Testing for symmetry and conditional symmetry using asymmetric kernels

Fernandes, M; Mendes, EF; Scaillet, O

For additional information about this publication click this link.
http://qmro.qmul.ac.uk/xmlui/handle/123456789/15803

Testing for symmetry and conditional symmetry using asymmetric kernels

Marcelo Fernandes

Sao Paulo School of Economics – FGV

E-mail: `marcelo.fernandes@fgv.br`

Queen Mary, University of London

E-mail: `m.fernandes@qmul.ac.uk`

Eduardo F. Mendes

School of Economics, University of New South Wales

E-mail: `eduardo.mendes@unsw.edu.au`

Olivier Scaillet

University of Geneva and Swiss Finance Institute

E-mail: `olivier.scaillet@unige.ch`

This version: April 9, 2014

Abstract: We derive nonparametric tests of symmetry using asymmetric kernels with either vanishing or fixed bandwidths. The idea is to split the sample around the symmetry point and then test whether the distributions to the right and to the left are the same. We show how to extend the approach to examine conditional symmetry by deriving conditions under which our tests are applicable to residuals from semiparametric models with a (sufficiently smooth) nonparametric link function. The latter setting is general enough to entertain as a particular case a unknown symmetry point, which we duely estimate by the sample median. The conditions we derive ensure that the resulting estimation error is asymptotically negligible. Simulations show that the asymptotic tests perform well even in very small samples, entailing better size and power properties than some of the existing symmetry tests.

1 Introduction

Symmetry and conditional symmetry play an important role in numerous situations. Conditional symmetry is part of the stochastic restrictions on unobservable errors used in semiparametric modeling. See, for instance, Powell (1994) for further discussion and references. It also implies constant conditional mean and median, and aids in the identification of models that can be symmetrized in their error terms. Further, adaptive estimation sometimes relies on the assumption of conditional symmetry. Bickel (1982) shows that, under conditional symmetry of the error term, it is possible to estimate adaptively the slope coefficients in linear regression models. Newey (1988) builds an adaptive estimator based on the generalized method of moments under the assumption of conditional symmetry. In macroeconomics, the symmetry of innovations also plays an important role (Campbell and Hentschel, 1992). Determining whether positive and negative shocks are equiprobable has crucial economic policy implications. In finance, knowing whether returns and risk factors exhibit symmetry may help in the choice of an adequate risk measure for the portfolio and risk management (Gouriéroux, Laurent and Scaillet, 2000). These few examples illustrate the relevance of contriving consistent tests of symmetry and conditional symmetry.

There are a number of nonparametric tests available in the literature. Fan and Gencay (1993) and Ahmad and Li (1997) propose a nonparametric test of symmetry based on Ahmad and van Belle (1974) affinity measure between two probability density functions. Fan and Gencay (1995) extend their result to deal with linear regression residuals, but fall into the same problem of requiring an arbitrary constant to avoid asymptotic degeneracy of the test statistic. Zheng (1998) derives tests of conditional symmetry by checking whether the conditional cumulative distribution function satisfies the restrictions imposed by symmetry. Bai and Ng (2001) show how to test whether residuals of nonlinear time-series models are symmetric, whereas Fan and Ullah (1999) propose a test of symmetry for weakly dependent data by gauging the closeness between $f(u)$ and $f(-u)$, where f is the stationary distribution of the process $\{X_t, \ t > 0\}$. More recently, Delgado and Escanciano (2007) propose a test for conditional symmetry based on empirical processes within a dynamic context.

Our testing strategy is somewhat different. Symmetry around zero implies that the shape of the density function to the right of the origin is a mirror image of the shape to the left of the origin.

Using the nonnegative data to estimate the density of the right part and the absolute value of the nonpositive data to estimate the density of the left part, we then check whether symmetry holds by looking at the closeness of these two probability density functions. To handle density functions whose supports are bounded from below, we rely on asymmetric kernels (Chen, 2000; Scaillet, 2004). These estimators are nonnegative, boundary-bias free, and achieve the optimal rate of convergence for the mean integrated error. Further, as opposed to the previous tests that use fixed kernels, our testing procedures do not require the continuity of the derivatives of the density function at the origin. Bouezmarni and Scaillet (2005) analyze the convergence of the asymmetric kernel estimator when the density is unbounded at the origin. Recent empirical applications of asymmetric kernel density estimators in economics and finance include Fernandes and Grammig (2005), Hagmann and Scaillet (2007), and Gustafsson, Hagmann, Nielsen and Scaillet (2009). These papers also report Monte Carlo results showing the nice finite-sample properties of asymmetric kernel estimation and testing.

We develop the asymptotic theory for the above tests considering both vanishing and fixed bandwidths. In the first case, we show that the integrated squared difference between the density estimates of the positive and the absolute value of the nonpositive data is a V-statistic. The asymptotic distribution is driven by a degenerate U-statistic that weakly converges to a Gaussian distribution with zero mean. The additional term that marks the difference between the V-statistic and the corresponding U-statistic gives way to the asymptotic bias in the limiting distribution of the V-statistic. We construct symmetry tests based on both V- and U-statistics within the context of vanishing bandwidths. In turn, our tests with fixed bandwidth rest on the V-statistic, weakly converging to a weighted sum of chi-squared random variables as in Anderson, Hall and Titterington (1994) and Fan (1998). Finally, we extend our approach to examine conditional symmetry by deriving conditions under which our tests are nuisance parameter free and hence applicable to residuals. This is important because we do not know in practice whether the symmetry point is indeed at the origin. The conditions are mild, allowing us to test the symmetry of residuals resulting even from semiparametric models that feature a (sufficiently smooth) nonparametric link function. This setup not only nests the case of a unknown symmetry point, but is also somewhat more general than the parametric settings that Zheng (1998), Bai and Ng (2001) and Lambert,

Laurent and Veredas (2012) consider for testing conditional symmetry.

We investigate through Monte Carlo simulations the performance of our conditional symmetry tests in small samples. In particular, we consider the gamma-kernel tests with vanishing bandwidths resulting from both V- and U-statistics. The latter has the advantage of avoiding the estimation of the asymptotic bias of the V-statistic. The results are encouraging, with the test based on the U-statistic improving considerably on the test rooted in the V-statistic. In stark contrast to many nonparametric tests, empirical size is very close to nominal size in most situations, even for sample sizes as small as 50 observations. This is particularly impressive in view that splitting the data into positive and negative values actually reduces the effective sample size by half. In addition, our tests also entail excellent power against a wide array of asymmetric distributions. Our methodology easily outperforms Zheng (1998) conditional symmetry test, whereas it competes well with the nonparametric tests put forth by Bai and Ng (2001) for sample sizes of at least 100 observations. This holds despite the fact that the latter tests display faster rates of convergence.

The outline for the remainder of the paper is as follows. Sections 2 and 3 derive the tests of symmetry and conditional symmetry, respectively. In particular, we show that our test statistic weakly converges to a standard normal distribution if the smoothing parameter shrinks to zero as the sample size grows. On the other hand, our test statistic weakly converges to an infinite sum of weighted chi-square variables for a fixed smoothing parameter. Section 4 reports some Monte carlo experiments addressing size and power in small samples. We focus on the nonparametric test of conditional symmetry based on a gamma kernel with a vanishing bandwidth. Section 5 offers some concluding remarks, whereas the Appendix collects the assumptions and technical proofs.

2   Testing symmetry

We check whether symmetry holds by looking at the closeness of the probability density distributions of the positive values and negative values. More formally, we derive a test statistic from the ℓ_2-distance with respect to the Lebesgue measure:

$$
\begin{aligned}
I &= \int_0^\infty [f(u) - g(u)]^2 \, \mathrm{d}u \\
&= \int_0^\infty f(u) \, \mathrm{d}F(u) + \int_0^\infty g(u) \, \mathrm{d}G(u) - 2 \int_0^\infty f(u)g(u) \, \mathrm{d}u,
\end{aligned}
\tag{1}
$$

where f and g are the density functions to the right and to the left of the origin, respectively. We impose without any loss of generality that the symmetry point is at the origin.[1] The integrated square distance that we adopt is convenient because it entails a consistent test, since it is always nonnegative, and takes value zero if and only if the null hypothesis holds, namely $\mathbb{H}_0: f(u) = g(u)$ almost everywhere. Bickel and Rosenblatt (1973), Aït-Sahalia, Bickel and Stoker (2001) and Aït-Sahalia, Fan and Peng (2009) rely on similar squared distance measures, though one could alternatively employ entropic pseudo-distance measures (Robinson, 1991; Hong and White, 2004).

From a random sample of N observations, it is straightforward to estimate the unknown density functions f and g using asymmetric kernel estimators. We first split the sample into a subsample with the nonnegative values $\{X_i; \ i = 1, \ldots, n_1\}$ and another subsample with the absolute values $\{Y_i; \ i = 1, \ldots, n_2 = N - n_1\}$ of the negative observations of the original sample. Next, we estimate the densities of X and Y using asymmetric kernels, so as to avoid the boundary bias that plagues fixed-kernel density estimation. For ease of exposition, we start with the simplifying assumption that $n = n_1 = n_2 = N/2$, so that

$$\hat{f}(u) = \frac{1}{n} \sum_{i=1}^{n} K_u(X_i, b) \qquad \text{and} \qquad \hat{g}(u) = \frac{1}{n} \sum_{i=1}^{n} K_u(Y_i, b),$$

where $K_u(\cdot, \cdot)$ is either the gamma kernel as in Chen (2000) or the (reciprocal) inverse Gaussian kernel as in Scaillet (2004), and b is a bandwidth tuning the amount of smoothing.

