

Yang, Lijian; Park, Byeong U.; Xue, Lan; Härdle, Wolfgang Karl

**Working Paper**

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

SFB 649 Discussion Paper, No. 2005-047

**Provided in Cooperation with:**

Collaborative Research Center 649: Economic Risk, Humboldt University Berlin

*Suggested Citation:* Yang, Lijian; Park, Byeong U.; Xue, Lan; Härdle, Wolfgang Karl (2005) : Estimation and testing for varying coefficients in additive models with marginal integration, SFB 649 Discussion Paper, No. 2005-047, Humboldt University of Berlin, Collaborative Research Center 649 - Economic Risk, Berlin

This Version is available at:

<https://hdl.handle.net/10419/25066>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

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

Lijian Yang\*  
Byeong U. Park\*\*  
Lan Xue\*\*\*  
Wolfgang Härdle\*\*\*\*



\* Department of Statistics and Probability, Michigan State University, USA

\*\* Department of Statistics, Seoul National University, Korea

\*\*\* Department of Statistics, Oregon State University, USA

\*\*\*\* Institute for Statistics and Econometrics, Humboldt-Universität zu Berlin, Germany

This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 "Economic Risk".

<http://sfb649.wiwi.hu-berlin.de>  
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin  
Spandauer Straße 1, D-10178 Berlin



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

Lijian YANG      Byeong U. PARK      Lan XUE      Wolfgang HÄRDLE

September 6, 2005

## 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 ( $\beta$ -mixing). Application of the test procedure to the West German real GNP 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

---

Lijian Yang is Associate Professor, Department of Statistics and Probability, Michigan State University, East Lansing, Michigan 48824 (E-mail: [yang@stt.msu.edu](mailto: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](mailto:bupark@stats.snu.ac.kr)). Lan Xue is Assistant Professor, Department of Statistics, Oregon State University, Corvallis, Oregon 97331-4501 (E-mail: [xuel@stat.oregonstate.edu](mailto:xuel@stat.oregonstate.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](mailto: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 partially supported by NSF grants DMS 9971186, DMS 0405330 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 partially supported by NSF grants BCS 0308420 and DMS 0405330. The authors thank the editor, the associate editor, and two referees for their insightful comments, which led a substantial improvement of the article.

# 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 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

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

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

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

is called a *varying-coefficient model* if

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

In Model I, all the variables  $\{X_s\}_{s=1}^d$  are different from each other. The model with all the variables  $\{X_s\}_{s=1}^d$  being the same, i.e.,  $m(X_i, \mathbf{T}_i) = \sum_{s=1}^d f_s(X_i)T_{is}$ , 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 (2000) and Cai, Fan and Yao (2000). 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  $X$ . 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 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 of the effectiveness of Model I, we consider a real time series data  $\{Y_t\}_{t=1}^n$  on West German GNP in Section 5. After taking first difference and de-seasonalization, the data is considered strictly stationary, as shown by the dotted curve in Figure 4. The varying coefficient models  $Y_t = f_1(Y_{t-1})Y_{t-2} + f_2(Y_{t-3})Y_{t-4} + (\text{noise})$  and  $Y_t = f_1(Y_{t-3})Y_{t-2} + f_2(Y_{t-1})Y_{t-4} + (\text{noise})$  are fitted and the estimates of the functions  $f_1$  and  $f_2$  are plotted in Figure 2. These varying coefficient AR models have 2.81 and 2.46 times, respectively, more prediction power than the simple linear AR model. See Table 3 to find  $0.00059/0.00021 = 2.81$  and  $0.00059/0.00024 = 2.46$ . More details about the data and the modelling procedures are found 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  $s = 1, \dots, d$ . In this respect, it would be of interest to compare Model I with some related FAR (functional coefficient autoregressive) models. For example, for the varying coefficient model  $Y_t = f_1(Y_{t-3})Y_{t-1} + f_2(Y_{t-4})Y_{t-2} + (\text{noise})$ , one may consider the following FAR model for a comparison:  $Y_t = f_1(Y_{t-3}, Y_{t-4})Y_{t-1} + f_2(Y_{t-3}, Y_{t-4})Y_{t-2} + (\text{noise})$ . 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 another special practical interests 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:

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

where now the coefficient functions  $f_{s1}, \dots, f_{sr_s}$  depends 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, one may have

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

in which  $Y_t$  denotes the implied volatility,  $r_t$  the interest rate,  $M_t$  the moneyness, and  $\tau_t$  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  $1/\sqrt{n}$ -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  $f_2$  can not.

