Yang, Lijian; Härdle, Wolfgang; Park, Byeong U.

Working Paper

Estimation and testing for varying coefficients in additive models with marginal integration

SFB 373 Discussion Paper, No. 2002,75

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Suggested Citation: Yang, Lijian; Härdle, Wolfgang; Park, Byeong U. (2002) : Estimation and testing for varying coefficients in additive models with marginal integration, SFB 373 Discussion Paper, No. 2002,75, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin, https://nbn-resolving.de/urn:nbn:de:kobv:11-10049411

This Version is available at:
https://hdl.handle.net/10419/65358

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

Lijian YANG* Wolfgang HÄRDLE Byeong U. PARK

October 2, 2002

Abstract

We propose marginal integration estimation and testing methods for the coefficients of varying coefficient multivariate regression model. 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 the West German real GNP data reveals that a partially linear varying coefficient model fits best 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 semi-parametric regression models allow the dependence to be of more general nonlinear forms. On the other hand, the appeal of simplicity and interpretation still motivates search for models that are nonparametric in nature but have special features that are appropriate for the data involved. Such are additive models (Chen and Tsay 1993a, Linton and Nielsen 1995, Masry and Tjøstheim

*Lijian Yang is associate professor, Department of Statistics and Probability, Michigan State University, East Lansing, Michigan 48824 (E-mail: *yang@stt.msu.edu*). Wolfgang Härdle is professor, Institut für Statistik und Ökonometrie, Humboldt-Universität zu Berlin, Spandauer Str.1, D-10178 Berlin, Germany (E-mail: *haerdle@wiwi.hu-berlin.de*). Byeong U. Park is professor, Department of Statistics, Seoul National University, Seoul 151-747, Korea (E-mail: *bupark@stats.snu.ac.kr*). This work was supported by Sonderforschungsbereich 373 "Quantifikation und Simulation Ökonomischer Prozesse" Deutsche Forschungsgemeinschaft at Humboldt-Universität zu Berlin. Yang's research was also partially supported by NSF grant DMS 9971186. Park's research was also supported by KOSEF through Statistical Research Center for Complex Systems at Seoul National University.

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), etc.

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

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

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

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

is called a varying-coefficient model if

$$m(\mathbf{X}_i, \mathbf{T}_i) = \sum_{s=1}^{d} f_s(X_{is})T_{is}. \tag{3}$$

One special case is when all the variables $\{X_s\}_{s=1}^d$ are the same X, which corresponds to the functional coefficient model of Chen and Tsay (1993b). 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, Fan and Li (2000), the model they studied was the same model proposed by Chen and Tsay (1993b), except with the additional feature of a possibly non-trivial link function. Cai, Fan and Li (2000) used local maximum likelihood estimation for all coefficient functions $\{f_s\}_{s=1}^d$, whose computing was no more than a univariate estimation, due to the fact that all these univariate functions depend on the same variable X. In the case of an identity link, the estimators are direct local polynomial estimators.

In practice, it is more realistic to allow some of the functions $\{f_s\}_{s=1}^d$ to depend on possibly different variables $\{X_s\}_{s=1}^d$. In such case, 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 paper, 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, consider a time series data $\{Y_t\}_{t=1}^n$ based on West German real GNP. After taking first difference and de-seasonalization, the data is considered strictly stationary, as shown by its plot, the dotted curve in Figure 4. A varying coefficient model $Y_t = f_1(Y_{t-2})Y_{t-1} + f_2(Y_{t-4})Y_{t-3} + \sigma_t\varepsilon_t$ is fitted and estimates of functions f_1, f_2 are plotted as solid curves in Figure 1, together with 95% point-wise confidence bands as dotted curves. More details about the data and its modelling are found in Section 4.

(Insert Figure 1 about here)

Although model (1) consists of additive bivariate functions, it is linear in the variables T_s. One interesting question one may ask is: are some of the coefficient functions $\{f_s\}_{s=1}^d$ constant? If the answer is yes for some but not all, then the model is partially linear in

some variables T_s; if the answer is yes to all, then the model is the classical linear regression model. Any constant f_s can then be estimated at $1/\sqrt{n}$-rate of convergence. A formal testing procedure is proposed in Section 3 for determining the constancy of coefficient functions f_s. For the German GNP data, it is found that f_2 can be set to a constant, while f_1 can not.

We organize the paper as follows. In Section 2, we describe a marginal estimation method for coefficient functions $\{f_s\}_{s=1}^d$ and derive asymptotic distribution theory of the estimator. In Section 3, a test procedure is proposed to test the hypothesis that f_s is a constant. In Section 4, we apply our estimation and testing methods to the West German real GNP data. All technical assumptions and proofs are in Appendix.

2 ESTIMATION OF VARYING COEFFICIENTS

In this section we formulate local polynomial integration estimators of the coefficient functions $\{f_s\}_{s=1}^d$. 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 $\{(\mathbf{X}_j, \mathbf{T}_j)\}_{j=1,...,i}$ for each $i = 1, ..., n$. This is sufficient for obtaining our main results on distribution theory as we assume $\{(\mathbf{X}_j, \mathbf{T}_j)\}_{j=1,...,n}$ to be strictly stationary and geometrically β-mixing in assumption A2 (see Appendix.), but weaker than the usual assumption that each ε_i is independent of the vectors $\{(\mathbf{X}_j, \mathbf{T}_j)\}_{j=1,...,n}$.

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

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

for some nonzero measurable functions r_s, then the regression function m in (3) equals

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

as well. Hence, for identifiability one assumes that

$$\sum_{s=1}^{d} r_s(X_{is})T_{is} = 0, a.s. \Rightarrow r_s(x) \equiv 0, s = 1, ..., d.$$

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

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

For a kernel function K, we write $K_h(u) = K(u/h)/h$. We fit p-th order local polynomials to estimate the varying coefficients. Write $\mathbf{Y} = (Y_i)_{1 \le i \le n}$, and let $\mathbf{p}(u) = (1, u, \ldots, u^p)^T$.

Define $\mathbf{Z}_s$ to be the $n \times (p+d)$ matrix which has $\left(\mathbf{p}\left\{(X_{is}-x_s)/h\right\}^T T_{is}, \mathbf{T}_{i,-s}^T\right)$ as its i-th 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

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

where $L_{\mathbf{g}}(\mathbf{u}) = (g_1\cdots g_{s-1}g_{s+1}\cdots g_d)^{-1}L(g_1^{-1}u_1,\ldots,g_{s-1}^{-1}u_{s-1},g_{s+1}^{-1}u_{s+1},\ldots,g_d^{-1}u_d)$, 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

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

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 zero except the $(l+1)$-th element which equals 1.

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

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

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 $\mathbf{X}_{i,-s}$'s has an unbounded support. Based on (4), one can predict Y given any realization $(\mathbf{x},\mathbf{t})$ of $(\mathbf{X},\mathbf{T})$ by the predictor

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

In the estimation procedure for f_s for a given s, we fit local constants for the other varying coefficients $f_{s'}$, $s' \neq s$. One 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.

Theorem 1 *Under the assumptions A1-A6 given in the appendix, we have, for any* $s = 1,...,d$,

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

as $n\to\infty$, *where* $b_s(x_s) = \kappa_s(x_s)/\eta_s$, $\sigma_s^2(x_s) = \tau_s^2(x_s)/\eta_s^2$, *and* κ_s, τ_s^2, η_s *are as defined at (A.14), (A.15) and (A.16), respectively.*

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 the assumptions A1-A6 given in the appendix, we have, for any* $s \neq s'$,

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

as $n \to \infty$, *and hence*

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

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 comment here that Theorems 1 and 2 hold only for local polynomial estimators of odd degree p, while similar results hold for p even as well. In particular, $p = 0$ corresponds to integrating the well-known Nadaraya-Watson type estimator. When an even p is used instead, the variance formula (A.15) remains the same while the bias formula (A.14) contains extra terms involving the derivatives of the design density.

As discussed in the introduction, our estimator (4) is designed for the model (1) when the regression function $m(\mathbf{x},\mathbf{t})$ is specified as in (3) with X_s's different to each other, while the estimators in Cai, Fan and Li (2000) can be applied only to the case where there is a common X for all $T_s, s = 1, ..., d$. Of special practical interests is the case where some but not all of the X_s's are the same. As an example, one may consider models such as

$$Y_t = c + a_1(r_t)M_t + a_2(r_t)M_t^2 + a_3(r)M_t^2 I_{\{M_t<0\}} + b_1(t)\tau_t + b_2(t)\tau_t^2 + \varepsilon_t,\ t = 1, ..., n$$

in which Y_t denotes the implied volatility, r_t the interest rate, M_t the moneyness, and τ_t the maturity at time t. Further research will be needed to obtain coefficient function estimators for such model.