Let $\int$ denote the integral over the support $[0, \infty)$. A sample analog of (1) is

$$I_n = \int \hat{f}(u)\, \mathrm{d}F_n(u) + \int \hat{g}(u)\, \mathrm{d}G_n(u) - \int \hat{f}(u)\, \mathrm{d}G_n(u) - \int \hat{g}(u)\, \mathrm{d}F_n(u),$$

where $F_n(\cdot)$ and $G_n(\cdot)$ are the empirical distribution functions based on the sample data $\{X_i, \ i = 1, \ldots, n\}$ and $\{Y_i, \ i = 1, \ldots, n\}$, respectively. Using the fact that

$$\int M(u)\, \mathrm{d}F_n(u) = \frac{1}{n} \sum_{j=1}^{n} M(X_j) \quad \text{and} \quad \int M(u)\, \mathrm{d}G_n(u) = \frac{1}{n} \sum_{j=1}^{n} M(Y_j),$$

---

[1] It is without any loss of generality only because we show in Section 3 that estimating the symmetry point by means of the sample median entails no asymptotic impact in the limiting distribution of the test statistic.

and omitting the dependence of K on the bandwidth, it follows that

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

It turns out that I_{1n} is a bias term, contributing only to the mean of the asymptotic distribution of the test statistic (see Lemma A3 in the Appendix). This means that I_{2n} drives the limiting distribution of I_n and hence it requires the application of a suitable central limit theorem. We first show that I_{2n} is a degenerate U-statistics by observing that $I_{2n} \equiv \frac{1}{n^2} \sum_{i<j} H_n(Z_i, Z_j)$, where $H_n(Z_i, Z_j) \equiv h_n(Z_i, Z_j) + h_n(Z_j, Z_i)$ with $Z_i = (X_i, Y_i)$ and $h_n(Z_i, Z_j) \equiv K_{X_j}(X_i) + K_{Y_j}(Y_i) - K_{Y_j}(X_i) - K_{X_j}(Y_i)$. Note that $H_n(Z_i, Z_j)$ is symmetric in Z_i and Z_j, and that $\mathbb{E}\left[H_n(Z_i, Z_j) \,\big|\, Z_i \right] = 0$ under the null. Accordingly, we may apply Koroljuk and Borovskich (1994) central limit theorem to show that if, for some $k > 1$,

$$
\frac{\mathbb{E}\left[M_n^k(Z_1, Z_2) \right] + n^{1-k} \mathbb{E}\left[H_n^{2k}(Z_1, Z_2) \right]}{\left\{ \mathbb{E}\left[H_n^2(Z_1, Z_2) \right] \right\}^k} \to 0, \tag{2}
$$

where $M_n(Z_1, Z_2) = \mathbb{E}[H_n(Z_1, Z_3) H_n(Z_2, Z_3)]$, I_{2n} is asymptotically normal with mean zero and variance $\frac{1}{2n^2} \mathbb{E}\left[H_n^2(Z_1, Z_2) \right]$. The latter reduces under the null to $2\,\mathbb{E}\left[X^{-\tau/2} f(X) \right]$, with $\tau = 1$ for the gamma and reciprocal inverse Gaussian kernels and $\tau = 3$ for the inverse Gaussian kernel (see Lemma A4 in the Appendix). We must choose the rate at which the bandwidth converges to zero so as to ensure that condition (2) holds. In particular, Fernandes and Monteiro (2005) show that it suffices to consider a bandwidth b of order $o\left(n^{-4/9}\right)$ if one restricts attention to the gamma kernel (see Lemma A5 in the Appendix for more details). The next result documents the asymptotic theory for the $n = n_1 = n_2$ case. As noted in the Appendix, it follows trivially from Lemmata A3 to A5 and the Slutsky theorem under some mild assumptions.

Proposition 1: *Assume that the bandwidth b is of order $o\left(n^{-4/9}\right)$ and that f and g are twice continuously differentiable, such that $\int \left[x^3 f''(x) \right]^2 \mathrm{d}x$ and $\int \left[x^3 g''(x) \right]^2 \mathrm{d}x$ are finite. Under the*

null $\mathbb{H}_0$, *it follows that*

(i) $(n\,b^{1/4}\,I_n - b^{-1/4}\,\hat{\mu}_1)/\hat{\sigma} \xrightarrow{d} N(0,1)$

(ii) $n\,b^{1/4}\,I_{2n}/\hat{\sigma} \xrightarrow{d} N(0,1)$,

where $\hat{\mu}_1$ is any consistent estimator of $\mu_1 \equiv \mathbb{E}(I_{1n})$ satisfying $|\hat{\mu}_1 - \mu_1| = o_p(b^{1/4})$ and $\hat{\sigma}^2$ is any consistent estimator of the asymptotic variance $\sigma_0^2 \equiv 2\pi^{-1/2}\,\mathbb{E}\left\{ X^{-\tau/2}\left[f(X) + g(X) \right] \right\}$.

Imposing the null hypothesis yields $\sigma_0^2 \equiv 4\pi^{-1/2}\,\mathbb{E}\left[X^{-\tau/2}\,f(X) \right]$. We now turn our attention to the general case in which the number of observations differs in the two subsamples (i.e., $n_1 \neq n_2$). A sample analog of (1) then is

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

with I_{1n_1,n_2} and I_{2n_1,n_2} respectively analogous to I_{1n} and I_{2n}. The differences in the summation indexes does not affect the limiting distribution since the underlying U-statistic remains degenerate. Further, to establish asymptotic results for $n_1 \to \infty$, it suffices to assume that $\lambda_n \equiv n_1/n_2 \to \lambda$ in the limit, where $0 < \lambda < \infty$ is a constant. In addition, the null of symmetry actually implies that $\lambda = 1$ and so the asymptotic distribution does not change under the null.

Proposition 2: *Assume the conditions in Proposition 1 hold. Under the null hypothesis, both $(n_1\,b^{1/4}\,I_{n_1,n_2} - b^{-1/4}\,\hat{\mu}_1)/\hat{\sigma}$ and $n_1\,b^{1/4}\,I_{2n_1,n_2}/\hat{\sigma}$ weakly converge to a standard Gaussian distribution.*

We derive Propositions 1 and 2 under the assumption that the smoothing parameter vanishes asymptotically. It is well known that asymptotic kernel-based tests are quite sensitive to the choice of the smoothing parameter; see e.g. Scaillet (2007) and the references therein. Thus, we state the corresponding asymptotic results for the fixed smoothing parameter case in the next two propositions. The limiting distribution follows readily from the results by Anderson et al. (1994) and Fan (1998). Since the asymmetric kernels are bounded, absolutely integrable and admit nonsingular Fourier transforms, the regularity conditions of Anderson et al. (1994) are automatically met.

Proposition 3: *Under the null $\mathbb{H}_0$, $n\, I_n \xrightarrow{d} \sum_{k=1}^{\infty} w_k Z_k^2$ and $n_1\, I_{n_1,n_2} \xrightarrow{d} \sum_{k=1}^{\infty} w_k \left(Z_{1k} - \lambda^{1/2} Z_{2k}\right)^2$ as the sample sizes grow, where $Z_1, Z_2, \ldots, Z_{11}, Z_{12}, \ldots, Z_{21}, Z_{22}, \ldots$ are independent standard normal random variables, and the weights w_k are the eigenvalues corresponding to an orthogonal expansion in the eigenfunctions of the kernel H_n with respect to f.*

Finding asymptotic-valid critical values requires the derivation of an infinite number of eigenvalues as the solution of integral equations. Accordingly, there are very few instances in which Proposition 3 yields feasible asymptotic-valid critical values. This means that, in general, we must rely on resampling methods as in Fan (1998) and Scaillet (2007).

We next study the asymptotic power of the above tests against a sequence of local alternatives $\mathbb{H}_1^{(n)}$. For simplicity, we entertain only the case in which $n = n_1 = n_2$. In particular, we assume that $\mathbb{H}_1^{(n)}$: $g(y) = f(y) + \epsilon_n\, h(y)$, with h satisfying $\int h(y) dy = 0$ and $0 < \int h^2(y) dy < \infty$. The next result considers how our tests behave according to the rate at which the perturbation ϵ_n converges to zero.

Proposition 4: *Let $\hat{Z}_n$ denote either $(n\, b^{1/4}\, I_n - b^{-1/4}\, \hat{\mu}_1)/\hat{\sigma}$ or $n\, b^{1/4}\, I_{2n}/\hat{\sigma}$. Under the conditions in Proposition 1, it follows under the local alternative $\mathbb{H}_1^{(n)}$ that, for any two-sided standard normal quantile z_α,*

$$Pr\left(|\hat{Z}_n| > z_\alpha\right) = \begin{cases} 1, & \text{if } \epsilon_n/b^2 \to \infty \\ 1 - \beta_\alpha, & \text{if } \epsilon_n/b^2 \to c \\ 0, & \text{if } \epsilon_n/b^2 \to 0 \end{cases}$$

for some $0 < \beta_\alpha = \beta_\alpha(c,h) < 1$, with c constant.

This means that our tests have asymptotic power against local alternatives of the form $\mathbb{H}_1^{(n)}$: $g(y) = f(y) + \epsilon_n\, h(y)$ as long as ϵ_n converges to zero at the same rate or faster than b^2. In particular, this implies consistency against fixed alternatives (i.e., $\epsilon_n = 1$ regardless of the sample size). Indeed, under $\mathbb{H}_1$: $g(y) = f(y) + h(y)$, our tests reject the null hypothesis with probability approaching one as the sample size grows without bound.

3 Conditional symmetry tests

In this section, we show how to extend our results to deal with tests for conditional symmetry of $V_1 \in \mathbb{R}$ given $V_2 \in \mathbb{R}^q$ within a semiparametric context. Given a parameter space Θ_1 and a function

$\xi_1 : \mathbb{R}^q \times \Theta_1 \to \mathbb{R}$, it suffices to check whether the conditional density function of V_1 given V_2 is symmetric around $\xi_1(V_2; \theta_1^0)$ for some $\theta_1^0 \in \Theta_1$. This is equivalent to test whether there exists $\theta_1^0 \in \Theta_1$ such that the conditional density of $V = V_1 - \xi_1\left(V_2; \theta_1^0\right)$ given V_2 is symmetric around zero.

To avoid the estimation of conditional densities (see the excellent discussion in Delgado and Escanciano, 2007), we assume that there exists a parameter space Θ_2 and a function $\xi_2 : \mathbb{R}^{q+1} \times \Theta_2 \to \mathbb{R}$ such that the marginal distribution of $U = \xi_2\left(V, V_2; \theta_2^0\right)$ is symmetric around zero for some $\theta_2^0 \in \Theta_2$. Given the dependence of V (and hence of U) on the parameter vector θ_1, it is convenient to summarize our semiparametric setup by considering a suitable parameter space Θ and function $\xi : \mathbb{R}^{q+1} \times \Theta \to \mathbb{R}$ such that the marginal density of $U = \xi(V_1, V_2; \theta_0)$ is symmetric around zero for some $\theta_0 \in \Theta$. This setup is well in line with Zheng (1998) and Bai and Ng (2001), for instance.