We organize the paper as the follows. In Section 2, we describe marginal estimation methods for Models I and II, and derive asymptotic distribution theory of the estimators. In Section 3, a test procedure is proposed 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 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  $\varepsilon_i$  is independent of the vectors  $\{(\mathbf{X}_j, \mathbf{T}_j)\}_{j=1, \dots, i}$  for each  $i = 1, \dots, n$ . This is sufficient for obtaining our main results on distribution theory as we assume  $\{(\mathbf{X}_j, \mathbf{T}_j)\}_{j=1, \dots, n}$  to be strictly stationary and geometrically  $\beta$ -mixing in assumption A2 (see appendix.), but weaker than the usual assumption that each  $\varepsilon_i$  is independent of the vectors  $\{(\mathbf{X}_j, \mathbf{T}_j)\}_{j=1, \dots, n}$ .

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 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

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

The condition (3) may be considered as an analogue of linear independence between covariates in linear models. It is a sufficient condition of avoiding *concurvity* as termed by Hastie and Tibshirani (1990). The term *concurvity* in additive models is understood as an analogue 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 below, see Subsection A.2 of the Appendix for more details.

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

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

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  $\mathbf{Y} = (Y_i)_{1 \leq i \leq n}$  and denote  $\mathbf{p}(u) = (1, u, \dots, u^p)^T$ . Define  $\mathbf{Z}_s$  be the  $n \times (p+d)$  matrix which has  $\left(\mathbf{p}\{(X_{is} - x_s)/h\}^T T_{is}, \mathbf{T}_{i,-s}^T\right)$  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

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

where  $L_{\mathbf{g}}(\mathbf{u}) = (g_1 \cdots g_{s-1} g_{s+1} \cdots g_d)^{-1} L(g_1^{-1} u_1, \dots, g_{s-1}^{-1} u_{s-1}, g_{s+1}^{-1} u_{s+1}, \dots, g_d^{-1} u_d)$ ,  $L$  is a  $(d-1)$ -variate kernel, and  $g_1, \dots, g_{s-1}, g_{s+1}, \dots, g_d$  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

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

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.

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

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

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

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. 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). One may sacrifice some rate of convergence in order to use a lower order kernel.

Let  $\varphi$ ,  $\varphi_{-s}$  and  $\varphi_s$  denote the densities of  $\mathbf{X}$ ,  $\mathbf{X}_{-s}$  and  $X_s$ , respectively. Define

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

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

where  $K_s^*$  is an equivalent kernel defined at (A.7).

**Theorem 1** *Under the assumptions A1-A7 given in the appendix, we have, for any  $s = 1, \dots, d$ , as  $n \rightarrow \infty$ ,*

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



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 the assumptions A1-A7 given in the appendix, we have, for any  $s \neq s'$ ,*

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

as  $n \rightarrow \infty$ , and hence

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

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 remains the same while 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 above theorems. Also, the modified multifold cross-validation criterion considered by Cai, Fan and Yao (2000) may be adapted for the above estimation. Theoretical development for these bandwidth selectors is beyond the scope of the paper. Below 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  $h_{\text{opt}}$  which minimizes the asymptotic mean integrated squared error of  $\hat{f}_s$  is given by

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

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

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

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

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

with the standard normal density  $\phi$  and the rule-of-the-thumb bandwidth

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

According to its definition given at (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

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

in which 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 the condition (A7) for Theorem 1 if  $q$ , the order of the kernel  $L$ , is greater than  $(d-1)/2$ . Thus one may take  $g_j \equiv (\log n)^{-1} h_{\text{opt}}^{(p+1)/q}$  for  $j = 1, \dots, s-1, s+1, \dots, d$ , where  $h_{\text{opt}}$  is the optimal bandwidth obtained from the above procedure.

## 2.2 Model II

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

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

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:

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

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

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

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

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

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

**Theorem 3** *Under the assumptions A1-A2, A3', A4', A5-A6 and A7' given in the appendix, we have, for any  $u = 1, \dots, r_s$  and  $s = 1, \dots, d_0$ ,*

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

as  $n \rightarrow \infty$ , 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 definition of  $\kappa_{su}$  and  $\tau_{su}^2$  are given at (A.19) and (A.20)

Each pair of the entries  $\hat{f}_{su}$  and  $\hat{f}_{s'u'}$  for  $1 \leq s, s' \leq d_0$  and  $1 \leq u, u' \leq r_s$  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 analogue of Theorem 2.