3 TESTING FOR VARYING COEFFICIENTS

Suppose we are interested in testing the hypothesis

$$f_s(x_s) \equiv \text{constant} \tag{9}$$

for a specific s. If this hypothesis is true, one would get $\min_\alpha E\{f_s(X_s) - \alpha\}^2 w_s(X_s) = 0$ where w_s is an arbitrary positive weight function with a compact support. This leads us to propose the following test statistic:

$$\begin{aligned} V_{ns} &= n^{-1}\min_\alpha \sum_{i=1}^n \{\hat{f}_s(X_{is}) - \alpha\}^2 w_s(X_{is}) \\ &= n^{-1}\sum_{i=1}^n \hat{f}_s(X_{is})^2 w_s(X_{is}) - n^{-1}\{\sum_{i=1}^n w_s(X_{is})\}^{-1}\{\sum_{i=1}^n \hat{f}_s(X_{is})w_s(X_{is})\}^2, \end{aligned} \tag{10}$$

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

$$\hat{\alpha}_s = \{\sum_{i=1}^n w_s(X_{is})\}^{-1}\{\sum_{i=1}^n \hat{f}_s(X_{is})w_s(X_{is})\}. \tag{11}$$

The next theorem describes the asymptotic distribution of the test statistic (10) under the null hypothesis (9).

Theorem 3 *Under the null hypothesis (9) and the assumptions A1-A6 given in the appendix, we have, for any* $s = 1, ..., d$,

$$nh^{1/2}(V_{ns} - n^{-1}h^{-1}v_s) \xrightarrow{\mathcal{L}} N\left\{0, \gamma_s^2\right\} \tag{12}$$

as $n \to \infty$, *where* v_s *and* γ_s *are as given in (A.20) and (A.19).*

For the practical implementation of the test, we suggest to use a bootstrap procedure instead of the asymptotic normal distribution theory in Theorem 3. The reason is that for a test statistic based on kernel type of 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, Tjøstheim and Yang (2002). Another reason is that the normal approximation given in Theorem 3 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 to use the wild bootstrap procedure which was first introduced by Wu (1986) and implemented in various settings by Liu (1988), Härdle and Mammen (1993), and Sperlich, Tjøstheim and Yang (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 to those of the single residual.

To describe the procedure in our setting, let $\tilde{m}(\mathbf{x}, \mathbf{t}) = \hat{\alpha}_s t_s + \sum_{k \neq s}^{d} \hat{f}_k(x_k) t_k$ is the regression estimator under the hypothesis (9), where $\hat{\alpha}_s$ is an estimate of the constant $\hat{f}_s(x_s)$ given by (11) while $\hat{f}_k(x_k)$ $(k \neq s)$ is the marginally integrated estimate of $f_k(x_k)$ in (4). The wild bootstrap procedure to estimate the sampling distribution of V_{ns} under the null hypothesis then consists of the following steps:

(i) Find the residuals $\tilde{\epsilon}_i = Y_i - \tilde{m}(\mathbf{X}_i, \mathbf{T}_i)$ for $i = 1, \ldots, n$.

(ii) Generate i.i.d. 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 $Y_i^* = \tilde{m}(\mathbf{X_i}, \mathbf{T_i}) + \tilde{\epsilon}_i Z_i^W$.

(iii) Compute the bootstrap test statistic V_{ns}^* according to (10) using the wild bootstrap sample $\{(Y_i^*, \mathbf{X}_i, \mathbf{T}_i)\}_{i=1}^n$.

(iv) Repeat the steps (ii) and (iii) M times, obtaining $V_{ns,1}^*, \ldots, V_{ns,M}^*$. Estimate the null distribution of V_{ns} by the empirical distribution of $V_{ns,1}^*, \ldots, V_{ns,M}^*$.

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 $M = 200$.

4 AN EMPIRICAL EXAMPLE

We illustrate our estimation and testing methods through the analysis of the quarterly (seasonally non-adjusted) West German real GNP from 1960:1 to 1990:4. The data $G_t, 1 \leq t \leq n = 124$, which was compiled by Wolters (1992, p. 424, note 4), is plotted in Figure 2(a). One may see clearly a linear trend and pattern of seasonality. Based on seasonal unit root testing of Franses (1996), we take the first differences of the logs, and obtain a time series data, $D_t, 1 \leq t \leq n = 124$, which is plotted in Figure 2(b). This time series no longer has any trend but is obviously seasonal. Following the de-seasonalization procedure of Yang and Tschernig (2002), the sample means of the four seasons are calculated, which are $-0.065116, 0.038595, 0.051829, 0.0089443$ respectively. By subtracting these seasonal means, as was done in Yang and Tschernig (2002), the de-seasonalized $Y_t, 1 \leq t \leq n = 124$, are 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 $Y_t, 1 \leq t \leq n = 124$ is plotted as the dotted curve.

(Insert Figure 2 about here)

According to the semiparametric lag selection performed in Yang and Tschernig (2002), it was clear that the significant variables for prediction of Y_t are Y_{t-4} and Y_{t-2}. Calculation of autocorrelation functions indicates that Y_t is more correlated with Y_{t-1} and Y_{t-3} than with other lagged values. Hence, we propose a varying coefficient model

$$Y_t = f_1\left(Y_{t-2}\right) Y_{t-1} + f_2\left(Y_{t-4}\right) Y_{t-3} + \sigma\varepsilon_t \tag{13}$$

which includes the special case of possible linear AR(2) model $Y_t = f_1 Y_{t-1} + f_2 Y_{t-3} + \sigma\varepsilon_t$.

According to the definition of marginal integration estimator (4), we have estimated the functions f_1, f_2, and computed the predicted values

$$\hat{Y}_t = \hat{f}_1\left(Y_{t-2}\right) Y_{t-1} + \hat{f}_2\left(Y_{t-4}\right) Y_{t-3}. \tag{14}$$

We have carried out local linear smoothing ($p = 1$). The kernels K and L we have used for smoothing are both the quartic kernel $L(x) = K(x) = 0.9375\left(1 - x^2\right)^2 I_{(|x|\leq 1)}$, while the bandwidths are $h = 0.05, g = 0.05/1.1 = 0.0454$. Also computed are the standardized residuals $\hat{\varepsilon}_t$. The independence of the error terms would indicate the goodness-of-fit of the proposed model (13). At a practical level, such independence can only be examined via the autocorrelation functions (ACFs) of powers of the absolute values of the residuals. In Figure 3, we have plotted the ACFs of both $|\hat{\varepsilon}_t|$ and $\hat{\varepsilon}_t^2$. 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 one can conclude that both $|\varepsilon_t|$ and ε_t^2 have no autocorrelation. The ACF plots for $|\hat{\varepsilon}_t|^3$, $\hat{\varepsilon}_t^4$, etc., have led to the same conclusion. Thus, the model (13) fits well the structure of the data Y_t. As further evidence, we have plotted in Figure 4 the overlay of Y_t together with the predicted series $\hat{Y}_t$ given by (14). The predicted series follows the actual series rather closely.

(Insert Figures 3 and 4 about here)

We have plotted $\left\{Y_{t-2}, \hat{f}_1\left(Y_{t-2}\right)\right\}_{t=5}^{124}$, $\left\{Y_{t-4}, \hat{f}_2\left(Y_{t-4}\right)\right\}_{t=5}^{124}$ in Figure 1 as the solid curve. Also plotted as dotted curves are 95% point-wise confidence bands of f_1 and f_2, based on 200 wild bootstrap samples. The solid horizontal lines have the average values of $\hat{f}_1\left(Y_{t-2}\right)$

and $\hat{f}_2(Y_{t-4})$, respectively. The function $\hat{f}_2$ looks rather unlike a constant function, as the constant fit represented by the horizontal line does not lie within the confidence bands. The function $\hat{f}_1$, however, has a confidence band that covers the horizontal line, and hence could be a constant function.

To determine the significance of this visual impression, we generated 200 wild bootstrap samples from the data, and calculated V_{ns} according to (10) for the data itself and V_{ns}^* for every generated bootstrap sample. The p-values for V_{n1}, V_{n2} have been calculated relative to the bootstrap distribution, and they are 0.05 and 0.455 respectively. Hence we can conclude that an appropriate model for the data is the partially linear model:

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

APPENDIX: PROOFS

Preliminaries

We shall need the following technical assumptions on the kernels:

A1: *The kernels K and L are symmetric, compactly supported and Lipschitz continuous with $\int K(u)\,du = \int L(u)\,du = 1$. While 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 s-th 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*

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

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

The following assumptions are for the functions involved in the estimation and testing.

A3: *The functions f_s'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 zero and infinity. 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.*

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

A6: $(\ln n)(nhg_{\text{prod}})^{-1/2} = O(n^{-a})$ *for some* $a > 0$ *and* $(nh\ln n)^{1/2} g_{\max}^{q} \to 0$ *as* $n \to \infty$ *where* $g_{\text{prod}} = g_1 \cdots g_{s-1} g_{s+1} \cdots g_d$ *and* $g_{\max} = \max(g_1, \ldots, g_{s-1}, g_{s+1}, \ldots, g_d)$, *and* h *is asymptotic to* $n^{-1/(2p+3)}$.

One should note here that for existence of the bandwidth vector $\mathbf{g}$ satisfying the assumption A6 it is necessary that q, the order of the kernel L, should be larger than $(d-1)/2$.

