Estimation and Testing for Varying Coefficients in Additive Models With Marginal Integration

Lijian YANG, Byeong U. PARK, Lan XUE, and Wolfgang HÄRDLE

We propose marginal integration estimation and testing methods for the coefficients of varying-coefficient multivariate regression models. Asymptotic distribution theory is developed for the estimation method, which enjoys the same rate of convergence as univariate function estimation. For the test statistic, asymptotic normal theory is established. These theoretical results are derived under the fairly general conditions of absolute regularity (β-mixing). Application of the test procedure to West German real GNP (gross national product) data reveals that a partially linear varying coefficient model is best parsimonious in fitting the data dynamics, a fact that is also confirmed with residual diagnostics.

KEY WORDS: Equivalent kernels; German real GNP; Local polynomial; Marginal integration; Rate of convergence.

1. INTRODUCTION

Parametric regression analysis usually assumes that the response variable Y depends linearly on a vector $\mathbf{X}$ of predictor variables. More flexible non- and semiparametric regression models allow the dependence to be of more general nonlinear forms. Conversely, the appeal of simplicity and interpretation still motivates the search for models that are nonparametric in nature but have special features that are appropriate for the data involved. These include additive models (Chen and Tsay 1993a; Linton and Nielsen 1995; Masry and Tjøstheim 1995, 1997; Mammen, Linton, and Nielsen 1999; Sperlich, Tjøstheim, and Yang 2002), generalized additive models (Linton and Härdle 1996), partially linear models (Härdle, Liang, and Gao 2000), and the like.

In this article we consider a form of flexible nonparametric regression model proposed by Hastie and Tibshirani (1993). The following model

$$Y_i = m(\mathbf{X}_i, \mathbf{T}_i) + \sigma(\mathbf{X}_i, \mathbf{T}_i)\varepsilon_i, \qquad i = 1, \ldots, n, \tag{1}$$

where $\{\varepsilon_i\}_{i\geq 1}$ are iid white noise, each ε_i independent of $(\mathbf{X}_i, \mathbf{T}_i)$, where

$$\mathbf{X}_i = (X_{i1}, \ldots, X_{id})^T, \qquad \mathbf{T}_i = (T_{i1}, \ldots, T_{id})^T, \tag{2}$$

is called a *varying-coefficient model* if

$$\text{(Model I)} \qquad m(\mathbf{X}_i, \mathbf{T}_i) = \sum_{s=1}^{d} f_s(X_{is})T_{is}.$$

In Model I, all the variables $\{X_s\}_{s=1}^d$ are different from each other. The model where all the variables $\{X_s\}_{s=1}^d$ are the same, that is, $m(X_i, \mathbf{T}_i) = \sum_{s=1}^d f_s(X_i)T_{is}$, is the *functional coefficient model* of Chen and Tsay (1993b) with univariate coefficient functions. The latter is different from Model I and was fully discussed by Cai, Fan, and Li (2000a) and Cai, Fan, and Yao (2000b). Indeed, Hastie and Tibshirani (1993) fitted real data examples exclusively with the functional coefficient model. Although the name *varying-coefficient model* was used by Cai et al. (2000a), the model they studied was the same model proposed by Chen and Tsay (1993b), except with the additional feature of a possibly nontrivial link function. Cai et al. (2000a) used local maximum likelihood estimation for all coefficient functions $\{f_s\}_{s=1}^d$, whose computing was no more than a univariate estimation, because all these univariate functions depend on the same variable X. The estimation method proposed for the functional coefficient model does not apply for Model I.

For Model I, the only existing estimation method was the backfitting method of Hastie and Tibshirani (1993), which has not been theoretically justified. Intuitively, inference about model (1) is no more complex than that of univariate models. In this article we develop a marginal integration-type estimator for each varying coefficient $\{f_s\}_{s=1}^d$ in the case when each varying coefficient can have a different variable. Our method achieves the optimal rate of convergence for univariate function estimation and has a simple asymptotic theory for the estimators.

As an illustration of the effectiveness of Model I, we consider real time series data $\{Y_t\}_{t=1}^n$ on West German gross national product (GNP) in Section 5. After taking the first difference and de-seasonalization, the data are considered strictly stationary, as shown by the dotted curve in Figure 4. The varying-coefficient models $Y_t = f_1(Y_{t-1})Y_{t-2} + f_2(Y_{t-3})Y_{t-4} +$ (noise) and $Y_t = f_1(Y_{t-3})Y_{t-2} + f_2(Y_{t-1})Y_{t-4} +$ (noise) are fitted, and the estimates of the functions f_1 and f_2 are plotted in Figure 2. These varying-coefficient autoregressive (AR) models have 2.81 and 2.46 times, respectively, more prediction power than the simple linear AR model. See Table 3 to find $.00059/.00021 = 2.81$ and $.00059/.00024 = 2.46$. More details about the data and the modeling procedures are given in Section 5.

Model I may be viewed as a special case of a functional coefficient model with *multivariate* coefficient functions $m(\mathbf{X}_i, \mathbf{T}_i) = \sum_{s=1}^d g_s(\mathbf{X}_i)T_{is}$, where $g_s(\mathbf{X}_i) = f_s(X_{is})$ for $s = 1, \ldots, d$. In this respect, it would be of interest to compare Model I with some related functional coefficient autoregressive (FAR) models. For example, for the varying-coefficient

Lijian Yang is Associate Professor, Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824 (E-mail: *yang@stt.msu.edu*). Byeong U. Park is Professor, Department of Statistics, Seoul National University, Seoul 151-747, Korea (E-mail: *bupark@stats.snu.ac.kr*). Lan Xue is Assistant Professor, Department of Statistics, Oregon State University, Corvallis, OR 97331-4501 (E-mail: *xuel@stat.oregonstate.edu*). Wolfgang Härdle is Professor, Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin, D-10178 Berlin, Germany (E-mail: *haerdle@wiwi.hu-berlin.de*). This work was supported by the Deutsche Forschungsgemeinschaft through SFB 373 "Quantifikation und Simulation Ökonomischer Prozesse" and SFB 649 "Economic Risk." Yang's research was also supported in part by NSF grants DMS-99-71186, DMS-04-05330, and SES-0127722. Park's research was also supported by the SRC/ERC program of MOST/KOSEF (grant R11-2000-073-00000). Xue's research was also supported in part by NSF grants BCS-0308420 and DMS-04-05330. The authors thank the editor, the associate editor, and two referees for their insightful comments, which led a substantial improvement of the article.

Journal of the American Statistical Association
September 2006, Vol. 101, No. 475, Theory and Methods
DOI 10.1198/016214506000000429

model $Y_t = f_1(Y_{t-3})Y_{t-1} + f_2(Y_{t-4})Y_{t-2} +$ (noise), one may consider the following FAR model for a comparison: $Y_t = f_1(Y_{t-3}, Y_{t-4})Y_{t-1} + f_2(Y_{t-3}, Y_{t-4})Y_{t-2} +$ (noise). In a simulation study that is presented in Section 4, we find that the mean average squared residuals and the mean average squared prediction errors of the FAR model are larger than those of the varying-coefficient model. More details on the simulation results are found in Section 4.2.

Of other special practical interest is the model that allows some of the X_s's to be the same. For this, we consider the following generalization of Model I:

$$\text{(Model II)} \qquad m(\mathbf{X}_i, \mathbf{T}_i) = \sum_{s=1}^{d_0} \sum_{u=1}^{r_s} f_{su}(X_{is}) T_{isu},$$

where now the coefficient functions $f_{s1}, \ldots, f_{sr_s}$ depend on the same variable X_s. In Model II, the dimension of $\mathbf{X}$ is d_0, which is less than $d = \sum_{s=1}^{d_0} r_s$, the dimension of $\mathbf{T}$, and all the variables $\{X_s\}_{s=1}^{d_0}$ are different from each other. An advantage of Model II is that it alleviates the dimensionality problem that the marginal integration method may have in fitting Model I. Furthermore, the functional coefficient model of Chen and Tsay (1993b) is a special case of Model II where $d_0 = 1$. As an example of Model II, we can write

$$\begin{aligned} Y_t &= c + a_1(r_t)M_t + a_2(r_t)M_t^2 \\ &\quad + a_3(r_t)M_t^2 I_{\{M_t<0\}} + b_1(t)\tau_t + b_2(t)\tau_t^2 \\ &\quad + \varepsilon_t, \qquad t = 1, \ldots, n, \end{aligned}$$

where Y_t denotes the implied volatility, r_t the interest rate, M_t the moneyness, and τ_t the maturity at time t.

Although our models consist of additive bivariate functions, they are linear in the variables T_s (T_{su}). One interesting question one may ask is: Are some of the coefficient functions f_s (f_{su}) constant? If the answer is yes for some but not all, then the model is partially linear in some variables T_s (T_{su}); if the answer is yes to all, then the model is the classical linear regression model. Any constant f_s (f_{su}) can then be estimated at the $1/\sqrt{n}$ rate of convergence. A formal testing procedure is proposed in Section 3 for determining the constancy of coefficient functions f_s (f_{su}). For the German GNP data, it is found that f_1 can be set to a constant, while f_2 cannot.

The article is organized as follows. In Section 2 we describe marginal estimation methods for Models I and II and derive the asymptotic distribution theory of the estimators. In Section 3 we propose a test procedure to test the hypothesis that f_s (f_{su}) is a constant. In Section 4 we illustrate the finite-sample properties of our proposals in the estimation and testing problems. In Section 5 we apply our estimation and testing methods to the West German real GNP data. All technical assumptions and proofs are given in the Appendix.

2. ESTIMATION OF VARYING COEFFICIENTS

2.1 Model I

In this section we formulate local polynomial integration estimators of the coefficient functions $\{f_s\}_{s=1}^d$ in Model I. For general background on the local polynomial method, see Stone (1977), Katkovnik (1979), Ruppert and Wand (1994), Wand and Jones (1995), and Fan and Gijbels (1996).

We assume that each ε_i is independent of the vectors $\{(\mathbf{X}_j, \mathbf{T}_j)\}_{j=1,\ldots,i}$ for each $i = 1, \ldots, n$. This is sufficient for obtaining our main results on distribution theory as we assume $\{(\mathbf{X}_j, \mathbf{T}_j)\}_{j=1,\ldots,n}$ is strictly stationary and geometrically β-mixing in assumption A2 (see the App.), but weaker than the usual assumption that each ε_i is independent of the vectors $\{(\mathbf{X}_j, \mathbf{T}_j)\}_{j=1,\ldots,n}$.

Note that if there exists nontrivial linear dependence among the variables T_s with corresponding functions of X_s as coefficients, then the functions f_s are unidentifiable. To be precise, if

$$\sum_{s=1}^{d} r_s(X_{is}) T_{is} = 0 \quad \text{a.s.}$$

for some nonzero measurable functions r_s, then the regression function m in Model I equals

$$\sum_{s=1}^{d} \{f_s(X_{is}) + r_s(X_{is})\} T_{is}$$

as well. Hence, for identifiability, we assume that

$$\sum_{s=1}^{d} r_s(X_{is}) T_{is} = 0 \quad \text{a.s.} \quad \Longrightarrow \quad r_s(x) \equiv 0, \qquad s = 1, \ldots, d. \tag{3}$$

Condition (3) may be considered an analog of linear independence between covariates in linear models. It is a sufficient condition for avoiding the *concurvity* referred by Hastie and Tibshirani (1990). The term *concurvity* in additive models is understood as an analog of *collinearity* in linear models. The condition is closely related to the invertibility of the matrix $\mathbf{Z}_s^T \mathbf{W}_s(\mathbf{X}_{-s}) \mathbf{Z}_s$ to be defined later; see Section A.2 for more details.

Now let $\mathbf{x} = (x_1, \ldots, x_d)^T \in \mathbb{R}^d$ be a point where we want to estimate the functions $\{f_s\}_{s=1}^d$. We denote by $(\mathbf{X}, \mathbf{T}) = (X_1, \ldots, X_d, T_1, \ldots, T_d)$ a generic random vector having the same distribution as $(\mathbf{X}_i, \mathbf{T}_i) = (X_{i1}, \ldots, X_{id}, T_{i1}, \ldots, T_{id})$ and define $\mathbf{X}_{-s}$ and $\mathbf{T}_{-s}$, as obtained from $\mathbf{X}$ and $\mathbf{T}$ by removing the sth components, by

$$\begin{aligned} \mathbf{X}_{-s} &= (X_1, \ldots, X_{s-1}, X_{s+1}, \ldots, X_d)^T, \qquad s = 1, \ldots, d, \\ \mathbf{T}_{-s} &= (T_1, \ldots, T_{s-1}, T_{s+1}, \ldots, T_d)^T, \qquad s = 1, \ldots, d. \end{aligned}$$

For a kernel function K, we write $K_h(u) = K(u/h)/h$. We fit pth-order local polynomials to estimate the varying coefficients. Write $\mathbf{Y} = (Y_i)_{1 \le i \le n}$ and $\mathbf{p}(u) = (1, u, \ldots, u^p)^T$. Define $\mathbf{Z}_s$ to be the $n \times (p + d)$ matrix that has $(\mathbf{p}\{(X_{is} - x_s)/h\}^T T_{is}, \mathbf{T}_{i,-s}^T)$ as its ith row. Let $\mathbf{W}_s(\mathbf{x}_{-s}) \equiv \mathbf{W}_s(x_s, \mathbf{x}_{-s})$ be the $n \times n$ diagonal matrix defined by

$$\mathbf{W}_s(\mathbf{x}_{-s}) = \operatorname{diag}\left\{ \frac{K_h(X_{js} - x_s) L_{\mathbf{g}}(\mathbf{X}_{j,-s} - \mathbf{x}_{-s})}{n} \right\}_{1 \le j \le n},$$

where $L_{\mathbf{g}}(\mathbf{u}) = (g_1 \cdots g_{s-1} g_{s+1} \cdots g_d)^{-1} L(g_1^{-1} u_1, \ldots, g_{s-1}^{-1} \times u_{s-1}, g_{s+1}^{-1} u_{s+1}, \ldots, g_d^{-1} u_d)$, L is a $(d-1)$-variate kernel, and $g_1, \ldots, g_{s-1}, g_{s+1}, \ldots, g_d$ are bandwidths that are allowed to be different from each other. Then the first component of the minimizer $\hat{\beta}$ of the weighted sum of squares

$$\sum_{j=1}^{n}\left\{Y_j-\sum_{l=0}^{p}\beta_{sl}(X_{js}-x_s)^l T_{js}-\sum_{k\neq s}\beta_k T_{jk}\right\}^2 \times K_h(X_{js}-x_s)L_{\mathbf{g}}(\mathbf{X}_{j,-s}-\mathbf{x}_{-s})$$

is given by

$$\hat{\beta}_{s0}\equiv\hat{\beta}_{s0}(\mathbf{x}_{-s})=e_0^T\left(\mathbf{Z}_s^T\mathbf{W}_s(\mathbf{x}_{-s})\mathbf{Z}_s\right)^{-1}\mathbf{Z}_s^T\mathbf{W}_s(\mathbf{x}_{-s})\mathbf{Y},$$

where e_l is the $(p+d)$-dimensional vector whose entries are 0 except the $(l+1)$th element, which equals 1.

The integration estimator of $f_s(x_s)$ is a weighted average of the $\beta_{s0}(\mathbf{X}_{i,-s})$'s, that is,

$$\hat{f}_s(x_s)=\sum_{i=1}^{n}w_{-s}(\mathbf{X}_{i,-s})\hat{\beta}_{s0}(\mathbf{X}_{i,-s})\Big/\sum_{i=1}^{n}w_{-s}(\mathbf{X}_{i,-s}), \qquad (4)$$

where the weight function $w_{-s}(\cdot)$ has a compact support with nonempty interior and is introduced here to avoid some technical difficulty that may arise when the density of the $\mathbf{X}_{i,-s}$'s has an unbounded support. Based on (4), we can predict Y given any realization $(\mathbf{x},\mathbf{t})$ of $(\mathbf{X},\mathbf{T})$ by the predictor

$$\hat{m}(\mathbf{x},\mathbf{t})=\sum_{s=1}^{d}\hat{f}_s(x_s)t_s. \qquad (5)$$

In the estimation procedure for f_s for a given s, we fit local constants for the other varying coefficients $f_{s'}$, $s'\neq s$. We could fit higher-order local polynomials for those varying coefficients, too. The theoretical performance of the resulting estimator would be the same as the present one, however. The smoothing bias of the present estimator due to the local averaging for $f_{s'}$, $s'\neq s$, can be made negligible by choosing the bandwidth vector $\mathbf{g}$ of smaller order than h and using a higher-order kernel L. See the conditions for the bandwidths and the kernel L given in the Appendix. In fact, the approach of taking a smaller bandwidth $\mathbf{g}$ and a higher-order kernel L for the directions not of interest was also adopted by Fan, Härdle, and Mammen (1998). We may sacrifice some rate of convergence in order to use a lower-order kernel.

Let φ, φ_{-s}, and φ_s denote the densities of $\mathbf{X}$, $\mathbf{X}_{-s}$, and X_s, respectively. Define

$$b_s(x_s)=\frac{f_s^{(p+1)}(x_s)\int u^{p+1}E\{w_{-s}(\mathbf{X}_{-s})T_sK_s^*(u;\mathbf{T},x_s,\mathbf{X}_{-s})\}\,du}{(p+1)!E\{w_{-s}(\mathbf{X}_{-s})\}},$$

$$\sigma_s^2(x_s)=E\left[\frac{w_{-s}^2(\mathbf{X}_{-s})}{\varphi^2(\mathbf{X})}\varphi_{-s}^2(\mathbf{X}_{-s})\sigma^2(\mathbf{X},\mathbf{T}) \times\int K_s^{*2}(u;\mathbf{T},\mathbf{X})\,du\Big|X_s=x_s\right]\frac{\varphi_s(x_s)}{E^2\{w_{-s}(\mathbf{X}_{-s})\}},$$

where K_s^* is the equivalent kernel defined in (A.7).