We next propose a two-step procedure in which we estimate $\xi(\cdot, \cdot; \theta)$ in the first step and then check whether the marginal density of $\widehat{U} = \widehat{\xi}(V_1, V_2; \widehat{\theta})$ is symmetric around zero in the second step. We believe that this approach is general enough in that we do not assume a particular functional form for ξ and hence it dwells in a semiparametric setting. It thus remains to derive the conditions under which the first-step estimation of (ξ, θ) does not affect the limiting distributions in Propositions 1 to 3, that is to say, the conditions under which the test is nuisance parameter free. The next result shows that this amounts to establishing the rate at which $(\widehat{\xi}, \widehat{\theta})$ converges to (ξ, θ_0) in the first step. As a by-product, we prove the consistency of the asymmetric kernel estimator of the derivative of the density function in Lemma A6 used for the proof of Proposition 5 in the Appendix.

Proposition 5: *The symmetry tests in Propositions 1 to 3 are nuisance parameter free as long as $(\widehat{\xi}, \widehat{\theta})$ converges to (ξ, θ_0) at a rate no slower than $n^{4/9}$.*

The choice of the semiparametric estimation method obviously depends on the problem under consideration, though the unknown function ξ should have enough derivatives if one hopes to find a nonparametric estimator $\widehat{\theta}$ that meets the above convergence rate. For instance, if $\widehat{\theta}$ converges at Stone (1982) optimal global rate, then it would suffice to require ξ to have at least 4 derivatives. See also Shen and Wong (1994) and Shen (1997) for a more general discussion based on sieve

estimation.

4   Monte Carlo study

Propositions 1 to 5 establish results of asymptotic nature. Typically there exists a large discrepancy between the empirical and nominal sizes of kernel-based tests. This reveals that asymptotic normality does not always provide much information about the finite-sample distribution of nonparametric test statistics. Also, it is well known that the performance of nonparametric tests in finite samples is sometimes very sensitive to the bandwidth choice (Fan, 1995).

We thus investigate in what follows how our nonparametric test of conditional symmetry fairs in small samples. We carry out Monte Carlo simulations to assess the size and power features for sample sizes of 50, 100, 200, and 400 observations. We draw 1,000 replications from both symmetric and asymmetric distributions and then evaluate our testing procedure using the asymptotic 5% and 10% critical values. For simplicity, we consider only tests using a gamma kernel with a vanishing bandwidth.

We generate the data as follows. Let $y_t = 1 + x_t + e_t$, where e_t and x_t denote two independent random variables. The regressor x_t is standard normal, whereas the error term e_t may come from different symmetric and asymmetric distributions according to the particular specification. The null hypothesis of interest corresponds to the conditional symmetry of y_t given x_t, which essentially boils down to testing the least-squares residuals $\widehat{e}_t$ for symmetry once we impose the linear regression structure. To ensure fair comparison between the different distributions, we standardize the least-squares residuals before running the nonparametric test in Proposition 2. The asymptotic critical values come from a standard normal distribution given that Proposition 5 ensures that the test is nuisance parameter free.

For comparison purposes, we employ exactly the same distributions as in Zheng (1998) to study the size and power of our testing procedures. Among the symmetric distributions, we sample e_t from the standard normal, t-student with 10 degrees of freedom, and two symmetric lambda distributions, whereas we consider four asymmetric lambda distributions to assess power. The lambda distribution family nests a wide array of symmetric and asymmetric distributions by defining the inverse of the cumulative distribution as $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$. The interesting feature of the lambda

family of distributions is that it covers a broad spectrum of skewness and kurtosis values. This is particularly important for applications in economics and finance. Table 1 lists the parameter vectors of all distributions and their corresponding skewness and kurtosis. We take the parameter values for all distributions from Zheng (1998). The parameter values are somewhat different in Bai and Ng (2001) essentially because they normalize λ_1 to zero. However, the resulting lambda distributions have similar, if not the same, skewness and kurtosis. The second symmetric lambda distribution in Table 1 is equivalent to their second symmetric lambda distribution, whereas A1, A2, and A4 correspond to their last three asymmetric lambda distributions. This allows for direct comparison with the Monte carlo results by Zheng (1998) and Bai and Ng (2001).

In the context of nonparametric testing, the choice of the kernel bandwidth normally plays a major role in the finite-sample properties of the test. There is an extensive literature on the optimal selection of the smoothing parameter (Silverman, 1986). To maximize power, we employ two different bandwidths, one for the estimation of the distribution of the nonnegative data and one for the estimation of the distribution of the absolute value of the nonpositive data. In both instances, we select the gamma-kernel bandwidths by means of a generalized cross-validation criterion. In view that the latter should yield a bandwidth of order $O(n_j^{-2/5})$ for asymmetric kernels (Bouezmarni and Rombouts, 2010), we multiply the resulting bandwidth by $n_j^{-2/45}/\ln n_j$ so as to ensure that $b_j = o(n_j^{-4/9})$ for $j = 1, 2$. Further analysis shows however that test results are not very sensitive to variations in the bandwidth (namely, either dividing or multiplying the resulting bandwidth by two). Monte Carlo results reported in other studies related to asymmetric kernels also show that weak sensitivity.

Table 2 reports the finite-sample performance of the gamma-kernel test based on I_n. It exhibits very little size distortions at the 5% and 10% levels for the standard normal and t-distributions. In contrast, there is significant size distortions for the symmetric lambda distributions, especially for the specification with negative excess kurtosis. This is consistent with most nonparametric symmetry tests in the literature, whose size depends heavily on kurtosis (Randles, Fligner, Policello and Wolf, 1980). As for statistical power, the results are also promising. In line with the asymptotic theory, power increases with the sample size. It is also very stable across the different asymmetric lambda distributions we consider. Moreover, our test seems not only to entail more power than

Zheng (1998), but also to stack up well with the nonparametric tests by Fan and Gencay (1995) and Bai and Ng (2001) for sample sizes with at least 200 observations. This is very encouraging not only because the latter tests have faster rates of convergence, but also because splitting the data into negative and positive values actually reduces the effective sample size by half.

Table 3 documents the size and power properties of a variant of the test that hinges on degenerate U-statistic I_{2n}. The idea is to avoid estimating the asymptotic bias of I_n due to I_{1n}, which is likely to be the source of the size distortions for the lambda distributions S3 and S4. The results indeed show a huge improvement. Size ameliorates for every distribution, with distortions shrinking substantially for the leptokurtic distribution S4. In particular, the difference between nominal and empirical size remains significant only for the platykurtic lambda distribution S3. This is in contrast to the extant nonparametric tests, whose size deteriorates with excess kurtosis as well. This is very reassuring given that it is uncommon, if not rare, to encounter data in economics and finance with negative excess kurtosis. Power also improves a lot especially for the smaller samples, with improvements of about 50% for $n = 50$ and $n = 100$. As a result, our nonparametric test starts comparing well with Bai and Ng (2001) test of conditional symmetry even for the samples of 100 observations. Moreover, adjusting power for size should also work in our favor given that our test displays less size distortion in the presence of high kurtosis.

5   Conclusion

This paper develops tests of symmetry using asymmetric kernels. The idea is to take benefit from the shape of the density function to the right of the symmetry point being a mirror image of the shape to the left. We thus gauge the closeness between density functions of the positive data and of the absolute value of the nonpositive data. We employ asymmetric kernels not only because they are nonnegative and free of boundary bias, but also because they usually entail tests with better finite-sample properties than fixed kernels (Fernandes and Grammig, 2005). One interesting feature of our tests is that they do not require the continuity of the derivatives of the density function at the symmetry point. Further, we also derive the conditions under which our tests are nuisance parameter free. It turns out that the conditions are mild in that we may apply our symmetry test to residuals from semiparametric models with a (sufficiently smooth) nonparametric link function.

This is important because it allows us to test for conditional symmetry in a setting somewhat more general than in Fan and Gençay (1995), Zheng (1998), and Bai and Ng (2001).

We spell out the asymptotic theory considering both fixed and vanishing bandwidths. In the first case, we document that the corresponding V-statistic weakly converges to a weighted sum of chi-squared random variables as in Anderson et al. (1994). As for the latter, we show that the test statistic we propose is a V-statistic, whose asymptotic normality is driven by a degenerate U-statistic with zero mean. Monte Carlo simulations reveal that symmetry tests based on the U-statistic outperform the tests based on the V-statistic most likely because they avoid the estimation of the asymptotic bias term. Their performance is also very promising relative to the extant tests in the literature. We find little size distortions for symmetric distributions with nonnegative excess kurtosis even for sample sizes as small as 50 observations. In addition, our test entails excellent power against asymmetric lambda distributions. The power is actually very robust in that it does not vary much with the parameter values we consider. All in all, employing asymmetric kernels pays off dearly as our tests entail clearly more power than Zheng (1998). Our testing procedures are also competitive relative to Bai and Ng (2001) nonparametric tests, even though the latter tests display faster rates of convergence. This is particularly impressive in view that splitting the data into positive and negative values actually reduces our effective sample size by half.

Appendix

We start by stating the assumptions we require on $(Z_1, \ldots, Z_n)$, f and g; and then collect some intermediate results in a sequence of technical lemmata. Lemmata A1 and A2 group the main identities we use throughout the Appendix, whereas Lemmata A3 to A5 establish some intermediate results we employ in the proof of Proposition 1. Finally, Lemma A6 gives the consistency of the asymmetric kernel estimator of the derivative of the density function we require in the proof of Proposition 5.

