=1}^{n_1} \sum_{i=1, i \neq j}^{n_2} \frac{1}{n_1 n_2} K_{X_j}(Y_i). \tag{3}
 \end{aligned}$$

Following Fan and Ullah [19], we deliver convergence results under the assumption that two sample sizes n_1 and n_2 diverge at the same rate. The asymptotic variance of $n_1 b^{1/4} I_{n_1, n_2}$ and its consistent estimate are also provided. Because the essential arguments are the same as those for Theorem 1 and Proposition 1, we omit the proofs of Theorem 2 and Proposition 2 and simply state the results. Observe that when $n_1 = n_2 = n$, these results collapse to Theorem 1 and Proposition 1, respectively.

Theorem 2. *Suppose that Assumptions 1–4 and $n_1/n_2 \rightarrow \lambda$ for some constant $\lambda \in (0, \infty)$ hold.*

(i) *Under H_0 , $n_1 b^{1/4} I_{n_1, n_2} \xrightarrow{d} N(0, \sigma_\lambda^2)$ as $n_1 \rightarrow \infty$, where*

$$\begin{aligned}
 \sigma_\lambda^2 = & 2V_I(2) \left[E \left\{ X^{-1/2} f(X) \right\} + \lambda E \left\{ X^{-1/2} g(X) \right\} \right. \\
 & \left. + \lambda E \left\{ Y^{-1/2} f(Y) \right\} + \lambda^2 E \left\{ Y^{-1/2} g(Y) \right\} \right],
 \end{aligned}$$

which reduces to $\sigma_\lambda^2 = 2(1 + \lambda)^2 V_I(2) E \left\{ X^{-1/2} f(X) \right\}$ under H_0 .

(ii) *A consistent estimator of σ_λ^2 is given by*

$$\begin{aligned}
 \hat{\sigma}_\lambda^2 = & 2V_I(2) \left\{ \frac{1}{n_1} \sum_{i=1}^{n_1} X_i^{-1/2} \hat{f}(X_i) + \left(\frac{n_1}{n_2}\right) \frac{1}{n_1} \sum_{i=1}^{n_1} X_i^{-1/2} \hat{g}(X_i) \right. \\
 & \left. + \left(\frac{n_1}{n_2}\right) \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i^{-1/2} \hat{f}(Y_i) + \left(\frac{n_1}{n_2}\right)^2 \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i^{-1/2} \hat{g}(Y_i) \right\}. \tag{4}
 \end{aligned}$$

Proposition 2. *If Assumptions 1–4 and $n_1/n_2 \rightarrow \lambda \in (0, \infty)$ hold, then under H_1 , $\Pr(T_{n_1, n_2} > B_{n_1}) := \Pr(n_1 b^{1/4} I_{n_1, n_2} / \hat{\sigma}_\lambda > B_{n_1}) \rightarrow 1$ as $n_1 \rightarrow \infty$ for any non-stochastic sequence B_{n_1} satisfying $B_{n_1} = o(n_1 b^{1/4})$.*

The next corollary is a natural outcome from Theorem 2 and comes from the fact that under H_0 , $n_1 \sim n_2$ or $\lambda = 1$ holds. Because N could be odd in this context, n should read $n = \lfloor N/2 \rfloor$.

Corollary 1. *If Assumptions 1–4 and $n_1/n_2 \rightarrow 1$ hold, then $n_1, n_2 \sim n = \lfloor N/2 \rfloor$ so that $n_1 b^{1/4} I_{n_1, n_2} = n b^{1/4} I_n + o_p(1)$.*

3.3. Extension to a Test for Conditional Symmetry

So far we have maintained the assumption that the random variable U is observable and has a distribution that is symmetric about zero. However, often U is unobservable or the axis of symmetry is not zero. The former is typical when we are interested in symmetry of the distribution of the disturbance conditional on regressors in regression analysis. In this scenario, the test is conducted after U is replaced with the residual. For the latter, the test should be based on location-adjusted observations, *i.e.*, transformed observations with an estimate of the axis of symmetry (e.g., the sample mean or the sample median) subtracted from U . These aspects motivate us to generalize the SSST to the testing for conditional symmetry.

Following FMS, we consider a testing for symmetry in the conditional distribution of $V_1 | V_2$ with $(V_1, V_2) \in \mathbb{R} \times \mathbb{R}^d$ within the framework of a semiparametric context. Specifically, for a parameter space Θ_1 and a function $\zeta_1 : \mathbb{R}^d \times \Theta_1 \rightarrow \mathbb{R}$, it suffices to check whether the conditional distribution of $V_1 | V_2$ is symmetric about $\zeta_1(V_2; \theta_1^0)$ for some $\theta_1^0 \in \Theta_1$. Observe that this is equivalent to test whether there is $\theta_1^0 \in \Theta_1$ such that the conditional distribution of $V | V_2 := V_1 - \zeta_1(V_2; \theta_1^0) | V_2$ is symmetric about zero.

However, implementing this type of testing strategy requires to estimate the conditional pdf of $V | V_2$ nonparametrically. This is cumbersome, considering the curse of dimensionality in V_2 and another smoothing parameter choice. Instead, as in Zheng [17], Bai and Ng [25] and Delgado and Escanciano [26], we assume that there are a parameter space Θ_2 and a function $\zeta_2 : \mathbb{R}^{d+1} \times \Theta_2 \rightarrow \mathbb{R}$ that can attain symmetry of the marginal distribution of $U := \zeta_2(V_1, V_2; \theta_2^0)$ about zero for some $\theta_2^0 \in \Theta_2$. Given the dependence of V (and thus U) on Θ_1 , we can finally rewrite our testing scheme as the one that tests, for a suitable parameter space Θ and a function $\zeta : \mathbb{R}^{d+1} \times \Theta \rightarrow \mathbb{R}$, symmetry of the marginal distribution of $U = \zeta(V_1, V_2; \theta_0)$ about zero for some $\theta_0 \in \Theta$.

Accordingly, the procedure of the conditional symmetry test takes the following two steps. First, we estimate $\zeta(\cdot, \cdot; \theta_0)$ given N observations $\{(V_{1i}, V_{2i})\}_{i=1}^N$ and denote a consistent estimator of (ζ, θ_0) as $(\hat{\zeta}, \hat{\theta})$. Second, the test is conducted using $\{\hat{U}_i\}_{i=1}^N := \{\hat{\zeta}(V_{1i}, V_{2i}; \hat{\theta})\}_{i=1}^N$. As before, the entire sample is split into two sub-samples $\{\hat{X}_i\}_{i=1}^{n_1} := \{\hat{U}_i : \hat{U}_i \geq 0\}_{i=1}^{n_1}$ and $\{\hat{Y}_i\}_{i=1}^{n_2} := \{-\hat{U}_i : \hat{U}_i < 0\}_{i=1}^{n_2}$. Then, the test statistics, namely, $I_n(\hat{\zeta}, \hat{\theta})$ and $I_{n_1, n_2}(\hat{\zeta}, \hat{\theta})$ for equal ($n_1 = n_2 = n$) and unequal ($n_1 \neq n_2$) sample sizes, can be obtained by replacing $(X, Y) = (X(\zeta, \theta_0), Y(\zeta, \theta_0))$ in $I_n(\zeta, \theta_0)$ and $I_{n_1, n_2}(\zeta, \theta_0)$ with (\hat{X}, \hat{Y}) , respectively.

Our remaining task is to demonstrate that there is no asymptotic cost in the test statistics with (ζ, θ_0) replaced by its estimator $(\hat{\zeta}, \hat{\theta})$, as long as $(\hat{\zeta}, \hat{\theta}) \xrightarrow{p} (\zeta, \theta_0)$ at a suitable rate of convergence. To control the convergence rate, we make Assumption 5 below. Observe that it allows for nonparametric rates of convergence; see Hansen [42], for instance, for uniform convergence rates of kernel estimators.

Assumption 5. $N^r \|\hat{\theta} - \theta_0\| \xrightarrow{p} 0$ and $N^r \|\hat{\zeta} - \zeta\| \xrightarrow{p} 0$ uniformly over $\mathbb{R}^{d+1} \times \Theta$ for some $r \in (0, 1/2]$.

Theorem 3 below provides combinations of the shrinking rate q for b and the convergence rate r for $(\hat{\zeta}, \hat{\theta})$ that can establish the first-order asymptotic equivalence between $nb^{1/4}I_n(\zeta, \theta_0)$ ($n_1b^{1/4}I_{n_1, n_2}(\zeta, \theta_0)$) and $nb^{1/4}I_n(\hat{\zeta}, \hat{\theta})$ ($n_1b^{1/4}I_{n_1, n_2}(\hat{\zeta}, \hat{\theta})$) when two sub-samples have equal (unequal) sample sizes.

Theorem 3. *If Assumptions 1–5 hold, then under H_0 ,*

$$nb^{1/4}I_n(\hat{\zeta}, \hat{\theta}) = nb^{1/4}I_n(\zeta, \theta_0) + o_p(1)$$

as $n \rightarrow \infty$ when $n_1 = n_2 = n$ and

$$n_1b^{1/4}I_{n_1, n_2}(\hat{\zeta}, \hat{\theta}) = n_1b^{1/4}I_{n_1, n_2}(\zeta, \theta_0) + o_p(1)$$

as $n_1 \rightarrow \infty$ when $n_1/n_2 \rightarrow \lambda \in (0, \infty)$, provided that (q, r) belong to the set $\{(q, r) : r > -5q/4 + 1, r > q/2, r \leq 1/2\}$.

The set given in the theorem can be expressed as the triangular region formed by the corners $(2/5, 1/2)$, $(4/7, 2/7)$ and $(1, 1/2)$ on the $q - r$ plane. The theorem also indicates that we must employ the sub-optimal smoothing parameter $b = o(n^{-2/5})$ or undersmooth the observations to avoid additional cost of estimating (ζ, θ_0) , as is the case with other kernel-smoothed tests. Moreover, FMS set $b = o(n^{-4/9})$ and obtain the lower bound of r as $4/9$. Indeed, the set provided in Theorem 3 overlaps the one derived by FMS $\{(q, r) : q > 4/9, r > 4/9\}$.

4. Smoothing Parameter Selection

How to choose the value of the smoothing parameter b is an important practical problem. Nonetheless, it appears that the issue has not been well addressed in the literature on testing problems using asymmetric kernels. While Fernandes and Grammig [33] adopt a method inspired by Silverman’s [43] rule-of-thumb, FMS adjust the value chosen via cross validation. Both methods choose the smoothing parameter value from the viewpoint of optimality for density estimation. Such choices cannot be justified in theory or practice, because estimation-optimal values may not be equally optimal for testing purposes. In contrast, there are a few works on test-oriented smoothing parameter selection. For the test of equality in two unknown regression curves, Kulasekera and Wang [34,35], analytically explore the idea of choosing the smoothing parameter value that maximizes the power with the size preserved. Gao and Gijbels [44] combine this idea with the Edgeworth expansion for a bootstrap specification test of parametric regression models.

