Econometrica, Vol. 73, No. 3 (May, 2005), 837–901

ASYMPTOTIC DISTRIBUTION THEORY FOR NONPARAMETRIC ENTROPY MEASURES OF SERIAL DEPENDENCE

BY YONGMIAO HONG AND HALBERT WHITE[1]

Entropy is a classical statistical concept with appealing properties. Establishing asymptotic distribution theory for smoothed nonparametric entropy measures of dependence has so far proved challenging. In this paper, we develop an asymptotic theory for a class of kernel-based smoothed nonparametric entropy measures of serial dependence in a time-series context. We use this theory to derive the limiting distribution of Granger and Lin's (1994) normalized entropy measure of serial dependence, which was previously not available in the literature. We also apply our theory to construct a new entropy-based test for serial dependence, providing an alternative to Robinson's (1991) approach. To obtain accurate inferences, we propose and justify a consistent smoothed bootstrap procedure. The naive bootstrap is not consistent for our test. Our test is useful in, for example, testing the random walk hypothesis, evaluating density forecasts, and identifying important lags of a time series. It is asymptotically locally more powerful than Robinson's (1991) test, as is confirmed in our simulation. An application to the daily S&P 500 stock price index illustrates our approach.

KEYWORDS: Density forecasts, entropy, invariance, jackknife kernel, nonlinear time series, random walk, serial dependence, smoothed bootstrap.

1. INTRODUCTION

MEASURING AND TESTING FOR SERIAL DEPENDENCE are central to time series analysis (e.g., Granger and Teräsvirta (1993), Robinson (1991), Tjøstheim (1996)). A conventional measure of serial dependence is the autocorrelation function, which may overlook essential nonlinear features of time series that have zero autocorrelation. As Granger and Teräsvirta (1993) pointed out, there are few simple suitable tools for analyzing nonlinear time series, although significant effort has been devoted to developing effective measures of and tests for serial dependence.

Of increasing interest recently are smoothed nonparametric entropy measures of and tests for serial dependence. These tools avoid restrictive parametric assumptions on the probability distribution generating the data and can capture all pairwise dependencies in the lags of the series. Further, they

[1]We thank the co-editor, three referees, Bin Chen, Max Chen, T. W. Epps, Clive W. J. Granger, Jinyong Hahn, Tae-Hwy Lee, Yoon-Jin Lee, Oliver Linton, Yanhui Liu, Joon Park, Sam Thompton, Aman Ullah, Jeffrey Wooldridge, Adonis Yatchew, and Wenjie Zhang, as well as seminar participants at Brown, the Chinese Academy of Science, PSU, Harvard–MIT, UC Riverside, Virginia, the BK21 International Conference in Econometrics at Sungkyunkwan University, Seoul, Korea, the Fifth International Conference of the International Chinese Statistical Association in Hong Kong, the International Symposium on Complexity Science and EconoPhysics in Hefei, China, and the Winter Meeting of the North American Econometric Society in New Orleans, for helpful comments and discussions. Hong and White's participations were supported by National Science Foundation Grants SES-0117649, SES-0111238, and SBR-9811562. Any remaining errors are solely ours.

have an appealing information-theoretic interpretation and are invariant under any continuous monotonic transformation of the data (Robinson (1991), Granger and Lin (1994), Skaug and Tjøstheim (1996)). Joe (1989a, 1989b) considered a smoothed nonparametric entropy measure of multivariate dependence of an independent and identically distributed (i.i.d.) random vector. Granger and Lin (1994) proposed a normalized smoothed nonparametric entropy measure of serial dependence to identify important lags in time series. Robinson (1991) developed a test for serial dependence using a modified entropy measure. Skaug and Tjøstheim (1993a, 1996) also considered a general class of smoothed density-based tests for serial dependence, which includes a test based on an entropy measure modified with a weight function. In a different but related context, White (1982, p. 17) suggested a smoothed nonparametric entropy-based approach to testing parametric conditional density specifications. For more discussion on entropy and its applications in econometrics, see Maasoumi (1993) and Ullah (1993).

As Granger and Lin (1994) pointed out, there is no asymptotic distribution theory available for smoothed nonparametric entropy measures of serial dependence. In fact, there is not even an asymptotic distribution theory for Joe's (1989a, 1989b) smoothed entropy measure of multivariate dependence in an i.i.d. context. Various smoothed nonparametric entropy estimators in an i.i.d. context have been considered in the literature (e.g., Ahmad and Lin (1976), Mokkadem (1989)). Consistency and in some cases convergence rates have been established, but asymptotic distributions for these entropy estimators are not available. This has hindered application of such otherwise appealing measures.

In pioneering work, Robinson (1991) first provided an asymptotic distribution theory for a smoothed nonparametric modified entropy measure of serial dependence, using a sample-splitting device. The great appeal of Robinson's approach is that it yields a limiting $N(0, 1)$ distribution under the i.i.d. hypothesis and that sample splitting does not render the modified entropy estimator inconsistent for the population entropy. Indeed, Robinson (1991) established the consistency of his test against a wide class of stationary ergodic processes. Nevertheless, Robinson's theory does not apply to smoothed nonparametric entropy estimators that do not use a sample-splitting device, such as the Granger–Lin normalized entropy measure. Further, the sample-splitting device involves some tuning parameters. As Robinson (1991) noted, the choice of these parameters remains open. Two practitioners using different tuning parameters may reach conflicting conclusions in finite samples. Another difficulty is that the sample-splitting device breaks down when the marginal distribution of the series is uniform. This case can arise, for example, in evaluating density forecasts, which are important for financial risk management, options pricing, and macroeconomic policy control (e.g., Diebold, Gunther, and Tay (1998)). Most importantly, as we show below, the sample-splitting device leads to suboptimal asymptotic local power.

In this paper we provide an asymptotic theory for a class of kernel-based smoothed nonparametric entropy estimators of serial dependence. Our theory yields the limit distribution of the Granger–Lin normalized entropy measure, which was previously unknown in the literature. We also use our theory to construct a new test for serial dependence, providing an alternative to Robinson's (1991) test. To obtain accurate inferences in finite samples, we propose and justify a consistent smoothed bootstrap procedure for our test. Interestingly, the naive bootstrap procedure is not consistent for our test, in spite of the i.i.d. null hypothesis. Our test is useful in (e.g.) testing the random walk hypothesis, evaluating density forecasts, and identifying important lags of a time series. It does not involve sample splitting and thus does not require choosing tuning parameters. We show that our test is asymptotically locally more powerful than Robinson's test. This is confirmed by simulation.

There are a number of other nonparametric measures of and tests for serial dependence in the literature. Besides smoothed density estimators, alternative tools are correlation integrals, empirical distribution functions, and empirical characteristic functions (Brock, Dechert, Scheinkman, and LeBaron (1996), Delgado (1996), Hong (1998, 1999), Pinkse (1998), Skaug and Tjøstheim (1993b)). Smoothed density-based entropy estimators for serial dependence, however, have their own particular merits (Granger and Lin (1994), Skaug and Tjøstheim (1993a, 1996)). In the context of density forecasts, for example, Diebold et al. (1998) showed that to assess whether a sequence of density forecasts is optimal, it suffices to check whether a series of probability integral transforms with respect to forecast densities is i.i.d. U[0, 1]. Here, it is natural to use a smoothed density-based test, and our approach provides a suitable test for this joint hypothesis.

In Section 2, we briefly review the concept of entropy and smoothed nonparametric entropy measures of serial dependence. In Section 3, we derive the limiting distribution for a class of kernel-based smoothed nonparametric entropy measures of serial dependence, which is then applied to derive the limiting distribution of the Granger–Lin normalized entropy measure. In Section 4, we use our theory to construct a new entropy-based test for serial dependence and to justify the consistency of a smoothed bootstrap procedure for our test. Asymptotic local power is studied in Section 5. In Section 6, we conduct a simulation study comparing our test, Robinson's (1991) test, and Skaug and Tjøstheim's (1996) test in finite samples. Section 7 presents an application to the S&P 500 stock index. GAUSS code for our test is available from the authors. All proofs are in the Appendixes. Throughout, all limits are taken as the sample size $n \to \infty$, and $C \in (0, \infty)$ denotes a generic bounded constant.

2. ENTROPY MEASURES OF SERIAL DEPENDENCE

Suppose $\{X_t\}$ is a strictly stationary time series with marginal density $g(\cdot)$ and pairwise joint density $f_j(\cdot)$ for $Z_{jt} \equiv (X_t, X_{t-j})'$, where $j \in \mathbb{N} \equiv \{1, 2, \dots\}$ is a given lag order. An important issue in time-series analysis (particularly nonlinear time-series analysis) is the measurement of serial dependence in $\{X_t\}$. For continuous distributions, X_t and X_{t-j} are independent if and only if $f_j(\cdot) = g(\cdot)g(\cdot)$ almost everywhere in the support of Z_{jt}. Any deviation of $f_j(\cdot)$ from $g(\cdot)g(\cdot)$ is evidence of serial dependence. To measure deviations of $f_j(\cdot)$ from $g(\cdot)g(\cdot)$, one can use the Kullback–Leibler information criterion

$$(2.1) \qquad \mathcal{I}(j) \equiv \int \ln\left[\frac{f_j(x, y)}{g(x)g(y)}\right] f_j(x, y)\, dx\, dy, \quad j \in \mathbb{N},$$

where the integral is taken over the support of Z_{jt}. Although $\mathcal{I}(j)$ is not a metric, it can characterize all pairwise serial dependencies because $\mathcal{I}(j) \geq 0$, and $\mathcal{I}(j) = 0$ if and only if X_t and X_{t-j} are independent.[2] Moreover, it has an appealing information-theoretic interpretation, and it is invariant under any continuous monotonic transformation of $\{X_t\}$. The invariance property of $\mathcal{I}(j)$ is attractive because $\{X_t\}$ is i.i.d. if and only if any series of its continuous monotonic transform is i.i.d.

Granger and Lin (1994) examined the properties of a normalized version of $\mathcal{I}(j)$,

$$(2.2) \qquad \gamma^2(j) \equiv 1 - \exp[-2\mathcal{I}(j)], \quad j \in \mathbb{N},$$

and interpreted it as a shadow autocorrelation of $\{X_t\}$. They proposed estimators

$$(2.3) \qquad \hat{\gamma}_n^2(j) \equiv 1 - \exp[-2\hat{\mathcal{I}}_n(j)] \qquad (j = 1, \dots, n-1),$$

where $\hat{\mathcal{I}}_n(j)$ is a smoothed nonparametric entropy estimator based on the sample $\mathcal{X} \equiv \{X_t\}_{t=1}^n$,

$$(2.4) \qquad \hat{\mathcal{I}}_n(j) \equiv (n-j)^{-1} \sum_{t \in S_n(j)} \ln\left[\frac{\hat{f}_{jt}(Z_{jt})}{\hat{g}_t(X_t)\hat{g}_{t-j}(X_{t-j})}\right] \qquad (j = 1, \dots, n-1),$$

where $\hat{f}_{jt}(\cdot)$ and $\hat{g}_t(\cdot)$ are kernel estimators for $f_j(\cdot)$ and $g(\cdot)$, and $S_n(j) \equiv$

[2] A similar measure that is a metric is the Hellinger distance, considered by Maasoumi and Racine (2002) and Maasoumi, Racine, and Granger (2004). Treatment of this measure is beyond the scope of our present analysis, although there are common elements, and our approach is applicable there.

$\{t \in \mathbb{N}: j < t \leq n, \hat{f}_{jt}(Z_{jt}) > 0, \hat{g}_t(X_t) > 0, \hat{g}_{t-j}(X_{t-j}) > 0\}$. They examined the finite sample performance of $\hat{\gamma}_n^2(j)$ in identifying important lags of a variety of linear and nonlinear time-series models. As Granger and Lin (1994) pointed out, however, no limiting distribution theory is available for $\hat{\mathcal{I}}_n(j)$ or $\hat{\gamma}_n^2(j)$. This has hindered application of their measure. In fact, Robinson (1991) first observed and elegantly explained why no scaling of $\hat{\mathcal{I}}_n(j)$ has a known limiting distribution under the null hypothesis $\mathbb{H}_0$ that $\{X_t\}$ is i.i.d. In particular, no scaling of $\hat{\mathcal{I}}_n(j)$ has a limiting null zero-mean normal distribution. Zheng (2000), in an i.i.d. context, also observed the difficulty of obtaining the limiting distribution of a White (1982, p. 17) entropy-based test statistic for parametric conditional density specification. To avoid this difficulty, Zheng used instead an alternative divergence measure which may be viewed as a modified first-order term of the Taylor series expansion of an entropy statistic. This measure, however, loses some appealing properties (e.g., invariance and the information-theoretic interpretation) of the entropy measure.

A main stumbling block to the asymptotic distribution theory for $\hat{\mathcal{I}}_n(j)$ is that under $\mathbb{H}_0$, $\hat{\mathcal{I}}_n(j) \xrightarrow{p} 0$ at a rate $n^{-1/2-\epsilon_n}$, where $\epsilon_n \geq c > 0$ depends on the smoothing parameters used in $\hat{f}_{jt}(\cdot)$ and $\hat{g}_t(\cdot)$. Consequently, the usual $n^{1/2}$ normalization leads to a degenerate statistic because $n^{1/2}\hat{\mathcal{I}}_n(j) \xrightarrow{p} 0$. To overcome this difficulty, Robinson (1991) proposed a modified entropy estimator that, when $j = 1$, has the form

$$\hat{\mathcal{I}}_{n,\gamma}(j) \equiv (n-j)^{-1} \sum_{t \in S_n(j)} C_t(\gamma) \ln\left[\frac{\hat{f}_{jt}(Z_{jt})}{\hat{g}_t^2(X_t)}\right] \qquad (j = 1, \ldots, n-1), \tag{2.5}$$

where the weight function $C_t(\gamma) = 1 + j\gamma$ if $t = 1, \operatorname{mod}(j+1)$ and $C_t(\gamma) = 1 - \gamma$ otherwise, with $\gamma \in (0, 1)$. Although Robinson's (1991) estimator has a different form when $j > 2$, (2.5) nevertheless defines a useful entropy estimator. As Robinson (1991) pointed out, the use of $C_t(\gamma)$ is essentially a form of sample splitting. It alters the convergence rate of $\hat{\mathcal{I}}_{n,\gamma}(j)$ under $\mathbb{H}_0$ so that a $n^{1/2}$ normalization yields a nondegenerate limiting normal distribution. Moreover, $\hat{\mathcal{I}}_{n,\gamma}(j)$ is still consistent for $\mathcal{I}(j)$ for any fixed $j > 0$ under stationary ergodicity. As noted above, however, this approach involves choosing the tuning parameter γ. Further, the approach breaks down when X_t has a uniform distribution under $\mathbb{H}_0$ because in this case one still has $n^{1/2}\hat{\mathcal{I}}_{n,\gamma}(j) \xrightarrow{p} 0$ under $\mathbb{H}_0$. This can arise in (e.g.) evaluating density forecasts (Diebold et al. (1998)). Most importantly, the test based on $\hat{\mathcal{I}}_{n,\gamma}(j)$ suffers from an asymptotic local power loss, as we show below.

3. ASYMPTOTIC DISTRIBUTION

We now develop an asymptotic distribution theory for $\hat{\mathcal{I}}_n(j)$ of (2.4). We show that after an adjustment for its asymptotic mean, a proper scaling of $\hat{\mathcal{I}}_n(j)$, which differs from $n^{1/2}$, has a limiting $\mathrm{N}(0,1)$ distribution under $\mathbb{H}_0$. Throughout, we impose the following condition on $\{X_t\}$.

ASSUMPTION A.1: *Take $\{X_t\}$ to be strictly stationary with X_t having support on $\mathbb{I} \equiv [0,1]$. Its marginal density $g:\mathbb{I} \to \mathbb{R}^+$ exists, is bounded away from 0, and is continuously twice differentiable on $\mathbb{I}$. Moreover, $|g^{(2)}(x_1) - g^{(2)}(x_2)| \le C|x_1 - x_2|$ for any $x_1, x_2 \in \mathbb{I}$.*

Throughout, we use the following convention to define derivatives at the end points of $\mathbb{I}$:

$$g^{(d)}(0) \equiv \lim_{x\to 0^+} \frac{g^{(d-1)}(0+x) - g^{(d-1)}(0)}{x} \qquad (d=1,2),$$

$$g^{(d)}(1) \equiv \lim_{x\to 0^-} \frac{g^{(d-1)}(1+x) - g^{(d-1)}(1)}{x} \qquad (d=1,2).$$

Assumption A.1, as assumed in Robinson (1991) and Hall (1988), avoids the awkward problem of treating entropy in the tails. It allows us to focus on the essentials and still maintain a relatively straightforward treatment. Compact support may at first look restrictive, but it can always be ensured by a continuous strictly monotonic transformation such as the logistic function

$$X_t = \frac{1}{1+\exp(-Y_t)}, \tag{3.1}$$

where $\{Y_t\}$ is the original time series with unbounded support. No information is lost in (3.1) because $\{X_t\}$ is i.i.d. if and only if $\{Y_t\}$ is i.i.d., and $\mathcal{I}(j)$ is invariant under any continuous monotonic transformation of the data. Moreover, as we show below, the asymptotic mean and variance of $\hat{\mathcal{I}}_n(j)$ in (2.4) are also distribution-free and thus are invariant under any continuous monotonic transformation. These features make the entropy measure attractive in practice.

To see that one can often easily ensure that $g(\cdot)$ is bounded away from zero, let Y_t have cumulative density function (CDF) $\tilde{G}(\cdot)$ with density $\tilde{g}(\cdot)$ and let $F(\cdot)$ be a prespecified CDF with density $f(\cdot)$. Then $X_t \equiv F(Y_t)$ has support on $\mathbb{I}$ and the CDF of X_t is given by

$$G(x) = P[F(Y_t) \le x] = \tilde{G}[F^{-1}(x)], \quad x \in \mathbb{I}.$$

It follows straightforwardly that

$$g(x) \equiv \frac{dG(x)}{dx} = \frac{\tilde{g}[F^{-1}(x)]}{f[F^{-1}(x)]}, \quad x \in \mathbb{I}.$$

To ensure $\min_{x\in\mathbb{I}} g(x) \geq c$, it suffices that

$$f(y) \leq c^{-1}\tilde{g}(y), \quad y \in \mathbb{R}.$$

Even though $\tilde{g}(\cdot)$ is typically unknown, we often know or are willing to assume enough about $\tilde{g}(\cdot)$, in particular its tail behavior, to specify a transforming CDF $F(\cdot)$ that satisfies $f(\cdot) \leq c^{-1}\tilde{g}(\cdot)$, thereby ensuring $g(\cdot) \geq c$. Moreover, it seems plausible that one could allow $g(x) \to 0$ at the end points with a sufficiently slow rate, and our theory would continue to hold under strengthened conditions on the bandwidth $h \equiv h_n$ used in kernel density estimation. As the involved technicality would be quite complicated and would detract from our main goal, we do not pursue this here. However, we will use simulation to examine the consequence of allowing $g(x) \to 0$ smoothly at the end points of $\mathbb{I}$; this confirms our conjecture.[3]

3.1. *Boundary Effects and Jackknife Kernels*

We estimate probability densities via a standard (i.e., second-order) kernel:

ASSUMPTION A.2: *Assume $k : [-1, 1] \to \mathbb{R}^+$ is a symmetric bounded probability density function.*

An example of $k(\cdot)$ is the "quartic kernel," namely

$$k(u) = \frac{15}{16}(1 - u^2)^2\mathbb{1}(|u| \leq 1), \tag{3.2}$$

where $\mathbb{1}(\cdot)$ is the indicator function. Assumption A.2 implies $\int_{-1}^{1} k(u)\,du = 1$, $\int_{-1}^{1} uk(u)\,du = 0$, and $\int_{-1}^{1} u^2k(u)\,du < \infty$. This and Assumption A.1 ensure that the biases of kernel density estimators are $O(h^2)$ in the interior region $(h, 1 - h)$ of support $\mathbb{I}$. Because X_t has bounded support, kernel density estimators are subject to boundary effects near the two ends of $\mathbb{I}$. Such boundary effects arise because there is no symmetric coverage of the data for $k(\cdot)$ in the boundary regions $[0, h] \cup [1 - h, 1]$. Consequently, the kernel density estimators in the boundary regions are not asymptotically unbiased as $h \to 0$. See Härdle (1990, pp. 130–133) for more discussion. We find that although these boundary regions are of the size of h and thus are vanishing as $n \to \infty$,

[3]An alternative to avoid Assumption A.1 is to use a weight function with compact support. This allows for $\{X_t\}$ to have unbounded support and/or for $g(\cdot)$ to vanish, but because the information in the tails is lost, it does not deliver a consistent measure or test for serial dependence. Alternatively, one could consider a sequence of weight functions with increasing compact supports as $n \to \infty$, as discussed by Robinson (1991). This delivers a consistent measure or test for serial dependence. However, the measure is expected to be sensitive in practice to the delicate choice of moving trimming.

for $\hat{\mathcal{I}}_n(j)$ of (2.4), the cumulative effect of the kernel density estimators in these vanishing regions can overwhelm the behavior of the kernel density estimators in the interior region $(h, 1-h)$ in terms of mean squared error (MSE). In particular, the cumulative bias effect of the kernel density estimators in the boundary regions is $O_P(h)$ rather than the usual $O_P(h^2)$. This slows the convergence of $\hat{\mathcal{I}}_n(j)$ to $\mathcal{I}(j)$ and complicates the analysis. To avoid this, we use $k(\cdot)$ only for the interior region $(h, 1-h)$; for the boundary regions $[1, h] \cup [1-h, 1]$, we use the jackknife kernel

$$k_b(u) \equiv (1+r)\frac{k(u)}{\omega_k(0,b)} - \frac{r}{\alpha}\frac{k(u/\alpha)}{\omega_k(0,b/\alpha)}, \tag{3.3}$$

where $\omega_k(l,b) \equiv \int_{-b}^{1} u^l k(u)\,du$ for $l = 0, 1$, and $r \equiv r(b)$ and $\alpha \equiv \alpha(b)$ depend on $b \in \mathbb{I}$, an index whose value depends on where the density is estimated and which will be integrated out in computing our entropy estimator.[4] We set

$$r \equiv \frac{\omega_k(1,b)/\omega_k(0,b)}{\alpha\omega_k(1,b/\alpha)/\omega_k(0,b/\alpha) - \omega_k(1,b)/\omega_k(0,b)}.$$

As suggested by Rice (1984), we choose $\alpha = 2 - b$. Given $b \in \mathbb{I}$, the support of $k_b(\cdot)$ is $[-\alpha, \alpha]$, rather than $[-1, 1]$. Consequently, for any $b \in \mathbb{I}$,

$$\int_{-\alpha}^{\alpha b} k_b(u)\,du = \int_{-\alpha b}^{\alpha} k_b(u)\,du = 1,$$

$$\int_{-\alpha}^{\alpha b} u k_b(u)\,du = -\int_{-\alpha b}^{\alpha} u k_b(u)\,du = 0,$$

$$\int_{-\alpha}^{\alpha b} u^2 k_b(u)\,du = \int_{-\alpha b}^{\alpha} u^2 k_b(u)\,du > 0,$$

$$\int_{-\alpha}^{\alpha b} k_b^2(u)\,du = \int_{-\alpha b}^{\alpha} k_b^2(u)\,du > 0.$$

By using $k_b(\cdot)$, we ensure that the asymptotic bias of the kernel density estimators in the boundary regions will be of the same order as that in the interior region. Thus, the cumulative effect of the kernel density estimators in the boundary regions $[1, h] \cup [1-h, 1]$ is at most the same order as that of the kernel density estimators in the interior region $(h, 1-h)$ in terms of MSE. We emphasize that the boundary effects are not particular to our kernel density estimators. They arise whenever $\{X_t\}$ has bounded support, and they do

[4]We note that the jackknife formula given by Härdle (1990, pp. 131–132) has a typo.

not depend on whether $k(\cdot)$ has bounded support. For example, the kernels used in Granger and Lin (1994), Robinson (1991), and Skaug and Tjøstheim (1993a, 1996) also suffer from the boundary problem.

Define the kernel-based "leave-one-out" marginal and bivariate density estimators

$$\hat{g}_t(X_t) \equiv (n-1)^{-1} \sum_{s=1}^{n} K_h(X_t, X_s)\mathbb{1}(s \neq t), \tag{3.4}$$

$$\hat{f}_{jt}(Z_{jt}) \equiv (n_j-1)^{-1} \sum_{s=j+1}^{n} K_h^{(2)}(Z_{jt}, Z_{js})\mathbb{1}(s \neq t), \tag{3.5}$$

where (and throughout) $n_j \equiv n-j$, $K_h^{(2)}(Z_{jt}, Z_{js}) \equiv K_h(X_t, X_s)K_h(X_{t-j}, X_{s-j})$, and

$$K_h(x,y) \equiv \begin{cases} h^{-1}k_{(x/h)}\left(\frac{x-y}{h}\right), & \text{if } x \in [0,h], \\ h^{-1}k\left(\frac{x-y}{h}\right), & \text{if } x \in (h, 1-h), \\ h^{-1}k_{[(1-x)/h]}\left(\frac{x-y}{h}\right), & \text{if } x \in [1-h, 1]. \end{cases} \tag{3.6}$$

Note that $K_h(x,y) \neq K_h(y,x)$ despite $k(\cdot)$ being symmetric. Although we allow using different bandwidths for $\hat{f}_{jt}(\cdot)$ and $\hat{g}_t(\cdot)$, here we use the same bandwidth h. This makes the biases of $\hat{f}_{jt}(\cdot)$ and $\hat{g}_t(\cdot)$ cancel each other to a higher order under $\mathbb{H}_0$, leading to weaker conditions on h and faster convergence of $\hat{\mathcal{I}}_n(j)$. In particular, the usual practice of undersmoothing is not needed to remove the biases, and the optimal bandwidth for $\hat{f}_{jt}(\cdot)$ or $\hat{g}_t(\cdot)$ is allowed. We also do not require a higher-order kernel. Note that our theory applies to the leave-one-out estimators (3.4) and (3.5). If, following Granger and Lin (1994), leave-one-out estimators are not used, the asymptotic mean of $\hat{\mathcal{I}}_n(j)$ must be modified; see the discussion following Theorem 3.2.

3.2. *Heuristics on the Asymptotic Expansion of the Entropy Estimators*

To derive the limiting distribution of $\hat{\mathcal{I}}_n(j)$, we will Taylor-expand $\hat{\mathcal{I}}_n(j)$ up to a second order and show that the first two terms jointly determine the limit distribution of $\hat{\mathcal{I}}_n(j)$. The trick is to remove a nonzero mean from $\hat{\mathcal{I}}_n(j)$ and to exploit the consequences of the cancellation of the biases from $\hat{f}_{jt}(\cdot)$ and $\hat{g}_t(\cdot)$. Both $\hat{f}_{jt}(\cdot)$ and $\hat{g}_t(\cdot)$ affect the limiting distribution of $\hat{\mathcal{I}}_n(j)$, although they have different convergence rates.

Given the well-known difficulty of obtaining the limiting distribution for $\hat{\mathcal{I}}_n(j)$, we first provide some heuristics to gain insight into our approach.

We write

$$(3.7)\qquad \hat{\mathcal{I}}_n(j) = n_j^{-1} \sum_{t\in S_n(j)} \left\{ \ln\left[\frac{f_j(Z_{jt})}{g(X_t)g(X_{t-j})}\right] + \ln\left[\frac{\hat{f}_{jt}(Z_{jt})}{f_j(Z_{jt})}\right] - \ln\left[\frac{\hat{g}_t(X_t)}{g(X_t)}\right] - \ln\left[\frac{\hat{g}_{t-j}(X_{t-j})}{g(X_{t-j})}\right]\right\}$$
$$\equiv \hat{I}_{jn}(f_j, g\circ g) + \hat{I}_{jn}(\hat{f}_j, f_j) - \hat{I}_{1jn}(\hat{g}, g) - \hat{I}_{2jn}(\hat{g}, g), \quad \text{say.}$$

For the first term in (3.7), we have $\hat{I}_{jn}(f_j, g\circ g) = 0$ a.s. under $\mathbb{H}_0$. For the second term in (3.7), using the inequality $|\ln(1+u) - u + \frac{1}{2}u^2| \le |u|^3$ for $|u| < 1$, we obtain

$$(3.8)\qquad \hat{I}_{jn}(\hat{f}_j, f_j) = \hat{W}_1(j) + \tfrac{1}{2}\hat{W}_2(j) + \text{remainder}$$

under $\mathbb{H}_0$, where $\hat{W}_1(j) \equiv n_j^{-1}\sum_{t=j+1}^n [\hat{f}_{jt}(Z_{jt}) - f_j(Z_{jt})]/f_j(Z_{jt})$ and $\hat{W}_2(j) \equiv n_j^{-1}\sum_{t=j+1}^n \{[\hat{f}_{jt}(Z_{jt}) - f_j(Z_{jt})]/f_j(Z_{jt})\}^2$. The term $\hat{W}_1(j)$ is the first-order term of our Taylor series expansion. It can be approximated as

$$(3.9)\qquad \hat{W}_1(j) = \tfrac{1}{2}\hat{H}_{1n}(j) + \text{remainder},$$

where $\hat{H}_{1n}(j)$ is a second-order U-statistic arising from the interaction between the sampling variation of the estimator $\hat{f}_{jt}(\cdot)$ and the averaging operation over the sample in $\hat{W}_1(j)$.

The term $\hat{W}_2(j)$ is the second-order term of the Taylor series expansion. It can be approximated by its integrated analog,

$$(3.10)\qquad \hat{W}_2(j) = n_j^{-1}\sum_{t=j+1}^n \int \left[\frac{\hat{f}_{jt}(z) - f_j(z)}{f_j(z)}\right]^2 f_j(z)\,dz + \text{remainder}$$
$$= n_j^{-1}\sum_{t=j+1}^n \int E\left[\frac{\hat{f}_{jt}(z) - f_j(z)}{f_j(z)}\right]^2 f_j(z)\,dz$$
$$+ n_j^{-1}\sum_{t=j+1}^n \int \left\{\left[\frac{\hat{f}_{jt}(z) - f_j(z)}{f_j(z)}\right]^2 - E\left[\frac{\hat{f}_{jt}(z) - f_j(z)}{f_j(z)}\right]^2\right\} f_j(z)\,dz + \text{remainder}$$
$$= L_n(j) + \hat{H}_{2n}(j) + \text{remainder},$$

where $L_n(j)$ is the integrated weighted MSE of $\hat{f}_{jt}(\cdot)$, and $\hat{H}_{2n}(j)$ is a second term U-statistic.

Collecting (3.8)–(3.10) together and putting $\hat{H}_n(j) \equiv \hat{H}_{1n}(j) - \hat{H}_{2n}(j)$, we have

$$(3.11) \qquad 2\hat{I}_{jn}(\hat{f}_j, f_j) = -L_n(j) + \hat{H}_n(j) + \text{remainder}.$$

Both $\hat{H}_{1n}(j)$ and $\hat{H}_{2n}(j)$ are of the same order and jointly determine the limiting distribution of $\hat{I}_{jn}(\hat{f}_j, f)$. Under $\mathbb{H}_0$, $\hat{H}_n(j) = O_P(n_j^{-1}h^{-1})$ is dominated by $L_n(j) = O(n_j^{-1}h^{-2} + h^4)$. If $nh^4 \to \infty$, $nh^8 \to 0$, then $n_j^{1/2}\hat{I}_{jn}(\hat{f}_j, f_j) \xrightarrow{p} 0$. When $h \propto n_j^{-1/6}$, $\hat{I}_{jn}(\hat{f}_j, f_j) = O_P(n_j^{-2/3})$ attains its best convergence rate.

