The Annals of Statistics
2012, Vol. 40, No. 3, 1376–1402
DOI: 10.1214/12-AOS1010

INFERENCE OF TIME-VARYING REGRESSION MODELS[1]

By Ting Zhang and Wei Biao Wu

University of Chicago

We consider parameter estimation, hypothesis testing and variable selection for partially time-varying coefficient models. Our asymptotic theory has the useful feature that it can allow dependent, nonstationary error and covariate processes. With a two-stage method, the parametric component can be estimated with a $n^{1/2}$-convergence rate. A simulation-assisted hypothesis testing procedure is proposed for testing significance and parameter constancy. We further propose an information criterion that can consistently select the true set of significant predictors. Our method is applied to autoregressive models with time-varying coefficients. Simulation results and a real data application are provided.

1. Introduction. Varying coefficient models have been extensively studied in the literature, and they are useful for characterizing nonconstancy relationship between predictors and responses in regression models; see, for example, [19, 20, 28, 29, 33, 44, 52, 55]. In this paper we consider the time-varying coefficient model

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta}_i + e_i, \qquad i = 1, \dots, n, \tag{1.1}$$

where y_i is the response, $\mathbf{x}_i$ is the predictor, $^\top$ is the transpose operator, $\boldsymbol{\beta}_i = \boldsymbol{\beta}(i/n)$ for some smooth function $\boldsymbol{\beta}: [0,1] \to \mathbb{R}^p$ and e_i is the error. We assume that $E(e_i|\mathbf{x}_i) = 0$. Estimation of the coefficient function $\boldsymbol{\beta}(\cdot)$ in model (1.1) has been considered by [6, 41, 45, 46, 56] among others. An important special example of (1.1) is the time-varying autoregressive model [13, 37] by letting $\mathbf{x}_i = (y_{i-1}, \dots, y_{i-p})^\top$. There are important differences between our model (1.1) and those under longitudinal or functional setting. Here we assume that only one realization $(\mathbf{x}_i, y_i)_{i=1}^n$ is available, while in the

Received February 2012; revised April 2012.
[1]Supported in part by NSF Grant DMS-09-06073.
AMS 2000 subject classifications. Primary 62G05, 62G10; secondary 62G20.
Key words and phrases. Information criterion, locally stationary processes, nonparametric hypothesis testings, time-varying coefficient models, variable selection.

longitudinal and functional setting, many subjects are measured at multiple times, so that one has multiple realizations.

A natural problem for model (1.1) is to test whether certain or all components in $\boldsymbol{\beta}_i$ are time-invariant. There is a huge literature on the problem of testing parameter stability; see, for example, [2, 5, 9, 14, 16, 18, 27, 31, 32, 38, 40, 42, 53]. For model (1.1), we are interested in testing

$$H_0 : \mathbf{A}\boldsymbol{\beta}(\cdot) \equiv \mathbf{a} \tag{1.2}$$

for some vector $\mathbf{a} \in \mathbb{R}^s$, where $\mathbf{A} \in \mathbb{R}^{s\times p}$ is a real-valued matrix. With an appropriately chosen $\mathbf{A}$, the null hypothesis (1.2) can be formulated to test whether a certain part of coefficients is zero or time-invariant. In the latter case, $\mathbf{a}$ needs to be replaced by an estimate $\hat{\mathbf{a}}$. Zhou and Wu [56] built simultaneous confidence tubes for the regression coefficient function $\boldsymbol{\beta}(\cdot)$, which can be used as an $\mathcal{L}^\infty$-test for (1.2). The latter test often does not have a good power if the alternative hypothesis consists of nonzero smooth functions. In Section 3.2 we propose a more powerful $\mathcal{L}^2$-test which is based on the weighted integrated squared errors. Our setting is much more general than the one in [8] where $(\mathbf{x}_i, e_i)$ is assumed to be β-mixing and stationary. In comparison, we allow nonstationary predictor and error processes which can be nonstrong mixing; see Section 2 for our nonstationary framework and basic assumptions.

If some of the coefficients are time-invariant, model (1.1) becomes the (semiparametric) partially time-varying coefficient model

$$y_i = \mathbf{x}_{D_1,i}^\top \boldsymbol{\beta}_{D_1} + \mathbf{x}_{D_2,i}^\top \boldsymbol{\beta}_{D_2}(i/n) + e_i, \qquad i = 1, \dots, n, \tag{1.3}$$

where $D_1, D_2 \subseteq D^* = \{1, \dots, p\}$ are groups of parametric and nonparametric components, respectively. Based on an estimate of (1.1), we simply take an integration/average over the parametric part to obtain an estimate of $\boldsymbol{\beta}_{D_1}$ that achieves the $n^{1/2}$-convergence rate. An asymptotic theory is given in Section 3.1. This method was previously used in [55] for estimating state-domain semivarying coefficient models. The latter paper assumed that $y_i = \mathbf{x}_i^\top \boldsymbol{\beta}(\mathbf{u}_i) + e_i$ where $(y_i, \mathbf{x}_i, \mathbf{u}_i)$, $i = 1, \dots, n$, are independent and identically distributed; see [51] for the case with stationary mixing processes. Our time-domain model (1.3) is very general, and it includes both (1.1) and usual linear regression models. Gao and Hawthorne [23] considered a special example of (1.3) with $D_1 = \{2, \dots, p\}$, $D_2 = 1$ and $\mathbf{x}_{D_2,i} \equiv 1$, so that only the intercept term in (1.3) is time-varying.

Section 3.3 deals with the problem of selecting significant predictors. Fan et al. [17] proposed an extended AIC for choosing locally significant variables. Abramovich et al. [1] considered the problem of order selection for time-varying autoregressive models by requiring that multiple realizations are available. Using the dependence framework in Section 2, we are able to

solve this problem under parameter instability and temporal dependence. In particular, we propose an information criterion, consisting of measures of goodness-of-fit and model complexity, that can consistently select the true set of relevant predictors based only on one realization. Section 4 provides simulation studies and an application. Proofs are given in the Appendix.

2. Model assumptions. Since the coefficient function $\boldsymbol{\beta}(\cdot)$ in (1.1) is smooth, we can naturally estimate it (along with its derivative) by

$$(\tilde{\boldsymbol{\beta}}(t), \tilde{\boldsymbol{\beta}}'(t)) = \underset{\boldsymbol{\eta}_0, \boldsymbol{\eta}_1 \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{n} \{y_i - \mathbf{x}_i^\top \boldsymbol{\eta}_0 - \mathbf{x}_i^\top \boldsymbol{\eta}_1 (i/n - t)\}^2 K\left(\frac{i/n - t}{b_n}\right), \tag{2.1}$$

where $K(\cdot)$ is the kernel function, and b_n is a bandwidth sequence satisfying $b_n \to 0$ and $nb_n \to \infty$. Throughout the paper we assume that the kernel function $K(\cdot)$ is a symmetric and bounded function in $\mathcal{C}^1[-1,1]$ with $\int_{-1}^{1} K(v)\,dv = 1$. For example, it can be the Epanechnikov kernel $K(v) = 3\max(0, 1 - v^2)/4$ or the Bartlett kernel $K(v) = \max(0, 1 - |v|)$. Observe that (2.1) has the closed form solution

$$\begin{pmatrix} \tilde{\boldsymbol{\beta}}(t) \\ b_n \tilde{\boldsymbol{\beta}}'(t) \end{pmatrix} = \begin{pmatrix} \mathbf{U}_{n,0}(t) & \mathbf{U}_{n,1}(t) \\ \mathbf{U}_{n,1}(t) & \mathbf{U}_{n,2}(t) \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{V}_{n,0}(t) \\ \mathbf{V}_{n,1}(t) \end{pmatrix} = \mathbf{U}_n(t)^{-1} \mathbf{V}_n(t), \tag{2.2}$$

where for $l \in \{0, 1, 2\}$,

$$\mathbf{U}_{n,l}(t) = (nb_n)^{-1} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^\top \{(i/n - t)/b_n\}^l K\{(i/n - t)/b_n\},$$

with the convention that $0^0 = 1$, and

$$\mathbf{V}_{n,l}(t) = (nb_n)^{-1} \sum_{i=1}^{n} \mathbf{x}_i y_i \{(i/n - t)/b_n\}^l K\{(i/n - t)/b_n\}.$$

To establish an asymptotic theory for $\tilde{\boldsymbol{\beta}}(\cdot)$, we need to impose appropriate regularity conditions on the covariates $(\mathbf{x}_i)$ and errors (e_i). For testing the hypothesis (1.2), [8] assumed that $(\mathbf{x}_i, e_i)$ is β-mixing and stationary. To allow nonstationary predictor and error processes that can be nonstrong mixing, we assume that

$$\mathbf{x}_i = \mathbf{G}(i/n; \boldsymbol{\mathcal{F}}_i) \quad \text{and} \quad e_i = H(i/n; \boldsymbol{\mathcal{F}}_i), \tag{2.3}$$

where $\boldsymbol{\mathcal{F}}_i = (\ldots, \boldsymbol{\varepsilon}_{i-1}, \boldsymbol{\varepsilon}_i)$ is a shift process of independent and identically distributed (i.i.d.) random variables $\boldsymbol{\varepsilon}_k$, $k \in \mathbb{Z}$ and $\mathbf{G}$ and H are measurable functions such that $\mathbf{G}(t; \boldsymbol{\mathcal{F}}_i)$ and $H(t; \boldsymbol{\mathcal{F}}_i)$ are well defined for each $t \in [0, 1]$. This setup is also used in [56].

For a random vector $\mathbf{Z}$, we write $\mathbf{Z} \in \mathcal{L}^q$, $q > 0$, if $\|\mathbf{Z}\|_q = \{E(|\mathbf{Z}|^q)\}^{1/q} < \infty$, where $|\cdot|$ is the Euclidean vector norm, and we denote $\|\cdot\| = \|\cdot\|_2$.

A process $\mathbf{J}(t;\boldsymbol{\mathcal{F}}_k)$ is said to be stochastically Lipschitz continuous ($\mathbf{J}\in\text{Lip}$ in short) if there exists $C>0$, such that $\|\mathbf{J}(t_1;\boldsymbol{\mathcal{F}}_k)-\mathbf{J}(t_2;\boldsymbol{\mathcal{F}}_k)\|\le C|t_1-t_2|$ holds uniformly for all $t_1,t_2\in[0,1]$. Then, under condition (A2) below, (2.3) defines locally stationary processes. Let $\{\boldsymbol{\varepsilon}'_j\}_{j\in\mathbb{Z}}$ be an i.i.d. copy of $\{\boldsymbol{\varepsilon}_j\}_{j\in\mathbb{Z}}$ and $\boldsymbol{\mathcal{F}}_{i,\{0\}}=(\ldots,\boldsymbol{\varepsilon}_{-1},\boldsymbol{\varepsilon}'_0,\boldsymbol{\varepsilon}_1,\ldots,\boldsymbol{\varepsilon}_i)$ be the coupled shift process. We define the functional dependence measure

$$\delta_{k,q}(\mathbf{J})=\sup_{t\in[0,1]}\|\mathbf{J}(t;\boldsymbol{\mathcal{F}}_k)-\mathbf{J}(t;\boldsymbol{\mathcal{F}}_{k,\{0\}})\|_q \quad\text{and}\quad \Theta_{m,q}(\mathbf{J})=\sum_{j=m}^{\infty}\delta_{j,q}(\mathbf{J}).$$

Let $\boldsymbol{\Lambda}(\mathbf{J},t)=\sum_{k\in\mathbb{Z}}\text{cov}\{\mathbf{J}(t;\boldsymbol{\mathcal{F}}_0),\mathbf{J}(t;\boldsymbol{\mathcal{F}}_k)\}$ be the long-run covariance matrix, and $\mathbf{M}(\mathbf{J},t)=E\{\mathbf{J}(t;\boldsymbol{\mathcal{F}}_0)\mathbf{J}(t;\boldsymbol{\mathcal{F}}_0)^\top\}$. Under the short-range dependence condition $\Theta_{0,2}(\mathbf{J})<\infty$, both of them are uniformly bounded over $t\in[0,1]$. Let $\mathbf{L}(t;\boldsymbol{\mathcal{F}}_k)=\mathbf{G}(t;\boldsymbol{\mathcal{F}}_k)H(t;\boldsymbol{\mathcal{F}}_k)$, and we shall make the following assumptions:

(A1) Smoothness: $\boldsymbol{\beta}\in\mathcal{C}^3[0,1]$;
(A2) Local stationarity: $\mathbf{G},\mathbf{L}\in\text{Lip}$;
(A3) Short-range dependence: $\Theta_{0,4}(\mathbf{G})+\Theta_{0,\iota}(\mathbf{L})<\infty$ for some $\iota>2$;
(A4) The smallest eigenvalue of $\mathbf{M}(\mathbf{G},\cdot)$ is bounded away from zero on $[0,1]$.

A sufficient condition for (A3) is that $\Theta_{0,2\iota}(\mathbf{G})+\Theta_{0,2\iota}(H)<\infty$ for some $\iota>2$.

3. Main results.

3.1. *Parameter estimation.* Let $\mathbf{A}$ be a pre-specified matrix and $\mathbf{a}=\int_0^1\mathbf{A}\boldsymbol{\beta}(t)\,dt$. Then

$$\hat{\mathbf{a}}=\int_0^1\mathbf{A}\tilde{\boldsymbol{\beta}}(t)\,dt$$

is an estimate of $\mathbf{a}$. For the partially time-varying coefficient model (1.3), let $\mathbf{A}_{D_1}\in\mathbb{R}^{p_1\times p}$ be a matrix with rows $\{\mathbf{z}_i^\top\}_{i\in D_1}$, where $\mathbf{z}_i\in\mathbb{R}^p$ is the vector with unit ith component and zeros elsewhere, then $\mathbf{A}_{D_1}\boldsymbol{\beta}(t)=\boldsymbol{\beta}_{D_1}(t)$, $t\in[0,1]$. Although $\boldsymbol{\beta}_{D_1}$ can be consistently estimated by $\tilde{\boldsymbol{\beta}}_{D_1}(t)$ for any $t\in(0,1)$, the smoothed estimate

$$\hat{\boldsymbol{\beta}}_{D_1}=\int_0^1\mathbf{A}_{D_1}\tilde{\boldsymbol{\beta}}(t)\,dt$$

can have a better rate of convergence.

THEOREM 3.1. *Assume* (A1)–(A4) *and* $\Theta_{n,\iota}(\mathbf{L})=\mathcal{O}(n^{-\nu})$ *for some* $\nu>1/2-1/\iota$. *Let* $\boldsymbol{\Xi}(t)=\mathbf{M}(\mathbf{G},t)^{-1}\boldsymbol{\Lambda}(\mathbf{L},t)\mathbf{M}(\mathbf{G},t)^{-1}$ *and* $\kappa_2=\int_{-1}^1 v^2K(v)\,dv$.

If $b_n \asymp n^{-c}$ *for some* $1/6 < c < \min\{1/3, 1/2 - 1/(2\iota)\}$, *then*

$$n^{1/2}(\hat{\mathbf{a}} - \mathbf{a} - \boldsymbol{\xi}_n) \Rightarrow N\left\{0, \int_0^1 \mathbf{A}\boldsymbol{\Xi}(t)\mathbf{A}^\top\, dt\right\} \quad \textit{where } \boldsymbol{\xi}_n = \frac{b_n^2\kappa_2}{2}\int_0^1 \mathbf{A}\boldsymbol{\beta}''(t)\, dt.$$

In Theorem 3.1, the term $\boldsymbol{\xi}_n$ can be interpreted as the bias due to nonparametric estimation, and it vanishes under the null hypothesis (1.2). Hence the parametric component $\boldsymbol{\beta}_{D_1}$ in the semi-parametric model (1.3) can have a $n^{1/2}$-consistent estimate $\hat{\boldsymbol{\beta}}_{D_1}$.

3.2. *Hypothesis testing.* For testing the null hypothesis (1.2), let $\mathbf{W}(\cdot)$ be a continuous mapping from $[0,1]$ to symmetric positive-definite matrices in $\mathbb{R}^{s\times s}$. Consider the weighted integrated squared error

$$T_n(\mathbf{A}, \mathbf{a}, \mathbf{W}) = \int_0^1 \{\mathbf{A}\tilde{\boldsymbol{\beta}}(t) - \mathbf{a}\}^\top \mathbf{W}(t)\{\mathbf{A}\tilde{\boldsymbol{\beta}}(t) - \mathbf{a}\}\, dt. \tag{3.1}$$

If $\mathbf{a}$ is unknown, an estimate can be used. For example we can use $\hat{\mathbf{a}} = \int_0^1 \mathbf{A}\tilde{\boldsymbol{\beta}}(t)\, dt$, which has the parametric convergence rate; see Theorem 3.1. Chen and Hong [8] considered the special case that $(\mathbf{x}_i, e_i)$ is a stationary β-mixing process. Their generalized Hausman test [26] relates to (3.1) with $\mathbf{A}$ being the identity matrix and $\mathbf{W}(t) = \mathbf{M}(\mathbf{G}, t)$. Such a choice of weight matrices should be used if we are interested in prediction. Alternatively, we can use $\mathbf{W}(t) \equiv \mathbf{I}_{s\times s}$, the identity matrix to form the integrated squared errors. Let $K_2 = \int_{-1}^1 K(v)^2\, dv$, by Theorem 1 in [56], $\mathbf{A}\tilde{\boldsymbol{\beta}}(t)$ has the asymptotic covariance $(nb_n)^{-1}K_2\mathbf{A}\boldsymbol{\Xi}(t)\mathbf{A}^\top$. Hence, we can choose $\mathbf{W}(t) = \{\mathbf{A}\boldsymbol{\Xi}(t)\mathbf{A}^\top\}^{-1}$ to serve as a normalizer. In this case, (3.1) is (proportionally) an integral of the squared local t-statistics.

For a matrix $\mathbf{A}$, define $\underline{\rho}(\mathbf{A}) = \inf\{|\mathbf{A}\mathbf{v}| : |\mathbf{v}| = 1\}$ and $\overline{\rho}(\mathbf{A}) = \sup\{|\mathbf{A}\mathbf{v}| : |\mathbf{v}| = 1\}$. Let

$$K^*(x) = \int_{-1}^{1-2|x|} K(v)K(v + 2|x|)\, dv,$$

and $K_2^* = \int_{-1}^1 K^*(v)^2\, dv$. Since $K \in \mathcal{K}$, we have $K^* \in \mathcal{C}^1[-1,1]$ and is symmetric. Let

$$\begin{aligned} \boldsymbol{\Xi}_{\mathbf{A},\mathbf{W}}(t) &= \mathbf{W}(t)^{1/2}\mathbf{A}\boldsymbol{\Xi}(t)\mathbf{A}^\top\mathbf{W}(t)^{1/2}, \\ \Xi_{\mathbf{A},\mathbf{W},l} &= \operatorname{tr}\left\{\int_0^1 \boldsymbol{\Xi}_{\mathbf{A},\mathbf{W}}(t)^l\, dt\right\}. \end{aligned} \tag{3.2}$$

Theorem 3.2 provides asymptotic normality for $T_n(\mathbf{A}, \mathbf{a}, \mathbf{W})$.

THEOREM 3.2. *Assume* (A1)–(A4), $\Theta_{0,4}(\mathbf{L}) < \infty$ *and* $\Theta_{n,\iota}(\mathbf{L}) = \mathcal{O}(n^{-\nu})$ *for some* $\nu > 1$. *If* $b_n \asymp n^{-c}$ *for some* $2/11 < c < \min\{1/3, 3/5 - 4/(5\iota)$,

$2-4/\iota\}$, *then*

$$nb_n^{1/2}\{T_n(\mathbf{A},\mathbf{a},\mathbf{W})-(nb_n)^{-1}K^*(0)\Xi_{A,\mathbf{W},1}\}\Rightarrow N(0,4K_2^*\Xi_{\mathbf{A},\mathbf{W},2}). \tag{3.3}$$

If in addition $\hat{\mathbf{a}}=\mathbf{a}+O_p(n^{-1/2})$, *then (3.3) holds for* $T_n(\mathbf{A},\hat{\mathbf{a}},\mathbf{W})$.

Let $\Phi(\cdot)$ be the cumulative standard normal distribution function and $q_{1-\alpha}$ be the corresponding $(1-\alpha)$th quantile. We reject the null hypothesis (1.2) at level α if

$$T_n(\mathbf{A},\hat{\mathbf{a}},\mathbf{W})>b_n^{-1/2}K^*(0)\Xi_{\mathbf{A},\mathbf{W},1}+n^{-1}b_n^{-1/2}(4K_2^*\Xi_{\mathbf{A},\mathbf{W},2})^{1/2}q_{1-\alpha}. \tag{3.4}$$

Let $\mathbf{f}:[0,1]\to\mathbb{R}^s$ be of class $\mathcal{C}^3$, and $\{d_n\}$ be a sequence of nonnegative real numbers. Proposition 3.1 provides the asymptotic power of the test (3.4) under the local alternative

$$\mathbf{A}\boldsymbol{\beta}(t)=\mathbf{a}+d_n\mathbf{f}(t). \tag{3.5}$$

PROPOSITION 3.1. *Assume conditions of Theorem 3.2. If* $nb_n^{1/2}d_n^2\to s>0$, *then the power of the test (3.4) satisfies*

$$\text{Power}\to\Phi\left\{q_\alpha+\frac{s\int_0^1\mathbf{f}(t)^\top\mathbf{W}(t)\mathbf{f}(t)\,dt}{(4K_2^*\Xi_{\mathbf{A},\mathbf{W},2})^{1/2}}\right\}. \tag{3.6}$$

3.3. *Variable selection.* In this section we shall propose an information criterion for time-varying coefficient models that can consistently identify the true set of relevant predictors. Recall that $D^*=\{1,\dots,p\}$ is the whole set of potential predictors, and $\tilde{\boldsymbol{\beta}}(\cdot)$ is the coefficient estimate. Let D_0 be the true set of relevant predictors. For a candidate subset $D\subseteq D^*$, we can compute the variable selection information criterion

$$\begin{aligned}\mathrm{VIC}(D)&=\log\{\mathrm{RSS}(D)\}+\chi_n|D|\\ \text{where } \mathrm{RSS}(D)&=\sum_{i=1}^n\{y_i-\mathbf{x}_{D,i}^\top\tilde{\boldsymbol{\beta}}_D(i/n)\}^2.\end{aligned} \tag{3.7}$$

Here χ_n is a tuning parameter. We select a subset D that minimizes $\mathrm{VIC}(D)$, thus balancing goodess-of-fit and model complexity. Smaller χ_n leads to more predictors, and vice versa. Theorem 3.3 provides theoretical properties of our procedure.