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

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

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

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

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

for some $\delta > 0$. Here, $S_c = \{(i_1, ..., i_m) | \#_r(i_1, ..., i_m) = c\}, c = 0, ..., m-1$, and for every $(i_1, ..., i_m), 1 \le i_1 \le \cdots \le i_m \le n$, $\#_r(i_1, ..., i_m) =$ the number of $j = 1, ..., m-1$ 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

$$\begin{aligned} \theta_n(F_n) &= \int g_n(x_1, ..., x_m)\, dF_n(x_1) \cdots dF_n(x_m) \\ &= \frac{1}{n^m} \sum_{i_1=1}^{n} \cdots \sum_{i_m=1}^{n} g_n(\xi_{i_1}, ..., \xi_{i_m}) \\ U_n &= \frac{1}{\binom{n}{m}} \sum_{1 \le i_1 < \cdots < i_m \le n} g_n(\xi_{i_1}, ..., \xi_{i_m}), \end{aligned}$$

respectively, allow decompositions as

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

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

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

where $U_n^{(c)} = \frac{(n-c)!}{n!} \sum_{1 \leq i_1 < \cdots < i_c \leq n} \int g_{n,c}(x_{i_1}, ..., x_{i_c}) \prod_{j=1}^{c} \left[dI_{R_+^d}(x_j - \xi_{i_j}) - dF(x_j) \right]$. Here, $g_{n,c}$ are the projections of g_n given by

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

so that $g_{n,0} = \theta_n(F), g_n = g_{n,m}$, and $I_{R_+^d}$ is the indicator function of the nonnegative part of R^d, $R_+^d = \left\{ (y_1, ..., y_d) \in R^d | y_j \geq 0, j = 1, ..., d \right\}$.

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

$$\begin{aligned} & EV_n^{(c)2} + EU_n^{(c)2} \qquad (A.1) \\ \leq\ & C(m, \delta, r)\, n^{-c} \left\{ M_n^{2/(2+\delta)} \sum_{k=r+1}^{n} k \beta^{\delta/(2+\delta)}(k) + \sum_{c'=0}^{m-1} n^{-c'} M_{n,c'}^{2/(2+\delta)} \sum_{k=1}^{r} k \beta^{\delta/(2+\delta)}(k) \right\} \end{aligned}$$

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

$$EV_n^{(c)2} + EU_n^{(c)2} \leq C(m, \delta, r)\, C_2 C(\rho) n^{-c} \left\{ M_n^{2/(2+\delta)} + \sum_{c'=0}^{m-1} n^{-c'} M_{n,c'}^{2/(2+\delta)} \right\}. \qquad (A.2)$$

Proof. The proof essentially is the same as Lemma 2 in Yoshihara (1976) which dealt with the special case where $g_n \equiv g, r = 1$ and $M_n = M_{n,c'}$. The inequalities in the proof of this lemma do not require all g_n's to be the same for $n = 1, 2, ...$, and the 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. ■

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

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

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

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

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,*

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

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

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

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

Lemma A.4 *Define*

$$\begin{aligned}
D_{s1}\left(x_s\right) &= \frac{1}{n}\sum_{i=1}^{n} w_{-s}(\mathbf{X}_{i,-s})R_s\left(x_s,\mathbf{X}_{i,-s}\right)\mathbf{Z}_s^T\mathbf{W}_{is}\mathbf{E}, \\
D_{s2}\left(x_s\right) &= \frac{1}{n}\sum_{i=1}^{n} w_{-s}(\mathbf{X}_{i,-s})R_s\left(x_s,\mathbf{X}_{i,-s}\right)\mathbf{Z}_s^T\mathbf{W}_{is}\left[\{f_s(X_{js})\}_{j=1}^{n} - \sum_{\nu=0}^{p}\frac{f_s^{(\nu)}(x_s)h^{\nu}}{\nu!}\mathbf{Z}_se_{\nu}\right], \\
D_{s3}\left(x_s\right) &= \frac{1}{n}\sum_{i=1}^{n} w_{-s}(\mathbf{X}_{i,-s})R_s\left(x_s,\mathbf{X}_{i,-s}\right)\mathbf{Z}_s^T\mathbf{W}_{is} \\
&\quad\times\left[\left\{\sum_{s'\neq s} f_{s'}(X_{js'})\right\}_{j=1}^{n} - \sum_{s'\neq s} f_{s'}(X_{is'})\mathbf{Z}_se_{p+s'}\right].
\end{aligned}$$

Then, as $n\to\infty$

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

Proof. The lemma follows directly from Lemmas A.3. ■

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

$$\begin{aligned}
R_{\ell 1}(x_s) &= \frac{1}{n}\sum_{i=1}^{n}\frac{w_{-s}(\mathbf{X}_{i,-s})}{\varphi(x_s,\mathbf{X}_{i,-s})}S_s^{-1}(x_s,\mathbf{X}_{i,-s})\left\{I_{p+d} - \frac{\mathbf{Z}_s^T\mathbf{W}_{is}\mathbf{Z}_sS_s^{-1}(x_s,\mathbf{X}_{i,-s})}{\varphi(x_s,\mathbf{X}_{i,-s})}\right\}^{\ell} \\
&\quad\times\mathbf{Z}_s^T\mathbf{W}_{is}\mathbf{E} \qquad (A.3) \\
R_{\ell 2}(x_s) &= \frac{1}{n}\sum_{i=1}^{n}\frac{w_{-s}(\mathbf{X}_{i,-s})}{\varphi(x_s,\mathbf{X}_{i,-s})}S_s^{-1}(x_s,\mathbf{X}_{i,-s})\left\{I_{p+d} - \frac{\mathbf{Z}_s^T\mathbf{W}_{is}\mathbf{Z}_sS_s^{-1}(x_s,\mathbf{X}_{i,-s})}{\varphi(x_s,\mathbf{X}_{i,-s})}\right\}^{\ell} \\
&\quad\times\mathbf{Z}_s^T\mathbf{W}_{isis}\left[\{f_s(X_{js})\}_{j=1}^{n} - \sum_{\nu=0}^{p}\frac{f_s^{(\nu)}(x_s)h^{\nu}}{\nu!}\mathbf{Z}_se_{\nu}\right] \qquad (A.4) \\
R_{\ell 3}(x_s) &= \frac{1}{n}\sum_{i=1}^{n}\frac{w_{-s}(\mathbf{X}_{i,-s})}{\varphi(x_s,\mathbf{X}_{i,-s})}S_s^{-1}(x_s,\mathbf{X}_{i,-s})\left\{I_{p+d} - \frac{\mathbf{Z}_s^T\mathbf{W}_{is}\mathbf{Z}_sS_s^{-1}(x_s,\mathbf{X}_{i,-s})}{\varphi(x_s,\mathbf{X}_{i,-s})}\right\}^{\ell} \\
&\quad\times\mathbf{Z}_s^T\mathbf{W}_{is}\left[\left\{\sum_{s'\neq s} f_{s'}(X_{js'})\right\}_{j=1}^{n} - \sum_{s'\neq s} f_{s'}(X_{is'})\mathbf{Z}_se_{p+s'}\right]. \qquad (A.5)
\end{aligned}$$

Then, as $n \to \infty$,

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

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

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

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

Denote $\xi_i = (\mathbf{X}_i, \mathbf{T}_i, Y_i)$, The term P_1 can be written as the von Mises' differentiable statistic $V_n = 1/2n^2 \sum^n$

$$_{i,j=1}\, g_n(\xi_i, \xi_j) \text{ where } g_n(\xi_i, \xi_j) \text{ equals } \frac{S(x_s, \mathbf{X}_{i,-s})}{\varphi(x_s, \mathbf{X}_{i,-s})} - \frac{E\left(\mathbf{Z}_s^T \mathbf{W}_{is} \mathbf{Z}_s | x_s, \mathbf{X}_{i,-s}\right)}{\varphi^2(x_s, \mathbf{X}_{i,-s})} \Bigg\}$$

$$w_{-s}(\mathbf{X}_{i,-s}) S_s^{-1}(x_s, \mathbf{X}_{i,-s}) \left\{ \begin{array}{l} T_{js} K_h(X_{js} - x_s) L_{\mathbf{g}}(\mathbf{X}_{j,-s} - \mathbf{X}_{i,-s}) \sigma(\mathbf{X}_j, \mathbf{T}_j) \varepsilon_j \\ \mathbf{p}\left\{(X_{js} - x_s)/h\right\} T_{js} K_h(X_{js} - x_s) L_{\mathbf{g}}(\mathbf{X}_{j,-s} - \mathbf{X}_{i,-s}) \sigma(\mathbf{X}_j, \mathbf{T}_j) \varepsilon_j \\ \mathbf{T}_{j,-s} K_h(X_{js} - x_s) L_{\mathbf{g}}(\mathbf{X}_{j,-s} - \mathbf{X}_{i,-s}) \sigma(\mathbf{X}_j, \mathbf{T}_j) \varepsilon_j \end{array} \right)$$

$$\times S_s^{-1}(x_s, \mathbf{X}_{i,-s})$$

$$+ w_{-s}(\mathbf{X}_{j,-s}) S_s^{-1}(x_s, \mathbf{X}_{j,-s}) \left\{ \frac{S(x_s, \mathbf{X}_{j,-s})}{\varphi(x_s, \mathbf{X}_{j,-s})} - \frac{E\left(\mathbf{Z}_s^T \mathbf{W}_{js} \mathbf{Z}_s | x_s, \mathbf{X}_{j,-s}\right)}{\varphi^2(x_s, \mathbf{X}_{j,-s})} \right\}$$

$$\times S_s^{-1}(x_s, \mathbf{X}_{j,-s}) \left(\begin{array}{l} T_{is} K_h(X_{is} - x_s) L_{\mathbf{g}}(\mathbf{X}_{i,-s} - \mathbf{X}_{j,-s}) \sigma(\mathbf{X}_i, \mathbf{T}_i) \varepsilon_i \\ \mathbf{p}\left\{(X_{is} - x_s)/h\right\} T_{js} K_h(X_{is} - x_s) L_{\mathbf{g}}(\mathbf{X}_{i,-s} - \mathbf{X}_{j,-s}) \sigma(\mathbf{X}_i, \mathbf{T}_i) \varepsilon_i \\ \mathbf{T}_{i,-s} K_h(X_{is} - x_s) L_{\mathbf{g}}(\mathbf{X}_{i,-s} - \mathbf{X}_{j,-s}) \sigma(\mathbf{X}_i, \mathbf{T}_i) \varepsilon_i \end{array} \right).$$

