Econometric Theory, **22**, 2006, 543–586. Printed in the United States of America.
DOI: 10.1017/S0266466606060282

A DATA-DRIVEN NONPARAMETRIC SPECIFICATION TEST FOR DYNAMIC REGRESSION MODELS

ALAIN GUAY
Université du Québec à Montréal

EMMANUEL GUERRE
LSTA, Université Paris 6

The paper introduces a new nonparametric specification test for dynamic regression models. The test combines chi-square statistics based on Fourier series regression. A data-driven choice of the regression order, which uses the square root of the number of Fourier coefficients, is proposed. The benefits of the new test are (1) the selection procedure produces explicit and chi-square critical values that give a finite-sample size close to the nominal size; (2) the test is adaptive rate-optimal and detects local alternatives converging to the null with a rate that can be made arbitrarily close to the parametric rate. Simulation experiments illustrate the practical relevance of the new test.

1. INTRODUCTION

Starting with Bierens (1984) and Robinson (1989), nonparametric specification testing for dependent data has received much attention in the econometric literature. The range of potential applications includes nonlinearity tests and time series model building as reviewed in Tjøstheim (1994) and Fan and Yao (2003), specification of a continuous-time diffusion model for interest rates (Aït-Sahalia, 1996), specification of the Phillips curve (Hamilton, 2001), rational expectations models and conditional portfolio efficiency (Chen and Fan, 1999; Robinson, 1989), and tests of the Black and Scholes formula (Aït-Sahalia, Bickel, and Stocker, 2001) among others.

An important branch of this literature has considered a nonparametric approach that uses a smoothing parameter, such as a bandwidth or the order of a series expansion. This has raised two important issues, the detection properties and the size accuracy. The former can be addressed with efficiency considerations,

The first author acknowledges financial support from the Fonds Québécois de la Recherche sur la Société et la Culture (FQRSC). The second author acknowledges financial support from LSTA. Address correspondence to Alain Guay, Université du Québec à Montréal, CIRPÉE and CIREQ; e-mail: guay.alain@uqam.ca. Emmanuel Guerre, LSTA (Université Paris 6); e-mail: eguerre@ccr.jussieu.fr.

 0266-4666/06 $12.00

as pioneered in Ingster (1992, 1993); see also Guerre and Lavergne (2002). This framework leads to calibration tests to detect alternatives, in a given smoothness class, that approach the null at the fastest possible rate. However, the proposed smoothing parameters depend upon the chosen smoothness class, which is too restrictive for practical applications because the choice of a smoothness class is often arbitrary. Regarding the size issue, the statistics considered in the literature are often quadratic, but the critical values are computed from a normal approximation that may be inaccurate; see Hong and White (1995) for nonparametric series and Tjøstheim (1994) for kernel methods. Recent work for independent and identically distributed (i.i.d.) observations, such as Fan, Zhang, and Zhang (2001), suggests that more sophisticated approximations should be used instead of the normal. Härdle and Mammen (1993) and Gozalo (1997), among others, have proposed bootstrapped critical values as a solution. This may be difficult when the parametric model under consideration is specified in continuous time and is therefore costly to simulate or to bootstrap. Bootstrapping is also a burden when the dynamic specification includes covariates that are not strongly exogenous and need to be simulated.

An important step for the detection issue was the development of the adaptive framework. Under this approach, the smoothness class containing the alternative is considered unknown. Adaptive tests combine several statistics, designed for a specific class, to build a test; see Hart (1997) for a review of earlier work in this direction. Spokoiny (1996) has developed an efficiency theory for the adaptive case. Various papers considered adaptive rate-optimal tests using the maximum of the statistics, including Fan (1996), Fan and Huang (2001), Horowitz and Spokoiny (2001), and Spokoiny (1996, 2001). More specifically, Horowitz and Spokoiny (2001) have proposed an adaptive rate-optimal kernel-based specification test for a general parametric regression model that has generated various extensions. Baraud, Huet, and Laurent (2003) consider some nonasymptotic refinements of the maximum approach for specification of a linear model. Poo, Sperlich, and Vieu (2004) are interested in a semiparametric null hypothesis, whereas Gayraud and Pouet (2005) considered a nonparametric null. Gao and King (2001, 2004) and Fan and Yao (2003) have proposed extending the scope of applications to dependent data.

However, the maximum approach produces statistics with unstable asymptotic null behavior, so that achieving an accurate size remains a difficult issue. Fan (1996) found that the null limit distribution of his test gives a poor approximation for finite samples. Horowitz and Spokoiny (2001) did not derive a null limit distribution and used simulated critical values. On the other hand, Guerre and Lavergne (2005) built on a data-driven selection procedure that, under the null, selects a prescribed statistic with a high probability. Compared to the maximum approach, this considerably reduces the complexity of the null behavior of the resulting test statistic, which asymptotic distribution is a standard normal given by a specific statistic. But the statistics of Guerre and Lavergne (2005) have a complicated quadratic structure, and so these authors used bootstrapped critical to achieve a level close to the nominal size. Hence, as mentioned earlier, such an approach may not be suitable for a dynamic model.

In this paper, a suitable modification of the Guerre and Lavergne (2005) test is proposed to derive an adaptive rate-optimal specification test with an accurate size in a dynamic setting. The null hypothesis considered is the specification of the conditional mean for a time series with heteroskedastic innovations. Nonparametric series methods are used to compute chi-square statistics of various orders, which, in case of low degrees of freedom, have an accurate chi-square approximation under the null. A selection criterion, using a low penalty term proportional to the square root of the number of coefficients, chooses a test statistic. Hence the rejection region of the test can use accurate chi-square critical values. The rest of the paper is organized as follows. Section 2 presents our test and the adaptive framework on a nontechnical level. Section 3 groups our main assumptions and our main results. After studying the null behavior of the test, adaptive rate-optimality is introduced, and the test is shown to be efficient. Detection of local alternatives, approaching the null with a rate close to the parametric one, is also considered. Section 4 illustrates the size and detection properties of the test with a simulation experiment, and Section 5 concludes the paper. The proofs are grouped in Section 6 and two Appendixes.

2. HEURISTICS OF THE DATA-DRIVEN TEST

Consider an autoregressive model with exogenous variables Z_t,

$$Y_t = \mu(Y_{t-1}, \ldots, Y_{t-q}, Z_t) + \varepsilon_t = \mu(X_t) + \varepsilon_t$$

with $X_t = [Y_{t-1}, \ldots, Y_{t-q}, Z_t']' \in \mathbb{R}^d$, $\mathbb{E}[\varepsilon_t | \mathcal{F}_t] = 0$, and $\text{Var}[\varepsilon_t | \mathcal{F}_t] = \sigma^2(X_t)$, where $\mathcal{F}_t$ is the past Borel field generated by $X_1, \ldots, X_t$. Given T observations $(Y_1, X_1), \ldots, (Y_T, X_T)$, we want to test that $\mu(\cdot)$ belongs to some parametric family $\{m(\cdot;\theta), \theta \in \Theta \in \mathbb{R}^p\}$, that is, the correct specification hypothesis

$$H_0: \mu(\cdot) = m(\cdot;\theta) \quad \text{for some } \theta \in \Theta.$$

The proposed procedure builds on the estimated residuals $\hat{U}_t = Y_t - m(X_t;\hat{\theta}_T)$, where $\hat{\theta}_T$ is a consistent estimator of θ under H_0, such as, for instance, the nonlinear least squares estimator

$$\hat{\theta}_T = \arg\min_{\theta \in \theta} \sum_{t=1}^{T} (Y_t - m(X_t;\theta))^2.$$

By $Y_t = \mu(X_t) + \varepsilon_t$, the residuals decompose as $\hat{U}_t = \hat{\Delta}(X_t) + \varepsilon_t$, where $\hat{\Delta}(\cdot) = \mu(\cdot) - m(\cdot;\hat{\theta}_T)$ indicates potential misspecification, which asymptotically vanishes under the null but not under the alternative. Our test combines nonparametric series statistics constructed by projecting the residuals to detect the presence of a significant $\hat{\Delta}(\cdot)$ over a compact $\Lambda = [-\lambda, \lambda]^d$. More specifically, we focus on multivariate Fourier series regression.[1] For $k = [k_1, \dots, k_d]' \in \mathbb{Z}^d$, define the kth trigonometric function over Λ as

$$\psi_k(x) = \prod_{\ell=1}^{d} F_{k_\ell}(x_\ell),$$

$$\text{where } F_n(z) = \begin{cases} \dfrac{\sqrt{2}}{\sqrt{\lambda}} \cos\left(\dfrac{\pi n z}{\lambda}\right) \mathbb{I}(-\lambda \le z \le \lambda) & n < 0, \\ \dfrac{1}{\sqrt{2\lambda}} \mathbb{I}(-\lambda \le z \le \lambda) & n = 0, \\ \dfrac{\sqrt{2}}{\sqrt{\lambda}} \sin\left(\dfrac{\pi n z}{\lambda}\right) \mathbb{I}(-\lambda \le z \le \lambda) & n > 0, \end{cases} \quad \textbf{(2.1)}$$

so that $\{\psi_k(\cdot), k \in \mathbb{Z}^d\}$ is an $L_2(dx)$-orthonormal system, that is, $\int_\Lambda \psi_k(x) \psi_{k'}(x)\, dx = 1$ if $k = k'$ and 0 otherwise. Let $|k| = \sum_{\ell=1}^{d} |k_\ell|$ be the degree of $\psi_k(\cdot)$. The series estimation of $\hat{\Delta}(\cdot)$ over Λ builds on trigonometric multivariate polynomial function $\sum_{|k|\le K} b_k \psi_k(\cdot)$ of degree K, with a number c_K of coefficients b_k proportional to K^d. To account for heteroskedasticity, assume that an estimator $\hat{\sigma}(\cdot)$ of $\sigma(\cdot)$ is given and consider the generalized least squares estimator $\hat{\beta}_K = [\hat{b}_k, |k| \le K]'$,

$$\hat{\beta}_K = (\Psi_K' \hat{\Omega}^{-1} \Psi_K)^{-1} \Psi_K' \hat{\Omega}^{-1} \hat{U} = \arg \min_{[b_k, |k| \le K]'} \sum_{t=1}^{T} \left(\frac{\hat{U}_t - \sum_{|k|\le K} b_k \psi_k(X_t)}{\hat{\sigma}(X_t)} \right)^2,$$

where $\hat{U} = [\hat{U}_1, \dots, \hat{U}_T]'$, $\hat{\Omega}^{1/2}$ is the diagonal matrix with entries $\hat{\sigma}(X_t)$, and Ψ_K is the $T \times c_K$ matrix $[\psi_k(X_t), 1 \le t \le T, |k| \le K]$. Suppose that $\hat{\Delta}(\cdot)$ is a trigonometric polynomial function of order K. A standard procedure to test the significance of Fourier coefficients would use the chi-square statistic

$$\hat{R}_K = \hat{U}' \hat{\Omega}^{-1} \Psi_K (\Psi_K' \hat{\Omega}^{-1} \Psi_K)^{-1} \Psi_K' \hat{\Omega}^{-1} \hat{U} = \sum_{t=1}^{T} \left(\frac{\sum_{|k|\le K} \hat{b}_k \psi_k(X_t)}{\hat{\sigma}(X_t)} \right)^2, \quad \textbf{(2.2)}$$

leading to rejection of H_0 when $\hat{R}_K$ is large. However, assuming that $\hat{\Delta}(\cdot)$ has a finite series expansion of known order K is too simplistic for practical applications. More generally, an arbitrary choice of K may affect the power, and a better understanding of the impact of K is important to build a proper specification test. Set $\hat{\Delta} = [\hat{\Delta}(X_1), \dots, \hat{\Delta}(X_T)]'$ and $\varepsilon = [\varepsilon_1, \dots, \varepsilon_T]'$ so that $\hat{U} = \hat{\Delta} + \varepsilon$ and $\hat{R}_K$ decomposes into three terms $\hat{R}_K = \hat{R}_{1K} + 2\hat{R}_{2K} + \hat{R}_{3K}$ with

$$\hat{R}_{1K} = \hat{\Delta}' \hat{\Omega}^{-1} \Psi_K (\Psi_K' \hat{\Omega}^{-1} \Psi_K)^{-1} \Psi_K' \hat{\Omega}^{-1} \hat{\Delta},$$
$$\hat{R}_{2K} = \hat{\Delta}' \hat{\Omega}^{-1} \Psi_K (\Psi_K' \hat{\Omega}^{-1} \Psi_K)^{-1} \Psi_K' \hat{\Omega}^{-1} \varepsilon,$$
$$\hat{R}_{3K} = \varepsilon' \hat{\Omega}^{-1} \Psi_K (\Psi_K' \hat{\Omega}^{-1} \Psi_K)^{-1} \Psi_K' \hat{\Omega}^{-1} \varepsilon. \quad \textbf{(2.3)}$$

The term $\hat{R}_{1K}$ is crucial regarding detection of potential misspecification. It is the squared norm of the orthogonal projection of $\hat{\Omega}^{-1/2}\hat{\Delta}$ on the columns of $\hat{\Omega}^{-1/2}\Psi_K$, which increases with K up to $\sum_{t=1}^{T}\hat{\Delta}^2(X_t)\mathbb{I}(X_t \in \Lambda)/\hat{\sigma}^2(X_t)$, achieved for $c_K \geq T$. Hence $\hat{R}_{1K}$ can be viewed as a downward-biased estimation of the empirical measure of misspecification $\sum_{t=1}^{T}\hat{\Delta}^2(X_t)\mathbb{I}(X_t \in \Lambda)/\hat{\sigma}^2(X_t)$, that is,

$$\hat{R}_{1K} = \sum_{t=1}^{T}\frac{\hat{\Delta}^2(X_t)}{\hat{\sigma}^2(X_t)}\mathbb{I}(X_t \in \Lambda) + \text{bias}_\mu(K),$$

where $\text{bias}_\mu(K) \leq 0$ depends upon the unknown $\mu(\cdot)$ and decreases with K. The other important term in the decomposition (2.3) of the statistic $\hat{R}_K$ is $\hat{R}_{3K}$, a pure noise term. It can be expected that $\hat{R}_{3K}$ is asymptotically a chi-square variable with c_K degree of freedom, with mean c_K and variance $2c_K$, so that $\hat{R}_{3K} = c_K + \sqrt{2c_K}O_{\mathbb{P}}(1)$. Neglecting[2] $\hat{R}_{2K}$ and substituting in (2.3) gives a bias variance type decomposition for $\hat{R}_K - c_K$[3]

$$\hat{R}_K - c_K = \sum_{t=1}^{T}\frac{\hat{\Delta}^2(X_t)}{\hat{\sigma}^2(X_t)}\mathbb{I}(X_t \in \Lambda) + \text{bias}_\mu(K) + \sqrt{2c_K}O_{\mathbb{P}}(1). \tag{2.4}$$

Looking for the best estimator $\hat{R}_K - c_K$ of the misspecification indicator suggests that an ideal choice of K should achieve the minimum of $|\text{bias}_\mu(K)| + \sqrt{2c_K}O_{\mathbb{P}}(1)$. However, this is infeasible in practice, at least because $\text{bias}_\mu(\cdot)$ depends upon the unknown $\mu(\cdot)$. Alternative feasible choices of K include the Akaike information criterion (AIC) and Bayesian information criterion (BIC) as reviewed in Hart (1997). These selection procedures consider a K achieving the maximum of $\hat{R}_K - \gamma c_K$ where γ is a penalty parameter. According to (2.4), this amounts to achieving the minimum of $|\text{bias}_\mu(K)| + (\gamma - 1)c_K(1 + o_{\mathbb{P}}(1))$. Therefore these selection procedures asymptotically balance $|\text{bias}_\mu(K)|$ with $(\gamma - 1)c_K$ in place of the ideal order $c_K^{1/2}$ in (2.4). This suggests using instead a lower penalty term of the form $c_k + \gamma c_K^{1/2}$ affecting the square root of the number of coefficients $c_K^{1/2}$ in place of c_k. More specifically, let $\mathcal{K}_T$ be a set of admissible degree K larger than or equal to $K_{\min}$. Our data-driven choice of K is

$$\begin{aligned}\hat{K}^\gamma &= \arg\max_{K\in\mathcal{K}_T}\{\hat{R}_K - c_K - \gamma_T(2(c_K - c_{K_{\min}}))^{1/2}\}\\ &= \arg\max_{K\in\mathcal{K}_T}\{\hat{R}_K - \hat{R}_{K_{\min}} - (c_K - c_{K_{\min}}) - \gamma_T(2(c_K - c_{K_{\min}}))^{1/2}\}\\ &\qquad\qquad \text{with } \gamma_T \geq 0.\end{aligned} \tag{2.5}$$

The introduction of $K_{\min}$ quantities in the penalty criterion reflects a preference for low degree as justified now from considerations on the null behavior of the retained $\hat{R}_{\hat{K}_\gamma}$.

As seen from Fan (1996) or Horowitz and Spokoiny (2001), finding an accurate approximation for the null distribution of a statistic that combines the $\hat{R}_K$'s as $\hat{R}_{\hat{K}^\gamma}$ is difficult. A first distinctive feature is that the selection procedure (2.5) is flexible enough to limit the contribution of the statistics with high K by taking γ_T large enough. Indeed, a limit case is $\gamma_T = +\infty$, which gives that $\hat{K}^\gamma = K_{\min}$. This continues to hold asymptotically provided γ_T diverges fast enough, as shown in Theorem 1 in Section 3. Moreover, as detailed now, an accurate approximation of the distribution of $\hat{R}_{\hat{K}^\gamma}$ is a standard chi-square. Because $\hat{\Delta}(\cdot)$ asymptotically vanishes under H_0, (2.3) shows that the null distribution of $\hat{R}_K$ is approximately that of $\hat{R}_{3K}$ and then, neglecting the effect of the variance estimation, of

$$R_{3K} = \varepsilon'\Omega^{-1}\Psi_K(\Psi_K'\Omega^{-1}\Psi_K)^{-1}\Psi_K'\Omega^{-1}\varepsilon,$$

where $\Omega^{1/2} = \text{Diag}[\sigma(X_1),\ldots,\sigma(X_T)]$. In the i.i.d. case and according to the Berry–Esseen bound in Hart (1997, Thm. 7.2), the distribution of the vector

$$\Psi_K'\Omega^{-1}\varepsilon = \left[\sum_{t=1}^{T}\frac{\psi_k(X_t)\varepsilon_t}{\sigma^2(X_t)}, |k| \le K\right]'$$

has a normal approximation up to an error $a(c_K)/T^{1/2}$ where $a(c_K)$ diverges with c_K. Therefore, the distribution of the chi-squared statistic R_{3K} should be close to a chi-square with c_K degree of freedom up to an error $a(c_K)/T^{1/2}$, which is smaller for moderate K.[4] Hence the test uses a chi-square critical value $z_\alpha = z_{\alpha,T}$ with

$$\mathbb{P}\left(\frac{\chi(c_{K_{\min}}) - c_{K_{\min}}}{\sqrt{2c_{K_{\min}}}} \ge z_\alpha\right) = \alpha,$$

where $\chi(c)$ is a chi-square with c degree of freedom and rejects H_0 if[5]

$$\frac{\hat{R}^\gamma - c_{K_{\min}}}{\sqrt{2c_{K_{\min}}}} \ge z_\alpha \quad \text{where } \hat{R}^\gamma = \hat{R}_{\hat{K}^\gamma}. \tag{2.6}$$

Consider now the power issue. The data-driven choice (2.5) of K combines the detection properties of each of the $\hat{R}_K$'s. Indeed, because $c_K \ge c_{K_{\min}}$ for any K in $\mathcal{K}_T$, we have

$$\begin{aligned}\hat{R}^\gamma - c_{K_{\min}} \ge \hat{R}_{\hat{K}^\gamma} - c_{\hat{K}^\gamma} &= \max_{K\in\mathcal{K}_T}\{\hat{R}_K - c_K - \gamma_T(2(c_K - c_{K_{\min}}))^{1/2}\} \\ &\quad + \gamma_T(2(c_{\hat{K}^\gamma} - c_{K_{\min}}))^{1/2} \\ &\ge \max_{K\in\mathcal{K}_T}\{\hat{R}_K - c_K - \gamma_T(2(c_K - c_{K_{\min}}))^{1/2}\} \\ &\ge \hat{R}_K - c_K - \gamma_T(2(c_K - c_{K_{\min}}))^{1/2}. \end{aligned} \tag{2.7}$$

https://doi.org/10.1017/S0266466606060282 Published online by Cambridge University Press

This gives the power lower bound

$$\mathbb{P}\left(\frac{\hat{R}^{\gamma} - c_{K_{\min}}}{\sqrt{2c_{K_{\min}}}} \geq z_\alpha\right) \geq \mathbb{P}(\hat{R}_K - c_K - \gamma_T(2(c_K - c_{K_{\min}}))^{1/2} - z_\alpha\sqrt{2c_{K_{\min}}} \geq 0)$$
$$= \mathbb{P}\left(\frac{\hat{R}_K - c_K}{\sqrt{2c_K}} \geq \frac{z_\alpha\sqrt{2c_{K_{\min}}} + \gamma_T(2(c_K - c_{K_{\min}}))^{1/2}}{\sqrt{2c_K}}\right), \tag{2.8}$$

which holds in particular for an optimal K that balances the bias with the penalty term. Taking $K = K_{\min}$ gives that

$$\mathbb{P}\left(\frac{\hat{R}^{\gamma} - c_{K_{\min}}}{\sqrt{2c_{K_{\min}}}} \geq z_\alpha\right) \geq \mathbb{P}\left(\frac{\hat{R}_{K_{\min}} - c_{K_{\min}}}{\sqrt{2c_{K_{\min}}}} \geq z_\alpha\right), \tag{2.9}$$

a power bound that shows that the test (2.6) improves on the one using the single statistic $\hat{R}_{K_{\min}}$. As seen from (2.4) and (2.8), consistency holds as soon as there is a degree K in $\mathcal{K}_T$ such that the misspecification measure $\sum_{t=1}^{T}\hat{\Delta}^2(X_t)\mathbb{I}(X_t \in \Lambda)/\hat{\sigma}^2(X_t)$ is asymptotically larger than the sum of $|\text{bias}_\mu(K)|$, $\gamma_T\sqrt{2(c_K - c_{K_{\min}})}$, and $z_\alpha\sqrt{2c_{K_{\min}}}$. Hence increasing γ_T too much should give a less powerful test. The form of the low penalty term in (2.5) is crucial to show adaptive rate-optimality; see Theorem 2 in Section 3. Theorem 3 in Section 3 shows that the test detects Pitman local alternatives with a rate arbitrarily close to the rate $T^{-1/2}$.