Below we tailor the procedure by Kulasekera and Wang [35] to the SSST. For a realistic setup, the case of $n_1 \neq n_2$ is exclusively considered. Their basic idea is from sub-sampling. Without loss of generality assume that $\{X_i\}_{i=1}^{n_1}$ and $\{Y_i\}_{i=1}^{n_2}$ are ordered samples. Then, the entire sample $\{\{X_i\}_{i=1}^{n_1}, \{Y_i\}_{i=1}^{n_2}\}$ can be split into M sub-samples, where $M = M_{n_1}$ is a non-stochastic sequence that satisfies $1/M + M/n_1 \rightarrow 0$ as $n_1 \rightarrow \infty$. Given such M and $(k_1, k_2) := (\lfloor n_1/M \rfloor, \lfloor n_2/M \rfloor)$, the m th sub-sample is defined as $\left\{ \left\{ X_{m+(i-1)M} \right\}_{i=1}^{k_1}, \left\{ Y_{m+(i-1)M} \right\}_{i=1}^{k_2} \right\}$, $m = 1, \dots, M$. This sub-sample yields the analogues to (3) and (4) as

$$I_{k_1, k_2}(m) = \sum_{j=1}^{k_1} \sum_{i=1, i \neq j}^{k_1} \frac{1}{k_1^2} K_{X_{m+(j-1)M}} \left(X_{m+(i-1)M} \right) + \sum_{j=1}^{k_2} \sum_{i=1, i \neq j}^{k_2} \frac{1}{k_2^2} K_{Y_{m+(j-1)M}} \left(Y_{m+(i-1)M} \right) - \sum_{j=1}^{k_2} \sum_{i=1, i \neq j}^{k_1} \frac{1}{k_1 k_2} K_{Y_{m+(j-1)M}} \left(X_{m+(i-1)M} \right) - \sum_{j=1}^{k_1} \sum_{i=1, i \neq j}^{k_2} \frac{1}{k_1 k_2} K_{X_{m+(j-1)M}} \left(Y_{m+(i-1)M} \right)$$

and

$$\hat{\sigma}_\lambda^2(m) = 2V_I(2) \left\{ \frac{1}{k_1} \sum_{i=1}^{k_1} X_{m+(i-1)M}^{-1/2} \hat{f}_m \left(X_{m+(i-1)M} \right) + \left(\frac{k_1}{k_2} \right) \frac{1}{k_1} \sum_{i=1}^{k_1} X_{m+(i-1)M}^{-1/2} \hat{g}_m \left(X_{m+(i-1)M} \right) + \left(\frac{k_1}{k_2} \right) \frac{1}{k_2} \sum_{i=1}^{k_2} Y_{m+(i-1)M}^{-1/2} \hat{f}_m \left(Y_{m+(i-1)M} \right) + \left(\frac{k_1}{k_2} \right)^2 \frac{1}{k_2} \sum_{i=1}^{k_2} Y_{m+(i-1)M}^{-1/2} \hat{g}_m \left(Y_{m+(i-1)M} \right) \right\},$$

where

$$\hat{f}_m(u) = \frac{1}{k_1} \sum_{i=1}^{k_1} K_{GG} \left(X_{m+(i-1)M}; u, b \right) \text{ and}$$

$$\hat{g}_m(u) = \frac{1}{k_2} \sum_{i=1}^{k_2} K_{GG} \left(Y_{m+(i-1)M}; u, b \right).$$

It follows that the test statistic using the m th sub-sample becomes

$$T_{k_1, k_2}(m) := \frac{k_1 b^{1/4} I_{k_1, k_2}(m)}{\hat{\sigma}_\lambda(m)}, m = 1, \dots, M.$$

Also denote the set of admissible values for $b = b_{n_1}$ as $H_{n_1} := [\underline{B}n_1^{-q}, \bar{B}n_1^{-q}]$ for some prespecified exponent $q \in (0, 1)$ and two constants $0 < \underline{B} < \bar{B} < \infty$. Moreover, let

$$\hat{\pi}_M(b_{k_1}) := \frac{1}{M} \sum_{m=1}^M \mathbf{1}\{T_{k_1, k_2}(m) > c_m(\alpha)\},$$

where $c_m(\alpha)$ is the critical value for the size α test using the m th sub-sample. We pick the power-maximized $\hat{b}_{k_1} = \hat{B}k_1^{-q} = \arg \max_{b_{k_1} \in H_{k_1}} \hat{\pi}_M(b_{k_1})$, and the smoothing parameter value $\hat{b}_{n_1} := \hat{B}n_1^{-q}$ follows.

The behavior of $\hat{\pi}_M(b_{k_1})$ can be examined by considering the local alternative

$$H'_1 : f(u) = g(u) + \frac{h(u)}{\sqrt{n_1 b^{1/4}}},$$

where $h(u)$ satisfies $\int_0^\infty h(u) du = 0$ and $I_h := \int_0^\infty h^2(u) du \in (0, \infty)$. Also let $\pi(b_{n_1}) := \Pr\{T_{n_1, n_2} > c(\alpha)\}$, where $c(\alpha)$ is the critical value for the size α test using the entire sample. For such $\pi(b_{n_1})$, define $b_{n_1}^* := B^*n_1^{-q} = \arg \max_{b_{n_1} \in H_{n_1}} \pi(b_{n_1})$. Then, \hat{b}_{n_1} is optimal in the sense of Proposition 3. The proof is omitted, because it is a minor modification of the one for Theorem 2.1 of Kulasekera and Wang [35]; indeed it can be established by recognizing that $T_{n_1, n_2} \xrightarrow{d} N(I_h/\sigma_\lambda, 1)$ under H'_1 , as in Proposition 3 of Fernandes and Grammig [33].

Proposition 3. *If Assumptions 1–4, $1/M + M/n_1 \rightarrow 0$ and $n_1/n_2 \rightarrow \lambda \in (0, \infty)$ hold, then $B^*/\hat{B} \xrightarrow{p} 1$ as $n_1 \rightarrow \infty$.*

We conclude this section by stating how to obtain \hat{b}_{n_1} in practice. Step 1 reflects that M should be divergent but smaller than both n_1 and n_2 in finite samples. Step 3 follows from the implementation methods in Kulasekera and Wang [34,35]. Finally, Step 4 considers that there may be more than one maximizer of $\hat{\pi}_M(b_{k_1})$.

- Step 1:** Choose some $\delta \in (0, 1)$ and specify $M = \min\{\lfloor n_1^\delta \rfloor, \lfloor n_2^\delta \rfloor\}$.
- Step 2:** Make M sub-samples of sizes $(k_1, k_2) = (\lfloor n_1/M \rfloor, \lfloor n_2/M \rfloor)$.
- Step 3:** Pick two constants $0 < \underline{H} < \bar{H} < 1$ and define $H_{k_1} = [\underline{H}, \bar{H}]$.
- Step 4:** Set $c_m(\alpha) \equiv z_\alpha$ and find $\hat{b}_{k_1} = \inf\{\arg \max_{b_{k_1} \in H_{k_1}} \hat{\pi}_M(b_{k_1})\}$ by a grid search.
- Step 5:** Obtain $\hat{B} = \hat{b}_{k_1} k_1^q$ and calculate $\hat{b}_{n_1} = \hat{B}n_1^{-q}$.

5. Finite-Sample Performance

5.1. Setup

It is widely recognized that asymptotic results on kernel-smoothed tests are not well transmitted to their finite-sample distributions, which reflects that omitted terms in the first-order asymptotics on the test statistics are highly sensitive to their smoothing parameter values in finite samples. On the other hand, Fernandes and Grammig [33] and FMS report superior finite-sample properties of asymmetric kernel-smoothed tests. To see which perspective dominates, this section investigates finite-sample performance of the test statistic for the SSST via Monte Carlo simulations.

To make a direct comparison with the results by FMS, we specialize in the conditional symmetry test using the same linear regression model $y = \beta_0 + \beta_1 x + u$ as used in FMS. The data are generated in the following manner. First, the regressor x is drawn from $N(0, 1)$. Second, the disturbance u ,

which is independent of x , is drawn from one of eight distributions with means of zero given in Table 1. Distributions with “S” (symmetric) and “A” (asymmetric) are used to investigate size and power properties of the test statistic, respectively. All the distributions are popularly chosen in the literature; the generalized lambda distribution (“GLD”) by Ramberg and Schmeiser [45], in particular, is known to nest a wide variety of symmetric and asymmetric distributions³. Finally, the dependent variable y is generated by setting $\beta_0 = \beta_1 = 1$.

Table 1. Distributions of the Disturbance u in the Simulation Study.

	Distribution	Skewness	Kurtosis
S1	$N(0,1)$	0.00	3.00
S2	t_{10}	0.00	4.00
S3	DE (0,1) or Standard Laplace	0.00	24.00
S4	$U[-1,1]$ or GLD with $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (0, 1, 1, 1)$	0.00	1.80
A1	$LN(0,1) - \exp(1/2)$	6.18	113.94
A2	$\chi_3^2 - 3$	1.63	7.00
A3	GLD with $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (12.601, -0.00980045, -0.11, -0.0001)$	-2.92	19.52
A4	GLD with $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (-9.7726, -0.0151878, -0.001, -0.13)$	3.16	23.75

Note: “S” and “A” stand for symmetric and asymmetric distributions, respectively. “GLD” denotes the generalized lambda distribution by Ramberg and Schmeiser [45]. The distribution is defined in terms of the inverse of the cumulative distribution function $F^{-1}(u) = \lambda_1 + \{u^{\lambda_3} + (1-u)^{\lambda_4}\} / \lambda_2$ for $u \in [0, 1]$.

We are interested in testing symmetry of the conditional distribution of y given x . For this purpose the SSST is applied for the least-squares residual $\hat{u}_i := y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$ using the sample $\{(y_i, x_i)\}_{i=1}^N$, where $(\hat{\beta}_0, \hat{\beta}_1)$ are least-squares estimates of (β_0, β_1) . Finite-sample size and power properties of the test statistic T_{n_1, n_2} for two sub-samples with unequal sample sizes are examined against nominal 5% and 10% levels. The MG and NM kernels (denoted as “ T_{n_1, n_2} -MG” and “ T_{n_1, n_2} -NM”, respectively) are employed as examples of the GG kernels.

Finite-sample properties of T_{n_1, n_2} -MG and T_{n_1, n_2} -NM are evaluated in comparison with other versions of the SSST. First, two versions of FMS’s original test statistic built on an equivalence to our I_n using the G kernel are considered. “FMS-G-O” is FMS’s truly original statistic, whereas “FMS-G-AltVar” is the one with the variance estimator replaced by $\hat{\sigma}_\lambda^2$ given in Theorem 2. Second, T_{n_1, n_2} using the G kernel (denoted as “ T_{n_1, n_2} -G”) is also calculated. Notice that FMS-G-AltVar and T_{n_1, n_2} -G take exactly the same form. The only difference is the method of choosing the smoothing parameter b , which will be discussed shortly. Effects of changing the variance estimator, the method of choosing b , and the kernel choice can be examined by weighing FMS-G-O with FMS-G-AltVar, FMS-G-AltVar with T_{n_1, n_2} -G, and T_{n_1, n_2} -G with T_{n_1, n_2} -MG or T_{n_1, n_2} -NM, respectively.