By similar (and simpler) reasoning, we can obtain

$$(3.12) \qquad 2\hat{I}_{ijn}(\hat{g}, g) = -l_n(j) + \hat{V}_{in}(j) + \text{remainder} \qquad (i = 1, 2)$$

under $\mathbb{H}_0$, where $l_n(j)$ is the integrated weighted MSE of $\hat{g}_t(\cdot)$, and $\hat{V}_{1n}(j)$ and $\hat{V}_{2n}(j)$ are second-order U-statistics. These are $O_P(n_j^{-1}h^{-1/2})$ under $\mathbb{H}_0$ and are dominated by $l_n(j)$. If $n_j h^2 \to \infty$, $n_j h^8 \to 0$, then $n^{1/2}\hat{I}_{ijn}(\hat{g}, g) \xrightarrow{p} 0$. When $h \propto n_j^{-1/5}$, then $\hat{I}_{ijn}(\hat{g}, g) = O_P(n_j^{-4/5})$ attains its optimal convergence rate. This is faster than the optimal rate $O_P(n_j^{-2/3})$ of the bivariate entropy estimator $\hat{I}_{jn}(\hat{f}_j, f_j)$.

Now, from (3.7), (3.11), and (3.12), we obtain

$$(3.13) \qquad 2\hat{\mathcal{I}}_n(j) = -n_j^{-1}d_n^0 + \hat{H}_n(j) + \text{remainder}$$

under $\mathbb{H}_0$, where the nonstochastic factor $d_n^0 = (A_n^0 - 1)^2$ and

$$(3.14) \qquad A_n^0 \equiv (h^{-1} - 2)\int_{-1}^{1} k^2(u)\,du + 2\int_0^1 \int_{-1}^{b} k_b^2(u)\,du\,db.$$

Thanks to the use of the same bandwidth h, the bias-squared terms from $\hat{f}_{jt}(\cdot)$ and $\hat{g}_t(\cdot)$ nicely cancel each other to a higher order within $\hat{\mathcal{I}}_n(j)$. As a consequence, d_n^0 depends only on the variance components of $\hat{f}_{jt}(\cdot)$ and $\hat{g}_t(\cdot)$, not on their biases. Undersmoothing is not needed to remove the biases, and the optimal bandwidth for $\hat{f}_{jt}(\cdot)$ or $\hat{g}_t(\cdot)$ is allowed. More importantly, the bias cancellation leads to a faster convergence rate for $\hat{\mathcal{I}}_n(j)$ than for $\hat{I}_{jn}(\hat{f}_j, f_j)$, so that we can have $\hat{\mathcal{I}}_n(j) = o_P(n_j^{-2/3})$. This occurs under and only under $\mathbb{H}_0$ and a class of local alternatives. On the other hand, we note that a correction term appears in the variance component A_n^0 in (3.14), due to the use of the jackknife kernel $k_b(\cdot)$ in the boundary regions $[0, h] \cup [1 - h, 1]$. This correction is not asymptotically negligible and it affects the asymptotic mean of $\hat{\mathcal{I}}_n(j)$.

3.3. *Asymptotic Normality of the Entropy Estimators*

Although the nonstochastic factor $n_j^{-1}d_n^0$ dominates the U-statistic $\hat{H}_n(j)$ in order of magnitude, it affects only the asymptotic mean of $\hat{\mathcal{I}}_n(j)$. The limiting distribution of $\hat{\mathcal{I}}_n(j)$, after centering, is determined by $\hat{H}_n(j)$. We now formally state the main result of this section.

THEOREM 3.1: *Suppose Assumptions* A.1 *and* A.2 *and* $\mathbb{H}_0$ *hold*, $nh^4/\ln n \to \infty$, *and* $nh^7 \to 0$.

(a) *Then* $2hn_j\hat{\mathcal{I}}_n(j) + hd_n^0 \xrightarrow{d} \mathrm{N}(0, \sigma^2)$ *for any lag order* $j = o(n)$, *where*

$$\sigma^2 \equiv 2\int_{-1}^{1}\int_{-1}^{1}\left[2k(u)k(u') - \int_{-1}^{1} k(u+v)k(v)\,dv \int_{-1}^{1} k(u'+v')k(v')\,dv'\right]^2 du\,du'.$$

(b) *Put* $\hat{\mathcal{I}}_n \equiv [2hn_1\hat{\mathcal{I}}_n(1) + hd_n^0, \ldots, 2hn_p\hat{\mathcal{I}}_n(p) + hd_n^0]'$, *where* $p \in \mathbb{N}^+$ *is fixed but may be arbitrarily large. Then* $\hat{\mathcal{I}}_n \xrightarrow{d} \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}_p)$, *where* $\mathbf{I}_p$ *is a* $p \times p$ *identity matrix.*

Thus, by adjusting for the mean hd_n^0 and by scaling $\hat{\mathcal{I}}_n(j)$ by n_jh, which tends to ∞ faster than $n_j^{1/2}$, we obtain a limiting normal distribution for $\hat{\mathcal{I}}_n(j)$. Note that we allow but do not require the lag order $j \to \infty$ as $n \to \infty$ for $\hat{\mathcal{I}}_n(j)$, and the condition on j is mild. Moreover, the condition on h is also not terribly restrictive. Both the optimal bandwidths ($h \propto n_j^{-1/6}$ and $n_j^{-1/5}$, respectively) for $\hat{f}_{jt}(\cdot)$ and $\hat{g}_t(\cdot)$ are allowed. Of course, these optimal bandwidths for the density estimators are not the same as the optimal bandwidth for the best convergence of the entropy estimator $\hat{\mathcal{I}}_n(j)$.

Theorem 3.1(b) shows that under $\mathbb{H}_0$ the finite-dimensional distribution of $\{2hn_j\hat{\mathcal{I}}_n(j), j \in \mathbb{N}^+\}$ is asymptotically multivariate normal for any set of distinct lag orders and that $\mathrm{cov}[hn_i\hat{\mathcal{I}}_n(i), hn_j\hat{\mathcal{I}}_n(j)] \to 0$ whenever $i \neq j$. This provides a basis for constructing a portmanteau test for $\mathbb{H}_0$; see Section 4.

A particularly appealing feature of $2hn_j\hat{\mathcal{I}}_n(j)$ is that its asymptotic mean hd_n^0 and variance σ^2 are distribution-free and thus are invariant under any continuous monotonic transformation. No estimation is required, because they are known given $k(\cdot)$ and h. This makes $\hat{\mathcal{I}}_n(j)$ rather attractive in practice. It greatly simplifies the calculation of confidence interval estimates. Moreover, to compute p-values for a smoothed bootstrap proposed in Section 4, one needs only to compare the unstandardized statistic $n_j\hat{\mathcal{I}}_n(j)$ with its bootstrap counterpart $n_j\hat{\mathcal{I}}_n^*(j)$. Numerically identical bootstrap p-values will be obtained as when one standardizes $n_j\hat{\mathcal{I}}_n(j)$ and $n_j\hat{\mathcal{I}}_n^*(j)$, and the appealing asymptotically pivotal character of our test is maintained. In contrast, other divergence measures, such as the squared L^2-norm

$$\hat{L}_n^2(j) \equiv n_j^{-1} \sum_{t=j+1}^{n} [\hat{f}_{jt}(Z_{jt}) - \hat{g}_t(X_t)\hat{g}_{t-j}(X_{t-j})]^2,$$

do not enjoy the invariance properties. The asymptotic mean and variance of $\hat{L}_n^2(j)$ are not distribution-free and must be estimated. Nevertheless, $\hat{L}_n^2(j)$ does allow $g(\cdot)$ to vanish to 0 smoothly. We note that our approach is applicable to $\hat{L}_n^2(j)$.

To the best of our knowledge, Theorem 3.1 is the first asymptotic distribution result for the smoothed entropy estimator $\hat{\mathcal{I}}_n(j)$. It can be used to obtain the limiting distribution of Granger and Lin's (1994) normalized entropy measure $\hat{\gamma}_n^2(j)$ of (2.3), which, as a shadow autocorrelation, is useful in identifying important lags and patterns of serial dependence in $\{X_t\}$.

THEOREM 3.2: *Suppose the conditions of Theorem* 3.1 *and* $\mathbb{H}_0$ *hold.* (a) *Then* $hn_j\hat{\gamma}_n^2(j) + hd_n^0 \xrightarrow{d} \mathrm{N}(0, \sigma^2)$ *for any lag order* $j = o(n)$. (b) *Put* $\hat{\boldsymbol{\gamma}}_n^2 \equiv [hn_1 \times \hat{\gamma}_n^2(1) + hd_n^0, \ldots, hn_p\hat{\gamma}_n^2(p) + hd_n^0]'$, *where* $p \in \mathbb{N}^+$ *is fixed but may be arbitrarily large. Then* $\hat{\boldsymbol{\gamma}}_n^2 \xrightarrow{d} \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}_p)$, *where* $\mathbf{I}_p$ *is a* $p \times p$ *identity matrix.*

Various attempts have been made in the literature to develop appropriate general dependence measures in nonlinear time-series contexts, but few have been as simple and informative as the sample autocorrelation function. Theorem 3.2 shows that the sample shadow autocorrelation function $\hat{\gamma}_n^2(j)$, after centering and scaling, is asymptotically $\mathrm{N}(0, 1)$ and is asymptotically independent across different lags. These properties are similar to those of the sample autocorrelation function. Thus, $\hat{\gamma}_n^2(j)$ can play an important role in nonlinear time-series analysis analogous to that of the sample autocorrelation in linear time-series analysis. We note that the asymptotic mean and variance of $\hat{\gamma}_n^2(j)$ under $\mathbb{H}_0$ are $-n_j^{-1}d_n^0 \propto n_j^{-1}h^{-2}$ and $(n_jh)^{-1}\sigma^2$. These rates differ from those obtained by Granger and Lin (1994) via simulation. Although the population measure $\gamma^2(j)$ in (2.2) is nonnegative, $\hat{\gamma}_n^2(j)$ may be negative under $\mathbb{H}_0$.

If, as in Granger and Lin (1994), leave-one-out kernel density estimators are not used, Theorems 3.1 and 3.2 still hold, except that the noncentrality factor d_n^0 must be replaced with

$$\tilde{d}_n^0 = d_n^0 - 2\{[h^{-1}k(0) - 1]^2 - 1\},$$

where the second term is contributed by the first-order term $\hat{W}_1(j)$ in (3.8).

4. HYPOTHESIS TESTING AND SMOOTHED BOOTSTRAP

4.1. *Testing the i.i.d. Hypothesis*

In an i.i.d. context, entropy has been used to test normality (Vasicek (1976)), uniform distribution (Dudewicz and van der Meulen (1981)), and goodness of fit (Gokhale (1983)). No limiting distributions are available for these tests, but they have been shown to have good power in simulations. Here, the ability of $\mathcal{I}(j)$ to capture all departures of $f_j(\cdot)$ from $g(\cdot)g(\cdot)$ makes it attractive for testing $\mathbb{H}_0$. Robinson (1991) has proposed an asymptotic $\mathrm{N}(0,1)$ test for $\mathbb{H}_0$ using $\hat{\mathcal{I}}_{n,\gamma}(j)$ of (2.5). We now apply our theory to construct some alternative $\mathrm{N}(0,1)$ tests for $\mathbb{H}_0$ using $\hat{\mathcal{I}}_n(j)$ of (2.4).

We first consider a test for $\mathbb{H}_0$ that is based on an individual lag j:

$$\mathcal{T}_n(j) \equiv \sigma^{-1}[2hn_j\hat{\mathcal{I}}_n(j) + hd_n^0], \quad j \ll n. \tag{4.1}$$

By Theorem 3.1, $\mathcal{T}_n(j) \xrightarrow{d} \mathrm{N}(0,1)$ under $\mathbb{H}_0$. We thus obtain an asymptotic $\mathrm{N}(0,1)$ test for $\mathbb{H}_0$ using $\hat{\mathcal{I}}_n(j)$. No sample splitting or tuning parameter is involved, and $\mathcal{T}_n(j)$ works even when X_t is i.i.d. uniform. As we show in Section 5, this test is asymptotically locally more efficient than Robinson's (1991) test.

Like Robinson's test, $\mathcal{T}_n(j)$ is a large sample test, and its finite sample level may differ substantially from the asymptotic level. The quality of the asymptotic approximation depends on the higher-order terms of the Taylor series expansion for $\hat{\mathcal{I}}_n(j)$ and the choice of the bandwidth h. Our analysis suggests that the asymptotic theory may not work well even for relatively large samples, because the asymptotically negligible higher-order terms in $\hat{\mathcal{I}}_n(j)$ are close in order of magnitude to the dominant U-statistic $\hat{H}_n(j)$ in (3.13), which determines the limiting distribution of $\mathcal{T}_n(j)$. The same problem was also documented by Skaug and Tjøstheim (1993a, 1996) for their tests.

Fortunately, because $\{X_t\}$ is i.i.d. under $\mathbb{H}_0$, the bootstrap is well suited and provides a simple way to obtain reasonable critical values for $\mathcal{T}_n(j)$. For the application of the bootstrap in econometrics, see (e.g.) Horowitz (2001). Interestingly, the naive bootstrap (i.e., resampling, with replacement, from the original sample $\mathcal{X} \equiv \{X_t\}_{t=1}^n$) does not deliver a consistent procedure for our test $\mathcal{T}_n(j)$, because it does not preserve the properties of the degenerate U-statistic $\hat{H}_n(j)$ in (3.13). Instead, we propose the following smoothed bootstrap procedure: (i) Draw a bootstrap sample $\mathcal{X}^* \equiv \{X_t^*\}_{t=1}^n$ from the smoothed kernel density

$$\hat{g}(x) \equiv n^{-1}\sum_{t=1}^{n} K_h(x, X_t), \quad x \in \mathbb{I}, \tag{4.2}$$

where $k(\cdot)$ and h are the same as those used in $\hat{\mathcal{I}}_n(j)$. (ii) Compute a bootstrap entropy statistic $\hat{\mathcal{I}}_n^*(j)$ in the same way as $\hat{\mathcal{I}}_n(j)$, with $\mathcal{X}^*$ replacing $\mathcal{X}$. The same $k(\cdot)$ and h are used in $\hat{\mathcal{I}}_n(j)$, $\hat{\mathcal{I}}_n^*(j)$, and $\hat{g}(\cdot)$. (iii) Repeat steps (i) and (ii) B times to obtain B bootstrap test statistics $\{\hat{\mathcal{I}}_{nl}^*(j)\}_{l=1}^B$. (iv) Compute the bootstrap p-value $p^* \equiv B^{-1}\sum_{l=1}^B \mathbb{1}[\hat{\mathcal{I}}_{nl}^*(j) > \hat{\mathcal{I}}_n(j)]$.

To obtain accurate bootstrap p-values, B must be sufficiently large. Note that we need only to compare the entropy estimator $\hat{\mathcal{I}}_n(j)$ with its bootstrap counterpart $\hat{\mathcal{I}}_n^*(j)$; there is no need to compute the asymptotic mean d_n^0 and the asymptotic variance σ^2 here. The obtained bootstrap p-values are numerically identical to those formed by comparing $\mathcal{T}_n(j)$ with $\mathcal{T}_n^*(j)$, and the asymptotically pivotal character of our test is maintained. All these follow from the asymptotic distribution-free property of d_n^0 and σ^2, and the use of the same bandwidth h in computing $\hat{\mathcal{I}}_n(j)$ and $\hat{\mathcal{I}}_n^*(j)$.

To show that the smoothed bootstrap procedure is consistent, we impose the following additional conditions on $k(\cdot)$.

ASSUMPTION A.3: *Suppose $k:[-1,1] \to \mathbb{R}^+$ is twice continuously differentiable on $[-1,1]$ with $k^{(d)}(-1) = k^{(d)}(1) = 0$ for $d = 0,1$ and $|k^{(2)}(u_1) - k^{(2)}(u_2)| \le C|u_1 - u_2|$ for $u_1, u_2 \in [-1,1]$.*

THEOREM 4.1: *Suppose Assumptions* A.1–A.3 *and $\mathbb{H}_0$ hold, $nh^5 = O(1)$, $nh^7 \ln^3 n \to 0$, and $j = o(n)$. Let $\mathcal{T}_n^*(j)$ be defined as $\mathcal{T}_n(j)$, with the bootstrap sample $\mathcal{X}^*$ defined above replacing the original sample $\mathcal{X}$ and with the same $k(\cdot)$ and h used in $\mathcal{T}_n(j)$, $\mathcal{T}_n^*(j)$, and $\hat{g}(\cdot)$. Then $\mathcal{T}_n^*(j) \xrightarrow{d}$* N(0, 1) *conditional on $\mathcal{X}$.*

Theorem 4.1 shows that the smoothed bootstrap provides an asymptotically valid approximation to the limit N(0, 1) distribution of $\mathcal{T}_n(j)$ under $\mathbb{H}_0$. Note that Theorem 4.1 implies that $\mathcal{T}_n^*(j) \xrightarrow{d} \mathrm{N}(0,1)$ unconditionally. However, it does not indicate the degree of improvement of the smoothed bootstrap upon the first-order asymptotic approximation. As $\mathcal{T}_n(j)$ is asymptotically pivotal, it is plausible that the smoothed bootstrap can achieve reasonably accurate levels for $\mathcal{T}_n(j)$. We have suggested using the same kernel $k(\cdot)$ and the same bandwidth h in computing $\mathcal{T}_n(j)$, $\mathcal{T}_n^*(j)$, and $\hat{g}(\cdot)$. This is not necessary, but it is expected to give a better finite sample approximation. We will examine the performance of the smoothed bootstrap in our simulation and we find that the smoothed bootstrap provides reasonable levels for our test in small samples.

Next, we consider the asymptotic behavior of $\mathcal{T}_n(j)$ under a global alternative to $\mathbb{H}_0$.

ASSUMPTION A.4: *Assume $\{X_t\}$ is a strictly stationary α-mixing process with mixing coefficient $\alpha(j) \le Cj^{-\nu}$ for some $\nu > 2$. For each $j \in \mathbb{N}^+$, the joint probability density $f_j(\cdot)$ of Z_{jt} has support $\mathbb{I}^2$, is bounded away from* 0, *and is twice continuously differentiable on $\mathbb{I}^2$.*

THEOREM 4.2: *Suppose Assumptions* A.1–A.4 *hold*, $nh^5 = O(1)$, $nh^7 \times \ln^3 n \to 0$, *and* $j = o(n)$. *Then* $P[\mathcal{T}_n(j) > C_n(j)] \to 1$ *for any sequence of constants* $\{C_n(j) = o(n_j h)\}$ *and* $P[\mathcal{T}_n(j) > \mathcal{T}_n^*(j)] \to 1$, *provided* $f_j(\cdot) \neq g(\cdot)g(\cdot)$.

Thus, the test based on $\mathcal{T}_n(j)$ is consistent against every alternative for which $f_j(\cdot) \neq g(\cdot)g(\cdot)$, no matter whether an asymptotic or bootstrap critical value is used. Theorem 4.2 implies $\mathcal{T}_n(j) \to +\infty$ with probability approaching 1 under any alternative with $f_j(\cdot) \neq g(\cdot)g(\cdot)$. Therefore, upper-tailed critical values are appropriate. The $N(0, 1)$ critical value at the 5% level, for example, is 1.645. We emphasize that Robinson (1991) established consistency of his test for a fixed lag order j under a milder condition of stationary ergodicity, which allows for processes with longer memory. In contrast, we impose an α-mixing condition, which implies ergodicity and rules out long memory processes. The mixing condition is convenient for nonlinear time-series analysis. We use it here because we allow the lag order $j \to \infty$ as $n \to \infty$.

The $\mathcal{T}_n(j)$ test is informative in revealing information about the lag(s) at which there exists significant serial dependence. However, for testing $\mathbb{H}_0$, it is possible that two different lag orders may give conflicting conclusions (see the empirical application in Section 7). It is thus desirable to have a portmanteau test that uses multiple lags. For this purpose, we consider

$$\mathcal{Q}_n(p) \equiv \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \mathcal{T}_n(j), \quad p \in \mathbb{N}^+. \tag{4.3}$$

For simplicity, we consider a fixed lag truncation number $p \in \mathbb{N}^+$. It is possible to allow $p \to \infty$ as $n \to \infty$ with a different weighting for each lag j, but we do not consider this here.

THEOREM 4.3: (i) *Suppose the conditions of Theorem* 4.1 *and* $\mathbb{H}_0$ *hold. Let* $\mathcal{Q}_n^*(p)$ *be defined as* $\mathcal{Q}_n(p)$, *with the bootstrap sample* $\mathcal{X}^*$ *replacing the original sample* $\mathcal{X}$ *but with the same* $k(\cdot)$ *and* h *used in* $\mathcal{Q}_n(p)$, $\mathcal{Q}_n^*(p)$, *and* $\hat{g}(\cdot)$. *Then for any fixed* $p \in \mathbb{N}^+$, $\mathcal{Q}_n(p) \xrightarrow{d} N(0, 1)$ *and* $\mathcal{Q}_n^*(p) \xrightarrow{d} N(0, 1)$ *conditional on* $\mathcal{X}$. (ii) *Suppose the conditions of Theorem* 4.2 *hold. Then* $P[\mathcal{Q}_n(p) > C_n] \to 1$ *for any sequence of constants* $\{C_n = o(nh)\}$ *and* $P[\mathcal{Q}_n(p) > \mathcal{Q}_n^*(p)] \to 1$, *provided* $f_j(\cdot) \neq g(\cdot)g(\cdot)$ *for some* $j \in \{1, 2, \ldots, p\}$.

Intuitively, because $\mathcal{T}_n(j) \xrightarrow{d} N(0, 1)$ and $\text{cov}[\mathcal{T}_n(i), \mathcal{T}_n(j)] \to 0$ for $i \neq j$ under $\mathbb{H}_0$, $\{\mathcal{T}_n(j)\}_{j=1}^p$ is a sequence of asymptotically i.i.d. $N(0, 1)$ random variables. Therefore, $\mathcal{Q}_n(p) \xrightarrow{d} N(0, 1)$ under $\mathbb{H}_0$ by the central limit theorem. Like many popular tests in time-series analysis, the power of $\mathcal{Q}_n(p)$ depends on the choice of p. However, the dependence of the power of $\mathcal{Q}_n(p)$ on p is expected to be less sensitive than the dependence of the power of $\mathcal{T}_n(j)$ on lag order j. One may view $\mathcal{Q}_n(p)$ as a portmanteau test for serial dependence that can capture linear and nonlinear dependence in $\{X_t\}$. Note that like the $\mathcal{T}_n(j)$ test, we can simply compare $\sum_{j=1}^{p} n_j\hat{\mathcal{I}}_n(j)$ with $\sum_{j=1}^{p} n_j\hat{\mathcal{I}}_n^*(j)$ in computing bootstrap p-values of $\mathcal{Q}_n(p)$.

Alternatively, it is possible to construct an asymptotic χ_p^2 test statistic $\kappa_n(p) \equiv \sum_{j=1}^{p} \mathcal{T}_n^2(j)$. However, following the asymptotic local power analysis in Section 5, we find that this test is asymptotically less efficient than $\mathcal{Q}_n(p)$, because $\kappa_n(p)$ does not exploit the one-sided nature (i.e., $\mathcal{T}_n(j) \to +\infty$) of individual test statistics $\mathcal{T}_n(j)$ under the alternative to $\mathbb{H}_0$.

4.2. *Testing for the i.i.d.* U[0, 1] *Hypothesis*

Diebold et al. (1998) recently proposed a graphical method to evaluate density forecasts, which is useful, for example, in financial risk management, options pricing, and macroeconomic policy control. They showed that if a series of one-step-ahead density forecasts correctly specifies the dynamic structure of the data generating process (though not necessarily correctly specifying the conditional density), then the series of probability integral transforms with respect to forecast densities is i.i.d. Thus, $\mathcal{T}_n(j)$ can be used to assess whether a sequence of density forecasts correctly specifies the dynamic structure. Diebold et al. (1998) further showed that if a series of one-step-ahead density forecasts coincides with the conditional density generating the data, then the series of probability integral transforms with respect to forecast densities is i.i.d. U[0, 1]. In this case, the density forecasts are optimal in terms of minimizing any expected loss criterion. Thus, to assess whether a series of density forecasts coincides with the corresponding series of true conditional densities, one needs only to assess whether the series of probability integral transforms is i.i.d. U[0, 1]. The $\mathcal{T}_n(j)$ test is not suitable for this hypothesis, because it does not incorporate the information about the U[0, 1] distribution. However, our theory yields a suitable test for this joint hypothesis. When an individual lag j is used, we define the test statistic

$$(4.4)\qquad \mathcal{T}_n^U(j) \equiv \sigma^{-1}\left[2h \sum_{t\in S_n(j)} \ln \hat{f}_{jt}(Z_{jt}) + h[(A_n^0)^2 - 1]\right],$$

where A_n^0 is in (3.14) and $S_n(j) \equiv \{t \in \mathbb{N} : j < t \leq n, \hat{f}_{jt}(Z_{jt}) > 0\}$. Note that we need not estimate the marginal density $g(\cdot)$, which is known under the i.i.d. U[0, 1] hypothesis. Because of this, the centering constant in $T_n^U(j)$ differs a bit from that for $\mathcal{T}_n(j)$. When multiple lags are considered, we can define

$$(4.5)\qquad \mathcal{Q}_n^U(p) \equiv \frac{1}{\sqrt{p}} \sum_{j=1}^{p} \mathcal{T}_n^U(j), \quad p \in \mathbb{N}^+.$$

We can also consider a bootstrap procedure for $\mathcal{T}_n^U(j)$ and $\mathcal{Q}_n^U(j)$. Here, we have to generate the bootstrap sample $\mathcal{X}^*$ from the U[0, 1] distribution.

Neither the naive bootstrap nor the smoothed bootstrap in (4.2) will deliver a consistent procedure for $\mathcal{T}_n^U(p)$ and $\mathcal{Q}_n^U(p)$. A much simpler procedure suffices; it is specified below. Like the $\mathcal{T}_n(j)$ and $\mathcal{Q}_n(p)$ tests, we can simply compare $\sum_{t\in S_n(j)} \ln \hat{f}_{jt}(Z_{jt})$ and $\sum_{j=1}^{p}\sum_{t\in S_n(j)} \ln \hat{f}_{jt}(Z_{jt})$ with their bootstrap counterparts in computing the bootstrap p-values of $\mathcal{T}_n^U(j)$ and $\mathcal{Q}_n^U(j)$, respectively.

THEOREM 4.4: *Suppose Assumptions* A.2 *and* A.4, *and the hypothesis* $\mathbb{H}_0^U$ *that* $\{X_t\} \sim$ *i.i.d.* U[0, 1] *hold and* $nh^5 = O(1)$, $h \to 0$. *Let* $\mathcal{T}_n^U(j)^*$ *and* $\mathcal{Q}_n^U(p)^*$ *be defined as* $\mathcal{T}_n^U(j)$ *and* $\mathcal{Q}_n^U(p)$, *respectively, with the bootstrap sample* $\mathcal{X}^*$ *replacing the original sample* $\mathcal{X}$, *where* $\mathcal{X}^*$ *is an i.i.d. sample drawn from the* U[0, 1] *distribution. The same* $k(\cdot)$ *and* h *are used in all test statistics. Then* (i) *for any lag order* $j = o(n)$, $\mathcal{T}_n^U(j) \xrightarrow{d} \mathrm{N}(0,1)$ *and* $\mathcal{T}_n^U(j)^* \xrightarrow{d} \mathrm{N}(0,1)$ *conditional on* $\mathcal{X}$; (ii) *for any fixed* $p \in \mathbb{N}^+$, $\mathcal{Q}_n^U(p) \xrightarrow{d} \mathrm{N}(0,1)$ *and* $\mathcal{Q}_n^U(p)^* \xrightarrow{d} \mathrm{N}(0,1)$ *conditional on* $\mathcal{X}$.

Although we do not state a formal result, the test based on $\mathcal{T}_n^U(j)$ can detect all pairwise dependence deviations from i.i.d. and all deviations from the U[0, 1] distribution. As pointed out by Diebold et al. (1998), there was no suitable statistical test for $\mathbb{H}_0^U$ in the previous literature, which is important for evaluating whether density forecasts are optimal.[5] Our $\mathcal{T}_n^U(j)$ and $\mathcal{Q}_n^U(p)$ are suitable tests for such a joint hypothesis. Note that there is no corresponding Robinson-type test for $\mathbb{H}_0^U$, because the sample-splitting device fails in this case.

5. ASYMPTOTIC LOCAL POWER

Robinson's (1991) test statistic and our test statistics are asymptotically $\mathrm{N}(0,1)$ under $\mathbb{H}_0$, and both are consistent against all j-dependent processes satisfying Assumption A.4. To examine their relative merits, we now study their asymptotic local power. As Tjøstheim (1996) pointed out, asymptotic local power analysis is rather difficult in nonparametric testing for serial dependence. For simplicity and tractability, we consider a class of locally j-dependent processes for which there exists serial dependence at lag j only, but j may grow to infinity as $n \to \infty$. We thus suppose that the joint probability density of Z_{jt} is given by

$$(5.1) \qquad \mathbb{H}_{jn}(a_n): f_j(z) = g(x)g(y)[1 + a_n q_j(z) + r_{jn}(z)], \quad z \equiv (x, y)' \in \mathbb{I}^2,$$

[5]In the time-series context, model-based density forecasts usually involve some model parameter estimators based on in-sample observations. Because the forecast setup is substantially different from our present framework, here we follow the usual practice in the forecast literature and do not consider the impact of parameter estimation uncertainty on our tests. However, we expect that parameter estimation uncertainty has no impact on the limiting distribution of $\mathcal{T}_n^U(j)$ and $\mathcal{Q}_n^U(p)$, because model parameter estimators typically converge to the true parameter values at a parametric rate that is faster than our nonparametric density estimators.

where $q_j : \mathbb{I}^2 \to \mathbb{R}$ is a function characterizing deviations from $\mathbb{H}_0$, $r_{jn}(\cdot)$ is a remainder term arising from the Taylor series expansion of $f_j(\cdot) \equiv f_{jn}(\cdot)$, and the constant $a_n \to 0$ governs the rate at which the local alternative $\mathbb{H}_{jn}(a_n)$ converges to the null hypothesis $\mathbb{H}_0$.

