Trending Time-Varying Coefficient Models With Serially Correlated Errors

Zongwu Cai*
Department of Mathematics
University of North Carolina
Charlotte, NC 28223
E-mail: zcai@uncc.edu

First Draft: December, 2002

In this paper we study time-varying coefficient models with time trend function and serially correlated errors to characterize nonlinear, nonstationary and trending phenomenon in time series. Compared with the Nadaraya-Watson method, the local linear approach is developed to estimate the time trend and coefficient functions. The consistency of the proposed estimators is obtained without any specification of the error distribution and the asymptotic normality of the proposed estimators is established under the α-mixing conditions. The explicit expressions of the asymptotic bias and variance are given for both estimators. The asymptotic bias is just in a regular nonparametric form but the asymptotic variance is shared by parametric estimators. Also, the asymptotic behaviors at both interior and boundary points are studied for both estimators and it shows that two estimators share the exact same asymptotic properties at the interior points but not at the boundaries. Moreover, proposed are a new bandwidth selector based on the nonparametric version of the Akaike information criterion, a consistent estimator of the asymptotic variance, and a simple nonparametric version of bootstrap (*i.e.* wild bootstrap) test for testing the misspecification and stationarity. Finally, we conduct some Monte Carlo experiments to examine the finite sample performances of the proposed modeling procedures and test.

KEY WORDS: Bandwidth selection; Boundary effects; Fixed design; Functional coefficient models; Local linear fitting; Misspecification test; Nadaraya-Watson estimation; Nonlinearity; Nonstationarity; Stationarity; Time series errors; Wild bootstrap.

*This research was supported, in part, by the National Science Foundation grant DMS-0072400 and funds provided by the University of North Carolina at Charlotte and Sonderforschungsbereich 373, Berlin, Germany. The author would like to thank Professors Philip Hans Franses, Ted Juhl, Xu-Feng Niu and Robert Shumway for helpful comments and suggestions.

1 Introduction

The analysis of nonlinear and nonstationary time series particularly with time trend has been very active during the last two decades because most of time series, in particular, observed from economics and business, are nonlinear or nonstationary or trending (Granger and Teräsvirta 1993; Franses 1996, 1998; Phillips 2001; Tsay 2002). For example, the market model in finance is an example that relates the return of an individual stock to the return of a market index or another individual stock. Another example is the term structure of interest rates in which the time evolution of the relationship between interest rates with different maturities is investigated. For more examples in macroeconomic activity, see the survey paper by Phillips (2001). To characterize these phenomena, during the recent years there have been proposed several nonlinear and nonstationary parametric, semiparametric and nonparametric time series models with/without time trend in the econometrics, finance and statistics literature. For more detailed discussions on this aspect, see the papers by Park and Phillips (1999, 2002) and Chang, Park and Phillips (1999), Phillips (2001), Karlsen and Tjøstheim (2001), and Karlsen, Myklebust and Tjøstheim (2001), and the books by Granger and Teräsvirta (1993), Franses (1996, 1998) and Tsay (2002) and the references therein. Although the literature is already vast and continues to grow swiftly, as pointed out by Phillips (2001), the research in this area is just beginning.

There are several ways to explore the nonlinearity and nonstationarity and one of the most attractive models is the time-varying coefficient time series models with time trend and serially correlated errors. Indeed, regression models with serially correlated (time series) errors are widely applicable in economics and finance, but it is one of most commonly misused econometric models because the serial dependence in the errors is often overlooked. Recently, there are some new developments. Roussas (1989) studied the following fixed design time series model without exogenous variables

$$Y_i = \beta_0(t_i) + u_i \tag{1}$$

under the assumption that $\{u_i\}$ is a sequence of α-mixing random variables and other types of dependence. He considered a linear estimator and obtained consistency of the proposed estimator. Indeed, this model deals with only the time trend function $\beta_0(t)$. Roussas, Tran and Ioannides (1992) derived the asymptotic normality but without giving the explicit expressions of the asymptotic bias and variance and Tran, Roussas, Yakowitz and Van (1996) considered the above model by assuming that $\{u_i\}$ is a linear process. For more references, see the paper by Tran, Roussas, Yakowitz and Van (1996). Actually, model (1) is very useful in various applied fields such as the longitudinal study in medical sciences. The reader is referred to the book by Müller (1988) for details.

To allow for the presence of exogenous (explanatory) variables (covariates) $\mathbf{X}_i = (X_{i1}, \ldots, X_{id})^T$ which might be important in an econometric context or other fields, we consider the following time-varying coefficient time series models with time trend

$$Y_i = \beta_0(t_i) + \sum_{j=1}^{d} \beta_j(t_i)\, X_{ij} + u_i = \widetilde{\mathbf{X}}_i^T\, \boldsymbol{\beta}(t_i) + u_i, \tag{2}$$

where $\widetilde{\mathbf{X}}_i = (1,\ \mathbf{X}_i^T)^T$, $\boldsymbol{\beta}(t) = (\beta_0(t), \beta_1(t), \ldots, \beta_d(t))^T$ and $E(u_i \,|\, \mathbf{X}_i) = 0$. Since the trend function $\beta_0(\cdot)$ might not be polynomial in many applications in econometrics, as pointed out by Phillips (2001), here $\boldsymbol{\beta}(\cdot)$ is unspecified but is assumed to be smooth. To include the heteroscedasticity in the model, $E(u_i^2 \,|\, \mathbf{X}_i)$ is allowed to be a function of $\mathbf{X}_i$. This is particularly appealing in economics and finance. Finally, $\{(u_i,\, \mathbf{X}_i)\}$ is assumed to be stationary. Clearly, both models (1) and (2) include the deterministic time trend function $\beta_0(t)$, the time series $\{Y_i\}$ is not stationary, and (2) covers (1) as a special case. The deterministic time trend function $\beta_0(t)$ might be an important ingredient in modeling the economic and financial data and it might not be polynomial as pointed out by Phillips (2001) who gave an excellent review on some present developments and future challenges about the trending time series models. Also, if all the coefficient functions including $\beta_0(t)$ are constant, the time series $\{Y_i\}$ generated by the above models is stationary if so is the time series $\{(u_i,\, \mathbf{X}_i)\}$. Finally, the model is closely related to the functional-coefficient time series regression models which allow the coefficient functions to depend on some random variables rather than time, studied by Chen and Tsay (1993), Xia and Li (1999) and Cai, Fan and Yao (2000), and the time-varying coefficient autoregressive models with $\mathbf{X}_i$ being a vector of lagged variables, explored by Cai and Tiwari (2000) and Kim (2001).

Robinson (1989) studied model (2) under the assumptions that the time series $\{\mathbf{X}_i\}$ is stationary α-mixing and the errors $\{u_i\}$ are iid and independent of $\mathbf{X}_i$ and developed the Nadaraya-Watson method to estimate the coefficient functions and studied the asymptotic properties of the proposed estimator. More importantly, he demonstrated that making $\boldsymbol{\beta}(t)$ to depend on the sample size n is necessary to provide the asymptotic justification for any nonparametric smoothing estimators. For more discussions on this point, see the paper by Robinson (1989) and Section 3.1. Later, Robinson (1991) considered a more general model and relaxed the iid assumption on $\{u_i\}$ to α-mixing.

The main contribution of this paper is to consider model (2) and to propose using the local linear estimation to estimate the coefficient functions and to make a comparison with the Nadaraya-Watson method. It is showed that the estimators based on both the local linear fitting and the Nadaraya-Watson method share the exact same asymptotic behavior at the interior points but not at boundaries. Also, it shows that the consistency of the proposed estimators can be obtained without specifying the error distribution and the asymptotic variance of the proposed estimator depends on not only the variance of the error but also the autocorrelations. This property is shared by parametric estimators. Further, to choose the data-driven fashioned bandwidth, we propose a new bandwidth selector based on the nonparametric version of the Akaike information criterion and we propose a consistent estimator of the asymptotic variances which can be used to construct the pointwise confidence intervals. Finally, an important econometric question in fitting model (2) arises whether the coefficient functions are actually varying (namely, if a linear model is adequate or the time series $\{Y_i\}$ is stationary) or more generally if a parametric model fits the given data or there is no time trend $\beta_0(t)$ at all or there are some exogenous variables not statistically significant. This amounts to testing whether the coefficient functions are constant or zero or in a certain parametric form. This is an important issue in the econometric misspecification and stationarity tests. We propose a new testing procedure based on the comparison of the residual sum of squares under the null and alternative models. This is related to the generalized likelihood ratio statistic of Fan, Zhang and Zhang (2001) and the nonparametric F-test of Hastie and Tibshirani (1990, Section 3.9) for the iid sample, Cai and Tiwari (2000) for the time-varying coefficient autoregressive time series model and Cai (2002a) for the additive time series model. The null distribution of the proposed test statistic is estimated by using a simple nonparametric version of bootstrap sampling scheme (*i.e.* wild bootstrap) which can include the heteroscedasticity in the model. Therefore, this paper provides some deeper insights into how to apply for the statistical tools to make econometric modeling of (2) and to make model (2) practically applicable and useful.

The rest of the paper is organized as follows. Section 2 is devoted to the presentation of the estimation methods and a new bandwidth selector based on the nonparametric version of the Akaike information criterion. In Section 3, the asymptotic theory is presented along with conditions and some remarks, a consistent estimator of the asymptotic variance is provided and a new test procedure for the misspecification and stationarity is proposed. Finally, Section 4 reports some results from numerical simulations and we conclude with a brief discussion in Section 5. All the technical proofs are relegated to the Appendix.

2 Modeling Procedures

2.1 Local Linear and Constant Estimation

For estimating $\{\beta_j(\cdot)\}$ in (2), a local linear method is employed, although a general local polynomial method is also applicable. Local (polynomial) linear methods have been widely used in nonparametric regression during recent years due to their attractive mathematical efficiency, bias reduction and adaptation of edge effects (see Fan 1993; Fan and Gijbels 1996). Assuming $\{\beta_j(\cdot)\}$ have a continuous second derivative, then $\{\beta_j(\cdot)\}$ can be approximated by a linear function at any fixed time point t as follows:

$$\beta_j(t_i) \simeq a_j + b_j\,(t_i - t), \quad 0 \le j \le d,$$

where $\simeq$ denotes the first order Taylor approximation and $a_j = \beta_j(t)$ and $b_j = \beta_j'(t)$. Hence (2) is approximated by

$$Y_i \simeq \mathbf{Z}_i^T\,\boldsymbol{\theta} + u_i,$$

where $\mathbf{Z}_i = (\widetilde{\mathbf{X}}_i^T, \widetilde{\mathbf{X}}_i^T (t_i - t))^T$ and $\boldsymbol{\theta} = \boldsymbol{\theta}(t) = (\boldsymbol{\beta}^T(t), \boldsymbol{\beta}'^T(t))^T$. Therefore, the locally weighted least square is

$$\sum_{i=1}^{n} \left\{ Y_i - \mathbf{Z}_i^T \boldsymbol{\theta} \right\}^2 K_h(t_i - t), \tag{3}$$

where $K_h(u) = K(u/h)/h$, $K(\cdot)$ is the kernel function and $h = h_n > 0$ is the bandwidth satisfying $h \to 0$ and $n\,h \to \infty$ as $n \to \infty$ which controls the amount of smoothing used in the estimation.