The smoothing parameter b for FMS-G-O and FMS-G-AltVar is determined via making an adjustment for the value chosen by a cross-validation criterion; see p. 657 of FMS for details. On the other hand, the values of b for T_{n_1, n_2} -G, T_{n_1, n_2} -MG and T_{n_1, n_2} -NM are selected by the power-optimality criterion in the previous section. Implementation details are as follows: (i) all critical values in $\hat{\pi}_M(b_{k_1})$ are set at $z_{0.05} = 1.645$; (ii) the shrinking rate of b is set at $q = 4/9$ because of $N^{1/2}$ -consistency of least-squares estimates and Theorem 3; (iii) three different values are considered for δ , namely, $\delta \in \{0.3, 0.5, 0.7\}$; and (iv) the interval for b_{k_1} is set equal to $H_{k_1} = [0.01, 0.64]$. The sample size is $N \in \{50, 100, 200\}$, and 1000 replications are drawn for each combination of the sample size N and the distribution of u .

³ Although the GLDs corresponding to A3 and A4 are used in Zheng [17] and FMS, they are found to have non-zero means. Therefore, we adjust the values of λ_1 and λ_2 with skewness and kurtosis maintained so that the resulting distributions have means of zero.

5.2. Simulation Results

Table 2 presents finite-sample rejection frequencies of each test statistic against nominal 5% and 10% levels across 1000 Monte Carlo samples. Critical values are simply based on the first-order normal limit, *i.e.*, 1.645 and 1.280 correspond to the 5% and 10% levels, respectively.

Table 2. Size and Power of the SSST.

(A) Size			Distribution								(%)
N	Test	δ	S1		S2		S3		S4		
			5%	10%	5%	10%	5%	10%	5%	10%	
50	FMS-G-O	–	4.8	9.4	4.9	9.5	6.0	10.8	7.4	12.6	
	FMS-G-AltVar	–	4.7	9.3	4.4	8.9	5.8	10.9	6.9	12.6	
	T_{n_1, n_2} -G	0.3	3.5	7.5	3.1	7.0	4.5	8.7	5.2	9.6	
		0.5	3.8	6.8	3.0	7.3	4.4	7.9	5.4	10.2	
		0.7	3.4	6.7	3.4	7.3	4.5	8.9	5.4	10.5	
	T_{n_1, n_2} -MG	0.3	3.6	7.7	3.9	7.5	4.9	9.5	5.3	9.7	
		0.5	4.0	7.2	3.7	7.2	4.9	8.7	5.4	10.2	
		0.7	3.3	6.8	3.8	7.3	4.7	9.1	5.3	10.7	
	T_{n_1, n_2} -NM	0.3	3.4	7.3	3.1	6.7	4.3	8.3	5.2	9.3	
		0.5	3.9	6.7	3.1	6.4	4.2	7.9	5.2	9.8	
		0.7	3.3	7.0	3.1	6.4	4.4	8.1	5.2	10.3	
	100	FMS-G-O	–	5.2	9.8	6.7	10.8	5.6	10.1	7.4	12.5
FMS-G-AltVar		–	4.9	8.9	6.4	10.6	5.3	9.9	6.6	12.6	
T_{n_1, n_2} -G		0.3	4.0	7.0	6.1	9.3	5.1	8.8	7.3	11.9	
		0.5	3.6	7.3	5.8	9.1	5.2	8.9	7.1	11.4	
		0.7	3.8	7.6	5.9	9.5	4.8	9.3	6.6	11.8	
T_{n_1, n_2} -MG		0.3	4.5	7.4	6.2	9.5	5.8	9.6	7.3	12.1	
		0.5	3.7	7.7	6.1	9.6	5.7	9.0	7.2	11.6	
		0.7	3.8	8.1	5.8	9.6	5.1	9.4	6.7	12.1	
T_{n_1, n_2} -NM		0.3	4.2	6.8	5.3	8.9	4.3	8.4	7.1	12.4	
		0.5	4.0	6.6	5.3	8.7	4.9	8.7	7.2	11.6	
		0.7	4.2	6.8	5.3	8.8	4.9	8.4	7.3	11.7	
200		FMS-G-O	–	4.1	7.3	6.0	9.3	5.8	8.9	8.9	14.1
	FMS-G-AltVar	–	4.1	7.0	6.1	8.9	5.4	9.3	8.7	15.1	
	T_{n_1, n_2} -G	0.3	3.3	6.3	4.9	8.0	4.9	8.7	8.9	12.9	
		0.5	3.2	6.4	4.9	8.6	5.3	8.7	8.7	13.3	
		0.7	3.5	6.8	5.2	9.2	5.4	8.9	8.4	13.9	
	T_{n_1, n_2} -MG	0.3	3.5	6.7	5.5	8.4	5.5	9.4	9.0	13.4	
		0.5	3.5	6.8	5.2	9.0	5.7	8.8	8.7	13.6	
		0.7	3.5	7.0	5.5	9.8	5.7	9.1	8.6	13.7	
	T_{n_1, n_2} -NM	0.3	3.5	6.0	4.0	7.5	4.4	8.5	8.7	13.3	
		0.5	3.6	5.9	4.4	7.6	4.4	8.4	8.7	13.3	
		0.7	3.6	5.8	4.6	7.4	4.5	8.4	8.8	13.1	

Table 2. Cont.

(B) Power			Distribution								(%)
N	Test	δ	A1		A2		A3		A4		
			5%	10%	5%	10%	5%	10%	5%	10%	
50	FMS-G-O	–	42.0	51.6	21.4	30.3	26.3	37.2	28.8	40.8	
			[43.7]	[52.4]	[22.7]	[31.0]	[27.6]	[38.1]	[30.5]	[42.0]	
	FMS-G-AltVar	–	24.9	37.0	13.3	22.4	39.8	52.2	43.1	55.3	
			[27.5]	[39.0]	[13.9]	[24.1]	[41.5]	[53.6]	[44.9]	[57.4]	
	T_{n_1, n_2} -G	0.3	31.3	41.5	17.3	24.7	30.8	41.3	33.8	44.3	
			[35.4]	[46.6]	[20.8]	[29.5]	[35.1]	[46.3]	[38.7]	[50.8]	
	T_{n_1, n_2} -MG	0.3	29.5	39.7	16.0	23.8	29.4	39.1	31.4	42.9	
			[35.2]	[45.7]	[20.9]	[29.3]	[35.0]	[45.8]	[39.1]	[49.5]	
	T_{n_1, n_2} -NM	0.3	27.8	38.7	14.5	22.6	28.0	36.8	30.2	40.6	
			[35.2]	[45.7]	[20.9]	[29.3]	[35.0]	[45.8]	[39.1]	[49.5]	
	T_{n_1, n_2} -MG	0.5	30.1	40.8	16.5	24.5	30.2	40.3	32.2	43.1	
			[35.2]	[45.7]	[20.9]	[29.3]	[35.0]	[45.8]	[39.1]	[49.5]	
T_{n_1, n_2} -NM	0.5	28.0	38.9	15.1	23.0	28.5	37.5	30.8	41.1		
		[35.2]	[45.7]	[20.9]	[29.3]	[35.0]	[45.8]	[39.1]	[49.5]		
T_{n_1, n_2} -NM	0.7	33.8	42.5	18.2	25.2	32.9	42.5	35.2	46.5		
		[36.6]	[48.8]	[20.7]	[30.4]	[36.8]	[48.5]	[39.2]	[53.5]		
T_{n_1, n_2} -NM	0.7	33.5	42.2	17.7	24.9	32.4	42.2	34.6	45.8		
		[36.6]	[48.8]	[20.7]	[30.4]	[36.8]	[48.5]	[39.2]	[53.5]		
T_{n_1, n_2} -NM	0.7	33.0	41.4	17.4	24.2	32.2	41.6	34.5	45.7		
		[36.6]	[48.8]	[20.7]	[30.4]	[36.8]	[48.5]	[39.2]	[53.5]		
100	FMS-G-O	–	73.0	81.5	41.4	51.4	59.1	71.2	64.2	74.6	
			[72.7]	[81.8]	[40.7]	[52.3]	[58.7]	[71.8]	[63.8]	[74.6]	
	FMS-G-AltVar	–	56.4	70.2	30.3	43.4	73.7	80.7	77.2	83.7	
			[58.1]	[71.6]	[32.2]	[44.4]	[74.4]	[81.8]	[78.1]	[84.4]	
	T_{n_1, n_2} -G	0.3	72.3	80.5	37.9	48.6	70.3	78.4	74.1	80.5	
			[75.4]	[84.4]	[40.2]	[54.9]	[72.8]	[82.3]	[76.1]	[84.1]	
	T_{n_1, n_2} -MG	0.3	67.1	77.2	33.9	45.0	65.2	75.7	69.5	78.2	
			[75.4]	[84.4]	[40.2]	[54.9]	[72.8]	[82.3]	[76.1]	[84.1]	
	T_{n_1, n_2} -MG	0.5	62.9	73.6	31.9	42.5	61.7	72.3	65.2	75.8	
			[75.4]	[84.4]	[40.2]	[54.9]	[72.8]	[82.3]	[76.1]	[84.1]	
	T_{n_1, n_2} -MG	0.7	73.1	80.2	38.3	49.3	70.5	78.3	73.6	81.0	
			[74.8]	[83.6]	[40.9]	[53.5]	[72.4]	[80.7]	[75.4]	[83.3]	
T_{n_1, n_2} -NM	0.3	67.4	77.5	34.6	45.4	65.6	75.4	69.5	77.9		
		[74.8]	[83.6]	[40.9]	[53.5]	[72.4]	[80.7]	[75.4]	[83.3]		
T_{n_1, n_2} -NM	0.5	62.9	73.5	32.0	42.3	62.2	72.5	65.3	75.2		
		[74.8]	[83.6]	[40.9]	[53.5]	[72.4]	[80.7]	[75.4]	[83.3]		
T_{n_1, n_2} -NM	0.7	76.8	84.0	41.7	51.9	75.5	82.1	76.8	84.0		
		[79.6]	[87.2]	[44.8]	[58.0]	[77.6]	[85.0]	[79.7]	[86.9]		
T_{n_1, n_2} -NM	0.5	77.0	84.0	40.1	51.0	75.1	82.0	75.7	83.0		
		[79.6]	[87.2]	[44.8]	[58.0]	[77.6]	[85.0]	[79.7]	[86.9]		
T_{n_1, n_2} -NM	0.7	76.4	83.7	39.6	50.1	74.9	81.9	75.6	82.9		
		[79.6]	[87.2]	[44.8]	[58.0]	[77.6]	[85.0]	[79.7]	[86.9]		
200	FMS-G-O	–	97.4	98.3	71.5	80.7	95.6	98.1	97.1	97.8	
			[97.8]	[98.6]	[75.0]	[84.2]	[96.9]	[98.6]	[97.4]	[98.4]	
	FMS-G-AltVar	–	93.4	96.3	60.4	72.8	98.7	99.0	98.4	99.1	
			[95.2]	[97.5]	[69.1]	[78.5]	[98.8]	[99.2]	[99.0]	[99.2]	
	T_{n_1, n_2} -G	0.3	97.7	99.1	77.0	84.8	98.6	99.1	98.6	99.2	
			[98.7]	[99.4]	[82.7]	[90.3]	[98.8]	[99.4]	[98.9]	[99.2]	
	T_{n_1, n_2} -MG	0.3	97.2	98.1	71.3	80.9	97.7	98.9	97.9	98.7	
			[98.7]	[99.4]	[82.7]	[90.3]	[98.8]	[99.4]	[98.9]	[99.2]	
	T_{n_1, n_2} -MG	0.5	96.4	97.6	65.4	75.8	96.7	98.8	97.2	98.3	
			[98.6]	[99.4]	[81.8]	[89.8]	[98.7]	[99.4]	[98.7]	[99.2]	
	T_{n_1, n_2} -MG	0.7	97.9	99.1	76.2	85.3	98.6	99.1	98.5	99.1	
			[98.6]	[99.4]	[81.8]	[89.8]	[98.7]	[99.4]	[98.7]	[99.2]	
T_{n_1, n_2} -NM	0.3	97.3	98.1	71.4	80.3	97.6	98.9	97.9	98.6		
		[99.0]	[99.2]	[91.1]	[95.8]	[99.4]	[99.7]	[99.5]	[99.7]		
T_{n_1, n_2} -NM	0.5	96.5	97.6	65.5	75.9	96.7	98.7	97.2	98.3		
		[99.0]	[99.2]	[91.1]	[95.8]	[99.4]	[99.7]	[99.5]	[99.7]		
T_{n_1, n_2} -NM	0.7	98.9	99.0	84.8	91.1	98.8	99.2	98.6	98.9		
		[99.0]	[99.2]	[91.1]	[95.8]	[99.4]	[99.7]	[99.5]	[99.7]		
T_{n_1, n_2} -NM	0.5	96.8	97.0	81.2	88.2	98.3	98.6	98.4	98.6		
		[99.0]	[99.2]	[91.1]	[95.8]	[99.4]	[99.7]	[99.5]	[99.7]		
T_{n_1, n_2} -NM	0.7	95.6	95.7	80.9	88.3	98.3	98.6	98.4	98.6		
		[99.0]	[99.2]	[91.1]	[95.8]	[99.4]	[99.7]	[99.5]	[99.7]		

