Journal of the
Royal Statistical Society
SERIES B
Statistical
Methodology

J. R. Statist. Soc. B (2010)
72, *Part* 4, *pp.* 513–531

Simultaneous inference of linear models with time varying coefficients

Zhou Zhou

University of Toronto, Canada

and Wei Biao Wu

University of Chicago, USA

[Received December 2008. Revised January 2010]

Summary. The paper considers construction of simultaneous confidence tubes for time varying regression coefficients in functional linear models. Using a Gaussian approximation result for non-stationary multiple time series, we show that the constructed simultaneous confidence tubes have asymptotically correct nominal coverage probabilities. Our results are applied to the problem of testing whether the regression coefficients are of certain parametric forms, which is a fundamental problem in the inference of functional linear models. As an application, we analyse an environmental data set and study the association between levels of pollutants and hospital admissions.

Keywords: Functional linear models; Gaussian approximation; Local linear regression; Local stationarity; Non-parametric goodness-of-fit tests; Non-stationary time series; Simultaneous confidence tubes

1. Introduction

Consider the time varying coefficients linear model

$$Y(t) = \mathbf{X}^{\mathrm{T}}(t)\,\boldsymbol{\beta}(t) + \varepsilon(t), \tag{1}$$

where t is the time index, $\mathbf{X}(\cdot)$, $\boldsymbol{\beta}(\cdot)$, $\varepsilon(\cdot)$ and $Y(\cdot)$ represent the $(p \times 1)$-dimensional covariate process, time varying regression coefficients, residual process and response process respectively. Here 'T' denotes matrix transpose. The model has been studied in Hoover *et al.* (1998), Wu *et al.* (1998), Fan and Zhang (2000), Lin and Ying (2001), Huang *et al.* (2004) and Ramsay and Silverman (2005) among others. Many of these references consider estimation of $\boldsymbol{\beta}(\cdot)$ in the longitudinal setting where many subjects are measured at multiple times. In our setting, however, we assume that only one realization is available and $\mathbf{X}(\cdot)$ and $Y(\cdot)$ are observed at time points $t_i = i/n, 1 \leqslant i \leqslant n$; see Section 6 for an application to the Hong Kong circulatory and respiratory data. Then model (1) becomes

$$y_i = \mathbf{x}_i^{\mathrm{T}} \boldsymbol{\beta}_i + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{2}$$

where $\mathbf{x}_i = \mathbf{X}(t_i)$, $\varepsilon_i = \varepsilon(t_i)$, $y_i = Y(t_i)$, $\boldsymbol{\beta}_i = \boldsymbol{\beta}(t_i)$ and $\boldsymbol{\beta}(\cdot)$ is a smooth function on $[0, 1]$. We assume that both $(\mathbf{x}_i)$ and (ε_i) are locally stationary processes: a special class of non-stationary processes. Our formulation is attractive in situations in which we expect that the underlying

Address for correspondence: Zhou Zhou, Department of Statistics, University of Toronto, 100 St George Street, Toronto, Ontario, M5S 3G3, Canada.
E-mail: zhou@utstat.toronto.edu

 1369–7412/10/72513

data-generating mechanisms change smoothly over time. We shall study asymptotic properties of estimates of $\boldsymbol{\beta}(\cdot)$ on the basis of the data $(\mathbf{x}_i^{\mathrm{T}}, y_i)_{i=1}^n$. The time varying model (2) has applications in various areas including environmental science, finance and econometrics. It has attracted considerable attention; see Robinson (1989, 1991), Orbe *et al.* (2005, 2006) and Cai (2007) among others.

With the regression parameter $\boldsymbol{\beta}(\cdot)$ being time varying, we shall be able to explore the dynamic associations between the response series (y_i) and the explanatory series $(\mathbf{x}_i)$. For example, we may be interested in testing whether a particular coefficient function $\beta_j(\cdot)$ is different from 0, or whether it is really time varying or whether it is increasing in time. To address the latter two questions, we shall construct simultaneous confidence regions instead of pointwise confidence intervals which do not reflect the overall pattern of the regression functions. Specifically, let $\mathbf{C}$ be a fixed $p \times s$ matrix with rank $s \leqslant p$, and $\boldsymbol{\beta}_{\mathbf{C}}(\cdot) = \mathbf{C}^{\mathrm{T}} \boldsymbol{\beta}(\cdot)$ be a linear combination of the regression function $\boldsymbol{\beta}(\cdot)$. For some preassigned significance level $\alpha \in (0, 1)$, we shall construct in $\mathbb{R}^s \times [0, 1]$ a $100(1-\alpha)\%$ asymptotic simultaneous confidence tube (SCT) $\{\Upsilon_\alpha(t), 0 \leqslant t \leqslant 1\}$ for $\boldsymbol{\beta}_{\mathbf{C}}(\cdot)$ in the sense that

$$\lim_{n\to\infty} [\mathbb{P}\{\boldsymbol{\beta}_{\mathbf{C}}(t) \in \Upsilon_\alpha(t), 0 \leqslant t \leqslant 1\}] = 1 - \alpha. \tag{3}$$

We can apply limit (3) to test hypotheses on patterns of $\boldsymbol{\beta}(\cdot)$. For example, to test whether $\beta_1(\cdot)$ is a constant function, we can let $s=1$ and $\mathbf{C}^{\mathrm{T}} = (1, 0, \ldots, 0)$. If a horizontal line can be embedded in the SCT $\{\Upsilon_\alpha(\cdot)\}$, then we accept the constancy hypothesis. Simultaneous confidence regions are more informative and they provide an important means to address the overall variability of the estimated curves. For the pointwise version, one constructs $\dot{\Upsilon}_\alpha(t)$ such that, for all $t \in [0, 1]$, $\lim_{n\to\infty}[\mathbb{P}\{\boldsymbol{\beta}_{\mathbf{C}}(t) \in \dot{\Upsilon}_\alpha(t)\}] = 1 - \alpha$.

The construction of simultaneous confidence regions has been a very difficult problem when dependence between errors is present. Earlier researchers obtained *conservative* simultaneous confidence bands since the Bonferroni correction procedure is used (Wu *et al.*, 1998; Huang *et al.*, 2004). Such simultaneous confidence bands are usually too wide and are thus of limited use. Assuming that the observations are independent, we can construct simultaneous confidence bands that have asymptotically correct coverage probabilities; see Bickel and Rosenblatt (1973), Eubank and Speckman (1993), Johnston (1982), Neumann and Polzehl (1998) and Fan and Zhang (2000) among others.

As a key step in the construction of simultaneous confidence bands that have asymptotically correct coverage probabilities, we need to obtain a limit theory for the normalized maximum deviations of the estimated functions from the true functions. This problem can be solved if we have

(a) an extreme value theory of Gaussian processes and
(b) a Gaussian approximation result for the partial sum process or the empirical process.

Wu *et al.* (1998) mentioned that the fundamental difficulty is (b), namely there had been no development on Gaussian approximations when dependence is present.

Recently Wu and Zhou (2009) obtained a Gaussian approximation principle for non-stationary multiple time series with nearly optimal rates. With their result, together with the deep extreme value theory of vector-valued Gaussian processes, we can construct SCTs that have asymptotically correct coverage probabilities.

The rest of the paper is organized as follows. Section 2 imposes model assumptions and dependence structures on $(\mathbf{x}_i)$ and (ε_i). Asymptotic theory for the local linear estimate of $\boldsymbol{\beta}(\cdot)$ is presented in Section 3. Section 4 deals with various issues in constructing the SCT, including bias correction, bootstrap simulations and estimation of covariance matrices. Section 5 provides a simulation study on the performance of our SCTs. Section 6 shows an application to a specification test and we study the relationship between levels of pollutants and hospital admissions in the Hong Kong pollution data. Proofs are given in Appendix A.

2. Model assumptions

We shall estimate the regression coefficient function $\beta(\cdot)$ by the local linear approach (Fan and Gijbels, 1996); see Section 3. To conduct asymptotic analysis of the estimates, we need to impose structural assumptions on the covariate process $(\mathbf{x}_i)$ and the error process (ε_i). Let ζ_i, $i \in \mathbb{Z}$, be independent and identically distributed (IID) random variables and $\mathcal{F}_i = (\ldots, \zeta_{i-1}, \zeta_i)$. We assume that both $(\mathbf{x}_i)$ and (ε_i) are locally stationary processes (Draghicescu *et al.*, 2009), a special class of non-stationary time series:

$$\mathbf{x}_i = \mathbf{G}(t_i, \mathcal{F}_i) \quad \text{and} \quad \varepsilon_i = H(t_i, \mathcal{F}_i), \qquad i = 1, \ldots, n, \tag{4}$$

where $\mathbf{G} := (G_1, \ldots, G_p)^{\mathrm{T}}$ and $H(\cdot, \cdot)$ are measurable functions such that $\mathbf{G}(t, \mathcal{F}_i)$ and $H(t, \mathcal{F}_i)$ are well defined for each $t \in [0, 1]$ and $\mathbb{E}(\varepsilon_i | \mathbf{x}_i) = 0$. To help to understand the formulation, we shall consider two special cases.

(a) Independent model: assume that $(\varepsilon_i)_{i \in \mathbb{Z}}$ are IID, $(\xi_i)_{i \in \mathbb{Z}}$ are also IID and $(\varepsilon_i)_{i \in \mathbb{Z}}$ is independent of $(\xi_i)_{i \in \mathbb{Z}}$. Let $\zeta_i = (\varepsilon_i, \xi_i)^{\mathrm{T}}$, $\mathbf{G}(t, \mathcal{F}_i) = \mathbf{G}_0\{t, (\ldots, \xi_{i-1}, \xi_i)\}$ and $H(t, \mathcal{F}_i) = H_0\{t, (\ldots, \varepsilon_{i-1}, \varepsilon_i)\}$, where $\mathbf{G}_0$ and H_0 are measurable functions. In this case the predictors and errors are two independent non-stationary processes. Under some further restrictions, this type of model was studied in Robinson (1989) and Orbe *et al.* (2005, 2006).

(b) Heteroscedastic model: define $\mathbf{x}_i = \mathbf{G}(t_i, \mathcal{F}_i)$ as in the independent model (a). Let $\varepsilon_i = B\{t_i, (\ldots, \xi_{i-1}, \xi_i)\}\, H_0\{t_i, (\ldots, \varepsilon_{i-1}, \varepsilon_i)\}$. Then the errors and the covariates are dependent. Such models are suitable when the errors exhibit heteroscedasticity with respect to time and independent variables. A special case was considered in Cai (2007). If $H_0\{t, (\ldots, \varepsilon_{i-1}, \varepsilon_i)\}$ has mean 0 and variance 1, then B^2 is the conditional variance of ε_i given (ε_j).

We can interpret series (4) as physical systems with $\mathcal{F}_i$ (and $\mathbf{x}_i$ and ε_i) being the inputs (and the outputs respectively), and $\mathbf{G}$ and H being the transforms that represent the underlying physical mechanism. By allowing $\mathbf{G}$ and H to vary smoothly in t, we have local stationarity (see conditions 2 and 3 in Section 3.1). Our formulation is different from the locally stationary processes in Dahlhaus (1997) who discussed time varying spectral representations.

To facilitate an asymptotic study of estimates of $\beta(\cdot)$, we shall introduce appropriate time series dependence measures. For a vector $\mathbf{v} = (v_1, \ldots, v_p) \in \mathbb{R}^p$, let $|\mathbf{v}| = (\Sigma_{j=1}^p v_j^2)^{1/2}$. For a random vector $\mathbf{V}$, write $\mathbf{V} \in \mathcal{L}^q$ $(q > 0)$ if $\|\mathbf{V}\|_q := \mathbb{E}(|\mathbf{V}|^q)^{1/q} < \infty$. Following Wu (2005), we have the following definition.