THEOREM 3.3. *Assume* (A1)–(A4), $\Theta_{0,4}(\mathbf{L})+\Theta_{0,2}(H)<\infty$, $\Theta_{n,\iota}(\mathbf{L})=\mathcal{O}(n^{-\nu})$ *for some* $\nu>1/2-1/\iota$. *Let* $\varphi_n=(nb_n)^{-1}\{n^{1/\iota}+(nb_n\log n)^{1/2}\}+b_n^2$ *and* $\rho_n=n^{-1/2}b_n^{-1}+b_n$. *If* $b_n\asymp n^{-c}$ *for some* $0<c<1-1/\iota$,

$$\chi_n\to 0\quad\textit{and}\quad\{\varphi_n(\varphi_n+\rho_n)\}^{-1}\chi_n\to\infty, \tag{3.8}$$

then, for any $D\neq D_0$, $\mathrm{pr}\{\mathrm{VIC}(D)>\mathrm{VIC}(D_0)\}\to 1$.

4. Implementation.

4.1. *Covariance matrix estimation.* Theorems 3.1 and 3.2 both involve unknown quantities depending on the covariance matrices $\mathbf{M}(\mathbf{G},t)$ and $\mathbf{\Lambda}(\mathbf{L},t)$, $t\in[0,1]$. The problem of estimating covariance matrices has been extensively studied; see [3, 36, 39] among others. Let ϖ_n, τ_n and ϱ_n be bandwidth sequences satisfying $\varpi_n\to 0$, $\tau_n\to 0$, $\varrho_n\to 0$ and $n\tau_n\varrho_n\to\infty$. Let $\mathcal{I}_{n,1}=[0,\tau_n\varrho_n]$, $\mathcal{I}_{n,2}=(\tau_n\varrho_n,1-\tau_n\varrho_n)$, $\mathcal{I}_{n,3}=[1-\tau_n\varrho_n,1]$ and

$$\boldsymbol{\lambda}_i(\mathbf{L},\tau_n\varrho_n)=\begin{cases}\mathbf{L}_i\mathbf{L}_i^\top+2\mathbf{L}_i\sum_{j=1}^n\mathbf{L}_j^\top\mathbb{1}_{\{0<j/n-i/n\le\tau_n\varrho_n\}}, & \text{if } i/n\in\mathcal{I}_{n,1};\\ \mathbf{L}_i\sum_{j=1}^n\mathbf{L}_j^\top\mathbb{1}_{\{|i/n-j/n|\le\tau_n\varrho_n\}}, & \text{if } i/n\in\mathcal{I}_{n,2};\\ \mathbf{L}_i\mathbf{L}_i^\top+2\mathbf{L}_i\sum_{j=1}^n\mathbf{L}_j^\top\mathbb{1}_{\{0<i/n-j/n\le\tau_n\varrho_n\}}, & \text{if } i/n\in\mathcal{I}_{n,3}.\end{cases}$$

For $t\in[0,1]$, we estimate $\mathbf{M}(\mathbf{G},t)$ and $\mathbf{\Lambda}(\mathbf{L},t)$, respectively, by

$$\hat{\mathbf{M}}(\mathbf{G},t)=\sum_{i=1}^n\mathbf{x}_i\mathbf{x}_i^\top\omega_{i,\varpi_n}(t)$$

and

$$\hat{\mathbf{\Lambda}}(\mathbf{L},t)=\sum_{i=1}^n\frac{\boldsymbol{\lambda}_i(\mathbf{L},\tau_n\varrho_n)+\boldsymbol{\lambda}_i(\mathbf{L},\tau_n\varrho_n)^\top}{2}\omega_{i,\tau_n}(t),$$

where $\omega_{i,b}(t)=K\{(i/n-t)/b\}\{P_{b,2}(t)-(t-i/n)P_{b,1}(t)\}/\{P_{b,2}(t)P_{b,0}(t)-P_{b,1}(t)^2\}$ are local linear weights with bandwidth b and $P_{b,l}(t)=\sum_{j=1}^n(t-j/n)^lK\{(j/n-t)/b\}$. Proposition 4.1 provides consistency of our covariance matrix estimates.

PROPOSITION 4.1. *Assume* (A2), $\Theta_{0,4}(\mathbf{G})+\Theta_{0,4}(\mathbf{L})<\infty$ *and* $\Theta_{n,2}(\mathbf{L})=O(n^{-\nu})$ *for some* $\nu>0$. *If both* $\mathbf{M}(\mathbf{G},\cdot)$ *and* $\mathbf{\Lambda}(\mathbf{L},\cdot)$ *are in class* $\mathcal{C}^2$, *then*

$$\sup_{t\in[0,1]}\|\hat{\mathbf{M}}(\mathbf{G},t)-\mathbf{M}(\mathbf{G},t)\|=O\{(n\varpi_n)^{-1/2}+\varpi_n^2\},\tag{4.1}$$

and

$$\sup_{t\in[0,1]}\|\hat{\mathbf{\Lambda}}(\mathbf{L},t)-\mathbf{\Lambda}(\mathbf{L},t)\|=O\{\varrho_n^{1/2}+(n\tau_n\varrho_n)^{-\nu}+(\tau_n\varrho_n)^{\nu/(1+\nu)}+\tau_n^2\}.\tag{4.2}$$

The bound in (4.1) is optimized and becomes $O(n^{-2/5})$ if $\varpi_n\asymp n^{-1/5}$. The optimal bound in (4.2) is complicated and it depends on ν, the decay rate

of dependence. In particular, if $\nu \geq 2/3$, then the optimal bound in (4.2) is $O\{n^{-2\nu/(5\nu+2)}\}$ if $\tau_n \asymp n^{-\nu/(5\nu+2)}$ and $\varrho_n \asymp n^{-4\nu/(5\nu+2)}$; otherwise it is $O\{n^{-\nu/(\nu+2)}\}$ if $\tau_n \asymp n^{-(1-\nu)/(\nu+2)}$ and $\varrho_n \asymp n^{-2\nu/(\nu+2)}$, or $\tau_n \asymp n^{-\nu/(2\nu+4)}$ and $\varrho_n \asymp n^{-1/2}$. In computing $\hat{\mathbf{\Lambda}}(\mathbf{L}, t)$, since (e_i) is usually unknown, we shall replace it by $(\tilde{e}_i)$, the estimated local linear residuals.

4.2. *A simulation-assisted testing procedure.* By the sandwich formula, let $\hat{\mathbf{\Xi}}(t) = \hat{\mathbf{M}}(\mathbf{G}, t)^{-1}\hat{\mathbf{\Lambda}}(\tilde{\mathbf{L}}, t)\hat{\mathbf{M}}(\mathbf{G}, t)^{-1}$ and, as in (3.2), correspondingly define $\hat{\mathbf{\Xi}}_{\mathbf{A},\mathbf{W}}(t)$ and $\hat{\Xi}_{\mathbf{A},\mathbf{W},l}$. By (3.4), we reject the null hypothesis (1.2) at level α if

$$\Delta_n(\mathbf{A}, \hat{\mathbf{a}}, \mathbf{W}) = \frac{nb_n^{1/2}\{T_n(\mathbf{A}, \hat{\mathbf{a}}, \mathbf{W}) - (nb_n)^{-1}K^*(0)\hat{\Xi}_{\mathbf{A},\mathbf{W},1}\}}{(4K_2^*\hat{\Xi}_{\mathbf{A},\mathbf{W},2})^{1/2}} > q_{1-\alpha}. \tag{4.3}$$

If $\mathbf{a}$ is known, then in (4.3) we can use $\mathbf{a}$ instead of $\hat{\mathbf{a}}$. The criterion (4.3) usually does not have a good performance because of the slow convergence in (3.3). Note that the statistic $\Delta_n(\mathbf{A}, \hat{\mathbf{a}}, \mathbf{W})$ is asymptotically pivotal, so we propose a simulation-assisted testing procedure that can substantially improve the finite-sample performance. In particular, we generate i.i.d. standard normal random variables y_i°, $i = 1, \ldots, n$, and i.i.d. standard multivariate normal random vectors $\mathbf{x}_i^\circ$, $i = 1, \ldots, n$, that are also independent of (y_k°). We compute the corresponding $\Delta_n^\circ(\mathbf{A}, \hat{\mathbf{a}}, \mathbf{W})$, and repeat this for many times to obtain its empirical quantile $\hat{q}_{1-\alpha}$. We reject the null hypothesis (1.2) at level α if $\Delta_n(\mathbf{A}, \hat{\mathbf{a}}, \mathbf{W}) > \hat{q}_{1-\alpha}$. Our procedure has a similar flavor as the Wilks type of phenomenon discussed in [18]. A major difference is that we allow dependent and nonstationary errors.

4.3. *Bandwidth selection.* Bandwidth selection for nonparametric hypothesis testing is a nontrivial problem, and it has been studied in [15, 21, 22, 30] among many others. As commented by Wang [48], there exists no uniform guidance for an optimal choice. On the positive side, our simulation results in Section 4.6 indicate that the empirical acceptance probabilities are not quite sensitive to the choice of bandwidth. Hence one can simply choose $b_n = n^{-1/5}$ that has the asymptotic mean integrated squared error (AMISE) optimal rate. As an alternative, we consider the generalized cross-validation (GCV) selector by [10], and estimate the covariance matrix $\mathbf{\Gamma}_n = \{E(e_ie_j)\}_{1\leq i,j\leq n}$ to correct for dependence [49]. Specifically, let $\mathbf{Y} = (y_1, \ldots, y_n)^\top$; then for any bandwidth $b \in (0, 1)$, one can write the local linear smoothed fitted values as $\hat{\mathbf{Y}}(b) = \mathbf{H}(b)\mathbf{Y}$, where $\mathbf{H}(b)$ is the corresponding hat matrix. We choose the bandwidth $\tilde{b}_n$ that minimizes

$$\mathrm{GCV}(b) = \frac{n^{-1}\{\hat{\mathbf{Y}}(b) - \mathbf{Y}\}^\top\hat{\mathbf{\Gamma}}_n^{-1}\{\hat{\mathbf{Y}}(b) - \mathbf{Y}\}}{[1 - \mathrm{tr}\{\mathbf{H}(b)\}/n]^2}. \tag{4.4}$$

An estimate of the covariance matrix $\mathbf{\Gamma}_n$ can be obtained by using the banding technique as in [4, 50]. The GCV selector (4.4) works reasonably well in our simulation studies.

We shall now provide data-driven choices of ϖ_n, τ_n and ϱ_n in the estimation of covariance matrices. From the construction in Section 4.1, we truncate the long-run covariance matrix estimate at lag $m_n = n\tau_n\varrho_n$ and, by the proof of Theorem 3.1,

$$n^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n-k}\mathbf{L}_i^\top\mathbf{L}_{i+k} - \operatorname{tr}\left[\int_0^1 \operatorname{cov}\{\mathbf{L}(t;\boldsymbol{\mathcal{F}}_0),\mathbf{L}(t;\boldsymbol{\mathcal{F}}_k)\}\,dt\right]\right) \Rightarrow N(0,\sigma_k^2),$$

where σ_k^2 is the integrated long-run variance of the process $\{\mathbf{L}_i^\top\mathbf{L}_{i+k}\}_{i=1}^{n-k}$. We propose to choose $\hat{m}_n = \max\{k \geq 0 : |n^{-1/2}\sum_{i=1}^{n-k}\mathbf{L}_i^\top\mathbf{L}_{i+k}| > 1.96\sigma_k\}$. Note that the final estimate is a local linear smoother of $\{\boldsymbol{\lambda}_i(\mathbf{L},\hat{m}_n/n) + \boldsymbol{\lambda}_i(\mathbf{L},\hat{m}_n/n)^\top\}/2$, $i = 1,\ldots,n$, and we can apply the GCV method to select $\hat{\tau}_n$. The latter can also be applied to $\mathbf{x}_i\mathbf{x}_i^\top$, $i = 1,\ldots,n$, to select $\tilde{\varpi}_n$, and we take $\hat{\varpi}_n = \max(\tilde{\varpi}_n, n^{-1/5})$ to avoid numerical singularities. These data-driven choices of bandwidths are able to capture dependence and nonstationarity and have a good performance in our simulation studies.

For the information criterion (3.7), if $\iota \geq 5/2$ and $b_n \asymp n^{-1/5}$, then (3.8) becomes

$$\chi_n \to 0 \quad \text{and} \quad \{n^{-3/5}(\log n)^{1/2}\}^{-1}\chi_n \to \infty. \tag{4.5}$$

Note that condition (4.5) is more restrictive than the traditional Bayesian information criterion (BIC) because of parameter instability. Under the latter setting, a heavier penalty on model complexity is usually needed to suppress the over-fitting problem; see [51] for a similar finding on cross-validation methods. As a rule of thumb, we suggest using $\hat{\chi}_n = n^{-2/5}$. This simple choice performs reasonably well as can be seen from our simulation study. The choice of bandwidth becomes further complicated due to model uncertainty. We suggest a two-stage selection procedure: let $\tilde{b}_n$ be the selected bandwidth by GCV with all available predictors, and we use the information criterion (3.7) to select a pilot set of relevant predictors; then we select the bandwidth $\hat{b}_n$ by applying the GCV method to this pilot set.

4.4. *Locally stationary autoregressive processes.* Modeling a nonstationary process by autoregressive models with time-varying coefficients has attracted considerable attention. A traditional approach is to project the coefficient function onto a basis of temporal functions, and estimates the basis coefficients; see, for example, [25, 43, 47]. Other contributions on parameter estimation can be found in [12, 13, 24, 37] among others. Abramovich et al. [1] considered the problem of order selection by requiring multiple realizations. Let $a_1(\cdot),\ldots,a_p(\cdot)$ be continuous functions. We shall prove that the

time-varying autoregressive process

$$y_i = a_1(i/n)y_{i-1} + \cdots + a_p(i/n)y_{i-p} + e_i, \qquad i = 1, \ldots, n, \tag{4.6}$$

has an approximate solution of form (2.3), and the difference is of a negligible order. Hence the results in Section 3 can be directly applied to address the problem of parameter estimation, hypothesis testing and order selection for time-varying autoregressive models.

Recall that $e_i = H(i/n; \boldsymbol{\mathcal{F}}_i)$. Let $\mathbf{x}_i = (y_i, \ldots, y_{i-p+1})^\top$,

$$\mathbf{A}(t) = \begin{pmatrix} a_1(t) & \cdots & a_{p-1}(t) & a_p(t) \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix} \in \mathbb{R}^{p\times p}$$

and

$$\mathbf{H}^\diamond(t; \boldsymbol{\mathcal{F}}_k) = \begin{pmatrix} H(t; \boldsymbol{\mathcal{F}}_k) \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^p.$$

Then (4.6) can be written as

$$\mathbf{x}_i = \mathbf{A}(i/n)\mathbf{x}_{i-1} + \mathbf{H}^\diamond(i/n; \boldsymbol{\mathcal{F}}_i). \tag{4.7}$$

We shall make the following assumptions:

(T1) The starting point $(y_p, \ldots, y_1)^\top \in \mathcal{L}^2$.

(T2) The coefficient functions $a_j(\cdot)$, $j = 1, \ldots, p$, are Lipschitz continuous on $[0, 1]$.

(T3) $\sum_{j=1}^p a_j(t)z^j \neq 1$ for all $|z| \leq 1 + c$ with $c > 0$ uniformly in $t \in [0, 1]$.

Conditions (T2) and (T3) entail local stationarity and short-range dependence, respectively; see also [11]. Proposition 4.2 states that the autoregressive process (4.7) can be approximated by (2.3) with a uniform approximation error of order $O_p(n^{-1})$.

PROPOSITION 4.2. *Assume* (T1)–(T3). *If* $H \in \mathrm{Lip}$, *then there exists a measurable function* $\mathbf{G} \in \mathrm{Lip}$ *and a constant* $C > 0$ *such that*

$$\max_{1\leq i\leq n} \|\mathbf{x}_i - \mathbf{G}(i/n; \boldsymbol{\mathcal{F}}_i)\| \leq Cn^{-1}. \tag{4.8}$$

In proving asymptotic results of Section 3, the key quantity is the partial sum process $\sum_{i=1}^k \mathbf{x}_i\mathbf{x}_i^\top$ and $\sum_{i=1}^k \mathbf{x}_i e_i$, $k = 1, \ldots, n$. By Proposition 4.2, there exists a measurable function $\mathbf{G} \in \mathrm{Lip}$ such that

$$\max_{1\leq k\leq n} \left| \sum_{i=1}^k \{\mathbf{x}_i\mathbf{x}_i^\top - \mathbf{G}(i/n; \boldsymbol{\mathcal{F}}_i)\mathbf{G}(i/n; \boldsymbol{\mathcal{F}}_i)^\top\} \right| = O_p(1)$$

and

$$\max_{1\le k\le n}\left|\sum_{i=1}^{k}\{\mathbf{x}_i-\mathbf{G}(i/n;\boldsymbol{\mathcal{F}}_i)\}e_i\right|=O_p(1).$$

A careful check of the proofs of our results in Section 3 indicates that, due to the above relation, they are still valid for the time-varying autoregressive process (4.6).

4.5. *A comparison with GLRT.* The generalized likelihood ratio test (GLRT, [18]) is a popular method for nonparametric hypothesis testing. It was used by Cai, Fan and Li [7] for testing the coefficient constancy, and generalized by Fan and Huang [16] to semiparametric models. Properties of GLRT have been extensively studied for i.i.d. samples. However, its validity for dependent data is not guaranteed. We shall here briefly review the GLRT and compare it with our method. For the null hypothesis

$$H_0:\boldsymbol{\beta}(\cdot)\equiv\boldsymbol{\beta}$$

for some vector $\boldsymbol{\beta}\in\mathbb{R}^p$, the GLRT statistic is defined as

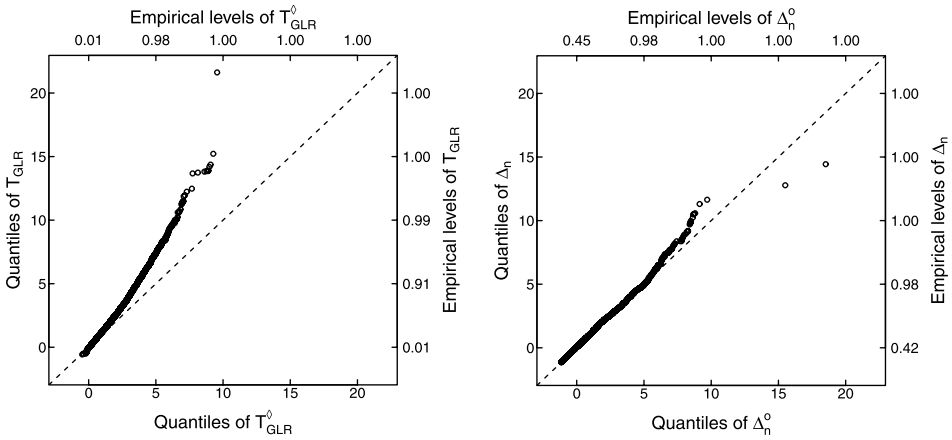
$$T_{\mathrm{GLR}}=\frac{n}{2}\log\frac{\mathrm{RSS}_0}{\mathrm{RSS}_1}=\frac{n}{2}\log\frac{\sum_{i=1}^{n}\{y_i-\mathbf{x}_i^\top\breve{\boldsymbol{\beta}}\}^2}{\sum_{i=1}^{n}\{y_i-\mathbf{x}_i^\top\tilde{\boldsymbol{\beta}}(i/n)\}^2},$$

where $\breve{\boldsymbol{\beta}}$ is the least squares estimate. To construct the null distribution of T_{GLR}, we use the conditional bootstrap as suggested by [7, 16]. Let $\tilde{\sigma}^2=n^{-1}\mathrm{RSS}_1$ and $\{e_i^\diamond\}_{i\in\mathbb{Z}}$ be i.i.d. $N(0,\tilde{\sigma}^2)$. We generate the bootstrap sample $y_i^\diamond=\mathbf{x}_i^\top\breve{\boldsymbol{\beta}}+e_i^\diamond$, $i=1,\dots,n$, and compute the test statistic $T_{\mathrm{GLR}}^\diamond$. We approximate the distribution of T_{GLR} by that of $T_{\mathrm{GLR}}^\diamond$.

Consider the AR-ARCH process with time-varying conditional variance

$$\begin{aligned}y_i&=0.5y_{i-1}+0.25[1+\{1+\exp(3-6i/n)\}^{-1}]e_i,\\ e_i&=(1+0.25e_{i-1}^2)^{1/2}\varepsilon_i,\end{aligned}$$

where $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ are i.i.d. $N(0,1)$. Let $n=500$ and the bandwidth $b_n=n^{-1/5}=0.289$. We consider testing whether the coefficient of y_{i-1} is a constant. For $\Delta_n(\mathbf{A},\hat{\mathbf{a}},\mathbf{W})$, we use the identity weights $\mathbf{W}=\mathbf{I}_{p\times p}$ and obtain its cut-off value by the simulation-assisted procedure in Section 4.2 with 5000 simulated $\Delta_n^\circ(\mathbf{A},\hat{\mathbf{a}},\mathbf{W})$. For T_{GLR}, the cut-off value is obtained by 5000 bootstrapped $T_{\mathrm{GLR}}^\diamond$. We generate 5000 realizations of the AR-ARCH process and use Q–Q plots to examine the performance. The results are presented in Figure 1. It shows that the GLRT fails to provide valid p-values in the presence of dependence and nonstationarity. For example, the empirical acceptance probabilities are 79%, 86.4% and 94.7% for the 90%, 95% and 99% nominal



(a) (b)

FIG. 1. *A comparison of the GLRT* (a) *with our dependence-adjusted testing procedure* (b). *Q–Q plots of* T_{GLR} *against* $T^{\diamond}_{\mathrm{GLR}}$ (a) *and* $\Delta_n(\mathbf{A}, \hat{\mathbf{a}}, \mathbf{W})$ *against* $\Delta^{\circ}_n(\mathbf{A}, \hat{\mathbf{a}}, \mathbf{W})$ (b). *The dashed lines in* (a) *and* (b) *have unit slope and zero intercept.*

levels, respectively. As shown in Figure 1(b), our dependence-adjusted procedure provides a satisfactory approximation of $\Delta_n(\mathbf{A}, \hat{\mathbf{a}}, \mathbf{W})$. At 90%, 95% and 99% nominal levels, our empirical acceptance probabilities are 89.5%, 95.0% and 98.7%, respectively.