First, one calculates that $g_{n,0} = 0$ and $g_{n,1}(\xi_j)$ equals

$$\int S_s^{-1}(x_s, \mathbf{z}_{-s}) w_{-s}(\mathbf{z}_{-s}) S_s^{-1}(x_s, \mathbf{z}_{-s}) \left\{ \frac{S(x_s, \mathbf{z}_{-s})}{\varphi(x_s, \mathbf{z}_{-s})} - \frac{E\left(\mathbf{Z}_s^T \mathbf{W}_{is} \mathbf{Z}_s | x_s, \mathbf{z}_{-s}\right)}{\varphi^2(x_s, \mathbf{z}_{-s})} \right\}$$

$$\times S_s^{-1}(x_s, \mathbf{z}_{-s}) \left(\begin{array}{l} T_{js} K_h(X_{js} - x_s) L_{\mathbf{g}}(\mathbf{X}_{j,-s} - \mathbf{z}_{-s}) \sigma(\mathbf{X}_j, \mathbf{T}_j) \varepsilon_j \\ \mathbf{p}\left\{(X_{js} - x_s)/h\right\} T_{js} K_h(X_{js} - x_s) L_{\mathbf{g}}(\mathbf{X}_{j,-s} - \mathbf{z}_{-s}) \sigma(\mathbf{X}_j, \mathbf{T}_j) \varepsilon_j \\ \mathbf{T}_{j,-s} K_h(X_{js} - x_s) L_{\mathbf{g}}(\mathbf{X}_{j,-s} - \mathbf{z}_{-s}) \sigma(\mathbf{X}_j, \mathbf{T}_j) \varepsilon_j \end{array} \right)$$

$$\times \varphi_{-s}(\mathbf{z}_{-s}) d\mathbf{z}_{-s}$$

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

greater than

$$Cb^{2+\delta}E\left|\begin{array}{l} w_{-s}(\mathbf{X}_{i,-s})T_{js}K_h\left(X_{js}-x_s\right)L_{\mathbf{g}}(\mathbf{X}_{j,-s}-\mathbf{X}_{i,-s})\sigma(\mathbf{X}_j,\mathbf{T}_j)\varepsilon_j \\ w_{-s}(\mathbf{X}_{i,-s})\mathbf{p}\left\{(X_{js}-x_s)/h\right\}T_{js}K_h\left(X_{js}-x_s\right)L_{\mathbf{g}}(\mathbf{X}_{j,-s}-\mathbf{X}_{i,-s})\sigma(\mathbf{X}_j,\mathbf{T}_j)\varepsilon_j \\ w_{-s}(\mathbf{X}_{i,-s})\mathbf{T}_{j,-s}K_h\left(X_{js}-x_s\right)L_{\mathbf{g}}(\mathbf{X}_{j,-s}-\mathbf{X}_{i,-s})\sigma(\mathbf{X}_j,\mathbf{T}_j)\varepsilon_j \end{array}\right|^{2+\delta},$$

which, by Lemma 1 of Yoshihara (1976), is less than or equal to

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

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

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

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. ■

Proof of main results

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

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

and define an equivalent kernel

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

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

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

This kernel satisfies the moment conditions as are 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*

$$\begin{array}{ll} E\left\{\int u^q T_s K_s^*(u;\mathbf{T},\mathbf{X})du|\mathbf{X}=\mathbf{x}\right\} = \delta_{0q}, & 0 \le q \le p\,; \\ E\left\{\int T_{s'} K_s^*(u;\mathbf{T},\mathbf{X})du|\mathbf{X}=\mathbf{x}\right\} = 0, & s' = 1,...,d, s' \ne s. \end{array} \tag{A.9}$$

In order to prove Theorem 1, we begin by observing the following simple equations:

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

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

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

Then, we obtain

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

which equals $Q_{2n}(x_s)$. 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

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

where

$$P_{1n}(x_s) = n^{-2}\sum_{i,j=1}^n \frac{w_{-s}(\mathbf{X}_{i,-s})}{\varphi(x_s, \mathbf{X}_{i,-s})} K_{s,h}^*\left(X_{js} - x_s; \mathbf{T}_j, x_s, \mathbf{X}_{i,-s}\right) L_{\mathbf{g}}(\mathbf{X}_{j,-s} - \mathbf{X}_{i,-s}) \sigma(\mathbf{X}_j, \mathbf{T}_j) \varepsilon_j \tag{A.11}$$

$$\begin{aligned} P_{2n}(x_s) &= n^{-2}\sum_{i,j=1}^n \frac{w_{-s}(\mathbf{X}_{i,-s})}{\varphi(x_s, \mathbf{X}_{i,-s})} K_{s,h}^*\left(X_{js} - x_s; \mathbf{T}_j, x_s, \mathbf{X}_{i,-s}\right) L_{\mathbf{g}}(\mathbf{X}_{j,-s} - \mathbf{X}_{i,-s}) \\ &\quad \times \left\{ f_s(X_{js}) - \sum_{\nu=0}^p \frac{1}{\nu!} f_s^{(\nu)}(x_s)(X_{js} - x_s)^\nu \right\} T_{js} \end{aligned} \tag{A.12}$$

$$\begin{aligned} P_{3n}(x_s) &= n^{-2}\sum_{i,j=1}^n \frac{w_{-s}(\mathbf{X}_{i,-s})}{\varphi(x_s, \mathbf{X}_{i,-s})} K_{s,h}^*\left(X_{js} - x_s; \mathbf{T}_j, x_s, \mathbf{X}_{i,-s}\right) L_{\mathbf{g}}(\mathbf{X}_{j,-s} - \mathbf{X}_{i,-s}) \\ &\quad \times \sum_{s' \ne s} \left\{ f_{s'}(X_{js'}) - f_{s'}(X_{is'}) \right\} T_{js'}. \end{aligned} \tag{A.13}$$

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

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

Proof. By the definition (A.11) and using Lemma A.1 for geometrically β-mixing processes, one has

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

which, after the change of variable $\mathbf{x}_{-s} = \mathbf{X}_{j,-s} - \mathbf{g}\mathbf{v}_{-s}$, equals

$$n^{-1} \sum_{j=1}^{n} \int \frac{w_{-s}(\mathbf{X}_{j,-s} - \mathbf{g}\mathbf{v}_{-s})}{\varphi(x_s, \mathbf{X}_{j,-s} - \mathbf{g}\mathbf{v}_{-s})} K^*_{s,h}\left(X_{js} - x_s; \mathbf{T}_j, x_s, \mathbf{X}_{j,-s} - \mathbf{g}\mathbf{v}_{-s}\right)$$

$$\times L(\mathbf{v}_{-s})\varphi_{-s}(\mathbf{X}_{j,-s} - \mathbf{g}\mathbf{v}_{-s})d\mathbf{v}_{-s}\sigma(\mathbf{X}_j, \mathbf{T}_j)\varepsilon_j + o_p\{(nh\log n)^{-1/2}\}.$$

Using again the fact that L is of order q, the above equals

$$n^{-1} \sum_{j=1}^{n} \frac{w_{-s}(\mathbf{X}_{j,-s})}{\varphi(x_s, \mathbf{X}_{j,-s})} K^*_{s,h}\left(X_{js} - x_s; \mathbf{T}_j, x_s, \mathbf{X}_{j,-s}\right) \varphi_{-s}(\mathbf{X}_{j,-s})\sigma(\mathbf{X}_j, \mathbf{T}_j)\varepsilon_j$$

$$+ o_p\{(nh\log n)^{-1/2}\}$$

which 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*

$$\kappa_s(x_s) = \frac{f_s^{(p+1)}(x_s)}{(p+1)!} \int u^{p+1} E\left\{w_{-s}(\mathbf{X}_{-s})T_s K^*_s(u; \mathbf{T}, x_s, \mathbf{X}_{-s})\right\} du. \tag{A.14}$$

Proof. By definition (A.12) and again using Lemma A.1, one derives

$$P_{2n}(x_s) = \int \frac{w_{-s}(\mathbf{x}_{-s})}{\varphi(x_s, \mathbf{x}_{-s})} K^*_{s,h}\left(z_s - x_s; \mathbf{t}, x_s, \mathbf{x}_{-s}\right) L_{\mathbf{g}}(\mathbf{z}_{-s} - \mathbf{x}_{-s})$$

$$\times \left\{ f_s(z_s) - \sum_{\nu=0}^{p} \frac{f_s^{(\nu)}(x_s)}{\nu!}(z_s - x_s)^\nu \right\} t_s\psi(\mathbf{z}, \mathbf{t})\varphi_{-s}(\mathbf{x}_{-s})d\mathbf{z}d\mathbf{t}d\mathbf{x}_{-s}\left\{1 + o_p(1)\right\}$$

which, after the changes of variables $z_s = x_s + hu$ and $\mathbf{z}_{-s} = \mathbf{x}_{-s} + \mathbf{g}\mathbf{v}_{-s}$, equals

$$\int \frac{w_{-s}(\mathbf{x}_{-s})}{\varphi(x_s, \mathbf{x}_{-s})} K^*_s\left(u; \mathbf{t}, x_s, \mathbf{x}_{-s}\right) L(\mathbf{v}_{-s}) \left\{ f_s(x_s + hu) - \sum_{\nu=0}^{p} \frac{f_s^{(\nu)}(x_s)h^\nu}{\nu!} u^\nu \right\}$$

$$\times t_s \psi(x_s + hu, \mathbf{x}_{-s} + \mathbf{g}\mathbf{v}_{-s}, \mathbf{t}) \varphi_{-s}(\mathbf{x}_{-s}) du d\mathbf{v}_{-s} d\mathbf{t} d\mathbf{x}_{-s} \{1 + o_p(1)\}.$$