**Theorem 4** *Under the assumptions of Theorem 3, we have as  $n \rightarrow \infty$*

(i) *when  $s' \neq s$*

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

(ii) *when  $s' = s$*

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

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

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

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

$$H_0 : f_s(x_s) \equiv \text{constant} \quad (11)$$

for a specific  $s$  in Model I. Testing the hypothesis (11) is a very important first step in model building procedure. 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:

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

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

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

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 the assumptions A1-A7 given in the appendix, we have, for any  $s = 1, \dots, d$ ,*

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

*as  $n \rightarrow \infty$ , where  $v_s$  and  $\gamma_s$  are as given in (A.17) and (A.16).*

For the practical implementation of the test, we suggest to use 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 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 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 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  $\tilde{m}(\mathbf{x}, \mathbf{t}) = \hat{\alpha}_s t_s + \sum_{k \neq s}^d \hat{f}_k(x_k) t_k$  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) 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  $\tilde{\varepsilon}_i = Y_i - \tilde{m}(\mathbf{X}_i, \mathbf{T}_i)$  for  $i = 1, \dots, n$ .
- (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  $Y_i^* = \tilde{m}(\mathbf{X}_i, \mathbf{T}_i) + \tilde{\varepsilon}_i Z_i^W$ .
- (iii) Compute the bootstrap test statistic  $V_{ns}^*$  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  $V_{ns,1}^*, \dots, V_{ns,M}^*$ . 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:

$$f_{su}(x_s) \equiv \text{constant}. \quad (15)$$

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

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

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  $s = 1, \dots, d_0$ ,*

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

as  $n \rightarrow \infty$ , where  $v_{su}$  and  $\gamma_{su}$  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 above for testing (11).

Some related work on this testing problem includes Chen and Liu (2001), and Cai, Fan and Yao (2000). The former paper treated testing, in the FAR model, whether all the coefficient functions are constant, i.e., 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 (i.i.d.), 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  $K(x) = L(x) = 0.9375(1 - x^2)^2 I_{(-1,1)}(x)$ , while the bandwidths were chosen as described Section 2.1.

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

$$\delta(\mathbf{X}, \mathbf{T}) = 0.5 + \frac{T_1^2 + T_2^2}{1 + T_1^2 + T_2^2} \exp(-2 + (X_1 + X_2)/2).$$

The particular form of  $\delta(\mathbf{X}, \mathbf{T})$  was considered to ensure the variance to be 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 zero and covariance matrix  $\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$ . The vectors  $\mathbf{X}$  and  $\mathbf{T}$  were generated independently. Finally, the error term  $\varepsilon$  was generated from the standard normal distribution independently of  $(\mathbf{X}, \mathbf{T})$ .

A total of 100 independent data sets with sizes  $n = 50, 100$  and  $250$  were generated. The estimated functions of  $f_s$ ,  $s = 1, 2, 3$  were evaluated on a grid of 91 equally-spaced points  $x_j$ ,  $j = 1, \dots, 91$  with  $x_1 = 0.05$ ,  $x_{91} = 0.95$ . 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

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

Here  $\hat{f}_{r,s}(x_j)$  denotes the estimated value of  $f_s$  at  $x_j$  for the  $r$ th data set,  $R = 100$  and  $g = 91$  are the numbers of data sets 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.

(Insert Table 1 about here)

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  $X_{i1}, X_{i2}, X_{i3}$ . Write them  $\hat{\beta}_{s0}(X_{i1}, X_{i2}, X_{i3})$ . These are the estimates before the marginal integration. We computed the mean average squared error

$$\begin{aligned} \text{MASE}_1 = & \frac{1}{R} \sum_{r=1}^R \frac{1}{n} \sum_{i=1}^n \left\{ \hat{\beta}_{r,10}(X_{i1}, X_{i2}, X_{i3}) + \hat{\beta}_{r,20}(X_{i1}, X_{i2}, X_{i3})T_{i1} \right. \\ & \left. + \hat{\beta}_{r,30}(X_{i1}, X_{i2}, X_{i3})T_{i2} - f_1(X_{i1}) - f_2(X_{i2})T_{i1} - f_3(X_{i3})T_{i2} \right\}^2, \end{aligned}$$

and also for the marginal integration estimate

$$\begin{aligned} \text{MASE}_2 = & \frac{1}{R} \sum_{r=1}^R \frac{1}{n} \sum_{i=1}^n \left\{ \hat{f}_{r,1}(X_{i1}) + \hat{f}_{r,2}(X_{i2})T_{i1} + \hat{f}_{r,3}(X_{i3})T_{i2} \right. \\ & \left. - f_1(X_{i1}) - f_2(X_{i2})T_{i1} - f_3(X_{i3})T_{i2} \right\}^2, \end{aligned}$$

where  $\hat{\beta}_{r,s0}(X_{i1}, X_{i2}, X_{i3})$  and  $\hat{f}_{r,s}(X_{is})$  are the estimates for the  $r$ th dataset. We found  $\text{MASE}_1 = 0.3164$  and  $\text{MASE}_2 = 0.2761$ .