Definition 1 (physical dependence measures). Assume, for all $t \in [0, 1]$, $\mathbf{L}(t, \mathcal{F}_j) \in \mathcal{L}^q$, $q > 0$. Let $(\zeta'_j)_{j \in \mathbb{Z}}$ be an IID copy of $(\zeta_j)_{j \in \mathbb{Z}}$. For $j \geqslant 0$ let $\mathcal{F}_j^* = (\mathcal{F}_{-1}, \zeta'_0, \zeta_1, \ldots, \zeta_{j-1}, \zeta_j)$. Define the physical dependence measure for the stochastic system $\mathbf{L}(t, \mathcal{F}_j)$ as

$$\delta_q(\mathbf{L}, j) = \sup_{t \in [0,1]} \{\|\mathbf{L}(t, \mathcal{F}_j) - \mathbf{L}(t, \mathcal{F}_j^*)\|_q\}. \tag{5}$$

If $\mathbf{L}(t, \mathcal{F}_j)$ does not functionally depend on the input ζ_0, then $\delta_q(\mathbf{L}, j) = 0$. So $\delta_q(\mathbf{L}, j)$ measures the dependence of the output $\mathbf{L}(t, \mathcal{F}_j)$ on the input ζ_0. If $j < 0$, $\delta_q(\mathbf{L}, j) = 0$. The above dependence measure is closely related to the data-generating mechanism and it is easy to work with. Section 4 of Zhou and Wu (2009) contains calculations of $\delta_q(\mathbf{L}, k)$ for some locally stationary linear and non-linear processes. Our input–output-based dependence measure is different from the classical strong mixing conditions and the measure that was proposed in Doukhan and Louhichi (1999) which concerns covariances of past and future values of a process. Under suitable conditions on physical dependence measures, we can have Gaussian approximations with nearly optimal rates (Wu, 2007; Liu and Lin, 2009).

3. Asymptotic theory

In this section we shall present an asymptotic theory for the local linear estimate of $\boldsymbol{\beta}(\cdot)$ from model (2). We first define the local linear estimate. Since $\boldsymbol{\beta}(s) \approx \boldsymbol{\beta}(t) + (s-t)\,\boldsymbol{\beta}'(t)$ for s close to t, it is natural to estimate $\boldsymbol{\beta}(\cdot)$ and $\boldsymbol{\beta}'(\cdot)$ by

$$(\hat{\boldsymbol{\beta}}_{b_n}(t), \hat{\boldsymbol{\beta}}'_{b_n}(t)) = \underset{\eta_0,\eta_1 \in \mathbb{R}^p}{\arg\min} \left[\sum_{i=1}^{n} \{y_i - \mathbf{x}_i^{\mathrm{T}}\eta_0 - \mathbf{x}_i^{\mathrm{T}}\eta_1(t_i - t)\}^2 K_{b_n}(t_i - t) \right], \tag{6}$$

where K is a kernel function, $b_n > 0$ is the bandwidth and $K_c(\cdot) = K(\cdot/c)$, $c > 0$. Throughout this paper we shall always assume that the kernel $K \in \mathcal{K}$, the collection of symmetric density functions K with support $[-1,1]$ and $K \in \mathcal{C}^1[-1,1]$. A popular choice is the Epanechnikov kernel $K(u)$ which is $3(1-u^2)/4$ if $|u| \leqslant 1$ and 0 if $|u| > 1$. We can interpret equation (6) as the weighted least squares estimate of the linear model $y_i = \mathbf{x}_i^{\mathrm{T}}\eta_0 + \mathbf{x}_i^{\mathrm{T}}\eta_1(t_i - t) + e_i$ with weights $K_{b_n}(t_i - t), n(t - b_n) \leqslant i \leqslant n(t + b_n)$. Define

$$\mathbf{S}_{n,l}(t) = (nb_n)^{-1} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^{\mathrm{T}} \{(t_i - t)/b_n\}^l K_{b_n}(t_i - t),$$

for $l = 0, 1, \ldots$; here we let $0^0 = 1$, and

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

Let $\hat{\boldsymbol{\eta}}_{b_n}(t) = (\hat{\boldsymbol{\beta}}^{\mathrm{T}}_{b_n}(t), b_n(\hat{\boldsymbol{\beta}}'_{b_n}(t))^{\mathrm{T}})^{\mathrm{T}}$. Then

$$\hat{\boldsymbol{\eta}}_{b_n}(t) = \begin{pmatrix} \mathbf{S}_{n,0}(t) & \mathbf{S}^{\mathrm{T}}_{n,1}(t) \\ \mathbf{S}_{n,1}(t) & \mathbf{S}_{n,2}(t) \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{R}_{n,0}(t) \\ \mathbf{R}_{n,1}(t) \end{pmatrix} := \mathbf{S}_n^{-1}(t)\,\mathbf{R}_n(t). \tag{7}$$

We shall omit the subscript b_n in $\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}}'$ hereafter if no confusion will be caused. Section 3.1 presents a central limit theorem for $\hat{\boldsymbol{\beta}}(\cdot)$. Previously when dealing with the time varying coefficients model (2), either $(\mathbf{x}_i)$ or (ε_i) or both were assumed to be stationary and strong mixing (Cai, 2007). Our dependence measure seems more convenient for constructing SCTs with locally stationary covariates and error processes. Section 3.3 concerns the uniform behaviour of $\{\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t), t \in [b_n, 1 - b_n]\}$.

3.1. Asymptotic normality

For a family of stochastic processes $(\mathbf{L}(t, \mathcal{F}_i))_{i \in \mathbb{Z}}$, we say that it is $\mathcal{L}^q$ stochastic Lipschitz continuous on [0, 1] if $\sup_{0 \leqslant s < t \leqslant 1}\{\|\mathbf{L}(t, \mathcal{F}_0) - \mathbf{L}(s, \mathcal{F}_0)\|_q/(t - s)\} < \infty$. Denote by Lip_q the collection of such systems. Let $\mathcal{C}^l\mathcal{I}$, $l \in \mathbb{N}$, be the collection of functions that have ith-order continuous derivatives on the interval $\mathcal{I} \subset \mathbb{R}$. We shall make the following assumptions.

Assumption 1. The smallest eigenvalue of $M(t) := \mathbb{E}\{\mathbf{G}(t, \mathcal{F}_0)\,\mathbf{G}(t, \mathcal{F}_0)^{\mathrm{T}}\}$ is bounded away from 0 on [0, 1].

Assumption 2. $\mathbf{G}(t, \mathcal{F}_i) \in \mathrm{Lip}_2$ and $\sup_{0 \leqslant t \leqslant 1}\{\|\mathbf{G}(t, \mathcal{F}_i)\|_4\} < \infty$.

Assumption 3. $\mathbf{U}(t,\mathcal{F}_j) := \mathbf{G}(t,\mathcal{F}_j)\,H(t,\mathcal{F}_j) \in \mathrm{Lip}_2$ and $\sup_{0\leqslant t\leqslant 1}\{\|\mathbf{U}(t,\mathcal{F}_j)\|_r\} < \infty, r \geqslant 2$.

Assumption 4. Short-range dependence condition: $\Sigma_{k=0}^{\infty}\{\delta_4(\mathbf{G},k)+\delta_2(\mathbf{U},k)\} < \infty$.

Assumption 5. The smallest eigenvalue of $\Lambda(t)$ is bounded away from 0 on [0, 1], where

$$\Lambda(t) = \sum_{j=-\infty}^{\infty} \mathrm{cov}\{\mathbf{U}(t,\mathcal{F}_0),\mathbf{U}(t,\mathcal{F}_j)\}. \tag{8}$$

Assumption 6. The coefficient function $\boldsymbol{\beta}(\cdot) \in \mathcal{C}^3[0,1]$, namely $\beta_j(\cdot) \in \mathcal{C}^3[0,1], j = 1,\ldots,p$.

Under assumption 2, $M(t)$ is well defined and is Lipschitz continuous on [0, 1]. Conditions 3 and 4 guarantee that $\Lambda(t)$ is positive definite and continuous on [0, 1]. Let

$$\mu_l = \mu_{l,K} = \int_{\mathbb{R}} x^l K(x)\,\mathrm{d}x \quad \text{and} \quad \phi_l = \phi_{l,K} = \int_{\mathbb{R}} x^l K^2(x)\,\mathrm{d}x, \qquad l = 0,1,\ldots.$$

Theorem 1. Let $\Sigma(t) = M^{-1}(t)\,\Lambda(t)\,M^{-1}(t)$ and $\boldsymbol{\beta}''(t)$ be the second-order derivative of $\boldsymbol{\beta}(t)$. Assume that conditions 1–6 hold with $r=2$, $nb_n \to \infty$ and $nb_n^7 \to 0$. Then, for any fixed $t \in (0,1)$,

$$(nb_n)^{1/2}\{\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t) - b_n^2\boldsymbol{\beta}''(t)\mu_2/2\} \Rightarrow N\{0,\phi_0\,\Sigma(t)\}. \tag{9}$$

We now comment on the regularity conditions of theorem 1. Condition 1 avoids asymptotic multicollinearity and assumption 5 prevents singularity of the limiting asymptotic covariance matrix. Condition 2 means local stationarity in the sense that, for a sequence $m \to \infty$ with $m/n \to 0$, the process $(\mathbf{x}_i)_{i=l-m}^{l+m}$ can be approximated by the *stationary* process $(\mathbf{G}(t_l,\mathcal{F}_i))_{i=l-m}^{l+m}$ in view of $\sup_{l-m\leqslant i\leqslant l+m}\|\mathbf{x}_i - \mathbf{G}(t_l,\mathcal{F}_i)\| = O(m/n) = o(1)$, by the $\mathcal{L}^2$ stochastic Lipschitz continuity. If $\mathbf{G}(t,\mathcal{F}_i)$ does not depend on t, then it becomes a stationary process. Here assumptions 2 and 3 can be checked by the results in Section 4 of Zhou and Wu (2009). The short-range dependence condition 4 means that the cumulative effect of ζ_0 on future values is bounded and it is easily verifiable for a large class of locally stationary processes.

3.2. Gaussian approximations

Let $\{\mathbf{Z}_i = \mathbf{W}(t_i,\mathcal{F}_i)\}_{i=1}^n$ be an s-dimensional locally stationary process with mean 0. We consider approximating the partial sum process $S_{\mathbf{Z}}(l) = \Sigma_{i=1}^{l}\mathbf{Z}_i$ by multiple Gaussian processes. This problem has an extensive history; see the references in Wu (2007) and Liu and Lin (2009). Wu and Zhou (2009) obtained a Gaussian approximation result with nearly optimal bounds for non-stationary multiple time series. Let Id_s denote the s-dimensional identity matrix. For a positive semidefinite matrix A with eigendecomposition $A = QDQ^{\mathrm{T}}$, where Q is orthonormal and D is a diagonal matrix, define $A^{1/2} = QD^{1/2}Q^{\mathrm{T}}$, where $D^{1/2}$ is the elementwise root of D.

Theorem 2. Assume

(a) $\sup_{0\leqslant t\leqslant 1}\{\|\mathbf{W}(t,\mathcal{F}_l)\|_4\} < \infty$,
(b) $\mathbf{W}(t,\mathcal{F}_l) \in \mathrm{Lip}_2$ and
(c) $\delta_4(\mathbf{W},k) = O(k^{-2})$.

Then, on a richer probability space, there is an IID $\mathbf{V}_1,\mathbf{V}_2,\ldots,\sim N(0,\mathrm{Id}_s)$ and a process $S_{\mathbf{Z}}^0(i)$ such that $\{S_{\mathbf{Z}}(i)\}_{i=0}^n =^{\mathcal{D}} \{S_{\mathbf{Z}}^0(i)\}_{i=0}^n$ and

$$\max_{i\leqslant n}\left|S_{\mathbf{Z}}^0(i) - \sum_{j=1}^{i}\Sigma_{\mathbf{W},j}\mathbf{V}_j\right| = o_{\mathbb{P}}\{n^{3/10}\log(n)\}, \tag{10}$$

where $\Sigma_{\mathbf{W},j} = \Sigma_{\mathbf{W}}(t_j)$,

$$\Sigma_{\mathbf{W}}(t) = \left[\sum_{l \in \mathbb{Z}} \operatorname{cov}\{\mathbf{W}(t, \mathcal{F}_0), \mathbf{W}(t, \mathcal{F}_l)\} \right]^{1/2}. \tag{11}$$

Theorem 2 follows from corollary 2 in Wu and Zhou (2009). Owing to the non-stationarity, the approximated Gaussian process $(\Sigma_{j=1}^{i} \Sigma_{\mathbf{W},j} \mathbf{V}_j)_{i=1}^{n}$ has independent but possibly non-identically distributed increments. The covariance matrix of the increments $\Sigma^2_{\mathbf{W},i}$ equals the long-run covariance matrices of series $(\mathbf{Z}_i)$ as defined in equation (11) and it accounts for the dependence of the series. The Gaussian approximation is the key theoretical tool for proving asymptotic properties of the SCT in theorem 3 in Section 3.3. It also suggests a bootstrap procedure for the construction of the SCT and non-parametric supremum-type tests; see Section 4.2.