By minimizing (3) with respect to $\boldsymbol{\theta}$, we obtain the local linear estimate of $\beta_j(t)$, denoted by $\widehat{\beta}_j(t)$, which is equal to the first $(d+1)$ elements of $\widehat{\boldsymbol{\theta}}$, and the local linear estimator of the derivative of $\beta_j(t)$, denoted by $\widehat{\beta}'_j(t)$, which is equal to the last $(d+1)$ elements of $\widehat{\boldsymbol{\theta}}$. It is easy to show that the minimizer of (3) is given by

$$\widehat{\boldsymbol{\theta}} = \begin{pmatrix} \mathbf{S}_{n,0}(t) & \mathbf{S}_{n,1}^T(t) \\ \mathbf{S}_{n,1}(t) & \mathbf{S}_{n,2}(t) \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{T}_{n,0}(t) \\ \mathbf{T}_{n,1}(t) \end{pmatrix} \equiv \mathbf{S}_n^{-1}(t)\, \mathbf{T}_n(t), \tag{4}$$

where

$$\mathbf{S}_{n,0}(t) = n^{-1} \sum_{i=1}^{n} \widetilde{\mathbf{X}}_i \widetilde{\mathbf{X}}_i^T K_h(t_i - t), \qquad \mathbf{S}_{n,1}(t) = n^{-1} \sum_{i=1}^{n} \widetilde{\mathbf{X}}_i \widetilde{\mathbf{X}}_i^T (t_i - t) K_h(t_i - t),$$

$$\mathbf{S}_{n,2}(t) = n^{-1} \sum_{i=1}^{n} \widetilde{\mathbf{X}}_i \widetilde{\mathbf{X}}_i^T (t_i - t)^2 K_h(t_i - t), \quad \text{and} \quad \mathbf{T}_{n,k}(t) = n^{-1} \sum_{i=1}^{n} \widetilde{\mathbf{X}}_i (t_i - t)^k K_h(t_i - t)\, Y_i$$

for $k = 0$ and 1. Note that the local linear estimator can be viewed as the least square estimator of the following working linear model

$$K_h^{1/2}(t_i - t)\, Y_i = K_h^{1/2}(t_i - t)\, \mathbf{Z}_i^T \boldsymbol{\theta}_1 + K_h^{1/2}(t_i - t)\, \mathbf{Z}_i^T (t_i - t)\, \boldsymbol{\theta}_2 + \varepsilon_i.$$

Therefore, the computational implementation can be easily carried out by any standard statistical softwares.

If $\beta_j(\cdot)$ is approximated by a constant at any fixed time point t, then (2) is approximated by $Y_i \simeq \widetilde{\mathbf{X}}_i^T \boldsymbol{\theta}^* + u_i$, where $\boldsymbol{\theta}^* = \boldsymbol{\beta}(t)$, and the locally weighted least square becomes

$$\sum_{i=1}^{n} \left\{ Y_i - \widetilde{\mathbf{X}}_i^T \boldsymbol{\theta}^* \right\}^2 K_h(t_i - t). \tag{5}$$

Minimizing (5) with respect to $\boldsymbol{\theta}^*$ gives the local constant (Nadaraya-Watson) estimator of $\boldsymbol{\beta}(t)$, denoted by $\widetilde{\boldsymbol{\beta}}(t)$, which can be expressed as follows

$$\widetilde{\boldsymbol{\beta}}(t) = \widehat{\boldsymbol{\theta}}^* = \mathbf{S}_{n,0}^{-1}(t)\, \mathbf{T}_{n,0}(t). \tag{6}$$

We remark that the local linear estimator given by (3) does not take into the account of autocorrelations of $\{u_i\}$. The efficiency improvements are possible by correcting the disturbance serial correlation. For example, given a known autocorrelation structure of $\{u_i\}$, we can replace (3) with

$$\{\mathbf{Y} - \mathbf{Z}\,\boldsymbol{\theta}\}^T\ \mathbf{K}_h^{1/2}\ \boldsymbol{\Sigma}^{-1}\,\mathbf{K}_h^{1/2}\ \{\mathbf{Y} - \mathbf{Z}\,\boldsymbol{\theta}\}$$

where $\mathbf{Y}$ and $\mathbf{Z}$ are obtained by stacking the Y_i and $\mathbf{Z}_i$, respectively, and $\boldsymbol{\Sigma}$ is the covariance matrix of $\{u_i\}$, and $\mathbf{K}_h^{1/2}$ is a diagonal matrix with i-th diagonal element being $K_h^{1/2}(t_i - t)$. When the autocorrelations of $\{u_i\}$ are unknown but can be modelled by certain parametric models such as the ARMA models, it is then possible to use an iterative procedure to estimate jointly the nonparametric function and the unknown parameters. The same remark can be applied to the local constant estimator given by (5). In this paper, we adopt the simple locally weighted least square approach and use equation (3) to construct our estimator.

Note that many other nonparametric smoothing methods can be used here. The locally weighted least square method or the Nadaraya-Watson approach is just one of the choices. There is a vast literature in theory and empirical study on the comparison of different nonparametric smoothing methods (see Härdle 1990; Fan and Gijbels 1996; Pagan and Ullah 1999).

The restriction to the locally weighted least square method suggests that the normality is at least being considered as a baseline. However, when the non-normality is clearly present, a robust approach would be considered. Cai and Ould-Said (2001) considered this aspect in nonparametric regression estimation for time series. If some of $\mathbf{X}_i$ are endogenous variables, the various instrumental variable type estimates of linear and nonlinear simultaneous equations and transformation models can be easily applied here with some modifications. For example, we can apply the two-stage local linear technique proposed by Cai, Das, Xiong and Wu (2002) and the nonparametric generalized method of moments proposed by Cai (2002b).

2.2 Bandwidth Selection

From the asymptotic results presented in Section 3, we can see that the bandwidth plays an essential role in the trade-off between reducing bias and variance. Therefore, the selection of the bandwidth is similar to the model selection for linear models. The practitioner is often able to choose the bandwidth satisfactorily by some ad hoc methods, however, it is desirable to have a reliable data-driven and easily implemented bandwidth selector, which is also very important issue in econometric modeling. In nonparametric regression setting with the Gaussian errors, there have been a considerable amount of methods devoted to selecting the optimal bandwidth, some of which can be adapted to handle dependence in time series. For the problem discussed in this paper, an easy way is to derive an analogue to the cross-validation or its variations, however, as pointed out by Fan, Heckman, and Wand (1995) and Fan and Gijbels (1996), the cross-validation performs poorly due to its large sample variation, even worse for dependent data, and its computation might be a burden.

Inspired by classical Akaike information criterion (AIC) discussed in Engle, Granger, Rice, and Weiss (1986) for time series data for the bandwidth selection, here we propose a simple and quick method to select bandwidth for the foregoing estimation procedures. Indeed, this procedure can be regarded as a nonparametric version of the AIC to be attentive to the structure of time series data and the over-fitting or under-fitting tendency. Note that the idea is also motivated by its analogue of Cai and Tiwari (2000) for the time-varying coefficient autoregressive models.

The basic idea is described as follows. For given observed values $\{Y_t\}_{t=1}^n$, the fitted values can be expressed as $\widehat{\mathbf{Y}} = \mathbf{H}_\lambda \mathbf{Y}$, where $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ and $\mathbf{H}_\lambda$ is called the $n \times n$ smoother (or hat) matrix associated with the smoothing parameter λ. Motivated by the classical AIC for linear models under the likelihood setting

$$-2\,(\text{maximized log likelihood}) + 2\,(\text{number of estimated parameters}), \tag{7}$$

here we propose the following nonparametric version of AIC to select h by minimizing

$$\text{AIC}(\lambda) = \log\{SSE\} + \psi(n_\lambda,\, n), \tag{8}$$

where $SSE = \sum_{t=1}^n (Y_t - \widehat{Y}_t)^2$, regarded as the replacement of the first term in (7) and n_λ is the trace of the hat matrix $\mathbf{H}_\lambda$, called to be the effective number of parameters or the nonparametric version of degrees of freedom by Hastie and Tibshirani (1990, Section 3.5) for nonparametric models. Particularly, we choose $\psi(n_\lambda,\, n)$ to be the form of the bias-corrected version of the AIC, due to Hurvich and Tsai (1989),

$$\psi(n_\lambda,\, n) = 2\,(n_\lambda + 1)/(n - n_\lambda - 2). \tag{9}$$

It has been suggested that we try (9), as recommended by Brockwell and Davis (1991, Section 9.3), which penalizes extra parameters for larger values of the number of parameters. For the nonparametric setting, n_λ would be very large since the parameter space is functional space with infinite dimension. Therefore, (9) is particularly suitable for this case. Indeed, (8) is a very general formulation. For example, when $\psi(n_\lambda,\, n) = -2\log(1 - n_\lambda/n)$, then (8) becomes the generalized cross-validation (GCV) criterion of Wahba (1977). When $\psi(n_\lambda,\, n) = 2\,n_\lambda/n$, then (8) is the classical AIC discussed in Engle, Granger, Rice, and Weiss (1986) for time series data. When $\psi(n_\lambda,\, n) = -\log(1 - 2\,n_\lambda/n)$, (8) is the T-criterion proposed and studied by Rice (1984) for iid samples. It is clear that when $n_\lambda/n \to 0$, then the nonparametric AIC, the GCV and the T-criterion are asymptotically equivalent. However, the T-criterion requires $n_\lambda/n < 1/2$, and, when n_λ/n is large, the GCV has relatively weak penalty. This is especially true for the nonparametric setting. Therefore, the criterion proposed here counteracts the over-fitting tendency of the GCV.

Alternatively, one might use some existing methods in the literature although they may require more computing, for example, see the papers by Robinson (1989), Yao and Tong (1994), View (1994), and Tschernig and Yang (2000).

3 Statistical Results

3.1 Asymptotic Theory

The estimation methods described in Section 2 can accommodate both fixed and random designs. Here, the main focus of this paper is on fixed-design. The reason of doing so is that it might be suitable for pure time series data, such as financial and economic data. Since data are generally observed in time order in many applications, we only consider the equal spaced design points $t_i = i/n$ for simplicity although the theoretical results developed later still hold for non-equal spaced design points. As pointed out by Robinson (1989), it is necessary to make $\boldsymbol{\beta}(t)$ to depend on the sample size n to provide the asymptotic justification for any nonparametric smoothing estimators. See Robinson (1989) for the detailed discussions on this aspect. This type of assumption is commonly used in fixed-design nonparametric regression contexts. Detailed discussions on this respect can be found in Müller (1988), Roussas (1989), Robinson (1989, 1991), Roussas, Tran, and Ioannides (1992), Tran, Roussas, Yakowitz and Van (1996), Cai and Chen (2002), among others for nonparametric regression estimation. For random design, it is commonly assumed in the statistics literature that the design points t_i, for $i = 1, \ldots, n$, are chosen independently according to some continuous and positive design density, and they are independent of $\{u_i\}$. Note that this type of assumption is particularly common in longitudinal data study to avoid the difficulties in theoretical derivations (Cai and Wu 2002).

The errors in a deterministic trend time series model such as (1) or (2) are usually assumed to follow certain linear time series models such as an ARMA process. Here we consider a more general structure – the α-mixing process, which includes many linear and nonlinear time series models as special cases. The asymptotic results here are derived under the α-mixing assumption. However, Roussas (1989) considered linear processes without satisfying the mixing condition. Potentially the theoretical results derived here can be extended to such cases. It is well-known in the econometrics and statistics literature that α-mixing is reasonably weak and is known to be fulfilled for many linear and nonlinear time series models under some regularity conditions. For example, Gorodetskii (1977) and Withers (1981) derived the conditions under which a linear process is α-mixing. In fact, under very mild assumptions, linear autoregressive and more generally bilinear time series models are α-mixing with mixing coefficients decaying exponentially. On the other hand, Auestad and Tjøstheim (1990) provided illuminating discussions on the role of α-mixing (including geometric ergodicity) for model identification in nonlinear time series analysis. Chen and Tsay (1993) showed that the functional autoregressive process is geometrically ergodic under certain conditions. Further, Masry and Tjøstheim (1995, 1997), Lu (1998) and Cai and Masry (2000) demonstrated that under some mild conditions, both autoregressive conditional heteroscedastic processes and nonlinear additive autoregressive models with exogenous variables, particularly popular in econometrics and finance, are stationary and α-mixing.

We first list all the assumptions needed for the asymptotic theory although some of them might not be the weakest possible.

ASSUMPTION A:

A1. The kernel $K(u)$ is symmetric and satisfies the Lipschitz condition and $u^2 K(u)$ is bounded.

A2. Assume that $\{(u_i, \mathbf{X}_i)\}$ is a strictly stationary α-mixing and there exists some $\delta > 0$ such that $E|\mathbf{X}_i|^{2(2+\delta)} < \infty$, $E|u_i \widetilde{\mathbf{X}}_i|^{2(1+\delta)} < \infty$ and the mixing coefficient $\alpha(i)$ satisfies $\alpha(i) = O\left(i^{-\tau}\right)$ with $\tau = (2+\delta)(1+\delta)/\delta$.

