Bernoulli **20**(1), 2014, 78–108
DOI: 10.3150/12-BEJ477

Nonparametric specification for non-stationary time series regression

ZHOU ZHOU

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

We investigate the behavior of the Generalized Likelihood Ratio Test (GLRT) (Fan, Zhang and Zhang [*Ann. Statist.* **29** (2001) 153–193]) for time varying coefficient models where the regressors and errors are non-stationary time series and can be cross correlated. It is found that the GLRT retains the minimax rate of local alternative detection under weak dependence and non-stationarity. However, in general, the Wilks phenomenon as well as the classic residual bootstrap are sensitive to either conditional heteroscedasticity of the errors, non-stationarity or temporal dependence. An averaged test is suggested to alleviate the sensitivity of the test to the choice of bandwidth and is shown to be more powerful than tests based on a single bandwidth. An alternative wild bootstrap method is proposed and shown to be consistent when making inference of time varying coefficient models for non-stationary time series.

Keywords: conditional heteroscedasticity; functional linear models; generalized likelihood ratio tests; local linear regression; local stationarity; weak dependence; wild bootstrap

1. Introduction

Specification tests are important in many nonparametric settings. Generally, one is interested in testing whether certain nonparametric components are significant, or whether they have a more parsimonious and efficient parametric representation. In the time series context, there is a large literature devoting to the latter topic, see for instance Hjellvik *et al.* [18], Fan and Li [16], Dette and Spreckelsen [9, 10], An and Cheng [1] and Paparoditis [31], among others. Many of the previous results perform specification for stationary time series.

The purpose of the paper is to develop specification tests for nonparametric regression of non-stationary time series. Specifically, consider the following time-varying coefficient model:

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta}(t_i) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{1}$$

where $t_i = i/n$, $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^\top$ are $p \times 1$ dimensional time series of regressors or predictors, ε_i are error series satisfying $\mathbb{E}(\varepsilon_i|\mathbf{x}_i) = 0$. Here $^\top$ denotes matrix or vector transpose. The processes $\{\mathbf{x}_i\}$ and $\{\varepsilon_i\}$ are allowed to be non-stationary and can be cross correlated. We assume that the regression parameters $\boldsymbol{\beta}(\cdot) := (\beta_1(\cdot), \ldots, \beta_p(\cdot))^\top$ is a smooth function on $[0,1]$. Nonparametric specification of model (1) boils down to testing whether $\boldsymbol{\beta}(\cdot)$ or a component of it has a certain parametric representation.

Due to their flexility and interpretability in investigating shifting association between the response and predictors over time, model (1) and its stochastic coefficient version have attracted considerable attention in various fields. See, for instance, Orbe *et al.* [29, 30], Cai [3], Brown *et al.* [2] an Stock and Watson [37] for applications in econometrics; Kitagawa and Gersch [23] and Gersch and Kitagawa [17] for applications in signal processing; Hoover *et al.* [20] and Ramsay and Silverman [34] for applications in longitudinal and functional data analysis. Most of the aforementioned literature on model (1) focused on parameter estimation. However, it seems that the important issue of model validation or specification of (1) have received little attention.

For varying coefficient models of i.i.d. samples, Fan, Zhang and Zhang [15] proposed the generalized likelihood ratio test (GLRT) as a general rule for nonparametric specification; see also Dette [6] for a closely related earlier test based on nonparametric analysis of variance (ANOVA). We also refer to the excellent review paper of Fan and Jiang [14] and the references cited therein for a more detailed discussion of the GLRT and related tests. The GLRT has three major advantages. First, it is of simple and intuitively appealing form. For instance, consider testing

$$H_0\colon\ \boldsymbol{\beta}(\cdot) = \boldsymbol{\beta}_0(\cdot) \quad \longleftrightarrow \quad H_a\colon\ \boldsymbol{\beta}(\cdot) \neq \boldsymbol{\beta}_0(\cdot), \tag{2}$$

where $\boldsymbol{\beta}_0(\cdot)$ is a known function on $[0,1]$. Then the GLRT statistic is proportional to $(\mathrm{RSS}_0 - \mathrm{RSS}_a)/\mathrm{RSS}_0$, where RSS_0 and RSS_a are residual sum of squares under the null and alternative hypothesis, respectively. Hence, it is similar in form to the classic analysis of variance. Second, the GLRT is powerful to apply. Fan, Zhang and Zhang [15] showed that the GLRT can detect local alternatives with the optimal rate in the sense of Ingster [22]. Third, the test is asymptotically nuisance parameter free; known as the Wilks phenomenon. The Wilks phenomenon insures that the residual wild bootstrap, that is, drawing i.i.d. samples from the centered empirical distribution of the residuals, is asymptotically consistent for the inference. In fact, the Wilks phenomenon is shown to hold for a wide range of nonparametric models when testing under the GLRT. See, for instance, Fan and Jiang [13] for additive models and Fan and Huang [12] for varying coefficient partially linear models. For state-domain nonparametric regression for stationary time series, Hong and Lee [19] showed that the Wilks phenomenon continue to hold when the errors are conditionally homogeneous.

In this paper, we shall prove that the Wilks phenomenon is sensitive to either conditional heteroscedasticity of the errors, non-stationarity or temporal dependence in model (1). In particular, the Wilks phenomenon fails for model (1) even when the errors and regressors are stationary and conditionally homogeneous. The latter finding is drastically different from the state domain regression case in Hong and Lee [19] where the Wilks phenomenon is shown to hold when the errors are conditionally homogeneous. As a consequence, the residual wild bootstrap fails for model (1) under dependence since the latter bootstrap generates (conditional) i.i.d. samples and hence mimics the Wilks type asymptotic behavior. A new robust methodology is needed when performing model specification for (1) under dependence and non-stationarity.

According to a result on Gaussian quadratic form approximation to the GLRT, we shall propose in this paper a new wild bootstrap method for the nonparametric specification of model (1). The latter bootstrap is shown to be consistent under non-stationarity and dependence. We further discover that the GLRT, though fails to be asymptotically pivotal, retains the minimax rate of local alternative detection under weak dependence and non-stationarity. Hence, the GLRT with the robust wild bootstrap is powerful to apply. Note that Zhou and Wu [43] discussed simultaneous confidence band (SCB) construction for model (1) which could be used for model specification. However, the SCB can detect local alternatives with inferior rates than that of the GLRT and hence is not a powerful tool for specification.

It is known that nonparametric specification is sensitive to the choice of smoothing bandwidth. To alleviate the problem, Horowitz and Spokoiny [21] and Fan, Zhang and Zhang [15], among others, proposed to maximize the test statistic over a wide range of bandwidths. However, for the GLRT test, the asymptotic behavior of the resulting statistic is unknown even for i.i.d. samples, which hampers the application of the latter test. It is worth mentioning that Zhang [41] derived the asymptotic null distribution of the maximum test for a bounded number of bandwidths. On the other hand, Müller [25] suggested to average the GLRT over a range of bandwidths as an alternative to the maximum test. The latter suggestion stems from surprising results, such as Lehmann [24], that the averaged likelihood ratio test can be more powerful than the maximum likelihood ratio test for complex alternatives. In this paper, we shall propose to use the averaged test for the specification of model (1) to alleviate the sensitivity of the test to the choice of bandwidth. We derive the asymptotic distribution and the local power of the averaged test. It is found that the averaged test is asymptotically at least as powerful as the best test based on a single bandwidth regardless of the shape of the alternative, the non-stationary dependence structure of the data or the kernel function. Our finding is potentially interesting for a wide range of nonparametric specification problems.

Recently, there have been many results on modeling non-stationary time series from the spectral domain. See, for instance, Dahlhaus [4], Nason *et al.* [26] and Ombao *et al.* [28], among others. At the same time, there is a great recent interest in specification of non-stationary time series in the spectral domain. Examples include, among others, Dahlhaus [5], Neumann and von Sachs [27], Paparoditis [32, 33], Sergides and Paparoditis [36] and Dette *et al.* [8]. However, for the varying coefficient regression (1), models from the spectral domain do not seem to be directly useful for an asymptotic theory. In this paper, we shall adopt the time domain modeling of locally stationary time series in Zhou and Wu [42]. The latter framework and the associated dependence measures directly facilitate the theory of the current paper.

The rest of the paper is organized as follows. Section 2 introduces the GLRT statistic and the non-stationary time series models for the error and regressor series. In Section 3, we shall derive the asymptotic null distribution and local power of the GLRT for parametric and semi-parametric null hypotheses. A detailed discussion on the failure of the Wilks phenomenon is included. In Section 4, we shall introduce the averaged test and the corresponding robust bootstrap and investigate their asymptotic behavior. In Section 5, we shall construct a monte carlo experiment to study the finite sample accuracy of the proposed averaged test. Proofs of the asymptotic results are placed in Section 6.

2. Preliminaries

2.1. The GLRT statistics

Consider the testing problem (2). The GLRT compares the residual sum of squares (RSS) under the null and alternative hypotheses, and a large difference indicates violation of the null. We refer to Fan, Zhang and Zhang [15] for a detailed derivation of the statistic. Specifically, the GLRT statistic

$$\lambda_n = \frac{n}{2}\log\frac{\mathrm{RSS}_0}{\mathrm{RSS}_a} \approx -\frac{n}{2}\frac{\mathrm{RSS}_a - \mathrm{RSS}_0}{\mathrm{RSS}_0}, \tag{3}$$

where $\mathrm{RSS}_0 = \sum_{i=1}^n (y_i - \mathbf{x}_i^\top \boldsymbol{\beta}_0(t_i))^2$ is the RSS under the null hypothesis and $\mathrm{RSS}_a = \sum_{i=1}^n (y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}(t_i))^2$ is the RSS under the nonparametric alternative. Here $\hat{\boldsymbol{\beta}}(\cdot)$ is the local linear kernel estimate of $\boldsymbol{\beta}(\cdot)$ (Fan and Gijbels, [11]), which is defined as

$$(\hat{\boldsymbol{\beta}}_{b_n}(t), \hat{\boldsymbol{\beta}}'_{b_n}(t)) = \operatorname*{argmin}_{\eta_0,\eta_1\in\mathbb{R}^p} \sum_{i=1}^n (y_i - \mathbf{x}_i^\top \eta_0 - \mathbf{x}_i^\top \eta_1 (t_i - t))^2 K_{b_n}(t_i - t), \tag{4}$$

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 $K \in \mathcal{C}^1[-1,1]$. Define

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

for $l = 0, 1, \ldots$, where $0^0 := 1$, 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 $\hat{\boldsymbol{\eta}}_{b_n}(t) = (\hat{\boldsymbol{\beta}}_{b_n}^\top(t), b_n(\hat{\boldsymbol{\beta}}'_{b_n}(t))^\top)^\top$. Then it can be shown that (Fan and Gijbels, [11])

$$\hat{\boldsymbol{\eta}}_{b_n}(t) = \begin{pmatrix} \mathbf{S}_{n,0}(t) & \mathbf{S}_{n,1}^\top(t) \\ \mathbf{S}_{n,1}(t) & \mathbf{S}_{n,2}(t) \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{R}_{n,0}(t) \\ \mathbf{R}_{n,1}(t) \end{pmatrix} := \mathbf{S}_n^{-1}(t)\mathbf{R}_n(t). \tag{5}$$

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.

2.2. Locally stationary time series models

Throughout this paper, we shall assume that both $(\mathbf{x}_i)$ and (ε_i) belong to a general class of locally stationary time series in the sense of Zhou and Wu [42] as follows,

$$\begin{aligned} \mathbf{x}_i &= \mathbf{G}(t_i, (\ldots, \epsilon_{i-1}, \epsilon_i)), \qquad i = 1, 2, \ldots, n, \\ \varepsilon_i &= H(t_i, (\ldots, \xi_{i-1}, \xi_i))V(t_i, (\ldots, \epsilon_{i-1}, \epsilon_i)), \qquad i = 1, 2, \ldots, n, \end{aligned} \tag{6}$$

where $\mathbf{G}(\cdot) = (G_1, G_2, \ldots, G_p)^\top(\cdot)$, $(\epsilon_i)_{i\in\mathbb{Z}}$ are i.i.d., $(\xi_i)_{i\in\mathbb{Z}}$ are also i.i.d. and $(\epsilon_i)_{i\in\mathbb{Z}}$ is independent of $(\xi_i)_{i\in\mathbb{Z}}$. Let $\mathcal{F}_i = (\ldots, \epsilon_{i-1}, \epsilon_i)$ and $\mathcal{G}_i = (\ldots, \xi_{i-1}, \xi_i)$. We assume that

$$\mathbb{E}(H(t, \mathcal{G}_i)) = 0 \quad \text{and} \quad \mathrm{Var}(H(t, \mathcal{G}_i)) = 1,$$

almost surely for all $t \in [0, 1]$, in which case $V^2(t_i, \mathcal{F}_i)$ is the conditional variance of ε_i given $\mathcal{F}_i$.

It is clear from (6) that $(\mathbf{x}_i)$ and (ε_i) are non-stationary. Formulation (6) can be interpreted as physical systems with $\mathcal{F}_i$ and $\mathcal{G}_i$ being the inputs and $\mathbf{x}_i$, ε_i being the outputs, respectively, and $\mathbf{G}$, H and V being the transforms or filters that represent the underlying physical mechanism. By allowing $\mathbf{G}$, H and V varying smoothly with respect to t, we have local stationarity of $(\mathbf{x}_i)$ and (ε_i). See also Zhou and Wu [42] for more discussions. The above formulation of covariates and error processes is very general and includes many settings in the existing time series regression literature as special cases. To help understand the formulation, we shall consider the following three cases:

(a) (I.i.d. model). Assume that $\mathbf{x}_i = \mathbf{G}_0(\epsilon_i)$ and $\varepsilon_i = H_0(\xi_i)$. Then $(\mathbf{x}_i^\top, \varepsilon_i)_{i=1}^n$ is a random sample and $(\varepsilon_i)_{i=1}^n$ is independent of $(\mathbf{x}_i)_{i=1}^n$. This type of design was discussed extensively in Fan, Zhang and Zhang [15] and Fan and Jiang [14], among others.

(b) (Exogenous model). In (6), we assume that $V(t_i, \mathcal{F}_i) = V_0(t_i)$. In this case, the regressors and errors are two independent locally stationary processes. Under further restrictions on the processes, this type of model was studied in Robinson [35], Orbe *et al.* [29, 30] among others.

(c) (Endogenous model). Assume (6). Note that in this case the errors are correlated with the regressors since they both depend on inputs $\mathcal{F}_i$. This type of model is suitable when the errors exhibit heteroscedasticity with respect to time and the regressors. When $\mathbf{x}_i$ and $H(t, \mathcal{G}_i)$ are stationary, the case was considered in Cai [3] among others.

Write $\chi_i = (\epsilon_i, \xi_i)^\top$ and $\mathcal{R}_i = (\ldots, \chi_{i-1}, \chi_i)$. For a generic locally stationary time series $Z_i = \mathbf{L}(t_i, \mathcal{R}_i)$. The strength of the temporal dependence in $\{Z_i\}$ can be measured by how strongly the 'current' observation of the time series, Z_i, is influenced by the innovation χ_0 which occurred i steps ahead. More specifically, we can define

$$\delta_p(\mathbf{L}, k) = \sup_{0\le t\le 1} \|\mathbf{L}(t, \mathcal{R}_k) - \mathbf{L}(t, \mathcal{R}_k^*)\|_p \qquad \text{where } \mathcal{R}_k^* = (\mathcal{R}_{-1}, \chi_0^*, \chi_1, \chi_2, \ldots, \chi_i) \tag{7}$$

and $\{\chi_i^*\}$ is an i.i.d. copy of $\{\chi_i\}$. Implementing the idea of coupling, $\delta_p(\mathbf{L}, k)$ measures the effect of χ_0 in generating observations that are k steps away. Therefore, if $\delta_p(\mathbf{L}, k)$

decays fast as k gets large, short range dependence is implied. We refer to Zhou and Wu [42] for more discussions and examples on the above dependence measures.