4.6. *Simulation studies.* We shall here carry out a simulation study to examine the finite-sample performance of our hypothesis testing procedure in Section 3.2 and the information criterion for variable selection in Section 3.3. Let $P_j(t)$ be the jth order Legendre polynomial and $\mathbf{P}(t) \in \mathbb{R}^{5\times 5}$ be the diagonal matrix with jth diagonal component $P_j(2t-1)/4$. Let $\boldsymbol{\varepsilon}_k = (\varepsilon_{k,1}, \ldots, \varepsilon_{k,6})^{\top}$, $k \in \mathbb{Z}$, be i.i.d. Rademacher random variables and $\mathbf{M}^{\diamond} = (0.2^{|i-j|})_{1\le i,j\le 5}$. Then $\boldsymbol{\xi}_k = \mathbf{M}^{\diamond}(\varepsilon_{k,1}, \ldots, \varepsilon_{k,5})^{\top}$, $k \in \mathbb{Z}$, forms a sequence of independent random vectors with correlated components. Let $\mathbf{x}_i = \sum_{j=0}^{\infty} \mathbf{P}(i/n)^j \boldsymbol{\xi}_{i-j}$ and $e_i = \sum_{j=0}^{\infty} P_6(i/n)^j \varepsilon_{i-j,6}$. Consider:

(i) a linear model with heteroscedastic errors: for $i = 1, \ldots, n$,

$$y_i = (2i/n-1)^2 + 2x_{i,1} + 2\log(i/n+1)x_{i,2} + 0.5(x_{i,2}^2 + x_{i,3}^2)^{1/2} e_i;$$

(ii) a linear model with autoregressive effects: for $i = 1, \ldots, n$,

$$y_i = 0.4\sin(2\pi i/n)y_{i-1} + 0.3x_{i,1} + 0.4(2i/n-1)^3 x_{i,2} + \exp(0.5i/n-2)\varepsilon_{i,6}.$$

Let $n = 500$ and the bandwidth $b = 0.1k$, $k = 1, \ldots, 9$. We use the information criterion (3.7) to estimate the set of relevant predictors. For model (ii), the whole set of potential predictors is taken to be $(\mathbf{x}_i)$ along with three lags of

TABLE 1

Percentages of under-fitted, correctly fitted and over-fitted models selected by the variable selection information criterion (3.7) for $n = 500$. Medians of the selected bandwidths are $\hat{b}_n(\mathrm{i}) = 0.25$ and $\hat{b}_n(\mathrm{ii}) = 0.18$ for models (i) and (ii), respectively, and $c = 2/3$

	VIC		
b	**Under-fitted**	**Correctly fitted**	**Over-fitted**
Model (i)			
0.1	0.0	100.0	0.0
$c\hat{b}_n(\mathrm{i})$	0.0	100.0	0.0
0.2	0.0	100.0	0.0
$\hat{b}_n(\mathrm{i})$	0.0	100.0	0.0
0.3	0.0	100.0	0.0
$\hat{b}_n(\mathrm{i})/c$	0.0	100.0	0.0
0.4	0.0	100.0	0.0
0.5	0.0	100.0	0.0
0.6	0.0	100.0	0.0
0.7	0.0	100.0	0.0
0.8	0.0	100.0	0.0
0.9	0.0	100.0	0.0
Model (ii)			
0.1	0.0	99.8	0.2
$c\hat{b}_n(\mathrm{ii})$	0.0	100.0	0.0
$\hat{b}_n(\mathrm{ii})$	0.0	100.0	0.0
0.2	0.0	100.0	0.0
$\hat{b}_n(\mathrm{ii})/c$	0.0	100.0	0.0
0.3	0.0	100.0	0.0
0.4	0.0	100.0	0.0
0.5	0.0	100.0	0.0
0.6	0.0	100.0	0.0
0.7	0.2	99.8	0.0
0.8	1.1	98.9	0.0
0.9	3.3	96.7	0.0

the response variable. A realization is categorized as under-fitting if we miss at least one relevant predictor, and over-fitting if the selected set contains at least one irrelevant predictor without under-fitting. The results are summarized in Table 1 based on 5000 realizations. Given models (i) and (ii), we use the hypothesis testing procedure to test whether $(x_{i,1})$ has time-invariant contributions. Three types of weight matrices are used: the identity weights $\mathbf{W}_1(t) = \mathbf{I}_{s\times s}$, the normalizer weights $\mathbf{W}_2(t) = \{\mathbf{A}\boldsymbol{\Xi}(t)\mathbf{A}^\top\}^{-1}$ and the prediction weights $\mathbf{W}_3(t) = \mathbf{A}\mathbf{M}(\mathbf{G}, t)\mathbf{A}^\top$. For each configuration, we use the simulation-assisted hypothesis testing procedure in Section 4.2 to obtain cut-off values $\hat{q}_{0.90}$ and $\hat{q}_{0.95}$ with 5000 simulated $\hat{\Delta}_n^\circ(\mathbf{A}, \hat{\mathbf{a}}, \mathbf{W})$. We then generate 5000 realizations of both models (i) and (ii), and calculate the corresponding

TABLE 2

Empirical acceptance probabilities (in percentage) of the simulation-assisted hypothesis testing procedure in Section 4.2 for $n = 500$. Medians of the selected bandwidths are $\hat{b}_n(\mathrm{i}) = 0.25$ and $\hat{b}_n(\mathrm{ii}) = 0.18$ for models (i) and (ii), respectively, and $c = 2/3$

	$\mathbf{W_1}(t)$		$\mathbf{W_2}(t)$		$\mathbf{W_3}(t)$	
b	**90%**	**95%**	**90%**	**95%**	**90%**	**95%**
Model (i)						
0.1	92.7	96.9	93.2	97.1	92.4	96.5
$c\hat{b}_n(\mathrm{i})$	92.2	96.5	92.9	96.7	92.0	96.4
0.2	91.7	96.0	92.4	96.2	91.7	95.9
$\hat{b}_n(\mathrm{i})$	90.9	95.8	91.4	96.4	90.3	95.5
0.3	90.3	95.2	90.9	95.7	90.0	94.8
$\hat{b}_n(\mathrm{i})/c$	90.5	95.1	90.9	95.8	90.3	95.1
0.4	90.9	95.5	91.7	96.1	91.1	95.5
0.5	90.4	94.8	91.3	95.2	90.4	95.0
0.6	90.7	95.3	91.2	95.5	90.7	95.2
0.7	91.6	95.6	91.7	95.7	91.5	95.7
0.8	89.1	94.9	89.0	94.8	89.0	94.8
0.9	89.9	95.0	90.4	95.4	89.8	94.9
Model (ii)						
0.1	93.1	97.0	93.2	97.0	92.9	96.9
$c\hat{b}_n(\mathrm{ii})$	92.6	96.6	92.2	96.1	92.1	96.2
$\hat{b}_n(\mathrm{ii})$	91.1	96.1	91.0	95.5	90.7	95.8
0.2	91.3	96.3	91.2	96.2	91.0	96.0
$\hat{b}_n(\mathrm{ii})/c$	90.9	95.4	90.8	95.7	90.7	95.2
0.3	90.8	95.1	89.9	94.8	90.4	95.1
0.4	90.0	95.1	89.8	94.9	89.9	95.1
0.5	89.5	94.6	88.2	93.9	89.2	94.5
0.6	88.6	93.6	87.7	93.1	88.5	93.6
0.7	88.6	93.8	87.4	93.6	88.1	93.8
0.8	87.4	92.6	87.0	92.2	87.4	92.7
0.9	88.3	94.4	88.1	94.0	88.2	94.3

test statistic $\hat{\Delta}_n(\mathbf{A}, \hat{\mathbf{a}}, \mathbf{W})$. Empirical acceptance probabilities are reported in Table 2.

It can be seen that the empirical acceptance probabilities are fairly close to their nominal levels (90% and 95%), and the information criterion (3.7) performs quite well. In addition, the results are not sensitive to choices of weight matrices and bandwidths. For models (i) and (ii), medians of the selected bandwidths based on the GCV criterion (4.4) are $\hat{b}_n(\mathrm{i}) = 0.25$ and $\hat{b}_n(\mathrm{ii}) = 0.18$, respectively. Observe that for model (i), the performance is also quite satisfactory if we choose bandwidths $0.25c$, 0.25 and $0.25/c$ with $c = 2/3$. A similar claim can be made for model (ii) as well.

4.7. *A real-data example.* We apply our model selection method to the Hong Kong circulatory and respiratory data which contains daily measurements of pollutants and hospital admissions in Hong Kong between January 1, 1994 and December 31, 1995 ($n = 730$). Four pollutants, sulphur dioxide (in $\mu g/m^3$), nitrogen dioxide (in $\mu g/m^3$), dust (in $\mu g/m^3$) and ozone (in $\mu g/m^3$), are considered here. The purpose is to understand the association between daily hospital admission (y_i) and levels of sulphur dioxide ($x_{i,2}$), nitrogen dioxide ($x_{i,3}$), dust ($x_{i,4}$) and ozone ($x_{i,5}$). Figure 2 provides their time series plots. In the analysis, we regularize the data so that each

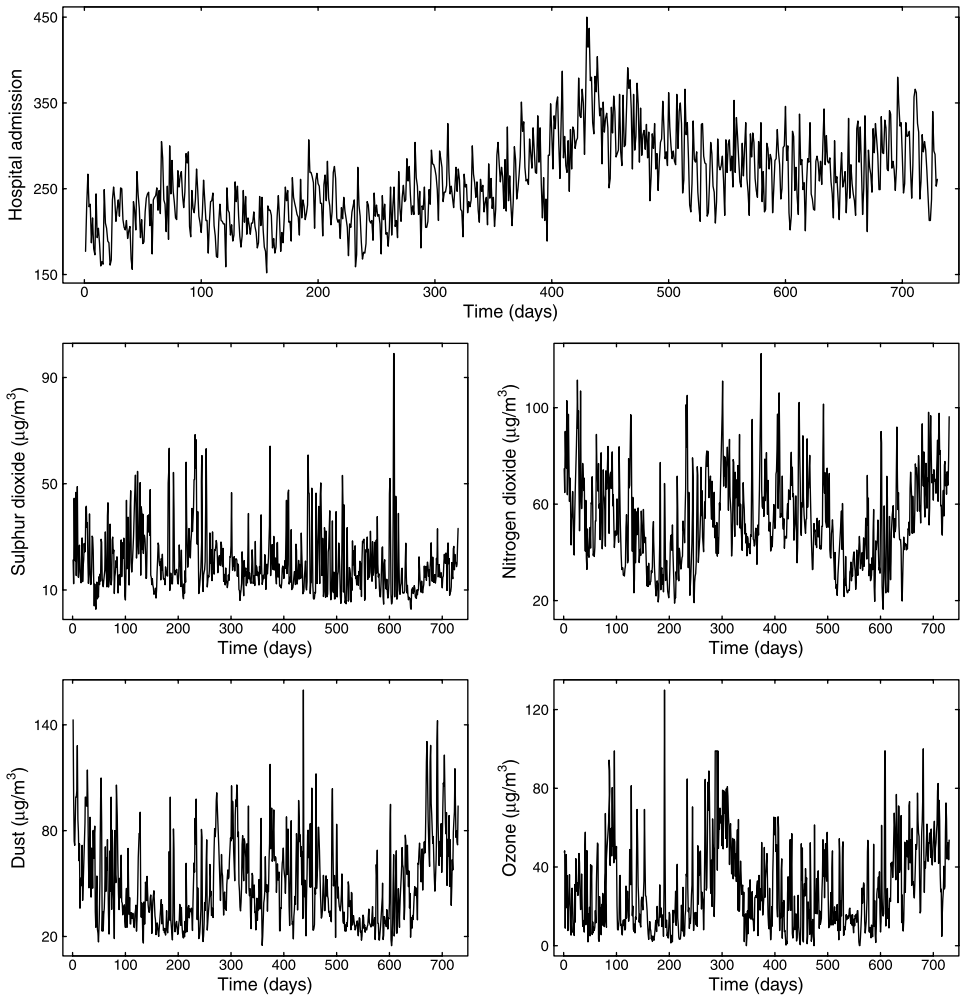


FIG. 2. *Time series plots for daily hospital admission (top) and levels of sulphur dioxide (middle left), nitrogen dioxide (middle right), dust (bottom left) and ozone (bottom right) from January 1, 1994 to December 31, 1995.*

TABLE 3
Summary of test statistics and corresponding p-values for testing parameter constancy with 5000 simulated $\Delta_n^\circ(\mathbf{A}, \hat{\mathbf{a}}, \mathbf{W})$

	$\mathbf{W_1}(t)$		$\mathbf{W_2}(t)$		$\mathbf{W_3}(t)$	
	$\boldsymbol{\Delta_n}$	p-value	$\boldsymbol{\Delta_n}$	p-value	$\boldsymbol{\Delta_n}$	p-value
$\beta_1(\cdot)$	69.77	0.00	120.77	0.02	69.77	0.00
$\beta_3(\cdot)$	6.88	0.14	12.47	0.19	6.85	0.15
$\beta_4(\cdot)$	16.27	0.02	30.13	0.09	23.06	0.01

variable has zero mean and unit variance. Letting $x_{i,1} \equiv 1$ be the intercept, we consider the time-varying coefficient model

$$y_i = \beta_1(i/n) + \sum_{j=2}^{5} \beta_j(i/n) x_{i,j} + e_i, \qquad i = 1, \dots, n. \tag{4.9}$$

The dataset has been studied by [7, 19, 20] by assuming that the observations are i.i.d., while Zhou and Wu [56] found substantial dependence among the fitted residuals. We shall here model the process by (2.3) and apply our model selection method in Section 3. The selected bandwidth and tuning parameter are $\hat{b}_n = 0.13$ and $\hat{\chi}_n = 0.072$, respectively. The information criterion (3.7) selects the intercept ($x_{i,1}$), nitrogen dioxide ($x_{i,3}$) and dust ($x_{i,4}$) as relevant predictors. Fan and Zhang [20] did not consider the ozone effect ($x_{i,5}$) and concluded that sulphur dioxide ($x_{i,2}$) is not statistically significant. We then apply the hypothesis testing procedure in Section 4.2 to examine whether the selected variables really have time-varying contributions. With 5000 simulated $\Delta_n^\circ(\mathbf{A}, \hat{\mathbf{a}}, \mathbf{W})$, the results are summarized in Table 3. Hence, at 10% significance level, we conclude that $\beta_1(\cdot)$ and $\beta_4(\cdot)$ are time-varying while $\beta_3(\cdot)$ can be treated as time-invariant, suggesting the model

$$y_i = \beta_1(i/n) + \beta_3 x_{i,3} + \beta_4(i/n) x_{i,4} + e_i, \qquad i = 1, \dots, n, \tag{4.10}$$

where $\hat{\beta}_3 = \int_0^1 \tilde{\beta}_3(t)\, dt = 0.15$, and $\tilde{\beta}_1(\cdot)$ and $\tilde{\beta}_4(\cdot)$ are plotted in Figure 3.

APPENDIX

For $a, b \in \mathbb{R}$, write $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$. For a matrix $\mathbf{A}$, recall that $\underline{\rho}(\mathbf{A}) = \inf\{|\mathbf{A}\mathbf{v}| : |\mathbf{v}| = 1\}$ and $\overline{\rho}(\mathbf{A}) = \sup\{|\mathbf{A}\mathbf{v}| : |\mathbf{v}| = 1\}$. The proofs of the following two propositions are straightforward, and the details are omitted.

PROPOSITION A.1. *Let* $\mathbf{A} = (a_{ij})_{1 \le i \le I, 1 \le j \le J}$ *be a real matrix. Then* (i) $\max_{i,j} |a_{ij}| \le \overline{\rho}(\mathbf{A}) \le \sqrt{IJ} \max_{i,j} |a_{ij}|$; (ii) *If* $\mathbf{B}$ *has same dimension as* $\mathbf{A}$, *then* $\overline{\rho}(\mathbf{A} + \mathbf{B}) \le \overline{\rho}(\mathbf{A}) + \overline{\rho}(\mathbf{B})$; (iii) *If* $\mathbf{B} = (b_{jk})_{1 \le j \le J, 1 \le k \le K} \in \mathbb{R}^{J \times K}$, *then*

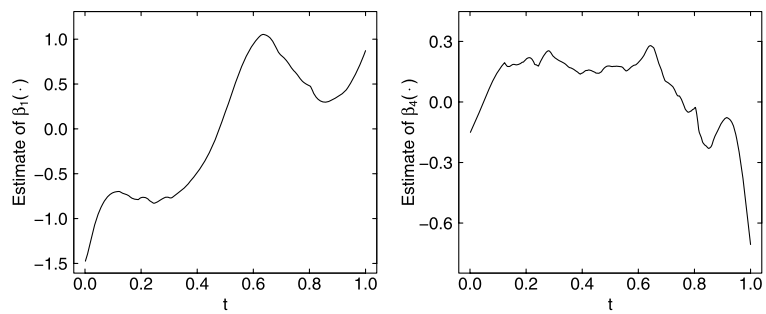


FIG. 3. *Plots of estimated coefficient functions, $\beta_1(\cdot)$ (left) and $\beta_4(\cdot)$ (right).*

$\overline{\rho}(\mathbf{AB}) \le \overline{\rho}(\mathbf{A})\overline{\rho}(\mathbf{B})$ *and* $\underline{\rho}(\mathbf{AB}) \ge \underline{\rho}(\mathbf{A})\underline{\rho}(\mathbf{B})$*; and* (iv) $\overline{\rho}(\mathbf{a}\mathbf{a}^\top) = |\mathbf{a}|^2$ *for any column vector* $\mathbf{a}$.

PROPOSITION A.2. *Assume that* $\mathbf{A}$ *is a nonsingular square matrix and that* $\mathbf{E}$ *is a matrix with the same dimension. If* $\overline{\rho}(\mathbf{A}^{-1}\mathbf{E}) < 1$*, then* $\mathbf{A}+\mathbf{E}$ *is nonsingular and* $\overline{\rho}\{(\mathbf{A}+\mathbf{E})^{-1} - \mathbf{A}^{-1}\} \le \overline{\rho}(\mathbf{E})\overline{\rho}(\mathbf{A}^{-1})^2/\{1-\overline{\rho}(\mathbf{A}^{-1}\mathbf{E})\}$.

Let $\boldsymbol{\mathcal{F}}_{i,j} = (\boldsymbol{\varepsilon}_i, \dots, \boldsymbol{\varepsilon}_j)$, $i \le j$. Define the projection operator

$$\mathcal{P}_k \cdot = E(\cdot|\boldsymbol{\mathcal{F}}_k) - E(\cdot|\boldsymbol{\mathcal{F}}_{k-1}), \qquad k \in \mathbb{Z}.$$

Let $\boldsymbol{\vartheta}_k(t) = \mathbf{J}(t; \boldsymbol{\mathcal{F}}_k)$ be a zero mean process. Write $t_{i,n} = i/n$, $i = 1, \dots, n$. Lemmas A.1 and A.2 provide $\mathcal{L}^q$-bounds for linear and quadratic forms of $\{\boldsymbol{\vartheta}_k(t_{k,n})\}_{k=1}^n$, respectively. To prove Theorem 3.1, we need Lemmas A.1 and A.3.

LEMMA A.1. *Assume* $\Theta_{0,q}(\mathbf{J}) < \infty$, $q > 1$. *Write* $q' = q \wedge 2$. *Let* $\{\mathbf{A}_{k,n}(t)\}_{k=1}^n$, $t \in [0,1]$, *be a sequence of real matrix functions, and define* $\mathbf{S}_n(t) = \sum_{k=1}^n \mathbf{A}_{k,n}(t)\boldsymbol{\vartheta}_k(t_{k,n})$. *Then:*

(i) $\|\mathbf{S}_n(t)\|_q \le C_q[\sum_{k=1}^n |\overline{\rho}\{\mathbf{A}_{k,n}(t)\}|^{q'}]^{1/q'}\Theta_{0,q}(\mathbf{J})$*;*
(ii) $\|\sup_{t\in[0,1]} |\mathbf{S}_n(t)|\|_q \le C_q n^{1/q'} \mathcal{A}_n \Theta_{0,q}(\mathbf{J})$,

where $\mathcal{A}_n = \sup_{t\in[0,1]}[\overline{\rho}\{\mathbf{A}_{1,n}(t)\} + \sum_{k=1}^{n-1}\overline{\rho}\{\mathbf{A}_{k+1,n}(t) - \mathbf{A}_{k,n}(t)\}]$.

PROOF. Let $\mathbf{D}_{k,l,n} = E\{\boldsymbol{\vartheta}_k(t_{k,n})|\boldsymbol{\mathcal{F}}_{k-l,k}\} - E\{\boldsymbol{\vartheta}_k(t_{k,n})|\boldsymbol{\mathcal{F}}_{k-l+1,k}\}$. Then $\boldsymbol{\vartheta}_k(t_{k,n}) - E\boldsymbol{\vartheta}_k(t_{k,n}) = \sum_{l=0}^\infty \mathbf{D}_{k,l,n}$ and $\mathbf{D}_{k,l,n}$, $k = 1, \dots, n$, form martingale differences. By the Burkholder and the Minkowski inequalities, we have

$$\left\|\sum_{k=1}^n \mathbf{A}_{k,n}(t)\mathbf{D}_{k,l,n}\right\|_q^{q'} \le C_q \sum_{k=1}^n |\overline{\rho}\{\mathbf{A}_{k,n}(t)\}|^{q'} \|\mathbf{D}_{k,l,n}\|_q^{q'}.$$

Since $\|\mathbf{D}_{k,l,n}\|_q = \|E\{\boldsymbol{\vartheta}_l(t_{k,n})|\boldsymbol{\mathcal{F}}_{0,l}\} - E\{\boldsymbol{\vartheta}_{l,\{0\}}(t_{k,n})|\boldsymbol{\mathcal{F}}_{0,l}\}\|_q \le \delta_{l,q}(\mathbf{J})$, (i) follows. We now prove (ii). By Doob's inequality and the summation by parts

formula, we have $\|\sup_{t\in[0,1]}|\sum_{k=1}^n \mathbf{A}_{k,n}(t)\mathbf{D}_{k,l,n}|\|_q \le C_q\mathcal{A}_n n^{1/q'}\delta_{l,q}(\mathbf{J})$, entailing (ii). □

LEMMA A.2. *Assume $\Theta_{0,2q}(\mathbf{J})<\infty$, $q\ge 2$. Let $\{\mathbf{Q}_{i,j,n}\}_{1\le i<j\le n}$ be real matrices and $L_n=\sum_{1\le i<j\le n}\boldsymbol{\vartheta}_i(t_{i,n})^\top\mathbf{Q}_{i,j,n}\boldsymbol{\vartheta}_j(t_{j,n})$. Then*

$$\|L_n-E(L_n)\|_q\le C_q n^{1/2}\mathcal{Q}_n\Theta_{0,2q}(\mathbf{J})^2,$$

where $\mathcal{Q}_n^2=(\max_i\sum_{j=i+1}^n|\overline{\rho}(\mathbf{Q}_{i,j,n})|^2)\vee(\max_j\sum_{i=1}^{j-1}|\overline{\rho}(\mathbf{Q}_{i,j,n})|^2)$.