A3. $n\, h^{1+4/\delta} \to \infty$.

It is clear that assumptions listed above are not strong. For example, the commonly used kernel functions such as the Gaussian density and Epanechnikov kernel $K(u) = 0.75\,(1-u^2)\,I(|u| \le 1)$ satisfy Assumption A1. Assumption A2 is a standard requirement for moments and the mixing coefficient for an α-mixing time series. If $\delta > 1$, then the optimal bandwidth $h_{opt} = O(n^{-1/5})$ (see below) satisfies Assumption A3.

Note that all the asymptotic results here assume that $n \to \infty$. Define, for $k \ge 0$, $\mu_k = \int u^k K(u)\, d\,u$ and $\nu_k = \int u^k K^2(u)\, d\,u$. Let $\mathbf{R}_k = \mathrm{cov}(u_i\, \widetilde{\mathbf{X}}_i, u_{i+k}\, \widetilde{\mathbf{X}}_{i+k})$ for any i and k. Set $\mathbf{\Sigma}_0 = \sum_{k=-\infty}^{\infty} \mathbf{R}_k$. Then $\mathbf{\Sigma}_0$ exists by Assumption A2 (see Lemma 1 later). If it is assumed that $\{\mathbf{X}_i\}$ is independent of $\{u_i\}$, then $\mathbf{R}_k = r_k\, \mathbf{\Omega}_k$, where $\mathbf{\Omega}_k = E(\widetilde{\mathbf{X}}_i\, \widetilde{\mathbf{X}}_{i+k}^T)$ and $r_k = \mathrm{cov}(u_i, u_{i+k})$ for any i and k. Let $\mathbf{H} = \mathrm{diag}\{\mathbf{I}_{d+1},\, h\,\mathbf{I}_{d+1}\}$, where $\mathbf{I}_{d+1}$ is the $(d+1)\times(d+1)$ identity matrix. Define

$$\mathbf{S} = \begin{pmatrix} \mathbf{\Omega}_0 & \mathbf{0} \\ \mathbf{0} & \mu_2\, \mathbf{\Omega}_0 \end{pmatrix}, \qquad \mathbf{\Sigma}_\beta = \mathbf{\Omega}_0^{-1}\, \mathbf{\Sigma}_0\, \mathbf{\Omega}_0^{-1} \qquad \text{and} \qquad \mathbf{\Delta} = \begin{pmatrix} \nu_0\, \mathbf{\Sigma}_0 & \mathbf{0} \\ \mathbf{0} & \nu_2\, \mathbf{\Sigma}_0 \end{pmatrix}.$$

Now we state the asymptotic properties of both the local linear and Nadaraya-Watson estimators $\widehat{\boldsymbol{\beta}}(t)$ and $\widetilde{\boldsymbol{\beta}}(t)$ at both the interior and boundary points, respectively. All the detailed proofs are relegated to the Appendix.

Theorem 1. Under Assumptions A1 and A2, for any $t \in (0,\ 1)$, we have,

$$\mathbf{H}^{-1}\left(\widehat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t)\right) - \frac{h^2}{2}\begin{pmatrix} \mu_2 \\ 0 \end{pmatrix} \otimes \boldsymbol{\beta}''(t) + o_p(h^2) = O_p((n\,h)^{-1/2}).$$

In particular,

$$\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t) - \frac{h^2}{2}\,\mu_2\,\boldsymbol{\beta}''(t) + o_p(h^2) = O_p((n\,h)^{-1/2}) \tag{10}$$

and (10) is true for $\widetilde{\boldsymbol{\beta}}(t)$.

Theorem 2. Under Assumptions A1 - A3, for any $t \in (0,\ 1)$, we have

$$\sqrt{n\,h}\,\left\{\mathbf{H}^{-1}\left(\widehat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t)\right) - \frac{h^2}{2}\begin{pmatrix} \mu_2 \\ 0 \end{pmatrix} \otimes \boldsymbol{\beta}''(t) + o_p(h^2)\right\} \longrightarrow N\left(\mathbf{0},\ \mathbf{S}^{-1}\,\mathbf{\Delta}\,\mathbf{S}^{-1}\right)$$

In particular,

$$\sqrt{n\,h}\,\left\{\widehat{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t) - \frac{h^2}{2}\,\mu_2\,\boldsymbol{\beta}''(t) + o_p(h^2)\right\} \longrightarrow N\left(\mathbf{0},\ \nu_0\,\mathbf{\Sigma}_\beta\right) \tag{11}$$

and (11) is true for $\widetilde{\boldsymbol\beta}(t)$.

It follows from Theorems 1 and 2 that both $\widehat{\boldsymbol\beta}(t)$ and $\widetilde{\boldsymbol\beta}(t)$ share the exact same asymptotic behavior at the interior points: they are consistent estimator of $\boldsymbol\beta(t)$ with the same convergence rate and they have the same asymptotic bias and variance so that the common mean square error (AMSE) is given by

$$\text{AMSE} = \frac{h^4}{4}\,\mu_2^2\;||\boldsymbol\beta''(t)||_2^2 + \frac{\nu_0\,\text{tr}(\boldsymbol\Sigma_\beta)}{n\,h}.$$

Minimizing the AMSE gives the optimal bandwidth

$$h_{opt} = \left\{\nu_0\,\text{tr}(\boldsymbol\Sigma_\beta)\,\mu_2^{-2}\;||\boldsymbol\beta''(t)||_2^{-2}\right\}^{-1/5}\; n^{-1/5}.$$

Hence, the optimal convergence rate of the AMSE for them is of the order of $n^{-4/5}$, as one would have expected. Also, the asymptotic variance of both estimators is independent of the time point t. Moreover, it is interesting to note that if $\{\mathbf X_i\}$ are mean zero and independent of $\{u_i\}$, then $\boldsymbol\Sigma_\beta = \sum_{k=-\infty}^{\infty} r_k \begin{pmatrix} 1 & \mathbf 0 \\ \mathbf 0 & \boldsymbol\Omega_{x,0}^{-1}\,\boldsymbol\Omega_{x,k}\,\boldsymbol\Omega_{x,0}^{-1}\end{pmatrix}$, where $\boldsymbol\Omega_{x,k} = E(\mathbf X_i\,\mathbf X_{i+k}^T)$. This implies that $\widehat\beta_0(t)$ and $\widehat\beta_j(t)$ $(1 \le j \le d)$ are asymptotically independent. Further, in addition, if $\{\mathbf X_i\}$ are iid , then $\boldsymbol\Sigma_\beta$ is reduced to $\boldsymbol\Sigma_\beta = \begin{pmatrix} \sum_{k=-\infty}^{\infty} r_k & \mathbf 0 \\ \mathbf 0 & r_0\,\boldsymbol\Omega_{x,0}^{-1}\end{pmatrix}$. This implies that the asymptotic variance of $\widehat\beta_j(t)$ $(1 \le j \le d)$ depends on only the variance of $\{u_i\}$ but not the autocorrelations $(\sum_{k=1}^{\infty} r_k)$.

More importantly, Theorem 2 shows that the asymptotic variance of both estimators depends on not only the variance of the error $(\mathbf R_0 = \text{var}(u_i\,\widetilde{\mathbf X}_i))$ but also the autocorrelations $(\sum_{k=1}^{\infty}\mathbf R_k)$. This property is shared by parametric estimators but it is different from that for random design nonparametric time series regression models (see Fan and Gijbels 1996, p.17) and the functional-coefficient time series models (see Cai, Fan and Yao 2000) for which the asymptotic variance of the estimators depends on only the variance of the error. The intuitive explanation is that for the random design case, the short term dependence does not have much effect on the local smoothing method. The reason is that for any two given random variables t_i and t_j and a point t, the random variables $K_h(t_i - t)$ and $K_h(t_j - t)$ are nearly uncorrelated as $h \to 0$. For more discussions, see Fan and Gijbels (1996, p.219). Finally, from Theorem 2, it might be strange that the asymptotic biases for $\widetilde{\boldsymbol\beta}(t)$ and $\widehat{\boldsymbol\beta}(t)$ are the exact same but they are different for random design time series regression models (see Fan and Gijbels 1996, p.17) and the functional-coefficient time series models (see Cai, Fan and Yao 2000). The intuitive is that for fixed design case, there does not exist the design density $f_t(t)$ so that the extra term $h^2\,\mu_2\,\boldsymbol\beta'(t)\,f_t'(t)/f_t(t)$ (see Fan and Gijbels 1996, p.17) in the asymptotic bias expression for the random design case disappears.

A natural question arises whether two estimators would still have the same asymptotic properties at the boundary points. To answer this question, we offer the following theorems for the asymptotic results for $\widehat{\boldsymbol\beta}(t)$ and $\widetilde{\boldsymbol\beta}(t)$ at the left end point $t = c\,h$ $(0 < c < 1)$ (say) and the similar results hold for the right end point $t = 1 - c\,h$. To this purpose, define, for $k \ge 0$, $\mu_{k,c} = \int_{-c}^{\infty} u^k K(u)\,du$,

$\nu_{k,c} = \int_{-c}^{\infty} u^k \, K^2(u) \, d\, u$,

$$\mathbf{S}_c = \boldsymbol{\mu}_c \otimes \boldsymbol{\Omega}_0 \quad \text{with} \quad \boldsymbol{\mu}_c = \begin{pmatrix} \mu_{0,c} & \mu_{1,c} \\ \mu_{1,c} & \mu_{2,c} \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Delta}_c = \begin{pmatrix} \nu_{0,c} & \nu_{1,c} \\ \nu_{1,c} & \nu_{2,c} \end{pmatrix} \otimes \boldsymbol{\Sigma}_0,$$

where $\otimes$ denotes the Kronecker product.

Theorem 3. Under Assumptions A1 and A2, we have,

$$\mathbf{H}^{-1}\left(\widehat{\boldsymbol{\theta}}(c\,h) - \boldsymbol{\theta}(c\,h)\right) - \frac{h^2}{2}\,\boldsymbol{\mu}_c^{-1}\begin{pmatrix} \mu_{2,c} \\ \mu_{3,c} \end{pmatrix} \otimes \boldsymbol{\beta}''(t) + o_p(h^2) = O_p((n\,h)^{-1/2}).$$

In particular,

$$\widehat{\boldsymbol{\beta}}(c\,h) - \boldsymbol{\beta}(c\,h) - \frac{h^2}{2}\,b_c\,\boldsymbol{\beta}''(0+) + o_p(h^2) = O_p((n\,h)^{-1/2}),$$

where $b_c = (\mu_{2,c}^2 - \mu_{1,c}\,\mu_{3,c})/(\mu_{0,c}\,\mu_{2,c} - \mu_{1,c}^2)$ *and* $\boldsymbol{\beta}''(0+) = \lim_{t\downarrow 0}\boldsymbol{\beta}''(t)$*, and*

$$\widetilde{\boldsymbol{\beta}}(c\,h) - \boldsymbol{\beta}(c\,h) - h\,\mu_{0,c}^{-1}\,\mu_{1,c}\,\boldsymbol{\beta}'(0+) + o_p(h) = O_p((n\,h)^{-1/2}).$$

Theorem 4. Under Assumptions A1 - A3, we have

$$\sqrt{n\,h}\,\left\{\mathbf{H}^{-1}\left(\widehat{\boldsymbol{\theta}}(c\,h) - \boldsymbol{\theta}(c\,h)\right) - \frac{h^2}{2}\boldsymbol{\mu}_c^{-1}\begin{pmatrix} \mu_{2,c} \\ \mu_{3,c} \end{pmatrix} \otimes \boldsymbol{\beta}''(t) + o_p(h^2)\right\} \longrightarrow N\left(\mathbf{0},\ \mathbf{S}_c^{-1}\,\boldsymbol{\Delta}_c\,\mathbf{S}_c^{-1}\right).$$

In particular,

$$\sqrt{n\,h}\,\left\{\widehat{\boldsymbol{\beta}}(c\,h) - \boldsymbol{\beta}(c\,h) - \frac{h^2}{2}\,b_c\,\boldsymbol{\beta}''(0+) + o_p(h^2)\right\} \longrightarrow N\left(\mathbf{0},\ a_c\,\boldsymbol{\Sigma}_\beta\right),$$

where $a_c = (\mu_{2,c}^2\,\nu_{0,c} - 2\,\mu_{1,c}\,\mu_{2,c}\,\nu_{1,c} + \mu_{1,c}^2\,\nu_{2,c})/(\mu_{0,c}\,\mu_{2,c} - \mu_{1,c}^2)^2$*, and*

$$\sqrt{n\,h}\,\left\{\widetilde{\boldsymbol{\beta}}(ch) - \boldsymbol{\beta}(ch) - h\,\mu_{0,c}^{-1}\,\mu_{1,c}\boldsymbol{\beta}'(0+) + o_p(h)\right\} \longrightarrow N\left(\mathbf{0},\ a_c^*\,\boldsymbol{\Sigma}_\beta\right),$$

where $a_c^* = \nu_{0,c}/\mu_{0,c}^2$.