3. MAIN RESULTS

3.1. Main Assumptions

Consider T observations (Y_t, X_t) with $Y_t = \mu(X_t) + \varepsilon_t$, $X_t = (Y_t, \ldots, Y_{t-q}, Z_t')' \in \mathbb{R}^d$, and where $\mu(\cdot)$ can depend upon T, in which case (Y_t, X_t) forms a triangular array (Y_{tT}, X_{tT}). Let $\underline{\mathcal{X}}_t$ and $\overline{\mathcal{X}}_t$ denote the Borel field generated by $X_1, \varepsilon_1, \ldots, X_t, \varepsilon_t$ and $X_t, \varepsilon_t, X_{t+1}, \varepsilon_{t+1}, \ldots$, respectively. The α-mixing coefficients of $\{X_t, \varepsilon_t\}_{t\in\mathbb{N}^*}$ are

$$\alpha(n) = \sup_{t\in\mathbb{N}^*}\ \sup_{A\in\underline{\mathcal{X}}_t,\, B\in\overline{\mathcal{X}}_{t+n}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|, \qquad n \in \mathbb{N}.$$

The next assumptions deal with the ε_t's, the mixing coefficients, and the parametric mean.

Assumption E. Let $\mathcal{F}_t$ be the Borel field generated by $(X_1, \varepsilon_0), \ldots, (X_t, \varepsilon_{t-1})$. The variables $\{\varepsilon_t\}_{t\in\mathbb{N}}$ are martingale difference with $\mathbb{E}[\varepsilon_t|\mathcal{F}_t] = 0$, $\mathbb{E}[\varepsilon_t^2|\mathcal{F}_t] = \sigma^2(X_{t-1})$, and $\sup_{t\in\mathbb{N}}\mathbb{E}[\varepsilon_t^8|\mathcal{F}_t] < \infty$ a.s. The standard deviation function, $\sigma(\cdot) = \text{Var}[\varepsilon_t|X_t = \cdot]$, is continuous and bounded away from 0 on $\mathbb{R}^d$.

Assumption X. The process $\{X_t, \varepsilon_t\}_{t \in \mathbb{N}^*}$ on $\mathbb{R}^d \times \mathbb{R}$ is stationary, with the following conditions holding.

(i) $\alpha(n) \leq An^{-1-a}$ for some constant $A, a > 0$.
(ii) The variable X_t has a density $f(\cdot)$ with respect to the Lebesgue measure on $\mathbb{R}^d$. The density $f(\cdot)$ is bounded away from 0 and infinity.

Assumption M. The parameter set Θ is a subset of $\mathbb{R}^p$, and the following conditions hold.

(i) The regression function $m(x;\theta)$ is twice continuously differentiable with respect to θ. The gradient $m^{(1)}(x;\theta)$ and Hessian matrix $m^{(2)}(x;\theta)$ are bounded over $\Lambda \times \Theta$.
(ii) For any sequence of regression functions $\mu_T(\cdot)$ with $\mathbb{E}\mu_T^2(X_t) < \infty$, there exists a sequence of parameter θ_T in Θ such that $T^{1/2}(\hat{\theta}_T - \theta_T) = O_{\mathbb{P}}(1)$, with $\theta_T = \theta$ if $\mu_T(\cdot) = m(\cdot;\theta)$ for some θ in Θ.

Assumption E ensures that the sums $\sum_{t=1}^T \psi_k(X_t)\varepsilon_t/\sigma^2(X_t)$ are martingales that are asymptotically normal under Assumption X(i). The polynomial mixing rate of X(i) is a minimal rate to achieve $T^{1/2}$-consistency in the weak law of large numbers for the empirical mean $T^{-1}\Psi_K'\Omega^{-1}\Psi_K$. Under Assumption X(ii), the limit of $T^{-1}\Psi_K'\Omega^{-1}\Psi_K$ has an inverse. Mixing conditions for Markovian (Y_t, X_t) as in Assumption X(i) can be derived using a drift condition; see Fan and Yao (2003, Thm. 2.4) and the references therein. When $\hat{\theta}_T = \arg\min_{\theta\in\theta} \sum_{t=1}^T (Y_t - m(X_t;\theta))^2$, the sequence θ_T in Assumption M(ii) is the pseudo–true value $\arg\min_{\theta\in\Theta} \mathbb{E}(\mu_T(X_t) - m(X_T;\theta))^2$, which is uniquely defined under identification of the parametric regression model; see Domowitz and White (1982). Assumption M(i) then ensures that $\hat{\Delta}(\cdot) = \mu_T(\cdot) - m(\cdot;\hat{\theta}_T)$ is close to $\Delta(\cdot) = \mu_T(\cdot) - m(\cdot;\theta_T)$ over Λ up to an $O_{\mathbb{P}}(T^{-1/2})$ term.

Let us now turn to the construction of the test. The first assumption specifies a set of admissible degrees $\mathcal{K}_T$ in the spirit of the dyadic bandwidth set of Horowitz and Spokoiny (2001).

Assumption K. Let a be as in Assumption X. Set $K_{\max} = 2^{J_{\max}} = O(T^{C_1/d})$ for some C_1 in $(0, \frac{3}{4}[(1+a)/(5+3a)])$, $K_{\min} = 2^{J_{\min}} \to \infty$ with $K_{\min}^d = O(\ln^{C_2} T)$ for $C_2 > 0$, where $J_{\min} \leq J_{\max}$ are integer numbers. The set of admissible degrees $\mathcal{K}_T$ is dyadic, that is,

$$\mathcal{K}_T = \{K = 2^J, J = J_{\min}, J_{\min}+1, \ldots, J_{\max}\}. \tag{3.1}$$

Note that (3.1) and the polynomial divergence rate of $K_{\max}$ imply that Card $\mathcal{K}_T$ is of exact order $\ln T$. Such a restriction is helpful to show that $\hat{K}^\gamma = K_{\min}$ asymptotically under the null but also has some practical justifications. Indeed, achieving a small $\mathbb{P}(\hat{K}^\gamma \neq K_{\min})$ is an important condition to get an accurate size. Because $\hat{R}_K - \hat{R}_{K_{\min}} - (c_K - c_{K_{\min}}) - \gamma_T(2(c_K - c_{K_{\min}}))^{1/2}$ vanishes if and only if $K = K_{\min}$, (2.5) yields that $\hat{K}^{\gamma} \neq K_{\min}$ if and only if one of these penalized statistics is strictly positive for a $K \neq K_{\min}$, or equivalently

$$\max_{K \in \mathcal{K}_T \setminus \{K_{\min}\}} \frac{\hat{R}_K - \hat{R}_{K_{\min}} - (c_K - c_{K_{\min}})}{(2(c_K - c_{K_{\min}}))^{1/2}} > \gamma_T.$$

Hence

$$\mathbb{P}(\hat{K}^{\gamma} \neq K_{\min}) = \mathbb{P}\left(\max_{K \in \mathcal{K}_T \setminus \{K_{\min}\}} \frac{\hat{R}_K - \hat{R}_{K_{\min}} - (c_K - c_{K_{\min}})}{(2(c_K - c_{K_{\min}}))^{1/2}} \geq \gamma_T \right), \tag{3.2}$$

so that $\mathbb{P}(\hat{K}^{\gamma} \neq K_{\min})$ increases with $\mathcal{K}_T$ and decreases with the penalty sequence γ_T. Therefore, using a parsimonious $\mathcal{K}_T$ can improve the size accuracy of the test. On the other hand, a dyadic $\mathcal{K}_T$ as in Assumption K contains sequences with any arbitrary order between $\ln^{C_2} T$ and T^{C_1} that is sufficient for adaptive rate-optimality. The constant C_1 of Assumption K must be smaller than $\frac{3}{4}[(1+a)/(5+3a)]$ where a comes from Assumption X(i), $\alpha(n) = O(n^{-1-a})$. This gives a $K_{\max}$ of order $T^{1/(4d)}$ at best, whereas, in the i.i.d. setup, Hong and White (1995) allowed for a better order $T^{1/(3d)}$ when using a single series statistic on which to base the test.

Let us now turn to variance estimation. The next condition allows us to approximate $T^{-1}\Psi_K'\hat{\Omega}^{-1}\Psi_K$ with $T^{-1}\Psi_K'\Omega^{-1}\Psi_K$ for degrees K depending on the sample size T, as in Assumption K.

Assumption V. Let $K_{\max} = 2^{J_{\max}} = \max\{K; K \in \mathcal{K}_T\}$. Then, for the considered sequence of regression models $Y_t = \mu_T(X_t) + \varepsilon_t$, $\sup_{x\in\Lambda}|\hat{\sigma}(x) - \sigma(x)| = O_{\mathbb{P}}(v_T)$ and, for some integer $\ell > d/2$ and all $(\ell_1, \ldots, \ell_d)$ with $\ell_1 + \cdots + \ell_d = \ell$, $\sup_{x\in\Lambda}|\partial^{\ell}\hat{\sigma}(x)/(\partial^{\ell_1}x_1 \ldots \partial^{\ell_d}x_d)| = O_{\mathbb{P}}(v_T)$, where $v_T = o(K_{\max}^{-3d/2}/\ln T)$ and $\liminf_{T\to\infty} T^{1/2}v_T > 0$.

Assumption V requires consistency of $\hat{\sigma}(\cdot)$ under the null and the alternative. Convergence of $\hat{\sigma}(\cdot)$ with the rate v_T requires that $\mu_T(\cdot)$ and $\sigma(\cdot)$ satisfy a minimal smoothness condition. As seen from Guerre and Lavergne (2002), consistency is not necessary under the alternative but can be useful to get a powerful test. Under homoskedasticity, a simple choice of $\hat{\sigma}(\cdot)$ is a constant difference-based estimator, in which case Assumption V holds with a best possible $v_T = T^{-1/2}$ so that $K_{\max} = o(T^{1/(3d)}\ln^{2/(3d)} T)$. The heteroskedastic case requires nonparametric variance estimation, such as kernel, sieves, series expansion; see, among others, Guerre and Lavergne (2002, 2005) and Horowitz and Spokoiny (2001). The rate v_T is then the consistency rate for the ℓth partial derivatives, which restricts the divergence rate of $K_{\max}$.

3.2. Asymptotic Behavior under the Null

As discussed following (3.2) and (2.9), a fast divergence rate for γ_T is useful to achieve an accurate size under the null but may negatively affect its power properties. Therefore, an important issue is to find a minimal divergence rate for γ_T ensuring that the test is asymptotically of level α or equivalently that $\mathbb{P}(\hat{K}^\gamma \neq K_{\min})$ asymptotically vanishes under H_0. The Bonferroni inequality gives, in (3.2),

$$\mathbb{P}(\hat{K}^\gamma \neq K_{\min}) \leq \sum_{K \in \mathcal{K}_T \setminus \{K_{\min}\}} \mathbb{P}\left(\frac{\hat{R}_K - \hat{R}_{K_{\min}} - (c_K - c_{K_{\min}})}{(2(c_K - c_{K_{\min}}))^{1/2}} > \gamma_T \right), \tag{3.3}$$

and showing that the last sum asymptotically vanishes for small γ_T necessitates precise uniform bounds for these probabilities, so that simple Chebychev-type inequalities may not be sufficient. Better Gaussian-type bounds in the spirit of Mill's ratio inequality $\mathbb{P}(\mathcal{N}(0,1) \geq \gamma) \leq \exp(-\gamma^2/2)/(\sqrt{2\pi}\gamma)$ are derived in Lemma A.3 in Appendix A. Because the exact order of Card $\mathcal{K}_T$ is $\ln T$, the next theorem ensures that the asymptotic size of the test is α provided that the penalty sequence γ_T diverges faster than $(\ln \ln T)^{1/2}$.

THEOREM 1. *Consider that the null hypothesis H_0 is true and assume that Assumptions E, K, M, V, and X hold. Then, if γ_T diverges with*

$$\gamma_T \geq (1+\epsilon)\sqrt{2 \ln \text{Card}\, \mathcal{K}_T}, \quad \textit{for some } \epsilon > 0, \tag{3.4}$$

$\lim_{T\to+\infty} \mathbb{P}(\hat{K}^\gamma = K_{\min}) = 1$, *and the test (2.6) is asymptotically of level α.*

The minimal divergence rate $(\ln \ln T)^{1/2}$ ensuring that the test is asymptotically of level α is surprisingly low compared to the penalty term of order $\ln T$ used in the BIC criterion. Such improvement comes from the Gaussian-type bounds used for the tails of the standardized $\hat{R}_K - \hat{R}_{K_{\min}}$'s. Indeed, this gives, up to remainder terms, a bound Card $\mathcal{K}_T \exp(-\gamma_T^2/2)/(\sqrt{2\pi}\gamma_T)$ in (3.3), which asymptotically vanishes provided that (3.4) holds. On the other hand, such a low rate is in line with previous findings for rate-optimal adaptive testing. Indeed, (3.2) shows that suitable γ_T should resemble the critical values of a maximum test such as that of Fan (1996), who found critical values with a typical rate of $(2 \ln \ln T)^{1/2}$. This suggests that our minimal rate condition (3.4) cannot be improved.

Another condition for Theorem 1 to hold is that $K_{\min}$ diverges with the sample size; see Assumption K. This is used to neglect the parametric estimation error $T^{1/2}(\hat{\theta}_T - \theta)$ in the chi-square approximation of the distribution of $\hat{R}_{K_{\min}}$. Accounting for such an effect would allow us to consider a fixed $K_{\min}$; see, for example, Hart (1997, Sect. 8.3.1).

3.3. Detection of Small Alternatives

As discussed following equation (2.9), the detection properties of the test depend upon a bias term from (2.4). Establishing formal adaptive rate-optimality of the test necessitates bounding this bias. The current mathematical approach to do so makes use of some smoothness restrictions. We consider here Hölder smoothness classes $\mathcal{C}(L,s)$ that we introduce now. Define the departure from the null as

$$\Delta_{\mu,T}(\cdot) = \mu(\cdot) - m(\cdot;\theta_T),$$

with a θ_T as in Assumption M. We restrict ourselves to departures $\Delta(\cdot)$ with a restriction to Λ that admits a (2λ)-periodic extension. Consider first the case $s \in (0,1]$, for which

$$\mathcal{C}(L,s) = \left\{\Delta(\cdot) : \sup_{x,x' \in \Lambda} \frac{|\Delta(x) - \Delta(x')|}{\|x - x'\|^s} \leq L\right\}.$$

For real $s > 0$, let $\lfloor s \rfloor$ be the lower integer part of s, that is, the unique integer number satisfying $\lfloor s \rfloor < s \leq \lfloor s \rfloor + 1$, so that $s - \lfloor s \rfloor$ is in $(0,1]$ with $s - \lfloor s \rfloor = s$ for $s \in (0,1]$. For any $s > 0$, the smoothness class $\mathcal{C}(L,s)$ is defined as

$$\mathcal{C}(L,s) = \{\Delta(\cdot) : \text{the } \lfloor s \rfloor\text{th partial derivatives of } \Delta(\cdot) \text{ are in } \mathcal{C}(L, s - \lfloor s \rfloor)\}.$$

Hence the smoothness class $\mathcal{C}(L,s)$ is defined for all $s > 0$ and $L > 0$. Lemma 1 in Section 6 gives, for the bias term of (2.4), the following bound:

$$|\text{bias}_\mu(K)| \leq O_{\mathbb{P}}\left[T^{1/2}K^{-s}\mathbb{E}^{1/2}\left(\frac{\Delta_{\mu,T}(X_t)\mathbb{I}(X_t \in \Lambda)}{\sigma(X_t)}\right)^2 + TK^{-2s}\right],$$

for any $\Delta_{\mu,T}(\cdot)$ in $\mathcal{C}(L,s)$ and any K. This gives, for small alternatives, which are the harder to detect,

$$|\text{bias}_\mu(K)| \leq O_{\mathbb{P}}(TK^{-2s}) \quad \text{provided } \mathbb{E}^{1/2}\left(\frac{\Delta_{\mu,T}(X_t)\mathbb{I}(X_t \in \Lambda)}{\sigma(X_t)}\right)^2 = O(K^{-s}). \tag{3.5}$$

Our minimax adaptive framework evaluates tests uniformly over alternatives at distance ρ from the null, that is, in

$$H_1(\rho;L,s) = \Bigg\{\mu(\cdot) = m(\cdot;\theta_T) + \Delta_{\mu,T}(\cdot); \\ \Delta_{\mu,T}(\cdot) \in \mathcal{C}(L,s), \mathbb{E}\left(\frac{\Delta_{\mu,T}(X_t)\mathbb{I}(X_t \in \Lambda)}{\sigma(X_t)}\right)^2 \geq \rho^2\Bigg\},$$

with unknown smoothness index (L,s). Such alternatives allow for a general shape of $\Delta_{\mu,T}(\cdot)$ with narrow peaks and valleys that may depend upon on T; see Horowitz and Spokoiny (2001). As pointed out in Guerre and Lavergne (2005), uniform consistency over $H_1(\tilde{\rho}_T;L,s)$ is equivalent to consistency against any sequence

$$\mu_T(\cdot) = m(\cdot;\theta_T) + \Delta_T(\cdot), \quad \text{where } \Delta_T(\cdot) = \Delta_{\mu_T,T}(\cdot),$$

in $H_1(\tilde{\rho}_T; L, s)$ as considered here. A crucial issue is the choice of a suitable asymptotically vanishing rate $\tilde{\rho}_T$. Indeed, some of the alternatives of $H_1(\tilde{\rho}_T; L, s)$ will not be detected by any tests if $\tilde{\rho}_T$ goes to 0 at too rapid a rate. On the other hand, detection can become straightforward if $H_1(\tilde{\rho}_T; L, s)$ remains far from the null. Hence a good candidate $\tilde{\rho}_T$ to evaluate a test is a frontier rate that separates these two extreme situations. In the adaptive approach, such a rate depends upon the unknown smoothness index s, and Spokoiny (1996) has shown that the optimal adaptive rate is[6]

$$\rho_T = \rho_T(s) = \left(\frac{\sqrt{\ln\ln T}}{T}\right)^{2s/(4s+d)},$$

which is slower than the parametric rate $T^{-1/2}$. Guerre and Lavergne (2002) derived an optimal rate for a known smoothness index s that improves ρ_T from the $(\ln\ln T)^{1/2}$ factor, so that the price to pay for rate adaptation is moderate. As is well known, the rate ρ_T decreases faster than the nonparametric estimation rate $T^{-s/(2s+d)}$. The adaptive rate-optimality of our test is stated in the next result.

THEOREM 2. *Consider a sequence of alternatives*

$$\mu_T(\cdot) = m(\cdot;\theta_T) + \Delta_T(\cdot) \quad \text{in } H_1\,(C_3 \cdot \rho_T; L, s) \quad \text{for some unknown } s \text{ and } L,$$

with $s \geq d(2/C_1 - 1)/4$, $L > 0$, $C_3 > 0$, *and* $\sup_{x\in\Lambda}|\Delta_T(x)| = O[\mathbb{E}^{1/2}(\Delta_T^2(X_t)\mathbb{I}(X_t \in \Lambda)/\sigma^2(X_t))]$. *Assume that Assumptions E, K, M, and V hold. Then, if* γ_T *is of exact order* $(\ln\ln T)^{1/2}$ *and provided* C_3 *is taken large enough, the test is consistent, that is,* $\lim_{T\to\infty}\mathbb{P}((\hat{R}^\gamma - c_{K_{\min}})/\sqrt{2c_{K_{\min}}} \geq z_\alpha) = 1$.

The proof of Theorem 2 builds on the lower power bound (2.8) and on the bias variance decomposition (2.4). In view of the bias order (3.5) for small alternatives, an optimal choice of K in (2.8) is such that the order of the penalty term $\gamma_T K^{d/2}$ is proportional to TK^{-2s}, that is, for

$$K_* = K_*(s) = 2^{[2/(4s+d)[\ln(T/\gamma_T)/(\ln 2)]]} \propto \left(\frac{T}{\gamma_T}\right)^{2/(4s+d)}, \tag{3.6}$$

where $[\cdot]$ is the integer part. Such K_* detects alternatives within the bias order divided by the sample size, $K_*^{-s} \propto (\gamma_T/T)^{2s/(4s+d)}$, which coincides with the optimal adaptive order ρ_T provided γ_T has the smallest possible order $(\ln\ln T)^{1/2}$ compatible with Theorem 1. Note that, under Assumption K, K_* is in $\mathcal{K}_T$ provided $s \geq d(2/C_1 - 1)/4$, which implies that $s > 7d/4$.

Because adaptation means detection over various smoothness classes $\mathcal{C}(L, s)$, it is crucial that the test combine several statistics, as seen from the optimal K_*

in (3.6), which depends on the smoothing index s. Therefore, tests that use a single statistic $\hat{R}_K$ generally fail to be rate-optimal adaptive. A more specific property of the test (2.6) is detection of small local alternatives.

THEOREM 3. *Consider a sequence of local alternatives $\mu_T(\cdot)$ satisfying*

$$\mu_T(X_t) = m(X_t;\theta_T) + r_T\,\Delta_{0T}(X_t)$$

$$\text{with } \Delta_{0T}(\cdot) \text{ in } \mathcal{C}(L,s) \quad \text{for some unknown } s, L > 0,$$

$$s \geq d(2/C_1 - 1), \text{ and}$$

$$\mathbb{E}\left(\frac{\Delta_{0T}(X_t)\mathbb{I}(X_t \in \Lambda)}{\sigma(X_t)}\right)^2 \geq 1, \qquad \sup_{x\in\Lambda}|\Delta_{0T}(x)| = O(1).$$

Then, under Assumptions E, K, M, V, and X, the test is consistent provided $1/r_T = o(\sqrt{T/K_{\min}^{d/2}})$.

Because $K_{\min}$ can diverge very slowly, the rate r_T can be arbitrarily close to the parametric detection rate $1/T^{1/2}$. This slightly improves on the results of Horowitz and Spokoiny (2001), who achieved a rate $(\ln\ln T)^{1/2}/T^{1/2}$. A key argument there is that the local alternatives of Theorem 3 are asymptotically very smooth, because the departure from the null $r_T\Delta_{0T}(\cdot)$ is in $\mathcal{C}(Lr_T, s)$, with a Lipschitz constant Lr_T that goes to 0. Hence these alternatives differ from the general ones in Theorem 2, and they are typically detected by trigonometric series with low degree such as $K_{\min}$, so that (2.9) yields consistency of the test. On the other hand, using the single statistic $\hat{R}_{K_{\min}}$ would give a test that is not consistent against the alternatives of Theorem 2, so that combining several statistics as in our procedure is crucial to achieve these opposite kinds of detection properties.