Assumptions

(IID) *The random variables $Z_1, \ldots, Z_n$ are independent and identically distributed.*

(BD) *Densities f and g are (a.e.) bounded with bounded derivatives.*

(NM) *Random variables X and Y satisfy $\mathbb{E}[X^{-\tau/2}]$ and $\mathbb{E}[Y^{-\tau/2}]$ are bounded, where $\tau \geq 1$.*

Assumptions (IID) and (BD) are standard in the literature and require no further explanation. Assumption (NM) requires existence of negative moments. Khuri and Casella (2002) discuss conditions under which the latter holds for $\tau = 2$. A simple modification of their result leads to the following necessary and sufficient condition:

$$\lim_{c \to 0^+} \int_a^b x^{-\tau/2} f(x) \, \mathrm{d}x = 0, \qquad \text{with } c > b > a > 0.$$

A simpler, but stronger, condition is to require that there exist $a > 0$ and $\alpha > 0$ such that for any $0 < x < a$, $f(x) < \alpha \, x^{\tau/2}$. It then follows directly from (BD) that $\mathbb{E}[f(X)^2]$ is bounded and from both (BD) and (NM) that $\mathbb{E}[f(X)X^{-\tau/2}]$ is also bounded. The next result requires these negative moment conditions.

Lemma A1: *Let X and Y denote independent random variables with bounded, differentiable density functions f and g, respectively. Assume $g(x) = f(x) + \epsilon_n h(x)$ for some real-valued function h satisfying $\int h(x) \, \mathrm{d}x = 0$ and $0 < \int h(x)^2 \, \mathrm{d}x < \infty$. Let (X_1, X_2) and (Y_1, Y_2) denote independent copies of X and Y, respectively. It then follows that:*

(a) $\mathbb{E}\left[K_v(X)\right] = f(v) + O(b)$ and $\mathbb{E}\left[K_v(Y)\right] = g(v) + O(b)$, for any fixed $v > 0$.

(b) $\mathbb{E}\left[K_Y(X)\right] = \mathbb{E}\left[g(X)\right] + O(b)$, $\mathbb{E}\left[K_X(Y)\right] = \mathbb{E}\left[g(X)\right] + O(b)$, $\mathbb{E}\left[K_{X_1}(X_2)\right] = \mathbb{E}\left[f(X)\right] + O(b)$,

and $\mathbb{E}\left[K_{Y_1}(Y_2)\right] = \mathbb{E}\left[g(X)\right] + O(\epsilon_n + b)$.

(c) $\mathbb{E}\left[K_Y^2(X)\right] = (4\pi b)^{-1/2}\mathbb{E}\left[X^{-\tau/2}g(X)\right]$, $\mathbb{E}\left[K_X^2(Y)\right] = (4\pi b)^{-1/2}\mathbb{E}\left[X^{-\tau/2}g(X)\right]$, $\mathbb{E}\left[K_{X_1}^2(X_2)\right] =$

$(4\pi b)^{-1/2}\mathbb{E}\left[X^{-\tau/2}f(X)\right]$, and $\mathbb{E}\left[K_{Y_1}^2(Y_2)\right] = (4\pi b)^{-1/2}\mathbb{E}\left[X^{-\tau/2}g(X)\right] + O(\epsilon_n b^{-1/2})$.

(d) $\mathbb{E}\left[K_{X_1}(X_2)K_{X_1}(Y)\right] = \mathbb{E}\left[f(X)g(X)\right] + O(b) = \mathbb{E}\left[f^2(X)\right] + O(\epsilon_n + b)$ and $\mathbb{E}\left[K_{Y_1}(Y_2)K_{Y_1}(X)\right] =$

$\mathbb{E}\left[g^2(X)\right] + O(b)$.

(e) $\mathbb{E}\left[K_{X_1}(X_2)K_Y(X_2)\right] = \frac{1}{2}\mathbb{E}\left[g^2(X)\right] + \{1 + o(1)\}$ and $\mathbb{E}\left[K_{Y_1}(Y_2)K_X(Y_2)\right] = \frac{1}{2}\mathbb{E}\left[f(X)g(X)\right]\{1 +$

$o(1)\} = \frac{1}{2}\mathbb{E}\left[f^2(X)\right]\{1 + o(1)\} + O(\epsilon_n)$.

(f) $\mathbb{E}\left[K_Y(X_1)K_{X_1}(X_2)\right] = \mathbb{E}\left[f(X)g(X)\right] + O(b) = \mathbb{E}\left[K_{X_1}(Y)K_{X_2}(X_1)\right]$ and $\mathbb{E}\left[K_X(Y_1)K_{Y_1}(Y_2)\right] =$

$\mathbb{E}\left[g(X)^2\right] + O(b) = \mathbb{E}\left[K_{Y_1}(X)K_{Y_2}(Y_1)\right]$.

Proof: We derive the results only for the gamma kernel, though it is straightforward to consider the inverse Gaussian and reciprocal inverse Gaussian kernels as well. Note that, in most results, we employ the following identities: $\mathbb{E}\left[f(Y)\right] = \mathbb{E}\left[g(X)\right] = \mathbb{E}\left[f(X)\right] + \epsilon_n \mathbb{E}\left[h(X)\right]$ and $\mathbb{E}\left[g(Y)\right] = \mathbb{E}\left[g(X)\right] + \epsilon_n \mathbb{E}\left[h(Y)\right]$.

(a) Denote ζ a gamma random variable with parameters $v/b + 1$ and b (and so with mean $v + b$). The result follows by noting that $\mathbb{E}\left[K_v(X)\right] = \mathbb{E}\left[f(\zeta)\right] = f(v) + [f'(v) + O(1)]\mathbb{E}\left[\zeta - v\right] = f(v) + O(b)$ provided that $f'(x) = \partial f(x)/\partial x$ is bounded.

(b) Given that $\mathbb{E}\left[K_X(Y)\right] = \mathbb{E}_Y\left\{\mathbb{E}\left[K_u(X)|Y = u\right]\right\}$, applying the previous identities yield the result.

(c) It follows from Chen (2000) and Scaillet (2004) that

$$\mathbb{E}\left[K_Y(X)\right]^2 = \int \int K_y^2(x)\, f(x)\, \mathrm{d}x\, \mathrm{d}G(y)$$

$$= \frac{1}{2\sqrt{\pi b}} \int x_2^{-\tau/2}\left[f(x_2) + O(b)\right]\mathrm{d}G(y)$$

$$= \frac{1}{2\sqrt{\pi b}}\, \mathbb{E}\left[Y^{-\tau/2}f(Y)\right],$$

where $\tau = 1$ for the gamma and reciprocal inverse Gaussian kernels and $\tau = 3$ for the inverse Gaussian kernel.

15

(d) As X_1 and X_2 are independent copies of X,

$$\mathbb{E}\left[K_{X_1}(X_2)K_{X_1}(Y)\right] = \mathbb{E}\left[\mathbb{E}\left[K_u(X_2)|X_1=u\right]\mathbb{E}\left[K_u(Y)|X_1=u\right]\right]$$

$$= \mathbb{E}\left[\left(f(X)+O(b)\right)\left(g(X)+O(b)\right)\right]$$

$$= \mathbb{E}\left[g(X)f(X)\right] + O(b)$$

$$= \mathbb{E}\left[f(X)^2\right] + \epsilon_n\mathbb{E}\left[f(X)h(X)\right] + O(b).$$

The corresponding result for Y follows similarly.

(e) For the gamma kernel, it suffices to apply Fernandes and Monteiro (2005) Lemma A.4 to see that

$$\mathbb{E}\left[K_{X_2}(X_1)\,K_Y(X_1)\right] = \int\int\int K_{x_2}(x_1)\,K_{y_2}(x_1)\,\mathrm{d}F(x_1)\,\mathrm{d}F(x_2)\,\mathrm{d}G(y)$$

$$= \int\int\int B_b(x_2,y)\,K_{(x_2+y)/b+1,b/2}(x_1)\,\mathrm{d}F(x_1)\,\mathrm{d}F(x_2)\,\mathrm{d}G(y)$$

$$= \int\int B_b(x_2,y)\left[\int f(x_1)\,\mathrm{d}K_{(x_2+y_2)/b+1,b/2}(x_1)\right]\mathrm{d}F(x_2)\,\mathrm{d}G(y)$$

$$= \int\int B_b(x_2,y)\left[f\left(\frac{x_2+y}{2}\right)+O(b)\right]\mathrm{d}F(x_2)\,\mathrm{d}G(y)$$

$$= \tfrac{1}{2}\,\mathbb{E}\left[f(Y)^2\right]\{1+o(1)\}$$

$$= \tfrac{1}{2}\left\{\mathbb{E}\left[f(X)^2\right]+\epsilon_n\mathbb{E}\left[f(X)h(X)\right]\right\}\{1+o(1)\},$$

where

$$B_b(x_2,y_2) = \frac{\Gamma[(x_2+y_2)/b+1]}{\Gamma(x_2/b+1)\,\Gamma(y_2/b+1)}\,\frac{b^{-1}}{2^{(x_2+y_2)/b+1}}.$$

The same holds true for the two other asymmetric kernels as well as for the second part relating to Y.