Theorem 1. Under assumptions A1–A7 in the Appendix, we have, for any $s=1,\dots,d$, as $n\to\infty$,

$$\sqrt{nh}\left\{\hat{f}_s(x_s)-f_s(x_s)-h^{p+1}b_s(x_s)\right\}\xrightarrow{\mathcal{L}}\mathrm{N}\{0,\sigma_s^2(x_s)\}. \qquad (6)$$

The estimator $\hat{m}(\mathbf{x},\mathbf{t})$ of the prediction function $m(\mathbf{x},\mathbf{t})$ enjoys the same rate of convergence as that of a single varying coefficient, and its asymptotic parameters are easily calculated from those of the $\hat{f}_s(x_s)$'s and the value of $\mathbf{t}$, as in the following theorem.

Theorem 2. Under assumptions A1–A7 in the Appendix, we have, for any $s\neq s'$,

$$\mathrm{cov}\left[\sqrt{nh}\{\hat{f}_s(x_s)-f_s(x_s)\},\sqrt{nh}\{\hat{f}_{s'}(x_{s'})-f_{s'}(x_{s'})\}\right]\longrightarrow 0 \qquad (7)$$

as $n\to\infty$. Hence,

$$\sqrt{nh}\left\{\hat{m}(\mathbf{x},\mathbf{t})-m(\mathbf{x},\mathbf{t})-h^{p+1}b_m(\mathbf{x},\mathbf{t})\right\}\xrightarrow{\mathcal{L}}\mathrm{N}\{0,\sigma_m^2(\mathbf{x},\mathbf{t})\}, \qquad (8)$$

where $b_m(\mathbf{x},\mathbf{t})=\sum_{s=1}^{d}b_s(x_s)t_s$ and $\sigma_m^2(\mathbf{x},\mathbf{t})=\sum_{s=1}^{d}\sigma_s^2(x_s)t_s^2$.

We note here that Theorems 1 and 2 hold only for local polynomial estimators of odd degree p, whereas similar results hold for p even as well. In particular, $p=0$ corresponds to integrating the well-known Nadaraya–Watson estimator. When an even p is used instead, the variance formula remains the same, whereas the bias formula contains extra terms involving the derivatives of the design density.

For selecting the bandwidths, following the idea of Ruppert, Sheather, and Wand (1995) in local least squares regression, several plug-in-type bandwidth selectors may be developed based on the asymptotic formulas given in the Theorems 1 and 2. Also, the modified multifold cross-validation criterion considered by Cai et al. (2000b) may be adapted for the preceding estimation. Theoretical development for these bandwidth selectors is beyond the scope of this article. In the following discussion, we describe a simple plug-in selection procedure for h and $\mathbf{g}$, which is employed in our numerical study in Sections 4 and 5.

The optimal bandwidth h_{opt}, which minimizes the asymptotic mean integrated squared error of $\hat{f}_s$, is given by

$$h_{\mathrm{opt}}=\left\{\frac{\int\sigma_s^2(x_s)\,dx_s}{2n(p+1)\int b_s^2(x_s)\,dx_s}\right\}^{1/(2p+3)}.$$

Now $\int b_s^2(x_s)\,dx_s$ and $\int\sigma_s^2(x_s)\,dx_s$ can be approximated, respectively, by

$$\left[(p+1)!n^{-1}\sum_{i=1}^{n}w_{-s}(\mathbf{X}_{i,-s})\right]^{-2} \times\int\left[f_s^{(p+1)}(x_s)\int u^{p+1}n^{-1} \times\sum_{i=1}^{n}\left\{w_{-s}(\mathbf{X}_{i,-s})T_{is}K_s^*(u,\mathbf{T}_i,x_s,\mathbf{X}_{i,-s})\right\}du\right]^2dx_s,$$

$$\left[n^{-1}\sum_{i=1}^{n}w_{-s}(\mathbf{X}_{i,-s})\right]^{-2}n^{-1} \times\sum_{i=1}^{n}w_{-s}^2(\mathbf{X}_{i,-s})\varphi^{-2}(\mathbf{X}_i)\varphi_{-s}^2(\mathbf{X}_{i,-s})\sigma^2(\mathbf{X}_i,\mathbf{T}_i) \times\int K_s^{*2}(u,\mathbf{T}_i,\mathbf{X}_i)\,du.$$

The unknown functions $f_s^{(p+1)}(x_s)$, $\sigma^2(\mathbf{x},\mathbf{t})$, $\varphi(\mathbf{x})$, $\varphi(\mathbf{x}_{-s})$, and K_s^* may be substituted with their estimators as follows.

The $(p+1)$th derivative function $f_s^{(p+1)}(x_s)$ is estimated by fitting a polynomial regression model of degree $(p+2)$:

$$m(\mathbf{X},\mathbf{T}) = \sum_{s=1}^{d}\sum_{k=0}^{p+2} a_{s,k} X_s^k T_s.$$

This leads to an estimator $\hat{f}_s^{(p+1)}(x_s) = (p+1)!\hat{a}_{s,p+1} + (p+2)!\hat{a}_{s,p+2}x_s$. As a by-product, the mean squared residual is used as an estimator of $\sigma^2(\mathbf{x},\mathbf{t})$. The density functions $\varphi(\mathbf{x})$ and $\varphi(\mathbf{x}_{-s})$ are estimated by

$$\hat{\varphi}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\prod_{s=1}^{d}\frac{1}{h(\mathbf{X},d)}\phi\left(\frac{X_{is}-x_s}{h(\mathbf{X},d)}\right),$$

$$\hat{\varphi}_{-s}(\mathbf{x}_{-s}) = \frac{1}{n}\sum_{i=1}^{n}\prod_{s'\neq s}^{d}\frac{1}{h(\mathbf{X}_{-s},d-1)}\phi\left(\frac{X_{is'}-x_{s'}}{h(\mathbf{X}_{-s},d_2-1)}\right),$$

with the standard normal density ϕ and the rule-of-the-thumb bandwidth

$$h(\mathbf{X},m) = \sqrt{\widehat{\operatorname{var}}(\mathbf{X})}\left\{\frac{4}{m+2}\right\}^{1/(m+4)} n^{-1/(m+4)}.$$

According to its definition given in (A.7), the dependence of the function $K_s^*(u,\mathbf{t},\mathbf{x})$ on u and $\mathbf{t}$ is completely known. The only unknown term $E(\mathbf{T}\mathbf{T}^T|\mathbf{X}=\mathbf{x})$ contained in $S_s^{-1}(\mathbf{x})$ is estimated by fitting a matrix polynomial regression

$$E(\mathbf{T}\mathbf{T}^T|\mathbf{X}=\mathbf{x}) = \mathbf{c} + \sum_{s=1}^{d}\sum_{k=1}^{p}\mathbf{c}_{s,k}x_s^k,$$

where the coefficients $\mathbf{c}$ and $\mathbf{c}_{s,k}$ are $d\times d$ matrices.

For the bandwidth vector $\mathbf{g}$, we note that the choice $g_1 = \cdots = g_{s-1} = g_{s+1} = \cdots = g_d = (\log n)^{-1}h^{(p+1)/q}$ with h asymptotic to $n^{-1/(2p+3)}$ satisfies condition (A.7) for Theorem 1 if q, the order of the kernel L, is greater than $(d-1)/2$. Thus, one may take $g_j \equiv (\log n)^{-1}h_{\mathrm{opt}}^{(p+1)/q}$ for $j=1,\dots,s-1,s+1,\dots,d$, where h_{opt} is the optimal bandwidth obtained from the preceding procedure.

2.2 Model II

In this section we describe local polynomial integration estimators of the coefficient functions $\{f_{su}, 1\le u\le r_s, 1\le s\le d_0\}$ in Model II. For the identifiability of the functions f_{su}, we assume that

$$\sum_{s=1}^{d_0}\sum_{u=1}^{r_s} r_{su}(X_{is})T_{isu} = 0 \quad \text{a.s.} \quad \Longrightarrow \quad r_{su}(x)\equiv 0, \qquad u=1,\dots,r_s, s=1,\dots,d_0.$$

Define $\mathbf{X}_{-s}$ and $\mathbf{x}_{-s}$ as in Section 2.1. Let $\hat{\beta}_{su0}(\mathbf{x}_{-s})$ be the first component of the minimizer $\hat{\beta}$ of the following weighted sum of squares:

$$\sum_{j=1}^{n}\left\{Y_j - \sum_{u=1}^{r_s}\sum_{l=0}^{p}\beta_{sul}(X_{js}-x_s)^l T_{jsu} - \sum_{s'\neq s}^{d_0}\sum_{u'=1}^{r_{s'}}\beta_{s'u'}T_{js'u'}\right\}^2 \times K_h(X_{js}-x_s)L_{\mathbf{g}}(\mathbf{X}_{j,-s}-\mathbf{x}_{-s}).$$

The integration estimator of $f_{su}(x_s)$ is given by a weighted average of $\beta_{su0}(\mathbf{X}_{i,-s})$'s, that is,

$$\hat{f}_{su}(x_s) = \sum_{i=1}^{n} w_{-s}(\mathbf{X}_{i,-s})\hat{\beta}_{su0}(\mathbf{X}_{i,-s}) \Big/ \sum_{i=1}^{n} w_{-s}(\mathbf{X}_{i,-s}).$$

As in Model I, we may predict Y given any realization $(\mathbf{x},\mathbf{t})$ of $(\mathbf{X},\mathbf{T})$ by the predictor

$$\hat{m}(\mathbf{x},\mathbf{t}) = \sum_{s=1}^{d_0}\sum_{u=1}^{r_s}\hat{f}_{su}(x_s)t_{su}. \tag{9}$$

We have the following theorem, which is an analog of Theorem 1.

Theorem 3. Under assumptions A1, A2, A3′, A4′, A5, A6, and A7′ in the Appendix, we have, for any $u=1,\dots,r_s$ and $s=1,\dots,d_0$,

$$\sqrt{nh}\{\hat{f}_{su}(x_s) - f_{su}(x_s) - h^{p+1}b_{su}(x_s)\} \xrightarrow{\mathcal{L}} \mathrm{N}\{0,\sigma_{su}^2(x_s)\} \tag{10}$$

as $n\to\infty$, where $b_{su}(x_s) = \kappa_{su}(x_s)/\eta_s$ and $\sigma_{su}^2(x_s) = \tau_{su}^2(x_s)/\eta_s^2$. The definitions of κ_{su} and τ_{su}^2 are given in (A.19) and (A.20).

Each pair of the entries $\hat{f}_{su}$ and $\hat{f}_{s'u'}$ for $1\le s,s'\le d_0$ and $1\le u,u'\le r_s$ has a negligible asymptotic covariance when $s\neq s'$. However, it has the same magnitude as the variance of each entry when $s=s'$. The following theorem is an analog of Theorem 2.

Theorem 4. Under the assumptions of Theorem 3, we have, as $n\to\infty$,

a. when $s'\neq s$,

$$\operatorname{cov}\left[\sqrt{nh}\{\hat{f}_{su}(x_s)-f_{su}(x_s)\}, \sqrt{nh}\{\hat{f}_{s'u'}(x_{s'})-f_{s'u'}(x_{s'})\}\right]\to 0;$$

b. when $s'=s$,

$$\operatorname{cov}\left[\sqrt{nh}\{\hat{f}_{su}(x_s)-f_{su}(x_s)\}, \sqrt{nh}\{\hat{f}_{su'}(x_s)-f_{su'}(x_s)\}\right] \to \tau_{suu'}(x_s)/\eta_s^2,$$

where $\tau_{suu'}$ is defined in (A.23). Hence,

$$\sqrt{nh}\{\hat{m}(\mathbf{x},\mathbf{t}) - m(\mathbf{x},\mathbf{t}) - h^{p+1}\tilde{b}_m(\mathbf{x},\mathbf{t})\} \xrightarrow{\mathcal{L}} \mathrm{N}\{0,\tilde{\sigma}_m^2(\mathbf{x},\mathbf{t})\},$$

where

$$\tilde{b}_m(\mathbf{x},\mathbf{t}) = \sum_{s=1}^{d_0}\sum_{u=1}^{r_s} b_{su}(x_s)t_{su},$$

$$\tilde{\sigma}_m^2(\mathbf{x},\mathbf{t}) = \sum_{s=1}^{d_0}\sum_{u=1}^{r_s}\sum_{u'=1}^{r_s}\sigma_{suu'}(x_s)t_{su}t_{su'}, \qquad \text{and}$$

$$\sigma_{suu'}(x_s) = \tau_{suu'}(x_s)/\eta_s^2.$$

3. TESTING FOR VARYING COEFFICIENTS

Suppose we are interested in testing the hypothesis

$$H_0 : f_s(x_s) \equiv \text{const} \tag{11}$$

for a specific s in Model I. Testing the hypothesis (11) is a very important first step in the model-building procedure. If this hypothesis were true, one would get $\min_\alpha E\{f_s(X_s) - \alpha\}^2 \times w_s(X_s) = 0$, where w_s is an arbitrary positive weight function with a compact support. This leads us to propose the following test statistic:

$$\begin{aligned} V_{ns} &= n^{-1}\min_\alpha \sum_{i=1}^n \{\hat{f}_s(X_{is}) - \alpha\}^2 w_s(X_{is}) \\ &= n^{-1}\sum_{i=1}^n \hat{f}_s(X_{is})^2 w_s(X_{is}) \\ &\quad - n^{-1}\left\{\sum_{i=1}^n w_s(X_{is})\right\}^{-1}\left\{\sum_{i=1}^n \hat{f}_s(X_{is}) w_s(X_{is})\right\}^2, \end{aligned} \tag{12}$$

where the obvious solution of the least squares problem is given by

$$\hat{\alpha}_s = \left\{\sum_{i=1}^n w_s(X_{is})\right\}^{-1}\left\{\sum_{i=1}^n \hat{f}_s(X_{is}) w_s(X_{is})\right\}. \tag{13}$$

The next theorem describes the asymptotic distribution of the test statistic (12) under the null hypothesis (11).

Theorem 5. Under the null hypothesis (11) and assumptions A1–A7 in the Appendix, we have, for any $s = 1, \dots, d$,

$$nh^{1/2}(V_{ns} - n^{-1}h^{-1}v_s) \xrightarrow{\mathcal{L}} \mathrm{N}\{0, \gamma_s^2\} \tag{14}$$

as $n \to \infty$, where v_s and γ_s are given in (A.17) and (A.16).

For the practical implementation of the test, we suggest using a bootstrap procedure instead of the asymptotic normal distribution theory in Theorem 5. The reason is that for a test statistic based on kernel-type smoothing, the normal approximation to the distribution of the test statistic is very poor, as shown in Härdle and Mammen (1993) and, more recently, confirmed by Sperlich et al. (2002). Another reason is that the normal approximation given in Theorem 5 involves too-complicated expressions, which makes the task of obtaining asymptotic critical values out of reach.

It is well known that the ordinary method of resampling residuals fails to work when the error variances are allowed to be different. See Wu (1986), Liu (1988), and Mammen (1992). Härdle and Mammen (1993) also pointed out that it breaks down even for homoscedastic errors in the case of the goodness-of-fit test statistic for testing a parametric hypothesis against the nonparametric alternative. As an alternative, we suggest using the wild bootstrap procedure, which was first introduced by Wu (1986) and implemented in various settings by Liu (1988), Härdle and Mammen (1993), Sperlich et al. (2002), among others. Basically, this approach attempts to mimic the conditional distribution of each response-given covariate using the corresponding *single* residual, in such a way that the first three moments of the bootstrap population equal those of the single residual.

To describe the procedure in our setting, let $\tilde{m}(\mathbf{x}, \mathbf{t}) = \hat{\alpha}_s t_s + \sum_{k\neq s}^d \hat{f}_k(x_k) t_k$ be the regression estimator under the hypothesis (11), where $\hat{\alpha}_s$ is an estimate of the constant $\hat{f}_s(x_s)$ given by (13), whereas $\hat{f}_k(x_k)$ $(k \neq s)$ is the marginally integrated estimate of $f_k(x_k)$ in (4). The wild bootstrap procedure for estimating the sampling distribution of V_{ns} under the null hypothesis then consists of the following steps:

1. Find the residuals $\tilde{\varepsilon}_i = Y_i - \tilde{m}(\mathbf{X}_i, \mathbf{T}_i)$ for $i = 1, \dots, n$.
2. Generate iid random variables Z_i^W such that $E(Z_i^W) = 0$, $E(Z_i^W)^2 = 1$, and $E(Z_i^W)^3 = 1$. Put $Y_i^* = \tilde{m}(\mathbf{X}_i, \mathbf{T}_i) + \tilde{\varepsilon}_i Z_i^W$.
3. Compute the bootstrap test statistic V_{ns}^* using the wild bootstrap sample $\{(Y_i^*, \mathbf{X}_i, \mathbf{T}_i)\}_{i=1}^n$.
4. Repeat steps 2 and 3 M times, obtaining $V_{ns,1}^*, \dots, V_{ns,M}^*$. Estimate the null distribution of V_{ns} by the empirical distribution of $V_{ns,1}^*, \dots, V_{ns,M}^*$.

For examples of Z_i^W satisfying the moment conditions, see Mammen (1992). For the empirical example in the next section, we used a two-point distribution: $Z_i^W = (1 - \sqrt{5})/2$ with probability $(5 + \sqrt{5})/10$, and $Z_i^W = (1 + \sqrt{5})/2$ with probability $(5 - \sqrt{5})/10$, with $M = 200$.