4. SIMULATION EXPERIMENTS

In this section we study the size and the power properties of the proposed procedure when testing for a null of linearity in the context of a Markov process of order 1. The resulting test is compared with the one developed by Hamilton (2001) to detect nonlinearity. First, to examine the size properties, we use the AR(1)

$$Y_t = \rho Y_{t-1} + \epsilon_t.$$

Three distributions are considered for the error term: standard normal, standardized student with five degrees of freedom, and a centered and standardized exponential. To examine the sensitivity of the tests to temporal dependence, we consider various values of the autoregressive parameter ρ, namely, $\rho = 0$, 0.25, 0.50, 0.75. To implement our test, we choose the interval (Λ in Section 2) for projecting the covariate Y_{t-1} onto the trigonometric expansion to be equal to 2 divided by standard error of Y_t under the null. This corresponds to approximately 95% of the observations. The set $\mathcal{K}_T$ is equal to $\{1,2,4,8,16\}$. The asymptotic critical value is given by $(\chi_{0.05}(1)-1)/\sqrt{2}$, where $\chi_{0.05}(1)$ is the critical value at 5% of a chi-square with one degree of freedom. We study the small-sample properties of the test for various values of the penalty parameter γ_T. We fix γ_T equal to $c\sqrt{2 \ln \text{Card}\, \mathcal{K}_T}$ where we set $c = 2,3,5$. The parameters are estimated by ordinary least squares (OLS). The sample size is set to 200, and the number of simulations is equal to 10,000.

The simulation results for the size, which are presented in Table 1, are encouraging. For $c = 2$ the test slightly overrejects in all cases. However, for $c = 3,5$, the size is accurate whatever the distribution, persistence, and number of observations considered. The Lagrange multiplier (LM) test developed by Hamilton (2001) shares these good size properties.

To study the effect on power of the penalty sequence γ_T, two alternative specifications of the linear autoregressive process are examined. The first specification is a threshold autoregressive model defined as

$$Y_t = \rho_1 Y_{t-1}\,\mathbb{I}_{\{Y_{t-1}>0\}} + \rho_2 Y_{t-1}\,\mathbb{I}_{\{Y_{t-1}<0\}} + \epsilon_t,$$

where ϵ_t is i.i.d. $N(0,1)$.[7] This representation contains two regimes delimited by a threshold equal to zero. When Y_{t-1} is greater than zero, the dynamic dependence is controlled by the parameter ρ_1. In the case where it is inferior to zero,

Table 1. Size properties (5%) of our test and Hamilton test (LM) (200 observations)

Distribution	ρ	c = 2	3	5	LM
normal	0	0.057	0.048	0.047	0.047
student	0	0.055	0.047	0.046	0.047
exponential	0	0.056	0.048	0.047	0.048
normal	0.25	0.057	0.048	0.047	0.045
student	0.25	0.059	0.052	0.051	0.051
exponential	0.25	0.060	0.051	0.049	0.049
normal	0.50	0.057	0.047	0.045	0.044
student	0.50	0.056	0.050	0.050	0.051
exponential	0.50	0.057	0.050	0.047	0.049
normal	0.75	0.052	0.044	0.043	0.044
student	0.75	0.059	0.051	0.050	0.050
exponential	0.75	0.058	0.048	0.046	0.055

the dynamic depends on the parameter ρ_2. Under the null of linearity $\rho_1 = \rho_2$. The distance from the null is a function of the absolute value of the difference between ρ_1 and ρ_2. To see this, we can rewrite the threshold autoregressive model as follows:

$$Y_t = \rho_1 Y_{t-1} + (\rho_2 - \rho_1) Y_{t-1} \mathbb{I}_{\{Y_{t-1}<0\}} + \epsilon_t.$$

Thus, under the null, $\mu(X_t) = \rho_1 Y_{t-1}$ whereas the nonlinear alternative is

$$\mu(X_t) = \rho_1 Y_{t-1} + \delta Y_{t-1} \mathbb{I}(Y_{t-1} < 0), \quad \text{where } \delta = \rho_2 - \rho_1.$$

To examine the sensitivity of the tests to temporal dependence, we consider various types of dependence for the process Y_t. We run the following experiments: (1) $\rho_1 = 0$ and $\rho_2 = 0.25, 0.50, 0.75$, (2) $\rho_1 = 0.25$ and $\rho_2 = 0.50, 0.75, -0.50$, (3) $\rho_1 = 0.50$ and $\rho_2 = 0.25, 0, -0.25$, and (4) $\rho_1 = 0.75$ and $\rho_2 = 0.50, 0.25, 0$. The values of ρ_2 under the alternative are chosen such that the parameter (δ) that governs the distance from the null is equal to 0.25, 0.50, and 0.75, respectively. Table 2 reports the power results. Our test is more powerful than Hamilton's for all cases. Our power gains increase with the degree of temporal dependence and the distance of the alternative from the null. The difference in the rejection rate can be as high as 38%.

TABLE 2. Power properties (5%) of our test and Hamilton test (LM): First experiment (200 observations)

			c			
ρ_1	ρ_2	$\|\rho_2 - \rho_1\|$	2	3	5	LM
0	0.25	0.25	0.236	0.224	0.222	0.118
	0.50	0.50	0.653	0.646	0.645	0.361
	0.75	0.75	0.849	0.846	0.846	0.682
0.25	0.5	0.25	0.237	0.227	0.226	0.128
	0.75	0.50	0.583	0.576	0.575	0.413
	−0.50	0.75	0.947	0.945	0.944	0.666
0.50	0.25	0.25	0.261	0.247	0.246	0.123
	0	0.50	0.725	0.715	0.713	0.360
	−0.25	0.75	0.967	0.965	0.964	0.652
0.75	0.50	0.25	0.312	0.298	0.295	0.160
	0.25	0.50	0.797	0.785	0.781	0.411
	0	0.75	0.979	0.976	0.975	0.673

The second experiment corresponds to an alternative for which the data-driven optimal test is specially designed. The alternative models have the following form:

$$Y_t = \rho Y_{t-1} + \frac{\rho}{\tau} f(Y_{t-1}/\tau) + \epsilon_t, \tag{4.1}$$

where $f(y/\tau) = (1/\sqrt{2\pi\sigma^2}) \times \exp(-(1/2\sigma^2)(y/\tau)^2)$, $\sigma^2 = 1/(1-\rho^2)$, and ϵ_t is i.i.d. $N(0,1)$. Figure 1 shows the function $f(\cdot)$ for $\tau = 1$, 0.50, and 0.25, $\rho = 0.50$, and values of Y_t between -10 and 10. The function $f(\cdot)$ is symmetric around zero and more concentrated for smaller values of τ. The function is bounded between zero and one, with $f(0) = 1$ and $\lim_{x\to\pm\infty} f(x) = 0$. We can easily show that the alternative (4.1) respects the drift condition of Fan and Yao (2003, Thm. 2.4) for geometric ergodicity. This alternative is then compatible with the assumptions in this paper.

We examine the sensitivity of the tests to the narrowness of the peak and temporal dependence. We consider the parameter values $\tau = 25$, 0.50, 0.75 and $\rho = 0.25$, 0.50, 0.75. Table 3 shows the results of the experiment. For $\tau = 1$,

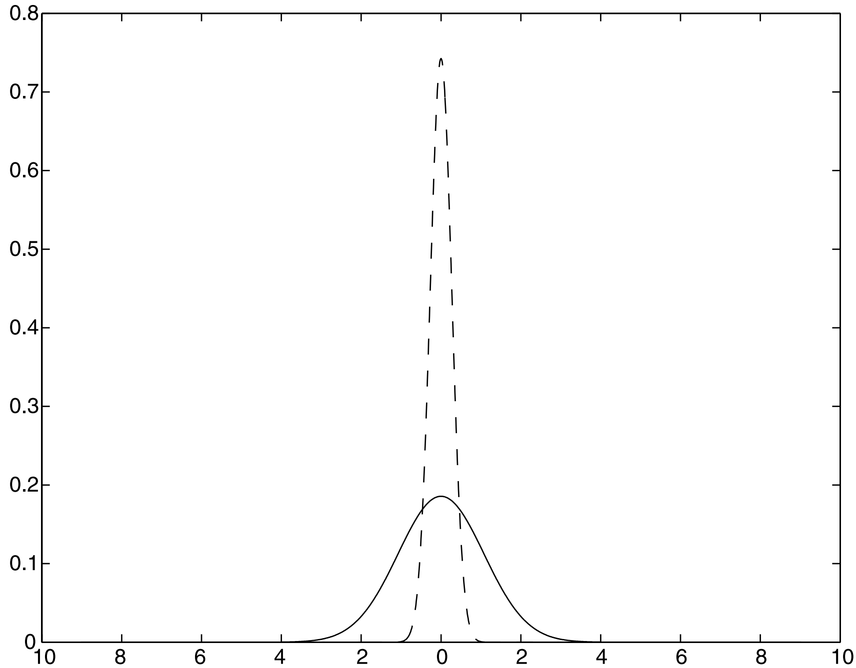

FIGURE 1. Alternative model ($\rho = 0.50$). Dashed line, $\tau = 0.25$; thick line, $\tau = 0.50$; and solid line, $\tau = 1$.

https://doi.org/10.1017/S0266466606060282 Published online by Cambridge University Press

TABLE 3. Power properties (5%) of our test and Hamilton test (LM): Second experiment (200 observations)

		c			
τ	ρ	2	3	5	LM
1	0.25	0.168	0.161	0.161	0.056
	0.50	0.426	0.421	0.420	0.072
	0.75	0.564	0.555	0.553	0.105
0.50	0.25	0.245	0.233	0.231	0.080
	0.50	0.639	0.605	0.595	0.213
	0.75	0.758	0.716	0.699	0.477
0.25	0.25	0.301	0.263	0.254	0.278
	0.50	0.751	0.664	0.622	0.716
	0.75	0.857	0.764	0.702	0.776

Hamilton's test is close to the nominal size. For 200 observations, our test rejects at a rate of 17% for $\rho = 0.25$ and 56% for $\rho = 0.75$. For $\tau = 0.50$, our test also clearly dominates the test proposed by Hamilton for all cases. For a narrow peak ($\tau = 0.25$), the rejection rate of both tests is quite similar. The better performance of the Hamilton test for this alternative compared to the one with a wider peak is probably due to the specification of the variance-covariance function of the random field underlying the test statistic. See Hamilton (2001) for further details on the construction of this test.

5. CONCLUDING REMARKS

This paper proposes a new adaptive rate-optimal specification test for time series. As in the maximum approach of Fan (1996) or Horowitz and Spokoiny (2001), the test combines several statistics to achieve adaptive rate-optimality. More specifically, the test builds on series regression chi-square statistics with increasing orders. A data-driven selection procedure, in the spirit of Guerre and Lavergne (2005), uses a penalty term proportional to the square root of the number of Fourier coefficients to choose the test statistic. Under the null, the retained statistic is, with high probability, a statistic with a distribution close to a chi-square. Therefore, standard chi-square critical values can be used, allowing for better control of the size of the test. This contrasts with the maximum approach, where using a null limit distribution performs poorly, as noted in Fan (1996), or is out of reach, as in Horowitz and Spokoiny (2001). Hence, the maximum approach necessitates the use of simulated critical values, limiting the scope of applications to time series models that can be easily simulated. A simulation experiment confirms the good level properties of the proposed test, which shows interesting power improvements compared to a simpler test using a single statistic such as that of Hamilton (2001). We also examine the power of the test that is adaptive rate-optimal and detects local alternatives approaching the null at a faster rate than in Horowitz and Spokoiny (2001). The simulation experiment shows that the choice of the penalty term has a moderate impact on the power. This positively illustrates the interest of our approach, which builds on the fact that the combination mechanism inherent to adaptive testing can also be designed to achieve a level close to the nominal size.

Although our results are stated for Fourier series methods, our approach also applies to wavelets or polynomial series regression. As noted in Guerre and Lavergne (2005), the series construction of the test statistic easily can be modified to cope with additive alternatives that are not affected by the curse of dimensionality. Obtaining an accurate size in the case of kernel or local polynomial methods is theoretically feasible. The scope of applications of the new data-driven selection procedure can also be extended as discussed in Hart (1997) for earlier adaptive procedures or as in Tjøstheim (1994) and Fan and Yao (2003) in the time series context, in addition to many other specification hypotheses of econometric interest.

6. PROOFS OF MAIN RESULTS

The proofs are organized as follows. Important intermediate results and proofs of the main statements are given in Section 6. Proofs of auxiliary results are gathered in Appendixes A and B. We now introduce some notation and conventions. All functions can be set to 0 outside Λ without loss of generality. We set $\sum_{t=0}^{-1} = \sum_{t=T+1}^{T} = 0$. The symbol $a_T \asymp b_T$ means that the two sequences a_T, b_T with the same sign are such that $c|a_T| \leq |b_T| \leq C|a_T|$ for some $0 < c \leq C < \infty$ and $T \geq 1$. Constants are denoted by the generic letter C and vary from expression to expression.

For notational convenience, we reindex the trigonometric functions (2.1) as $\{\psi_k(\cdot)\}_{k\in\mathbb{N}^*}$ and set $c_K = \kappa$. We assume that the new ordering is such that $\Psi_K = [\psi_1, \ldots, \psi_\kappa]$ and uses the notation Ψ_κ for Ψ_K. Here ψ_k, $k \in \mathbb{N}^*$, is a column vector with $\psi_k = [\psi_k(X_1), \ldots, \psi_k(X_T)]' \in \mathbb{R}^T$. Therefore Ψ_κ is a $T \times \kappa$ matrix and $\kappa \asymp K^d$. With little abuse of notation, $\mathcal{K}_T$ denotes both the set of admissible K or κ with κ between $\kappa_{\min} \asymp 2^{J_{\min} d}$ and $\kappa_{\max} \asymp 2^{J_{\max} d}$. The term $\hat{\kappa}^\gamma$ corresponds to $\hat{K}^\gamma$. The variance estimation rate in Assumption V is such that $v_T = o(\kappa_{\max}^{-3/2}/\ln T)$.

Let $\|\cdot\|$ be the euclidean norm of $\mathbb{R}^T$ or $\mathbb{R}^\kappa$, that is, if $u = [u_1, \ldots, u_\kappa]' \in \mathbb{R}^\kappa$, $\|u\| = (\sum_{k=1}^{\kappa} u_k^2)^{1/2} = (u'u)^{1/2}$. If $m = [m(X_1), \ldots, m(X_T)]'$ where $m(\cdot)$ maps $\mathbb{R}^d$ to $\mathbb{R}$, $\|m\| \leq T^{1/2} \sup_{x\in\mathbb{R}^d}|m(x)|$. Under Assumption E, $\|\varepsilon\| = O_{\mathbb{P}}(T^{1/2})$. For a $\kappa \times \kappa$ matrix $\Sigma = [\Sigma_{k\ell}]_{1\leq k,\ell\leq\kappa}$, $\|\Sigma\|$ is the spectral radius $\|\Sigma\| = \sup_{u\neq 0\in\mathbb{R}^\kappa}\|\Sigma u\|/\|u\|$. Recall that $\|\Sigma u\| \leq \|\Sigma\|\|u\|$, $|u_1'\Sigma u_2| \leq \|\Sigma\|\|u_1|\|u_2\|$. It follows that the entries of Σu are bounded by $\kappa^{1/2}\|\Sigma\|\max_{1\le k\le\kappa}|u_k|$. If Σ is a symmetric matrix, $\|\Sigma\| = \sup_{\|u\|=1}|u'\Sigma u|$ is the largest eigenvalue in absolute value of Σ. Because $\hat{\Omega}^{-1/2}\Psi_\kappa(\Psi_\kappa'\hat{\Omega}^{-1}\Psi_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1/2}$ is the orthogonal projection on the space spanned by the columns of $\hat{\Omega}^{-1/2}\Psi_\kappa$, we have

$$|u'\hat{\Omega}^{-1}\Psi_\kappa(\Psi_\kappa'\hat{\Omega}^{-1}\Psi_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}u| \le \|\hat{\Omega}^{-1/2}u\|^2 \text{ and}$$

$$\|\hat{\Omega}^{-1/2}\Psi_\kappa(\Psi_\kappa'\hat{\Omega}^{-1}\Psi_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1/2}\| \le 1.$$

In what follows, we bound variance of sums using the Wolkonski–Rozanov inequality (see Fan and Yao, 2003, Prop. 2.5(ii)), which states that

$$|\mathrm{Cov}(g_1(X_t), g_2(X_{t-n}))| \le 4\alpha(n)\sup_{x\in\mathbb{R}^d}|g_1(x)|\sup_{x\in\mathbb{R}^d}|g_2(x)|$$

for any real-valued bounded $g_1(\cdot)$ and $g_2(\cdot)$. This gives

$$\begin{aligned}\mathrm{Var}\left(\frac{1}{T}\sum_{t=1}^{T} g(X_t)\right) &= \frac{1}{T}\left(\mathrm{Var}(g(X_1)) + 2\sum_{n=1}^{T}\frac{T-n}{T}\mathrm{Cov}(g(X_1), g(X_{n-1}))\right)\\ &\le \frac{8}{T}\sup_{x\in\mathbb{R}^d}|g(x)|^2\sum_{n=0}^{\infty}\alpha(n).\end{aligned} \tag{6.1}$$

6.1. Estimation Errors

We consider first the parametric and variance estimation errors induced by $\hat{\theta}_T - \theta_T$ and $\hat{\sigma}(\cdot) - \sigma(\cdot)$, respectively. For $\Delta_T(\cdot) = \mu_T(\cdot) - m(x;\theta_T)$, set $U = \Delta_T + \varepsilon$ and let $\Omega^{1/2}$ be the $T\times T$ diagonal matrix with entries $\sigma(X_t)$. Set

$$\Sigma_\kappa = \mathbb{E}\left[\frac{\Psi_\kappa'(X_t)\Psi_\kappa(X_t)}{\sigma^2(X_t)}\right] = \left[\mathbb{E}\left(\frac{\psi_k(X_t)\psi_\ell(X_t)}{\sigma^2(X_t)}\right)\right]_{1\le k,\ell\le\kappa},$$

$$\hat{\Sigma}_\kappa = \hat{\Sigma}_\kappa(\hat{\Omega}) = \left[\overline{\frac{\psi_k(X)\psi_\ell(X)}{\hat{\sigma}^2(X)}}\right]_{1\le k,\ell\le\kappa}, \tag{6.2}$$

where

$$\overline{\frac{\psi_k(X)\psi_\ell(X)}{\hat{\sigma}^2(X)}} = \frac{1}{T}\sum_{t=1}^{T}\frac{\psi_k(X_t)\psi_\ell(X_t)}{\hat{\sigma}^2(X_t)},$$

$$\text{so that } T\hat{\Sigma}_\kappa = \Psi_\kappa'\hat{\Omega}^{-1}\Psi_\kappa \quad\text{and}\quad \hat{R}_\kappa = U'\hat{\Omega}^{-1}\Psi_\kappa(T\hat{\Sigma}_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}U.$$

PROPOSITION 1. *Consider a departure from the null such that* $\sup_{x\in\Lambda}|\Delta_T(x)| = O[\mathbb{E}^{1/2}(\Delta_T(X_t)/\sigma(X_t))^2]$. *Under Assumptions E, M, V, and X, and if* $\kappa_{\min}\to\infty$, $\kappa_{\max} = O(T^{1/3}/\ln^2 T)$, *we have*

$$\max_{\kappa \in \mathcal{K}_T} \frac{|\hat{R}_\kappa - \varepsilon'\Omega^{-1}\Psi_\kappa(T\Sigma_\kappa)^{-1}\Psi'_\kappa\Omega^{-1}\varepsilon - 2\Delta_T\hat{\Omega}^{-1}\Psi_\kappa(T\hat{\Sigma}_\kappa)^{-1}\Psi'_\kappa\hat{\Omega}^{-1}\varepsilon - \Delta_T\hat{\Omega}^{-1}\Psi_\kappa(T\hat{\Sigma}_\kappa)^{-1}\Psi'_\kappa\hat{\Omega}^{-1}\Delta_T|}{\kappa^{1/2}}$$

$$= O_{\mathbb{P}}\left(\frac{T^{1/2}}{\kappa_{\min}^{1/2}}\mathbb{E}^{1/2}\left(\frac{\Delta_T(X_t)}{\sigma(X_t)}\right)^2\right).$$

Proof of Proposition 1. See Appendix A.

6.2. Proof of Theorem 1

The next proposition is the key tool to establish Theorem 1.

PROPOSITION 2. *Assume that H_0 holds, that is, $\Delta_T(\cdot) = 0$. Then under Assumptions E, K, M, V, and X, make the following assumptions.*

(i) Let $\chi(\kappa)$ be a chi-square variable with κ degree of freedom. Then, for any $\kappa = \kappa_T$ in $\mathcal{K}_T$,

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}\left(\frac{\varepsilon'\Omega^{-1}\Psi_\kappa(T\Sigma_\kappa)^{-1}\Psi'_\kappa\Omega^{-1}\varepsilon - \kappa}{\sqrt{2\kappa}} \geq z\right) - \mathbb{P}\left(\frac{\chi(\kappa) - \kappa}{\sqrt{2\kappa}} \geq z\right)\right| = o(1)$$

and

$$\sup_{z\in\mathbb{R}}\left|\mathbb{P}\left(\frac{\hat{R}_{\kappa_{\min}} - \kappa_{\min}}{\sqrt{2\kappa_{\min}}} \geq z\right) - \mathbb{P}\left(\frac{\chi(\kappa_{\min}) - \kappa_{\min}}{\sqrt{2\kappa_{\min}}} \geq z\right)\right| = o(1).$$

(ii) Assume that (3.4) holds, that is, that for some $\epsilon > 0$, $\gamma_T \geq (1+\epsilon)\sqrt{2\ln \operatorname{Card}\mathcal{K}_T}$. Then

$$\mathbb{P}\left(\max_{\kappa\in\mathcal{K}_T\setminus\{\kappa_{\min}\}} \frac{\hat{R}_\kappa - \hat{R}_{\kappa_{\min}} - (\kappa - \kappa_{\min})}{\sqrt{2(\kappa-\kappa_{\min})}} \geq \gamma_T\right) = o(1).$$

Proof of Proposition 2. See Appendix A.

Proof of Theorem 1. Equation (3.2) and Proposition 2(ii) yield

$$\mathbb{P}(\hat{R}^\gamma \neq \hat{R}_{\kappa_{\min}}) \leq \mathbb{P}(\hat{\kappa}^\gamma \neq \kappa_{\min})$$

$$= \mathbb{P}\left(\max_{\kappa\in\mathcal{K}_T\setminus\{\kappa_{\min}\}} \frac{\hat{R}_\kappa - \hat{R}_{\kappa_{\min}} - (\kappa - \kappa_{\min})}{\sqrt{2(\kappa-\kappa_{\min})}} \geq \gamma_T\right) = o(1).$$

Then the definition of z_α in (2.6) and Proposition 2(i) yield

$$\mathbb{P}\left(\frac{\hat{R}^{\gamma}-\kappa_{\min}}{\sqrt{2\kappa_{\min}}}\geq z_\alpha\right)=\mathbb{P}\left(\frac{\hat{R}_{\kappa_{\min}}-\kappa_{\min}}{\sqrt{2\kappa_{\min}}}\geq z_\alpha\right)+o(1)$$

$$=\mathbb{P}\left(\frac{\chi(\kappa_{\min})-\kappa_{\min}}{\sqrt{2\kappa_{\min}}}\geq z_\alpha\right)+o(1)\to\alpha. \quad \blacksquare$$

6.3. Proof of Theorems 2 and 3

The next lemma is crucial for the consistency properties of the test and is used for the item $\hat{R}_{1K}$ in (2.3).