From Theorems 3 and 4, we can see that $\lim_{c\to 1} b_c = \mu_2$ and $\lim_{c\to 1} a_c = \nu_0$. More importantly, from Theorem 4, it can be seen clearly that the asymptotic biases and variances for $\widehat{\boldsymbol{\beta}}(c\,h)$ and $\widetilde{\boldsymbol{\beta}}(c\,h)$ are different. Indeed, the significant difference is that the convergence rate for the asymptotic bias of $\widetilde{\boldsymbol{\beta}}(c\,h)$ is only of the order h but not h^2, the order for $\widehat{\boldsymbol{\beta}}(c\,h)$. Based on the above discussions, it concludes that the local linear estimators do not suffer from boundary effects but the Nadaraya-Watson estimator does.

3.2 Estimation of Variance

In practice, it is desirable to have a quick and easy implementation to estimate the asymptotic variance of $\widehat{\boldsymbol{\beta}}(t)$ to construct pointwise confidence intervals. The explicit expression of the asymptotic variances in Theorems 2 and 4 provides a direct estimator. To estimate $\boldsymbol{\Sigma}_\beta$, we can construct the estimation of $\boldsymbol{\Sigma}_0$ by using the sample auto-covariances to estimate $\{\mathbf{R}_k\}$ and $\boldsymbol{\Omega}_0$ by using the method of moments.

3.3 Test for Misspecification and Stationarity

In econometrics, it is interesting to consider the following testing hypothesis

$$H_0 : \beta_j(t) = \alpha_j(t, \boldsymbol{\gamma}), \qquad 0 \le j \le d, \tag{12}$$

where $\alpha_j(t, \boldsymbol{\gamma})$ is a given family of functions indexed by unknown parameter vector $\boldsymbol{\gamma}$. This is to test whether model (2) holds with a specified parametric form, particularly, the stationary time series regression model (that is, all coefficient functions in (2) are constant) or no time trend. This kind of test problem has been considered in the literature. For example, Fan, Zhang and Zhang (2001) proposed the generalized likelihood ratio test for independent samples, Cai, Fan and Yao (2000) and Hong and Lee (1999) considered the comparison of the residual sum of squares (RSS) from both parametric and nonparametric fittings for the functional-coefficient time series regression models and used a simple version of nonparametric bootstrap to estimate the null distribution of the test statistic whereas Cai and Tiwari (2000) and Cai (2002a) proposed the F-type test statistic for the time-varying coefficient autoregressive and additive time series models, Kim (2001) utilized the Wald type test statistic for the locally stationary processes, and Juhl (2002) employed the t-ratio test for unit root for the functional coefficient models under unit root behavior.

For easy implementation purpose, here we adapt a misspecification test based on comparing the residual sum of squares from both parametric and nonparametric fittings. This method is closely related to the generalized likelihood ratio test method proposed by Fan, Zhang and Zhang (2001) who demonstrated the optimality of this kind of procedures for independent samples. The empirical work conducted by the aforementioned papers shows that the resulting testing procedure is indeed powerful and the bootstrap procedure does give the correct null distribution. This is consistent with the Wilks phenomenon observed by Fan, Zhang and Zhang (2001).

The testing method is described as follows. Let $\widehat{\boldsymbol{\gamma}}$ be an estimator of $\boldsymbol{\gamma}$ (say MLE). The RSS under the null hypothesis is $\text{RSS}_0 = n^{-1}\sum_{i=1}^n \widehat{u}_{i,0}^2$, where $u_{i,0} = Y_i - \widetilde{\mathbf{X}}_i^T\, \boldsymbol{\alpha}(t_i\, \widehat{\boldsymbol{\gamma}})$ and the RSS under H_a is $\text{RSS}_1 = n^{-1}\sum_{i=1}^n \widehat{u}_{i,1}^2$, where $\widehat{u}_{i,1} = Y_i - \widetilde{\mathbf{X}}_i^T\, \widehat{\boldsymbol{\beta}}(t_i)$. We define the test statistic is defined as $T_n = (\text{RSS}_0 - \text{RSS}_1)/\text{RSS}_1 = \text{RSS}_0/\text{RSS}_1 - 1$ and we reject the null hypothesis (12) for large value of T_n. For simplicity, we evaluate the p-value by using the following nonparametric wild bootstrap approach which can accommodate the heteroscedasticity in the model.

First, we generate the wild bootstrap residuals $\{u_i^*\}_{i=1}^n$ from the centered nonparametric residuals $\{\widehat{u}_i\}_{i=1}^n$, where $\widehat{u}_i = \widehat{u}_{i,1} - \bar{\widehat{u}}_{i,1}$ with $\bar{\widehat{u}}_{i,1} = n^{-1}\sum_{i=1}^n \widehat{u}_{i,1}$, and define the bootstrap sample $Y_i^* = \widetilde{\mathbf{X}}_i^T\, \boldsymbol{\alpha}(t_i, \widehat{\boldsymbol{\gamma}}) + u_i^*$. In practice, we can define $u_i^* = \widehat{u}_i \cdot \eta_i$, where $\{\eta_i\}$ is a sequence of iid random variables with mean zero and unit variance. See Kreiss, Neumann, and Yao (1998) for the detailed discussions. Next, we calculate the bootstrap test statistic T_n^* based on the bootstrap sample $\{(\mathbf{X}_i,\, Y_i^*)\}_{i=1}^n$ and we reject the null hypothesis H_0 when T_n is greater than the upper-α point of the conditional distribution of T_n^* given $\{(\mathbf{X}_i,\, Y_i)\}_{i=1}^n$. Finally, the p-value of the test is simply evaluated based on the relative frequency of the event $\{T_n^* \ge T_n\}$ in the replications of the bootstrap sampling.

For the sake of simplicity, we use the same bandwidth in calculating T_n^* as that in T_n. Note that we bootstrap the centralized residuals from the nonparametric fit instead of the parametric fit, because the nonparametric estimate of residuals is always consistent, no matter whether the null or the alternative hypothesis is correct. The method should provide a consistent estimator of the null distribution even when the null hypothesis does not hold. Kreiss, Neumann, and Yao (1998) considered nonparametric bootstrap tests in a general nonparametric regression setting and proved that the conditional distribution of the bootstrap test statistic is indeed asymptotically the distribution of the test statistic under the null hypothesis. It may be proven that the similar result holds here as long as $\widehat{\boldsymbol{\gamma}}$ converges to $\boldsymbol{\gamma}$ at the rate $n^{-1/2}$.

4 Monte Carlo Experiments

Throughout this section, we use the Epanechnikov kernel and the bandwidth selector proposed in Section 2.2. We illustrate the finite sample performances of the local linear estimator and test with a simulated example of the time-varying coefficient time series model. For this simulated example, the performance of the estimators is evaluated by the mean absolute deviation error (MADE):

$$\mathcal{E}_j = n_0^{-1} \sum_{k=1}^{n_0} \left| \widehat{\beta}_j(u_k) - \beta_j(u_k) \right|$$

for $\beta_j(\cdot)$, where $\{u_k,\, k = 1, \ldots, n_0\}$ are the grid points from $(0,\ 1]$.

In this simulated example, we consider the following time-varying coefficient time series model:

$$Y_i = \beta_0(i/n) + \beta_1(i/n)\, X_i + u_i, \qquad i = 1, \ldots, n,$$

where $\beta_0(x) = 0.2\, \exp(-0.7 + 3.5\, x)$, $\beta_1(x) = 2\, x + \exp(-16\, (x - 0.5)^2) - 1$, X_i is simulated from the AR(1) model $X_i = \rho_x\, X_{i-1} + \varepsilon_{1i}$ with ε_{1i} generated from $N(0,\ 2^{-2})$ independently, the error u_i is generated from the AR(1) model $u_i = \rho_u\, u_{i-1} + \varepsilon_{2i}$ with ε_{2i} generated from $N(0,\ 4^{-2})$ independently, and $\{X_i\}$ and $\{u_i\}$ are independent. The simulation is repeated 500 times for each of the sample sizes $n = 200$, 400, and 700. For the sample size $n = 400$, Figure 1(a) presents the time series plot for the time series $\{Y_i\}$ with the true trend function $\beta_0(\cdot)$ (solid line) for a typical example. The typical sample is selected in such a way that its total MADE value ($= \mathcal{E} + \mathcal{E}_0 + \mathcal{E}_1$) equals to the median in the 500 replications. We choose the optimal bandwidth $h_n = 0.27$ based on the criterion described in Section 2.2. We can compute easily $\mathbf{\Sigma}_\beta$ for this model, which is given by

$$\mathbf{\Sigma}_\beta = \begin{pmatrix} \frac{\sigma_u^2}{(1-\rho_u)^2} & 0 \\ 0 & \frac{\sigma_u^2}{1-\rho_u^2}\,\frac{1-\rho_x^2}{\sigma_x^2}\,\frac{1+\rho_u\,\rho_x}{1-\rho_u\,\rho_x} \end{pmatrix} \qquad \text{and} \qquad \mathbf{\Sigma}_\beta = \begin{pmatrix} 1.5626 & 0 \\ 0 & 0.8105 \end{pmatrix}$$

for $\rho_u = 0.8$ and $\rho_x = 0.9$. Therefore, for $\rho_u = 0.8$ and $\rho_x = 0.9$, we compute the true standard errors for $\beta_0(\cdot)$ and $\beta_1(\cdot)$ which are 0.0932 and 0.0883, respectively. Figures 1(b) and 1(c) display, respectively, the estimated $\beta_0(\cdot)$ and $\beta_1(\cdot)$ (dotted line) from the typical example, together with their true values (solid line) and the 95% confidence interval with the bias ignored (dashed lines) and they show that the estimated values are very close to the true values and all the estimated values are within the 95% confidence interval. The median and standard deviation (in parentheses) of the 500 MADE values are summarized in Table 1, which shows that all the MADE values decrease

Table 1: The median and standard deviation of the 500 *MADE values.*

	$\rho_x = 0$ and $\rho_u = 0.8$		$\rho_x = 0.9$ and $\rho_u = 0.8$	
n	$\mathcal{E}_0$	$\mathcal{E}_1$	$\mathcal{E}_0$	$\mathcal{E}_1$
200	0.122(0.047)	0.089(0.032)	0.142(0.054)	0.108(0.035)
400	0.098(0.036)	0.073(0.024)	0.102(0.039)	0.085(0.025)
700	0.081(0.027)	0.058(0.019)	0.085(0.027)	0.070(0.020)

as n increases and the values for the case that $\{\mathbf{X}_i\}$ are iid ($\rho_x = 0$) are slightly smaller than those for the dependent case ($\rho_x = 0.9$), as one would have expected. This makes the asymptotic theory more relevant. Overall, the proposed estimation procedure performs fairly well.