Next, we give some numerical results for the testing method. For each of the simulated data sets above, 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  $\alpha = 0.05$  among the 100 replications.

(Insert Table 2 about here)

## 4.2 The time series case

In this simulation,  $R = 200$  time series were generated. Each time, 1000 observations were generated from the following varying coefficient AR (VCAR) model, among which only the last 250 observations were used:

$$Y_t = f_1(Y_{t-3})Y_{t-1} + f_2(Y_{t-4})Y_{t-2} + 0.2\varepsilon_t \quad (16)$$

where  $f_1(Y_{t-3}) = 0.4 + (0.1 + Y_{t-3}) \exp(-3Y_{t-3}^2)$ ,  $f_2(Y_{t-4}) = -0.2 - (0.6 + Y_{t-4}) \exp(-3Y_{t-4}^2)$ , and  $\varepsilon_t$  are i.i.d. standard normal random variates. Again, the performance of the estimators of  $f_1$  and  $f_2$  were assessed by MISE. We obtained  $\text{MISE}(\hat{f}_1) = 0.0137$  and  $\text{MISE}(\hat{f}_2) = 0.0151$ .



We found that the Monte Carlo average over 200 time series of  $\sqrt{\sum_{t=1}^{250} (Y_t - \bar{Y})^2 / 250}$  equals 0.6374 with the standard error 0.0026, where  $\bar{Y} = \sum_{t=1}^{250} Y_t / 250$ . The obtained values of MISEs 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  $Y$ .

Similarly as in the i.i.d. case, we report here a numerical result for testing  $H_{10} : f_1 = \text{constant}$  and  $H_{20} : f_2 = \text{constant}$ . For each of the simulated time series we applied the wild bootstrap method with  $M = 500$  and used the significant level 0.05. We found that the proportion of the cases where the null hypothesis was rejected among the 200 replications was 0.57 for  $H_{10}$  and 0.943 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, Fan and Li (2000) and Cai, Fan and Yao (2000), where all the coefficient functions depend on the same variable(s). For this purpose, we considered the following three FAR models:

$$Y_t = g_1(Y_{t-3})Y_{t-1} + g_2(Y_{t-3})Y_{t-2} + 0.2\varepsilon_t, \quad (17)$$

$$Y_t = g_1(Y_{t-4})Y_{t-1} + g_2(Y_{t-4})Y_{t-2} + 0.2\varepsilon_t, \quad (18)$$

$$Y_t = g_1(Y_{t-3}, Y_{t-4})Y_{t-1} + g_2(Y_{t-3}, Y_{t-4})Y_{t-2} + 0.2\varepsilon_t. \quad (19)$$

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

$$\text{MASR} = \sum_{r=1}^{200} \sum_{t=1}^{250} (y_{r,t} - \hat{y}_{r,t})^2 / (200 \times 250),$$

where  $y_{r,t}$  denotes the  $t$ th observation in the  $r$ th replication, and  $\hat{y}_{r,t}$  is the corresponding fitted value based on the underlying model. We note that 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 to compare 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 the MASR values 1.020, 0.343 and 0.081, respectively, whereas the VCAR model (16) gave a much smaller 0.075.

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

and computed

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

where  $\hat{y}_{r,t}$  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 the MASPE values 0.075, 0.071 and 0.062, respectively, while the VCAR model (16) gave 0.059.

## 5 AN EMPIRICAL EXAMPLE

We illustrate our estimation and testing methods with an analysis of the quarterly (seasonally non-adjusted) West German real GNP data collected 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 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 a time series data,  $D_t$ ,  $1 \leq t \leq n = 124$ , which is 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  $-0.065116$ ,  $0.038595$ ,  $0.051829$  and  $0.008944$ , respectively, were calculated and subtracted from  $D_t$  so that the de-seasonalized  $Y_t$ ,  $1 \leq t \leq n = 124$ , 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  $Y_t$ ,  $1 \leq t \leq n = 124$ , is plotted as the dotted curve.