LEMMA 1. *Consider a departure from the null such that* $\sup_{x\in\Lambda}|\Delta_T(x)| = O[\mathbb{E}^{1/2}(\Delta_T(X_t)/\sigma(X_t))^2]$. *Assume that Assumptions E, V, and X hold and that* $\kappa=\kappa_T$ *diverges with* $\kappa=o(T^{1/3}/\ln^2 T)$.

Then there exists a constant $C_5>0$, *depending upon s, L, and* Λ, *such that for any* $\kappa\in\mathcal{K}_T$, *any* $\Delta(\cdot)$ *from* Λ *to* $\mathbb{R}$ *in* $\mathcal{C}(L,s)$, *we have*

$$[\Delta_T'\hat{\Omega}^{-1}\Psi_\kappa(\Psi_\kappa'\hat{\Omega}^{-1}\Psi_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}\Delta_T]^{1/2}$$
$$\geq T^{1/2}\left[\mathbb{E}^{1/2}\left(\frac{\Delta_T(X_t)}{\sigma(X_t)}\right)^2-C_5\kappa^{-s/d}\right](1+o_{\mathbb{P}}(1)), \tag{6.3}$$

$$|\Delta_T'\hat{\Omega}^{-1}\Psi_\kappa(\Psi_\kappa'\hat{\Omega}^{-1}\Psi_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}\varepsilon|$$
$$=T^{1/2}O_{\mathbb{P}}\left[\mathbb{E}^{1/2}\left(\frac{\Delta_T(X_t)}{\sigma(X_t)}\right)^2+\kappa^{-s/d}\right]. \tag{6.4}$$

Proof of Lemma 1. See Appendix A.

Proof of Theorem 2. Let $s\leq d(2/C_1-1)$ and L be some unknown smoothness indexes. Let K_* be as in (3.6), so that K_* corresponds to a κ_* in the new indexation. Observe that this κ_* is such that

$$T\kappa_*^{-2s/d}\asymp T\rho_T^2=(\sqrt{\ln\ln T})^{4s/(4s+d)}T^{d/(4s+d)}\asymp\gamma_T\kappa_*^{1/2}\asymp\gamma_T\sqrt{2(\kappa_*-\kappa_{\min})}, \tag{6.5}$$

because the exact order of γ_T is $\ln^{1/2}\ln T$, $s>0$, and $\kappa_{\min}$ is smaller than a power of $\ln T$.

Consider now a sequence of alternatives $\mu_T(\cdot)$ in $H_1(C_3.\rho_T)$ with $C_3\rho_T> 2C_5\kappa_*^{-s/d}$, where C_5 is from Lemma 1. This gives that $\mathbb{E}^{1/2}(\Delta_T(X_t)/\sigma(X_t))^2 - C_5\kappa_*^{-s/d}\geq\frac{1}{2}\mathbb{E}^{1/2}(\Delta_T(X_t)/\sigma(X_t))^2$ and that $T\mathbb{E}(\Delta_T(X_t)/\sigma(X_t))^2$ diverges. Hence Lemma 1 gives

$$\Delta_T'\hat{\Omega}^{-1}\Psi_{\kappa_*}(\Psi_{\kappa_*}'\hat{\Omega}^{-1}\Psi_{\kappa_*})^{-1}\Psi_{\kappa_*}'\hat{\Omega}^{-1}\Delta_T$$

$$\geq \left(\frac{1}{4}+o_{\mathbb{P}}(1)\right)T\mathbb{E}\left(\frac{\Delta_T(X_t)\mathbb{I}(X_t\in\Lambda)}{\sigma(X_t)}\right)^2,$$

$$\Delta_T'\hat{\Omega}^{-1}\Psi_{\kappa_*}(\Psi_{\kappa_*}'\hat{\Omega}^{-1}\Psi_{\kappa_*})^{-1}\Psi_{\kappa_*}'\hat{\Omega}^{-1}\varepsilon$$

$$= O_{\mathbb{P}}\left[T^{1/2}\mathbb{E}^{1/2}\left(\frac{\Delta_T(X_t)\mathbb{I}(X_t\in\Lambda)}{\sigma(X_t)}\right)^2\right]$$

$$= o_{\mathbb{P}}(1)\Delta_T'\hat{\Omega}^{-1}\Psi_{\kappa_*}(\Psi_{\kappa_*}'\hat{\Omega}^{-1}\Psi_{\kappa_*})^{-1}\Psi_{\kappa_*}'\hat{\Omega}^{-1}\Delta_T.$$

Observe also that Proposition 2(i) shows that

$$\varepsilon'\hat{\Omega}^{-1}\Psi_{\kappa_*}(\Psi_{\kappa_*}'\hat{\Omega}^{-1}\Psi_{\kappa_*})^{-1}\Psi_{\kappa_*}'\hat{\Omega}^{-1}\varepsilon-\kappa_T$$

$$= O_{\mathbb{P}}(\kappa_*^{1/2}) = o_{\mathbb{P}}(1)(\gamma_T\sqrt{2(\kappa_*-\kappa_{\min})})$$

$$= o_{\mathbb{P}}(1)\Delta_T'\hat{\Omega}^{-1}\Psi_{\kappa_*}(\Psi_{\kappa_*}'\hat{\Omega}^{-1}\Psi_{\kappa_*})^{-1}\Psi_{\kappa_*}'\hat{\Omega}^{-1}\Delta_T.$$

Hence, (6.5), applying Proposition 1 for $\mathcal{K}_T=\{\kappa_*\}$ (so that $\kappa_{\max}=\kappa_{\min}=\kappa_T$), and substituting yield

$$\hat{R}_{\kappa_*}-\kappa_*-\gamma_T\sqrt{2(\kappa_T-\kappa_{\min})}-z_\alpha\sqrt{2\kappa_{\min}}$$

$$= [\Delta_T'\hat{\Omega}^{-1}\Psi_{\kappa_*}(\Psi_{\kappa_*}'\hat{\Omega}^{-1}\Psi_{\kappa_*})^{-1}\Psi_{\kappa_*}'\hat{\Omega}^{-1}\Delta_T-\gamma_T\sqrt{2(\kappa_T-\kappa_{\min})}](1+o_{\mathbb{P}}(1))$$

$$\geq T\left[\frac{1}{4}\mathbb{E}\left(\frac{\Delta_T(X_t)\mathbb{I}(X_t\in\Lambda)}{\sigma(X_t)}\right)^2-C\rho_T^2\right](1+o_{\mathbb{P}}(1))$$

$$\geq T\rho_T^2\left(\frac{C_3^2}{4}-C\right)(1+o_{\mathbb{P}}(1))\xrightarrow{\mathbb{P}}+\infty$$

provided C_3 is large enough. The lower power bound (2.8) then shows that Theorem 2 is proved. ∎

Proof of Theorem 3. Because the proof of Theorem 3 is similar to the proof of Theorem 2 up to the fact that detection is achieved through $\kappa_{\min}$, we just give the main steps. Expression (2.7) yields that $\hat{R}^{\gamma}-\kappa_{\min}\geq\hat{R}_{\kappa_{\min}}-\kappa_{\min}$, so that it is sufficient to show that $\hat{R}_{\kappa_{\min}}-\kappa_{\min}-\sqrt{2\kappa_{\min}}$ diverges to $+\infty$ in probability. Building on Propositions 1 and 2(i) and Lemma 1 as for Theorem 2 now gives, because $\kappa_{\min}\asymp K_{\min}^d\to\infty$,

https://doi.org/10.1017/S0266466606060282 Published online by Cambridge University Press

$$
\begin{aligned}
&\hat{R}_{\kappa_{\min}} - \kappa_{\min} - \sqrt{2\kappa_{\min}} \\
&\quad \geq r_T^2 \Delta'_{0T} \hat{\Omega}^{-1} \Psi_{\kappa_T} (\Psi'_{\kappa_T} \hat{\Omega}^{-1} \Psi_{\kappa_T})^{-1} \Psi'_{\kappa_T} \hat{\Omega}^{-1} \Delta_{0T} (1 + o_{\mathbb{P}}(1)) - O_{\mathbb{P}}(\kappa_{\min}^{1/2}) \\
&\quad \geq Tr_T^2 \left[\mathbb{E}\left(\frac{\Delta_{0T}(X_t)\mathbb{I}(X_t \in \Lambda)}{\sigma(X_t)} \right)^2 - C_5 \kappa_{\min}^{-s/d} \right] (1 + o_{\mathbb{P}}(1)) - O_{\mathbb{P}}(\kappa_{\min}^{1/2}) \\
&\quad = Tr_T^2 - O_{\mathbb{P}}(K_{\min}^{d/2}) \xrightarrow{\mathbb{P}} +\infty
\end{aligned}
$$

provided Tr_T^2 diverges with $\lim_{T\to\infty} K_{\min}^{d/2}/(Tr_T^2) = 0$ as assumed in Theorem 3. ■

NOTES

1. Using other series approximation methods, as, for instance, polynomial functions or wavelets, is possible but leads to a more involved theoretical study. Indeed, the Fourier system satisfies $\sup_{k\in\mathbb{Z}^d}\sup_{x\in\Lambda}|\psi_k(x)| < \infty$, a condition that simplifies algebraic manipulations under dependence mixing conditions. Another interest of Fourier methods is that using wavelets may limit the scope of applications to alternatives with a maximal smoothness given by the choice of the wavelet basis; see the wavelet tests considered in Spokoiny (1996) and Theorem 2.4 therein.

2. Assume that H_0 is $\mu(\cdot) = 0$ and that $\sigma(\cdot)$ is known so that $\hat{\Delta}(\cdot) = \mu(\cdot)$ and the choice $\hat{\sigma}(\cdot) = \sigma(\cdot)$ is possible. In the case of Gaussian i.i.d. ε_t independent of the X_t's, $\hat{R}_{2K}$ would be an $\mathcal{N}(0, \hat{R}_{1K}) = O_{\mathbb{P}}(\hat{R}_{1K}^{1/2})$, which can be neglected with respect to $\hat{R}_{1K}$ when this variable diverges. Note also that the distribution of $\hat{R}_{3K}$ coincides with its chi-square approximation for such $\hat{\Delta}(\cdot)$, $\hat{\sigma}(\cdot)$, and ε.

3. Note that $\hat{R}_K - c_K$ is a better misspecification indicator than $\hat{R}_K$, which is affected by an additional systematic bias term c_K. Guerre and Lavergne (2005) proposed a different bias correction that makes asymptotic inference less accurate in finite sample, so that the bootstrap is used.

4. This continues to hold in the dependent setup where the bound (B.9) in Appendix B gives a more complicated error term, which is $K^{2d}/T^{1/2}$ at best. A normal approximation would be affected with a bigger $K^{2d}/T^{1/2} + K^{-d/2}$ error term.

5. A second distinctive feature of the selection procedure (2.5) is standardization with $c_{K_{\min}}$ in the critical region $\{\hat{R}^\gamma \geq c_{K_{\min}} + z_\alpha\sqrt{2c_{K_{\min}}}\}$; see (2.6). Because $\hat{K}^\gamma = K_{\min}$ asymptotically, an alternative α-level critical region would use $c_{\hat{K}^\gamma}$ in place of $c_{K_{\min}}$. But such a choice would asymptotically reduce power because $c_{\hat{K}^\gamma} + z_\alpha\sqrt{2c_{\hat{K}^\gamma}} \geq c_{K_{\min}} + z_\alpha\sqrt{2c_{K_{\min}}}$. This also contrasts with a maximum procedure that would use the test statistic $(\hat{R}_{\hat{K}^*} - c_{\hat{K}^*})/\sqrt{2c_{\hat{K}^*}} = \max_{K\in\mathcal{K}_T}(\hat{R}_K - c_K)/\sqrt{2c_K}$ with a $c_{\hat{K}^*}$ larger than $c_{K_{\min}}$. The simulation experiments of Guerre and Lavergne (2005) revealed that such a construction of the critical region (2.6) gives a test that improves on its adaptive rate-optimal competitors.

6. Spokoiny (1996) studied the continuous time white noise model (CTWN) $Y_n(t) = m(t)\,dt + (\sigma/\sqrt{n})\,dW(t)$, $t \in [0,1]$, where $\{W(t)\}_{t\in[0,1]}$ is a standard Brownian motion. Although this model is mainly of theoretical interest, results established for the CTWN model extend to more common models through model equivalence; see Brown and Low (1996).

7. Results for the normal distribution are only reported here because the results for the two other distributions are very similar. Of course, those results can be obtained upon request.

REFERENCES

Aït-Sahalia, Y. (1996) Testing continuous-time models of the spot interest rate. *Review of Financial Studies* 9, 385–426.

Aït-Sahalia, Y., P.J. Bickel, & T.M. Stocker (2001) Goodness of fit tests for kernel regression with an application to option-implied volatility. *Journal of Econometrics* 105, 363–412.

Baraud, Y., S. Huet, & B. Laurent (2003) Adaptive tests of linear hypotheses by model selection. *Annals of Statistics* 31, 225–251.

Bierens, H.J. (1984) Model specification testing of time series regressions. *Journal of Econometrics* 26, 323–353.

Billingsley, P. (1968) *Convergence of Probability Measures*. Wiley.

Brown, L.D. & M.G. Low (1996) Asymptotic equivalence of nonparametric regression and white noise. *Annals of Statistics* 24, 2384–2398.

Chen, X. & Y. Fan (1999) Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series. *Journal of Econometrics* 91, 373–401.

Chow, S.C. & H. Teicher (1988) *Probability Theory: Independence, Interchangeability, Martingales*. Springer.

Domowitz, I. & H. White (1982) Misspecified models with dependent observations. *Journal of Econometrics* 20, 35–58.

Fan, J. (1996) Test of significance based on wavelet thresholding and Neyman's truncation. *Journal of the American Statistical Association* 91, 674–688.

Fan, J. & L.S. Huang (2001) Goodness-of-fits test for parametric regression models. *Journal of the American Statistical Association* 96, 640–652.

Fan, J. & Q. Yao (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer.

Fan, J., C. Zhang, & J. Zhang (2001) Generalized likelihood ratio statistics and Wilks phenomenon. *Annals of Statistics* 29, 153–193.

Gao, J. & M. King (2001) Estimation and Model Specification Testing in Nonparametric and Semiparametric Regression. Working paper, School of Mathematics and Statistics, the University of Western Australia.

Gao, J. & M. King (2004) Adaptive testing in continuous-time diffusion models. *Econometric Theory* 20, 844–883.

Gayraud, G. & C. Pouet (2005) Adaptive minimax testing in the discrete regression scheme. *Probability Theory and Related Fields* 133, 531–58.

Gozalo, P.L. (1997) Nonparametric bootstrap analysis with applications to demographic effects in demand functions. *Journal of Econometrics* 81, 357–393.

Guerre, E. & P. Lavergne (2002) Optimal minimax rates for nonparametric specification testing in regression models. *Econometric Theory* 18, 1139–1171.

Guerre, E. & P. Lavergne (2005) Data-driven rate-optimal specification testing in regression models. *Annals of Statistics* 33, 840–870.

Hamilton, J.D. (2001) A parametric approach to flexible nonlinear inference. *Econometrica* 69, 537–573.

Härdle, W. & E. Mammen (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926–1947.

Hart, J.D. (1997) *Nonparametric Smoothing and Lack-of-fit Tests*. Springer.

Hong, Y. & H. White (1995) Consistent specification testing via nonparametric series regression. *Econometrica* 63, 1133–1159.

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

Ingster, Y.I. (1992, 1993) Asymptotically minimax hypothesis testing for nonparametric alternatives, parts I, II, and III. *Mathematical Methods of Statistics* 2, 85–114, 171–189, and 249–268.

Poo, J.R., S. Sperlich, & P. Vieu (2004) An Adaptive Specification Test for Semiparametric Models. Manuscript, Universidad de Zaragoza.

Rio, E. (2000) *Théorie asymptotique des processus aléatoires faiblement dépendants*. Mathématiques and Applications 31. Springer.

Robinson, P.M. (1989) Hypothesis testing in semiparametric and nonparametric models for econometric time series. *Review of Economic Studies* 56, 511–534.

Shorack, G.R. & J.A. Wellner (1986) *Empirical Processes with Applications to Statistics*. Wiley.
Spokoiny, V.G. (1996) Adaptive hypothesis testing using wavelets. *Annals of Statistics* 24, 2477–2498.
Spokoiny, V.G. (2001) Data-driven testing the fit of linear models. *Mathematical Methods of Statistics* 10, 465–497.
Timan, A.F. (1994) *Theory of Approximation of Functions of a Real Variable*. Dover.
Tjøstheim, D. (1994) Non-linear time series: A selective review. *Scandinavian Journal of Statistics* 21, 97–130.

APPENDIX A: Proofs of Propositions 1 and 2 and Lemma 1

A.1. Preliminary Lemmas. We begin with the estimation errors $\hat{\Sigma}_k - \Sigma_k$ (see (6.2)) and preliminary bounds. Define

$$\tilde{\psi}_k(\cdot) = \psi_k(\cdot) - \Psi_{\kappa_{\min}}(\cdot)\Sigma_{\kappa_{\min}}^{-1}\mathbb{E}\left[\frac{\psi_k(X_t)\Psi'_{\kappa_{\min}}(X_t)}{\sigma^2(X_t)}\right],$$

$$\tilde{\Sigma}_\kappa = \left[\mathbb{E}\left(\frac{\tilde{\psi}_k(X_t)\tilde{\psi}_\ell(X_t)}{\sigma^2(X_t)}\right)\right]_{\kappa_{\min}<k,\,\ell\le\kappa}, \tag{A.1}$$

which are used to study the difference $\hat{R}_\kappa - \hat{R}_{\kappa_{\min}}$ in the proof of Proposition 2(ii). The next lemmas hold for general orthonormal systems $\{\psi_k(\cdot)\}_{k\in\mathbb{N}^*}$ of $L^2(\Lambda, dx)$ with $\sup_{k\in\mathbb{N}^*}\sup_{x\in\Lambda}|\psi_k(x)| < \infty$. Recall that v_T is such that $\sup_{x\in\Lambda}|\hat{\sigma}(x) - \sigma(x)| = O_{\mathbb{P}}(v_T)$ with $v_T = o(\kappa_{\max}^{-3/2}/\ln T)$; see Assumption V.

LEMMA A.1. *Let Σ_κ, $\hat{\Sigma}_\kappa$ be as in (6.2) and $\{\tilde{\psi}_k(\cdot)\}_{k>\kappa_{\min}}$, $\tilde{\Sigma}_\kappa$ as in (A.1). Then, under Assumptions E, V, and X,*

(i) $\sup_{\kappa\in\mathbb{N}^*}\max(\|\Sigma_\kappa^{-1}\|, \|\Sigma_\kappa\|) < \infty$, $\sup_{k>\kappa_{\min}}\sup_{x\in\Lambda}|\tilde{\psi}_k(x)| < C\kappa_{\min}^{1/2}$ *and* $\sup_{\kappa\in\mathbb{N}^*}\max(\|\tilde{\Sigma}_\kappa^{-1}\|, \|\tilde{\Sigma}_\kappa\|) < \infty$.

(ii) If $\kappa_{\max} = o(T^{1/2})$, the matrices $\hat{\Sigma}_\kappa$, $1 \le \kappa \le \kappa_{\max}$, have an inverse with a probability tending to 1 and

$$\max_{\kappa\in\mathcal{K}_T}\max[\|\hat{\Sigma}_\kappa - \Sigma_\kappa\|, \|\hat{\Sigma}_\kappa^{-1} - \Sigma_\kappa^{-1}\|] = O_{\mathbb{P}}\left[\left(\frac{\kappa_{\max}^2}{T}\right)^{1/2} + \kappa_{\max}v_T\right] = o_{\mathbb{P}}(1).$$

(iii) If $\kappa_{\max} = o(T^{1/2})$, $\max_{1\le\kappa\le\kappa_{\max}}\max(\|\hat{\Sigma}_\kappa\|, \|\hat{\Sigma}_\kappa^{-1}\|) = O_{\mathbb{P}}(1)$.

https://doi.org/10.1017/S0266466606060282 Published online by Cambridge University Press

LEMMA A.2. *Let* $m_T(\cdot)$ *and* $\mu_T(\cdot)$ *from* $\mathbb{R}^d$ *to* $\mathbb{R}$ *be some functions with support* Λ. *Then, under Assumptions E, V, and X and if* Card $\mathcal{K}_T = O(\ln T)$, $\kappa_{\max} = o(T^{1/3}/\ln^{2/3} T)$,

$$\max_{\kappa\in\mathcal{K}_T} |\mu_T'\hat{\Omega}^{-1}\Psi_\kappa(\Psi_\kappa'\hat{\Omega}^{-1}\Psi_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}m_T| \le \|\hat{\Omega}^{-1/2}\mu_T\|\,\|\hat{\Omega}^{-1/2}m_T\|, \tag{A.2}$$

$$\max_{\kappa\in\mathcal{K}_T} \frac{\|\Psi_\kappa'\hat{\Omega}^{-1}\varepsilon\|}{\sqrt{T\kappa}} = \text{Card}^{1/2}\mathcal{K}_T O_{\mathbb{P}}(1), \tag{A.3}$$

$$\max_{\kappa\in\mathcal{K}_T} \frac{|\varepsilon'\hat{\Omega}^{-1}\Psi_\kappa(\Psi_\kappa'\hat{\Omega}^{-1}\Psi_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}\varepsilon - \varepsilon'\Omega^{-1}\Psi_\kappa(T\Sigma_\kappa)^{-1}\Psi_\kappa'\Omega^{-1}\varepsilon|}{\kappa^{1/2}}$$

$$= \text{Card}\,\mathcal{K}_T O_{\mathbb{P}}\left[\left(\frac{\kappa_{\max}^3}{T}\right)^{1/2} + \kappa_{\max}^{3/2} v_T\right] = o_{\mathbb{P}}(1), \tag{A.4}$$

$$\max_{\kappa\in\mathcal{K}_T} \frac{|m_T'\hat{\Omega}^{-1}\Psi_\kappa(\Psi_\kappa'\hat{\Omega}^{-1}\Psi_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}\varepsilon|}{\kappa^{1/2}}$$

$$= \frac{T^{1/2}}{\kappa_{\min}^{1/2}} O_{\mathbb{P}}\left[\mathbb{E}^{1/2}\left(\frac{m_T(X_t)}{\sigma(X_t)}\right)^2 + \inf_{\pi(\cdot)\in\Pi_{\kappa_{\min}}} \sup_{x\in\Lambda}|m_T(x) - \pi(x)| + v_T \sup_{x\in\Lambda}|m_T(x)|\right]. \tag{A.5}$$

The functions $m_T(\cdot)$ *and* $\mu_T(\cdot)$ *may depend upon* $(X_1,\varepsilon_1),\ldots,(X_T,\varepsilon_T)$ *in (A.2) but not in (A.5).*

Proofs of Lemmas A.1 and A.2. See Appendix B.