(f) The first identity results from

$$\mathbb{E}\left[K_Y(X_1)K_{X_1}(X_2)\right] = \int\int\left[\int f(x_2)K_{x_2}(x_1)\,\mathrm{d}x_2\right]K_y(x_1)f(x_1)\,\mathrm{d}x_1 g(y)dy$$

$$= \int\int\left[f(x_1)+O(b)\right]K_y(x_1)f(x_1)\,\mathrm{d}x_1 g(y)dy$$

$$= \mathbb{E}\left[f^2(Y)\right]+O(b)$$

$$= \mathbb{E}\left[f(X)g(X)\right]+O(b).$$

The remaining identities follow along the same line as above, completing the proof. ∎

Lemma A2: *Let X and Y have bounded density functions f and g on $[0, \infty)$, respectively. It then follows that* $\Lambda_b = \mathbb{E}\left[K_X(Y) K_Y(X)\right] = \frac{1}{2\sqrt{\pi\, b}} \mathbb{E}_X\left[X^{-\tau/2} g(X)\right].$

Proof: In the following, we derive the result only for the gamma kernel, though it is straightforward to consider the inverse Gaussian and reciprocal inverse Gaussian kernels. The Stirling approximation states that

$$\Gamma(x+1) = \sqrt{2\pi}\, x^{x+\frac{1}{2}}\, \exp(-x),$$

which implies that

$$\Lambda_b = \frac{1}{2\,b\,\pi} \int_0^\infty \int_0^\infty (x/y)^{\frac{y-x}{b}}\, \frac{f(x)\, g(y)}{\sqrt{x\,y}}\, \mathrm{d}x\, \mathrm{d}y.$$

We then split the integral over y into four pieces, giving way to

$$\Lambda_b = \frac{1}{2\,b\,\pi} \int_0^\infty \left[\int_0^{(1-t)x} + \int_{(1-t)x}^x + \int_x^{(1+t)x} + \int_{(1+t)x}^\infty\right] (x/y)^{\frac{y-x}{b}}\, \frac{f(x)\, g(y)}{\sqrt{x\,y}}\, \mathrm{d}y\, \mathrm{d}x$$
$$= \Lambda_b^{(1)} + \Lambda_b^{(2)} + \Lambda_b^{(3)} + \Lambda_b^{(4)},$$

with $0 < t < 1$. We next compute coincident upper and inferior bounds for each one of the above nonnegative terms, so as to obtain the result. These bounds depend on the fact that

$$\exp\left[-\frac{(y-x)^2}{x\,b}\right] \le (x/y)^{\frac{y-x}{b}} \le \exp\left[-\frac{(y-x)^2}{x\,b} + \frac{(y-x)^3}{2\,x^2\,b}\right], \qquad \text{if } y > x,$$

whereas

$$\exp\left[-\frac{(y-x)^2}{x\,b} + \frac{(y-x)^3}{2\,y^2\,b}\right] \le (x/y)^{\frac{y-x}{b}} \le \exp\left[-\frac{(y-x)^2}{x\,b}\right], \qquad \text{if } y < x.$$

It then follows that

$$\Lambda_b^{(1)} = \frac{1}{2\,b\,\pi} \int_0^\infty \int_0^{(1-t)x} (x/y)^{\frac{y-x}{b}}\, \frac{f(x)\, g(y)}{\sqrt{x\,y}}\, \mathrm{d}y\, \mathrm{d}x$$
$$\le \frac{1}{2\,b\,\pi} \int_0^\infty \int_0^{(1-t)x} \exp\left[-\frac{(y-x)^2}{x\,b}\right] \frac{f(x)\, g(y)}{\sqrt{x\,y}}\, \mathrm{d}y\, \mathrm{d}x,$$

whence we substitute for $z = \frac{y-x}{\sqrt{x\,b}}$ yielding

$$\Lambda_b^{(1)} \le \frac{1}{2\,\pi\,\sqrt{b}} \int_0^\infty \int_{-\sqrt{x/b}}^{-t\sqrt{x/b}} e^{-z^2}\, \frac{f(x)\, g(x + z\sqrt{x\,b})}{\sqrt{x + z\sqrt{x\,b}}}\, \mathrm{d}z\, \mathrm{d}x = 0,$$

as $b \to 0$, by Lebesgue dominated convergence theorem. This ultimately means that $\Lambda_b^{(1)} = 0$ given that it is nonnegative. We now turn our attention to the second term for which we derive the

17

following upper limit:

$$
\Lambda_b^{(2)} = \frac{1}{2\,b\,\pi} \int_0^\infty \int_{(1-t)x}^x (x/y)^{\frac{y-x}{b}} \frac{f(x)\,g(y)}{\sqrt{x\,y}}\,\mathrm{d}y\,\mathrm{d}x
$$

$$
\leq \frac{1}{2\,b\,\pi} \int_0^\infty \int_{(1-t)x}^x \exp\left[-\frac{(y-x)^2}{x\,b}\right] \frac{f(x)\,g(y)}{\sqrt{x\,y}}\,\mathrm{d}y\,\mathrm{d}x
$$

$$
= \frac{1}{2\,\pi\,\sqrt{b}} \int_0^\infty \int_{-t\sqrt{x/b}}^0 e^{-z^2} \frac{f(x)\,g(x+z\,\sqrt{x\,b})}{\sqrt{x+z\,\sqrt{x\,b}}}\,\mathrm{d}z\,\mathrm{d}x
$$

$$
= \frac{1}{4\sqrt{\pi\,b}}\,\mathbb{E}\left[X^{-1/2}\,g(X)\right],
$$

where the last equality holds by Lebesgue dominated convergence theorem (as $b \to 0$) and by the

fact that $\int_0^\infty e^{-z^2}\,\mathrm{d}z = \sqrt{\pi}/2$. As for the inferior bound, observe that

$$
\Lambda_b^{(2)} \geq \frac{1}{2\,b\,\pi} \int_0^\infty \int_{(1-t)x}^x \exp\left\{-\frac{(y-x)^2}{x\,b}\left[1-\frac{(y-x)x}{2y^2}\right]\right\} \frac{f(x)\,g(y)}{\sqrt{x\,y}}\,\mathrm{d}y\,\mathrm{d}x
$$

$$
\geq \frac{1}{2\,b\,\pi} \int_0^\infty \int_{(1-t)x}^x \exp\left[-\frac{(y-x)^2}{x\,b}(1+\xi)\right] \frac{f(x)\,g(y)}{\sqrt{x\,y}}\,\mathrm{d}y\,\mathrm{d}x,
$$

where $\xi = (1-t)^{-2}\,t/2$. Substituting for $z = (y-x)\sqrt{\frac{1+\xi}{x\,b}}$ then results in

$$
\Lambda_b^{(2)} \geq \frac{1}{2\,\pi\,\sqrt{b}} \int_0^\infty \int_{-t\sqrt{\frac{1+\xi}{x\,b}}}^0 e^{-z^2} \frac{f(x)\,g(x+z\,\sqrt{x\,b/(1+\xi)})}{\sqrt{x+z\,\sqrt{x\,b/(1+\xi)}}\,\sqrt{1+\xi}}\,\mathrm{d}z\,\mathrm{d}x
$$

$$
= \frac{1}{4\sqrt{\pi\,b}}\,\mathbb{E}\left[X^{-1/2}\,g(X)\right](1+\xi)^{-1/2}.
$$

It then suffices to let t shrink to zero to appreciate that

$$
\liminf_{b\to 0} b^{-1/2}\,\Lambda_b^{(2)} \geq \frac{1}{4\sqrt{\pi}}\,\mathbb{E}\left[X^{-1/2}\,g(X)\right],
$$

and hence $\Lambda_b^{(2)} = \frac{1}{4\sqrt{\pi\,b}}\,\mathbb{E}\left[X^{-1/2}\,g(X)\right]$ since the upper and lower limits are equal. Applying the

same argument to $\Lambda_b^{(3)}$ and $\Lambda_b^{(4)}$ shows that they coincide with $\Lambda_b^{(2)}$ and $\Lambda_b^{(1)}$, respectively. This

means that $\Lambda_b = \Lambda_b^{(2)} + \Lambda_b^{(3)} = \frac{1}{2\sqrt{\pi\,b}}\,\mathbb{E}\left[X^{-1/2}\,g(X)\right]$, which completes the proof. ∎

Lemma A3: *Under the null* $\mathbb{H}_0$, I_{1n} *is such that* $\mathbb{E}(I_{1n}) = 2(2\pi b)^{-1/2}\,n^{-1}\,\mathbb{E}(X^{-\tau/2}) + O(n^{-1})$ *and*

$Var(I_{1n}) = O(b^{-1}n^{-3})$, *respectively.*

Proof: We derive the result only for the gamma kernel, but we can apply the same method to

the (reciprocal) inverse Gaussian kernel. For the gamma kernel, one can derive using the Stirling

approximation that $K_u(u) = (2\pi bu)^{-1/2}$. Hence

$$\mathbb{E}[I_{1n}] = n^{-1}(2\pi b)^{-1/2}\mathbb{E}\left[X^{-1/2} + Y^{-1/2}\right] - n^{-1}\mathbb{E}\left[g(X) + f(Y)\right] + O(bn^{-1})$$

$$= n^{-1}2(2\pi b)^{-1/2}\mathbb{E}\left[X^{-1/2}\right] + n^{-1}2(2\pi b)^{-1/2}\epsilon_n\left\{\int h(x)x^{-1/2}dx\right\} - n^{-1}2\mathbb{E}[g(X)] + O(bn^{-1})$$

$$= n^{-1}2(2\pi b)^{-1/2}\mathbb{E}\left[X^{-1/2}\right] + O\left(n^{-1}(1 + b^{-1/2}\epsilon_n)\right).$$

To calculate the variance of I_{1n}, set $l(Z, Z) = [K_X(X) + K_Y(Y)] - [K_X(Y) + K_Y(X)]$ and write $I_n = n^{-2}\sum_i h_n(Z_i, Z_i)$.

$$\mathbb{E}[I_{1n}]^2 = n^{-4}\sum_{i=1}^{n}\mathbb{E}\left[h_n(Z_i, Z_i)\right]^2 + n^{-4}\sum_{i\neq j}\mathbb{E}\left[h_n(Z_i, Z_i)h_n(Z_j, Z_j)\right]$$

$$= n^{-3}\mathbb{E}\left[h_n^2(Z, Z)\right] + n^{-2}\mathbb{E}^2\left[h_n(Z, Z)\right]$$

$$= n^{-3}\mathbb{E}\left[h_n^2(Z, Z)\right] + \mathbb{E}^2\left[I_{1n}\right]$$

$$= O(n^{-3}b^{-1}) + \mathbb{E}^2\left[I_{1n}\right].$$

The last line follows by writing

$$\mathbb{E}\left[h_n(Z, Z)^2\right] = \mathbb{E}\left[(K_X(X) + K_Y(Y))^2\right] + \mathbb{E}\left[(K_X(Y) + K_Y(X))^2\right]$$

$$- 2\mathbb{E}\left[(K_X(X) + K_Y(Y))(K_X(Y) + K_Y(X))\right]$$

$$= \mathbb{E}\left[K_X^2(X)\right] + \mathbb{E}\left[K_Y^2(Y)\right] + 2\mathbb{E}\left[K_X(X)\right]\mathbb{E}\left[K_Y(Y)\right]$$

$$+ \mathbb{E}\left[K_X^2(Y)\right] + \mathbb{E}\left[K_Y^2(X)\right] + 2\mathbb{E}\left[K_Y(X)K_X(Y)\right]$$

$$- 2\mathbb{E}\left[K_X(X)K_X(Y)\right] + 2\mathbb{E}\left[K_Y(Y)K_Y(X)\right]$$

$$- 2\mathbb{E}\left[K_X(X)K_Y(X)\right] + 2\mathbb{E}\left[K_Y(Y)K_X(Y)\right]$$

$$= O(b^{-1}),$$

where the last line follows by Lemmata A1 and A2, and the Stirling approximation. Note that we implicitly require $\mathbb{E}\left[X^{-\tau} + Y^{-\tau}\right] < \infty$. This means that

$$\text{Var}(I_{1n}) = \mathbb{E}\left[I_{1n}\right]^2 - \mathbb{E}^2\left[I_{1n}\right]$$

$$= O(b^{-1}n^{-3}),$$

proving the result. ∎

Lemma A4: *Assume that f and g are bounded density functions on $[0, \infty)$. Under the null $\mathbb{H}_0$, the variance of I_{2n} is $\text{Var}(I_{2n}) = n^{-2}b^{-1/2}\pi^{-1/2}\mathbb{E}\left\{X^{-\tau/2}\left[f(X) + g(X)\right]\right\}$.*

