The Annals of Statistics
2003, Vol. 31, No. 6, 2059–2095

TESTING CONDITIONAL MOMENT RESTRICTIONS

BY GAUTAM TRIPATHI[1] AND YUICHI KITAMURA[2]

University of Wisconsin and University of Pennsylvania

Let (x, z) be a pair of observable random vectors. We construct a new "smoothed" empirical likelihood-based test for the hypothesis $\mathbb{E}\{g(z, \theta)|x\} = 0$ w.p.1, where g is a vector of known functions and θ an unknown finite-dimensional parameter. We show that the test statistic is asymptotically normal under the null hypothesis and derive its asymptotic distribution under a sequence of local alternatives. Furthermore, the test is shown to possess an optimality property in large samples. Simulation evidence suggests that it also behaves well in small samples.

1. Introduction. In a series of papers, Owen (1988, 1990, 1991) studied the use of inference based on the nonparametric likelihood ratio. This approach is particularly useful when testing hypotheses that can be expressed as moment restrictions. However, the attention of most of the literature seems to have been confined to dealing with hypotheses expressed as unconditional moment restrictions. In this paper, we extend the empirical likelihood paradigm to handle the testing of conditional moment restrictions. Let x denote a continuously distributed random vector. Throughout the paper, we will treat x as the conditioning variable. In this paper, we extend the empirical likelihood approach to test

$$\text{H}_0 : \Pr\{\mathbb{E}[g(z, \theta)|x] = 0\} = 1 \qquad \text{for some } \theta \in \Theta \tag{1.1}$$

against the alternative that H_0 is false.

Much progress has been made in the area of testing conditional moment restrictions. See, among others, Newey (1985), Bierens (1990), de Jong and Bierens (1994) and the references therein. Related to this literature is the work on specification testing of a parametric regression function against a nonparametric alternative. See, for instance, Eubank and Spiegelman (1990), Wooldridge (1992), Yatchew (1992), Härdle and Mammen (1993), Whang and Andrews (1993), Hong and White (1995), Fan and Li (1996), Zheng (1996), Andrews (1997), Bierens and Ploberger (1997), Ellison and Ellison (2000), Aït-Sahalia, Bickel and Stoker (2001) and Horowitz and Spokoiny (2001). Unlike these papers, we examine a general class of conditional moment restrictions that nests conditional mean regression as a special case. For example, our approach is capable of handling the case where g is a vector of residuals from a system of static nonlinear

Received September 2000; revised December 2002.

[1]Supported by NSF Grants SES-01-11917 and SES-02-14081.

[2]Supported by an Alfred P. Sloan Foundation Research Fellowship and NSF Grants SBR-96-32101 and SES-99-05247.

AMS 2000 subject classifications. Primary 62G10; secondary 62J20.

Key words and phrases. Conditional moment restrictions, empirical likelihood, smoothing.

simultaneous equations. We show that a test for H_0 based on Owen's empirical likelihood provides a useful alternative to the procedures developed in the above-mentioned papers. Our test is easy to construct and straightforward to implement. Its distribution is asymptotically normal under the null, and it is able to detect local alternatives that converge to the null at rates only slightly slower than the parametric rate. A distinguishing feature of the proposed test is that it is asymptotically optimal in terms of an average power criterion as used by Wald (1943) and Andrews and Ploberger (1994). Moreover, it also appears to work well in finite samples.

The following notation is used throughout the paper. By a "vector" we mean a column vector. We do not make any notational distinction between a random variable and the value taken by it. The difference should be clear from the context. The symbol S denotes a subset of $\mathbb{R}^s$ which may be unbounded, $\mathbb{I}\{A\}$ is the indicator function of set A, and for a matrix V the symbol $\|V\| = \sqrt{\operatorname{tr}(VV')}$ denotes its Frobenius norm; $\|V\|$ reduces to the usual Euclidean norm when V happens to be a vector. Unless stated otherwise, all limits are taken as the number of observations $n \uparrow \infty$.

2. The smoothed empirical likelihood approach. This section develops an empirical likelihood-based test of conditional moment restriction $\mathbb{E}\{g(z, \theta)|x\} = 0$. Our main tool is empirical likelihood (EL), though a kernel smoothing technique plays an important part in formulating our test procedure. Recall that smoothing arises naturally in the theory of local likelihood estimation by considering the expected log-likelihood. See, for example, Brillinger (1977), Owen (1984), Hastie and Tibshirani (1986), Staniswalis (1987) and Staniswalis and Severini (1991). Our empirical likelihood ratio-based test can also be motivated using an expected log-likelihood criterion.

Smoothing is necessary in our case because the conventional EL approach fails when testing conditional moment restrictions for which the conditioning variables are continuously distributed. The problem is analogous to the failure of likelihood-based function estimation described in Hastie and Tibshirani (1986), Section 5. The remedy they suggest is to maximize the expected log-likelihood instead. Applying this idea to our situation, consider solving

$$\max_{\{p_{ij}:\, i,j=1,\dots,n\}} \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\log p_{ij} \quad \text{s.t.}$$
$$(2.1)\qquad p_{ij} \ge 0,\ \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij} = 1,\ \frac{\sum_{j=1}^{n} g(z_j,\hat{\theta})p_{ij}}{\sum_{j=1}^{n} p_{ij}} = 0,$$

where $\hat{\theta}$ is a preliminary estimator of θ, p_{ij} denotes the probability mass placed at (x_i, z_j) by a discrete distribution with support $\{x_1, \dots, x_n\} \times \{z_1, \dots, z_n\}$,

$$w_{ij} = \frac{K((x_i - x_j)/b_n)}{\sum_{j=1}^{n} K((x_i - x_j)/b_n)} = \frac{K_{ij}}{\sum_{j=1}^{n} K_{ij}}$$

and the function K is chosen to satisfy Assumption 3.7. The w_{ij}'s are kernel weights familiar from the nonparametric regression literature. The bandwidth b_n is a null sequence of positive numbers satisfying certain conditions described later in the paper.

In a b_n-neighborhood of x_i, w_{ij} assigns smaller weights to those x_j's that are farther away from x_i. This has the effect of smoothing the empirical log-likelihood at each x_i. Since the objective function depends on p_{ij} only through $\log p_{ij}$, the nonnegativity constraint does not bind. Hence, (2.1) is solved by maximizing the Lagrangian

$$\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\log p_{ij} - \mu\left(\sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij} - 1\right) - \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i' g(z_j,\hat{\theta}) p_{ij},$$

where μ is the Lagrange multiplier for the second constraint and $\{\lambda_i \in \mathbb{R}^q : i = 1,\dots,n\}$ is the set of multipliers for the third constraint. It is easy to verify that the solution to this problem is given by

$$\hat{p}_{ij} = \frac{w_{ij}}{n + \lambda_i' g(z_j,\hat{\theta})},$$

where each λ_i solves

$$\sum_{j=1}^{n} \frac{w_{ij} g(z_j,\hat{\theta})}{n + \lambda_i' g(z_j,\hat{\theta})} = 0, \qquad i = 1,\dots,n. \tag{2.2}$$

Note that λ_i is shorthand for $\lambda(x_i,\hat{\theta})$. Its dependence on $\hat{\theta}$ is suppressed for notational convenience. Hence, we can rewrite the restricted (i.e., under H_0) SEL as

$$\mathrm{SEL}^r = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\log \hat{p}_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\log\left\{\frac{w_{ij}}{n + \lambda_i' g(z_j,\hat{\theta})}\right\}.$$

Next, we look at the unrestricted problem, which is similar to (2.1) except that the conditional moment constraint is absent; that is, we solve

$$\max_{\{p_{ij} : i,j=1,\dots,n\}} \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\log p_{ij} \quad \text{s.t.} \quad p_{ij} \ge 0, \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij} = 1.$$

The solution to this is $\tilde{p}_{ij} = w_{ij}/n$, and we can write the unrestricted SEL as

$$\mathrm{SEL}^{ur} = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\log \tilde{p}_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\log\left\{\frac{w_{ij}}{n}\right\}.$$

An analog of the parametric likelihood ratio test statistic would then be

$$2(\mathrm{SEL}^{ur} - \mathrm{SEL}^r) = 2\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\log\left\{1 + \frac{\lambda_i' g(z_j,\hat{\theta})}{n}\right\}, \tag{2.3}$$

where λ_i solves (2.2). Heuristically speaking, $\mathrm{SEL}^{ur} - \mathrm{SEL}^{r}$ will be small if (1.1) holds. Therefore, it seems sensible to base the test for H_0 on (2.3). However, we use a modified version of (2.3) for our test because we now restrict ourselves to a situation where we are interested in the behavior of $x \mapsto \mathbb{E}\{g(z,\theta)|x\}$ only on a certain fixed subset (say S_*) of S, the support of x. So define the smoothed empirical likelihood ratio (SELR) as

$$\mathrm{SELR} = 2\sum_{i=1}^{n} \mathbb{I}\{x_i \in S_*\} \sum_{j=1}^{n} w_{ij} \log\left\{1 + \frac{\lambda_i' g(z_j,\hat{\theta})}{n}\right\}, \tag{2.4}$$

where each λ_i solves (2.2). Our test for H_0 is based on (2.4); namely, we reject the null hypothesis for large values of SELR.

Note that S_* is identical to the fixed trimming set used in Aït-Sahalia, Bickel and Stoker (2001). Fixed trimming is useful for practical and technical reasons. As Aït-Sahalia, Bickel and Stoker (2001) point out, a practical benefit is that we can focus specification testing on regions in x-space which may be empirically more relevant. Technically, it lets us avoid the usual edge effects associated with kernel estimators [Härdle and Marron (1990), page 66].

Before proceeding any further, we mention some additional papers in the empirical likelihood literature which may be relevant to us. The basic references are, of course, the seminal papers by Owen cited earlier. Using i.i.d. observations, Qin and Lawless (1994, 1995) and Imbens (1997) look at efficiently estimating finite-dimensional parameters under unconditional moment restrictions. Kitamura (1997) extends the treatment to weakly dependent data. Kitamura (2001) also describes an optimal property of empirical likelihood-based tests for unconditional moment restrictions. Not much work seems to have been done as far as applying empirical likelihood to conditional moment restrictions is concerned. Some exceptions include LeBlanc and Crowley (1995), Brown and Newey (1998) and Kitamura, Tripathi and Ahn (2002). LeBlanc and Crowley (1995) and Kitamura, Tripathi and Ahn (2002) are mainly concerned with estimation, while Brown and Newey (1998) consider the bootstrap under a conditional moment restriction. Some earlier papers in the econometrics literature that may be related to the empirical likelihood approach include Cosslett (1981a, b) and Chamberlain (1987, 1992). None of these papers contains the results obtained here.

Finally, in a recent study, Chen, Härdle and Kleinow (2001) propose a method that is closely related to our approach. They consider nonparametric specification testing using empirical likelihood in a time series context. They use a version of sample moments, localized at each point of a lattice over the space of conditioning variable. This yields a sequence of localized empirical likelihood ratios defined on the lattice. The fact that the user has to choose a lattice for the test brings some arbitrariness into their method. In contrast, our test does not require choosing such a lattice. Also, in this paper we demonstrate that our test has an asymptotic optimality property.

3. Basic assumptions and notation. Let $\mathbb{I}_i = \mathbb{I}\{x_i \in S_*\}$, $\mathbb{S}^a = \{\xi \in \mathbb{R}^a : \|\xi\| = 1\}$, $V(x_i, \theta) = \mathbb{E}\{g(z_i, \theta)g'(z_i, \theta)|x_i\}$ and $\hat{V}(x_i, \theta) = \sum_{j=1}^n w_{ij} g(z_j, \theta) \times g'(z_j, \theta)$. Here $x^{(i)}$ is the ith component of a vector x and $M^{(ij)}$ the (i, j)th element of a matrix M. Furthermore, $\text{vol}(S_*) = \int_{S_*} dx$ denotes the Lebesgue measure of S_*, $\partial g(z, \theta)/\partial\theta$ is the $q \times p$ Jacobian matrix, $D(x_i, \theta) = \mathbb{E}\{\partial g(z_i, \theta)/\partial\theta|x_i\}$ and "w.p.a.1" stands for "with probability approaching 1." The following regularity conditions help us determine the asymptotic behavior of our test.

ASSUMPTION 3.1. (i) $\{x_i, z_i\}_{i=1}^n$ is a random sample on $S \times \mathbb{R}^d$. (ii) x is continuously distributed with Lebesgue density h, while z can be continuous, discrete or mixed. (iii) $\Theta \subseteq \mathbb{R}^p$ and $g : \mathbb{R}^d \times \Theta \to \mathbb{R}^q$ is known. (iv) $\mathbb{E}\{\sup_{\theta\in\Theta} \|g(z, \theta)\|^m\} < \infty$ for some $m \geq 6$.

Note that $m = 6$ will be required in the proof of Lemma A.1. The next assumption describes the nature of S_*.

ASSUMPTION 3.2. The set S_* is compact and contained in the interior of S such that $\inf_{x\in S_*} h(x) > 0$.

This lets us avoid the boundary problems associated with kernel estimators. Compactness of S_* is required when we use uniform rates of convergence for kernel estimators of conditional expectations to handle remainder terms in the proofs. A consequence of this assumption is that our test will be consistent only against those alternatives that differ from the null on S_*. As suggested by a referee, it would be of interest (though technically challenging) to know how our results change if we let S_* expand so that the amount of trimming decreases with sample size.

ASSUMPTION 3.3. There exists a $\theta_0 \in \text{int}(\Theta)$ for which (1.1) holds.

We assume that θ_0 can be estimated by an $n^{1/2}$-consistent estimator.

ASSUMPTION 3.4. $\hat{\theta}$ is an estimator of θ_0 such that $\|\hat{\theta} - \theta_0\| = O_p(n^{-1/2})$.

The $n^{1/2}$-consistency of $\hat{\theta}$ guarantees that replacing $\hat{\theta}$ by θ_0 does not change the asymptotic behavior of our test statistics. Other details about the limiting distribution of $\hat{\theta}$ are not required.

ASSUMPTION 3.5. (i) $h(x)$ and $V(x, \theta_0)$ are twice continuously differentiable on S. (ii) $h(x)$ and $\mathbb{E}\{\|g(z, \theta_0)\|^m|x\}h(x)$ are uniformly bounded on S. (iii) $D(x, \theta_0)$ is continuous on S. (iv) $(\xi, x) \mapsto \xi' V(x, \theta_0)\xi$ is bounded away from 0 on $\mathbb{S}^q \times S_*$.

Conditions (i) and (ii) are used to obtain uniform (over $x \in S_*$) rates of convergence for kernel estimators. (iii) will be used in the proof of Lemma A.9. Condition (iv) implies that $\|V^{-1}(x, \theta_0)\|$ is bounded on S_*.

ASSUMPTION 3.6. For $1 \le i \le q$ and $1 \le j, k \le p$, there exists an open ball $\mathcal{N}_0$ around θ_0 on which $\theta \mapsto g(z, \theta)$ is twice continuously differentiable w.p.1 such that $\sup_{\theta \in \mathcal{N}_0} |\partial g^{(i)}(z, \theta)/\partial \theta^{(j)}| \le d(z)$ and $\sup_{\theta \in \mathcal{N}_0} |\partial^2 g^{(i)}(z, \theta)/\partial \theta^{(j)}\, \partial \theta^{(k)}| \le l(z)$ hold w.p.1 for some real-valued functions $d(z)$ and $l(z)$, where $\mathbb{E}\, d^\eta(z) < \infty$ for $\eta \ge 6$ and $\mathbb{E} l^2(z) < \infty$.

Under this assumption, the mean value approximations $\|g(z, \theta) - g(z, \theta_0)\| \le d(z)\|\theta - \theta_0\|$ and $\|g(z, \theta) - g(z, \theta_0) - \partial g(z, \theta_0)/\partial \theta(\theta - \theta_0)\| \le l(z)\|\theta - \theta_0\|^2$ hold w.p.1 for $\theta \in \mathcal{N}_0$. Note that $\eta = 6$ will be used in the proof of Lemma A.1, and $\mathbb{E} l^2(z) < \infty$ is required in the proof of Lemma A.5.

ASSUMPTION 3.7. $K(x) = \prod_{i=1}^s \kappa(x^{(i)})$, where κ is a continuously differentiable p.d.f. with support $[-1, 1]$. The function κ is symmetric about the origin and for some $a \in (0, 1)$ is bounded away from 0 on $[-a, a]$.