3.3. *Maximum deviations*

Theorem 3 is for constructing SCTs in the sense of expression (3). Let $\mathbf{C}$ be a fixed $p \times s$ matrix with rank $s \leqslant p$, $\mathbf{A}_{\mathbf{C}}(t) = M^{-1}(t)\mathbf{C}$, $\Sigma^2_{\mathbf{C}}(t) = \mathbf{A}^{\mathrm{T}}_{\mathbf{C}}(t)\,\Lambda(t)\mathbf{A}_{\mathbf{C}}(t)$ and $\hat{\boldsymbol{\beta}}_{\mathbf{C}}(t) = \mathbf{C}^{\mathrm{T}}\,\hat{\boldsymbol{\beta}}(t)$.

Theorem 3. Assume that conditions 1–6 hold with $r = 4$. Further assume that

(a) $\delta_4(\mathbf{U}, k) = O(k^{-2})$,
(b) $\Sigma_{k=0}^{\infty} \delta_4(\mathbf{G}, k) < \infty$,
(c) $\Lambda(t)$ is Lipschitz continuous on [0, 1] and
(d) $\log^3(n)/n^{2/5} b_n + n b_n^7 \log(n) \to 0$.

Then, as $n \to \infty$, we have

$$\mathbb{P}\left[\frac{\surd(n b_n)}{\surd\phi_0} \sup_{t \in \mathcal{T}} \left| \Sigma^{-1}_{\mathbf{C}}(t) \left\{ \hat{\boldsymbol{\beta}}_{\mathbf{C}}(t) - \boldsymbol{\beta}_{\mathbf{C}}(t) - \frac{\mu_2 b_n^2 \boldsymbol{\beta}''_{\mathbf{C}}(t)}{2} \right\} \right| - B_K(m^*) \leqslant \frac{u}{\surd\{2\log(m^*)\}} \right] = \exp\{-2\exp(-u)\}, \tag{12}$$

where $\boldsymbol{\beta}''_{\mathbf{C}}(\cdot) = \mathbf{C}^{\mathrm{T}}\,\boldsymbol{\beta}''(\cdot)$, $\mathcal{T} = [b_n, 1 - b_n]$, $m^* = 1/b_n$ and

$$B_K(m^*) = \surd\{2\log(m^*)\} + \frac{\log(C_K) + (s/2 - \frac{1}{2})\log\{\log(m^*)\} - \log(2)}{\surd\{2\log(m^*)\}} \tag{13}$$

with

$$C_K = \frac{\left\{ \int_{-1}^{1} |K'(u)|^2 \,\mathrm{d}u / \phi_0 \pi \right\}^{1/2}}{\Gamma(s/2)}.$$

4. Construction of simultaneous confidence tube

We now apply theorem 3 to construct an SCT. Let $\hat{\boldsymbol{\beta}}''_{\mathbf{C}}(t)$ and $\hat{\Sigma}_{\mathbf{C}}(t)$ be uniformly consistent estimates of $\boldsymbol{\beta}''_{\mathbf{C}}(t)$ and $\Sigma_{\mathbf{C}}(t)$ respectively. Let $\alpha \in (0, 1)$. Then the SCT of $\boldsymbol{\beta}_{\mathbf{C}}(t)$

$$\hat{\boldsymbol{\beta}}_{\mathbf{C}}(t) - \frac{\mu_2 b_n^2 \hat{\boldsymbol{\beta}}''_{\mathbf{C}}(t)}{2} + \frac{\surd\phi_0}{\surd(n b_n)} \left[B_K(m^*) - \frac{\log[\log\{(1-\alpha)^{-1/2}\}]}{\surd\{2\log(m^*)\}} \right] \hat{\Sigma}_{\mathbf{C}}(t) \mathcal{B}_s, \tag{14}$$

where $\mathcal{B}_s = \{\mathbf{u} \in \mathbb{R}^s : |\mathbf{u}| \leqslant 1\}$ is the unit ball, has asymptotic coverage probability $1 - \alpha$. Sections 4.1–4.4 concern some implementational issues.

4.1. Bias correction

To use expression (12) or (14), we need to deal with the bias term that involves $\beta''(t)$. It is generally not easy to obtain a good estimate of $\beta''(t)$. One way out is to use undersmoothing by choosing a bandwidth $b_n = o(n^{-1/5})$; see for example the discussion in Neumann and Polzehl (1998). However, as mentioned in Neumann and Polzehl (1998), it is unclear how to choose such a bandwidth and most automatic bandwidth selectors choose the mean-squared error optimal bandwidth $b_n \sim cn^{-1/5}$ with some $c > 0$. Here, without estimating $\beta''(t)$, we shall use the simple jackknife bias-corrected estimator:

$$\tilde{\boldsymbol{\beta}}_{b_n}(t) = 2\hat{\boldsymbol{\beta}}_{b_n/\sqrt{2}}(t) - \hat{\boldsymbol{\beta}}_{b_n}(t). \tag{15}$$

The bias of $\tilde{\boldsymbol{\beta}}(t)$ is of order $o(b_n^3)$ and is asymptotically negligible under the conditions of theorem 3. Implementing estimator (15) is asymptotically equivalent to using the fourth-order kernel $K^*(x) = 2\sqrt{2}K(\sqrt{2}x) - K(x)$. Then the mean-squared error optimal bandwidth is of the form $cn^{-1/9}$ for some $c > 0$. In our data analysis, we recommend using $b_n' = 2\hat{b}_n$, for this biased-corrected estimator, where $\hat{b}_n$ is the bandwidth selected for the original local linear estimator.

4.2. Convergence issues

We see from equation (12) that the convergence therein is of logarithmic rate and it is very slow. To circumvent the problem, we shall adopt a bootstrap method which can have a better performance (Härdle and Marron, 1991; Hall, 1991; Neumann and Kreiss, 1998). The key ingredient is the following proposition. Let $\tilde{\boldsymbol{\beta}}_{\mathbf{C}}(t) = \mathbf{C}^{\mathrm{T}}\tilde{\boldsymbol{\beta}}(t)$.

Proposition 1. Assume that conditions 1–3 and 5 hold with $r = 4$. Further assume conditions (a)–(c) of theorem 3 and $b_n = O(n^{-\theta})$ with $1/7 < \theta < 2/5$. Then, on a richer probability space, there are IID $\mathbf{V}_1, \mathbf{V}_2, \ldots, \sim N(0, \mathrm{Id}_s)$ such that

$$\sup_{t\in\mathcal{T}} |\tilde{\boldsymbol{\beta}}_{\mathbf{C}}(t) - \boldsymbol{\beta}_{\mathbf{C}}(t) - \Xi(t)| = O_{\mathbb{P}}\left\{ \frac{n^{-\nu}}{\sqrt{(nb_n)}\log^{1/2}(n)} \right\}, \tag{16}$$

where $\nu = \min(3\theta/4, 7\theta/2 - 1/2, 1/5 - \theta/2) > 0$ and

$$\Xi(t) = \Sigma_{\mathbf{C}}(t)\,\boldsymbol{\mu}_{b_n}(t),$$

where

$$\boldsymbol{\mu}_{b_n}(t) = \sum_{i=1}^{n} \mathbf{V}_i\, K_{b_n}^*(t_i - t)/nb_n.$$

Proposition 1 follows from a careful check of the proof of theorem 3 (see expressions (34) and (35) and lemma 2 in Appendix A). Details have been omitted. For the mean-squared error optimal bandwidth with $\theta = 1/5$, $\nu = 1/10$. Equation (16) implies that simultaneous stochastic variation of $\Sigma_{\mathbf{C}}^{-1}(t)\{\tilde{\boldsymbol{\beta}}_{\mathbf{C}}(t) - \boldsymbol{\beta}_{\mathbf{C}}(t)\}$ can be well approximated by that of $\boldsymbol{\mu}_{b_n}(t)$. Hence the distribution of $\sup_{t\in\mathcal{T}} |\Sigma_{\mathbf{C}}^{-1}(t)\{\tilde{\boldsymbol{\beta}}_{\mathbf{C}}(t) - \boldsymbol{\beta}_{\mathbf{C}}(t)\}|$ can be approximated by $\sup_{t\in\mathcal{T}} |\boldsymbol{\mu}_{b_n}(t)|$, which can be obtained by generating a large number of IID copies

$$\boldsymbol{\mu}_{b_n}^{\dagger}(t) = \sum_{i=1}^{n} \mathbf{V}_i^{\dagger}\, K_{b_n}^*(t_i - t)/nb_n$$

via the wild bootstrap. Here $\mathbf{V}_i^{\dagger}$ are IID $N(0, \mathrm{Id}_s)$. In proposition 1, $\mathbf{V}_i$ are not wild bootstrap random variables. Note that, at a fixed point t, $\mathrm{cov}\{\boldsymbol{\mu}_{b_n}(t)\}$ is proportional to the identity matrix.

Summarizing the above discussion, we propose the following simulation-based procedure for constructing the SCT.

(a) Find an appropriate bandwidth $\hat{b}_n$ for estimating $\boldsymbol{\beta}_{\mathbf{C}}(\cdot)$ by using methods that are proposed in Section 4.4. Let $b'_n = 2\hat{b}_n$ and calculate $\tilde{\boldsymbol{\beta}}_{\mathbf{C},b'_n}(t)$.
(b) Generate IID Gaussian vectors $\mathbf{V}_1^\dagger, \mathbf{V}_2^\dagger, \ldots, \sim N(0, \mathrm{Id}_s)$ and calculate $\sup_{0\leqslant t\leqslant 1} |\boldsymbol{\mu}^\dagger_{b'_n}(t)|$.
(c) Repeat step (b) for 10^4 (say) times and obtain the estimated $(1-\alpha)$th quantile $\hat{q}_{1-\alpha}$ of $\sup_{0\leqslant t\leqslant 1} |\boldsymbol{\mu}_{b'_n}(t)|$.
(d) Calculate $\hat{\Sigma}_{\mathbf{C}}(t) = \{\mathbf{C}^{\mathrm{T}} \hat{M}^{-1}(t) \hat{\Lambda}(t) \hat{M}^{-1}(t) \mathbf{C}\}^{1/2}$ by using the method in Section 4.3.
(e) Construct the $(1-\alpha)$th SCT of $\boldsymbol{\beta}_{\mathbf{C}}(t)$ as $\tilde{\boldsymbol{\beta}}_{\mathbf{C},b'_n}(t) + \hat{\Sigma}_{\mathbf{C}}(t)\hat{q}_{1-\alpha}\mathcal{B}_s$.

4.3. Estimation of covariance matrices

To apply theorems 1 and 3, we need to estimate $M(t)$ and $\Lambda(t)$, the long-run covariance matrix function of the locally stationary time series $(\mathbf{x}_i\varepsilon_i)$ that is given in equation (8). We estimate $M(t)$ by $\hat{M}(t) = \mathbf{S}_{n,0}(t^*)$, where $t^* = \max\{b_n, \min(t, 1-b_n)\}$. Since $M(\cdot)$ is Lipschitz continuous, by lemma 6, in Appendix A, $\sup_{t\in[0,1]} |\hat{M}(t) - M(t)| = O_{\mathbb{P}}(b_n + n^{-1/2} b_n^{-1}) = o_{\mathbb{P}}(1)$ if $nb_n^2 \to \infty$. The problem of estimating $\Lambda(t)$ is not easy and in our case it is further complicated by the fact that the errors ε_i cannot be observed but need to be estimated instead.

