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 Iowa and 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 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

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

Received February 2012; revised April 2012.

[1]Supported in part by NSF Grant DMS-09-06073.

MSC2010 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.

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

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

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

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

where $\mathcal{F}_i = (\ldots, \boldsymbol{\varepsilon}_{i-1}, \boldsymbol{\varepsilon}_i)$ 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 $\mathbf{G}(t; \mathcal{F}_i)$ and $H(t; \mathcal{F}_i)$ 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 $\mathbf{J}(t; \mathcal{F}_k)$ is said to be stochastically Lipschitz continuous ($\mathbf{J} \in \text{Lip}$ in short) if there exists $C > 0$, such that $\|\mathbf{J}(t_1; \mathcal{F}_k) - \mathbf{J}(t_2; \mathcal{F}_k)\| \le C|t_1 - t_2|$ 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 $\mathcal{F}_{i,\{0\}} = (\ldots, \boldsymbol{\varepsilon}_{-1}, \boldsymbol{\varepsilon}'_0, \boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_i)$ be the coupled shift process. We define the functional dependence measure

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

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

(A1) Smoothness: $\boldsymbol{\beta} \in \mathcal{C}^3[0, 1]$;
(A2) Local stationarity: $\mathbf{G}, \mathbf{L} \in \mathrm{Lip}$;
(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* $\kappa_2 = \int_{-1}^1 v^2 K(v)\,dv$. *If* $b_n \asymp n^{-c}$ *for some* $1/6 < c < \min\{1/3, 1/2 - 1/(2\iota)\}$, *then*

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

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

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

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*

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

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

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

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

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

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* $0 < c < 1 - 1/\iota$,

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

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 $\boldsymbol{\Lambda}(\mathbf{L}, t)$, $t \in [0, 1]$. 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 $\boldsymbol{\Lambda}(\mathbf{L}, t)$, 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 $P_{b,l}(t)=\sum_{j=1}^{n}(t-j/n)^l K\{(j/n-t)/b\}$. 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 $\nu\ge 2/3$, 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°, $i=1,\dots,n$, and i.i.d. standard multivariate normal random vectors $\mathbf{x}_i^\circ$, $i=1,\dots,n$, 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 $\boldsymbol{\Gamma}_n = \{E(e_i e_j)\}_{1\le i,j\le n}$ 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

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

An estimate of the covariance matrix $\boldsymbol{\Gamma}_n$ 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,

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

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 $\hat{m}_n = \max\{k \ge 0 : |n^{-1/2}\sum_{i=1}^{n-k}\mathbf{L}_i^\top\mathbf{L}_{i+k}| > 1.96\sigma_k\}$. 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 $\iota \ge 5/2$ 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 $a_1(\cdot), \dots, a_p(\cdot)$ be continuous functions. We shall prove that the time-varying autoregressive process

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

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

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

and

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

Then (4.6) can be written as

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

We shall make the following assumptions:

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

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

(T3) $\sum_{j=1}^p a_j(t)z^j \neq 1$ for all $|z| \le 1 + c$ 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*

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

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 $\sum_{i=1}^k \mathbf{x}_i e_i$, $k = 1, \dots, n$. By Proposition 4.2, there exists a measurable function $\mathbf{G} \in \mathrm{Lip}$ such that

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

and

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

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).

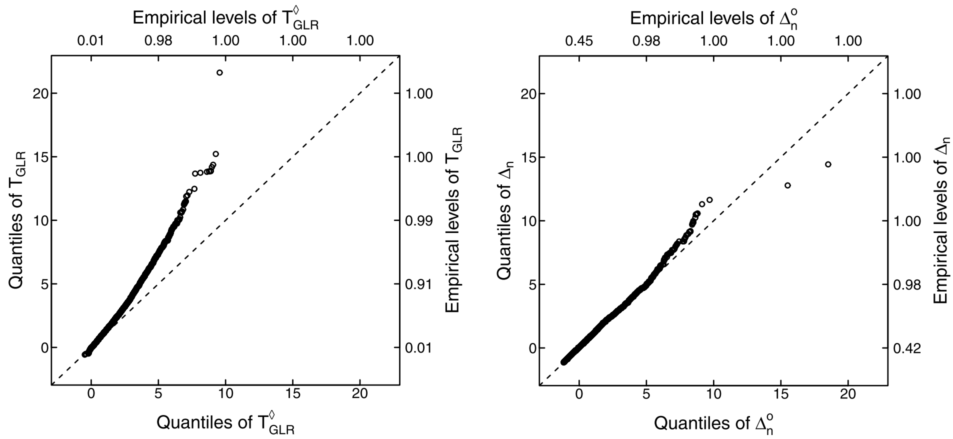
4.5. *A comparison with GLRT.* The generalized likelihood ratio test (GLRT, [18]) is a popular method for nonparametric hypothesis testing. It was used by Cai, Fan and Li [7] for testing the coefficient constancy, and generalized by Fan and Huang [16] to semiparametric models. Properties of GLRT have 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 $T_{\mathrm{GLR}}^\diamond$.