Proof: It follows from Lemma A1 that the expected value of I_{2n} is $O(\epsilon_n)$. To compute the asymptotic variance of I_{2n}, observe that

$$
\mathbb{E}\left[H_n(Z_1, Z_2)\right]^2 = \mathbb{E}\left[h_n(Z_1, Z_2) + h_n(Z_2, Z_1)\right]^2
$$

$$
= \mathbb{E}\left[h_n(Z_1, Z_2)\right]^2 + \mathbb{E}\left[h_n(Z_2, Z_1)\right]^2 + 2\,\mathbb{E}\left[h_n(Z_1, Z_2)\,h_n(Z_2, Z_1)\right]
$$

$$
= 2\,\mathbb{E}\left[h_n(Z_1, Z_2)\right]^2 + 2\,\mathbb{E}\left[h_n(Z_1, Z_2)\,h_n(Z_2, Z_1)\right], \tag{3}
$$

because both Z_1 and Z_2 are distributed with density $f(x)g(y)$. The squared-term in the righthand side of (3) decomposes into

$$
\mathbb{E}\left[h_n(Z_1, Z_2)\right]^2 = \mathbb{E}\left[K_{X_2}(X_1) + K_{Y_2}(Y_1) - K_{X_2}(Y_1) - K_{Y_2}(X_1)\right]^2
$$

$$
= \mathbb{E}\left[K_{X_2}(X_1)\right]^2 + \mathbb{E}\left[K_{Y_2}(Y_1)\right]^2 + \mathbb{E}\left[K_{X_2}(Y_1)\right]^2 + \mathbb{E}\left[K_{Y_2}(X_1)\right]^2
$$

$$
+ 2\,\mathbb{E}\left[K_{X_2}(X_1)\,K_{Y_2}(Y_1)\right] + 2\,\mathbb{E}\left[K_{X_2}(Y_1)\,K_{Y_2}(X_1)\right]
$$

$$
- 2\,\mathbb{E}\left[K_{X_2}(X_1)\,K_{Y_2}(X_1)\right] - 2\,\mathbb{E}\left[K_{X_2}(Y_1)\,K_{Y_2}(Y_1)\right]
$$

$$
- 2\,\mathbb{E}\left[K_{X_2}(X_1)\,K_{X_2}(Y_1)\right] - 2\,\mathbb{E}\left[K_{Y_2}(Y_1)\,K_{Y_2}(X_1)\right].
$$

Applying Lemmata A1 and A2, we conclude that,

$$
\mathbb{E}\left[h_n(Z_1, Z_2)\right]^2 = \frac{2}{\sqrt{\pi\, b}}\,\mathbb{E}\left[X^{-\tau/2} f(X)\right] + O(n^{-2} b^{-1/2} \epsilon_n).
$$

Now, it remains to bound the cross-term in (3), which under the null reads

$$
\mathbb{E}\left[h_n(Z_1, Z_2)\,h_n(Z_2, Z_1)\right] = \mathbb{E}\left[K_{X_1}(X_2) K_{X_2}(X_1)\right] + \mathbb{E}\left[K_{Y_1}(Y_2) K_{Y_2}(Y_1)\right]
$$

$$
+ 2\mathbb{E}\left[K_{X_1}(X_2)\right]\mathbb{E}\left[K_{Y_1}(Y_2)\right]
$$

$$
+ 2\mathbb{E}\left[K_X(Y) K_Y(X)\right] + 2\mathbb{E}\left[K_X(Y)\right]\mathbb{E}\left[K_Y(X)\right]
$$

$$
- 2\mathbb{E}\left[K_{X_1}(X_2) K_{X_2}(Y)\right] - 2\mathbb{E}\left[K_Y(X_2) K_{X_2}(X_1)\right]
$$

$$
- 2\mathbb{E}\left[K_{Y_1}(Y_2) K_{Y_2}(X)\right] - 2\mathbb{E}\left[K_X(Y_2) K_{Y_2}(Y_1)\right].
$$

We bound the right-hand side using Lemmata A1 and A2. It then ensues that,

$$
\mathrm{Var}(I_{2n}) = 2\, n^{-2}\, b^{-1/2}\, \pi^{-1/2}\,\mathbb{E}\left\{X^{-\tau/2}\left[f(X) + g(X)\right]\right\} + O\left(n^{-2} b^{-1/2} \epsilon_n\right),
$$

completing the proof. ∎

Lemma A5: *Let $A_n = \int [\hat{f}(x) - f(x)]^2\, \mathrm{d}x$. Assume also that the bandwidth b is of order $o(n^{-4/9})$ and that f is twice continuously differentiable and such that $\int [x^3\, f''(x)]^2\, \mathrm{d}x < \infty$. It then follows*

20

that $nb^{1/4} A_n - \frac{b^{-1/4}}{2\sqrt{\pi}} \mathbb{E}(X^{-\tau/2}) \xrightarrow{d} N(0, \sigma_f^2)$, *where* $\sigma_f^2 \equiv (2\pi)^{-1/2} \mathbb{E}[X^{-\tau/2} f(X)]$.

Proof: It readily follows along the same lines as in the proof of Theorem 1.1 in Fernandes and Monteiro (2005). ∎

Lemma A6: *Let the random sample* $(X_1, \ldots, X_n)$ *have a bounded density function* f *on* $[0, \infty)$. *Suppose that* $\lim_{x \to \infty} K_u'(x, b) f'(x) = 0$, *where* $K_u'(x, b)$ *and* $f'(x)$ *are the first derivatives with respect to* x, *and that the bandwidth is such that* $b \to 0$ *and* $n b^{3/2} \to \infty$. *It then follows that* $\left\| \hat{f}'(u) - f'(u) \right\| = O_p\left(n^{-1/2} b^{-3/4}\right)$, *where* $\hat{f}'(u) = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial u} K_u(X_i, b)$.

Proof: In the following, we derive the result only for the gamma kernel, though it is straightforward to consider the inverse Gaussian and reciprocal inverse Gaussian kernels. We first observe that

$$\frac{\partial}{\partial u} K_u(x, b) = K_u(x, b) \frac{\ln(x/b) - \psi(u/b + 1)}{b},$$

where $\psi(\cdot)$ is the digamma function, i.e., the logarithmic derivative of the gamma function (Abramowitz and Stegun, 1972, pp. 258-259). It then ensues that

$$\hat{f}'(u) = \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial u} K_u(X_i, b)$$

$$= \frac{1}{n} \sum_{i=1}^n K_u(X_i, b) \frac{\ln(X_i/b) - \psi(u/b + 1)}{b}.$$

Taking expectations then yields

$$\mathbb{E}\left[\hat{f}'(u)\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}\left[K_u(X_i, b) \frac{\ln(X_i/b) - \psi(u/b + 1)}{b}\right]$$

$$= b^{-1} \int_0^\infty K_u(x_1, b) \left[\ln(x_1/b) - \psi(u/b + 1)\right] f(x_1) \, dx_1.$$

The digamma function satisfies

$$\psi(z + 1) = \ln z + \frac{1}{2z} - \sum_{j=1}^\infty \frac{B_{2j}}{2j z^{2j}},$$

where B_{2j} are the Bernoulli numbers (Abramowitz and Stegun, 1972, pp. 804-806), and so

$$\mathbb{E}\left[\hat{f}'(u)\right] = b^{-1} \int_0^\infty K_u(x_1, b) \left[\ln(x_1/b) - \ln(u/b) - b/(2u) + O\left(b^2\right)\right] f(x_1) \, dx_1$$

$$= b^{-1} \int_0^\infty K_u(x_1, b) \ln(x_1/u) f(x_1) \, dx_1$$

$$\quad - \int_0^\infty K_u(x_1, b) f(x_1) \, dx_1 \{1/(2u) + O(b)\}$$

$$= b^{-1} \mathbb{E}_\zeta\left[\tilde{h}(\zeta)\right] - \mathbb{E}_\zeta\left[f(\zeta)\right] \{1/(2u) + O(b)\},$$

where $\tilde{h}(\zeta) \equiv \ln(\zeta/u) f(\zeta)$ and ζ has a gamma distribution $\mathcal{G}(u/b + 1, b)$, with mean $u + b$ and variance $(u + b)b$. A third-order Taylor expansion then yields

$$
\begin{aligned}
\mathbb{E}_\zeta \left[\tilde{h}(\zeta) \right] &= \tilde{h}(u) + b \left[\tilde{h}'(u) + \frac{1}{2} \tilde{h}''(u) u \right] + O(b^2) \\
&= \left[f'(u) + f(u)/(2u) \right] b + O(b^2),
\end{aligned}
$$

given that $\tilde{h}(u) = 0$, $\tilde{h}'(u) = f(u)/u$, and $\tilde{h}''(u) = 2 f'(u)/u - f(u)/u^2$. It follows from a second-order Taylor expansion that $\mathbb{E}_\zeta \left[f(\zeta) \right] = f(u) + O(b)$. It thus ensues that $\mathbb{E} \left[\hat{f}'(u) \right] = f'(u) + O(b)$. It now remains to compute the variance of the derivative estimator so as to prove convergence in the mean square sense. The variance of the derivative estimator is at most of the same order of its second moment, viz.