Note: Numbers in brackets are size-adjusted powers.

Panel (A) reports size properties. At first glance, we can find that the results of FMS-G-O are close to what is reported in Table 3 of FMS. It has the tendency of over-rejecting the null slightly against the nominal size. Comparing FMS-G-O with FMS-G-AltVar reveals that replacing the variance formula is likely to decrease the rejection frequencies. Changing the choice method of b further reduces the rejection frequencies, and T_{n_1, n_2} -G tends to result in mild under-rejection of the null. Effects of alternative kernel choices are mixed. While T_{n_1, n_2} -G and T_{n_1, n_2} -MG have similar size properties, T_{n_1, n_2} -NM looks more conservative in the sense that its rejection frequencies are slightly

smaller. Impacts of varying δ are found to be minor at best. A concern is that all test statistics exhibit size distortions for S4. However, the distribution is platykurtic and has sharp boundaries at ± 1 . A platykurtic distribution is an exception rather than a rule in economics and finance, and a distribution with a compact support violates Assumption 1. In sum, all test statistics exhibit good size properties, although their convergence rates are nonparametric ones, effective sample sizes are (roughly) a half of the entire sample size N , and no size correction devices such as bootstrapping are used.

Panel (B) refers to power properties. We can immediately see that the rejection frequencies of each test statistic approach to one with the sample size N , which confirms consistency of the SSST. There is substantial improvement in power as the sample size increases from $N = 50$ to 100. Most of rejection frequencies become nearly one for as small as $N = 200$. After a closer look, we can find it hard to judge whether changing the variance formula from FMS-G-O to FMS-G-AltVar may affect power properties favorably or adversely. However, once the smoothing parameter value is chosen via the power-optimality criterion, power properties are improved in general. Power properties of T_{n_1, n_2} -G and T_{n_1, n_2} -MG again look alike, whereas T_{n_1, n_2} -NM appears to be more powerful than these two. Because the power tends to decrease with δ , it could be safe to choose $\delta = 0.3$ from the viewpoint of power-maximization. Indeed, for $N = 200$ and $\delta = 0.3$, each of T_{n_1, n_2} -G, T_{n_1, n_2} -MG and T_{n_1, n_2} -NM exhibits better power properties than FMS-G-O and FMS-G-AltVar.

For convenience, Panel (B) presents size-adjusted powers, where the best case scenario (*i.e.*, $\delta = 0.3$) is considered for T_{n_1, n_2} -G, T_{n_1, n_2} -MG and T_{n_1, n_2} -NM. These three test statistics again outperform FMS's original statistics in terms of size-adjusted powers, and T_{n_1, n_2} -NM appears to have the best power properties among three. All in all, Monte Carlo results indicate superior size and power properties of the SSST with the GG kernels plugged in.

6. Conclusions

The SSST developed by FMS is built on the idea of gauging the closeness between right and left sides of the axis of symmetry of an unknown pdf. To implement the test, we split the entire sample into two sub-samples and estimate both sides of the pdf nonparametrically using asymmetric kernels with support on $[0, \infty)$. This paper has improved the SSST by combining it with the newly proposed GG kernels. The test statistic can be interpreted as a standardized version of a degenerate U -statistic. We deliver convergence properties of the test statistic and provide the asymptotic variance formulae for the cases of two sub-samples with equal and unequal sample sizes separately. It is demonstrated that the SSST smoothed by the GG kernels has a normal limit under the null of symmetry and is consistent under the alternative. As a part of the implementation method we also propose to select the smoothing parameter in a power-optimality criterion. Monte Carlo simulations indicate that the GG kernel-smoothed SSST with the power-maximized smoothing parameter value plugged in enjoys superior finite-sample properties. It should be stressed that the good performance of the SSST is grounded on the first-order normal limit and a small number of observations, despite its nonparametric convergence rate and sample-splitting procedure.

Appendix A. Appendix

Appendix A.1. Proof of Lemma 1

Because the proof for the MG kernel is basically the same as those for Lemmata 1(e) and 2 of FMS, we prove the case of the NM kernel. Among all statements, we concentrate on demonstrating that

$$E \{ K_{X_2}(X_1) K_{Y_2}(X_1) \} \sim E \{ f(X) g(X) \}, \text{ and} \quad (\text{A1})$$

$$E \{ K_{X_2}(Y_1) K_{Y_1}(X_2) \} \sim b^{-1/2} V_f(2) E \{ X^{-1/2} g(X) \}. \quad (\text{A2})$$

All the remaining statements can be shown in the same manner. To approximate the gamma function, we frequently refer to the following well-known formulae:

1. *Stirling’s formula (“SF”)*:

$$\Gamma(z + 1) = \sqrt{2\pi} z^{z+1/2} e^{-z} \left\{ 1 + \frac{1}{12z} + \frac{1}{288z^2} + O(z^{-3}) \right\} \text{ as } z \rightarrow \infty.$$

2. *Legendre’s duplication formula (“LDF”)*:

$$\Gamma(z) \Gamma\left(z + \frac{1}{2}\right) = \frac{\sqrt{\pi}}{2^{2z-1}} \Gamma(2z) \text{ for } z > 0.$$

In addition, proofs of the above statements require the following lemma. Its proof is virtually the same as those for Lemmata A.1 and A.2 of Fernandes and Monteiro [46], and thus it is omitted.

Lemma A1. For a constant $D > 0$ and two numbers $x, y > 0$,

$$\exp\left\{-\frac{(y-x)^2}{Dx}\right\} \leq \left(\frac{x}{y}\right)^{\frac{y-x}{D}} \leq \exp\left\{-\frac{(y-x)^2}{Dx} + \frac{(y-x)^3}{2Dx^2}\right\} \tag{A3}$$

if $x \leq y$, and

$$\exp\left\{-\frac{(y-x)^2}{Dx} + \frac{(y-x)^3}{2Dy^2}\right\} \leq \left(\frac{x}{y}\right)^{\frac{y-x}{D}} \leq \exp\left\{-\frac{(y-x)^2}{Dx}\right\} \tag{A4}$$

if $y < x$.

Proof of (A1). We apply the trimming argument as on p. 476 of Chen [30]. For some $\epsilon \in (0, 1/2)$,

$$\begin{aligned} & E\{K_{X_2}(X_1) K_{Y_2}(X_1)\} \\ &= \int_{b^{1-\epsilon}}^{\infty} \left[\int_{b^{1-\epsilon}}^{\infty} \left\{ \int_0^{\infty} L_b(u; x, y) f(x_1) dx_1 \right\} g(y) dy \right] f(x) dx + O(b^{1-\epsilon}), \end{aligned}$$

where $L_b(u; x, y) := K_{NM}(u; x, b) K_{NM}(u; y, b)$ for interior x, y . Then, the proof takes a multi-step approach including the following steps:

Step 1: approximating $J(x, y) := \int_0^{\infty} L_b(u; x, y) f(u) du$.

Step 2: approximating $J := \int_{b^{1-\epsilon}}^{\infty} \left\{ \int_{b^{1-\epsilon}}^{\infty} J(x, y) g(y) dy \right\} f(x) dx$.

Step 1: Define

$$P_z(b) := \frac{z\Gamma\left(\frac{z}{2b}\right)}{\Gamma\left(\frac{z}{2b} + \frac{1}{2}\right)} \tag{A5}$$

for $z = x, y, x + y$. Then,

$$\begin{aligned} L_b(u; x, y) &= \frac{2 \left\{ P_x(b) P_y(b) / \sqrt{P_x^2(b) + P_y^2(b)} \right\}^{\frac{x+y}{b}-1} \Gamma\left(\frac{x+y}{2b} - \frac{1}{2}\right)}{P_x^{x/b}(b) \Gamma\left(\frac{x}{2b}\right) P_y^{y/b}(b) \Gamma\left(\frac{y}{2b}\right)} \\ &\quad \times \frac{2u^{\left(\frac{x+y}{b}-1\right)-1} \exp\left[-\left\{\frac{u}{P_x(b)P_y(b)/\sqrt{P_x^2(b)+P_y^2(b)}}\right\}^2\right]}{\left\{ P_x(b) P_y(b) / \sqrt{P_x^2(b) + P_y^2(b)} \right\}^{\frac{x+y}{b}-1} \Gamma\left(\frac{x+y}{2b} - \frac{1}{2}\right)}, \end{aligned} \tag{A6}$$

where the first term is denoted as $B_b(x, y)$, and the second term can be viewed as the pdf of $GG\left(\frac{x+y}{b} - 1, P_x(b) P_y(b) / \sqrt{P_x^2(b) + P_y^2(b)}, 2\right)$. Moreover, $B_b(x, y)$ can be further rewritten as

$$B_{1b} B_{2b} B_{3b} := \frac{2\Gamma\left(\frac{x+y}{2b} - \frac{1}{2}\right) \{P_x^2(b) + P_y^2(b)\}^{1/2}}{\Gamma\left(\frac{x}{2b}\right) \Gamma\left(\frac{y}{2b}\right)} \frac{P_x^{y/b}(b) P_y^{x/b}(b)}{P_x(b) P_y(b) \{P_x^2(b) + P_y^2(b)\}^{\frac{x+y}{2b}}}, \tag{A7}$$

and an approximation to each of B_{1b} , B_{2b} and B_{3b} is provided separately.