The next lemma is used for Proposition 2. It is stated for general maps $\varphi_k(\cdot)$ from $\mathbb{R}^d$ to $\mathbb{R}$, $k \ge 1$. Consider the row vector $\Phi_\kappa(X_t) = [\varphi_1(X_t),\ldots,\varphi_\kappa(X_t)]$ and the $\kappa \times T$ matrix $\Phi_\kappa = [\Phi_\kappa(X_1)',\ldots,\Phi_\kappa(X_T)']'$. Define

$$V_\kappa = \mathbb{E}\left[\frac{\Phi_\kappa'(X_t)\Phi_\kappa(X_t)}{\sigma(X_t)}\right].$$

We make the following assumption.

Assumption B. The matrices V_κ have an inverse with $\sup_{\kappa\in\mathbb{N}^*}\|V_\kappa^{-1}\| < \infty$, and the functions $\varphi_k(\cdot)$ are such that $\max(\sup_{1\le k\le\kappa}\sup_{x\in\mathbb{R}^d}|\varphi_k(x)|,1) = \varphi_\infty < \infty$.

Define

$$S_T = S_{\kappa T} = V_\kappa^{-1/2}\sum_{t=1}^T \frac{\Phi_\kappa'(X_t)}{\sigma(X_t)}\,\frac{\varepsilon_t}{\sigma(X_t)}, \qquad Q_T = Q_{\kappa T} = \frac{T^{-1}S_T'S_T - \kappa}{\sqrt{2\kappa}} = \frac{T^{-1}\|S_T\|^2 - \kappa}{\sqrt{2\kappa}}.$$

We now study the tail probability of Q_T.

LEMMA A.3. *Let $Q_T = Q_{\kappa T}$ be as before. Then, under Assumptions E, X(i), B, and $\kappa = \kappa_T = o(T^{(3/4)[(1+a)/(5+3a)]})$, make the following assumptions.*

(i) Let $\chi(\kappa)$ be a chi-square variable with κ degree of freedom. Then

$$\lim_{T\to\infty} \sup_{\gamma\in\mathbb{R}} \left| \mathbb{P}(Q_T \geq \gamma) - \mathbb{P}\left(\frac{\chi(\kappa)-\kappa}{\sqrt{2\kappa}} \geq \gamma \right) \right| = 0.$$

(ii) Consider $\epsilon > 0$. Then there exists a constant C_ϵ, which does not depend upon κ and γ, such that for any $\gamma > \epsilon$ and κ,

$$\mathbb{P}(Q_T \geq \gamma) \leq \frac{1}{\sqrt{2\pi}(\gamma-\epsilon)} \exp\left(-\frac{(\gamma-\epsilon)^2}{2}\right)$$

$$+ C_\epsilon \left[\varphi_\infty^6 \kappa^2 T^{-(3/2)[(1+a)/(5+3a)]} + \frac{1}{\sqrt{\kappa}} \right].$$

Proof of Lemma A.3. See Appendix B.

A.2. Proof of Propositions 1 and 2.

Proof of Proposition 1. For brevity of notation, the proof is made for $p = \dim\theta = 1$. Define

$$e(\theta) = [e_1(\theta),\ldots,e_T(\theta)]' \quad \text{where } e_t = m(X_t;\theta_T) - m(X_t;\hat\theta_T) \quad \text{so that } \hat U = U + e(\theta).$$

This gives

$$\hat R_\kappa = U'\hat\Omega^{-1}\Psi_\kappa(T\hat\Sigma_\kappa)^{-1}\Psi_\kappa'\hat\Omega^{-1}U + 2A_\kappa + B_\kappa$$

$$\text{with } \; A_\kappa = U'\hat\Omega^{-1}\Psi_\kappa(T\hat\Sigma_\kappa)^{-1}\Psi_\kappa'\hat\Omega^{-1}e(\theta)$$

$$\text{and } \; B_\kappa = e(\theta)'\hat\Omega^{-1}\Psi_\kappa(T\hat\Sigma_\kappa)^{-1}\Psi_\kappa'\hat\Omega^{-1}e(\theta).$$

Under Assumption M, $\max_{1\leq t\leq T}|e_t(\theta)| = O_{\mathbb{P}}(T^{-1/2})$, which gives $\|e(\theta)\| = O_{\mathbb{P}}(1)$ and $\max_{\kappa\in\mathcal{K}_T}|B_\kappa| = O_{\mathbb{P}}(1)$, so that $\max_{1\leq\kappa\leq\kappa_{\max}}\kappa^{-1/2}|B_\kappa| = O_{\mathbb{P}}(\kappa_{\min}^{-1/2})$. Consider now A_κ. Under Assumption M, the Taylor formula gives

$$e_t(\theta) = (\hat\theta_T - \theta_T)\frac{\partial m(X_t;\theta_T)}{\partial\theta} + \frac{1}{2}(\hat\theta_T-\theta_T)^2 \frac{\partial^2 m(X_t;\theta_{tT}^*)}{\partial^2\theta}$$

so that $e(\theta) = (\hat\theta_T - \theta_T)m_1 + \frac{1}{2}(\hat\theta_T-\theta_T)^2 m_2$, with a θ_{tT}^* between θ_T and $\hat{\theta}_T$ and where m_1 and m_2 are $\mathbb{R}^T$ column vectors with bounded entries given by the first- and second-order derivatives. Because $U = \Delta_T + \varepsilon$, this gives

$$A_\kappa = A_{1\kappa} + A_{2\kappa} + \frac{1}{2} A_{3\kappa} \quad \text{with } A_{1\kappa} = e'(\theta)\hat{\Omega}^{-1}\Psi_\kappa(T\hat{\Sigma}_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}\Delta_T,$$

$$A_{2\kappa} = (\hat{\theta}_T - \theta_T)m_1'\hat{\Omega}^{-1}\Psi_\kappa(T\hat{\Sigma}_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}\varepsilon,$$

$$A_{3\kappa} = (\hat{\theta}_T - \theta_T)^2 m_2'\hat{\Omega}^{-1}\Psi_\kappa(T\hat{\Sigma}_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}\varepsilon.$$

The Cauchy–Schwarz inequality gives $|A_{1\kappa}| \leq \|e(\theta)\| \|\Delta_T\|$ with $\|e(\theta)\| = O_{\mathbb{P}}(1)$, so that

$$\max_{\kappa \in \mathcal{K}_T} \frac{|A_{1\kappa}|}{\kappa^{1/2}} = O_{\mathbb{P}}\left(\frac{\|\Delta_T\|}{\kappa_{\min}^{1/2}}\right) = O_{\mathbb{P}}\left(\frac{T^{1/2}}{\kappa_{\min}^{1/2}} \mathbb{E}^{1/2}\left(\frac{\Delta_T(X_t)}{\sigma(X_t)}\right)^2\right),$$

because $\|\Delta_T\|^2 = O_{\mathbb{P}}(T)\mathbb{E}(\Delta_T(X_t)/\sigma(X_t))^2$ by the Markov inequality and Assumption E. Because $T^{1/2}(\hat{\theta}_T - \theta_T) = O_{\mathbb{P}}(1)$ and under Assumption M, applying (A.5) for $A_{2\kappa}$ and the Cauchy–Schwarz inequality for $A_{3\kappa}$ give

$$\max_{\kappa \in \mathcal{K}_T} \frac{|A_{2\kappa}|}{\kappa^{1/2}} = O_{\mathbb{P}}(\kappa_{\min}^{-1/2}),$$

$$\max_{\kappa \in \mathcal{K}_T} \frac{|A_{3\kappa}|}{\kappa^{1/2}} = O_{\mathbb{P}}\left(\frac{1}{T\kappa_{\min}^{1/2}}\right)\|m_2\| \|\varepsilon\| = O_{\mathbb{P}}(\kappa_{\min}^{-1/2}).$$

Substituting in the expression of A_κ and $\hat{R}_\kappa$ give

$$\max_{\kappa \in \mathcal{K}_T} \frac{|\hat{R}_\kappa - U'\hat{\Omega}^{-1}\Psi_\kappa(T\hat{\Sigma}_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}U|}{\kappa^{1/2}} = \frac{1}{\kappa_{\min}^{1/2}} O_{\mathbb{P}}\left[1 + T^{1/2}\mathbb{E}^{1/2}\left(\frac{\Delta_T(X_t)}{\sigma(X_t)}\right)^2\right]. \tag{A.6}$$

But

$$U'\hat{\Omega}^{-1}\Psi_\kappa(T\hat{\Sigma}_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}U = \varepsilon'\hat{\Omega}^{-1}\Psi_\kappa(T\hat{\Sigma}_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}\varepsilon$$
$$+ 2\Delta_T'\hat{\Omega}^{-1}\Psi_\kappa(T\hat{\Sigma}_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}\varepsilon$$
$$+ \Delta_T'\hat{\Omega}^{-1}\Psi_\kappa(T\hat{\Sigma}_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}\Delta_T$$

so that substituting (A.4) in the preceding equation and (A.6) give the desired result. ■

Proof of Proposition 2. Define

$$R_\kappa^0 = \varepsilon'\Omega^{-1}\Psi_\kappa(T\Sigma_\kappa)^{-1}\Psi_\kappa'\Omega^{-1}\varepsilon \quad \text{and} \quad Q_\kappa^0 = Q_{\kappa T}^0 = \frac{R_\kappa^0 - \kappa}{\sqrt{2\kappa}}.$$

Under the null, Proposition 1 yields

$$\max_{\kappa\in\mathcal{K}_T} \frac{|\hat{R}_\kappa - R_\kappa^0|}{\kappa^{1/2}} = o_{\mathbb{P}}(1) \quad \text{or, equivalently,} \quad \max_{\kappa\in\mathcal{K}_T} \left| \frac{\hat{R}_\kappa - \kappa}{\sqrt{2\kappa}} - Q_\kappa^0 \right| = o_{\mathbb{P}}(1). \tag{A.7}$$

Hence Proposition 2(i) follows from taking $\kappa = \kappa_{\min}$ in Lemma A.3(i) and (A.7). Consider now Proposition 2(ii). Let ϵ be as in (3.4), so that $\gamma_T \geq \sqrt{2 \ln \operatorname{Card} \mathcal{K}_T} + \epsilon$ for T large enough. Therefore (A.7) yields that Proposition 2(ii) is a consequence of

$$\mathbb{P}\left(\max_{\kappa\in\mathcal{K}_T\setminus\{\kappa_{\min}\}} \frac{R_\kappa^0 - R_{\kappa_{\min}}^0 - (\kappa - \kappa_{\min})}{\sqrt{2(\kappa - \kappa_{\min})}} \geq \sqrt{2 \ln \operatorname{Card} \mathcal{K}_T} + \epsilon \right) = o(1). \tag{A.8}$$

To prove (A.8), we first rewrite $R_\kappa^0 - R_{\kappa_{\min}}^0$ as a suitable quadratic form. For $k, \kappa > \kappa_{\min}$, let $\tilde{\psi}_k(\cdot)$ and $\tilde{\Sigma}_\kappa$ be as in (A.1) and consider the row vectors $\tilde{\Psi}_{\kappa_{\min}}^\kappa(X_t) = [\tilde{\psi}_{\kappa_{\min}+1}(X_t), \ldots, \tilde{\psi}_\kappa(X_t)]$,

$$\tilde{\Psi}_\kappa(X_t) = [\Psi_{\kappa_{\min}}(X_t), \tilde{\Psi}_{\kappa_{\min}}^\kappa(X_t)] = \Psi_\kappa(X_t)\beta_\kappa \quad \text{so that } \tilde{\Psi}_\kappa = \Psi_\kappa \beta_\kappa,$$

for some regular $\kappa \times \kappa$ matrix β_κ. Elementary algebra gives

$$R_\kappa^0 = T^{-1}\varepsilon'\Omega^{-1}\tilde{\Psi}_\kappa \begin{bmatrix} \Sigma_{\kappa_{\min}}^{-1} & 0 \\ 0 & \tilde{\Sigma}_\kappa^{-1} \end{bmatrix} \tilde{\Psi}_\kappa' \Omega^{-1}\varepsilon \quad \text{and}$$

$$R_{\kappa_{\min}}^0 = T^{-1}\varepsilon'\Omega^{-1}\tilde{\Psi}_\kappa \begin{bmatrix} \Sigma_{\kappa_{\min}}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \tilde{\Psi}_\kappa' \Omega^{-1}\varepsilon.$$

Hence

$$\begin{aligned} R_\kappa^0 - R_{\kappa_{\min}}^0 &= T^{-1}\varepsilon'\Omega^{-1}\tilde{\Psi}_\kappa \begin{bmatrix} 0 & 0 \\ 0 & \tilde{\Sigma}_\kappa^{-1} \end{bmatrix} \tilde{\Psi}_\kappa' \Omega^{-1}\varepsilon = T^{-1}\varepsilon'\Omega^{-1}\tilde{\Psi}_{\kappa_{\min}}^\kappa \tilde{\Sigma}_\kappa^{-1} [\tilde{\Psi}_{\kappa_{\min}}^\kappa]'\Omega^{-1}\varepsilon' \\ &= \sqrt{2(\kappa - \kappa_{\min})}\tilde{Q}_\kappa + \kappa - \kappa_{\min}. \end{aligned}$$

We now verify that the quadratic form $\tilde{Q}_\kappa$ obeys the conditions of Lemma A.3. Lemma A.1(i) yields that $\sup_k \sup_{x\in\mathbb{R}^d} |\tilde{\psi}_k(x)| \leq C\kappa_{\min}^{1/2}$, so that Assumption B holds taking $\varphi_\infty = O(\kappa_{\min}^{1/2}) = O(\ln^{C_2 d/2} T)$. Recall that $\kappa - \kappa_{\min} \asymp 2^{jd} - 2^{J_{\min} d}$ by the definition (3.1) of $\mathcal{K}_T$. Hence Lemma A.3(ii) yields, for (A.8),

$$\begin{aligned} &\mathbb{P}\left(\max_{\kappa\in\mathcal{K}_T\setminus\{\kappa_{\min}\}} \tilde{Q}_\kappa \geq \sqrt{2 \ln \operatorname{Card} \mathcal{K}_T} + \epsilon \right) \\ &\quad\leq \sum_{\kappa\in\mathcal{K}_T\setminus\{\kappa_{\min}\}} \mathbb{P}(\tilde{Q}_\kappa \geq \sqrt{2 \ln \operatorname{Card} \mathcal{K}_T} + \epsilon) \\ &\quad\leq \operatorname{Card} \mathcal{K}_T \frac{\exp(-\ln \operatorname{Card} \mathcal{K}_T)}{2\pi\sqrt{2 \ln \operatorname{Card} \mathcal{K}_T}} + C\varphi_\infty^6 \sum_{\kappa\in\mathcal{K}_T\setminus\{\kappa_{\min}\}} (\kappa - \kappa_{\min})^2 T^{-(3/2)[(1+a)/(5+3a)]} \\ &\quad\quad + C \sum_{\kappa\in\mathcal{K}_T\setminus\{\kappa_{\min}\}} (\kappa - \kappa_{\min})^{-1/2} \\ &\quad= o(1) + C\varphi_\infty^6 \operatorname{Card} \mathcal{K}_T \kappa_{\max} T^{-(3/2)[(1+a)/(5+3a)]} + C2^{-dJ_{\min}/2} \sum_{j=1}^{+\infty} 2^{-jd/2} = o(1). \qquad \blacksquare \end{aligned}$$

A.3. Proof of Lemma 1. In this proof, we apply Lemmas A.1 and A.2 for $\mathcal{K}_T = \{\kappa\}$, which is such that $\kappa = \kappa_{\min} = \kappa_{\max} = o(T^{1/3}/\ln^2 T)$. The Jackson theorem (see Timan, 1994, eqn. (8), p. 278) yields that there is a trigonometric polynomial function $\Pi(\cdot) = \Pi_{\Delta_T,\kappa}(\cdot)$ with degree $\asymp \kappa^{1/d}$ such that

$$\Pi(x) = \sum_{k=1}^{\kappa} \beta_k \psi_k(x)\mathbb{I}(x \in \Lambda) \quad \text{such that } \sup_{x\in\Lambda}|\Delta_T(x) - \Pi(x)| \leq C\kappa^{-s/d}. \tag{A.9}$$

Because $\hat{\sigma}(\cdot)$ is bounded away from 0 over Λ in probability, (A.9) implies that

$$\max_{1\leq t\leq T}\left|\frac{\Delta_T(X_t) - \Pi(X_t)}{\hat{\sigma}(X_t)}\right| = O_{\mathbb{P}}(\kappa^{-s/d}).$$

Note that $|m'\hat{\Omega}^{-1}\Psi_\kappa(\Psi_\kappa'\hat{\Omega}^{-1}\Psi_\kappa)\Psi_\kappa'\hat{\Omega}^{-1}m| \leq \|m\| \leq T^{1/2}\sup_{x\in\mathbb{R}^d}|m(x)|$. Let $\Pi = [\Pi(X_1),\ldots,\Pi(X_T)]'$, which is such that $\Pi'\hat{\Omega}^{-1}\Psi_\kappa(\Psi_\kappa'\hat{\Omega}^{-1}\Psi_\kappa)\Psi_\kappa'\hat{\Omega}^{-1}\Pi = \Pi'\hat{\Omega}^{-1}\Pi = \|\hat{\Omega}^{-1/2}\Pi\|$ because $\hat{\Omega}^{-1/2}\Pi$ is in the space spanned by the columns of $\hat{\Omega}^{-1/2}\Psi_\kappa$. Hence the triangular inequality and (A.9) give

$$\begin{aligned}
&[\Delta_T'\hat{\Omega}^{-1}\Psi_\kappa(\Psi_\kappa'\hat{\Omega}^{-1}\Psi_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}\Delta_T]^{1/2} \\
&\quad\geq [\Pi'\hat{\Omega}^{-1}\Psi_\kappa(\Psi_\kappa'\hat{\Omega}^{-1}\Psi_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}\Pi]^{1/2} \\
&\qquad - [(\Pi-\Delta_T)'\hat{\Omega}^{-1}\Psi_\kappa(\Psi_\kappa'\hat{\Omega}^{-1}\Psi_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}(\Pi-\Delta_T)]^{1/2} \\
&\quad\geq \|\hat{\Omega}^{-1/2}\Pi\| - CT^{1/2}\kappa^{-s/d}.
\end{aligned}$$

In the expression (A.9) of $\Pi(\cdot)$, write $\beta = [\beta_1,\ldots,\beta_\kappa]'$, so that the definitions of $\hat{\Sigma}_\kappa$, Σ_κ in (6.2) and Lemma A.1(ii) give

$$\begin{aligned}
\|\hat{\Omega}^{-1/2}\Pi\| &= \left(\sum_{t=1}^{T}\frac{\Pi^2(X_t)}{\hat{\sigma}^2(X_t)}\right)^{1/2} = (T\beta'\hat{\Sigma}_\kappa\beta)^{1/2} \\
&= T(\beta'\Sigma_\kappa\beta)^{1/2}\left(1 + O_{\mathbb{P}}\left[\left(\frac{\kappa^2}{T}\right)^{1/2} + \kappa v_T\right]\right) \\
&= T^{1/2}\mathbb{E}^{1/2}\left(\frac{\Pi(X_t)}{\sigma(X_t)}\right)^2(1+o_{\mathbb{P}}(1)) \geq T^{1/2}\left[\mathbb{E}^{1/2}\left(\frac{\Delta_T(X_t)}{\sigma(X_t)}\right)^2 - C\kappa^{-s/d}\right].
\end{aligned}$$

Substituting shows that (6.3) is proved. Equation (6.4) follows from (A.5) and Assumption V, which gives

$$|\Delta_T'\hat{\Omega}^{-1}\Psi_\kappa(\Psi_\kappa'\hat{\Omega}^{-1}\Psi_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}\varepsilon| = T^{1/2}O_{\mathbb{P}}\left[\mathbb{E}^{1/2}\left(\frac{\Delta_T(X_t)}{\sigma(X_t)}\right)^2 + \kappa^{-s/d}\right]. \qquad \blacksquare$$

APPENDIX B: Proof of Lemmas A.1–A.3

B.1. Proof of Lemma A.1. We begin with Lemma A.1(i), $\sup_{\kappa\in\mathbb{N}^*}\max(\|\Sigma_\kappa^{-1}\|,\|\Sigma_\kappa\|)<\infty$. Because $u'\Sigma_\kappa u=\mathbb{E}(\sum_{k=1}^{\kappa}u_k\psi_k(X_t)/\sigma(X_t))^2$, $\|\Sigma_\kappa\|$ is the largest eigenvalue of the symmetric Σ_κ and $\|\Sigma_\kappa^{-1}\|$ is the inverse of the smallest eigenvalue of Σ_κ. Hence

$$\|\Sigma_\kappa\|=\sup_{\|u\|=1}\mathbb{E}\left(\sum_{k=1}^{\kappa}u_k\frac{\psi_k(X_t)}{\sigma(X_t)}\right)^2,\qquad \frac{1}{\|\Sigma_\kappa^{-1}\|}=\inf_{\|u\|=1}\mathbb{E}\left(\sum_{k=1}^{\kappa}u_k\frac{\psi_k(X_t)}{\sigma(X_t)}\right)^2.$$

Because $f(\cdot)$ and $\sigma(\cdot)$ are bounded away from 0 and infinity over Λ by Assumptions E and X(ii), and because $\{\psi_k(\cdot)\}_{k\in\mathbb{N}^*}$ is an orthonormal system of $L^2(\Lambda,dx)$, we have uniformly in κ

$$\mathbb{E}\left(\sum_{k=1}^{\kappa}u_k\frac{\psi_k(X_t)}{\sigma(X_t)}\right)^2=\int_\Lambda\left(\sum_{k=1}^{\kappa}u_k\psi_k(x)\right)^2\frac{f(x)}{\sigma^2(x)}\,dx\asymp\int_\Lambda\left(\sum_{k=1}^{\kappa}u_k\psi_k(x)\right)^2dx=\|u\|^2.$$