PROOF. Let $\tilde{\boldsymbol{\vartheta}}_k(t)=E\{\boldsymbol{\vartheta}_k(t)|\boldsymbol{\mathcal{F}}_{k-m,k}\}$ be the m-dependent approximated process and $\tilde{L}_n$ be the corresponding quadratic form. If $l>2m$, $\mathcal{P}_{j-l}\{\tilde{\boldsymbol{\vartheta}}_i(t_{i,n})^\top\times Q_{i,j,n}\tilde{\boldsymbol{\vartheta}}_j(t_{j,n})\}=0$. Hence

$$\|\tilde{L}_n-E(\tilde{L}_n)\|_q\le\sum_{l=0}^{2m}\left\|\sum_{j=2}^n\mathcal{P}_{j-l}\sum_{i=1}^{j-1}\tilde{\boldsymbol{\vartheta}}_i(t_{i,n})^\top Q_{i,j,n}\tilde{\boldsymbol{\vartheta}}_j(t_{j,n})\right\|_q,$$

where

$$\left\|\sum_{j=2}^n\mathcal{P}_{j-l}\left(\sum_{i=1}^{j-l-1}+\sum_{i=j-l}^{j-1}\right)\tilde{\boldsymbol{\vartheta}}_i(t_{i,n})^\top Q_{i,j,n}\tilde{\boldsymbol{\vartheta}}_j(t_{j,n})\right\|_q^2\le C_q n\mathcal{Q}_n^2 l^2\Theta_{0,2q}(\mathbf{J})^2.$$

By Lemma A.1 and the arguments of Proposition 1 in [34], we have $\|\{L_n-E(L_n)\}-\{\tilde{L}_n-E(\tilde{L}_n)\}\|_q\le C_q\sqrt{n}\mathcal{Q}_n\Theta_{0,2q}(\mathbf{J})^2$. So Lemma A.2 follows. □

LEMMA A.3. *Assume $\sup_{t\in[0,1]}\|\mathbf{J}(t;\boldsymbol{\mathcal{F}}_0)\|_\iota<\infty$, $\iota>2$, and $\Theta_{n,\iota}(\mathbf{J})=O(n^{-\nu})$ for some $\nu>1/2-1/\iota$. Let $\mathbf{S}_{K,n}(t)=(nb_n)^{-1}\sum_{k=1}^n K\{(k/n-t)/b_n\}\times\boldsymbol{\vartheta}_k(t_{k,n})$. Then*

$$\sup_{b_n\le t\le 1-b_n}|\mathbf{S}_{K,n}(t)|=\frac{O_p(v_n)}{nb_n}\qquad\text{where } v_n=n^{1/\iota}+(nb_n\log n)^{1/2}. \tag{A.1}$$

PROOF. Let $S_n^*=nb_n\sup_{b_n\le t\le 1-b_n}|\mathbf{S}_{K,n}(t)|$. By Theorem 2(ii) in [35], there exist constants $C_1,C_2>0$ such that, for all $\lambda\ge 1$ and l,

$$\begin{aligned}&\operatorname{pr}\left(\max_{0\le j\le nb_n}\left|\sum_{i=l}^{l+j}\boldsymbol{\vartheta}_i(t_{i,n})\right|\ge\lambda v_n\right)\\&\qquad\le C_1\frac{nb_n}{(\lambda v_n)^\iota}+2\exp\{-(\lambda v_n)^2/(nb_nC_2)\}.\end{aligned}\tag{A.2}$$

Note that $[b_n,1-b_n]\subseteq\bigcup_{j\le 1/b_n}[jb_n,(j+1)b_n]$. Using the summation by parts formula, since K has support $[-1,1]$, we have (A.1) in view of (A.2) and

$$\operatorname{pr}(S_n^*\ge\lambda v_n)=O(b_n^{-1})\left[C_1\frac{nb_n}{(\lambda v_n)^\iota}+2\exp\{-(\lambda v_n)^2/(nb_nC_2)\}\right]$$

by choosing a sufficiently large λ. □

For $l \in \{0,1,2\}$, let

$$\mathbf{R}_{n,l}(t) = (nb_n)^{-1} \sum_{i=1}^{n} \mathbf{x}_i e_i \{(i/n - t)/b_n\}^l K\{(i/n - t)/b_n\}.$$

Proof of Theorem 3.1. By Lemma A.1, we have

$$\begin{aligned}
&\int_0^1 \mathbf{A}\mathbf{M}(\mathbf{G}, t)^{-1} \mathbf{R}_{n,0}(t)\, dt \\
&\quad = \frac{1}{nb_n} \sum_{i=1}^{n} \mathbf{A} \left\{ \int_0^1 \mathbf{M}(\mathbf{G}, t)^{-1} K\left(\frac{i/n - t}{b_n}\right) dt \right\} \mathbf{x}_i e_i \\
&\quad = \frac{1}{n} \sum_{i=1}^{n} \mathbf{A}\mathbf{M}(\mathbf{G}, i/n)^{-1} \mathbf{x}_i e_i + O_p\left\{ \frac{(nb_n)^{1/2}}{n} + \frac{b_n n^{1/2}}{n} \right\}.
\end{aligned}$$

By m-dependence approximation, under Conditions (A2), (A3) and (A4), we obtain

$$n^{-1/2} \sum_{i=1}^{n} \mathbf{A}\mathbf{M}(\mathbf{G}, i/n)^{-1} \mathbf{x}_i e_i \Rightarrow N\left\{0, \int_0^1 \mathbf{A}\boldsymbol{\Xi}(t)\mathbf{A}^\top\, dt\right\}.$$

By Lemmas A.1 and A.3, and the argument in the proof of Theorem 3 in [56], we have $\sup_{t\in[0,1]} |\tilde{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)| = O_p(\varphi_n)$ and

$$\begin{aligned}
&\sup_{b_n \le t \le 1-b_n} |\mathbf{M}(\mathbf{G}, t)\{\tilde{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t) - 2^{-1} b_n^2 \kappa_2 \boldsymbol{\beta}''(t)\} - \mathbf{R}_{n,0}(t)| \\
&\qquad = O_p(\varphi_n \rho_n).
\end{aligned} \tag{A.3}$$

Therefore,

$$\hat{\mathbf{a}} - \mathbf{a} - \boldsymbol{\xi}_n = \int_0^1 \mathbf{A}\mathbf{M}(\mathbf{G}, t)^{-1} \mathbf{R}_{n,0}(t)\, dt + O_p(\varphi_n \rho_n + b_n \varphi_n + b_n^3).$$

Under our bandwidth conditions, $\varphi_n \rho_n + b_n \varphi_n + b_n^3 = o(n^{-1/2})$. So Theorem 3.1 follows. □

Let $\gamma_{k,2}(\mathbf{J}) = \sum_{i=0}^{\infty} \delta_{i,2}(\mathbf{J}) \delta_{i+|k|,2}(\mathbf{J})$. Lemma A.4 provides continuity properties of long-run covariance matrices for stochastically Lipschitz continuous processes.

Lemma A.4. *Assume* $\mathbf{J} \in \mathrm{Lip}$ *and* $\Theta_{0,2}(\mathbf{J}) < \infty$. *Then:* (i) *for any nonnegative sequence* $a_n \to 0$, $\sup_{|t_1 - t_2| \le a_n} \overline{\rho}\{\boldsymbol{\Lambda}(\mathbf{J}, t_1) - \boldsymbol{\Lambda}(\mathbf{J}, t_2)\} = o(1)$*;* (ii) *if, in addition,* $\Theta_{n,2}(\mathbf{J}) = O(n^{-\nu})$ *for some* $\nu > 0$, *then* $\sup_{|t_1 - t_2| \le a_n} \overline{\rho}\{\boldsymbol{\Lambda}(\mathbf{J}, t_1) - \boldsymbol{\Lambda}(\mathbf{J}, t_2)\} = O\{a_n^{\nu/(1+\nu)}\}$*; and* (iii) *if* $\inf_{t\in[0,1]} \underline{\rho}\{\boldsymbol{\Lambda}(\mathbf{J}, t)\} > 0$, *then* (i) *and* (ii) *hold for the inverse* $\Lambda^{-1}(\mathbf{J}, t)$.

PROOF. We first observe that

$$\overline{\rho}[E\{\boldsymbol{\vartheta}_i(t_1)\boldsymbol{\vartheta}_j(t_2)^\top\}] \le \sum_{s\in\mathbb{Z}} \|\mathcal{P}_s\boldsymbol{\vartheta}_i(t_1)\|\|\mathcal{P}_s\boldsymbol{\vartheta}_j(t_2)\| \le \sum_{s\in\mathbb{Z}} \delta_{i-s,2}(\mathbf{J})\delta_{j-s,2}(\mathbf{J}) = \gamma_{|j-i|,2}(\mathbf{J}).$$

The Lipschitz continuity implies

$$\overline{\rho}(E[\boldsymbol{\vartheta}_i(t_1)\{\boldsymbol{\vartheta}_j(t_2)-\boldsymbol{\vartheta}_j(t_1)\}^\top]) \le C\{\gamma_{|i-j|,2}(\mathbf{J}) \wedge |t_2-t_1|\} \tag{A.4}$$

uniformly. Hence

$$\sup_{|t_1-t_2|\le a_n} \overline{\rho}\{\boldsymbol{\Lambda}(\mathbf{J},t_1)-\boldsymbol{\Lambda}(\mathbf{J},t_2)\} \le C\sum_{k\in\mathbb{Z}}\{\gamma_{k,2}(\mathbf{J})\wedge a_n\},$$

which entails (i) by the dominated convergence theorem. Let $r_n = a_n^{-1/(1+\nu)}$ which goes to infinity as $n\to\infty$. Since $\sum_{k=l}^\infty \gamma_{k,2}(\mathbf{J}) \le \Theta_{l,2}(\mathbf{J})\Theta_{0,2}(\mathbf{J})$, we have $\sum_{k=0}^\infty\{\gamma_{k,2}(\mathbf{J})\wedge a_n\} = \mathcal{O}(r_na_n + r_n^{-\nu})$, (ii) follows. Then (iii) follows by Proposition A.2. □

Let $\mathbf{W}_0(\cdot)$ be a continuous mapping from $[0,1]$ to symmetric matrices in $\mathbb{R}^{p\times p}$. For $l\in\{1,2\}$, define $\Lambda_{\mathbf{W}_0,l} = \mathrm{tr}[\int_0^1\{\mathbf{W}_0(t)\boldsymbol{\Lambda}(\mathbf{L},t)\}^l\,dt]$. Before we prove Theorem 3.2, we shall first establish a parallel result for

$$T_n^\diamond(\mathbf{W}_0) = \int_0^1 \mathbf{R}_{n,0}(t)^\top\mathbf{W}_0(t)\mathbf{R}_{n,0}(t)\,dt. \tag{A.5}$$

Let $r_{1,n} = (nb_n)^{-1}\sum_{k=0}^\infty\{\gamma_{k,2}(\mathbf{L})\wedge b_n\}$ and $r_{2,n} = (nb_n)^{-1}\sum_{k=0}^\infty[\{k/(nb_n)\}\wedge 1]\gamma_{k,2}(\mathbf{L})$.

LEMMA A.5. *Assume* $\mathbf{L}\in\mathrm{Lip}$ *and* $\Theta_{0,4}(\mathbf{L})<\infty$. *If* $b_n\to 0$ *and* $nb_n^{3/2}\to\infty$, *then*

$$nb_n^{1/2}[T_n^\diamond(\mathbf{W}_0) - E\{T_n^\diamond(\mathbf{W}_0)\}] \Rightarrow N(0, 4K_2^*\Lambda_{\mathbf{W}_0,2}), \tag{A.6}$$

and

$$E\{T_n^\diamond(\mathbf{W}_0)\} = (nb_n)^{-1}K^*(0)\Lambda_{\mathbf{W}_0,1} + O(r_{1,n}+r_{2,n}) + o(n^{-1}b_n^{-1/2}). \tag{A.7}$$

PROOF. Let $\boldsymbol{\zeta}_k(t) = \mathbf{L}(t;\boldsymbol{\mathcal{F}}_k)$ and $\tilde{\boldsymbol{\zeta}}_k(t) = E\{\boldsymbol{\zeta}_k(t)|\boldsymbol{\mathcal{F}}_{k-m,k}\}$ be its m-dependent counterpart and $\boldsymbol{\Lambda}(\tilde{\mathbf{L}},t)$ be the corresponding long-run covariance matrix. Then $\boldsymbol{\Lambda}(\tilde{\mathbf{L}},t)\to\boldsymbol{\Lambda}(\mathbf{L},t)$ uniformly as $m\to\infty$. Let $w_{k,n}(t) = (nb_n)^{-1}\times K\{(k/n-t)/b_n\}$, $k=1,\dots,n$, and $\mathbf{Q}_{i,j,n} = \int_0^1 w_{i,n}(t)\mathbf{W}_0(t)w_{j,n}(t)\,dt$. The central limit theorem (A.6) is a multivariate generalization of Theorem A1 in [54] by using Propositions A.1 and A.2. We shall only detail steps

that require special attention on the dimensionality. Essentially, we need to show that

$$\lim_{m\to\infty}\lim_{n\to\infty}\left[n^2 b_n \sum_{j=2m+1}^{n}\sum_{i=1}^{j-2m} E\{\tilde{\mathbf{D}}_{i,n}^{*\top}\mathbf{Q}_{i,j,n}E(\tilde{\mathbf{D}}_{j,n}^{*}\tilde{\mathbf{D}}_{j,n}^{*\top})\mathbf{Q}_{i,j,n}^{\top}\tilde{\mathbf{D}}_{i,n}^{*}\}\right]$$
$$= K_2^{*}\Lambda_{\mathbf{W}_0,2},$$

where $\tilde{\mathbf{D}}_{k,n}^{*} = \mathcal{P}_k\sum_{l=0}^{\infty}\tilde{\boldsymbol{\zeta}}_{k+l}(t_{k,n})$. Since $E(\tilde{\mathbf{D}}_{k,n}^{*}\tilde{\mathbf{D}}_{k,n}^{*\top}) = \boldsymbol{\Lambda}(\tilde{\mathbf{L}}, t_{k,n})$, by Lemma A.4, we have

$$\sum_{j=2m+1}^{n}\sum_{i=1}^{j-2m} E\{\tilde{\mathbf{D}}_{i,n}^{*\top}\mathbf{Q}_{i,j,n}E(\tilde{\mathbf{D}}_{j,n}^{*}\tilde{\mathbf{D}}_{j,n}^{*\top})\mathbf{Q}_{i,j,n}^{\top}\tilde{\mathbf{D}}_{i,n}^{*}\}$$
$$= \sum_{1\le i<j\le n}\operatorname{tr}[\{\mathbf{W}_0(t_{i,n})\boldsymbol{\Lambda}(\mathbf{L}, t_{i,n})\}^2]\left(\int_0^1 w_{i,n}(t)w_{j,n}(t)\,dt\right)^2$$
$$+ o\{n^2 b_n(n^2 b_n)^{-2}\} + O\{\ell(m)n^2 b_n(n^2 b_n)^{-2}\} + O\{mn(n^2 b_n)^{-2}\}$$

for some function $\ell(m)\to 0$ as $m\to\infty$. Then (A.6) follows. For (A.7), by the proof of Theorem 1 in [54], we have

$$E\{T_n^{\diamond}(\mathbf{W}_0)\} = \sum_{i=1}^{n}\operatorname{tr}\{\mathbf{Q}_{i,i,n}\boldsymbol{\Lambda}(\mathbf{L}, t_{i,n})\} + O(r_{1,n} + r_{2,n} + r_{3,n}),$$

where

$$r_{3,n} = \sum_{i=1}^{n}\left(\sum_{j=-\infty}^{0} + \sum_{j=n+1}^{\infty}\right)\overline{\rho}(Q_{i,j,n})\gamma_{|j-i|,2}(\mathbf{J}) \le Cnb_n(n^2 b_n)^{-1}\Theta_{0,2}(\mathbf{L})^2.$$

Since $\sum_{i=1}^{n}\operatorname{tr}\{\mathbf{Q}_{i,i,n}\boldsymbol{\Lambda}(\mathbf{L}, t_{i,n})\} = (nb_n)^{-1}K^{*}(0)\Lambda_{\mathbf{W}_0,1} + o(n^{-1}b_n^{-1/2})$, (A.7) follows. □

Proof of Theorem 3.2. Let $\mathcal{B}_n = [0, b_n]\cup[1-b_n, 1]$. Lemma A.1 implies $\sup_{t\in\mathcal{B}_n}|\mathbf{R}_{n,0}(t)| = O_p\{(nb_n)^{-1/2}\}$ and $\sup_{t\in\mathcal{B}_n}|\mathbf{U}_n(t) - E\{\mathbf{U}_n(t)\}| = O_p\{(nb_n)^{-1/2}\}$. By the proof of Theorem 1 in [56], we have $\sup_{t\in\mathcal{B}_n}|\tilde{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)| = O_p\{(nb_n)^{-1/2} + b_n^2\}$. Hence,

$$\int_{\mathcal{B}_n}\{\mathbf{A}\tilde{\boldsymbol{\beta}}(t) - \mathbf{a}\}^{\top}\mathbf{W}(t)\{\mathbf{A}\tilde{\boldsymbol{\beta}}(t) - \mathbf{a}\}\,dt = O_p(n^{-1} + b_n^5).$$

By (A.3) and Lemma A.3, we have $\sup_{b_n\le t\le 1-b_n}|\mathbf{A}\tilde{\boldsymbol{\beta}}(t) - \mathbf{a}| = O_p(\varphi_n)$ and

$$\sup_{b_n\le t\le 1-b_n}|\mathbf{A}\tilde{\boldsymbol{\beta}}(t) - \mathbf{a} - \mathbf{A}\mathbf{M}(\mathbf{G}, t)^{-1}\mathbf{R}_{n,0}(t)| = O_p(\varphi_n\rho_n). \tag{A.8}$$

Let $\mathbf{W}_0(t) = \mathbf{M}(\mathbf{G},t)^{-1}\mathbf{A}^\top\mathbf{W}(t)\mathbf{A}\mathbf{M}(\mathbf{G},t)^{-1}$. Since $\sup_{t\in\mathcal{B}_n}|\mathbf{R}_{n,0}(t)| = O_p\{(nb_n)^{-1/2}\}$,

$$\int_{\mathcal{B}_n}\mathbf{R}_{n,0}(t)^\top\mathbf{W}_0(t)\mathbf{R}_{n,0}(t)\,dt = O_p(n^{-1}).$$

Under our bandwidth conditions, $nb_n^{1/2}\varphi_n^2\rho_n = o(1)$. For (3.3), by Lemma A.5, it suffices to show that both $r_{1,n}$ and $r_{2,n}$ are of order $o(n^{-1}b_n^{-1/2})$. By the proof of Lemma A.4, $nb_nr_{1,n} = O\{b_n^{\nu/(1+\nu)}\} = o(b_n^{1/2})$ since $\nu > 1$. Let $r_n = (nb_n)^{1/(2+\nu)}$. Then

$$\sum_{k=0}^{\infty}[\{k/(nb_n)\}\wedge 1]\gamma_{k,2}(\mathbf{L}) \le \frac{r_n(r_n+1)}{2nb_n}\Theta_{0,2}(\mathbf{L})^2 + \Theta_{r_n,2}(\mathbf{L})\Theta_{0,2}(\mathbf{L}) = O(r_n^{-\nu}).$$

Hence, we have $nb_nr_{2,n} = O\{(nb_n)^{-\nu/(2+\nu)}\} = o(b_n^{1/2})$ since $\nu > 1$, (3.3) follows. Note that

$$T_n(\mathbf{A},\hat{\mathbf{a}},\mathbf{W}) - T_n(\mathbf{A},\mathbf{a},\mathbf{W}) = I_n - 2II_n,$$

where $I_n = \int_0^1(\hat{\mathbf{a}}-\mathbf{a})^\top\mathbf{W}(t)(\hat{\mathbf{a}}-\mathbf{a})\,dt = O_p(n^{-1})$, and by (A.8) and Lemmas A.1 and A.3,

$$\begin{aligned} II_n &= \int_0^1(\hat{\mathbf{a}}-\mathbf{a})^\top\mathbf{W}(t)\{\mathbf{A}\tilde{\boldsymbol{\beta}}(t)-\mathbf{a}\}\,dt \\ &= (\hat{\mathbf{a}}-\mathbf{a})^\top\left\{\int_{b_n}^{1-b_n}\mathbf{W}(t)\mathbf{A}\mathbf{M}(\mathbf{G},t)^{-1}\mathbf{R}_{n,0}(t)\,dt + O_p(\varphi_n\rho_n + b_n\varphi_n)\right\} \\ &= O_p\{(n^{-1/2}(n^{-1/2}+\varphi_n\rho_n)\}. \end{aligned}$$

Note that $(nb_n)^{1/2}\varphi_n\rho_n = o(1)$, and Theorem 3.2 follows. □

Proof of Proposition 3.1. Under the local alternative (3.5), we have $\mathbf{A}\boldsymbol{\beta}''(t) = d_n\mathbf{f}''(t)$ and

$$T_n(\mathbf{A},\mathbf{a},\mathbf{W}) - T_n^\diamond(\mathbf{W}_0) = d_n^2\int_0^1\mathbf{f}(t)^\top\mathbf{W}(t)\mathbf{f}(t)\,dt + I_n + 2II_n,$$

where by (A.3) and Lemmas A.1 and A.3,

$$\begin{aligned} I_n &= \int_0^1\{\mathbf{A}\tilde{\boldsymbol{\beta}}(t)-\mathbf{A}\boldsymbol{\beta}(t)\}^\top\mathbf{W}(t)\{\mathbf{A}\tilde{\boldsymbol{\beta}}(t)-\mathbf{A}\boldsymbol{\beta}(t)\}\,dt - T_n^\diamond(\mathbf{W}_0) \\ &= O_p\{(\varphi_n+d_nb_n^2)(\varphi_n\rho_n+d_nb_n^2)+(\varphi_n\rho_n+d_nb_n^2)\varphi_n\}, \end{aligned}$$

the weight matrix $\mathbf{W}_0(t) = \mathbf{M}(\mathbf{G},t)^{-1}\mathbf{A}^\top\mathbf{W}(t)\mathbf{A}\mathbf{M}(\mathbf{G},t)^{-1}$ and

$$II_n = d_n\int_0^1\mathbf{f}(t)^\top\mathbf{W}(t)\{\mathbf{A}\tilde{\boldsymbol{\beta}}(t)-\mathbf{A}\boldsymbol{\beta}(t)\}\,dt = O_p\{d_n(\varphi_n\rho_n+n^{-1/2}+d_nb_n^2)\}.$$

Since $nb_n^{1/2}\varphi_n^2\rho_n = o(1)$, (3.6) follows from Lemma A.5. □

Recall that $D \subseteq D^*$ is a subset with complement $\bar{D}$, and D_0 is the true set of relevant predictors. Let $\tilde{e}_{D,i} = y_i - \mathbf{x}_{D,i}^\top \tilde{\boldsymbol{\beta}}(i/n)$, $1 \le i \le n$. Then $\mathrm{RSS}(D) = \sum_{i=1}^n \tilde{e}_{D,i}^2$. Lemma A.6 provides bounds for $\mathrm{RSS}(D) - \sum_{i=1}^n e_i^2$ for both cases $D_0 \subseteq D$ and $D_0 \not\subseteq D$.