(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.*

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

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

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 $\Delta^{\circ}_n(\mathbf{A}, \hat{\mathbf{a}}, \mathbf{W})$. For T_{GLR}, the cut-off value is obtained by 5000 bootstrapped $T^{\diamond}_{\mathrm{GLR}}$. 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 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 $i = 1, \ldots, n$,

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

(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 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 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 $\hat{b}_n(\text{i}) = 0.25$ and $\hat{b}_n(\text{ii}) = 0.18$, 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

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

data so that each 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

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 $\hat{b}_n(\text{i}) = 0.25$ and $\hat{b}_n(\text{ii}) = 0.18$ 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
$c\hat{b}_n(\text{i})$	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
$\hat{b}_n(\text{i})$	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
$\hat{b}_n(\text{i})/c$	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
$c\hat{b}_n(\text{ii})$	92.6	96.6	92.2	96.1	92.1	96.2
$\hat{b}_n(\text{ii})$	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
$\hat{b}_n(\text{ii})/c$	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

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.

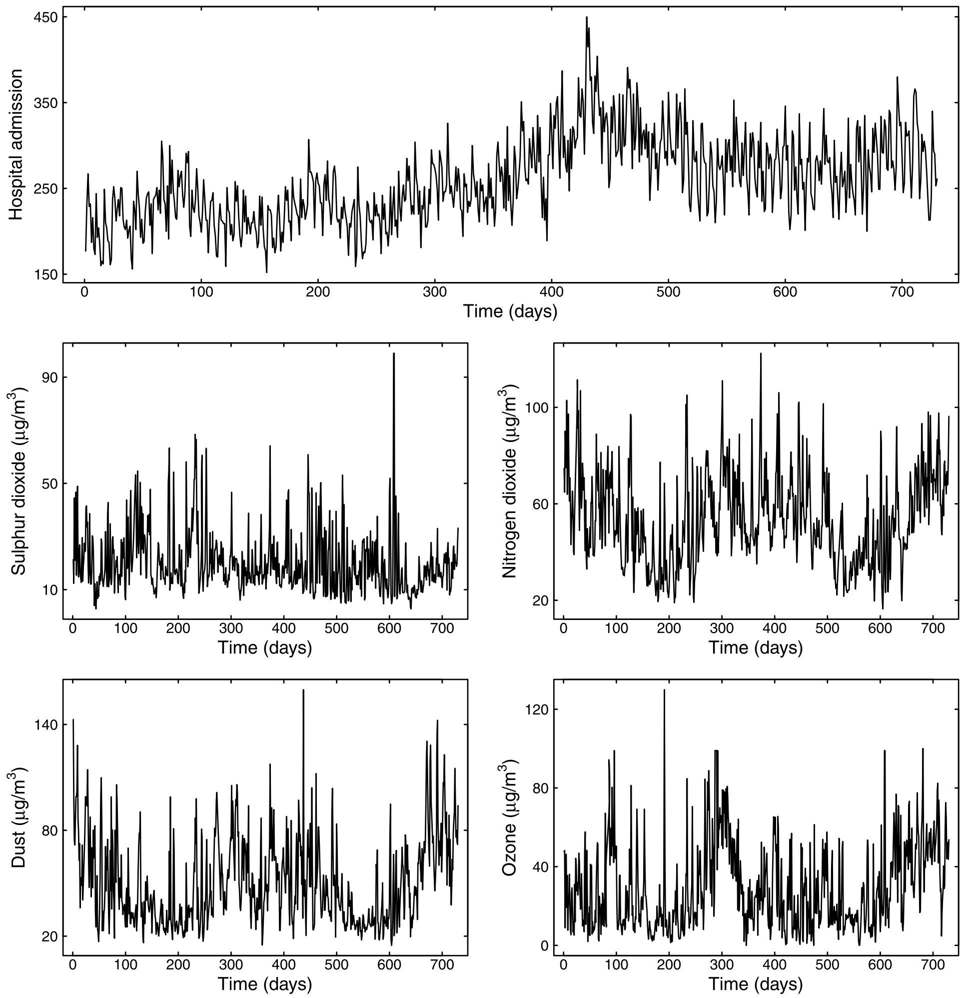


FIG. 2. *Time series plots for daily hospital admission* (*top*) *and levels of sulphur dioxide* (*middle left*), *nitrogen dioxide* (*middle right*), *dust* (*bottom left*) *and ozone* (*bottom right*) *from January* 1, 1994 *to December* 31, 1995.