Since the kernels in Assumption 3.7 are employed to estimate probabilities, the use of kernels with order greater than 2 is ruled out. The nonnegativity of K is also explicitly used several times in the proofs. Continuous differentiability of κ means that K satisfies a Lipschitz condition. This allows us to use the uniform convergence rates for kernel estimators in Newey (1994). The requirement that K be bounded away from 0 on a closed ball centered at the origin allows us to use a result of Devroye and Wagner (1980) in Lemma C.5. For later use, define $\mathrm{R}(K) = \int_{[-1,1]^s} K^2(u)\, du$, $K^*(x) = \int_{[-1,1]^s} K(v)K(x - v)\, dv$ and $K^{**} = \int_{[-2,2]^s} \{K^*(u)\}^2\, du$.

4. The test statistics and their distributions under the null. In this section, we construct two statistics to test H_0. Both statistics, subsequently denoted by $\zeta_{1,n}$ and $\zeta_{2,n}$, are based on SELR. The first step is to transform SELR so that we can apply a CLT due to de Jong (1987). So let $b_n = n^{-\alpha}$ for $0 < \alpha < \min\{\frac{1}{s}(1 - \frac{4}{m}), \frac{1}{3s}, \frac{1}{s}(1 - \frac{2}{m} - \frac{2}{\eta})\}$. Then, following Lemma A.1, we can write

$$\mathrm{SELR} = \hat{\mathrm{T}} + o_p(1), \tag{4.1}$$

where

$$\hat{\mathrm{T}} = \sum_{i=1}^n \mathbb{I}\{x_i \in S_*\} \left\{ \sum_{j=1}^n w_{ij} g'(z_j, \hat{\theta}) \right\} \hat{V}^{-1}(x_i, \hat{\theta}) \left\{ \sum_{j=1}^n w_{ij} g(z_j, \hat{\theta}) \right\}.$$

Now decompose $\hat{\mathrm{T}} = \hat{\mathrm{T}}_1 + \hat{\mathrm{T}}_2 + \hat{\mathrm{T}}_3 + \hat{\mathrm{T}}_4 + \hat{\mathrm{T}}_5$, where

$$\hat{\mathrm{T}}_1 = K^2(0)\sum_{i=1}^{n}\mathbb{I}_i \frac{g'(z_i,\hat{\theta})\hat{V}^{-1}(x_i,\hat{\theta})g(z_i,\hat{\theta})}{\{\sum_{u=1}^{n} K_{iu}\}^2},$$

$$\hat{\mathrm{T}}_2 = \sum_{i=1}^{n}\sum_{j=1,\, j\neq i}^{n} \mathbb{I}_i w_{ij}^2 g'(z_j,\hat{\theta})\hat{V}^{-1}(x_i,\hat{\theta})g(z_j,\hat{\theta}),$$

$$\hat{\mathrm{T}}_3 = K(0)\sum_{i=1}^{n}\sum_{j=1,\, j\neq i}^{n}\mathbb{I}_i \frac{g'(z_i,\hat{\theta})\hat{V}^{-1}(x_i,\hat{\theta})g(z_j,\hat{\theta})w_{ij}}{\sum_{u=1}^{n} K_{iu}}, \qquad \hat{\mathrm{T}}_4 = \hat{\mathrm{T}}_3,$$

$$\hat{\mathrm{T}}_5 = \sum_{i=1}^{n}\sum_{j=1,\, j\neq i}^{n}\sum_{t=1,\, t\neq j\neq i}^{n} \mathbb{I}_i w_{ij} g'(z_j,\hat{\theta})\hat{V}^{-1}(x_i,\hat{\theta})g(z_t,\hat{\theta})w_{it}.$$

Define $\sigma^2 = 2qK^{**}\,\mathrm{vol}(S_*)$. Then, under H_0 and our choice of bandwidth, $b_n^{s/2}\hat{\mathrm{T}}_1 = o_p(1)$ follows by Lemma A.2, $b_n^{s/2}\hat{\mathrm{T}}_2 = b_n^{-s/2} q\,\mathrm{R}(K)\,\mathrm{vol}(S_*) + O_p(b_n^{2-s/2}) + o_p(1)$ by Lemma A.3, $b_n^{s/2}\hat{\mathrm{T}}_3 = o_p(1)$ by Lemma A.4 and $b_n^{s/2}\hat{\mathrm{T}}_5 \xrightarrow{d} \mathrm{N}(0,\sigma^2)$ by Lemma A.5. Although $b_n^{s/2}\hat{\mathrm{T}}_1$ and $b_n^{s/2}\hat{\mathrm{T}}_3$ are asymptotically negligible in probability, $b_n^{s/2}\hat{\mathrm{T}}_2$ explodes as $n \uparrow \infty$. Therefore, SELR has to be properly centered if we want a test statistic with a valid asymptotic distribution. This is done by subtracting $b_n^{s/2}\hat{\mathrm{T}}_2$ from SELR. Subtracting $b_n^{s/2}\hat{\mathrm{T}}_2$ does not lead to any loss of information as far as testing H_0 is concerned. This follows from Lemmas A.3 and B.3, which show that the asymptotic behavior of $\hat{\mathrm{T}}_2$ remains unchanged under H_0 and the sequence of local alternatives in (6.1).

We are now ready to construct $\zeta_{1,n}$ and $\zeta_{2,n}$. So define

$$\zeta_{1,n} = \frac{b_n^{s/2}\mathrm{SELR} - b_n^{s/2}\hat{\mathrm{T}}_2}{\sigma}. \tag{4.2}$$

By (4.1) and the above facts, $\zeta_{1,n} = b_n^{s/2}\hat{\mathrm{T}}_5/\sigma + o_p(1)$. Hence, the following result is immediate.

THEOREM 4.1. *Let Assumptions* 3.1–3.7 *hold. Furthermore, assume that* $b_n = n^{-\alpha}$ *for* $0 < \alpha < \min\{\frac{1}{s}(1-\frac{4}{m}), \frac{1}{3s}, \frac{1}{s}(1-\frac{2}{m}-\frac{2}{\eta})\}$. *Then* $\zeta_{1,n} \xrightarrow{d} \mathrm{N}(0,1)$ *under* H_0.

A size-γ test for H_0 can be obtained by comparing $\zeta_{1,n}$ with critical values obtained from a standard normal distribution. Notice that σ does not depend on any unknown parameters and can be calculated analytically. However, to use $\zeta_{1,n}$, we do need to calculate $\hat{\mathrm{T}}_2$. Even this calculation can be eliminated when we have at most three conditioning variables. To see this, observe that if $s \leq 3$, then

$b_n^{s/2}\hat{\mathrm{T}}_2 = b_n^{-s/2} q\,\mathrm{R}(K)\,\mathrm{vol}(S_*) + o_p(1)$. Hence, we can use

$$\zeta_{2,n} = \frac{b_n^{s/2}\mathrm{SELR} - b_n^{-s/2} q\,\mathrm{R}(K)\,\mathrm{vol}(S_*)}{\sigma} \tag{4.3}$$

to test H_0 when $s \leq 3$. This leads to the following result.

COROLLARY 4.1. *Let Assumptions* 3.1–3.7 *hold. Furthermore, assume that* $s \leq 3$ *and* $b_n = n^{-\alpha}$ *for* $0 < \alpha < \min\{\frac{1}{s}(1 - \frac{4}{m}), \frac{1}{3s}, \frac{1}{s}(1 - \frac{2}{m} - \frac{2}{\eta})\}$. *Then* $\zeta_{2,n} \xrightarrow{d} \mathrm{N}(0, 1)$ *under* H_0.

In practice, $\zeta_{2,n}$ seems more useful than $\zeta_{1,n}$ because $s \leq 3$ is a reasonable bound for most applications of nonparametric regression. A nice interpretation of Corollary 4.1 can be obtained by observing that we can express its result as

$$\frac{\mathrm{SELR} - c_1\gamma_n}{c_2\sqrt{2\gamma_n}} \xrightarrow{d} \mathrm{N}(0, 1), \tag{4.4}$$

where $c_1 = \mathrm{R}(K)$, $c_2 = \sqrt{K^{**}}$ and $\gamma_n = b_n^{-s} q\,\mathrm{vol}(S_*)$. Equation (4.4) can be regarded as a nonparametric analog of Wilks' theorem: If SELR were distributed as a χ^2 random variable with $c_1\gamma_n$ degrees of freedom and we had used a K for which $\mathrm{R}(K) = K^{**}$, then (4.4) would be interpreted as the normal approximation of a χ^2 random variable with large degrees of freedom. See Fan, Zhang and Zhang (2001) for more discussion regarding a nonparametric analog of Wilks' theorem.

5. Practical considerations. Implementing our SELR-based test is straightforward. To see this, first observe that since $\lambda \mapsto \log(1 + \lambda' g(z_j, \hat{\theta})/n)$ is well defined and strictly concave for large enough n, the λ_i's in (2.2) are numerical solutions of

$$\max_{\lambda \in \mathbb{R}^q} \sum_{j=1}^{n} w_{ij} \log\left\{1 + \frac{\lambda' g(z_j, \hat{\theta})}{n}\right\}. \tag{5.1}$$

This optimization problem can be uniquely solved for λ_i by a standard Newton–Raphson procedure. Therefore, we can rewrite (2.4) as

$$\mathrm{SELR} = 2\sum_{i=1}^{n} \mathbb{I}\{x_i \in S_*\} \max_{\bar{\lambda}_i \in \mathbb{R}^q} \sum_{j=1}^{n} w_{ij} \log\left\{1 + \frac{\bar{\lambda}_i' g(z_j, \hat{\theta})}{n}\right\}. \tag{5.2}$$

A useful feature of SELR is that it is invariant to nonsingular linear transformations of the moment conditions. Let $C(x, \theta)$ be a $q \times q$ matrix which is nonsingular w.p.1 for every $\theta \in \Theta$. Clearly, $\mathbb{E}\{g(z, \theta_0)|x\} = 0$ if and only if $\mathbb{E}\{C(x, \theta_0) g(z, \theta_0)|x\} = 0$. If the preliminary estimator $\hat{\theta}$ is invariant to this linear transformation [e.g., the maximum smoothed empirical likelihood estimator proposed by Kitamura, Tripathi and Ahn (2002) satisfies this requirement], then it is easy to show that SELR (hence, $\zeta_{1,n}$ and $\zeta_{2,n}$) is also invariant.

Calculating SELR in (5.2) may be computationally demanding as it requires n maximizations. As suggested by a referee, one way to circumvent this problem is to use a one-step approximation for λ_i in constructing SELR. Since $\hat{V}(x_i, \hat{\theta})$ is invertible on S_* w.p.a.1, it is straightforward to verify that when n is large enough and $x_i \in S_*$, a one-step approximation (starting from 0) for the solution to (2.2) is given by $\lambda_{i,(1)} = n\hat{V}^{-1}(x_i, \hat{\theta}) \sum_{j=1}^{n} w_{ij} g(z_j, \hat{\theta})$. Hence, a one-step version of SELR is

$$\text{SELR}_{(1)} = 2\sum_{i=1}^{n} \mathbb{I}\{x_i \in S_*\} \sum_{j=1}^{n} w_{ij} \log\left\{1 + \frac{\lambda'_{i,(1)} g(z_j, \hat{\theta})}{n}\right\}. \tag{5.3}$$

The asymptotic theory for $\zeta_{1,n}$ and $\zeta_{2,n}$ remains unchanged if we substitute $\text{SELR}_{(1)}$ for SELR in (4.2) and (4.3). To see this, examine the proof of Lemma A.1. Note that if we set the remainder term $r_{1,i}$ identically equal to 0 in (A.1), we obtain the one-step approximation $\lambda_{i,(1)}$. Hence, following the rest of the proof, it is easily seen that

$$\text{SELR}_{(1)} = \hat{\mathrm{T}} + O_p\left(\left\{\frac{\log n}{n^{1/3} b_n^s}\right\}^{3/2}\right).$$

Therefore, our asymptotic results for $\zeta_{1,n}$ and $\zeta_{2,n}$ do not change.

To summarize, implementing the test involves the following steps. Step 1: Obtain $\hat{\theta}$, a preliminary estimator of θ_0. Step 2: Pick a bandwidth b_n. Step 3: Use (5.2) to calculate SELR or (5.3) to calculate $\text{SELR}_{(1)}$. Step 4: Depending on the dimension of x, construct $\zeta_{1,n}$ or $\zeta_{2,n}$ as defined in (4.2) and (4.3). Once $\hat{\theta}$ and b_n have been chosen, no other parameters need be estimated. Obtaining $\hat{\theta}$ is straightforward. However, as in any other nonparametric procedure, the choice of b_n requires a little more effort. Suppose, for example, we wish to carry out specification testing for a parametric regression function. In this case, it is natural to cross-validate the average squared errors or a similar goodness-of-fit measure for nonparametric regression to select an appropriate bandwidth. See, for example, Härdle (1990). This strategy covers a large majority of practically interesting situations, and has been used widely in the nonparametric specification testing literature. See, for instance, Hart (1997). If, however, the model is not in a regression form, we need to find an alternative loss function for a bandwidth selector. One possible avenue for exploration is described in LeBlanc and Crowley (1995), Section 3.2. A detailed analysis of automatic or data-driven bandwidth choice for SELR is beyond the scope of the current paper and is left for future research.

Finally, notice that the above discussion makes sense only if (5.1), or, equivalently (2.2), has a solution. A look at (2.1) reveals that a necessary and sufficient condition for the solution to exist is that the origin is contained in the convex hull of $\{g(z_1, \hat{\theta}), \ldots, g(z_n, \hat{\theta})\}$. We now show that this condition holds w.p.a.1 if we assume that $\mathbb{E}\{g(z, \theta_0) g'(z, \theta_0)\}$ exists and has full rank and that

$$\Pr\{z : \xi' g(z, \theta) = 0\} = 0 \qquad \text{for each } (\xi, \theta) \in \mathbb{S}^q \times \mathscr{B}_0, \tag{5.4}$$

where $\mathscr{B}_0$ is some compact neighborhood of θ_0. For example, (5.4) holds whenever $g(z, \theta)$ has a density with respect to the Lebesgue measure for each $\theta \in \mathscr{B}_0$. Let $a(z, \xi, \theta) = \mathbb{I}\{\xi' g(z, \theta) > 0\}$. Since $\hat{\theta}$ is consistent for θ_0, $\sup_{\xi \in \mathbb{S}^q} |n^{-1} \sum_{j=1}^n \mathbb{I}\{\xi' g(z_j, \hat{\theta}) > 0\} - \Pr\{\xi' g(z, \theta_0) > 0\}| \leq (1) + (2)$ holds w.p.a.1, where $(1) = \sup_{(\xi,\theta) \in \mathbb{S}^q \times \mathscr{B}_0} |n^{-1} \sum_{j=1}^n a(z_j, \xi, \theta) - \mathbb{E} a(z, \xi, \theta)|$ and $(2) = \sup_{\xi \in \mathbb{S}^q} |\mathbb{E} a(z, \xi, \hat{\theta}) - \mathbb{E} a(z, \xi, \theta_0)|$. But, under (5.4), $(\xi, \theta) \mapsto a(z, \xi, \theta)$ is continuous on $\mathbb{S}^q \times \mathscr{B}_0$ w.p.1. Hence, by Newey and McFadden [(1994), Lemma 2.4], it follows that $(1) = o_p(1)$ and $(\xi, \theta) \mapsto \mathbb{E} a(z, \xi, \theta)$ is continuous on $\mathbb{S}^q \times \mathscr{B}_0$. Since $\mathbb{S}^q$ is compact, the latter fact implies that $\theta \mapsto \max_{\xi \in \mathbb{S}^q} |\mathbb{E} a(z, \xi, \theta) - \mathbb{E} a(z, \xi, \theta_0)|$ is continuous on $\mathscr{B}_0$. Hence, $(2) = o_p(1)$ by the continuous mapping theorem. Thus, we have shown that $(1) + (2) = o_p(1)$. Owen [(1990), Lemma 2] shows that $\inf_{\xi \in \mathbb{S}^q} \Pr\{\xi' g(z, \theta_0) > 0\} > 0$ provided $\mathbb{E}\{g(z, \theta_0) g'(z, \theta_0)\}$ exists and has full rank. Therefore, $\inf_{\xi \in \mathbb{S}^q} n^{-1} \sum_{j=1}^n \mathbb{I}\{\xi' g(z_j, \hat{\theta}) > 0\} > 0$ holds w.p.a.1. As a consequence, the origin lies in the convex hull of $\{g(z_1, \hat{\theta}), \ldots, g(z_n, \hat{\theta})\}$ w.p.a.1.