Here, we write $\mathbf{g}\mathbf{v}_{-s} = (g_1 v_1, \ldots, g_{s-1} v_{s-1}, g_{s+1} v_{s+1}, \ldots, g_d v_d)$. Thus,

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

This completes the proof of the lemma. ■

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

Proof. By definition (A.13) and applying Lemma A.1, one has

$$P_{3n}(x_s) = \int \frac{w_{-s}(\mathbf{x}_{-s})}{\varphi(x_s, \mathbf{x}_{-s})} K_{s,h}^* (z_s - x_s; \mathbf{t}, x_s, \mathbf{x}_{-s}) L_{\mathbf{g}}(\mathbf{z}_{-s} - \mathbf{x}_{-s})$$

$$\times \left[\sum_{s' \neq s} \{ f_{s'}(z_{s'}) - f_{s'}(x_{s'}) \} t_{s'} \right] \psi(\mathbf{z}, \mathbf{t}) \varphi_{-s}(\mathbf{x}_{-s}) d\mathbf{z} d\mathbf{t} d\mathbf{x}_{-s} \{1 + o_p(1)\}.$$

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

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

since L is of order q by the 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 $x_s \in supp(w_s)$ is asymptotically normal with mean 0 and variance

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

After the change of variable $z_s = x_s + hu$, this equals

$$\int \frac{w_{-s}^2(\mathbf{z}_{-s})}{\varphi^2(x_s, \mathbf{z}_{-s})} K_s^{*2} (u; \mathbf{t}, x_s, \mathbf{z}_{-s}) \varphi_{-s}^2(\mathbf{z}_{-s}) \sigma^2(x_s + hu, \mathbf{z}_{-s}, \mathbf{t})$$
$$\times \psi(x_s + hu, \mathbf{z}_{-s}, \mathbf{t}) du d\mathbf{z}_{-s} d\mathbf{t} \{1 + o(1)\},$$

the leading term of which equals

$$\tau_s^2(x_s) = \int \frac{w_{-s}^2(\mathbf{z}_{-s})}{\varphi^2(x_s, \mathbf{z}_{-s})} K_s^{*2}(u; \mathbf{t}, x_s, \mathbf{z}_{-s})\, \varphi_{-s}^2(\mathbf{z}_{-s}) \sigma^2(x_s, \mathbf{z}_{-s}, \mathbf{t}) \psi(x_s, \mathbf{z}_{-s}, \mathbf{t}) du d\mathbf{z}_{-s} d\mathbf{t}. \tag{A.15}$$

The theorem now follows immediately from Lemmas A.7, A.8, the conditions on the bandwidths as given in A6, and the fact that $Q_{1n} = \eta_s + O_p(n^{-1/2})$ where

$$\eta_s = \int w_{-s}(\mathbf{z}_{-s}) \varphi_{-s}(\mathbf{z}_{-s}) d\mathbf{z}_{-s}. \tag{A.16}$$

■

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

$$\hat{f}_s(x_s) - f_s(x_s) = b_s(x_s) h^{p+1} + n^{-1} \eta_s^{-1} \sum_{j=1}^{n} p_{js}(x_s) \varepsilon_j + o_p(h^{p+1}). \tag{A.17}$$

Applying (A.17), one only needs 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 i.i.d. white noise and each ε_i is independent of the vectors $(\mathbf{X}_j, \mathbf{T}_j), j = 1, ..., i$ for each $i = 1, ..., n$, we need only to show that

$$E\left\{ p_{js}(x_s) p_{js'}(x_{s'}) \right\} = o(h^{-1}). \tag{A.18}$$

By change of variables technique for X_s and $X_{s'}$ which are contained in $p_{js}(x_s)$ and $p_{js'}(x_{s'})$ respectively, one may show that the left hand side of (A.18) is actually $O(1)$, which proves the theorem. ■

Proof of Theorem 3. For this proof, we use (A.10) again. Under the hypothesis (9), $P_{2n}(x_s) = R_{l2}(x_s) = D_{s2}(x_s) = 0$ and so

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

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 $\sum_{k=1}^{n} w_s(X_{ks})$ $P_{1n}^2(X_{ks})/n$. By the definition (A.11), the latter equals

$$\begin{aligned} & n^{-5} \sum_{k=1}^{n} w_s(X_{ks}) \left\{ \sum_{i,j=1}^{n} \frac{w_{-s}(\mathbf{X}_{i,-s})}{\varphi(X_{ks}, \mathbf{X}_{i,-s})} K_{s,h}^{*}(X_{js} - X_{ks}; \mathbf{T}_j, X_{ks}, \mathbf{X}_{i,-s}) \right. \\ & \qquad \left. \times L_{\mathbf{g}}(\mathbf{X}_{j,-s} - \mathbf{X}_{i,-s}) \sigma(\mathbf{X}_j, \mathbf{T}_j) \varepsilon_j \right\}^2 \\ = \; & n^{-5} \sum_{i,j,k,l,m=1}^{n} \widetilde{g}_n(\xi_i, \xi_j, \xi_k, \xi_l, \xi_m), \end{aligned}$$

where $\xi_i = (\mathbf{X}_i, \mathbf{T}_i, Y_i)$ and

$$\begin{aligned}\widetilde{g}_n(\xi_i,\xi_j,\xi_k,\xi_l,\xi_m) &= w_s(X_{ks})\frac{w_{-s}(\mathbf{X}_{i,-s})}{\varphi(X_{ks},\mathbf{X}_{i,-s})}K^*_{s,h}(X_{js}-X_{ks};\mathbf{T}_j,X_{ks},\mathbf{X}_{i,-s})\\ &\quad\times L_{\mathbf{g}}(\mathbf{X}_{j,-s}-\mathbf{X}_{i,-s})\sigma(\mathbf{X}_j,\mathbf{T}_j)\varepsilon_j\\ &\quad\times\frac{w_{-s}(\mathbf{X}_{l,-s})}{\varphi(X_{ks},\mathbf{X}_{l,-s})}K^*_{s,h}(X_{ms}-X_{ks};\mathbf{T}_m,X_{ks},\mathbf{X}_{l,-s})\\ &\quad\times L_{\mathbf{g}}(\mathbf{X}_{m,-s}-\mathbf{X}_{l,-s})\sigma(\mathbf{X}_m,\mathbf{T}_m)\varepsilon_m.\end{aligned}$$

Next, we define

$$g_n(\xi_i,\xi_j,\xi_k,\xi_l,\xi_m) = \frac{1}{5!}\sum_{(i',j',k',l',m')}\widetilde{g}_n(\xi_{i'},\xi_{j'},\xi_{k'},\xi_{l'},\xi_{m'})$$

where the sum is over all possible permutations of i,j,k,l,m. Then $\sum_{k=1}^n w_s(X_{ks})P^2_{1n}(X_{ks})/n$ is expressed as a V statistic

$$n^{-5}\sum_{i,j,k,l,m=1}^n g_n(\xi_i,\xi_j,\xi_k,\xi_l,\xi_m).$$

It is easy to see that $g_{n,0}=0, g_{n,1}=0$, and

$$\begin{aligned}g_{n,2}(\xi_j,\xi_m) &= \sigma(\mathbf{X}_j,\mathbf{T}_j)\sigma(\mathbf{X}_m,\mathbf{T}_m)\varepsilon_j\varepsilon_m\int w_s(x_{ks})\frac{w_{-s}(\mathbf{x}_{i,-s})}{\varphi(x_{ks},\mathbf{x}_{i,-s})}\frac{w_{-s}(\mathbf{x}_{l,-s})}{\varphi(x_{ks},\mathbf{x}_{l,-s})}\\ &\quad\times K^*_{s,h}(X_{js}-x_{ks};\mathbf{t}_j,x_{ks},\mathbf{x}_{i,-s})\,L_{\mathbf{g}}(\mathbf{X}_{j,-s}-\mathbf{x}_{i,-s})\\ &\quad\times K^*_{s,h}(X_{ms}-x_{ks};\mathbf{T}_m,x_{ks},\mathbf{x}_{l,-s})\,L_{\mathbf{g}}(\mathbf{X}_{m,-s}-\mathbf{x}_{l,-s})\\ &\quad\times\psi(\mathbf{x}_i,\mathbf{t}_i)\psi(\mathbf{x}_l,\mathbf{t}_l)\psi(\mathbf{x}_k,\mathbf{t}_k)d\mathbf{x}_i d\mathbf{x}_l d\mathbf{x}_k d\mathbf{t}_i d\mathbf{t}_l d\mathbf{t}_k.\end{aligned}$$