This gives $\sup_{\kappa\in\mathbb{N}^*}\max(\|\Sigma_\kappa^{-1}\|,\|\Sigma_\kappa\|)<\infty$, and we now prove that $\sup_{\kappa\in\mathbb{N}^*}\max(\|\tilde{\Sigma}_\kappa^{-1}\|,\|\tilde{\Sigma}_\kappa\|)<\infty$. Let $\Psi_{\kappa_{\min}}^{\kappa}(X_t)=[\psi_{\kappa_{\min}+1}(X_t),\dots,\psi_\kappa(X_t)]$ and note that

$$\begin{aligned}\tilde{\Sigma}_\kappa&=\left[\mathbb{E}\left(\frac{\psi_k(X_t)\psi_\ell(X_t)}{\sigma^2(X_t)}\right)\right]_{\kappa_{\min}<k,\,\ell\le\kappa}\\&\quad-\mathbb{E}\left(\frac{\Psi_{\kappa_{\min}}(X_t)\Psi_{\kappa_{\min}}^{\kappa'}(X_t)}{\sigma^2(X_t)}\right)\Sigma_{\kappa_{\min}}^{-1}\mathbb{E}\left(\frac{\Psi_{\kappa_{\min}}^{\kappa}(X_t)\Psi_{\kappa_{\min}}'(X_t)}{\sigma^2(X_t)}\right).\end{aligned}$$

It then follows that $\tilde{\Sigma}_\kappa\preccurlyeq[\mathbb{E}\psi_k(X_t)\psi_\ell(X_t)/\sigma^2(X_t)]_{\kappa_{\min}<k,\,\ell\le\kappa}$ where $A\preccurlyeq B$ means that $A-B$ is a symmetric nonnegative matrix. This gives that $\|\tilde{\Sigma}_\kappa\|\le\|\Sigma_\kappa\|$ because the upper bound is a diagonal block submatrix of Σ_κ. Observe that $\tilde{\Sigma}_\kappa^{-1}$ is also a diagonal block of Σ_κ^{-1} by the partitioned inverse formula, so that $\|\tilde{\Sigma}_\kappa^{-1}\|\le\|\Sigma_\kappa^{-1}\|$. This gives $\sup_{\kappa\in\mathbb{N}^*}\max(\|\tilde{\Sigma}_\kappa^{-1}\|,\|\tilde{\Sigma}_\kappa\|)<\infty$. To show that $\sup_{k>\kappa_{\min}}\sup_{x\in\Lambda}|\tilde{\psi}_k(x)|<\infty$, note that $\Psi_{\kappa_{\min}}(\cdot)\Sigma_{\kappa_{\min}}^{-1}\mathbb{E}[\psi_k(X_t)\Psi_{\kappa_{\min}}'(X_t)/\sigma^2(X_t)]$ is the $L_2(\Lambda,f(x)\,dx/\sigma^2(x))$-orthogonal projection of $\psi_k(\cdot)$ on $\psi_1(\cdot),\dots,\psi_{\kappa_{\min}}(\cdot)$. The Pythagore inequality gives, uniformly in $k\ge 1$,

$$\mathbb{E}\left(\frac{\Psi_{\kappa_{\min}}(X_t)}{\sigma(X_t)}\Sigma_{\kappa_{\min}}^{-1}\mathbb{E}\left[\frac{\psi_k(X_t)\Psi_{\kappa_{\min}}'(X_t)}{\sigma^2(X_t)}\right]\right)^2\le\mathbb{E}\left[\frac{\psi_k^2(X_t)}{\sigma^2(X_t)}\right]\le C.$$

Therefore, the Cauchy–Schwarz inequality gives for all x and $\kappa\ge 1$,

$$\begin{aligned}|\tilde{\psi}_k(x)|&\le\sup_{\kappa\ge1}\sup_{x\in\Lambda}|\psi_\kappa(x)|+\sup_{x\in\Lambda}\|\Sigma_{\kappa_{\min}}^{-1/2}\Psi_{\kappa_{\min}}'(x)\|\left\|\Sigma_{\kappa_{\min}}^{-1/2}\mathbb{E}\left[\frac{\psi_k(X_t)\Psi_{\kappa_{\min}}'(X_t)}{\sigma^2(X_t)}\right]\right\|\\&\le C+C\|\Sigma_{\kappa_{\min}}^{-1/2}\|\kappa_{\min}^{1/2}\times\mathbb{E}\left[\frac{\psi_k(X_t)\Psi_{\kappa_{\min}}(X_t)}{\sigma^2(X_t)}\right]\Sigma_{\kappa_{\min}}^{-1}\mathbb{E}\left[\frac{\psi_k(X_t)\Psi_{\kappa_{\min}}'(X_t)}{\sigma^2(X_t)}\right]\\&\le C\kappa_{\min}^{1/2}\mathbb{E}\psi_k^2(X_t)=C\kappa_{\min}^{1/2}.\end{aligned}$$

https://doi.org/10.1017/S0266466606060282 Published online by Cambridge University Press

Consider now Lemma A.1(ii) and (iii). Define

$$\bar{\Sigma}_\kappa = \hat{\Sigma}_\kappa(\Omega) = \left[\overline{\frac{\psi_k(X)\psi_\ell(X)}{\sigma^2(X)}} = \bar{\Sigma}_{k\ell} \right]_{1\le k,\ell\le\kappa}.$$

Assumptions E and X(i) and (6.1) give

$$\mathbb{E}\bar{\Sigma}_{k\ell} = \mathbb{E}\left[\frac{\psi_k(X_t)\psi_\ell(X_t)}{\sigma^2(X)} \right] = \Sigma_{k\ell}, \qquad \operatorname{Var}(\bar{\Sigma}_{k\ell}) \le \frac{C}{T}\sum_{n=0}^{\infty}\alpha(n),$$

and then, by the Cauchy–Schwarz inequality

$$\begin{aligned} \mathbb{E}\max_{1\le\kappa\le\kappa_{\max}} \|\bar{\Sigma}_\kappa - \Sigma_\kappa\|^2 &\le \mathbb{E}\|\bar{\Sigma}_{\kappa_{\max}} - \Sigma_{\kappa_{\max}}\|^2 = \mathbb{E}\sup_{\|u\|=1}\left[\sum_{k=1}^{\kappa_{\max}}\left(\sum_{\ell=1}^{\kappa_{\max}}(\bar{\Sigma}_{k\ell}-\Sigma_{k\ell})u_\ell\right)^2\right] \\ &\le \sup_{\|u\|=1}\sum_{k=1}^{\kappa_{\max}}\sum_{\ell=1}^{\kappa_{\max}}\mathbb{E}(\bar{\Sigma}_{k\ell}-\Sigma_{k\ell})^2\|u\|^2 = O\left(\frac{\kappa_{\max}^2}{T}\right) \qquad \textbf{(B.1)} \end{aligned}$$

and then $\max_{1\le\kappa\le\kappa_{\max}}\|\bar{\Sigma}_\kappa - \Sigma_\kappa\| = O_{\mathbb{P}}(\kappa_{\max}^2/T)^{1/2}$, and we now bound $\max_{1\le\kappa\le\kappa_{\max}}$ $\|\hat{\Sigma}_\kappa - \bar{\Sigma}_\kappa\|$. We have, uniformly in $k \le \kappa_{\max}$,

$$\begin{aligned} \|\hat{\Sigma}_\kappa - \bar{\Sigma}_\kappa\| &= \sup_{\|u\|=1}\left|\frac{u'\Psi_\kappa'(\hat{\Omega}^{-1}-\Omega^{-1})\Psi_\kappa u}{T}\right| \\ &= \sup_{\|u\|=1}\left|\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{k=1}^{\kappa}u_k\psi_k(X_t)\right)^2\left(\frac{1}{\hat{\sigma}^2(X_t)}-\frac{1}{\sigma^2(X_t)}\right)\right| \\ &\le O_{\mathbb{P}}\left(\max_{1\le t\le T}|\hat{\sigma}(X_t)-\sigma(X_t)|\right)\frac{1}{T}\sum_{t=1}^{T}\sum_{k=1}^{\kappa_{\max}}\psi_k^2(X_t) = O_{\mathbb{P}}(\kappa_{\max}v_T). \end{aligned}$$

Because $\kappa_{\max} \asymp K_{\max}^d$, Assumption V and $\kappa_{\max}^2/T = o(1)$ yield

$$\max_{1\le\kappa\le\kappa_{\max}}\|\hat{\Sigma}_\kappa - \Sigma_\kappa\| = O_{\mathbb{P}}\left[\left(\frac{\kappa_{\max}^2}{T}\right)^{1/2} + \kappa_{\max}v_T\right] = o_{\mathbb{P}}(1).$$

Therefore the smallest eigenvalue of $\hat{\Sigma}_\kappa$ is bounded away from 0 and these matrices have an inverse for $1 \le \kappa \le \kappa_{\max}$ with a probability tending to 1. The order of $\max_{1\le\kappa\le\kappa_{\max}}\|\hat{\Sigma}_\kappa^{-1} - \Sigma_\kappa^{-1}\|$ comes from the series expansion

$$\begin{aligned} \|\hat{\Sigma}_\kappa^{-1} - \Sigma_\kappa^{-1}\| &= \|\Sigma_\kappa^{-1}[(\mathrm{Id}_\kappa + (\hat{\Sigma}_\kappa - \Sigma_\kappa)\Sigma_\kappa^{-1})^{-1} - \mathrm{Id}_\kappa]\| = \left\|\sum_{n=1}^{\infty}\Sigma_\kappa^{-1}((\hat{\Sigma}_\kappa - \Sigma_\kappa)\Sigma_\kappa^{-1})^n\right\| \\ &\le \sum_{n=1}^{\infty}\|\hat{\Sigma}_\kappa - \Sigma_\kappa\|^n\left(\sup_{\kappa\in\mathbb{N}^*}\|\Sigma_\kappa^{-1}\|\right)^{n+1}, \end{aligned}$$

which ends the proof of Lemma A.1(i) and (iii) because $\sup_\kappa\|\Sigma_\kappa^{-1}\| < \infty$. ■

https://doi.org/10.1017/S0266466606060282 Published online by Cambridge University Press

B.2. Proof of Lemma A.2. Let us recall some results from an empirical process useful to establish some preliminary bounds. Consider the class of functions $\mathcal{G}_T$ from Λ to $\mathbb{R}$ with

$$\mathcal{G}_T = \Big\{ g(\cdot) : \sup_{x\in\Lambda} |g(x) - \sigma(x)| \le M_T, \sup_{x\in\Lambda} \left| \frac{\partial^{\ell} g(x)}{\partial^{\ell_1} x_1 \ldots \partial^{\ell_d} x_d} \right| \le M_T$$

$$\text{for all } d\text{-uple with } \ell_1 + \cdots + \ell_d = \ell \Big\},$$

with ℓ as in Assumption V. Under Assumption V, there is an $M_T \asymp v_T$ such that

$\liminf_{T\to\infty} \mathbb{P}(\hat{\sigma}^{-2}(\cdot) \in \mathcal{G}_T) \ge 1 - \epsilon, \quad$ for any ϵ.

Then, to establish Lemma A.2, we can view $\hat{\sigma}^{-2}(\cdot)$ as a member of a $\mathcal{G}_T$. Consider now a sequence of functions from Λ to $\mathbb{R}$ and define the empirical process $Z_T^k(\cdot) = \{Z_T^k(g), g \in \mathcal{G}_T\}$ as

$$Z_T^k(g) = \frac{1}{T^{1/2}} \sum_{t=1}^{T} (m_T(X_t)\psi_k(X_t)g(X_t) - \mathbb{E}[m_T(X_t)\psi_k(X_t)g(X_t)]) \quad \text{or}$$

$$Z_T^k(g) = \frac{1}{T^{1/2}} \sum_{t=1}^{T} m_T(X_t)\psi_k(X_t)g(X_t)\varepsilon_t.$$

Modifications of bounds (8.3), (8.7), and (8.9) in Rio (2000) to account for multiplication by $m_T(\cdot)$ and $\psi_k(\cdot)$ with $\sup_{x\in\Lambda}|\psi_k(x)| = 1$ show that

$$\sup_{k\ge 1} \mathbb{E}\left(\sup_{g\in\mathcal{G}_T} |Z_T^k(g) - Z_T^k(\sigma^{-2})|^2 \right) \le O(v_T^2) \sup_{x\in\Lambda} |m_T(x)|^2. \tag{B.2}$$

Define

$$e_\kappa(\varepsilon) = \Psi_\kappa'(\hat{\Omega}^{-1} - \Omega^{-1})\varepsilon, \qquad e_\kappa(m) = \Psi_\kappa'(\hat{\Omega}^{-1} - \Omega^{-1})m_T, \qquad e_\kappa(\Sigma) = \hat{\Sigma}_\kappa^{-1} - \Sigma_\kappa^{-1},$$

so that $\Psi_\kappa'\hat{\Omega}^{-1}\varepsilon = \Psi_\kappa'\Omega^{-1}\varepsilon + e_\kappa(\varepsilon)$, $\Psi_\kappa'\hat{\Omega}^{-1}m_T = \Psi_\kappa'\Omega^{-1}m_T + e_\kappa(m_T)$, and $\hat{\Sigma}_\kappa^{-1} = \Sigma_\kappa^{-1} + e_\kappa(\Sigma)$. The Chebyshev inequality, (B.2), and Lemma A.1(ii) give

$$\begin{aligned} \max_{\kappa\in\mathcal{K}_T} \frac{\|e_\kappa(\varepsilon)\|^2}{T\kappa} &= \max_{\kappa\in\mathcal{K}_T} \frac{1}{\kappa} \sum_{k=1}^{\kappa} \left(\frac{1}{T^{1/2}} \sum_{t=1}^{T} \psi_k(X_t)(\hat{\sigma}^{-2}(X_t) - \sigma^{-2}(X_t))\varepsilon_t \right)^2 \\ &= O_{\mathbb{P}}(1) \sum_{\kappa\in\mathcal{K}_T} \frac{1}{\kappa} \sum_{k=1}^{\kappa} \max_{g(\cdot)\in\mathcal{G}_T} \left(\frac{1}{T^{1/2}} \sum_{t=1}^{T} \psi_k(X_t)(g(X_t) - \sigma^{-2}(X_t))\varepsilon_t \right)^2 \\ &= O_{\mathbb{P}}(v_T^2 \operatorname{Card} \mathcal{K}_T), \end{aligned} \tag{B.3}$$

https://doi.org/10.1017/S0266466606060282 Published online by Cambridge University Press

$$\max_{\kappa \in \mathcal{K}_T} \frac{\|e_\kappa(m)\|^2}{T^2\kappa} = O_{\mathbb{P}}(1) \max_{\kappa \in \mathcal{K}_T} \frac{1}{\kappa} \sum_{k=1}^{\kappa} \max_{g(\cdot) \in \mathcal{G}_T} \left(\frac{1}{T} \sum_{t=1}^{T} m_T(X_t) \psi_k(X_t)(g(X_t) - \sigma^{-2}(X_t)) \right)^2$$

$$\leq O_{\mathbb{P}}(1) \sum_{\kappa \in \mathcal{K}_T} \frac{2}{\kappa} \sum_{k=1}^{\kappa} \max_{g \in \mathcal{G}_T} \left[\left(\frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[m_T(X_t)\psi_k(X_t)(g(X_t) - \sigma^{-2}(X_t))] \right)^2 + \frac{(Z_n^k(g) - Z_n^k(\sigma^{-2}))^2}{T^{1/2}} \right]$$

$$= O_{\mathbb{P}}(v_T^2 \operatorname{Card} \mathcal{K}_T) \sup_{x \in \Lambda} |m_T(x)|^2, \qquad \textbf{(B.4)}$$

$$\max_{\kappa \in \mathcal{K}_T} \|e_\kappa(\Sigma)\| = O_{\mathbb{P}} \left[\left(\frac{\kappa_{\max}^2}{T} \right)^{1/2} + \kappa_{\max} v_T \right]. \qquad \textbf{(B.5)}$$

Observe also that the martingale structure of the ε_t's, Assumption E, and (6.1) yield that

$$\mathbb{E}\left[\max_{\kappa \in \mathcal{K}_T} \frac{\|\Psi_\kappa \Omega^{-1} \varepsilon\|^2}{T\kappa} \right] \leq \sum_{\kappa \in \mathcal{K}_T} \frac{1}{T\kappa} \sum_{k=1}^{\kappa} \mathbb{E}\left(\sum_{t=1}^{T} \frac{\psi_k(X_t)\varepsilon_t}{\sigma^2(X_t)} \right)^2$$

$$\leq C \operatorname{Card} \mathcal{K}_T,$$

$$\mathbb{E}\left[\max_{\kappa \in \mathcal{K}_T} \frac{\|\Psi_\kappa \Omega^{-1} m_T - \mathbb{E}[\Psi_\kappa \Omega^{-1} m_T]\|^2}{T\kappa} \right] \leq \sum_{\kappa \in \mathcal{K}_T} \frac{1}{T\kappa} \sum_{k=1}^{\kappa} \operatorname{Var}\left(\sum_{t=1}^{T} \frac{m_T(X_t)\psi_k(X_t)}{\sigma^2(X_t)} \right)^2$$

$$\leq C \operatorname{Card} \mathcal{K}_T \sup_{x \in \Lambda} |m_T(x)|^2.$$

It follows that

$$\max_{\kappa \in \mathcal{K}_T} \frac{\|\Psi_\kappa \Omega^{-1} \varepsilon\|}{\sqrt{T\kappa}} = O_{\mathbb{P}}(\operatorname{Card}^{1/2} \mathcal{K}_T),$$

$$\max_{\kappa \in \mathcal{K}_T} \frac{\|\Psi_\kappa \Omega^{-1} m_T - \mathbb{E}[\Psi_\kappa \Omega^{-1} m_T]\|}{\sqrt{T\kappa}} = O_{\mathbb{P}}(\operatorname{Card}^{1/2} \mathcal{K}_T) \sup_{x \in \Lambda} |m_T(x)|. \qquad \textbf{(B.6)}$$

Note that (A.2) is due to Cauchy–Schwarz inequality and $\|\hat{\Omega}^{-1/2}\Psi_\kappa(\Psi_\kappa'\hat{\Omega}^{-1}\Psi_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1/2}\| = 1$. Expression (A.3) follows from (B.3) and (B.6). We now prove (A.4). We have

https://doi.org/10.1017/S0266466606060282 Published online by Cambridge University Press

$$
\begin{aligned}
&\varepsilon'\hat{\Omega}^{-1}\Psi_\kappa(\Psi_\kappa'\hat{\Omega}^{-1}\Psi_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}\varepsilon \\
&\quad = \frac{(\varepsilon'\Omega^{-1}\Psi_\kappa + e_\kappa'(\varepsilon))(\Sigma_\kappa^{-1} + e_\kappa(\Sigma))(\Psi_\kappa'\Omega^{-1}\varepsilon + e_\kappa(\varepsilon))}{T} \\
&\quad = \frac{\varepsilon'\Omega^{-1}\Psi_\kappa\Sigma_\kappa^{-1}\Psi_\kappa'\Omega^{-1}\varepsilon}{T} + \frac{2\varepsilon'\Omega^{-1}\Psi_\kappa\Sigma_\kappa^{-1}e_\kappa(\varepsilon) + \varepsilon'\Omega^{-1}\Psi_\kappa e_\kappa(\Sigma)\Psi_\kappa'\Omega^{-1}\varepsilon}{T} \\
&\qquad + \frac{2\varepsilon'\Omega^{-1}\Psi_\kappa e_\kappa(\Sigma)e_\kappa(\varepsilon) + e_\kappa'(\varepsilon)\Sigma_\kappa^{-1}e_\kappa(\varepsilon) + e_\kappa'(\varepsilon)e_\kappa(\Sigma)e_\kappa(\varepsilon)}{T}.
\end{aligned}
$$

By (B.3), (B.5), (B.6), Lemma A.1(i), Assumption V, $\kappa_{\max} = o(T^{1/3}/\ln^{2/3} T)$, and Card $\mathcal{K}_T = O(\ln T)$, we have

$$
\begin{aligned}
\max_{\kappa\in\mathcal{K}_T}\frac{|\varepsilon'\Omega^{-1}\Psi_\kappa e_\kappa(\Sigma)\Psi_\kappa'\Omega^{-1}\varepsilon|}{T\kappa^{1/2}} &\le \kappa_{\max}^{1/2}\max_{\kappa\in\mathcal{K}_T}\frac{|\varepsilon'\Omega^{-1}\Psi_\kappa e_\kappa(\Sigma)\Psi_\kappa'\Omega^{-1}\varepsilon|}{T\kappa} \\
&\le \kappa_{\max}^{1/2}\max_{\kappa\in\mathcal{K}_T}\frac{\|e_\kappa(\Sigma)\|\,\|\Psi_\kappa'\Omega^{-1}\varepsilon\|^2}{T\kappa} \\
&\le \kappa_{\max}^{1/2}\max_{\kappa\in\mathcal{K}_T}\frac{\|\Psi_\kappa'\Omega^{-1}\varepsilon\|^2}{T\kappa} \times \max_{\kappa\in\mathcal{K}_T}\|e_\kappa(\Sigma)\| \\
&= O_{\mathbb{P}}\left[\left(\frac{\kappa_{\max}^3}{T}\right)^{1/2} + \kappa_{\max}^{3/2}v_T\right]\mathrm{Card}\,\mathcal{K}_T = o_{\mathbb{P}}(1), \\
\max_{\kappa\in\mathcal{K}_T}\frac{|\varepsilon'\Omega^{-1}\Psi_\kappa\Sigma_\kappa^{-1}e_\kappa(\varepsilon)|}{T\kappa^{1/2}} &= O_{\mathbb{P}}(\kappa_{\max}^{1/2})\max_{\kappa\in\mathcal{K}_T}\frac{\|\varepsilon'\Omega^{-1}\Psi_\kappa\|}{\sqrt{T\kappa}} \times \max_{\kappa\in\mathcal{K}_T}\frac{\|e_\kappa(\varepsilon)\|}{\sqrt{T\kappa}} \\
&= O_{\mathbb{P}}(\kappa_{\max}v_T\,\mathrm{Card}\,\mathcal{K}_T),
\end{aligned}
$$

the other remainder terms being negligible. This gives (A.3).