ASSUMPTION A.5: (a) $1 + a_n q_j(z) + r_{jn}(z) \geq 0$ *for all* $z \in \mathbb{I}^2$, *all* n, $j \in \mathbb{N}$; (b) $\int_{\mathbb{I}^2} q_j(z) g(x) g(y)\, dz = 0$ *and* $\int_{\mathbb{I}^2} r_{jn}(z) g(x) g(y)\, dz = 0$ *for all* n, $j \in \mathbb{N}^+$; (c) $g(\cdot)$ *and* $q_j(\cdot)$ *are twice continuously differentiable on* $\mathbb{I}$ *and* $\mathbb{I}^2$, *respectively*, $|g^{(2)}(x_1) - g^{(2)}(x_2)| \leq C|x_1 - x_2|$ *for any* x_1, $x_2 \in \mathbb{I}$ *and some* $C \in (0, \infty)$; (d) $\int_{\mathbb{I}^2} |q_j(z)|^3 g(x) g(y)\, dz \leq C$ *and* $\int_{\mathbb{I}^2} |r_{jn}(z)|^3 g(x) g(y)\, dz = o(a_n^2)$ *uniformly in* $j \in \mathbb{N}^+$.

Assumptions A.5(a) and (b) ensure that $f_j(\cdot)$ is a valid bivariate probability density for all n, $j \in \mathbb{N}^+$. Assumption A.5(d) ensures that the remainder term $r_{jn}(\cdot)$ has no impact on the limiting distribution of our tests. Note that the marginal density $g_n(\cdot)$ of X_t may depend on n under $\mathbb{H}_{jn}(a_n)$ and may not coincide with $g(\cdot)$, the marginal density of X_t under $\mathbb{H}_0$.

Two examples of $\mathbb{H}_{1n}(a_n)$ are an MA(1) process

$$X_t = a_n \varepsilon_{t-1} + \varepsilon_t \tag{5.2}$$

and an ARCH(1)-type process

$$X_t = \varepsilon_t \sqrt{1 + a_n \varepsilon_{t-1}^2}, \tag{5.3}$$

where $\{\varepsilon_t\}$ is i.i.d. $(0, \sigma_\varepsilon^2)$ with marginal density bounded away from 0. We have $q_1(z) = xy$ for (5.2) and $q_1(z) = (x^2 - \sigma_\varepsilon^2)(y^2 - \sigma_\varepsilon^2)$ for (5.3).

For brevity, we consider only $\mathcal{T}_n(j)$. The conclusions for the local power of $\mathcal{Q}_n(p)$ are similar.

THEOREM 5.1: *Suppose Assumptions* A.2 *and* A.5 *hold*, $nh^4/\ln n \to \infty$, $nh^7 \to 0$, *and* $j = o(n)$. *Then* $\mathcal{T}_n(j) - \mu_j \xrightarrow{d} \mathrm{N}(0,1)$ *under* $\mathbb{H}_{jn}(n^{-1/2}h^{-1/2})$, *where* $\mu_j \equiv \sigma^{-1} \int_{\mathbb{I}^2} q_j^2(z) g(x) g(y)\, dz$.

Thus, $\mathcal{T}_n(j)$ has nontrivial power against $\mathbb{H}_{jn}(n^{-1/2}h^{-1/2})$ whenever $q_j(\cdot) \neq 0$. The rate $a_n = n^{-1/2}h^{-1/2}$ is slower than the parametric rate $n^{-1/2}$ as $h \to 0$, but is faster than $n^{-1/4}$, because $nh^3 \to \infty$. For example, when $h \propto n^{-1/6}$ (the optimal rate for $\hat{f}_{jt}(\cdot)$) is used, we have $n^{-1/2}h^{-1/2} \propto n^{-5/12}$, which is slightly slower than $n^{-1/2}$, but is faster than $n^{-1/4}$. In fact, the admissible rates for a_n could be further improved toward the parametric rate $n^{-1/2}$ by relaxing the conditions on bandwidth h, which could be achieved by using a higher-order kernel. The rate a_n could thus be made arbitrarily close to $n^{-1/2}$, but it is always slower than $n^{-1/2}$, due to smoothing. In practice we need to choose h to balance the level and power in finite samples. We will use a data-driven method to choose h in our simulation and empirical application.

Robinson (1991) proposed a test statistic that, when $j=1$, has the form

$$(5.4)\qquad \mathcal{R}_n(j) \equiv \left[\frac{n_j}{2(j+1)\gamma^2 \hat{V}_\delta}\right]^{1/2} \hat{\mathcal{I}}_{n,\gamma}(j),$$

where $\hat{\mathcal{I}}_{n,\gamma}(j)$ is the modified entropy estimator (2.5) and $\hat{V}_\delta \equiv n^{-1}\sum_{t\in S_n} C_t(\delta)\times \ln^2 \hat{g}_t(X_t) - [n^{-1}\sum_{t\in S_n} C_t(\delta)\ln \hat{g}_t(X_t)]^2$, $S_n \equiv \{t\in\mathbb{N}: 1\le t\le n, \hat{g}_t(X_t)>0\}$, and $C_t(\delta)$ is as in (2.5), with $\delta\in[0,1)$.[6]

Robinson (1991) required using different bandwidths $h_1 \neq h_2$ for $\hat{f}_{jt}(\cdot)$ and $\hat{g}_t(\cdot)$ and a higher-order kernel $k(\cdot)$ for $\hat{f}_{jt}(\cdot)$.[7] Undersmoothing is needed to remove the effect of the bias terms of $\hat{f}_{jt}(\cdot)$ and $\hat{g}_t(\cdot)$ on the limit distribution of $\mathcal{R}_n(j)$. Given suitable conditions, Robinson (1991, Theorem 3.2) showed that $\mathcal{R}_n(j) \xrightarrow{d} N(0,1)$ under $\mathbb{H}_0$. In fact, our analysis in Section 3 suggests that when the same bandwidth h and the same kernel are used for $\hat{f}_{jt}(\cdot)$ and $\hat{g}_t(\cdot)$, undersmoothing and higher-order kernels can be avoided in (5.4), thanks to the cancellation of the bias-squared terms of $\hat{f}_{jt}(\cdot)$ and $\hat{g}_t(\cdot)$.

Robinson (1991) did not examine the asymptotic local power of his test. Below, we show that the $\mathcal{R}_n(j)$ test can only detect $\mathbb{H}_{jn}(n^{-1/4})$. To see this, we note that $\mathcal{R}_n(j)$ has nontrivial power if its limit noncentrality parameter under $\mathbb{H}_{jn}(a_n)$ obeys

$$(5.5)\qquad \lim_{n\to\infty} \frac{1}{\sqrt{2(j+1)\gamma^2 V}} n_j^{-1/2} \sum_{t=j+1}^{n} C_t(\gamma) E \ln\left[\frac{f_{jn}(Z_{jt})}{g_n(X_t)g_n(X_{t-j})}\right]$$
$$= \frac{1}{2\sqrt{2(j+1)\gamma^2 V}} \sigma\mu_j \lim_{n\to\infty}(n^{1/2}a_n^2) > 0,$$

where $V \equiv \operatorname{var}[\ln g(X_t)]$, and the equality follows from the inequality $|\ln(1+u) - u + \frac{1}{2}u^2| \le |u|^3$ for $|u|<1$ and Assumption A.5(d). Note that (5.5) holds if and only if $\lim_{n\to\infty} n^{1/2}a_n^2 = c > 0$ or $a_n \propto n^{-1/4}$. Thus, Robinson's test can detect $\mathbb{H}_{jn}(n^{-1/4})$ only, which converges to $\mathbb{H}_0$ more slowly than $\mathbb{H}_{jn}(n^{-1/2}h^{-1/2})$. In other words, $\mathcal{T}_n(j)$ is asymptotically locally more powerful than Robinson's (1991) test. We emphasize, however, that Robinson (1991) pointed out that

[6]A factor of 2 is missing in Robinson's (1991) formula (2.21). We note that when multiple lags are considered, Robinson (1991) considered an alternative entropy test statistic that is based on the multivariate density estimator $\hat{f}(X_t, X_{t-1}, \dots, X_{t-p})$. Our pairwise dependence test statistic $\mathcal{Q}_n(p)$ avoids the curse of dimensionality difficulty associated with $\hat{f}(X_t, X_{t-1}, \dots, X_{t-p})$ when p is large.

[7]For the marginal density estimator $\hat{g}_t(\cdot)$, Robinson (1991) used a second-order kernel $m(\cdot)$.

one should not expect tests using his weighting device to generally dominate all rival tests.

Skaug and Tjøstheim (1996) proposed a test for $\mathbb{H}_0$, using the functional

$$(5.6)\qquad \mathcal{J}_n(j) \equiv \hat{S}^{-1} n_j^{-1/2} \sum_{t=j+1}^{n} [\hat{f}_{jt}(Z_{jt}) - \hat{g}_t(X_t)\hat{g}_{t-j}(X_{t-j})]w(Z_{jt}),$$

where $\hat{S}^2$ is a consistent asymptotic variance estimator and $w(\cdot): \mathbb{I}^2 \to \mathbb{R}^+$ is a weight function used to remove extreme observations at the tail.[8] Under suitable conditions that allow $\{X_t\}$ to have unbounded support and $g(\cdot)$ to vanish to 0 smoothly, Skaug and Tjøstheim (1996) showed that $\mathcal{J}_n(j) \xrightarrow{d} \mathrm{N}(0,1)$ under $\mathbb{H}_0$ and suggested using upper-tailed critical values. As Skaug and Tjøstheim (1996) pointed out, this test is not consistent against a global fixed alternative to $\mathbb{H}_0$, because $n_j^{-1/2}\mathcal{J}_n(j) \xrightarrow{p} S^{-1}\int [f_j(z) - g(x)g(y)]w(z)f_j(z)\,dz$, which may be 0 or negative even if $f_j(\cdot) \neq g(\cdot)g(\cdot)$, as nicely illustrated by an example in Skaug and Tjøstheim (1993a, p. 209). However, this test has been shown to beat Robinson's (1991) test and has excellent power in finite samples against a variety of alternatives (Skaug and Tjøstheim (1993a, 1996)). It can be shown that it has nontrivial power against $\mathbb{H}_{jn}(a_n)$ with suitable rate a_n because its limit noncentrality parameter under $\mathbb{H}_{jn}(a_n)$ obeys

$$\begin{aligned}
&\lim_{n\to\infty} n_j^{-1/2} \sum_{t=1}^{n} E\{[f_{jn}(Z_{jt}) - g_n(X_t)g_n(X_{t-j})]w(Z_{jt})\} \\
&\quad = \int_{\mathbb{I}^2} q_j(z)g^2(x)g^2(y)w(z)\,dz \lim_{n\to\infty}(n^{1/2}a_n) \\
&\qquad + \int_{\mathbb{I}^2} q_j^2(z)g^2(x)g^2(y)w(z)\,dz \lim_{n\to\infty}(n^{1/2}a_n^2)
\end{aligned}$$

given Assumption A.5. Suppose $\int_{\mathbb{I}^2} q_j(z)g^2(x)g^2(y)w(z)\,dz > 0$. Then the $\mathcal{J}_n(j)$ test has nontrivial power against $\mathbb{H}_{jn}(n^{-1/2})$, thus dominating both the $\mathcal{T}_n(j)$ and $\mathcal{R}_n(j)$ tests. If, however, $\int_{\mathbb{I}^2} q_j(z)g^2(x)g^2(y)w(z)\,dz = 0$, then it can detect $\mathbb{H}_{jn}(n^{-1/4})$, the same rate as Robinson's (1991) test $\mathcal{R}_n(j)$. See a similar analysis in Skaug and Tjøstheim (1996, p. 366) for this second case. Of course, the relative finite sample performance may tell a different story, as our simulation study and empirical application below demonstrate.

[8]The weight function $w(\cdot)$ is generally not needed for the $\mathcal{J}_n(j)$ test. When $g(\cdot)$ is uniform, however, $w(\cdot)$ is needed to prevent degeneracy, because otherwise the asymptotic variance of $\mathcal{J}_n(j)$ would vanish to 0.

6. MONTE CARLO EVIDENCE

We now compare the finite sample performance of tests based on $\mathcal{T}_n(j)$, $\mathcal{R}_n(j)$, and $\mathcal{J}_n(j)$ under the data generating processes (DGPs)

$$
\begin{array}{ll}
\text{DGP 0 I.I.D.:} & X_t = \varepsilon_t, \\
\text{DGP 1 AR(1):} & X_t = 0.3X_{t-1} + \varepsilon_t, \\
\text{DGP 2 ARCH(1):} & X_t = \varepsilon_t h_t^{1/2},\ h_t = 1 + 0.8X_{t-1}^2, \\
\text{DGP 3 Threshold GARCH(1, 1):} & \begin{cases} X_t = \varepsilon_t h_t^{1/2}, \\ h_t = 0.25 + 0.6h_{t-1} \\ \qquad + 0.5X_{t-1}^2 \mathbb{1}(\varepsilon_t < 0) \\ \qquad + 0.2X_{t-1}^2 \mathbb{1}(\varepsilon_t \geq 0), \end{cases} \\
\text{DGP 4 Bilinear AR(1):} & X_t = 0.8X_{t-1}\varepsilon_{t-1} + \varepsilon_t, \\
\text{DGP 5 Nonlinear MA(1):} & X_t = 0.8\varepsilon_{t-1}^2 + \varepsilon_t, \\
\text{DGP 6 Threshold AR(1):} & X_t = \begin{cases} -0.5X_{t-1} + \varepsilon_t, & \text{if } X_{t-1} \leq 1, \\ 0.4X_{t-1} + \varepsilon_t, & \text{if } X_{t-1} > 1, \end{cases} \\
\text{DGP 7 Fractional AR(1):} & X_t = 0.8|X_{t-1}|^{0.5} + \varepsilon_t, \\
\text{DGP 8 Sign AR(1):} & X_t = \operatorname{sign}(X_{t-1}) + 0.43\varepsilon_t,
\end{array}
$$

where we consider two innovation processes: (i) $\{\varepsilon_t\} \sim$ i.i.d. $N(0, 1)$ and (ii) $\{\varepsilon_t\} \sim$ i.i.d. lognormal$(0, 1)$, normalized to have zero mean and unit variance. For all DGPs, the condition that the marginal density $g(\cdot) \geq c > 0$ in Assumption A.1 fails even after the logistic transformation in (3.1). This allows us to examine the consequence of such violations. DGP 0 allows us to examine the level of the tests. DGPs 1–8 cover a variety of commonly used linear and nonlinear time-series processes in the literature. We consider two sample sizes: $n = 100, 200$. For each DGP, we first generate $n + 100$ observations and then discard the first 100 to mitigate the impact of initial values. A preliminary experiment shows that the asymptotic theory provides rather poor approximation for the levels of all tests for sample sizes we consider. To ensure a fair comparison of all tests, throughout we use the smoothed bootstrap proposed in Section 4.[9]

To compute $\mathcal{T}_n(j)$, $\mathcal{R}_n(j)$, and $\mathcal{J}_n(j)$, we first transform the data via the logistic function in (3.1) and then rescale them so that the sample $\mathcal{X} \equiv \{X_t\}_{t=1}^n$ has support on $\mathbb{I}$. This induces dependence of each X_t on the whole sample $\mathcal{X}$, but this is a higher-order effect and is expected to be negligible asymptotically. The entropy $\mathcal{I}(j)$ is invariant under such scaling. We need to choose

[9]Although not formally justified, the naive bootstrap procedure is expected to be applicable to the $\mathcal{R}_n(j)$ and $\mathcal{J}_n(j)$ tests, because their distributions are not based on degenerate U-statistics. However, we use the smoothed bootstrap procedure for $\mathcal{T}_n(j)$, $\mathcal{R}_n(j)$, and $\mathcal{J}_n(j)$ for comparability.

tuning parameters (γ, δ) for $\mathcal{R}_n(j)$, and a weighting function $w(\cdot)$ for $\mathcal{J}_n(j)$. We set $(\gamma, \delta) = (0.5, 0)$ and choose the Beta(2, 2) density function for $w(\cdot)$ given that the transformed series has support $\mathbb{I}$. For all tests, we use the quartic kernel (3.2) and the same bandwidth h. We use the following data-driven method to choose h, which is more objective than an arbitrary choice or a simple rule-of-thumb.[10] By (3.12), an asymptotically optimal bandwidth that yields the optimal convergence rate for the Kullback–Leibler information criterion for $\hat{g}_t(\cdot)$ is

$$(6.1)\qquad h^o = \left\{\left[\int_{-1}^{1} k^2(u)\,du\right]^{-1}\left[\int_{-1}^{1} u^2 k(u)\,du\right]^2 \times \int_0^1 \left[\frac{g^{(2)}(x)}{g(x)}\right]^2 g(x)\,dx\right\}^{-1/5} n^{-1/5},$$

where $k(\cdot)$ is the kernel used to estimate $g(\cdot)$. For the quartic kernel (3.2), we have

$$(6.2)\qquad h_Q^o = 2.0236\left\{\int_0^1 \left[\frac{g^{(2)}(x)}{g(x)}\right]^2 g(x)\,dx\right\}^{-1/5} n^{-1/5}.$$

This optimal bandwidth is infeasible because it involves the unknown $g(\cdot)$ and $g^{(2)}(\cdot)$. We therefore use a "plug-in" method to obtain a data-driven bandwidth

$$(6.3)\qquad \hat{h}_Q = 2.0236\left\{n^{-1}\sum_{t\in\tilde{S}_n}\left[\frac{\tilde{g}_{nt}^{(2)}(X_t)}{\tilde{g}_{nt}(X_t)}\right]^2\right\}^{-1/5} n^{-1/5},$$

where $\tilde{g}_{nt}(\cdot)$ and $\tilde{g}_{nt}^{(2)}(\cdot)$ are preliminary estimators for $g(\cdot)$ and $g^{(2)}(\cdot)$, $\tilde{S}_n \equiv \{t \in \mathbb{N} : 1 \le t \le n : X_t \in [h_0, 1 - h_0]\}$ is the set of indices where X_t falls within the interior region $[h_0, 1 - h_0]$, and $h_0 \equiv h_0(n)$ is the preliminary bandwidth used in $\tilde{g}_{nt}(\cdot)$ and $\tilde{g}_{nt}^{(2)}(\cdot)$. The use of $\tilde{S}_n$ avoids the boundary effect for $\tilde{g}_{nt}^{(2)}(\cdot)$, which still exists despite the fact that the boundary effect of $\tilde{g}_{nt}(\cdot)$ has been taken care of when the jackknife kernel $k_b(\cdot)$ is used. Different preliminary estimators $\tilde{g}_{nt}(\cdot)$ and $\tilde{g}_{nt}^{(2)}(\cdot)$ are equivalent to different choices of a tuning parameter. If $\tilde{g}_{nt}(\cdot)$ and $\tilde{g}_{nt}^{(2)}(\cdot)$ are consistent for $g(\cdot)$ and $g^{(2)}(\cdot)$, $\hat{h}_Q$ will

[10]Strictly speaking, our theory does not cover the use of a data-driven bandwidth ($\hat{h}$ say), as is the case for the bulk of smoothed nonparametric testing procedures for serial dependence. However, we expect that with additional conditions on the smoothness of $k(\cdot)$ and the convergence rate of $\hat{h}$ to a nonstochastic bandwidth h, we could extend our theory to cover the use of $\hat{h}$. We leave this to further work. In this paper, we use simulation only to examine the performance of the tests using $\hat{h}_Q$ in (6.3).

be consistent for h_Q^0. For convenience, we also use the quartic kernel (3.2) for $\tilde{g}_{nt}(\cdot)$ and $\tilde{g}_{nt}^{(2)}(\cdot)$. To examine the sensitivity of all tests to the choice of the preliminary bandwidth h_0, we set $h_0 = \hat{S}_X n^{-1/(1+\lambda)}$ for $\lambda = 1, 2, 3, 4, 5$, where $\hat{S}_X$ is the sample standard deviation of sample $\mathcal{X}$. This covers a sufficiently wide range of rates for h_0. These rates ensure consistency of $\tilde{g}_{nt}(\cdot)$ and $\tilde{g}_{nt}^{(2)}(\cdot)$ for $g(\cdot)$ and $g^{(2)}(\cdot)$ if $g(\cdot)$ is fourth-order continuously differentiable on $\mathbb{I}$. For brevity, we report only the results for $\lambda = 5$.

To examine the bootstrap levels of the tests, we generate 1,000 realizations of sample $\mathcal{X}$ under DGP 0 (for both normal and lognormal distributions), using the GAUSS Windows version random number generator. We set $B = 100$, the number of bootstrap iterations for each simulation iteration. Tables I and II report the empirical rejection rates (in percentages) of $\mathcal{T}_n(j)$, $\mathcal{R}_n(j)$, and $\mathcal{J}_n(j)$ for $j = 1, \ldots, 10$, as well as their portmanteau tests (the last two rows). The portmanteau test statistics are $\mathcal{Q}_n(p)$ in (4.3), $p^{-1/2}\sum_{j=1}^{p}\mathcal{R}_n(j)$, and $p^{-1/2}\sum_{j=1}^{p}\mathcal{J}_n(j)$; we consider two lag truncation orders $p = 5, 10$. We first examine the individual tests $\mathcal{T}_n(j)$, $\mathcal{R}_n(j)$, and $\mathcal{J}_n(j)$. All three tests have reasonable levels at all three (10%, 5%, and 1%) significance levels, for all 10 lags, both sample sizes, and both normal and lognormal data. This indicates that the smoothed bootstrap procedure can effectively capture higher-order corrections to the first-order asymptotic approximation. Next, we examine the

TABLE I

BOOTSTRAP LEVELS OF TESTS UNDER I.I.D. N(0, 1) SAMPLE

	$n=100$									$n=200$								
	$\mathcal{T}_n(j)$			$\mathcal{R}_n(j)$			$\mathcal{J}_n(j)$			$\mathcal{T}_n(j)$			$\mathcal{R}_n(j)$			$\mathcal{J}_n(j)$		
j	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
1	12.6	6.5	1.4	10.5	4.7	1.0	10.5	5.1	0.7	12.6	6.6	1.6	12.1	6.5	1.9	11.0	5.8	0.9
2	11.7	6.4	0.8	11.2	6.7	1.2	10.4	5.5	1.0	12.4	5.9	0.7	10.8	5.3	0.7	10.7	5.7	1.3
3	12.0	5.8	1.1	10.2	4.6	1.3	10.6	4.8	0.5	11.7	5.6	1.0	8.9	4.3	1.0	10.7	5.1	0.8
4	12.2	6.8	1.1	10.0	4.6	0.8	9.8	4.6	0.7	10.8	5.9	1.4	9.8	4.6	0.9	9.5	4.3	1.4
5	11.1	6.2	1.3	8.9	4.2	1.2	10.7	5.3	1.2	12.2	6.4	1.3	11.4	5.9	0.9	10.3	6.0	1.9
6	12.2	5.6	2.0	9.8	5.0	0.9	9.0	4.8	1.1	12.0	5.9	0.8	11.7	5.4	1.3	11.1	6.1	1.3
7	12.5	6.4	1.4	10.9	5.4	1.7	10.7	5.5	1.1	13.8	6.9	1.1	12.0	6.2	1.9	10.6	4.9	0.6
8	10.5	5.7	0.8	9.8	4.6	0.6	8.9	3.6	0.8	11.0	5.5	1.0	11.2	5.5	0.6	10.6	5.2	0.7
9	12.1	5.5	0.8	10.1	5.2	0.8	9.5	5.1	1.1	11.0	6.2	0.8	12.4	6.6	1.0	9.9	5.2	0.8
10	11.2	5.4	1.6	10.8	5.5	0.8	10.8	5.2	1.0	10.9	6.2	0.8	10.2	5.0	1.0	10.4	4.8	1.0
p																		
5	13.6	6.0	1.1	11.2	4.9	1.1	10.7	5.2	0.7	12.9	6.1	0.7	12.3	7.0	1.7	11.5	6.0	1.5
10	12.8	6.5	1.1	11.5	5.7	1.3	10.1	5.3	0.7	14.0	5.8	0.9	14.5	8.8	1.9	9.4	5.1	0.8

Notes: (i) Data generating process, $X_t \sim$ i.i.d. N(0, 1).

(ii) $\mathcal{T}_n(j)$ denotes the new entropy test; $\mathcal{R}_n(j)$ denotes Robinson's (1991) test, and $\mathcal{J}_n(j)$ denotes Skaug and Tjøstheim's (1996) test.

(iii) 1,000 simulation iterations and 100 bootstrap iterations for each simulation iteration.

TABLE II

BOOTSTRAP LEVELS OF TESTS UNDER LOGNORMAL SAMPLE

	$n=100$									$n=200$								
	$\mathcal{T}_n(j)$			$\mathcal{R}_n(j)$			$\mathcal{J}_n(j)$			$\mathcal{T}_n(j)$			$\mathcal{R}_n(j)$			$\mathcal{J}_n(j)$		
j	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
1	12.4	5.8	0.5	10.6	4.3	0.6	10.4	5.8	1.0	10.9	5.0	0.7	11.3	6.2	1.3	10.5	4.6	1.1
2	9.8	4.7	0.3	10.1	4.4	0.7	9.6	4.4	1.1	10.7	6.3	1.7	11.1	5.6	1.1	10.7	5.3	1.0
3	10.8	5.8	1.2	9.6	4.3	0.4	8.6	4.5	0.2	8.3	4.2	0.7	9.7	4.3	0.8	10.5	4.7	0.7
4	11.3	6.0	1.2	10.0	4.9	1.4	8.9	4.3	0.6	12.2	5.8	1.0	10.7	5.8	1.0	8.8	3.9	0.8
5	12.8	6.9	1.5	9.1	4.8	1.8	8.8	3.6	0.8	10.4	5.4	1.4	10.0	4.5	0.9	9.6	5.3	1.0
6	13.0	7.1	1.7	8.4	4.3	0.9	10.0	4.7	1.2	9.9	5.1	0.7	10.2	4.0	0.9	10.9	5.9	1.3
7	13.0	5.5	1.2	9.5	5.0	1.3	11.1	4.3	0.5	11.7	5.7	1.3	10.5	4.6	0.8	10.7	4.6	0.9
8	11.7	6.0	1.5	10.5	5.3	0.8	9.1	4.2	0.8	10.8	5.3	0.7	8.4	4.2	0.9	10.0	4.9	1.1
9	12.8	5.8	1.2	10.7	5.2	1.4	10.0	4.4	1.0	9.5	4.1	0.6	11.4	5.1	1.4	9.4	4.7	0.8
10	12.9	6.9	1.1	10.0	5.0	1.3	8.6	3.4	0.5	9.3	4.1	0.4	10.4	4.5	1.1	8.5	4.9	0.7
p																		
5	10.3	3.8	0.5	9.4	4.8	0.8	10.9	4.7	0.5	9.5	3.9	0.8	10.9	6.7	1.1	10.0	4.3	0.5
10	11.2	4.6	0.6	10.6	5.2	1.0	10.2	4.2	0.8	7.9	3.3	0.3	12.0	5.9	0.9	9.0	4.0	0.7

Notes: (i) Data generating process, $X_t \sim$ i.i.d. lognormal(0, 1), scaled to have zero mean and unit variance.

(ii) $\mathcal{T}_n(j)$ denotes the new entropy test, $\mathcal{R}_n(j)$ denotes Robinson's (1991) test, and $\mathcal{J}_n(j)$ denotes Skaug and Tjøstheim's (1996) test.

(iii) 1,000 simulation iterations and 100 bootstrap iterations for each simulation iteration.

portmanteau tests. In almost all cases, our $\mathcal{Q}_n(p)$ test has reasonable levels at three levels and for both sample sizes, both lag truncation orders, and both normal and lognormal data. One exception is that it displays a bit of overrejection at the 10% level under the normal random sample. The portmanteau test based on the Robinson-type statistics $\mathcal{R}_n(j)$ also performs reasonably well, except that it displays some overrejections at all three levels under the normal random sample with $n = 200$. Skaug and Tjøstheim's (1996) portmanteau test has reasonable bootstrap levels at all three levels, for both sample sizes, both lag truncation orders, and both normal and lognormal data.

Now we turn to examining the powers of the tests under DGPs 1–8. We generate 500 realizations of sample $\mathcal{X}$ under each DGP. Again, we consider $n = 100, 200$, and set $B = 100$. Since all eight alternatives are first-order linear and nonlinear time-series processes, we examine only tests based on the first lag order ($j = 1$), which delivers the best power for each test among the 10 lags in most cases. Tests based on higher orders have little or low power in most cases.

Table III reports the empirical rejection rates of $\mathcal{T}_n(1)$, $\mathcal{R}_n(1)$, and $\mathcal{J}_n(1)$ under DGPs 1–8 with i.i.d. N(0, 1) innovations, for both $n = 100$ and 200. When $n = 100$, $\mathcal{T}_n(1)$ is as powerful as $\mathcal{R}_n(1)$ under DGPs 1, 6, and 8 at all three significance levels, and is more powerful than $\mathcal{R}_n(1)$ under DGPs 2, 3, 4, 5, and 7 at all three levels. When $n = 200$, $\mathcal{T}_n(1)$ is, to various extents (from sub-

TABLE III

POWERS OF TESTS UNDER I.I.D. N(0, 1) INNOVATION

DGP		$\mathcal{T}_n(j)$			$\mathcal{R}_n(j)$			$\mathcal{J}_n(j)$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
$n = 100$										
1	AR(1)	24.8	14.0	3.4	23.4	13.8	3.6	22.4	12.4	4.0
2	ARCH(1)	50.0	37.6	17.8	41.4	26.4	11.4	74.2	61.2	28.0
3	Threshold GARCH(1)	29.4	20.6	7.4	23.2	15.0	4.6	42.2	27.8	8.8
4	Bilinear AR(1)	79.0	69.6	45.6	72.6	59.8	33.4	91.2	81.6	52.0
5	Nonlinear MA(1)	50.6	34.0	14.0	45.2	31.8	12.8	51.0	34.8	12.0
6	Threshold AR(1)	38.6	25.6	9.2	38.0	24.6	9.6	39.6	25.8	8.0
7	Fractional AR(1)	25.6	17.0	5.0	22.8	14.2	4.8	21.4	13.4	3.8
8	Sign AR(1)	64.4	60.8	56.2	64.0	60.2	54.2	60.8	55.8	48.8
$n = 200$										
1	AR(1)	41.8	27.0	7.4	37.8	25.4	7.6	34.4	22.0	5.8
2	ARCH(1)	75.2	67.6	41.0	64.8	52.2	25.2	95.6	90.0	68.0
3	Threshold GARCH(1)	48.0	35.2	13.8	37.4	24.2	7.2	68.4	52.0	23.8
4	Bilinear AR(1)	97.0	95.6	86.8	94.4	91.4	75.4	99.2	98.4	92.4
5	Nonlinear MA(1)	85.4	74.0	49.8	76.6	65.2	37.2	83.0	72.8	43.2
6	Threshold AR(1)	90.8	85.4	25.2	80.8	71.0	23.8	94.4	86.8	18.2
7	Fractional AR(1)	37.8	26.2	10.0	34.6	22.2	8.4	38.2	23.8	7.4
8	Sign AR(1)	85.8	84.6	82.6	85.6	84.0	81.2	81.8	79.8	74.8

Notes: (i) DGP 1, $X_t = 0.3X_{t-1} + \varepsilon_t$; DGP 2, $X_t = \varepsilon_t h_t^{1/2}$, $h_t = 1 + 0.8X_{t-1}^2$; DGP 3, $X_t = \varepsilon_t h_t^{1/2}$, $h_t = 0.25 + 0.5h_{t-1} + 0.5X_{t-1}^2 \mathbb{1}(\varepsilon_t < 0) + 0.2X_{t-1}^2 \mathbb{1}(\varepsilon_t \geq 0)$; DGP 4, $X_t = 0.8X_{t-1}\varepsilon_{t-1} + \varepsilon_t$; DGP 5, $X_t = 0.8\varepsilon_{t-1}^2 + \varepsilon_t$; DGP 6, $X_t = -0.5X_{t-1}\mathbb{1}(X_{t-1} \leq 1) + 0.4X_{t-1}\mathbb{1}(X_{t-1} > 1) + \varepsilon_t$; DGP 7, $X_t = 0.8|X_{t-1}|^{0.5} + \varepsilon_t$; DGP 8, $X_t = \text{sign}(X_{t-1}) + 0.43\varepsilon_t$, where $\varepsilon_t \sim$ i.i.d. N(0, 1).