For Model II, we consider the following hypothesis:

$$f_{su}(x_s) \equiv \text{const}. \tag{15}$$

The corresponding test statistic for the hypothesis (15) is given by

$$\begin{aligned} V_{nsu} &= n^{-1}\sum_{i=1}^n \hat{f}_{su}(X_{is})^2 w_s(X_{is}) \\ &\quad - n^{-1}\left\{\sum_{i=1}^n w_s(X_{is})\right\}^{-1}\left\{\sum_{i=1}^n \hat{f}_{su}(X_{is}) w_s(X_{is})\right\}^2. \end{aligned}$$

The next theorem describes the asymptotic distribution of the test statistic V_{nsu} under the null hypothesis (15).

Theorem 6. Under the null hypothesis (15) and the assumptions of Theorem 3, we have, for any $u = 1, \dots, r_s$ and $s = 1, \dots, d_0$,

$$nh^{1/2}(V_{nsu} - n^{-1}h^{-1}v_{su}) \xrightarrow{\mathcal{L}} \mathrm{N}\{0, \gamma_{su}^2\}$$

as $n \to \infty$, where v_{su} and γ_{su} are as given in (A.22) and (A.21).

For testing the hypothesis (15), let $\bar{m}(\mathbf{x}, \mathbf{t}) = \hat{\alpha}_{su} t_{su} + \sum_{s'\neq s}^{d_0} \sum_{u'=1}^{r_{s'}} \hat{f}_{s'u'}(x_{s'}) t_{s'u'}$, where

$$\hat{\alpha}_{su} = \left\{\sum_{i=1}^n w_s(X_{is})\right\}^{-1}\left\{\sum_{i=1}^n \hat{f}_{su}(X_{is}) w_s(X_{is})\right\}.$$

A wild bootstrap procedure may be obtained by simply replacing $\tilde{m}$, V_{ns}, and V_{ns}^* by $\bar{m}$, V_{nsu} and V_{nsu}^*, respectively, in the four steps described previously for testing (11).

Some related work on this testing problem includes Chen and Liu (2001) and Cai et al. (2000b). The former article treated testing, in the FAR model, whether all the coefficient functions are constant, that is, whether the underlying process is simply a linear AR model. The latter proposed a testing procedure for the hypothesis that all the coefficient functions have known parametric forms. We think testing for a parametric form in our models is also an interesting topic for future research.

4. SIMULATION STUDY

In this section we investigate the finite-sample properties of the estimation and testing methods through two simulated examples. One is the case where $(\mathbf{X}_i, \mathbf{T}_i)$ are independent and identically distributed, and the other is the case where they are endogenous and are lagged observations of the response Y. We employed local linear smoothing ($p = 1$) in all cases. Both of the kernels K and L were the quartic kernel $K(x) = L(x) = .9375(1 - x^2)^2 I_{(-1,1)}(x)$, whereas the bandwidths were chosen as described Section 2.1.

4.1 The iid Case

In this case, we generated the data from the following varying-coefficient model:

$$Y = f_1(X_1) + f_2(X_2)T_1 + f_3(X_3)T_2 + \delta(\mathbf{X}, \mathbf{T})\varepsilon,$$

where $f_1(X_1) = 1 + \exp(2X_1 - 1)$, $f_2(X_2) = \cos(2\pi X_2)$, and $f_3(X_3) = 2$. The heteroscedastic conditional standard deviation was set to be

$$\delta(\mathbf{X}, \mathbf{T}) = .5 + \frac{T_1^2 + T_2^2}{1 + T_1^2 + T_2^2} \exp\left(-2 + \frac{X_1 + X_2}{2}\right).$$

The particular form of $\delta(\mathbf{X}, \mathbf{T})$ was considered to ensure that the variance is bounded. The vector $\mathbf{X} = (X_1, X_2, X_3)^T$ was generated from the uniform distribution over the unit cube $[0, 1]^3$, and $\mathbf{T} = (T_1, T_2)^T$ was generated from the bivariate normal with mean 0 and covariance matrix $\begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix}$. The vectors $\mathbf{X}$ and $\mathbf{T}$ were generated independently. Finally, the error term ε was generated from the standard normal distribution independently of $(\mathbf{X}, \mathbf{T})$.

A total of 100 independent datasets with sizes $n = 50$, 100, and 250 was generated. The estimated functions of f_s, $s = 1, 2, 3$, were evaluated on a grid of 91 equally spaced points $x_j, j = 1, \ldots, 91$, with $x_1 = .05$ and $x_{91} = .95$. To assess the performance of $\hat{f}_s$ for $s = 1, 2, 3$, we calculated the mean integrated squared error (MISE) of $\hat{f}_s$, which is defined by

$$\mathrm{MISE}(\hat{f}_s) = \frac{1}{R}\sum_{r=1}^{R} \mathrm{ISE}(\hat{f}_{r,s}) = \frac{1}{R}\sum_{r=1}^{R} \frac{1}{g}\sum_{j=1}^{g} \{\hat{f}_{r,s}(x_j) - f_s(x_j)\}^2.$$

Here $\hat{f}_{r,s}(x_j)$ denotes the estimated value of f_s at x_j for the rth dataset, and $R = 100$ and $g = 91$ are the numbers of datasets and grid points, respectively. Table 1 summarizes the MISE values of the function estimators. This simulation study numerically supports our theoretical results for the estimation method as given in Section 2.

To see how the marginal integration improves the three-dimensional function estimators, we also computed the mean average squared errors for the case where $n = 50$. Consider $\hat{\beta}_{s0}$, as defined in Section 2.1, evaluated at the observed X_{i1}, X_{i2}, and X_{i3}, and write them as $\hat{\beta}_{s0}(X_{i1}, X_{i2}, X_{i3})$. These are the estimates before the marginal integration. We computed the mean average squared error

$$\mathrm{MASE}_1 = \frac{1}{R}\sum_{r=1}^{R} \frac{1}{n}\sum_{i=1}^{n} \{\hat{\beta}_{r,10}(X_{i1}, X_{i2}, X_{i3}) + \hat{\beta}_{r,20}(X_{i1}, X_{i2}, X_{i3})T_{i1} + \hat{\beta}_{r,30}(X_{i1}, X_{i2}, X_{i3})T_{i2} - f_1(X_{i1}) - f_2(X_{i2})T_{i1} - f_3(X_{i3})T_{i2}\}^2$$

and the marginal integration estimate

$$\mathrm{MASE}_2 = \frac{1}{R}\sum_{r=1}^{R} \frac{1}{n}\sum_{i=1}^{n} \{\hat{f}_{r,1}(X_{i1}) + \hat{f}_{r,2}(X_{i2})T_{i1} + \hat{f}_{r,3}(X_{i3})T_{i2} - f_1(X_{i1}) - f_2(X_{i2})T_{i1} - f_3(X_{i3})T_{i2}\}^2,$$

where $\hat{\beta}_{r,s0}(X_{i1}, X_{i2}, X_{i3})$ and $\hat{f}_{r,s}(X_{is})$ are the estimates for the rth dataset. We found $\mathrm{MASE}_1 = .3164$ and $\mathrm{MASE}_2 = .2761$.

Next, we give some numerical results for the testing method. For each of the simulated datasets given previously, we applied the proposed wild bootstrap method with $M = 500$ to test the null hypothesis $H_{s0}: f_s = c_s$ for some constants c_s. Table 2 provides for each s the proportion of the cases where the null hypothesis H_{s0} was rejected at the significance level $\alpha = .05$ among the 100 replications.

Table 1. MISEs of the Estimators $\hat{f}_1$, $\hat{f}_2$, and $\hat{f}_3$ for the iid Case

	f_1	f_2	f_3
$n = 50$	.0559	.1144	.1336
$n = 100$	.0300	.0515	.0617
$n = 250$	.0108	.0223	.0225

Table 2. Proportions Among the 100 Replications of Rejecting the Null Hypotheses H_{s0}, $s = 1, 2, 3$, at the Significance Level .05 for the iid Case

	H_{10}	H_{20}	H_{30}
$n = 50$	.94	.85	.02
$n = 100$	1	1	.04
$n = 250$	1	1	.08

4.2 The Time Series Case

In this simulation, $R = 200$ time series were generated. Each time, 1,000 observations were generated from the following varying-coefficient autoregressive (VCAR) model, among which only the last 250 observations were used:

$$Y_t = f_1(Y_{t-3})Y_{t-1} + f_2(Y_{t-4})Y_{t-2} + .2\varepsilon_t, \tag{16}$$

where $f_1(Y_{t-3}) = .4 + (.1 + Y_{t-3})\exp(-3Y_{t-3}^2)$, $f_2(Y_{t-4}) = -.2 - (.6 + Y_{t-4})\exp(-3Y_{t-4}^2)$, and ε_t are iid standard normal random variates. Again, the performance of the estimators of f_1 and f_2 was assessed by MISE. We obtained $\mathrm{MISE}(\hat{f}_1) = .0137$ and $\mathrm{MISE}(\hat{f}_2) = .0151$. We found that the Monte Carlo average over 200 time series of $\sqrt{\sum_{t=1}^{250}(Y_t - \bar{Y})^2/250}$ equals .6374 with a standard error of .0026, where $\bar{Y} = \sum_{t=1}^{250} Y_t/250$. The MISE values are much smaller than the variation of Y, which means that the fitted model with $\hat{f}_1$ and $\hat{f}_2$ is useful to explain the variation of Y.

As in the iid case, we report a numerical result for testing $H_{10}: f_1 =$ const and $H_{20}: f_2 =$ const. For each of the simulated time series, we applied the wild bootstrap method with $M = 500$ and used the significant level .05. We found that the proportion of cases where the null hypothesis was rejected among the 200 replications was .57 for H_{10} and .943 for H_{20}.

It is also of interest to examine the effectiveness of the varying-coefficient model (16) in comparison with some related FAR models, discussed in Cai et al. (2000a,b), where all the coefficient functions depend on the same variable(s). For this purpose, we considered the following three FAR models:

$$Y_t = g_1(Y_{t-3})Y_{t-1} + g_2(Y_{t-3})Y_{t-2} + .2\varepsilon_t, \tag{17}$$

$$Y_t = g_1(Y_{t-4})Y_{t-1} + g_2(Y_{t-4})Y_{t-2} + .2\varepsilon_t, \tag{18}$$

$$Y_t = g_1(Y_{t-3}, Y_{t-4})Y_{t-1} + g_2(Y_{t-3}, Y_{t-4})Y_{t-2} + .2\varepsilon_t. \tag{19}$$

We fitted the three FAR models with the same series generated by (16). For comparison, we computed the mean average squared residuals (MASR) defined by

$$\mathrm{MASR} = \sum_{r=1}^{200}\sum_{t=1}^{250}\frac{(y_{r,t} - \widehat{y}_{r,t})^2}{200 \times 250},$$

where $y_{r,t}$ denotes the tth observation in the rth replication, and $\widehat{y}_{r,t}$ is the corresponding fitted value based on the underlying model. We note that the average squared residuals (ASR), as a statistic that can be computed from any data, real or simulated, is a very useful measure of goodness of fit. This is illustrated in the next section where ASR is used to select an optimal forecasting model. Thus, MASR is a sensible criterion for comparing different models. Although it varies with the bandwidth, an incorrect model would have an MASR asymptotically greater than that of a correct model by a positive constant, which is of larger magnitude than any variation caused by bandwidth tuning. The three FAR models (17), (18), and (19) gave MASR values of 1.020, .343, and .081, respectively, whereas the VCAR model (16) gave a much smaller value of .075.

We also compared the mean average squared prediction errors (MASPE) of these models. For this, we generated an additional 50 observations for each of the 200 time series of size 250 and computed

$$\mathrm{MASPE} = \sum_{r=1}^{200}\sum_{t=251}^{300}\frac{(y_{r,t} - \widehat{y}_{r,t})^2}{200 \times 50},$$

where $\widehat{y}_{r,t}$ is the predicted value of $y_{r,t}$ based on the estimated model from the first 250 observations. The three FAR models (17), (18), and (19) gave MASPE values of .075, .071, and .062, respectively, whereas the VCAR model (16) gave .059.

5. AN EMPIRICAL EXAMPLE

We illustrate our estimation and testing methods with an analysis of the quarterly (seasonally nonadjusted) West German real GNP data collected from 1960:1 to 1990:4. The data G_t, $1 \le t \le n = 124$, which was compiled by Wolters (1992, p. 424, note 4), is plotted in Figure 1(a). One sees clearly a linear trend and a seasonal pattern. Based on the seasonal unit root test of Franses (1996), we took the first differences of the logs and obtained time series data, D_t, $1 \le t \le n = 124$, which are plotted in Figure 1(b). This time series no longer reveals any linear or higher-order trends, but is obviously seasonal. Following the de-seasonalization procedure of Yang and Tschernig (2002), the sample means of the four seasons, −.065116, .038595, .051829, and .008944, respectively, were calculated and subtracted from D_t so that the de-seasonalized Y_t, $1 \le t \le n = 124$, became the growth rates with respect to the spring season. As such, it is reasonable to assume that the Y_t's satisfy our strict stationarity and mixing conditions. In Figure 4 the data Y_t, $1 \le t \le n = 124$, are plotted as the dotted curve.

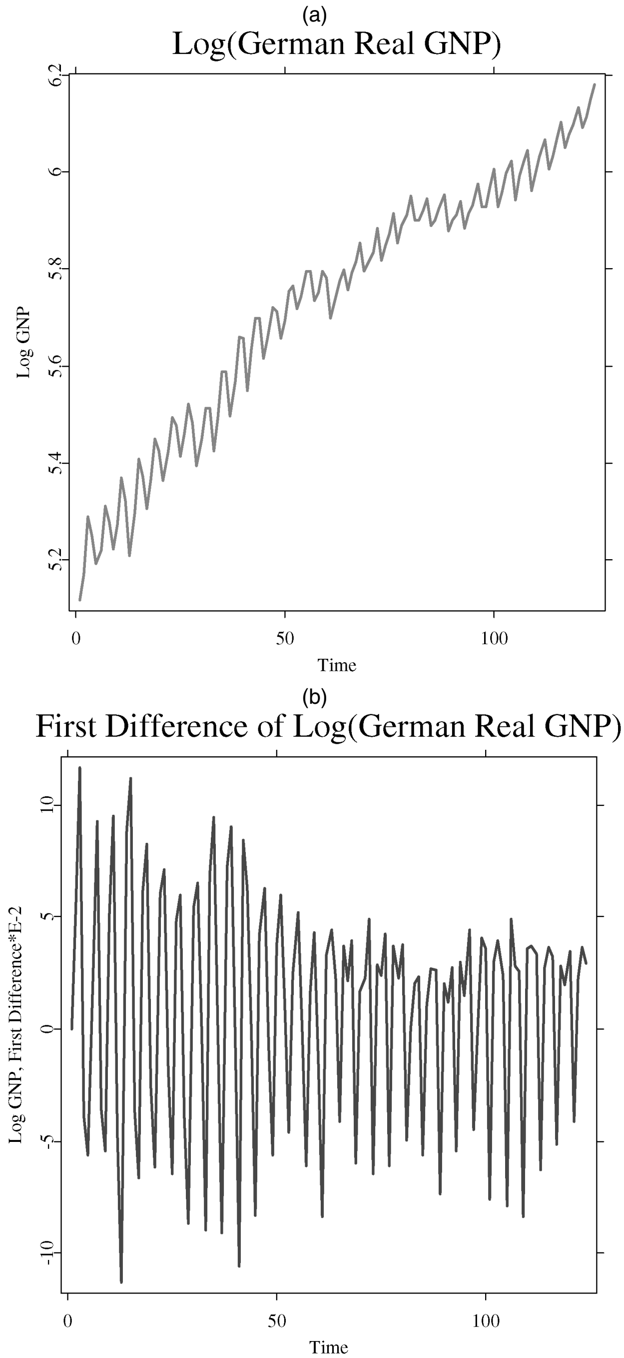

Figure 1. Plots of the West German Real GNP Quarterly Data From 1960:1 to 1990:4. Panel (a) shows log(GNP) over time, and (b) depicts the first difference of log(GNP).