By LDF, B_{1b} becomes

$$B_{1b} = \left(\frac{2}{\frac{x+y}{2b} - \frac{1}{2}}\right) \frac{\Gamma\left(\frac{x+y}{2b} + \frac{1}{2}\right)}{\Gamma\left(\frac{x}{2b}\right) \Gamma\left(\frac{y}{2b}\right)} = \frac{b\sqrt{\pi}}{2^{\frac{x+y}{b}-3} (x+y-b)} \frac{\Gamma\left(\frac{x+y}{2b}\right)}{\Gamma\left(\frac{x}{2b}\right) \Gamma\left(\frac{y}{2b}\right) \Gamma\left(\frac{x+y}{2b}\right)}.$$

Then, by SF, an approximation to B_{1b} is given by

$$B_{1b} = \sqrt{\frac{2}{\pi}} \left(\frac{\sqrt{xy}}{x+y}\right) \left(\frac{x+y}{x}\right)^{\frac{x}{2b}} \left(\frac{x+y}{y}\right)^{\frac{y}{2b}} \{1 + o(1)\}. \tag{A8}$$

Next, it follows from LDF and SF that

$$P_z(b) = \frac{2^{z/b-1} z \Gamma^2\left(\frac{z}{2b}\right)}{\sqrt{\pi} \Gamma\left(\frac{z}{b}\right)} = (2bz)^{1/2} \left\{1 + \frac{b}{4z} + O(b^2)\right\}. \tag{A9}$$

Hence,

$$B_{2b}^2 = \frac{1}{P_x^2(b)} + \frac{1}{P_y^2(b)} = \frac{b^{-1}}{2} \left(\frac{x+y}{xy}\right) \{1 + o(1)\},$$

and thus

$$B_{2b} = \frac{b^{-1/2}}{\sqrt{2}} \left(\frac{\sqrt{x+y}}{\sqrt{xy}}\right) \{1 + o(1)\}. \tag{A10}$$

Furthermore, (A9) also implies that

$$\begin{aligned} P_x^{y/b}(b) &= (2bx)^{\frac{y}{2b}} \exp\left(\frac{y}{4x}\right) \{1 + o(1)\}; \\ P_y^{x/b}(b) &= (2by)^{\frac{x}{2b}} \exp\left(\frac{x}{4y}\right) \{1 + o(1)\}; \text{ and} \\ \{P_x^2(b) + P_y^2(b)\}^{\frac{x+y}{2b}} &= \{2b(x+y)\}^{\frac{x+y}{2b}} \exp\left(\frac{1}{2}\right) \{1 + o(1)\}. \end{aligned}$$

Then,

$$B_{3b} = x^{\frac{y}{2b}} y^{\frac{x}{2b}} (x+y)^{-\frac{x+y}{2b}} \exp\left\{\frac{1}{4}\left(\frac{y}{x} - 1\right)\right\} \exp\left\{\frac{1}{4}\left(\frac{x}{y} - 1\right)\right\} \{1 + o(1)\}. \tag{A11}$$

Substituting (A8), (A10) and (A11) into (A7) finally yields

$$B_b(x, y) := b^{-1/2} \tilde{B}_b(x, y) \left(\frac{x}{y}\right)^{\frac{y-x}{2b}} \{1 + o(1)\},$$

where

$$\tilde{B}_b(x, y) = \frac{1}{\sqrt{\pi}\sqrt{x+y}} \exp\left\{\frac{1}{4}\left(\frac{y}{x} - 1\right)\right\} \exp\left\{\frac{1}{4}\left(\frac{x}{y} - 1\right)\right\}.$$

Then, for a random variable $\zeta_x \stackrel{d}{=} GG \left((x + y) / b - 1, P_x(b) P_y(b) / \sqrt{P_x^2(b) + P_y^2(b)}, 2 \right)$,

$$\begin{aligned} J(x, y) &= \int_0^\infty L_b(u; x, y) f(u) du \{1 + o(1)\} \\ &= b^{-1/2} \tilde{B}_b(x, y) \left(\frac{x}{y} \right)^{\frac{y-x}{2b}} E \{f(\zeta_x)\} \{1 + o(1)\}. \end{aligned}$$

By the property of GG random variables, (A5), (A9), and (A10),

$$E(\zeta_x) = B_{2b}^{-1} \frac{\Gamma\left(\frac{x+y}{2b}\right)}{\Gamma\left(\frac{x+y}{2b} - \frac{1}{2}\right)} = \sqrt{xy} \{1 + O(b)\}.$$

In the end, a first-order Taylor expansion of $f(\zeta_x)$ around $\zeta_x = \sqrt{xy}$ gives

$$J(x, y) = b^{-1/2} \tilde{B}_b(x, y) f(\sqrt{xy}) \left(\frac{x}{y} \right)^{\frac{y-x}{2b}} \{1 + o(1)\},$$

which completes Step 1.

Step 2: For some $t \in (0, 1)$, we split the interval for y into four subintervals as follows:

$$\begin{aligned} J &= \int_{b^{1-\epsilon}}^\infty \left\{ \int_{b^{1-\epsilon}}^{(1-t)x} + \int_{(1-t)x}^x + \int_x^{(1+t)x} + \int_{(1+t)x}^\infty J(x, y) g(y) dy \right\} f(x) dx \\ &= J_1 + J_2 + J_3 + J_4 \text{ (say)}. \end{aligned}$$

Also denote $h(x, y) := \tilde{B}_b(x, y) f(\sqrt{xy}) g(y)$. Then, by (A4) and the change of variable $v := (y - x) / \sqrt{2bx}$,

$$\begin{aligned} J_1 &\leq \int_{b^{1-\epsilon}}^\infty \left[\int_0^{(1-t)x} b^{-1/2} h(x, y) \exp\left\{-\frac{(y-x)^2}{2bx}\right\} \{1 + o(1)\} dy \right] f(x) dx \\ &\leq \int_{b^{1-\epsilon}}^\infty \sqrt{2x} \left[\int_{-\frac{1}{\sqrt{2}}\sqrt{\frac{x}{b}}}^{-\frac{t}{\sqrt{2}}\sqrt{\frac{x}{b}}} h\left(x, x + v\sqrt{2bx}\right) e^{-v^2} \{1 + o(1)\} dv \right] f(x) dx \rightarrow 0 \end{aligned}$$

as $b \rightarrow 0$.

Next, it follows from (A4) that

$$\begin{aligned} J_2 &\geq \int_{b^{1-\epsilon}}^\infty \left[\int_{(1-t)x}^x b^{-1/2} h(x, y) \exp\left\{-\frac{(y-x)^2}{2bx} + \frac{(y-x)^3}{4by^2}\right\} \{1 + o(1)\} dy \right] f(x) dx \\ &\geq \int_{b^{1-\epsilon}}^\infty \left[\int_{(1-t)x}^x b^{-1/2} h(x, y) \exp\left\{-\frac{(y-x)^2}{2bx} (1 + \tau_1)\right\} \{1 + o(1)\} dy \right] f(x) dx, \end{aligned}$$

where $\tau_1 := t(2-t) / \{2(1-t)^2\}$. By the change of variable $w := (y - x) \sqrt{(1 + \tau_1) / (2bx)}$, the right-hand side becomes

$$\int_{b^{1-\epsilon}}^\infty \sqrt{\frac{2x}{1 + \tau_1}} \left[\int_{-t\sqrt{\frac{1+\tau_1}{2}}\sqrt{\frac{x}{b}}}^0 h\left(x, x + w\sqrt{\frac{2bx}{1 + \tau_1}}\right) e^{-w^2} \{1 + o(1)\} dw \right] f(x) dx.$$

Because $\int_{-\infty}^0 e^{-v^2} dv = \sqrt{\pi}/2$ and $h(x, x) = f(x)g(x) / \sqrt{2\pi x}$, we have

$$\liminf_{b \rightarrow 0} J_2 = \frac{1}{2} \sqrt{\frac{1}{1 + \tau_1}} \int_0^\infty f(x)g(x) dF(x) \rightarrow \frac{1}{2} E\{f(X)g(X)\}$$

by letting t shrink toward zero. On the other hand, again by (A4) and the change of variable $v = (y - x) / \sqrt{2bx}$,

$$\begin{aligned} J_2 &\leq \int_{b^{1-\epsilon}}^\infty \left[\int_{(1-t)x}^x b^{-1/2} h(x, y) \exp\left\{-\frac{(y-x)^2}{2bx}\right\} \{1 + o(1)\} dy \right] f(x) dx \\ &\leq \int_{b^{1-\epsilon}}^\infty \sqrt{2x} \left[\int_{-\frac{t}{\sqrt{2}}\sqrt{\frac{x}{b}}}^0 h\left(x, x + v\sqrt{2bx}\right) e^{-v^2} \{1 + o(1)\} dv \right] f(x) dx, \end{aligned}$$

so that

$$\limsup_{b \rightarrow 0} J_2 = \frac{1}{2} \int_0^\infty f(x)g(x) dF(x) = \frac{1}{2} E\{f(X)g(X)\}.$$

Hence, we can conclude that $J_2 \rightarrow (1/2) E\{f(X)g(X)\}$.

It can be also demonstrated that $J_3 \rightarrow (1/2) E\{f(X)g(X)\}$ and $J_4 \rightarrow 0$ with the assistance of (A3). Therefore, $J \rightarrow E\{f(X)g(X)\}$, and thus (A1) is established. □

Proof of (A2). Again for some $\epsilon \in (0, 1/2)$,

$$E\{K_{X_2}(Y_1)K_{Y_1}(X_2)\} = \int_{b^{1-\epsilon}}^\infty \left[\int_{b^{1-\epsilon}}^\infty M_b(x, y)g(y) dy \right] f(x) dx + O(b^{-\epsilon}),$$

where $\Lambda_b(x, y) := K_{NM}(y; x, b)K_{NM}(x; y, b)$ for interior x, y and the order of the remainder term is $O(b^{-\epsilon}) = o(b^{-1/2})$ by construction. Observe that

$$\Lambda_b(x, y) = \frac{4y^{x/b-1}x^{y/b-1}}{P_x^{x/b}(b)P_y^{y/b}(b)\Gamma\left(\frac{x}{2b}\right)\Gamma\left(\frac{y}{2b}\right)} \exp\left\{-\frac{y^2}{P_x^2(b)}\right\} \exp\left\{-\frac{x^2}{P_y^2(b)}\right\}. \tag{A12}$$

It follows from (A9) that

$$P_z^{z/b}(b) = (2bz)^{\frac{z}{2b}} \exp\left(\frac{1}{4}\right) \{1 + o(1)\} \tag{A13}$$

for $z = x, y$. Similarly,

$$\exp\left\{-\frac{y^2}{P_x^2(b)}\right\} = \exp\left(-\frac{y^2}{2bx}\right) \exp\left\{\frac{1}{4}\left(\frac{y}{x}\right)^2\right\} \{1 + o(1)\}, \text{ and} \tag{A14}$$

$$\exp\left\{-\frac{x^2}{P_y^2(b)}\right\} = \exp\left(-\frac{x^2}{2by}\right) \exp\left\{\frac{1}{4}\left(\frac{x}{y}\right)^2\right\} \{1 + o(1)\}. \tag{A15}$$