3. Asymptotic results

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

(A1) Let $M(t)$ be the $p\times p$ matrix with (i,j)th entry $m_{ij}(t)=\mathbb{E}[G_i(t,\mathcal{F}_0)G_j(t,\mathcal{F}_0)]$. Assume that the smallest eigenvalue of $M(t)$ is bounded away from 0 on $[0,1]$ and $M(t)\in\mathcal{C}^2[0,1]$.
(A2) $\mathbf{G}(t,\mathcal{F}_i)\in\mathcal{U}^{32}\cap\mathrm{Lip}_2$ for some $r>0$.
(A3) $\mathbf{U}(t,\mathcal{R}_i):=\mathbf{G}(t,\mathcal{F}_i)V(t,\mathcal{F}_i)H(t,\mathcal{G}_i)\in\mathcal{U}^4\cap\mathrm{Lip}_2$.
(A4) $\sum_{k=0}^{\infty}\delta_{32}(\mathbf{G},k)<\infty$.
(A5) $\delta_4(V,k)+\delta_4(H,k)=\mathrm{O}((k+1)^{-2})$.
(A6) $\delta_4(\mathbf{U},k)=\mathrm{O}(\chi^k)$ for some $\chi\in(0,1)$.
(A7) The smallest eigenvalue of $\Lambda(t)$ is bounded away from 0 on $[0,1]$, where

$$\Lambda(t)=\sum_{i=-\infty}^{\infty}\mathrm{cov}(\mathbf{U}(t,\mathcal{R}_0),\mathbf{U}(t,\mathcal{R}_i)). \tag{8}$$

(A8) The coefficient functions $\beta_j(\cdot)\in\mathcal{C}^2[0,1]$, $j=1,\ldots,p$.

A few remarks on the regularity conditions are in order. Conditions (A1), (A2) and (A4) insures local stationarity and short memory of the regressor process $\mathbf{x}_i$. The existence of the 32rd moment is for technical convenience only and may be relaxed. The eigenvalue constraint in condition (A1) insures the non-singularity of the design. Conditions (A3), (A5) and (A6) guarantees the smoothness and short range dependence of the error process ε_i. Furthermore, condition (A7) means that the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}(t)$ is non-singular.

3.1. The null distributions

Theorem 1. *Assume that condition* (A) *holds and that* $nb_n^{9/2}=\mathrm{O}(1)$ *and* $nb_n^4/(\log n)^6\to\infty$. *Then under* H_0, *we have*

$$\sqrt{b_n}\left\{2\lambda_n+\frac{\tilde{K}(0)}{b_n\mathcal{V}}\int_0^1\mathrm{tr}[H(t)]\,\mathrm{d}t+\frac{nb_n^4\mu_2^2}{4\mathcal{V}}\int_0^1[\boldsymbol{\beta}''(t)]^\top M(t)\boldsymbol{\beta}''(t)\,\mathrm{d}t\right\}\Rightarrow N(0,\sigma^2/\mathcal{V}^2), \tag{9}$$

where

$$\sigma^2 = \int_{\mathbb{R}} \tilde{K}^2(t)\,\mathrm{d}t \int_0^1 \mathrm{tr}[H(t)^2]\,\mathrm{d}t,$$

$\tilde{K}(\cdot) = K * K(\cdot) - 2K(\cdot)$, $H(\cdot) = \Lambda^{1/2}(\cdot)M^{-1}(\cdot)\Lambda^{1/2}(\cdot)$, $\mathcal{V} = \int_0^1 \mathbb{E}[V(t,\mathcal{F}_0)]^2\,\mathrm{d}t$, $\mu_2 = \int_{-1}^1 x^2K(x)\,\mathrm{d}x$, *'$*$' is the convolution operator and '*tr*' denotes the trace of a matrix.*

Theorem 1 reveals the asymptotic behavior of the GLRT for a very wide class of predictor and error processes. In particular, the latter Theorem explains when and why the Wilks phenomenon fails. In the following, we will consider four special cases to see how endogeneity, non-stationarity and temporal dependence influence the Wilks phenomenon. To simplify the discussion, we will assume in the examples below that the asymptotic bias effect, $\frac{nb_n^4\mu_2^2}{4\mathcal{V}}\int_0^1[\boldsymbol{\beta}''(t)]^\top M(t)\boldsymbol{\beta}''(t)\,\mathrm{d}t$, is asymptotically negligible in (9). In practice, the latter task can be achieved by pre-whitening. We will discuss bias reduction techniques for GLRT in Section 4.2.

Example 1 (I.i.d. sample without endogeneity). Consider the case when $\mathbf{x}_i = \mathbf{G}(\epsilon_i)$ and $\varepsilon_i = CH(\zeta_i)$, where C is a positive constant. In this case, the covarites and errors are two independent i.i.d. sequences and the conditions in Fan, Zhang and Zhang [15] are satisfied. Note that $\mathcal{V} = C^2$, $\Lambda(t) = M(t)C^2$ and $H(t) = C^2\mathbf{I}_p$, where $\mathbf{I}_p$ is the $p\times p$ identity matrix. In particular,

$$\int_0^1 \mathrm{tr}[H(t)]\,\mathrm{d}t/\mathcal{V} = p \quad \text{and} \quad \int_0^1 \mathrm{tr}[H(t)^2]\,\mathrm{d}t/\mathcal{V}^2 = p \tag{10}$$

in (9). Hence, it is easy to check that

$$\sqrt{b_n}\left\{2\lambda_n + \frac{p\tilde{K}(0)}{b_n}\right\} \Rightarrow N\left(0, p\int_{\mathbb{R}}\tilde{K}^2(t)\,\mathrm{d}t\right),$$

which coincides with Theorem 5 of Fan, Zhang and Zhang [15] and the Wilks phenomenon holds.

Example 2 (The effect of temporal dependence). In this case $\mathbf{x}_i = \mathbf{G}(\mathcal{F}_i)$ and $\varepsilon_i = CH(\mathcal{G}_i)$, where C is a positive constant. Hence, $\{\mathbf{x}_i\}$ and $\{\varepsilon_i\}$ are two stationary processes which are independent of each other. In particular, neither endogeneity nor non-stationary is assumed in the model. It is easy to see that, in this case,

$$\Lambda(t) = C^2\sum_{i=-\infty}^{\infty}\mathbb{E}[\mathbf{G}(\mathcal{F}_0)\mathbf{G}^\top(\mathcal{F}_i)]\mathbb{E}[H(\mathcal{G}_0)H(\mathcal{G}_i)], \tag{11}$$

$\mathcal{V} = C^2$ and $M(t) = \mathbb{E}[\mathbf{G}(\mathcal{F}_0)\mathbf{G}^\top(\mathcal{F}_0)]$. An important observation is that

$$\int_0^1 \mathrm{tr}[H(t)]\,\mathrm{d}t/\mathcal{V} = \mathrm{tr}\left(\{\mathbb{E}[\mathbf{G}(\mathcal{F}_0)\mathbf{G}^\top(\mathcal{F}_0)]\}^{-1}\sum_{i=-\infty}^{\infty}\mathbb{E}[\mathbf{G}(\mathcal{F}_0)\mathbf{G}^\top(\mathcal{F}_i)]\mathbb{E}[H(\mathcal{G}_0)H(\mathcal{G}_i)]\right),$$

$$\int_0^1 \mathrm{tr}[H(t)^2]\,\mathrm{d}t/\mathcal{V}^2 = \mathrm{tr}\left(\left[\{\mathbb{E}[\mathbf{G}(\mathcal{F}_0)\mathbf{G}^\top(\mathcal{F}_0)]\}^{-1}\sum_{i=-\infty}^{\infty}\mathbb{E}[\mathbf{G}(\mathcal{F}_0)\mathbf{G}^\top(\mathcal{F}_i)]\mathbb{E}[H(\mathcal{G}_0)H(\mathcal{G}_i)]\right]^2\right)$$

are no longer nuisance parameter free compared with the results in (10). As a consequence, the Wilks phenomenon fails to hold in this case. Additionally, it is easy to see that the latter loss of pivotality is due to the fact that the summands in (11) are generally nonzero for $i \neq 0$, which is caused by the temporal dependence. Indeed, if the summands are zero for $i \neq 0$ in (11), then $\Lambda(t) = C^2\mathbb{E}[\mathbf{G}(\mathcal{F}_0)\mathbf{G}^\top(\mathcal{F}_0)]$ and we have (10). Like in many pivotal tests such as the Wald test, the term $\mathrm{RSS}_0/n \approx \mathcal{V}$ in the GLRT serves as a scaling device which cancels out the variance factor in $\mathrm{RSS}_1 - \mathrm{RSS}_0$ and makes the test pivotal in the i.i.d. case. However, as shown above, RSS_0/n fails to fulfill the latter scaling task under dependence.

Example 3 (The effect of non-stationarity). Let $\mathbf{x}_i = \mathbf{G}(t_i, \epsilon_i)$ and $\varepsilon_i = V(t_i)H(t_i, \zeta_i)$. Here $\{\mathbf{x}_i\}$ and $\{\varepsilon_i\}$ are two independent but non-stationary sequences which are independent of each other. In this case, we have

$$\int_0^1 \mathrm{tr}[H(t)]\,\mathrm{d}t/\mathcal{V} = p \quad \text{and} \quad \int_0^1 \mathrm{tr}[H(t)^2]\,\mathrm{d}t/\mathcal{V}^2 = p\frac{\int_0^1 V^4(t)\,\mathrm{d}t}{(\int_0^1 V^2(t)\,\mathrm{d}t)^2}. \tag{12}$$

Note that the second term in (12) depends on the time-varying variance $V^2(t)$ and hence the Wilks phenomenon fails to hold in this case. Additionally, observe that $\frac{\int_0^1 V^4(t)\,\mathrm{d}t}{(\int_0^1 V^2(t)\,\mathrm{d}t)^2} \geq 1$ and the equation holds if and only if $V(t)$ is a constant function. Compared with the results in (10), we conclude that, in this case, non-stationarity in the errors tends to inflate the variance of GLRT. Furthermore, if $\{\varepsilon_i\}$ has constant variance, then the Wilks phenomenon holds even if $\{\mathbf{x}_i\}$ is a non-stationary sequence.

Example 4 (The effect of endogeneity). Suppose that $\mathbf{x}_i = \mathbf{G}(\epsilon_i)$ and $\varepsilon_i = V(\epsilon_i)H(\zeta_i)$. In this case $\{\mathbf{x}_i\}$ and $\{\varepsilon_i\}$ are two i.i.d. sequences which are dependent of each other. We obtain

$$\int_0^1 \mathrm{tr}[H(t)]\,\mathrm{d}t/\mathcal{V} = \mathrm{tr}(\{\mathbb{E}[\mathbf{G}(\epsilon_0)\mathbf{G}^\top(\epsilon_0)]\}^{-1}\mathbb{E}[\mathbf{G}(\epsilon_0)\mathbf{G}^\top(\epsilon_0)V^2(\epsilon_0)])/\mathbb{E}[V^2(\epsilon_0)],$$

$$\int_0^1 \mathrm{tr}[H(t)^2]\,\mathrm{d}t/\mathcal{V}^2 = \mathrm{tr}([\{\mathbb{E}[\mathbf{G}(\epsilon_0)\mathbf{G}^\top(\epsilon_0)]\}^{-1}\mathbb{E}[\mathbf{G}(\epsilon_0)\mathbf{G}^\top(\epsilon_0)V^2(\epsilon_0)]]^2)/(\mathbb{E}[V^2(\epsilon_0)])^2.$$

Note that if $\mathbb{E}[\mathbf{G}(\epsilon_0)\mathbf{G}^\top(\epsilon_0)V^2(\epsilon_0)] = \mathbb{E}[\mathbf{G}(\epsilon_0)\mathbf{G}^\top(\epsilon_0)]\mathbb{E}[V^2(\epsilon_0)]$, then we have (10) and hence the Wilks phenomenon. Due to the dependence of $\mathbf{G}(\epsilon_0)$ and $V(\epsilon_0)$, the latter factorization generally fails and hence the Wilks phenomenon fails to hold in this case.

In many real applications, one is interested in specifying a component of $\boldsymbol{\beta}(\cdot)$. For instance, one may want to test whether $\beta_j(\cdot)$ is significantly different from zero. This leads us to consider the following hypothesis testing problem where both H_{01} and H_{a1}

are nonparametric:

$$H_{01}\colon\ \boldsymbol{\beta}^{(1)}(\cdot)=\boldsymbol{\beta}_0^{(1)}(\cdot)\quad\longleftrightarrow\quad H_{a1}\colon\ \boldsymbol{\beta}^{(1)}(\cdot)\neq\boldsymbol{\beta}_0^{(1)}(\cdot), \tag{13}$$

where

$$\boldsymbol{\beta}(t)=\begin{pmatrix}\boldsymbol{\beta}^{(1)}(t)\\ \boldsymbol{\beta}^{(2)}(t)\end{pmatrix},\qquad \boldsymbol{\beta}_0(t)=\begin{pmatrix}\boldsymbol{\beta}_0^{(1)}(t)\\ \boldsymbol{\beta}_0^{(2)}(t)\end{pmatrix}\quad\text{and}\quad \mathbf{x}_i=\begin{pmatrix}\mathbf{x}_i^{(1)}\\ \mathbf{x}_i^{(2)}\end{pmatrix},$$

$\boldsymbol{\beta}^{(1)}(t)$, $\boldsymbol{\beta}_0^{(1)}(t)$ and $\mathbf{x}_i^{(1)}$ are $p_1<p$ dimensional and $\boldsymbol{\beta}_0^{(1)}(t)$ is a known function. Define $y_i^*=y_i-[\boldsymbol{\beta}_0^{(1)}(t_i)]^\top\mathbf{x}_i^{(1)}$. Then under H_{01} the functions $\beta_j(\cdot)$, $j=p_1+1,\ldots,p$ can be estimated by the local linear regression of y_i^* on $\mathbf{x}_i^{(2)}$ with bandwidth b_n. Throughout the paper we assume that the bandwidth b_n used under H_{01} is the same as that under H_{a1}. Asymptotic results can be easily obtained using the arguments of the paper when the two bandwidths are different. However, the resulting asymptotic bias and variance are much more complicated. For the sake of presentational clarity, we will only consider the case of equal bandwidth.

The GLRT statistic for testing H_{01} against H_{a1} is defined as

$$\lambda_{1n}=\frac{n}{2}\log\frac{\mathrm{RSS}_1}{\mathrm{RSS}_a}=\frac{n}{2}\left[\log\frac{\mathrm{RSS}_1}{\mathrm{RSS}_0}-\log\frac{\mathrm{RSS}_a}{\mathrm{RSS}_0}\right]\approx-\frac{n}{2}\frac{\mathrm{RSS}_a-\mathrm{RSS}_1}{\mathrm{RSS}_0}, \tag{14}$$

where RSS_1 is the RSS under H_{01}.