(ii) $\mathcal{T}_n(1)$ denotes the new entropy test, $\mathcal{R}_n(1)$ denotes Robinson's (1991) test, and $\mathcal{J}_n(1)$ denotes Skaug and Tjøstheim's (1996) test.

(iii) 500 simulation iterations and 100 bootstrap iterations for each simulation iteration.

stantially to slightly), more powerful than $\mathcal{R}_n(1)$ at the three levels under all eight DGPs, particularly at the 10% and 5% levels. On the other hand, when $n = 100$, $\mathcal{T}_n(1)$ is as powerful as $\mathcal{J}_n(1)$ under DGPs 1, 5, and 6, and is slightly more powerful than $\mathcal{J}_n(1)$ under DGPs 7 and 8. However, $\mathcal{J}_n(1)$ is significantly more powerful than $\mathcal{T}_n(1)$ under DGPs 2, 3, and 4 at the three levels. When $n = 200$, $\mathcal{T}_n(1)$ is more powerful than $\mathcal{J}_n(1)$ under DGPs 1, 5, 7, and 8, and is less powerful than $\mathcal{J}_n(1)$ under DGPs 2, 3, and 4 at the three levels. Under DGP 6, $\mathcal{J}_n(1)$ slightly outperforms $\mathcal{T}_n(1)$ at the 10% and 5% levels, but is dominated by $\mathcal{T}_n(1)$ at the 1% level.

Table IV reports the empirical rejection rates of the tests under DGPs 1–8 with i.i.d. lognormal innovations. For both $n = 100, 200$ and DGPs 1–7, $\mathcal{T}_n(1)$ outperforms $\mathcal{R}_n(1)$ at all three levels. Under DGP 8, $\mathcal{T}_n(1)$ is slightly more powerful than $\mathcal{R}_n(1)$ at the 10% and 5% levels, but they are roughly equally powerful at the 1% level. On the other hand, there is a less clear ranking between the powers of $\mathcal{T}_n(1)$ and $\mathcal{J}_n(1)$. When $n = 100$, $\mathcal{T}_n(1)$ outperforms $\mathcal{J}_n(1)$ under DGPs 1 and 5, is equally powerful to $\mathcal{J}_n(1)$ under DGPs 2, 3, and 8,

TABLE IV

POWERS OF TESTS UNDER I.I.D. LOGNORMAL INNOVATIONS

		$\mathcal{T}_n(j)$			$\mathcal{R}_n(j)$			$\mathcal{J}_n(j)$		
DGP		10%	5%	1%	10%	5%	1%	10%	5%	1%
$n = 100$										
1	AR(1)	71.8	60.8	32.2	41.6	30.4	10.6	70.6	52.4	21.6
2	ARCH(1)	52.8	39.4	15.6	36.2	22.0	7.8	53.6	35.0	12.4
3	Threshold GARCH(1)	31.0	21.0	7.0	19.6	10.8	4.0	33.0	18.0	4.2
4	Bilinear AR(1)	89.8	79.4	52.4	62.0	46.4	18.6	95.0	90.8	74.2
5	Nonlinear MA(1)	69.4	57.8	29.2	44.0	30.4	11.8	67.2	52.4	21.4
6	Threshold AR(1)	90.8	85.4	66.0	80.8	71.0	44.0	94.4	86.8	55.4
7	Fractional AR(1)	92.2	86.4	62.4	76.6	61.4	33.0	92.6	86.6	58.0
8	Sign AR(1)	17.6	10.4	4.4	14.6	8.8	5.2	16.2	8.2	4.4
$n = 200$										
1	AR(1)	96.8	92.8	75.8	82.2	69.8	37.8	97.2	90.8	69.0
2	ARCH(1)	79.8	69.6	44.8	61.2	46.6	22.6	81.6	70.6	34.4
3	Threshold GARCH(1)	47.6	34.4	15.6	34.0	22.4	7.2	46.6	32.2	11.4
4	Bilinear AR(1)	98.8	98.0	92.6	89.4	82.0	57.6	98.6	98.0	93.8
5	Nonlinear MA(1)	94.2	89.8	70.2	80.6	67.0	31.2	94.2	88.4	68.4
6	Threshold AR(1)	99.8	99.4	97.0	98.6	97.8	88.8	100.0	99.8	97.6
7	Fractional AR(1)	100.0	99.6	96.4	98.6	95.6	82.4	99.8	99.8	97.2
8	Sign AR(1)	15.0	7.8	2.8	12.0	6.4	3.0	13.2	7.2	2.0

Notes: (i) DGP 1, $X_t = 0.3X_{t-1} + \varepsilon_t$; DGP 2, $X_t = \varepsilon_t h_t^{1/2}$, $h_t = 1 + 0.8X_{t-1}^2$; DGP 3, $X_t = \varepsilon_t h_t^{1/2}$, $h_t = 0.25 + 0.5h_{t-1} + 0.5X_{t-1}^2 \mathbb{1}(\varepsilon_t < 0) + 0.2X_{t-1}^2 \mathbb{1}(\varepsilon_t \geq 0)$; DGP 4, $X_t = 0.8X_{t-1}\varepsilon_{t-1} + \varepsilon_t$; DGP 5, $X_t = 0.8\varepsilon_{t-1}^2 + \varepsilon_t$; DGP 6, $X_t = -0.5X_{t-1}\mathbb{1}(X_{t-1} \leq 1) + 0.4X_{t-1}\mathbb{1}(X_{t-1} > 1) + \varepsilon_t$; DGP 7, $X_t = 0.8|X_{t-1}|^{0.5} + \varepsilon_t$; DGP 8, $X_t = \text{sign}(X_{t-1}) + 0.43\varepsilon_t$, where $\varepsilon_t \sim$ i.i.d. lognormal(0, 1), standardized to have zero mean and unit variance.

(ii) $\mathcal{T}_n(1)$ denotes the new entropy test, $\mathcal{R}_n(1)$ denotes Robinson's (1991) test, and $\mathcal{J}_n(1)$ denotes Skaug and Tjøstheim's (1996) test.

(iii) 500 simulation iterations and 100 bootstrap iterations for each simulation iteration.

and is less powerful than $\mathcal{J}_n(1)$ under DGP 4. Under DGP 6, $\mathcal{J}_n(1)$ slightly outperforms $\mathcal{T}_n(1)$ at the 10% and 5% levels, but is dominated by $\mathcal{T}_n(1)$ at the 1% level. Under DGP 7, $\mathcal{T}_n(1)$ is as powerful as $\mathcal{J}_n(1)$ at the 10% and 5% levels, but it outperforms $\mathcal{J}_n(1)$ at the 1% level. When $n = 200$, $\mathcal{T}_n(1)$ is roughly equally powerful to $\mathcal{J}_n(1)$ at the 10% and 5% levels under DGPs 4–8. Under DGPs 1–3, $\mathcal{T}_n(1)$ is as powerful as $\mathcal{J}_n(1)$ at the 10% and 5% levels, and it outperforms $\mathcal{J}_n(1)$ at the 1% level. Interestingly, all three tests have excellent power against DGP 8 (Sign AR(1)) with i.i.d. N(0, 1) innovations, but they all have little power when the innovations are lognormal.

To sum up: (i) All three tests $\mathcal{T}_n(j), \mathcal{R}_n(j)$, and $\mathcal{J}_n(j)$ and their portmanteau versions have reasonable bootstrap levels in small samples, for various lag orders and both normal and lognormal data. The smoothed bootstrap procedure can effectively capture higher-order corrections to the first-order asymptotic approximation. (ii) The power of the $\mathcal{T}_n(1)$ test always is better than or equal to the power of the $\mathcal{R}_n(1)$ test for all eight DGPs (especially with i.i.d. lognormal innovations). This result corroborates our asymptotic analysis that

$\mathcal{T}_n(1)$ is asymptotically locally more powerful than $\mathcal{R}_n(1)$. (iii) There is a less clear ranking between the powers of $\mathcal{T}_n(1)$ and $\mathcal{J}_n(1)$. The $\mathcal{J}_n(1)$ test tends to outperform in most cases when the innovations are i.i.d. N(0, 1) and the sample size is smaller. In contrast, $\mathcal{T}_n(1)$ tends to outperform $\mathcal{J}_n(1)$ in more cases when the innovations are lognormal, or when the sample size is larger, or when the significance level is smaller.

7. APPLICATION TO THE S&P 500 STOCK INDEX

7.1. *Testing the Random Walk Hypothesis*

We now use our tests to explore possible nonlinear serial dependence in the daily S&P 500 stock price index. It has long been hypothesized that stock prices follow a (geometric) random walk possibly with a drift. We are first interested in testing this hypothesis and in identifying important lags. From DATASTREAM, we obtained data on the S&P 500 daily closing price index (P_t) from January 1, 1992 to December 31, 2003, for a total of 3,136 observations. We define $X_t \equiv 100\ln(P_t/P_{t-1})$, transform it by the logistic function (3.1), and rescale it to have support on $\mathbb{I}$. Then we can test the random walk hypothesis by checking if $\{X_t\}$ is i.i.d. Panel A in Table V reports the boot-

TABLE V

BOOTSTRAP p-VALUES OF TESTS FOR DAILY S&P 500 STOCK RETURNS

	A: S&P 500 Daily Returns				B: AR(3)–GARCH(1, 1) Residuals			
j	$\mathcal{T}_n(j)$	$\mathcal{R}_n(j)$	$\mathcal{J}_n(j)$	$\mathcal{C}_n(j)$	$\mathcal{T}_n(j)$	$\mathcal{R}_n(j)$	$\mathcal{J}_n(j)$	$\mathcal{C}_n(j)$
1	0.108	0.071	0.000	0.920	0.102	0.016	0.688	0.864
2	0.000	0.018	0.000	0.069	0.029	0.155	0.241	0.011
3	0.001	0.085	0.000	0.006	0.720	0.644	0.786	0.014
4	0.000	0.293	0.000	0.584	0.227	0.094	0.007	0.245
5	0.000	0.022	0.000	0.304	0.002	0.033	0.156	0.115
6	0.000	0.302	0.000	0.185	0.923	0.740	0.486	0.157
7	0.068	0.033	0.000	0.096	0.192	0.103	0.535	0.465
8	0.001	0.587	0.000	0.925	0.625	0.424	0.985	0.668
9	0.003	0.110	0.000	0.828	0.323	0.016	0.077	0.798
10	0.012	0.645	0.000	0.040	0.197	0.586	0.023	0.226
p								
5	0.000	0.005	0.000	0.029	0.020	0.021	0.074	0.083
10	0.000	0.004	0.000	0.018	0.109	0.023	0.066	0.223

Notes: (i) The sample period for the S&P 500 daily price index is from January 1, 1992 to December 31, 2003, with a total of 3,151 observations.

(ii) $\mathcal{T}_n(j)$ denotes the new entropy test, $\mathcal{R}_n(j)$ denotes Robinson's (1991) modified entropy test, $\mathcal{J}_n(j)$ denotes Skaug and Tjøstheim's (1996) test, and $\mathcal{C}_n(j)$ denotes the autocorrelation test. The last two rows are the corresponding portmanteau tests, with truncation lag orders $p = 5, 10$, respectively.

(iii) Panel A shows the bootstrap p-values for observed raw S&P 500 daily price changes, using the smoothed bootstrap procedure proposed in Section 4. Panel B is the bootstrap p-values for estimated standardized residuals from the AR(3)–GARCH(1, 1) model in (7.1), using the smoothed bootstrap procedure described in Section 7. $B = 1{,}000$ bootstrap iterations are used in computing the bootstrap p-values.

strap p-values of $\mathcal{T}_n(j)$, $\mathcal{R}_n(j)$, and $\mathcal{J}_n(j)$ for lag orders j from 1 to 10, and their corresponding portmanteau tests for truncation lag orders $p = 5, 10$. The bootstrap p-values, based on $B = 1{,}000$ bootstrap iterations, are computed as described in Section 4. For all three tests, we use the quartic kernel in (3.2) and the plug-in bandwidth in (6.3), with the preliminary bandwidth $h_0 = \hat{S}_X n^{-1/6}$, where $\hat{S}_X$ is the sample standard deviation of $\mathcal{X} \equiv \{X_t\}_{t=1}^n$. We set $(\gamma, \delta) = (0.5, 0)$ for $\mathcal{R}_n(j)$ and use a Beta(2, 2) density function for $w(\cdot)$ in $\mathcal{J}_n(j)$. We also report the autocorrelation statistic $\mathcal{C}_n(j) \equiv n_j^{1/2}\hat{\rho}_n(j)$ and the Box–Pierce portmanteau statistic $\mathrm{BP}(p) \equiv \sum_{j=1}^p \mathcal{C}_n^2(j)$, where $\hat{\rho}_n(j)$ is the sample autocorrelation function of $\mathcal{X}$. Under $\mathbb{H}_0$, $\mathcal{C}_n(j) \xrightarrow{d} \mathrm{N}(0, 1)$ and $\mathrm{BP}(p) \xrightarrow{d} \chi_p^2$. Unlike $\mathcal{T}_n(j)$, $\mathcal{R}_n(j)$, and $\mathcal{J}_n(j)$, however, two-sided bootstrap critical values should be used for $\mathcal{C}_n(j)$.

We first examine the portmanteau tests. Both our portmanteau test and Skaug and Tjøstheim's (1996) portmanteau test have a zero p-value for both $p = 5$ and 10, and the portmanteau test based on Robinson's (1991) statistics has p-values of 0.5% and 0.4% for $p = 5$ and 10, respectively. Moreover, the Box–Pierce portmanteau test has p-values of 2.9% and 1.8% for $p = 5$ and 10, respectively. All these results indicate that the S&P 500 daily stock price index does not follow a random walk, and nonparametric density-based tests give much stronger evidence than the Box–Pierce test. Next, we use individual tests to examine possible lag structure of serial dependence in $\{X_t\}$. The $\mathcal{J}_n(j)$ test has a zero p-value for all 10 lags. The $\mathcal{T}_n(j)$ test has p-values ranging from 0 to 1.2% for all lags except lags 1 and 7. At lags 1 and 7, $\mathcal{T}_n(j)$ is insignificant at the 5% level. The $\mathcal{R}_n(j)$ test has p-values less than or equal to 3.3% only at lags 2, 5, and 7. For all other lags, $\mathcal{R}_n(j)$ is insignificant at the 5% level. The $\mathcal{C}_n(j)$ test has p-values of 0.6% and 4% at lags 3 and 10, respectively, and is insignificant at the 5% level for all other eight lags. The results of $\mathcal{C}_n(j)$ suggest that there might exist mild low-order autocorrelation in $\{X_t\}$.

7.2. *Testing the i.i.d. Hypothesis for Standardized Model Residuals*

It is well known that there exists strong volatility clustering in stock returns; this, together with significant autocorrelation at lag 3, may well have contributed to our rejection of the random walk hypothesis. To check whether the rejection is due to GARCH effects and low-order autocorrelation in $\{X_t\}$, we fit an AR(3)–GARCH(1, 1) model to $\{X_t\}$ by the quasi-maximum likelihood estimation (QMLE) method using a Gaussian likelihood. The estimated AR(3)–GARCH(1, 1) parameters are indeed highly significant:

$$\begin{aligned}(7.1)\qquad X_t &= \underset{(0.0137)}{0.0549} + \underset{(0.0189)}{0.0075}X_{t-1} - \underset{(0.0186)}{0.0060}X_{t-2} - \underset{(0.0187)}{0.0309}X_{t-3} + \hat{\varepsilon}_t\hat{h}_t^{1/2},\end{aligned}$$

$$\hat{h}_t = 0.0045 + 0.0589\hat{\varepsilon}_{t-1}^2\hat{h}_{t-1} + 0.9387\hat{h}_{t-1},$$
$$(0.0015)\ (0.0081) \qquad\qquad (0.0080)$$
$$n = 3135, \quad \text{mean log-likelihood} = -1.31516,$$

where the numbers inside the parentheses are robust QMLE standard errors. We now apply the tests to the standardized residuals $\{\hat{\varepsilon}_t\}_{t=1}^n$. We emphasize that like Robinson (1991) and Skaug and Tjøstheim (1996), our theory is based on observed raw data $\mathcal{X}$ rather than the estimated residuals $\{\hat{\varepsilon}_t\}_{t=1}^n$, but it is plausible that parameter estimation uncertainty has no impact asymptotically, as the convergence rate of parameter estimators is $n^{-1/2}$ (e.g., Lee and Hansen (1994)), more rapid than that for our $\hat{f}_{jt}(\cdot)$ and $\hat{g}_t(\cdot)$. Moreover, the smoothed bootstrap procedure described below is expected to take into account the impact of parameter estimation uncertainty in finite samples.

The smoothed bootstrap is implemented here as follows: (i) Transform $\{\hat{\varepsilon}_t\}_{t=1}^n$ via the logistic function in (3.1) and scale it to have unit support on $\mathbb{I}$. Denote the transformed standardized residual sample as $\{e_t\}_{t=1}^n$. Then use $\{e_t\}_{t=1}^n$ to obtain the smoothed density estimator $\hat{g}_e(e) = n^{-1}\sum_{t=1}^n K_h(e - e_t)$ for the standardized residuals, where h is based on the "plug-in" method in (6.3). (ii) Draw a bootstrap sample $\{e_t^*\}_{t=1}^n$ from the smoothed density estimator $\hat{g}_e(\cdot)$. Transform $\{e_t^*\}_{t=1}^n$ via the inverse logistic function transformation, normalized to have zero mean and unit variance, and denote the transformed sample as $\{\varepsilon_t^*\}_{t=1}^n$. (iii) Construct a bootstrap sample $\{X_t^*\}_{t=1}^n$ using $\{\varepsilon_t^*\}_{t=1}^n$ and the estimated parameters in (7.1). (iv) Estimate an AR(3)–GARCH(1, 1) model for $\{X_t^*\}_{t=1}^n$ via QMLE and save the resulting estimated standardized residuals $\{\hat{\varepsilon}_t^*\}_{t=1}^n$. (v) Transform $\{\hat{\varepsilon}_t^*\}_{t=1}^n$ via (3.1) and scale it to have unit support on $\mathbb{I}$. Then compute a bootstrap statistic, say $\hat{\mathcal{I}}_n^*(j)$, using $\{\hat{\varepsilon}_t^*\}_{t=1}^n$. (vi) Repeat Steps (ii)–(v) B times, and thus obtain B bootstrap test statistics $\{\hat{\mathcal{I}}_{nl}^*(j)\}_{l=1}^B$. (vii) The bootstrap p-value is $p^* = B^{-1}\sum_{l=1}^N \mathbb{1}[\hat{\mathcal{I}}_{nl}^*(j) > \hat{\mathcal{I}}_n(j)]$, where $\hat{\mathcal{I}}_n(j)$ is the entropy estimator based on $\{\hat{\varepsilon}_t\}_{t=1}^n$. We use $B = 1{,}000$.

Panel B in Table V reports the bootstrap p-values of $\mathcal{T}_n(j)$, $\mathcal{R}_n(j)$, $\mathcal{J}_n(j)$, and $\mathcal{C}_n(j)$ for $j = 1, \ldots, 10$, as well as their portmanteau tests. We first examine the portmanteau tests. Our portmanteau test $\mathcal{Q}_n(p)$ has p-values of 2.0% and 10.9% for $p = 5$ and 10, respectively, giving some evidence of serial dependence in $\{\hat{\varepsilon}_t\}$. From the individual $\mathcal{T}_n(j)$ tests, we see that the rejection of $\mathcal{Q}_n(5)$ mainly comes from serial dependence at lags 2 and 5. Next, the portmanteau test based on Robinson's (1991) statistics has p-values of 2.1% and 2.3% for $p = 5$ and 10, respectively, giving significant evidence of serial dependence in $\{\hat{\varepsilon}_t\}$ at the 5% level for both $p = 5$ and 10. The individual $\mathcal{R}_n(j)$ statistic is significant at the 5% level for lags 1, 5, and 9. On the other hand, Skaug and Tjøstheim's (1996) portmanteau test has p-values of 7.4% and 6.6% for $p = 5$ and 10, although $\mathcal{J}_n(j)$ is significant at the 5% level for lag 4, and the Box–Pierce test has p-values of 8.3% and 22.3% for $p = 5$ and 10, respectively, indicating no serial dependence in $\{\hat{\varepsilon}_t\}$. The individual $C_n(j)$ test, however, is significant at the 5% level for lags 2 and 3. Overall, Robinson's (1991) tests and our tests suggest that the AR(3)–GARCH(1, 1) model cannot fully capture the dynamics of the S&P 500 daily returns. This seems to be consistent with the empirical findings of Hansen (1994) and Jondeau and Rockinger (2003) that higher-order conditional moments (e.g., skewness and kurtosis) of financial time series are time-varying. Of course, this may also be due to misspecification of the conditional mean and conditional variance. It would be interesting to explore the extent to which the remaining serial dependence can be used to forecast the distribution of the S&P 500 daily returns, which could be useful for improving (e.g.) financial risk management, hedging, and derivatives pricing. This requires estimation and out-of-sample forecast evaluation of suitable nonlinear time-series models. We leave this to subsequent research.

8. CONCLUSION

We have provided asymptotic theory for a class of kernel-based smoothed nonparametric entropy estimators of serial dependence. We used our theory to derive the limiting distribution of Granger and Lin's (1994) normalized entropy measure of serial dependence, which was previously not available in the literature. We also constructed a new entropy-based test for serial dependence, providing an alternative to Robinson's (1991) test. To obtain accurate inferences in finite samples, we proposed and justified a consistent smoothed bootstrap procedure. The naive bootstrap is not consistent for our test. Our test is useful in, for example, testing the random walk hypothesis, evaluating density forecasts, and identifying important lags of a time series. It is asymptotically locally more powerful than Robinson's (1991) test, as is confirmed in a simulation study. An empirical application to the daily S&P 500 index illustrates our approach, revealing potential opportunities for improving the modeling of the S&P 500 dynamics.

Dept. of Economics and Dept. of Statistical Science, Cornell University, 424 Uris Hall, Ithaca, NY 14853-7601, U.S.A.; Dept. of Economics, School of Economics and Management, Tsinghua University, Beijing 100084, China; yh20@cornell.edu

and

Project for Econometric Analysis, Dept. of Economics, University of California, San Diego, CA 92093, U.S.A.; hwhite@weber.ucsd.edu.

Manuscript received August, 2001; final revision received May, 2004.

APPENDIX A: PROOFS OF THEOREMS

PROOF OF THEOREM 3.1: (a) To show Theorem 3.1(a), we first state Theorems A.1 and A.2.

THEOREM A.1: *Suppose Assumptions* A.1 *and* A.2 *hold*, $nh^4/\ln n \to \infty$, $nh^7 \to 0$, *and* $j = o(n)$. *For* $z_1, z_2 \in \mathbb{I}^2$, *put* $\tilde{K}_h^{(2)}(z_1, z_2) \equiv K_h^{(2)}(z_1, z_2) - \int_{\mathbb{I}^2} K_h^{(2)}(z, z_2)\,dz$, *where* $K_h^{(2)}(\cdot, \cdot)$ *is given in* (3.5). *Define*

$$A_{jn}(z_1, z_2) \equiv \left[K_h^{(2)}(z_1, z_2) - \int_{\mathbb{I}^2} K_h^{(2)}(z_1, z)f_j(z)\,dz\right] \Big/ f_j(z_1),$$

$$\tilde{A}_{jn}(z_1, z_2) \equiv \left[\tilde{K}_h^{(2)}(z_1, z_2) - \int_{\mathbb{I}^2} \tilde{K}_h^{(2)}(z_1, z)f_j(z)\,dz\right] \Big/ f_j(z_1),$$

$$B_{jn}(z_1) \equiv \left[\int_{\mathbb{I}^2} K_h^{(2)}(z_1, z)f_j(z)\,dz - f_j(z_1)\right] \Big/ f_j(z_1),$$

$$H_{1jn}(z_1, z_2) \equiv \tilde{A}_{jn}(z_1, z_2) + \tilde{A}_{jn}(z_2, z_1),$$

$$H_{2jn}(z_1, z_2) \equiv \int_{\mathbb{I}^2} A_{jn}(z, z_1)A_{jn}(z, z_2)f_j(z)\,dz.$$

Then under $\mathbb{H}_0$,

$$2\hat{I}_{jn}(\hat{f}_j, f_j) = -L_n(j) + \hat{H}_n(j) + 2[\hat{B}_n(j) - \hat{C}_n(j)] + o_P(n_j^{-1}h^{-1}),$$

where $L_n(j) \equiv n_{j+1}^{-1} EA_{jn}^2(Z_3, Z_1) + EB_{jn}^2(Z_1)$, $Z_3 \equiv (X_3, X_2)'$, $Z_1 \equiv (X_1, X_0)'$, $\hat{H}_n(j) \equiv \hat{H}_{1n}(j) - \hat{H}_{2n}(j)$,

$$\hat{H}_{1n}(j) = \binom{n_j}{2}^{-1} \sum_{t=j+2}^{n} \sum_{s=j+1}^{t-1} H_{1jn}(Z_{jt}, Z_{js}),$$

$$\hat{H}_{2n}(j) = \binom{n_j}{2}^{-1} \sum_{t=j+2}^{n} \sum_{s=j+1}^{t-1} H_{2jn}(Z_{jt}, Z_{js}),$$

$$\hat{B}_n(j) = n_j^{-1} \sum_{t=j+1}^{n} B_{jn}(Z_{jt}), \quad \textit{and}$$

$$\hat{C}_n(j) = n_j^{-1} \sum_{t=j+1}^{n} \int_{\mathbb{I}^2} A_{jn}(z, Z_{jt})B_{jn}(z)f_j(z)\,dz.$$

By construction, we have $\int_{\mathbb{I}^2} H_{ijn}(z_1, z)f_j(z)\,dz = \int_{\mathbb{I}^2} H_{ijn}(z, z_2)f_j(z)\,dz = 0$ for all $z_1, z_2 \in \mathbb{I}^2$, $i = 1, 2$. Note that $H_{2jn}(\cdot, \cdot)$ is defined in terms of $A_{jn}(\cdot, \cdot)$ rather than $\tilde{A}_{jn}(\cdot, \cdot)$.

THEOREM A.2: *Suppose Assumptions* A.1 *and* A.2 *hold*, $nh/\ln n \to \infty$, $nh^7 \to 0$, *and* $j = o(n)$. *For* $x_1, x_2 \in \mathbb{I}$, *put* $\tilde{K}_h(x_1, x_2) \equiv K_h(x_1, x_2) -$

$\int_0^1 K_h(x, x_2)\, dx$, *where* $K_h(\cdot,\cdot)$ *is given in* (3.6). *Define*

$$a_n(x_1, x_2) \equiv \left[K_h(x_1, x_2) - \int_0^1 K_h(x_1, x) g(x)\, dx\right] \Big/ g(x_1),$$

$$\tilde{a}_n(x_1, x_2) \equiv \left[\tilde{K}_h(x_1, x_2) - \int_0^1 \tilde{K}_h(x_1, x) g(x)\, dx\right] \Big/ g(x_1),$$

$$b_n(x_1) \equiv \left[\int_0^1 K_h(x_1, x) g(x)\, dx - g(x_1)\right] \Big/ g(x_1),$$

$$v_{1n}(x_1, x_2) \equiv \tilde{a}_n(x_1, x_2) + \tilde{a}_n(x_2, x_1),$$

$$v_{2n}(x_1, x_2) \equiv \int_0^1 a_n(x, x_1) a_n(x, x_2) g(x)\, dx.$$

Then under $\mathbb{H}_0$,

$$2\hat{I}_{ijn}(\hat{g}, g) = -l_n(j) + \hat{V}_{in}(j) + 2[\hat{b}_{in}(j) - \hat{c}_{in}(j)] + o_P(n_j^{-1} h^{-1/2}) \qquad (i = 1, 2),$$

where $l_n(j) \equiv n_{j+1}^{-1} E a_n^2(X_2, X_1) + E b_n^2(X_1)$,

$$\hat{V}_{1n}(j) \equiv \binom{n_j}{2}^{-1} \sum_{t=j+2}^{n} \sum_{s=j+1}^{t-1} [v_{1n}(X_t, X_s) - v_{2n}(X_t, X_s)],$$

$$\hat{V}_{2n}(j) \equiv \binom{n_j}{2}^{-1} \sum_{t=j+2}^{n} \sum_{s=j+1}^{t-1} [v_{1n}(X_{t-j}, X_{s-j}) - v_{2n}(X_{t-j}, X_{s-j})],$$

$\hat{b}_{1n}(j) = n_j^{-1} \sum_{t=j+1}^{n} b_n(X_t)$, $\hat{b}_{2n}(j) = n_j^{-1} \sum_{t=j+1}^{n} b_n(X_{t-j})$, $\hat{c}_{1n}(j) = n_j^{-1} \times \sum_{t=j+1}^{n} \int_0^1 a_n(x, X_t) b_n(x) g(x)\, dx$, *and* $\hat{c}_{2n}(j) = n_j^{-1} \sum_{t=j+1}^{n} \int_0^1 a_n(x, X_{t-j}) b_n(x) \times g(x)\, dx$.

The proofs of Theorems A.1 and A.2 are deferred to the end of this proof. We first use these theorems to show Theorem 3.1(a). By Theorems A.1 and A.2, (3.7), $\hat{I}_{jn}(f_j, g \circ g) = 0$ a.s., and $h \to 0$, we have

$$\text{(A1)} \qquad \begin{aligned} 2\hat{\mathcal{I}}_n(j) = {} & -n_{j+1}^{-1}[E A_{jn}^2(Z_3, Z_1) - 2E a_n^2(X_2, X_1)] \\ & - [E B_{jn}^2(Z_1) - 2E b_n^2(X_1)] \\ & + \{\hat{H}_n(j) - [\hat{V}_{1n}(j) + \hat{V}_{2n}(j)]\} + 2[\hat{B}_n(j) - \hat{b}_{1n}(j) - \hat{b}_{2n}(j)] \\ & - 2[\hat{C}_n(j) - \hat{c}_{1n}(j) - \hat{c}_{2n}(j)] + o_P(n_j^{-1} h^{-1}). \end{aligned}$$

It follows that $2\hat{\mathcal{I}}_n(j) = -n_{j+1}^{-1}d_n^0 + \hat{H}_n(j) + o_P(n_j^{-1}h^{-1})$ by Lemmas A.1–A.4 and $\hat{V}_{1n}(j) + \hat{V}_{2n}(j) = O_P(n_j^{-1}h^{-1/2})$ under $\mathbb{H}_0$ by Chebyshev's inequality. The desired asymptotic normality then follows because $hn_j\hat{H}_n(j) \xrightarrow{d} \mathrm{N}(0, \sigma^2)$ under $\mathbb{H}_0$. The proof of the latter is omitted here because it is a special case of Theorems A.6–A.9, which show $hn_j\hat{H}_n(j) \xrightarrow{d} \mathrm{N}(0, \sigma^2)$ under a class of local alternatives.