$$
\begin{aligned}
\mathbb{E} \left[\hat{f}'(u)^2 \right] &= n^{-1} b^{-2} \int_0^\infty K_u^2(x_1, b) \left[\ln(x_1/u) + O(b) \right]^2 f(x_1) \, \mathrm{d}x_1 \\
&= n^{-1} b^{-2} \int_0^\infty K_u^2(x_1, b) \left[\ln(x_1/u) \right]^2 f(x_1) \, \mathrm{d}x_1 \left\{ 1 + O(b) \right\} \\
&= n^{-1} b^{-2} B_b(u) \, \mathbb{E}_\eta \left\{ \left[\ln(\eta/u) \right]^2 f(\eta) \right\} \left\{ 1 + O(b) \right\},
\end{aligned}
$$

where η is a random variate with gamma distribution $\mathcal{G}(2u/b + 1, b/2)$, and

$$
B_b(u) = \frac{\Gamma(2u/b + 1)/b}{2^{2u/b+1} \Gamma^2(u/b + 1)}
$$

as in Fernandes and Monteiro (2005). It then follows from the properties of the standard gamma distribution that the above expectation is of order $O(b)$, whereas Chen (2000) shows that $\sqrt{b\, u}\, B_b(u) \leq 1/(2\sqrt{\pi})$. This means that

$$
\mathbb{E} \left[\hat{f}'(u)^2 \right] \leq n^{-1} \frac{b^{-5/2} \, u^{-1/2}}{2\sqrt{\pi}} O(b) \left\{ 1 + O(b) \right\} = O \left(n^{-1} b^{-3/2} \right).
$$

This shows that the derivative estimator that employs the gamma kernel is consistent as long as $b \to 0$ and $n\, b^{3/2} \to \infty$. ∎

Proof of Proposition 1: Note that we may write the test statistic as

$$
\frac{n\, b^{1/4} I_n - b^{1/4} \hat{\mu}_1}{\hat{\sigma}} = \left[\frac{n\, b^{1/4} I_{1n} - b^{-1/4} \mu_1}{\sigma_0} + b^{-1/4} \frac{\mu_1 - \hat{\mu}_1}{\sigma_0} \right] \frac{\sigma_0}{\hat{\sigma}}.
$$

The second term within brackets is $o_p(1)$ given the assumption that $\hat{\mu}_1 - \mu_1 = o_p(b^{1/4})$, whereas the consistency of $\hat{\sigma}$ ensures that $\sigma_0/\hat{\sigma}$ converges to one in probability. Altogether, this means that

$$
\frac{n\, b^{1/4} I_n - b^{1/4} \hat{\mu}_1}{\hat{\sigma}} = \frac{n\, b^{1/4} I_n - b^{1/4} \mu_1}{\hat{\sigma}} + o_p(1)
$$

22

by Slutsky theorem. It then follows from Lemmata A3 to A5 that the first term on the right-hand side converges to a standard Gaussian distribution, completing the proof of (i). It is easy to show that (ii) holds along the same lines. ∎

Proof of Proposition 2: The result follows along the same lines as in Proposition 1 by noting that $\lambda_n \to 1$ under the null as the sample sizes grow. ∎

Proof of Proposition 3: This is a particular case of the general results in Anderson et al. (1994). ∎

Proof of Proposition 4: It follows from the proof of Lemma A4 that under the local alternative hypothesis $\mathbb{H}_1^{(n)}: \ g(y) = f(y) + \epsilon_n h(y)$, $\mathbb{E}[nb^{1/4} I_{2n}] = n\, b^{1/4}\, \epsilon_n\, \mu_h$ and $\mathrm{Var}[n\, b^{1/2}\, I_{2n}] = \sigma_0^2 + \sigma_h^2$. As in Lemma A5, this implies that $n\, b^{1/4}\, I_{2n} - n\, b^{1/4}\, \epsilon_n \mu_h \xrightarrow{d} N\left(0, \sqrt{\sigma_0^2 + \sigma_h^2}\right)$. The result then follows trivially by replacing $n = o(b^{-9/4})$. ∎

Proof of Proposition 5: By assumption, there exists $\widehat{U} = U(\widehat{\xi}, \widehat{\theta}) = \widehat{\xi}(V_1, V_2; \widehat{\theta})$ that converges in probability to $U = \xi(V_1, V_2; \theta_0)$ at a rate N^d, with $d \geq 4/9$. The test is nuisance parameter free if the statistic evaluated at $(\widehat{\xi}, \widehat{\theta})$ converges to the same distribution of the statistic evaluated at (ξ, θ_0). For ease of exposition, we assume in what follows that $n_1 = n_2 = n$. The test statistic then reads

$$I_n(\widehat{\xi}, \widehat{\theta}) = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j\neq i, j=1}^{n} K_{\widehat{X}_j}(\widehat{X}_i) + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j\neq i, j=1}^{n} K_{\widehat{Y}_j}(\widehat{Y}_i)$$
$$- \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j\neq i, j=1}^{n} K_{\widehat{Y}_j}(\widehat{X}_i) - \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j\neq i, j=1}^{n} K_{\widehat{X}_j}(\widehat{Y}_i),$$

where $\left(\widehat{X}_1 \ldots, \widehat{X}_n\right)$ and $\left(\widehat{Y}_1 \ldots, \widehat{Y}_n\right)$ respectively correspond to the nonnegative values and the absolute value of the nonpositive values of $\left(\widehat{U}_1 \ldots, \widehat{U}_N\right)$, with $n = N/2$. A second-order Taylor functional expansion of $I_n(\widehat{\xi}, \widehat{\theta})$ around (ξ, θ_0) yields

$$I_n(\widehat{\xi}, \widehat{\theta}) - I_n(\xi, \theta_0) = \Delta_1(\xi, \theta_0)(\widehat{\xi} - \xi) + \Delta_2(\xi, \theta_0)(\widehat{\theta} - \theta_0) + R_n(\xi_*, \theta_*),$$

where $\Delta_1(\xi, \theta_0)$ and $\Delta_2(\xi, \theta_0)$ respectively denote the functional derivative of I_n with respect to the first and second arguments evaluated at (ξ, θ_0), and $R_n(\xi_*, \theta_*)$ denotes the residual term of the expansion. The latter depends on the second functional derivative of I_n evaluated at (ξ_*, θ_*), with $\xi_* \in [\xi, \widehat{\xi}]$ and $\theta_* \in [\theta_0, \widehat{\theta}]$, and on both $(\widehat{\xi} - \xi)^2$ and $(\widehat{\theta} - \theta_0)^2$. The limiting distributions of

$n\, b^{1/4}\, I_n(\widehat{\xi}, \widehat{\theta})$ and $n\, b^{1/4}\, I_n(\xi, \theta_0)$ then coincide if and only if

$$\Delta_1(\xi, \theta_0)(\widehat{\xi} - \xi) + \Delta_2(\xi, \theta_0)(\widehat{\theta} - \theta_0) = o_p\left(n^{-1}\, b^{-1/4}\right) = o_p\left(n^{-8/9}\right),$$

given that b is of order $o\left(n^{-4/9}\right)$. Both Δ_1 and Δ_2 depend linearly on the product of the density functions f and g and their first functional derivatives. In addition, they are equal to zero under the null hypothesis. The norm of the density estimation error is of order $O_p\left(n^{-1/2}\, b^{-1/2}\right)$, whereas Lemma A6 shows that the norm of the first-derivative estimation error is of order $O_p\left(n^{-1/2}\, b^{-3/4}\right)$. This means that both Δ_1 and Δ_2 are of order $O_p\left(n^{-1}\, b^{-5/4}\right) = o_p\left(n^{-4/9}\right)$ and hence

$$\Delta_1(\xi, \theta_0)(\widehat{\xi} - \xi) + \Delta_2(\xi, \theta_0)(\widehat{\theta} - \theta_0) = o_p\left(n^{-4/9-d}\right).$$

We now move to the residual term of the expansion. It is straightforward to show that the supremum of the second functional derivative of I_n over a neighborhood of (ξ_*, θ_*) is also of order $O\left(n^{-1}\, b^{-3}\right)$ and so $n\, b^{1/4}\, R_n(\xi_*, \theta_*) = O_p\left(n^{-2d}\, b^{-11/4}\right) = o_p\left(n^{-2d+11/9}\right)$. The latter condition is not binding in view that d must already exceed $4/9$. ∎

References

Abramowitz, M. and Stegun, I. A. (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover.

Ahmad, I. A. and Li, Q. (1997). Testing symmetry by kernel methods, *Journal of Nonparametric Statistics* **7**: 279–293.

Ahmad, I. A. and van Belle, G. (1974). Mesuring affinity of distributions, *in* F. Proschan and R. J. Serfling (eds), *Reliability and Biometry: Statistical Analysis of Life Testing*, SIAM, Philadelphia.

Aït-Sahalia, Y., Bickel, P. J. and Stoker, T. M. (2001). Goodness-of-fit tests for kernel regression with an application to option implied volatilities, *Journal of Econometrics* **105**: 363412.

Aït-Sahalia, Y., Fan, J.-i. and Peng, H. (2009). Nonparametric transition-based tests for diffusions, *Journal of the American Statistical Association* **104**: 1102–1116.

Anderson, N. H., Hall, P. and Titterington, D. M. (1994). Two-sample test statistics for measuring discrepancies between two multivariate probability density functions using kernel-based density estimates, *Journal of Multivariate Analysis* **50**: 41–54.

Bai, J. and Ng, S. (2001). A consistent test for conditional symmetry in time series models, *Journal of Econometrics* **103**: 225–258.

Bickel, P. J. (1982). On adaptive estimation, *Annals of Statistics* **10**: 647–671.

Bickel, P. J. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates, *Annals of Statistics* **1**: 1071–1095.