LEMMA A.6. *Assume* (A1)–(A4), $\Theta_{0,4}(\mathbf{L}) < \infty$, $\Theta_{n,\iota}(\mathbf{L}) = \mathcal{O}(n^{-\nu})$ *for some* $\nu > 1/2 - 1/\iota$, $b_n \to 0$ *and* $nb_n \to \infty$. *Then* (i) *if* $D_0 \subseteq D$, *then*

$$\mathrm{RSS}(D) = \sum_{i=1}^n e_i^2 + O_p\{n\varphi_n(\varphi_n + \rho_n)\};$$

and (ii) *if* $D_0 \not\subseteq D$, *then*

$$\mathrm{RSS}(D) = \sum_{i=1}^n e_i^2 + \sum_{i=1}^n \boldsymbol{\beta}_{\bar{D}}(i/n)^\top E\{\mathbf{x}_{\bar{D},i}\mathbf{x}_{\bar{D},i}^\top\}\boldsymbol{\beta}_{\bar{D}}(i/n) + O_p\{n^{1/2} + n\varphi_n(\varphi_n + \rho_n)\}.$$

PROOF. For (i), since $D_0 \subseteq D$, we have $\tilde{e}_{D,i} = e_i + \mathbf{x}_{D,i}^\top\{\boldsymbol{\beta}_D(i/n) - \tilde{\boldsymbol{\beta}}_D(i/n)\}$ and

$$\mathrm{RSS}(D) = \sum_{i=1}^n \tilde{e}_{D,i}^2 = \sum_{i=1}^n e_i^2 + I_n - 2II_n,$$

where, by the proof of Theorem 3.1, $I_n = \sum_{i=1}^n [\mathbf{x}_{D,i}^\top\{\tilde{\boldsymbol{\beta}}_D(i/n) - \boldsymbol{\beta}_D(i/n)\}]^2 = O_p(n\varphi_n^2)$ and, by (A.3) and Lemmas A.1 and A.3,

$$\begin{aligned} II_n &= \sum_{i=1}^n (\mathbf{x}_i e_i)^\top \mathbf{A}_D^\top \mathbf{A}_D\{\tilde{\boldsymbol{\beta}}(i/n) - \boldsymbol{\beta}(i/n)\} \\ &= \sum_{i=1}^n (\mathbf{x}_i e_i)^\top \mathbf{A}_D^\top \mathbf{A}_D \mathbf{M}(\mathbf{G}, i/n)^{-1}\mathbf{R}_{n,0}(i/n) \\ &\quad + O_p(nb_n\varphi_n + n\varphi_n\rho_n + n^{1/2}b_n^2). \end{aligned}$$

Since, by Lemma A.2,

$$\begin{aligned} &\frac{1}{nb_n}\sum_{i=1}^n\sum_{j=1}^n (\mathbf{x}_i e_i)^\top \mathbf{A}_D^\top \mathbf{A}_D \mathbf{M}(\mathbf{G}, i/n)^{-1}(\mathbf{x}_j e_j) K\left(\frac{j/n - i/n}{b_n}\right) \\ &\quad = O_p(b_n^{-1/2} + b_n^{-1}), \end{aligned}$$

we have $II_n = O_p(b_n^{-1} + n\varphi_n\rho_n + n^{1/2}b_n^2)$. Since $b_n^{-1} + n^{1/2}b_n^2 = o\{n\varphi_n(\varphi_n + \rho_n)\}$, (i) follows. For (ii), since $\tilde{e}_{D,i} = e_i + \mathbf{x}_{D,i}^\top\{\boldsymbol{\beta}_D(i/n) - \tilde{\boldsymbol{\beta}}_D(i/n)\} +$

$\mathbf{x}_{\bar{D},i}^\top \boldsymbol{\beta}_{\bar{D}}(i/n)$, we have

$$\mathrm{RSS}(D) = \sum_{i=1}^n e_i^2 + I_n^\circ + 2II_n^\circ + III_n^\circ,$$

where, by (i),

$$I_n^\circ = \sum_{i=1}^n [e_i + \mathbf{x}_{D,i}^\top \{\boldsymbol{\beta}_D(i/n) - \tilde{\boldsymbol{\beta}}_D(i/n)\}]^2 - \sum_{i=1}^n e_i^2 = O_p\{n\varphi_n(\varphi_n + \rho_n)\}$$

and, by Lemma A.1 and the argument on the quantity II_n in (i),

$$\begin{aligned} II_n^\circ &= \sum_{i=1}^n [e_i + \mathbf{x}_{D,i}^\top \{\boldsymbol{\beta}_D(i/n) - \tilde{\boldsymbol{\beta}}_D(i/n)\}]\mathbf{x}_{\bar{D},i}^\top \boldsymbol{\beta}_{\bar{D}}(i/n) \\ &= O_p(n^{1/2} + b_n^{-1} + n\varphi_n\rho_n + n^{1/2}b_n^2). \end{aligned}$$

In addition, by Lemma A.1,

$$III_n^\circ = \sum_{i=1}^n \{\mathbf{x}_{\bar{D},i}^\top \boldsymbol{\beta}_{\bar{D}}(i/n)\}^2 = \sum_{i=1}^n \boldsymbol{\beta}_{\bar{D}}(i/n)^\top E\{\mathbf{x}_{\bar{D},i}\mathbf{x}_{\bar{D},i}^\top\}\boldsymbol{\beta}_{\bar{D}}(i/n) + O_p(n^{1/2}),$$

Lemma A.6 follows. □

Proof of Theorem 3.3. By Lemma A.1, $\sum_{i=1}^n (e_i^2 - Ee_i^2) = O_p(n^{1/2})$. Lemma A.6 implies

$$\log\{\mathrm{RSS}(D)\} = \log\left(\sum_{i=1}^n e_i^2\right) + O_p\{\varphi_n(\varphi_n + \rho_n)\}$$

for $D_0 \subseteq D$, and

$$\log\{\mathrm{RSS}(D)\} = \log\left[\sum_{i=1}^n e_i^2 + \sum_{i=1}^n \boldsymbol{\beta}_{\bar{D}}(i/n)^\top E\{\mathbf{x}_{\bar{D},i}\mathbf{x}_{\bar{D},i}^\top\}\boldsymbol{\beta}_{\bar{D}}(i/n)\right] + o_p(1)$$

for $D_0 \not\subseteq D$. Since $\chi_n = o(1)$ and $\varphi_n(\varphi_n + \rho_n) = o(\chi_n)$, Theorem 3.3 follows. □

Proof of Proposition 4.1. By Lemma A.1,

$$\sup_{t\in[0,1]} \|\hat{\mathbf{M}}(\mathbf{G}, t) - E\{\hat{\mathbf{M}}(\mathbf{G}, t)\}\| = O\{(n\varpi_n)^{-1/2}\},$$

and, by Lemma A.2,

$$\sup_{t\in[0,1]} \|\hat{\boldsymbol{\Lambda}}(\mathbf{L}, t) - E\{\hat{\boldsymbol{\Lambda}}(\mathbf{L}, t)\}\| = O(\varrho_n^{1/2}).$$

By (A.4), we have

$$\begin{aligned}
&\max_{1\le i\le n}|E\{\boldsymbol{\lambda}_i(\mathbf{L},\tau_n\varrho_n)\}-\boldsymbol{\Lambda}(\mathbf{L},t_{i,n})| \\
&\quad\le C\sum_{k=0}^{\infty}\{\gamma_k(\mathbf{L})\wedge(\tau_n\varrho_n)\}+\Theta_{n\tau_n\varrho_n,2}(\mathbf{L}) \\
&\quad= O\{(\tau_n\varrho_n)^{\nu/(1+\nu)}+(n\tau_n\varrho_n)^{-\nu}\}.
\end{aligned}$$

Proposition 4.1 follows by properties of local linear estimates. □

Proof of Proposition 4.2. Consider the process $\{\mathbf{z}_{t,i}\}_{i\in\mathbb{Z}}$ that satisfies the recursion

$$\mathbf{z}_{t,i}=\mathbf{A}(t)\mathbf{z}_{t,i-1}+\mathbf{H}^{\diamond}(t;\boldsymbol{\mathcal{F}}_i),\qquad i\in\mathbb{Z}.$$

Then, for each $t\in[0,1]$, the process $\{\mathbf{z}_{t,i}\}_{i\in\mathbb{Z}}$ is stationary, and there exists a measurable function $\mathbf{G}$ such that $\mathbf{z}_{t,i}=\mathbf{G}(t;\boldsymbol{\mathcal{F}}_i)$, $i\in\mathbb{Z}$. By condition (T3), $\rho_{\mathbf{A}}=\sup_{t\in[0,1]}\overline{\rho}\{\mathbf{A}(t)\}<1$. Hence, by condition (T2) and induction, we have

$$\begin{aligned}
\max_{1\le i\le n}\|\mathbf{x}_i-\mathbf{z}_{i/n,i}\|&\le\rho_{\mathbf{A}}^k\max_{1\le i\le n}\|\mathbf{x}_{i-k}-\mathbf{z}_{i/n,i-k}\| \\
&\quad+C\sum_{j=1}^{k-1}\frac{j\rho_{\mathbf{A}}^j}{n},\qquad k\ge 2.
\end{aligned}\tag{A.9}$$

Since $\rho_{\mathbf{A}}<1$ and $\sum_{j=1}^{\infty}j\rho_{\mathbf{A}}^j<\infty$, (4.8) follows by letting $k\to\infty$. It suffices to show that $\mathbf{G}\in\mathrm{Lip}$. For this, by a similar argument of (A.9), we have for any $k\ge 2$,

$$\sup_{t_1,t_2\in[0,1]}\|\mathbf{z}_{t_1,i}-\mathbf{z}_{t_2,i}\|\le\rho_{\mathbf{A}}^k\sup_{t_1,t_2\in[0,1]}\|\mathbf{z}_{t_1,i-k}-\mathbf{z}_{t_2,i-k}\|+C|t_1-t_2|\sum_{j=0}^{k-1}\rho_{\mathbf{A}}^j.$$

Since $\rho_{\mathbf{A}}<1$ and $\sum_{j=1}^{\infty}\rho_{\mathbf{A}}^j<\infty$, Proposition 4.2 follows by letting $k\to\infty$. □

Acknowledgments. We are grateful to the Editor, an Associate Editor, and two anonymous referees for their helpful comments and suggestions.

REFERENCES

[1] Abramovich, Y. I., Spencer, N. K. and Turley, M. D. E. (2007). Order estimation and discrimination between stationary and time-varying (TVAR) autoregressive models. *IEEE Trans. Signal Process.* **55** 2861–2876. MR2473559

[2] Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica* **61** 821–856. MR1231678

[3] Andrews, D. W. K. and Monahan, J. C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* **60** 953–966. MR1168742

[4] Bickel, P. J. and Levina, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist.* **36** 199–227. MR2387969

[5] Brown, R. L., Durbin, J. and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *J. Roy. Statist. Soc. Ser. B* **37** 149–192. MR0378310

[6] Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *J. Econometrics* **136** 163–188. MR2328589

[7] Cai, Z., Fan, J. and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *J. Amer. Statist. Assoc.* **95** 888–902. MR1804446

[8] Chen, B. and Hong, Y. (2012). Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica* **80** 1157–1183.

[9] Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica* **28** 591–605. MR0141193

[10] Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions. Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.* **31** 377–403. MR0516581

[11] Dahlhaus, R. (1996). On the Kullback–Leibler information divergence of locally stationary processes. *Stochastic Process. Appl.* **62** 139–168. MR1388767

[12] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37. MR1429916

[13] Dahlhaus, R., Neumann, M. H. and von Sachs, R. (1999). Nonlinear wavelet estimation of time-varying autoregressive processes. *Bernoulli* **5** 873–906. MR1715443

[14] Davis, R. A., Huang, D. W. and Yao, Y.-C. (1995). Testing for a change in the parameter values and order of an autoregressive model. *Ann. Statist.* **23** 282–304. MR1331669

[15] Dette, H. and Spreckelsen, I. (2004). Some comments on specification tests in nonparametric absolutely regular processes. *J. Time Series Anal.* **25** 159–172. MR2045571

[16] Fan, J. and Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli* **11** 1031–1057. MR2189080

[17] Fan, J., Yao, Q. and Cai, Z. (2003). Adaptive varying-coefficient linear models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **65** 57–80. MR1959093

[18] Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29** 153–193. MR1833962

[19] Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *Ann. Statist.* **27** 1491–1518. MR1742497

[20] Fan, J. and Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scand. J. Stat.* **27** 715–731. MR1804172

[21] Fan, Y. and Linton, O. (2003). Some higher-order theory for a consistent nonparametric model specification test. *J. Statist. Plann. Inference* **109** 125–154. MR1946644

[22] Gao, J. and Gijbels, I. (2008). Bandwidth selection in nonparametric kernel testing. *J. Amer. Statist. Assoc.* **103** 1584–1594. MR2504206

[23] Gao, J. and Hawthorne, K. (2006). Semiparametric estimation and testing of the trend of temperature series. *Econom. J.* **9** 332–355. MR2324973

[24] Gençağa, D., Kuruoğlu, E. E., Ertüzün, A. and Yildirim, S. (2008). Estimation of time-varying AR $S\alpha S$ processes using Gibbs sampling. *Signal Processing* **88** 2564–2572.

[25] Grenier, Y. (1983). Time-dependent ARMA modelling of nonstationary signals. *IEEE Trans. Acoust. Speech* **31** 899–911.

[26] Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica* **46** 1251–1271. MR0513692

[27] He, C., Teräsvirta, T. and González, A. (2009). Testing parameter constancy in stationary vector autoregressive models against continuous change. *Econometric Rev.* **28** 225–245. MR2655626

[28] Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85** 809–822. MR1666699

[29] Huang, J. Z., Wu, C. O. and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statist. Sinica* **14** 763–788. MR2087972

[30] Kulasekera, K. B. and Wang, J. (1997). Smoothing parameter selection for power optimality in testing of regression curves. *J. Amer. Statist. Assoc.* **92** 500–511. MR1467844

[31] Leybourne, S. J. and McCabe, B. P. M. (1989). On the distribution of some test statistics for coefficient constancy. *Biometrika* **76** 169–177. MR0991435

[32] Lin, C. F. J. and Teräsvirta, T. (1994). Testing the constancy of regression parameters against continuous structural change. *J. Econometrics* **62** 211–228.

[33] Lin, D. Y. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data. *J. Amer. Statist. Assoc.* **96** 103–126. MR1952726

[34] Liu, W. and Wu, W. B. (2010). Asymptotics of spectral density estimates. *Econometric Theory* **26** 1218–1245. MR2660298

[35] Liu, W., Xiao, H. and Wu, W. B. (2012). Probability and moment inequalities under dependence. Working paper.

[36] Lumley, T. and Heagerty, P. (1999). Weighted empirical adaptive variance estimators for correlated data regression. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61** 459–477. MR1680302

[37] Moulines, E., Priouret, P. and Roueff, F. (2005). On recursive estimation for time varying autoregressive processes. *Ann. Statist.* **33** 2610–2654. MR2253097

[38] Nabeya, S. and Tanaka, K. (1988). Asymptotic theory of a test for the constancy of regression coefficients against the random walk alternative. *Ann. Statist.* **16** 218–235. MR0924867

[39] Newey, W. K. and West, K. D. (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55** 703–708. MR0890864

[40] Nyblom, J. (1989). Testing for the constancy of parameters over time. *J. Amer. Statist. Assoc.* **84** 223–230. MR0999682

[41] Orbe, S., Ferreira, E. and Rodriguez-Poo, J. (2005). Nonparametric estimation of time varying parameters under shape restrictions. *J. Econometrics* **126** 53–77. MR2118278

[42] Ploberger, W., Krämer, W. and Kontrus, K. (1989). A new test for structural stability in the linear regression model. *J. Econometrics* **40** 307–318. MR0994952

[43] Rajan, J. J., Rayner, P. J. W. and Godsill, S. J. (1997). Bayesian approach to parameter estimation and interpolation of time-varying autoregressive processes using the Gibbs sampler. *IEE P-Vis. Image Sign.* **144** 249–256.

[44] Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, 2nd ed. Springer, New York. MR2168993

[45] Robinson, P. M. (1989). Nonparametric estimation of time-varying parameters. In *Statistical Analysis and Forecasting of Economic Structural Change* (P. Hackl, ed.) 253–264. Springer, Berlin.

[46] Robinson, P. M. (1991). Time-varying nonlinear regression. In *Economic Structure Change Analysis and Forecasting* (P. Hackl and A. H. Westland, eds.) 179–190. Springer, Berlin.

[47] Subba Rao, T. (1970). The fitting of non-stationary time-series models with time-dependent parameters. *J. Roy. Statist. Soc. Ser. B* **32** 312–322. MR0269065

[48] Wang, L. (2008). Nonparametric test for checking lack of fit of the quantile regression model under random censoring. *Canad. J. Statist.* **36** 321–336. MR2431684

[49] Wang, Y. D. (1998). Smoothing spline models with correlated random errors. *J. Amer. Statist. Assoc.* **93** 341–348.

[50] Wu, W. B. and Pourahmadi, M. (2009). Banding sample autocovariance matrices of stationary processes. *Statist. Sinica* **19** 1755–1768. MR2589209

[51] Xia, Y., Zhang, W. and Tong, H. (2004). Efficient estimation for semivarying-coefficient models. *Biometrika* **91** 661–681. MR2090629

[52] Zeger, S. L. and Diggle, P. J. (1994). Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* **50** 689–699.

[53] Zhang, C. and Dette, H. (2004). A power comparison between nonparametric regression tests. *Statist. Probab. Lett.* **66** 289–301. MR2045474

[54] Zhang, T. and Wu, W. B. (2011). Testing parametric assumptions of trends of a nonstationary time series. *Biometrika* **98** 599–614. MR2836409

[55] Zhang, W., Lee, S.-Y. and Song, X. (2002). Local polynomial fitting in semivarying coefficient model. *J. Multivariate Anal.* **82** 166–188. MR1918619

[56] Zhou, Z. and Wu, W. B. (2010). Simultaneous inference of linear models with time varying coefficients. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 513–531. MR2758526

Department of Statistics
and Actuarial Science
University of Iowa
Iowa City, Iowa 52242-1409
USA
E-mail: ting-zhang@uiowa.edu

Department of Statistics
University of Chicago
5734 S. University Ave.
Chicago, Illinois 60637
USA
E-mail: wbwu@galton.uchicago.edu

Journal: BIOMETRIKA

Article Id: ASR017

Article Title: Testing parametric assumptions of trends of nonstationary time series

First Author: T. Zhang

Corr. Author: T. Zhang

AUTHOR QUERIES – TO BE ANSWERED BY THE CORRESPONDING AUTHOR

The following queries have arisen during the typesetting of your manuscript. Please answer these queries by marking the required corrections at the appropriate point in the text. Failure to do so could result in delayed publication.

Q1	Please check that all names have been spelled correctly and appear in the correct order. Please also check that all initials are present. Please check that the author surnames (family name) have been correctly identified by a pink background. If this is incorrect, please identify the full surname of the relevant authors. Occasionally, the distinction between surnames and forenames can be ambiguous, and this is to ensure that the authors' full surnames and forenames are tagged correctly, for accurate indexing online. Please also check all author affiliations.	
Q2	Please review the suggested running head.	
Q3	Figures have been placed as close as possible to their first citation. Please check that they have no missing sections and that the correct figure legend is present.	
Q4	Please supply supplementary material.	
Q5	Please update reference Wu and Zhou (2010).	
Q6	Cheng, B & Tong, H (1998) not cited in text.	
Q7	Horowitz, J-L & Spokoing, V.G (2001) not cited in text	

Biometrika (2011), *pp.* 1–17
Printed in Great Britain

doi: 10.1093/biomet/asr017
Advance Access publication 00 Month 2011

Testing parametric assumptions of trends of nonstationary time series

BY TING ZHANG AND WEI BIAO WU Q1

Department of Statistics, The University of Chicago, Chicago, Illinois 60637, U.S.A.
tzhang@galton.uchicago.edu wbwu@galton.uchicago.edu

SUMMARY

The paper considers testing whether the mean trend of a nonstationary time series is of certain parametric forms. A central limit theorem for the integrated squared error is derived, and with that a hypothesis-testing procedure is proposed. The method is illustrated in a simulation study, and is applied to assess the mean pattern of lifetime-maximum wind speeds of global tropical cyclones from 1981 to 2006. We also revisit the trend pattern in the central England temperature series.

Some key words: Bias correction; Central limit theorem; Integrated squared error; Local linear estimation; Locally stationary process; Nonparametric hypothesis testing.