By changes of variables

$$\mathbf{X}_{j,-s}-\mathbf{x}_{i,-s}=\mathbf{g}\mathbf{u}_{i,-s},\mathbf{X}_{m,-s}-\mathbf{x}_{l,-s}=\mathbf{g}\mathbf{u}_{l,-s},X_{js}-x_{ks}=hu_{ks},$$

$g_{n,2}(\xi_j,\xi_m)$ becomes

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

By applying the martingale central limit theorem, one may show that the off-diagonal sum $2\,n^{-2}\sum_{1\le j<m\le n}^n g_{n,2}(\xi_j,\xi_m)$ is asymptotically normal with mean 0 and asymptotic variance given by

$$\frac{2}{n^2}\int\left\{\sigma(\mathbf{x}_j,\mathbf{t}_j)\sigma(\mathbf{x}_m,\mathbf{t}_m)\int\frac{w_s(x_{js})w_{-s}(\mathbf{x}_{j,-s})w_{-s}(\mathbf{x}_{m,-s})}{\varphi(x_{js},\mathbf{x}_{j,-s})\varphi(x_{js},\mathbf{x}_{m,-s})}\right.$$

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

After making another change of variable $x_{ms}=x_{js}+hv_s$, the above variance becomes

$$
\begin{aligned}
&\frac{2\left\{1+O\left(h^{p+1}+g^q\right)\right\}}{n^2h}\\
&\times\int\left\{\sigma(\mathbf{x}_j,\mathbf{t}_j)\sigma(x_{js}+hv_s,\mathbf{x}_{m,-s},\mathbf{t}_m)\frac{w_s(x_{js})w_{-s}(\mathbf{x}_{j,-s})w_{-s}(\mathbf{x}_{m,-s})}{\varphi(\mathbf{x}_j)\varphi(x_{js},\mathbf{x}_{m,-s})}\right\}^2\\
&\times\left\{\int K_s^*\left(u_{ks};\mathbf{t}_j,\mathbf{x}_j\right)K_s^*\left(v_s+u_{ks};\mathbf{t}_m,x_{js},\mathbf{x}_{m,-s}\right)du_{ks}\right.\\
&\qquad\left.\times\varphi_{-s}(\mathbf{x}_{j,-s})\varphi_{-s}(\mathbf{x}_{m,-s})\varphi(x_{js})\right\}^2\\
&\times\psi(\mathbf{x}_j,\mathbf{t}_j)\psi(x_{js}+hv_s,\mathbf{x}_{m,-s},\mathbf{t}_m)dv_sd\mathbf{x}_{m,-s}d\mathbf{x}_jd\mathbf{t}_jd\mathbf{t}_m,
\end{aligned}
$$

or $\left\{1+O\left(h^{p+1}+g^q\right)\right\}n^{-2}h^{-1}\eta_s^4\gamma_s^2$ where

$$
\begin{aligned}
\gamma_s^2\;=\;&\frac{2}{\eta_s^4}\int\frac{w^2(\mathbf{x}_{-s})w^2(\mathbf{z}_{-s})w_s^2(x_s)}{\varphi^2(x_s,\mathbf{x}_{-s})\varphi^2(x_s,\mathbf{z}_{-s})}\left\{K_s^{*(c)}\left(u;\mathbf{t}_1,\mathbf{t}_2,x_s,\mathbf{x}_{-s},\mathbf{z}_{-s}\right)\right\}^2\\
&\times\varphi_{-s}^2(\mathbf{x}_{-s})\varphi_{-s}^2(\mathbf{z}_{-s})\sigma^2(x_s,\mathbf{x}_{-s},\mathbf{t}_1)\sigma^2(x_s,\mathbf{z}_{-s},\mathbf{t}_2)\varphi_s^2(x_s)\\
&\times\psi(x_s,\mathbf{x}_{-s},\mathbf{t}_1)\psi(x_s,\mathbf{z}_{-s},\mathbf{t}_2)dudx_sd\mathbf{x}_{-s}d\mathbf{z}_{-s}d\mathbf{t}_1d\mathbf{t}_2
\end{aligned}
\tag{A.19}
$$

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

Meanwhile, by the martingale central limit theorem again, the diagonal sum $n^{-2}\sum_{1\le j\le n}^n g_{n,2}\left(\xi_j,\xi_j\right)$ is asymptotically normal with mean

$$
\begin{aligned}
&\frac{\left\{1+O\left(h^{p+1}+g^q\right)\right\}}{n}\int\sigma^2(\mathbf{x}_j,\mathbf{t}_j)\int\frac{w_s(x_{js})w_{-s}^2(\mathbf{x}_{j,-s})}{\varphi^2(x_{js},\mathbf{x}_{j,-s})}K_s^*\left(u_{ks};\mathbf{t}_j,x_{js},\mathbf{x}_{j,-s}\right)\\
&\quad\times L(\mathbf{u}_{i,-s})L(\mathbf{u}_{l,-s})K_{s,h}^*\left(hu_{ks};\mathbf{t}_j,x_{js},\mathbf{x}_{j,-s}\right)\psi(x_{is},\mathbf{x}_{j,-s},\mathbf{t}_i)\psi(x_{ls},\mathbf{x}_{j,-s},\mathbf{t}_l)\\
&\quad\times\psi(x_{js},\mathbf{x}_{k,-s},\mathbf{t}_k)dx_{is}d\mathbf{u}_{i,-s}dx_{ls}d\mathbf{u}_{l,-s}du_{ks}d\mathbf{x}_{k,-s}d\mathbf{t}_id\mathbf{t}_ld\mathbf{t}_k\psi(\mathbf{x}_j,\mathbf{t}_j)d\mathbf{x}_jd\mathbf{t}_j\\
=\;&\frac{\left\{1+O\left(h^{p+1}\right)\right\}}{nh}\int\sigma^2(\mathbf{x}_j,\mathbf{t}_j)\frac{w_s(x_{js})w_{-s}^2(\mathbf{x}_{j,-s})}{\varphi^2(\mathbf{x}_j)}\varphi_{-s}^2(\mathbf{x}_{j,-s})\varphi(x_{js})\\
&\qquad\times K_s^{*2}\left(u_{ks};\mathbf{t}_j,\mathbf{x}_j\right)du_{ks}\psi(\mathbf{x}_j,\mathbf{t}_j)d\mathbf{x}_jd\mathbf{t}_j\\
=\;&\frac{\eta_s^2}{nh}v_s\left\{1+O\left(h^{p+1}\right)\right\},
\end{aligned}
$$

where v_s is given by

$$v_s = \int \frac{w_{-s}^2(\mathbf{x}_{-s})w_s(x_s)}{\eta_s^2\varphi^2(\mathbf{x})} K_s^{*2}(u;\mathbf{t},\mathbf{x})\,\varphi_{-s}^2(\mathbf{x}_{-s})\sigma^2(\mathbf{x},\mathbf{t})\psi(\mathbf{x},\mathbf{t})\varphi_s(x_s)du d\mathbf{x} d\mathbf{t}. \tag{A.20}$$

The asymptotic variance of $n^{-2}\sum_{1\le j\le n}^{n} g_{n,2}(\xi_j,\xi_j)$ is likewise calculated, and may be shown to be of order $n^{-3}h^{-2}$. Therefore, we establish

$$n\sqrt{h}\left\{n^{-2}\sum_{j,m=1}^{n} g_{n,2}(\xi_j,\xi_m) - \frac{\eta_s^2}{nh}v_s\right\} \xrightarrow{\mathcal{L}} N\left(0,\eta_s^4\gamma_s^2\right). \tag{A.21}$$

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

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

This completes the proof of Theorem 3. ■

REFERENCES

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

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

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

Chen, R. and Tsay, R. S. (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 and Hall.

Franses, H. F. (1996), *Periodicity and Stochastic Trends in Economic Time Series*, Oxford: 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", *Annals of Statistics*, 21, 1926-1947.

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

Hastie, T. J. and Tibshirani, R. J. (1993), "Varying-Coefficient Models," *Journal of the Royal Statistical Society Series B*, 55, 757-796.

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", *Annals of Statistics*, 16, 1696-1708.

Mammen, E. (1992), *When Does Bootstrap Work: Asymptotic Results and Simulations*, Lecture Notes in Statistics 77, Berlin: Springer-Verlag.

Mammen, E. Linton, O. and Nielsen, J. (1999), "The Existence and Asymptotic Properties of a Backfitting Projection Algorithm under Weak Conditions," *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.

Masry, E. and Tjøstheim, D. (1997), "Additive Nonlinear ARX Time Series and Projection Estimates," *Econometric Theory*, 13, 214-252.

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," *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 and Hall.

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

Wu, C. F. J. (1986), "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis (with discussion)", *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 (1976), "Limiting Behavior of U-statistics for Stationary, Absolutely Regular Processes," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 35, 237-252.