6. Limiting behavior under local alternatives. We now derive the asymptotic power function of $\zeta_{1,n}$ and $\zeta_{2,n}$ under a sequence of alternatives that approach the null hypothesis as $n \uparrow \infty$. To generate these local alternatives, we follow the approach of Hong and White [(1995), Section 3]; namely, we keep the joint distribution of (x, z) fixed and assume that there exists a nonstochastic sequence $\theta_{n,0} \in \Theta$ such that

$$\mathrm{H}_{1n} : \mathbb{E}\{g(z, \theta_{n,0}) | x\} = \frac{\delta(x)}{n^{1/2} b_n^{s/4}} \tag{6.1}$$

holds w.p.1 for some $\delta : S \to \mathbb{R}^q$.

Notice that the null hypothesis is obtained if $\delta(x) = 0$. We need some additional assumptions in order to obtain the asymptotic distribution of $\zeta_{1,n}$ and $\zeta_{2,n}$ under the sequence of local alternatives defined in (6.1). For the next assumption, recall the definition of $\mathscr{N}_0$ as given in Assumption 3.6.

ASSUMPTION 6.1. (i) $h(x)$ and $V(x, \theta)$ are twice continuously differentiable on S for $\theta \in \mathscr{N}_0$. (ii) $h(x)$ and $\sup_{\theta \in \mathscr{N}_0} \mathbb{E}\{\|g(z, \theta)\|^m | x\} h(x)$ are uniformly bounded on S. (iii) $D(x, \theta)$ and $V(x, \theta)$ are continuous on $S \times \mathscr{N}_0$. (iv) $\inf_{(\xi, x, \theta) \in \mathbb{S}^q \times S_* \times \mathscr{N}_0} \xi' V(x, \theta) \xi > 0$ and $\sup_{(\xi, x, \theta) \in \mathbb{S}^q \times S_* \times \mathscr{N}_0} \xi' V(x, \theta) \xi < \infty$.

Assumption 6.1 is a generalization of Assumption 3.5.

ASSUMPTION 6.2. (i) $\theta_{n,0}$ is a nonstochastic sequence such that (6.1) holds, and $\|\theta_{n,0} - \theta_0\| \downarrow 0$ as $n \uparrow \infty$. (ii) $\delta : S \to \mathbb{R}^q$ is continuous and $\mathbb{E}\|\delta(x)\|^m < \infty$. (iii) $\hat{\theta}$ is $n^{1/2}$-consistent for $\theta_{n,0}$, that is, $\|\hat{\theta} - \theta_{n,0}\| = O_p(n^{-1/2})$.

Condition (i) ensures that $\theta_{n,0} \in \mathcal{N}_0$ for large enough n so that the regularity conditions in Assumptions 3.6 and 6.1 hold. Continuity of δ and existence of moments in (ii) are required for technical reasons and are used in the proofs. Condition (iii) guarantees that replacing $\hat{\theta}$ by $\theta_{n,0}$ in $\zeta_{1,n}$ and $\zeta_{2,n}$ does not change their asymptotic distribution under H_{1n}.

By Lemma B.1, (4.1) remains valid under H_{1n}. Hence, by Lemmas B.2 and B.4, $\zeta_{1,n} = b_n^{s/2}\hat{\mathrm{T}}_5/\sigma + o_p(1)$ as before. Define $\mu = \mathbb{E}[\mathbb{I}\{x_1 \in S_*\}\delta'(x_1)V^{-1}(x_1, \theta_0) \times \delta(x_1)]$. Using Lemma B.5, we can show the next result.

THEOREM 6.1. *Let Assumptions* 3.1, 3.2, 3.6, 3.7, 6.1 *and* 6.2 *hold. Furthermore, assume that* $b_n = n^{-\alpha}$ *for* $0 < \alpha < \min\{\frac{1}{s}(1 - \frac{4}{m}), \frac{1}{3s}, \frac{1}{s}(1 - \frac{2}{m} - \frac{2}{\eta})\}$. *Then* $\zeta_{1,n} \xrightarrow{d} \mathrm{N}(\mu/\sigma, 1)$ *under* H_{1n}.

Therefore, the asymptotic local power function of a size-γ test using $\zeta_{1,n}$ is given by $1 - \Phi(c_\gamma - \frac{\mu}{\sigma})$, where $\Phi(t) = \Pr\{\mathrm{N}(0, 1) \le t\}$ and $\Phi(c_\gamma) = 1 - \gamma$. When $s \le 3$, a similar result holds for $\zeta_{2,n}$.

COROLLARY 6.1. *Let Assumptions* 3.1, 3.2, 3.6, 3.7, 6.1 *and* 6.2 *hold. Furthermore, assume that* $s \le 3$ *and* $b_n = n^{-\alpha}$ *for* $0 < \alpha < \min\{\frac{1}{s}(1 - \frac{4}{m}), \frac{1}{3s}, \frac{1}{s}(1 - \frac{2}{m} - \frac{2}{\eta})\}$. *Then* $\zeta_{2,n} \xrightarrow{d} \mathrm{N}(\mu/\sigma, 1)$ *under* H_{1n}.

7. Asymptotic optimality of the SELR test. As noted in the Introduction, there are alternative tests for conditional moment restrictions available in the literature. All of these tests are nonparametric and are consistent against general alternatives. There is, of course, a price one pays for this generality: nonparametric tests tend to have lower power than parametric ones. Therefore, it is important to find a nonparametric test with good power properties.

This section identifies an optimal test among a class of conditional moment restrictions tests. Aït-Sahalia, Bickel and Stoker (2001) provide a convenient framework for this purpose. They consider a testing procedure based on a weighted sum of squared residuals from kernel regression. Many earlier tests, at least asymptotically, can be regarded as a special case of this test with a particular choice of weighting function. Härdle and Mammen (1993), Fan and Li (1996), Zheng (1996) and our SELR test, for example, fall into this category. Hong and White (1995) apply a similar principle, though they use series instead of kernels.

To simplify our argument, let $q = 1$, $s = 1$ and $S_* = [0, 1]$. In implementing the test of Aït-Sahalia, Bickel and Stoker, the researcher chooses a piecewise smooth, bounded and square integrable weight function $a : [0, 1] \to \mathbb{R}_+$ and calculates

$G(a) = b_n \sum_{i=1}^n \hat{\mathbb{E}}^2\{g(z,\hat\theta)|x_i\}a(x_i)$, where $\hat{\mathbb{E}}\{g(z,\hat\theta)|x_i\} = \sum_{j=1}^n w_{ij} g(z_j,\hat\theta)$. The statistic for testing H_0 proposed by Aït-Sahalia, Bickel and Stoker is

$$\tau(a) = \frac{b_n^{-1/2}\{G(a) - \mathrm{R}(K)\int_0^1 V(x,\theta_0)a(x)\,dx\}}{\sqrt{2K^{**}\int_0^1 V^2(x,\theta_0)a^2(x)\,dx}}. \tag{7.1}$$

We can replace $V(x,\theta_0)$ with an appropriate consistent estimator without affecting the asymptotic properties of the test. Since $\tau(ca) = \tau(a)$ for any $c \neq 0$, without loss of generality we assume that $\int_0^1 a^2(x)\,dx = 1$. Now let

$$M(a,\delta) = \frac{\int_0^1 \delta^2(x)a(x)h(x)\,dx}{\sqrt{2K^{**}\int_0^1 V^2(x,\theta_0)a^2(x)\,dx}}. \tag{7.2}$$

As Aït-Sahalia, Bickel and Stoker show, under H_{1n},

$$\tau(a) \xrightarrow{d} \mathrm{N}\big(M(a,\delta), 1\big). \tag{7.3}$$

The asymptotic power of their test with critical value c_γ is thus given by

$$\pi(a,\delta) = 1 - \Phi\big(c_\gamma - M(a,\delta)\big). \tag{7.4}$$

Comparing (7.4) and Theorem 6.1, we can see that our SELR test is asymptotically equivalent to the $\tau(a)$ test with the weighting scheme

$$a_{\mathrm{SELR}}(x) = \frac{1}{V(x,\theta_0)\sqrt{\int_0^1 V^{-2}(x,\theta_0)\,dx}}. \tag{7.5}$$

We shall demonstrate that this choice of weighting, which is implicitly achieved by the SELR test, is optimal in a certain sense.

If δ is known counterfactually, it is easy to derive the optimal weighting function that maximizes (7.2). For a known δ, an application of the Cauchy–Schwarz inequality on (7.2) shows that (7.4) is maximized by choosing

$$a(x,\delta) = \frac{\delta^2(x)h(x)}{V^2(x,\theta_0)\sqrt{\int_0^1 \delta^4(x)V^{-4}(x,\theta_0)h^2(x)\,dx}}. \tag{7.6}$$

The notation $a(x,\delta)$ indicates that the optimal choice of a depends on δ. This result is not terribly useful since δ is unknown in practice. It is also clear from (7.6) that there is no uniformly (in δ) optimal test. This resembles the multiparameter optimal testing problem considered in the seminal paper of Wald (1943).

Wald shows that the likelihood ratio test, and other asymptotically equivalent tests, for a hypothesis about finite-dimensional parameters is optimal in terms of an average power criterion. Loosely put, he considers a weighted average of the power function where uniform weights are given along each probability contour of the distribution of the estimator he uses (MLE). This criterion is natural and attractive since it is impartial—it puts heavy (light) weights in directions where the detection of departures from the null is difficult (easy). This approach has been used in the literature quite effectively. For example, Andrews and Ploberger (1994) consider optimal inference in a nonstandard testing problem. They derive a test that is optimal with respect to a Wald-type average power criterion. Their optimal test performs well in finite samples [see Andrews and Ploberger (1996)], indicating the practical relevance of Wald's approach.

Our testing problem is different from the ones considered by Wald in that instead of being finite dimensional, our parameter of interest is an unknown function. A natural extension of Wald's approach is to consider a probability measure on an appropriate space of functions and let the measure mimic the distribution of the "estimator." Then the local average power criterion is obtained by integrating (7.4) against the probability measure. Note that the tests we are comparing rely on the kernel regression estimator $\hat{\mathbb{E}}\{g(z,\hat{\theta})|x\}$, either explicitly or implicitly. Therefore, we propose to use a probability measure that approximates the asymptotic distribution of the sample path of $\hat{\mathbb{E}}\{g(z,\hat{\theta})|x\}$.

So let $\tilde{\delta}$ be a C([0, 1])-valued random variable given by $\tilde{\delta}(x) = V^{1/2}(x,\theta_0) \times h^{-1/2}(x)y(x)$, where $y(x) = \int_0^1 k(\frac{x}{\beta} - z)\, dW(z - \lfloor z \rfloor)$, W is the standard Brownian motion on [0, 1], $k(\cdot)$ an appropriate weighting function, β a positive adjustable parameter and $\lfloor z \rfloor$ the integer part of z. For each x in [0, 1], $y(x)$ is a stochastic integral. Note the use of $dW(z - \lfloor z \rfloor)$ as the integrator. This implies that the covariance kernel $r(s) = \mathbb{E}[y(x)y(x+s)]$ of the Gaussian process y is circular, that is, $r(s) = r(1-s)$. Circular processes are widely used for analyzing stationary processes on a finite interval [see, e.g., Hannan (1970) and Priestley (1981)]. In our case, it lets us avoid treating $y(x)$'s close to the end points of the interval [0, 1] differently from the ones in the middle. Consequently, for an arbitrary function f such that the integral $\int_0^1 f(y(x))\, dx$ is well defined, the joint distribution of the bivariate random vector $(\int_0^1 f(y(x))\, dx, y(x_0))$ does not depend on the location $x_0 \in [0, 1]$. Other properties of $\tilde{\delta}$, such as its Gaussianity, are not important in our argument below.

Note that the variance function of $\tilde{\delta}(x)$ coincides with the asymptotic variance function of $\hat{\mathbb{E}}\{g(z,\hat{\theta})|x\}$ up to scale. This is one of the features we intend to replicate by using $\tilde{\delta}$. The Gaussian process $\tilde{\delta}$ is based on an approximation of $\hat{\mathbb{E}}\{g(z,\hat{\theta})|x\}$ derived by Liero (1982). Also see Johnston (1982) and Härdle (1989) for related results. In our theory, however, k and β do not have to be the same as K and b_n. Here k determines the pattern of autocorrelations of $y(x)$ and β is used for scaling x. A large β and a spread-out k correspond to stronger dependence, yielding paths of y and $\tilde{\delta}$ that look smoother. Our optimality result does not depend on the choice of β and k.

We are now ready to define our average power concept. Let Q be the probability measure induced by $\tilde{\delta}$ on C([0, 1]). Using the definition of $\tilde{\delta}$, rewrite the random variable $M(a, \tilde{\delta})$ as

$$M(a, \tilde{\delta}) = \frac{\int_0^1 V(x, \theta_0) y^2(x) a(x)\, dx}{\sqrt{2K^{**} \int_0^1 V^2(x, \theta_0) a^2(x)\, dx}} = \frac{1}{\sqrt{2K^{**}}} \int_0^1 A(x) y^2(x)\, dx,$$

where

$$A(x) = \frac{V(x, \theta_0) a(x)}{\sqrt{\int_0^1 V^2(x, \theta_0) a^2(x)\, dx}}. \tag{7.7}$$

Note that $\int_0^1 A^2(x)\, dx = 1$ and it is sometimes convenient to deal with A rather than a. Note also that $M(a, \tilde{\delta}) = M(A/V, \tilde{\delta})$. Let F_A be the c.d.f. of $M(A/V, \tilde{\delta})$. The average asymptotic power of the test proposed by Aït-Sahalia, Bickel and Stoker (2001) [see (7.4)] is the following functional of A:

$$\bar{\pi}(A) = \int \pi(A/V, \tilde{\delta})\, dQ(\tilde{\delta}) = \int_0^\infty [1 - \Phi(c_\gamma - m)] F_A(dm). \tag{7.8}$$

Observe that the integrand in (7.8) is strictly increasing in m. So if there exists a piecewise smooth, bounded, square integrable function $A^* : [0, 1] \mapsto \mathbb{R}_+$ such that $\int_0^1 A^{*2}(x)\, dx = 1$ and for all A the c.d.f. F_{A^*} first-order stochastically dominates F_A [i.e., $F_A(m) \geq F_{A^*}(m)$ for all m], then A^* maximizes $\bar{\pi}(A)$. By (7.7), the optimal weighting function a^* is given by

$$a^*(x) = \frac{A^*(x)}{V(x, \theta_0) \sqrt{\int A^{*2}(x)/V^2(x, \theta_0)\, dx}}.$$

To find A^*, fix $m \in \mathbb{R}$ arbitrarily and consider solving the following variational problem over all piecewise smooth, bounded, square integrable functions from $[0, 1] \to \mathbb{R}_+$:

$$\min_A F_A(m) \quad \text{s.t.} \quad \int_0^1 A^2(x)\, dx = 1. \tag{7.9}$$

For any $x_0 \in [0, 1]$, let $F_A(m|y(x_0))$ be the conditional c.d.f. of $M(A/V, \tilde{\delta})$ given $y(x_0)$. Let $f_A(m|y(x_0))$ be the conditional p.d.f. corresponding to $F_A(m|y(x_0))$. Now it is clear that $F_A(m) = \mathbb{E}_{y(x_0)}[F_A(m|y(x_0))]$, where the symbol $\mathbb{E}_{y(x_0)}$ indicates that the expectation is over $y(x_0)$. Furthermore,

$$\frac{\partial F_A(m|y(x_0))}{\partial A(x_0)} = \frac{\partial \mathbb{E}[\mathbb{I}\{\int_0^1 A(x) y^2(x)\, dx < m\} | y(x_0)]}{\partial A(x_0)} = y^2(x_0) f_A\big(m|y(x_0)\big).$$

These results imply that

$$\frac{\partial F_A(m)}{\partial A(x_0)} = \mathbb{E}_{y(x_0)}\big[y^2(x_0) f_A\big(m|y(x_0)\big)\big] \qquad \text{for all } x_0 \in [0, 1].$$

Thus, the Euler–Lagrange equation for the variational problem (7.9) is

$$\mathbb{E}_{y(x_0)}\big[y^2(x) f_{A^*}\big(m|y(x_0)\big)\big] = 2\lambda A^*(x_0) \qquad \text{for all } x_0 \in [0,1], \tag{7.10}$$

where λ is the Lagrange multiplier for the constraint in (7.9) and A^* the solution. To solve (7.10), we use a guess-and-verify approach. So suppose that $A^*(x) = \mathbb{I}\{x \in [0,1]\}$. Clearly, this is a feasible guess. As noted in our earlier discussion on the nature of the random process y, the joint distribution of $M(A^*/V, \tilde{\delta}) = (1/\sqrt{2K^{**}}) \int_0^1 y^2(x)\,dx$ and $y(x_0)$ does not depend on $x_0 \in [0,1]$. Therefore, $\mathbb{E}_{y(x_0)}[y^2(x_0) f_{A^*}(m|y(x_0))] \overset{\text{def}}{=} K$ (say) does not depend on $x_0 \in [0,1]$. So (7.10) is satisfied with $A^*(x) = \mathbb{I}\{x \in [0,1]\}$ and $\lambda = K/2$. We have verified that $A^*(x) = \mathbb{I}\{x \in [0,1]\}$ solves (7.9). The optimal a corresponding to $A^*(x) = \mathbb{I}\{x \in [0,1]\}$ is

$$a^*(x) = \frac{\mathbb{I}\{x \in [0,1]\}}{V(x,\theta_0)\sqrt{\int_0^1 V^{-2}(x,\theta_0)\,dx}}.$$

Comparing this with (7.5), we immediately obtain that the weight a_{SELR} is optimal.