Substituting (A13)–(A15) into (A12) and using SF, we have

$$\begin{aligned} \Lambda_b(x, y) &= \frac{b^{-1}}{\pi\sqrt{x}\sqrt{y}} \left(\frac{x}{y}\right)^{(y-x)/b} \exp\left\{-\frac{(y+x)(y-x)^2}{2bxy}\right\} \\ &\quad \times \exp\left[\frac{1}{4}\left\{\left(\frac{y}{x}\right)^2 - 1\right\}\right] \exp\left[\frac{1}{4}\left\{\left(\frac{x}{y}\right)^2 - 1\right\}\right] \{1 + o(1)\}. \end{aligned}$$

As before, for some $t \in (0, 1)$, consider

$$\begin{aligned} \Lambda &= \int_{b^{1-\epsilon}}^{\infty} \left\{ \int_{b^{1-\epsilon}}^{(1-t)x} + \int_{(1-t)x}^x + \int_x^{(1+t)x} + \int_{(1+t)x}^{\infty} \Lambda_b(x, y) g(y) dy \right\} f(x) dx \\ &= \Lambda_1 + \Lambda_2 + \Lambda_3 + \Lambda_4 \text{ (say)}. \end{aligned}$$

It follows from (A4) that

$$\begin{aligned} \Lambda_1 &\leq \int_{b^{1-\epsilon}}^{\infty} \left[\int_0^{(1-t)x} \frac{b^{-1}}{\pi\sqrt{x}\sqrt{y}} \exp \left\{ -\frac{(y-x)^2}{bx} - \frac{(y+x)(y-x)^2}{2bxy} \right\} \right. \\ &\quad \times \exp \left\{ \frac{1}{4} \left(\left(\frac{y}{x} \right)^2 - 1 \right) \right\} \exp \left\{ \frac{1}{4} \left(\left(\frac{x}{y} \right)^2 - 1 \right) \right\} \{1 + o(1)\} g(y) dy \Big] f(x) dx \\ &\leq \int_{b^{1-\epsilon}}^{\infty} \left[\int_0^{(1-t)x} \frac{b^{-1}}{\pi\sqrt{x}\sqrt{y}} \exp \left\{ -\frac{\tau_2 (y-x)^2}{2bx} \right\} \right. \\ &\quad \times \exp \left\{ \frac{1}{4} \left(\left(\frac{y}{x} \right)^2 - 1 \right) \right\} \exp \left\{ \frac{1}{4} \left(\left(\frac{x}{y} \right)^2 - 1 \right) \right\} \{1 + o(1)\} g(y) dy \Big] f(x) dx, \end{aligned}$$

where $\tau_2 := (4 - t) / (1 - t)$. Then, by the change of variable $\eta := (y - x) \sqrt{\tau_2 / (2bx)}$,

$$\begin{aligned} b^{1/2} \Lambda_1 &\leq \int_{b^{1-\epsilon}}^{\infty} \frac{1}{\pi} \sqrt{\frac{2}{\tau_2}} \left[\int_{-\sqrt{\frac{\tau_2}{2}} \sqrt{\frac{x}{b}}}^{-t\sqrt{\frac{\tau_2}{2}} \sqrt{\frac{x}{b}}} \frac{1}{\sqrt{x + \eta \sqrt{\frac{2bx}{\tau_2}}}} e^{-\eta^2} \exp \left\{ \frac{1}{4} \left(\left(\frac{x + \eta \sqrt{\frac{2bx}{\tau_2}}}{x} \right)^2 - 1 \right) \right\} \right. \\ &\quad \times \exp \left\{ \frac{1}{4} \left(\left(\frac{x}{x + \eta \sqrt{\frac{2bx}{\tau_2}}} \right)^2 - 1 \right) \right\} \{1 + o(1)\} g \left(x + \eta \sqrt{\frac{2bx}{\tau_2}} \right) d\eta \Big] f(x) dx \\ &\rightarrow 0, \end{aligned}$$

or $\Lambda_1 = o(b^{-1/2})$.

Next, (A4) implies that

$$\begin{aligned} \Lambda_2 &\geq \int_{b^{1-\epsilon}}^{\infty} \left[\int_{(1-t)x}^x \frac{b^{-1}}{\pi\sqrt{x}\sqrt{y}} \exp \left\{ -\frac{(y-x)^2}{bx} + \frac{(y-x)^3}{2by^2} - \frac{(y+x)(y-x)^2}{2bxy} \right\} \right. \\ &\quad \times \exp \left\{ \frac{1}{4} \left(\left(\frac{y}{x} \right)^2 - 1 \right) \right\} \exp \left\{ \frac{1}{4} \left(\left(\frac{x}{y} \right)^2 - 1 \right) \right\} \{1 + o(1)\} g(y) dy \Big] f(x) dx \\ &\geq \int_{b^{1-\epsilon}}^{\infty} \left[\int_{(1-t)x}^x \frac{b^{-1}}{\pi\sqrt{x}\sqrt{y}} \exp \left\{ -\frac{(y-x)^2}{2bx} (3 + \tau_3) \right\} \right. \\ &\quad \times \exp \left\{ \frac{1}{4} \left(\left(\frac{y}{x} \right)^2 - 1 \right) \right\} \exp \left\{ \frac{1}{4} \left(\left(\frac{x}{y} \right)^2 - 1 \right) \right\} \{1 + o(1)\} g(y) dy \Big] f(x) dx, \end{aligned}$$

where $\tau_3 := (1 - t)^{-2}$. By the change of variable $\mu := (y - x) \sqrt{(3 + \tau_3) / (2bx)}$, the right-hand side becomes

$$\int_{b^{1-\epsilon}}^{\infty} \frac{b^{-1/2}}{\pi} \sqrt{\frac{2}{3+\tau_3}} \left[\int_{\sqrt{\frac{3+\tau_3}{2}} \sqrt{\frac{x}{b}}}^0 \frac{1}{\sqrt{x + \mu \sqrt{\frac{2bx}{3+\tau_3}}}} e^{-\mu^2} \exp \left\{ \frac{1}{4} \left(\left(\frac{x + \mu \sqrt{\frac{2bx}{3+\tau_3}}}{x} \right)^2 - 1 \right) \right\} \right. \\ \left. \times \exp \left\{ \frac{1}{4} \left(\left(\frac{x}{x + \mu \sqrt{\frac{2bx}{3+\tau_3}}} \right)^2 - 1 \right) \right\} \{1 + o(1)\} g \left(x + \mu \sqrt{\frac{2bx}{3+\tau_3}} \right) d\mu \right] f(x) dx,$$

so that

$$\liminf_{b \rightarrow 0} b^{1/2} \Lambda_2 = \frac{1}{2\sqrt{\pi}} \sqrt{\frac{2}{3+\tau_3}} \int_0^{\infty} x^{-1/2} g(x) dF(x) \rightarrow \frac{1}{2} V_{I,NM}(2) E \{ X^{-1/2} g(X) \}$$

by letting t shrink toward zero, where $V_{I,NM}(2) := 1/\sqrt{2\pi}$. Notice that we may safely assume that $E \{ X^{-1/2} g(X) \} < \infty$: Assumption 2 ensures that f and g are bounded, and thus it must be the case that $x^{-1/2} f(x) g(x) \leq cx^{-1/2}$ in the vicinity of the origin. On the other hand, (A4) also yields

$$\Lambda_2 \leq \int_{b^{1-\epsilon}}^{\infty} \left[\int_{(1-t)x}^x \frac{b^{-1}}{\pi \sqrt{x} \sqrt{y}} \exp \left\{ -\frac{(y-x)^2}{bx} - \frac{(y+x)(y-x)^2}{2bxy} \right\} \right. \\ \left. \times \exp \left\{ \frac{1}{4} \left(\left(\frac{y}{x} \right)^2 - 1 \right) \right\} \exp \left\{ \frac{1}{4} \left(\left(\frac{x}{y} \right)^2 - 1 \right) \right\} \{1 + o(1)\} g(y) dy \right] f(x) dx \\ \leq \int_{b^{1-\epsilon}}^{\infty} \left[\int_{(1-t)x}^x \frac{b^{-1}}{\pi \sqrt{x} \sqrt{y}} \exp \left\{ -\frac{2(y-x)^2}{bx} \right\} \right. \\ \left. \times \exp \left\{ \frac{1}{4} \left(\left(\frac{y}{x} \right)^2 - 1 \right) \right\} \exp \left\{ \frac{1}{4} \left(\left(\frac{x}{y} \right)^2 - 1 \right) \right\} \{1 + o(1)\} g(y) dy \right] f(x) dx.$$

By the change of variable $\omega := (y-x) \sqrt{2/(bx)}$,

$$\Lambda_2 \leq \int_{b^{1-\epsilon}}^{\infty} \frac{b^{-1/2}}{\sqrt{2\pi}} \left[\int_{-t\sqrt{2}\sqrt{\frac{x}{b}}}^0 \frac{1}{\sqrt{x + \omega \sqrt{\frac{bx}{2}}}} e^{-\omega^2} \exp \left\{ \frac{1}{4} \left(\left(\frac{x + \omega \sqrt{\frac{bx}{2}}}{x} \right)^2 - 1 \right) \right\} \right. \\ \left. \times \exp \left\{ \frac{1}{4} \left(\left(\frac{x}{x + \omega \sqrt{\frac{bx}{2}}} \right)^2 - 1 \right) \right\} \{1 + o(1)\} g \left(x + \omega \sqrt{\frac{bx}{2}} \right) d\omega \right] f(x) dx,$$

and thus

$$\limsup_{b \rightarrow 0} b^{1/2} \Lambda_2 = \frac{1}{2\sqrt{2\pi}} \int_0^{\infty} x^{-1/2} g(x) dF(x) = \frac{1}{2} V_{I,NM}(2) E \{ X^{-1/2} g(X) \}.$$

Hence, we can conclude that $\Lambda_2 \sim b^{-1/2} (1/2) V_{I,NM}(2) E \{ X^{-1/2} g(X) \}$.

It also follows from (A3) that $\Lambda_3 \sim b^{-1/2} V_{I,NM}(2) (1/2) E \{ X^{-1/2} g(X) \}$ and $\Lambda_4 = o(b^{-1/2})$. Therefore, $\Lambda \sim b^{-1/2} V_{I,NM}(2) E \{ X^{-1/2} g(X) \}$, and thus (A2) is also established. \square

Appendix A.2. Proof of Theorem 1

Because (ii) is obvious given that (i) is true, we concentrate only on (i). The proof strategy for (i) largely follows the one for Theorem 1.1 of Fernandes and Monteiro [46]. The proof of (i) also requires three lemmata below.