We now state Lemmas A.1–A.4 under the conditions of Theorem 3.1. The proofs of these lemmas are given in Appendix B.

LEMMA A.1: *Under* $\mathbb{H}_0$, $EA_{jn}^2(Z_3, Z_1) - 2Ea_n^2(X_2, X_1) = d_n^0 + O(1)$, *where* $Z_3 = (X_3, X_2)'$, $Z_1 = (X_1, X_0)'$, *and* d_n^0 *is as in* (3.14).

LEMMA A.2: *Under* $\mathbb{H}_0$, $EB_{jn}^2(Z_1) - 2Eb_n^2(X_1) = 2[Eb_n(X_1)]^2 + O(h^6)$.

LEMMA A.3: *Under* $\mathbb{H}_0$, $\hat{B}_n(j) - \hat{b}_{1n}(j) - \hat{b}_{2n}(j) = [Eb_n(X_1)]^2 + O_P(n_j^{-1/2}h^4)$.

LEMMA A.4: *Under* $\mathbb{H}_0$, $\hat{C}_n(j) - \hat{c}_{1n}(j) - \hat{c}_{2n}(j) = O_P(n_j^{-1/2}h^4)$.

(b) Next, we show Theorem 3.1(b) by the Cramer–Wold device. Let $\lambda \in \mathbb{R}^p$ be an arbitrary nonzero vector such that $\lambda'\lambda = 1$, where $p \in \mathbb{N}$ is fixed. Define $\hat{\mathcal{M}}_\lambda = \sum_{j=1}^p \lambda_j[2hn_j\hat{\mathcal{I}}_n(j) + hd_n^0]$ and $\hat{\mathcal{H}}_\lambda = \sum_{j=1}^p \lambda_j hn_j\hat{H}_n(j)$. Then we have $\hat{\mathcal{M}}_\lambda = \hat{\mathcal{H}}_\lambda + o_P(1)$ given that $2\hat{\mathcal{I}}_n(j) + n_{j+1}^{-1}d_n^0 = \hat{H}_n(j) + o_P(n_j^{-1}h^{-1})$ and $\max_{1\le j\le p}|\lambda_j| \le 1$. Using reasoning analogous to Theorems A.6–A.9, we have $\hat{\mathcal{H}}_\lambda \xrightarrow{d} \mathrm{N}(0, \sigma^2)$ under $\mathbb{H}_0$ for any arbitrary λ with $\lambda'\lambda = 1$. Here, $\mathrm{avar}(\hat{\mathcal{H}}_\lambda) \to \sigma^2$ for all $\lambda \neq 0$ such that $\lambda'\lambda = 1$, because $\mathrm{avar}[hn_j\hat{H}_n(j)] \to \sigma^2$ for all $j = o(n)$ and $\mathrm{cov}[hn_i\hat{H}_n(i), hn_j\hat{H}_n(j)] \to 0$ for $i \neq j$, and $i, j = o(n)$. The latter can be shown in the same way as the former in the proof of Theorem A.8. It follows that $\hat{\mathcal{M}}_\lambda \xrightarrow{d} \mathrm{N}(0, \sigma^2)$ for any λ with $\lambda'\lambda = 1$ and thus $\hat{\mathcal{I}}_n \xrightarrow{d} \mathrm{N}(\mathbf{0}, \sigma^2\mathbf{I}_p)$ by the Cramer–Wold device. This completes the proof of Theorem 3.1(b).

The proof of Theorem 3.1 will be completed provided Theorems A.1 and A.2 are proven, which we turn to next.

PROOF OF THEOREM A.1: We show Theorem A.1 using Lemmas A.5–A.10, which are proved in Appendix B. First, by the inequality that $|\ln(1+u) - u + \frac{1}{2}u^2| \le |u|^3$ for $|u| < 1$, we obtain the following lemma.

LEMMA A.5: *Suppose Assumptions* A.1 *and* A.2 *hold*, $nh^4/\ln n \to \infty$, $h \to 0$, *and* $j = o(n)$. *Then* $\hat{I}_{jn}(\hat{f}_j, f_j) - \hat{W}_1(j) + \frac{1}{2}\hat{W}_2(j) = O_P(n_j^{-3/2}h^{-3}\ln n_j + h^6)$ *under* $\mathbb{H}_0$, *where* $\hat{W}_1(j)$ *and* $\hat{W}_2(j)$ *are as in* (3.8).

We now consider the first-order term $\hat{W}_1(j)$. Put $\bar{f}_j(z_1) \equiv \int_{\mathbb{I}^2} K_h^{(2)}(z_1, z_2) \times f_j(z_2)\,dz_2$ and $\gamma_{jn}(z_1, z_2) = \int_{\mathbb{I}^2}[K_h^{(2)}(z, z_2) - \int_{\mathbb{I}^2} K_h^{(2)}(z, z')f_j(z')\,dz']\,dz/f_j(z_1)$. Then we can write

$$
\begin{aligned}
\text{(A2)} \qquad \hat{W}_1(j) &= n_j^{-1} \sum_{t=j+1}^{n} \left[\frac{\hat{f}_{jt}(Z_{jt}) - \bar{f}_j(Z_{jt})}{f_j(Z_{jt})} + \frac{\bar{f}_j(Z_{jt}) - f_j(Z_{jt})}{f_j(Z_{jt})} \right] \\
&= \frac{1}{2} \binom{n_j}{2}^{-1} \sum_{t=j+2}^{n} \sum_{s=j+1}^{t-1} [\tilde{A}_{jn}(Z_{jt}, Z_{js}) + \tilde{A}_{jn}(Z_{js}, Z_{jt})] \\
&\quad + \frac{1}{2} \binom{n_j}{2}^{-1} \sum_{t=j+2}^{n} \sum_{s=j+1}^{t-1} [\gamma_{jn}(Z_{jt}, Z_{js}) + \gamma_{jn}(Z_{js}, Z_{jt})] \\
&\quad + n_j^{-1} \sum_{t=j+1}^{n} B_{jn}(Z_{jt}) \\
&= \frac{1}{2}\hat{H}_{1n}(j) + \frac{1}{2}\hat{\Gamma}_n(j) + \hat{B}_n(j).
\end{aligned}
$$

LEMMA A.6: *Suppose Assumptions* A.1 *and* A.2 *hold*, $h \to 0$, *and* $j = o(n)$. *Then with probability one*, $\hat{\Gamma}_n(j) = 0$ *for all n sufficiently large under* $\mathbb{H}_0$.

We now consider the second-order term $\hat{W}_2(j)$ in Lemma A.5. We write

$$
\begin{aligned}
\text{(A3)} \qquad \hat{W}_2(j) &= n_j^{-1} \sum_{t=j+1}^{n} \left[\frac{\hat{f}_{jt}(Z_{jt}) - \bar{f}_j(Z_{jt})}{f_j(Z_{jt})} \right]^2 + n_j^{-1} \sum_{t=j+1}^{n} \left[\frac{\bar{f}_j(Z_{jt}) - f_j(Z_{jt})}{f_j(Z_{jt})} \right]^2 \\
&\quad + 2n_j^{-1} \sum_{t=j+1}^{n} \left[\frac{\hat{f}_{jt}(Z_{jt}) - \bar{f}_j(Z_{jt})}{f_j(Z_{jt})} \right] \left[\frac{\bar{f}_j(Z_{jt}) - f_j(Z_{jt})}{f_j(Z_{jt})} \right] \\
&\equiv \hat{W}_{21}(j) + \hat{W}_{22}(j) + 2\hat{W}_{23}(j), \quad \text{say.}
\end{aligned}
$$

We now consider each of the terms in (A3). For the first term $\hat{W}_{21}(j)$ in (A3), we can write

$$
\text{(A4)} \qquad \hat{W}_{21}(j) = \frac{1}{2} n_{j+1}^{-1} \hat{D}_n(j) + \frac{1}{3} n_{j+2} n_{j+1}^{-1} \tilde{H}_{2n}(j),
$$

where $\hat{D}_n(j) \equiv \binom{n_j}{2}^{-1} \sum_{t=j+2}^{n} \sum_{s=j+1}^{t-1} D_{jn}(Z_{jt}, Z_{js})$,

$$
\tilde{H}_{2n}(j) \equiv \binom{n_j}{3}^{-1} \sum_{t=j+3}^{n} \sum_{s=j+2}^{t-1} \sum_{\tau=j+1}^{s-1} \tilde{H}_{2jn}(Z_{jt}, Z_{js}, Z_{j\tau}),
$$

and $D_{jn}(z_1, z_2) \equiv A_{jn}^2(z_1, z_2) + A_{jn}^2(z_2, z_1)$,

$$\tilde{H}_{2jn}(z_1, z_2, z_3) \equiv A_{jn}(z_1, z_2)A_{jn}(z_1, z_3) + A_{jn}(z_2, z_3)A_{jn}(z_2, z_1) + A_{jn}(z_3, z_1)A_{jn}(z_3, z_2).$$

Using Lemma B.1 in Appendix B, a projection theorem for second-order U-statistics of a j-dependent process whose j may grow as $n \to \infty$, we have the following result.

LEMMA A.7: *Suppose Assumptions* A.1 *and* A.2 *hold*, $nh^2 \to \infty$, $h \to 0$, *and* $j = o(n)$. *Then* $\hat{D}_n(j) = 2EA_{jn}^2(Z_3, Z_1) + O_P(n_j^{-1}h^{-3})$ *under* $\mathbb{H}_0$, *where* $Z_3 \equiv (X_3, X_2)'$ *and* $Z_1 \equiv (X_1, X_0)'$.

The second term in (A4), $\tilde{H}_{2n}(j)$, is a third-order U-statistic. Using Lemma B.2 in Appendix B, a projection theorem for third-order dependent U-statistics, we obtain the following result.

LEMMA A.8: *Suppose Assumptions* A.1 *and* A.2 *hold*, $nh^2 \to \infty$, $h \to 0$, *and* $j = o(n)$. *Then* $\tilde{H}_{2n}(j) = 3\hat{H}_{2n}(j) + O_P(n_j^{-3/2}h^{-2})$ *under* $\mathbb{H}_0$, *where* $\hat{H}_{2n}(j)$ *is defined in Theorem* A.1.

This shows that $\tilde{H}_{2n}(j)$ is asymptotically equivalent to the second-order U-statistic $3\hat{H}_{2n}(j)$. We have now dealt with the first term $\hat{W}_{21}(j)$ in (A3).

The second term $\hat{W}_{22}(j)$ in (A3) is related to a bias-squared term, as stated next.

LEMMA A.9: *Suppose Assumptions* A.1 *and* A.2 *hold*, $nh^2 \to \infty$, $h \to 0$, *and* $j = o(n)$. *Then* $\hat{W}_{22}(j) = EB_{jn}^2(Z_1) + O_P(n_j^{-1/2}h^4)$ *under* $\mathbb{H}_0$, *where* $Z_1 = (X_1, X_0)'$.

Finally, for the last term $\hat{W}_{23}(j)$ in (A3), we can have the following result by Lemma B.1.

LEMMA A.10: *Suppose Assumptions* A.1 *and* A.2 *hold*, $nh^2 \to \infty$, $h \to 0$, *and* $j = o(n)$. *Then* $\hat{W}_{23}(j) = \hat{C}_n(j) + O_P(n_j^{-1}h)$ *under* $\mathbb{H}_0$, *where* $\hat{C}_n(j)$ *is defined in Theorem* A.1.

Now, by (A1)–(A4), Lemmas A.5–A.10, $nh^4/\ln n \to \infty$, and $nh^7 \to 0$, we have $2\hat{I}_{jn}(\hat{f}_j, f_j) = -n_{j+1}^{-1}L_n(j) + \hat{H}_n(j) + 2[\hat{B}_n(j) - \hat{C}_n(j)] + o_P(n_j^{-1}h^{-1})$ under $\mathbb{H}_0$. This completes the proof of Theorem A.1. *Q.E.D.*

PROOF OF THEOREM A.2: The proof of Theorem A.2 is similar to and simpler than the proof of Theorem A.1, because only the univariate density estimator $\hat{g}_t(\cdot)$ in (3.4) is involved. *Q.E.D.*

The proof of Theorem 3.1 is completed. *Q.E.D.*

PROOF OF THEOREM 3.2: (a) Because $\hat{\mathcal{I}}_n(j) = O_P(n_j^{-1}h^{-2})$ under $\mathbb{H}_0$ by Theorem 3.1(a) and $d_n^0 = O(h^{-2})$ given (3.14), we have $\hat{\mathcal{I}}_n^2(j) = O_P(n_j^{-2}h^{-4})$. By the inequality that $|\exp(x) - 1 - x| \le 2x^2$ for small $x \in \mathbb{R}$ and $nh^4/\ln n \to \infty$, we have $|\hat{\gamma}_n^2(j) - 2\hat{\mathcal{I}}_n(j)| = |1 - \exp[-2\hat{\mathcal{I}}_n(j)] - 2\hat{\mathcal{I}}_n(j)| \le 8\hat{\mathcal{I}}_n^2(j) = o_P(n_j^{-1}h^{-1})$. Hence, $hn_j\hat{\gamma}_n^2(j) + hd_n^0 = 2hn_j\hat{\mathcal{I}}_n(j) + hd_n^0 + o_P(1) \xrightarrow{d} \mathrm{N}(0, \sigma^2)$ for all $j = o(n)$ by Theorem 3.1(a).

(b) Let $\lambda \in \mathbb{R}^p$ be a nonzero vector such that $\lambda'\lambda = 1$. For any fixed $p \in \mathbb{N}^+$, we have $|\sum_{j=1}^p \lambda_j[hn_j\hat{\gamma}_n^2(j) + hd_n^0] - \sum_{j=1}^n \lambda_j[2hn_j\hat{\mathcal{I}}_n(j) + hd_n^0]| \le 8\sum_{j=1}^p \lambda_j hn_j\hat{\mathcal{I}}_n^2(j) \xrightarrow{p} 0$ given $nh^4/\ln n \to \infty$ and $\max_{1\le j\le p}|\lambda_j| \le 1$. It follows that $\sum_{j=1}^p \lambda_j[hn_j\hat{\gamma}_n^2(j) + hd_n^0] \xrightarrow{d} \mathrm{N}(0, \sigma^2)$ because $\sum_{j=1}^n \lambda_j[2hn_j\hat{\mathcal{I}}_n(j) + hd_n^0] \xrightarrow{d} \mathrm{N}(0, \sigma^2)$, as is shown in the proof of Theorem 3.1(b). The desired result follows by the Cramer–Wold device. *Q.E.D.*

PROOF OF THEOREM 4.1: To show Theorem 4.1, we first state a lemma, which is proven in Appendix B.

LEMMA A.11: *Suppose Assumptions* A.1–A.4 *hold*, $nh^4/\ln n \to \infty$, *and* $nh^7 \to 0$. *Then* (a) $\sup_{x\in\mathbb{I}}|\hat{g}(x) - g(x)| = O_P(n^{-1/2}h^{-1/2}\ln n + h^2)$; (b) $\sup_{(x_1,x_2)\in\mathbb{N}(\delta)}|\hat{g}(x_1) - \hat{g}(x_2)| = O_P(n^{-1/2}h^{-1/2}\ln n + h^2 + \delta)$, *where* $\mathbb{N}(\delta) \equiv \{(x_1, x_2) \in \mathbb{I}^2 : |x_1 - x_2| \le \delta\}$; (c) $\sup_{x\in\mathbb{I}}|E^*[\hat{g}_t^*(x)|\mathcal{X}] - \hat{g}(x)| \le Ch^2 \times \sup_{x\in\mathbb{I}}|\hat{g}^{(2)}(x)|$, *where and throughout* $E^*(\cdot|\mathcal{X})$ *is the expectation with respect to the smoothed bootstrap distribution* $\hat{g}(\cdot)$ *of the resample* $\mathcal{X}^*$ *conditional on the original sample* $\mathcal{X}$; (d) $\sup_{x\in\mathbb{I}}|\hat{g}^{(d)}(x)| = O_P(1)$ *for* $d = 0, 1$ *and* $\sup_{x\in\mathbb{I}}|\hat{g}^{(2)}(x)| = O_P(\ln n)$.

We shall show $P[\mathcal{T}_n^*(j) \le u|\mathcal{X}] \to \Phi(u)$ for all $u \in \mathbb{R}$ with probability approaching 1, where $\Phi(\cdot)$ is the N(0, 1) CDF. Put $\mathcal{E}_1 \equiv \{\sup_{x\in\mathbb{I}}|\hat{g}(x) - g(x)| \le C(n^{-1/2}h^{-1/2}\ln n + h^2), \sup_{x_1,x_2\in\mathbb{N}(\delta)}|\hat{g}(x_1) - \hat{g}(x_2)| \le C(n^{-1/2}h^{-1/2}\ln n + h^2 + \delta)$, where $\mathbb{N}(\delta) \equiv \{(x_1, x_2) \in \mathbb{I}^2 : |x_1 - x_2| \le \delta\}$, $\sup_{x\in\mathbb{I}}|\hat{g}^{(d)}(x)| \le C$ for $d = 0, 1$, and $\sup_{x\in\mathbb{I}}|\hat{g}^{(2)}(x)| \le C\ln n\}$, where C is a large positive constant. Let $\mathcal{E}_2$ be the complement of $\mathcal{E}_1$. Then

$$\begin{aligned} &P[\mathcal{T}_n^*(j) \le u|\mathcal{X}] \\ &\quad = P[\mathcal{T}_n^*(j) \le u|\mathcal{X} \cap \mathcal{E}_1]P(\mathcal{E}_1) + P[\mathcal{T}_n^*(j) \le u|\mathcal{X} \cap \mathcal{E}_2]P(\mathcal{E}_2). \end{aligned}$$

Given Lemma A.11, we have for any $\epsilon > 0$ that there exists a sufficiently large constant C such that $P(\mathcal{E}_2) \le \epsilon$ for n sufficiently large. Therefore, the second probability $P[\mathcal{T}_n^*(j) \le u|\mathcal{X} \cap \mathcal{E}_2]P(\mathcal{E}_2) \le P(\mathcal{E}_2) \le \epsilon$. It suffices to show $P[\mathcal{T}_n^*(j) \le u|\mathcal{X} \cap \mathcal{E}_1] \to \Phi(u)$ for all $u \in \mathbb{R}$.

Let $\hat{g}_t^*(X_t^*)$ and $\hat{f}_{jt}^*(Z_{jt}^*)$ be defined as $\hat{g}_t(X_t)$ and $\hat{f}_{jt}(Z_{jt})$ in (3.4) and (3.5), with $\mathcal{X}^*$ replacing $\mathcal{X}$. Then we can decompose the resampled entropy estimator as

$$\text{(A5)} \qquad \hat{\mathcal{I}}_n^*(j) = n_j^{-1} \sum_{t\in S_n^*(j)} \ln\left[\frac{\hat{f}_j(Z_{jt}^*)}{\hat{g}(X_t^*)\hat{g}(X_{t-j}^*)}\right] + n_j^{-1} \sum_{t\in S_n^*(j)} \ln\left[\frac{\hat{f}_{jt}^*(Z_{jt}^*)}{\hat{f}_j(Z_{jt}^*)}\right]$$
$$- n_j^{-1} \sum_{t\in S_n^*(j)} \ln\left[\frac{\hat{g}_t^*(X_t^*)}{\hat{g}(X_t^*)}\right] - n_j^{-1} \sum_{t\in S_n^*(j)} \ln\left[\frac{\hat{g}_{t-j}^*(X_{t-j}^*)}{\hat{g}(X_{t-j}^*)}\right]$$
$$= \hat{I}_{jn}^*(\hat{f}_j, \hat{g}\circ\hat{g}) + \hat{I}_{jn}^*(\hat{f}_j^*, \hat{f}_j) - \hat{I}_{1nj}^*(\hat{g}^*, \hat{g}) - \hat{I}_{2nj}^*(\hat{g}^*, \hat{g}), \quad \text{say,}$$

where $\hat{f}_j(\cdot)$ is the density of Z_{jt}^* conditional on $\mathcal{X}$. Because conditional on $\mathcal{X}$, $\mathcal{X}^*$ is i.i.d. with marginal density $\hat{g}(\cdot)$, we have $\hat{f}_j(z) = \hat{g}(x)\hat{g}(y)$. It follows that $\hat{\mathcal{I}}_{jn}^*(\hat{f}_j, \hat{g}\circ\hat{g}) = 0$ a.s. conditional on $\mathcal{X}$ and $\mathcal{E}_1$.

Next, we consider the second term $\hat{I}_{jn}^*(\hat{f}_j^*, \hat{f}_j)$ in (A5). We first state two theorems.

THEOREM A.3: *Suppose Assumptions* A.1–A.4 *hold*, $nh^4/\ln n \to \infty$, $nh^7\ln^3 n \to 0$, *and* $j = o(n)$. *Then* $2\hat{I}_{jn}^*(\hat{f}_j^*, \hat{f}_j) = -L_n^*(j) + \hat{H}_n^*(j) + 2[\hat{B}_n^*(j) - \hat{C}_n^*(j)] + o_P(n_j^{-1}h^{-1})$ *conditional on* $\mathcal{X}\cap\mathcal{E}_1$, *where* $L_n^*(j) \equiv n_{j+1}^{-1}E^*[A_{jn}^*(Z_3^*, Z_1^*)^2 | \mathcal{X}\cap\mathcal{E}_1] + E^*[B_{jn}^*(Z_1^*)^2|\mathcal{X}\cap\mathcal{E}_1]$, *and* $\hat{I}_{jn}^*(\hat{f}_j^*, \hat{f}_j)$, $\hat{H}_n^*(j)$, $\hat{B}_n^*(j)$, *and* $\hat{C}_n^*(j)$ *are defined as* $\hat{I}_{jn}(\hat{f}_j, f_j)$, $\hat{H}_n(j)$, $\hat{B}_n(j)$, *and* $\hat{C}_n(j)$, *respectively, with* $\{\hat{f}_j(z) = \hat{g}(x)\hat{g}(y), \mathcal{X}^*\}$ *replacing* $\{f_j(z), \mathcal{X}\}$.

THEOREM A.4: *Suppose Assumptions* A.1–A.4 *hold*, $nh^4/\ln n \to \infty$, $nh^7 \times \ln^3 n \to 0$, *and* $j = o(n)$. *Then for* $i = 1, 2$, $2\hat{I}_{ijn}^*(\hat{g}^*, \hat{g}) = -l_n^*(j) + \hat{V}_{in}^*(j) + 2[\hat{b}_{in}^*(j) - \hat{c}_{in}^*(j)] + o_P(n_j^{-1}h^{-1/2})$ *conditional on* $\mathcal{X}\cap\mathcal{E}_1$, *where* $l_n^*(j) \equiv n_{j+1}^{-1} \times E^*[a_n^*(X_1^*, X_0^*)^2|\mathcal{X}\cap\mathcal{E}_1] + E^*[b_n^*(X_1^*)^2|\mathcal{X}\cap\mathcal{E}_1]$, *and* $\hat{I}_{ijn}^*(\hat{g}^*, \hat{g})$, $\hat{V}_{in}^*(j)$, $\hat{b}_{in}^*(j)$, *and* $\hat{c}_{in}^*(j)$ *are defined as* $\hat{I}_{ijn}(\hat{g}, g)$, $\hat{V}_{in}(j)$, $\hat{b}_{in}(j)$, *and* $\hat{c}_{in}(j)$, *respectively, with* $\{\hat{g}(\cdot), \mathcal{X}^*\}$ *replacing* $\{g(\cdot), \mathcal{X}\}$.

Theorems A.3 and A.4 imply that, conditional on $\mathcal{X}\cap\mathcal{E}_1$,

$$\text{(A6)} \qquad \hat{\mathcal{I}}_n^*(j) = -n_{j+1}^{-1}\big\{E^*[A_{jn}^*(Z_3^*, Z_1^*)^2|\mathcal{X}\cap\mathcal{E}_1]$$
$$- 2E^*[a_{jn}^*(X_1^*, X_0^*)^2|\mathcal{X}\cap\mathcal{E}_1]\big\}$$
$$- \big\{E^*[B_{jn}^*(Z_1^*)^2|\mathcal{X}\cap\mathcal{E}_1] - 2E^*[b_n^*(X_1^*)^2|\mathcal{X}\cap\mathcal{E}_1]\big\}$$

$$+\left\{\hat{H}_n^*(j)-[\hat{V}_{1n}^*(j)+\hat{V}_{2n}^*(j)]\right\}+2[\hat{B}_n^*(j)-\hat{b}_{1n}^*(j)-\hat{b}_{2n}^*(j)]$$
$$-2[\hat{C}_n^*(j)-\hat{c}_{1n}^*(j)-\hat{c}_{2n}^*(j)]+o_P(n_j^{-1}h^{-1}).$$

We now consider each term in (A6). First, observe that under the smoothed bootstrap procedure, we have $\hat{f}_j(z)=\hat{g}(x)\hat{g}(y)$ and

$$\text{(A7)}\qquad \begin{aligned} A_{jn}^*(z_1,z_2)&=a_n^*(x_1,x_2)a_n^*(y_1,y_2)\\ &\quad+a_n^*(x_1,x_2)\frac{\bar{K}_h^*(y_1)}{\hat{g}(y_1)}+\frac{\bar{K}_h^*(x_1)}{\hat{g}(x_1)}a_n^*(y_1,y_2),\end{aligned}$$

where $\bar{K}_h^*(x)\equiv\int_0^1 K_h(x,y)\hat{g}(y)\,dy$. Following reasoning analogous to the proof of Lemma A.1, we have

$$\text{(A8)}\qquad \begin{aligned} &E^*[A_{jn}^*(Z_3^*,Z_1^*)^2|\mathcal{X}\cap\mathcal{E}_1]-2E^*[a_n^*(X_1^*,X_0^*)^2|\mathcal{X}\cap\mathcal{E}_1]\\ &\quad=\left\{E^*[a_n^*(X_3^*,X_2^*)^2|\mathcal{X}\cap\mathcal{E}_1]\right\}^2\\ &\qquad+2E^*[a_n^*(X_1^*,X_0^*)^2|\mathcal{X}\cap\mathcal{E}_1]\left\{E^*\left[\frac{\bar{K}^*(X_1^*)}{\hat{g}(X_1^*)}\Big|\mathcal{X}\cap\mathcal{E}_1\right]^2-1\right\}\\ &\quad=(A_n^0-1)^2+O(1),\end{aligned}$$

where

$$\begin{aligned} &\left\{E^*[a_n^*(X_3^*,X_2^*)^2|\mathcal{X}\cap\mathcal{E}_1]\right\}^2\\ &\quad=E^*\left[\frac{K_h^2(X_2^*,X_1^*)}{\hat{g}^2(X_2^*)}\Big|\mathcal{X}\cap\mathcal{E}_1\right]-\left\{E^*\left[\frac{K_h(X_2^*,X_1^*)}{\hat{g}(X_2^*)}\Big|\mathcal{X}\cap\mathcal{E}_1\right]\right\}^2\\ &\quad=\int_0^1\frac{K_h^2(x_2,x_1)}{\hat{g}(x_2)}\hat{g}(x_1)\,dx_1\,dx_2-[1+O(h)]\\ &\quad=(A_n^0-1)+O(h)\end{aligned}$$

by first-order Taylor series expansions. Note that A_n^0, defined in (3.14), is $O(h^{-1})$.

Next, observe that when $\hat{f}_j(z)=\hat{g}(x)\hat{g}(y)$, we have

$$\text{(A9)}\qquad B_{jn}^*(z)=b_n^*(x)b_n^*(y)+b_n^*(x)+b_n^*(y).$$

It follows from (A9) and the fact that $\sup_{x\in\mathbb{I}}|b_n^*(x)|=O(h^2\ln n)$ conditional on $\mathcal{E}_1$ that

$$\text{(A10)}\qquad \begin{aligned} &E^*[B_{jn}^*(Z_1^*)^2|\mathcal{X}\cap\mathcal{E}_1]-2E^*[b_n^*(X_1^*)^2|\mathcal{X}\cap\mathcal{E}_1]\\ &\quad=2\left\{E^*[b_n^*(X_1^*)|\mathcal{X}\cap\mathcal{E}_1]\right\}^2\end{aligned}$$

$$+4E^*[b_n^*(X_1^*)^2|\mathcal{X}\cap\mathcal{E}_1]E^*[b_n^*(X_1^*)|\mathcal{X}\cap\mathcal{E}_1]$$
$$+\left\{E^*[b_n^*(X_1^*)^2|\mathcal{X}\cap\mathcal{E}_1]\right\}^2$$
$$=2\left\{E^*[b_n^*(X_1^*)|\mathcal{X}\cap\mathcal{E}_1]\right\}^2+O(h^6\ln^3 n).$$

Next, we consider the fourth term in (A6). Using (A9), the i.i.d. property of $\mathcal{X}^*$ given $\mathcal{X}$, and reasoning similar to the proof of Lemma A.7, we have

$$\text{(A11)}\quad \begin{aligned}&\hat{B}_n^*(j)-\hat{b}_{1n}^*(j)-\hat{b}_{2n}^*(j)\\&=\left\{E^*[b_n^*(X_1^*)|\mathcal{X}\cap\mathcal{E}_1]\right\}^2\\&\quad+n_j^{-1}\sum_{t=j+1}^{n}\left[b_n^*(X_t^*)b_n^*(X_{t-j}^*)-\left\{E^*[b_n^*(X_1^*)|\mathcal{X}\cap\mathcal{E}_1]\right\}^2\right]\\&=\left\{E^*[b_n^*(X_1^*)|\mathcal{X}\cap\mathcal{E}_1]\right\}^2+O_P(n_j^{-1/2}h^4\ln^2 n)\end{aligned}$$

by Chebychev's inequality, independence between Z_{jt}^* and Z_{js}^* whenever $t\neq s$, $s\pm j$ conditional on $\mathcal{X}$, and $\sup_{x\in\mathbb{I}}|b_n^*(x)|=O(h^2\ln n)$ conditional on $\mathcal{E}_1$.

Moreover, using (A7), (A9), and reasoning similar to the proof of Lemma A.8, we have

$$\text{(A12)}\quad \hat{C}_n^*(j)-\hat{c}_{1n}^*(j)-\hat{c}_{2n}^*(j)=O_P(n_j^{-1/2}h^4\ln^{3/2}n)$$

conditional on $\mathcal{X}\cap\mathcal{E}_1$.

Collecting (A6), (A8), (A10)–(A12), and the fact that $\hat{V}_{1n}^*(j)+\hat{V}_{2n}^*(j)=O_P(n_j^{-1}h^{-1/2})$ conditional on $\mathcal{X}\cap\mathcal{E}_1$ by Chebyshev's inequality, we obtain $2hn_j\hat{\mathcal{I}}_n^*(j)=-d_n^0+hn_j\hat{H}_n^*(j)+o_P(1)$ conditional on $\mathcal{X}\cap\mathcal{E}_1$. Following reasoning similar to the proof of Theorems A.6–A.9, we can obtain $hn_j\hat{H}_n^*(j)\xrightarrow{d}\mathrm{N}(0,\sigma^2)$ conditional on $\mathcal{X}\cap\mathcal{E}_1$. Here $E^*[(hn_j\hat{H}_n^*(j))^2|\mathcal{X}\cap\mathcal{E}_1]\to\sigma^2$ by using the conditions that $\hat{f}_j(z)=\hat{g}(x)\hat{g}(y)$ and $\hat{g}(\cdot)$ is continuous on $\mathbb{I}$ conditional on $\mathcal{X}\cap\mathcal{E}_1$.