Write

$$M(t)=\begin{pmatrix}M_{11}(t) & M_{12}(t)\\ M_{21}(t) & M_{22}(t)\end{pmatrix}\quad\text{and}\quad \Lambda(t)=\begin{pmatrix}\Lambda_{11}(t) & \Lambda_{12}(t)\\ \Lambda_{21}(t) & \Lambda_{22}(t)\end{pmatrix},$$

where $M_{11}(t)$ and $\Lambda_{11}(t)$ are of dimension $p_1\times p_1$.

Define $p\times p$ matrix $H_2(t)=\operatorname{diag}(\mathbf{0}_{p_1},\Lambda_{22}^{1/2}(t)M_{22}^{-1}(t)\Lambda_{22}^{1/2}(t))$. We have the following theorem.

Theorem 2. *Assume that condition* (A) *holds and that* $nb_n^{9/2}=\mathrm{O}(1)$ *and* $nb_n^4/(\log n)^6\to\infty$. *Then under* H_{01}, *we have*

$$\sqrt{b_n}\left\{2\lambda_{1n}+\frac{\tilde{K}(0)}{b_n\mathcal{V}}\int_0^1\operatorname{tr}[H^*(t)]\,\mathrm{d}t+\frac{nb_n^4\mu_2^2}{4\mathcal{V}}\int_0^1\Upsilon(t)\,\mathrm{d}t\right\}\Rightarrow N(0,\sigma_1^2/\mathcal{V}^2),$$

where $H^*(\cdot)=H(\cdot)-H_2(\cdot)$, $\Upsilon(t)=[\boldsymbol{\beta}''(t)]^\top M(t)\boldsymbol{\beta}''(t)-\{[\boldsymbol{\beta}^{(2)}(t)]''\}^\top M_{22}(t)[\boldsymbol{\beta}^{(2)}(t)]''$ *and*

$$\sigma_1^2=\int_R\tilde{K}^2(t)\,\mathrm{d}t\int_0^1\operatorname{tr}[\{H^*(t)\}^2]\,\mathrm{d}t.$$

Theorem 2 unveils the asymptotic null distribution of the test under H_{01}. Following very similar arguments as those in Examples 1–4, the Wilks phenomenon can be shown to be sensitive to non-stationary, temporal dependence and endogeneity in this case as well.

Practitioners and researchers often encounter testing problems where the null is specified up to a parametric part. For instance, one may want to test whether $\boldsymbol{\beta}(\cdot)$ is really time varying in model (1), which amounts to testing $\boldsymbol{\beta}(\cdot) = C$ for some unspecified constant vector C. Heuristically, since the convergence rate of the local linear estimates is always slower than the $\sqrt{n}$ parametric rate, it is expected that the null distribution will not be altered as long as we plug in a $\sqrt{n}$ consistent estimate of the unspecified parametric part. The following discussion rigorously confirms the intuition. Consider testing

$$\tilde{H}_{01}\colon\ \boldsymbol{\beta}^{(1)}(\cdot) = \boldsymbol{\beta}_0^{(1)}(\cdot, \theta_0) \qquad \text{for some unknown } \theta_0 \in \Omega \subset \mathbb{R}^q,$$

where $\{\boldsymbol{\beta}_0^{(1)}(\cdot,\theta)\colon\ \theta \in \Omega\}$ is a parametric family of smooth functions. Let $\tilde{y}_i^* = y_i - (\boldsymbol{\beta}_0^{(1)})^\top \times (t_i, \hat{\theta})\mathbf{x}_i^{(1)}$ and $\widetilde{\mathrm{RSS}}_1$ be the residual sum of squares of the local linear regression of $\tilde{y}_i^*$ on $\mathbf{x}_i^{(2)}$ with bandwidth b_n. We shall make the following assumptions on the parametric family $\boldsymbol{\beta}_0^{(1)}(\cdot,\theta)$ and the estimate $\hat{\theta}$:

(B1) For each $t \in [0,1]$, $\boldsymbol{\beta}_0^{(1)}(t,\theta)$ is $\mathcal{C}^2$ in θ in a neighborhood Θ of θ_0. Additionally,

$$\sup_{t\in[0,1],\theta\in\Theta} \left\{ \left| \frac{\partial \boldsymbol{\beta}_0^{(1)}(t,\theta)}{\partial\theta} \right| + \left| \frac{\partial^2 \boldsymbol{\beta}_0^{(1)}(t,\theta)}{\partial\theta^2} \right| \right\} < \infty.$$

(B2) Under $\tilde{H}_{01}$, $\|\hat{\theta} - \theta_0\|_4 = \mathrm{O}(1/\sqrt{n})$.

Proposition 1. *Under $\tilde{H}_{01}$, condition* (B) *and the assumptions of Theorem 2, we have*

$$\widetilde{\mathrm{RSS}}_1 - \mathrm{RSS}_1 - \mathrm{O}_{\mathbb{P}}(\sqrt{n} b_n^2) = \mathrm{O}_{\mathbb{P}}(1). \tag{15}$$

The $\mathrm{O}_{\mathbb{P}}(\sqrt{n} b_n^2)$ term on the left-hand side of (15) corresponds to the extra bias introduced by the estimation error of θ. And the $\mathrm{O}_{\mathbb{P}}(1)$ term on the right-hand side of (15) corresponds to the extra variance caused by the latter error. Both terms are asymptotically negligible compared to the $\mathrm{O}_{\mathbb{P}}(nb_n^4)$ bias and $\mathrm{O}_{\mathbb{P}}(1/b_n)$ variance of RSS_1. As a consequence, the results of Theorems 1 and 2 continues to hold if θ is replaced by $\hat{\theta}$.

3.2. Local power of the GLRT

Proposition 2. *Assume the alternative $H_{a,n}\colon\ \boldsymbol{\beta}(\cdot) = \boldsymbol{\beta}_0(\cdot) + n^{-4/9}\mathbf{f}_n(\cdot)$, where $\mathbf{f}_n(\cdot) \in \mathcal{C}^2[0,1]$. Further assume that $b_n = cn^{-2/9}$ for some $c > 0$, that $\int_0^1 |\mathbf{f}_n''(t)|\,\mathrm{d}t = \mathrm{o}(n^{4/9})$ and that*

$$\int_0^1 \mathbf{f}_n^\top(t) M(t) \mathbf{f}_n(t)\,\mathrm{d}t \to F_1, \qquad n^{-8/9} \int_0^1 [\mathbf{f}_n''(t)]^\top M(t) \mathbf{f}_n''(t)\,\mathrm{d}t \to F_2$$

for some finite constants F_1 *and* F_2*. Then under condition* (A)*, we have*

$$\sqrt{b_n}\left\{2\lambda_n + \frac{\tilde{K}(0)}{b_n\mathcal{V}}\int_0^1 \mathrm{tr}[H(t)]\,\mathrm{d}t\right\} + \frac{c^{9/2}\mu_2^2}{4\mathcal{V}}\int_0^1 [\boldsymbol{\beta}''(t)]^\top M(t)\boldsymbol{\beta}''(t)\,\mathrm{d}t + \frac{c^{9/2}\mu_2^2}{4\mathcal{V}}F_2 - \frac{c^{1/2}}{\mathcal{V}}F_1$$
$$\Rightarrow N(0, \sigma^2/\mathcal{V}^2).$$

When the errors and regressors are weakly dependent locally stationary time series, Proposition 2 claims that the GLRT can still detect local alternatives with the optimal rate $\mathrm{O}(n^{-4/9})$ in the sense of Ingster [22]. As a consequence, the GLRT is powerful to apply for nonparametric model validation of model (1) under non-stationarity and dependence. However, it should be noted that the GLRT may not be the most powerful among all rate optimal tests. In the literature, among other examples, Zhang and Dette [40] discovered that other tests may yield smaller variance than the GLRT for independent samples. From Proposition 2, the asymptotic local power of the GLRT with level α

$$\beta_\alpha(c) = \Phi(R_1 - z_{1-\alpha}) \qquad \text{where } R_1 = \frac{c^{1/2}F_1 - c^{9/2}\mu_2^2F_2/4}{\sigma}, \tag{16}$$

$\Phi(\cdot)$ and $z_{1-\alpha}$ denote the cumulative distribution function and the $1-\alpha$ quantile of the standard normal distribution. Assume that $F_1 \neq 0$ and $F_2 \neq 0$, then simple calculations show that the bandwidth which maximizes the above power is

$$\tilde{b}_n = \tilde{c}n^{-2/9} \qquad \text{where } \tilde{c} = \left(\frac{4F_1}{9\mu_2^2F_2}\right)^{1/4}.$$

Remark 1. A typical example which satisfies $F_1 \neq 0$ and $F_2 \neq 0$ is when $\mathbf{f}_n(t) = a_n\mathbf{f}(a_n^2(t-t_0))$, where $\mathbf{f} \in \mathcal{C}^2[-1,1]$, $t_0 \in (0,1)$ and $a_n = n^{1/9}$. Simple calculations show that

$$F_1 = \int_{-1}^1 \mathbf{f}^\top(t)M(t_0)\mathbf{f}(t)\,\mathrm{d}t, \qquad F_2 = \int_{-1}^1 [\mathbf{f}''(t)]^\top M(t_0)\mathbf{f}''(t)\,\mathrm{d}t. \tag{17}$$

Hence $F_1 \neq 0$ and $F_2 \neq 0$ as long as the corresponding terms in (17) are nonzero.

4. Tests for locally stationary time series

4.1. The test

Consider the testing problem (2). Two important observations lead to the following modifications of the original GLRT when testing for non-stationary time series. First, as shown in Examples 2–4, the denominator RSS_0/n is redundant when testing for non-stationary time series. Second, as we discussed in the Introduction, averaging the test over a range of bandwidths can reduce the sensitivity of the test with respect to the selection of bandwidth and may also gain power over tests based on a single (optimal) bandwidth. Based on the above discussions, we suggest using the following averaged test when specifying model (1) for non-stationary time series:

$$\lambda_n^* = \int_{c_{\min}}^{c_{\max}} (\mathrm{RSS}_0 - \mathrm{RSS}_a(zn^{-\gamma}))\,\mathrm{d}z, \tag{18}$$

where $\mathrm{RSS}_a(b)$ is the RSS under H_a when bandwidth is chosen as b, $0 < c_{\min} < c_{\max} < \infty$. Large λ_n^* indicates evidence against H_0. In the literature, nonparametric ANOVA tests ignoring the denominator were first proposed in Dette [6] for independent samples. Dette and Hetzler [7] also considered averaged nonparametric specification tests over a range of bandwidths. The following theorem derives the asymptotic null distribution of the averaged test.

Theorem 3. *Assume that condition* (A) *holds and that* $2/9 \le \gamma < 1/4$. *Then under* H_0, *we have*

$$\begin{aligned}\sqrt{n^{-\gamma}}\bigg\{&\lambda_n^* + n^{\gamma}\tilde{K}(0)[\log(c_{\max}) - \log(c_{\min})]\int_0^1 \mathrm{tr}[H(t)]\,\mathrm{d}t \\ &+ \frac{n^{1-4\gamma}\mu_2^2(c_{\max}^5 - c_{\min}^5)}{20}\int_0^1 [\boldsymbol{\beta}''(t)]^{\top} M(t)\boldsymbol{\beta}''(t)\,\mathrm{d}t\bigg\} \Rightarrow N(0, (\sigma^*)^2),\end{aligned}$$

where

$$(\sigma^*)^2 = \int_{\mathbb{R}} Q^2(c_{\max}, t)\,\mathrm{d}t \int_0^1 \mathrm{tr}[H(t)^2]\,\mathrm{d}t \quad \textit{and}$$
$$Q(x,y) = \int_{c_{\min}}^{x} [2K(y/z) - K * K(y/z)]/z\,\mathrm{d}z.$$

Now we consider the local power of λ_n^* under the alternative $H_{a,n}$ specified in Proposition 2. By Theorem 3 and similar arguments as those of Proposition 2, it is easy to show that the asymptotic local power of λ_n^* with level α

$$\begin{aligned}&\beta_\alpha^*(c_{\min}, c_{\max}) = \Phi(R_2 - z_{1-\alpha}) \\ &\text{where } R_2 = \frac{(c_{\max} - c_{\min})F_1 - (c_{\max}^5 - c_{\min}^5)\mu_2^2 F_2/20}{\sigma^*}.\end{aligned} \tag{19}$$

Suppose that λ_n is asymptotically unbiased; namely $R_1 > 0$. From (19) and (16), we observe that λ_n^* is asymptotically more powerful than λ_n if and only if $R_2/R_1 > 1$. Simple calculations show that

$$R_2/R_1 = \frac{[(c_{\max} - c_{\min})F_1 - (c_{\max}^5 - c_{\min}^5)\mu_2^2 F_2/20]\sqrt{\int_{\mathbb{R}} \tilde{K}^2(t)\,\mathrm{d}t}}{[c^{1/2}F_1 - c^{9/2}\mu_2^2 F_2/4]\sqrt{\int_{\mathbb{R}} Q^2(c_{\max}, t)\,\mathrm{d}t}}.$$

An interesting observation from the above equation is that R_2/R_1 does not depend on the dependence or the non-stationarity structure of the data. Furthermore, we have the following result.

Proposition 3. *Under $H_{a,n}$ and the assumptions of Proposition 2, we have*

$$\sup_{0<c_{\min}<c_{\max}<\infty} \beta^*_\alpha(c_{\min}, c_{\max}) \geq \sup_{0<c<\infty} \beta_\alpha(c). \tag{20}$$

Proposition 3 claims that, asymptotically, the averaged test λ^*_n is at least as powerful as the test which is based on the maximum generalized likelihood ratio. The result is very general in the sense that it does not depend on the nature of the local alternative $\mathbf{f}_n(\cdot)$, the dependence structure of the data or the kernel function. When we restrict ourselves to a specific kernel function, the power comparison can be more exact. Let us consider the following example:

Example 5. Suppose that λ_n is asymptotically unbiased and that the bandwidth for λ_n is chosen as $cn^{-2/9}$. Let $c_{\min} = \tilde{c}_{\min} c$ for some fixed $\tilde{c}_{\min} \leq 1$ and let $c_{\max} = \tilde{c}_{\max} c$ such that $\tilde{c}_{\max}$ solves the equation $x^4 + \tilde{c}_{\min} x^3 + (\tilde{c}_{\min})^2 x^2 + (\tilde{c}_{\min})^3 x + (\tilde{c}_{\min})^4 = 5$. Choosing $c_{\max}$ in the latter way insures that F_1 and F_2 do not enter the ratio R_2/R_1 and hence the power comparison is relatively simple. Now simple calculations show that

$$R_2/R_1 = \frac{(\tilde{c}_{\max} - \tilde{c}_{\min})\sqrt{\int_{\mathbb{R}} \tilde{K}^2(t)\,\mathrm{d}t}}{\sqrt{\int_{\mathbb{R}} (\int_{\tilde{c}_{\min}}^{\tilde{c}_{\max}} [2K(y/z) - K * K(y/z)]/z\,\mathrm{d}z)^2\,\mathrm{d}y}}. \tag{21}$$

An application of the Cauchy–Schwarz inequality similar to the proof of Proposition 3 shows that $\sup_{0<\tilde{c}_{\min}\leq 1} R_2/R_1 \geq 1$ regardless of the kernel function. Now let us consider the uniform kernel $K(x) = I\{|x| \leq 1\}/2$. Figure 1 shows R_2/R_1 as a function of $\tilde{c}_{\min}$. We

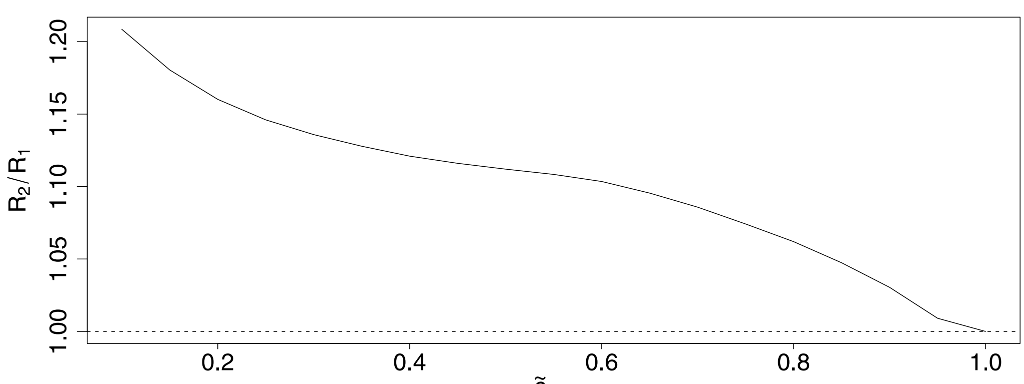

Figure 1. Ratio R_2/R_1 as a function of $\tilde{c}_{\min}$ in Example 5. The uniform kernel is used.

observe from the figure that the averaged test λ_n^* is asymptotically more powerful than λ_n on $(0,1)$ regardless of the shape of the alternative. Figure 1 further supports the use of the averaged test.