*(Insert Figure 1 and Table 3 about here)*

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 the twelve 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-of-sample predictions for the last 10 observations. Their average squared residuals (ASR) and average squared prediction errors (ASPE) are reported in Table 3. One may expect the ASRs should 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

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

while the model with the smallest ASPE is

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

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

$$Y_t = c_1Y_{t-2} + c_2Y_{t-4} + (\text{noise}). \quad (22)$$

In Table 3, the ASR and ASPE of the linear AR model (22) are also included. Both the optimal VCAR models (20) and (21) have much smaller ASR and ASPE than the linear AR model. These two VCAR models have similar values of ASR and ASPE. 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 procedures. For the model (20), the p-values were 0.80 for  $f_1$  and 0.01 for  $f_2$ , while for the model (21) they were 0.22 and 0.48, 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 for that  $f_2$  is not a constant. Thus one may conclude that a parsimonious model is the partially linear model:

$$Y_t = f_1Y_{t-2} + f_2(Y_{t-3})Y_{t-4} + (\text{noise}).$$

We further computed the ASR and ASPE of this semiparametric partially linear model, which are 0.00032 and 0.00024 respectively, as seen 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 while 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 BIC works for linear AR time series where ASR is similar to AIC. For linear AR time series, it is well known that AIC is optimal for forecasting while BIC is consistent in identifying a correct AR model. It should be noted also that ASR can be compared across models not necessarily nested within each other, while 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 it is another way of assessing goodness-of-fit for the models. At a practical level, such independence can be checked using the autocorrelation

functions (ACF) 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  $|\hat{\varepsilon}_t|^3$ ,  $\hat{\varepsilon}_t^4$ , etc., led to the same conclusion. Thus, the models (20) and (21) fit well the structure of the data  $Y_t$ . As a 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.

*(Insert Figures 2, 3 and 4 about here)*

## APPENDIX: PROOFS

A longer version of the paper with proofs of greater detail may be found at <http://ace.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. 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  $\beta$ -mixing with mixing coefficients  $\beta(k) \leq C_2 \rho^k$ ,  $0 < \rho < 1$ . 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  $\sigma$ -algebra generated by  $(\mathbf{X}_t, \mathbf{T}_t), (\mathbf{X}_{t+1}, \mathbf{T}_{t+1}), \dots, (\mathbf{X}_{t'}, \mathbf{T}_{t'})$  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  $f_s$ 's have bounded continuous  $(p+1)$ -th derivatives for all  $1 \leq s \leq d$ , and  $p \geq q-1$
- A4: The distribution of  $(\mathbf{X}, \mathbf{T})$  has a density  $\psi$  and  $\mathbf{X}$  has a marginal density  $\varphi$ . On the supports of weight functions  $w_{-s}$  and  $w_s$ , the densities  $\varphi_{-s}$  of  $\mathbf{X}_{-s}$  and  $\varphi_s$  of  $X_s$ , respectively, are uniformly bounded away from zero and infinity. The marginal density  $\varphi$  and  $E(T_s T_{s'} | \mathbf{X} = \cdot)$  for  $1 \leq s, s' \leq d$  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.
- A6: The error term  $\varepsilon_t$  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  $\mathcal{X}_\epsilon = \{\mathbf{x} : \text{there exists } \mathbf{z} \in \mathcal{X} \text{ such that } \|\mathbf{z} - \mathbf{x}\| \leq \epsilon\}$ . 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  $\sigma(\mathbf{x}, \mathbf{t}) \leq \tilde{\sigma}(\mathbf{t})$  for all  $\mathbf{x} \in \mathcal{X}_\epsilon$ ,  $\psi_{j,k,l,m}(\mathbf{x}_j, \mathbf{t}_j; \mathbf{x}_k, \mathbf{t}_k; \mathbf{x}_l, \mathbf{t}_l; \mathbf{x}_m, \mathbf{t}_m) \leq \tilde{\varphi}_{j,k,l,m}(\mathbf{t}_j, \mathbf{t}_k, \mathbf{t}_l, \mathbf{t}_m)$  for all  $\mathbf{x}_j, \mathbf{x}_k, \mathbf{x}_l, \mathbf{x}_m$  in  $\mathcal{X}_\epsilon$ , and  $\int (\|\mathbf{t}_j\| \|\mathbf{t}_k\| \|\mathbf{t}_l\| \|\mathbf{t}_m\|)^{2+c} |\tilde{\sigma}(\mathbf{t}_j) \tilde{\sigma}(\mathbf{t}_k) \tilde{\sigma}(\mathbf{t}_l) \tilde{\sigma}(\mathbf{t}_m)|^{2+c} \tilde{\varphi}_{j,k,l,m}(\mathbf{t}_j, \mathbf{t}_k, \mathbf{t}_l, \mathbf{t}_m) d\mathbf{t}_j d\mathbf{t}_k d\mathbf{t}_l d\mathbf{t}_m \leq C < \infty$  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  $(nh \ln n)^{1/2} g_{\max}^a \rightarrow 0$  as  $n \rightarrow \infty$  where  $g_{\text{prod}} = g_1 \cdots g_{s-1} g_{s+1} \cdots g_d$  and  $g_{\max} = \max(g_1, \dots, g_{s-1}, g_{s+1}, \dots, g_d)$ , and  $h$  is asymptotic to  $n^{-1/(2p+3)}$ .