To demonstrate the power of the proposed misspecification test, we consider the null hypothesis $H_0 : \beta_j(u) = \theta_j$ for $j = 0,\ 1$, namely a stationary linear time series model, versus the alternative $H_a : \beta_j(u) \neq \theta_j$ for at least one j. The power function is evaluated under a family of the alternative models indexed by α, $H_a : \beta_j(u) = \theta_j + \alpha\,(\beta_j^*(u) - \theta_j)$ for $j = 0$, 1 and $0 \leq \alpha \leq 1$, where $\{\beta_j^*(u)\}$ are the solid curves given in Figures 1(b) and 1(c) and θ_j is the average height of $\beta_j^*(u)$ (indeed, $\theta_0 = 0.991$ and $\theta_1 = 0.441$). The other type of tests can be considered in a same way. For the sample size $n = 400$, we apply the misspecification test described in Section 3.3 in a simulation with 500 replications and we repeat the bootstrap sampling 1000 times for each realization. Figure 1(d) plots the simulated power function against α. When $\alpha = 0$, the specified alternative hypothesis collapses into the null hypothesis. The power is 0.048, which is close to the significance level of 5%. This demonstrates that the bootstrap estimate of the null distribution is approximately correct. The power function shows that our test is indeed powerful. To appreciate why, consider the specific alternative with $\alpha = 0.175$. The functions $\{\beta_j(u)\}$ under H_a are shown in Figure 2 (solid lines). The null hypothesis is essentially the constant curves (dotted lines) in Figure 2. Even with such a small difference under our noise level, we can correctly detect the alternative over 81.4% among the 500 simulations. The power increases rapidly to 1 when $\alpha = 0.325$. When $\alpha = 1$, we test the constant functions in Figure 2 against the coefficient functions in Figures 1(b) and 1(c).

5 Discussions

In this paper we developed a useful class of time series models, the time-varying coefficient time series models with time trend and serially correlated errors, for modeling nonlinear, nonstationary and trending time series. We developed nonparametric methods for estimating the trend function and coefficient functions and studied their asymptotic properties at both the interior and boundary points. We obtained some insights about the modeling methods and we demonstrated that the local linear estimator is superior than the Nadaraya-Watson estimator. Also, the usefulness of the models was demonstrated by a simulated example. To make the model practically useful, we proposed an easily implemented bandwidth selector and a new testing procedure to test the misspecification and stationarity, based on the comparison of the residual sum of squares and suggested using a wild bootstrap to estimate the p-value. However, the models considered here did not allow the $\mathbf{X}_i$ to contain lagged Y_i, a major drawback in view of some econometric applications. It is not difficult to see that although the proposed modeling procedures can be used when there are lagged variables, the difficulty is to establish the asymptotic theory under reasonably attractive and primitive conditions. Also, in some applications, it is likely that a semi-time-varying coefficient time series model might be suitable. For these models, a local linear modeling technique might still be applicable. See Zhang, Lee and Song (2002) for the detailed discussions. Extensions to other models in the regression family are apparent. Indeed, many of the modeling techniques and asymptotic theory developed in this paper are relevant to a similar analysis of more general models, that are not necessarily of regression type (say, instrument variable type). Finally, the predictive utility of using the time-varying coefficient time series models studied in this paper needs definitely a further investigation due to its importance in various applications in economics and finance.

APPENDIX: PROOFS

Throughout this appendix, we use the same notation as used in Sections 2 and 3 and we denote by C a generic constant, which may take different values at different appearances.

Lemma 1. *Let $r_{jm}(i)$ denote the (j,m)-th element of $\mathbf{R}_i$. If Assumption A2 is satisfied, then, $\sum_{i=-\infty}^{\infty} |r_{jm}(i)| < \infty$ so that $\mathbf{\Sigma}_0$ exists.*

Proof: By the Davydov's inequality (see, e.g., Corollary A.2 in Hall and Heyde 1980),

$$r_{jm}(k) = \operatorname{cov}(\operatorname{cov}(u_i\, \widetilde{X}_{ij},\, u_{i+k}\, \widetilde{X}_{(i+k)m}) \le C\, \alpha^{\delta/(2+\delta)}(k)\, \left\{E|u_i\, \widetilde{X}_{ij}|^{2+\delta}\right\}^{2/(2+\delta)} \le C\, \alpha^{\delta/(2+\delta}(k)$$

so that $\sum_k |r_{jm}(k)| < \infty$. □

To prove the theorems, we first define $\mathbf{M}(t_i) = \boldsymbol{\beta}(t_i) - \left\{\boldsymbol{\beta}(t) + (t_i - t)\, \boldsymbol{\beta}'(t) + \frac{1}{2}\, (t_i - t)^2\, \boldsymbol{\beta}''(t)\right\}$, $\mathbf{T}^*_{n,k}(t) = n^{-1}\, \sum_{i=1}^n \widetilde{\mathbf{X}}_i\, (t_i - t)^k\, K_h(t_i - t)\, u_i$, $\mathbf{R}_{n,k}(t) = n^{-1}\, \sum_{i=1}^n \widetilde{\mathbf{X}}_i\, \widetilde{\mathbf{X}}_i^T\, (t_i - t)^k\, K_h(t_i - t)\, \mathbf{M}(t_i)$, and $\mathbf{B}_{n,k}(t) = \frac{1}{2}\, \mathbf{S}_{n,k+2}(t)\, \boldsymbol{\beta}''(t)$. Then,

$$\mathbf{T}_{n,k}(t) = \mathbf{S}_{n,k}(t)\, \boldsymbol{\beta}(t) + \mathbf{S}_{n,k+1}(t)\, \boldsymbol{\beta}'(t) + \mathbf{T}^*_{n,k}(t) + \mathbf{B}_{n,k}(t) + \mathbf{R}_{n,k}(t).$$

In conjunction with (4) and (6), we have

$$\widehat{\boldsymbol{\theta}}(t) - \boldsymbol{\theta}(t) - \mathbf{S}_n^{-1}(t)\, \mathbf{B}_n(t) - \mathbf{S}_n^{-1}(t)\, \mathbf{R}_n(t) = \mathbf{S}_n^{-1}(t)\, \mathbf{T}^*_n(t) \qquad \text{(A.1)}$$

and

$$\widetilde{\boldsymbol{\beta}}(t) - \boldsymbol{\beta}(t) - \mathbf{S}_{n,0}^{-1}(t)\, \left\{\mathbf{B}_{n,0}(t) + \mathbf{S}_{n,1}(t)\, \boldsymbol{\beta}'(t)\right\} - \mathbf{S}_{n,0}^{-1}(t)\, \mathbf{R}_{n,0}(t) = \mathbf{S}_{n,0}^{-1}(t)\, \mathbf{T}^*_{n,0}(t). \qquad \text{(A.2)}$$

Lemma 2. *Under Assumptions A1 and A2, for $t \in (0,\ 1)$, we have*

$$h^{-k}\,\mathbf{S}_{n,k}(t) = \mu_k\,\mathbf{\Omega}_0\,\{1+o_p(1)\}, \qquad \text{and} \qquad h^{-k}\,\mathbf{S}_n(c\,h) = \mu_{k,c}\,\mathbf{\Omega}_0\,\{1+o_p(1)\}.$$

Also, for either $t \in (0,\ 1)$ or $t = c\,h$, $h^{-k}\,\mathbf{R}_{n,k}(t) = o_p(h^2)$.

Proof: It follows by the Riemann sum approximation of an integral that for $0 \le k \le 2$,

$$\begin{aligned}
h^{-k}\,E(\mathbf{S}_{n,k}(t)) &= n^{-1}\sum_{i=1}^n E\left\{\widetilde{\mathbf{X}}_i\widetilde{\mathbf{X}}_i^T\right\}\left(\frac{t_i-t}{h}\right)^k\;K_h(t_i-t)\\
&= \mathbf{\Omega}_0\,n^{-1}\sum_{i=1}^n\left(\frac{t_i-t}{h}\right)^k\;K_h(t_i-t)\\
&\approx \mathbf{\Omega}_0\int_0^1\left(\frac{u-t}{h}\right)^k\;K_h(u-t)\,du\\
&= \mathbf{\Omega}_0\int_{-t/h}^{(1-t)/h} u^k\;K(u)\,du\\
&\approx \begin{cases}\mathbf{\Omega}_0\;\mu_k & \text{if } t\in(0,\ 1),\\ \mathbf{\Omega}_0\;\mu_{k,c} & \text{if } t = c\,h.\end{cases}
\end{aligned}$$

Let η_{jm} denote the (j,m)-th element of $h^{-k}\,\mathbf{S}_{n,k}(t)$ for either $t\in(0,\ 1)$ or $t = c\,h$. That is,

$$\eta_{jm} = n^{-1}\sum_{i=1}^n \widetilde{X}_{ij}\,\widetilde{X}_{im}\left(\frac{t_i-t}{h}\right)^k\;K_h(t_i-t).$$

Then, consider the variance of η_{jm}

$$\begin{aligned}
\text{var}(\eta_{jm}) &= n^{-2}\sum_{i=1}^n \text{var}\left\{\widetilde{X}_{ij}\,\widetilde{X}_{im}\right\}\left(\frac{t_i-t}{h}\right)^{2k}\;K_h^2(t_i-t)\\
&\quad +2\,n^{-2}\sum_{1\le i<l\le n}\text{cov}(\widetilde{X}_{ij}\widetilde{X}_{im},\widetilde{X}_{lj}\widetilde{X}_{lm})\left(\frac{t_i-t}{h}\right)^k K_h(t_i-t)\left(\frac{t_{nl}-t}{h}\right)^k K_h(t_{nl}-t)\\
&\equiv I_1 + I_2.
\end{aligned}$$

Obviously,

$$I_1 \approx (n\,h)^{-1}\,\text{var}\left\{\widetilde{X}_{ij}\,\widetilde{X}_{im}\right\}\int_{-t/h}^{(1-t)/h} u^{2k}\,K^2(u)du \le C\,(n\,h)^{-1} = o(1).$$

Since $u^2\,K(u)$ is bounded, then, for any $0\le k\le 2$, $|u/h|^k\,K_h(u)\le C/h$ and $n^{-1}\sum_{i=1}^n|(t_i-t)/h|^k K_h(t_i-t)\le C$ by the Riemann sum approximation of an integral. In conjunction with the Davydov's inequality (see, e.g., Corollary A.2 in Hall and Heyde 1980), we have,

$$|I_2| \;\le\; C\,n^{-2}\sum_{1\le i<l\le n}|\text{cov}(\widetilde{X}_{ij}\widetilde{X}_{im},\widetilde{X}_{lj}\widetilde{X}_{lm})|\left|\frac{t_i-t}{h}\right|^k K_h(t_i-t)\left|\frac{t_{nl}-t}{h}\right|^k K_h(t_{nl}-t)$$

$$\begin{aligned}
&\le C\, n^{-2} \sum_{1\le i<l\le n} \alpha^{\delta/(2+\delta)}(l-i) \left|\frac{t_i-t}{h}\right|^k K_h(t_i-t) \left|\frac{t_{nl}-t}{h}\right|^k K_h(t_{nl}-t) \\
&\le C\,(n\,h)^{-1}\, n^{-1} \sum_{i=1}^{n} \left|\frac{t_i-t}{h}\right|^k K_h(t_i-t) \sum_{l>1} \alpha^{\delta/(2+\delta)}(l) \\
&\le C\,(n\,h)^{-1} \sum_{l>1} \alpha^{\delta/(2+\delta)}(l) \le C\,(n\,h)^{-1} \to 0
\end{aligned}$$

by Assumption A2. Similarly,

$$E\left\{h^{-(2+k)}\, \mathbf{R}_{n,k}(t)\right\} \approx \mathbf{\Omega}_0 \int_{-t/h}^{(1-t)/h} u^k\, K(u)\, h^{-2}\, \mathbf{M}(t+h\,u)du \to \mathbf{0}$$

by the fact that $h^{-2}\,\mathbf{M}(t+h\,u)\to\mathbf{0}$ for any u and

$$\mathrm{var}\left\{h^{-(2+k)}\, \mathbf{R}_{n,k}(t)\right\} \to \mathbf{0}.$$

This proves the lemma. □

Lemma 3. *Under Assumptions A1 and A2, for $t\in(0,\ 1)$, we have*

$$n\,h\,var\Big(\mathbf{H}^{-1}\,\mathbf{T}_n^*(t)\Big) = \mathbf{\Delta}+o(1) \quad \textit{and} \quad n\,h\,var\Big(\mathbf{H}^{-1}\,\mathbf{T}_n^*(c\,h)\Big) = \mathbf{\Delta}_c+o(1)$$