According to the semiparametric lag selection performed in Yang and Tschernig (2002), the significant variables for the prediction of Y_t are Y_{t-4} and Y_{t-2}. Calculation of the autocorrelation functions indicated that Y_t is more correlated with Y_{t-1} and Y_{t-3} than other lagged values. Hence, we fitted all 12 VCAR models of Model I type, consisting of the lagged variables Y_{t-1}, Y_{t-2}, Y_{t-3}, and Y_{t-4}. According to the definition (4) of the marginal integration estimator, we estimated all VCAR models using the first 114 observations and made out-

Table 3. Average Squared Residuals (ASRs) and Average Squared Prediction Errors (ASPEs) Obtained From Fitting 12 VCAR Models With the German Real GNP Data

Model	ASR	ASPE
1234	**.00021**	.00011
1243	.00040	.00019
1324	.00025	.00013
1342	.00039	.00016
1423	.00026	.00014
1432	.00024	**.00009**
2134	.00023	.00017
2143	.00051	.00037
2341	.00049	.00032
2431	.00024	.00015
3142	.00041	.00023
3241	.00038	.00017
Linear AR	.00059	.00041
Partially linear VCAR	.00032	.00024

NOTE: Each model is identified by the four digits that indicate the order in which the lagged variables Y_{t-1}, Y_{t-2}, Y_{t-2}, and Y_{t-4} enter the VCAR model. For example, the model "1234" means $Y_t = f_1(Y_{t-1})Y_{t-2} + f_2(Y_{t-3})Y_{t-4} +$ (noise). The partially linear VCAR model at the bottom is $Y_t = f_1 Y_{t-2} + f_2(Y_{t-3})Y_{t-4} +$ (noise).

of-sample predictions for the last 10 observations. Their average squared residuals (ASRs) and average squared prediction errors (ASPEs) are reported in Table 3. One may expect the ASRs to be smaller than the ASPEs. But we found in the residual plots that there were some very large residual terms that made all the ASRs larger than their corresponding ASPEs. The model with the smallest ASR is

$$Y_t = f_1(Y_{t-1})Y_{t-2} + f_2(Y_{t-3})Y_{t-4} + \text{(noise)}, \qquad (20)$$

whereas the model with the smallest ASPE is

$$Y_t = f_1(Y_{t-3})Y_{t-2} + f_2(Y_{t-1})Y_{t-4} + \text{(noise)}. \qquad (21)$$

Both of these models include as a special case the following linear AR(2) model:

$$Y_t = c_1 Y_{t-2} + c_2 Y_{t-4} + \text{(noise)}. \qquad (22)$$

Table 3 also shows the ASR and ASPE of the linear AR model (22). Both optimal VCAR models (20) and (21) have much smaller ASRs and ASPEs than the linear AR model. These two VCAR models have similar ASR and ASPE values. Figure 2 depicts the estimates of the functions f_1 and f_2 for each model. To test if these functions are significantly different from a constant, we carried out the wild bootstrap procedure. For the model (20), the p values were .80 for f_1 and .01 for f_2, whereas for the model (21), they were .22 and .48, respectively. This means that for the model (20) the function f_1 is not significantly different from a constant, but there is a strong evidence in the data that f_2 is not a constant. Thus, one may conclude that a parsimonious model is the partially linear model:

$$Y_t = f_1 Y_{t-2} + f_2(Y_{t-3})Y_{t-4} + \text{(noise)}.$$

We further computed the ASR and ASPE of this semiparametric partially linear model, which are .00032 and .00024, respectively, as shown in Table 3. In terms of these estimation and forecasting errors, the semiparametric model is much inferior to its nonparametric parent model (20). Thus, the simpler semiparametric model is preferred only for its parsimony, whereas the nonparametric model (20) should be used if optimal forecasting is the goal. The testing for coefficient functions, therefore, works in a similar fashion as the Bayesian information criterion (BIC) works for linear AR time series where ASR is similar to the Akaike information criterion (AIC). For linear AR time series, it is well known that the AIC is optimal for forecasting, whereas the BIC is consistent in identifying a correct

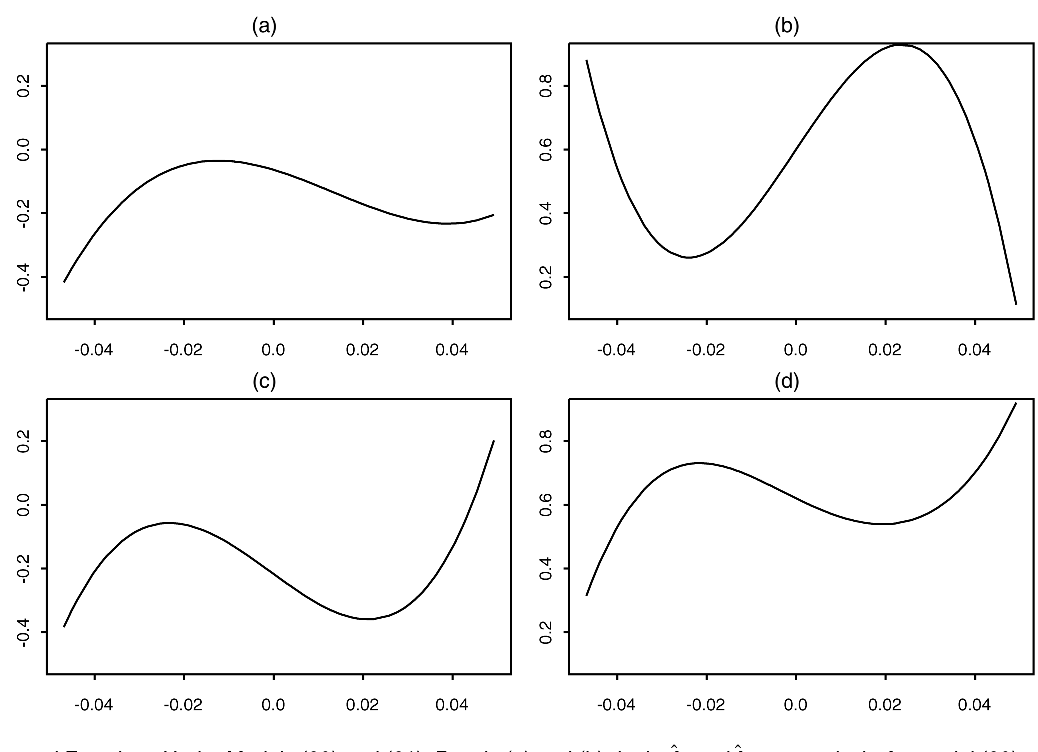

Figure 2. Estimated Functions Under Models (20) and (21). Panels (a) and (b) depict $\hat{f}_1$ and $\hat{f}_2$, respectively, for model (20), whereas (c) and (d) are for model (21).

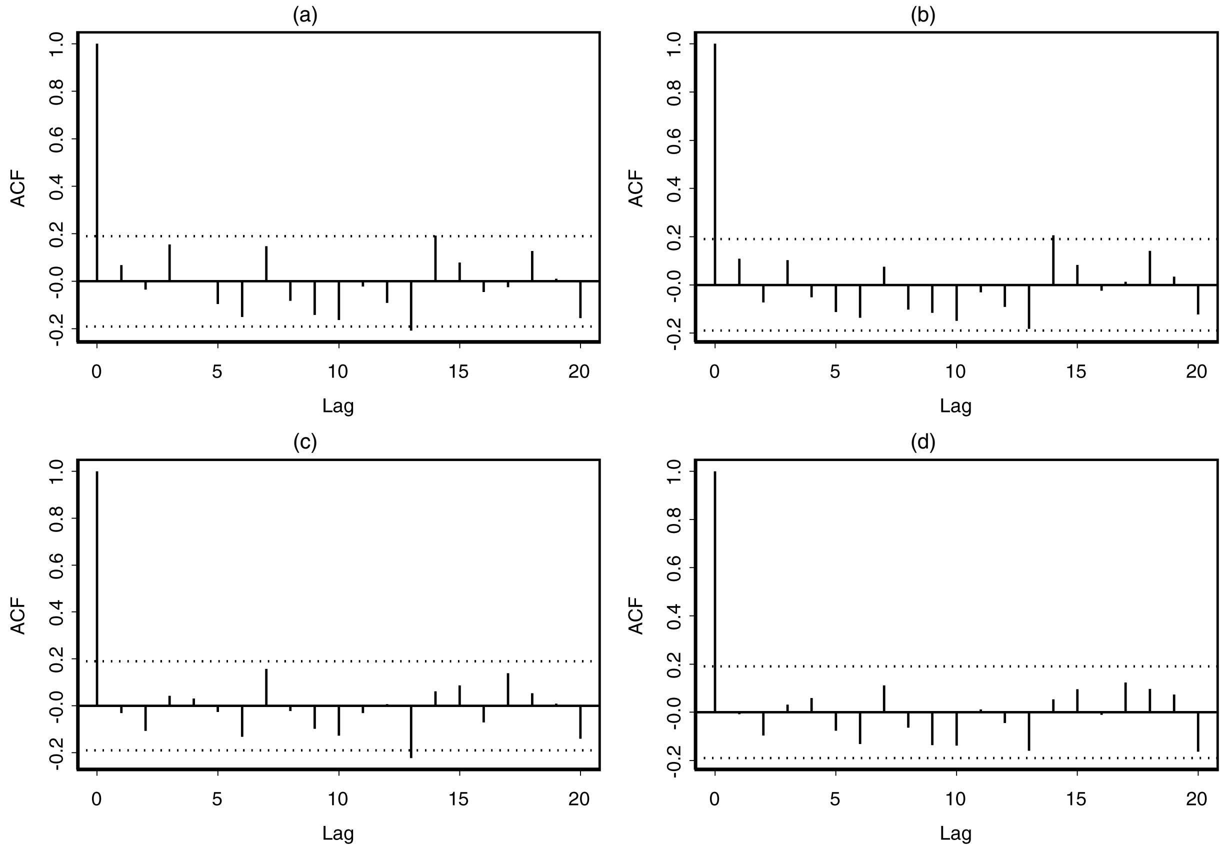

Figure 3. Autocorrelations of Standardized Residuals $\hat{\epsilon}_t$. Panels (a) and (b) are for model (20) and depict the autocorrelations of $|\hat{\varepsilon}_t|$ and $\hat{\varepsilon}_t^2$, respectively, whereas (c) and (d) are for model (21). The dotted horizontal lines at levels $\pm 2 \times n^{-1/2}$ represent the 95.44% confidence bands of the autocorrelation functions.

AR model. It should be noted also that ASRs can be compared across models not necessarily nested within each other, whereas the testing procedure selects the most parsimonious model from a nested hierarchy of models.

To further verify the validity of the models (20) and (21), we examined the residuals $\hat{\varepsilon}_t$ to check the independence of the error terms as another way of assessing the goodness of fit of the models. At a practical level, such independence can be checked using the autocorrelation functions (ACFs) of powers of $|\hat{\varepsilon}_t|$. Figure 3 shows the ACFs of both $|\hat{\varepsilon}_t|$ and $\hat{\varepsilon}_t^2$ for the models (20) and (21). As can be seen from the plots, within the confidence levels of $\pm 2 \times n^{-1/2}$ lie more than 95% of all the sample ACFs, and, hence, we can conclude that both $|\hat{\varepsilon}_t|$ and $\hat{\varepsilon}_t^2$ have no autocorrelation. The ACF plots for $|\hat{\varepsilon}_t|^3$, $\hat{\varepsilon}_t^4$, and so forth, led to the same conclusion. Thus, the models (20) and (21) fit well the structure of the data Y_t. As further evidence, Figure 4 shows the overlay of Y_t together with the predicted series $\hat{Y}_t$ obtained from fitting the models (20) and (21). The predicted series follows the actual series very closely.

APPENDIX: PROOFS

A longer version of this article with proofs of greater detail may be found at *http://stat.snu.ac.kr/theostat/papers/jasa-ypxh.pdf*.

A.1 Preliminaries

We shall need the following technical assumptions on the kernels:

A1. The kernels K and L are symmetric, Lipschitz continuous with $\int K(u)\,du = \int L(u)\,du = 1$ and have compact supports with nonempty interiors. Whereas K is nonnegative, the kernel L is of order q.

When estimating the function f_s for a particular s, a multiplicative kernel is used consisting of K for the sth variable and L for all other variables. To accommodate dependent data, such as those from varying-coefficient autoregression models, we assume that

A2. The vector process $\{(\mathbf{X}_i, \mathbf{T}_i)\}_{i=1}^n$ is strictly stationary and β-mixing with mixing coefficients $\beta(k) \le C_2 \rho^k, 0 < \rho < 1$. Here

$$\beta(n) = \sup_k E \sup\{|P(A|\mathcal{F}_{-\infty}^k) - P(A)| : A \in \mathcal{F}_{n+k}^\infty\},$$

where $\mathcal{F}_t^{t'}$ is the σ-algebra generated by $(\mathbf{X}_t, \mathbf{T}_t), (\mathbf{X}_{t+1}, \mathbf{T}_{t+1}), \ldots, (\mathbf{X}_{t'}, \mathbf{T}_{t'})$ for $t < t'$.

The following assumptions are on the smoothness of the functions involved in the estimation and testing and on the moments of the process for the proofs of Theorems 1, 2, and 5.

A3. The functions f_s have bounded, continuous $(p+1)$th derivatives for all $1 \le s \le d$ and $p \ge q - 1$.

A4. The distribution of $(\mathbf{X}, \mathbf{T})$ has a density ψ and $\mathbf{X}$ has a marginal density φ. On the supports of weight functions w_{-s} and w_s, the densities φ_{-s} of $\mathbf{X}_{-s}$ and φ_s of X_s, respectively, are uniformly bounded away from 0 and ∞. The marginal density φ and $E(T_s T_{s'}|\mathbf{X} = \cdot)$ for $1 \le s, s' \le d$ are Lipschitz continuous. Also, $\sigma^2(\cdot, \mathbf{t})$ and $\psi(\cdot, \mathbf{t})$ are equicontinuous.

A5. The weight functions w_{-s} and w_s are nonnegative, have compact supports with nonempty interiors, and are continuous on their supports.

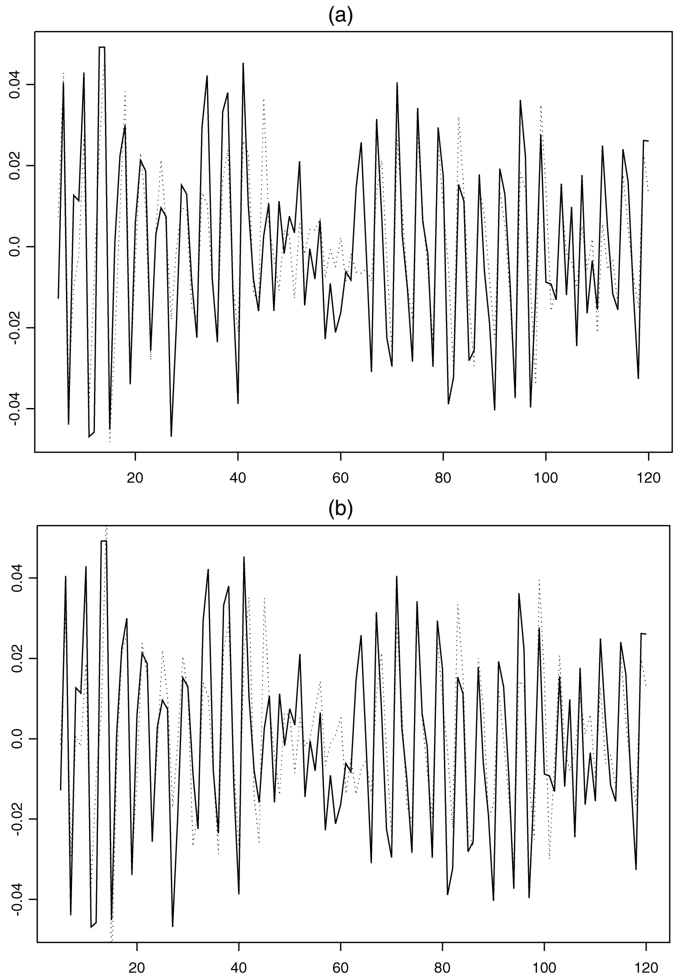

Figure 4. Prediction for the West German Real GNP Quarterly Data Based on the Marginal Integration Fits of the Varying Coefficient Models (20) and (21). Panel (a) is for model (20), and (b) is for (21). Solid lines represent the predicted values $\hat{Y}_t$, whereas the dotted are for the observed values Y_t.

A6. The error term ε_t satisfies $E|\varepsilon_t|^{4+\delta} < \infty$ for some $\delta > 0$. For $j < k < l < m$, there exists a joint probability density function $\psi_{j,k,l,m}$ of $(\mathbf{X}_j, \mathbf{T}_j; \mathbf{X}_k, \mathbf{T}_k; \mathbf{X}_l, \mathbf{T}_l; \mathbf{X}_m, \mathbf{T}_m)$. Let $\mathcal{X} = \{\mathbf{x} : x_s \in \text{supp}(w_s), \mathbf{x}_{-s} \in \text{supp}(w_{-s})\}$, and for $\epsilon > 0$ define $\mathcal{X}_\epsilon = \{\mathbf{x} : \text{there exists } \mathbf{z} \in \mathcal{X} \text{ such that } \|\mathbf{z} - \mathbf{x}\| \le \epsilon\}$. There exist $\epsilon > 0$, $\tilde{\sigma}(\mathbf{t})$, and $\tilde{\varphi}_{j,k,l,m}(\mathbf{t}_j, \mathbf{t}_k, \mathbf{t}_l, \mathbf{t}_m)$ such that $\sigma(\mathbf{x}, \mathbf{t}) \le \tilde{\sigma}(\mathbf{t})$ for all $\mathbf{x} \in \mathcal{X}_\epsilon$, $\psi_{j,k,l,m}(\mathbf{x}_j, \mathbf{t}_j; \mathbf{x}_k, \mathbf{t}_k; \mathbf{x}_l, \mathbf{t}_l; \mathbf{x}_m, \mathbf{t}_m) \le \tilde{\varphi}_{j,k,l,m}(\mathbf{t}_j, \mathbf{t}_k, \mathbf{t}_l, \mathbf{t}_m)$ for all $\mathbf{x}_j, \mathbf{x}_k, \mathbf{x}_l, \mathbf{x}_m$ in $\mathcal{X}_\epsilon$, and $\int (\|\mathbf{t}_j\| \times \|\mathbf{t}_k\| \|\mathbf{t}_l\| \|\mathbf{t}_m\|)^{2+c} |\tilde{\sigma}(\mathbf{t}_j)\tilde{\sigma}(\mathbf{t}_k)\tilde{\sigma}(\mathbf{t}_l)\tilde{\sigma}(\mathbf{t}_m)|^{2+c} \tilde{\varphi}_{j,k,l,m}(\mathbf{t}_j, \mathbf{t}_k, \mathbf{t}_l, \mathbf{t}_m)\, d\mathbf{t}_j\, d\mathbf{t}_k\, d\mathbf{t}_l\, d\mathbf{t}_m \le C < \infty$ for some $c > 0$ and $C > 0$.