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

- A3': The functions  $f_{su}$ 's have bounded continuous  $(p+1)$ -th derivatives for all  $1 \leq s \leq d_0$ ,  $1 \leq u \leq r_s$  and  $p \geq q-1$ .
- A4': It is the same as A4 except that now we require  $E(T_{su} T_{s'u'} | \mathbf{X} = \cdot)$  for  $1 \leq s, s' \leq d_0$  and  $1 \leq u, u' \leq r_s$  are Lipschitz continuous.
- A7': It is also the same as A7 except that  $d$  is replaced by  $d_0$ .

One should note here that for existence of the bandwidth vector  $\mathbf{g}$  satisfying the assumption A7 and A7' it is necessary that  $q$ , the order of the kernel  $L$ , should 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  $U$ -statistic and von Mises statistic of dependent variables derived from Yoshihara (1976). Let  $\xi_i, 1 \leq i \leq n$  be a strictly stationary sequence of random variables with values in  $R^d$  and  $\beta$ -mixing coefficients  $\beta(k), k = 1, 2, \dots$ , and  $r$  a fixed positive integer. Let  $\{\theta_n(F)\}$  denote the functionals of the distribution function  $F$  of  $\xi_i$

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

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

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

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

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

The von Mises' differentiable statistic and the  $U$ -statistic

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

allow decompositions as

$$\begin{aligned} \theta_n(F_n) &= \theta_n(F) + \sum_{c=1}^m \binom{m}{c} V_n^{(c)}, \\ U_n &= \theta_n(F) + \sum_{c=1}^m \binom{m}{c} U_n^{(c)}. \end{aligned}$$

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

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

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

$$\begin{aligned} V_n^{(c)} &= \int g_{n,c}(x_1, \dots, x_c) \prod_{j=1}^c [dF_n(x_j) - dF(x_j)], \\ U_n^{(c)} &= \frac{(n-c)!}{n!} \sum_{1 \leq i_1 < \dots < i_c \leq n} \int g_{n,c}(x_{i_1}, \dots, x_{i_c}) \prod_{j=1}^c [dI_{R_+^d}(x_j - \xi_{i_j}) - dF(x_j)] \end{aligned}$$

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

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

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

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

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

**Proof.** The proof essentially is the same as 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 at (3) is closely related to the invertibility of the matrix  $S_s(\mathbf{X})$ . To see this, we note that for vectors  $\lambda_1$  and  $\lambda_2$  of dimensions  $p+1$  and  $d-1$ , respectively,

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

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  $u \in \text{supp}(K)$ . Since  $K$  has a nonempty interior, the identifiability condition (3) implies  $\lambda_1 \equiv 0$  and  $\lambda_2 \equiv 0$  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  $n \rightarrow \infty$*

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

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  $\beta$ -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  $k$ ,*

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

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

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

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

**Lemma A.4** *Define*

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



Then, as  $n \rightarrow \infty$

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

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

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

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

$$R_{\ell 2}(x_s) = \frac{1}{n} \sum_{i=1}^n \frac{w_{-s}(\mathbf{X}_{i,-s})}{\varphi(x_s, \mathbf{X}_{i,-s})} S_s^{-1}(x_s, \mathbf{X}_{i,-s}) \left\{ I_{p+d} - \frac{\mathbf{Z}_s^T \mathbf{W}_{is} \mathbf{Z}_s S_s^{-1}(x_s, \mathbf{X}_{i,-s})}{\varphi(x_s, \mathbf{X}_{i,-s})} \right\}^\ell \times \mathbf{Z}_s^T \mathbf{W}_{is} \left[ \{f_s(X_{js})\}_{j=1}^n - \sum_{\nu=0}^p \frac{f_s^{(\nu)}(x_s) h^\nu}{\nu!} \mathbf{Z}_s e_\nu \right] \quad (\text{A.4})$$