1. INTRODUCTION

The problem of testing whether the mean trend of a time series follows certain parametric forms has attracted considerable attention; see Dette (1999), Bissantz et al. (2005), Percival & Rothrock (2005), Wu & Zhao (2007) and Pawlak & Stadmüller (1996, 2007), among others. Parametric models have the advantage of ease of interpretation and prediction, but may suffer from misspecification, leading to erroneous conclusions. Hypothesis testing in nonparametric regression under independence has been discussed by Härdle & Mammen (1993), Hart (1997), Fan et al. (2001) and Van Keilegom et al. (2008). Other contributions can be found in Azzalini et al. (1989), Eubank & LaRiccia (1992), Aerts et al. (1999), Eubank (1999), Horowitz & Spokoiny (2001) and Fan & Jiang (2007). In this paper we adopt the following formulation: suppose we observe

$$Y_i = \mu(i/n) + e_i \quad (i = 1, \ldots, n), \tag{1}$$

where $\mu(t)$, $t \in [0, 1]$, is an unknown signal or trend function and $(e_i)_{i=1}^n$ is a zero-mean noise sequence which can be nonstationary. We are interested in testing the hypothesis

$$H_0 : \mu(t) = f(\theta, t), \tag{2}$$

where the function $f(\cdot, \cdot)$ has a known form and $\theta \in \mathbb{R}^d$ is a parameter vector of f being identically zero, constant, and special cases $f \equiv 0$, $f \equiv$ a constant and $f(\theta, t) = \theta_0 + \theta_1 t$ for some $\theta = (\theta_0, \theta_1) \in \mathbb{R}^2$ correspond to testing whether a signal exists, is time-varying and nonlinear, respectively. A natural approach would be to compare a nonparametric estimator of $\mu(\cdot)$ and the fitted parametric trend $f(\hat{\theta}_n, t)$, where $\hat{\theta}_n$ might, for example, be the least squares estimator, which minimizes

$$\ell_n(\theta) = \sum_{i=1}^{n} \{Y_i - f(\theta, i/n)\}^2. \tag{3}$$

Nonparametric estimation of μ can be performed in a number of ways including the Priestley–Chao (1972), the Nadaraya–Watson, local polynomial, spline and wavelet methods. Here we shall use the local linear estimator, which has a good boundary performance (Fan & Gijbels, 1996),

$$\hat{\mu}_n(t) = \sum_{i=1}^{n} Y_i w_i(t), \tag{4}$$

where $w_i(t) = K\{(i/n - t)/b_n\}\{S_2(t) - (t - i/n)S_1(t)\}/\{S_2(t)S_0(t) - S_1^2(t)\}$ are the local linear weights, b_n is the bandwidth, $K(\cdot)$ is a kernel function and $S_j(t) = \sum_{i=1}^{n}(t - i/n)^j K\{(i/n - t)/b_n\}$. To test H_0, we shall develop central limit theory for the integrated squared error (ISE)

$$\text{ISE} = \int_0^1 \{\hat{\mu}_n(t) - \mu(t)\}^2 dt.$$

Under H_0, ISE can be estimated by Δ_2^2, where Δ_2 is the $\mathcal{L}^2$ distance

$$\Delta_2 = \left[\int_0^1 \{\hat{\mu}_n(t) - f(\hat{\theta}_n, t)\}^2 dt\right]^{1/2}.$$

We reject H_0 if Δ_2 is too large. The asymptotic normality of ISE has been studied under different settings. See for example Bickel & Rosenblatt (1973), Hall (1984), Ioannides (1992) and Alcalá et al. (1999), among others. However, in those papers the errors are independent. González-Manteiga & Vilar Fernández (1995) and Biedermann & Dette (2000) considered the same problem for linear processes with independent and identically distributed innovations. The ISE is a quadratic form in the errors e_i. As commented in Pawlak & Stadtmüller (2007), existing results on quadratic forms for dependent processes are mostly confined to linear processes. It is unclear whether similar results hold for general nonlinear processes. Here we shall substantially generalize earlier results by allowing nonlinear and nonstationary error processes. Hence our central limit theory should be widely applicable.

2. PRELIMINARIES

In (1) we allow nonstationary noise processes, on which there is a huge literature. Priestley (1965, 1988) considered processes with time-varying spectral representations. Dahlhaus (1997) defined a class of locally stationary processes for which a rigorous asymptotic theory can be obtained. Mallat et al. (1998) modelled locally stationary processes with pseudo-differential operators that are time-varying convolutions. Cheng & Tong (1998) applied wavelet representations. Nason et al. (2000) proposed to use a set of discrete non-decimated wavelets rather than the Fourier complex exponentials as in Dahlhaus (1997). Giurcanu & Spokoiny (2004) treated nonstationarity by assuming that correlation functions could be well approximated by those of stationary processes. Ombao et al. (2005) generalized the framework of Dahlhaus (1997) by utilizing the smooth localized complex exponentials. Here we shall follow the framework in Draghicescu et al. (2009) and Zhou & Wu (2009) and assume that the error sequence $\{e_i\}_{i=1}^n$ is generated from the model

$$e_i = G(i/n; \mathcal{F}_i), \tag{5}$$

where $\mathcal{F}_i = (\ldots, \epsilon_{i-1}, \epsilon_i)$ is a shift process of independent and identically distributed shocks ϵ_k, $k \in \mathbb{Z}$, and $G : [0, 1] \times \mathbb{R}^\infty \to \mathbb{R}$ is a measurable function such that e_i is well defined. The

framework (5) covers a wide range of nonstationary processes and naturally extends many existing stationary time series models to their nonstationary counterparts. Following Wu (2005), we interpret $\mathcal{F}_i$ and e_i as the input and output, and G as the transform that represents the underlying physical mechanism. Let $\zeta_i(t) = G(t; \mathcal{F}_i)$. Then $G(t; \cdot)$ is the data-generating mechanism at t, possibly time-varying. If $G(t; \cdot)$ does not depend on t, then (5) provides a very general framework for stationary processes; see Priestley (1988) and Tong (1990). Under a stochastic continuity condition, cf. Condition 2, equation (5) defines a locally stationary process.

We now introduce dependence measures that will be needed for our asymptotic theory. Let ϵ_i', ϵ_j $(i, j \in \mathbb{Z})$ be independent and identically distributed and define the coupled process $\mathcal{F}_{k,\{0\}} = (\ldots, \epsilon_{-1}, \epsilon_0', \epsilon_1, \ldots, \epsilon_k)$ and $\zeta_{k,\{0\}}(t) = G(t, \mathcal{F}_{k,\{0\}})$. For $p > 0$, assume that $c_p = \sup_{t\in[0,1]} \|\zeta_0(t)\|_p < \infty$, where $\|X\|_p = E(|X|^p)^{1/p}$, and define

$$\delta_{k,p} = \sup_{t\in[0,1]} \|\zeta_k(t) - \zeta_{k,\{0\}}(t)\|_p.$$

If $G(t, \mathcal{F}_k)$ does not depend functionally on ϵ_0, then $\zeta_k(t) - \zeta_{k,\{0\}}(t) = 0$. So $\delta_{k,p}$ measures the dependence of $G(t; \mathcal{F}_k)$ on the single input ϵ_0 over $t \in [0, 1]$. Let $p' = \min(p, 2)$ and define

$$\Theta_{m,p} = \sum_{j=m}^{\infty} \delta_{j,p}, \quad \Psi_{m,p} = \Big(\sum_{j=m}^{\infty} \delta_{j,p}^{p'} \Big)^{1/p'}.$$

We can interpret $\Theta_{m,p}$ as the cumulative dependence of ϵ_0 on $\{\zeta_j(t)\}_{j=m}^{\infty}$. Throughout the paper we assume that the short-range dependence condition holds for some $p \geqslant 2$:

$$\Theta_{0,p} = \sum_{j=0}^{\infty} \delta_{j,p} < \infty. \tag{6}$$

If (6) holds with $p = 2$, then the long-run variance function of the stationary process $\{\zeta_k(t)\}_k$ is bounded:

$$g(t) = \sum_{k\in\mathbb{Z}} \operatorname{cov}\{\zeta_0(t), \zeta_k(t)\} < \infty. \tag{7}$$

Write $\delta_k = \delta_{k,2}$ and $\|\cdot\| = \|\cdot\|_2$. We shall impose the following regularity conditions:

Condition 1. Let $\mu \in \mathcal{C}^3[0, 1]$;

Condition 2. There exists $C > 0$ such that $\|\zeta_i(t_1) - \zeta_i(t_2)\| \leqslant C|t_1 - t_2|$ holds for all $t_1, t_2 \in [0, 1]$;

Condition 3. Let $g \in \mathcal{C}^2[0, 1]$.

The key, Condition (2) indicates that the underlying data-generating mechanism $\zeta_i(t) = G(t; \mathcal{F}_i)$ changes smoothly in time, thus suggesting local stationarity. More specifically, a length k subsequence $(e_j)_{j=i}^{i+k-1}$ can be approximated by the stationary process $e_j^{\circ} = G(i/n, \mathcal{F}_j)$, $j = i, i + 1, \ldots, i + k - 1$, in that $\|e_j - G(i/n, \mathcal{F}_j)\| \leqslant Ck/n = o(1)$ if $k = o(n)$.

3. Main results

3·1. *Asymptotic normality*

Throughout the paper we assume that in (4) the kernel function $K(\cdot) \in \mathcal{K}$, the collection of symmetric, bounded functions in $\mathcal{C}^1[-1, 1]$ with $\int_{-1}^{1} K(v)dv = 1$. For example, K can be the

Epanechnikov kernel $K(v) = 3\max(0, 1 - v^2)/4$, the Bartlett kernel $K(v) = \max(0, 1 - |v|)$, or the rectangle kernel $K(v) = \mathrm{I}(|v| \leqslant 1)/2$, where $\mathrm{I}(\cdot)$ is the indicator function. Define

$$K^*(x) = \int_{-1}^{1-2|x|} K(v)K(v + 2|x|)dv.$$

Since $K \in \mathcal{C}^1[-1, 1]$, we have $K^* \in \mathcal{K}$, which is continuous on $\mathbb{R}$. We present below central limit theorems for

$$I_n = \mathrm{ISE} - E(\mathrm{ISE}) = \int_0^1 (\{\hat{\mu}_n(t) - \mu(t)\}^2 - E[\{\hat{\mu}_n(t) - \mu(t)\}^2])dt$$

when $\mu \equiv 0$ and $\mu \not\equiv 0$, respectively. It turns out that the asymptotic distributions are different for these two cases.

THEOREM 1. *Let $\mu \equiv 0$. Assume Condition 2, Condition 3, $\Theta_{0,4} < \infty$, $K \in \mathcal{K}$, $b_n \to 0$ and $nb_n \to \infty$ as $n \to \infty$.*

(i) *Let $g_1 = \int_0^1 g(t)dt$, $\Gamma_k = \sum_{i=0}^{\infty} \delta_i \delta_{i+|k|}$ and $l_n = \lfloor 2nb_n \rfloor$. Then*

$$nb_n \int_0^1 E\{\hat{\mu}_n(t)^2\}dt - g_1 K^*(0) = \sum_{k=0}^{\infty} O\{\min(\Gamma_k, b_n) + \min(1, k/l_n)\Gamma_k\}. \tag{8}$$

(ii) *Let $K_2^* = \int_{-1}^1 \{K^*(v)\}^2 dv$ and $g_2 = \int_0^1 g^2(t)dt$. If $nb_n^{3/2} \to \infty$, then as $n \to \infty$,*

$$nb_n^{1/2} I_n \to N(0, 4g_2 K_2^*) \tag{9}$$

in distribution.

Theorem 1(i) deals with the mean integrated squared error $E(\mathrm{ISE}) = \int_0^1 E\{\hat{\mu}_n(t)\}^2 dt$. Since $\sum_{k=0}^{\infty} \Gamma_k \leqslant (\sum_{i=0}^{\infty} \delta_i)^2 < \infty$, $b_n \to 0$ and $l_n \to \infty$, by the Lebesgue dominated convergence theorem, the right-hand side of (8) goes to zero as $n \to \infty$. The quantity g_1 is the integrated long-run variance. For stationary processes we have $g(t) \equiv g(0)$ for all t, and $g_1 = g(0)$, $g_2 = g_1^2$. Based on Theorem 1, we present in § 3·2 a simulation-based testing procedure.

THEOREM 2. *Assume Conditions 1–3 and $\Theta_{0,4} < \infty$. Also assume $K \in \mathcal{K}$, $b_n \to 0$ and $nb_n^{3/2} \to \infty$ as $n \to \infty$.*

(i) *If $nb_n^5 \to 0$, then (9) holds.*
(ii) *Let $\kappa_2 = \int K(v)v^2 dv$. If $nb_n^5 \to \infty$, then*

$$\frac{n^{1/2}}{b_n^2} I_n \to \kappa_2 N\left\{0, \int_0^1 g(t)\mu''(t)^2 dt\right\}$$

in distribution.

For both big and small bandwidths, I_n is asymptotically normal. However, for those two cases, both the normalization sequences and the asymptotic variances are different.

3·2. *Implementation*

We now present a testing procedure based on results in §3·1. Assume at the outset that the long-run variance function $g(\cdot)$ is known and we want to test the hypothesis of nonexistence of a signal

$$H_0' : \mu \equiv 0. \tag{10}$$

Section 3·4 concerns testing of the hypothesis (2) with general nonzero parametric forms, and §3·6 presents an estimator for the long-run variance function $g(\cdot)$. To test (10), by Theorem 1, it is natural to use $\text{ISE} = \int_0^1 \hat{\mu}_n(t)^2 dt$ and we reject H_0' at level α, $0 < \alpha < 1$, if

$$\text{ISE} > g_1 K^*(0)(nb_n)^{-1} + z_{1-\alpha} n^{-1} b_n^{-1/2} (4 g_2 K_2^*)^{1/2},$$

where $z_{1-\alpha}$ is the $(1-\alpha)$th quantile of a standard normal distribution. However, our simulation study in § 4·1 shows that convergence in Theorem 1 can be quite slow. To overcome this, we propose a simulation-based procedure. Let Z_i $(i \in \mathbb{Z})$ be independent standard normal random variables, $Y_i^\diamond = e_i^\diamond = g(i/n)^{1/2} Z_i$, and let $\hat{\mu}_n^\diamond(\cdot)$ be the corresponding local linear estimator (4) with Y_i therein replaced by $Y_i^\diamond$. By Theorem 1, we have

$$nb_n \int_0^1 E\{\hat{\mu}_n^\diamond(t)^2\} dt - g_1 K^*(0) = O(b_n)$$

and, for $\text{ISE}^\diamond = \int_0^1 \hat{\mu}_n^\diamond(t)^2 dt$, the central limit theorem holds: as $n \to \infty$,

$$nb_n^{1/2}\{\text{ISE}^\diamond - E(\text{ISE}^\diamond)\} \to N(0, 4 g_2 K_2^*) \tag{11}$$

in distribution. Hence $\text{ISE} = \int_0^1 \hat{\mu}_n(t)^2 dt$ and $\text{ISE}^\diamond$ have the same asymptotic normal distribution, with mean $g_1 K^*(0) n^{-1} b_n^{-1}$ and variance $4 g_2 K_2^* n^{-2} b_n^{-1}$ if the bound in the right-hand side of (8) is of order $o(b_n^{1/2})$. The latter observation suggests that, instead of using the central limit theorem, the cutoff value of ISE can be obtained by simulating $\text{ISE}^\diamond$.

3·3. *Asymptotic power*

We consider the power of our test for the local alternative model of the form

$$Y_i = a_n h(i/n) + e_i,$$

where h is a known nonzero $\mathcal{C}^2$ function on $[0, 1]$ and $(a_n)_{n \geqslant 1}$ is a positive sequence with $a_n \to 0$. Here a_n indicates the magnitude of departure from the null hypothesis. By Proposition 1, the power goes to 1 if $nb_n^{1/2} a_n^2 \to \infty$.

PROPOSITION 1. *Assume Conditions* 2 *and* 3 *and* $\Theta_{0,4} < \infty$. *Also assume that* $K \in \mathcal{K}$, $b_n \to 0$, $nb_n^{3/2} \to \infty$ *and* $nb_n^{1/2} a_n^2 \to c > 0$ *as* $n \to \infty$. *Let* Φ *be the standard normal distribution function. Then the power of our testing procedure converges to* $\Phi\{z_\alpha + (4 g_2 K_2^*)^{-1/2} c \int_0^1 h(t)^2 dt\}$.

Proof. Let $\hat{\mu}_n^0(t) = \sum_{i=1}^n w_i(t) e_i$ and $h_n(t) = \sum_{i=1}^n w_i(t) h(i/n)$. Then

$$\int_0^1 \{\hat{\mu}_n(t)\}^2 - \{\hat{\mu}_n^0(t)\}^2 dt = 2a_n \int_0^1 \hat{\mu}_n^0(t) h_n(t) dt + a_n^2 \int_0^1 h_n(t)^2 dt. \tag{12}$$

Since $h \in \mathcal{C}^2[0,1]$, $|\int_0^1 w_i(t)h_n(t)dt| \leqslant C/n$ for some $C > 0$. By Lemma 1,

$$nb_n^{1/2}a_n \left\| \int_0^1 \hat{\mu}_n^0(t)h_n(t)dt \right\| = nb_n^{1/2}a_n O(n^{-1/2}) \to 0.$$

So, Proposition 1 follows from Theorem 1 and (12) since $\int_0^1 h_n(t)^2 dt \to \int_0^1 h(t)^2 dt$. □

3·4. *Hypothesis testing with general parametric forms*

Under the null hypothesis (2), let θ_0 be the true value and $\hat{\theta}_n$ be the minimizer of (3). Then a natural test statistic would be $\int_0^1 \{\hat{\mu}_n(t) - f(\hat{\theta}_n, t)\}^2 dt$. However, it is not convenient to use this directly since by Theorem 2, $\int_0^1 \{\hat{\mu}_n(t) - f(\theta_0, t)\}^2 dt$ can have different asymptotic normal distributions for different bandwidths. It is nontrivial to determine which central limit theorem to use. To solve this problem, we use the modified version

$$\text{ISE}_M = \int_0^1 \{\hat{\mu}_n(t) - \mu_M(t)\}^2 dt,$$

where $\mu_M(t) = \sum_{i=1}^n f(\theta_0, i/n)w_i(t)$ is the local linear smoothed version of $f(\theta_0, t)$. Since $\hat{\mu}_n(t) - \mu_M(t) = \sum_{i=1}^n w_i(t)e_i$, ISE_M reduces to ISE with $\mu \equiv 0$. The bias then disappears on replacing $\mu(t)$ by $\mu_M(t)$. Such a bias correction scheme was previously used in Härdle & Mammen (1993). Note that ISE_M is not directly usable since it depends on the unknown function μ, which under H_0 depends on the parameter θ_0. It can be estimated by

$$\text{ISE}_{\hat{M}} = \int_0^1 \{\hat{\mu}_n(t) - \mu_{\hat{M}}(t)\}^2 dt = \int_0^1 \{\hat{\mu}_n^{(e)}(t)\}^2 dt, \tag{13}$$

where $\mu_{\hat{M}}(t) = \sum_{i=1}^n f(\hat{\theta}_n, i/n)w_i(t)$ and $\hat{\mu}_n^{(e)}(t) = \hat{\mu}_n(t) - \mu_{\hat{M}}(t)$ is the local linear smoother of the estimated residuals $\hat{e}_i = Y_i - f(\hat{\theta}_n, i/n)$.

PROPOSITION 2. *Assume that $\dot{f}(\theta, t) = \partial f(\theta, t)/\partial\theta$ exists at a neighbourhood of θ_0, that*

$$\sup_{t\in[0,1]} \sup_{|\theta-\theta_0|\leqslant c} |\dot{f}(\theta, t)| < \infty \tag{14}$$

holds for some $c > 0$, and that $f(\theta, t)$ admits the uniform Taylor expansion: as $\theta \to \theta_0$,

$$\sup_{t\in[0,1]} |f(\theta, t) - f(\theta_0, t) - (\theta - \theta_0)^{\mathrm{T}} \dot{f}(\theta_0, t)| = \mathcal{O}(|\theta - \theta_0|^2). \tag{15}$$

Then under conditions of Theorem 1, if $\hat{\theta}_n - \theta_0 = O_p(n^{-1/2})$, we have

$$n(\text{ISE}_{\hat{M}} - \text{ISE}_M) = O_p(1). \tag{16}$$

Proposition 2 implies that the statistic $\text{ISE}_{\hat{M}}$ can approximate ISE_M well, and it is also asymptotically normally distributed with same asymptotic mean and variance, given in Theorem 1.

3·5. *Bandwidth selection*

Choosing a bandwidth such that the test procedure based on $\text{ISE}_{\hat{M}}$ has a good performance, is nontrivial, and in our case it is further complicated by the presence of dependence and nonstationarity. In the regression setting with independent errors, the problem was considered

by Härdle & Marron (1985), Härdle et al. (1988), Park & Marron (1990), Ruppert et al. (1995), Wand & Jones (1995) and Xia (1998), among others. Herrmann et al. (1992) and Wu & Zhao (2007) dealt with models with stationary errors. Hall & Hart (1990), Kulasekera & Wang (1997) and Gao & Gijbels (2008) considered bandwidth selection in the context of nonparametric hypothesis testing.

We propose using the asymptotic mean squared error optimal bandwidth $b_n = cn^{-1/5}$, where $c > 0$ is a constant. Due to dependence and nonstationarity, it is difficult to estimate c. On the positive side, our simulation studies carried out in §4 suggests that the performance of our simulation-based test is relatively robust to the choice of c. Hence in our simulation and data analysis, we simply choose $b_n = n^{-1/5}$ in computing $\text{ISE}_{\hat{M}}$, and interestingly, this simple choice performs quite well.

3·6. *Estimation of variance functions*

In the implementation of our testing procedure, a key issue is to estimate the pointwise long-run variance function, $g(t), t \in [0, 1]$. If the errors e_i were independent and identically distributed, then $g(t) \equiv \sigma_e^2 = \|e_i\|^2$, the variance of e_i. In this case we can apply difference-based variance estimators and there is a huge literature on the estimation of σ_e^2; see Rice (1984), Hall et al. (1990) and Dette et al. (1998), among others. In our setting, however, due to the dependence, the difference-based approach is generally invalid. For example, assuming that $\mu \in \mathcal{C}^2[0, 1]$ and e_i are stationary, as $n \to \infty$, Rice's (1984) estimator $(2n-2)^{-1} \sum_{i=2}^n (Y_i - Y_{i-1})^2 \to \gamma_0 - \gamma_1$, by the ergodic theorem. Here $\gamma_k = \text{cov}(e_0, e_k)$ is the auto-covariance function of (e_i). Note that $\sigma^2 = \sum_{k \in \mathbb{Z}} \gamma_k$, which is generally different from $\gamma_0 - \gamma_1$.

To account for dependence and nonstationarity, we estimate $g(t)$ by

$$\hat{g}(t) = \frac{\sum_{i=1}^n Q_i \mathrm{I}(|i/n - t| \leqslant b_n)}{\sum_{i=1}^n \mathrm{I}(|i/n - t| \leqslant b_n)}, \tag{17}$$

where $Q_i = e_i \sum_{|j-i| \leqslant m_n} e_j$. By Theorem 3, the above estimator is consistent.

THEOREM 3. *Assume Conditions* 2 *and* 3, $\Theta_{0,4} < \infty$, $b_n \to 0$ *and* $m_n^{-1} + m_n(nb_n)^{-1} \to 0$. *Then as* $n \to \infty$,

$$(nb_n/m_n)^{1/2}[\hat{g}(t) - E\{\hat{g}(t)\}] \to N\{0, 2g^2(t)\} \tag{18}$$

in distribution for any $t \in (0, 1)$. *Also, uniformly over* $t \in [b_n, 1 - b_n]$, *the bias*

$$E\{\hat{g}(t)\} - g(t) = O(b_n^2) + \sum_{k=0}^{m_n} O\{\min(\Gamma_k, m_n/n)\} + \sum_{k > m_n} O(\Gamma_k). \tag{19}$$

If there exists $\rho \in (0, 1)$ such that $\Gamma_k = O(\rho^k)$, letting $m_n = \lfloor \log n / \log \rho^{-1} \rfloor$, we have by (19) that $E\{\hat{g}(t)\} - g(t) = O(b_n^2) + O(m_n^2/n)$. Hence, by (18), the mean squared error of $\hat{g}(t)$ is of order $O(b_n^4 + m_n^4/n^2) + O\{m_n(nb_n)^{-1}\} = O\{(n^{-1}\log n)^{4/5}\}$ if $b_n \asymp (n^{-1}\log n)^{1/5}$.