Here we shall first establish a convergence result on the estimation of $\Lambda(t)$ by assuming that ε_i are known. The result is important in its own right since it provides convergence on covariance matrices estimates of multivariate locally stationary time series. Let $\mathbf{L}_i = \mathbf{x}_i\varepsilon_i$, $i = 1, \ldots, n$, and $\mathbf{Q}_i = \Sigma_{j=-m}^{m} \mathbf{L}_{i+j}$. Note that $\mathbb{E}(\mathbf{L}_i) = 0$. If the series $(\mathbf{L}_i)$ were stationary and hence $\Lambda(\cdot)$ were not time varying, then for each i, as $m\to\infty$, $\Delta_i := \mathbf{Q}_i\mathbf{Q}_i^{\mathrm{T}}/(2m+1)$ would converge to a distribution with expectation Λ, by the central limit theorem. In the locally stationary case, we could make use of the fact that a block of $(\mathbf{L}_i)$ is approximately stationary when its length is small compared with n. Hence $\mathbb{E}(\Delta_i) \approx \Lambda(t_i)$ as $m\to\infty$ and $m/n\to 0$. Since $\Lambda(\cdot) \in \mathcal{C}^2[0,1]$, we can use the Nadaraya–Watson-type estimator. Let τ_n be the bandwidth and $\gamma_n = \tau_n + (m+1)/n$. For $t \in \mathcal{I} = [\gamma_n, 1-\gamma_n] \subset (0,1)$, let

$$\hat{\Lambda}(t) = \sum_{i=1}^{n} \omega(t,i)\Delta_i, \qquad \omega(t,i) = K_{\tau_n}(t_i - t) \Big/ \sum_{k=1}^{n} K_{\tau_n}(t_k - t). \tag{17}$$

We define $\tilde{\Lambda}(t)$ over the whole interval $[0,1]$ by letting $\tilde{\Lambda}(t) = \tilde{\Lambda}(\gamma_n)$ if $t\in[0,\gamma_n]$ and $\tilde{\Lambda}(t) = \tilde{\Lambda}(1-\gamma_n)$ if $t\in[1-\gamma_n, 1]$. Note that $\hat{\Lambda}(t)$ is always positive semidefinite. In practice ε_i are not known. In theorem 5 we use $\tilde{\Lambda}(t) = \Sigma_{i=1}^{n} \omega(t,i)\bar{\Delta}_i$, where $\bar{\Delta}_i$ is defined as Δ_i with $\mathbf{L}_i$ therein replaced by $\bar{\mathbf{L}}_i := \mathbf{x}_i\hat{\varepsilon}_i$. Write $\|\cdot\| = \|\cdot\|_2$.

Theorem 4. Assume condition 3 with $r=4$, $\Lambda(t)\in\mathcal{C}^2[0,1]$, $\delta_4(\mathbf{U},k) = O[\{k\log(k)\}^{-2}]$, $m = m_n\to\infty$, $m = O(n^{1/3})$, $\tau_n\to 0$ and $n\tau_n\to\infty$. Then

(a) for any fixed $t\in(0,1)$,

$$\|\hat{\Lambda}(t) - \Lambda(t)\| = O\left\{\sqrt{\left(\frac{m}{n\tau_n}\right)} + \frac{1}{m} + \tau_n^2\right\}, \tag{18}$$

(b) for $\mathcal{I} = [\gamma_n, 1-\gamma_n] \subset (0,1)$, where $\gamma_n = \tau_n + (m+1)/n$,

$$\left\|\sup_{t\in\mathcal{I}} |\hat{\Lambda}(t) - \Lambda(t)|\right\| = O\left\{\sqrt{\left(\frac{m}{n\tau_n^2}\right)} + \frac{1}{m} + \tau_n^2\right\}. \tag{19}$$

Theorem 5. Assume that conditions (a)–(d) of theorem 3 and the conditions of theorem 4 hold. Further assume conditions 1 and 5, $\mathbf{G}(t,\mathcal{F}_i)\in\mathrm{Lip}_2$, $\sup_{0\leqslant t\leqslant 1}\{\|\mathbf{G}(t,\mathcal{F}_i)\|_\kappa\} < \infty$, for some $\kappa > 4$, and $\vartheta_n = o(1)$, where $\vartheta_n = n^{2/\kappa}\sqrt{m}\log(n)^2\{1/\sqrt{(nb_n)} + b_n^2\}$. Then

$$\sup_{t\in\mathcal{I}} |\tilde{\Lambda}(t)-\Lambda(t)| = O_{\mathbb{P}}\left\{\vartheta_n + \sqrt{\left(\frac{m}{n\tau_n^2}\right)} + \frac{1}{m} + \tau_n^2\right\}. \tag{20}$$

The bound in equation (18) is minimized and it becomes $O(n^{-2/7})$ if $m \asymp n^{2/7}$ and $\tau_n \asymp n^{-1/7}$. Here for two positive sequences (r_n) and (s_n) we write $r_n \asymp s_n$ if $s_n/r_n + r_n/s_n$ are bounded for all large n. If $m \asymp n^{1/4}$ and $\tau_n \asymp n^{-1/8}$, then the uniform bound in equation (19) is $O(n^{-1/4})$. Theorem 5 requires stronger conditions: if $m \asymp n^{q_1}$, $b_n \asymp n^{-q_2}$ and $\tau_n \asymp n^{-q_3}$, where $q_1, q_2, q_3 > 0$, then we need $q_1/2 + 2/\kappa < \min(2q_2, \frac{1}{2} - q_2/2)$ and $q_1 < 1 - 2q_3$. Under the latter conditions, since $\Lambda(t) \in \mathcal{C}^2[0,1]$, we have by theorem 5 that $\sup_{t\in[0,1]} |\tilde{\Lambda}(t) - \Lambda(t)| = O_{\mathbb{P}}(n^{-\nu})$ for some $\nu > 0$, which by lemma 6 in Appendix A implies that the estimate $\hat{\Sigma}_{\mathbf{C}}(t)$ in step (d) satisfies $\sup_{t\in[0,1]} |\hat{\Sigma}_{\mathbf{C}}(t) - \Sigma_{\mathbf{C}}(t)| = O_{\mathbb{P}}(n^{-\lambda})$ for some $\lambda > 0$. So the convergence in expression (12) still holds if $\Sigma_{\mathbf{C}}(t)$ therein is replaced by $\hat{\Sigma}_{\mathbf{C}}(t)$.

4.4. Selection of smoothing parameters

In algorithmic implementation of the foregoing procedures, we need to choose smoothing parameters b_n, m and τ_n. Theorem 1 suggests that the minimum asymptotic mean integrated squared error bandwidth for estimating $\boldsymbol{\beta}(\cdot)$ is

$$b_n^* = \left[\frac{\phi_0 \int_0^1 \mathrm{tr}\{\Sigma(t)\}\,\mathrm{d}t}{\mu_2^2 \int_0^1 |\boldsymbol{\beta}''(t)|^2\,\mathrm{d}t}\right]^{1/5} n^{-1/5}.$$

Hence we can estimate the second derivative $\boldsymbol{\beta}''$ and $\mathrm{tr}\{\Sigma(\cdot)\}$ and then plug them in b_n^*. Another selector, which we adopt in our simulations and data analysis, is the generalized cross-validation (GCV) method (Craven and Wahba, 1979). For estimating $\boldsymbol{\beta}(\cdot)$, we can write $\hat{\mathbf{Y}} = Q(b)\mathbf{Y}$ for some square matrix Q, where $\mathbf{Y}$ and $\hat{\mathbf{Y}}$ denote the vector of observed values and estimated values respectively, and b is the bandwidth. We can choose

$$\hat{b}_n = \arg\min_b \{\mathrm{GCV}(b)\}, \qquad \mathrm{GCV}(b) = \frac{n^{-1}|\mathbf{Y} - \hat{\mathbf{Y}}|^2}{[1 - \mathrm{tr}\{Q(b)\}/n]^2}. \tag{21}$$

The GCV selector works reasonably well in our simulations.

We now discuss the choice of m and τ_n for estimating the long-run covariance matrix. By equation (18) of theorem 4, for an easy implementation, we could simply choose $m^* = \lfloor n^{2/7} \rfloor$ and $\tau_n^* = n^{-1/7}$. For refinements, we recommend the following extended minimum volatility method which is an extension of the minimum volatility method that was proposed in chapter 9 of Politis *et al.* (1999). The idea behind the extended minimum volatility method is that, if a pair of block size and bandwidth is in a reasonable range, then confidence regions for the local mean constructed by $\tilde{\Lambda}(t)$ should be stable when considered as a function of block size and bandwidth. Hence we could first propose a grid of possible block sizes and bandwidths and then choose the pair that minimizes the volatility of the boundary points (curves) of the confidence regions near this pair. More specifically, let the grid of possible block sizes and bandwidths be $\{m_1, \ldots, m_{M_1}\}$ and $\{\tau_1, \ldots, \tau_{M_2}\}$ respectively and let the estimated long-run covariance functions be $\{\tilde{\Lambda}_{m_h,\tau_j}(t)\}$, $h = 1, \ldots, M_1$, $j = 1, \ldots, M_2$. For each pair (m_h, τ_j), calculate

$$\mathrm{ise}[\cup_{r=-2}^{2} \{\tilde{\Lambda}_{m_h,\tau_{j+r}}(t)\} \cup \cup_{r=-2}^{2} \{\tilde{\Lambda}_{m_{h+r},\tau_j}(t)\}],$$

where ise denotes the integrated standard error

$$\mathrm{ise}[\{\tilde{\Lambda}_l(t)\}_{l=1}^k] = \int_0^1 \left\{ \frac{1}{k-1} \sum_{l=1}^k |\tilde{\Lambda}_l(t) - \bar{\tilde{\Lambda}}(t)|^2 \right\}^{1/2} \mathrm{d}t$$

with $\bar{\tilde{\Lambda}}(t) = \Sigma_{l=1}^k \tilde{\Lambda}_l(t)/k$. Then we choose the pair (m_h^*, τ_j^*) that minimizes ise. In our simulations, the extended minimum volatility selector performs reasonably well and it is also found that the estimated covariance functions are not sensitive to the choice of (m, τ) as long as this pair is not very different from the pair that is chosen by the extended minimum volatility method.

4.5. Simultaneous confidence tubes with minimal volumes

Our SCT is optimal in the sense that it has asymptotically smallest average volume. To see this, we apply the Lagrange multiplier argument. For illustration, we focus on the one-dimensional case and write $\boldsymbol{\beta}$, $\mathbf{C}$, $\mathbf{A}$ and Σ as β, $\mathbf{c}$, $\mathbf{a}$ and σ respectively. Let $s_l = (2l-1)b_n\sqrt{2}$, $l = 1, \ldots, g_n$, where $g_n = \lfloor 1/2b_n\sqrt{2} \rfloor$. From equation (16), $\tilde{\beta}_{\mathbf{c}}(s_l) - \beta_{\mathbf{c}}(s_l)$ are asymptotically independent $N\{0, \phi_{0,K^*}\sigma_{\mathbf{c}}^2(s_l)/nb_n\}$. Suppose that a band $[l(s_l) \leqslant \tilde{\beta}_{\mathbf{c}}(s_l) - \beta_{\mathbf{c}}(s_l) \leqslant u(s_l)]$, $l = 1, \ldots, g_n$, achieves the preassigned coverage probability $1-\alpha$. Owing to the asymptotic independence, we have, asymptotically,

$$c(n, b_n) := \prod_{l=1}^{g_n} \left[\Phi\left\{ \frac{u(s_l)\sqrt{(nb_n)}}{\sigma_{\mathbf{c}}(s_l)\sqrt{\phi_{0,K^*}}} \right\} - \Phi\left\{ \frac{l(s_l)\sqrt{(nb_n)}}{\sigma_{\mathbf{c}}(s_l)\sqrt{\phi_{0,K^*}}} \right\} \right] = 1-\alpha,$$

where $\Phi(\cdot)$ is the normal distribution function. To achieve the minimum average length, we choose $l(s_l)$ and $u(s_l)$ that minimize the target function

$$\sum_{l=1}^{g_n} \{u(s_l) - l(s_l)\} - \lambda[\log\{c(n, b_n)\} - \log(1-\alpha)]. \tag{22}$$

Simple calculations show that the minimum is achieved at $u(s_l) = -l(s_l) = g(n, b_n, \alpha)\,\sigma_{\mathbf{c}}(s_l)$, where $g(n, b_n, \alpha)$ is a deterministic function. It suggests that the asymptotically optimal SCT should have a length that is proportional to the long-run standard deviation at each time point, which is satisfied by our construction procedure.

5. Simulation study

In this section we shall perform a simulation to study the finite sample coverage probabilities of our SCT under scenarios (a) and (b) in Section 2. We are particularly interested in investigating whether heteroscedasticity of errors would result in inaccurate coverage probabilities. For this purpose, consider the time varying coefficient model

$$y_i = \beta_1(i/n) + \beta_2(i/n)x_i + \varepsilon_i, \qquad i = 1, \ldots, n. \tag{23}$$

where $\beta_1(t) = \sin(2\pi t)/4$ and $\beta_2(t) = \exp\{-(t-\frac{1}{2})^2\}/2$. We shall consider the following two cases which correspond to scenarios (a) and (b) in Section 2 respectively.