4.2. Bias reduction and bandwidth range selection

As we see from Theorem 3, the asymptotic bias of λ_n^* involves the second derivative of $\boldsymbol{\beta}(t)$ and the estimation of the latter quantity is generally highly nontrivial. Following the idea of Fan and Jiang [14], a prewhitening technique can be used to alleviate the problem. More specifically, consider the following null hypothesis:

$$\tilde{H}_0\colon\ \boldsymbol{\beta}(\cdot)=\boldsymbol{\beta}_0(\cdot,\theta) \qquad \text{for some unknown } \theta_0\in\Omega\subset\mathbb{R}^q,$$

where $\{\boldsymbol{\beta}_0(\cdot,\theta)\colon\ \theta\in\Omega\}$ is a parametric family of smooth functions. Let $\hat{\theta}_0$ be a $\sqrt{n}$ consistent estimator of θ_0 and define $\boldsymbol{\beta}^*(t)=\boldsymbol{\beta}(\cdot)-\boldsymbol{\beta}_0(t,\hat{\theta}_0)$. Then by the similar arguments as those of Proposition 1, the asymptotic bias and variance of estimating θ_0 is negligible in the current setting and hence testing $\tilde{H}_0$ is equivalent to testing

$$\breve{H}_0\colon\ \boldsymbol{\beta}^*(\cdot)=0 \quad \text{versus} \quad \breve{H}_a\colon\ \boldsymbol{\beta}^*(\cdot)\neq 0.$$

Then we can perform λ_n^* to testing $\breve{H}_0$ with transformed regression coefficients $\boldsymbol{\beta}^*(\cdot)$ and response $\breve{y}_i=y_i-\mathbf{x}_i^\top\boldsymbol{\beta}_0(t,\hat{\theta}_0)$. Note that the local linear estimator of $\boldsymbol{\beta}^*(\cdot)$ has no bias under $\breve{H}_0$ and we can avoid the notorious problem of bias estimation .

As mentioned in Fan and Jiang [14], a choice of larger bandwidth favors smoother alternatives and a smaller bandwidth tends to detect less smooth alternatives. Thanks to the introduction of the averaged test, the sensitivity of the test to the choice of bandwidth is alleviated due to the introduction of a group of bandwidths. On the other hand, the correlation of λ_n between nearby bandwidths are usually quite high and hence in practice one only needs to average the test over a grid of relatively separated bandwidths. Zhang [41] found that the correlation between $\lambda_n(h)$ and $\lambda_n(ch)$ is quite high for $c=1.3$. As suggested by Fan and Jiang [14], here we recommend choosing the grid of three bandwidths $\tilde{b}_n/1.5$, $\tilde{b}_n$ and $\tilde{b}_n\times 1.5$ to represent small, medium and large bandwidths and average the test over the latter grid. Here $\tilde{b}_n=b_n^*\times n^{-1/45}$ and b_n^* is the optimal bandwidth for nonparametric curve estimation.

4.3. The robust wild bootstrap

A direct implementation of the asymptotic distribution in Theorem 3 may not perform satisfactorily in practice due to the following two reasons. First, the convergence rate of test statistic equals $\mathrm{O}(n^{-1/9})$ when bandwidth b_n is chosen optimally. The rate is quite slow and hence the asymptotic approximation may not be accurate for moderate samples. Second, as we can see from the proof of Lemma 7 in Section 6, the asymptotic normal approximation is particularly rough at the boundaries of the time interval for finite samples. As a remedy, we observe the following proposition.

Proposition 4. *Let the bandwidth range be* $[c_{\min} n^{-\gamma}, c_{\max} n^{-\gamma}]$ *for some* $0 < c_{\min} < c_{\max} < \infty$. *Suppose that either (1):* $\boldsymbol{\beta}_0(\cdot)$ *is a linear function or (2):* $\gamma > 2/9$. *Then under* H_0, *condition* (A) *and the assumption that* $\gamma < 1/4$, *on a possibly richer probability space, there exist i.i.d. p-dimensional standard Gaussian random vectors* $V_1, \ldots, V_n$, *such that*

$$\lambda_n^* = \Phi_n + o_{\mathbb{P}}(\sqrt{n^\gamma}), \tag{22}$$

where

$$\Phi_n = \int_{c_{\min}}^{c_{\max}} \left\{ 2 \sum_{i=1}^n \tilde{V}_i^\top [\mathbb{E}\mathbf{S}_{n,n(s)}(t_i)]^{-1} \tilde{\mathbf{T}}_{n,n(s)}(t_i) - \sum_{i=1}^n [\mathbf{z}_i^\top [\mathbb{E}\mathbf{S}_{n,n(s)}(t_i)]^{-1} \tilde{\mathbf{T}}_{n,n(s)}(t_i)]^2 \right\} \mathrm{d}s$$

with $n(s) = sn^{-\gamma}$, $\mathbf{z}_i = (\mathbf{x}_i^\top, \mathbf{0}_p^\top)^\top$, $\tilde{V}_i = (V_i^\top \Lambda^{1/2}(t_i), \mathbf{0}_p^\top)^\top$, $\tilde{\mathbf{T}}_{n,b}(t) = (\tilde{\mathbf{T}}_{n,0,b}^\top(t), \tilde{\mathbf{T}}_{n,1,b}^\top(t))^\top$ *and*

$$\tilde{\mathbf{T}}_{n,l,b}(t) = (nb)^{-1} \sum_{i=1}^n \Lambda^{1/2}(t_i) V_i [(t_i - t)/b]^l K_b(t_i - t), \qquad l = 0, 1. \tag{23}$$

Proposition 4 follows easily from (30) and Lemma 5 in Section 6. Details are omitted. The latter proposition claims that λ_n^* can be well approximated by a Gaussian quadratic form Φ_n. In particular, we observe from the proofs in Section 6 that the approximation is accurate at the boundaries due to the fact that it directly mimics the form of the test statistic. When implementing λ_n^*, we recommend generating a large (say of size 1000) sample of i.i.d. copies of Φ_n and use the resulting empirical distribution to approximate that of λ_n^* under the null hypothesis and obtain the p-value of the test.

As we suggested in Section 4.2, in practice, one usually uses a grid of bandwidths $\mathcal{B} = \{c_{\min} n^{-\gamma} = b_1 < b_2 < \cdots < b_M = c_{\max} n^{-\gamma}\}$ and calculate $\lambda_n^*(\mathcal{B}) = \sum_{i=1}^M (\mathrm{RSS}_0 - \mathrm{RSS}_a(b_i))$. To perform wild bootstrap in those cases, one compares $\lambda_n^*(\mathcal{B})$ to the simulated quantiles of

$$\Phi_n(\mathcal{B}) := \sum_{j=1}^M \left\{ 2 \sum_{i=1}^n \tilde{V}_i^\top [\mathbb{E}\mathbf{S}_{n,b_j}(t_i)]^{-1} \tilde{\mathbf{T}}_{n,b_j}(t_i) - \sum_{i=1}^n [\mathbf{z}_i^\top [\mathbb{E}\mathbf{S}_{n,b_j}(t_i)]^{-1} \tilde{\mathbf{T}}_{n,b_j}(t_i)]^2 \right\}$$

to calculate the p-value of the test. In Section 5, we shall conduct a simulation study to compare the finite sample performance of the wild bootstrap and the direct implementation of the asymptotic distribution.

If one is interested in the semiparametric testing problem H_{01} versus H_{a1} in (13), then the corresponding averaged test is

$$\lambda_{1n}^* = \int_{c_{\min}}^{c_{\max}} (\mathrm{RSS}_1(zn^{-\gamma}) - \mathrm{RSS}_a(zn^{-\gamma}))\, \mathrm{d}z. \tag{24}$$

Write $\varepsilon_i = ([\varepsilon_i^{(1)}]^\top, [\varepsilon_i^{(2)}]^\top)^\top$ and $V_i = ([V_i^{(1)}]^\top, [V_i^{(2)}]^\top)^\top$, where $\varepsilon_i^{(1)}$ and $V_i^{(1)}$ are p_1 dimensional. Define $\mathbf{S}_{n,b}^{(2)}$, $\mathbf{S}_{n,l,b}^{(2)}$, $\mathbf{z}_i^{(2)}$, $\tilde{V}_i^{(2)}$, $\mathbf{T}_n^{(2)} \mathbf{T}_{nl}^{(2)}$, $\tilde{\mathbf{T}}_n^{(2)}$, $\tilde{\mathbf{T}}_{nl}^{(2)}$ and $\Phi_n^{(2)}$ in the same way as their counterparts without the superscript $^{(2)}$ with $\mathbf{x}_i$, ε_i, $\Lambda(t)$ and V_i therein replaced by $\mathbf{x}_i^{(2)}$, $\varepsilon_i^{(2)}$, $\Lambda_{22}(t)$ and $V_i^{(2)}$, respectively. We have the following proposition.

Proposition 5. *Suppose that* $1/4 > \gamma > 2/9$. *Then under* H_{01} *and condition* (A), *on a possibly richer probability space, there exist i.i.d. p-dimensional standard Gaussian random vectors* $V_1, \dots, V_n$, *such that*

$$\lambda_{1n}^* = \Phi_n - \Phi_n^{(2)} + \mathrm{o}_{\mathbb{P}}(\sqrt{n^\gamma}). \tag{25}$$

Note that $\Phi_n - \Phi_n^{(2)}$ is a quadratic form of $V_1, \dots, V_n$. By Proposition 5, in practice, one could generate a large sample of i.i.d. copies of $\Phi_n - \Phi_n^{(2)}$ to obtain the p-value of testing H_{01}.

4.4. Long-run covariance matrix estimation

By Lemma 9 in Section 6, $\mathbb{E}\mathbf{S}_{n,n(s)}(t_i)$ in Proposition 4 can be well approximated by $\mathbf{S}_{n,n(s)}(t_i)$. Therefore, in order to implement the wild bootstrap, one only needs to estimate the long-run covariance matrix $\Lambda(\cdot)$. Here we suggest using the local lag window estimate of $\Lambda(\cdot)$ proposed in Zhou and Wu [43]. For the sake of completeness, we will briefly introduce the estimator here. We refer to the latter paper for more details including the derivation of convergence rates of the estimator and the choice of smoothing parameters.

Define $\hat{\mathbf{L}}_i := \mathbf{x}_i \hat{\varepsilon}_i$, where $\hat{\varepsilon}_i$'s are the residuals under the alternative. For a window size m and a bandwidth τ_n, $\Lambda(\cdot)$ can be estimated by

$$\hat{\Lambda}(\cdot) = \sum_{i=1}^n \omega(\cdot, i)\Delta_i \qquad \text{where } \omega(\cdot, i) = \frac{K_{\tau_n}(t_i - \cdot)}{\sum_{j=1}^n K_{\tau_n}(t_j - \cdot)}$$

and $\Delta_i = (\sum_{j=-m}^m \hat{\mathbf{L}}_{i+j})(\sum_{j=-m}^m \hat{\mathbf{L}}_{i+j}^\top)/(2m+1)$. Zhou and Wu [43] showed that $\hat{\Lambda}(t)$ is always positive semidefinite and has convergence rate $\mathrm{O}(n^{-2/7})$ when $m = \mathrm{O}(n^{2/7})$ and $\tau_n = \mathrm{O}(n^{-1/7})$.

5. Simulation studies

In this section, we shall design simulations to study the accuracy of the wild bootstrap procedure of the paper and compare it with that of the bootstrap procedure of Fan and Jiang [14] and the method of direct implementation of the asymptotic distribution in (9). Let us consider the following model

$$y_i = \beta_1(t_i) + \beta_2(t_i)x_{2i} + \varepsilon_i \tag{26}$$

and the test H_0: $\beta_1(\cdot) = \beta_2(\cdot) = 0$. The following four scenarios are considered in order to investigate the effects of endogeneity, non-stationarity and temporal dependence.

Scenario (a). In this case x_{2i}'s are i.i.d. exponential random variables with mean 1 and ε_i's are i.i.d. standard normal. The two processes $\{x_{2i}\}$ and $\{\varepsilon_i\}$ are independent. The latter design satisfies the conditions in Fan, Zhang and Zhang [15] and hence it is expected that the bootstrap procedure in Fan and Jiang [14] will work in this case.

Scenario (b). In this scenario x_{2i}'s are i.i.d. exponential random variables with mean 1 and $\varepsilon_i = x_{2i}\zeta_i$, where ζ_i's are i.i.d. standard normal and are independent of $\{x_{2i}\}$. In scenario (b) we are interested in investigating the effect of endogeneity on the behavior of GLRT.

Scenario (c). Let x_{2i}'s be independent student t random variables and the degrees of freedom of $x_{2i} = 5 + 10t_i$. Let $\varepsilon_i = \exp(-1/t_i)/(100t_i^4)\zeta_i$, where ζ_i's are i.i.d. standard normal. Further let x_{2i}'s and ε_i's be independent. Note that $\{\varepsilon_i\}$ is a locally stationary process with time-varying variance and $\{x_{2i}\}$ is locally stationary process with smoothly varying tail index. In this case, we are investigating the effect of non-stationarity on the behavior of GLRT.

Scenario (d). Let $x_{2i} = \epsilon_i\epsilon_{i-1}$, where ϵ_i's are i.i.d. standard normal. Let $\varepsilon_i = 0.5\varepsilon_{i-1} + \zeta_i$, where ζ_i's are i.i.d. standard normal. Further let $\{\epsilon_i\}$ be independent of $\{\zeta_i\}$. Note $\{x_{2i}\}$ and $\{\varepsilon_i\}$ are two stationary weakly dependent processes. In this case we are interested in investigating the effect of temporal dependence on the behavior of GLRT.