Proof: By the stationarity of $\{u_i\,\widetilde{\mathbf{X}}_i\}$,

$$\begin{aligned}
n\,h\,\mathrm{var}(\mathbf{T}_{n,0}^*(t)) &= n^{-1}\,h \sum_{1\le k,\,l\le n} \mathbf{R}_{k-l}\, K_h(t_{nk}-t)\, K_h(t_{nl}-t) \\
&= n^{-1}\,h\,\mathbf{R}_0 \sum_{k=1}^{n} K_h^2(t_{nk}-t) + 2\,n^{-1}\,h \sum_{1\le l<k\le n} \mathbf{R}_{k-l}\, K_h(t_{nk}-t)\, K_h(t_{nl}-t) \\
&\equiv \mathbf{I}_3+\mathbf{I}_4.
\end{aligned}$$

Clearly, by the Riemann sum approximation of an integral,

$$\mathbf{I}_3 \approx \mathbf{R}_0\, h \int_0^1 K_h^2(u-t)du = \mathbf{R}_0 \int_{-t/h}^{(1-t)/h} K^2(u)du \approx \begin{cases} \nu_0\,\mathbf{R}_0 & \text{if } t\in(0,\ 1), \\ \nu_{0,c}\,\mathbf{R}_0 & \text{if } t=c\,h. \end{cases}$$

Since $n\,h\to\infty$, there exists $d_n\to\infty$ such that $d_n/(n\,h)\to 0$. Let $S_1=\{(k,l):1\le k-l\le d_n;\ 1\le l<k\le n\}$ and $S_2=\{(k,l):\ 1\le l<k\le n\}-S_1$. Then, $\mathbf{I}_4$ is split into two terms as $\sum_{S_1}(\cdots)$, denoted by $\mathbf{I}_{41}$, and $\sum_{S_2}(\cdots)$, denoted by $\mathbf{I}_{42}$. Since $K(\cdot)$ is bounded, then, $K_h(\cdot)\le C/h$ and $n^{-1}\sum_{k=1}^{n} K_h(t_{nk}-t)\le C$. In conjunction with the Davydov's inequality (see, e.g., Corollary A.2 in Hall and Heyde 1980), we have, for the (j,m)-th element of $\mathbf{I}_{42}$,

$$\begin{aligned}
|\mathbf{I}_{42(jm)}| &\le C\, n^{-1} h \sum_{S_2} |r_{jm}(k-l)|\, K_h(t_{nk}-t)\, K_h(t_{nl}-t) \\
&\le C\, n^{-1}\, h \sum_{S_2} \alpha^{\delta/(2+\delta)}(k-l) K_h(t_{nk}-t)\, K_h(t_{nl}-t)
\end{aligned}$$

$$
\begin{aligned}
&\le C\, n^{-1} \sum_{k=1}^{n} K_h(t_{nk}-t) \sum_{k_1>d_n} \alpha^{\delta/(2+\delta)}(k_1) \\
&\le C \sum_{k_1>d_n} \alpha^{\delta/(2+\delta)}(k_1) \\
&\le C\, c_n^{-\delta} \to 0
\end{aligned}
$$

by Assumption A2 and the fact that $d_n \to \infty$. For any $(k,l) \in S_1$, by Assumption A1

$$
|K_h(t_{nk}-t) - K_h(t_{nl}-t)| \le C\, h^{-1}\, (t_{nk}-t_{nl})/h \le C\, c_n/(n\, h^2),
$$

which implies that

$$
\begin{aligned}
|\mathbf{I}_{412(jm)}| \;&\equiv\; \left| 2\, n^{-1}\, h \sum_{l=1}^{n-1} \sum_{1\le k-l\le c_n} r_{jm}(k-l)\, \{K_h(t_{nk}-t) - K_h(t_{nl}-t)\}\, K_h(t_{nl}-t) \right| \\
&\le C\, d_n\, n^{-2}\, h^{-1} \sum_{l=1}^{n-1} \sum_{1\le k-l\le d_n} |r_{jm}(k-l)|\, K_h(t_{nl}-t) \\
&\le C\, d_n\, n^{-2}\, h^{-1} \sum_{l=1}^{n-1} K_h(t_{nl}-t) \sum_{k\ge 1} |r_{jm}(k)| \\
&\le C\, d_n/(n\, h) \to 0
\end{aligned}
$$

by Lemma 1 and the fact that $d_n/(n\, h) \to 0$. Therefore,

$$
\begin{aligned}
\mathbf{I}_{41(jm)} &= 2\, n^{-1}\, h \sum_{l=1}^{n-1} \sum_{1\le k-l\le d_n} r_{jm}(k-l)\, K_h(t_{nk}-t)\, K_h(t_{nl}-t) \\
&= 2\, n^{-1}\, h \sum_{l=1}^{n-1} K_h^2(t_{nl}-t) \sum_{1\le k-l\le d_n} r_{jm}(k-l) + \mathbf{I}_{412(jm)} \\
&\to 2 \sum_{k=1}^{\infty} r_{jm}(k) \begin{cases} \nu_0 & \text{if } t \in (0,\ 1), \\ \nu_{0,c} & \text{if } t = c\, h. \end{cases}
\end{aligned}
$$

Thus,

$$
n\, h\, \mathrm{var}(\mathbf{T}^*_{n,0}(t)) \to \left(\mathbf{R}_0 + 2 \sum_{k=1}^{\infty} \mathbf{R}_k \right) \begin{cases} \nu_0 & \text{if } t \in (0,\ 1), \\ \nu_{0,c} & \text{if } t = c\, h. \end{cases}
$$

Similarly

$$
\begin{aligned}
n\, h\, \mathrm{var}\left(h^{-1} \mathbf{T}^*_{n,1}(t) \right) &= n^{-1}\, h \sum_{1\le k,\, l\le n} \mathbf{R}_{k-l} \left(\frac{t_{nk}-t}{h} \right) \left(\frac{t_{nl}-t}{h} \right) K_h(t_{nk}-t)\, K_h(t_{nl}-t) \\
&\to \mathbf{\Sigma}_0 \begin{cases} \nu_2 & \text{if } t \in (0,\ 1), \\ \nu_{2,c} & \text{if } t = c\, h. \end{cases}
\end{aligned}
$$

and

$$
\begin{aligned}
n\, h\, \mathrm{cov}\left(\mathbf{T}^*_{n,0}(t),\ h^{-1} \mathbf{T}^*_{n,1}(t) \right) &= n^{-1}\, h \sum_{1\le k,\, l\le n} \mathbf{R}_{k-l} \left(\frac{t_{nl}-t}{h} \right) K_h(t_{nk}-t)\, K_h(t_{nl}-t) \\
&\to \mathbf{\Sigma}_0 \begin{cases} \nu_1 & \text{if } t \in (0,\ 1), \\ \nu_{1,c} & \text{if } t = c\, h. \end{cases} .
\end{aligned}
$$

This proves the lemma. $\square$

Proofs of Theorems 1 - 4: First, it is easy to see from Lemma 2 that

$$\mathbf{H}^{-1}\,\mathbf{B}_n(t) = \frac{h^2}{2}\begin{pmatrix}\mu_2\,\boldsymbol{\Omega}_0\\ \mathbf{0}\end{pmatrix}\otimes\boldsymbol{\beta}''(t)+o_p(h^2)$$

for any $t\in(0,\ 1)$, and

$$\mathbf{H}^{-1}\,\mathbf{B}_n(c\,h) = \frac{h^2}{2}\begin{pmatrix}\mu_{2,c}\,\boldsymbol{\Omega}_0\\ \mu_{3,c}\,\boldsymbol{\Omega}_0\end{pmatrix}\otimes\boldsymbol{\beta}''(0+)+o_p(h^2).$$

Therefore, it follows from (A.1) and Lemmas 2 and 3 that

$$\begin{aligned}
&\sqrt{n\,h}\,\left\{\mathbf{H}\left(\widehat{\boldsymbol{\theta}}(t)-\boldsymbol{\theta}(t)\right)-\frac{h^2}{2}\begin{pmatrix}\mu_2\\ 0\end{pmatrix}\otimes\boldsymbol{\beta}''(t)+o_p(h^2)\right\}\\
=&\ \left(\mathbf{H}^{-1}\,\mathbf{S}_n(t)\,\mathbf{H}^{-1}\right)^{-1}\ \sqrt{n\,h}\ \mathbf{H}^{-1}\,\mathbf{T}_n^*(t)\\
=&\ \mathbf{S}^{-1}\,\sqrt{n\,h}\,\mathbf{H}^{-1}\,\mathbf{T}_n^*(t)\{1+o_p(1)\} \qquad\qquad (\text{A.3})\\
=&\ O_p(1),
\end{aligned}$$

for any $t\in(0,\ 1)$, and

$$\begin{aligned}
&\sqrt{n\,h}\,\left\{\mathbf{H}\left(\widehat{\boldsymbol{\theta}}(c\,h)-\boldsymbol{\theta}(c\,h)\right)-\frac{h^2}{2}\,\boldsymbol{\mu}_c^{-1}\begin{pmatrix}\mu_{2,c}\\ \mu_{3,c}\end{pmatrix}\otimes\boldsymbol{\beta}''(0+)+o_p(h^2)\right\}\\
=&\ \left(\mathbf{H}^{-1}\,\mathbf{S}_n(c\,h)\,\mathbf{H}^{-1}\right)^{-1}\ \sqrt{n\,h}\ \mathbf{H}^{-1}\,\mathbf{T}_n^*(c\,h)\\
=&\ \mathbf{S}_c^{-1}\,\sqrt{n\,h}\ \mathbf{H}^{-1}\,\mathbf{T}_n^*(c\,h)\{1+o_p(1)\} \qquad\qquad (\text{A.4})\\
=&\ O_p(1),
\end{aligned}$$

which establish the consistency with a convergence rate for $\widehat{\boldsymbol{\theta}}(t)$ at both interior and boundary points, which is the first part of Theorems 1 and 3. Next,

$$\mathbf{S}_{n,0}^{-1}(t)\ \{\mathbf{B}_{n,0}(t)+\mathbf{S}_{n,1}(t)\,\boldsymbol{\beta}'(t)\} = \frac{h^2}{2}\,\mu_2\,\boldsymbol{\Omega}_0\,\boldsymbol{\beta}''(t)\{1+o_p(1)\}$$

for any $t\in(0,\ 1)$, and

$$\mathbf{S}_{n,0}^{-1}(c\,h)\ \{\mathbf{B}_{n,0}(c\,h)+\mathbf{S}_{n,1}(c\,h)\,\boldsymbol{\beta}'(c\,h)\} = \mu_{0,c}^{-1}\left\{\frac{h^2}{2}\,\mu_{2,c}\,\boldsymbol{\beta}''(0+)+h\,\mu_{1,c}\,\boldsymbol{\beta}'(0+)\right\}\{1+o_p(1)\}.$$

Then, by (A.2) and Lemmas 2 and 3, we have,

$$\begin{aligned}
&\sqrt{n\,h}\,\left\{\widetilde{\boldsymbol{\beta}}(t)-\boldsymbol{\beta}(t)-\frac{h^2}{2}\,\mu_2\,\boldsymbol{\beta}''(t)+o_p(h^2)\right\} = \mathbf{S}_{n,0}^{-1}(t)\ \sqrt{n\,h}\ \mathbf{T}_{n,0}^*(t)\\
=&\ \boldsymbol{\Omega}_0^{-1}\,\sqrt{n\,h}\,\mathbf{T}_{n,0}^*(t)\{1+o_p(1)\} \qquad\qquad (\text{A.5})\\
=&\ O_p(1)
\end{aligned}$$

for any $t \in (0,\ 1)$, and

$$\begin{aligned}
& \sqrt{n\,h}\ \left\{\widetilde{\boldsymbol{\beta}}(c\,h) - \boldsymbol{\beta}(c\,h) - \mu_{0,c}^{-1}\left(\frac{h^2}{2}\,\mu_{2,c}\,\boldsymbol{\beta}''(0+) + h\,\mu_{1,c}\,\boldsymbol{\beta}'(0+)\right) + o_p(h^2)\right\} \\
= \ & \mathbf{S}_{n,0}^{-1}(c\,h)\ \sqrt{n\,h}\ \mathbf{T}_{n,0}^*(c\,h) = \mu_{0,c}^{-1}\ \boldsymbol{\Omega}_0^{-1}\ \sqrt{n\,h}\ \mathbf{T}_{n,0}^*(c\,h)\{1 + o_p(1)\} \qquad \text{(A.6)} \\
= \ & O_p(1).
\end{aligned}$$

Therefore, this provides the consistency with a convergence rate for $\widetilde{\boldsymbol{\beta}}(c\,h)$ at both interior and boundary points, which is the second part of Theorems 1 and 3.