(a) Let $H(t, \mathcal{F}_i) = 4^{-1}\Sigma_{j=0}^\infty a(t)^j \zeta_{i-j}$, $\varepsilon_i = H(i/n, \mathcal{F}_i)$, $\mathbf{G}(t, \mathcal{F}_i) = (1, \Sigma_{j=0}^\infty c(t)^j \epsilon_{i-j})$ and $\mathbf{x}_i = (1, x_i) = \mathbf{G}(i/n, \mathcal{F}_i)$, $i = 1, \ldots, n$, where $a(t) = \frac{1}{2} - (t-\frac{1}{2})^2$, $c(t) = \frac{1}{4} + t/2$, and $\epsilon_h, \zeta_l, h, l \in \mathbb{Z}$, are IID $N(0, 1)$.

(b) Let x_i, ϵ_h, ζ_l and $a(\cdot)$ be the same as in (a) and $\varepsilon_i = H(t_i, \mathcal{F}_i)$, where $H(t, \mathcal{F}_i) = 8^{-1}[\Sigma_{j=0}^\infty \{a(t)^j \times \zeta_{i-j}\}]\{\Sigma_{j=0}^\infty c(t)^j \epsilon_{i-j}\}$. Then $\mathbb{E}(\varepsilon_i | \epsilon_i, \epsilon_{i-1}, \ldots) = 0$.

In these two examples, the variances of ε_i are roughly of the same order. In our simulations, for each of the above two cases we generate 5000 samples of size $n = 500$. For each sample the

Table 1. Simulated coverage probabilities of the SCT of $\beta_2(\cdot)$ and $(\beta_1(\cdot), \beta_2(\cdot))^{\mathrm{T}}$ at 90% and 95% nominal levels for model (23) with two cases (a) and (b)

b	*Results for case (a)*				*Results for case (b)*			
	$\beta_2(t)$		$(\beta_1(t), \beta_2(t))^{\mathrm{T}}$		$\beta_2(t)$		$(\beta_1(t), \beta_2(t))^{\mathrm{T}}$	
	90%	*95%*	*90%*	*95%*	*90%*	*95%*	*90%*	*95%*
0.1	0.864	0.917	0.784	0.875	0.873	0.92	0.836	0.915
0.125	0.875	0.940	0.821	0.886	0.874	0.925	0.865	0.930
0.15	0.905	0.948	0.861	0.921	0.895	0.940	0.895	0.941
0.175	0.901	0.943	0.869	0.922	0.889	0.940	0.903	0.940
0.2	0.900	0.949	0.896	0.945	0.899	0.942	0.884	0.955
0.225	0.904	0.947	0.893	0.945	0.898	0.943	0.910	0.955
0.25	0.910	0.955	0.900	0.943	0.908	0.956	0.899	0.946
0.275	0.902	0.950	0.902	0.946	0.908	0.949	0.912	0.956
0.3	0.911	0.957	0.912	0.953	0.909	0.953	0.914	0.946
0.325	0.909	0.957	0.910	0.956	0.903	0.946	0.914	0.952
0.35	0.904	0.952	0.913	0.952	0.908	0.944	0.908	0.949

local linear estimation together with the jackknife bias reduction are performed with bandwidths $b_j = 0.025j$, $j = 4, \ldots, 14$, to obtain estimates $\tilde{\beta}_{i,b_j}(t)$, $i = 1, 2$. We then calculate $\hat{\sigma}_{(0,1)}(t)$ and $\hat{\Sigma}(t)$ with smoothing parameters selected by the extended minimum volatility method. We use 3000 bootstrap samples to estimate $\hat{q}_{1-\alpha}$ for $\alpha = 0.1$ and $\alpha = 0.05$. SCTs for $\beta_2(t)$ and $(\beta_1(t), \beta_2(t))^{\mathrm{T}}$ are then constructed for two levels: 90% and 95%. Table 1 reports the simulated coverage probabilities. The coverage probabilities depend on the bandwidth in a complicated way, and in finite samples it is difficult to know whether they are conservative or anticonservative. GCV selects bandwidth 0.25 and 0.17 for models (a) and (b) respectively. Table 1 suggests that for both models the simulated coverage probabilities are reasonably close to the nominal levels for the GCV-selected bandwidths.

6. Specification tests

Let $\{\mathbf{f}(t, \theta)\}$ be a parametric family of functions with $t \in [0, 1]$ and $\theta \in \Theta \subset \mathbb{R}^k$. To test the null hypothesis $H_0 : \boldsymbol{\beta}_{\mathbf{C}}(\cdot) = \mathbf{f}(\cdot, \theta)$ for some unknown $\theta \in \Theta$ at level α, we construct a $1 - \alpha$ SCT for $\boldsymbol{\beta}_{\mathbf{C}}(\cdot)$ and check whether, for some $\theta \in \Theta$, $\mathbf{f}(\cdot, \theta)$ is fully contained in the SCT. If there is one such θ, then the null hypothesis is accepted. If the structure of $\{\mathbf{f}(t, \theta)\}$ is complicated, then the checking may be tedious. However, under the null hypothesis, model (2) is essentially a semiparametric model and we could expect to obtain a root n consistent estimate $\mathbf{f}(t, \hat{\theta})$ of $\boldsymbol{\beta}_{\mathbf{C}}(t)$. For example, if $\mathbf{f}(t, \theta) = (\theta_1^{\mathrm{T}} \mathbf{g}_1(t), \ldots, \theta_s^{\mathrm{T}} \mathbf{g}_s(t))^{\mathrm{T}}$, where $\theta_l \in \mathbb{R}^{d_l}$ and $\mathbf{g}_l : [0, 1] \to \mathbb{R}^{d_l}$ are fixed functions $l = 1, \ldots, s$, with known forms, the profile least squares method (Fan and Huang, 2005) can be used to obtain a root n consistent estimate of $\hat{\theta}$. The convergence rate of our SCT is always slower than root n. Hence, if the null hypothesis is true, $\mathbf{f}(t, \hat{\theta})$ can be treated as the true value of $\boldsymbol{\beta}_{\mathbf{C}}(t)$ and the null is rejected if the SCT does not fully contain $\mathbf{f}(t, \hat{\theta})$.

As an application, we consider the Hong Kong circulatory and respiratory data. They consist of daily measurements of pollutants and daily hospital admissions in Hong Kong between

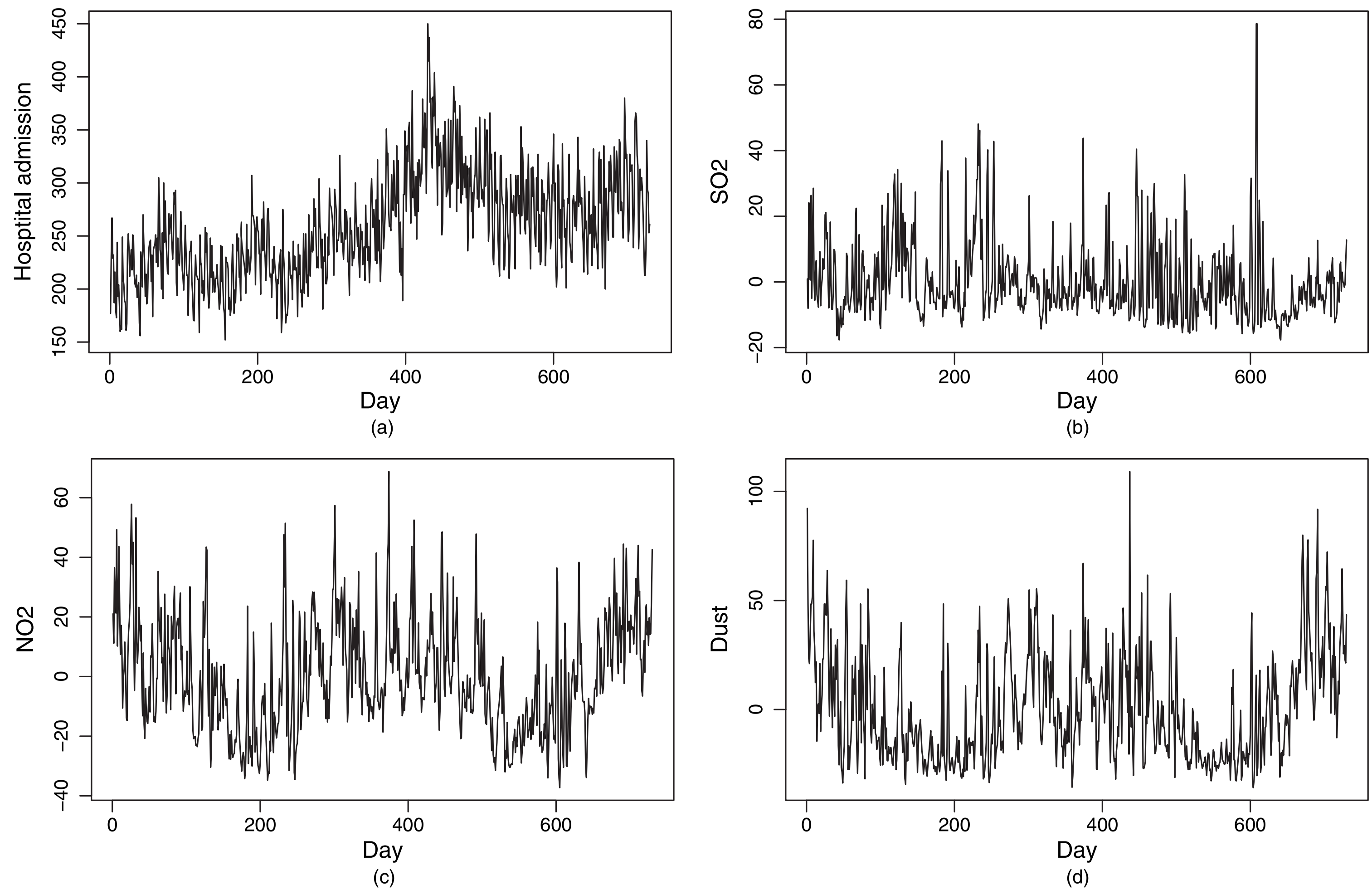

Fig. 1. Time series plots of the response and the covariates for the Hong Kong circulatory and respiratory data: (a) total number of hospital admissions; (b) SO_2 level; (c) NO_2 level; (d) dust level