We consider two different sample sizes, $n = 200$ and 400. We compare three different methods, namely the robust wild bootstrap test (22) (WILD), test based on the asymptotic distribution (9) (ASYM) and the residual bootstrap test of Fan and Jiang [14] (IID). Both the single bandwidth test λ_n in (3) and the suggested averaged test λ_n^* in (18) are considered. For the averaged test, the bandwidth ranges are selected as $[\tilde{b}_n/1.5, 1.5\tilde{b}_n]$ according to the discussion in Section 4.2. To investigate the sensitivity of the accuracy of the wild bootstrap method on the choice of bandwidth, three different bandwidths, namely $0.15, 0.25$ and 0.35 are considered in the simulation. Based on 500 replications, the simulated type I error rates at 10% nominal level are summarized in Table 1 below.

We observe from Table 1 that, for the robust wild bootstrap, the simulated type I errors of the averaged test and the single bandwidth test are reasonably close to the nominal and the performance is stable for all four cases when $n = 400$. For $n = 200$, the robust bootstrap is slightly anti-conservative in cases (a), (b) and (d) for small bandwidths. As we expected, the averaged test performs more stably than the single bandwidth test. On the other hand, we observe that tests based on the asymptotic distribution do not perform well for moderately large samples. As we discussed in Section 4.3, the reason is due to the slow convergence of the test statistic and the rough approximation of the asymptotic distribution at the boundaries. The residual wild bootstrap performs slightly better than our robust wild bootstrap for i.i.d. data without endogeneity. However, we observe that the residual bootstrap is no longer consistent under non-stationarity, temporal dependence or endogeneity, which is consistent with our theoretical findings.

Table 1. Simulated type I error rates (in percentage) for the wild bootstrap test (22) (WILD), test based on the asymptotic distribution (9) (ASYM) and the bootstrap test of Fan and Jiang [14] (IID) with nominal level 10% under scenarios (a), (b), (c) and (d). For the averaged test λ_n^*, the bandwidth range is selected as $[\tilde{b}_n/1.5, 1.5\tilde{b}_n]$. Series length $n = 200$ and 400 with 500 replicates

		$n = 200$				$n = 400$			
Method		(a)	(b)	(c)	(d)	(a)	(b)	(c)	(d)
Averaged test λ_n^*									
WILD	$\tilde{b}_n = 0.15$	7.5	7.4	10.4	7.1	8.1	8	9.7	9.1
WILD	$\tilde{b}_n = 0.25$	8.5	8.15	10.2	7.7	8.5	8.4	9.8	9.7
WILD	$\tilde{b}_n = 0.35$	8.9	8.7	10	7.7	8.7	9.1	9.2	9.8
ASYM	$\tilde{b}_n = 0.15$	35.4	14.4	18.8	28.2	38.3	18.5	15.0	33
ASYM	$\tilde{b}_n = 0.25$	39.1	18.5	19.4	33.3	39.9	21.2	17.8	36.3
ASYM	$\tilde{b}_n = 0.35$	44.1	21.4	18.0	36.2	44.5	23.8	20.7	38.4
IID	$\tilde{b}_n = 0.15$	10.4	83.6	20.5	68.8	11.9	87.7	15.7	73.3
IID	$\tilde{b}_n = 0.25$	11.4	79.6	19.1	61.9	9.9	82.7	17.9	63.8
IID	$\tilde{b}_n = 0.35$	11.0	74.3	17.8	55.9	10.2	78.8	19.8	56.8
Single bandwidth test λ_n									
WILD	$b_n = 0.15$	5.0	5.8	10.2	5.8	8.6	7.2	11.2	9.4
WILD	$b_n = 0.25$	8.2	7.8	9.4	8.8	9.2	8.2	10.2	11.6
WILD	$b_n = 0.35$	9.8	9.2	9.0	8.2	11.2	9.6	11.2	11.4
ASYM	$b_n = 0.15$	32.2	13.2	17.8	27.8	27.4	16.8	13.8	30
ASYM	$b_n = 0.25$	36.2	19.6	20.4	36.8	29	20.2	16.8	36.6
ASYM	$b_n = 0.35$	43.6	21.2	20.4	38.8	34	22	18	38
IID	$b_n = 0.15$	8.2	86.8	20.8	73.2	10.8	89	15.2	76.2
IID	$b_n = 0.25$	7.8	82.2	20.6	63	9.4	80.2	18	63.4
IID	$b_n = 0.35$	10.4	76.2	17.4	55.6	12	77.2	17.8	56.6

6. Proofs

Note that under the null hypothesis H_0,

$$\mathrm{RSS}_a - \mathrm{RSS}_0 = 2\sum_{i=1}^{n} \mathbf{x}_i^\top \varepsilon_i(\boldsymbol{\beta}(t_i) - \hat{\boldsymbol{\beta}}(t_i)) + \sum_{i=1}^{n} \{\mathbf{x}_i^\top(\boldsymbol{\beta}(t_i) - \hat{\boldsymbol{\beta}}(t_i))\}^2 := 2I_n + II_n. \quad (27)$$

On the other hand, by (5),

$$\mathbf{S}_n(t)(\hat{\boldsymbol{\eta}}(t) - \boldsymbol{\eta}(t)) = \begin{pmatrix} b_n^2 \mathbf{S}_{n,2}(t)(\boldsymbol{\beta}''(t) + \mathrm{o}(1))/2 \\ b_n^2 \mathbf{S}_{n,3}(t)(\boldsymbol{\beta}''(t) + \mathrm{o}(1))/2 \end{pmatrix} + \begin{pmatrix} \mathbf{T}_{n,0}(t) \\ \mathbf{T}_{n,1}(t) \end{pmatrix} := \mathbf{B}_n(t) + \mathbf{T}_n(t), \quad (28)$$

where $\boldsymbol{\eta}(t) = (\boldsymbol{\beta}^\top(t), b_n \boldsymbol{\beta}'^\top(t))^\top$, and

$$\mathbf{T}_{n,l}(t) = r_n^2 \sum_{i=1}^{n} \mathbf{x}_i \varepsilon_i [(t_i - t)/b_n]^l K_{b_n}(t_i - t), \qquad l = 0, 1, \ldots$$

with $r_n := 1/\sqrt{nb_n}$. In (28), $\mathbf{B}_n(t)$ corresponds to the bias of the local linear estimate at time t. Lemmas 1 and 2 below control the asymptotic influence of the bias term $\mathbf{B}_n(\cdot)$ on $\mathrm{RSS}_a - \mathrm{RSS}_0$.

Lemma 1. *Define* $\mathbf{z}_i = (\mathbf{x}_i^\top, \mathbf{0}_p^\top)^\top$*, where* $\mathbf{0}_p$ *is the column vector of* p *zeros. Under condition* (A)*, we have* $-I_n = D_{n1} + \mathrm{O}_{\mathbb{P}}(\sqrt{n}b_n^2)$*, where* $D_{n1} := \sum_{i=1}^n \mathbf{z}_i^\top \varepsilon_i \mathbf{S}_n^{-1}(t_i)\mathbf{T}_n(t_i)$*.*

Proof. By (27) and (28), we have

$$-I_n - D_{n1} = \sum_{i=1}^n \mathbf{z}_i^\top \varepsilon_i \mathbf{S}_n^{-1}(t_i)\mathbf{B}_n(t_i).$$

Define $ID_{n1} = \mathbb{E}[(-I_n - D_{n1})^2 | \mathcal{F}_n]$ and $\mathcal{P}_i(\cdot) = \mathbb{E}(\cdot|\mathcal{G}_i) - \mathbb{E}(\cdot|\mathcal{G}_{i-1})$. Recall that $\mathcal{G}_i = (\ldots, \xi_{i-1}, \xi_i)$. Using the facts that $H(t, \mathcal{G}_i) = \sum_{j=-\infty}^i \mathcal{P}_j H(t, \mathcal{G}_i)$ and $\mathcal{P}_i$ and $\mathcal{P}_j$ are orthogonal for $i \neq j$, elementary calculations show that

$$\begin{aligned} ID_{n1} = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=-\infty}^n & \mathbb{E}[\mathcal{P}_k H(t_i, \mathcal{G}_i) \mathcal{P}_k H(t_j, \mathcal{G}_j)] \\ & \times V(t_i, \mathcal{F}_i) \mathbf{S}_n^{-1}(t_i) \mathbf{B}_n(t_i) V(t_j, \mathcal{F}_j) \mathbf{S}_n^{-1}(t_j) \mathbf{B}_n(t_j). \end{aligned}$$

Let $\delta_H(k, p) = 0$ if $k < 0$. Note that

$$\begin{aligned} \sum_{k=-\infty}^n |\mathbb{E}[\mathcal{P}_k H(t_i, \mathcal{G}_i) \mathcal{P}_k H(t_j, \mathcal{G}_j)]| &\le \sum_{k=-\infty}^n \|\mathcal{P}_k H(t_i, \mathcal{G}_i)\| \|\mathcal{P}_k H(t_j, \mathcal{G}_j)\| \\ &\le \sum_{k=-\infty}^n \delta_H(i-k, 2)\delta_H(j-k, 2) \\ &\le C(|i-j|+1)^{-2}. \end{aligned}$$

On the other hand, by Lemma 9, the Hölder's inequality and similar arguments as those of Lemma 6 in Zhou and Wu [43], we have

$$\begin{aligned} &\mathbb{E}|V(t_i, \mathcal{F}_i)\mathbf{S}_n^{-1}(t_i)\mathbf{B}_n(t_i)V(t_j, \mathcal{F}_j)\mathbf{S}_n^{-1}(t_j)\mathbf{B}_n(t_j)| \\ &\quad \le \|V(t_i, \mathcal{F}_i)\|_4 \|\mathbf{S}_n^{-1}(t_i)\|_8 \|\mathbf{B}_n(t_i)\|_8 \|V(t_j, \mathcal{F}_j)\|_4 \|\mathbf{S}_n^{-1}(t_j)\|_8 \|\mathbf{B}_n(t_j)\|_8 \le Cb_n^4. \end{aligned}$$

Therefore, $\mathbb{E}ID_{n1} \le C\sum_{i=1}^n \sum_{j=1}^n (|i-j|+1)^{-2} b_n^4 \le Cnb_n^4$. Note that $\mathbb{E}(-I_n - D_{n1})^2 = \mathbb{E}ID_{n1}$. Therefore, this lemma follows. □

Lemma 2. *Under condition* (A) *and the assumption that* $nb_n^{5/2} \to \infty$*, we have*

$$II_n = D_{n2} + \frac{nb_n^4 \mu_2^2}{4} \int_0^1 [\boldsymbol{\beta}''(t)]^\top M(t) \boldsymbol{\beta}''(t)\,\mathrm{d}t + \mathrm{o}_{\mathbb{P}}(nb_n^4),$$

where $D_{n2} := \sum_{i=1}^n \{\mathbf{z}_i^\top \mathbf{S}_n^{-1}(t_i)\mathbf{T}_n(t_i)\}^2$.

Proof. By (27) and (28), we have

$$\begin{aligned} II_n - D_{n2} &= \sum_{i=1}^{n} (\mathbf{z}_i^\top \mathbf{S}_n^{-1}(t_i)\mathbf{B}_n(t_i))^2 + 2\sum_{i=1}^{n} \mathbf{z}_i^\top \mathbf{S}_n^{-1}(t_i)\mathbf{B}_n(t_i)\mathbf{z}_i^\top \mathbf{S}_n^{-1}(t_i)\mathbf{T}_n(t_i) \\ &:= ID_{n2}^{*} + 2ID_{n2}^{**}. \end{aligned}$$

By Lemma 9 and the Hölder's inequality, it follows that

$$ID_{n2}^{*} - \sum_{i=1}^{n} \{\mathbf{z}_i^\top [\mathbb{E}\mathbf{S}_n(t_i)]^{-1}\mathbf{B}_n(t_i)\}^2 = \mathrm{O}_{\mathbb{P}}(nb_n^4/\sqrt{nb_n}).$$

By condition (A4) and the similar arguments as those in the proof of Lemma 1, we have

$$\sum_{i=1}^{n} \{\mathbf{z}_i^\top [\mathbb{E}\mathbf{S}_n(t_i)]^{-1}\mathbf{B}_n(t_i)\}^2 - \mathbb{E}\left[\sum_{i=1}^{n} \{\mathbf{z}_i^\top [\mathbb{E}\mathbf{S}_n(t_i)]^{-1}\mathbf{B}_n(t_i)\}^2\right] = \mathrm{O}_{\mathbb{P}}(\sqrt{n}b_n^4).$$

It is easy to see that, for $i = 1, 2, \ldots, n$,

$$\mathbb{E}(\mathbf{z}_i^\top [\mathbb{E}\mathbf{S}_n(t_i)]^{-1}\mathbf{B}_n(t_i))^2 - b_n^4 \mathbb{E}\left(\mathbf{z}_i^\top [\mathbb{E}\mathbf{S}_n(t_i)]^{-1} \begin{pmatrix} \mathbf{S}_{n,2}(t_i)\boldsymbol{\beta}''(t_i)/2 \\ \mathbf{S}_{n,3}(t_i)\boldsymbol{\beta}''(t_i)/2 \end{pmatrix}\right)^2 = \mathrm{o}(b_n^4).$$

Additionally, by Lemma 9 and simple algebra, we have

$$\sum_{i=1}^{n} \mathbb{E}\left(\mathbf{z}_i^\top [\mathbb{E}\mathbf{S}_n(t_i)]^{-1} \begin{pmatrix} \mathbf{S}_{n,2}(t_i)\boldsymbol{\beta}''(t_i)/2 \\ \mathbf{S}_{n,3}(t_i)\boldsymbol{\beta}''(t_i)/2 \end{pmatrix}\right)^2 = n\mu_2^2 \int_0^1 [\boldsymbol{\beta}''(t)]^\top M(t)\boldsymbol{\beta}''(t)\,\mathrm{d}t/4 + \mathrm{o}(n).$$

Therefore, $ID_{n2}^{*} = nb_n^4\mu_2^2 \int_0^1 [\boldsymbol{\beta}''(t)]^\top M(t)\boldsymbol{\beta}''(t)\,\mathrm{d}t/4 + \mathrm{o}_p(nb_n^4)$. Furthermore,

$$ID_{n2}^{**} = r_n^2 \sum_{j=1}^{n}\sum_{i=1}^{n} \mathbf{z}_i^\top \mathbf{S}_n^{-1}(t_i)\mathbf{B}_n(t_i)\mathbf{z}_i^\top \mathbf{S}_n^{-1}(t_i)\mathbf{x}_j K_{b_n}(t_i - t_j)\varepsilon_j.$$

Recall that $r_n = 1/\sqrt{nb_n}$. Following the similar arguments as those in the proof of Lemma 1, we have $ID_{n2}^{**} = \mathrm{O}_{\mathbb{P}}(\sqrt{b_n^3}) = \mathrm{o}_{\mathbb{P}}(nb_n^4)$. Details are omitted. Hence, the lemma follows. □

Lemma 3. *Under condition* (A) *and the assumption that $nb_n^3 \to \infty$, we have*

$$D_{n1} = \bar{D}_{n1} + \mathrm{o}_{\mathbb{P}}(1/\sqrt{b_n}),$$

where $\bar{D}_{n1} = \sum_{i=1}^{n} \mathbf{z}_i^\top \varepsilon_i [\mathbb{E}\mathbf{S}_n(t_i)]^{-1}\mathbf{T}_n(t_i)$.