To establish the asymptotic normality for $\widehat{\boldsymbol{\theta}}(t)$ and $\widetilde{\boldsymbol{\beta}}(t)$ at both the interior and boundary points, it suffices to show that

$$\sqrt{n\,h}\,\mathbf{H}^{-1}\,\mathbf{T}_n^*(t) \ \longrightarrow\ N(\mathbf{0},\ \boldsymbol{\Delta}) \qquad \text{and} \qquad \sqrt{n\,h}\,\mathbf{H}^{-1}\,\mathbf{T}_n^*(c\,h) \ \longrightarrow\ N(\mathbf{0},\ \boldsymbol{\Delta}_c) \qquad \text{(A.7)}$$

from (A.4) and (A.5), and that

$$\sqrt{n\,h}\ \mathbf{T}_{n,0}^*(t) \ \longrightarrow\ N(\mathbf{0},\ \nu_0\,\boldsymbol{\Sigma}_0) \qquad \text{and} \qquad \sqrt{n\,h}\ \mathbf{T}_{n,0}^*(c\,h) \ \longrightarrow\ N(\mathbf{0},\ a_c^*\,\boldsymbol{\Sigma}_0) \qquad \text{(A.8)}$$

from (A.6) and (A.7). Since the proof of (A.8) is simpler than that for (A.7), next we will present only the proof of the first result in (A.7).

To prove (A.7), we use the Cramér-Wold device. That is to show that for any unit vector $\mathbf{d}$ in $\Re^{2(d+1)}$,

$$\sqrt{n\,h}\ \mathbf{d}^T\ \mathbf{H}^{-1}\ \mathbf{T}_n^*(t) \ \longrightarrow\ N(\mathbf{0},\ \mathbf{d}^T\,\boldsymbol{\Delta}\,\mathbf{d}).$$

To this end, let $Z_{n,i} = n^{-1/2}\ h^{1/2}\ \mathbf{d}^T\ \mathbf{H}^{-1}\ \mathbf{Z}_i\,u_i\,K_h(t_i - t)$. Then, $\sqrt{n\,h}\ \mathbf{d}^T\ \mathbf{H}^{-1}\ \mathbf{T}_n^*(t) = \sum_{i=1}^n Z_{n,i}$ and by Lemma 3,

$$\operatorname{var}\left(\sum_{i=1}^n Z_{n,i}\right) = \mathbf{d}^T\,\boldsymbol{\Delta}\,\mathbf{d}\ \{1 + o(1)\} \equiv \theta_d^2\ \{1 + o(1)\}.$$

Next we use the Doob's small-block and large-block technique. Namely, partition $\{1,\ \ldots,\ n\}$ into $2\,q_n+1$ subsets with large-block of size $r_n = \left\lfloor (n\,h)^{1/2} \right\rfloor$ and small-block of size $s_n = \left\lfloor (n\,h)^{1/2}/\log n \right\rfloor$, where $q_n = \lfloor n/(r_n + s_n) \rfloor$. Let $r_j^* = j\,(r_n + s_n)$ and define the random variables, for $0 \le j \le q_n - 1$,

$$\eta_j = \sum_{i=r_j^*+1}^{r_j^*+r_n} Z_{n,i}, \qquad \zeta_j = \sum_{i=r_j^*+r_n+1}^{r_{j+1}^*} Z_{n,i}, \qquad \text{and} \qquad Q_{n,3} = \sum_{i=r_{q_n}^*+1}^{n} Z_{n,i}.$$

Then, $\sqrt{n\,h}\ \mathbf{d}^T\ \mathbf{H}^{-1}\,\mathbf{T}_n^*(t) = Q_{n,1} + Q_{n,2} + Q_{n,3}$, where $Q_{n,1} = \sum_{j=0}^{q_n-1} \eta_j$ and $Q_{n,2} = \sum_{j=0}^{q_n-1} \zeta_j$. Next we prove the followings: as $n \to \infty$,

$$E(Q_{n,2})^2 \to 0, \qquad\qquad E(Q_{n,3})^2 \to 0, \qquad \text{(A.9)}$$

$$\left| E\left[\exp(i\,s\,Q_{n,1})\right] - \prod_{j=0}^{q_n-1} E\left[\exp(i\,s\,\eta_j)\right] \right| \to 0 \qquad \text{(A.10)}$$

for any s and

$$\mathrm{var}(Q_{n,1}) \to \theta_d^2 \qquad \text{and} \qquad \sum_{j=0}^{q_n-1} E|\eta_j|^{2+\delta} \to 0. \tag{A.11}$$

(A.9) implies that $Q_{n,2}$ and $Q_{n,3}$ are asymptotically negligible in probability. (A.10) shows that the summands $\{\eta_j\}$ in $Q_{n,1}$ are asymptotically independent, and (A.11) is the standard Lindeberg-Feller and Lyapounov conditions for asymptotic normality of $Q_{n,1}$ for the independent setup.

Next, we verify the above four equations (A.9) - (A.11). First, to establish (A.10), we make use of Lemma 1.1 in Volkonskii and Rozanov (1959) (see also Ibragimov and Linnik 1971, p.338) to obtain

$$\left| E\left[\exp\left(i\, s\, Q_{n,1}\right)\right] - \prod_{j=0}^{q_n-1} E\left[\exp(i\, s\, \eta_j)\right] \right| \le 16\, q_n\, \alpha(s_n) \le C\; n^{-(\tau-1)/2}\, h^{-(\tau+1)/2}\, \log^\tau n \to 0$$

by Assumption A3. To prove (A.9), we observe that

$$E(Q_{n,2})^2 = \sum_{0\le i,j\le q_n-1}\ \sum_{1\le k_1,k_2\le s_n} \mathrm{cov}(Z_{n,r_i^*+r_n+k_1},\, Z_{n,r_j^*+r_n+k_2}) \equiv F_1 + F_2 + F_3 + F_4,$$

where with $\mathbf{d}^T = (\mathbf{d}_1^T, \mathbf{d}_2^T)$ and $S = \{((i,j,k_1,k_2) : 0 \le i,j \le q_n - 1,\ 1 \le k_1, k_2 \le s_n\}$,

$$F_1 = n^{-1}\, h\, \sum_S \mathbf{d}_1^T\ \mathbf{R}_{r_i^*+k_1-r_j^*-k_2}\, \mathbf{d}_1\, K_h(t_{n,r_i^*+r_n+k_1} - t) K_h(t_{n,r_j^*+r_n+k_2} - t),$$

$$F_2 = n^{-1}\, h\, \sum_S \mathbf{d}_1^T\ \mathbf{R}_{r_i^*+k_1-r_j^*-k_2}\, \mathbf{d}_2 \left(\frac{t_{n,r_j^*+r_n+k_2} - t}{h}\right) K_h(t_{n,r_i^*+r_n+k_1} - t) K_h(t_{n,r_j^*+r_n+k_2} - t),$$

$$F_3 = n^{-1}\, h\, \sum_S \mathbf{d}_1^T\ \mathbf{R}_{r_i^*+k_1-r_j^*-k_2}\, \mathbf{d}_2 \left(\frac{t_{n,r_i^*+r_n+k_1} - t}{h}\right) K_h(t_{n,r_i^*+r_n+k_1} - t) K_h(t_{n,r_j^*+r_n+k_2} - t),$$

and

$$\begin{aligned} F_4 \quad &= \quad n^{-1}\, h\, \sum_S \mathbf{d}_2^T\ \mathbf{R}_{r_i^*+k_1-r_j^*-k_2}\, \mathbf{d}_2 \left(\frac{t_{n,r_i^*+r_n+k_1} - t}{h}\right) \left(\frac{t_{n,r_j^*+r_n+k_2} - t}{h}\right) \\ &\quad\ \times K_h(t_{n,r_i^*+r_n+k_1} - t) K_h(t_{n,r_j^*+r_n+k_2} - t). \end{aligned}$$

By Assumption A1,

$$|F_1| \le C\ (n\, h)^{-1} \sum_{0\le i,j\le q_n-1}\ \sum_{1\le k_1,k_2\le s_n} |\mathbf{d}_1^T\ \mathbf{R}_{r_i^*+k_1-r_j^*-k_2}\, \mathbf{d}_1| \le C\ q_s\ s_n\ (n\, h)^{-1} \to 0.$$

Similarly, one can show that $F_k \to 0$ for $2 \le k \le 4$, so that $E(Q_{n,2})^2 \to 0$. By the same arguments, we have

$$\mathrm{var}(Q_{n,3}) = \sum_{r_q^*+1\le j_1,j_2\le n} \mathrm{cov}(Z_{n,j_1},\, Z_{n,j_2}) \le C\ (n - r_{q_n}^*)\, (n\, h)^{-1} \le C\ r_n\, (n\, h)^{-1} \to 0.$$

This proves (A.9). An application of (A.9) gives the first assertion in (A.11). It remains to establish the second assertion in (A.11). To this end, it follows from Theorem 4.1 of Shao and Yu (1996) that

$$\begin{aligned} E|\eta_j|^{2+\delta} &\leq C\,(n^{-1}\,h\,r_n)^{(2+\delta)/2}\,\max_i\left(E\left|\mathbf{d}^T\,\mathbf{H}^{-1}\,\mathbf{Z}_i\,u_i\,K_h(t_i-t)\right|^{2(1+\delta)}\right)^{(2+\delta)/2(1+\delta)} \\ &\leq C\,(n\,h)^{-(2+\delta)/4}. \end{aligned}$$

Then,

$$\sum_{j=0}^{q_n-1} E|\eta_j|^{2+\delta} \leq C\,q_n\,(n\,h)^{-(2+\delta)/4} \leq C\,(n\,h^{1+4/\delta})^{-\delta/4} \to 0$$

by Assumption A3. This proves (A.7). □

REFERENCES

Auestad, B. and Tjøstheim, D. (1990). Identification of nonlinear time series: First order characterization and order determination. *Biometrika*, **77**, 669-687.

Brockwell, P.J. and Davis, R.A. (1991). *Time Series Theory and Methods*. New York: Springer.

Cai, Z. (2002a). A two-stage approach to additive time series models. *Statistica Neerlandica*, **56**, No.4, 1-19.

Cai, Z. (2002b). Varying coefficient dynamic panel data models. *Working Paper*, Department of Mathematics, University of North Carolina at Charlotte.

Cai, Z. and Chen, R. (2002). Nonlinear seasonal time series models. Revised for *Journal of the American Statistical Association*.

Cai, Z., M. Das, H. Xiong and X. Wu (2002). Functional-coefficient instrumental variables models. *Working Paper*, Department of Economics, Columbia University.

Cai, Z., Fan, J. and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, **95**, 941-956.

Cai, Z. and Masry, E. (2000). Nonparametric estimation in nonlinear ARX time series models: Projection and linear fitting. *Econometric Theory*, **16**, 465-501.

Cai, Z. and Ould-Said, E. (2001). Local robust regression estimation for time series. *Submitted for publication*.

Cai, Z. and Tiwari, R.C. (2000). Application of a local linear autoregressive model to BOD time series. *Environmetrics*, **11**, 341-350.

Cai, Z. and Wu, H. (2002). Local quasi-likelihood method for generalized random curve models with longitudinal data. *Submitted for publication*.

Chang, Y., Park, J.Y. and Phillips, P.C.B. (2001). Nonlinear econometric models with cointegrated and deterministically trending regressors. *Econometrics Journal*, **4**, 1-36.

Chen, R. and Tsay, R.S. (1993). Functional-coefficient autoregressive models. *Journal of the American Statistical Association*, **88**, 298-308.

Engle, R.F., Granger, C.W.J., Rice, J. and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of The American Statistical Association*, **81**, 310-320.