Also, we assume that the bandwidths, $\mathbf{g}$ for the kernel L and h for the kernel K, satisfy

A7. $(\ln n)(nhg_{\text{prod}})^{-1/2} = O(n^{-a})$ for some $a > 0$ and $(nh \ln n)^{1/2} g_{\max}^q \to 0$ as $n \to \infty$, where $g_{\text{prod}} = g_1 \cdots g_{s-1} \times g_{s+1} \cdots g_d$ and $g_{\max} = \max(g_1, \ldots, g_{s-1}, g_{s+1}, \ldots, g_d)$, and h is asymptotic to $n^{-1/(2p+3)}$.

For Theorems 3, 4, and 6, we need to modify assumptions A3, A4, and A7 slightly as follows:

A3′. The functions f_{su} have bounded, continuous $(p+1)$th derivatives for all $1 \le s \le d_0$, $1 \le u \le r_s$, and $p \ge q - 1$.

A4′. This is the same as A4 except that now we require $E(T_{su}T_{s'u'} | \mathbf{X} = \cdot)$ for $1 \le s, s' \le d_0$, and $1 \le u, u' \le r_s$ are Lipschitz continuous.

A7′. This is the same as A7 except that d is replaced by d_0.

Note that for the existence of the bandwidth vector $\mathbf{g}$ satisfying assumptions A7 and A7′, it is necessary that q, the order of the kernel L, be larger than $(d-1)/2$ and $(d_0 - 1)/2$, respectively.

To prove many of our results, we make use of some inequalities about the U statistic and the von Mises statistic of dependent variables derived from Yoshihara (1976). Let ξ_i, $1 \le i \le n$, be a strictly stationary sequence of random variables with values in $\mathbb{R}^d$ and β-mixing coefficients $\beta(k)$, $k = 1, 2, \ldots$, and let r be a fixed positive integer. Let $\{\theta_n(F)\}$ denote the functionals of the distribution function F of ξ_i:

$$\theta_n(F) = \int g_n(x_1, \ldots, x_m)\, dF(x_1) \cdots dF(x_m),$$

where $\{g_n\}$ are measurable functions symmetric in their m arguments such that

$$\int |g_n(x_1, \ldots, x_m)|^{2+\delta}\, dF(x_1) \cdots dF(x_m) \le M_n < +\infty,$$

$$\sup_{(i_1,\ldots,i_m)\in S_c} \int |g_n(x_1, \ldots, x_m)|^{2+\delta}\, dF_{\xi_{i_1},\ldots,\xi_{i_m}}(x_1, \ldots, x_m) \le M_{n,c} < +\infty, \qquad c = 0, \ldots, m-1,$$

for some $\delta > 0$, where $S_c = \{(i_1, \ldots, i_m) | \#_r(i_1, \ldots, i_m) = c\}$, $c = 0, \ldots, m-1$, and for every $(i_1, \ldots, i_m)$, $1 \le i_1 \le \cdots \le i_m \le n$, $\#_r(i_1, \ldots, i_m) =$ the number of $j = 1, \ldots, m-1$ satisfying $i_{j+1} - i_j \le r$. Clearly, the cardinality of each set S_c is less than n^{m-c}.

The von Mises differentiable statistic and the U statistic

$$\theta_n(F_n) = \int g_n(x_1, \ldots, x_m)\, dF_n(x_1) \cdots dF_n(x_m) = \frac{1}{n^m} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} g_n(\xi_{i_1}, \ldots, \xi_{i_m}),$$

$$U_n = \frac{1}{\binom{n}{m}} \sum_{1 \le i_1 < \cdots < i_m \le n} g_n(\xi_{i_1}, \ldots, \xi_{i_m}),$$

allow decompositions as

$$\theta_n(F_n) = \theta_n(F) + \sum_{c=1}^{m} \binom{m}{c} V_n^{(c)},$$

$$U_n = \theta_n(F) + \sum_{c=1}^{m} \binom{m}{c} U_n^{(c)}.$$

Here $g_{n,c}$ are the projections of g_n defined by

$$g_{n,c}(x_1, \ldots, x_c) = \int g_n(x_1, \ldots, x_m)\, dF(x_{c+1}) \cdots dF(x_m), \qquad c = 0, 1, \ldots, m,$$

so that $g_{n,0} = \theta_n(F)$, $g_n = g_{n,m}$, and

$$V_n^{(c)} = \int g_{n,c}(x_1, \ldots, x_c) \prod_{j=1}^{c} [dF_n(x_j) - dF(x_j)],$$

$$U_n^{(c)} = \frac{(n-c)!}{n!} \sum_{1 \le i_1 < \cdots < i_c \le n} \int g_{n,c}(x_{i_1}, \ldots, x_{i_c}) \times \prod_{j=1}^{c} \left[dI_{\mathbb{R}_+^d}(x_j - \xi_{i_j}) - dF(x_j)\right],$$

where $I_{\mathbb{R}_+^d}$ is the indicator function of $\mathbb{R}_+^d = \{(y_1,\dots,y_d)\in\mathbb{R}^d | y_j \ge 0, j=1,\dots,d\}$.

Lemma A.1. If $\beta(k)\le C_1 k^{-(2+\delta')/\delta'}$ for $\delta>\delta'>0$, then

$$EV_n^{(c)2}+EU_n^{(c)2} \le C(m,\delta,r)n^{-c}\left\{M_n^{2/(2+\delta)}\sum_{k=r+1}^{n} k\beta^{\delta/(2+\delta)}(k) + \sum_{c'=0}^{m-1} n^{-c'}M_{n,c'}^{\prime 2/(2+\delta)}\sum_{k=1}^{r}k\beta^{\delta/(2+\delta)}(k)\right\} \quad (A.1)$$

for some constant $C(m,\delta,r)>0$. In particular, if $\beta(k)\le C_2\rho^k$ for $0<\rho<1$, then

$$EV_n^{(c)2}+EU_n^{(c)2}\le C(m,\delta,r)C_2C(\rho)n^{-c} \times\left\{M_n^{2/(2+\delta)}+\sum_{c'=0}^{m-1}n^{-c'}M_{n,c'}^{\prime 2/(2+\delta)}\right\}. \quad (A.2)$$

Proof. The proof is essentially the same as that of Lemma 2 in Yoshihara (1976), which dealt with the special case of $g_n\equiv g$, $r=1, M_n=M_n'$ and yielded (A.1). The inequalities in the proof of this lemma do not require all g_n's to be the same for $n=1,2,\dots$, and terms in $U_n^{(c)}$ where exactly c' pairs of neighboring indices differ by at most r form a subset of terms with cardinality of order $n^{c-c'}$. Elementary arguments then establish (A.2) under geometric mixing conditions.

A.2 Proofs of Theorems 1, 2, and 5

Define the following square matrix of dimension $(p+d)$:

$$S_s(\mathbf{x})=\begin{bmatrix}\int\mathbf{p}(u)\mathbf{p}^T(u)K(u)\,du\,E(T_s^2|\mathbf{X}=\mathbf{x}) & \int\mathbf{p}(u)K(u)\,du\,E(T_s\mathbf{T}_{-s}^T|\mathbf{X}=\mathbf{x})\\ E(T_s\mathbf{T}_{-s}|\mathbf{X}=\mathbf{x})\int\mathbf{p}^T(u)K(u)\,du & E(\mathbf{T}_{-s}\mathbf{T}_{-s}^T|\mathbf{X}=\mathbf{x})\end{bmatrix}.$$

The identifiability condition given in (3) is closely related to the invertibility of the matrix $S_s(\mathbf{X})$. To see this, note that, for vectors λ_1 and λ_2 of dimensions $p+1$ and $d-1$, respectively,

$$(\lambda_1^T,\lambda_2^T)S_s(\mathbf{x})(\lambda_1^T,\lambda_2^T)^T = \int E\big[\{\lambda_1^T\mathbf{p}(u)T_s+\lambda_2^T\mathbf{T}_{-s}\}^2|\mathbf{X}=\mathbf{x}\big]K(u)\,du.$$

Thus, if $[\lambda_1(X_s)^T,\lambda_2(\mathbf{X}_{-s})^T]S_s(\mathbf{X})[\lambda_1(X_s)^T,\lambda_2(\mathbf{X}_{-s})^T]^T=0$, a.s., then $\lambda_1(X_s)^T\mathbf{p}(u)T_s+\lambda_2(\mathbf{X}_{-s})^T\mathbf{T}_{-s}=0$ a.s. $(\mathbf{X},\mathbf{T})$ and $u\in\operatorname{supp}(K)$. Because K has a nonempty interior, the identifiability condition (3) implies $\lambda_1\equiv0$ and $\lambda_2\equiv0$ by the uniqueness of polynomial expansion.

The next lemma shows that the matrix $S_s(\mathbf{x})$ is proportional to the limiting dispersion matrix.

Lemma A.2. As $n\to\infty$,

$$\sup_{x_s\in\operatorname{supp}(w_s),\mathbf{x}_{-s}\in\operatorname{supp}(w_{-s})}\left|\mathbf{Z}_s^T\mathbf{W}_s(\mathbf{x}_{-s})\mathbf{Z}_s-\varphi(x_s,\mathbf{x}_{-s})S(x_s,\mathbf{x}_{-s})\right| = o(b) \quad \text{a.s.},$$

where $b=\ln n(h+g_{\max}^q+1/\sqrt{nhg_{\text{prod}}})$.

Proof. The conclusion follows by directly using the covering technique and exponential inequalities for β-mixing processes, as in the proof of theorem 2.2 of Bosq (1998).

Now let c be an integer such that $b^{c+1}=o(h^{p+2})$. The next lemma decomposes the dispersion matrix.

Lemma A.3. For any integer k,

$$\big(\mathbf{Z}_s^T\mathbf{W}_s(\mathbf{x}_{-s})\mathbf{Z}_s\big)^{-1}-\frac{S_s^{-1}(x_s,\mathbf{x}_{-s})}{\varphi(x_s,\mathbf{x}_{-s})} = \frac{S_s^{-1}(x_s,\mathbf{x}_{-s})}{\varphi(x_s,\mathbf{x}_{-s})}\sum_{\ell=1}^{c}\left\{I_{p+d}-\frac{\mathbf{Z}_s^T\mathbf{W}_s(\mathbf{x}_{-s})\mathbf{Z}_sS_s^{-1}(x_s,\mathbf{x}_{-s})}{\varphi(x_s,\mathbf{x}_{-s})}\right\}^{\ell} + R_s(x_s,\mathbf{x}_{-s})$$

as $n\to\infty$, where the matrix $R_s(x_s,\mathbf{x}_{-s})$ satisfies

$$\sup_{x_s\in\operatorname{supp}(w_s),\mathbf{x}_{-s}\in\operatorname{supp}(w_{-s})}|R_s(x_s,\mathbf{x}_{-s})|=o(h^{p+2}) \quad \text{a.s.}$$

Proof. By a Taylor expansion of the matrix inversion operation, Lemma A.2 immediately yields the result.

Lemma A.4. Define

$$D_{s1}(x_s)=\frac1n\sum_{i=1}^n w_{-s}(\mathbf{X}_{i,-s})R_s(x_s,\mathbf{X}_{i,-s})\mathbf{Z}_s^T\mathbf{W}_{is}\mathbf{E},$$

$$D_{s2}(x_s)=\frac1n\sum_{i=1}^n w_{-s}(\mathbf{X}_{i,-s})R_s(x_s,\mathbf{X}_{i,-s})\mathbf{Z}_s^T\mathbf{W}_{is} \times\left[\{f_s(X_{js})\}_{j=1}^n-\sum_{\nu=0}^{p}\frac{f_s^{(\nu)}(x_s)h^\nu}{\nu!}\mathbf{Z}_se_\nu\right],$$

$$D_{s3}(x_s)=\frac1n\sum_{i=1}^n w_{-s}(\mathbf{X}_{i,-s})R_s(x_s,\mathbf{X}_{i,-s})\mathbf{Z}_s^T\mathbf{W}_{is} \times\left[\left\{\sum_{s'\ne s}f_{s'}(X_{js'})\right\}_{j=1}^n-\sum_{s'\ne s}f_{s'}(X_{is'})\mathbf{Z}_se_{p+s'}\right].$$

Then, as $n\to\infty$,

$$\sup_{x_s\in\operatorname{supp}(w_s)}\big\{|D_{s1}(x_s)|+|D_{s2}(x_s)|+|D_{s3}(x_s)|\big\}=o(h^{p+2}) \quad \text{a.s.}$$

Proof. The lemma follows directly from Lemma A.3.

Lemma A.5. Write $\mathbf{W}_{is}=\mathbf{W}_s(\mathbf{X}_{i,-s})$ and $\mathbf{E}=\{\sigma(\mathbf{X}_1,\mathbf{T}_1)\varepsilon_1,\dots,\sigma(\mathbf{X}_n,\mathbf{T}_n)\varepsilon_n\}^T$. For $\ell=1,2,\dots$, define

$$R_{\ell1}(x_s)=\frac1n\sum_{i=1}^n\frac{w_{-s}(\mathbf{X}_{i,-s})}{\varphi(x_s,\mathbf{X}_{i,-s})}S_s^{-1}(x_s,\mathbf{X}_{i,-s}) \times\left\{I_{p+d}-\frac{\mathbf{Z}_s^T\mathbf{W}_{is}\mathbf{Z}_sS_s^{-1}(x_s,\mathbf{X}_{i,-s})}{\varphi(x_s,\mathbf{X}_{i,-s})}\right\}^{\ell}\mathbf{Z}_s^T\mathbf{W}_{is}\mathbf{E}, \quad (A.3)$$

$$R_{\ell2}(x_s)=\frac1n\sum_{i=1}^n\frac{w_{-s}(\mathbf{X}_{i,-s})}{\varphi(x_s,\mathbf{X}_{i,-s})}S_s^{-1}(x_s,\mathbf{X}_{i,-s}) \times\left\{I_{p+d}-\frac{\mathbf{Z}_s^T\mathbf{W}_{is}\mathbf{Z}_sS_s^{-1}(x_s,\mathbf{X}_{i,-s})}{\varphi(x_s,\mathbf{X}_{i,-s})}\right\}^{\ell} \times\mathbf{Z}_s^T\mathbf{W}_{isis}\left[\{f_s(X_{js})\}_{j=1}^n-\sum_{\nu=0}^{p}\frac{f_s^{(\nu)}(x_s)h^\nu}{\nu!}\mathbf{Z}_se_\nu\right], \quad (A.4)$$

$$R_{\ell3}(x_s)=\frac1n\sum_{i=1}^n\frac{w_{-s}(\mathbf{X}_{i,-s})}{\varphi(x_s,\mathbf{X}_{i,-s})}S_s^{-1}(x_s,\mathbf{X}_{i,-s}) \times\left\{I_{p+d}-\frac{\mathbf{Z}_s^T\mathbf{W}_{is}\mathbf{Z}_sS_s^{-1}(x_s,\mathbf{X}_{i,-s})}{\varphi(x_s,\mathbf{X}_{i,-s})}\right\}^{\ell} \times\mathbf{Z}_s^T\mathbf{W}_{is}\left[\left\{\sum_{s'\ne s}f_{s'}(X_{js'})\right\}_{j=1}^n-\sum_{s'\ne s}f_{s'}(X_{is'})\mathbf{Z}_se_{p+s'}\right]. \quad (A.5)$$

Then, as $n \to \infty$,

$$|R_{\ell 1}(x_s)| + |R_{\ell 2}(x_s)| + |R_{\ell 3}(x_s)| = o_p(b^\ell/\sqrt{nh}). \qquad \text{(A.6)}$$

Proof. For simplicity, consider the case $R_{\ell 1}(x_s)$ and only $\ell = 1$. The term $R_{\ell 1}(x_s)$ equals $P_1 - P_2$, where

$$\begin{aligned} P_1 &= \frac{1}{n}\sum_{i=1}^n w_{-s}(\mathbf{X}_{i,-s}) S_s^{-1}(x_s, \mathbf{X}_{i,-s}) \\ &\quad \times \left\{ \frac{S(x_s, \mathbf{X}_{i,-s})}{\varphi(x_s, \mathbf{X}_{i,-s})} - \frac{E(\mathbf{Z}_s^T \mathbf{W}_{is} \mathbf{Z}_s | x_s, \mathbf{X}_{i,-s})}{\varphi^2(x_s, \mathbf{X}_{i,-s})} \right\} \\ &\quad \times S_s^{-1}(x_s, \mathbf{X}_{i,-s}) \mathbf{Z}_s^T \mathbf{W}_{is} \mathbf{E}, \\ P_2 &= \frac{1}{n}\sum_{i=1}^n w_{-s}(\mathbf{X}_{i,-s}) S_s^{-1}(x_s, \mathbf{X}_{i,-s}) \\ &\quad \times \left\{ \frac{\mathbf{Z}_s^T \mathbf{W}_{is} \mathbf{Z}_s}{\varphi(x_s, \mathbf{X}_{i,-s})} - \frac{E(\mathbf{Z}_s^T \mathbf{W}_{is} \mathbf{Z}_s | x_s, \mathbf{X}_{i,-s})}{\varphi^2(x_s, \mathbf{X}_{i,-s})} \right\} \\ &\quad \times S_s^{-1}(x_s, \mathbf{X}_{i,-s}) \mathbf{Z}_s^T \mathbf{W}_{is} \mathbf{E}. \end{aligned}$$