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

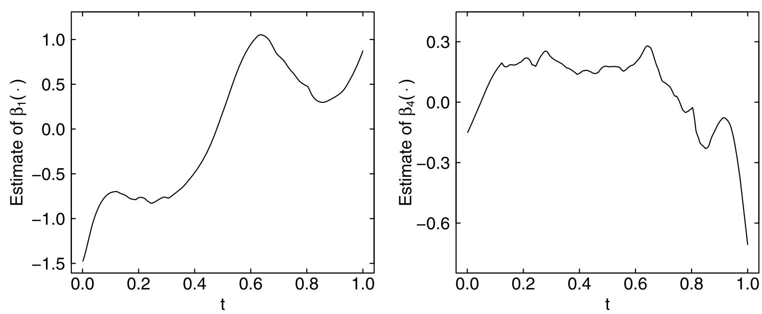


FIG. 3. *Plots of estimated coefficient functions*, $\beta_1(\cdot)$ (*left*) *and* $\beta_4(\cdot)$ (*right*).

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 $\underline{\rho}(\mathbf{A}) = \inf\{|\mathbf{Av}| : |\mathbf{v}| = 1\}$ and $\overline{\rho}(\mathbf{A}) = \sup\{|\mathbf{Av}| : |\mathbf{v}| = 1\}$. 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* $\overline{\rho}(\mathbf{AB}) \le \overline{\rho}(\mathbf{A})\overline{\rho}(\mathbf{B})$ *and* $\underline{\rho}(\mathbf{AB}) \ge \underline{\rho}(\mathbf{A})\underline{\rho}(\mathbf{B})$; *and* (iv) $\overline{\rho}(\mathbf{aa}^\top) = |\mathbf{a}|^2$ *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 $\boldsymbol{\mathcal{F}}_{i,j} = (\boldsymbol{\varepsilon}_i, \ldots, \boldsymbol{\varepsilon}_j)$, $i \le j$. 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 $t_{i,n} = i/n$, $i = 1, \ldots, n$. 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 $\mathbf{D}_{k,l,n} = E\{\boldsymbol{\vartheta}_k(t_{k,n})|\mathcal{F}_{k-l,k}\} - E\{\boldsymbol{\vartheta}_k(t_{k,n})|\mathcal{F}_{k-l+1,k}\}$. 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 $\|\mathbf{D}_{k,l,n}\|_q = \|E\{\boldsymbol{\vartheta}_l(t_{k,n})|\mathcal{F}_{0,l}\} - E\{\boldsymbol{\vartheta}_{l,\{0\}}(t_{k,n})|\mathcal{F}_{0,l}\}\|_q \le \delta_{l,q}(\mathbf{J})$, (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 $\tilde{\boldsymbol{\vartheta}}_k(t) = E\{\boldsymbol{\vartheta}_k(t)|\mathcal{F}_{k-m,k}\}$ 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 $\sup_{t\in[0,1]}\|\mathbf{J}(t;\mathcal{F}_0)\|_\iota < \infty$, $\iota > 2$, 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*

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

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,

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

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

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

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

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

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

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

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

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

Therefore,

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

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

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

The Lipschitz continuity implies

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

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 $\sum_{k=0}^{\infty}\{\gamma_{k,2}(\mathbf{J}) \wedge a_n\} = \mathcal{O}(r_n a_n + r_n^{-\nu})$, (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*

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

and

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

PROOF. Let $\boldsymbol{\zeta}_k(t) = \mathbf{L}(t; \mathcal{F}_k)$ and $\tilde{\boldsymbol{\zeta}}_k(t) = E\{\boldsymbol{\zeta}_k(t) | \mathcal{F}_{k-m,k}\}$ 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

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

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

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

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

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

where

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

Since $\sum_{i=1}^{n} \mathrm{tr}\{\mathbf{Q}_{i,i,n}\boldsymbol{\Lambda}(\mathbf{L}, t_{i,n})\} = (nb_n)^{-1} K^*(0)\Lambda_{\mathbf{W}_0,1} + o(n^{-1}b_n^{-1/2})$, (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, $nb_n r_{1,n} = O\{b_n^{\nu/(1+\nu)}\} = o(b_n^{1/2})$ 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 $nb_n r_{2,n} = O\{(nb_n)^{-\nu/(2+\nu)}\} = o(b_n^{1/2})$ 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,

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

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

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

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,

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

Since, by Lemma A.2,

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

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),

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

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

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

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,

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

By (A.4), we have

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

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 non-parametric 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