The above result shows that the SELR test attains maximum average local power. An alternative way of achieving this optimality is to estimate a^* by

$$\hat{a}^*(x) = \frac{\mathbb{I}\{x \in [0,1]\}}{\hat{V}(x,\hat{\theta})\sqrt{\int_0^1 \hat{V}^{-2}(x,\hat{\theta})\,dx}}.$$

We then use $\hat{a}^*$ to calculate G for the test statistic in (7.1). While this approach is valid asymptotically, such a "plug-in" method often leads to poor finite-sample behavior. At the very least, it would require a good nonparametric estimator of $V(x,\theta_0)$. An advantage of our statistic over plug-in statistics is that this optimal weighting is carried out automatically and implicitly, eliminating the need for estimating $V(x,\theta_0)$. This feature is similar to the "internal Studentization" property of other empirical likelihood ratio statistics emphasized in the literature. Empirical evidence suggests that internal Studentization often improves finite-sample properties of the tests substantially. See, for example, Fisher, Hall, Jing and Wood (1996).

8. Simulation experiments. This section reports some experimental evidence on the finite-sample performance of the SELR test against two well-known competitors.

8.1. *Scope of the simulation study.* We compare the SELR test with two tests considered in the Horowitz and Spokoiny (2001) simulation study, namely, the tests by Härdle and Mammen (1993) and Horowitz and Spokoiny (2001).

The Härdle–Mammen test is a kernel-based test. It is widely used and is often considered as a benchmark of nonparametric conditional mean specification tests. Also, their test and the SELR tests can be put in the asymptotic framework used in Section 7, where we have demonstrated that the SELR test is optimal in Wald's sense. It is therefore interesting to investigate the performance of our test relative to the Härdle–Mammen test in finite samples.

The Horowitz–Spokoiny test is also kernel based. It is based on nonparametric goodness-of-fit statistics calculated over a range of bandwidths. Horowitz and Spokoiny show that their test is adaptive and rate optimal (i.e., it is uniformly consistent at the fastest possible rate). Our test complements, rather than substitutes, for their test. The result in Section 7 suggests that SELR has a desirable theoretical property under a sequence of local alternatives, though it is not rate optimal. In contrast, Horowitz and Spokoiny obtain a test that is adaptive and rate optimal, though they do not discuss the local power of their test. By adopting Horowitz and Spokoiny's strategy and using many bandwidths, it may be possible to construct an adaptive and rate-optimal version of the SELR test. However, such an extension is beyond the scope of the current paper.

Finally, it should be noted that the tests of Andrews (1997) and Bierens and Ploberger (1997) (which do not require any nonparametric smoothing) are consistent against alternatives of the form $n^{-1/2}\delta(x)$. Other Cramér–von Mises-type and Kolmogorov–Smirnov-type tests typically share this property as well. However, as Härdle and Mammen [(1993), page 1931] point out, "(t)hese tests ...are of more parametric nature—in the sense that they look into certain one-dimensional directions." This point, that is, the relative merits of specification tests with smoothing over tests without smoothing, has also been emphasized by Hart (1997) and other researchers. Indeed, Horowitz and Spokoiny report that the Andrews test is dominated by the Härdle–Mammen test in terms of power, uniformly over their experimental designs. But, as we shall see immediately, the Härdle–Mammen test is, in turn, dominated by the SELR test uniformly in the same experimental designs. This fact provides useful evidence on the finite-sample performance of the SELR test compared with tests like that of Andrews.

8.2. *Simulation design.* Our simulation design is nearly identical to the design used by Horowitz and Spokoiny (2001). The null hypothesis specification takes the form $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, where $\beta_0 = \beta_1 = 1$ and $x_i \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, 25)$ with its 5% upper and lower tails truncated. In our simulation study, a series of x_i (for $i = 1, 2, \ldots, 250$) is drawn for each Monte Carlo replication. This is the main difference between our experiments and those of Horowitz and Spokoiny, who generated a series of i.i.d. draws $\{x_i\}_{i=1}^{250}$ from $\mathrm{N}(0, 25)$ once and then kept it fixed throughout the simulations. Note that ε_i is i.i.d. and independent of x_i. We experiment with three specifications for ε_i as used by Horowitz and Spokoiny: normal with mean 0 and variance 4, mixture of normals [$\mathrm{N}(0, 1.56)$ with probability $9/10$ and $\mathrm{N}(0, 25)$ with probability $1/10$] and Type I extreme value distribution with variance 4.

We also investigate finite-sample power properties of the three tests under the alternatives $y_i = \beta_0 + \beta_1 x_i + (c/\tau)\phi(x_i/\tau) + \varepsilon_i$, where ϕ denotes the standard normal density, $\tau = 0.25$, 1 or 2 and $c = 2.5$ or 5. This is the same specification of alternatives used by Horowitz and Spokoiny, though they did not consider the cases $\tau = 2$ and $c = 2.5$. The parameters τ and c control the shape of the deviation from the linear null model. For example, it is narrowly peaked for small values of τ.

The OLS estimator is used to estimate β_0 and β_1; then the three tests are carried out. The Gaussian kernel is used for all of the three tests. A bandwidth needs to be specified to calculate the Härdle–Mammen statistic and SELR. To make our experiment comparable to Horowitz and Spokoiny's, and to reduce the computational burden at the same time, we set $b_n = 3.5$, which is the bandwidth value used by Horowitz and Spokoiny. The critical values for the Härdle–Mammen and SELR tests are obtained using the wild bootstrap procedure described in Härdle and Mammen (1993). The number of bootstrap replications is 99. The Horowitz–Spokoiny statistic is obtained by taking the maximum of a Studentized goodness-of-fit statistic over the set of bandwidths $\{2.5, 3, 3.5, 4, 4.5\}$. Its critical values are obtained via simulations; see Horowitz and Spokoiny (2001) for details on the implementation of their test. The number of observations is set to 250 throughout the experiments. The number of Monte Carlo replications is 1000 for the null, and 250 for each alternative.

8.3. *Simulation results.* Our simulation results are summarized in Table 1. The reported figures are simulated rejection probabilities of the three tests at the 5% significance level. The first panel shows simulation results under the null. The three tests perform well in size. All of the rejection frequencies are within two simulation standard errors from the nominal size of 0.05.

The middle panel tabulates the results for alternatives with $c = 5$. The distributional specification of ε has some impact on rejection rates, though rankings among the three tests are robust with respect to the distribution of ε. When the alternative hypothesis consists of a smooth bump ($\tau = 2$), SELR is most powerful among the three tests. For a narrowly peaked alternative ($\tau = 0.25$), the Horowitz–Spokoiny (H–S) test performs very well, though the power of the SELR test is satisfactory and it is more powerful than the Härdle–Mammen (H–M) test.

A similar observation applies to the alternatives with $c = 2.5$ (the bottom panel). In this case, the peak of the alternative is quite spread out even for $\tau = 1$. SELR and the Horowitz–Spokoiny test are equally powerful for this case, and the Härdle–Mammen test is considerably less powerful. For $\tau = 0.25$, the Horowitz–Spokoiny test ranks first, SELR second and the Härdle–Mammen test third.

The computational burden of the simulation exercise limits its scope. Nevertheless, the finite-sample behavior of the SELR test documented above is encouraging. Our test is more powerful than the Härdle–Mammen test for all of the alternatives considered here and in the Horowitz–Spokoiny paper. As the SELR and Härdle–Mammen tests belong to the same class of nonparametric specification tests, this comparison is informative. The Horowitz–Spokoiny test works

TABLE 1
Monte Carlo results: *nominal size* = 0.05; $n = 250$; $b_n = 3.5$

		Probability of rejecting H_0		
Distribution of ε	τ	**SELR test**	**H–M test**	**H–S test**
Null hypothesis is true (1000 reps.)				
Normal	—	0.057	0.058	0.053
Mixture	—	0.060	0.050	0.050
Extreme value	—	0.043	0.046	0.048
Null hypothesis is false (250 reps.); $c = 5$				
Normal	2.00	0.716	0.688	0.676
Mixture	2.00	0.760	0.688	0.708
Extreme value	2.00	0.756	0.684	0.704
Normal	1.00	0.964	0.932	0.976
Mixture	1.00	0.968	0.912	0.980
Extreme value	1.00	0.996	0.948	1.000
Normal	0.25	0.948	0.940	0.984
Mixture	0.25	0.948	0.908	0.984
Extreme value	0.25	0.956	0.908	0.992
Null hypothesis is false (250 reps.); $c = 2.5$				
Normal	1.00	0.508	0.420	0.496
Mixture	1.00	0.536	0.404	0.488
Extreme value	1.00	0.548	0.428	0.552
Normal	0.25	0.584	0.468	0.660
Mixture	0.25	0.600	0.492	0.704
Extreme value	0.25	0.604	0.488	0.740

very well, especially for peaked alternatives, though it is less powerful than the SELR test for smooth alternatives. As noted previously, the SELR test and the Horowitz–Spokoiny test are not substitutes but rather complements. Recall that the Horowitz–Spokoiny test is based on the maximum of a version of the Härdle–Mammen statistic calculated over a set of bandwidths. The good performance of the SELR test relative to the Härdle–Mammen test indicates the potential usefulness of SELR in the context of Horowitz–Spokoiny-type rate-optimal testing.

9. Conclusion. The results obtained in this paper show that the SELR test is easy to construct and straightforward to implement. It is asymptotically normal under the null hypothesis, has nontrivial local power under a sequence of local alternatives and is asymptotically optimal in terms of an average power criterion. Simulation evidence suggests that our test behaves well in finite samples.

APPENDIX A

Asymptotic theory under the null. For the remainder of the paper, c denotes a generic constant, $g_*(z_j) = \sup_{\theta\in\Theta} \|g(z_j,\theta)\|$, $I_* = \{1 \le i \le n : x_i \in S_*\}$, $\hat{h}(x_i) = \sum_{j=1}^n K_{ij}/(nb_n^s)$, $\hat{\Omega}(x_i,\theta) = \sum_{j=1}^n K_{ij} g(z_j,\theta) g'(z_j,\theta)/(nb_n^s)$, $\tilde{H}_n(x_i,\theta_0) = \mathbb{E}\{\hat{\Omega}(x_i,\theta_0)|x_i\}\mathbb{E}\{\hat{h}(x_i)|x_i\}$, $\hat{H}(x_i,\theta) = \hat{V}(x_i,\theta)\hat{h}^2(x_i)$, $H(x_i,\theta) = V(x_i,\theta) \times h^2(x_i)$ and $A_{tj} = \sum_{i=1, i\neq j\neq t}^n \mathbb{I}_i K_{ij} \tilde{H}_n^{-1}(x_i,\theta_0) K_{it}$.

LEMMA A.1. *Let Assumptions* 3.1–3.7 *hold. Assume that* $b_n = n^{-\alpha}$ *for* $0 < \alpha < \frac{1}{s}(1 - \frac{4}{m})$. *Then*

$$\mathrm{SELR} = \hat{\mathrm{T}} + o_p\left(\left\{\frac{\log n}{n^{1/2-1/m} b_n^s}\right\}^2\right) + o_p\left(\frac{1}{n^{1-2/m}}\right) + O_p\left(\left\{\frac{\log n}{n^{1/3} b_n^s}\right\}^{3/2}\right) \qquad \textit{under } \mathrm{H}_0,$$

where $\hat{\mathrm{T}} = \sum_{i=1}^n \mathbb{I}_i\{\sum_{j=1}^n w_{ij} g'(z_j,\hat{\theta})\}\hat{V}^{-1}(x_i,\hat{\theta})\{\sum_{j=1}^n w_{ij} g(z_j,\hat{\theta})\}$.

PROOF. Our proof follows Owen [(1990), pages 100–102]. However, unlike Owen, we obtain nonparametric (i.e., slower than $n^{1/2}$) rates of convergence. Since λ_i solves (2.2),

$$\begin{aligned} 0 &= \sum_{j=1}^n \frac{w_{ij} g(z_j,\hat{\theta})}{n + \lambda_i' g(z_j,\hat{\theta})} \\ &= \frac{1}{n}\sum_{j=1}^n w_{ij} g(z_j,\hat{\theta}) - \frac{1}{n^2}\hat{V}(x_i,\hat{\theta})\lambda_i + \frac{1}{n}\sum_{j=1}^n \frac{w_{ij} g(z_j,\hat{\theta})(\lambda_i' g(z_j,\hat{\theta})/n)^2}{1 + (\lambda_i' g(z_j,\hat{\theta})/n)}. \end{aligned}$$

By Lemma C.2(ii), $\hat{V}(x_i,\hat{\theta})$ is invertible on S_* w.p.a.1. Consequently,

$$\mathbb{I}_i\lambda_i = n\mathbb{I}_i \hat{V}^{-1}(x_i,\hat{\theta})\sum_{j=1}^n w_{ij} g(z_j,\hat{\theta}) + \mathbb{I}_i \hat{V}^{-1}(x_i,\hat{\theta}) r_{1,i} \tag{A.1}$$

holds w.p.a.1, where

$$r_{1,i} = \sum_{j=1}^n \frac{w_{ij} g(z_j,\hat{\theta})(\lambda_i' g(z_j,\hat{\theta}))^2}{n + \lambda_i' g(z_j,\hat{\theta})}.$$

Equation (2.2) also shows that

$$\sum_{j=1}^n \frac{w_{ij}(\lambda_i' g(z_j,\hat{\theta}))^2}{n + \lambda_i' g(z_j,\hat{\theta})} = \sum_{j=1}^n w_{ij}\lambda_i' g(z_j,\hat{\theta}). \tag{A.2}$$

Hence, as $n + \lambda_i' g(z_j, \hat{\theta}) \geq 0$ (because $\hat{p}_{ij} \geq 0$),

$$\|r_{1,i}\| \leq \max_{1 \leq j \leq n} \|g(z_j, \hat{\theta})\| \sum_{j=1}^{n} w_{ij} \lambda_i' g(z_j, \hat{\theta})$$

$$\stackrel{\text{Lemma C.4}}{=} o(n^{1/m}) \left\| \sum_{j=1}^{n} w_{ij} g(z_j, \hat{\theta}) \right\| \|\lambda_i\|,$$

where the $o(n^{1/m})$ term does not depend on i, j or $\theta \in \Theta$. Now assume that n is large enough so that $\hat{\theta} \in \mathcal{N}_0$ and our regularity conditions hold. By Assumption 3.6,

$$g(z_j, \hat{\theta}) = g(z_j, \theta_0) + \operatorname{rem}(z_j, \hat{\theta} - \theta_0) \text{ w.p.1}, \quad \text{where } \|\operatorname{rem}(z_j, \hat{\theta} - \theta_0)\| \leq d(z_j)\|\hat{\theta} - \theta_0\|. \tag{A.3}$$