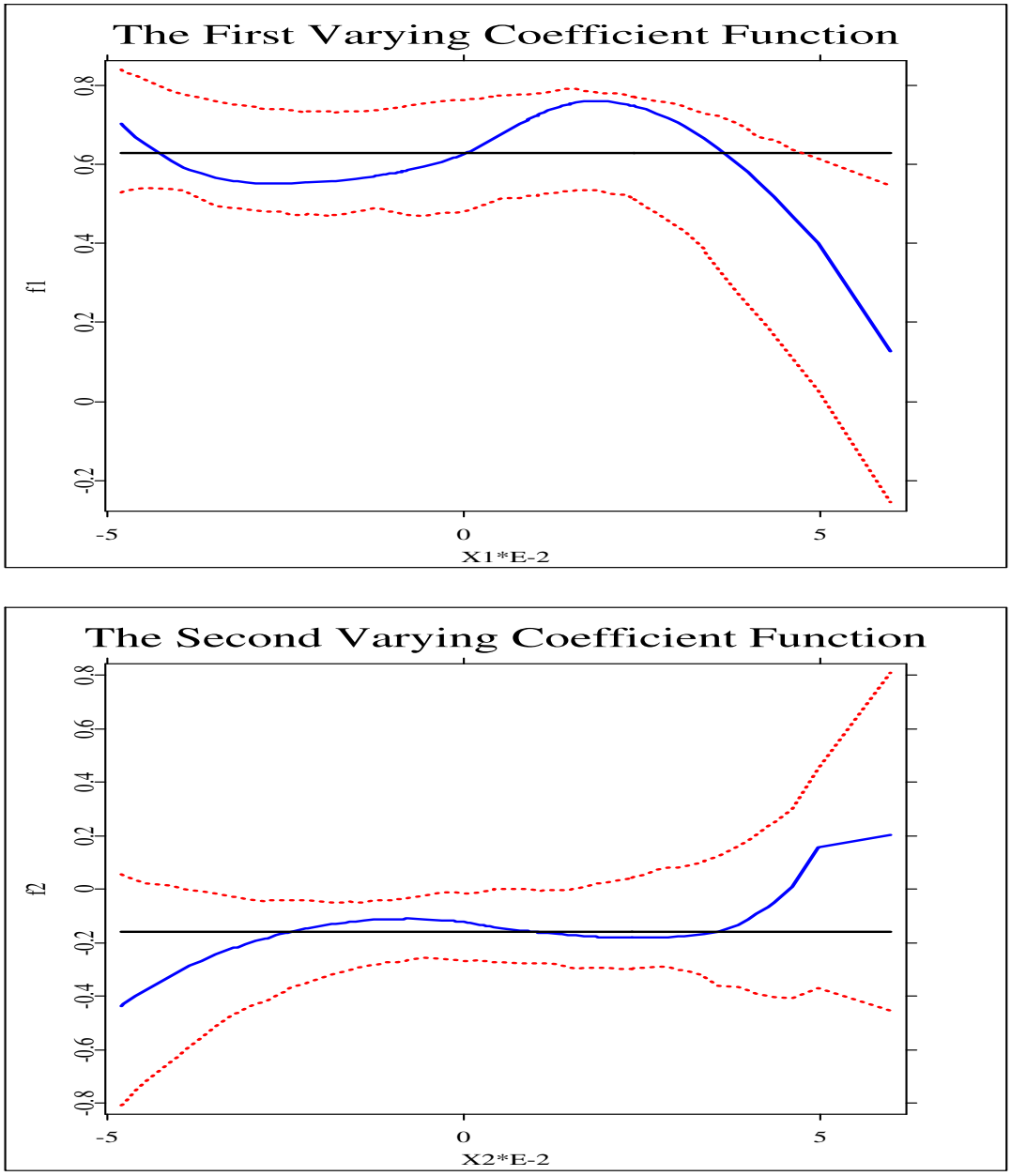
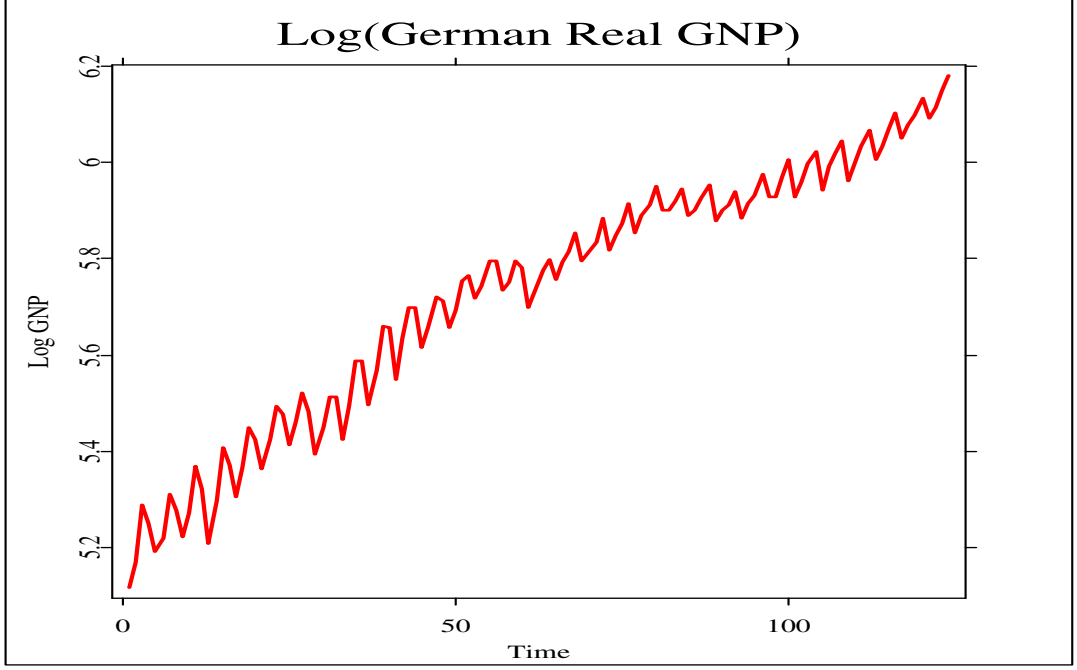
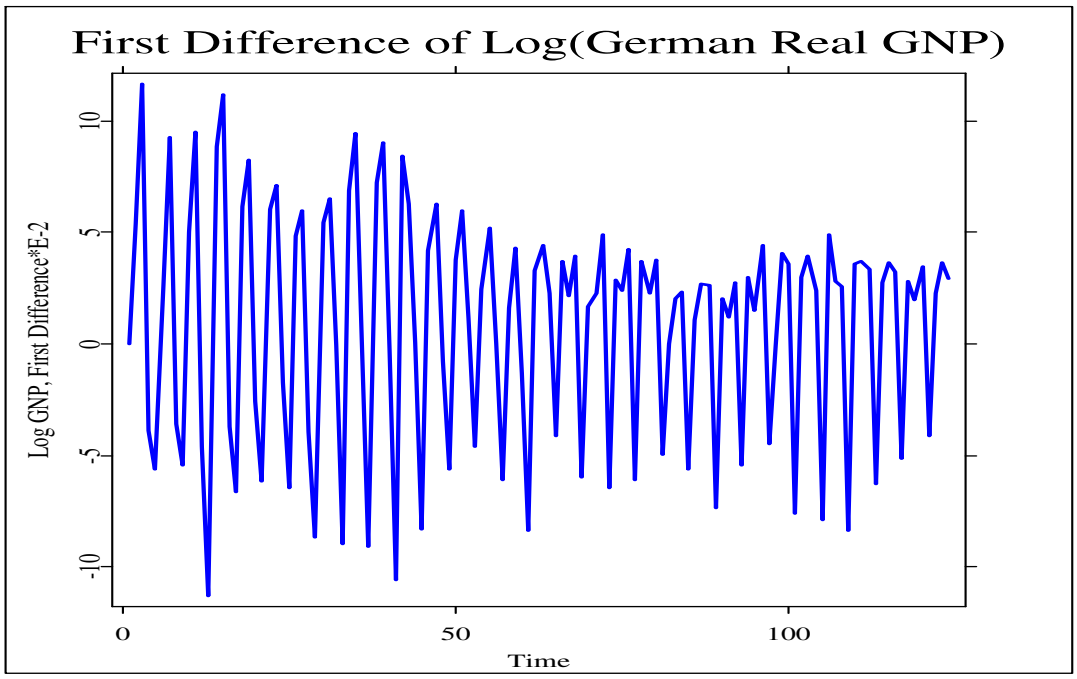


Figure 1: Marginal integration estimates of varying coefficient functions for the West German real GNP quarterly data from 1960:1 to 1990:4. The model is $Y_t = f_1(Y_{t-2})\,Y_{t-1} + f_2(Y_{t-4})\,Y_{t-3} + \sigma\varepsilon_t$, in which the time series Y_t consists of de-seasonalized first differences of the logarithm GNP. The response variable is $Y = Y_t$, while the predictors are $X_1 = Y_{t-2}$, $X_2 = Y_{t-4}$, $T_1 = Y_{t-1}$ and $T_2 = Y_{t-3}$. Solid curves are function estimates while the dotted curves are point-wise 95% confidence bands. The horizontal lines represent the means of the coefficient functions over the compact ranges.