The proof of Theorem 4.1 will be completed provided we show Theorems A.3 and A.4.

PROOF OF THEOREM A.3: Following reasoning analogous to the proof of Lemma A.5, we have

$$\text{(A13)}\quad \hat{I}_{jn}^*(\hat{f}_j^*,\hat{f}_j)=\hat{W}_1^*(j)-\tfrac{1}{2}\hat{W}_2^*(j)+O_P(n_j^{-3/2}h^{-3}\ln n_j+h^6\ln^3 n)$$

conditional on $\mathcal{X}\cap\mathcal{E}_1$, where $\hat{W}_1^*(j)$ and $\hat{W}_2^*(j)$ are as $\hat{W}_1(j)$ and $\hat{W}_2(j)$ in (3.8), respectively, with $\{\hat{f}_j(z)=\hat{g}(x)\hat{g}(y),\mathcal{X}^*\}$ replacing $\{f_j(z),\mathcal{X}\}$. Here, we have used the fact that, conditional on $\mathcal{X} \cap \mathcal{E}_1$,

$$\max_{1\le t\le n}\sup_{z\in\mathbb{I}^2}|\hat{f}_{jt}^*(z)-\hat{f}_j(z)| = O_P(n_j^{-1/2}h^{-1}\ln n_j + h^2\ln n),$$

$$n_j^{-1}\sum_{t=j+1}^{n} E^*\{[\hat{f}_{jt}^*(Z_{jt}^*)-\hat{f}_j(Z_{jt}^*)]^2|\mathcal{X}\cap\mathcal{E}_1\} = O(n_j^{-1}h^{-2}+h^4\ln^2 n),$$

which can be shown by a standard variance-bias argument for kernel estimators (e.g., Fan and Yao (2003)), the i.i.d. properties of $\mathcal{X}^*$ conditional on $\mathcal{X}$, $\hat{f}_j(z)=\hat{g}(x)\hat{g}(y)$, as well as the continuity of $\hat{g}(\cdot)$ and $\sup_{x\in\mathbb{I}}|E^*[\hat{f}_{jt}^*(z)|\mathcal{X}\cap\mathcal{E}_1]-\hat{f}_j(z)| = O(h^2\ln n)$ conditional on $\mathcal{E}_1$.

Now we consider $\hat{W}_1^*(j)$ in (A13). As in treating $\hat{W}_1(j)$, we can write

$$\text{(A14)}\quad \begin{aligned}\hat{W}_1^*(j) &= \frac{1}{2}\binom{n_j}{2}^{-1}\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1}[\tilde{A}_{jn}^*(Z_{jt}^*,Z_{js}^*)+\tilde{A}_{jn}^*(Z_{js}^*,Z_{jt}^*)]\\ &\quad+\frac{1}{2}\binom{n_j}{2}^{-1}\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1}[\gamma_{jn}^*(Z_{jt}^*,Z_{js}^*)+\gamma_{jn}^*(Z_{js}^*,Z_{jt}^*)]\\ &\quad+n_j^{-1}\sum_{t=j+1}^{n}B_{jn}^*(Z_{jt}^*),\\ &=\frac{1}{2}\hat{H}_{1n}^*(j)+\frac{1}{2}\hat{\Gamma}_n^*(j)+\hat{B}_n^*(j),\end{aligned}$$

where $\tilde{A}_{jn}^*(z_1,z_2)$, $\gamma_{jn}^*(z_1,z_2)$, and $B_{jn}^*(z)$ are defined as $A_{jn}(z_1,z_2)$, $\gamma_j(z_1,z_2)$, and $B_{jn}(z)$ in Theorem A.1, respectively, with $\hat{f}_j(z)=\hat{g}(z)\hat{g}(y)$ replacing $f_j(z)$. Note that $E^*[\tilde{A}_{jn}^*(Z_{jt}^*,z_2)|\mathcal{X}] = E^*[\tilde{A}_{jn}^*(z_1,Z_{js}^*)|\mathcal{X}]=0$, thanks to the use of the smoothed bootstrap. In contrast, if the naive bootstrap (i.e., resampling $\mathcal{X}^*$ with replacement from $\mathcal{X}$) were used, we would have

$$E^*[\tilde{A}_{jn}^*(Z_{jt}^*,z_2)|\mathcal{X}] = \int_{\mathbb{I}^2}\tilde{A}_{jn}^*(z,z_2)\,d\hat{F}_{jn}(z)\neq 0.$$

Using reasoning analogous to the proof of Lemma A.6, we have that, with probability one,

$$\text{(A15)}\quad \hat{\Gamma}_n^*(j)=0 \qquad \text{for all } n \text{ sufficiently large}$$

conditional on $\mathcal{X}\cap\mathcal{E}_1$.

Next, we consider $\hat{W}_2^*(j)$ in (A13). As when treating $\hat{W}_2(j)$, we can write

$$\text{(A16)}\quad \hat{W}_2^*(j) = n_j^{-1}\sum_{t=j+1}^{n}\left[\frac{\hat{f}_{jt}^*(Z_{jt}^*) - \bar{f}_j^*(Z_{jt}^*)}{\hat{f}_j(Z_{jt}^*)}\right]^2$$
$$+ n_j^{-1}\sum_{t=j+1}^{n}\left[\frac{\bar{f}_j^*(Z_{jt}^*) - \hat{f}_j(Z_{jt}^*)}{\hat{f}_j(Z_{jt}^*)}\right]^2$$
$$+ 2n_j^{-1}\sum_{t=j+1}^{n}\left[\frac{\hat{f}_{jt}^*(Z_{jt}^*) - \bar{f}_j^*(Z_{jt}^*)}{\hat{f}_j(Z_{jt}^*)}\right]\left[\frac{\bar{f}_j^*(Z_{jt}^*) - \hat{f}_j(Z_{jt}^*)}{\hat{f}_j(Z_{jt}^*)}\right]$$
$$= \hat{W}_{21}^*(j) + \hat{W}_{22}^*(j) + 2\hat{W}_{23}^*(j), \quad \text{say.}$$

For the first term in (A16), we have

$$\text{(A17)}\quad \hat{W}_{21}^*(j) = \frac{1}{2}n_{j+1}^{-1}\hat{D}_n^*(j) + \frac{1}{3}n_{j+2}n_{j+1}^{-1}\tilde{H}_{2n}^*(j),$$

where $\hat{D}_n^*(j)$ and $\tilde{H}_{2n}^*(j)$ are defined as $\hat{D}_n(j)$ and $\tilde{H}_{2n}(j)$ in (A4), respectively, with $\{\hat{f}_j(z) = \hat{g}(x)\hat{g}(y), \mathcal{X}^*\}$ replacing $\{f_j(z), \mathcal{X}\}$. Following reasoning analogous to the proof of Lemma A.7, we have

$$\text{(A18)}\quad \hat{D}_n^*(j) = 2E^*[A_{jn}^*(Z_{jt}^*, Z_{js}^*)^2|\mathcal{X}\cap\mathcal{E}_1] + O_P(n_j^{-1}h^{-3})$$

conditional on $\mathcal{X}\cap\mathcal{E}_1$. Moreover, following reasoning similar to the proof of Lemma A.8, we have

$$\text{(A19)}\quad \tilde{H}_{2n}^*(j) = 3\hat{H}_{2n}^*(j) + O_P(n_j^{-3/2}h^{-2})$$

conditional on $\mathcal{X}\cap\mathcal{E}_1$, where $\hat{H}_{2n}^*(j)$ is defined as $\hat{H}_{2n}(j)$ with $\{\hat{f}_j(z) = \hat{g}(x)\hat{g}(y), \mathcal{X}^*\}$ replacing $\{f_j(z), \mathcal{X}\}$.

For the term $\hat{W}_{22}^*(j)$ in (A16), recalling $B_{jn}^*(z) = [\bar{f}_j^*(z) - \hat{f}_j(z)]/\hat{f}_j(z)$, we have

$$\text{(A20)}\quad \hat{W}_{22}^*(j) = E[B_{jn}^*(Z_{jt}^*)^2|\mathcal{X}\cap\mathcal{E}_1] + O_P(n_j^{-1/2}h^4\ln^2 n)$$

conditional on $\mathcal{X}\cap\mathcal{E}_1$. For $\hat{W}_{23}^*(j)$ in (A16), by reasoning analogous to the proof of Lemma A.10, we have

$$\text{(A21)}\quad \hat{W}_{23}^*(j) = \hat{C}_n^*(j) + O_P(n_j^{-1}h\ln^{1/2} n)$$

conditional on $\mathcal{X}\cap\mathcal{E}_1$, where $\hat{C}_n^*(j)$ is defined as $\hat{C}_n(j)$ with $\{\hat{f}_j(z) = \hat{g}(x)\hat{g}(y), \mathcal{X}^*\}$ replacing $\{f_j(z), \mathcal{X}\}$.

Collecting (A13)–(A21), we obtain $2\hat{I}^*_{jn}(\hat{f}^*_j, \hat{f}_j) = -L^*_n(j) + \hat{H}^*_n(j) + 2[\hat{B}^*_n(j) - \hat{C}^*_n(j)] + o_P(n_j^{-1}h^{-1})$ conditional on $\mathcal{X} \cap \mathcal{E}_1$. This completes the proof of Theorem A.3. *Q.E.D.*

PROOF OF THEOREM A.4: The proof is similar and a bit simpler than the proof of Theorem A.3 because only the univariate "leave-one-out" $\mathcal{X}^*$-based density estimator $\hat{g}^*_t(X^*_t)$ is involved. Note that again, the smoothed bootstrap (but not the naive bootstrap) ensures that $E^*[\tilde{a}^*_n(X^*_t, y)|\mathcal{X}] = E^*[\tilde{a}^*_n(x, X^*_s)|\mathcal{X}] = 0$, where $\tilde{a}^*_n(\cdot,\cdot)$ is defined as $\tilde{a}_n(\cdot,\cdot)$ in Theorem A.2, with $\hat{g}(\cdot)$ replacing $g(\cdot)$. *Q.E.D.*

The proof of Theorem 4.1 is completed. *Q.E.D.*

PROOF OF THEOREM 4.2: (a) We first show $P[\mathcal{T}_n(j) > C_n(j)] \to 1$ for $C_n(j) = o(n_j h)$. Using a variance-bias decomposition technique (i.e., $\hat{g}_t(\cdot) = [\hat{g}_t(\cdot) - E\hat{g}_t(\cdot)] + [E\hat{g}_t(\cdot) - g(\cdot)]$), the mixing condition in Assumption A.4, second-order Taylor series expansions, and the jackknife kernel $k_b(\cdot)$ in (3.3), we can show

$$\text{(A22)} \qquad \sup_{1\le t\le n} E[\hat{g}_t(X_t) - g(X_t)]^2 \to 0,$$

$$\text{(A23)} \qquad \sup_{j<t\le n} E[\hat{f}_{jt}(Z_{jt}) - f_j(Z_{jt})]^2 \to 0,$$

given Assumptions A.1, A.2, and A.4, $nh^2 \to \infty$, $h \to 0$, and $j = o(n)$. It follows from the inequality that $|\ln(1+x)| \le 2|x|$ for small $x \in \mathbb{R}$, $f_j(\cdot) \ge c > 0$, (A23), the Cauchy–Schwarz inequality, and Markov's inequality that $|\hat{I}_{jn}(\hat{f}_j, f_j)| \le n_j^{-1}\sum_{t=j+1}^n |\hat{f}_{jt}(Z_{jt})/f_j(Z_{jt}) - 1| \le c^{-1}n_j^{-1}\sum_{t=j+1}^n |\hat{f}_{jt}(Z_{jt}) - f_j(Z_{jt})| \xrightarrow{p} 0$. Similarly, we have $\hat{I}_{1jn}(\hat{g}, g) + \hat{I}_{2jn}(\hat{g}, g) \xrightarrow{p} 0$ given (A22). Hence, $\hat{\mathcal{I}}_n(j) = \hat{I}_{jn}(f_j, g^2) + o_P(1) = \mathcal{I}(j) + o_P(1)$, where $\hat{I}_{jn}(f_j, g^2) - \mathcal{I}(j) \xrightarrow{p} 0$ by Chebyshev's inequality and the α-mixing condition in Assumption A.4. Thus, $(n_j h)^{-1}\mathcal{T}_n(j) - 2\sigma^{-1}\mathcal{I}(j) \xrightarrow{p} 0$, where $\mathcal{I}(j) \ge c > 0$ whenever $f(\cdot,\cdot) \ne g(\cdot)g(\cdot)$. It follows that $P[\mathcal{T}_n(j) > C_n(j)] \to 1$ for any $C_n(j) = o_P(n_j h)$.

(b) Next, we show $P[\mathcal{T}_n(j) > \mathcal{T}^*_n(j)|\mathcal{X}] \to 1$ with probability approaching 1. Write

$$P[\mathcal{T}_n(j) > \mathcal{T}^*_n(j)|\mathcal{X}] = P[\mathcal{T}_n(j) > \mathcal{T}^*_n(j)|\mathcal{X} \cap \mathcal{E}_1]P(\mathcal{E}_1) + P[\mathcal{T}_n(j) > \mathcal{T}^*_n(j)|\mathcal{X} \cap \mathcal{E}_2]P(\mathcal{E}_2),$$

where $\mathcal{E}_1$ is defined as in the proof of Theorem 4.1, and $\mathcal{E}_2$ is the complement of $\mathcal{E}_1$. Because $P(\mathcal{E}_2) \to 0$ as $C \to \infty$ by Lemma A.11, it suffices to show $P[\mathcal{T}_n(j) > \mathcal{T}^*_n(j)|\mathcal{X} \cap \mathcal{E}_1] \to 1$ under the alternative to $\mathbb{H}_0$. Conditional on $\mathcal{X} \cap \mathcal{E}_1$, we still have $\mathcal{T}_n^*(j) \xrightarrow{d} \mathrm{N}(0,1)$ by Theorem 4.1. This holds regardless of whether $\{X_t\}$ is i.i.d. or α-mixing. Therefore, conditional on $\mathcal{X} \cap \mathcal{E}_1$, we have $\mathcal{T}_n^*(j) = O_P(1)$ and so $P[\mathcal{T}_n(j) > \mathcal{T}_n^*(j) | \mathcal{X} \cap \mathcal{E}_1] \to 1$. This completes the proof of Theorem 4.2. *Q.E.D.*

PROOF OF THEOREM 4.3: (a) Theorem 3.1(b) immediately implies $\mathcal{Q}_n(p) \xrightarrow{d} \mathrm{N}(0,1)$ under $\mathbb{H}_0$. On the other hand, using reasoning analogous to the proof of Theorem 4.1 in combination with the Cramer–Wold device, we can show $[2hn_1\hat{\mathcal{I}}_n^*(1) + hd_n^0, \ldots, 2hn_p\hat{\mathcal{I}}_n^*(p) + hd_n^0]' \xrightarrow{d} \mathrm{N}(\mathbf{0}, \mathbf{I}_p)$ conditional on $\mathcal{X}$. It follows that $\mathcal{Q}_n^*(p) \xrightarrow{d} \mathrm{N}(0,1)$ conditional on $\mathcal{X}$.

(b) The proofs for $P[\mathcal{Q}_n(j) > C_n] \to 1$ for $C_n = o(nh)$ and $P[\mathcal{Q}_n(j) > \mathcal{Q}_n(j)^*] \to 1$ when $f_j(\cdot) \neq g(\cdot)g(\cdot)$ for some $j \in \{1, \ldots, p\}$ are similar to Theorem 4.2. *Q.E.D.*

PROOF OF THEOREM 4.4: (a) Under $\mathbb{H}_0^U$, we have $g(x) = 1$ for all $x \in \mathbb{I}$. It follows that $f_j(z) = g(x)g(y) = 1$ and the bias $B_{jn}(z) = 0$ for all $z \in \mathbb{I}^2$ under $\mathbb{H}_0^U$, where $B_{jn}(\cdot)$ is given in Theorem A.1. Hence, $\hat{B}_n(j) = \hat{C}_n(j) = 0$ a.s. Furthermore, the remainder term in Lemma A.5 becomes $O_P(n_j^{-3/2}h^{-3} \times \ln n_j)$ because the bias $B_{jn}(z) = 0$. Collecting all of these, we obtain $2n_j^{-1} \times \sum_{t\in S_n(j)} \ln \hat{f}_{jt}(Z_{jt}) = 2\hat{I}_{jn}(\hat{f}_j, f_j) = -L_n(j) + \hat{H}_n(j) + O_P(n_j^{-3/2}h^{-3}\ln n_j)$, where, given $B_{jn}(z) = 0$ and $g(x) = 1$,

$$\begin{aligned} L_n(j) &\equiv n_{j+1}^{-1} E A_{jn}^2(Z_3, Z_1) + E B_{jn}^2(Z_1) \\ &= n_{j+1}^{-1}\left\{[Ea_n^2(X_3, X_2) + 1]^2 - 1\right\} \\ &= n_{j+1}^{-1}\left\{h^{-2}\left[(1-2h)\int_{-1}^{1} k^2(u)\,du \right.\right. \\ &\qquad\qquad \left.\left. + 2h\int_0^1\int_{-1}^{b} k_b^2(u)\,du\,db\right]^2 - 1\right\} \\ &= n_{j+1}^{-1}[(A_n^0)^2 - 1] \end{aligned}$$

by change of variables. Therefore, $2h\sum_{t\in S_n(j)} \ln \hat{f}_{jt}(Z_{jt}) + h[(A_n^0)^2 - 1] = hn_j\hat{H}_n(j) + o_P(1)$ given $nh^4/\ln n \to \infty$. It follows that $\mathcal{T}_n^U(j) \xrightarrow{d} \mathrm{N}(0,1)$ under $\mathbb{H}_0^U$ because $hn_j\hat{H}_n(j) \xrightarrow{d} \mathrm{N}(0, \sigma^2)$ under $\mathbb{H}_0$.

(b) The proofs for the asymptotic normality of $\mathcal{T}_n^U(j)^*$ and $\mathcal{Q}_n^U(p)^*$ are similar to the proofs of Theorems 4.1 and 4.3, respectively. Note that the bootstrap population density $\hat{g}(\cdot)$ of $\mathcal{X}^*$ conditional on $\mathcal{X}$ in Theorem 4.1 is now replaced with $\hat{g}(x) = 1$ for all $x \in \mathbb{I}$, because $\mathcal{X}^*$ is now generated from a U[0, 1] distribution. *Q.E.D.*

PROOF OF THEOREM 5.1: Following reasoning analogous to but more tedious than that of Theorem 3.1, we can show that $2hn_j\hat{\mathcal{I}}_n(j) + hd_n^0 = 2hn_j\hat{I}_{jn}(f_j, g \circ g) + hn_j\hat{H}_n(j) + o_P(1)$ under $\mathbb{H}_{jn}(n^{-1/2}h^{-1/2})$. The desired result follows from Theorems A.5–A.9.

The following results hold under the conditions of Theorem 5.1.

THEOREM A.5: *We have* $\sigma^{-1}2hn_j\hat{I}_{jn}(f_j, g \circ g) - \mu_j \xrightarrow{p} 0$, *where* μ_j *is defined in Theorem* 5.1.

THEOREM A.6: *We have* $hn_j[\hat{H}_n(j) - \hat{U}_n(j)] \xrightarrow{p} 0$, *where* $\hat{U}_n(j) = \binom{n_j}{2}^{-1} \times \sum_{t=3j+2}^{n}\sum_{s=j+1}^{t-2j-1} H_{jn}(Z_{jt}, Z_{js})$, *and* $H_{jn}(z_1, z_2) = H_{1jn}(z_1, z_2) - H_{2jn}(z_1, z_2)$, $H_{1jn}(\cdot,\cdot)$ *and* $H_{2jn}(\cdot,\cdot)$ *are as in Theorem* A.1.

THEOREM A.7: *We have* $hn_j[\hat{U}_n(j) - \tilde{U}_n(j)] \xrightarrow{p} 0$, *where* $\tilde{U}_n(j) = \binom{n_j}{2}^{-1} \times \sum_{t=3j+2}^{n}\sum_{s=j+1}^{t-2j-1} U_{jn}(Z_{jt}, Z_{js})$, $U_{jn}(Z_{jt}, Z_{js}) \equiv H_{jn}(Z_{jt}, Z_{js}) - E[H_{jn}(Z_{jt}, Z_s) | \mathcal{F}_{t-1}]$, *and* $\{\mathcal{F}_t\}$ *is the sequence of the sigma fields generated by* $\{X_s, s \le t\}$.

THEOREM A.8: *We have* $\text{var}[hn_j\tilde{U}_n(j)] \to \sigma^2$.

THEOREM A.9: *We have* $hn_j\tilde{U}_n(j) \xrightarrow{d} \text{N}(0, \sigma^2)$.

Theorems A.6–A.9 imply $hn_j\hat{H}_n(j) \xrightarrow{d} \text{N}(0, \sigma^2)$ under $\mathbb{H}_n(n^{-1/2}h^{-1/2})$. Note that $\hat{H}_n(j)$ is a U-statistic of a j-dependent process $\{Z_{jt}\}$ where j is allowed to grow as $n \to \infty$. In addition, the dependence between Z_{jt} and $Z_{j(t-j)}$ never decays to 0 as $j \to \infty$. However, Z_{jt} and Z_{js} are mutually independent if $t \notin \{s, s \pm j, s \pm 2j\}$. We will explore this particular structure in the proofs of Theorems A.5–A.9. We note that Härdle and Horowitz (1993), Härdle and Mammen (1993), and Fan and Li (1996) also considered degenerate U-statistics for independent observations. Hjellvik, Yao, and Tjøstheim (1998) considered degenerate U-statistics for β-mixing dependent processes. These results do not apply to our test statistics, however.

PROOF OF THEOREM A.5: Put $\delta_{jn}(z) \equiv g(x)g(y)/f_j(z) - 1$ and $\xi_{nt}(j) \equiv \ln[1 + \delta_{jn}(Z_{jt})] - E\ln[1 + \delta_{jn}(Z_{jt})]$. Then we can write

$$\text{(A24)} \quad \hat{I}_{jn}(f_j, g \circ g) = -n_j^{-1}\sum_{t=j+1}^{n}\xi_{nt}(j) - n_j^{-1}\sum_{t=j+1}^{n}E\ln[1 + \delta_{jn}(Z_{jt})]$$
$$\equiv -\hat{\xi}_n(j) + E\hat{I}_{jn}(f_j, g \circ g).$$

Under $\mathbb{H}_n(a_n)$, $\{Z_{jt}\}$ is a $2j$-dependent process, and Z_{jt} and Z_{js} are independent if $t \notin \{s, s \pm j, s \pm 2j\}$. Thus, $\{\xi_{nt}(j)\}$ is a zero-mean $2j$-dependent process with $E[\xi_{nt}(j)\xi_{ns}(j)]=0$ if $t\notin\{s, s\pm j, s\pm 2j\}$. It follows that for the first term in (A24),

$$\text{(A25)}\quad E\hat{\xi}_n^2(j) = n_j^{-2}\sum_{t=j+1}^{n}\sum_{s=j+1}^{n} E[\xi_{nt}(j)\xi_{ns}(j)]\mathbb{1}(t\in\{s, s\pm j, s\pm 2j\})$$
$$= O(n_j^{-1}a_n^2)$$

by the Cauchy–Schwarz inequality and $E\xi_{nt}^2(j)\le 4E\delta_{jn}^2(Z_{jt}) = O(a_n^2)$. The latter follows from the inequality that $|\ln(1+x)|\le 2|x|$ for small $x\in\mathbb{R}$ and $E[\xi_{jn}(Z_{jt})]=0$.

Next, we consider the second term $E\hat{I}_{jn}(f_j, g\circ g)$ in (A24). Using the inequality that $|\ln(1+x)-x+\frac{1}{2}x^2|\le|x|^3$ for small $x\in\mathbb{R}$ and $E[\delta_{jn}(Z_{jt})]=0$, we obtain

$$\text{(A26)}\quad 2\sigma^{-1}E\hat{I}_{jn}(f_j, g\circ g)$$
$$= \sigma^{-1}n_j^{-1}\sum_{t=j+1}^{n} E\delta_{nj}^2(Z_{jt}) + O\left[n_j^{-1}\sum_{t=j+1}^{n} E|\delta_{jn}(Z_{jt})|^3\right]$$
$$= a_n^2\mu_j + o(a_n^2),$$

where $a_n^{-2}\sigma^{-1}E[\delta_{jn}^2(Z_{jt})]\to\sigma^{-1}\int_{\mathbb{I}^2} q_j^2(z)g(x)g(y)\,dx\,dy$. The desired results then follow from (A24)–(A26), $a_n=(nh)^{-1/2}$, $h\to 0$, and $j=o(n)$. *Q.E.D.*

PROOF OF THEOREM A.6: Given the definitions of $\hat{H}_n(j)$ and $\hat{U}_n(j)$, we obtain

$$\text{(A27)}\quad \hat{H}_n(j)-\hat{U}_n(j) = \binom{n_j}{2}^{-1}\sum_{t=j+2}^{n}\sum_{s=\max(j+1,t-2j)}^{t-1} H_{jn}(Z_{jt}, Z_{js})$$
$$\equiv \hat{M}_{1n}(j)+\hat{M}_{2n}(j),$$

where

$$\hat{M}_{1n}(j) = \binom{n_j}{2}^{-1}\sum_{t=j+2}^{n}\sum_{s=\max(j+1,t-2j)}^{t-1} H_{jn}(Z_{jt}, Z_{js})\mathbb{1}(t\notin\{s+j, s+2j\}),$$

$$\hat{M}_{2n}(j) = \binom{n_j}{2}^{-1}\sum_{t=j+2}^{n}\sum_{s=\max(j+1,t-2j)}^{t-1} H_{jn}(Z_{jt}, Z_{js})\mathbb{1}(t\in\{s+j, s+2j\}).$$

Recall that Z_{jt} and Z_{js} are independent if $t \notin \{s \pm j, s \pm 2j\}$. Thus, for the term $\hat{M}_{1n}(j)$ in (A27), we have

$$\text{(A28)} \quad E\hat{M}_{1n}^2(j) = \binom{n_j}{2}^{-2} \sum_{t=j+2}^{n} \sum_{s=\max(j+1,t-2j)}^{t-1} \sum_{t'=j+2}^{n} \sum_{s'=\max(j+1,t'-2j)}^{t'-1} E\big[H_{jn}(Z_{jt}, Z_{js}) \times H_{jn}(Z_{jt'}, Z_{js'})\big] \times \mathbb{1}(t \notin \{s+j, s+2j\}, t' \notin \{s'+j, s'+2j\}) \times \mathbb{1}(t, s \in \{t', t' \pm j, t' \pm 2j, s', s' \pm j, s' \pm 2j\}) = O(n_j^{-3} j h^{-2}),$$

where the expectation in the summand is 0 if Z_{jt} is independent of $(Z_{jt'}, Z_{js'})$ (when $t \notin \{t', t' \pm j, t' \pm 2j, s', s' \pm j, s' \pm 2j\}$) or if Z_{js} is independent of $(Z_{jt'}, Z_{js'})$ (when $s \notin \{t', t' \pm j, t' \pm 2j, s', s' \pm j, s' \pm sj\}$). We also used the Cauchy–Schwarz inequality and $EH_{jn}^2(Z_{jt}, Z_{js}) = O(h^{-2})$ for $t > s$, $j > 0$.

Next, we consider the second term $\hat{M}_{2n}(j)$ in (A27), where Z_{jt} and Z_{js} are not independent given $t \in \{s + j, s + 2j\}$. Put $\tilde{U}_{jn}(z_1, z_2) \equiv H_{jn}(z_1, z_2) - EH_{jn}(z_1, z_2)$. Then

$$\text{(A29)} \quad \hat{M}_{2n}(j) = \binom{n_j}{2}^{-1} \sum_{t=j+2}^{n} \sum_{s=\max(j+1,t-2j)}^{t-1} \tilde{U}_{jn}(Z_{jt}, Z_{js}) \mathbb{1}(t \in \{s+j, s+2j\}) + \binom{n_j}{2}^{-1} \sum_{t=j+2}^{n} \sum_{s=\max(j+1,t-2j)}^{t-1} EH_{jn}(Z_{jt}, Z_{js}) \times \mathbb{1}(t \in \{s+j, s+2j\}) \equiv \hat{M}_{21n}(j) + \hat{M}_{22n}(j), \quad \text{say.}$$

The first term $\hat{M}_{21}(j)$ in (A29) essentially consists of two single sums over t, with the summands equal to $\tilde{U}_{jn}(Z_{jt}, Z_{j(t-j)})$ and $\tilde{U}_{jn}(Z_{jt}, Z_{j(t-2j)})$, respectively. Under $\mathbb{H}_{jn}(a_n)$ the summand $\tilde{U}_{jn}(Z_{jt}, Z_{j(t-j)})$ is a $3j$-dependent process with mean 0 and $E[\tilde{U}_{jn}(Z_{jt}, Z_{j(t-j)})\tilde{U}_{jn}(Z_{jt'}, Z_{j(t'-j)})] = 0$ if $t \notin \{t', t' \pm j, t' \pm 2j, t' \pm 3j\}$, which follows from the definition of $\tilde{U}_{jn}(z_1, z_2)$. This implies that for any $j \in \mathbb{N}$, at most only seven pairs of (t, t') yield nonzero expectations. Similarly, the summand $\tilde{U}_{jn}(Z_{jt}, Z_{j(t-2j)})$ is a $4j$-dependent process with mean 0 and $E[\tilde{U}_{jn}(Z_{jt}, Z_{j(t-2j)})\tilde{U}_{jn}(Z_{jt'}, Z_{j(t'-2j)})] = 0$ if $t \notin \{t', t' \pm j, t' \pm 2j, t' \pm 3j, t' \pm 4j\}$. Thus, for any $j \in \mathbb{N}$, at most only nine pairs of (t, t') yield nonzero expectations.