Proof. Let $\overline{ID}_{n1} = D_{n1} - \bar{D}_{n1}$ and $\mathbf{IS}_n(t) = \mathbf{S}_n^{-1}(t) - [\mathbb{E}\mathbf{S}_n(t)]^{-1}$. Then

$$\overline{ID}_{n1} = \sum_{i=1}^{n} \mathbf{z}_i^\top \varepsilon_i \mathbf{IS}_n(t_i)\mathbf{T}_n(t_i).$$

Let $\mathbf{A}_{n,k} = \sum_{i=1}^{k} \mathbf{z}_i^\top \varepsilon_i \mathbf{I}\mathbf{S}_n(t_i)$ and $\mathbf{A}_{n,0} = 0$. Then by Lemma 9 and the similar arguments as those of Lemma 1, it is easy to show that $\max_{1\le k\le n} \|\mathbf{A}_{n,k}\|_4 \le C r_n \sqrt{n}$. Note that

$$\overline{ID}_{n1} = \sum_{i=1}^{n} (\mathbf{A}_{n,i} - \mathbf{A}_{n,i-1}) \mathbf{T}_n(t_i) = \sum_{i=1}^{n-1} \mathbf{A}_{n,i} (\mathbf{T}_n(t_i) - \mathbf{T}_n(t_{i-1})) + \mathbf{A}_{n,n} \mathbf{T}_n(t_n).$$

By the similar arguments as those of Lemma 1, we have

$$\max_{1\le i\le n} \|\mathbf{T}_n(t_i) - \mathbf{T}_n(t_{i-1})\|_4 \le C r_n^3 \tag{29}$$

and $\|\mathbf{T}_n(t_n)\|_4 = \mathrm{O}(r_n)$. Therefore,

$$\begin{aligned} \|\overline{ID}_{n1}\| &\le \sum_{i=1}^{n-1} \|\mathbf{A}_{n,i}\|_4 \|\mathbf{T}_n(t_i) - \mathbf{T}_n(t_{i-1})\|_4 + \|\mathbf{A}_{n,n}\|_4 \|\mathbf{T}_n(t_n)\|_4 \\ &\le C\left(\sum_{i=1}^{n-1} r_n \sqrt{n} r_n^3 + r_n \sqrt{n} r_n\right) = \mathrm{O}(1/(\sqrt{n} b_n^2)) = \mathrm{o}(1/\sqrt{b_n}). \end{aligned}$$

Therefore, the lemma follows. □

Lemma 4. *Under condition* (A) *and the assumption that* $nb_n^3 \to \infty$, *we have*

$$D_{n2} = \bar{D}_{n2} + \mathrm{o}_{\mathbb{P}}(1/\sqrt{b_n}),$$

where $\bar{D}_{n2} = \sum_{i=1}^{n} \{\mathbf{z}_i^\top [\mathbb{E}\mathbf{S}_n(t_i)]^{-1} \mathbf{T}_n(t_i)\}^2$.

Proof. Note that $D_{n2} - \bar{D}_{n2} = \sum_{i=1}^{n} \Gamma_1(i)\Gamma_2(i)$, where $\Gamma_1(i) = \mathbf{z}_i^\top (\mathbf{S}_n^{-1}(t_i) + [\mathbb{E}\mathbf{S}_n(t_i)]^{-1}) \times \mathbf{T}_n(t_i)$ and $\Gamma_2(i) = \mathbf{z}_i^\top (\mathbf{S}_n^{-1}(t_i) - [\mathbb{E}\mathbf{S}_n(t_i)]^{-1}) \mathbf{T}_n(t_i)$.

Let $S\Gamma_1(i) = \sum_{j=1}^{i} \Gamma_1(i)$ for $1 \le i \le n$ and $S\Gamma_1(0) = 0$. Then

$$D_{n2} - \bar{D}_{n2} = \sum_{i=1}^{n} (S\Gamma_1(i) - S\Gamma_1(i))\Gamma_2(i) = \sum_{i=1}^{n-1} S\Gamma_1(i)(\Gamma_2(i) - \Gamma_2(i+1)) + S\Gamma_1(n)\Gamma_2(n).$$

Note that

$$\begin{aligned} S\Gamma_1(i) &= r_n^2 \sum_{k=1}^{n} \sum_{j=1}^{i} \mathbf{z}_j^\top (\mathbf{S}_n^{-1}(t_j) + [\mathbb{E}\mathbf{S}_n(t_j)]^{-1}) K_{b_n}(t_k - t_j) \begin{pmatrix} \mathbf{x}_k \varepsilon_k \\ \mathbf{x}_k \varepsilon_k [(t_k - t_j)/b_n] \end{pmatrix} \\ &= r_n^2 \sum_{k=1}^{n} \Xi_1(i,k)\varepsilon_k + r_n^2 \sum_{k=1}^{n} \Xi_2(i,k)\varepsilon_k, \end{aligned}$$

where

$$\Xi_1(i,k) = \sum_{j=1}^{i} \mathbf{z}_j^\top (\mathbf{S}_n^{-1}(t_j) + [\mathbb{E}\mathbf{S}_n(t_j)]^{-1}) K_{b_n}(t_k - t_j) \mathbf{z}_k^\top,$$

$$\Xi_2(i,k) = \sum_{j=1}^{i} \mathbf{z}_j^\top (\mathbf{S}_n^{-1}(t_j) + [\mathbb{E}\mathbf{S}_n(t_j)]^{-1}) K_{b_n}(t_k - t_j)(\mathbf{0}_p^\top, \mathbf{x}_k^\top)^\top.$$

By Lemma 9 and the Hölder's inequality, $\max_i \|\Xi_1(i,k)\| \le Cnb_n$. Hence by similar conditioning arguments as those in the proof Lemma 1,

$$r_n^2 \max_i \left\| \sum_{k=1}^{n} \Xi_1(i,k)\varepsilon_k \right\| = \mathrm{O}(\sqrt{n}).$$

Similarly, $r_n^2 \max_i \| \sum_{k=1}^{n} \Xi_2(i,k)\varepsilon_k\| = \mathrm{O}(\sqrt{n})$. Hence, $\max_i \|S\Gamma_1(i)\| = \mathrm{O}(\sqrt{n})$. By similar arguments, we have

$$\max_i \|\Gamma_2(i) - \Gamma_2(i+1)\| = \mathrm{O}(r_n^4) \quad \text{and} \quad \|\Gamma_2(n)\| = \mathrm{O}(r_n^2).$$

Therefore

$$\begin{aligned}\mathbb{E}|D_{n2} - \bar{D}_{n2}| &\le \sum_{i=1}^{n-1} \|S\Gamma_1(i)\| \|\Gamma_2(i) - \Gamma_2(i+1)\| + \|S\Gamma_1(n)\| \|S\Gamma_2(n)\| \\ &= \mathrm{O}(1/(\sqrt{n} b_n^2)) = \mathrm{o}(1/\sqrt{b_n}).\end{aligned}$$

The lemma follows. □

Lemma 5. *Under condition* (A) *and the assumption that* $nb_n^3 \to \infty$, *we have*

$$\bar{D}_{n2} = \Theta_n + \mathrm{o}_{\mathbb{P}}(1/\sqrt{b_n}),$$

where $\Theta_n = \sum_{i=1}^{n} \mathbf{T}_n^\top(t_i)[\mathbb{E}\mathbf{S}_n(t_i)]^{-1}\mathbb{E}[\mathbf{z}_i\mathbf{z}_i^\top][\mathbb{E}\mathbf{S}_n(t_i)]^{-1}\mathbf{T}_n(t_i)$.

Proof. Note that $\bar{D}_{n2} = \sum_{i=1}^{n} \mathbf{T}_n^\top(t_i)[\mathbb{E}\mathbf{S}_n(t_i)]^{-1}\mathbf{z}_i\mathbf{z}_i^\top[\mathbb{E}\mathbf{S}_n(t_i)]^{-1}\mathbf{T}_n(t_i)$. Therefore

$$\bar{D}_{n2} - \Theta_n = \sum_{i=1}^{n} \mathbf{T}_n^\top(t_i)\Theta_n(i),$$

where $\Theta_n(i) = [\mathbb{E}\mathbf{S}_n(t_i)]^{-1}\{\mathbf{z}_i\mathbf{z}_i^\top - \mathbb{E}[\mathbf{z}_i\mathbf{z}_i^\top]\}[\mathbb{E}\mathbf{S}_n(t_i)]^{-1}\mathbf{T}_n(t_i)$. Note that

$$\begin{aligned}\sum_{j=1}^{i} \Theta_n(j) = r_n^2 \sum_{k=1}^{n}\sum_{j=1}^{i} &[\mathbb{E}\mathbf{S}_n(t_j)]^{-1}\{\mathbf{z}_j\mathbf{z}_j^\top - \mathbb{E}[\mathbf{z}_j\mathbf{z}_j^\top]\}[\mathbb{E}\mathbf{S}_n(t_j)]^{-1} \\ &\times K_{b_n}(t_k - t_j)\begin{pmatrix} \mathbf{x}_k\varepsilon_k \\ \mathbf{x}_k\varepsilon_k[(t_k - t_j)/b_n] \end{pmatrix}.\end{aligned}$$

By the short memory property of $\mathbf{x}_i$ in condition (A4) and similar arguments as those in the proof of Lemma 1, we have

$$\max_i \left\| \sum_{j=1}^{i} [\mathbb{E}\mathbf{S}_n(t_j)]^{-1} \{\mathbf{z}_j\mathbf{z}_j^\top - \mathbb{E}[\mathbf{z}_j\mathbf{z}_j^\top]\} [\mathbb{E}\mathbf{S}_n(t_j)]^{-1} K_{b_n}(t_k - t_j) \right\| = \mathrm{O}(\sqrt{nb_n}).$$

Hence by similar conditioning arguments as those in the proof of Lemma 1, we have

$$\max_i \left\| \sum_{j=1}^{i} \Theta_n(j) \right\| = \mathrm{O}(\sqrt{n} r_n).$$

Together with (29) and the summation by parts technique used in Lemma 3, it follows that $\mathbb{E}|\bar{D}_{n2} - \Theta_n| = \mathrm{O}(1/(\sqrt{n} b_n^2)) = \mathrm{o}(1/\sqrt{b_n})$. The lemma follows. □

Lemma 6. *Assume condition* (A)*. Then on a possibly richer probability space, there exist i.i.d standard p dimensional Gaussian random vectors $V_1, \ldots, V_n$, such that*

$$|\Theta_n - \Theta_n^*| + |\bar{D}_{n1} - \bar{D}_{n1}^*| = \mathrm{O}_{\mathbb{P}}((\log n)^{3/2}/(n^{1/4} b_n^{3/2})), \tag{30}$$

where

$$\Theta_n^* = \sum_{i=1}^{n} \tilde{\mathbf{T}}_n^\top(t_i) [\mathbb{E}\mathbf{S}_n(t_i)]^{-1} \mathbb{E}[\mathbf{z}_i\mathbf{z}_i^\top] [\mathbb{E}\mathbf{S}_n(t_i)]^{-1} \tilde{\mathbf{T}}_n(t_i),$$
$$\bar{D}_{n1}^* = \sum_{i=1}^{n} \tilde{V}_i^\top [\mathbb{E}\mathbf{S}_n(t_i)]^{-1} \tilde{\mathbf{T}}_n(t_i).$$

Proof. Recall the definitions of $\tilde{V}_i$, $\tilde{\mathbf{T}}_n(t)$ and $\tilde{\mathbf{T}}_{n,l}(t)$ in Proposition 4. We will only prove $\Theta_n - \Theta_n^* = \mathrm{O}_{\mathbb{P}}((\log n)^{3/2}/(n^{1/4} b_n^{3/2}))$ since $\bar{D}_{n1} - \bar{D}_{n1}^* = \mathrm{O}_{\mathbb{P}}((\log n)^{3/2}/(n^{1/4} b_n^{3/2}))$ follows by similar arguments. Note that

$$\Theta_n = \sum_{i=1}^{n} \mathbf{T}_n^\top(t_i) [\mathbb{E}\mathbf{S}_n(t_i)]^{-1} \mathbb{E}[\mathbf{z}_i\mathbf{z}_i^\top] [\mathbb{E}\mathbf{S}_n(t_i)]^{-1} \mathbf{T}_n(t_i) := \sum_{i=1}^{n} \mathbf{T}_n^\top(t_i) \tilde{\Theta}_n(i).$$

By Corollaries 1 and 2 of Wu and Zhou [39], on a possibly richer probability space, there exist i.i.d p dimensional standard Gaussian random vectors $V_1, \ldots, V_n$, such that

$$\max_{1 \le i \le n} |\Delta_i| = \mathrm{O}_{\mathbb{P}}(n^{1/4} (\log n)^{3/2}), \tag{31}$$

where $\Delta_i = \sum_{j=1}^{i} (\varepsilon_j \mathbf{x}_j - \Lambda^{1/2}(t_j) V_j)$. Write $\Theta_n^{(1)} = \sum_{i=1}^{n} \tilde{\mathbf{T}}_n^\top(t_i) \tilde{\Theta}_n(i)$. Then

$$\begin{aligned} &|\Theta_n - \Theta_n^{(1)}| \\ &= \left| \sum_{i=1}^{n} [\mathbf{T}_n^\top(t_i) - \tilde{\mathbf{T}}_n^\top(t_i)] \tilde{\Theta}_n(i) \right| \end{aligned}$$

$$= \left| \sum_{i=1}^{n} [(\mathbf{T}_{n,0}^{\top}(t_i), \mathbf{0}_p^{\top}) - (\tilde{\mathbf{T}}_{n,0}^{\top}(t_i), \mathbf{0}_p^{\top})]\tilde{\Theta}_n(i) + [(\mathbf{0}_p^{\top}, \mathbf{T}_{n,1}^{\top}(t_i)) - (\mathbf{0}_p^{\top}, \tilde{\mathbf{T}}_{n,1}^{\top}(t_i))]\tilde{\Theta}_n(i) \right|$$

$$:= \left| \sum_{i=1}^{n} [W_{n,0}^{\top}(t_i)\tilde{\Theta}_n(i) + W_{n,1}^{\top}(t_i)\tilde{\Theta}_n(i)] \right|.$$

Write $\tilde{\Delta}_i = (\Delta_i^{\top}, \mathbf{0}_p^{\top})^{\top}$ and $\tilde{\Delta}_0 = 0$. Note that

$$\begin{aligned}
\sum_{i=1}^{n} W_{n,0}^{\top}(t_i)\tilde{\Theta}_n(i) &= r_n^2 \sum_{i=1}^{n}\sum_{k=1}^{n} (\tilde{\Delta}_k - \tilde{\Delta}_{k-1}) K_{b_n}(t_k - t_i)\tilde{\Theta}_n(i) \\
&= r_n^2 \sum_{k=1}^{n} (\tilde{\Delta}_k - \tilde{\Delta}_{k-1}) \sum_{i=1}^{n} K_{b_n}(t_k - t_i)\tilde{\Theta}_n(i) \\
&:= r_n^2 \sum_{k=1}^{n} (\tilde{\Delta}_k - \tilde{\Delta}_{k-1}) \Omega_n(k).
\end{aligned}$$