Denote $\boldsymbol{\xi}_i = (\mathbf{X}_i, \mathbf{T}_i, Y_i)$. The term P_1 can be written as the von Mises differentiable statistic $V_n = (2n^2)^{-1} \sum_{i,j=1}^n g_n(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j)$, where $g_n(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j)$ equals

$$\begin{aligned} & w_{-s}(\mathbf{X}_{i,-s}) S_s^{-1}(x_s, \mathbf{X}_{i,-s}) \\ &\quad \times \left\{ \frac{S(x_s, \mathbf{X}_{i,-s})}{\varphi(x_s, \mathbf{X}_{i,-s})} - \frac{E(\mathbf{Z}_s^T \mathbf{W}_{is} \mathbf{Z}_s | x_s, \mathbf{X}_{i,-s})}{\varphi^2(x_s, \mathbf{X}_{i,-s})} \right\} \\ &\quad \times S_s^{-1}(x_s, \mathbf{X}_{i,-s}) \\ &\quad \times \begin{pmatrix} T_{js} K_h(X_{js} - x_s) L_{\mathbf{g}}(\mathbf{X}_{j,-s} - \mathbf{X}_{i,-s}) \sigma(\mathbf{X}_j, \mathbf{T}_j)\varepsilon_j \\ \mathbf{p}\{(X_{js} - x_s)/h\} T_{js} K_h(X_{js} - x_s) L_{\mathbf{g}}(\mathbf{X}_{j,-s} - \mathbf{X}_{i,-s}) \sigma(\mathbf{X}_j, \mathbf{T}_j)\varepsilon_j \\ \mathbf{T}_{j,-s} K_h(X_{js} - x_s) L_{\mathbf{g}}(\mathbf{X}_{j,-s} - \mathbf{X}_{i,-s}) \sigma(\mathbf{X}_j, \mathbf{T}_j)\varepsilon_j \end{pmatrix} \\ &\quad + w_{-s}(\mathbf{X}_{j,-s}) S_s^{-1}(x_s, \mathbf{X}_{j,-s}) \\ &\quad \times \left\{ \frac{S(x_s, \mathbf{X}_{j,-s})}{\varphi(x_s, \mathbf{X}_{j,-s})} - \frac{E(\mathbf{Z}_s^T \mathbf{W}_{js} \mathbf{Z}_s | x_s, \mathbf{X}_{j,-s})}{\varphi^2(x_s, \mathbf{X}_{j,-s})} \right\} \\ &\quad \times S_s^{-1}(x_s, \mathbf{X}_{j,-s}) \\ &\quad \times \begin{pmatrix} T_{is} K_h(X_{is} - x_s) L_{\mathbf{g}}(\mathbf{X}_{i,-s} - \mathbf{X}_{j,-s}) \sigma(\mathbf{X}_i, \mathbf{T}_i)\varepsilon_i \\ \mathbf{p}\{(X_{is} - x_s)/h\} T_{js} K_h(X_{is} - x_s) L_{\mathbf{g}}(\mathbf{X}_{i,-s} - \mathbf{X}_{j,-s}) \sigma(\mathbf{X}_i, \mathbf{T}_i)\varepsilon_i \\ \mathbf{T}_{i,-s} K_h(X_{is} - x_s) L_{\mathbf{g}}(\mathbf{X}_{i,-s} - \mathbf{X}_{j,-s}) \sigma(\mathbf{X}_i, \mathbf{T}_i)\varepsilon_i \end{pmatrix}. \end{aligned}$$

First, we calculate that $g_{n,0} = 0$ and $g_{n,1}(\boldsymbol{\xi}_j)$ equals

$$\begin{aligned} & \int S_s^{-1}(x_s, \mathbf{z}_{-s}) w_{-s}(\mathbf{z}_{-s}) S_s^{-1}(x_s, \mathbf{z}_{-s}) \\ &\quad \times \left\{ \frac{S(x_s, \mathbf{z}_{-s})}{\varphi(x_s, \mathbf{z}_{-s})} - \frac{E(\mathbf{Z}_s^T \mathbf{W}_{is} \mathbf{Z}_s | x_s, \mathbf{z}_{-s})}{\varphi^2(x_s, \mathbf{z}_{-s})} \right\} \\ &\quad \times S_s^{-1}(x_s, \mathbf{z}_{-s}) \\ &\quad \times \begin{pmatrix} T_{js} K_h(X_{js} - x_s) L_{\mathbf{g}}(\mathbf{X}_{j,-s} - \mathbf{z}_{-s}) \sigma(\mathbf{X}_j, \mathbf{T}_j)\varepsilon_j \\ \mathbf{p}\{(X_{js} - x_s)/h\} T_{js} K_h(X_{js} - x_s) L_{\mathbf{g}}(\mathbf{X}_{j,-s} - \mathbf{z}_{-s}) \sigma(\mathbf{X}_j, \mathbf{T}_j)\varepsilon_j \\ \mathbf{T}_{j,-s} K_h(X_{js} - x_s) L_{\mathbf{g}}(\mathbf{X}_{j,-s} - \mathbf{z}_{-s}) \sigma(\mathbf{X}_j, \mathbf{T}_j)\varepsilon_j \end{pmatrix} \\ &\quad \times \varphi_{-s}(\mathbf{z}_{-s})\, d\mathbf{z}_{-s}, \end{aligned}$$

which has mean 0 and variance of order b^2/nh. So $V_n^{(1)} = n^{-1} \times \sum_{j=1}^n g_{n,1}(\boldsymbol{\xi}_j) = o_p(b/\sqrt{nh})$. Next, take a small constant $\delta > 0$. Then the $(2+\delta)$th moment of $g_n(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j)$, $i < j$, is not greater than

$$\begin{aligned} & Cb^{2+\delta} C(\rho) \\ &\quad \times \left(\frac{1}{g_{\text{prod}}^{1+2\delta}} \times E \left| \begin{matrix} T_{js} K_h(X_{js} - x_s) \sigma(\mathbf{X}_j, \mathbf{T}_j)\varepsilon_j \\ \mathbf{p}\{(X_{js} - x_s)/h\} T_{js} K_h(X_{js} - x_s) \sigma(\mathbf{X}_j, \mathbf{T}_j)\varepsilon_j \\ \mathbf{T}_{j,-s} K_h(X_{js} - x_s) \sigma(\mathbf{X}_j, \mathbf{T}_j)\varepsilon_j \end{matrix} \right|^{2+2\delta} \right)^{(2+\delta)/(2+2\delta)} \\ &\leq Cb^{2+\delta} C(\rho) \left(\frac{1}{h^{1+2\delta} g_{\text{prod}}^{1+2\delta}} \right)^{(2+\delta)/(2+2\delta)} \end{aligned}$$

by lemma 1 of Yoshihara (1976).

Hence, we can take $M_n = M_{n,0} = Cb^{2+\delta}(h^{1+2\delta} \times g_{\text{prod}}^{1+2\delta})^{-(2+\delta)/(2+2\delta)}$ in the context of Lemma A.1 with $m = c = 2$ and $r = 1$. Similarly, we can show that $M_{n,1} = Cb^{2+\delta} h^{-(1+\delta)} g_{\text{prod}}^{-(2+\delta)}$. By applying Lemma A.1 with $m = c = 2$ and $r = 1$, (A.2) gives

$$\begin{aligned} EP_1^2 &\leq Cn^{-2} b^2 (hg_{\text{prod}})^{-2(1+2\delta)/(2+2\delta)} \\ &\qquad + Cn^{-3} b^2 h^{-(1+\delta)2/(2+\delta)} g_{\text{prod}}^{-(2+\delta)2/(2+\delta)} + Cb^2/nh \\ &\leq Cn^{-1} h^{-1} b^2 \end{aligned}$$

by making δ sufficiently small. Similar arguments establish that $EP_2^2 \leq Cn^{-1}h^{-1}b^2$. Hence, $P_1 - P_2 = o_p(b/\sqrt{nh})$. We have thus concluded the proof of the lemma.

Now write $\mathbf{q}_s(u; \mathbf{t})$ for the $(p+d)$-dimensional vector given by

$$\mathbf{q}_s(u; \mathbf{t})^T = (\mathbf{p}(u)^T t_s, \mathbf{t}_{-s}^T) = (t_s, ut_s, \ldots, u^p t_s, \mathbf{t}_{-s}^T),$$

and define an equivalent kernel

$$K_s^*(u; \mathbf{t}, \mathbf{x}) = e_0^T S_s^{-1}(\mathbf{x}) \mathbf{q}_s(u; \mathbf{t}) K(u). \qquad \text{(A.7)}$$

Write $K_{s,h}^*(u; \mathbf{t}, \mathbf{x}) = (1/h) K_s^*(u/h; \mathbf{t}, \mathbf{x})$, that is,

$$K_{s,h}^*(u; \mathbf{t}, \mathbf{x}) = (1/h) e_0^T S_s^{-1}(\mathbf{x}) \mathbf{q}_s(u/h; \mathbf{t}) K(u/h). \qquad \text{(A.8)}$$

This kernel satisfies the moment conditions given in the following lemma, which follows directly from the definition of $S_s(\mathbf{x})$ and $S_s^{-1}(\mathbf{x})$.

Lemma A.6. Let δ_{jk} equal 1 if $j = k$ and 0 otherwise. Then

$$\begin{aligned} E\left\{ \int u^q T_s K_s^*(u; \mathbf{T}, \mathbf{X})\, du \middle| \mathbf{X} = \mathbf{x} \right\} &= \delta_{0q}, \qquad 0 \leq q \leq p, \\ E\left\{ \int T_{s'} K_s^*(u; \mathbf{T}, \mathbf{X})\, du \middle| \mathbf{X} = \mathbf{x} \right\} &= 0, \qquad s' = 1, \ldots, d, s' \neq s. \end{aligned} \qquad \text{(A.9)}$$

To prove Theorem 1, we begin by observing

$$e_0^T (\mathbf{Z}_s^T \mathbf{W}_{is} \mathbf{Z}_s)^{-1} \mathbf{Z}_s^T \mathbf{W}_{is} \mathbf{Z}_s e_l = \delta_{0l}, \qquad l = 0, \ldots, p+d-1.$$

Define $Q_{1n} = \sum_{i=1}^n w_{-s}(\mathbf{X}_{i,-s})/n$ and

$$\begin{aligned} Q_{2n}(x_s) &= n^{-1} \sum_{i=1}^n w_{-s}(\mathbf{X}_{i,-s}) e_0^T (\mathbf{Z}_s^T \mathbf{W}_{is} \mathbf{Z}_s)^{-1} \mathbf{Z}_s^T \mathbf{W}_{is} \\ &\qquad \times \left\{ \mathbf{Y} - \sum_{\nu=0}^p \frac{f_s^{(\nu)}(x_s) h^\nu}{\nu!} \mathbf{Z}_s e_\nu - \sum_{s' \neq s}^d f_{s'}(X_{is'}) \mathbf{Z}_s e_{p+s'} \right\}. \end{aligned}$$

Then we obtain $Q_{1n}\{\hat{f}_s(x_s) - f_s(x_s)\} = Q_{2n}(x_s)$. By Lemmas A.5, A.4, and A.3 and by the definition of $K_{s,h}^*(u, \mathbf{t}; \mathbf{x})$ in (A.8), we now write

$$Q_{2n}(x_s) = \sum_{a=1}^3 \left\{ P_{an}(x_s) + \sum_{l=1}^c R_{la}(x_s) + D_{sa}(x_s) \right\}, \qquad \text{(A.10)}$$

where, for $a = 1, 2, 3$,

$$P_{an}(x_s) = n^{-2} \sum_{i,j=1}^{n} \frac{w_{-s}(\mathbf{X}_{i,-s})}{\varphi(x_s, \mathbf{X}_{i,-s})} K^*_{s,h}(X_{js} - x_s; \mathbf{T}_j, x_s, \mathbf{X}_{i,-s}) \times L_{\mathbf{g}}(\mathbf{X}_{j,-s} - \mathbf{X}_{i,-s}) H_{js}, \quad \text{(A.11)}$$

with H_{js} being $\sigma(\mathbf{X}_j, \mathbf{T}_j)\varepsilon_j$ for $a = 1$, $\{f_s(X_{js}) - \sum_{\nu=0}^{p} f_s^{(\nu)}(x_s)(X_{js} - x_s)^\nu/\nu!\}T_{js}$ for $a = 2$, and $\sum_{s'=1, s' \neq s}^{d} \{f_{s'}(X_{js'}) - f_{s'}(X_{is'})\}T_{js'}$.

In the following three lemmas, we derive the asymptotics for P_{1n}, P_{2n}, and P_{3n}.

Lemma A.7. As $n \to \infty$,

$$P_{1n}(x_s) = n^{-1} \sum_{j=1}^{n} p_{js}(x_s)\varepsilon_j + o_p\{(nh \log n)^{-1/2}\},$$

where $p_{js}(x_s) = w_{-s}(\mathbf{X}_{j,-s}) K^*_{s,h}(X_{js} - x_s; \mathbf{T}_j, x_s, \mathbf{X}_{j,-s}) \varphi_{-s}(\mathbf{X}_{j,-s}) \times \sigma(\mathbf{X}_j, \mathbf{T}_j)/\varphi(x_s, \mathbf{X}_{j,-s})$.

Proof. By definition (A.11) and using Lemma A.1 for geometrically β-mixing processes, we have

$$P_{1n}(x_s) = n^{-1} \sum_{j=1}^{n} \int \frac{w_{-s}(\mathbf{x}_{-s})}{\varphi(x_s, \mathbf{x}_{-s})} K^*_{s,h}(X_{js} - x_s; \mathbf{T}_j, x_s, \mathbf{x}_{-s}) \times L_{\mathbf{g}}(\mathbf{X}_{j,-s} - \mathbf{x}_{-s}) \varphi_{-s}(\mathbf{x}_{-s})\, d\mathbf{x}_{-s} \sigma(\mathbf{X}_j, \mathbf{T}_j)\varepsilon_j + o_p\{(nh \log n)^{-1/2}\}.$$

By the change of variable $\mathbf{x}_{-s} = \mathbf{X}_{j,-s} - \mathbf{g}\mathbf{v}_{-s}$ and the fact that L is of order q, it equals

$$n^{-1} \sum_{j=1}^{n} \frac{w_{-s}(\mathbf{X}_{j,-s})}{\varphi(x_s, \mathbf{X}_{j,-s})} K^*_{s,h}(X_{js} - x_s; \mathbf{T}_j, x_s, \mathbf{X}_{j,-s}) \times \varphi_{-s}(\mathbf{X}_{j,-s}) \sigma(\mathbf{X}_j, \mathbf{T}_j)\varepsilon_j + o_p\{(nh \log n)^{-1/2}\}.$$

This completes the proof of the lemma.

Lemma A.8. As $n \to \infty$, $P_{2n}(x_s) = \kappa_s(x_s)h^{p+1} + o_p(h^{p+1})$, where

$$\kappa_s(x_s) = (p+1)!^{-1} f_s^{(p+1)}(x_s) \times \int u^{p+1} E\{w_{-s}(\mathbf{X}_{-s}) T_s K^*_s(u; \mathbf{T}, x_s, \mathbf{X}_{-s})\}\, du.$$

Proof. By definition (A.11) and again using Lemma A.1, we derive

$$P_{2n}(x_s) = \int \frac{w_{-s}(\mathbf{x}_{-s})}{\varphi(x_s, \mathbf{x}_{-s})} K^*_{s,h}(z_s - x_s; \mathbf{t}, x_s, \mathbf{x}_{-s}) L_{\mathbf{g}}(\mathbf{z}_{-s} - \mathbf{x}_{-s}) \times \left\{ f_s(z_s) - \sum_{\nu=0}^{p} f_s^{(\nu)}(x_s)(z_s - x_s)^\nu/\nu! \right\} \times t_s \psi(\mathbf{z}, \mathbf{t}) \varphi_{-s}(\mathbf{x}_{-s})\, d\mathbf{z}\, d\mathbf{t}\, d\mathbf{x}_{-s} \{1 + o_p(1)\}.$$

By the changes of variables $z_s = x_s + hu$ and $\mathbf{z}_{-s} = \mathbf{x}_{-s} + \mathbf{g}\mathbf{v}_{-s}$, we obtain

$$\begin{aligned} P_{2n}(x_s) &= h^{p+1}(p+1)!^{-1} \times \int \frac{w_{-s}(\mathbf{x}_{-s})}{\varphi(x_s, \mathbf{x}_{-s})} K^*_s(u; \mathbf{t}, x_s, \mathbf{x}_{-s}) f_s^{(p+1)}(x_s) u^{p+1} t_s \\ &\quad \times \varphi_{-s}(\mathbf{x}_{-s}) \psi(x_s, \mathbf{x}_{-s}, \mathbf{t})\, du\, d\mathbf{x}_{-s}\, d\mathbf{t} \{1 + o_p(1)\} \\ &= h^{p+1}(p+1)!^{-1} f_s^{(p+1)}(x_s) \times E\left[w_{-s}(\mathbf{X}_{-s}) \int K^*_s(u; \mathbf{t}, x_s, \mathbf{X}_{-s}) u^{p+1} t_s \times \psi(\mathbf{t}|x_s, \mathbf{X}_{-s})\, du\, d\mathbf{t} \right] + o_p(h^{p+1}). \end{aligned}$$

This completes the proof of the lemma.

Lemma A.9. As $n \to \infty$, $P_{3n}(x_s) = O_p(g^q_{\max})$.