It follows from the Cauchy–Schwarz inequality and $EH_{jn}^2(Z_{jt}, Z_{js}) = O(h^{-2})$ for $t > s$ and $j > 0$ that

$$\text{(A30)} \qquad E\hat{M}_{21n}^2(j) = O(n_j^{-3}h^{-2}).$$

For the second term $\hat{M}_{22n}(j)$ in (A29), we can obtain $EH_{jn}(Z_{jt}, Z_{j(t-j)}) = O(1)$ and $EH_{jn}(Z_{jt}, Z_{j(t-2)}) = O(1)$ under $\mathbb{H}_{jn}(n^{-1/2}h^{-1/2})$ by change of variables. It follows that

$$\text{(A31)} \qquad \hat{M}_{22n}(j) = O(n_j^{-1}).$$

Combining (A27)–(A31), Chebyshev's inequality, and Markov's inequality, we obtain $hn_j[\hat{H}_n(j) - \hat{U}_n(j)] = O_P(n_j^{-1/2}j^{1/2} + h) = o_P(1)$ given $j = o(n)$ and $h \to 0$. *Q.E.D.*

PROOF OF THEOREM A.7: Given the definitions of $\hat{U}_n(j)$ and $\tilde{U}_n(j)$, we obtain $\hat{U}_n(j) - \tilde{U}_n(j) = \binom{n_j}{2}^{-1} \sum_{t=3j+2}^{n} \sum_{s=j+1}^{t-2j-1} E_{t-1}[H_{jn}(Z_{jt}, Z_{js})]$, where Z_{jt} and Z_{js} are independent given $t > s+2j$. Note that for $t > s+2j$, the conditional expectation $E_{t-1}[H_{jn}(Z_{jt}, Z_{js})]$ is a function of (X_{t-j}, Z_{js}) under $\mathbb{H}_{jn}(a_n)$, so we denote $G_{jn}(X_{t-j}, Z_{js}) \equiv E_{t-1}[H_{jn}(Z_{jt}, Z_{js})]$. Observe that $G_{jn}(x, Z_{js})$ is a zero-mean $2j$-dependent process with $E[G_{jn}(x_1, Z_{js_1})G_{jn}(x_2, Z_{js_2})] = 0$ if $s_1 \notin \{s_2, s_2 \pm j, s_2 \pm 2j\}$. Similarly, $G_{jn}(X_{t-j}, z)$ is a zero-mean j-dependent process with mean 0 and $E[G_{jn}(X_{t_1-j}, z_1)G_{jn}(X_{t_2-j}, z_2)] = 0$ if $t_1 \notin \{t_2, t_2 \pm j\}$. It follows that

$$\begin{aligned} &E[\hat{U}_n(j) - \tilde{U}_n(j)]^2 \\ &\quad = \binom{n_j}{2}^{-2} \\ &\qquad \times \sum_{t_1=3j+2}^{n} \sum_{t_2=3j+2}^{n} \sum_{s_1=j+1}^{t_1-2j-1} \sum_{s_2=j+1}^{t_2-2j-1} E\big[G_{jn}(X_{t_1-j}, Z_{js_1})G_{jn}(X_{t_2-j}, Z_{js_2})\big] \\ &\qquad \times \mathbb{1}(t_1 \in \{t_2, t_2 \pm j\}, s_1 \in \{s_2, s_2 \pm j, s_2 \pm 2j\}) \\ &\quad = O(n_j^{-2}h^{-1}), \end{aligned}$$

where the expectation in the summand is 0 if X_{t_1-j} and X_{t_2-j} are independent or if Z_{js_1} and Z_{js_2} are independent. We also used the fact that $EG_{jn}^2(X_{t-j}, Z_{js}) = O(h^{-1})$ for $t > s + 2j$, which is proven below. Therefore, $hn_j[\hat{U}_n(j) - \tilde{U}_n(j)] = O_P(h^{1/2}) = o_P(1)$ by Chebyshev's inequality and $h \to 0$.

It remains to show $EG_{jn}^2(X_{t-j}, Z_{js}) = O(h^{-1})$ for $t > s+2j$. By the definition of $G_{jn}(X_{t-j}, Z_{js})$, we have

$$\text{(A32)}\quad EG_{jn}^2(X_{t-j}, Z_{js}) \leq 2E\{E[H_{1jn}(Z_{jt}, Z_{js})|\mathcal{F}_{t-1}]\}^2 + 2E\{E[H_{2jn}(Z_{jt}, Z_{js})|\mathcal{F}_{t-1}]\}^2.$$

For the first term in (A32), we have that, for all n sufficiently large,

$$\begin{aligned} &E[H_{1jn}(Z_{jt}, Z_{js})|\mathcal{F}_{t-1}] \\ &= E[\tilde{A}_{jn}(Z_{jt}, Z_{js})|\mathcal{F}_{t-1}] + E[\tilde{A}_{jn}(Z_{js}, Z_{jt})|\mathcal{F}_{t-1}] \\ &= a_n(X_{t-j}, X_{s-j}) \\ &\quad + a_n(X_{t-j}, X_{s-j}) \frac{\int_0^1 K_h(x, X_s) g_j(x|X_{s-j})\, dx}{g_j(X_s|X_{s-j})}, \end{aligned}$$

where $g_j(x|y) \equiv g_{jn}(x|y)$ is the conditional density of $X_t = x$ given $X_{t-j} = y$. Under $\mathbb{H}_{jn}(a_n)$ and Assumption A.5, $\int_0^1 K_h(x_1, x_2) g_j(x_1|y_2)\, dx_1 / g_j(x_2|y_2) \to 1$ uniformly in $(x_2, y_2) \in \mathbb{I}^2$. It follows that

$$\text{(A33)}\quad \begin{aligned} E\{E[H_{1jn}(Z_{jt}, Z_{js})|\mathcal{F}_{t-1}]\}^2 &\leq CEa_n^2(X_{t-j}, X_{s-j}) \\ &= Ch^{-1} \int_{-1}^1 k^2(u)\, du[1+o(1)] \end{aligned}$$

for n sufficiently large. Similarly, we have for all n sufficiently large

$$\text{(A34)}\quad \begin{aligned} &E\{E[H_{2jn}(Z_{jt}, Z_{js})|\mathcal{F}_{t-1}]\}^2 \\ &\leq Ch^{-1} \int_{-1}^1 \left[\int_{-1}^1 k(v)k(u+v)\, dv\right]^2 du[1+o(1)]. \end{aligned}$$

Combining (A32)–(A34) yields $EG_{jn}^2(X_{t-j}, Z_{js}) = O(h^{-1})$. This completes the proof. *Q.E.D.*

PROOF OF THEOREM A.8: Write

$$\tilde{U}_n(j) = \binom{n_j}{2}^{-1} \sum_{t=3j+1}^{n} \tilde{U}_{nt}(j) \quad \text{and} \quad \tilde{U}_{nt}(j) \equiv \sum_{s=j+1}^{t-2j-1} U_{jn}(Z_{jt}, Z_{js}).$$

Because $\{\tilde{U}_{nt}(j), \mathcal{F}_{t-1}\}$ is a martingale difference sequence (m.d.s.), we have $E[\tilde{U}_n^2(j)] = \binom{n_j}{2}^{-2} \sum_{t=3j+2}^{n} E\tilde{U}_{nt}^2(j)$, where

$$(\text{A35}) \qquad E\tilde{U}_{nt}^2(j) = \sum_{s_2=j+1}^{t-2j-1} \sum_{s_1=j+1}^{t-2j-1} E[U_{jn}(Z_{jt}, Z_{js_1})U_{jn}(Z_{jt}, Z_{js_2})] \times \mathbb{1}(s_1 \in \{s_2, s_2 \pm j, s_2 \pm 2j\}),$$

where the expectation in the summand is 0 if $s_1 \notin \{s_2, s_2 \pm j, s_2 \pm 2j\}$.

We first consider the term with $s_1 = s_2$. Recalling the definition of $U_{jn}(\cdot, \cdot)$ and using the law of iterated expectations, we have for $t > s + 2j$,

$$(\text{A36}) \qquad EU_{jn}^2(Z_{jt}, Z_{js}) = EH_{jn}^2(Z_{jt}, Z_{js}) - EG_{jn}^2(X_{t-j}, Z_{js}).$$

For the first term in (A36), we have

$$(\text{A37}) \qquad EH_{jn}^2(Z_{jt}, Z_{js}) = EH_{1jn}^2(Z_{jt}, Z_{js}) + EH_{2jn}^2(Z_{jt}, Z_{js}) - 2E[H_{1jn}(Z_{jt}, Z_{js})H_{2jn}(Z_{jt}, Z_{js})].$$

By straightforward algebra (mainly change of variables), we have for $t > s+2j$,

$$(\text{A38}) \qquad EH_{1jn}^2(Z_{jt}, Z_{js}) = 4\int_{\mathbb{I}^4} \tilde{A}_{jn}^2(z_1, z_2)f_j(z_1)f_j(z_2)\,dz_1\,dz_2 = 4h^{-2}\left[\int_{-1}^{1} k^2(u)\,du\right]^2[1+o(1)],$$

$$(\text{A39}) \qquad EH_{2jn}^2(Z_{jt}, Z_{js}) = \int_{\mathbb{I}^4}\left[\int_{\mathbb{I}^2} A_{jn}(z, z_1)A_{jn}(z, z_2)f_j(z)\,dz\right]^2 f_j(z_1)f_j(z_2)\,dz_1\,dz_2 = h^{-2}\left\{\int_{-1}^{1}\left[\int_{-1}^{1} k(v)k(u+v)\,dv\right]^2 du\right\}^2[1+o(1)],$$

$$(\text{A40}) \qquad E[H_{1jn}(Z_{jt}, Z_{js})H_{2jn}(Z_{jt}, Z_{js})] = 2\int_{\mathbb{I}^6} \tilde{A}_{jn}(z_1, z_2)A_{jn}(z, z_1)A_{jn}(z, z_2)f_j(z)f_j(z_1)f_j(z_2)\,dz_1\,dz_2\,dz = 2h^{-2}\left[\int_{-1}^{1}\int_{-1}^{1} k(u)k(v)k(u+v)\,du\,dv\right]^2[1+o(1)].$$

Note that the boundary correction due to the use of the jackknife kernel $k_b(\cdot)$ contributes to the negligible $o(1)$ terms in (A38)–(A40).

Combining (A36)–(A40) and $EG_{jn}^2(X_{t-j}, Z_{js}) = O(h^{-1})$ for $t > s + 2j$, we have

$$\text{(A41)} \quad h^2 EU_{jn}^2(Z_{jt}, Z_{js})$$
$$\to 4\left[\int_{-1}^{1} k^2(u)\,du\right]^2 + \left[\int_{-1}^{1}\left[\int_{-1}^{1} k(v)k(u+v)\,dv\right]^2 du\right]^2$$
$$- 4\left[\int_{-1}^{1}\int_{-1}^{1} k(u)k(v)k(u+v)\,du\,dv\right]^2$$
$$= \int_{-1}^{1}\int_{-1}^{1}\Big[2k(u)k(u')$$
$$- \int_{-1}^{1} k(u+v)k(v)\,dv \int_{-1}^{1} k(u'+v')k(v')\,dv'\Big]^2 du\,du'$$
$$= \frac{\sigma^2}{2}.$$

By analogous reasoning, we can obtain that under $\mathbb{H}_n(a_n)$ and $s_1, s_2 < t - 2j$,

$$\text{(A42)} \quad h^2 E\big[U_{jn}(Z_{jt}, Z_{js_1}), U_{jn}(Z_{jt}, Z_{js_2})\big] = O(a_n) \quad \text{if } s_1 \in \{s_2 \pm j, s_2 \pm 2j\}.$$

It follows from (A35), (A41)–(A42), and $j = o(n)$ that $E[n_j h\tilde{U}_n(j)]^2 \to \sigma^2$. *Q.E.D.*

PROOF OF THEOREM A.9: Let $\tilde{U}_{nt}(j)$ be defined as in the proof of Theorem A.8. Because $\{\tilde{U}_{nt}(j), \mathcal{F}_{t-1}\}$ is an m.d.s., we use Brown's (1971) martingale theorem, which states that $\text{var}[\tilde{U}_n(j)]^{-1/2}\tilde{U}_n(j) \xrightarrow{d} \text{N}(0, 1)$ if

$$\text{(A43)} \quad \text{var}^{-1}[\tilde{U}_n(j)]\binom{n_j}{2}^{-2}$$
$$\times \sum_{t=3j+2}^{n} E\big\{\tilde{U}_{nt}^2(j)\mathbb{1}[|\tilde{U}_{nt}(j)| > \epsilon n_j^2 \text{var}^{1/2}[\tilde{U}_n(j)]]\big\} \to 0 \quad \forall \epsilon > 0,$$

and

$$\text{(A44)} \quad \text{var}^{-1}[\tilde{U}_n(j)]\binom{n_j}{2}^{-2} \sum_{t=3j+2}^{n} E[\tilde{U}_{nt}^2(j)|\mathcal{F}_{t-1}] \xrightarrow{p} 1.$$

We now verify these two conditions. Given Theorem A.8, we can verify condition (A43) by showing that $(hn_j)^4\binom{n_j}{2}^{-4}\sum_{t=3j+2}^{n} E\tilde{U}_{nt}^4(j) \to 0$. Observe that

$$\text{(A45)} \quad \tilde{U}_{nt}^4(j) \leq 2\Bigg[\sum_{s_1=j+1}^{t-2j-1}\sum_{s_2=j+1}^{t-2j-1} U_{jn}(Z_{jt}, Z_{js_1})U_{jn}(Z_{jt}, Z_{js_2}) \times \mathbb{1}(s_1 \in \{s_2, s_2 \pm j, s_2 \pm 2j\})\Bigg]^2$$

$$+ 2\Bigg[\sum_{s_1=j+1}^{t-2j-1}\sum_{s_2=j+1}^{t-2j-1} U_{jn}(Z_{jt}, Z_{js_1})U_{jn}(Z_{jt}, Z_{js_2}) \times \mathbb{1}(s_1 \notin \{s_2, s_2 \pm j, s_2 \pm 2j\})\Bigg]^2$$

$$\equiv 2\tilde{U}_{1nt}^4(j) + 2\tilde{U}_{2nt}^4(j), \quad \text{say.}$$

For the first term $\tilde{U}_{1nt}^4(j)$ in (A45), we have

$$\text{(A46)} \quad E\tilde{U}_{1nt}^4(j) \leq \Bigg\{\sum_{s_1=j+1}^{t-2j-1}\sum_{s_2=j+1}^{t-2j-1} \big[EU_{jn}^4(Z_{jt}, Z_{js_1})EU_{jn}^4(Z_{jt}, Z_{js_2})\big]^{1/2} \times \mathbb{1}(s_1 \in \{s_2, s_2 \pm j, s_2 \pm 2j\})\Bigg\}^2$$

$$= O[(t-2j)^2h^{-6}]$$

by Minkowski's inequality and $EU_{jn}^4(Z_{jt}, Z_{js}) \leq EH_{jn}^4(Z_{jt}, Z_{js}) = O(h^{-6})$ for $t > s + 2j$, which follows by change of variables.

For the term $\tilde{U}_{2nt}^4(j)$ in (A45), where Z_{js_1} and Z_{js_2} are independent given $s_1 \notin \{s_2, s_2 \pm j, s_2 \pm 2j\}$, we have

$$\text{(A47)} \quad E[\tilde{U}_{2nt}^4(j)] = \sum_{s_1=j+1}^{t-2j-1}\sum_{s_2=j+1}^{t-2j-1}\sum_{s_1'=j+1}^{t-2j-1}\sum_{s_2'=j+1}^{t-2j-1} E\big[U_{jn}(Z_{jt}, Z_{js_1})U_{jn}(Z_{jt}, Z_{js_2}) \times U_{jn}(Z_{jt}, Z_{js_1'})U_{jn}(Z_{jt}, Z_{js_2'})\big]$$

$$\times \mathbb{1}(s_1 \notin \{s_2, s_2 \pm j, s_2 \pm 2j\}, s_1' \notin \{s_2', s_2' \pm j, s_2' \pm 2j\})$$

$$\times \mathbb{1}(s_1, s_2 \in \{s_1', s_1' \pm j, s_1' \pm 2j, s_2', s_2' \pm j, s_2' \pm 2j\})$$

$$= O[(t-2j)^2h^{-6}],$$

where the expectation in the summand is 0 if Z_{js_1} is independent of $(Z_{js'_1}, Z_{js'_2})$ or if Z_{js_2} is independent of $(Z_{js'_1}, Z_{js'_2})$. We also made use of the Cauchy–Schwarz inequality and $EH_n^4(Z_{jt}, Z_{js}) = O(h^{-6})$ for $t > s + 2j$ by change of variables. Thus, from (A45)–(A47), $nh^4/\ln n \to 0$, and $j = o(n)$, we have $(hn_j)^4 \binom{n_j}{2}^{-4} \sum_{t=3j+2}^{n} E\tilde{U}_{nt}^4(j) = O(n_j^{-1}h^{-2}) \to 0$. This ensures that condition (A43) holds.

We now verify (A44) by showing $(hn_j)^4 E\{\binom{n_j}{2}^{-2} \sum_{t=3j+2}^{n} E[\tilde{U}_{nt}^2(j)|\mathcal{F}_{t-1}] - E\tilde{U}_{nt}^2(j)\}^2 \to 0$. We write

$$\text{(A48)}\quad \begin{aligned} E_{t-1}[\tilde{U}_{nt}^2(j)] &= \sum_{s_1=j+1}^{t-2j-1} \sum_{s_2=j+1}^{t-2j-1} E_{t-1}\big[U_{jn}(Z_{jt}, Z_{js_1})U_{jn}(Z_{jt}, Z_{js_2})\big] \\ &\quad \times \mathbb{1}(s_1 \in \{s_2, s_2 \pm j, s_2 \pm 2j\}) \\ &\quad + \sum_{s_2=j+1}^{t-2j-1} \sum_{s_1=j+1}^{t-2j-1} E_{t-1}\big[U_{jn}(Z_{jt}, Z_{js_1})U_{jn}(Z_{jt}, Z_{js_2})\big] \\ &\quad \times \mathbb{1}(s_1 \notin \{s_2, s_2 \pm j, s_2 \pm 2j\}) \\ &\equiv \hat{M}_{3nt}(j) + Q_{1nt}(j), \quad \text{say,} \end{aligned}$$

where $EQ_{1nt}(j) = 0$ because Z_{jt}, Z_{js_1}, and Z_{js_2} are mutually independent. Put

$$\tilde{\Omega}_{jn}(X_{t-j}, z_1, z_2) \equiv E_{t-1}[U_{jn}(Z_{jt}, z_1)U_{jn}(Z_{jt}, z_2)] - \Omega_{jn}(z_1, z_2),$$

where $\Omega_{jn}(z_1, z_2) \equiv E[U_{jn}(Z_{jt}, z_1)U_{jn}(Z_{jt}, z_2)]$. Then we can write

$$\text{(A49)}\quad \hat{M}_{3nt}(j) = Q_{2nt}(j) + Q_{3nt}(j) + Q_{4nt}(j),$$

where

$$\begin{aligned} Q_{2nt}(j) &\equiv \sum_{s_1=j+1}^{t-2j-1} \sum_{s_2=j+1}^{t-2j-1} \tilde{\Omega}_{jn}(X_{t-j}, Z_{js_1}, Z_{js_2})\mathbb{1}(s_1 \in \{s_2, s_2 \pm j, s_2 \pm 2j\}), \\ Q_{3nt}(j) &\equiv \sum_{s_1=j+1}^{t-2j-1} \sum_{s_2=j+1}^{t-2j-1} \big[\Omega_{jn}(Z_{js_1}, Z_{js_2}) - E\Omega_{jn}(Z_{js_1}, Z_{js_2})\big] \\ &\qquad \times \mathbb{1}(s_1 \in \{s_2, s_2 \pm j, s_2 \pm 2j\}), \\ Q_{4nt}(j) &\equiv \sum_{s_1=j+1}^{t-2j-1} \sum_{s_2=j+1}^{t-2j-1} E\Omega_{jn}(Z_{js_1}, Z_{js_2})\mathbb{1}(s_1 \in \{s_2, s_2 \pm j, s_2 \pm 2j\}). \end{aligned}$$

Because $EQ_{cnt}(j)=0$ for $c=1, 2, 3$, we have $E\tilde{U}_{nt}^2(j)=Q_{4nt}(j)$. This, (A48), and (A49) imply $E_{t-1}[\tilde{U}_{nt}^2(j)]-E\tilde{U}_{nt}^2(j)=\sum_{c=1}^3 Q_{cnt}(j)$. It follows that

$$\text{(A50)}\quad E\left[\sum_{t=3j+2}^{n}\left\{E_{t-1}[\tilde{U}_{nt}^2(j)]-E\tilde{U}_{nt}^2(j)\right\}\right]^2 \le 16\sum_{c=1}^{3}E\left[\sum_{t=3j+2}^{n}Q_{cnt}(j)\right]^2.$$

For the first term Q_{1jnt} in (A48), we have

$$\begin{aligned}\text{(A51)}\quad & EQ_{1nt}^2(j)\\ &= E\sum_{s_1=j+1}^{t-2j-1}\sum_{s_2=j+1}^{t-2j-1}\sum_{s_1'=j+1}^{t-2j-1}\sum_{s_2'=j+1}^{t-2j-1}E_{t-1}\big[U_{jn}(Z_{jt},Z_{js_1})U_{jn}(Z_{jt},Z_{js_2})\big]\\ &\quad\times E_{t-1}\big[U_{jn}(Z_{jt},Z_{js_1'})U_{jn}(Z_{jt},Z_{js_2'})\big]\\ &\quad\times \mathbb{1}(s_1\notin\{s_2,s_2\pm j,s_2\pm 2j\},s_1'\notin\{s_2',s_2'\pm j,s_2'\pm 2j\})\\ &\quad\times \mathbb{1}(s_1,s_2\in\{s_1',s_1'\pm j,s_1'\pm 2j,s_2',s_2'\pm j,s_2'\pm 2j\})\\ &= O[(t-2j)^2h^{-3}],\end{aligned}$$

where the expectation in the summand is 0 if Z_{js_1} is independent of $(Z_{js_1'},Z_{js_2'})$ or if Z_{js_2} is independent of $(Z_{js_1'},Z_{js_2'})$. We also have used the fact that $E\{E_{t-1}[U_{jn}(Z_{jt},Z_{js_1})U_{jn}(Z_{jt},Z_{js_2})]\}^2=O(h^{-3})$, where Z_{jt}, Z_{js_1}, and Z_{js_2} are mutually independent, which follows by change of variables.

Next, we consider the term $Q_{2nt}(j)$ in (A49). Noting that $\{\tilde{\Omega}_{jn}(X_{t-j},z_1,z_2)\}$ is a j-dependent process with mean 0 and $E[\tilde{\Omega}_{jn}(X_{t-j},z_1,z_2)\tilde{\Omega}_{jn}(X_{t'-j},z_1,z_2)]=0$ if $t\notin\{t',t'\pm j\}$, we have, by Minkowski's inequality,

$$\begin{aligned}\text{(A52)}\quad & E\left[\sum_{t=3j+2}^{n}Q_{2nt}(j)\right]^2\\ &\le\left\{\sum_{s_1=j+1}^{n-2j-1}\sum_{s_2=j+1}^{n-2j-1}\left[E\left(\sum_{t=\min(s_1,s_2)+2j+1}^{n}\tilde{\Omega}_{jn}(X_{t-j},Z_{js_1},Z_{js_2})\right)^2\right]^{1/2}\right.\\ &\qquad\left.\times\,\mathbb{1}(s_1\in\{s_2,s_2\pm j,s_2\pm 2j\})\right\}^2\\ &= O(n_j^3h^{-6}),\end{aligned}$$

where

$$E\left[\sum_{t=\min(s_1,s_2)+2j+1}^{n} \tilde{\Omega}_{jn}(X_{t-j}, Z_{js_1}, Z_{js_2})\right]^2$$
$$= \sum_{t=\min(s_1,s_2)+2j+1}^{n} \sum_{t'=\min(s_1,s_2)+2j+1}^{n} E\big[\tilde{\Omega}_{jn}(X_{t-j}, Z_{js_1}, Z_{js_2}) \times \tilde{\Omega}_{jn}(X_{t'-j}, Z_{js_1}, Z_{js_2})\big]$$
$$\times \mathbb{1}(t \in \{t', t' \pm j\})$$
$$= O[(n-2j)h^{-6}],$$

where the expectation in the summand is 0 if X_{t-j} and $X_{t'-j}$ are independent. We also used the fact that $E\tilde{\Omega}^2_{jn}(X_{t-j}, Z_{js_1}, Z_{js_2}) = O(h^{-6})$ by the Cauchy–Schwarz inequality, Jensen's inequality, and $EH^4_{jn}(Z_{jt}, Z_{js}) = O(h^{-6})$ for $t > s + 2j$.

Finally, observe that $Q_{3nt}(j)$ in (A49) essentially consists of five single sums over s_1 with the summands equal to $\Omega_{jn}(Z_{js_1}, Z_{js_2}) - E\Omega_{jn}(Z_{js_1}, Z_{js_2})$, where $s_2 = s_1, s_1 \pm j, s_1 \pm 2j$, respectively. This has a structure similar to that of $\hat{M}_{21n}(j)$ in (A29). Thus, following reasoning analogous to that ((A30) and related argument) for $E\hat{M}^2_{21n}(j)$, we can obtain $EQ^2_{3nt}(j) = O[(t-2j)h^{-6}]$, where we make use of the fact that $E\tilde{\Omega}^2_{nt}(j) = O(h^{-6})$. It follows by Minkowski's inequality that

$$\text{(A53)} \quad E\left[\sum_{t=3j+2}^{n} Q_{3nt}(j)\right]^2 \le E\left\{\sum_{t=3j+2}^{n} [EQ^2_{3nt}(j)]^{1/2}\right\}^2 = O(n_j^3 h^{-6}).$$

Combining (A50)–(A53) yields $(hn_j)^4\binom{n_j}{2}^{-4} E\{\sum_{t=3j+2}^{n}[E_{t-1}[\tilde{U}^2_{nt}(j)] - E\tilde{U}^2_{nt}(j)]\}^2 = O(h + n_j^{-1}h^{-2}) \to 0$ given $h \to 0$, $nh^4/\ln n \to 0$, and $j = o(n)$. Thus, (A44) holds, so that $n_j h\tilde{U}_n(j) \xrightarrow{d} \mathrm{N}(0, \sigma^2)$ by Brown's theorem. *Q.E.D.*

The proof of Theorem 5.1 is completed. *Q.E.D.*

APPENDIX B: PROOFS OF TECHNICAL LEMMAS

LEMMA B.1: *Put $Z_{jt} \equiv (X_t, X_{t-j})'$, where $j = o(n)$ and $\{X_t\}$ is i.i.d. with CDF $G(\cdot)$. Consider a second-order U-statistic*

$$\hat{\phi}_n(j) \equiv \binom{n_j}{2}^{-1} \sum_{t=j+2}^{n} \sum_{s=j+1}^{t-1} \phi_{jn}(Z_{jt}, Z_{js}),$$

where $\phi_{jn}(\cdot,\cdot)$ *is a kernel function such that* $\phi_{jn}(z_1,z_2)=\phi_{jn}(z_2,z_1)$ *and* $E\phi_{jn}^2(Z_{jt},Z_{js})=O(c_{n_j}^2)$. *Put* $\phi_{jn0}\equiv\int_{\mathbb{I}^2}\phi_{jn}(z_1,z_2)\,dF_j(z_1)\,dF_j(z_2)$ *and* $\phi_{jn1}(z)\equiv\int_{\mathbb{I}^2}\phi_{jn}(z,z')\,dF_j(z')$, *where* $F_j(z)=G(x)G(y)$ *and* $z=(x,y)'$. *Then* $\hat{\phi}_n(j)=\phi_{jn0}+2n_j^{-1}\sum_{t=j+1}^{n}[\phi_{jn1}(Z_{jt})-\phi_{jn0}]+O_P(n_j^{-1}c_{n_j})$. *If in addition* $E\phi_{jn1}^2(Z_{jt})\le C$ *and* $c_{n_j}=O(n_j)$, *then* $\hat{\phi}_n(j)=\phi_{jn0}+O_P(n_j^{-1/2})$.

PROOF: Put $\tilde{\phi}_{jn}(z_1,z_2)\equiv\phi_{jn}(z_1,z_2)-\phi_{jn1}(z_1)-\phi_{jn1}(z_2)+\phi_{jn0}$. Then

$$\int_{\mathbb{I}^2}\tilde{\phi}_{jn}(z_1,z)\,dF_j(z)=\int_{\mathbb{I}^2}\tilde{\phi}_{jn}(z,z_2)\,dF_j(z)=0\quad\forall z_1,z_2\in\mathbb{I}^2.\tag{B1}$$

By straightforward algebra, we can write

$$\begin{aligned}\hat{\phi}_n(j)&=\phi_{jn0}+2n_j^{-1}\sum_{t=j+1}^{n}[\phi_{jn1}(Z_{jt})-\phi_{jn0}]\\&\quad+\binom{n_j}{2}^{-1}\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1}\tilde{\phi}_{jn}(Z_{jt},Z_{js})\\&=\phi_{jn0}+2n_j^{-1}\sum_{t=j+1}^{n}[\phi_{jn1}(Z_{jt})-\phi_{jn0}]+\tilde{\phi}_n(j),\quad\text{say.}\end{aligned}\tag{B2}$$

For the last term $\tilde{\phi}_n(j)$ in (B2), we write

$$\begin{aligned}\tilde{\phi}_n(j)&=\binom{n_j}{2}^{-1}\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1}\tilde{\phi}_{jn}(Z_{jt},Z_{js})\mathbb{1}(t\neq s+j)\\&\quad+\binom{n_j}{2}^{-1}\sum_{t=j+2}^{n}\tilde{\phi}_{jn}\big(Z_{jt},Z_{j(t-j)}\big)\\&=\tilde{\phi}_{n1}(j)+\tilde{\phi}_{n2}(j),\quad\text{say,}\end{aligned}\tag{B3}$$

where the second term $\tilde{\phi}_{n2}(j)$ corresponds to $t=s+j$. We have $E|\tilde{\phi}_{n2}(j)|=O(n_j^{-1}c_{n_j})$ by the Cauchy–Schwarz inequality and $E\phi_{jn}^2(Z_{jt},Z_{js})=O(c_{n_j}^2)$. For the first term $\tilde{\phi}_{1n}(j)$, observing that Z_{jt} and Z_{js} are independent given $t\neq s+j$, we obtain

$$\begin{aligned}E\tilde{\phi}_{n1}^2(j)&=\binom{n_j}{2}^{-2}\sum_{t=j+2}^{n}\sum_{s=j+1}^{t-1}\sum_{t'=j+2}^{n}\sum_{s'=j+1}^{t'-1}E[\tilde{\phi}_{jn}(Z_{jt},Z_{js})\tilde{\phi}_{jn}(Z_{jt'},Z_{js'})]\\&\quad\times\mathbb{1}(t\neq s+j,t'\neq s'+j)\end{aligned}$$

$$\times \mathbb{1}(t, s \in \{t', t' \pm j, s', s' \pm j\})$$
$$= O(n_j^{-2} c_{n_j}^2),$$

where the last equality follows from (B1) and the Cauchy–Schwarz inequality. It follows that $\tilde{\phi}_n(j) = O_P(n_j^{-1} c_{n_j})$ by (B3), Chebyshev's inequality, and Markov's inequality. Hence, we have $\hat{\phi}_n(j) = \phi_{jn0} + 2n_j^{-1} \sum_{t=j+1}^{n} [\phi_{jn1}(Z_{jt}) - \phi_{jn0}] + O_P(n_j^{-1} c_{n_j})$.