To use (17), we suggest using the automatic bandwidth selector in Ruppert et al. (1995) to obtain a local linear fit $\tilde{\mu}_n(\cdot)$ of the mean function and then the estimated residuals $\tilde{e}_i = Y_i - \tilde{\mu}_n(i/n)$. Then we replace e_i in (17) by $\tilde{e}_i$.

3·7. *A simulation-based testing procedure*

We summarize the testing procedure as follows. Its validity is justified by noting that ISE and ISE° have the same asymptotic distribution, as argued in (11). Section 4 presents a simulation

study of its finite-sample performance.

(i) Select a bandwidth b_n^* by the procedure in Ruppert et al. (1995); perform a local linear fit for $\mu(\cdot)$ and obtain the estimated residuals $\tilde{e}_i = Y_i - \tilde{\mu}_n(i/n)$; compute $\hat{g}(t)$, $t \in [0, 1]$, via (17).

(ii) Obtain an estimator $\hat{\theta}_n$ for the parameter θ; compute $\tilde{e}_i = Y_i - f(\hat{\theta}_n, i/n)$ and $\text{ISE}_{\hat{M}}$ via (13) with the local linear estimator (4) where $b_n = n^{-1/5}$ is used.

(iii) Generate independent $Z_1, \ldots, Z_n \sim N(0, 1)$ and let $Y_i^\diamond = \hat{g}(i/n)^{1/2} Z_i$. Then compute the corresponding $\text{ISE}_{\hat{M}}^\diamond$ in the same manner as (ii).

(iv) Let $\alpha \in (0, 1)$ be the significance level. Repeat step (iii) and obtain the estimated quantile $\hat{q}_{1-\alpha}$ of $\text{ISE}_{\hat{M}}^\diamond$, the bootstrap cutoff value.

(v) Reject the null hypothesis at level α if $\text{ISE}_{\hat{M}} > \hat{q}_{1-\alpha}$.

4. A SIMULATION STUDY

4·1. *Approximations of distributions of test statistics*

Consider model (1) with $e_i = \zeta_i(i/n)$, where for any $t \in [0, 1]$, $\{\zeta_i(t)\}_{i\in\mathbb{Z}}$ follows the recursion

$$\zeta_i(t) = \rho(t)\zeta_{i-1}(t) + \sigma\epsilon_i. \tag{20}$$

Here ϵ_i, $i \in \mathbb{Z}$, are independent random variables with $\text{pr}(\epsilon_i = -1) = \text{pr}(\epsilon_i = 1) = 1/2$. Thus, $(e_i)_{i\in\mathbb{Z}}$ is a first-order autoregressive process with time-varying coefficient. Calculations show that $E\{\zeta_i(t)\} = 0$, $g_0(t) = \text{var}\{\zeta_i(t)\} = \sigma^2/\{1 - \rho(t)^2\}$ and the long-run variance function $g(t) = \sigma^2/\{1 - \rho(t)\}^2$. We use the Epanechnikov kernel $K(v) = 3\max(0, 1 - v^2)/4$. Then $K^*(0) = 3/5$ and $K_2^* = 167/770$. We consider the problem of testing $H_0' : \mu \equiv 0$. Choose $\rho(t) = 0{\cdot}1 + 0{\cdot}4t$, $\sigma = 1$ and $n = 500$, so the bandwidth $b_n = n^{-1/5} = 0{\cdot}289$. We simulate 50 000 realizations of $\text{ISE}_{\hat{M}}$. Three different approximations of $\text{ISE}_{\hat{M}}$ are considered: the normal approximation in Theorem 1; $\text{ISE}^\sharp = \int_0^1 \hat{\mu}_n^\sharp(t)^2 dt$, where $\hat{\mu}_n^\sharp(t) = \sum_{i=1}^n Y_i^\sharp w_i(t)$, $Y_i^\sharp = g_0(i/n)^{1/2} Z_i$ and Z_i are independent standard normal variables; and $\text{ISE}^\diamond = \int_0^1 \hat{\mu}_n^\diamond(t)^2 dt$, introduced in §3.2, which differs from the second in that the long-run variance function g is used instead of the marginal variance function g_0. In the second scheme, the dependence is ignored. We use Q–Q plot to compare the distributions. The results are presented in Fig. 1, which shows that the normal approximation does not have a satisfactory finite-sample performance. A similar phenomenon was observed by Härdle & Mammen (1993). If we ignore the inherited dependence structure, then one may obtain an erroneous conclusion; see Fig. 1(b) of Fig. 1. As shown in Fig. 1(c), the dependence-adjusted procedure provides a very good approximation of $\text{ISE}_{\hat{M}}$. The same conclusion applies to other parametric forms. Q3

For a theoretical justification of the superiority of the simulation-based method, we use the Gaussian approximation principle in Wu & Zhou (2010). Consider the linear process $X_i = \sum_{j=0}^\infty a_j(i/n)\eta_{i-j}$, where η_i are independent and identically distributed with mean 0 and $\eta_i \in \mathcal{L}^p$ $(p > 2)$, and $a_j(\cdot)$ are differentiable functions satisfying $\sum_{j=0}^\infty \sup_{t\in[0,1]} |a_j'(t)| < \infty$ and $\sum_{j=m}^\infty \sup_{t\in[0,1]} |a_j(t)| = O(m^{1/p-1/2})$. Then on a richer probability space one can construct $e_1^\star, \ldots, e_n^\star$ and independent standard normal random variables $Z_1^\star, \ldots, Z_n^\star$ such that $(e_i^\star)_{i=1}^n$ and $(e_i)_{i=1}^n$ have the same distribution and, for $S_i^\star = \sum_{j=1}^i e_i^\star$ and $T_i^\star = \sum_{j=1}^i g(i/n)^{1/2} Z_i^\star$,

$$\max_{i\leqslant n} |S_i^\star - T_i^\star| = O_p(n^{1/p}\log n). \tag{21}$$

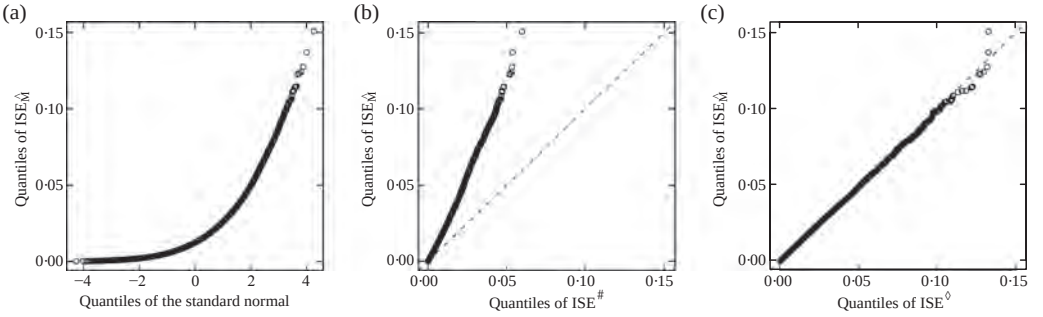


Fig. 1. Title Q–Q plots of the test statistic $\text{ISE}_{\hat{M}}$ against normal approximation (a), approximation without dependence, $\text{ISE}^{\sharp}$, (b) and approximation with dependence, $\text{ISE}^{\diamond}$, (c). The dashed lines in (b) and (c) have unit slope and zero intercept. Q3

Let $\mu_n^\star(t) = \sum_{i=1}^n w_i(t) e_i^\star$ and $\nu_n^\star(t) = \sum_{i=1}^n w_i(t) g(i/n)^{1/2} Z_i^\star$. For the local linear weights $w_i(t)$ in (4), if K is Lipschitz continuous, by (21), we obtain

$$\sup_{0 \leqslant t \leqslant 1} |\mu_n^\star(t) - \nu_n^\star(t)| = O_p\{n^{1/p} \log n/(nb_n)\} \tag{22}$$

by a summation by parts technique. By Theorem 1, $\int_0^1 \mu_n^\star(t)^2 dt = O_p\{(nb_n)^{-1}\}$. So, by (22),

$$\int_0^1 |\mu_n^\star(t)^2 - \nu_n^\star(t)^2| dt = O_p\{n^{1/p} \log n (nb_n)^{-3/2}\}. \tag{23}$$

When $\mu \equiv 0$, since $(e_i^\star)_{i=1}^n$ and $(e_i)_{i=1}^n$ have the same distribution, $\int_0^1 \mu_n^\star(t)^2 dt$ is identically distributed as $\int_0^1 \hat{\mu}_n(t)^2 dt$. Note that $\int_0^1 \nu_n^\star(t)^2 dt$ corresponds to $\text{ISE}_{\hat{M}}^{\diamond}$ in step (iii) in §3·7 if $\hat{g}$ therein is replaced by the true g. For the mean squared error optimal bandwidth $b_n \asymp n^{-1/5}$, the error bound in (22) is $O_p(n^{\tau - 6/5})$. Since τ can be arbitrarily small if p is large, $\int_0^1 \mu_n^\star(t)^2 dt$ and $\int_0^1 \nu_n^\star(t)^2 dt$ can be very close; recalling Proposition 2 that the difference between ISE_M and the realized version $\text{ISE}_{\hat{M}}$ has a larger order $O_p(n^{-1})$. So, we expect that the simulation-based method can have an excellent performance.

4·2. *Effect of bandwidths*

We consider model (1) with error structure (20). With true mean function $\mu(t) = 4t^2 - 4t + 3$, $t \in [0, 1]$, we consider testing whether the mean function has a quadratic form. To study how bandwidths affect the performance of our test, we choose $\rho(t) = 0{\cdot}3 - 0{\cdot}5t$, $\sigma = 1$ and 10 levels of b: $b = 0{\cdot}05j$ $(j = 1, \ldots, 10)$. For each b, we calculate $\hat{q}_{0{\cdot}95}$ by repeating step (iii) in §3·7 50 000 times. Then we generate 50 000 realizations of the time-varying AR(1) process and compute the corresponding $\text{ISE}_{\hat{M}}$ values for each realization. Simulated empirical rejection proportions with $n = 100$, 200 and 500, presented in Table 1, are reasonably close to the nominal level. If we choose $b_n = n^{-1/5}$, then $b_{100} = 0{\cdot}398$, $b_{200} = 0{\cdot}347$ and $b_{500} = 0{\cdot}289$, and the corresponding empirical rejection probabilities are about 5·01, 4·97 and 5·03%, respectively. In addition, they become more robust to the change in bandwidths as n gets larger. So, in practice, we recommend using $b_n = n^{-1/5}$.

Table 1. *Empirical rejection percentages with different bandwidths b and sample sizes n. The significance level is* 5%.

	b									
n	0·05	0·1	0·15	0·2	0·25	0·3	0·35	0·4	0·45	0·5
100	2·8	4·0	4·5	4·7	4·7	5·0	5·2	5·0	5·1	5·1
200	3·8	4·6	4·6	5·1	4·6	5·1	5·0	5·2	5·1	5·2
500	4·6	4·7	5·0	4·9	5·1	5·0	5·0	5·1	5·2	5·1

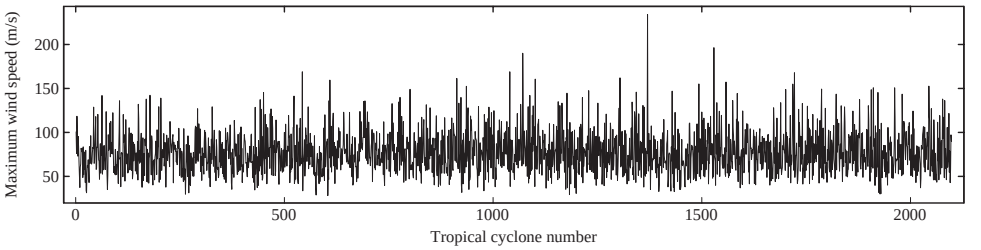


Fig. 2. Satellite-derived lifetime-maximum wind speeds of tropical cyclones during 1981–2006.

5. APPLICATIONS

5·1. *Tropical cyclone data*

According to Emanuel (1991), Holland (1997) and Bengtsson (2007), global warming is leading to increasing ocean temperatures and consequently to more tropical cyclones. Elsner et al. (2008) analysed patterns of tropical cyclone winds by fitting linear trends for quantiles of satellite-derived lifetime-maximum wind speeds of 2098 tropical cyclones over the globe during 1981–2006. The data set is available in the Supplementary material see Elsner et al. (2008) for a detailed description. Figure 2 shows these data.

We shall model the wind speed data by (1) and test the null hypothesis $H_0 : \mu(\cdot)$ is linear. Zhou (2010) argued that the error process (e_i) is nonstationary. Hence $g(t)$ is not a constant function. Using the procedure in Ruppert et al. (1995), we select $b_n^* = 0{\cdot}081$ and estimate the long-run variance function $g(t)$ by using (17) with $m_n = (nb_n^*)^{1/3}$. For hypothesis testing, we choose $b_n = n^{-1/5} = 0{\cdot}217$. The test statistic $\text{ISE}_{\hat{M}} = 1{\cdot}011$ with p-value 0·12 after 50 000 repetitions of step (iii) in §3.7. So, at the 5% significance level, we accept the linear trend hypothesis. The fitted trend, with standard error, is $\mu(t) = 75{\cdot}3(1{\cdot}1) + 3{\cdot}8(1{\cdot}8)t$, $t \in [0, 1]$. For testing the hypothesis $H_0^* : \mu$ is constant, using the same method we obtain $\text{ISE}_{\hat{M}} = 2{\cdot}091$ with p-value 0·04. Thus the mean constancy hypothesis is rejected at the 5% level. Zhou (2010) applied an $\mathcal{L}^\infty$-based method, and failed to reject H_0^*. Our $\mathcal{L}^2$-based testing procedure appears to be more powerful.

5·2. *Central england temperature data*

We consider the annual central England temperature series from year 1659 to 2009 by using model (1). The time series is plotted in Fig. 3 and the data are available in the Supplementary material. It was first constructed by Manley (1974) and is now routinely updated by the Hadley Center, U.K. See Jones & Hulme (1997) for a more detailed description. Jones & Hulme (1997) and Jones & Bradley (1992a) fitted linear trends, while Benner (1999) and Harvey & Mills (2003) fitted quadratic curves. Realizing that the quadratic trend assumption might not be appropriate, Harvey & Mills (2003) also tried local polynomial regression. Here we shall test the

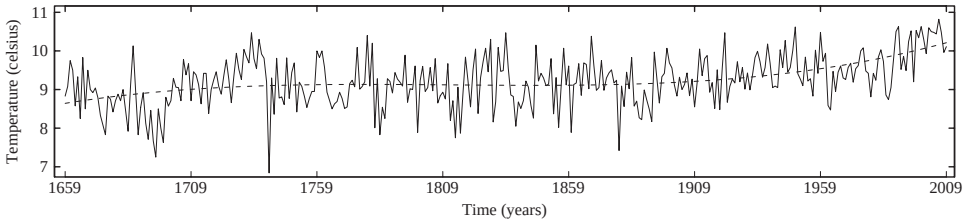


Fig. 3. Annual central England temperature series from 1659 to 2009. The dashed curve is the global cubic fit.

hypothesis $H_0 : \mu(\cdot)$ is quadratic. As in the analysis of the tropical cyclone data, we follow the procedure in §3.7 and obtain the test statistic $\text{ISE}_{\hat{M}} = 0{\cdot}0059$ with $p = 0{\cdot}00002$. Hence, the quadratic trend assumption is rejected at the 5% level. Interestingly, the cubic trend hypothesis is accepted: the test statistic $\text{ISE}_{\hat{M}} = 0{\cdot}000096$ and the corresponding p-value is 0·47. The fitted equation with standard error in parenthesis, is $\mu(t) = 8{\cdot}63(0{\cdot}13) + 3{\cdot}7(1{\cdot}1)t - 8{\cdot}9(2{\cdot}5)t^2 + 6{\cdot}8(1{\cdot}7)t^3$, where $t \in [0, 1]$. The cubic trend fit accords well with Benner's observation that the whole series has roughly three periods: the earliest part corresponds to the coldest weather which may represent the little ice age (Jones & Bradley, 1992b), the middle part fluctuates around the mean, while the last part exhibits a warming trend. This cannot be described by a quadratic fit.

Supplementary material

Supplementary Material is available at *Biometrika* online. Q4

Acknowledgement

We are grateful to the editor, an associate editor and three referees for their helpful comments and suggestions which render a much improved version. We thank Professor Wand for clarifications on bandwidth selection procedures. The research was supported by the US National Science Foundation.

Appendix

Recall that $\zeta_k(t) = G(t, \mathcal{F}_k)$. We assume that $E\{G(t, \mathcal{F}_k)\} = 0$ for any $t \in [0, 1]$. Define

$$\tilde{\zeta}_k(t) = E\{\zeta_k(t) \mid \mathcal{F}_{k-m,k}\}, \quad m \in \mathbb{N}.$$

Then for all $t \in [0, 1]$, $\{\tilde{\zeta}_k(t)\}_k$ is m-dependent with mean zero. Define the projection operator

$$\mathcal{P}_k \cdot = E(\cdot \mid \mathcal{F}_k) - E(\cdot \mid \mathcal{F}_{k-1}), \quad k \in \mathbb{Z}.$$

For a set $T \subseteq \mathbb{Z}$, let $\epsilon_{i,T} = \epsilon_i'$ if $i \in T$, and $\epsilon_{i,T} = \epsilon_i$ otherwise. Let $\mathcal{F}_{i,j,T} = (\epsilon_{k,T}, k = i, \ldots, j)$. In the proof C denotes constants whose value may vary from place to place.

To prove Theorem 1, we shall apply the technique in Liu & Wu (2010). Lemmas 1, 2 and 3 provide bounds for m-dependent and martingale approximations for linear and quadratic forms. They can be proved by using the arguments in Lemma 1, Propositions 1 and 2 in Liu & Wu (2010), respectively. That paper deals with stationary processes, but there are no essential additional difficulties involved for generalization to nonstationary processes. Detailed proofs can be found in the online Supplementary Material.

LEMMA A1. *Assume* $\Theta_{0,p} < \infty$, $p \geqslant 2$. *Let* $\alpha_1, \alpha_2, \ldots, \in \mathbb{R}$, $A_n = (\sum_{i=1}^n \alpha_i^2)^{1/2}$, *and* $C_p = 18p^{3/2}(p-1)^{-1/2}$. *Then (i)* $\|\sum_{i=1}^n \alpha_i e_i\|_p \leqslant C_p A_n \Theta_{0,p}$; *and (ii)* $\|\sum_{i=1}^n \alpha_i (e_i - \tilde{e}_i)\|_p \leqslant C_p A_n \Theta_{m+1,p}$.

LEMMA A2. *Assume* $\Theta_{0,4} < \infty$. *Let* $\alpha_j \in \mathbb{R}$,

$$L_n = \sum_{1 \leqslant i < j \leqslant n} \alpha_{j-i} e_i e_j, \quad \tilde{L}_n = \sum_{1 \leqslant i < j \leqslant n} \alpha_{j-i} \tilde{e}_i \tilde{e}_j. \tag{A1}$$

Let $d_{m,4} = \sum_{k=0}^\infty \min(\delta_{k,4}, \Psi_{m+1,4})$ *and* $A_n = (\sum_{i=1}^{n-1} \alpha_i^2)^{1/2}$. *Then*

$$\|\{L_n - E(L_n)\} - \{\tilde{L}_n - E(\tilde{L}_n)\}\| \leqslant C_4 \Theta_{0,4} d_{m,4} n^{1/2} A_n.$$

LEMMA A3. *Let* $\alpha_1, \ldots, \alpha_{n-1} \in \mathbb{R}$ *and* $V_m(\alpha) = \max_{i<n} \alpha_i^2 + m \sum_{i=1}^{n-1} |\alpha_i - \alpha_{i-1}|^2$; *let*

$$\tilde{D}_{k,n} = \sum_{l=0}^\infty \mathcal{P}_k \tilde{e}_{k+l}, \quad \tilde{M}_n = \sum_{1 \leqslant i < j \leqslant n} \alpha_{j-i} \tilde{D}_{i,n} \tilde{D}_{j,n}.$$

Assume $\Theta_{0,4} < \infty$. *Then* $\|\tilde{L}_n - E(\tilde{L}_n) - \tilde{M}_n\|^2 \leqslant C m^3 n V_m(\alpha)$.