Proof. By definition (A.11) and applying Lemma A.1, we obtain

$$P_{3n}(x_s) = \int \frac{w_{-s}(\mathbf{x}_{-s})}{\varphi(x_s, \mathbf{x}_{-s})} K^*_{s,h}(z_s - x_s; \mathbf{t}, x_s, \mathbf{x}_{-s}) \times L_{\mathbf{g}}(\mathbf{z}_{-s} - \mathbf{x}_{-s}) \times \left[\sum_{s' \neq s} \{f_{s'}(z_{s'}) - f_{s'}(x_{s'})\} t_{s'} \right] \psi(\mathbf{z}, \mathbf{t}) \varphi_{-s}(\mathbf{x}_{-s})\, d\mathbf{z}\, d\mathbf{t}\, d\mathbf{x}_{-s} \times \{1 + o_p(1)\}.$$

After the changes of variables $\mathbf{z}_{-s} = \mathbf{x}_{-s} + \mathbf{g}\mathbf{v}_{-s}$ and $z_s = x_s + hu$, we get

$$\begin{aligned} P_{3n}(x_s) &= \int \frac{w_{-s}(\mathbf{x}_{-s})}{\varphi(x_s, \mathbf{x}_{-s})} K^*_s(u; \mathbf{t}, x_s, \mathbf{x}_{-s}) L(\mathbf{v}_{-s}) \times \left[\sum_{s' \neq s} \{f_{s'}(x_{s'} + g_{s'} v_{s'}) - f_{s'}(x_{s'})\} t_{s'} \right] \\ &\quad \times \psi(x_s + hu, \mathbf{x}_{-s} + \mathbf{g}\mathbf{v}_{-s}, \mathbf{t}) \varphi_{-s}(\mathbf{x}_{-s})\, du\, d\mathbf{v}_{-s}\, d\mathbf{t}\, d\mathbf{x}_{-s} \times \{1 + o_p(1)\} \\ &= O_p(g^q_{\max}) \end{aligned}$$

because L is of order q by assumption A1. Thus, we have proved the lemma.

Proof of Theorem 1. By Lemma A.7 and the martingale central limit theorem of Liptser and Shirjaev (1980), $\sqrt{nh} P_{1n}(x_s)$ for each $x_s \in \operatorname{supp}(w_s)$ is asymptotically normal with mean 0 and variance

$$h \int \frac{w^2_{-s}(\mathbf{z}_{-s})}{\varphi^2(x_s, \mathbf{z}_{-s})} K^{*2}_{s,h}(z_s - x_s; \mathbf{t}, x_s, \mathbf{z}_{-s}) \times \varphi^2_{-s}(\mathbf{z}_{-s}) \sigma^2(\mathbf{z}, \mathbf{t}) \psi(\mathbf{z}, \mathbf{t})\, d\mathbf{z}\, d\mathbf{t} \{1 + o(1)\}.$$

By the change of variable $z_s = x_s + hu$, the leading term of this equals

$$\tau_s^2(x_s) = \int \frac{w^2_{-s}(\mathbf{z}_{-s})}{\varphi^2(x_s, \mathbf{z}_{-s})} K^{*2}_s(u; \mathbf{t}, x_s, \mathbf{z}_{-s}) \varphi^2_{-s}(\mathbf{z}_{-s}) \times \sigma^2(x_s, \mathbf{z}_{-s}, \mathbf{t}) \psi(x_s, \mathbf{z}_{-s}, \mathbf{t})\, du\, d\mathbf{z}_{-s}\, d\mathbf{t}.$$

The theorem now follows immediately from Lemmas A.7 and A.8, the conditions on the bandwidths as given in assumption A7, and the fact that $Q_{1n} = \int w_{-s}(\mathbf{z}_{-s}) \varphi_{-s}(\mathbf{z}_{-s})\, d\mathbf{z}_{-s} + O_p(n^{-1/2})$.

Proof of Theorem 2. First, note that (8) follows directly from (7), so we will only show the latter. Now, from Lemmas A.7–A.9 and the conditions on the bandwidths, we obtain

$$\hat{f}_s(x_s) - f_s(x_s) = b_s(x_s)h^{p+1} + n^{-1}\eta_s^{-1} \sum_{j=1}^{n} p_{js}(x_s)\varepsilon_j + o_p(h^{p+1}). \quad \text{(A.12)}$$

Applying (A.12), we only need to show that the two stochastic terms $n^{-1}\sum_{j=1}^{n} p_{js}(x_s)\varepsilon_j$ and $n^{-1}\sum_{j=1}^{n} p_{js'}(x_{s'})\varepsilon_j$ for $s \neq s'$ have covariance of order $o(n^{-1}h^{-1})$. Noting that the ε_j's are iid white noise and each ε_i is independent of the vectors $(\mathbf{X}_j, \mathbf{T}_j), j = 1, \ldots, i$, for each $i = 1, \ldots, n$, we need only show that

$$E\{p_{js}(x_s) p_{js'}(x_{s'})\} = o(h^{-1}). \quad \text{(A.13)}$$

By a change-of-variable technique for X_s and $X_{s'}$ which are contained in $p_{js}(x_s)$ and $p_{js'}(x_{s'})$, respectively, we can show that the left side of (A.13) is actually $O(1)$, which proves the theorem.

Proof of Theorem 5. For this proof, we again use (A.10). Under the hypothesis (11), $P_{2n}(x_s) = R_{l2}(x_s) = D_{s2}(x_s) = 0$, and, thus,

$$Q_{1n}\{\hat{f}_s(x_s) - \alpha\} = P_{1n}(x_s) + \sum_{l=1}^{c} R_{l1}(x_s) + D_{s1}(x_s) + P_{3n}(x_s) + \sum_{l=1}^{c} R_{l3}(x_s) + D_{s3}(x_s).$$

Hence, to study $\sum_{k=1}^{n} \hat{f}_s(X_{ks})^2 w_s(X_{ks})/n$, we derive the asymptotics of such as $\sum_{k=1}^{n} w_s(X_{ks})\, P_{1n}^2(X_{ks})/n$. Let $\boldsymbol{\xi}_i = (\mathbf{X}_i, \mathbf{T}_i, Y_i)$ and define

$$\begin{aligned}
&\widetilde{g}_n(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j, \boldsymbol{\xi}_k, \boldsymbol{\xi}_l, \boldsymbol{\xi}_m) \\
&= w_s(X_{ks}) \frac{w_{-s}(\mathbf{X}_{i,-s})}{\varphi(X_{ks}, \mathbf{X}_{i,-s})} K_{s,h}^*(X_{js} - X_{ks}; \mathbf{T}_j, X_{ks}, \mathbf{X}_{i,-s}) \\
&\quad \times L_{\mathbf{g}}(\mathbf{X}_{j,-s} - \mathbf{X}_{i,-s}) \sigma(\mathbf{X}_j, \mathbf{T}_j)\varepsilon_j \\
&\quad \times \frac{w_{-s}(\mathbf{X}_{l,-s})}{\varphi(X_{ks}, \mathbf{X}_{l,-s})} K_{s,h}^*(X_{ms} - X_{ks}; \mathbf{T}_m, X_{ks}, \mathbf{X}_{l,-s}) \\
&\quad \times L_{\mathbf{g}}(\mathbf{X}_{m,-s} - \mathbf{X}_{l,-s}) \sigma(\mathbf{X}_m, \mathbf{T}_m)\varepsilon_m.
\end{aligned}$$

Then, by the definition (A.11),

$$\sum_{k=1}^{n} \frac{w_s(X_{ks}) P_{1n}^2(X_{ks})}{n} = n^{-5} \sum_{i,j,k,l,m=1}^{n} \widetilde{g}_n(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j, \boldsymbol{\xi}_k, \boldsymbol{\xi}_l, \boldsymbol{\xi}_m).$$

Next, we define $g_n(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j, \boldsymbol{\xi}_k, \boldsymbol{\xi}_l, \boldsymbol{\xi}_m) = \sum \widetilde{g}_n(\boldsymbol{\xi}_{i'}, \boldsymbol{\xi}_{j'}, \boldsymbol{\xi}_{k'}, \boldsymbol{\xi}_{l'}, \boldsymbol{\xi}_{m'})/5!$, where the sum is over all possible permutations (i', j', k', l', m') of (i, j, k, l, m). Then $\sum_{k=1}^{n} w_s(X_{ks}) P_{1n}^2(X_{ks})/n$ is expressed as a V statistic $n^{-5} \sum_{i,j,k,l,m=1}^{n} g_n(\boldsymbol{\xi}_i, \boldsymbol{\xi}_j, \boldsymbol{\xi}_k, \boldsymbol{\xi}_l, \boldsymbol{\xi}_m)$. It is easy to see that $g_{n,0} = 0$, $g_{n,1} = 0$, and by changes of variables $g_{n,2}(\boldsymbol{\xi}_j, \boldsymbol{\xi}_m)$ equals

$$\begin{aligned}
&\sigma(\mathbf{X}_j, \mathbf{T}_j)\sigma(\mathbf{X}_m, \mathbf{T}_m)\varepsilon_j\varepsilon_m \\
&\times \int \frac{w_s(X_{js} - hu_{ks}) w_{-s}(\mathbf{X}_{j,-s} - \mathbf{g}\mathbf{u}_{i,-s}) w_{-s}(\mathbf{X}_{m,-s} - \mathbf{g}\mathbf{u}_{l,-s})}{\varphi(X_{js} - hu_{ks}, \mathbf{X}_{j,-s} - \mathbf{g}\mathbf{u}_{i,-s}) \varphi(X_{js} - hu_{ks}, \mathbf{X}_{m,-s} - \mathbf{g}\mathbf{u}_{l,-s})} \\
&\times K_s^*(u_{ks}; \mathbf{T}_j, X_{js} - hu_{ks}, \mathbf{X}_{j,-s} - \mathbf{g}\mathbf{u}_{i,-s}) L(\mathbf{u}_{i,-s}) L(\mathbf{u}_{l,-s}) \\
&\times K_{s,h}^*(X_{ms} - X_{js} + hu_{ks}; \mathbf{T}_m, X_{js} - hu_{ks}, \mathbf{X}_{m,-s} - \mathbf{g}\mathbf{u}_{l,-s}) \\
&\times \psi(x_{is}, \mathbf{X}_{j,-s} - \mathbf{g}\mathbf{u}_{i,-s}, \mathbf{t}_i) \psi(x_{ls}, \mathbf{X}_{m,-s} - \mathbf{g}\mathbf{u}_{l,-s}, \mathbf{t}_l) \\
&\times \psi(X_{js} - hu_{ks}, \mathbf{x}_{k,-s}, \mathbf{t}_k) \\
&\times dx_{is}\, d\mathbf{u}_{i,-s}\, dx_{ls}\, d\mathbf{u}_{l,-s}\, du_{ks}\, d\mathbf{x}_{k,-s}\, d\mathbf{t}_i\, d\mathbf{t}_l\, d\mathbf{t}_k.
\end{aligned}$$

To establish the asymptotic normality of the off-diagonal sum $2n^{-2} \sum_{1 \le j < m \le n} g_{n,2}(\boldsymbol{\xi}_j, \boldsymbol{\xi}_m)$, we use lemma 3.2 of Hjellvik, Yao, and Tjøstheim (1998). Let δ_n^2 be their σ_n^2, that is, $\delta_n^2 = \sum_{1 \le j < m \le n} \operatorname{var}\{2n^{-2} g_{n,2}(\boldsymbol{\xi}_j, \boldsymbol{\xi}_m)\}$. Define λ_{nk} in the same way as their M_{nk} for $k = 1, \ldots, 6$ with $2n^{-2} g_{n,2}(\boldsymbol{\xi}_j, \boldsymbol{\xi}_m)$ taking the role of their φ_{jm}. If we prove that, for some $\zeta > 0$,

$$n^2 \delta_n^{-2} \bigl(\lambda_{n1}^{1/(\zeta+1)} + \lambda_{n5}^{1/(\zeta+1)} + \lambda_{n6}^{1/2}\bigr) \longrightarrow 0, \quad \text{(A.14)}$$

$$n^{3/2} \delta_n^{-2} \bigl(\lambda_{n2}^{1/\{2(\zeta+1)\}} + \lambda_{n3}^{1/2} + \lambda_{n4}^{1/\{2(\zeta+1)\}}\bigr) \longrightarrow 0, \quad \text{(A.15)}$$

then we establish that $2n^{-2} \sum_{1 \le j < m \le n} g_{n,2}(\boldsymbol{\xi}_j, \boldsymbol{\xi}_m)$ is asymptotically normal with mean 0 and variance δ_n^2.

We compute δ_n^2 first. Note that

$$\begin{aligned}
\delta_n^2 = \frac{2}{n^2} \int \biggl\{ &\sigma(\mathbf{x}_j, \mathbf{t}_j)\sigma(\mathbf{x}_m, \mathbf{t}_m) \\
&\times \int \frac{w_s(x_{js}) w_{-s}(\mathbf{x}_{j,-s}) w_{-s}(\mathbf{x}_{m,-s})}{\varphi(x_{js}, \mathbf{x}_{j,-s}) \varphi(x_{js}, \mathbf{x}_{m,-s})} K_s^*(u_{ks}; \mathbf{t}_j, x_{js}, \mathbf{x}_{j,-s}) \\
&\times K_{s,h}^*(x_{ms} - x_{js} + hu_{ks}; \mathbf{t}_m, x_{js}, \mathbf{x}_{m,-s}) L(\mathbf{u}_{i,-s}) L(\mathbf{u}_{l,-s}) \\
&\times \psi(x_{is}, \mathbf{x}_{j,-s}, \mathbf{t}_i) \psi(x_{ls}, \mathbf{x}_{m,-s}, \mathbf{t}_l) \psi(x_{js}, \mathbf{x}_{k,-s}, \mathbf{t}_k) \\
&\times dx_{is}\, d\mathbf{u}_{i,-s}\, dx_{ls}\, d\mathbf{u}_{l,-s}\, du_{ks}\, d\mathbf{x}_{k,-s}\, d\mathbf{t}_i\, d\mathbf{t}_l\, d\mathbf{t}_k \biggr\}^2 \\
&\times \psi(\mathbf{x}_j, \mathbf{t}_j) \psi(\mathbf{x}_m, \mathbf{t}_m)\, d\mathbf{x}_j\, d\mathbf{x}_m\, d\mathbf{t}_j\, d\mathbf{t}_m \\
&\times \{1 + O(h^{p+1} + g^q)\}.
\end{aligned}$$

By a change of variable $x_{ms} = x_{js} + hv_s$ and further approximations of the functions, we obtain $\delta_n^2 = \{1 + O(h^{p+1} + g^q)\} n^{-2} h^{-1} \eta_s^4 \gamma_s^2$, where

$$\begin{aligned}
\gamma_s^2 = \frac{2}{\eta_s^4} \int &\frac{w_{-s}^2(\mathbf{x}_{-s}) w_{-s}^2(\mathbf{z}_{-s}) w_s^2(x_s)}{\varphi^2(x_s, \mathbf{x}_{-s}) \varphi^2(x_s, \mathbf{z}_{-s})} \\
&\times \bigl\{K_s^{*(c)}(u; \mathbf{t}_1, \mathbf{t}_2, x_s, \mathbf{x}_{-s}, \mathbf{z}_{-s})\bigr\}^2 \\
&\times \varphi_{-s}^2(\mathbf{x}_{-s}) \varphi_{-s}^2(\mathbf{z}_{-s}) \sigma^2(x_s, \mathbf{x}_{-s}, \mathbf{t}_1) \sigma^2(x_s, \mathbf{z}_{-s}, \mathbf{t}_2) \varphi_s^2(x_s) \\
&\times \psi(x_s, \mathbf{x}_{-s}, \mathbf{t}_1) \\
&\times \psi(x_s, \mathbf{z}_{-s}, \mathbf{t}_2)\, du\, dx_s\, d\mathbf{x}_{-s}\, d\mathbf{z}_{-s}\, d\mathbf{t}_1\, d\mathbf{t}_2 \qquad \text{(A.16)}
\end{aligned}$$

and

$$K_s^{*(c)}(w; \mathbf{t}_1, \mathbf{t}_2, x_s, \mathbf{x}_{-s}, \mathbf{z}_{-s}) = \int K_s^*(u; \mathbf{t}_1, x_s, \mathbf{x}_{-s}) K_s^*(w + u; \mathbf{t}_2, x_s, \mathbf{z}_{-s})\, du.$$