Bouezmarni, T. and Rombouts, J. V. K. (2010). Nonparametric density estimation for multivariate bounded data, *Journal of Statistical Planning and Inference* **140**: 139–152.

Bouezmarni, T. and Scaillet, O. (2005). Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data, *Econometric Theory* **21**: 390–412.

Campbell, J. Y. and Hentschel, L. (1992). No news is good news, *Journal of Financial Economics* **31**: 281–318.

Chen, S. X. (2000). Probability density function estimation using gamma kernels, *Annals of the Institute of Statistical Mathematics* **52**: 471–480.

Delgado, M. A. and Escanciano, J. C. (2007). Nonparametric tests for conditional symmetry in dynamic models, *Journal of Econometrics* **141**: 652–682.

Fan, Y. (1995). Bootstrapping a consistent nonparametric goodness-of-fit test, *Econometric Reviews* **14**: 367–382.

Fan, Y. (1998). Goodness-of-fit tests based on kernel density estimators with fixed smoothing parameters, *Econometric Theory* **14**: 604–621.

Fan, Y. and Gencay, R. (1993). Hypotheses testing based on modified nonparametric estimation of an affinity measure between two distributions, *Journal of Nonparametric Statistics* **2**: 389–403.

Fan, Y. and Gencay, R. (1995). A consistent nonparametric test of symmetry in linear regression models, *Journal of the American Statistical Association* **90**: 551–557.

Fan, Y. and Ullah, A. (1999). On goodness-of-fit tests for weakly dependent processes using kernel methods, *Journal of Nonparametric Statistics* **11**: 337–360.

Fernandes, M. and Grammig, J. (2005). Nonparametric specification tests for conditional duration models, *Journal of Econometrics* **127**: 35–68.

Fernandes, M. and Monteiro, P. K. (2005). Central limit theorem for asymmetric kernel functionals, *Annals of the Institute of Statistical Mathematics* **57**: 425–442.

Gouriéroux, C., Laurent, J. P. and Scaillet, O. (2000). Sensitivity analysis of Value at Risk, *Journal of Empirical Finance* **7**: 225–245.

Gustafsson, J., Hagmann, M., Nielsen, J. P. and Scaillet, O. (2009). Local transformation kernel density estimation of loss distributions, *Journal of Business and Economic Statistics* **27**: 161–175.

Hagmann, M. and Scaillet, O. (2007). Local multiplicative bias correction for asymmetric kernel density estimators, *Journal of Econometrics* **141**: 213–249.

Hong, Y. and White, H. (2004). Asymptotic distribution theory for nonparametric entropy measures of serial dependence, *Econometrica* **73**: 837–901.

Khuri, A. and Casella, G. (2002). The existence of the first negative moment revisited, *The American Statistician* **56**: 44–47.

Koroljuk, V. S. and Borovskich, Y. V. (1994). *Theory of U-statistics*, Kluwer Academic Publishers.

Lambert, P., Laurent, S. and Veredas, D. (2012). Testing conditional asymmetry: A residual-based approach, *Journal of Economic Dynamics and Control* **36**: 1229–1247.

Newey, W. K. (1988). Adaptive estimation of regression models via moment restrictions, *Journal of Econometrics* **38**: 301–339.

Powell, J. L. (1994). Estimation of semiparametric models, *in* R. F. Engle and D. L. MacFadden (eds), *Handbook of Econometrics IV*, Elsevier Science, Amsterdam, chapter 41, pp. 2443–2521.

Randles, R. H., Fligner, M. A., Policello, G. E. and Wolf, D. A. (1980). An asymptotically distribution-free test for symmetry versus asymmetry, *Journal of the American Statistical Association* **75**: 168–172.

Robinson, P. M. (1991). Consistent nonparametric entropy-based testing, *Review of Economic Studies* **58**: 437–453.

Scaillet, O. (2004). Density estimation using inverse and reciprocal inverse Gaussian kernels, *Journal of Nonparametric Statistics* **16**: 217–226.

Scaillet, O. (2007). Kernel based goodness-of-fit tests for copulas with fixed smoothing parameters, *Journal of Multivariate Analysis* **98**: 533–543.

Shen, X. (1997). On methods of sieves and penalization, *Annals of Statistics* **25**: 2555–2591.

Shen, X. and Wong, W. H. (1994). Convergence rates of sieve estimates, *Annals of Statistics* **22**: 580–615.

Silverman, B. W. (1986). *Density Estimation in Statistics and Data Analysis*, Chapman and Hall, London.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression, *Annals of Statistics* **10**: 1040–1053.

Zheng, J. X. (1998). Consistent specification testing for conditional symmetry, *Econometric Theory* **14**: 139–149.

Table 1

Distributions in the Monte Carlo study

The quantile function of the lambda distribution with parameter vector $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ is $F^{-1}(u) = \lambda_1 + [u^{\lambda_3} - (1-u)^{\lambda_4}]/\lambda_2$.

	distribution	skewness	kurtosis
S1	standard normal	0.000	3.00
S2	t-student with 10 degrees of freedom	0.000	4.00
S3	lambda with $\boldsymbol{\lambda} = (0, 0.1666667, 1, 1)$	0.000	1.80
S4	lambda with $\boldsymbol{\lambda} = (0, -0.0141264, -0.08, -0.08)$	0.000	5.99
A1	lambda with $\boldsymbol{\lambda} = (-12.601, -0.00980045, -0.11, -0.0001)$	-2.924	19.52
A2	lambda with $\boldsymbol{\lambda} = (9.7726, 0.0151878, -0.001, -0.13)$	3.160	23.80
A3	lambda with $\boldsymbol{\lambda} = (-7.84595, -0.0223643, -0.15, -0.001)$	-3.478	30.24
A4	lambda with $\boldsymbol{\lambda} = (6.43871, 0.00317949, -0.001, -0.17)$	3.880	40.70

Table 2

Performance of the gamma-kernel symmetry test using I_n with vanishing bandwidth

We report the rejection frequency of the asymptotic test based on I_n within 1,000 replications. In particular, we test whether the standardized least-squares residuals of a linear regression with both intercept and slope coefficients equal to one are symmetric for sample sizes of 50, 100, 200, and 400 observations. We select the bandwidth by means of a cross-validation approach. To assess size and power, we draw the error term from both symmetric and asymmetric distributions. In particular, we consider the standard Gaussian (S1), t-student with 10 degrees of freedom (S2), two symmetric lambda distributions (S3 and S4), as well as four asymmetric lambda distributions (A1 to A4). We also report within parentheses and brackets the corresponding values for Zheng (1998) and Bai and Ng (2001), respectively. See Table 1 for the values of the parameters in the above distributions.

distribution	50		100		200		400	
	5%	10%	5%	10%	5%	10%	5%	10%
S1	0.048	0.090	0.060	0.090	0.067	0.139	0.090	0.155
	(0.052)	(0.086)	(0.057)	(0.101)	(0.070)	(0.093)		
	[0.037]		[0.051]		[0.049]			
S2	0.048	0.108	0.063	0.081	0.072	0.120	0.156	0.210
	(0.069)	(0.114)	(0.089)	(0.134)	(0.070)	(0.103)		
S3	0.100	0.166	0.125	0.195	0.174	0.267	0.181	0.284
	(0.034)	(0.069)	(0.039)	(0.061)	(0.029)	(0.055)		
S4	0.093	0.145	0.073	0.107	0.109	0.168	0.093	0.163
	(0.068)	(0.101)	(0.074)	(0.131)	(0.078)	(0.125)		
	[0.078]		[0.087]		[0.075]			
A1	0.185	0.315	0.435	0.625	0.870	0.933	0.996	0.999
			(0.149)	(0.229)	(0.320)	(0.439)	(0.474)	(0.616)
A2	0.176	0.314	0.450	0.621	0.875	0.944	0.993	0.997
			(0.168)	(0.257)	(0.363)	(0.479)	(0.541)	(0.645)
	[0.932]		[0.999]		[1.000]			
A3	0.195	0.334	0.457	0.654	0.889	0.951	0.996	0.997
			(0.184)	(0.286)	(0.400)	(0.513)	(0.560)	(0.692)
A4	0.194	0.320	0.491	0.671	0.895	0.952	0.993	0.995
			(0.215)	(0.303)	(0.436)	(0.541)	(0.610)	(0.722)
	[0.961]		[1.000]		[1.000]			

Table 3

Performance of the gamma-kernel symmetry test using I_{2n} with vanishing bandwidth

We report the rejection frequency of the asymptotic test based on I_{2n} within 1,000 replications. See Table 2 for more details.

distribution	50		100		200		400	
	5%	10%	5%	10%	5%	10%	5%	10%
S1	0.054	0.084	0.030	0.090	0.055	0.091	0.048	0.096
			(0.052)	(0.086)	(0.057)	(0.101)	(0.070)	(0.093)
	[0.037]		[0.051]		[0.049]			
S2	0.048	0.072	0.051	0.093	0.042	0.072	0.060	0.138
			(0.069)	(0.114)	(0.089)	(0.134)	(0.070)	(0.103)
S3	0.073	0.107	0.093	0.149	0.096	0.175	0.108	0.190
			(0.034)	(0.069)	(0.039)	(0.061)	(0.029)	(0.055)
S4	0.060	0.096	0.046	0.074	0.061	0.111	0.056	0.107
			(0.068)	(0.101)	(0.074)	(0.131)	(0.078)	(0.125)
	[0.078]		[0.087]		[0.075]			
A1	0.324	0.449	0.715	0.820	0.984	0.994	0.999	0.999
			(0.149)	(0.229)	(0.320)	(0.439)	(0.474)	(0.616)
A2	0.332	0.450	0.674	0.805	0.980	0.990	1.000	1.000
			(0.168)	(0.257)	(0.363)	(0.479)	(0.541)	(0.645)
	[0.932]		[0.999]		[1.000]			
A3	0.337	0.464	0.712	0.821	0.985	0.994	1.000	1.000
			(0.184)	(0.286)	(0.400)	(0.513)	(0.560)	(0.692)
A4	0.346	0.480	0.731	0.831	0.989	0.995	0.998	0.999
			(0.215)	(0.303)	(0.436)	(0.541)	(0.610)	(0.722)
	[0.961]		[1.000]		[1.000]			