Fan, J. (1993). Local linear smoothers and their minimax efficiency. *The Annals of Statistics*, **21**, 196-216.

Fan, J., Heckman, N.E. and Wand, M.P. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *Journal of the American Statistical Association*, **90**, 141-150.

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

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

Franses, P.H. (1996). *Periodicity and Stochastic Trends in Economic Time Series*. New York: Cambridge University Press.

Franses, P.H. (1998). *Time Series Models for Business and Economic Forecasting*. New York: Cambridge University Press.

Gorodetskii, V.V. (1977). On the strong mixing property for linear sequences. *Theory of Probability and Its Applications*, **22**, 411-413.

Granger, C.L.J. and Teräsvirta, T. (1993). *Modelling Nonlinear Economic Relationships*. Oxford: Oxford University Press.

Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and its Applications*. New York: Academic Press.

Härdle, W. (1990). *Applied Nonparametric Regression*. New York: Cambridge University Press.

Hastie, T.J. and Tibshirani, R. (1990). *Generalized Additive Models*. London: Chapman and Hall.

Hong, Y. and Lee, T.-H. (1999). Inference and forecast of exchange rates via generalized spectrum and nonlinear time series models. Forthcoming in *The Review of Economics and Statistics*.

Hurvich, C.M., Simonoff, J.S. and Tsai, C.-L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society, Series B*, **60**, 271-293.

Hurvich, C.M. and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, **76**, 297-307.

Ibragimov, I.A. and Yu.V. Linnik (1971). *Independent and Stationary Sequences of Random Variables*. Groningen, the Netherlands: Walters-Noordhoff.

Juhl, T. (2002). Functional coefficient models under unit root behavior. *Working Paper*, Department of Economics, Kansas University.

Karlsen, H. A. and Tjøstheim, D. (2001). Nonparametric estimation in null recurrent time series. *The Annals of Statistics*, **29**, 372-416.

Karlsen, H. A., Myklebust, T. and Tjøstheim, D. (2001). Nonparametric estimation in a nonlinear cointergration model. *Submitted for publication.*

Kim, W. (2001). Nonparametric kernel estimation of evolutionary autoregressive processes. *Working Paper*, Institute of Statistics and Economics, Humboldt University, Germany.

Kreiss, J.P., Neumann, M. and Yao, Q. (1998). Bootstrap tests for simple structures in nonparametric time series regression. *Working Paper*, Department of Statistics, London School of Economics, UK.

Lu, Z. (1998). On the ergodicity of non-linear autoregressive model with an autoregressive conditional heteroscedastic term. *Statistica Sinica*, **8**, 1205-1217.

Müller, H.-G. (1988). *Nonparametric Regression Analysis of Longitudinal Data.* Lecture Notes in Statistics, No. 46. Berlin: Springer-Verlag.

Pagan, A. and Ullah, A. (1999). *Nonparametric Econometrics.* New York: Cambridge University Press.

Park, J.Y. and Phillips, P.C.B. (1999). Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory*, **15**, 269-298.

Park, J.Y. and Phillips, P.C.B. (2002). Nonlinear regressions with integrated time series. *Econometrica*, **69**, 117-161.

Phillips, P.C.B. (2001). Trending time series and macroeconomic activity: Some present and future challenges. *Journal of Econometrics*, **100**, 21-27.

Rice, J. (1984). Bandwidth selection for nonparametric regression. *The Annals of Statistics*, **12**, 1215-1230.

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

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

Roussas, G.G. (1989). Consistent regression estimation with fixed design points under dependence conditions. *Statistics and Probability Letters*, **8**, 41-50.

Roussas, G.G., Tran, L.T. and Ioannides, D.A. (1992). Fixed design regression for time series: Asymptotic normality. *Journal of Multivariate Analysis*, **40**, 262-291.

Shao, Q. and H. Yu (1996). Weak convergence for weighted empirical processes of dependent sequences. *The Annals of Probability*, **24**, 2098-2127.

Tran, L., Roussas, G., Yakowitz, S. and Van, B.T. (1996). Fixed-design regression for linear time series. *The Annals of Statistics*, **24**, 975-991.

Tsay, R.S. (2002). *Analysis of Financial Time Series.* New York: John Wiley & Sons.

Tschernig, R. and Yang, L. (2000). Nonparametric lag selection for time series. *Journal of Time Series Analysis*, **21**, 457-487.

View, P. (1994). Order choice in nonlinear autoregressive models. *Statistics*, **24**, 1-22.

Volkonskii, V.A. and Yu.A. Rozanov (1959). Some limit theorems for random functions. I. *Theory of Probability and Its Applications*, **4**, 178-197.

Wahba, G. (1977). A survey of some smoothing problems and the method of generalized cross-validation for solving them. In *Applications of Statistics* (P.R. Krisnaiah, ed.), 507-523, Amsterdam, North Holland.

Withers, C.S. (1981). Conditions for linear processes to be strong mixing. *Zeitschrift fur Wahrscheinlichkeitstheorie verwandte Gebiete*, **57**, 477-480.

Xia, Y. and Li, W.K. (1999). On the estimation and testing of functional-coefficient linear models. *Statistica Sinica*, **9**, 735-757.

Yao, Q. and Tong, H. (1994). On subset selection in nonparametric stochastic regression. *Statistica Sinica*, **4**, 51-70.

Zhang, W., Lee, S.-Y. and Song, X. (2002). Local polynomial fitting in semivarying coefficient model. *Journal of Multivariate Analysis*, **82**, 166-188.

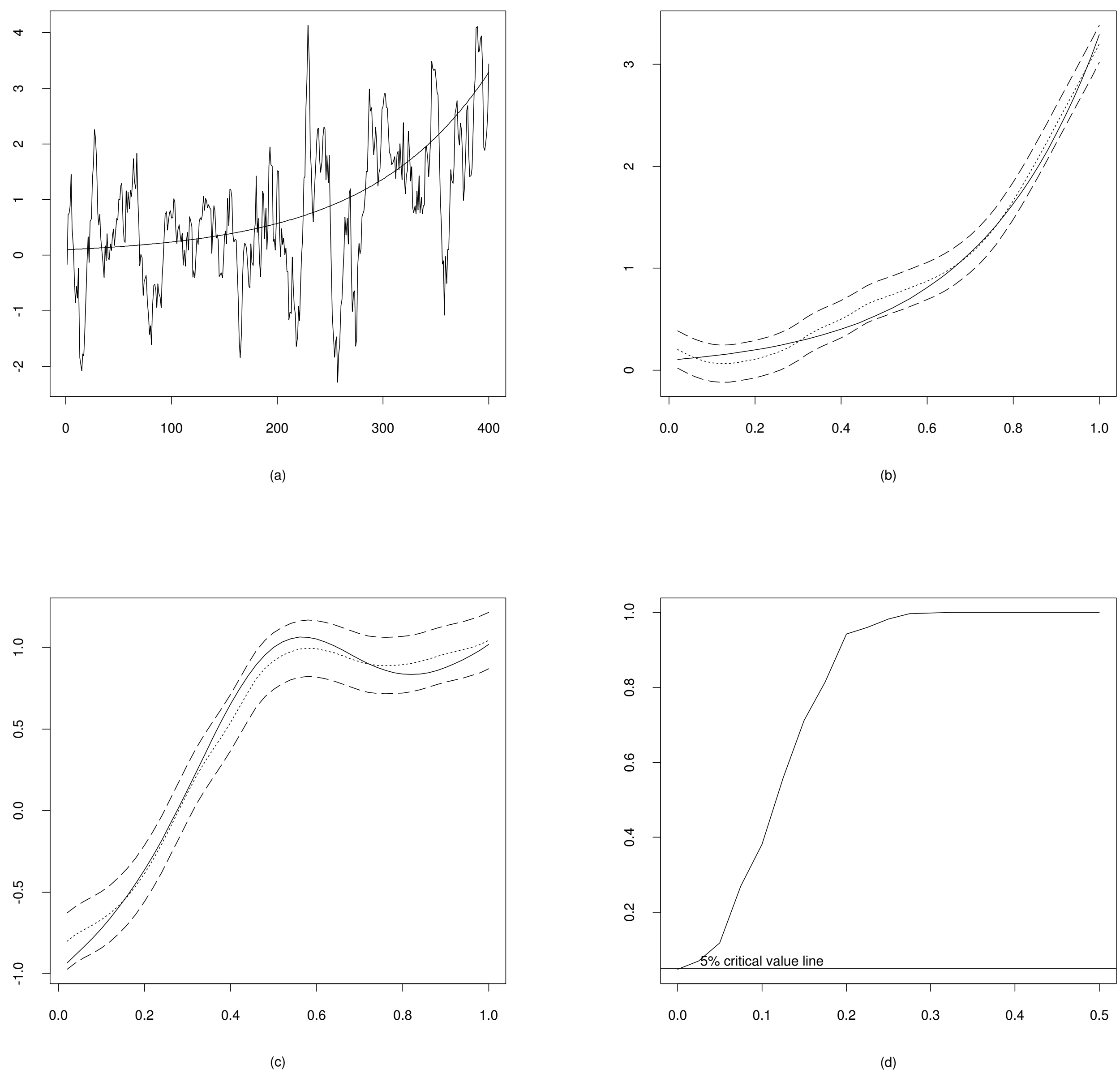

Figure 1: Simulation Results for Example 1. (a) The time series plot of the time series $\{y_i\}$ with the true trend function $\beta_0(\cdot)$; (b) the local linear estimator (dotted line) of the trend function $\beta_0(\cdot)$ (solid line) with the 95% confidence interval with the bias ignored (dashed lines); (c) the local linear estimator (dotted line) of the coefficient function $\beta_1(\cdot)$ (solid line) with the 95% confidence interval with the bias ignored (dashed lines); (d) the plot of the power curve against α for the misspecification test.

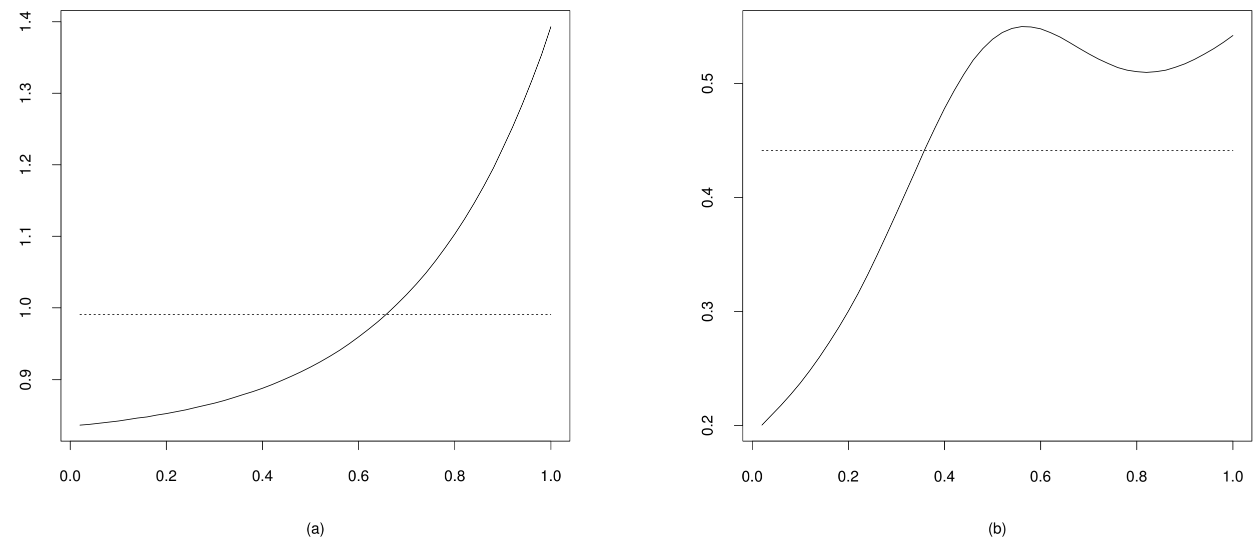

Figure 2: Simulation Results for Example 1. The trend function $\beta_0(\cdot)$ in (a) and the coefficient function $\beta_1(\cdot)$ in (b). The solid curves are under H_a with $\alpha = 0.175$ and the dotted lines under H_0.