We now turn to (A.5). Let $\pi_\kappa(\cdot) = \pi_{\kappa,T}(\cdot)$ be a trigonometric polynomial function of Π_κ with $\sup_{x\in\Lambda}|m_T(x) - \pi_\kappa(x)| \le 2\inf_{\pi(\cdot)\in\Pi_\kappa}\sup_{x\in\Lambda}|m_T(x) - \pi(x)|$. Because $\hat{\Omega}^{-1/2}\pi_{\kappa_{\min}}$ is a linear combination of the columns of $\hat{\Omega}^{-1/2}\Psi_\kappa$ for all $\kappa \ge \kappa_{\min}$, it follows that $\pi_{\kappa_{\min}}'\hat{\Omega}^{-1}\Psi_\kappa(\Psi_\kappa'\hat{\Omega}^{-1}\Psi_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}\varepsilon = \pi_{\kappa_{\min}}'\hat{\Omega}^{-1}\varepsilon$. This gives

$$
m_T'\hat{\Omega}^{-1}\Psi_\kappa(\Psi_\kappa'\hat{\Omega}^{-1}\Psi_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}\varepsilon = \pi_{\kappa_{\min}}'\hat{\Omega}^{-1}\varepsilon + (m_T - \pi_{\kappa_{\min}})'\hat{\Omega}^{-1}\Psi_\kappa(T\hat{\Sigma}_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}\varepsilon \tag{B.7}
$$

with

$$
\begin{aligned}
&(m_T - \pi_{\kappa_{\min}})'\hat{\Omega}^{-1}\Psi_\kappa(T\hat{\Sigma}_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}\varepsilon \\
&\quad = \mathbb{E}\left[\frac{(m_T - \pi_{\kappa_{\min}})'\Omega^{-1}\Psi_\kappa}{T}\right]\Sigma_\kappa^{-1}\Psi_\kappa'\Omega^{-1}\varepsilon \\
&\qquad + \mathbb{E}\left[\frac{(m_T - \pi_{\kappa_{\min}})'\Omega^{-1}\Psi_{\kappa_{\min}}}{T}\right](\hat{\Sigma}_\kappa^{-1}\Psi_\kappa'\hat{\Omega}^{-1} - \Sigma_\kappa^{-1}\Psi_\kappa'\Omega^{-1})\varepsilon \\
&\qquad + \left(\frac{(m_T - \pi_{\kappa_{\min}})'\hat{\Omega}^{-1}\Psi_\kappa}{T} - \mathbb{E}\left[\frac{(m_T - \pi_{\kappa_{\min}})'\Omega^{-1}\Psi_\kappa}{T}\right]\right)\hat{\Sigma}_\kappa^{-1}\Psi_\kappa'\hat{\Omega}^{-1}\varepsilon. \qquad \textbf{(B.8)}
\end{aligned}
$$

Consider first the leading term $\pi'_{\kappa_{\min}}\hat{\Omega}^{-1}\varepsilon$ of (B.7). Because $\sup_{\kappa_{\min}} \sup_{x\in\Lambda}|\pi_{\kappa_{\min}}(x)| < \infty$ and taking $\psi_1(\cdot) = 1$ gives, in (B.2),

$$|\pi'_{\kappa_{\min}}(\hat{\Omega}^{-1} - \Omega^{-1})\varepsilon| = O_{\mathbb{P}}(T^{1/2}v_T)\sup_{x\in\Lambda}|\pi_{\kappa_{\min}}(x)|$$

$$= O_{\mathbb{P}}(T^{1/2}v_T)\left(\sup_{x\in\Lambda}|m_T(x)| + \sup_{x\in\Lambda}|m_T(x) - \pi_{\kappa_{\min}}(x)|\right).$$

The definition of $\pi_{\kappa_{\min}}(\cdot)$ yields, under Assumption E,

$$\mathbb{E}(\pi'_{\kappa_{\min}}\Omega^{-1}\varepsilon)^2 = \mathbb{E}(\pi'_{\kappa_{\min}}\Omega^{-1}\pi_{\kappa_{\min}})$$

$$= T\mathbb{E}\left(\frac{\pi_{\kappa_{\min}}(X_t)}{\sigma(X_t)}\right)^2 \leq T\left(\mathbb{E}^{1/2}\left(\frac{\pi_{\kappa_{\min}}(X_t)}{\sigma(X_t)}\right)^2 + \sup_{x\in\Lambda}|m_T(x) - \pi_{\kappa_{\min}}(x)|\right)^2.$$

This gives, for the leading term of (B.7),

$$\max_{\kappa\in\mathcal{K}_T}\left|\frac{\pi'_{\kappa_{\min}}\hat{\Omega}^{-1}\varepsilon}{\kappa^{1/2}}\right| = \frac{T^{1/2}}{\kappa_{\min}^{1/2}}O_{\mathbb{P}}\left[\mathbb{E}^{1/2}\left(\frac{m_T(X_t)}{\sigma(X_t)}\right)^2 + (1+v_T)\right.$$

$$\left.\times \inf_{\pi(\cdot)\in\Pi_{\kappa_{\min}}}\sup_{x\in\Lambda}|m_T(x) - \pi(x)| + v_T\sup_{x\in\Lambda}|m_T(x)|\right].$$

For the first item of (B.8), note that Assumption E gives that $\mathrm{Var}(\Psi'_\kappa\Omega^{-1}\varepsilon) = T\Sigma_\kappa = \mathbb{E}[\Psi'_\kappa\Omega^{-1}\Psi_\kappa]$; see (6.2). Because orthogonal projection decreases the mean squared norm, this gives, for the first term in (B.8),

$$\mathbb{E}\left(\mathbb{E}\left[\frac{(m_T - \pi_{\kappa_{\min}})'\Omega^{-1}\Psi_\kappa}{T}\right]\Sigma_\kappa^{-1}\Psi'_\kappa\Omega^{-1}\varepsilon\right)^2$$

$$= \frac{1}{T}\mathbb{E}[(m_T - \pi_{\kappa_{\min}})'\Omega^{-1}\Psi_\kappa][\mathbb{E}(\Psi'_\kappa\Omega^{-1}\Psi_\kappa)]^{-1}\mathbb{E}[\Psi'_\kappa\Omega^{-1}(m_T - \pi_{\kappa_{\min}})]$$

$$\leq \frac{1}{T}\mathbb{E}[(m_T - \pi_{\kappa_{\min}})'\Omega^{-1}(m_T - \pi_{\kappa_{\min}})]$$

$$= \mathbb{E}\left(\frac{m_T(X_t) - \pi_{\kappa_{\min}}(X_t)}{\sigma(X_t)}\right)^2 \leq C\sup_{x\in\Lambda}|m_T(x) - \pi_{\kappa_{\min}}(x)|,$$

so that

$$\max_{\kappa\in\mathcal{K}_T}\left|\frac{1}{T\kappa^{1/2}}\mathbb{E}[(m_T - \pi_{\kappa_{\min}})'\Omega^{-1}\Psi_\kappa]\Sigma_\kappa^{-1}\Psi'_\kappa\Omega^{-1}\varepsilon\right|$$

$$= \inf_{\pi(\cdot)\in\Pi_{\kappa_{\min}}}\sup_{x\in\Lambda}|m_T(x) - \pi(x)|O_{\mathbb{P}}\left(\frac{\mathrm{Card}^{1/2}\,\mathcal{K}_T}{\kappa_{\min}^{1/2}}\right).$$

For the second term in (B.8), observe that

$$\left\| \mathbb{E}\left[\frac{(m_T - \pi_{\kappa_{\min}})'\Omega^{-1}\Psi_\kappa}{T^{1/2}} \right]' \right\|^2 \le C \left\| \Sigma_\kappa^{-1/2} \mathbb{E}\left[\frac{(m_T - \pi_{\kappa_{\min}})'\Omega^{-1}\Psi_\kappa}{T^{1/2}} \right]' \right\|^2$$

$$= \frac{C}{T} \mathbb{E}[(m_T - \pi_{\kappa_{\min}})'\Omega^{-1}\Psi_\kappa][\mathbb{E}(\Psi_\kappa'\Omega^{-1}\Psi_\kappa)]^{-1}$$

$$\times \mathbb{E}[\Psi_\kappa'\Omega^{-1}(m_T - \pi_{\kappa_{\min}})]$$

$$\le C\mathbb{E}\left(\frac{m_T(X_t) - \pi_{\kappa_{\min}}(X_t)}{\sigma(X_t)} \right)^2$$

$$\le C \sup_{x\in\Lambda} |m_T(x) - \pi_{\kappa_{\min}}(x)|, \quad \text{and then}$$

$$\left| \mathbb{E}\left[\frac{(m_T - \pi_{\kappa_{\min}})'\Omega^{-1}\Psi_{\kappa_{\min}}}{T} \right] (\hat{\Sigma}_\kappa^{-1}\Psi_\kappa'\hat{\Omega}^{-1} - \Sigma_\kappa^{-1}\Psi_\kappa'\Omega^{-1})\varepsilon \right|$$

$$\le \frac{C \sup_{x\in\Lambda} |m_T(x) - \pi_{\kappa_{\min}}(x)|}{T^{1/2}} [\|\hat{\Sigma}_\kappa^{-1} - \Sigma_\kappa^{-1}\| \|\Psi_\kappa'\Omega^{-1}\varepsilon\| + \|\hat{\Sigma}_\kappa^{-1}\| \|e_\kappa(\varepsilon)\|].$$

Therefore Lemma A.1, (B.3), and (B.6) yield

$$\max_{\kappa\in\mathcal{K}_T} \frac{1}{\kappa^{1/2}} \left| \mathbb{E}\left[\frac{(m_T - \pi_{\kappa_{\min}})'\Omega^{-1}\Psi_{\kappa_{\min}}}{T} \right] (\hat{\Sigma}_\kappa^{-1}\Psi_\kappa'\hat{\Omega}^{-1} - \Sigma_\kappa^{-1}\Psi_\kappa'\Omega^{-1})\varepsilon \right|$$

$$= \inf_{\pi(\cdot)\in\Pi_{\kappa_{\min}}} \sup_{x\in\Lambda} |m_T(x) - \pi(x)| O_{\mathbb{P}}\left[\left(\frac{\kappa_{\max}^2}{T} \right)^{1/2} + \kappa_{\max} v_T \right] \text{Card}^{1/2}\,\mathcal{K}_T.$$

For the last item of (B.8), (B.3), (B.4), (B.6), and Lemma A.1 give that

$$\max_{\kappa\in\mathcal{K}_T} \frac{1}{\sqrt{\kappa}} \left| \left(\frac{(m_T - \pi_{\kappa_{\min}})'\hat{\Omega}^{-1}\Psi_\kappa}{T} - \mathbb{E}\left[\frac{(m_T - \pi_{\kappa_{\min}})'\Omega^{-1}\Psi_\kappa}{T} \right] \right) \hat{\Sigma}_\kappa^{-1}\Psi_\kappa'\hat{\Omega}^{-1}\varepsilon \right|$$

$$= \inf_{\pi(\cdot)\in\Pi_{\kappa_{\min}}} \sup_{x\in\Lambda} |m_T(x) - \pi(x)| \kappa_{\max}^{1/2} O_{\mathbb{P}}(v_T \,\text{Card}^{1/2}\,\mathcal{K}_T)$$

$$\times \sqrt{T\kappa_{\max}} O_{\mathbb{P}}[(1 + v_T)\text{Card}^{1/2}\,\mathcal{K}_T]$$

$$= \inf_{\pi(\cdot)\in\Pi_{\kappa_{\min}}} \sup_{x\in\Lambda} |m_T(x) - \pi(x)| T^{1/2} O_{\mathbb{P}}(\kappa_{\max}^{1/2} v_T \,\text{Card}\,\mathcal{K}_T).$$

https://doi.org/10.1017/S0266466606060282 Published online by Cambridge University Press

Substituting in (B.8) and (B.7) yields

$$\max_{\kappa\in\mathcal{K}_T} \frac{|m_T'\hat{\Omega}^{-1}\Psi_\kappa(\Psi_\kappa'\hat{\Omega}^{-1}\Psi_\kappa)^{-1}\Psi_\kappa'\hat{\Omega}^{-1}\varepsilon|}{\kappa^{1/2}}$$

$$= \frac{T^{1/2}}{\kappa_{\min}^{1/2}} O_{\mathbb{P}}\left[\mathbb{E}^{1/2}\left(\frac{m_T(X_t)}{\sigma(X_t)}\right)^2 + (1+v_T) \inf_{\pi(\cdot)\in\Pi_{\kappa_{\min}}} \sup_{x\in\Lambda}|m_T(x)-\pi(x)| + v_T \sup_{x\in\Lambda}|m_T(x)|\right]$$

$$+ \inf_{\pi(\cdot)\in\Pi_{\kappa_{\min}}} \sup_{x\in\Lambda}|m_T(x)-\pi(x)| O_{\mathbb{P}}$$

$$\times\left[\left(\kappa_{\min}^{-1/2} + \left(\frac{\kappa_{\max}^2}{T}\right)^{1/2} + \kappa_{\max} v_T\right)\mathrm{Card}^{1/2}\,\mathcal{K}_T + T^{1/2}\kappa_{\max}^{1/2} v_T\,\mathrm{Card}\,\mathcal{K}_T\right]$$

$$= \frac{T^{1/2}}{\kappa_{\min}^{1/2}} O_{\mathbb{P}}\left[\mathbb{E}^{1/2}\left(\frac{m_T(X_t)}{\sigma(X_t)}\right)^2 + \inf_{\pi(\cdot)\in\Pi_{\kappa_{\min}}} \sup_{x\in\Lambda}|m_T(x)-\pi(x)| + v_T \sup_{x\in\Lambda}|m_T(x)|\right].$$

■

B.3. Proof of Lemma A.3. Abbreviate $V_\kappa^{-1/2}\Phi_\kappa'(X_t)\varepsilon_t$ into η_t. Consider a sequence $\{\tilde{\eta}_t\}_{t\in\mathbb{N}}$ of i.i.d. $N(0,\mathrm{Id}_\kappa)$ variables independent of $\{\varepsilon_t\}_{t\in\mathbb{N}}$ and $\{X_t\}_{t\in\mathbb{N}}$, where Id_κ is the identity matrix of dimension $\kappa\times\kappa$. Let $\mathcal{I}(\cdot)$ be a three time differentiable real function. Define $\tilde{S}_{t+1}^T = \sum_{i=t+1}^T \tilde{\eta}_t$, $\tilde{Q}_T = (T^{-1}\|\tilde{S}_1^T\|^2 - \sigma^2\kappa)/\sqrt{2\kappa}$. The proof of Lemma A.3 is divided into three steps. The main step aims to establish that for $C(\mathcal{I}) = \max(1,\sup_{z\in\mathbb{R}}|\mathcal{I}'(z)|,\sup_{z\in\mathbb{R}}|\mathcal{I}''(z)|)$ and some $C>0$ independent of κ and T,

$$|\mathbb{E}[\mathcal{I}(Q_T)] - \mathbb{E}[\mathcal{I}(\tilde{Q}_T)]| \leq C\cdot C(\mathcal{I})\cdot\varphi_\infty^6\kappa^2 T^{-(3/2)[(1+a)/(5+3a)]}. \quad \textbf{(B.9)}$$

Step 1. Proof of (B.9). We build on arguments used in the proof of the Lindeberg central limit theorem as given in Billingsley (1968, Thm. 7.2); see Horowitz and Spokoiny (2001, Lem. 10) for a similar approach in the context of adaptive testing. It consists of successive changes of the η_t into their Gaussian counterparts $\hat{\eta}_t$, as seen from (B.10), which follows. However, a important difference is due to the use of nonparametric series methods and dependence. Define

$$S_{tT}(\eta) = S_{t-1} + \eta + \tilde{S}_{t+1}^T, \qquad Q_{tT}(\eta) = \frac{T^{-1}S_{tT}'(\eta)S_{tT}(\eta) - \kappa}{\sqrt{2\kappa}},$$

$$\mathcal{J}_{tT}(\eta) = \mathcal{I}(Q_{tT}(\eta)) \quad \text{for } \eta\in\mathbb{R}^\kappa.$$

This gives

$$|\mathbb{E}[\mathcal{I}(Q_T)] - \mathbb{E}[\mathcal{I}(\tilde{Q}_T)]| = |\mathbb{E}[\mathcal{J}_{TT}(\eta_T)] - \mathbb{E}[\mathcal{J}_{1T}(\tilde{\eta}_1)]| = \left|\sum_{t=1}^T (\mathbb{E}[\mathcal{J}_{tT}(\eta_t)] - \mathbb{E}[\mathcal{J}_{tT}(\tilde{\eta}_t)])\right|$$

$$\leq \sum_{t=1}^T |\mathbb{E}[\mathcal{J}_{tT}(\eta_t)] - \mathbb{E}[\mathcal{J}_{tT}(\tilde{\eta}_t)]|. \quad \textbf{(B.10)}$$

Define, for $z \in \mathbb{R}$ and $\eta \in \mathbb{R}^k$, $\mathcal{J}_{tT}(z;\eta) = \mathcal{J}_{tT}(z\eta)$. A third-order Taylor expansion of $\mathcal{J}_{tT}(z;\eta)$ with integral remainder yields

$$\mathcal{J}_{tT}(\eta_t) - \mathcal{J}_{tT}(0) = \frac{d\mathcal{J}_{tT}(0;\eta_t)}{dz} + \frac{1}{2}\frac{d^2\mathcal{J}_{tT}(0;\eta_t)}{d^2z} + \int_0^1 \frac{(1-z)^2}{2}\frac{d^3\mathcal{J}_{tT}(z;\eta_t)}{d^3z}\,dz,$$

with

$$\begin{cases} \dfrac{d\mathcal{J}_{tT}(0;\eta_t)}{dz} = \dfrac{2}{T\sqrt{2\kappa}}\,\eta_t' S_{tT}(0)\mathcal{I}'(Q_{tT}(0)), \\ \dfrac{d^2\mathcal{J}_{tT}(0;\eta_t)}{d^2z} = \dfrac{2}{T\sqrt{2\kappa}}\,\|\eta_t\|^2\mathcal{I}'(Q_{tT}(0)) + \dfrac{4}{T^2\kappa}\,(\eta_t' S_{tT}(0))^2\mathcal{I}''(Q_{tT}(0)), \\ \dfrac{d^3\mathcal{J}_{tT}(z;\eta_t)}{d^3z} = \dfrac{10}{T^2\kappa}\,\|\eta_t\|^2\eta_t' S_{tT}(z\eta_t)\mathcal{I}''(Q_{tT}(z\eta_t)) + \dfrac{8}{T^3\kappa^{3/2}}\,(\eta_t' S_{tT}(z\eta_t))^3\mathcal{I}'''(Q_{tT}(z\eta_t)). \end{cases} \quad \textbf{(B.11)}$$

Let $\tilde{\mathcal{F}}_t$ be the sigma field generated by $\ldots,\eta_{t-2},\eta_{t-1},\tilde{\eta}_{t+1},\tilde{\eta}_{t+2},\ldots$ and note that $S_{tT}(0)$ and $Q_{tT}(0)$ are $\tilde{\mathcal{F}}_t$-measurable. Because η_t and $\tilde{\eta}_t$ are centered given $\tilde{\mathcal{F}}_t$, we have

$$\mathbb{E}\left[\frac{d\mathcal{J}_{tT}(0;\eta_t)}{dz} - \frac{d\mathcal{J}_{tT}(0;\tilde{\eta}_t)}{dz}\right] = \mathbb{E}\left[\frac{2}{T\sqrt{2\kappa}}\,S_{tT}'(0)\mathcal{I}'(Q_{tT}(0))\mathbb{E}[(\eta_t - \tilde{\eta}_t)|\tilde{\mathcal{F}}_t]\right] = 0.$$

Substituting the Taylor expansion in (B.10) yields

$$|\mathbb{E}[\mathcal{I}(Q_T)] - \mathbb{E}[\mathcal{I}(\tilde{Q}_T)]| \le \frac{1}{2}\sum_{t=1}^T \left|\mathbb{E}\left[\frac{d^2\mathcal{J}_{tT}(0;\eta_t)}{d^2z} - \frac{d^2\mathcal{J}_{tT}(0;\tilde{\eta}_t)}{d^2z}\right]\right| \quad \textbf{(B.12)}$$

$$+ \frac{1}{2}\sum_{t=1}^T \int_0^1 (1-z)^2\left[\left|\mathbb{E}\,\frac{d^3\mathcal{J}_{tT}(z;\eta_t)}{d^3z}\right| + \left|\mathbb{E}\,\frac{d^3\mathcal{J}_{tT}(z;\tilde{\eta}_t)}{d^3z}\right|\right]dz, \quad \textbf{(B.13)}$$

and we now bound each of these two sums.

We begin by establishing a preliminary inequality. Let n_1 and n_2 be two positive real numbers with $2 \le n_1 + n_2 \ge 8$. Then for any t, t' and $z \in [0,1]$,

$$\max(\mathbb{E}\|S_{tT}(z\eta_t)\|^{n_1}\|\eta_{t'}\|^{n_2}, \mathbb{E}\|S_{tT}(z\tilde{\eta}_t)\|^{n_1}\|\tilde{\eta}_{t'}\|^{n_2}) \le C\varphi_\infty^{n_1+n_2}\kappa^{(n_1+n_2)/2}T^{n_1/2}. \quad \textbf{(B.14)}$$

We give a proof for $\mathbb{E}\|S_{tT}(z\eta_t)\|^{n_1}\|\eta_{t'}\|^{n_2}$, the other bound being similarly established. The Hölder inequality implies that

$$\begin{aligned} \mathbb{E}\|S_{tT}(z\eta_t)\|^{n_1}\|\eta_{t'}\|^{n_2} &\le \mathbb{E}^{n_1/(n_1+n_2)}\|S_{tT}(z\eta_t)\|^{n_1+n_2}\mathbb{E}^{n_2/(n_1+n_2)}\|\eta_{t'}\|^{n_1+n_2} \\ &\le \mathbb{E}^{n_1/(n_1+n_2)}(\|S_t + z\eta_t\| + \|\tilde{S}_{t+1}^T\|)^{n_1+n_2} \\ &\quad \times \left(\|V_\kappa^{-1/2}\|^{n_2}\mathbb{E}^{n_2/(n_1+n_2)}\left[\sum_{k=1}^{\kappa}\varphi_k^2(X_{t'})\varepsilon_{t'}^2\right]^{(n_1+n_2)/2}\right) \\ &\le 2^{n_1+n_2-1}(\mathbb{E}^{n_1/(n_1+n_2)}\|S_t + z\eta_t\|^{n_1+n_2} + \mathbb{E}^{n_1/(n_1+n_2)}\|\tilde{S}_{t+1}^T\|^{n_1+n_2}) \\ &\quad \times (\|V_\kappa^{-1/2}\|^{n_2}\varphi_\infty^{n_2}\kappa^{n_2/2}\mathbb{E}^{n_2/(n_1+n_2)}|\varepsilon_{t'}|^{n_1+n_2}). \end{aligned}$$

https://doi.org/10.1017/S0266466606060282 Published online by Cambridge University Press

Because $\tilde{\eta}_t$ is an $N(0,\sigma^2 \mathrm{Id}_\kappa)$, it is easily seen that $\mathbb{E}^{n_1/(n_1+n_2)}\|\tilde{S}_{t+1}^T\|^{n_1+n_2} \leq C\sigma^{n_1+n_2}(\kappa T)^{n_1/2}$, and we now bound $\mathbb{E}^{n_1/(n_1+n_2)}\|S_t + z\eta_t\|^{n_1+n_2}$. We have, by convexity, the Burkholder inequality (see Chow and Teicher, 1988, p. 396, noticing that $\sum_{i=1}^{t-1}\varphi_k(X_i)\varepsilon_i + z\varphi_k(X_t)\varepsilon_t$ is a sum of difference of martingale), and the Minkowski inequality

$$
\begin{aligned}
&\mathbb{E}\|S_t + z\eta_t\|^{n_1+n_2}\\
&\quad\leq \|V_\kappa^{-1/2}\|^{n_1+n_2}\kappa^{(n_1+n_2)/2}\mathbb{E}\left[\frac{1}{\kappa}\sum_{k=1}^{\kappa}\left(\sum_{i=1}^{t-1}\varphi_k(X_i)\varepsilon_i + z\varphi_k(X_t)\varepsilon_t\right)^2\right]^{(n_1+n_2)/2}\\
&\quad\leq \|V_\kappa^{-1/2}\|^{n_1+n_2}\kappa^{[(n_1+n_2)/2]-1}\sum_{k=1}^{\kappa}\left[\mathbb{E}^{1/(n_1+n_2)}\left|\sum_{i=1}^{t-1}\varphi_k(X_i)\varepsilon_i + z\varphi_k(X_t)\varepsilon_t\right|^{n_1+n_2}\right]^{n_1+n_2}\\
&\quad\leq \|V_\kappa^{-1/2}\|^{n_1+n_2}\kappa^{[(n_1+n_2)/2]-1}\\
&\qquad\times C\sum_{k=1}^{\kappa}\left[\mathbb{E}^{2/(n_1+n_2)}\left(\sum_{i=1}^{t-1}\varphi_k^2(X_i)\varepsilon_i^2 + z^2\varphi_k^2(X_t)^2\varepsilon_t\right)^{(n_1+n_2)/2}\right]^{(n_1+n_2)/2}\\
&\quad\leq C\|V_\kappa^{-1/2}\|^{n_1+n_2}\kappa^{[(n_1+n_2)/2]-1}\sum_{k=1}^{\kappa}\left[\sum_{i=1}^{t}\mathbb{E}^{2/(n_1+n_2)}|\varphi_k^2(X_i)\varepsilon_i^2|^{(n_1+n_2)/2}\right]^{(n_1+n_2)/2}\\
&\quad\leq C\|V_\kappa^{-1/2}\|^{n_1+n_2}(\kappa T)^{(n_1+n_2)/2}\varphi_\infty^{n_1+n_2}\mathbb{E}|\varepsilon_t|^{n_1+n_2}.
\end{aligned}
$$

This gives $\mathbb{E}^{n_1/(n_1+n_2)}\|S_t + z\eta_t\|^{n_1+n_2} \leq C(\kappa T)^{n_1/2}\varphi_\infty^{n_1}$ and then (B.14).