Hence,

$$\mathbb{I}_i \left\| \sum_{j=1}^{n} w_{ij} g(z_j, \hat{\theta}) \right\| \leq \max_{i \in I_*} \left\| \sum_{j=1}^{n} w_{ij} g(z_j, \theta_0) \right\| + \|\hat{\theta} - \theta_0\| \sum_{j=1}^{n} d(z_j) w_{ij}, \tag{A.4}$$

which implies $\mathbb{I}_i \|r_{1,i}\| = o(n^{1/m}) \{\max_{i \in I_*} \|\sum_{j=1}^n w_{ij} g(z_j, \theta_0)\| + \|\hat{\theta} - \theta_0\| \times \sum_{j=1}^n d(z_j) w_{ij}\} \mathbb{I}_i \|\lambda_i\|$. Next, let $\lambda_i = \rho_i \xi_i$, where $\rho_i \geq 0$ and $\xi_i \in \mathbb{S}^q$. Observe that

$$0 \leq n + \lambda_i' g(z_j, \hat{\theta}) \leq n + \rho_i \|g(z_j, \hat{\theta})\| \stackrel{\text{Lemma C.4}}{=} n + \rho_i o(n^{1/m}).$$

Under our choice of b_n, $\max_{1 \leq i \leq n} |\xi_i' \hat{V}(x_i, \hat{\theta}) \xi_i - \xi_i' V(x_i, \theta_0) \xi_i| = o_p(1)$ by Lemma C.2(i). Hence, as $\xi_i' V(x_i, \theta_0) \xi_i$ is bounded away from 0 on $(\xi_i, x_i) \in \mathbb{S}^q \times S_*$, by (A.2) and (A.4),

$$\frac{\mathbb{I}_i \rho_i}{n + \rho_i o(n^{1/m})} \leq \frac{\mathbb{I}_i \sum_{j=1}^n w_{ij} \xi_i' g(z_j, \hat{\theta})}{\xi_i' \hat{V}(x_i, \hat{\theta}) \xi_i}$$

$$= O_p(1) \mathbb{I}_i \left\{ \max_{i \in I_*} \left\| \sum_{j=1}^{n} w_{ij} g(z_j, \theta_0) \right\| + \|\hat{\theta} - \theta_0\| \sum_{j=1}^{n} d(z_j) w_{ij} \right\},$$

where the $O_p(1)$ term does not depend on $i \in I_*$. By Lemma C.1,

$$\max_{i \in I_*} \left\| \sum_{j=1}^{n} w_{ij} g(z_j, \theta_0) \right\| = O_p(c_n),$$

where $c_n \stackrel{\text{def}}{=} \sqrt{\log n / n b_n^s}$. By Lemma C.6, $\max_{1 \leq i \leq n} \sum_{j=1}^n d(z_j) w_{ij} = o(n^{1/\eta})$ holds w.p.1 as $n \uparrow \infty$. But $n^{1/m} c_n \downarrow 0$ and $1/m + 1/\eta \leq 1/2$ under our assumptions. Hence, solving for ρ_i, we obtain

$$\mathbb{I}_i \rho_i = O_p(n)\left\{\max_{i\in I_*}\left\|\sum_{j=1}^n w_{ij} g(z_j,\theta_0)\right\| + \|\hat\theta-\theta_0\|\sum_{j=1}^n d(z_j)w_{ij}\right\}, \tag{A.5}$$

where the $O_p(n)$ term does not depend on $i \in I_*$. Thus, by Jensen's inequality,

$$\mathbb{I}_i\|r_{1,i}\| = o_p(n^{1+1/m})\left\{\left[\max_{i\in I_*}\left\|\sum_{j=1}^n w_{ij} g(z_j,\theta_0)\right\|\right]^2 + \|\hat\theta-\theta_0\|^2\sum_{j=1}^n d^2(z_j)w_{ij}\right\},$$

where the $o_p(n^{1+1/m})$ term does not depend on $i \in I_*$. Since $\max_{i\in I_*}\|\hat V^{-1}(x_i, \hat\theta)\| = O_p(1)$ by Lemma C.2(ii), (A.1) can be written as

$$\mathbb{I}_i\lambda_i = n\mathbb{I}_i \hat V^{-1}(x_i,\hat\theta)\sum_{j=1}^n w_{ij} g(z_j,\hat\theta) + \mathbb{I}_i r_{2,i}, \tag{A.6}$$

where

$$\begin{aligned}\mathbb{I}_i\|r_{2,i}\| &= o_p(n^{1+1/m})\\ &\times\left\{\left[\max_{i\in I_*}\|\left\|\sum_{j=1}^n w_{ij} g(z_j,\theta_0)\right\|\right]^2 + \|\hat\theta-\theta_0\|^2\sum_{j=1}^n d^2(z_j)w_{ij}\right\}.\end{aligned} \tag{A.7}$$

For $u > -1$, $\log(1+u) = u - u^2/2 + \bar\eta$ holds by a Taylor expansion, and the remainder term $|\bar\eta| \le c|u|^3$ if $|u|$ is bounded away from 1. By (A.5) and Lemmas C.4 and C.6, $\max_{1\le i,j\le n}|\lambda_i' g(z_j,\hat\theta)/n| = o_p(1)$. Hence, w.p.a.1, we can write

$$\log\left(1+\frac{\lambda_i' g(z_j,\hat\theta)}{n}\right) = \frac{\lambda_i' g(z_j,\hat\theta)}{n} - \frac{1}{2}\left(\frac{\lambda_i' g(z_j,\hat\theta)}{n}\right)^2 + \bar\eta_{ij}, \tag{A.8}$$

where

$$|\bar\eta_{ij}| \le c\left|\frac{\lambda_i' g(z_j,\hat\theta)}{n}\right|^3 \le cn^{-3}\|\lambda_i\|^3\,\|g(z_j,\hat\theta)\|^3 \le cn^{-3}\rho_i^3 g_*^3(z_j). \tag{A.9}$$

Using (2.4), (A.6) and (A.8), a little algebra shows that, w.p.a.1,

$$\mathrm{SELR} = \hat{\mathrm{T}} - \frac{1}{n^2}\sum_{i=1}^n \mathbb{I}_i r_{2,i}'\hat V(x_i,\hat\theta)r_{2,i} + 2\sum_{i=1}^n\sum_{j=1}^n \mathbb{I}_i w_{ij}\bar\eta_{ij}. \tag{A.10}$$

Since $\max_{i\in I_*}\|\hat V(x_i,\hat\theta)\| = O_p(1)$ by Lemma C.2(i), $\|\sum_{i=1}^n \mathbb{I}_i r_{2,i}'\hat V(x_i,\hat\theta)r_{2,i}\| = O_p(1)\|\sum_{i=1}^n \mathbb{I}_i\|r_{2,i}\|^2$. But $\sum_{i=1}^n \mathbb{I}_i\|r_{2,i}\|^2 = o_p(n^{2+2/m})\{O_p(nc_n^4) + \|\hat\theta-\theta_0\|^4 \times \sum_{i=1}^n\sum_{j=1}^n d^4(z_j)w_{ij}\}$ by (A.7), Lemma C.1 and Jensen's inequality. Lemma C.5 shows that $\sum_{i=1}^n\sum_{j=1}^n d^4(z_j)w_{ij} = O_p(n)$. Therefore,

$$\frac{1}{n^2}\sum_{i=1}^n \mathbb{I}_i r_{2,i}'\hat V(x_i,\hat\theta)r_{2,i} = o_p(n^{1+2/m}c_n^4) + o_p\left(\frac{1}{n^{1-2/m}}\right). \tag{A.11}$$

Next, by (A.5), (A.9), Lemma C.1 and Jensen's inequality,

$$\left|\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{I}_i w_{ij}\bar{\eta}_{ij}\right| = O_p(nc_n^3)\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} g_*^3(z_j)w_{ij} + O_p(n\|\hat{\theta}-\theta_0\|^3)\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{n} d^3(z_j)w_{ij}\right)\left(\sum_{j=1}^{n} g_*^3(z_j)w_{ij}\right).$$

But by the Cauchy–Schwarz and Jensen inequalities,

$$\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{j=1}^{n} d^3(z_j)w_{ij}\right)\left(\sum_{j=1}^{n} g_*^3(z_j)w_{ij}\right) \le \left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} d^6(z_j)w_{ij}\right\}^{1/2}\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} g_*^6(z_j)w_{ij}\right\}^{1/2}.$$

Hence, by Lemma C.5, $|\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{I}_i w_{ij}\bar{\eta}_{ij}| = O_p(nc_n^3) + O_p(n^{-1/2}) = O_p(nc_n^3)$. The desired result now follows by (A.10) and (A.11). □

LEMMA A.2. *Let Assumptions* 3.1–3.7 *hold. Then* $\hat{\mathrm{T}}_1 = O_p(1/nb_n^{2s})$ *under* H_0.

PROOF. Since

$$|\hat{\mathrm{T}}_1| \le \max_{i\in I_*}\|\hat{H}^{-1}(x_i,\hat{\theta})\|\frac{K^2(0)}{n^2 b_n^{2s}}\sum_{i=1}^{n} g_*^2(z_i),$$

the desired result follows from that fact that $\max_{i\in I_*}\|\hat{H}^{-1}(x_i,\hat{\theta})\| = O_p(1)$. □

LEMMA A.3. *Let Assumptions* 3.1–3.7 *hold. Assume that* $b_n = n^{-\alpha}$ *for* $0<\alpha<\frac{1}{s}(1-\frac{4}{m})$. *Then*

$$\hat{\mathrm{T}}_2 = b_n^{-s}\left\{q\,\mathrm{R}(K)\,\mathrm{vol}(S_*) + O_p\left(\sqrt{\frac{\log n}{nb_n^s}} + b_n^2\right) + o_p(n^{-1/2+1/m+1/\eta})\right\} \qquad \textit{under } \mathrm{H}_0.$$

PROOF. Assume that n is large enough so that $\hat{\theta}\in\mathcal{N}_0$ and our regularity conditions hold. By (A.3), we can write $\hat{\mathrm{T}}_2 = \hat{\mathrm{T}}_2^{(1)} + R_2$, where

$$\hat{\mathrm{T}}_2^{(1)} = \sum_{i=1}^{n}\sum_{j=1,\, j\ne i}^{n}\mathbb{I}_i w_{ij}^2 g'(z_j,\theta_0)\hat{V}^{-1}(x_i,\hat{\theta})g(z_j,\theta_0)$$

and R_2 denotes the remaining terms. Using Lemmas C.3(ii) and C.7, we can show that $R_2 = O_p(n^{-1/2}b_n^{-s})$. Next, write $\hat{\mathrm{T}}_2^{(1)} = (1)_\mathrm{a} + (1)_\mathrm{b}$, where

$$(1)_\mathrm{a} = n^{-2}b_n^{-2s}\sum_{i=1}^{n}\sum_{j=1,\, j\neq i}^{n} \mathbb{I}_i K_{ij}^2 g'(z_j,\theta_0)H^{-1}(x_i,\theta_0)g(z_j,\theta_0),$$

$$(1)_\mathrm{b} = \frac{1}{n^2 b_n^{2s}}\sum_{i=1}^{n}\sum_{j=1,\, j\neq i}^{n} \mathbb{I}_i K_{ij}^2 g'(z_j,\theta_0)\{\hat{H}^{-1}(x_i,\hat{\theta}) - H^{-1}(x_i,\theta_0)\}g(z_j,\theta_0).$$

Let

$$O_p(\nu_n) \overset{\mathrm{def}}{=} O_p\left(\sqrt{\frac{\log n}{nb_n^s}} + b_n^2\right) + o_p(n^{-1/2+1/m+1/\eta}).$$

Then, by Lemmas C.3(i) and C.7,

$$(1)_\mathrm{b} \le c \max_{i\in I_*} \|\hat{H}^{-1}(x_i,\hat{\theta}) - H^{-1}(x_i,\theta_0)\| \frac{1}{n^2 b_n^{2s}}\sum_{i=1}^{n}\sum_{j=1,\, j\neq i}^{n} K_{ij} g_*^2(z_j)$$
$$= O_p(\nu_n)O_p(b_n^{-s}).$$

Now define $\tau_n = \sqrt{\log n/nb_n^s} + b_n^2$ and observe that

$$(1)_\mathrm{a} = \frac{1}{nb_n^s}\,\mathrm{tr}\sum_{i=1}^{n}\mathbb{I}_i\left\{\frac{1}{nb_n^s}\sum_{j=1,\, j\neq i}^{n} K_{ij}^2 g(z_j,\theta_0)g'(z_j,\theta_0)\right\}H^{-1}(x_i,\theta_0)$$
$$= \frac{1}{nb_n^s}\,\mathrm{tr}\sum_{i=1}^{n}\mathbb{I}_i\left\{\mathrm{R}(K)V(x_i,\theta_0)h(x_i) + R_a(x_i)\right\}H^{-1}(x_i,\theta_0),$$

where $\sup_{x_i\in S_*}\|R_a(x_i)\| = O_p(\tau_n)$ follows from the uniform consistency of kernel estimators. Since $\sup_{x_i\in S_*}\|H^{-1}(x_i,\theta_0)\| < \infty$ by Assumption 3.5(iv),

$$(1)_\mathrm{a} = \frac{q\,\mathrm{R}(K)}{nb_n^s}\sum_{i=1}^{n}\frac{\mathbb{I}_i}{h(x_i)} + \frac{O_p(\tau_n)}{b_n^s} = b_n^{-s}\{q\,\mathrm{R}(K)\,\mathrm{vol}(S_*) + O_p(\tau_n)\},$$

where the second equality follows because $n^{-1}\sum_{i=1}^{n}\mathbb{I}_i h^{-1}(x_i) = \mathrm{vol}(S_*) + O_p(n^{-1/2})$ by the central limit theorem. The desired result follows by combining the results for $(1)_\mathrm{a}$ and $(1)_\mathrm{b}$. □

LEMMA A.4. *Let Assumptions* 3.1–3.7 *hold. Assume that* $b_n = n^{-\alpha}$ *for* $0 < \alpha < \frac{1}{s}(1 - \frac{4}{m})$. *Then*

$$\hat{\mathrm{T}}_3 = \left\{O_p\left(\sqrt{\frac{\log n}{nb_n^s}}\right) + o_p(n^{-1/2+1/m+1/\eta})\right\}O_p\left(\sqrt{\frac{1}{nb_n^{3s}}}\right)$$
$$+ O_p\left(\sqrt{\frac{1}{nb_n^{2s}}}\right) \qquad \textit{under}\ \mathrm{H}_0.$$

PROOF. Assume that n is large enough so that $\hat{\theta} \in \mathcal{N}_0$ and our regularity conditions hold. Hence, $\hat{\mathrm{T}}_3 = \hat{\mathrm{T}}_3^{(1)} + R_3$ by (A.3), where

$$\hat{\mathrm{T}}_3^{(1)} = K(0) \sum_{i=1}^{n} \sum_{j=1,\, j\neq i}^{n} \mathbb{I}_i \frac{g'(z_i, \theta_0)\hat{V}^{-1}(x_i, \hat{\theta}) g(z_j, \theta_0) w_{ij}}{\sum_{u=1}^{n} K_{iu}}$$

and R_3 denotes the remaining terms. Now $\hat{\mathrm{T}}_3^{(1)} = (K(0)/n^{1/2} b_n^s) \sum_{l=1}^{q} \sum_{v=1}^{q} \hat{P}_{lv}$, where

$$\hat{P}_{lv} = \frac{1}{n^{3/2} b_n^s} \sum_{i=1}^{n} \sum_{j=1,\, j\neq i}^{n} \mathbb{I}_i g^{(l)}(z_i, \theta_0) \hat{F}^{(lv)}(x_i) g^{(v)}(z_j, \theta_0) K_{ij}$$

and $\hat{F}^{(lv)}(x_i)$ is the (lv)th element of $\hat{H}^{-1}(x_i, \hat{\theta})$. Let $G^{(lv)}(x_i)$ be the (lv)th element of $\tilde{H}_n^{-1}(x_i, \theta_0)$. Write $\hat{P}_{lv} = P_{lv}^{(1)} + \hat{P}_{lv}^{(2)}$, where

$$P_{lv}^{(1)} = \frac{1}{n^{3/2} b_n^s} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathbb{I}_i g^{(l)}(z_i, \theta_0) G^{(lv)}(x_i) g^{(v)}(z_j, \theta_0) K_{ij},$$

$$\hat{P}_{lv}^{(2)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{I}_i g^{(l)}(z_i, \theta_0) [\hat{F}^{(lv)}(x_i) - G^{(lv)}(x_i)] Q_{n,i}$$

and

$$Q_{n,i} = \frac{1}{n b_n^s} \sum_{j=1,\, j\neq i}^{n} g^{(v)}(z_j, \theta_0) K_{ij}.$$