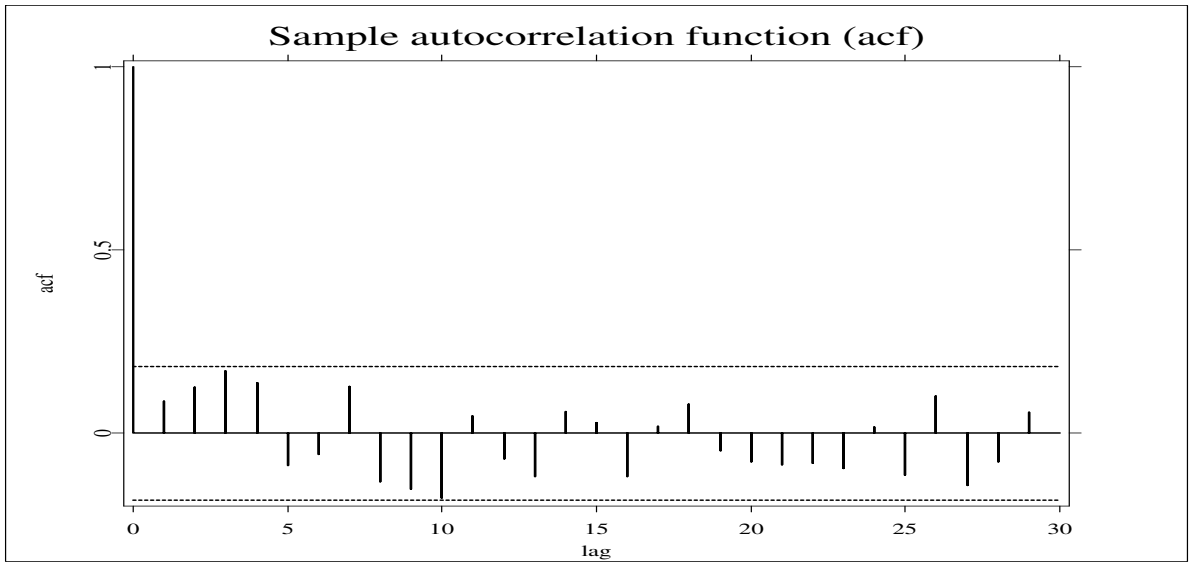


(a)

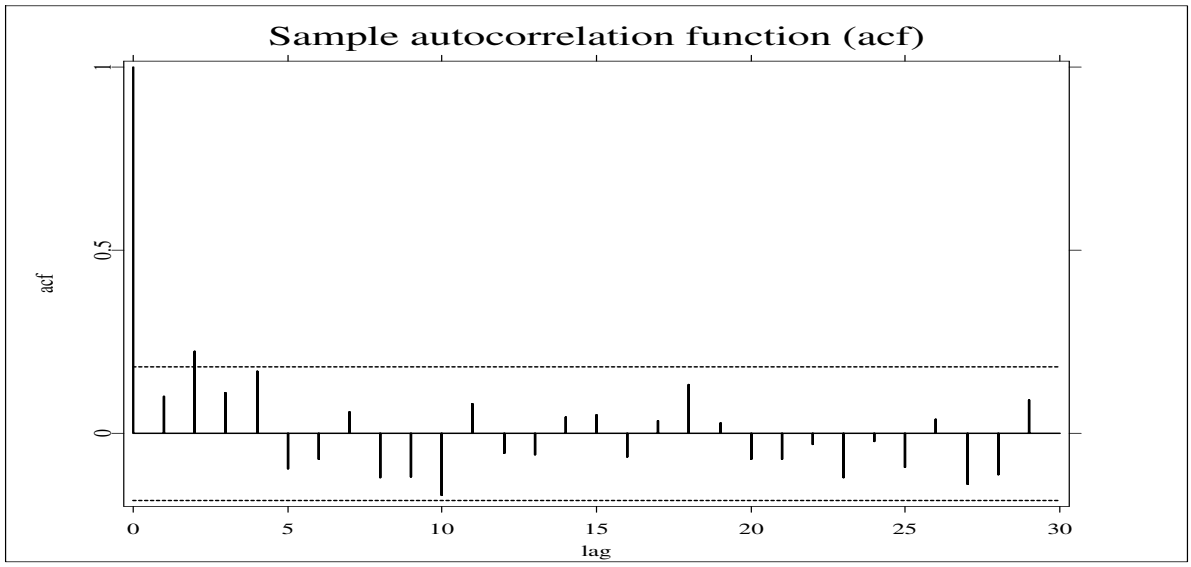


(b)

Figure 2: Plots of the West German real GNP quarterly data from 1960:1 to 1990:4: (a) the logarithm of GNP (b) the first difference of the logarithm GNP.



(a)



(b)

Figure 3: Autocorrelation functions over 30 lags of: (a) the absolute values and (b) the squares, of the standardized residuals $\hat{\varepsilon}_t$ for fitted model $Y_t = f_1(Y_{t-2})\, Y_{t-1} + f_2(Y_{t-4})\, Y_{t-3} + \sigma\varepsilon_t$, in which the time series Y_t consists of de-seasonalized first differences of the logarithm West German real GNP quarterly data from 1960:1 to 1990:4. The solid horizontal lines at levels $\pm 2 \times n^{-1/2}$, represent the 95.44% confidence bands of the autocorrelation functions.

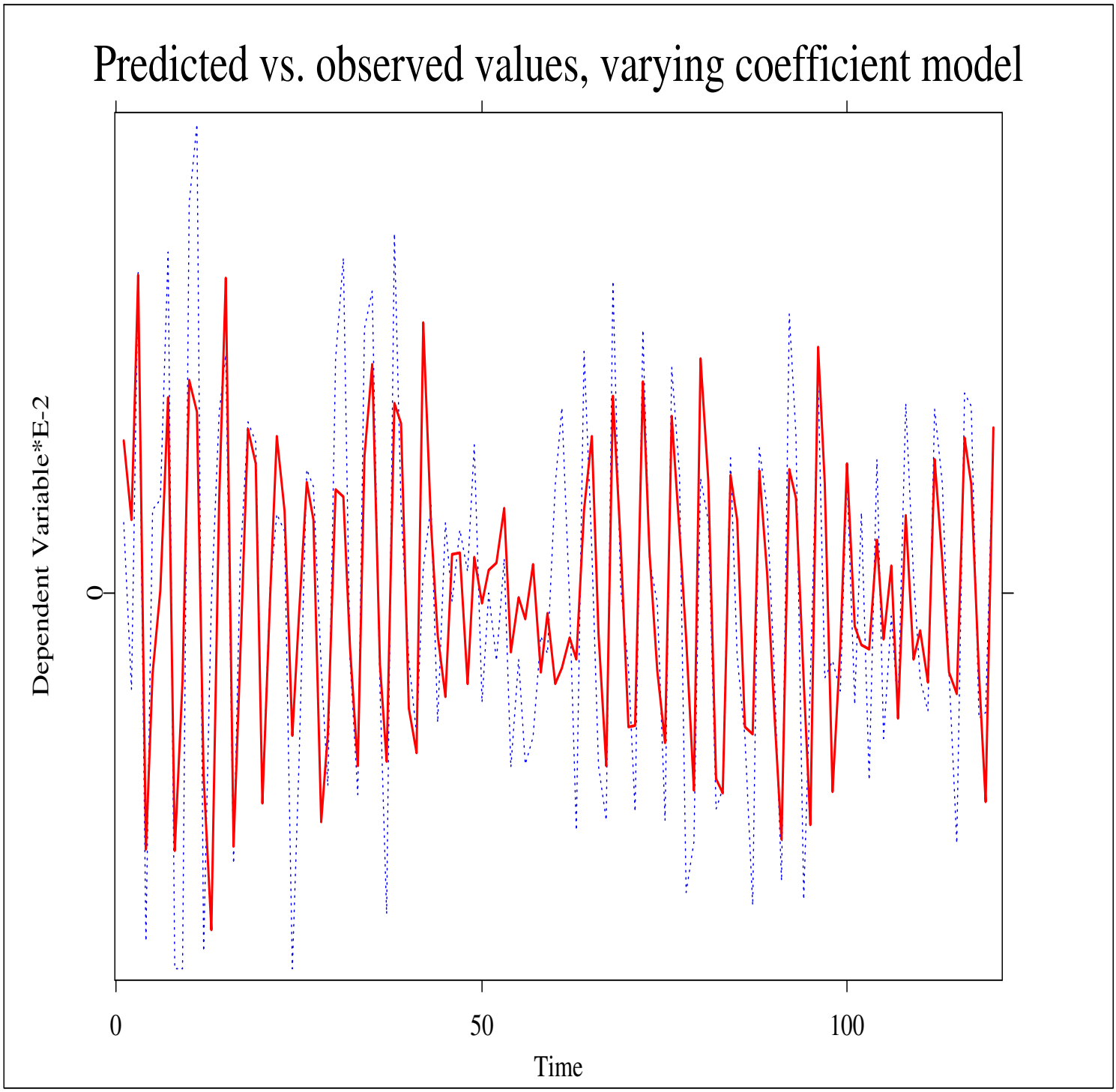


Figure 4: Prediction for the West German real GNP quarterly data from 1960:1 to 1990:4 based on marginal integration fit of varying coefficient model $Y_t = f_1(Y_{t-2})\,Y_{t-1} + f_2(Y_{t-4})\,Y_{t-3} + \sigma\varepsilon_t$. The time series Y_t consists of de-seasonalized first differences of the logarithm GNP. The response variable is $Y = Y_t$, while the predictors are $X_1 = Y_{t-2}$, $X_2 = Y_{t-4}$, $T_1 = Y_{t-1}$ and $T_2 = Y_{t-3}$. Solid curve consists of predicted values $\hat{Y}_t$ given by (14) while the dotted curve consists of the observed values Y_t.