By the summation by parts formula,

$$\begin{aligned}
\left| \sum_{k=1}^{n} (\tilde{\Delta}_k - \tilde{\Delta}_{k-1}) \Omega_n(k) \right| &= \left| \sum_{k=1}^{n-1} \tilde{\Delta}_k (\Omega_n(k) - \Omega_n(k+1)) + \tilde{\Delta}_n \Omega_n(n) \right| \\
&\leq \max_{1 \leq i \leq n} |\tilde{\Delta}_i| \left(\sum_{k=1}^{n-1} |\Omega_n(k) - \Omega_n(k+1)| + |\Omega_n(n)| \right).
\end{aligned}$$

By the smoothness of $K(\cdot)$ and the similar arguments as those in the proof of Lemma 1, it follows that

$$\max_{1 \leq k \leq n-1} \|\Omega_n(k) - \Omega_n(k+1)\| = \mathrm{O}(r_n), \qquad \|\Omega_n(n)\| = \mathrm{O}(1/r_n).$$

Therefore by (31), we have

$$\left| \sum_{i=1}^{n} W_{n,0}^{\top}(t_i)\tilde{\Theta}_n(i) \right| = \mathrm{O}_{\mathbb{P}}\{n^{1/4} \log^{3/2} n (n r_n^3 + r_n)\} = \mathrm{O}_{\mathbb{P}}((\log n)^{3/2}/(n^{1/4} b_n^{3/2})).$$

Similarly,

$$\left| \sum_{i=1}^{n} W_{n,1}^{\top}(t_i)\tilde{\Theta}_n(i) \right| = \mathrm{O}_{\mathbb{P}}((\log n)^{3/2}/(n^{1/4} b_n^{3/2})).$$

Hence, $|\Theta_n - \Theta_n^{(1)}| = \mathrm{O}_{\mathbb{P}}((\log n)^{3/2}/(n^{1/4} b_n^{3/2}))$. Note that

$$\left| \Theta_n^{(1)} - \sum_{i=1}^{n} \tilde{\mathbf{T}}_n^{\top}(t_i)[\mathbb{E}\mathbf{S}_n(t_i)]^{-1}\mathbb{E}[\mathbf{z}_i\mathbf{z}_i^{\top}][\mathbb{E}\mathbf{S}_n(t_i)]^{-1}\tilde{\mathbf{T}}_n(t_i) \right| = \sum_{i=1}^{n} \hat{\Theta}_n(t_i)[\mathbf{T}_n(t_i) - \tilde{\mathbf{T}}_n(t_i)],$$

where $\hat{\Theta}_n(t_i) = \tilde{\mathbf{T}}_n^\top(t_i)[\mathbb{E}\mathbf{S}_n(t_i)]^{-1}\mathbb{E}[\mathbf{z}_i\mathbf{z}_i^\top][\mathbb{E}\mathbf{S}_n(t_i)]^{-1}$. Hence by similar arguments, it follows that

$$\left|\sum_{i=1}^n \hat{\Theta}_n(t_i)[\tilde{\mathbf{T}}_n(t_i) - \mathbf{T}_n(t_i)]\right| = \mathrm{O}_{\mathbb{P}}((\log n)^{3/2}/(n^{1/4}b_n^{3/2})).$$

The lemma follows. □

Lemma 7. *Under condition* (A) *and the assumption that* $b_n \to 0$, $nb_n \to \infty$, *we have*

$$\sqrt{b_n}\left\{\Theta_n^* - 2\bar{D}_{n1}^* - \tilde{K}(0)\int_0^1 \mathrm{tr}[H(t)H^\top(t)]\,\mathrm{d}t/b_n\right\} \Rightarrow N(0,\sigma^2).$$

Proof. Note that both Θ_n^* and D_{n1}^* are quadratic forms of i.i.d. standard Gaussian random vectors. By Lemma 9 and similar arguments as those in the proof of Lemma 5, it can be shown that $\Theta_n^* - \Theta_n^{**} = \mathrm{O}_{\mathbb{P}}(1)$ and $\bar{D}_{n1}^* - \bar{D}_{n1}^{**} = \mathrm{O}_{\mathbb{P}}(1)$, where

$$\Theta_n^{**} = \sum_{i=1}^n \tilde{\mathbf{T}}_{n,0}^\top(t_i)M^{-1}(t_i)\tilde{\mathbf{T}}_{n,0}(t_i),$$

$$\bar{D}_{n1}^{**} = \sum_{i=1}^n V_i^\top \Lambda^{1/2}(t_i)M^{-1}(t_i)\tilde{\mathbf{T}}_{n,0}(t_i).$$

Note that

$$\Theta_n^{**} = r_n^4 \sum_{k=1}^n\sum_{r=1}^n V_k^\top \Lambda^{1/2}(t_k)\left[\sum_{i=1}^n M^{-1}(t_i)K_{b_n}(t_k - t_i)K_{b_n}(t_r - t_i)\right]\Lambda^{1/2}(t_r)V_r$$

and that $M^{-1}(t_i)K_{b_n}(t_k - t_i)K_{b_n}(t_r - t_i) = 0$ if $|t_k - t_r| \ge 2b_n$ or $\min\{|t_i - t_r|, |t_i - t_k|\} \ge b_n$. Hence by Lemma 9 and similar arguments as those in the proof of Lemma 5, it follows that

$$\Theta_n^{**} - \Theta_n^{***} = \mathrm{O}(1) \qquad \text{where } \Theta_n^{***} = r_n^2\sum_{k=1}^n\sum_{r=1}^n V_k^\top \tilde{H}(t_k)K * K_{b_n}(t_k - t_r)\tilde{H}^\top(t_r)V_r,$$

where $\tilde{H}(\cdot) = \Lambda^{1/2}(\cdot)M^{-1/2}(\cdot)$. Similarly,

$$D_{n1}^{**} - D_{n1}^{***} = \mathrm{O}(1) \qquad \text{where } D_{n1}^{***} = r_n^2\sum_{k=1}^n\sum_{r=1}^n V_k^\top \tilde{H}(t_k)K_{b_n}(t_k - t_r)\tilde{H}^\top(t_r)V_r.$$

Using the fact that V_i's are i.i.d. standard Gaussian, elementary calculations show that

$$\sqrt{b_n}\left\{\Theta_n^{***} - 2\bar{D}_{n1}^{***} - \tilde{K}(0)\int_0^1 \mathrm{tr}[H(t)]\,\mathrm{d}t/b_n\right\} \Rightarrow N(0,\sigma^2).$$

The lemma follows. □

Lemma 8. *Under conditions* (A1)–(A7), *we have*

$$\frac{\sum_{i=1}^n \varepsilon_i^2}{n} = \int_0^1 \vartheta^2(t)\,\mathrm{d}t + \mathrm{O}_{\mathbb{P}}(1/\sqrt{n}),$$

where $\vartheta^2(t) = \mathbb{E}[V(t,\mathcal{F}_0)]^2$.

Proof. Note that $\mathbb{E}\varepsilon_i^2 = \vartheta^2(t_i)$. Therefore

$$\sum_{i=1}^n [\varepsilon_i^2 - \vartheta^2(t_i)] = \sum_{k=-\infty}^n \sum_{i=1}^n \mathcal{P}_k^* \varepsilon_i^2,$$

where $\mathcal{P}_i^*(\cdot) = \mathbb{E}(\cdot|\mathcal{R}_i) - \mathbb{E}(\cdot|\mathcal{R}_{i-1})$. Since $\mathcal{P}_i^*$ and $\mathcal{P}_j^*$ are orthogonal for $i \neq j$, we have

$$\left\|\sum_{i=1}^n [\varepsilon_i^2 - \vartheta^2(t_i)]\right\|^2 = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=-\infty}^n \mathbb{E}[\mathcal{P}_k^*\varepsilon_i^2 \mathcal{P}_k^* \varepsilon_j^2] \le \sum_{i=1}^n \sum_{j=1}^n \sum_{k=-\infty}^n \|\mathcal{P}_k^*\varepsilon_i^2\| \|\mathcal{P}_k^*\varepsilon_j^2\|.$$

Let (χ_k^*) be an i.i.d. copy of (χ_k). By Theorem 1 in Wu [38], $\|\mathcal{P}_k^*\varepsilon_i^2\| \le \|\varepsilon_i^2 - \varepsilon_{i,k}^2\|$, where $\varepsilon_{i,k} = (\mathcal{R}_{k-1}, \chi_k^*, \chi_{k+1}, \ldots, \chi_i)$ if $k \le i$ and $\varepsilon_{i,k} = \varepsilon_i$ otherwise. By the Cauchy–Schwarz inequality, we have for $i \ge k$

$$\begin{aligned}
&\|\varepsilon_i^2 - \varepsilon_{i,k}^2\| \\
&\quad \le \|\varepsilon_i + \varepsilon_{i,k}\|_4 \|\varepsilon_i - \varepsilon_{i,k}\|_4 \le C\|H(t_i,\mathcal{G}_i)V(t_i,\mathcal{F}_i) - H(t_i,\mathcal{G}_{i,k})V(t_i,\mathcal{F}_{i,k})\|_4 \\
&\quad \le C\{\|H(t_i,\mathcal{G}_i)\|_4 \|(V(t_i,\mathcal{F}_i) - V(t_i,\mathcal{F}_{i,k})\|_4 + \|V(t_i,\mathcal{F}_{i,k})\|_4 \|H(t_i,\mathcal{G}_i) - H(t_i,\mathcal{G}_{i,k})\|_4\} \\
&\quad \le C(i-k+1)^{-2}.
\end{aligned}$$

Therefore,

$$\left\|\sum_{i=1}^n [\varepsilon_i^2 - \vartheta^2(t_i)]\right\|^2 \le C \sum_{i=1}^n \sum_{j=1}^n \sum_{k=-\infty}^{\min(i,j)} (i-k+1)^{-2}(j-k+1)^{-2} \le Cn.$$

Hence, $\|\sum_{i=1}^n [\varepsilon_i^2 - \vartheta^2(t_i)]\| = \mathrm{O}(\sqrt{n})$. Note that $\sum_{i=1}^n \vartheta^2(t_i) = n\int_0^1 \vartheta^2(t)\,\mathrm{d}t + \mathrm{O}(1)$. The lemma follows. □

Lemma 9. *Recall that* $\mu_h = \int_{-1}^1 x^h K(x)\,\mathrm{d}x$. *Under condition* (A), *we have*

$$\sup_{0\le t\le 1} \|\mathbf{S}_n^{-1}(t) - [\mathbb{E}\mathbf{S}_n(t)]^{-1}\|_8 = \mathrm{O}\left(\frac{1}{\sqrt{nb_n}}\right).$$

Additionally, $\sup_{0\le t\le 1} |[\mathbb{E}\mathbf{S}_n(t)]^{-1}| = \mathrm{O}(1)$. *For* $h = 0, 2$, *we have*

$$\sup_{b_n \le t \le 1-b_n} |[\mathbb{E}\mathbf{S}_{n,h}(t)]^{-1} - [\mu_h M(t)]^{-1}| = \mathrm{O}(b_n^2).$$

Proof. The proof follows by the similar arguments as those of Lemma 6 in Zhou and Wu [43]. Details are omitted. □

Proof of Theorem 1. Theorem 1 follows from Lemmas 1–8 above and the Slutsky's theorem. □

Proof of Theorem 2. Recall that $\varepsilon_i = ([\varepsilon_i^{(1)}]^\top, [\varepsilon_i^{(2)}]^\top)^\top$ and $V_i = ([V_i^{(1)}]^\top, [V_i^{(2)}]^\top)^\top$, where $\varepsilon_i^{(1)}$ and $V_i^{(1)}$ are p_1 dimensional. Note that, under H_{01}, we have a local linear regression of y_i^* on $\mathbf{x}_i^{(2)}$. Recall again the definitions of $\mathbf{S}_n^{(2)}$, $\mathbf{S}_{nl}^{(2)}$, $\mathbf{z}_i^{(2)}$, $\tilde{V}_i^{(2)}$, $\mathbf{T}_n^{(2)}$, $\mathbf{T}_{nl}^{(2)}$, $\tilde{\mathbf{T}}_n^{(2)}$ and $\tilde{\mathbf{T}}_{nl}^{(2)}$ in Section 4.3.

Following very similar arguments as those in Lemmas 1 to 8, it can be shown that

$$\mathrm{RSS}_1 - \mathrm{RSS}_0 - \mathbf{B}_n^{(2)} = \Theta_n^{(2)*} - 2\bar{D}_{n1}^{(2)*} + \mathrm{o}_{\mathbb{P}}(1/\sqrt{b_n}), \tag{32}$$

where $\mathbf{B}_n^{(2)} = \frac{nb_n^4\mu_2^2}{4}\int_0^1 \{[\boldsymbol{\beta}^{(2)}(t)]''\}^\top M_{22}(t)[\boldsymbol{\beta}^{(2)}(t)]''\,\mathrm{d}t + \mathrm{o}_{\mathbb{P}}(nb_n^4)$,

$$\Theta_n^{(2)*} = \sum_{i=1}^n [\tilde{\mathbf{T}}_n^{(2)}(t_i)]^\top [\mathbb{E}\mathbf{S}_n^{(2)}(t_i)]^{-1}\mathbb{E}[\mathbf{z}_i^{(2)}\mathbf{z}_i^{(2)\top}][\mathbb{E}\mathbf{S}_n^{(2)}(t_i)]^{-1}\tilde{\mathbf{T}}_n^{(2)}(t_i),$$

$$\bar{D}_{n1}^{(2)*} = \sum_{i=1}^n \tilde{V}_i^{(2)\top}[\mathbb{E}\mathbf{S}_n^{(2)}(t_i)]^{-1}\tilde{\mathbf{T}}_n^{(2)}(t_i).$$

Note that $\Theta_n^{(2)*}$ and $\bar{D}_{n1}^{(2)*}$ are quadratic forms of i.i.d. Gaussian vectors $V_1, \ldots, V_n$. Theorem 2 follows easily from (30) and (32). □

Proof of Proposition 1. Define $\mathbf{Y}^* = (y_1^*, \ldots, y_n^*)^\top$ and $\tilde{\mathbf{Y}}^* = (\tilde{y}_1^*, \ldots, \tilde{y}_n^*)^\top$. Let $\hat{\varepsilon}_i$ and $\tilde{\varepsilon}_i$ be the ith residual of the local linear regression of y_i^* and $\tilde{y}_i^*$ on $\mathbf{x}_i^{(2)}$, respectively. From (5), we can write $\hat{\varepsilon}_i = y_i^* - R_i\mathbf{Y}^*$ and $\tilde{\varepsilon}_i = \tilde{y}_i^* - R_i\tilde{\mathbf{Y}}^*$, where R_i is a $1 \times n$ vector which can be written in a closed form (5). Note also that R_i is functionally independent of the errors ε_i. Hence,

$$\widetilde{\mathrm{RSS}}_1 - \mathrm{RSS}_1 = \sum_{i=1}^n (\tilde{\varepsilon}_i^2 - \hat{\varepsilon}_i^2) = \sum_{i=1}^n (\tilde{\varepsilon}_i - \hat{\varepsilon}_i)^2 + 2\sum_{i=1}^n \hat{\varepsilon}_i(\tilde{\varepsilon}_i - \hat{\varepsilon}_i) := I + 2II.$$

Let $\delta_i = -(\mathbf{x}_i^{(1)})^\top(\boldsymbol{\beta}_0^{(1)}(t_i,\hat{\theta}) - \boldsymbol{\beta}_0^{(1)}(t_i,\theta_0))$ and $\Delta_n = (\delta_1, \ldots, \delta_n)$. Hence,

$$\mathbb{E}(I) = \sum_{i=1}^n \|\tilde{\varepsilon}_i - \hat{\varepsilon}_i\|^2 = \sum_{i=1}^n \|\delta_i - R_i\Delta_n\|^2.$$