Since $\mathbb{I}_i g^{(l)}(z_i, \theta_0) G^{(lv)}(x_i) g^{(v)}(z_j, \theta_0) K_{ij}$ and $\mathbb{I}_i g^{(l)}(z_i, \theta_0) G^{(lv)}(x_i) g^{(v)}(z_k, \theta_0) \times K_{ik}$ are uncorrelated for $i \neq j \neq k$, by the Cauchy–Schwarz inequality,

$$\mathbb{E}\{P_{lv}^{(1)}\}^2 \leq \frac{2}{n^3 b_n^{2s}} \sum_{i=1}^{n} \sum_{j=1,\, j\neq i}^{n} \mathbb{E}\{\mathbb{I}_i g^{(l)}(z_i, \theta_0) G^{(lv)}(x_i) g^{(v)}(z_j, \theta_0) K_{ij}\}^2.$$

Thus, $P_{lv}^{(1)} = O_p(\sqrt{1/n b_n^s})$ since $\sup_{x_i \in S_*} G^{(lv)}(x_i) < \infty$ for large enough n. Next, by the Cauchy–Schwarz inequality,

$$|\hat{P}_{lv}^{(2)}|^2 \leq \left\{\max_{i \in I_*} |\hat{F}^{(lv)}(x_i) - G^{(lv)}(x_i)|\right\}^2 \frac{1}{n} \sum_{i=1}^{n} [g^{(l)}(z_i, \theta_0)]^2 \sum_{i=1}^{n} Q_{n,i}^2.$$

But $\mathbb{E} Q_{n,i}^2 = O(1/n b_n^s)$ because $g^{(v)}(z_j, \theta_0) K_{ij}$ and $g^{(v)}(z_k, \theta_0) K_{ik}$ are uncorrelated for $i \neq j \neq k$. Hence, $\hat{P}_{lv}^{(2)} = O_p(d_n) O_p(b_n^{-s/2})$ by Lemma C.3(ii), where $O_p(d_n) \overset{\text{def}}{=} O_p(\sqrt{\log n / n b_n^s}) + o_p(n^{-1/2+1/m+1/\eta})$. Combining the results for $P_{lv}^{(1)}$ and $\hat{P}_{lv}^{(2)}$, we get $\hat{\mathrm{T}}_3^{(1)} = O_p(d_n) O_p(\sqrt{1/n b_n^{3s}})$. Finally, $R_3 = O_p(\sqrt{1/n b_n^{2s}})$ by Lemma C.5 and the Cauchy–Schwarz and Jensen inequalities. The desired result follows. $\square$

LEMMA A.5. *Let Assumptions* 3.1–3.7 *hold. Furthermore, assume that* $b_n = n^{-\alpha}$ *for* $0 < \alpha < \min\{\frac{1}{s}(1-\frac{4}{m}), \frac{1}{3s}, \frac{1}{s}(1-\frac{2}{m}-\frac{2}{\eta})\}$. *Then* $b_n^{s/2}\hat{\mathrm{T}}_5 \xrightarrow{d} \mathrm{N}(0, 2qK^{**}\operatorname{vol}(S_*))$ *under* H_0.

PROOF. Assume that n is large enough so that $\hat{\theta} \in \mathcal{N}_0$ and our regularity conditions hold. By Assumption 3.6,

$$g(z, \hat{\theta}) = g(z, \theta_0) + \frac{\partial g(z, \theta_0)}{\partial \theta}(\hat{\theta} - \theta_0) + \operatorname{Rem}(z, \hat{\theta} - \theta_0)$$

holds w.p.1, where $\|\operatorname{Rem}(z, \hat{\theta} - \theta_0)\| \le l(z)\|\hat{\theta} - \theta_0\|^2$. Hence, we can write $\hat{\mathrm{T}}_5 = \hat{\mathrm{T}}_5^{(1)} + 2\hat{\mathrm{T}}_5^{(2)} + R_5$, where

$$\hat{\mathrm{T}}_5^{(1)} = \sum_{i=1}^{n} \sum_{j=1,\, j\ne i}^{n} \sum_{t=1,\, t\ne j\ne i}^{n} \mathbb{I}_i w_{ij} g'(z_j, \theta_0) \hat{V}^{-1}(x_i, \hat{\theta}) g(z_t, \theta_0) w_{it},$$

$$\hat{\mathrm{T}}_5^{(2)} = \sum_{i=1}^{n} \sum_{j=1,\, j\ne i}^{n} \sum_{t=1,\, t\ne j\ne i}^{n} \mathbb{I}_i w_{ij} g'(z_j, \theta_0) \hat{V}^{-1}(x_i, \hat{\theta}) \frac{\partial g(z_t, \theta_0)}{\partial \theta}(\hat{\theta} - \theta_0) w_{it},$$

and R_5 denotes the remaining terms. Now $b_n^{s/2}\hat{\mathrm{T}}_5^{(1)} \xrightarrow{d} \mathrm{N}(0, 2qK^{**}\operatorname{vol}(S_*))$ by Lemma A.6 and $b_n^{s/2}\hat{\mathrm{T}}_5^{(2)} = o_p(1)$ by Lemma A.9. Next, since $\max_{i\in I_*}\|\hat{V}^{-1}(x_i, \hat{\theta})\| = O_p(1)$ by Lemma C.2(ii), the Cauchy–Schwarz and Jensen inequalities reveal that $R_5 = O_p(1)$. The desired result follows. □

LEMMA A.6. $b_n^{s/2}\hat{\mathrm{T}}_5^{(1)} \xrightarrow{d} \mathrm{N}(0, 2qK^{**}\operatorname{vol}(S_*))$ *under the conditions of Lemma* A.5.

PROOF. Write $\hat{\mathrm{T}}_5^{(1)} = \hat{\mathrm{T}}_5^* + (\hat{\mathrm{T}}_5^{(1)} - \hat{\mathrm{T}}_5^*)$, where $\hat{\mathrm{T}}_5^* = \tilde{\mathrm{T}}_5^*/(n^2 b_n^{2s})$ and

$$\tilde{\mathrm{T}}_5^* = \sum_{i=1}^{n} \sum_{j=1,\, j\ne i}^{n} \sum_{t=1,\, t\ne j\ne i}^{n} \mathbb{I}_i K_{ij} g'(z_j, \theta_0) \tilde{H}_n^{-1}(x_i, \theta_0) g(z_t, \theta_0) K_{it}. \tag{A.12}$$

Since $b_n^{s/2}\{\hat{\mathrm{T}}_5^{(1)} - \hat{\mathrm{T}}_5^*\} = o_p(1)$ by Lemma A.7, it suffices to show that $b_n^{s/2}\hat{\mathrm{T}}_5^* \xrightarrow{d} \mathrm{N}(0, 2dK^{**}\operatorname{vol}(S_*))$. To do so, we use a CLT for generalized quadratic forms due to de Jong (1987). First, change the order of summation in (A.12) to write $\tilde{\mathrm{T}}_5^* = \sum_{t=1}^{n}\sum_{j=1,\, j\ne t}^{n} g'(z_t, \theta_0) A_{tj} g(z_j, \theta_0)$. Next, define

$$W_{tj} = g'(z_t, \theta_0) A_{tj} g(z_j, \theta_0) + g'(z_j, \theta_0) A_{tj} g(z_t, \theta_0) = 2g'(z_t, \theta_0) A_{tj} g(z_j, \theta_0).$$

Using iterated expectations and the independence of observations, it is straightforward to verify that $\mathbb{E}(W_{tj}|x_t, z_t) = \mathbb{E}(W_{tj}|x_j, z_j) = 0$ for $1 \le t, j \le n$; that is, W_{tj} is "clean" in the terminology of de Jong [(1987), page 263]. Hence, in de Jong's notation, $\tilde{\mathrm{T}}_5^* = \sum_{t=1}^{n-1} \sum_{j=t+1}^{n} W_{tj}$. We now determine s_n^2, the variance of $\tilde{\mathrm{T}}_5^*$. Note that

$$s_n^2 = \operatorname{var} \tilde{\mathrm{T}}_5^* = \sum_{t=1}^{n-1} \sum_{j=t+1}^{n} \mathbb{E} W_{tj}^2 = 4 \sum_{t=1}^{n-1} \sum_{j=t+1}^{n} \mathbb{E}\{g'(z_t, \theta_0) A_{tj} g(z_j, \theta_0)\}^2,$$

where any cross terms vanish due to the uncorrelatedness of W_{tj} and W_{tk} for $t \neq j \neq k$. By Lemma A.8, $s_n^2 = 2n(n-1)(n-2)(n-3) q b_n^{3s} K^{**} \operatorname{vol}(S_*)\{1 + o(1)\}$. As in de Jong [(1987), page 266], let

$$G_I = \sum_{t=1}^{n-1} \sum_{j=t+1}^{n} \mathbb{E} W_{tj}^4,$$

$$G_{II} = \sum_{t=1}^{n-2} \sum_{j=t+1}^{n-1} \sum_{k=j+1}^{n} (\mathbb{E} W_{tj}^2 W_{tk}^2 + \mathbb{E} W_{jt}^2 W_{jk}^2 + \mathbb{E} W_{kt}^2 W_{kj}^2)$$

and

$$G_{IV} = \sum_{t=1}^{n-3} \sum_{j=t+1}^{n-2} \sum_{k=j+1}^{n-1} \sum_{l=k+1}^{n} (\mathbb{E} W_{tj} W_{tk} W_{lj} W_{lk} + \mathbb{E} W_{tj} W_{tl} W_{kj} W_{kl} + \mathbb{E} W_{tk} W_{tl} W_{jk} W_{jl}).$$

Since $G_I = 16 \sum_{t=1}^{n-1} \sum_{j=t+1}^{n} \mathbb{E}\{g'(z_t, \theta_0) A_{tj} g(z_j, \theta_0)\}^4$,

$$\begin{aligned}
&\mathbb{E}\{g'(z_t, \theta_0) A_{tj} g(z_j, \theta_0)\}^4 \\
&\quad = \sum_{i=1,\, i \neq j \neq t}^{n} \mathbb{E}\{\mathbb{I}_i K_{ij} g'(z_t, \theta_0) \tilde{H}_n^{-1}(x_i, \theta_0) g(z_j, \theta_0) K_{it}\}^4 \\
&\qquad + 3 \sum_{i=1,\, i \neq j \neq t}^{n} \sum_{k=1,\, k \neq i \neq j \neq t}^{n} \mathbb{E}\{\mathbb{I}_i K_{ij} g'(z_t, \theta_0) \tilde{H}_n^{-1}(x_i, \theta_0) g(z_j, \theta_0) K_{it}\}^2 \\
&\qquad \qquad \times \{\mathbb{I}_k K_{kj} g'(z_t, \theta_0) \tilde{H}_n^{-1}(x_k, \theta_0) g(z_j, \theta_0) K_{kt}\}^2.
\end{aligned}$$

But $\mathbb{E}\{\mathbb{I}_i K_{ij} g'(z_t, \theta_0) \tilde{H}_n^{-1}(x_i, \theta_0) g(z_j, \theta_0) K_{it}\}^4 < \infty$ by $\sup_{x_i \in S_*} \|\tilde{H}_n^{-1}(x_i, \theta_0)\| < \infty$ and the fact that z_t is independent of z_j for $t \neq j$. Hence, $G_I = O(n^4)$ by the Cauchy–Schwarz inequality. Similarly, $G_{II} = O(n^5)$ and $G_{IV} = O(n^6)$. If $b_n = n^{-\alpha}$ for $0 < \alpha < \frac{1}{3s}$, then G_I, G_{II} and G_{IV} are $o(s_n^4)$. Hence, by de Jong [(1987), Proposition 3.2], $s_n^{-1} \tilde{\mathrm{T}}_5^* \xrightarrow{d} \mathrm{N}(0, 1)$. Therefore, $b_n^{s/2} \hat{\mathrm{T}}_5^* \xrightarrow{d} \mathrm{N}(0, 2q K^{**} \operatorname{vol}(S_*))$. □

LEMMA A.7. *$b_n^{s/2}\{\hat{\mathrm{T}}_5^{(1)} - \hat{\mathrm{T}}_5^*\} = o_p(1)$ under the conditions of Lemma* A.5.

PROOF. Observe that

$$|\hat{\mathrm{T}}_5^{(1)} - \hat{\mathrm{T}}_5^*| = \frac{1}{n^2 b_n^{2s}} \left| \operatorname{tr} \sum_{i=1}^n \mathbb{I}_i \{\hat{H}^{-1}(x_i, \hat{\theta}) - \tilde{H}_n^{-1}(x_i, \theta_0)\} \right.$$
$$\left. \times \left\{ \sum_{j=1,\, j\neq i}^n K_{ij} g(z_j, \theta_0) \right\} \left\{ \sum_{t=1,\, t\neq j\neq i}^n K_{it} g'(z_t, \theta_0) \right\} \right|$$
$$\leq n \max_{i \in I_*} \|\hat{H}^{-1}(x_i, \hat{\theta}) - \tilde{H}_n^{-1}(x_i, \theta_0)\|$$
$$\times \left\{ \max_{i \in I_*} \left\| \frac{1}{n b_n^s} \sum_{j=1,\, j\neq i}^n K_{ij} g(z_j, \theta_0) \right\| \right\}^2.$$

Hence, by (C.1) and Lemma C.3, $b_n^{s/2} |\hat{\mathrm{T}}_5^{(1)} - \hat{\mathrm{T}}_5^*| = o_p(1)$ under our choice of b_n. $\square$

LEMMA A.8. $\mathbb{E}\{g'(z_t, \theta_0) A_{tj} g(z_j, \theta_0)\}^2 = (n-2)(n-3) q b_n^{3s} K^{**} \operatorname{vol}(S_*) \times \{1 + o(1)\}$ *under the conditions of Lemma* A.5.

PROOF. By iterated expectations and the independence of observations, it is straightforward to show that $\mathbb{E}\{g'(z_t, \theta_0) A_{tj} g(z_j, \theta_0)\}^2 = \operatorname{tr} \mathbb{E}\{A_{tj} V(x_j, \theta_0) A_{tj} \times V(x_t, \theta_0)\}$. Hence, we can write

$$\mathbb{E}\{g'(z_t, \theta_0) A_{tj} g(z_j, \theta_0)\}^2 = \sum_{i=1,\, i\neq j\neq t}^n \operatorname{tr} P_1 + \sum_{i=1,\, i\neq j\neq t}^n \sum_{u=1,\, u\neq i\neq j\neq t}^n \operatorname{tr} P_2,$$

where

$$P_1 = \mathbb{E}\left\{ \frac{\mathbb{I}_i K_{ij}^2 K_{it}^2 [\mathbb{E}\{\hat{\Omega}(x_i, \theta_0) | x_i\}]^{-1}}{\mathbb{E}^2\{\hat{h}(x_i) | x_i\}} \right.$$
$$\left. \times V(x_j, \theta_0) [\mathbb{E}\{\hat{\Omega}(x_i, \theta_0) | x_i\}]^{-1} V(x_t, \theta_0) \right\}$$

and

$$P_2 = \mathbb{E}\left\{ \frac{\mathbb{I}_i \mathbb{I}_u K_{ij} K_{it} K_{uj} K_{ut} [\mathbb{E}\{\hat{\Omega}(x_i, \theta_0) | x_i\}]^{-1}}{\mathbb{E}\{\hat{h}(x_u) | x_u\} \mathbb{E}\{\hat{h}(x_i) | x_i\}} \right.$$
$$\left. \times V(x_j, \theta_0) [\mathbb{E}\{\hat{\Omega}(x_u, \theta_0) | x_u\}]^{-1} V(x_t, \theta_0) \right\}.$$

It is straightforward, albeit tedious, to show that

$$\operatorname{tr} P_1 = q b_n^{2s} \mathrm{R}^2(K) \mathbb{E}\left\{ \frac{\mathbb{I}_1}{h^2(x_1)} \right\} [1 + O(b_n^2)]$$

and

$$\operatorname{tr} P_2 = q b_n^{3s} K^{**} \operatorname{vol}(S_*)\{1 + O(b_n)\}.$$

The desired result follows. □

LEMMA A.9. *$\hat{\mathrm{T}}_5^{(2)} = O_p(1)$ under the conditions of Lemma* A.5.