Lemma A2. Let (X_1, X_2) and (Y_1, Y_2) be two independent copies of X and Y , respectively. Then, under Assumptions 1–3, the followings hold:

- (a) $E \left\{ K_{X_2}^2(X_1) \right\} \sim b^{-1/2} V_I(2) E \left\{ X^{-1/2} f(X) \right\}$; $E \left\{ K_{X_2}^2(Y_1) \right\} \sim b^{-1/2} V_I(2) E \left\{ X^{-1/2} g(X) \right\}$;
 $E \left\{ K_{Y_2}^2(X_1) \right\} \sim b^{-1/2} V_I(2) E \left\{ Y^{-1/2} f(Y) \right\}$; and $E \left\{ K_{Y_2}^2(Y_1) \right\} \sim b^{-1/2} V_I(2) E \left\{ Y^{-1/2} g(Y) \right\}$,
 where $V_I(2)$ is given in Condition 5 of Definition 1.
- (b) $E \left\{ K_{X_2}(X_1) K_{Y_2}(Y_1) \right\} \sim E \left\{ f(X) \right\} E \left\{ g(Y) \right\}$; $E \left\{ K_{X_2}(Y_1) K_{Y_2}(X_1) \right\} \sim E \left\{ g(X) \right\} E \left\{ f(Y) \right\}$;
 $E \left\{ K_{X_2}(Y_1) K_{X_1}(Y_2) \right\} \sim E^2 \left\{ g(X) \right\}$; and $E \left\{ K_{Y_2}(X_1) K_{Y_1}(X_2) \right\} \sim E^2 \left\{ f(Y) \right\}$.
- (c) $E \left\{ K_{X_2}(X_1) K_{X_2}(Y_1) \right\} \sim E \left\{ f(X) g(X) \right\}$; and $E \left\{ K_{Y_2}(Y_1) K_{Y_2}(X_1) \right\} \sim E \left\{ g(Y) f(Y) \right\}$.
- (d) $E \left\{ K_{X_2}(X_1) K_{X_1}(Y_2) \right\} \sim E \left\{ f(X) g(X) \right\}$; $E \left\{ K_{Y_1}(X_2) K_{X_2}(X_1) \right\} \sim E \left\{ f^2(Y) \right\}$;
 $E \left\{ K_{X_1}(Y_2) K_{Y_2}(Y_1) \right\} \sim E \left\{ g^2(X) \right\}$; and $E \left\{ K_{Y_2}(Y_1) K_{Y_1}(X_2) \right\} \sim E \left\{ f(Y) g(Y) \right\}$.

Lemma A3. If Assumptions 1–4 and $n_1 = n_2 = n$ hold, then

$$E \left\{ \Phi_n^2(Z_1, Z_2) \right\} \sim \frac{4V_I(2)}{n^4 b^{1/2}} E \left[X^{-1/2} \{f(X) + g(X)\} + Y^{-1/2} \{f(Y) + g(Y)\} \right].$$

Lemma A4. If Assumptions 1–4 and $n_1 = n_2 = n$ hold, then $E \left\{ \Phi_n^{2k}(Z_1, Z_2) \right\} < \infty$ and

$$\frac{E \left\{ Y_n^k(Z_1, Z_2) \right\} + n^{1-k} E \left\{ \Phi_n^{2k}(Z_1, Z_2) \right\}}{E^k \left\{ \Phi_n^2(Z_1, Z_2) \right\}} \rightarrow 0$$

for some $k \in (1, 3/2)$, where $Y_n(x, y) := E \left\{ \Phi_n(Z_1, x) \Phi_n(y, Z_1) \right\}$.

Appendix A.2.1. Proof of Lemma A2

The variance approximation in Theorem 1 of Hirukawa and Sakudo [32] and the trimming argument on p.476 of Chen [30] yield (a). On the other hand, the bias approximation in Theorem 1 of Hirukawa and Sakudo [32] is applied to (b)–(d). As a consequence, (b) can be established by recognizing that $E \left\{ K_{X_2}(X_1) K_{Y_2}(Y_1) \right\} = E \left\{ K_{X_2}(X_1) \right\} E \left\{ K_{X_2}(X_1) \right\}$, for instance. Moreover, (c) and (d) follow from the proofs for (d) and (f) in Lemma A1 of FMS. □

Appendix A.2.2. Proof of Lemma A3

Because $\phi_n(Z_1, Z_2) \stackrel{d}{=} \phi_n(Z_2, Z_1)$, we have

$$E \left\{ \Phi_n^2(Z_1, Z_2) \right\} = \frac{2}{n^4} \left[E \left\{ \phi_n^2(Z_1, Z_2) \right\} + E \left\{ \phi_n(Z_1, Z_2) \phi_n(Z_2, Z_1) \right\} \right].$$

With the assistance of Assumption 4 and Lemma A2, we can pick out the leading terms of $E \left\{ \phi_n^2(Z_1, Z_2) \right\}$ and $E \left\{ \phi_n(Z_1, Z_2) \phi_n(Z_2, Z_1) \right\}$ as:

$$\begin{aligned}
 & E \left\{ \phi_n^2 (Z_1, Z_2) \right\} \\
 &= E \left\{ K_{X_2}^2 (X_1) \right\} + E \left\{ K_{X_2}^2 (Y_1) \right\} + E \left\{ K_{Y_2}^2 (X_1) \right\} + E \left\{ K_{Y_2}^2 (Y_1) \right\} + O(1) \\
 &= b^{-1} V_I(2) E \left[X^{-1/2} \{f(X) + g(X)\} + Y^{-1/2} \{f(Y) + g(Y)\} \right] + o(b^{-1}); \text{ and} \\
 & E \left\{ \phi_n (Z_1, Z_2) \phi_n (Z_2, Z_1) \right\} \\
 &= E \left\{ K_{X_2} (X_1) K_{X_1} (X_2) \right\} + E \left\{ K_{X_2} (Y_1) K_{Y_1} (X_2) \right\} \\
 &+ E \left\{ K_{Y_2} (X_1) K_{X_1} (Y_2) \right\} + E \left\{ K_{Y_2} (Y_1) K_{Y_1} (Y_2) \right\} + O(1) \\
 &= b^{-1} V_I(2) E \left[X^{-1/2} \{f(X) + g(X)\} + Y^{-1/2} \{f(Y) + g(Y)\} \right] + o(b^{-1}).
 \end{aligned}$$

The result immediately follows. □

Appendix A.2.3. Proof of Lemma A4

It follows from Lemma A3 that

$$E^k \left\{ \Phi_n^2 (Z_1, Z_2) \right\} = O \left(n^{-4k} b^{-k/2} \right). \tag{A16}$$

Next, by Jensen’s and C_r -inequalities,

$$\begin{aligned}
 & E \left\{ Y_n^k (Z_1, Z_2) \right\} \\
 &= E_{Z_1, Z_2} \left[E_{Z_3}^k \left\{ \Phi_n (Z_3, Z_1) \Phi_n (Z_2, Z_3) \right\} \right] \\
 &\leq E_{Z_1, Z_2} \left[E_{Z_3} \left\{ \Phi_n (Z_3, Z_1) \Phi_n (Z_2, Z_3) \right\}^k \right] \\
 &\leq n^{-4k} E_{Z_1, Z_2} \left\{ E_{Z_3} |\phi_n (Z_3, Z_1) \phi_n (Z_2, Z_3) + \phi_n (Z_3, Z_1) \phi_n (Z_3, Z_2) \right. \\
 &\quad \left. + \phi_n (Z_1, Z_3) \phi_n (Z_2, Z_3) + \phi_n (Z_1, Z_3) \phi_n (Z_3, Z_2) \right\}^k \\
 &\leq n^{-4k} 2^{2(k-1)} E_{Z_1, Z_2} \left[E_{Z_3} |\phi_n (Z_3, Z_1) \phi_n (Z_2, Z_3)|^k + E_{Z_3} |\phi_n (Z_3, Z_1) \phi_n (Z_3, Z_2)|^k \right. \\
 &\quad \left. + E_{Z_3} |\phi_n (Z_1, Z_3) \phi_n (Z_2, Z_3)|^k + E_{Z_3} |\phi_n (Z_1, Z_3) \phi_n (Z_3, Z_2)|^k \right] \\
 &= n^{-4k} 2^{2(k-1)} (Y_1 + Y_2 + Y_3 + Y_4) \text{ (say)}.
 \end{aligned}$$

Furthermore, applying C_r -inequality repeatedly yields

$$Y_1 \leq 2^{4(k-1)} \cdot 8 \left[E \left\{ K_{X_1}^k (X_3) K_{X_3}^k (X_2) \right\} + E^2 \left\{ K_{X_1}^k (X_3) \right\} \right]$$

under H_0 . Essentially the same arguments as in the proofs of Lemmata 1 and A2 establish that $E \left\{ K_{X_1}^k (X_3) K_{X_3}^k (X_2) \right\}$ is bounded by $cb^{1-k} \int_0^\infty x^{1-k} f^3(x) dx$. It follows from $k < 3/2$ that $x^{1-k} f^3(x) \leq cx^{-1/2}$ in the neighborhood of the origin, and thus $\int_0^\infty x^{1-k} f^3(x) dx < \infty$ holds. Hence, $E \left\{ K_{X_1}^k (X_3) K_{X_3}^k (X_2) \right\} \leq O(b^{1-k})$. Similarly, $E^2 \left\{ K_{X_1}^k (X_3) \right\} \leq O \left\{ \left(b^{(1-k)/2} \right)^2 \right\} = O(b^{1-k})$, and thus $Y_1 \leq O(b^{1-k})$. It can be also shown that each of Y_2, Y_3 and Y_4 is bounded by $O(b^{1-k})$. As a result,

$$E \left\{ Y_n^k (Z_1, Z_2) \right\} \leq O \left(n^{-4k} b^{1-k} \right). \tag{A17}$$

Using C_r -inequality and $\phi_n (Z_1, Z_2) \stackrel{d}{=} \phi_n (Z_2, Z_1)$, we also have

$$\begin{aligned} E \left\{ \Phi_n^{2k} (Z_1, Z_2) \right\} &\leq n^{-4k} E \left| \phi_n (Z_1, Z_2) + \phi_n (Z_2, Z_1) \right|^{2k} \\ &\leq n^{-4k} 2^{2k-1} \left\{ E \left| \phi_n (Z_1, Z_2) \right|^{2k} + E \left| \phi_n (Z_2, Z_1) \right|^{2k} \right\} \\ &= n^{-4k} 2^{2k} E \left| \phi_n (Z_1, Z_2) \right|^{2k}. \end{aligned}$$

Again, by C_r -inequality,

$$E \left| \phi_n (Z_1, Z_2) \right|^{2k} \leq 4 \cdot 2^{2(2k-1)} E \left\{ K_{X_2}^{2k} (X_1) \right\} \leq cb^{\frac{1-2k}{2}} \int_0^\infty x^{\frac{1-2k}{2}} f^2(x) dx,$$

where $x^{(1-2k)/2} f^2(x) \leq cx^{-(1-\varepsilon)}$ for some $\varepsilon \in (0, 1/2)$ as $x \rightarrow 0$ so that $\int_0^\infty x^{(1-2k)/2} f^2(x) dx < \infty$ is ensured. Therefore,

$$E \left\{ \Phi_n^{2k} (Z_1, Z_2) \right\} \leq O \left(n^{-4k} b^{\frac{1-2k}{2}} \right) = o(1) \tag{A18}$$

and thus $E \left\{ \Phi_n^{2k} (Z_1, Z_2) \right\} < \infty$ is demonstrated.