$$R_{\ell 3}(x_s) = \frac{1}{n} \sum_{i=1}^n \frac{w_{-s}(\mathbf{X}_{i,-s})}{\varphi(x_s, \mathbf{X}_{i,-s})} S_s^{-1}(x_s, \mathbf{X}_{i,-s}) \left\{ I_{p+d} - \frac{\mathbf{Z}_s^T \mathbf{W}_{is} \mathbf{Z}_s S_s^{-1}(x_s, \mathbf{X}_{i,-s})}{\varphi(x_s, \mathbf{X}_{i,-s})} \right\}^\ell \times \mathbf{Z}_s^T \mathbf{W}_{is} \left[ \left\{ \sum_{s' \neq s} f_{s'}(X_{js'}) \right\}_{j=1}^n - \sum_{s' \neq s} f_{s'}(X_{is'}) \mathbf{Z}_s e_{p+s'} \right]. \quad (\text{A.5})$$

Then, as  $n \rightarrow \infty$ ,

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

**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

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

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

Denote  $\xi_i = (\mathbf{X}_i, \mathbf{T}_i, Y_i)$ , The term  $P_1$  can be written as the von Mises' differentiable statistic  $V_n = (2n^2)^{-1} \sum_{i,j=1}^n g_n(\xi_i, \xi_j)$  where  $g_n(\xi_i, \xi_j)$  equals

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

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

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

which has mean zero and variance of order  $b^2/nh$ . So  $V_n^{(1)} = n^{-1} \sum_{j=1}^n g_{n,1}(\xi_j) = o_p(b/\sqrt{nh})$ . 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

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

by Lemma 1 of Yoshihara (1976).

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

Now applying Lemma A.1 with  $m = c = 2$  and  $r = 1$ , (A.2) gives

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

by making  $\delta$  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

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

and define an equivalent kernel

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

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

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

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  $\delta_{jk}$  equal 1 if  $j = k$  and 0 otherwise. Then*

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

In order to prove Theorem 1, we begin by observing

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

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

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

Then, we obtain  $Q_{1n} \left\{ \hat{f}_s(x_s) - f_s(x_s) \right\} = Q_{2n}(x_s)$ . 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

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

where for  $a = 1, 2, 3$

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

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  $n \rightarrow \infty$ ,

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

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

**Proof.** By the definition (A.11) and using Lemma A.1 for geometrically  $\beta$ -mixing processes,

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

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

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

This completes the proof of the lemma. ■

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

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

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

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

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

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

This completes the proof of the lemma. ■

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

**Proof.** By definition (A.11) and applying Lemma A.1, one has

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

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  $x_s \in \text{supp}(w_s)$  is asymptotically normal with mean 0 and variance

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

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

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

The theorem now follows immediately from Lemmas A.7, A.8, the conditions on the bandwidths as given in 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.** 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

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

Applying (A.12), 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  $\varepsilon_j$ 's are i.i.d. white noise and each  $\varepsilon_i$  is independent of the vectors  $(\mathbf{X}_j, \mathbf{T}_j)$ ,  $j = 1, \dots, i$  for each  $i = 1, \dots, n$ , we need only to show that

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

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.13) is actually  $O(1)$ , which proves the theorem. ■

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

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

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  $\sum_{k=1}^n w_s(X_{ks}) P_{1n}^2(X_{ks})/n$ . Let  $\xi_i = (\mathbf{X}_i, \mathbf{T}_i, Y_i)$  and define

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

Then, by the definition (A.11)

$$\sum_{k=1}^n w_s(X_{ks}) P_{1n}^2(X_{ks})/n = n^{-5} \sum_{i,j,k,l,m=1}^n \tilde{g}_n(\xi_i, \xi_j, \xi_k, \xi_l, \xi_m).$$

Next, we define  $g_n(\xi_i, \xi_j, \xi_k, \xi_l, \xi_m) = \sum \tilde{g}_n(\xi_{i'}, \xi_{j'}, \xi_{k'}, \xi_{l'}, \xi_{m'})/5!$ , 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  $n^{-5} \sum_{i,j,k,l,m=1}^n g_n(\xi_i, \xi_j, \xi_k, \xi_l, \xi_m)$ . It is easy to see that  $g_{n,0} = 0$ ,  $g_{n,1} = 0$ , and by changes of variables  $g_{n,2}(\xi_j, \xi_m)$  equals