We now return to (B.13). The expression (B.11) of the third derivative of $\mathcal{J}_{tT}(z;\eta_t)$ and (B.14) yield

$$
\begin{aligned}
&\sum_{t=1}^{T}\int_0^1(1-z)^2\left[\left|\mathbb{E}\,\frac{d^3\mathcal{J}_{tT}(z;\eta_t)}{d^3z}\right| + \left|\mathbb{E}\,\frac{d^3\mathcal{J}_{tT}(z;\tilde{\eta}_t)}{d^3z}\right|\right]dz\\
&\quad\leq C(\mathcal{I})\sum_{t=1}^{T}\int_0^1(1-z)^2\left\{\frac{10}{T^2\kappa}\sum_{t=1}^{T}\mathbb{E}[\|\eta_t\|^3\|S_{tT}(z\eta_t)\| + \|\tilde{\eta}_t\|^3\|S_{tT}(z\tilde{\eta}_t)\|]\right.\\
&\qquad\left. + \frac{8}{T^3\kappa^{3/2}}\mathbb{E}[\|\eta_t\|^3\|S_{tT}(z\eta_t)\|^3 + \|\tilde{\eta}_t\|^3\|S_{tT}(z\tilde{\eta}_t)\|^3]\right\}dz\\
&\quad\leq C\cdot\varphi_\infty^6\cdot C(\mathcal{I})(\kappa\cdot T^{-1/2} + \kappa^{3/2}\cdot T^{-1/2}) \leq C\cdot\varphi_\infty^6\cdot C(\mathcal{I})\left(\frac{\kappa^3}{T}\right)^{1/2}. \qquad \textbf{(B.15)}
\end{aligned}
$$

To study (B.12), let $\bar{\Phi}_\kappa(X_t) = V_\kappa^{-1/2}\Phi_\kappa'(X_t) = [\bar{\varphi}_1(X_t),\ldots,\bar{\varphi}_\kappa(X_t)]'$, $S_{tT} = S_{tT}(0) = [S_{1tT},\ldots,S_{\kappa tT}]'$, $Q_{tT} = Q_{tT}(0)$. The definitions of η_t, $\tilde{\eta}_t$, and $\tilde{\mathcal{F}}_t$ show that $\mathbb{E}[\eta_{kt}\eta_{\ell t} - \tilde{\eta}_{kt}\tilde{\eta}_{\ell t}|\tilde{\mathcal{F}}_t] = \bar{\varphi}_k(X_t)\bar{\varphi}_\ell(X_t) - \mathbb{I}(\kappa=\ell) = \bar{\varphi}_k(X_t)\bar{\varphi}_\ell(X_t) - \mathbb{E}[\bar{\varphi}_k(X_t)\bar{\varphi}_\ell(X_t)]$. Therefore because Q_{tT} and S_{tT} are $\tilde{\mathcal{F}}_t$ measurable, conditioning with respect to $\tilde{\mathcal{F}}_t$ yields, using the expression of the second-order derivative of $\mathcal{J}_{tT}(0;\eta_t)$ given in (B.11),

$$\mathbb{E}\left[\frac{d^2\mathcal{J}_{tT}(0;\eta_t)}{d^2z}-\frac{d^2\mathcal{J}_{tT}(0;\tilde{\eta}_t)}{d^2z}\right]$$

$$=\frac{2}{T\sqrt{2\kappa}}\sum_{k=1}^{\kappa}\mathbb{E}[(\bar{\varphi}_k^2(X_t)-\mathbb{E}\bar{\varphi}_k^2(X_t))\mathcal{I}'(Q_{tT})]$$

$$+\frac{4}{T^2\kappa}\sum_{1\le k,\ell\le\kappa}\mathbb{E}[(\bar{\varphi}_k(X_t)\bar{\varphi}_\ell(X_t)-\mathbb{E}[\bar{\varphi}_k(X_t)\bar{\varphi}_\ell(X_t)])S_{ktT}S_{\ell tT}\mathcal{I}''(Q_{tT})]$$

$$=\frac{2}{T\sqrt{2\kappa}}\sum_{k=1}^{\kappa}\mathrm{Cov}(\bar{\varphi}_k^2(X_t),\mathcal{I}'(Q_{tT}))$$

$$+\frac{4}{T^2\kappa}\sum_{1\le k,\ell\le\kappa}\mathrm{Cov}(\bar{\varphi}_k(X_t)\bar{\varphi}_\ell(X_t),S_{ktT}S_{\ell tT}\mathcal{I}''(Q_{tT})).$$

Let n be an integer and define

$$\breve{S}_{tT}=S_{t-n-1}+\tilde{S}_{t+1}^{T},$$

$$\breve{Q}_{tT}=\frac{T^{-1}\|\breve{S}_{tT}\|^2-\sigma^2\kappa}{\sqrt{2\kappa}}=Q_{tT}-\frac{2\breve{S}_{tT}'(S_t-S_{t-n-1})+\|S_t-S_{t-n-1}\|^2}{T\sqrt{2\kappa}}.$$

The variables $\breve{Q}_{tT}$ and $\breve{S}_{tT}$ depend upon $\tilde{\eta}_{t+1},\dots,\tilde{\eta}_T$ and $\eta_1,\dots,\eta_{t-n-1}$, which are $n+1$ time periods far from the $\bar{\varphi}_k^2(X_t)$'s. Because $\sup_{x\in\mathbb{R}^d}|\bar{\varphi}_k(x)|\le\sup_{x\in\mathbb{R}^d}\|\bar{\Phi}_\kappa(x)\|\le\|V_\kappa^{-1/2}\|\varphi_\infty\sqrt{\kappa}$, the Wolkonski–Rozanov inequality yields

$$\left|\sum_{k=1}^{\kappa}\mathrm{Cov}(\bar{\varphi}_k^2(X_t),\mathcal{I}'(\breve{Q}_{tT}))\right|$$

$$\le 4C(\mathcal{I})\|V_\kappa^{-1/2}\|^2\varphi_\infty^2\kappa\alpha(n), \tag{B.16}$$

$$\left|\sum_{1\le k,\ell\le\kappa}\mathrm{Cov}[\bar{\varphi}_k(X_t)\bar{\varphi}_\ell(X_t),\breve{S}_{ktT}\breve{S}_{\ell tT}\mathcal{I}''(\breve{Q}_{tT})]\right|$$

$$\le 8C(\mathcal{I})\|V_\kappa^{-1/2}\|^2\varphi_\infty^2\kappa\alpha^{3/4}(n)\sum_{1\le k,\ell\le\kappa}\mathbb{E}^{1/4}|\breve{S}_{ktT}\breve{S}_{\ell tT}|^4$$

$$\le 8C(\mathcal{I})\|V_\kappa^{-1/2}\|^2\varphi_\infty^2\kappa\alpha^{3/4}(n)\sum_{1\le k,\ell\le\kappa}\mathbb{E}^{1/8}|\breve{S}_{ktT}|^8\mathbb{E}^{1/8}|\breve{S}_{\ell tT}|^8$$

$$\le C\cdot C(\mathcal{I})\|V_\kappa^{-1/2}\|^4\varphi_\infty^4\kappa^3T\alpha^{3/4}(n), \tag{B.17}$$

by first integrating out with respect to the $\tilde{\eta}_{t+1},\dots,\tilde{\eta}_T$, which are independent from the η_t's, and using (B.14). Note that $\mathbb{E}\bar{\varphi}_k(X_t)^4\le\|V_\kappa^{-1/2}\|^2\varphi_\infty^2\kappa\mathbb{E}\bar{\varphi}_k(X_t)^2=\|V_\kappa^{-1/2}\|^2\varphi_\infty^2\kappa$ and $\mathrm{Var}^{1/2}(\bar{\varphi}_k^2(X_t)\bar{\varphi}_k^2(X_t))\le(\mathbb{E}\bar{\varphi}_k^4(X_t)\mathbb{E}\bar{\varphi}_k^4(X_t))^{1/4}\le\|V_\kappa^{-1/2}\|\varphi_\infty\sqrt{\kappa}$. This together with the definition of $\breve{Q}_{tT}$ and (B.14) gives

$$\left|\sum_{k=1}^{\kappa} \mathrm{Cov}[\bar{\varphi}_k^2(X_t), \mathcal{I}'(Q_{tT}) - \mathcal{I}'(\check{Q}_{tT})]\right|$$

$$\leq \frac{C(\mathcal{I})}{T\sqrt{2\kappa}} \sum_{k=1}^{\kappa} \mathbb{E}[|\bar{\phi}_k^2(X_t) - 1|(2\|\check{S}_{tT}\|\|S_{t-1} - S_{t-n-1}\| + \|S_{t-1} - S_{t-n-1}\|^2)]$$

$$\leq \frac{C(\mathcal{I})}{T\sqrt{2\kappa}} \sum_{k=1}^{\kappa} \mathbb{E}^{1/2}|\bar{\phi}_k^2(X_t) - 1|^2$$

$$\times (\mathbb{E}^{1/4}\|\check{S}_{tT}\|^4 \times \mathbb{E}^{1/4}\|S_{t-1} - S_{t-n-1}\|^4 + \mathbb{E}^{1/2}\|S_{t-1} - S_{t-n-1}\|^4)$$

$$\leq C\frac{C(\mathcal{I})}{T\sqrt{2\kappa}} (\kappa \times \varphi_\infty \kappa^{1/2} \times (\varphi_\infty\sqrt{\kappa T}\varphi_\infty\sqrt{\kappa n} + \varphi_\infty^2 \kappa n))$$

$$= C \cdot C(\mathcal{I}) \cdot \varphi_\infty^3 \kappa^2 \frac{\sqrt{Tn} + n}{T},$$

$$\left|\sum_{1\leq k,\ell\leq\kappa} \mathrm{Cov}[\bar{\varphi}_k(X_t)\bar{\varphi}_\ell(X_t), S_{\ell tT} S_{ktT}(\mathcal{I}''(Q_{tT}) - \mathcal{I}''(\check{Q}_{tT}))]\right|$$

$$\leq \frac{C(\mathcal{I})}{T\sqrt{2\kappa}} (\mathbb{E}^{1/8}\|\check{S}_{tT}\|^8 \mathbb{E}^{1/8}\|S_{t-1} - S_{t-n-1}\|^8 + \mathbb{E}^{1/4}\|S_{t-1} - S_{t-n-1}\|^8)$$

$$\times \sum_{1\leq k,\ell\leq\kappa} \mathrm{Var}^{1/2}(\bar{\varphi}_k^2(X_t)\bar{\varphi}_\ell^2(X_t))\mathbb{E}^{1/8} S_{\ell tT}^8 \mathbb{E}^{1/8} S_{ktT}^8$$

$$\leq C\frac{C(\mathcal{I})}{T\sqrt{\kappa}} \kappa\varphi_\infty^2(\sqrt{Tn} + n) \times \kappa^2\varphi_\infty\sqrt{\kappa}\varphi_\infty^2 T = C \cdot C(\mathcal{I}) \cdot \varphi_\infty^5 \kappa^3(\sqrt{Tn} + n),$$

$$\left|\sum_{1\leq k,\ell\leq\kappa} \mathrm{Cov}[\bar{\varphi}_k(X_t)\bar{\varphi}_\ell(X_t), (S_{\ell tT} - \check{S}_{\ell tT})(S_{ktT} - \check{S}_{ktT})\mathcal{I}''(\check{Q}_{tT})]\right|$$

$$\leq C(\mathcal{I}) \sum_{1\leq k,\ell\leq\kappa} \mathrm{Var}^{1/2}(\bar{\varphi}_k^2(X_t)\bar{\varphi}_\ell^2(X_t))\mathbb{E}^{1/4}(S_{\ell tT} - \check{S}_{\ell tT})^4 \mathbb{E}^{1/4}(S_{ktT} - \check{S}_{ktT})^4$$

$$\leq C \cdot C(\mathcal{I}) \cdot \varphi_\infty^3 \kappa^{5/2} n,$$

$$\left|\sum_{1\leq k,\ell\leq\kappa} \mathrm{Cov}[\bar{\varphi}_k(X_t)\bar{\varphi}_\ell(X_t), \check{S}_{ktT}(S_{\ell tT} - \check{S}_{\ell tT})\mathcal{I}''(\check{Q}_{tT})]\right|$$

$$\leq C(\mathcal{I}) \sum_{1\leq k,\ell\leq\kappa} \mathrm{Var}^{1/2}(\bar{\varphi}_k^2(X_t)\bar{\varphi}_\ell^2(X_t))\mathbb{E}^{1/4}\check{S}_{ktT}^4 \mathbb{E}^{1/4}(S_{\ell tT} - \check{S}_{\ell tT})^4$$

$$\leq C \cdot C(\mathcal{I}) \cdot \varphi_\infty^3 \kappa^{5/2}\sqrt{nT}.$$

Therefore, (B.16), (B.17), and these inequalities give

$$\left|\sum_{k=1}^{\kappa} \mathrm{Cov}(\bar{\varphi}_k^2(X_t), \mathcal{I}'(Q_{tT}))\right|$$

$$\leq C\cdot C(\mathcal{I})\cdot \varphi_\infty^3 \kappa\left(\alpha_X(n) + \kappa\,\frac{(\sqrt{Tn}+n)}{T}\right)$$

$$\left|\sum_{1\leq k,\ell\leq\kappa} \mathrm{Cov}(\bar{\varphi}_k(X_t)\bar{\varphi}_\ell(X_t), S_{ktT}S_{\ell tT}\mathcal{I}''(Q_{tT}))\right|$$

$$\leq \left|\sum_{1\leq k,\ell\leq\kappa} \mathrm{Cov}[\bar{\varphi}_k(X_t)\bar{\varphi}_\ell(X_t), \breve{S}_{ktT}\breve{S}_{\ell tT}\mathcal{I}''(\breve{Q}_{tT})]\right|$$

$$+ \left|\sum_{1\leq k,\ell\leq\kappa} \mathrm{Cov}[\bar{\varphi}_k(X_t)\bar{\varphi}_\ell(X_t), S_{ktT}S_{\ell tT}(\mathcal{I}''(Q_{tT}) - \mathcal{I}''(\breve{Q}_{tT}))]\right|$$

$$+ 2\left|\sum_{1\leq k,\ell\leq\kappa} \mathrm{Cov}[\bar{\varphi}_k(X_t)\bar{\varphi}_\ell(X_t), \breve{S}_{ktT}(S_{\ell tT} - \breve{S}_{\ell tT})\mathcal{I}''(\breve{Q}_{tT})]\right|$$

$$+ \left|\sum_{1\leq k,\ell\leq\kappa} \mathrm{Cov}[\bar{\varphi}_k(X_t)\bar{\varphi}_\ell(X_t), (S_{\ell tT} - \breve{S}_{\ell tT})(S_{ktT} - \tilde{S}_{ktT})\mathcal{I}''(\breve{Q}_{tT})]\right|$$

$$\leq C\cdot C(\mathcal{I})\cdot \varphi_\infty^5 \kappa^3 (T\alpha_X^{3/4}(n) + \sqrt{nT} + n)\cdot$$

Summing over t gives in (B·12)

$$\sum_{t=1}^{T}\left|\mathbb{E}\left[\frac{d^2\mathcal{J}_{tT}(0;\eta_t)}{d^2z} - \frac{d^2\mathcal{J}_{tT}(0;\tilde{\eta}_t)}{d^2z}\right]\right|$$

$$\leq C\cdot C(\mathcal{I})\cdot \varphi_\infty^3 \sqrt{\kappa}\left(\alpha_X(n) + \kappa\,\frac{(\sqrt{Tn}+n)}{T}\right)$$

$$+ C\cdot C(\mathcal{I})\cdot \varphi_\infty^5 \frac{\kappa^2}{T}(T\alpha_X^{3/4}(n) + \sqrt{nT} + n)$$

$$\leq C\cdot C(\mathcal{I})\cdot \varphi_\infty^5\left[\sqrt{\kappa}\alpha_X(n) + \kappa^2\alpha_X^{3/4}(n) + (\kappa^{3/2} + \kappa^2)\,\frac{(\sqrt{Tn}+n)}{T}\right]$$

$$\leq C\cdot C(\mathcal{I})\cdot \varphi_\infty^5 \kappa^2\left(\alpha_X^{3/4}(n) + \frac{(\sqrt{Tn}+n)}{T}\right)$$

$$\leq C\cdot C(\mathcal{I})\cdot \varphi_\infty^5 \kappa^2\left(n^{-(3/4)(1+a)} + \frac{(\sqrt{Tn}+n)}{T}\right), \tag{B.18}$$

under Assumption M(i). An optimal choice of the order of n in (B.18) is $T^{2/(5+3a)}$, which gives the upper bound $C\cdot C(\mathcal{I})\cdot\varphi_\infty^5\kappa^2 T^{-(3/2)[(1+a)/(5+3a)]}$. Therefore (B.18) and (B.12), (B.15), and (B.13) yield that (B.9) is proved.

Step 2. Proof of Lemma A.3(i). Now choose a three time continuously differentiable $\mathcal{I}_\epsilon(z)$ with $\mathcal{I}_\epsilon(z) = 0$ if $z \le -\epsilon$, $\mathcal{I}_\epsilon(z) = 1$ if $z > 0$. This gives, for any $\gamma \in \mathbb{R}$,

$$\mathbb{I}(z \ge \gamma) \le \mathcal{I}_\epsilon(z - \gamma) \le \mathbb{I}(z \ge \gamma - \epsilon), \tag{B.19}$$

and then, by (B.9),

$$\begin{aligned}\mathbb{P}(Q_T \ge \gamma) &\le \mathbb{E}\mathcal{I}_\epsilon(Q_T - \gamma) \le \mathbb{E}\mathcal{I}_\epsilon(\tilde{Q}_T - \gamma) + C_\epsilon \cdot \varphi_\infty^6 \kappa^2 T^{-(3/2)[(1+a)/(5+3a)]} \qquad \textbf{(B.20)}\\ &\le \mathbb{P}(\tilde{Q}_T \ge \gamma - \epsilon) + o(1),\\ \mathbb{P}(Q_T \ge \gamma) &\ge \mathbb{E}\mathcal{I}_\epsilon(Q_T - \gamma - \epsilon) \ge \mathbb{E}\mathcal{I}_\epsilon(\tilde{Q}_T - \gamma - \epsilon) + o(1)\\ &\ge \mathbb{P}(\tilde{Q}_T \ge \gamma + \epsilon) + o(1).\end{aligned}$$

Note that $\tilde{Q}_T$ is a $(\chi(\kappa) - \kappa)/\sqrt{2\kappa}$ that has a continuous density and converges in distribution to a standard normal if κ goes to infinity. Therefore taking ϵ small enough gives Lemma A.3(i).

Step 3. Proof of Lemma A.3(ii). The proof is done by bounding $\mathbb{E}\mathcal{I}_\epsilon(\tilde{Q}_T - \gamma)$ in (B.20). Observe that $\tilde{Q}_T$ has the same distribution as

$$\frac{1}{\sqrt{\kappa}} \sum_{k=1}^{\kappa} \frac{\zeta_k^2 - 1}{\sqrt{2}},$$

where the ζ_k's are i.i.d. $N(0,1)$ random variables. As established in the proof of Theorem 7.2 of Billingsley (1968) and changing the $(\zeta_k^2 - 1)/\sqrt{2}$ into standard $N(0,1)$ variables, there is a constant C_ϵ with

$$|\mathbb{E}\mathcal{I}_\epsilon(\tilde{Q}_T - \gamma) - \mathbb{E}\mathcal{I}_\epsilon(N(0,1) - \gamma)| \le \frac{C_\epsilon}{\sqrt{\kappa}}.$$

Then (B.19) and (B.20) show

$$\mathbb{P}(Q_T \ge \gamma) \le \mathbb{P}(N(0,1) \ge \gamma - \epsilon) + C_\epsilon[\kappa^2 T^{-(3/2)[(1+a)/(5+3a)]} + (1/\sqrt{\kappa})].$$

Applying the Mill's ratio inequality (see Shorack and Wellner, 1986, p. 850) to $\mathbb{P}(N(0,1) \ge \gamma - \epsilon)$ shows that Lemma A.3(ii) is proved. ■