In the end, by (A16)–(A18),

$$\begin{aligned} \frac{E \left\{ Y_n^k (Z_1, Z_2) \right\}}{E^k \left\{ \Phi_n^2 (Z_1, Z_2) \right\}} &= O \left(b^{1-k/2} \right) \rightarrow 0, \text{ and} \\ \frac{n^{1-k} E \left\{ \Phi_n^{2k} (Z_1, Z_2) \right\}}{E^k \left\{ \Phi_n^2 (Z_1, Z_2) \right\}} &= O \left\{ \left(nb^{1/2} \right)^{1-k} \right\} \rightarrow 0, \end{aligned}$$

as long as $1 < k < 3/2$. This completes the proof. \square

Appendix A.2.4. Proof of Theorem 1

It follows from Lemma A4 that a martingale central limit theorem for a degenerate U -statistic (Theorem 4.7.3 of Koroljuk and Borovskich [41], to be precise) applies. Moreover, by Lemma A3, the asymptotic variance of the normal limit becomes

$$\begin{aligned} \sigma^2 &= \lim_{n \rightarrow \infty} n^2 b^{1/2} \text{Var} (I_n) \\ &= \lim_{n \rightarrow \infty} n^2 b^{1/2} \frac{n(n-1)}{2} E \left\{ \Phi_n^2 (Z_1, Z_2) \right\} \\ &= 2V_I(2) E \left[X^{-1/2} \{f(X) + g(X)\} + Y^{-1/2} \{f(Y) + g(Y)\} \right] \\ &= 8V_I(2) E \left\{ X^{-1/2} f(X) \right\} \text{ under } H_0. \end{aligned}$$

\square

Appendix A.3. Proof of Proposition 1

The proof closely follows the one for Theorem 2.2 of Fan and Ullah [19]. Under H_1 , $E(I_n) = \int_0^\infty \{f(u) - g(u)\}^2 du + O(b) = I + O(b)$. Moreover, $\text{Var}(I_n) = O(n^{-2}b^{-1/2})$ and $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$, regardless of whether H_0 or H_1 may be true. Therefore, $I_n = I + O(b) + O_p(n^{-1}b^{-1/4}) \xrightarrow{p} I > 0$, and thus $nb^{1/4}I_n/\hat{\sigma}$ is a divergent stochastic sequence with an expansion rate of $nb^{1/4}$. The result immediately follows. \square

Appendix A.4. Proof of Theorem 3

For brevity, we focus only on the case of equal sample sizes in two sub-samples. The proof largely follow the one for Proposition 5 of FMS. FMS consider the Taylor expansion

$$I_n(\hat{\xi}, \hat{\theta}) - I_n(\xi, \theta_0) = \Delta_1(\xi, \theta_0)(\hat{\xi} - \xi) + \Delta_2(\xi, \theta_0)(\hat{\theta} - \theta_0) + R_n,$$

where $\Delta_1(\xi, \theta_0)$ and $\Delta_2(\xi, \theta_0)$ are partial derivatives of I_n with respect to the first and second arguments evaluated at (ξ, θ_0) , respectively, and R_n is the remainder term of a smaller order. The only difference between their proof and ours is that we derive the range of (q, r) within which

$$nb^{1/4} \{ \Delta_1(\xi, \theta_0)(\hat{\xi} - \xi) + \Delta_2(\xi, \theta_0)(\hat{\theta} - \theta_0) \} = o_p(1)$$

is the case. Because each of Δ_1 and Δ_2 is $O(b) + O_p(n^{-1}b^{-3/4})$, the left-hand side is bounded by

$$O(n^{1-r}b^{5/4}) + O_p(n^{-r}b^{-1/2}) = O(n^{1-r-5q/4}) + O_p(n^{-r+q/2}).$$

This becomes $o_p(1)$ if (q, r) satisfy $r > -5q/4 + 1$, $r > q/2$ and $r \leq 1/2$. \square

Acknowledgments: We would like to thank the editor Kerry Patterson, four anonymous referees, Yohei Yamamoto, and the participants of seminars at Hitotsubashi University and the Development Bank of Japan for their constructive comments and suggestions. We are also grateful to Marcelo Fernandes, Eduardo Mendes and Olivier Scaillet for providing us with the computer codes used for Monte Carlo simulations in Fernandes, Mendes and Scaillet [29]. This research was supported, in part, by the grant from Japan Society of the Promotion of Science (grant number 15K03405). The views expressed herein and those of the authors do not necessarily reflect the views of the Development Bank of Japan.

Author Contributions: The authors contributed equally to the paper as a whole.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bacon, R.W. Rockets and feathers: The asymmetric speed of adjustment of UK retail gasoline prices to cost changes. *Energy Econ.* **1991**, *13*, 211–218.
2. Campbell, J.Y.; Hentschel, L. No news is good news: An asymmetric model of changing volatility in stock returns. *J. Financ. Econ.* **1992**, *31*, 281–318.
3. Clarida, R.; Gertler, M. How the Bundesbank conducts monetary policy. In *Reducing Inflation: Motivation and Strategy*; Romer, C.D., Romer, D.H., Eds.; University of Chicago Press: Chicago, IL, USA, 1997; pp. 363–412.
4. Chamberlain, G. A characterization of the distributions that imply mean-variance utility functions. *J. Econ. Theory* **1983**, *29*, 185–201.
5. Owen, J.; Rabinovitch, R. On the class of elliptical distributions and their applications to the theory of portfolio choice. *J. Finance* **1983**, *38*, 745–752.
6. Ingersoll, J.E., Jr. *Theory of Financial Decision Making*; Rowman & Littlefield: Savage, MD, USA, 1987.
7. Bickel, P.J. On adaptive estimation. *Ann. Stat.* **1982**, *10*, 647–671.
8. Newey, W.K. Adaptive estimation of regression models via moment restrictions. *J. Econom.* **1988**, *38*, 301–339.
9. Carroll, R.J.; Welsh, A.H. A note on asymmetry and robustness in linear regression. *Am. Stat.* **1988**, *42*, 285–287.
10. Lee, M.-J. Mode regression. *J. Econom.* **1989**, *42*, 337–349.
11. Lee, M.-J. Quadratic mode regression. *J. Econom.* **1993**, *57*, 1–19.
12. Zinde-Walsh, V. Asymptotic theory for some high breakdown point estimators. *Econom. Theory* **2002**, *18*, 1172–1196.
13. Bondell, H.D.; Stefanski, L.A. Efficient robust regression via two-stage generalized empirical likelihood. *J. Am. Stat. Assoc.* **2013**, *108*, 644–655.

14. Baldauf, M.; Santos Silva, J.M.C. On the use of robust regression in econometrics. *Econ. Lett.* **2012**, *114*, 124–127.
15. Fan, Y.; Gencay, R. A consistent nonparametric test of symmetry in linear regression models. *J. Am. Stat. Assoc.* **1995**, *90*, 551–557.
16. Ahmad, I.A.; Li, Q. Testing symmetry of unknown density functions by kernel method. *J. Nonparametr. Stat.* **1997**, *7*, 279–293.
17. Zheng, J.X. Consistent specification testing for conditional symmetry. *Econom. Theory* **1998**, *14*, 139–149.
18. Diks, C.; Tong, H. A test for symmetries of multivariate probability distributions. *Biometrika* **1999**, *86*, 605–614.
19. Fan, Y.; Ullah, A. On goodness-of-fit tests for weakly dependent processes using kernel method. *J. Nonparametr. Stat.* **1999**, *11*, 337–360.
20. Randles, R.H.; Fligner, M.A.; Policello, G.E., II; Wolfe, D.A. An asymptotically distribution-free test for symmetry versus asymmetry. *J. Am. Stat. Assoc.* **1980**, *75*, 168–172.
21. Godfrey, L.G.; Orme, C.D. Testing for skewness of regression disturbances. *Econ. Lett.* **1991**, *37*, 31–34.
22. Bai, J.; Ng, S. Tests for skewness, kurtosis, and normality for time series data. *J. Bus. Econ. Stat.* **2005**, *23*, 49–58.
23. Premaratne, G.; Bera, A. A test for symmetry with leptokurtic financial data. *J. Financ. Econom.* **2005**, *3*, 169–187.
24. Newey, W.K.; Powell, J.L. Asymmetric least squares estimation and testing. *Econometrica* **1987**, *55*, 819–847.
25. Bai, J.; Ng, S. A consistent test for conditional symmetry in time series models. *J. Econom.* **2001**, *103*, 225–258.
26. Delgado, M.A.; Escanciano, J.C. Nonparametric tests for conditional symmetry in dynamic models. *J. Econom.* **2007**, *141*, 652–682.
27. Chen, T.; Tripathi, G. Testing conditional symmetry without smoothing. *J. Nonparametr. Stat.* **2013**, *25*, 273–313.
28. Fang, Y.; Li, Q.; Wu, X.; Zhang, D. A data-driven test of symmetry. *J. Econom.* **2015**, *188*, 490–501.
29. Fernandes, M.; Mendes, E.F.; Scaillet, O. Testing for symmetry and conditional symmetry using asymmetric kernels. *Ann. Inst. Stat. Math.* **2015**, *67*, 649–671.
30. Chen, S.X. Probability density function estimation using gamma kernels. *Ann. Inst. Stat. Math.* **2000**, *52*, 471–480.
31. Gospodinov, N.; Hirukawa, M. Nonparametric estimation of scalar diffusion models of interest rates using asymmetric kernels. *J. Empir. Finance* **2012**, *19*, 595–609.
32. Hirukawa, M.; Sakudo, M. Family of the generalised gamma kernels: A generator of asymmetric kernels for nonnegative data. *J. Nonparametr. Stat.* **2015**, *27*, 41–63.
33. Fernandes, M.; Grammig, J. Nonparametric specification tests for conditional duration models. *J. Econom.* **2005**, *127*, 35–68.
34. Kulasekera, K.B.; Wang, J. Smoothing parameter selection for power optimality in testing of regression curves. *J. Am. Stat. Assoc.* **1997**, *92*, 500–511.
35. Kulasekera, K.B.; Wang, J. Bandwidth selection for power optimality in a test of equality of regression curves. *Stat. Probab. Lett.* **1998**, *37*, 287–293.
36. Stacy, E.W. A generalization of the gamma distribution. *Ann. Math. Stat.* **1962**, *33*, 1187–1192.
37. Abramson, I.S. On bandwidth variation in kernel estimates—A square root law. *Ann. Stat.* **1982**, *10*, 1217–1223.
38. Stone, C.J. Optimal rates of convergence for nonparametric estimators. *Ann. Stat.* **1980**, *8*, 1348–1360.
39. Anderson, N.H.; Hall, P.; Titterton, D.M. Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates. *J. Multivar. Anal.* **1994**, *50*, 41–54.
40. Hall, P. Central limit theorem for integrated square error of multivariate nonparametric density estimators. *J. Multivar. Anal.* **1984**, *14*, 1–16.
41. Koroljuk, V.S.; Borovskich, Y.V. *Theory of U-Statistics*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1994.
42. Hansen, B.E. Uniform convergence rates for kernel estimation with dependent data. *Econom. Theory* **2008**, *24*, 726–748.
43. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Chapman & Hall: London, UK, 1986.
44. Gao, J.; Gijbels, I. Bandwidth selection in nonparametric kernel testing. *J. Am. Stat. Assoc.* **2008**, *103*, 1584–1594.

45. Ramberg, J.S.; Schmeiser, B.W. An approximate method for generating asymmetric random variables. *Commun. ACM* **1974**, *17*, 78–82.
46. Fernandes, M.; Monteiro, P.K. Central limit theorem for asymmetric kernel functionals. *Ann. Inst. Stat. Math.* **2005**, *57*, 425–442.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).