PROOF. Let

$$\tilde{\mathrm{T}}_5^{(2)} = \sum_{i=1}^n \sum_{j=1, j\neq i}^n \sum_{t=1, t\neq j\neq i}^n \mathbb{I}_i w_{ij} g'(z_j, \theta_0) \hat{V}^{-1}(x_i, \hat{\theta}) \frac{\partial g(z_t, \theta_0)}{\partial \theta} w_{it}.$$

Then $\hat{\mathrm{T}}_5^{(2)} = \tilde{\mathrm{T}}_5^{(2)}(\hat{\theta} - \theta_0)$. Since $\|\hat{\theta} - \theta_0\| = O_p(n^{-1/2})$, we show that $\|\tilde{\mathrm{T}}_5^{(2)}\| = O_p(n^{1/2})$. So let $\zeta \in \mathbb{S}^p$ be arbitrary and look at $\tilde{\mathrm{T}}_5^{(2)}\zeta$. Write $\tilde{\mathrm{T}}_5^{(2)}\zeta = \text{(a)} + \text{(b)}$, where

$$\text{(a)} = \frac{1}{n^2 b_n^{2s}} \sum_{i=1}^n \sum_{j=1,\, j\neq i}^n \sum_{t=1, t\neq j\neq i}^n \mathbb{I}_i K_{ij} g'(z_j, \theta_0) \tilde{H}_n^{-1}(x_i, \theta_0) \frac{\partial g(z_t, \theta_0)}{\partial \theta} \zeta K_{it}$$

and

$$\text{(b)} = \frac{1}{n^2 b_n^{2s}} \sum_{i=1}^n \sum_{j=1,\, j\neq i}^n \sum_{t=1, t\neq j\neq i}^n \mathbb{I}_i K_{ij} g'(z_j, \theta_0) \{\hat{H}^{-1}(x_i, \hat{\theta}) - \tilde{H}_n^{-1}(x_i, \theta_0)\} \times \frac{\partial g(z_t, \theta_0)}{\partial \theta} \zeta K_{it}.$$

Since $\mathbb{E}\{g(z_j, \theta_0)|x_j\} = 0$, some tedious but straightforward algebra shows that $\mathbb{E}\{(a)\}^2 = O(n)$, that is, (a) $= O_p(n^{1/2})$. Next, as in the proof of Lemma A.7, we can show that (b) $= o_p(n^{1/2})$. Therefore, $\|\tilde{\mathrm{T}}_{n,5}^{(2)}\| = O_p(n^{1/2})$. The desired result follows. □

APPENDIX B

Asymptotic theory under local alternatives.

LEMMA B.1. *Let Assumptions* 3.1, 3.2, 3.6, 3.7, 6.1 *and* 6.2 *hold. Assume that $b_n = n^{-\alpha}$ for $0 < \alpha < \frac{1}{s}(1 - \frac{4}{m})$. Then*

$$\mathrm{SELR} = \hat{\mathrm{T}} + o_p\left(\left\{\frac{\log n}{n^{1/2-1/m} b_n^s}\right\}^2\right) + o_p\left(\frac{1}{n^{1-2/m}}\right) + O_p\left(\left\{\frac{\log n}{n^{1/3} b_n^s}\right\}^{3/2}\right) \qquad \textit{under } \mathrm{H}_{1n},$$

where $\hat{\mathrm{T}} = \sum_{i=1}^n \mathbb{I}_i \{\sum_{j=1}^n w_{ij} g'(z_j, \hat{\theta})\} \hat{V}^{-1}(x_i, \hat{\theta}) \{\sum_{j=1}^n w_{ij} g(z_j, \hat{\theta})\}$.

PROOF. Since Lemmas B.1–B.3 of Newey (1994) remain valid when θ_0 is replaced by $\theta_{n,0}$ [because $g(z_1, \theta_{n,0}), \dots, g(z_n, \theta_{n,0})$ are i.i.d. for each n], the proofs of Lemmas C.2 and C.3 go through without any change. Hence, we can follow the proof of Lemma A.1 leading up to (A.6) and (A.7) to show that $\mathbb{I}_i \lambda_i = n\mathbb{I}_i \hat{V}^{-1}(x_i, \hat{\theta}) \sum_{j=1}^n w_{ij} g(z_j, \hat{\theta}) + \mathbb{I}_i r_{2,i}$ w.p.a.1, where

$$\mathbb{I}_i \|r_{2,i}\| = o_p(n^{1+1/m}) \left\{ \left[\max_{i \in I_*} \left\| \sum_{j=1}^n w_{ij} g(z_j, \theta_{n,0}) \right\| \right]^2 + \|\hat{\theta} - \theta_{n,0}\|^2 \sum_{j=1}^n d^2(z_j) w_{ij} \right\}$$

and the $o_p(n^{1+1/m})$ term does not depend on $i \in I_*$. Using the continuity of $\delta(x)$ and $h(x)$ and the compactness of S_*, it is straightforward to show that $\max_{i \in I_*} \| \sum_{j=1}^n w_{ij} g(z_j, \theta_{n,0}) \| = O_p(\sqrt{\log n / n b_n^s})$. Now proceed as in Lemma A.1 to obtain the desired result. $\square$

LEMMA B.2. *Let Assumptions* 3.1, 3.2, 3.6, 3.7, 6.1 *and* 6.2 *hold. Then* $\hat{\mathrm{T}}_1 \overset{\mathrm{H}_{1n}}{=} O_p(1/n b_n^{2s})$.

PROOF. Same as the proof of Lemma A.2. $\square$

LEMMA B.3. *Let Assumptions* 3.1, 3.2, 3.6, 3.7, 6.1 *and* 6.2 *hold. Assume that* $b_n = n^{-\alpha}$ *for* $0 < \alpha < \frac{1}{s}(1 - \frac{4}{m})$. *Then*

$$\hat{\mathrm{T}}_2 \overset{\mathrm{H}_{1n}}{=} b_n^{-s} \left\{ q\, \mathrm{R}(K) \operatorname{vol}(S_*) + O_p\left(\sqrt{\frac{\log n}{n b_n^s}} + b_n^2 \right) + o_p(n^{-1/2 + 1/m + 1/\eta}) \right\}.$$

PROOF. Assume that n is large enough so that $\hat{\theta}$ and $\theta_{n,0}$ lie in $\mathcal{N}_0$ and our regularity conditions hold. By Assumption 3.6,

$$g(z_j, \hat{\theta}) = g(z_j, \theta_{n,0}) + \operatorname{rem}(z_j, \hat{\theta} - \theta_{n,0}) \text{ w.p.1}, \quad \text{where } \|\operatorname{rem}(z_j, \hat{\theta} - \theta_{n,0})\| \le d(z_j) \|\hat{\theta} - \theta_{n,0}\|. \tag{B.1}$$

As in Lemma C.3, we can show that if $b_n = n^{-\alpha}$ for $0 < \alpha < \frac{1}{s}(1 - \frac{4}{m})$, then

$$\sup_{x_i \in S_*} \|\hat{H}^{-1}(x_i, \hat{\theta}) - H^{-1}(x_i, \theta_{n,0})\| = O_p\left(\sqrt{\frac{\log n}{n b_n^s}} + b_n^2 \right) + o_p(n^{-1/2+1/m+1/\eta}). \tag{B.2}$$

Therefore, using (B.1) and the way we dealt with R_2 in the proof of Lemma A.3,

$$\hat{\mathrm{T}}_2 = \sum_{i=1}^n \sum_{j=1,\, j \ne i}^n \mathbb{I}_i w_{ij}^2 g'(z_j, \theta_{n,0}) \hat{V}^{-1}(x_i, \hat{\theta}) g(z_j, \theta_{n,0}) + O_p(n^{-1/2} b_n^{-s}). \tag{B.3}$$

As in Lemma C.2, we can show that if $b_n = n^{-\alpha}$ for $0 < \alpha < \frac{1}{s}(1 - \frac{4}{m})$, then

$$\sup_{x_i \in S_*} \|\hat{V}^{-1}(x_i, \hat{\theta}) - V^{-1}(x_i, \theta_{n,0})\| = O_p\left(\sqrt{\frac{\log n}{nb_n^s}} + b_n^2\right) + o_p(n^{-1/2+1/m+1/\eta}).$$

The desired result follows by (B.3) and the way we handled $\hat{\mathrm{T}}_2^{(1)}$ in the proof of Lemma A.3. □

LEMMA B.4. *Let Assumptions* 3.1, 3.2, 3.6, 3.7, 6.1 *and* 6.2 *hold. Assume that* $b_n = n^{-\alpha}$ *for* $0 < \alpha < \frac{1}{s}(1 - \frac{4}{m})$. *Then*

$$\hat{\mathrm{T}}_3 \overset{\mathrm{H}_{1n}}{=} \left\{O_p\left(\sqrt{\frac{\log n}{nb_n^s}}\right) + o_p(n^{-1/2+1/m+1/\eta})\right\} O_p\left(\sqrt{\frac{1}{nb_n^{3s}}}\right) + O_p\left(\sqrt{\frac{1}{nb_n^{5s/2}}}\right).$$

PROOF. Using (B.1), the proof is very similar to that of Lemma A.4. □

LEMMA B.5. *Let Assumptions* 3.1, 3.2, 3.6, 3.7, 6.1 *and* 6.2 *hold. Furthermore, assume that* $b_n = n^{-\alpha}$ *for* $0 < \alpha < \min\{\frac{1}{s}(1 - \frac{4}{m}), \frac{1}{3s}, \frac{1}{s}(1 - \frac{2}{m} - \frac{2}{\eta})\}$. *Then* $b_n^{s/2}\hat{\mathrm{T}}_5 \xrightarrow{d} \mathrm{N}(\mu, 2qK^{**}\,\mathrm{vol}(S_*))$ *under* H_{1n}, *where* $\mu = \mathbb{E}[\mathbb{I}\{x_1 \in S_*\}\delta'(x_1) \times V^{-1}(x_1, \theta_0)\delta(x_1)]$.

PROOF. Assume that n is large enough so that $\hat{\theta}$ and $\theta_{n,0}$ lie in $\mathcal{N}_0$ and our regularity conditions hold. By Assumption 3.6,

$$g(z, \hat{\theta}) = g(z, \theta_{n,0}) + \frac{\partial g(z, \theta_{n,0})}{\partial \theta}(\hat{\theta} - \theta_{n,0}) + \mathrm{Rem}(z, \hat{\theta} - \theta_{n,0})$$

holds w.p.1, where $\|\mathrm{Rem}(z, \hat{\theta} - \theta_{n,0})\| \le l(z)\|\hat{\theta} - \theta_{n,0}\|^2$. Hence, as we handled R_5 in the proof of Lemma A.5, we can show that $b_n^{s/2}\hat{\mathrm{T}}_5 = b_n^{s/2}(\mathrm{C}) + 2b_n^{s/2}(\mathrm{D}) + o_p(1)$, where

$$(\mathrm{C}) = \sum_{i=1}^{n} \sum_{j=1,\, j\neq i}^{n} \sum_{t=1,\, t\neq j\neq i}^{n} \mathbb{I}_i w_{ij} g'(z_j, \theta_{n,0}) \hat{V}^{-1}(x_i, \hat{\theta}) g(z_t, \theta_{n,0}) w_{it},$$

$$(\mathrm{D}) = \sum_{i=1}^{n} \sum_{j=1,\, j\neq i}^{n} \sum_{t=1,\, t\neq j\neq i}^{n} \mathbb{I}_i w_{ij} g'(z_j, \theta_{n,0}) \hat{V}^{-1}(x_i, \hat{\theta}) \frac{\partial g(z_t, \theta_{n,0})}{\partial \theta}(\hat{\theta} - \theta_{n,0}) w_{it}.$$

Let $\varepsilon_n = n^{-1/2} b_n^{-s/4}$, $f_n(z, \theta) = g(z, \theta) - \varepsilon_n \delta(x)$. Since $\mathbb{E}\{f_n(z, \theta_{n,0})|x\} = 0$ and $\|\hat{\theta} - \theta_{n,0}\| = O_p(n^{-1/2})$ under H_{1n}, we can use Lemma C.5 to show that $(\mathrm{D}) = O_p(b_n^{-s/4})$. Therefore, $b_n^{s/2}\hat{\mathrm{T}}_5 = b_n^{s/2}(\mathrm{C}) + o_p(1)$. Now write $(\mathrm{C}) = (\mathrm{C}_1)+$

$(\mathrm{C}_2) + R_{\mathrm{C}}$, where

$$(\mathrm{C}_1) = \frac{1}{n^2 b_n^{2s}} \sum_{i=1}^{n} \sum_{j=1,\, j\neq i}^{n} \sum_{t=1,\, t\neq j\neq i}^{n} \mathbb{I}_i K_{ij} f_n'(z_j, \theta_{n,0}) \hat{H}^{-1}(x_i, \hat{\theta}) f_n(z_t, \theta_{n,0}) K_{it},$$

$$(\mathrm{C}_2) = \frac{\varepsilon_n^2}{n^2 b_n^{2s}} \sum_{i=1}^{n} \sum_{j=1,\, j\neq i}^{n} \sum_{t=1,\, t\neq j\neq i}^{n} \mathbb{I}_i K_{ij} \delta'(x_j) \hat{H}^{-1}(x_i, \hat{\theta}) \delta(x_t) K_{it},$$

and R_{C} denotes the remaining terms. As in Lemma A.5, $b_n^{s/2}(\mathrm{C}_1) \xrightarrow{d} \mathrm{N}(0, 2q \times K^{**} \operatorname{vol}(S_*))$. Next, the continuity of $\delta(x)$ and $h(x)$ on S implies that $\sup_{x_i \in S_*} \|(1/nb_n^s) \sum_{t=1}^n \delta(x_t) K_{it} - \delta(x_i) h(x_i)\| = o_p(1)$. Hence, $b_n^{s/2}(\mathrm{C}_2) = n^{-1} \sum_{i=1}^n \mathbb{I}_i \times \delta'(x_i) \hat{H}^{-1}(x_i, \hat{\theta}) \delta(x_i) h^2(x_i) + o_p(1)$. By (B.2) and dominated convergence, $n^{-1} \sum_{i=1}^n \mathbb{I}_i \delta'(x_i) H^{-1}(x_i, \theta_{n,0}) \delta(x_i) h^2(x_i) = \mu + o_p(1)$. Therefore, $b_n^{s/2}(\mathrm{C}_2) = \mu + o_p(1)$. Finally, as we handled $\tilde{\mathrm{T}}_{n,5}^{(2)}$ in the proof of Lemma A.9, $R_{\mathrm{C}} = O_p(b_n^{-s/4})$. Combining these results, we obtain $b_n^{s/2}(\mathrm{C}) \xrightarrow{d} \mathrm{N}(\mu, 2qK^{**} \operatorname{vol}(S_*))$. The desired result follows. □

APPENDIX C

Some useful results.