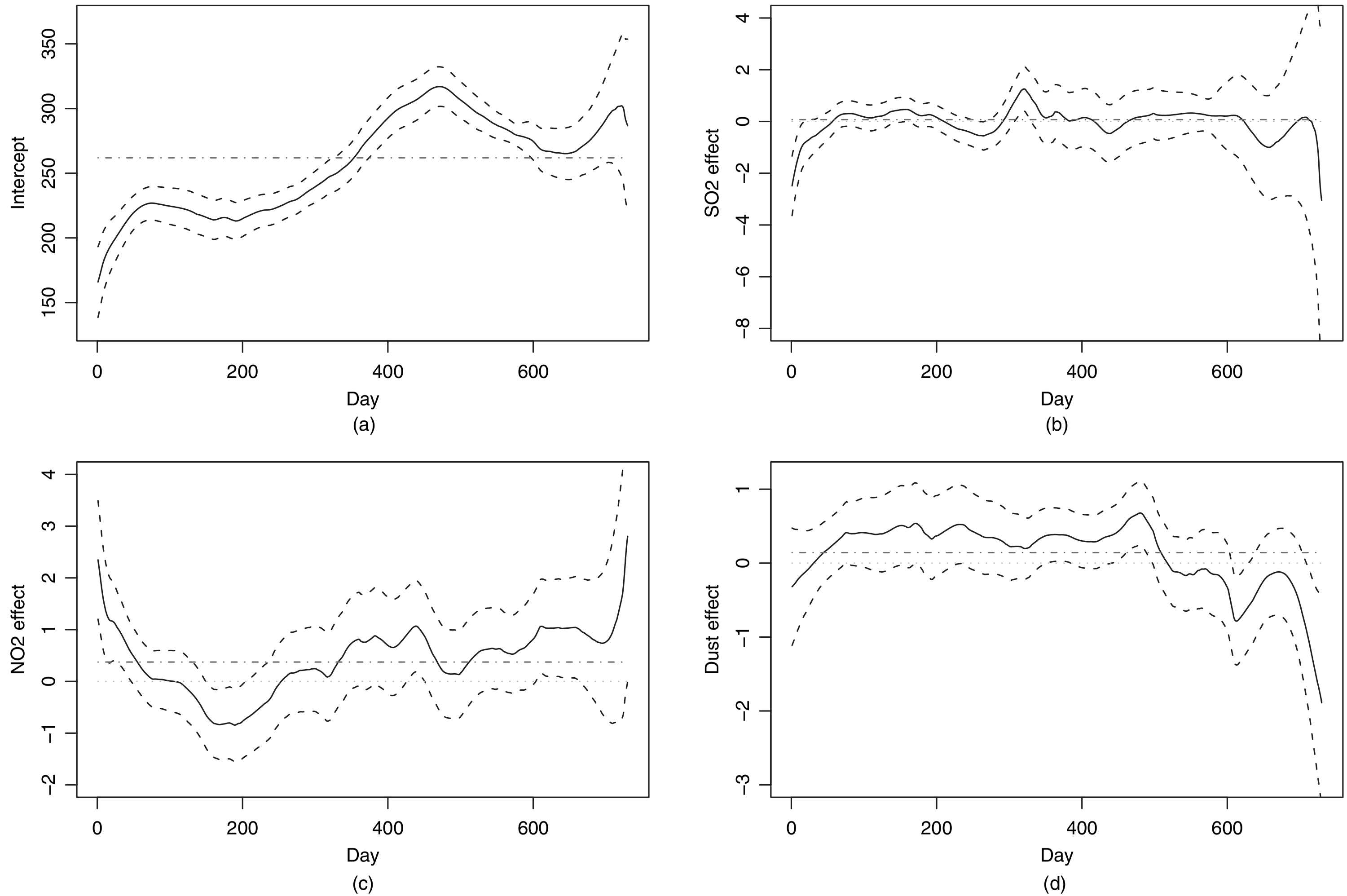

Fig. 2. 95% simultaneous confidence bands for $\beta_h(t)$, $h = 1, 2, 3, 4$, in model (24) ($\cdots\cdots$, $\beta_h(t) = 0, h = 2, 3, 4$; $\cdot - \cdot -$, fitted constant trends $\beta_h(t) \equiv \hat{c}_h$, $\hat{c}_1 = 261.91$, $\hat{c}_2 = 0.070$, $\hat{c}_3 = 0.37$, $\hat{c}_4 = 0.14$): (a) intercept; (b) SO$_2$ effect; (c) NO$_2$ effect; (d) dust effect

January 1st, 1994, and December 31st, 1995. The purpose is to investigate the dynamic association between the levels of pollutants and the total number of hospital admissions of circulation and respiration. In the context of varying-coefficient models, this data set has been studied by Fan and Zhang (1999, 2000) and Cai *et al.* (2000). Most of the previous results assumed that the observations were IID. Here we shall investigate the data under framework (4) where both the covariates and the errors are modelled as non-stationary time series and we shall also compare our findings with the previous results. More specifically, consider the model

$$y_i = \beta_1(i/n) + \sum_{p=2}^{4} \beta_p(i/n) x_{ip} + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{24}$$

where (y_i) is the series of daily total number of hospital admissions of circulation and respiration and (x_{ip}), $p = 2, 3, 4$, represent the series of daily levels of sulphur dioxide (SO_2) (in micrograms per cubic metre), nitrogen dioxide (NO_2) (in micrograms per cubic metre) and dust (in micrograms per cubic metre) respectively. Here $n = 2 \times 365 = 730$. Following Fan and Zhang (2000), we first centre each of the three pollutants by their averages so that the intercept $\beta_1(\cdot)$ can be interpreted as the expected number of admissions when the pollutants are set at their averages. Fig. 1 summarizes the data and it suggests that neither the response nor the covariates are IID. In particular, seasonal patterns can be found in the NO_2 series and the dust series.

In the data analysis we use the Epanechnikov kernel. Following the procedures in Section 4.4, the smoothing parameters b_n, τ_n and m are selected as 0.21, 0.2 and 14 respectively. For each of the three pollutants, we are interested in testing whether it is significantly associated with the number of hospital admissions. This amounts to testing $\beta_h(t) \equiv 0$ for $h = 2, 3, 4$. Furthermore, we also check whether the intercept or the pollutants' effect are really time varying. So the four hypotheses $\beta_h(t) \equiv c_h$, $h = 1, 2, 3, 4$, are tested. To obtain estimates of c_h, the profile least squares method (Fan and Huang, 2005) is used with bandwidth $b_n = 0.15$. It turns out that all hypotheses are rejected at the 0.01 level. Therefore we conclude that all three pollutants are associated with the response and the pollutants' effect varies significantly with time. Under the assumption of IID observations, Fan and Zhang (2000) claimed that the effect of SO_2 was not significant. Note that their simultaneous confidence bands were constructed according to the asymptotic distributions which converge at the slow logarithmic rate. Additionally, the autocorrelation function plots show that there is substantial dependence between the fitted residuals $\hat{\varepsilon}_i$. In comparison with our Fig. 2, the bands in their Fig. 2 are generally wider.

Acknowledgements

We are grateful to the referee, the Associate Editor and the Joint Editor for their many helpful comments. We thank Professor Jianqing Fan and Professor Wenyang Zhang for providing the circulatory and respiratory data for Hong Kong. The research is supported by National Science Foundation grants NSF/DMS/STAT-04478704 and 0906073.

Appendix A

A.1. Proof of theorem 1

By assumption 6 and Taylor's expansion, if $|t_j - t| \leqslant b_n$, $\boldsymbol{\beta}(t_j) = \boldsymbol{\beta}(t) + \boldsymbol{\beta}'(t)(t - t_j) + \{\boldsymbol{\beta}''(t)/2 + O(b_n)\}(t - t_j)^2$. Since K has support $[-1, 1]$, by expression (7),

$$\mathbf{S}_n(t)\{\hat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}(t)\} = \begin{pmatrix} b_n^2 \mathbf{S}_{n,2}(t)\{\boldsymbol{\beta}''(t) + O(b_n)\}/2 \\ b_n^2 \mathbf{S}_{n,3}(t)\{\boldsymbol{\beta}''(t) + O(b_n)\}/2 \end{pmatrix} + \begin{pmatrix} \mathbf{T}_{n,0}(t) \\ \mathbf{T}_{n,1}(t) \end{pmatrix}, \tag{25}$$

where $\boldsymbol{\eta}(t) = (\boldsymbol{\beta}^{\mathrm{T}}(t), b_n \boldsymbol{\beta}'^{\mathrm{T}}(t))^{\mathrm{T}}$, $\mathbf{T}_n(t) = (\mathbf{T}_{n,0}^{\mathrm{T}}(t), \mathbf{T}_{n,1}^{\mathrm{T}}(t))^{\mathrm{T}}$ and

$$\mathbf{T}_{n,l}(t) = (nb_n)^{-1} \sum_{i=1}^{n} \mathbf{x}_i \varepsilon_i \{(t_i - t)/b_n\}^l K_{b_n}(t_i - t), \qquad l = 0, 1, \ldots.$$

Let $\tilde{M}(t) = \mathrm{diag}\{M(t), \mu_2 M(t)\}$, $\chi_n = (nb_n)^{-1/2} + b_n$ and $\rho_n = (nb_n)^{-1/2} + b_n^2$. By lemma 6, $\mathbf{S}_n(t) - \tilde{M}(t) = O_{\mathbb{P}}(\chi_n)$. Note that $\mathbb{E}\{\mathbf{T}_n(t)\} = 0$. By assumptions 3 and 4 and the Cramer–Wold device, proposition 6 in Zhou and Wu (2009) implies the central limit theorem

$$(nb_n)^{1/2} \mathbf{T}_n(t) \Rightarrow N\{0, \tilde{\Lambda}(t)\}, \qquad \tilde{\Lambda}(t) = \mathrm{diag}\{\phi_0 \Lambda(t), \phi_2 \Lambda(t)\}. \tag{26}$$

By expressions (25) and (26) we have $\hat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}(t) = O_{\mathbb{P}}(\rho_n)$, and

$$\mathbf{S}_n(t)\{\hat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}(t)\} = \begin{pmatrix} \mu_2 M(t) \boldsymbol{\beta}''(t) b_n^2/2 + O_{\mathbb{P}}(b_n^2 \chi_n) \\ O_{\mathbb{P}}(b_n^2 \chi_n) \end{pmatrix} + \mathbf{T}_n(t). \tag{27}$$

Since $nb_n^7 \to 0$, $(nb_n)^{1/2}\{\mathbf{S}_n(t) - \tilde{M}(t)\}\rho_n = o_{\mathbb{P}}(1)$. By expressions (26) and (27), expression (9) follows. □

Lemma 1. Let $F_n(t) = \Sigma_{i=1}^n \mathbf{V}_i K_{b_n}(t_i - t)$, where $\mathbf{V}_i$, $i \in \mathbb{Z}$, are IID $N(0, \mathrm{Id}_s)$. Suppose that $K \in \mathcal{K}$, $b_n \to 0$ and $nb_n/\log^2(n) \to \infty$. Let $m^* = 1/b_n$. Then

$$\lim_{n\to\infty} \left(\mathbb{P}\left[\frac{1}{\sqrt{(\phi_0 n b_n)}} \sup_{t \in \mathcal{T}} |F_n(t)| - B_K(m^*) \leqslant \frac{u}{\sqrt{\{2\log(m^*)\}}} \right] \right) = \exp\{-2\exp(-u)\}. \tag{28}$$

Proof. Let $\{\mathbb{B}(t), t \in \mathbb{R}\}$ be an s-dimensional standard Brownian motion. Then

$$\mathbf{Y}(t) := \int_{\mathbb{R}} K(t-u) \, \mathrm{d}\mathbb{B}(u)/\sqrt{\phi_0}$$

is a stationary Gaussian process. Note that $Y_j(\cdot), 1 \leqslant j \leqslant s$, the jth components of $\mathbf{Y}(\cdot)$, are IID with $\mathrm{var}\{Y_j(t)\} = 1, \mathrm{cov}\{Y_j(0), Y_j(t)\} = 1 - \lambda_2 t^2/2 + o(t^2)$, where $\lambda_2 = \int_{-1}^{1} |K'(x)|^2 \,\mathrm{d}x/\phi_0$. Since the support of K is within $[-1, 1]$, $\mathrm{cov}\{Y_j(0), Y_j(t)\} = 0$ if $|t| \geqslant 2$. By theorem 3.1 in Lindgren (1980) and Slutsky's theorem, we have

$$\lim_{n\to\infty} \left(\mathbb{P}\left[\max_{t \in [1, m^*-1]} |\mathbf{Y}(t)| - B_K(m^*) \leqslant \frac{u}{\sqrt{\{2\log(m^*)\}}} \right] \right) = \exp\{-2\exp(-u)\}. \tag{29}$$

Let $\tilde{\mathbf{Y}}(t) = \int_0^{m^*} K(t - \lfloor 1 + k_n u \rfloor / k_n) \,\mathrm{d}\mathbb{B}(u)/\sqrt{\phi_0}$, where $k_n = nb_n$. Then, by the argument in the proof of lemma 2 of Wu and Zhao (2007), we have

$$\max_{t \in [1, m^*-1]} |\mathbf{Y}(t) - \tilde{\mathbf{Y}}(t)| = O_{\mathbb{P}}\left[\sqrt{\left\{ \frac{\log(n)}{k_n} \right\}} \right] = o_{\mathbb{P}}\left\{ \frac{1}{\sqrt{\log(n)}} \right\}. \tag{30}$$

Since the process $F_n(t/m^*)/\sqrt{(\phi_0 k_n)}$ is identically distributed as $\tilde{\mathbf{Y}}(t)$ over $t \in [1, m^* - 1]$, lemma 1 follows.