Next, noting that $\phi_{jn1}(Z_{jt}) - \phi_{jn0}$ is a j-dependent process with mean 0, $E[\phi_{jn1}(Z_{jt}) - \phi_{jn0}]^2 \leq C$, and $E\{[\phi_{jn1}(Z_{jt}) - \phi_{jn0}][\phi_{jn1}(Z_{js}) - \phi_{jn0}]\} = 0$ for $t \notin \{s, s \pm j\}$, we have $E\{n_j^{-1} \sum_{t=j+1}^{n} [\phi_{jn1}(Z_{jt}) - \phi_{jn0}]\}^2 = O(n_j^{-1})$ because only the summands with $t = s$, $s \pm j$ are nonzero. Thus, we have $\hat{\phi}_n(j) = \phi_n(j) + O_P(n_j^{-1/2})$ by Chebyshev's inequality. This completes the proof. *Q.E.D.*

LEMMA B.2: *Put $Z_{jt} \equiv (X_t, X_{t-j})'$, where $j = o(n)$ and $\{X_t\}$ is i.i.d. with CDF $G(\cdot)$. Consider a third-order U-statistic*

$$\hat{\phi}_n(j) \equiv \binom{n_j}{3}^{-1} \sum_{t=j+3}^{n} \sum_{s=j+2}^{t-1} \sum_{r=j+1}^{s-1} \phi_{jn}(Z_{jt}, Z_{js}, Z_{jr}),$$

where $\phi_{jn}(\cdot, \cdot, \cdot)$ is a kernel symmetric in its arguments, with $E\phi_{jn}^2(Z_{jt}, Z_{js}, Z_{jr}) = O(c_{jn}^2)$ and

$$\text{(B4)} \qquad \int_{\mathbb{I}^4} \phi_{jn}(z_1, z_2, z_3)\, dF_j(z_2)\, dF_j(z_3) = 0 \quad \forall\, z_1 \in \mathbb{I}^2,$$

where $F_j(z) = G(x)G(y)$ and $z = (x, y)'$. Put $\phi_{jn2}(z_1, z_2) \equiv \int_{\mathbb{I}^2} \phi_{jn}(z_1, z_2, z_3)\, dF_j(z_3)$. Then we have $\hat{\phi}_n(j) = 3\binom{n_j}{2}^{-1} \sum_{t=j+2}^{n} \sum_{s=j+1}^{t-1} \phi_{jn2}(Z_{jt}, Z_{js}) + O_P(n_j^{-3/2} c_{jn})$.

PROOF: Put $\tilde{\phi}_{jn}(z_1, z_2, z_3) \equiv \phi_{jn}(z_1, z_2, z_3) - \phi_{jn2}(z_1, z_2) - \phi_{jn2}(z_2, z_3) - \phi_{jn2}(z_3, z_1)$. Then $\tilde{\phi}_{jn}(\cdot, \cdot, \cdot)$ is symmetric in its arguments, with

$$\text{(B5)} \qquad \int_{\mathbb{I}^2} \tilde{\phi}_{jn}(z_1, z_2, z_3)\, dF_j(z_3) = 0 \quad \forall\, z_1, z_2 \in \mathbb{I}^2$$

given (B4). By straightforward algebra, we can write

$$\text{(B6)} \qquad \hat{\phi}_n(j) = 3\binom{n_j}{2}^{-1} \sum_{t=j+2}^{n} \sum_{s=j+1}^{t-1} \phi_{jn2}(Z_{jt}, Z_{js}) + \binom{n_j}{3}^{-1} \sum_{t=j+3}^{n} \sum_{s=j+2}^{t-1} \sum_{r=j+1}^{s-1} \tilde{\phi}_{jn}(Z_{jt}, Z_{js}, Z_{jr})$$

$$= 3\hat{\phi}_{n2}(j) + \tilde{\phi}_n(j), \quad \text{say.}$$

For the last term $\tilde{\phi}_n(j)$ in (B6), we have the decomposition

$$\begin{aligned}
\text{(B7)} \qquad \tilde{\phi}_n(j) &= \sum_{t>s>r} \tilde{\phi}_{jn}(Z_{jt}, Z_{js}, Z_{jr})\mathbb{1}(t = s + j) \\
&\quad + \sum_{t>s>r} \tilde{\phi}_{jn}(Z_{jt}, Z_{js}, Z_{jr})\mathbb{1}(t \neq s + j, s = r + j) \\
&\quad + \sum_{t>s>r} \tilde{\phi}_{jn}(Z_{jt}, Z_{js}, Z_{jr})\mathbb{1}(t \neq s + j, s \neq r + j, t = r + j) \\
&\quad + \sum_{t>s>r} \tilde{\phi}_{jn}(Z_{jt}, Z_{js}, Z_{jr})\mathbb{1}(t \neq s + j, s \neq r + j, t \neq r + j) \\
&= \tilde{\phi}_{n1}(j) + \tilde{\phi}_{n2}(j) + \tilde{\phi}_{n3}(j) + \tilde{\phi}_{n4}(j), \quad \text{say.}
\end{aligned}$$

Noting that $\tilde{\phi}_{n1}(j)$ is a double sum over (s, r) and that Z_{jr} is independent of $(Z_{j(s+j)}, Z_{js})$ if $s \neq r + j$, we have $E\tilde{\phi}_{n1}^2(j) = o(n_j^3 c_{jn}^2)$ given (B5), $E\phi_{jn}^2(Z_{jt}, Z_{js}, Z_{jr}) = O(c_{jn}^2)$, and the Cauchy–Schwarz inequality. Similarly, we have $E\tilde{\phi}_{n2}^2(j) = o(n_j^3 c_{jn}^2)$ and $E\tilde{\phi}_{n3}^2(j) = o(n_j^3 c_{jn}^2)$. Finally, we consider

$$\begin{aligned}
E\tilde{\phi}_{n4}^2(j) = \sum_{t_1>s_1>r_1} \sum_{t_2>s_2>r_2} & E\left[\tilde{\phi}_{jn}(Z_{jt_1}, Z_{js_1}, Z_{jr_1})\tilde{\phi}_{jn}(Z_{jt_2}, Z_{js_2}, Z_{jr_2})\right] \\
& \times \mathbb{1}(t_1 \neq s_1 + j, s_1 \neq r_1 + j, t_1 \neq r_1 + j) \\
& \times \mathbb{1}(t_2 \neq s_2 + j, s_2 \neq r_2 + j, t_2 \neq r_2 + j).
\end{aligned}$$

Because Z_{jt_1}, Z_{js_1}, and Z_{jr_1} are independent given $t_1 > s_1 > r_1$, $t_1 \neq s_1 + j$, $s_1 \neq r_1 + j$, and $t_1 \neq r_1 + j$, each of them must be dependent on at least one of $(Z_{jt_2}, Z_{js_2}, Z_{jr_2})$ to have a nonzero expectation. This occurs if and only if each of (t_1, s_1, r_1) is equal to ι or $\iota \pm j$, where ι is one of (t_2, s_2, r_2). Thus, $E\tilde{\phi}_{n4}^2(j)$ consists of only finitely many triple sums over (t_2, s_2, r_2), so it is bounded by $O(n_j^3 c_{jn}^2)$. It follows that $E\tilde{\phi}_{n4}^2(j) = o(n_j^3 c_{jn}^2)$, whence, from (B7), we obtain $E\tilde{\phi}_n^2(j) = o(n_j^3 c_{jn}^2)$. This and (B6) yield the desired result. *Q.E.D.*

We now prove Lemmas A.1–A.11.

PROOF OF LEMMA A.1: Put $\bar{K}_h(x) \equiv EK_h(x, X_1)$. Then when $f_j(z) = g(x)g(y)$, we have

$$\begin{aligned}
\text{(B8)} \qquad A_{jn}(z_1, z_2) &= a_n(x_1, x_2)a_n(y_1, y_2) \\
&\quad + a_n(x_1, x_2)\frac{\bar{K}_h(y_1)}{g(y_1)} + \frac{\bar{K}_h(x_1)}{g(x_1)}a_n(y_1, y_2),
\end{aligned}$$

where $z_1 = (x_1, y_1)$ and $z_2 = (x_2, y_2)$. Because Z_1 and Z_3 are independent, $f_j(\cdot) = g(\cdot)g(\cdot)$ under $\mathbb{H}_0$, and $Ea_n(X_3, X_1) = 0$, we have from (B8) that

$$\text{(B9)}\qquad \begin{aligned} &EA_{jn}^2(Z_3, Z_1) - 2Ea_n^2(X_3, X_1) \\ &= \left\{E[a_n^2(X_3, X_1)]\right\}^2 + 2E[a_n^2(X_3, X_1)]\left\{E\left[\frac{\bar{K}_h(X_3)}{g(X_3)}\right]^2 - 1\right\} \\ &= \left\{E[a_n^2(X_3, X_1)]\right\}^2 + O(h), \end{aligned}$$

where $E[a_n^2(X_3, X_1)] = O(h^{-1})$ by change of variables, and continuity of $g(\cdot)$ and $E[\bar{K}_h(X_3)/g(X_3)]^2 - 1 = [Eb_n^2(X_3)]^2 + 2Eb_n(X_3) = O(h^2)$ given $\sup_{x\in\mathbb{I}} |b_n(x)| = O(h^2)$, which follows by change of variables, a second-order Taylor series expansion, boundedness of $g^{(2)}(x)$, and the use of the jackknife kernel $k_b(\cdot)$.

By change of variables $x_1 = x_2 + hu$ and Taylor series expansions (first-order Taylor series expansion when $x_3 \in [0, h) \cup (1-h, 1]$ and second-order Taylor series expansion when $x_3 \in [h, 1-h]$), we can obtain $E[a_n^2(X_3, X_1)] = E[K_h(X_3, X_1)/g(X_3)]^2 - \{E[\bar{K}_h(X_3)/g(X_3)]\}^2 = A_n^0 - 1 + O(h)$. It follows that $\{E[a_n^2(X_3, X_1)]\}^2 = d_n^0 + O(1)$. This and (B9) imply Lemma A.1. *Q.E.D.*

PROOF OF LEMMA A.2: When $f_j(z) = g(x)g(y)$ under $\mathbb{H}_0$, we have

$$\text{(B10)}\qquad B_{jn}(z) = b_n(x)b_n(y) + b_n(x) + b_n(y).$$

This and $\sup_{x\in\mathbb{I}} |b_n(x)| = O(h^2)$ imply Lemma A.2, namely

$$\begin{aligned} &EB_{jn}^2(Z_1) - 2Eb_n^2(X_1) \\ &= 2[Eb_n(X_1)]^2 + [Eb_n^2(X_1)]^2 + 4E[b_n^2(X_1)]E[b_n(X_1)] \\ &= 2[Eb_n(X_1)]^2 + O(h^6). \end{aligned}$$

Q.E.D.

PROOF OF LEMMA A.3: Using (B10), we have

$$\begin{aligned} &\hat{B}_n(j) - \hat{b}_{1n}(j) - \hat{b}_{2n}(j) \\ &= [Eb_n(X_1)]^2 + n_j^{-1} \sum_{t=j+1}^{n} \left\{b_n(X_t)b_n(X_{t-j}) - E[b_n(X_t)b_n(X_{t-j})]\right\} \\ &= [Eb_n(X_1)]^2 + O_P(n_j^{-1/2}h^4), \end{aligned}$$

where the last equality follows by Chebyshev's inequality, the fact that Z_{jt} and Z_{js} are mutually independent for $t \neq s$, $s \pm j$ under $\mathbb{H}_0$, and $\sup_{x\in\mathbb{I}} |b_n(x)| = O(h^2)$. *Q.E.D.*

PROOF OF LEMMA A.4: Put $c_n(X_t) \equiv \int_0^1 a_n(x, X_t) b_n(x) g(x)\,dx$. Using (B8), (B10), and $f_j(z) = g(x)g(y)$ under $\mathbb{H}_0$, we have

$$\begin{aligned}
&\int_{\mathbb{I}^2} A_{jn}(z, Z_{jt}) B_{jn}(z) f_j(z)\,dz \\
&\quad = c_n(X_t) + c_n(X_{t-j}) + c_n(X_t)c_n(X_{t-j}) \\
&\qquad + [c_n(X_t) + c_n(X_{t-j})] \int_0^1 b_n(y) g(y)\,dy \\
&\qquad + \left[c_n(X_t) \int_0^1 a_n(y, X_{t-j}) g(y)\,dy \right. \\
&\qquad\quad \left. + c_n(X_{t-j}) \int_0^1 a_n(x, X_t) g(x)\,dx \right] \\
&\qquad + [c_n(X_t) + c_n(X_{t-j})] \int_0^1 \bar{K}_h(y) b_n(y)\,dy \\
&\qquad + \left[\int_0^1 a_n(x, X_t) g(x)\,dx + \int_0^1 a_n(y, X_{t-j}) g(y)\,dy \right] \\
&\qquad\quad \times \int_0^1 \bar{K}_h(x) b_n(x)\,dx \\
&\quad \equiv c_n(X_t) + c_n(X_{t-j}) + \sum_{i=1}^{5} \delta_{in}(Z_{jt}), \quad \text{say.}
\end{aligned}$$

It follows that $\hat{C}_n(j) - \hat{c}_{1n}(j) - \hat{c}_{2n}(j) = \sum_{i=1}^{5} \hat{\delta}_i(j) = O_P(n_j^{-1/2} h^4)$, where $\hat{\delta}_i(j) \equiv n_j^{-1} \sum_{t=j+1}^{n} \delta_{in}(Z_{jt})$ and the second equality follows by Chebyshev's inequality, $\sup_{x\in\mathbb{I}} |b_n(x)| = O(h^2)$, and the fact that for any given $y \in (0, 1)$ and for all n sufficiently large (so that $y/h \to +\infty$ and $(1-y)/h \to +\infty$), we have

$$\begin{aligned}
&\int_0^1 K_h(x, y)\,dx \\
&\quad = \int_0^h K_h(x, y)\,dx + \int_{1-h}^1 K_h(x, y)\,dx + \int_h^{1-h} K_h(x, y)\,dx = 1
\end{aligned}$$

by change of variables and Assumption A.2, where, for n sufficiently large, the first two terms are identically zero given that $k(\cdot)$ has bounded support on $[-1, 1]$, and the last term is equal to unity identically given $\int_{-1}^1 k(u)\,du = 1$. This latter fact implies that $\int_0^1 a_n(x, y) g(x)\,dx = 0$ for $y \in (0, 1)$ and $\int_0^1 b_n(x) \times g(x)\,dx = 0$ for n sufficiently large. As a result, we have $\hat{\delta}_2(j) = \hat{\delta}_3(j) = \hat{\delta}_5(j) = 0$ a.s. for n sufficiently large. *Q.E.D.*

PROOF OF LEMMA A.5: By a standard variance-bias argument for kernel estimators (e.g., Fan and Yao (2003)), we have that under Assumptions A.1 and A.2 and $\mathbb{H}_0$,

$$\text{(B11)}\qquad \max_{1\le t\le n}\sup_{z\in\mathbb{I}}|\hat{f}_{jt}(z)-f(z)| = O_P(n_j^{-1/2}h^{-1}\ln n_j + h^2),$$

$$\text{(B12)}\qquad n_j^{-1}\sum_{t=j+1}^{n} E[\hat{f}_{jt}(Z_{it})-f_j(Z_{jt})]^2 = O(n_j^{-1}h^{-2}+h^4).$$

Note that the fact that j may grow as $n\to\infty$ does not affect the convergence rates because Z_{jt} and Z_{js} are independent if $t\notin\{s, s\pm j\}$. It follows from (B11) and (B12) that

$$\begin{aligned}\text{(B13)}\qquad & n_j^{-1}\sum_{t=j+1}^{n}|\hat{f}_{jt}(Z_{it})-f_j(Z_{jt})|^3\\ &\le \max_{1\le t\le n}\sup_{z\in\mathbb{I}}|\hat{f}_{jt}(z)-f_j(z)|\left\{n_j^{-1}\sum_{t=j+1}^{n}[\hat{f}_{jt}(Z_{it})-f_j(Z_{jt})]^2\right\}\\ &= O_P(n_j^{-3/2}h^{-3}\ln n_j + h^6).\end{aligned}$$

Now, using the inequality that $|\ln(1+x)-x+\frac{1}{2}x^2|\le|x|^3$ for small $x\in\mathbb{R}$, we obtain

$$\begin{aligned}\text{(B14)}\qquad & \left|\hat{I}_{jn}(\hat{f}_j, f_j) - n_j^{-1}\sum_{t\in S_n(j)}\frac{\hat{f}_{jt}(Z_{jt})-f_j(Z_{jt})}{f_j(Z_{jt})} + \frac{1}{2}\sum_{t\in S_n(j)}\left[\frac{\hat{f}_{jt}(Z_{jt})-f_j(Z_{jt})}{f_j(Z_{jt})}\right]^2\right|\\ &\le n_j^{-1}\sum_{t\in S_n(j)}\left|\frac{\hat{f}_{jt}(Z_{jt})-f_j(Z_{jt})}{f_j(Z_{jt})}\right|^3 = O_P(n_j^{-3/2}h^{-3}\ln n_j + h^6)\end{aligned}$$

by (B13). Also, recalling the definitions of $\hat{W}_1(j)$ and $\hat{W}_2(j)$ in (3.8), we have

$$\begin{aligned}\text{(B15)}\qquad & n_j^{-1}\sum_{t\in S_n(j)}\frac{\hat{f}_{jt}(Z_{jt})-f_j(Z_{jt})}{f_j(Z_{jt})}\\ &\quad -\frac{1}{2}\sum_{t\in S_n(j)}\left[\frac{\hat{f}_{jt}(Z_{jt})-f_j(Z_{jt})}{f_j(Z_{jt})}\right]^2 - \left[\hat{W}_1(j)-\frac{1}{2}\hat{W}_2(j)\right]\\ &= O_P\left[n_j^{-1}\sum_{t=j+1}^{n}P(t\in S_n(j))\right]\end{aligned}$$

$$= O_P\left[n_j^{-1} \sum_{t=j+1}^{n} E\left| \frac{\hat{f}_{jt}(Z_{jt}) - f_j(Z_{jt})}{f_j(Z_{jt})} \right|^3 \right]$$
$$= O_P(n_j^{-3/2} h^{-3} \ln n_j + h^6),$$

where we used (B11), (B12) and the fact that

$$P(t \in S_n(j)) = P(\hat{f}_{jt}(Z_{jt}) \le 0) \le P\big[|\hat{f}_{jt}(Z_{jt}) - f_j(Z_{jt})| > f_j(Z_{jt})\big]$$
$$\le E\left| \frac{\hat{f}_{jt}(Z_{jt}) - f_j(Z_{jt})}{f_j(Z_{jt})} \right|^3.$$

It follows from (B14) and (B15) that $\hat{I}_{jn}(\hat{f}_j, f_j) - \hat{W}_1(j) + \frac{1}{2}\hat{W}_2(j) = O_P(n_j^{-3/2} \times h^{-3}/\ln n_j + h^6)$. *Q.E.D.*

PROOF OF LEMMA A.6: As shown in the proof of Lemma A.4, $\int_0^1 K_h(x, y)\,dx = 1$ for $y \in (0,1)$ and for all n sufficiently large. In follows that $P[\gamma_{jn}(Z_{jt}, Z_{js}) = 0$ for all n sufficiently large$] = 1$ under $\mathbb{H}_0$, whence, we have that with probability one, $\hat{\Gamma}_n(j) = 0$ for all n sufficiently large. *Q.E.D.*

PROOF OF LEMMA A.7: We apply Lemma B.1, with $\phi_{jn}(z_1, z_2) = h^2 D_{jn}(z_1, z_2)$. By the Cauchy–Schwarz inequality, Jensen's inequality, and $c \le f_j(z) \le C$, we have $E[h^2 D_{jn}(Z_{jt}, Z_{js})]^2 \le Ch^4 \int_{\mathbb{I}^4} A_{jn}^4(z_1, z_2)\,dz_1\,dz_2 = O(h^{-2}) = o(n_j)$, by change of variables and $n_j h^4/\ln n_j \to \infty$. Also, we can verify that $E\phi_{jn1}^2(Z_{jt}) \le C$. Hence, the condition of Lemma B.1 holds and so $h^2\hat{D}_n(j) = h^2 \int_{\mathbb{I}^4} D_{jn}(z_1, z_2) f_j(z_1) f_j(z_2)\,dz_1\,dz_2 + O_P(n_j^{-1/2}) = 2h^2 E A_{jn}^2(Z_3, Z_1) + O_P(n_j^{-1/2})$. *Q.E.D.*

PROOF OF LEMMA A.8: We apply Lemma B.2, putting $\phi_{jn}(z_1, z_2, z_3) = \tilde{H}_{2jn}(z_1, z_2, z_2)$. Then $\phi_{2jn}(z_1, z_2) \equiv \int_{\mathbb{I}^2} \phi_{jn}(z_1, z_2, z_3) f_j(z_3)\,dz_3 = H_{2jn}(z_1, z_2)$, and the result of Lemma A.8 follows immediately from Lemma B.2 with $c_{jn} = h^{-2}$ because $E\phi_{jn}^2(Z_{jt}, Z_{js}, Z_{jr}) \le c^{-1}h^2 \int_{\mathbb{I}^6} A_{jn}^2(z_1, z_2) A_{jn}^2(z_1, z_3)\,dz_1\,dz_2\,dz_3 = o(h^{-2})$. *Q.E.D.*

PROOF OF LEMMA A.9: By Chebyshev's inequality, $EB_{jn}^4(Z_1) = O(h^8)$, and independence between Z_{jt} and Z_{js} unless $t \in \{s, s \pm j\}$ under $\mathbb{H}_0$, we immediately obtain $\hat{W}_{22}(j) = EB_{jn}^2(Z_1) + n_j^{-1} \sum_{t=j+1}^{n} [B_{jn}^2(Z_{jt}) - EB_{jn}^2(Z_1)] = EB_{jn}^2(Z_1) + O_P(n_j^{-1/2} h^4)$. *Q.E.D.*

PROOF OF LEMMA A.10: Put $\tilde{C}_{jn}(z_1, z_2) \equiv A_{jn}(z_1, z_2) B_{jn}(z_1) + A_{jn}(z_2, z_1) \times B_{jn}(z_2)$. Then $\hat{W}_{23}(j) = \frac{1}{2}\binom{n_j}{2}^{-1} \sum_{t=j+2}^{n} \sum_{s=j+1}^{t-1} \tilde{C}_{jn}(Z_{jt}, Z_{js})$. We now apply Lemma B.1, putting $\phi_{jn}(z_1, z_2) = h^{-2}\tilde{C}_{jn}(z_1, z_2)$. Because $E\phi_{jn}^2(Z_{jt}, Z_{js}) \le 2h^{-4} \times$

$\int_{\mathbb{I}^4} A_{jn}^2(z_1, z_2)B_{jn}^2(z_1)\,dz_1\,dz_2 = O(h^{-2}) = o(n_j)$ by $\sup_{z\in\mathbb{I}^2}|B_{jn}(z)| = O(h^2)$ and change of variables, we have $h^{-2}\hat{W}_{23}(j) = h^{-2}\hat{C}_n(j) + o_P(n_j^{-1/2})$ by Lemma B.1. *Q.E.D.*

PROOF OF LEMMA A.11: (a) Part (a) is a standard result available in most standard nonparametric density textbooks (e.g., Fan and Yao (2003)).

(b) We now show part (b). Given $\hat{g}(x_1) - \hat{g}(x_2) = [\hat{g}(x_1) - g(x_1)] - [\hat{g}(x_2) - g(x_2)] + [g(x_1) - g(x_2)]$, we have $\sup_{x_1,x_2\in\mathbb{N}(\delta)}|\hat{g}(x_1) - \hat{g}(x_2)| \le 2\sup_{x\in\mathbb{I}}|\hat{g}(x) - g(x)| + \sup_{x_1,x_2\in\mathbb{N}(\delta)}|g(x_1) - g(x_2)| = O_P(n^{-1/2}h^{-1/2}\ln n + h^2) + O(\delta)$, where the last equality follows from part (a) and the continuous differentiability of $g(\cdot)$ on $\mathbb{I}$.

(c) By a second-order Taylor series expansion and the properties of the jackknife kernel $k_b(\cdot)$, we have for all $x \in \mathbb{I}$,

$$
\begin{aligned}
E^*[\hat{g}_t^*(x)|\mathcal{X}] - \hat{g}(x) &= \int_0^1 K_h(x, x^*)\hat{g}(x^*)\,dx^* - \hat{g}(x) \\
&= \frac{1}{2}h^2 \int_{-1}^1 u^2 K_{b(x)}(u)\hat{g}^{(2)}(x + \lambda uh)\,du,
\end{aligned}
$$

where $b(x) = x/h$ for $x \in [0, h)$, $b(x) = (1 - x)/h$ for $x \in (1 - h, 1]$, and $b(x) = 1$ for $x \in [h, 1 - h]$.

(d) The desired result follows given $nh^5 = O(1)$ because

$$
\begin{aligned}
\sup_{x\in\mathbb{I}}|\hat{g}^{(2)}(x)| &\le \sup_{x\in\mathbb{I}}|\hat{g}^{(2)}(x) - E\hat{g}^{(2)}(x)| + \sup_{x\in\mathbb{I}}|E\hat{g}^{(2)}(x)| \\
&= O_P(n^{-1/2}h^{-5/2}\ln n) + O(1) = O_P(\ln n),
\end{aligned}
$$

where the first term follows from an argument similar to the proof for $\sup_{x\in\mathbb{I}}|\hat{g}(x) - E\hat{g}(x)|$. *Q.E.D.*

REFERENCES

AHMAD, I., AND P. LIN (1976): "A Nonparametric Estimation of the Entropy for Absolutely Continuous Distributions," *IEEE Transactions on Information Theory*, 22, 372–375.

BROCK, W., D. DECHERT, J. SCHEINKMAN, AND B. LEBARON (1996): "A Test for Independence Based on the Correlation Dimension," *Econometric Reviews*, 15, 197–235.

BROWN, B. (1971): "Martingale Limit Theorems," *The Annals of Mathematical Statistics*, 42, 59–66.

DELGADO, M. (1996): "Testing Serial Independence Using the Sample Distribution Function," *Journal of Time Series Analysis*, 17, 271–285.

DIEBOLD, F. X., T. GUNTHER, AND A. TAY (1998): "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review*, 39, 863–883.

DUDEWICZ, E. J., AND E. C. VAN DER MEULEN (1981): "Entropy-Based Tests of Uniformity," *Journal of the American Statistical Association*, 76, 967–974.

FAN, J., AND Q. YAO (2003): *Nonlinear Time Series: Parametric and Nonparametric Approaches*. New York: Springer-Verlag.

FAN, Y., AND Q. LI (1996): "Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms," *Econometrica*, 64, 865–890.

GOKHALE, D. V. (1983): "On Entropy-Based Goodness-of-Fit Tests," *Computational Statistics and Data Analysis*, 1, 157–165.

GRANGER, C. W. J., AND J. L. LIN (1994): "Using the Mutual Information Coefficient to Identify Lags in Nonlinear Models," *Journal of Time Series Analysis*, 15, 371–384.

GRANGER, C. W. J., AND T. TERÄSVIRTA (1993): *Modelling Nonlinear Economic Relationships*. New York: Oxford University Press.

HALL, P. (1988): "Estimating the Direction in Which a Data Set Is Most Interesting," *Probability Theory and Related Fields*, 80, 51–77.

HANSEN, B. (1994): "Autoregressive Conditional Density Estimation," *International Economic Review*, 35, 705–730.

HÄRDLE, W. (1990): *Applied Nonparametric Regression*. New York: Cambridge University Press.

HÄRDLE, W., AND J. HOROWITZ (1993): "Testing a Parametric Model Against a Semiparametric Alternative," *Econometric Theory*, 10, 821–848.

HÄRDLE, W., AND E. MAMMEN (1993): "Comparing Nonparametric versus Parametric Regression Fits," *The Annals of Statistics*, 21, 1926–1947.

HJELLVIK, V., Q. YAO, AND D. TJØSTHEIM (1998): "Linearity Testing Using Local Polynomial Approximation," *Journal of Statistical Planning and Inference*, 68, 295–321.

HONG, Y. (1998): "Testing for Pairwise Independence via the Empirical Distribution Function," *Journal of the Royal Statistical Society*, Ser. B, 60, 429–453.

——— (1999): "Hypothesis Testing in Time Series via the Empirical Characteristic Function: A Generalized Spectral Density Approach," *Journal of the American Statistical Association*, 84, 1201–1220.

HOROWITZ, J. (2001): "Bootstrap," in *Handbook of Econometrics*, Vol. 5, ed. by J. J. Heckman and E. E. Leamer. Amsterdam: Elsevier, Chap. 52, 3159–3228.

JOE, H. (1989a): "Relative Entropy Measures of Multivariate Dependence," *Journal of the American Statistical Association*, 84, 157–164.

——— (1989b): "Estimation of Entropy and Other Functionals of a Multivariate Density," *Annals of the Institute of Statistical Mathematics*, 41, 683–697.

JONDEAU, E., AND M. ROCKINGER (2003): "Conditional Volatility, Skewness, and Kurtosis: Existence, Persistence, and Comovements," *Journal of Economic Dynamics and Control*, 27, 1699–1737.

LEE, S., AND B. HANSEN (1994): "Asymptotic Theory for the GARCH(1, 1) Quasi-Maximum Likelihood Estimator," *Econometric Theory*, 10, 29–52.

MAASOUMI, E. (1993): "A Compendium to Information Theory in Economics and Econometrics," *Econometric Reviews*, 12, 137–181.

MAASOUMI, E., AND J. RACINE (2002): "Entropy and Predictability of Stock Returns," *Journal of Econometrics*, 107, 175–189.

MAASOUMI, E., J. RACINE, AND C. GRANGER (2004): "A Dependence Metric for Possibly Nonlinear Processes," *Journal of Time Series Analysis*, 25, 649–699.

MOKKADEM, A. (1989): "Estimation of the Entropy and Information of Absolutely Continuous Random Variables," *IEEE Transactions on Information Theory*, 35, 193–196.

PINKSE, J. (1998): "Consistent Nonparametric Testing for Serial Independence," *Journal of Econometrics*, 84, 205–231.

RICE, J. (1984): "Boundary Modification for Kernel Regression," *Communications in Statistics*, Ser. A, 13, 893–900.

ROBINSON, P. M. (1991): "Consistent Nonparametric Entropy-Based Testing," *Review of Economic Studies*, 58, 437–453.

SKAUG, H. J., AND D. TJØSTHEIM (1993a): "Nonparametric Tests of Serial Independence," in *Developments in Time Series Analysis: The Priestley Birthday Volume*, ed. by T. Subba Rao. London: Chapman & Hall, 207–229.

——— (1993b): "A Nonparametric Test of Serial Independence Based on the Empirical Distribution Function," *Biometrika*, 80, 591–602.

——— (1996): "Measures of Distance Between Densities with Application to Testing for Serial Independence," in *Time Series Analysis in Memory of E. J. Hannan*, ed. by P. Robinson and M. Rosenblatt. New York: Springer-Verlag, 363–377.

TJØSTHEIM, D. (1996): "Measures and Tests of Independence: A Survey," *Statistics*, 28, 249–284.

ULLAH, A. (1993): "Entropy, Divergence and Distance Measures with Econometric Applications," *Journal of Statistical Planning and Inference*, 49, 137–162.

VASICEK, O. (1976): "A Test for Normality Based on Sample Entropy," *Journal of the Royal Statistical Society*, Ser. B, 38, 54–59.

WHITE, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.

ZHENG, X. (2000): "A Consistent Test of Conditional Parametric Distributions," *Econometric Theory*, 16, 667–691.