LEMMA C.1. *Let Assumptions* 3.1–3.3, 3.5 *and* 3.7 *hold. If* $\log n / n^{1-2/m} b_n^s \downarrow 0$, *then*

$$\sup_{x_i \in S_*} \left\| \sum_{j=1}^{n} w_{ij} g(z_j, \theta_0) \right\| = O_p\left(\sqrt{\frac{\log n}{nb_n^s}} \right) \qquad \textit{under } \mathrm{H}_0.$$

PROOF. By Newey [(1994), Lemma B.1],

$$\sup_{x_i \in S_*} |\hat{h}(x_i) - \mathbb{E}\hat{h}(x_i)| = O_p\left(\sqrt{\frac{\log n}{nb_n^s}} \right) \quad \text{and}$$
$$\sup_{x_i \in S_*} \left\| \frac{1}{nb_n^s} \sum_{j=1}^{n} g(z_j, \theta_0) K_{ij} \right\| \overset{\mathrm{H}_0}{=} O_p\left(\sqrt{\frac{\log n}{nb_n^s}} \right). \tag{C.1}$$

The desired result follows since $\mathbb{E}\hat{h}(x_i)$ is bounded away from 0 on S_* for large enough n. □

LEMMA C.2. *Let Assumptions* 3.1–3.7 *hold. If* $\log n / n^{1-4/m} b_n^s \downarrow 0$, *then*

(i)
$$\sup_{x_i \in S_*} \|\hat{V}(x_i, \hat{\theta}) - V(x_i, \theta_0)\| = O_p\left(\sqrt{\frac{\log n}{nb_n^s}} + b_n^2\right) + o_p(n^{-1/2+1/m+1/\eta}),$$

(ii)
$$\sup_{x_i \in S_*} \|\hat{V}^{-1}(x_i, \hat{\theta}) - V^{-1}(x_i, \theta_0)\| = O_p\left(\sqrt{\frac{\log n}{nb_n^s}} + b_n^2\right) + o_p(n^{-1/2+1/m+1/\eta}).$$

PROOF. Assume n is large enough so that $\hat{\theta} \in \mathcal{N}_0$ and our regularity conditions hold. By Assumption 3.6, $g(z, \hat{\theta}) = g(z, \theta_0) + \mathrm{rem}(z, \hat{\theta} - \theta_0)$ w.p.1, where $\|\mathrm{rem}(z, \hat{\theta} - \theta_0)\| \le d(z)\|\hat{\theta} - \theta_0\|$. Hence, $\|\hat{V}(x_i, \hat{\theta}) - \hat{V}(x_i, \theta_0)\| \le 2A(x_i) + B(x_i)$, where $A(x_i) = \sum_{j=1}^n \|g(z_j, \theta_0)\| \, \|\mathrm{rem}(z_j, \hat{\theta} - \theta_0)\| w_{ij}$ and $B(x_i) = \sum_{j=1}^n \|\mathrm{rem}(z_j, \hat{\theta} - \theta_0)\|^2 w_{ij}$. By Lemmas C.4 and C.6,

$$\sup_{x_i \in S_*} A(x_i) \le \|\hat{\theta} - \theta_0\| \max_{1 \le j \le n} \|g(z_j, \theta_0)\| \sup_{x_i \in S_*} \sum_{j=1}^n d(z_j) w_{ij} = o_p(n^{-1/2+1/m+1/\eta}).$$

Similarly, $\sup_{x_i \in S_*} B(x_i) = o_p(n^{-1+2/\eta})$. Hence, as $\eta \ge 2$,

$$\sup_{x_i \in S_*} \|\hat{V}(x_i, \hat{\theta}) - \hat{V}(x_i, \theta_0)\| = o_p(n^{-1/2+1/m+1/\eta}). \tag{C.2}$$

Let $\tau_n \stackrel{\text{def}}{=} \sqrt{\log n / nb_n^s} + b_n^2$. By Newey [(1994), Lemma B.3], $\sup_{x_i \in S_*} |\hat{h}(x_i) - h(x_i)| = O_p(\tau_n)$ and $\sup_{x_i \in S_*} \|1/nb_n^s \sum_{j=1}^n K_{ij} g(z_j, \theta_0) g'(z_j, \theta_0) - V(x_i, \theta_0) \times h(x_i)\| = O_p(\tau_n)$. Hence, as h is bounded away from 0 on S_*, $\sup_{x_i \in S_*} \|\hat{V}(x_i, \theta_0) - V(x_i, \theta_0)\| = O_p(\tau_n)$. Therefore, (i) follows by (C.2); (ii) follows from (i) since $\inf_{(\xi, x_i) \in \mathbb{S}^q \times S_*} \xi' V(x_i, \theta_0) \xi > 0$. □

LEMMA C.3. *Let Assumptions* 3.1–3.7 *hold. If* $\log n / n^{1-4/m} b_n^s \downarrow 0$, *then*

(i)
$$\sup_{x_i \in S_*} \|\hat{H}^{-1}(x_i, \hat{\theta}) - H^{-1}(x_i, \theta_0)\| = O_p\left(\sqrt{\frac{\log n}{nb_n^s}} + b_n^2\right) + o_p(n^{-1/2+1/m+1/\eta}),$$

(ii)
$$\sup_{x_i \in S_*} \|\hat{H}^{-1}(x_i, \hat{\theta}) - \tilde{H}_n^{-1}(x_i, \theta_0)\| = O_p\left(\sqrt{\frac{\log n}{nb_n^s}}\right) + o_p(n^{-1/2+1/m+1/\eta}).$$

PROOF. Similar to the proof of Lemma C.2. □

LEMMA C.4. *Let* $z_1, \dots, z_n$ *be identically distributed. If* $\mathbb{E}\{\sup_{\theta\in\Theta} \|g(z, \theta)\|^m\} < \infty$, *then we have* $\Pr\{\max_{1\le j\le n} \sup_{\theta\in\Theta} \|g(z_j, \theta)\| = o(n^{1/m})\} = 1$ *as* $n \uparrow \infty$.

PROOF. Our proof is based on the idea in Owen [(1990), Lemma 3]. Let $\varepsilon > 0$. Since $\sum_{n=1}^{\infty} \Pr\{[\sup_{\theta\in\Theta} \|g(z_n, \theta)\|]^m/\varepsilon^m \ge n\} < \infty$, by the Borel–Cantelli lemma $\{[\sup_{\theta\in\Theta} \|g(z_n, \theta)\|]^m/\varepsilon^m \ge n\}$ happens infinitely often w.p.0. Equivalently, the event $\{\sup_{\theta\in\Theta} \|g(z_n, \theta)\|/\varepsilon < n^{1/m}\}$ happens for all but finitely many n w.p.1. Since $n^{1/m}$ eventually exceeds the largest element in the finite collection of $\sup_{\theta\in\Theta} \|g(z_k, \theta)\|/\varepsilon$'s that exceed $k^{1/m}$, $\Pr\{\max_{1\le j\le n} \sup_{\theta\in\Theta} \|g(z_j, \theta)\| < n^{1/m}\varepsilon\} = 1$ for large enough n. The desired result follows since ε can be chosen arbitrarily small. □

LEMMA C.5. *Let* $\{x_i, z_i\}_{i=1}^n$ *be a random sample, let* $f(z)$ *be a real-valued function such that* $\mathbb{E}|f(z_1)| < \infty$ *and let Assumption* 3.7 *hold. Then* $\mathbb{E}\{\sum_{j=1}^n |f(z_j)|w_{ij}\} \le c\mathbb{E}|f(z_1)|$, *where the constant* c *depends only upon the kernel.*

PROOF. See Devroye and Wagner [(1980), Lemma 2]. □

LEMMA C.6. *Let* $f(z)$ *be a real-valued function such that* $\mathbb{E}|f(z)|^a < \infty$ *and let Assumption* 3.7 *hold. Then* $\Pr\{\sup_{x_i\in\mathbb{R}^s} |\sum_{j=1}^n f(z_j)w_{ij}| = o(n^{1/a})\} = 1$ *as* $n \uparrow \infty$.

PROOF. Observe that $|\sum_{j=1}^n f(z_j)w_{ij}| \le \max_{1\le j\le n} |f(z_j)|$. Now use Lemma C.4. □

LEMMA C.7. *Let* $\{x_i, z_i\}_{i=1}^n$ *be a random sample such that the p.d.f. of* x_1 *is bounded, let* $f(z)$ *be a real-valued function such that* $\mathbb{E}f^2(z_1) < \infty$ *and let Assumption* 3.7 *hold. Then*

$$\mathbb{E}\left\{\frac{1}{nb_n^s}\sum_{j=1}^n |f(z_j)|K_{ij}\right\} \le c\left\{\mathbb{E}f^2(z_1) + \frac{1}{n^2 b_n^{2s}} + \frac{1}{nb_n^s} + 1\right\},$$

where c *depends only upon* K *and* h.

PROOF. Since $(1/nb_n^s)\sum_{j=1}^n |f(z_j)|K_{ij} = \sum_{j=1}^n |f(z_j)|w_{ij}\hat{h}(x_i)$, by Jensen's inequality

$$\frac{1}{nb_n^s}\sum_{j=1}^n |f(z_j)|K_{ij} \le \frac{1}{2}\left\{\left[\sum_{j=1}^n |f(z_j)|w_{ij}\right]^2 + \hat{h}^2(x_i)\right\}$$
$$\le \frac{1}{2}\left\{\sum_{j=1}^n f^2(z_j)w_{ij} + \hat{h}^2(x_i)\right\}.$$

It is easy to show that

$$\mathbb{E}\hat{h}^2(x_i) \le c\left\{\frac{1}{n^2 b_n^{2s}} + \frac{1}{nb_n^s} + 1\right\}.$$

The desired result follows by Lemma C.5. □

LEMMA C.8. *Let $f(z)$ be a real-valued function such that $\mathbb{E}|f(z)|^a < \infty$ and let Assumptions* 3.5 *and* 3.7 *hold. If* $\log n/nb_n^s \downarrow 0$, *then*

$$\sup_{x_i\in S_*}\left|\frac{1}{nb_n^s}\sum_{j=1}^n f(z_j)K_{ij}\right| = o_p(n^{1/a}).$$

PROOF. Since

$$\sup_{x_i\in S_*}\frac{1}{nb_n^s}\sum_{j=1}^n |f(z_j)|K_{ij} \le \max_{1\le j\le n}|f(z_j)| \sup_{x_i\in S_*}\hat{h}(x_i),$$

the desired result follows by Lemma C.4 and the fact that $\sup_{x_i\in S_*}\hat{h}(x_i) = O_p(1)$. □

Acknowledgments. We thank an Associate Editor and two anonymous referees for comments that greatly improved this paper. We also thank Don Andrews, Bruce Hansen, Wolfgang Härdle, Yongmiao Hong, Joel Horowitz, Oliver Linton, George Tauchen and Ken West for helpful suggestions. Shane Sherlund provided excellent research assistance. The first author would also like to thank the hospitality of Professor Wolfgang Härdle of the Institute of Statistics and Econometrics, Humboldt-University Berlin, where part of this research was carried out.

REFERENCES

AÏT-SAHALIA, Y., BICKEL, P. and STOKER, T. (2001). Goodness-of-fit tests for kernel regression with an application to option implied volatilities. *J. Econometrics* **105** 363–412.

ANDREWS, D. (1997). A conditional Kolmogorov test. *Econometrica* **65** 1097–1128.

ANDREWS, D. W. and PLOBERGER, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* **62** 1383–1414.

ANDREWS, D. W. and PLOBERGER, W. (1996). Testing for serial correlation against an ARMA(1, 1) process. *J. Amer. Statist. Assoc.* **91** 1331–1342.

BIERENS, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica* **58** 1443–1458.

BIERENS, H. J. and PLOBERGER, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica* **65** 1129–1151.

BRILLINGER, D. (1977). Discussion of "Consistent nonparametric regression" by C. J. Stone. *Ann. Statist.* **5** 622–623.

BROWN, B. W. and NEWEY, W. K. (1998). Efficient bootstrapping for semiparametric models. Unpublished manuscript.

CHAMBERLAIN, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *J. Econometrics* **34** 305–334.

CHAMBERLAIN, G. (1992). Efficiency bounds for semiparametric regression. *Econometrica* **60** 567–596.

CHEN, S., HÄRDLE, W. and KLEINOW, T. (2001). An empirical likelihood goodness-of-fit test for time series. Discussion Paper 1, Sonderforschungsbereich 373, Humboldt-Univ. Berlin.

COSSLETT, S. (1981a). Efficient estimation of discrete choice models. In *Structural Analysis of Discrete Data with Econometric Applications* (C. F. Manski and D. McFadden, eds.) 51–111. MIT Press.

COSSLETT, S. (1981b). Maximum likelihood estimator for choice-based samples. *Econometrica* **49** 1289–1316.

DE JONG, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Related Fields* **75** 261–277.

DE JONG, R. M. and BIERENS, H. J. (1994). On the limit behavior of a chi-square type test if the number of conditional moments tested approaches infinity. *Econometric Theory* **10** 70–90.

DEVROYE, L. P. and WAGNER, T. J. (1980). Distribution-free consistency results in nonparametric discrimination and regression function estimation. *Ann. Statist.* **8** 231–239.

ELLISON, G. and ELLISON, S. F. (2000). A simple framework for nonparametric specification testing. *J. Econometrics* **96** 1–23.

EUBANK, R. and SPIEGELMAN, C. (1990). Testing the goodness of fit of a linear model via nonparametric regression techniques. *J. Amer. Statist. Assoc.* **85** 387–392.

FAN, Y. and LI, Q. (1996). Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica* **64** 865–890.

FAN, J., ZHANG, C. and ZHANG, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29** 153–193.

FISHER, N. I., HALL, P., JING, B.-Y. and WOOD, A. T. A. (1996). Improved pivotal methods for constructing confidence regions with directional data. *J. Amer. Statist. Assoc.* **91** 1062–1070.

HANNAN, E. (1970). *Time Series Analysis*. Wiley, New York.

HÄRDLE, W. (1989). Asymptotic maximal deviation of M-smoothers. *J. Multivariate Anal.* **29** 163–179.

HÄRDLE, W. (1990). *Applied Nonparametric Regression*. Cambridge Univ. Press.

HÄRDLE, W. and MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21** 1926–1947.

HÄRDLE, W. and MARRON, J. (1990). Semiparametric comparison of regression curves. *Ann. Statist.* **18** 63–89.

HART, J. D. (1997). *Nonparametric Smoothing and Lack-of-Fit Tests*. Springer, New York.

HASTIE, T. and TIBSHIRANI, R. (1986). Generalized additive models (with discussion). *Statist. Sci.* **1** 297–318.

HONG, Y. and WHITE, H. (1995). Consistent specification testing via nonparametric series regression. *Econometrica* **63** 1133–1159.

HOROWITZ, J. L. and SPOKOINY, V. G. (2001). An adaptive rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* **69** 599–631.

IMBENS, G. W. (1997). One-step estimators for over-identified generalized method of moments models. *Rev. Econom. Stud.* **64** 359–383.

JOHNSTON, G. (1982). Probabilities of maximal deviations for nonparametric regression function estimates. *J. Multivariate Anal.* **12** 402–414.

KITAMURA, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Ann. Statist.* **25** 2084–2102.

KITAMURA, Y. (2001). Asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica* **69** 1661–1672.

KITAMURA, Y., TRIPATHI, G. and AHN, H. (2002). Empirical likelihood based inference in conditional moment restriction models. Manuscript, Dept. Economics, Univ. Wisconsin, Madison.

LEBLANC, M. and CROWLEY, J. (1995). Semiparametric regression functionals. *J. Amer. Statist. Assoc.* **90** 95–105.

LIERO, H. (1982). On the maximal deviation of the kernel regression function estimate. *Math. Operationsforsch. Statist. Ser. Statist.* **13** 171–182.

NEWEY, W. K. (1985). Maximum likelihood specification testing and conditional moment tests. *Econometrica* **53** 1047–1070.

NEWEY, W. K. (1994). Kernel estimation of partial means and a general variance estimator. *Econometric Theory* **10** 233–253.

NEWEY, W. K. and MCFADDEN, D. (1994). Large sample estimation and hypothesis testing. In *Handbook of Econometrics* (R. Engle and D. McFadden, eds.) **4** 2111–2245. North-Holland, Amsterdam.

OWEN, A. (1984). The estimation of smooth curves. Stanford Linear Accelerator Center Publication 3394.

OWEN, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75** 237–249.

OWEN, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.* **18** 90–120.

OWEN, A. (1991). Empirical likelihood for linear models. *Ann. Statist.* **19** 1725–1747.

PRIESTLEY, M. (1981). *Spectral Analysis and Time Series*. Academic Press, New York.

QIN, J. and LAWLESS, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22** 300–325.

QIN, J. and LAWLESS, J. (1995). Estimating equations, empirical likelihood and constraints on parameters. *Canad. J. Statist.* **23** 145–159.

STANISWALIS, J. G. (1987). A weighted likelihood motivation for kernel estimators of a regression function with biomedical applications. Technical report, Virginia Commonwealth Univ.

STANISWALIS, J. G. and SEVERINI, T. A. (1991). Diagnostics for assessing regression models. *J. Amer. Statist. Assoc.* **86** 684–692.

WALD, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* **54** 426–482.

WHANG, Y. and ANDREWS, D. (1993). Tests of specification for parametric and semiparametric models. *J. Econometrics* **57** 277–318.

WOOLDRIDGE, J. (1992). A test for functional form against nonparametric alternatives. *Econometric Theory* **8** 452–475.

YATCHEW, A. (1992). Nonparametric regression tests based on least squares. *Econometric Theory* **8** 435–451.

ZHENG, J. (1996). A consistent test of functional form via nonparametric estimation techniques. *J. Econometrics* **75** 263–289.

DEPARTMENT OF ECONOMICS
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706
E-MAIL: gtripath@ssc.wisc.edu

DEPARTMENT OF ECONOMICS
UNIVERSITY OF PENNSYLVANIA
PHILADELPHIA, PENNSYLVANIA 19104
E-MAIL: ykitamur@ssc.upenn.edu