Lemma 2. Let $\mathbf{D}_{\mathbf{W}}(t) = (nb_n)^{-1} \Sigma_{i=1}^n \mathbf{Z}_i K_{b_n}(t_i - t)$, where $\mathbf{W}(\cdot, \cdot)$ and $(\mathbf{Z}_i)_1^n$ are as defined in theorem 2. Assume that $\Sigma_{\mathbf{W}}(t)$ is Lipschitz continuous and bounded away from 0 on $[0, 1]$ and $\log^3(n)/n^{2/5} b_n + b_n \log^2(n) = o(1)$. Then, under conditions of theorem 2, we have

$$\lim_{n\to\infty} \left(\mathbb{P}\left[\sup_{t \in \mathcal{T}} \left\{ \frac{\sqrt{(nb_n)}}{\sqrt{\phi_0}} |\Sigma_{\mathbf{W}}^{-1}(t) \mathbf{D}_{\mathbf{W}}(t)| \right\} - B_K(m^*) \leqslant \frac{u}{\sqrt{\{2\log(m^*)\}}} \right] \right) = \exp\{-2\exp(-u)\}. \tag{31}$$

Proof. By theorem 2 and the summation-by-parts formula, simple calculations show that there are IID s-dimensional standard Gaussian random vectors $\{\mathbf{V}_i\}$ such that

$$\sup_{t \in \mathcal{T}} |\mathbf{D}_{\mathbf{W}}(t) - \Xi_{\mathbf{W}}(t)| = O_{\mathbb{P}}\left\{ \frac{n^{3/10} \log(n)}{nb_n} \right\} = o_{\mathbb{P}}\left\{ \frac{1}{\sqrt{(nb_n)} \log^{1/2}(n)} \right\}, \tag{32}$$

where $\Xi_{\mathbf{W}}(t) = (nb_n)^{-1} \Sigma_{i=1}^n \Sigma_{\mathbf{W},i} \mathbf{V}_i K_{b_n}(t_i - t)$. Since $\Sigma_{\mathbf{W}}(\cdot)$ is Lipschitz, we have

$$\sup_{t \in \mathcal{T}} |\Xi_{\mathbf{W}}(t) - (nb_n)^{-1} \Sigma_{\mathbf{W}}(t) \sum_{i=1}^{n} \mathbf{V}_i K_{b_n}(t_i - t)| = O_{\mathbb{P}}\left\{ \frac{b_n \log(n)}{\sqrt{(nb_n)}} \right\}. \tag{33}$$

So equation (31) follows from equations (32) and (33) and lemma 1 since $\log^3(n)/n^{2/5} b_n + b_n \log^2(n) \to 0$.

A.2. Proof of theorem 3

Applying lemma 2 to the process $\mathbf{Z}_i = \mathbf{x}_i\varepsilon_i$, we have $\sup_{t\in\mathcal{T}} |\mathbf{T}_n(t)| = O_{\mathbb{P}}\{(nb_n)^{-1/2}\log(n)\}$. Let $\chi_n' = n^{-1/2}b_n^{-1} + b_n$ and $\rho_n' = (nb_n)^{-1/2}\log(n) + b_n^2$. By equation (25) and lemma 6, we have $\sup_{t\in\mathcal{T}} |\hat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}(t)| = O_{\mathbb{P}}(\rho_n')$, and hence

$$\sup_{t\in\mathcal{T}} \left| M(t)\{\hat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t)\} - \frac{b_n^2\mu_2\, M(t)\,\boldsymbol{\beta}''(t)}{2} - \mathbf{T}_{n,0}(t) \right| = O_{\mathbb{P}}(\rho_n'\chi_n'). \tag{34}$$

Let

$$\Omega_{\mathbf{C}}(t) = (nb_n)^{-1} \sum_{i=1}^n \mathbf{A}_{\mathbf{C}}^{\mathrm{T}}(t_i)\mathbf{x}_i\varepsilon_i\, K_{b_n}(t_i - t).$$

By the proof of proposition 7 in Zhou and Wu (2009) and the differentiability of $\mathbf{A}_{\mathbf{C}}(t)$, we have

$$\sup_{t\in\mathcal{T}} |\mathbf{A}_{\mathbf{C}}^{\mathrm{T}}(t)\,\mathbf{T}_{n,0}(t) - \Omega_{\mathbf{C}}(t)| = O_{\mathbb{P}}\{(nb_n)^{-1/2}b_n^{3/4}\} = o_{\mathbb{P}}\left\{\frac{1}{\sqrt{(nb_n)}\log^{1/2}(n)}\right\}. \tag{35}$$

Note that $(\mathbf{A}_{\mathbf{C}}^{\mathrm{T}}(t_i)\mathbf{x}_i\varepsilon_i)$ is a locally stationary process of the form $(\mathbf{W}(t_i, \mathcal{F}_i))$ with long-run variance $\Sigma_{\mathbf{W},i}^2 = \Sigma_{\mathbf{C}}^2(t_i)$. Since the conditions of lemma 2 are satisfied for this $\mathbf{W}$, by equations (34) and (35) and lemma 2, theorem 3 follows. □

To prove theorem 4, we need lemmas 3–5. Recall that $\mathbf{Q}_i = \Sigma_{j=-m}^m \mathbf{L}_{i+j}$. Define

$$\mathbf{Q}_i^\diamond = \sum_{j=-m}^m \mathbf{U}(t_i, \mathcal{F}_{i+j})$$

and

$$\Delta_i^\diamond = \frac{\mathbf{Q}_i^\diamond(\mathbf{Q}_i^\diamond)^{\mathrm{T}}}{2m+1}.$$

Note that $\Delta_i = \mathbf{Q}_i\mathbf{Q}_i^{\mathrm{T}}/(2m+1)$. For $k \in \mathbb{Z}$ define the projection operator

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

Lemma 3. Under conditions of theorem 4, we have $\|\hat{\Lambda}(t) - \mathbb{E}\{\hat{\Lambda}(t)\}\| = O\{(mn^{-1}\tau_n^{-1})^{1/2}\}$.

Proof. Since $\mathbf{Q}_i$ is $\mathcal{F}_{i+m}$ measurable, we can find a measurable function f such that $\mathbf{Q}_i = f(\mathcal{F}_{i+m})$. Let $(\zeta_i')_{i\in\mathbb{Z}}$ be an IID copy of $(\zeta_i)_{i\in\mathbb{Z}}$, and, for $j, l \in \mathbb{Z}$, define

$$\mathbf{Q}_{i,\{l\}}^* = f(\mathcal{F}_{i+m,\{l\}}), \qquad \mathcal{F}_{j,\{l\}} = (\ldots, \zeta_{l-2}, \zeta_{l-1}, \zeta_l', \zeta_{l+1}, \ldots, \zeta_j).$$

$\mathcal{F}_{j,\{l\}}$ is obtained by replacing ζ_l in $\mathcal{F}_j$ by an IID copy ζ_l', and $\mathcal{F}_{j,\{l\}} = \mathcal{F}_j$ if $l > j$. Since $\Sigma_{k=0}^\infty \delta_4(\mathbf{U}, k) < \infty$, using the argument of theorem 1 in Wu (2007), we have $\sup_i(\|\mathbf{Q}_i\|_4) = O(\sqrt{m})$. Since

$$\|\mathbf{Q}_i - \mathbf{Q}_{i,\{i-l\}}\|_4 \leqslant \sum_{j=-m}^m \delta_4(\mathbf{U}, l+j),$$

$$\begin{aligned}\|\mathbf{Q}_i\mathbf{Q}_i^{\mathrm{T}} - \mathbf{Q}_{i,\{i-l\}}\mathbf{Q}_{i,\{i-l\}}^{\mathrm{T}}\| &\leqslant \|\mathbf{Q}_i\|_4\|\mathbf{Q}_i^{\mathrm{T}} - \mathbf{Q}_{i,\{i-l\}}^{\mathrm{T}}\|_4 + \|\mathbf{Q}_i - \mathbf{Q}_{i,\{i-l\}}\|_4\|\mathbf{Q}_{i,\{i-l\}}^{\mathrm{T}}\|_4 \\ &= O(\sqrt{m}) \sum_{j=-m}^m \delta_4(\mathbf{U}, l+j).\end{aligned} \tag{36}$$

By theorem 1 in Wu (2005), $\|\mathcal{P}_{i-l}(\mathbf{Q}_i\mathbf{Q}_i^{\mathrm{T}})\| \leqslant \|\mathbf{Q}_i\mathbf{Q}_i^{\mathrm{T}} - \mathbf{Q}_{i,\{i-l\}}\mathbf{Q}_{i,\{i-l\}}^{\mathrm{T}}\|$. Define $\Psi_l = \Sigma_{i=1}^n \omega(t,i)\mathcal{P}_{i-l}\Delta_i$. Since $\mathcal{P}_{i-l}\Delta_i, 1 \leqslant i \leqslant n$, are martingale differences, we have

$$\|\Psi_l\|^2 = \sum_{i=1}^n \omega^2(t,i)\|\mathcal{P}_{i-l}\Delta_i\|^2 = \frac{O(1)}{nm\tau_n}\left\{\sum_{j=-m}^m \delta_4(\mathbf{U}, l+j)\right\}^2.$$

Since $\hat{\Lambda}(t) - \mathbb{E}\{\hat{\Lambda}(t)\} = \Sigma_{k=0}^\infty \Psi_{k-m}$ and $\Sigma_{k=0}^\infty \delta_4(\mathbf{U}, k) < \infty$, the lemma follows.

Lemma 4. Under the conditions of theorem 4, we have

$$|\mathbb{E}\{\hat{\Lambda}(t)\} - \mathbb{E}\{\hat{\Lambda}^\diamond(t)\}| = O\{\sqrt{(m/n)}\}, \qquad \hat{\Lambda}^\diamond(t) = \sum_{i=1}^n \omega(t,i)\Delta_i^\diamond. \tag{37}$$

Proof. Let $\mathcal{N}(t)=[\max(t-\tau_n,0),\ \min(t+\tau_n,1)]$. For $i/n\in\mathcal{N}(t)$, we have

$$|\mathbb{E}(\Delta_i^\diamond)-\mathbb{E}(\Delta_i)|\leqslant\frac{\|\mathbf{Q}_i\|\|\mathbf{Q}_i^{\mathrm{T}}-(\mathbf{Q}_i^\diamond)^{\mathrm{T}}\|+\|\mathbf{Q}_i-\mathbf{Q}_i^\diamond\|\|\mathbf{Q}_i^\diamond\|}{2m+1}. \tag{38}$$

As mentioned in the proof of lemma 3, $\sup_i(\|\mathbf{Q}_i\|_4)=O(\surd m)$. Let $\mathbf{R}_j=\mathbf{U}(t_{i+j},\mathcal{F}_{i+j})-\mathbf{U}(t_i,\mathcal{F}_{i+j})$, $|j|\leqslant m$. By assumption 3, $\|\mathbf{R}_j\|=O(m/n)$. Also $\|\mathcal{P}_{i+j-k}\mathbf{R}_j\|\leqslant 2\,\delta_2(\mathbf{U},k)$, $k\geqslant 0$. So

$$\|\mathbf{Q}_i-\mathbf{Q}_i^\diamond\|\leqslant\sum_{k=0}^{\infty}\left\|\sum_{j=-m}^{m}\mathcal{P}_{i+j-k}\mathbf{R}_j\right\|=\sum_{k=0}^{\infty}O(\surd m)\min\{m/n,\delta_2(\mathbf{U},k)\} \tag{39}$$

uniformly over i with $i/n\in\mathcal{N}(t)$. Since $\delta_2(\mathbf{U},k)=O[\{k\log(k)\}^{-2}]$, $\|\mathbf{Q}_i-\mathbf{Q}_i^\diamond\|=O(mn^{-1/2})$. By inequality (38), expression (37) follows.

Lemma 5. Under the conditions of theorem 4, we have $|\Lambda(t)-\mathbb{E}\{\hat{\Lambda}^\diamond(t)\}|=O(m^{-1}+\tau_n^2)$.