Next, we approximate λ_{nj}. We only illustrate the calculation of λ_{n4}. For $j < k$ and $l < m$ with all j, k, l, m different, we obtain

$$\begin{aligned}
&E|g_{n,2}(\xi_j, \xi_k) g_{n,2}(\xi_l, \xi_m)|^{2(1+\zeta)} \\
&\le \text{const}\, h^{-4(1+\zeta)+2} \bigl(E|\varepsilon_1|^{2(1+\zeta)}\bigr)^4 \\
&\quad \times \int \bigl|\sigma(\mathbf{x}_j, \mathbf{t}_j) \sigma(x_{js} + hv, \mathbf{x}_{k,-s}, \mathbf{t}_k) \\
&\quad \times \sigma(\mathbf{x}_l, \mathbf{t}_l) \sigma(x_{ls} + hv', \mathbf{x}_{m,-s}, \mathbf{t}_m)\bigr|^{2(1+\zeta)} \\
&\quad \times \bigl|K_s^{*(c)}(v; \mathbf{t}_j, \mathbf{t}_k, x_{js}, \mathbf{x}_{j,-s}, \mathbf{x}_{k,-s}) \\
&\quad \times K_s^{*(c)}(v'; \mathbf{t}_l, \mathbf{t}_m, x_{ls}, \mathbf{x}_{l,-s}, \mathbf{x}_{m,-s})\bigr|^{2(1+\zeta)} \\
&\quad \times \psi_{j,k,l,m}(\mathbf{x}_j, \mathbf{t}_j; x_{js} + hv, \mathbf{x}_{k,-s}, \mathbf{t}_k; \mathbf{x}_l, \mathbf{t}_l; x_{ls} + hv', \mathbf{x}_{m,-s}, \mathbf{t}_m) \\
&\quad \times d\mathbf{x}_j\, d\mathbf{t}_j\, dv\, d\mathbf{x}_{k,-s}\, d\mathbf{t}_k\, d\mathbf{x}_l\, d\mathbf{t}_l\, dv'\, d\mathbf{x}_{m,-s}\, d\mathbf{t}_m,
\end{aligned}$$

where the integrations with respect to $\mathbf{x}_j, dv, d\mathbf{x}_{k,-s}, \mathbf{x}_l, dv', d\mathbf{x}_{m,-s}$ are over compact sets. By assumption A6, the right side of the preceding inequality is bounded by

$$\begin{aligned}
&\text{const}\, h^{-4(1+\zeta)+2} \\
&\quad \times \int (\|\mathbf{t}_j\| \|\mathbf{t}_k\| \|\mathbf{t}_l\| \|\mathbf{t}_m\|)^{2(1+\zeta)} |\tilde{\sigma}(\mathbf{t}_j) \tilde{\sigma}(\mathbf{t}_k) \tilde{\sigma}(\mathbf{t}_l) \tilde{\sigma}(\mathbf{t}_m)|^{2(1+\zeta)} \\
&\quad \times \tilde{\varphi}_{j,k,l,m}(\mathbf{t}_j, \mathbf{t}_k, \mathbf{t}_l, \mathbf{t}_m)\, d\mathbf{t}_j\, d\mathbf{t}_k\, d\mathbf{t}_l\, d\mathbf{t}_m \\
&\le \text{const}\, h^{-4(1+\zeta)+2}.
\end{aligned}$$

This shows

$$\begin{aligned}
n^{3/2} \delta_n^{-2} \lambda_{n4}^{1/\{2(\zeta+1)\}} &\asymp n^{3/2} \times n^2 h \times n^{-4} h^{-(1+2\zeta)/(1+\zeta)} \\
&= n^{-(2p+2p\zeta+3+\zeta)/\{2(1+\zeta)(2p+3)\}}.
\end{aligned}$$

Similarly, we can establish

$$n^2 \delta_n^{-2} \lambda_{n1}^{1/(\zeta+1)} \asymp n^2 \times n^2 h \times n^{-4} h^{-2\zeta/(\zeta+1)} = h^{(1-\zeta)/(1+\zeta)},$$

$$\begin{aligned}
n^{3/2} \delta_n^{-2} \lambda_{n2}^{1/\{2(\zeta+1)\}} &\asymp n^{3/2} \times n^2 h \times n^{-4} h^{-(1+2\zeta)/(1+\zeta)} \\
&= n^{-(2p+2p\zeta+3+\zeta)/\{2(1+\zeta)(2p+3)\}},
\end{aligned}$$

$$n^{3/2}\delta_n^{-2}\lambda_{n3}^{1/2} \asymp n^{3/2} \times n^2 h \times n^{-4} h^{-3/2} = (nh)^{-1/2},$$

$$n^2\delta_n^{-2}\lambda_{n5}^{1/\{2(\zeta+1)\}} \asymp n^2 \times n^2 h \times n^{-4} h^{-(1+2\zeta)/\{2(\zeta+1)\}} = h^{1/\{2(1+\zeta)\}},$$

$$n^2\delta_n^{-2}\lambda_{n6}^{1/2} \asymp n^2 \times n^2 h \times n^{-4} h^{-1/2} = h^{1/2}.$$

Thus, if we take ζ such that $0 < \zeta < 1$, the convergences (A.14) and (A.15) hold.

By the martingale central limit theorem again, the diagonal sum $n^{-2}\sum_{j=1}^n g_{n,2}(\boldsymbol{\xi}_j, \boldsymbol{\xi}_j)$ is also asymptotically normal with mean $\eta_s^2 v_s n^{-1} h^{-1}\{1 + O(h^{p+1})\}$, where v_s is given by

$$v_s = \int \frac{w_{-s}^2(\mathbf{x}_{-s}) w_s(x_s)}{\eta_s^2 \varphi^2(\mathbf{x})} K_s^{*2}(u; \mathbf{t}, \mathbf{x}) \varphi_{-s}^2(\mathbf{x}_{-s}) \times \sigma^2(\mathbf{x}, \mathbf{t}) \psi(\mathbf{x}, \mathbf{t}) \varphi_s(x_s)\, du\, d\mathbf{x}\, d\mathbf{t}. \quad \text{(A.17)}$$

The asymptotic variance of $n^{-2}\sum_{j=1}^n g_{n,2}(\boldsymbol{\xi}_j, \boldsymbol{\xi}_j)$ is likewise calculated and may be shown to be of order $n^{-3}h^{-2}$.

Therefore, we establish

$$nh^{1/2}\left\{n^{-2}\sum_{j=1}^n \sum_{m=1}^n g_{n,2}(\boldsymbol{\xi}_j, \boldsymbol{\xi}_m) - \frac{\eta_s^2}{nh} v_s\right\} \xrightarrow{\mathcal{L}} \mathrm{N}(0, \eta_s^4 \gamma_s^2). \quad \text{(A.18)}$$

Application of Lemma A.1 reveals that $n^{-c}\sum_{j_1,\dots,j_c=1}^n g_{n,c}(\boldsymbol{\xi}_{j_1}, \dots, \boldsymbol{\xi}_{j_c}) = o(n^{-1}h^{-1/2})$ for $c = 3, 4, 5$. Using Lemma A.1 again, now applied to terms such as $\sum_{k=1}^n w_s(X_{ks}) P_{3n}^2(X_{ks})/n$, $\sum_{k=1}^n w_s(X_{ks}) \times R_{\ell 1}^2(X_{ks})/n$, and $\sum_{k=1}^n w_s(X_{ks}) R_{\ell 3}^2(X_{ks})/n$, we can show that they are all of order $o(n^{-1}h^{-1/2})$ as well. Using Lemma A.4, one may also prove $\sum_{k=1}^n w_s(X_{ks}) D_{s1}^2(X_{ks})/n$ and $\sum_{k=1}^n w_s(X_{ks}) D_{s3}^2(X_{ks})/n$ are both of order $o(h^{2p+4}) = o(n^{-1}h^{-1/2})$. Similar arguments establish that $\{\sum_{i=1}^n \hat{f}_s(X_{is}) w_s(X_{is})\}^2 = o(n^{-1}h^{-1/2})$. Hence,

$$V_{ns} = Q_{1n}^{-2}\sum_{k=1}^n \frac{P_{1n}^2(X_{ks}) w_s(X_{ks})}{n} + o(n^{-1}h^{-1/2}).$$

This completes the proof of Theorem 5.

A.3 Proofs of Theorems 3, 4, and 6

Define $\mathbf{q}_{su}(v; \mathbf{t})^T = (\mathbf{p}(v)^T t_{su}, \mathbf{t}_{-(su)}^T)$ and $S_{su}(\mathbf{x})$ in the same way as $S_s(\mathbf{x})$ with T_s and $\mathbf{T}_{-(s)}$ being replaced by T_{su} and $\mathbf{T}_{-(su)}$, respectively. Define an equivalent kernel $K_{(su)}^*(v; \mathbf{t}, \mathbf{x}) = e_0^T S_{su}^{-1}(\mathbf{x}) \mathbf{q}_{su}(v; \mathbf{t}) \times K(v)$. Let $K_{(su)}^{*(c)}$ denote the twofold convolution of $K_{(su)}^*$. Theorems 3 and 6 can be proved in the same way as the proofs of Theorems 1 and 5 with the following definitions of κ_{su}, τ_{su}^2, γ_{su}, and v_{su}:

$$\kappa_{su}(x_s) = \frac{f_{su}^{(p+1)}(x_s)}{(p+1)!} \times \int u^{p+1} E\{w_{-s}(\mathbf{X}_{-s}) T_{su} \times K_{(su)}^*(v; \mathbf{T}, x_s, \mathbf{X}_{-s})\}\, dv, \quad \text{(A.19)}$$

$$\tau_{su}^2(x_s) = \int \frac{w_{-s}^2(\mathbf{z}_{-s})}{\varphi^2(x_s, \mathbf{z}_{-s})} K_{(su)}^{*2}(v; \mathbf{t}, x_s, \mathbf{z}_{-s}) \varphi_{-s}^2(\mathbf{z}_{-s}) \times \sigma^2(x_s, \mathbf{z}_{-s}, \mathbf{t}) \psi(x_s, \mathbf{z}_{-s}, \mathbf{t})\, dv\, d\mathbf{z}_{-s}\, d\mathbf{t}, \quad \text{(A.20)}$$

$$\gamma_{su}^2 = \frac{2}{\eta_s^4} \int \frac{w_{-s}^2(\mathbf{x}_{-s}) w_{-s}^2(\mathbf{z}_{-s}) w_s^2(x_s)}{\varphi^2(x_s, \mathbf{x}_{-s}) \varphi^2(x_s, \mathbf{z}_{-s})} \times \{K_{(su)}^{*(c)}(v; \mathbf{t}_1, \mathbf{t}_2, x_s, \mathbf{x}_{-s}, \mathbf{z}_{-s})\}^2 \times \varphi_{-s}^2(\mathbf{x}_{-s}) \varphi_{-s}^2(\mathbf{z}_{-s}) \sigma^2(x_s, \mathbf{x}_{-s}, \mathbf{t}_1) \times \sigma^2(x_s, \mathbf{z}_{-s}, \mathbf{t}_2) \varphi_s^2(x_s) \psi(x_s, \mathbf{x}_{-s}, \mathbf{t}_1) \times \psi(x_s, \mathbf{z}_{-s}, \mathbf{t}_2)\, dv\, dx_s\, d\mathbf{x}_{-s}\, d\mathbf{z}_{-s}\, d\mathbf{t}_1\, d\mathbf{t}_2, \quad \text{(A.21)}$$

$$v_{su} = \int \frac{w_{-s}^2(\mathbf{x}_{-s}) w_s(x_s)}{\eta_s^2 \varphi^2(\mathbf{x})} K_{(su)}^{*2}(v; \mathbf{t}, \mathbf{x}) \varphi_{-s}^2(\mathbf{x}_{-s}) \times \sigma^2(\mathbf{x}, \mathbf{t}) \psi(\mathbf{x}, \mathbf{t}) \varphi_s(x_s)\, dv\, d\mathbf{x}\, d\mathbf{t}. \quad \text{(A.22)}$$

The proof of Theorem 4(a) is the same as that of the first part of Theorem 2. For the proof of Theorem 4(b), define

$$p_{jsu}(x_s) = \frac{w_{-s}(\mathbf{X}_{j,-s})}{\varphi(x_s, \mathbf{X}_{j,-s})} K_{(su),h}^*(X_{js} - x_s; \mathbf{T}_j, x_s, \mathbf{X}_{j,-s}) \times \varphi_{-s}(\mathbf{X}_{j,-s}) \sigma(\mathbf{X}_j, \mathbf{T}_j),$$

where $K_{(su),h}^*(v; \mathbf{t}, \mathbf{x}) = (1/h) K_{(su)}^*(v/h; \mathbf{t}, \mathbf{x})$. We observe

$$\hat{f}_{su}(x_s) - f_{su}(x_s) = b_{su}(x_s) h^{p+1} + n^{-1}\eta_s^{-1}\sum_{j=1}^n p_{jsu}(x_s)\varepsilon_j + o_p(h^{p+1}).$$

Thus, for the case $s = s'$, we obtain $\operatorname{cov}(\hat{f}_{su}(x_s), \hat{f}_{su'}(x_s)) = \eta_s^{-2} \times \tau_{suu'}(x_s) n^{-1} h^{-1}\{1 + o(1)\}$, where

$$\tau_{suu'}(x_s) = \int \frac{w_{-s}^2(\mathbf{z}_{-s})}{\varphi^2(x_s, \mathbf{z}_{-s})} K_{(su)}^*(v; \mathbf{t}, x_s, \mathbf{z}_{-s}) \times K_{(su')}^*(v; \mathbf{t}, x_s, \mathbf{z}_{-s}) \varphi_{-s}^2(\mathbf{z}_{-s}) \times \sigma^2(x_s, \mathbf{z}_{-s}, \mathbf{t}) \psi(x_s, \mathbf{z}_{-s}, \mathbf{t})\, dv\, d\mathbf{z}_{-s}\, d\mathbf{t}. \quad \text{(A.23)}$$

We note that $\tau_{su}^2(x_s) = \tau_{suu}(x_s)$. This completes the proof of Theorem 4.

[Received October 2002. Revised December 2005.]

REFERENCES

Bosq, D. (1998), *Nonparametric Statistics for Stochastic Processes*, New York: Springer-Verlag.

Cai, Z., Fan, J., and Li, R. Z. (2000a), "Efficient Estimation and Inferences for Varying-Coefficient Models," *Journal of the American Statistical Association*, 95, 888–902.

Cai, Z., Fan, J., and Yao, Q. (2000b), "Functional-Coefficient Regression Models for Nonlinear Times Series," *Journal of the American Statistical Association*, 95, 941–956.

Chen, R., and Liu, L. M. (2001), "Functional Coefficient Autoregressive Models: Estimation and Tests of Hypotheses," *Journal of Time Series Analysis*, 22, 151–173.

Chen, R., and Tsay, R. S. (1993a), "Nonlinear Additive ARX Models," *Journal of the American Statistical Association*, 88, 955–967.

——— (1993b), "Functional-Coefficient Autoregressive Models," *Journal of the American Statistical Association*, 88, 298–308.

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

Fan, J., Härdle, W., and Mammen, E. (1998), "Direct Estimation of Low-Dimensional Components in Additive Models," *The Annals of Statistics*, 26, 943–971.

Franses, H. F. (1996), *Periodicity and Stochastic Trends in Economic Time Series*, Oxford, U.K.: Oxford University Press.

Härdle, W., Hlavka, Z., and Klinke, S. (2000), *XploRe Application Guide*, Heidelberg: Springer-Verlag.

Härdle, W., Liang, H., and Gao, J. T. (2000), *Partially Linear Models*, Heidelberg: Springer-Verlag.

Härdle, W., and Mammen, E. (1993), "Comparing Nonparametric versus Parametric Regression Fits," *The Annals of Statistics*, 21, 1926–1947.

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

——— (1993), "Varying-Coefficient Models," *Journal of the Royal Statistical Society*, Ser. B, 55, 757–796.

Hjellvik, V., Yao, Q., and Tjøstheim, D. (1998), "Linearity Testing Using Local Polynomial Approximation," *Journal of Statistical Planning and Inference*, 68, 295–321.

Katkovnik, V. Ya. (1979), "Linear and Nonlinear Methods of Nonparametric Regression Analysis," *Soviet Automatic Control*, 93, 35–46 [Russian].

Linton, O. B. (1997), "Efficient Estimation of Additive Nonparametric Regression Models," *Biometrika*, 84, 469–473.

Linton, O. B., and Härdle, W. (1996), "Estimation of Additive Regression Models With Known Links," *Biometrika*, 83, 529–540.

Linton, O. B., and Nielsen, J. P. (1995), "A Kernel Method of Estimating Structured Nonparametric Regression Based on Marginal Integration," *Biometrika*, 82, 93–100.

Liptser, R. Sh., and Shirjaev, A. N. (1980), "A Functional Central Limit Theorem for Martingales," *Theory of Probability and Applications*, 25, 667–688.

Liu, R. (1988), "Bootstrap Procedures Under Some Non I.I.D. Models," *The Annals of Statistics*, 16, 1696–1708.

Mammen, E. (1992), *When Does Bootstrap Work: Asymptotic Results and Simulations*, Berlin: Springer-Verlag.

Mammen, E., Linton, O., and Nielsen, J. (1999), "The Existence and Asymptotic Properties of a Backfitting Projection Algorithm Under Weak Conditions," *The Annals of Statistics*, 5, 1443–1490.

Masry, E., and Tjøstheim, D. (1995), "Non-Parametric Estimation and Identification of ARCH Nonlinear Time Series: Strong Convergence and Asymptotic Normality," *Econometric Theory*, 11, 258–289.

——— (1997), "Additive Nonlinear ARX Time Series and Projection Estimates," *Econometric Theory*, 13, 214–252.

Ruppert, D., Sheather, S. J., and Wand, M. P. (1995), "An Effective Bandwidth Selection for Local Least Squares Regression," *Journal of the American Statistical Association*, 90, 1257–1270.

Ruppert, D., and Wand, M. P. (1994), "Multivariate Locally Weighted Least Squares Regression," *The Annals of Statistics*, 22, 1346–1370.

Sperlich, S., Tjøstheim, D., and Yang, L. (2002), "Nonparametric Estimation and Testing of Interaction in Additive Models," *Econometric Theory*, 18, 197–251.

Stone, C. J. (1977), "Consistent Nonparametric Regression," *The Annals of Statistics*, 5, 595–645.

Tjøstheim, D., and Auestad, B. (1994), "Nonparametric Identification of Nonlinear Time Series: Projections," *Journal of the American Statistical Association*, 89, 1398–1409.

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

Wolters, J. (1992), "Persistence and Seasonality in Output and Employment of the Federal Republic of Germany," *Recherches Economiques de Louvain*, 58, 421–439.

Wu, C. F. J. (1986), "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis" (with discussion), *The Annals of Statistics*, 14, 1261–1350.

Yang, L., and Tschernig, R. (2002), "Non- and Semiparametric Identification of Seasonal Nonlinear Autoregression Models," *Econometric Theory*, 18, 1408–1448.

Yoshihara, K. (1976), "Limiting Behavior of U-Statistics for Stationary, Absolutely Regular Processes," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 35, 237–252.