From condition (B), it is easy to see that, for sufficiently large n,

$$\Big\|\max_{1\le i\le n}|\boldsymbol{\beta}_0^{(1)}(t_i,\hat{\theta}) - \boldsymbol{\beta}_0^{(1)}(t_i,\theta_0)|\Big\|_4 = \mathrm{O}(1/\sqrt{n}).$$

Therefore, it is easy to derive from condition (A) that

$$\max_{1\le i\le n}\|\delta_i\| = \mathrm{O}(1/\sqrt{n}) \quad \text{and} \quad \max_{1\le i\le n}\|R_i\Delta_n\| = \mathrm{O}(1/\sqrt{n}). \tag{33}$$

Hence, $I = \mathrm{O}_{\mathbb{P}}(1)$. We now deal with II. Note that, by (28),

$$\hat{\varepsilon}_i = \varepsilon_i - (\mathbf{z}_i^{(2)})^\top(\hat{\boldsymbol{\eta}}^{(2)}(t_i) - \boldsymbol{\eta}^{(2)}(t_i)) = \varepsilon_i - (\mathbf{z}_i^{(2)})^\top(\mathbf{S}_n^{(2)}(t_i))^{-1}[\mathbf{B}_n^{(2)}(t_i) + \mathbf{T}_n^{(2)}(t_i)].$$

Hence,

$$\begin{aligned} II &= \sum_{i=1}^n (\mathbf{z}_i^{(2)})^\top(\mathbf{S}_n^{(2)}(t_i))^{-1}\mathbf{B}_n^{(2)}(t_i)[\tilde{\varepsilon}_i - \hat{\varepsilon}_i] + \sum_{i=1}^n \varepsilon_i[\tilde{\varepsilon}_i - \hat{\varepsilon}_i] \\ &\quad + \sum_{i=1}^n (\mathbf{z}_i^{(2)})^\top(\mathbf{S}_n^{(2)}(t_i))^{-1}\mathbf{T}_n^{(2)}(t_i)[\tilde{\varepsilon}_i - \hat{\varepsilon}_i] \\ &:= II^* + II^{**} + II^{***}. \end{aligned}$$

By Hölder inequality, condition (A) and (33), the bias term

$$\mathbb{E}|II^*| \le \sum_{i=1}^n \|\mathbf{z}_i^{(2)}\|_6 \|\mathbf{S}_n^{(2)}(t_i)^{-1}\|_6 \|\mathbf{B}_n^{(2)}(t_i)\|_6 [\|\delta_i\| + \|R_i\Delta_n\|] = \mathrm{O}(\sqrt{n}b_n^2).$$

Write $J_i = -(\mathbf{x}_i^{(1)})^\top \frac{\partial \boldsymbol{\beta}_0^{(1)}(t_i,\theta_0)}{\partial\theta}$ and let $\mathbf{J} = (J_1^\top, \ldots, J_n^\top)^\top$. By second order Taylor expansion of $\boldsymbol{\beta}_0^{(1)}(t_i, \hat{\theta})$ at θ_0 and condition (B), it is easy to see that

$$\tilde{\varepsilon}_i - \hat{\varepsilon}_i = (J_i - R_i\mathbf{J})(\hat{\theta} - \theta_0) + r_i, \tag{34}$$

with the reminder term r_i satisfying $\max_{1\le i\le n}\|r_i\| = \mathrm{O}(1/n)$. Therefore,

$$\mathbb{E}|II^{**}| \le \left\|\sum_{i=1}^n \varepsilon_i(J_i - R_i\mathbf{J})\right\| \|(\hat{\theta} - \theta_0)\| + \max_{1\le i\le n}\|r_i\| \sum_{i=1}^n \|\varepsilon_i\|$$

By the similar conditioning arguments as those in the proof of Lemma 1, it is easy to show that $\|\sum_{i=1}^n \varepsilon_i(J_i - R_i\mathbf{J})\| = \mathrm{O}(\sqrt{n})$. Hence $\mathbb{E}|II^{**}| = \mathrm{O}(1)$. By similar arguments and elementary but tedious calculations, it follows that $\mathbb{E}|II^{***}| = \mathrm{O}(1)$. Therefore, the proposition follows. □

Proof of Proposition 2. Let $\overline{\mathrm{RSS}}_0 = \sum_{i=1}^n \varepsilon_i^2$. Then $\mathrm{RSS}_a - \mathrm{RSS}_0 = \mathrm{RSS}_a - \overline{\mathrm{RSS}}_0 - (\mathrm{RSS}_0 - \overline{\mathrm{RSS}}_0)$. Under the local alternative $\boldsymbol{\beta}(\cdot) = \boldsymbol{\beta}_0(\cdot) + n^{-4/9}\mathbf{f}_n(\cdot)$, we have

$$\mathrm{RSS}_0 - \overline{\mathrm{RSS}}_0 = n^{-4/9}\sum_{i=1}^n \mathbf{f}_n^\top(t_i)\mathbf{x}_i\varepsilon_i + n^{-8/9}\sum_{i=1}^n [\mathbf{f}_n^\top(t_i)\mathbf{x}_i]^2.$$

By the similar arguments as those in the proof of Lemma 1, it is easy to show that

$$\sqrt{b_n} n^{-8/9} \sum_{i=1}^{n} [\mathbf{f}_n^\top(t_i)\mathbf{x}_i]^2 = c^{1/2} \int_0^1 \mathbf{f}_n^\top(t) M(t) \mathbf{f}_n(t)\, \mathrm{d}t + \mathrm{o}_{\mathbb{P}}(1),$$

$$\sum_{i=1}^{n} \mathbf{f}_n^\top(t_i)\mathbf{x}_i \varepsilon_i = \mathrm{O}_{\mathbb{P}}(n^{1/2}).$$

On the other hand, by Lemmas 1–8 and the fact that $\boldsymbol{\beta}(\cdot) = \boldsymbol{\beta}_0(\cdot) + n^{-4/9}\mathbf{f}_n(\cdot)$, it is easy to show that

$$\sqrt{b_n}\left\{\mathrm{RSS}_a - \overline{\mathrm{RSS}}_0 - \frac{\tilde{K}(0)}{b_n}\int_0^1 \mathrm{tr}[H(t)]\, \mathrm{d}t\right\} - \frac{c^{9/2}\mu_2^2}{4}\int_0^1 [\boldsymbol{\beta}''(t)]^\top M(t) \boldsymbol{\beta}''(t)\, \mathrm{d}t - \frac{c^{9/2}\mu_2^2}{4} F_2$$
$$\Rightarrow N(0, \sigma^2)$$

and $\mathrm{RSS}_0/n = \mathcal{V} + \mathrm{o}_{\mathbb{P}}(1)$. Therefore, the proposition follows. □

Proof of Theorem 3. A careful check of Lemmas 1 and 2 shows that the asymptotic bias of λ_n^*

$$\mathbf{B}_n^* = \int_{c_{\min}}^{c_{\max}} \frac{n(zn^{-\gamma})^4\mu_2^2}{4}\, \mathrm{d}z \int_0^1 [\boldsymbol{\beta}''(t)]^\top M(t) \boldsymbol{\beta}''(t)\, \mathrm{d}t + \mathrm{o}_{\mathbb{P}}(n^{1-4\gamma}). \tag{35}$$

Another careful check of Lemmas 3 to 8 and using Lemma 9 show that

$$\begin{aligned}\lambda_n^* - \mathbf{B}_n^* = \int_{c_{\min}}^{c_{\max}} \sum_{k=1}^{n}\sum_{r=1}^{n} V_k^\top \tilde{H}(t_k)[2K_{zn^{-\gamma}}(t_k - t_r) \\ - K * K_{zn^{-\gamma}}(t_k - t_r)]\tilde{H}^\top(t_r) V_r/(nzn^{-\gamma})\, \mathrm{d}z + \mathrm{o}_{\mathbb{P}}(n^{-\gamma/2}).\end{aligned} \tag{36}$$

Since V_i's are i.i.d. standard Gaussian, a central limit theorem for $\lambda_n^* - \mathbf{B}_n^*$ can be easily derived. Now Theorem 3 follows from (35) and (36). Details are omitted. □

Proof of Proposition 3. By the Cauchy–Schwarz inequality,

$$\begin{aligned}\int_{\mathbb{R}} Q(c_{max}, y)^2\, \mathrm{d}y &= \int_{\mathbb{R}} \left[\int_{c_{\min}}^{c_{\max}} ([2K(y/z) - K * K(y/z)]/\sqrt{z}) \times 1/\sqrt{z}\, \mathrm{d}z\right]^2 \mathrm{d}y \\ &\le \int_{\mathbb{R}} \left[\int_{c_{\min}}^{c_{\max}} [2K(y/z) - K * K(y/z)]^2/z\, \mathrm{d}z \int_{c_{\min}}^{c_{\max}} 1/z\, \mathrm{d}z\right] \mathrm{d}y \\ &= (\log(c_{\max}) - \log(c_{\min})) \int_{c_{\min}}^{c_{\max}} \int_{\mathbb{R}} [2K(y/z) - K * K(y/z)]^2/z\, \mathrm{d}y\, \mathrm{d}z \\ &= (\log(c_{\max}) - \log(c_{\min}))(c_{\max} - c_{\min}) \int_{\mathbb{R}} \tilde{K}^2(t)\, \mathrm{d}t.\end{aligned}$$

Consider any fixed $c \in (0,\infty)$. Plugging the above inequality into (19) and letting $c_{\max} \downarrow c$ and $c_{\min} \uparrow c$, it follows that $\sup_{0<c_{\min}<c_{\max}<\infty} \beta^*_\alpha(c_{\min}, c_{\max}) \geq \beta_\alpha(c)$. Hence, the proposition follows. □

Acknowledgements

I am grateful to the two anonymous referees for their many helpful comments which greatly improved the quality of the original version of the paper. The research was supported in part by NSERC of Canada.

References

[1] An, H. and Cheng, B. (1991). A Kolmogorov–Smirnov type statistic with application to test for nonlinearity in time series. *International Statistical Review* **59** 287–307.

[2] Brown, J.P., Song, H. and McGillivray, A. (1997). Forecasting UK house prices: A time varying coefficient approach. *Economic Modeling* **14** 529–548.

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

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

[5] Dahlhaus, R. (2009). Local inference for locally stationary time series based on the empirical spectral measure. *J. Econometrics* **151** 101–112. MR2559818

[6] Dette, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimators. *Ann. Statist.* **27** 1012–1040. MR1724039

[7] Dette, H. and Hetzler, B. (2007). Specification tests indexed by bandwidths. *Sankhyā* **69** 28–54. MR2385277

[8] Dette, H., Preuss, P. and Vetter, M. (2011). A measure of stationarity in locally stationary processes with applications to testing. *J. Amer. Statist. Assoc.* **106** 1113–1124. MR2894768

[9] Dette, H. and Spreckelsen, I. (2003). A note on a specification test for time series models based on spectral density estimation. *Scand. J. Stat.* **30** 481–491. MR2002223

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

[11] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications. Monographs on Statistics and Applied Probability* **66**. London: Chapman & Hall. MR1383587

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

[13] Fan, J. and Jiang, J. (2005). Nonparametric inferences for additive models. *J. Amer. Statist. Assoc.* **100** 890–907. MR2201017

[14] Fan, J. and Jiang, J. (2007). Nonparametric inference with generalized likelihood ratio tests. *TEST* **16** 409–444. MR2365172

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

[16] Fan, Y. and Li, Q. (1999). Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing. *J. Nonparametr. Stat.* **10** 245–271. MR1708583

[17] Gersch, W. and Kitagawa, G. (1985). A time varying AR coefficient model for modelling and simulating earthquake ground motion. *Earthquake Engineering & Structural Dynamics* **13** 243–254.

[18] Hjellvik, V., Yao, Q. and Tjøstheim, D. (1998). Local polynomial estimation of conditional quantities with application to linearity testing. *J. Statist. Plann. Inference* **68** 295–321.

[19] Hong, Y. and Lee, Y. (2009). A loss function approach to model specification testing and its relative efficiency to the GLR test. Unpublished manuscript.

[20] 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

[21] Horowitz, J.L. and Spokoiny, V.G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* **69** 599–631. MR1828537

[22] Ingster, Y.I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. I. *Math. Methods Statist.* **2** 85–114. MR1257978

[23] Kitagawa, G. and Gersch, W. (1985). A smoothness priors time-varying AR coefficient modeling of nonstationary covariance time series. *IEEE Trans. Automat. Control* **30** 48–56. MR0777076

[24] Lehmann, E.L. (2006). On likelihood ratio tests, 2nd ed. In *Optimality. Institute of Mathematical Statistics Lecture Notes—Monograph Series* **49** 1–8. Beachwood, OH: IMS. MR2337826

[25] Müller, H.G. (2007). Comments on: Nonparametric inference with generalized likelihood ratio tests. *TEST* **16** 450–452. MR2415642

[26] Nason, G.P., von Sachs, R. and Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **62** 271–292. MR1749539

[27] Neumann, M.H. and von Sachs, R. (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. *Ann. Statist.* **25** 38–76. MR1429917

[28] Ombao, H., von Sachs, R. and Guo, W. (2005). SLEX analysis of multivariate nonstationary time series. *J. Amer. Statist. Assoc.* **100** 519–531. MR2160556

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

[30] Orbe, S., Ferreira, E. and Rodriguez-Poo, J. (2006). On the estimation and testing of time varying constraints in econometric models. *Statist. Sinica* **16** 1313–1333. MR2327493

[31] Paparoditis, E. (2000). Spectral density based goodness-of-fit tests for time series models. *Scand. J. Stat.* **27** 143–176. MR1774049

[32] Paparoditis, E. (2009). Testing temporal constancy of the spectral structure of a time series. *Bernoulli* **15** 1190–1221. MR2597589

[33] Paparoditis, E. (2010). Validating stationarity assumptions in time series analysis by rolling local periodograms. *J. Amer. Statist. Assoc.* **105** 839–851. MR2724865

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

[35] Robinson, P.M. (1989). Nonparametric estimation of time-varying parameters. In *Statistical Analysis and Forecasting of Economic Structural Change* (P. Hackl, ed.) 164–253. Berlin: Springer.

[36] Sergides, M. and Paparoditis, E. (2009). Frequency domain tests of semi-parametric hypotheses for locally stationary processes. *Scand. J. Stat.* **36** 800–821. MR2573309

[37] Stock, J.H. and Watson, M.W. (1998). Median unbiased estimation of coefficient variance in a time-varying parameter model. *J. Amer. Statist. Assoc.* **93** 349–358. MR1614585

[38] Wu, W.B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci. USA* **102** 14150–14154. MR2172215

[39] Wu, W.B. and Zhou, Z. (2011). Gaussian approximations for non-stationary multiple time series. *Statist. Sinica* **21** 1397–1413. MR2827528

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

[41] Zhang, C.M. (2003). Adaptive tests of regression functions via multiscale generalized likelihood ratios. *Canad. J. Statist.* **31** 151–171. MR2016225

[42] Zhou, Z. and Wu, W.B. (2009). Local linear quantile estimation for nonstationary time series. *Ann. Statist.* **37** 2696–2729. MR2541444

[43] 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

Received January 2012 and revised August 2012