Proof. Let $\Gamma_k=\mathbb{E}\{\mathbf{U}(t_i,\mathcal{F}_0)\,\mathbf{U}^{\mathrm{T}}(t_i,\mathcal{F}_k)\}, k\in\mathbb{Z}$. Since $\mathbf{U}(t_i,\mathcal{F}_k)=\Sigma_{j\in\mathbb{Z}}\,\mathcal{P}_j\mathbf{U}(t_i,\mathcal{F}_k)$ and $\mathcal{P}_j$ are orthogonal, we have

$$\begin{aligned}|\Gamma_k|&=\left|\mathbb{E}\left[\sum_{j\in\mathbb{Z}}\{\mathcal{P}_j\,\mathbf{U}(t_i,\mathcal{F}_0)\}\times\{\mathcal{P}_j\mathbf{U}^{\mathrm{T}}(t_i,\mathcal{F}_k)\}\right]\right|\leqslant\sum_{j\in\mathbb{Z}}\|\mathcal{P}_j\,\mathbf{U}(t_i,\mathcal{F}_0)\|\|\mathcal{P}_j\,\mathbf{U}(t_i,\mathcal{F}_k)\|\\&\leqslant\sum_{j\in\mathbb{Z}}\delta_2(\mathbf{U},\,-j)\,\delta_2(\mathbf{U},k-j)\\&=O\{(|k|\log|k|)^{-2}\},\end{aligned}$$

in view of $\delta_2(\mathbf{U},k)=O[\{k\log(k)\}^{-2}]$. So we have uniformly over $i\in\mathcal{N}(t)$ that

$$(2m+1)|\mathbb{E}(\tilde{\Delta}_i)-\Lambda(t_i)|=\sum_{j\in\mathbb{Z}}\min(|j|,2m+1)|\Gamma_j|=O(1). \tag{40}$$

Since $\mathbb{E}\{\hat{\Lambda}^\diamond(t)\}=\Sigma_{i=1}^n\omega(t,i)\,\mathbb{E}(\Delta_i^\diamond)$ and $\Lambda(t)\in\mathcal{C}^2[0,1]$, by equation (40), lemma 5 holds.

A.3. Proof of theorem 4

(a) The first part of theorem 4 follows from lemmas 3–5 in view of $\surd(m/n)=O(1/m)$.
(b) Following the chaining argument in lemma 6 of Zhou and Wu (2009) as well as the proof of lemma 3, we have

$$\|\sup_{t\in\mathcal{I}}|\hat{\Lambda}(t)-\mathbb{E}\{\hat{\Lambda}(t)\}|\|=O(m^{1/2}n^{-1/2}\tau_n^{-1}).$$

It is easily seen that lemmas 4 and 5 hold uniformly on $\mathcal{I}$. Hence the second part of theorem 4 follows.

A.4. Proof of theorem 5

Let $\mathcal{I}_1$ be a closed interval in (0, 1) such that $\mathcal{I}\subset\mathcal{I}_1$ and the two intervals do not share common end points. Since $\sup_{t\in[0,1]}\{\|\mathbf{G}(t,\mathcal{F}_i)\|_\kappa\}<\infty$, we have $\sup_i|\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}|=O_{\mathbb{P}}(n^{2/\kappa})$. Let ρ_n' be as in the proof of theorem 3. Then

$$\sup_{i/n\in\mathcal{I}_1}|\bar{\mathbf{L}}_i-\mathbf{L}_i|=\sup_{i/n\in\mathcal{I}_1}|\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\{\boldsymbol{\beta}(t_i)-\hat{\boldsymbol{\beta}}(t_i)\}|=O_{\mathbb{P}}(n^{2/\kappa}\rho_n'). \tag{41}$$

Note that $\mathbf{Q}_i/(2m+1)$ is the Nadaraya–Watson smoother of the series $(\mathbf{L}_i)$ at i with the rectangle kernel and bandwidth m/n. Hence, by the proof of theorem 3, it follows that

$$\sup_{i/n\in\mathcal{I}_1}|\mathbf{Q}_i|=O_{\mathbb{P}}[\surd\{m\log(n)\}+m^3/n^2]=O_{\mathbb{P}}[\surd\{m\log(n)\}]. \tag{42}$$

Define $\bar{\mathbf{Q}}_i=\Sigma_{j=-m}^m\bar{\mathbf{L}}_{i+j}$ and $\bar{\Delta}_i=\bar{\mathbf{Q}}_i\bar{\mathbf{Q}}_i^{\mathrm{T}}/(2m+1)$. Write

$$(2m+1)(\Delta_i-\bar{\Delta}_i)=(\mathbf{Q}_i-\bar{\mathbf{Q}}_i)\mathbf{Q}_i^{\mathrm{T}}+\mathbf{Q}_i(\mathbf{Q}_i-\bar{\mathbf{Q}}_i)^{\mathrm{T}}-(\mathbf{Q}_i-\bar{\mathbf{Q}}_i)(\mathbf{Q}_i-\bar{\mathbf{Q}}_i)^{\mathrm{T}}.$$

Plugging equations (41) and (42) into this equation, since $\vartheta_n\rightarrow 0$, we have $\sup_{i/n\in\mathcal{I}_1}|\Delta_i-\bar{\Delta}_i|=O_{\mathbb{P}}(\vartheta_n)$. By the definitions of $\tilde{\Lambda}(t)$ and $\hat{\Lambda}(t)$, we obtain $\sup_{t\in\mathcal{I}}|\hat{\Lambda}(t)-\tilde{\Lambda}(t)|=O_{\mathbb{P}}(\vartheta_n)$. Together with results in theorem 4, theorem 5 follows.

Lemma 6. Let $\chi_n=(nb_n)^{-1/2}+b_n$, $\chi_n'=n^{-1/2}b_n^{-1}+b_n$ and $h\geqslant 0$. Under conditions 2 and 4, we have

(a) for any fixed $t\in(0,1)$, $\mathbf{S}_{n,h}(t)-\mu_h M(t)=O_{\mathbb{P}}(\chi_n)$, and
(b) $\sup_{b_n\leqslant t\leqslant 1-b_n}|\mathbf{S}_{n,h}(t)-\mu_h M(t)|=O_{\mathbb{P}}(\chi_n')$.

A.5. Proof of lemma 6

As in expression (36), for $l\geqslant 0$, we have $\sup_j\|\mathcal{P}_{j-l}\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}\|\leqslant 2\,\delta_4(\mathbf{G},l)\kappa$, where $\kappa=\Sigma_{l=0}^{\infty}\delta_4(\mathbf{G},l)$. Let $\mathbf{s}_k=\Sigma_{j=1}^{k}\{\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}-\mathbb{E}(\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}})\}$ and $\mathbf{m}_{k,l}=\Sigma_{j=1}^{k}\mathcal{P}_{j-l}\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}$. Then $(\mathbf{m}_{k,l})_{k=1}^{n}$ is a martingale. By Doob's inequality, $\|\max_{k\leqslant n}|\mathbf{m}_{k,l}|\|\leqslant 2\|\mathbf{m}_{n,l}\|$. So

$$\|\max_{k\leqslant n}|\mathbf{s}_k|\|\leqslant\sum_{l=0}^{\infty}\|\max_{k\leqslant n}|\mathbf{m}_{k,l}|\|\leqslant\sum_{l=0}^{\infty}4\,\delta_4(\mathbf{G},l)\kappa\sqrt{n}=O(\sqrt{n}).\tag{43}$$

Let $w_j(t)=(nb_n)^{-1}K_{b_n}(t_j-t)$. By the summation-by-parts formula,

$$\sup_t|\mathbf{S}_{n,0}(t)-\mathbb{E}\{\mathbf{S}_{n,0}(t)\}|=\sup_t\left|\sum_{k=1}^{n}(\mathbf{s}_k-\mathbf{s}_{k-1})\,w_k(t)\right|\leqslant\max_{j\leqslant n}|\mathbf{s}_j|\frac{K_0}{nb_n},\tag{44}$$

where

$$K_0=2\sup_u|K(u)|+\int_{-1}^{1}|K'(u)|\,\mathrm{d}u.$$

Condition 2 implies that $M(\cdot)$ is Lipschitz continuous. Hence $\mathbb{E}\{\mathbf{S}_{n,h}(t)\}-\mu_h M(t)=O\{b_n+(nb_n)^{-1}\}$ holds uniformly over $b_n\leqslant t\leqslant 1-b_n$. By expressions (43) and (44), (b) holds with $h=0$. Since $\mathbf{S}_{n,0}(t)-\mathbb{E}\{\mathbf{S}_{n,0}(t)\}=\Sigma_{l=0}^{\infty}\mathbf{h}_l$, where $\mathbf{h}_l=\Sigma_{j=1}^{n}\mathcal{P}_{j-l}\,\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}}\,w_j(t)$ satisfies $\|\mathbf{h}_l\|=O\{(nb_n)^{-1/2}\}\,\delta_4(\mathbf{G},l)$, (a) follows. The general case with $h\geqslant 1$ can be similarly dealt with.

References

Bickel, P. J. and Rosenblatt, M. (1973) On some global measures of the deviations of density function estimates. *Ann. Statist.*, **1**, 1071–1095.

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

Cai, Z., Fan, J. and Li, R. Z. (2000) Efficient estimation and inferences for varying-coefficient models. *J. Am. Statist. Ass.*, **95**, 888–902.

Craven, P. and Wahba, G. (1979) Smoothing noisy data with spline functions. *Numer. Math.*, **31**, 377–403.

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

Doukhan, P. and Louhichi, S. (1999) A new weak dependence condition and applications to moment inequalities. *Stoch. Processes Appl.*, **84**, 313–342.

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

Eubank, R. L. and Speckman, P. L. (1993) Confidence bands in nonparametric regression. *J. Am. Statist. Ass.*, **88**, 1287–1301.

Fan, J. and Gijbels, I. (1996) *Local Polynomial Modelling and Its Applications*. New York: Chapman and Hall.

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

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

Fan, J. and Zhang, W. Y. (2000) Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scand. J. Statist.*, **27**, 715–731.

Hall, P. (1991) On the distribution of suprema. *Probab. Theor. Reltd Flds*, **89**, 447–455.

Härdle, W. and Marron, J. S. (1991) Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.*, **19**, 778–796.

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.

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

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

Lin, D. Y. and Ying, Z. (2001) Semiparametric and nonparametric regression analysis of longitudinal data (with discussion). *J. Am. Statist. Ass.*, **96**, 103–126.

Lindgren, G. (1980) Extreme values and crossings for the χ^2-process and other functions of multidimensional gaussian processes, with reliability applications. *Adv. Appl. Probab.*, **12**, 746–774.

Liu, W. and Lin, Z. (2009) Strong approximation for a class of stationary processes. *Stoch. Processes Appl.*, **119**, 249–280.
Neumann, M. H. and Kreiss, J. P. (1998) Regression-type inference in nonparametric autoregression. *Ann. Statist.*, **26**, 1570–1613.
Neumann, M. H. and Polzehl, J. (1998) Simultaneous bootstrap confidence bands in nonparametric regression. *J. Nonparam. Statist.*, **9**, 307–333.
Orbe, S., Ferreira, E. and Rodriguez-Poo, J. (2005) Nonparametric estimation of time varying parameters under shape restrictions. *J. Econometr.*, **126**, 53–77.
Orbe, S., Ferreira, E. and Rodriguez-Poo, J. (2006) On the estimation and testing of time varying constraints in econometric models. *Statist. Sin.*, **16**, 1313–1333.
Politis, D. N., Romano, J. P. and Wolf, M. (1999) *Subsampling*. New York: Springer.
Ramsay, J. and Silverman, B. W. (2005) *Functional Data Analysis*. New York: Springer.
Robinson, P. M. (1989) Nonparametric estimation of time-varying parameters. In *Statistical Analysis and Forecasting of Economic Structural Change* (ed. P. Hackl), pp. 164–253. Berlin: Springer.
Robinson, P. M. (1991) Time-varying nonlinear regression. In *Economic Structure Change Analysis and Forecasting* (eds P. Hackl and A. H. Westland), pp. 179–190. Berlin: Springer.
Wu, W. B. (2005) Nonlinear system theory: another look at dependence. *Proc. Natn. Acad. Sci. USA*, **102**, 14150–14154.
Wu, W. B. (2007) Strong invariance principles for dependent random variables. *Ann. Probab.*, **35**, 2294–2320.
Wu, C. O., Chiang, C. T. and Hoover, D. R. (1998) Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *J. Am. Statist. Ass.*, **93**, 1388–1402.
Wu, W. B. and Zhao, Z. (2007) Inference of trends in time series. *J. R. Statist. Soc.* B, **69**, 391–410.
Wu, W. B. and Zhou, Z. (2009) Gaussian approximations for non-stationary multiple time series. *Statist. Sin.*, to be published.
Zhou, Z. and Wu, W. B. (2009) Local linear quantile estimation of nonstationary time series. *Ann. Statist.*, **37**, 2696–2729.