THEOREM A1. *Assume Condition* 2 *and Condition* 3, *(A3)*, $\Theta_{0,4} < \infty$, $K \in \mathcal{K}$, $b_n \to 0$ *and* $n b_n^{3/2} \to \infty$. *Recall (A1) for* L_n *and let* $\alpha_j = n^{-1} b_n^{-1/2} K^*\{j/(2nb_n)\}$. *Then as* $n \to \infty$, $L_n - E(L_n) \to N(0, g_2 K_2^*)$ *in distribution.*

Proof. Recall Lemma A2 for A_n and Lemma A3 for $V_m(\alpha)$. Since $K^* \in \mathcal{K}$, $V_m(\alpha) = O(mn^3 b_n^2)$ and $A_n = O(n^{-1/2})$. By Lemmas A2 and A3,

$$\lim_{m \to \infty} \limsup_{n \to \infty} \|L_n - E(L_n) - \tilde{M}_n\| = 0. \tag{A2}$$

In (7) we replace $\zeta_i(t)$ by $\tilde{\zeta}_i(t)$ and let $\tilde{g}(t)$ be the long-run variance function for the latter m-dependent process. Set $\tilde{g}_2 = \int_0^1 \tilde{g}^2(t)dt$, by (A2) it suffices to verify that $\tilde{M}_n \to N(0, \tilde{g}_2 K_2^*)$ in distribution. This can be proved by using the argument of Theorem 6 in Liu & Wu (2010), where the case of stationary processes is dealt with. Here we shall only detail the step, cf. (A3) below, that requires special attention of nonstationarity since all other steps similarly follow. A complete proof is available as supplementary material. We shall show that

$$\sum_{j=2m+1}^n E(J_{1,j,n}^2) E(\tilde{D}_{j,n}^2) \to \tilde{g}_2 K_2^*, \tag{A3}$$

where $J_{1,j,n} = \sum_{i=1}^{j-2m} \alpha_{j-i} \tilde{D}_{i,n}$. Let $\tilde{D}_{k,n}^* = \sum_{l=0}^\infty \mathcal{P}_k \tilde{\zeta}_{k+l}(k/n)$. By Condition (2),

$$|\|\tilde{D}_{k,n}\|^2 - \|\tilde{D}_{k,n}^*\|^2| \leqslant (\|\tilde{D}_{k,n}\| + \|\tilde{D}_{k,n}^*\|)\|\tilde{D}_{k,n} - \tilde{D}_{k,n}^*\| \leqslant C m^3 n^{-1} = o(1).$$

Observe that $\|\tilde{D}_{k,n}^*\|^2 = \tilde{g}(k/n)$. Then (A3) follows from

$$\begin{aligned}
\sum_{j=2m+1}^n \sum_{i=1}^{j-2m} \alpha_{j-i}^2 \|\tilde{D}_{i,n}^*\|^2 \|\tilde{D}_{j,n}^*\|^2 &= \sum_{1 \leqslant i < j \leqslant n} \alpha_{j-i}^2 \tilde{g}(i/n) \tilde{g}(j/n) + o(1) \\
&= \sum_{i=1}^{n-1} \tilde{g}(i/n) \sum_{j=1+i}^n \alpha_{j-i}^2 \tilde{g}(j/n) + o(1) \\
&= \sum_{i=1}^{n-1} \tilde{g}^2(i/n) \sum_{j=1+i}^n \alpha_{j-i}^2 + o(1) = \tilde{g}_2 K_2^* + o(1),
\end{aligned}$$

since $\sum_{j=1+i}^n \alpha_{j-i}^2 = n^{-1} K_2^*\{1 + o(1)\}$ and $\sum_{j=1+i}^n \alpha_{j-i}^2 \{\tilde{g}(j/n) - \tilde{g}(i/n)\} = o(n^{-1})$. □

Proof of Theorem 1. Let $\breve{\mu}_n(t) = \sum_{i=1}^n v_i(t) Y_i$ be the Priestley–Chao estimator, where $v_i(t) = (nb_n)^{-1} K\{(i/n - t)/b_n\}$. Note that $w_i(t) = v_i(t) + \mathcal{O}\{(nb_n)^{-2}\}$ uniformly over $(b_n, 1 - b_n)$. Both $\int_0^{b_n} E\{\hat{\mu}_n(t)^2\} dt$ and $\int_{1-b_n}^1 E\{\hat{\mu}_n(t)^2\} dt$ are of order $O(n^{-1})$. Then

$$\left\| \int_0^1 \hat{\mu}_n(t)^2 dt - \int_0^1 \breve{\mu}_n(t)^2 dt \right\|_1 \leqslant C\{n^{-1} + (nb_n)^{-2}\}.$$

Hence it suffices to prove the same results for the Priestley–Chao estimator $\breve{\mu}_n(t)$. For (i), write $l = l_n$. If $|i - j| \leqslant l$, by Condition (2) and the Cauchy–Schwarz inequality, $|E[e_i\{e_j - \zeta_j(i/n)\}]| \leqslant \|e_i\| \|\zeta_j(j/n) - \zeta_j(i/n)\| = O(l/n)$. Let $\gamma_k(t) = E\{\zeta_0(t)\zeta_k(t)\}$. Then $E(e_i e_j) - \gamma_{j-i}(i/n) = O(l/n)$. Since $\mathcal{P}_s$, $s \in \mathbb{Z}$, are orthogonal,

$$\begin{aligned} |E(e_i e_j)| &= \left| E\left(\sum_{s \in \mathbb{Z}} \mathcal{P}_s e_i \sum_{s' \in \mathbb{Z}} \mathcal{P}_{s'} e_j \right) \right| \leqslant \sum_{s \in \mathbb{Z}} |E\{(\mathcal{P}_s e_i)(\mathcal{P}_s e_j)\}| \\ &\leqslant \sum_{s \in \mathbb{Z}} \|\mathcal{P}_s e_i\| \|\mathcal{P}_s e_j\| \leqslant \sum_{s \in \mathbb{Z}} \delta_{i-s} \delta_{j-s} = \Gamma_{i-j}. \end{aligned} \tag{A4}$$

Similarly, $|\gamma_{j-i}(i/n)| \leqslant \Gamma_{j-i}$. Then

$$\frac{1}{n} \sum_{i=1}^n \sum_{k=-l}^l \left| K^*\left(\frac{k}{2nb_n}\right) \{E(e_i e_{i+k}) - \gamma_k(i/n)\} \right| = \sum_{k=0}^l O\{\min(\Gamma_k, b_n)\}. \tag{A5}$$

Let $\bar{\gamma}_k = \int_0^1 \gamma_k(t) dt$. Since $\gamma_k(\cdot)$ is Lipschitz continuous and $K^* \in \mathcal{C}[-1, 1]$,

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^n \sum_{k=-l}^l K^*\left(\frac{k}{2nb_n}\right) \gamma_k(i/n) - K^*(0) \sum_{k=-l}^l \bar{\gamma}_k \\ &\quad = \sum_{k=-l}^l K^*\left(\frac{k}{2nb_n}\right) \{\bar{\gamma}_k + O(n^{-1})\} - K^*(0) \sum_{k=-l}^l \bar{\gamma}_k = O(l/n) + \sum_{k=0}^l O(k/l)\Gamma_k. \end{aligned} \tag{A6}$$

By Lemma A1(i), $\sup_t \|\hat{\mu}_n(t)\|^2 = O\{(nb_n)^{-1}\}$. Note that $\int_{-b}^{1+b} K(x - t/b) K(y - t/b) dt = bK^*(y/2 - x/2)$ if $0 \leqslant x \leqslant y \leqslant b^{-1}$. Hence

$$\begin{aligned} nb_n \int_0^1 E\{\hat{\mu}_n(t)^2\} dt &= nb_n \int_{-b_n}^{1+b_n} E\{\hat{\mu}_n(t)^2\} dt + O(b_n) \\ &= \frac{1}{n} \sum_{1 \leqslant i, j \leqslant n} K^*\left(\frac{i-j}{2nb_n}\right) E(e_i e_j) + O(b_n). \end{aligned} \tag{A7}$$

Since $g_1 = \sum_{k \in \mathbb{Z}} \bar{\gamma}_k$, by (A5), (A6) and (A7), (8) follows in view of

$$\frac{1}{n} \sum_{i=1}^n \left\{ \sum_{k=-l}^l - \sum_{\max(1-i,l)}^{\min(n-i,l)} \right\} K^*\left(\frac{k}{2nb_n}\right) E(e_i e_{i+k}) = \frac{1}{n} O(l) = O(b_n).$$

For (ii), by Theorem A1, we have

$$nb_n^{1/2} \int_{-b_n}^{1+b_n} [\hat{\mu}_n(t)^2 - E\{\hat{\mu}_n(t)^2\}] dt \to N(0, 4g_2 K_2^*)$$

in distribution. So, (ii) follows in view of $nb_n^{1/2} \int_{-b_n}^0 E\{\hat{\mu}_n(t)^2\} dt = nb_n^{1/2} O\{b_n (nb_n)^{-1}\} \to 0$ and similarly, for the right tail, $nb_n^{1/2} \int_1^{1+b_n} E\{\hat{\mu}_n(t)^2\} dt \to 0$. □

Proof of Theorem 2. Observe that $I_n = I I_n + I I I_n$, where

$$I I_n = \int_0^1 \{[\hat{\mu}_n(t) - E\{\hat{\mu}_n(t)\}]^2 - E([\hat{\mu}_n(t) - E\{\hat{\mu}_n(t)\}]^2)\}dt,$$

$$I I I_n = \int_0^1 2[\hat{\mu}_n(t) - E\{\hat{\mu}_n(t)\}][E\{\hat{\mu}_n(t)\} - \mu(t)]dt.$$

By Theorem 1(ii), $nb_n^{1/2} I I_n \to N(0, 4g_2 K_2^*)$ in distribution. Then it suffices to show that

$$\frac{n^{1/2}}{b_n^2} I I I_n = \sum_{i=1}^n q_{n,i} e_i \to \kappa_2 N(0, \sigma^2) \tag{A8}$$

in distribution, where $\sigma^2 = \int_0^1 g(t)\mu''(t)^2 dt$ and

$$q_{n,i} = \frac{2n^{1/2}}{b_n^2} \int_0^1 w_i(t)[E\{\hat{\mu}_n(t)\} - \mu(t)]dt.$$

Under Condition 1, the bias $E\{\hat{\mu}_n(t)\} - \mu(t) = b_n^2 \kappa_2 \mu''(t)/2 + o(b_n^2)$. Since $K \in \mathcal{K}$, by elementary calculations, $r_{n,i} := q_{n,i} - n^{-1/2}\kappa_2 \mu''(i/n)$ satisfies $\sum_{i=1}^n |r_{n,i}|^2 = o(1)$. By Lemma A1(i), $\| \sum_{i=1}^n r_{n,i} e_i \| = o(1)$, and (A8) follows if

$$\sum_{i=1}^n n^{-1/2} \mu''(i/n) e_i \to N(0, \sigma^2) \tag{A9}$$

in distribution. To prove (A9) we apply the m-dependence approximation method. By Lemma A1(ii),

$$\lim_{m\to\infty} \limsup_{n\to\infty} \left\| \sum_{i=1}^n n^{-1/2} \mu''(i/n)(e_i - \tilde{e}_i) \right\| = 0.$$

Note that $\tilde{e}_i = E(e_i \mid \mathcal{F}_{i-m,i})$ are m-dependent. Let $\tilde{\sigma}^2 = \int_0^1 \tilde{g}(t)\mu''(t)^2 dt$. Hence (A9) follows from Hoeffding & Robbins' (1948) central limit theorem for $\sum_{i=1}^n n^{-1/2}\mu''(i/n)\tilde{e}_i$ for m-dependent random variables $\tilde{e}_i$ in view of

$$E\left\{ \sum_{i=1}^n n^{-1/2} \mu''(i/n)\tilde{e}_i \right\}^2 = \frac{1}{n} \sum_{|i-j| \leqslant m} \mu''(i/n)\mu''(j/n) E(\tilde{e}_i \tilde{e}_j) \to \tilde{\sigma}^2.$$

To see the above relation, we note that $E(\tilde{e}_i \tilde{e}_j) - \tilde{\gamma}_{i-j}(i/n) = o(1)$ if $|i-j| \leqslant m$, and $n^{-1} \sum_{i=1}^n \mu''(i/n)^2 \tilde{\gamma}_{i-j}(i/n) \to \int_0^1 \mu''(t)^2 \tilde{\gamma}_{i-j}(t)dt$, by the continuity of $\tilde{\gamma}_{i-j}(\cdot)$ and $\mu''(\cdot)$. □

Proof of Proposition 2. Write

$$\begin{aligned} \text{ISE}_{\hat{M}} - \text{ISE}_M &= 2\int_0^1 \{\hat{\mu}_n(t) - \mu_M(t)\}\{\mu_M(t) - \mu_{\hat{M}}(t)\}dt \\ &\quad + \int_0^1 \{\mu_M(t) - \mu_{\hat{M}}(t)\}^2 dt = 2A_n + B_n. \end{aligned}$$

So, (16) follows if both A_n and B_n are of order n^{-1}. To this end, by (15),

$$\begin{aligned} \mu_M(t) - \mu_{\hat{M}}(t) &= \sum_{i=1}^n w_i(t)\{f(\theta_0, t) - f(\hat{\theta}_n, t)\} \\ &= (\theta_0 - \hat{\theta}_n)^\top \sum_{i=1}^n w_i(t) \dot{f}(\theta_0, t) + O_p(n^{-1}) \end{aligned} \tag{A10}$$

holds uniformly over $t \in [0, 1]$ since $\hat{\theta}_n$ is $n^{1/2}$-consistent. Then by (14), we have $B_n = O_p(n^{-1})$. For A_n, note that $\hat{\mu}_n(t) - \mu_M(t) = \sum_{j=1}^{n} w_j(t) e_j$. By (A10), since $\|\hat{\mu}_n(t) - \mu_M(t)\| n^{-1} = O(n^{-3/2} b_n^{-1/2})$ and, for some constant $C > 0$, by (14),

$$\left| \int_0^1 w_j(t) \sum_{i=1}^{n} w_i(t) \dot{f}(\theta_0, t) dt \right| \leqslant C \int_0^1 |w_j(t)| dt \leqslant \frac{C}{n},$$

then we also have $A_n = O_p(n^{-1})$ in view of Lemma A1(i). □

Proof of Theorem 3. Let $\mathcal{I}_n(t) = \{i : |i/n - t| \leqslant b_n\}$. Observe that

$$\sum_{i=1}^{n} Q_i \mathrm{I}(|i/n - t| \leqslant b_n) = \sum_{i, j \in \mathcal{I}_n(t)} e_i e_j \mathrm{I}(|i - j| \leqslant m_n) + R_n,$$

where by (A4),

$$R_n = \sum_{i \in \mathcal{I}_n(t), j \notin \mathcal{I}_n(t)} e_i e_j \mathrm{I}(|i - j| \leqslant m_n) = O_p(m_n).$$

With elementary manipulations, (18) follows by applying the argument of Theorem 1 to $(e_i)_{i \in \mathcal{I}_n(t)}$ with $\alpha_{i-j} = \mathrm{I}(|i - j| \leqslant m_n)$. For (19), write

$$E(Q_i) - g(i/n) = \sum_{j=i-m_n}^{i+m_n} E[\zeta_i(i/n)\{\zeta_j(j/n) - \zeta_j(i/n)\}] + \sum_{j:|j-i|>m_n} E\{\zeta_i(i/n)\zeta_j(i/n)\}.$$

By (A6) and (Condition 2), the first and second terms above are of order $\sum_{k \leqslant m_n} O\{\min(\Gamma_k, m_n/n)\}$ and $\sum_{k > m_n} O(\Gamma_k)$, respectively. Then (19) follows from Condition 3. □

References

Aerts, M., Claeskens, G. & Hart, J. (1999). Testing the fit of a parametric function. *J. Am. Statist. Assoc.* **94**, 869–79.

Alcalá, J. T., Cristóbal, J. A. & González-Manteiga, W. (1999). Goodness-of-fit test for linear models based on local polynomials. *Statist. Prob. Lett.* **42**, 39–46.

Azzalini, A., Bowman, A. W. & Härdle, W. (1989). On the use of nonparametric regression for model checking. *Biometrika* **76**, 1–11.

Bengtsson, L., Hodges, K. I., Esch, M., Keenlyside, N., Kornblueh, L., Luo, J. J. & Yamagata, T. (2007). How may tropical cyclones change in a warmer climate? *Tellus* A **59**, 539–61.

Benner, T. C. (1999). Central England temperatures: Long-term variability and teleconnections. *Int. J. Climatol.* **19**, 391–403.

Bickel, P. J. & Rosenblatt, M. (1973). On some global measures of the deviations of density function estimate. *Ann. Statist.* **1**, 1071–95.

Bissantz, N., Holzmann, H., & Munk, A. (2005). Testing parametric assumptions on band- or time-limited signals under noise. *IEEE Trans. Info. Theory.* **51**, 3796–805.

Biedermann, S. & Dette, H. (2000). Testing linearity of regression models with dependent errors by kernel based methods. *Test* **9**, 417–38.

Cheng, B. & Tong, H. (1998). k-stationarity and wavelets. *J. Statist. Plan. Infer.* **68**, 129–144. Q6

Dahlhaus, R. (1997). Fitting time series models to non-stationary processes. *Ann. Statist.* **25**, 1–37.

Dette, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimators. *Ann. Statist.* **27**, 1012–40.

Dette, H., Munk, A. & Wagner, T. (1998). Estimating the variance in nonparametric regression by quadratic forms-what is a reasonable choice? *J. R. Statist. Soc.* B. **60**, 751–64.

Draghicescu, D., Guillas, S. & Wu, W. B. (2009). Quantile curve estimation and visualization for non-stationary time series. *J. Comp. Graph. Statist.* **18**, 1–20.

Elsner, J. B., Kossin, J. P. & Jagger, T. H. (2008). The increasing intensity of the strongest tropical cyclones. *Nature* **455**, 92–5.

Emanuel, K. A. (1991). The theory of hurricanes. *Annu. Rev. Fluid Mech.* **23**, 179–196.

Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing*, 2nd ed. New York: Marcel Dekker.

EUBANK, R. L. & LARICCIA, V. M. (1992). Asymptotic comparison of Cramér–von Mises and nonparametric function estimation techniques for testing goodness-of-fit. *Ann. Statist.* **20**, 2071–86.

FAN, J. Q. & GIJBELS, I. (1996). *Local Polynomial Modeling and Its Applications*. London, U.K.: Chapman & Hall.

FAN, J. Q. & JIANG, J. C. (2007). Nonparametric inference with generalized likelihood ratio tests. *Test* **16**, 409–44.

FAN, J. Q., ZHANG, C. M. & ZHANG, J. (2001). Generalized likelihood ratio statistics and Wilks phenomenon. *Ann. Statist.* **29**, 153–93.

GAO, J. & GIJBELS, I. (2008). Bandwidth selection in nonparametric kernel testing. *J. Am. Statist. Assoc.* **484**, 1584–94.

GIURCANU, M. & SPOKOINY V. G. (2004). Confidence estimation of the covariance function of stationary and locally stationary processes. *Statist. Decis.* **22**, 283–300.

GONZÁLEZ-MANTEIGA, W. & VILAR FERNÁNDEZ, J. M. (1995). Testing linear regression models using non-parametric regression estimators when the errors are correlated. *Comput. Statist. Data Anal.* **20**, 521–41.

HALL, P. (1984). Integrated square error properties of kernel estimators of regression functions. *Ann. Statist.* **12**, 241–60.

HALL, P. & HART, J. D. (1990). Bootstrap test for difference between means in nonparametric regression. *J. Am. Statist. Assoc.* **85**, 1039–49.

HALL, P., KAY, J. W. & TITTERINGTON, D. M. (1990). Asymptotically optimal difference-based estimation of variance in nonparametric regression. *Biometrika* **77**, 521–28.

HÄRDLE, W., HALL, P. & MARRON, J. S. (1988). How far are automatically chosen regression smoothing parameters from their optimum? *J. Am. Statist. Assoc.* **83**, 86–95.

HÄRDLE, W. & MARRON, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist.* **13**, 1465–81.

HÄRDLE, W. & MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21**, 1926–47.

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

HARVEY, D. I. & MILLS, T. C. (2003). Modelling trends in central England temperatures. *J. Forecasting* **22**, 35–47.

HERRMANN, E., GASSER, T. & KNEIP, A. (1992). Choice of bandwidth for kernel regression when residuals are correlated. *Biometrika* **79**, 783–95.

HOEFFDING, W. & ROBBINS, H. (1948). The central limit theorem for dependent random variables. *Duke Math. J.* **15**, 773–780.

HOLLAND, G. J. (1997). The maximum potential intensity of tropical cyclones. *J. Atmosph. Sci.* **54**, 2519–41.

HOROWITZ J. L. & SPOKOINY, V. G. (2001) An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* **69**, 599–631 Q7

IOANNIDES, D. A. (1992). Integrated square error of nonparametric estimators of regression function: The fixed design case. *Statist. Prob. Lett.* **15**, 85–94.

JONES, P. D. & BRADLEY, R. S. (1992a). Climatic variations in the longest instrumental records. In *Climate Since A.D. 1500*, Ed. R. S. Bradley and P. D. Jones, London: Routledge. pp. 246–68.

JONES, P. D. & BRADLEY, R. S. (1992b). Climatic variations over the last 500 years. In *Climate Since A.D. 1500*, Ed. R. S. Bradley and P. D. Jones, London: Routledge. pp. 649–65.

JONES, P. D. & HULME, M. (1997). The changing temperature of Central England. In *Climates of the British Isles, Present, Past and Future*, Ed. M. Hulme and E. Barrow, London: Routledge. pp. 173–95.

KULASEKERA, K. B. & WANG, J. (1997). Smoothing parameter selection for power optimality in testing of regression curves. *J. Am. Statist. Assoc.* **92**, 500–11.

LIU, W. D. & WU, W. B. (2010). Asymptotics of spectral density estimates. *Economet. Theory* **26**, 1218–45.

MALLAT, S., PAPANICOLAOU, G. & ZHANG, Z. (1998). Adaptive covariance estimation of locally stationary processes. *Ann. Statist*. **26**, 1–47.

MANLEY, G. (1974). Central England temperatures: monthly means 1659 to 1973. *Quart. J. R. Meteorol. Soc.* **100**, 389–405.

NASON, G. P., VON SACHS, R. & KROISANDT, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Statist. Soc.* B. **62**, 271–92.

OMBAO, H., VON SACHS, R. & GUO, W. (2005). SLEX analysis of multivariate nonstationary time series. *J. Am. Statist. Assoc.* **100**, 519–31.

PARK, B. U. & MARRON, J. S. (1990). Comparison of data-driven bandwidth selectors. *J. Am. Statist. Assoc.* **85**, 66–72.

PAWLAK, M. & STADTMÜLLER, U. (1996). Recovering bank-limited signals under noise. *IEEE Trans. Inf. Theory* **42**, 1425–38.

PAWLAK, M. & STADTMÜLLER, U. (2007). Signal sampling and recovery under dependent errors. *IEEE Trans. Info. Theory* **53**, 2526–41.

PERCIVAL, D. B. & ROTHROCK, D. A. (2005). "Eyeballing" trends in climate time series: A cautionary note. *J. Climate* **18**, 886–91.

PRIESTLEY, M. B. (1965). Evolutionary spectra and non-stationary processes. *J. R. Statist. Soc.* B. **27**, 204–237.

PRIESTLEY, M. B. (1988). *Nonlinear and Nonstationary Time Series Analysis*. London: Academic Press.

PRIESTLEY, M. B. & CHAO, M. T. (1972). Non-parametric function fitting. *J. R. Statist. Soc.* B. **34**, 385–92.

RICE, J. (1984). Bandwidth choice for nonparametric regression. *Ann. Statist.* **12**, 1215–1230.

RUPPERT, D., SHEATHER, S. J. & WAND, M. P. (1995). An effective bandwidth selector for local least squares regression. *J. Am. Statist. Assoc.* **90**, 1257–70.

TONG, H. (1990). *Non-linear Time Series: A Dynamic System Approach*. Oxford: Oxford University Press.

VAN KEILEGOM, I., MANTEIGA, W. G. & SELLERO, C. S. (2008). Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *Test* **17**, 401–15.

WAND, M. P. & JONES, M. C. (1995). *Kernel Smoothing*. London: Chapman and Hall.

WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Nat. Acad. Sci.* **102**, 14150–54.

WU, W. B. & ZHAO, Z. B. (2007). Inference of trends in time series. *J. R. Statist. Soc.* B. **69**, 391–410.

WU, W. B. & ZHOU, Z. (2010). Gaussian approximations for non-stationary multiple time series. *Statist. Sinica*, To Appear. Q5

XIA, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *J. R. Statist. Soc.* B. **60**, 797–811.

ZHOU, Z. (2010). Nonparametric inference of quantile curves for nonstationary time series. *Ann. Statist.* **38**, 2187–217.

ZHOU, Z. & WU, W. B. (2009). Local linear quantile estimation for non-stationary time series. *Ann. Statist.* **37**, 2696–29.

[*Received April* 2010. *Revised February* 2011]