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

To establish the asymptotic normality of the off-diagonal sum  $2n^{-2} \sum_{1 \leq j < m \leq n} g_{n,2}(\xi_j, \xi_m)$ , we use Lemma 3.2 of Hjellvik, Yao and Tjøstheim (1998). Let  $\delta_n^2$  be their  $\sigma_n^2$ , i.e.,  $\delta_n^2 = \sum_{1 \leq j < m \leq n} \text{var} \{2n^{-2} g_{n,2}(\xi_j, \xi_m)\}$ . Define  $\lambda_{nk}$  in the same way as their  $M_{nk}$  for  $k = 1, \dots, 6$  with  $2n^{-2} g_{n,2}(\xi_j, \xi_m)$  taking the role of their  $\varphi_{jm}$ . If we prove that for some  $\zeta > 0$

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

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

then we establish that  $2n^{-2} \sum_{1 \leq j < m \leq n} g_{n,2}(\xi_j, \xi_m)$  is asymptotically normal with mean 0 and variance  $\delta_n^2$ .

We compute  $\delta_n^2$  first. Note that

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

By the 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

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

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

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

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 the assumption A6 the right hand side of the above inequality is bounded by

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

This shows

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

Similarly, we can establish

$$\begin{aligned}
n^2 \delta_n^{-2} \lambda_{n1}^{1/(\zeta+1)} & \asymp n^2 \times n^2 h \times n^{-4} h^{-2\zeta/(\zeta+1)} = h^{(1-\zeta)/(1+\zeta)}, \\
n^{3/2} \delta_n^{-2} \lambda_{n2}^{1/\{2(\zeta+1)\}} & \asymp n^{3/2} \times n^2 h \times n^{-4} h^{-(1+2\zeta)/(1+\zeta)} = n^{-(2p+2p\zeta+3+\zeta)/\{2(1+\zeta)(2p+3)\}}, \\
n^{3/2} \delta_n^{-2} \lambda_{n3}^{1/2} & \asymp n^{3/2} \times n^2 h \times n^{-4} h^{-3/2} = (nh)^{-1/2}, \\
n^2 \delta_n^{-2} \lambda_{n5}^{1/\{2(\zeta+1)\}} & \asymp n^2 \times n^2 h \times n^{-4} h^{-(1+2\zeta)/\{2(\zeta+1)\}} = h^{1/\{2(1+\zeta)\}}, \\
n^2 \delta_n^{-2} \lambda_{n6}^{1/2} & \asymp n^2 \times n^2 h \times n^{-4} h^{-1/2} = h^{1/2}.
\end{aligned}$$

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

By the martingale central limit theorem again, the diagonal sum  $n^{-2} \sum_{j=1}^n g_{n,2}(\xi_j, \xi_j)$  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

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

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



Therefore, we establish

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

Application of Lemma A.1 reveals that  $n^{-c} \sum_{j_1, \dots, j_c=1}^n g_{n,c}(\xi_{j_1}, \dots, \xi_{j_c}) = o(n^{-1}h^{-1/2})$  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  $o(n^{-1}h^{-1/2})$  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,

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

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  $K_{(su)}^*(v; \mathbf{t}, \mathbf{x}) = e_0^T S_{su}^{-1}(\mathbf{x}) \mathbf{q}_{su}(v; \mathbf{t}) K(v)$ . Let  $K_{(su)}^{*(c)}$  denote the two-folded convolution of  $K_{(su)}^*$ . Theorems 3 and 6 can be proved in the same way as in the proofs of Theorems 1 and 5 with the following definitions of  $\kappa_{su}$ ,  $\tau_{su}^2$ ,  $\gamma_{su}$  and  $v_{su}$ :

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

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

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

$$\times \psi(x_s, \mathbf{x}_{-s}, \mathbf{t}_1) \psi(x_s, \mathbf{z}_{-s}, \mathbf{t}_2) dv dx_s d\mathbf{x}_{-s} d\mathbf{z}_{-s} d\mathbf{t}_1 d\mathbf{t}_2, \\ v_{su} = \int \frac{w_{-s}^2(\mathbf{x}_{-s}) w_s(x_s)}{\eta_s^2 \varphi^2(\mathbf{x})} K_{(su)}^{*2}(v; \mathbf{t}, \mathbf{x}) \varphi_{-s}^2(\mathbf{x}_{-s}) \sigma^2(\mathbf{x}, \mathbf{t}) \psi(\mathbf{x}, \mathbf{t}) \times \varphi_s(x_s) dv d\mathbf{x} d\mathbf{t}. \quad (\text{A.22})$$

The proof of Theorem 4(i) is the same as that of the first part of Theorem 2. For the

proof of Theorem 4(ii), define

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

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  $\text{cov}(\hat{f}_{su}(x_s), \hat{f}_{su'}(x_s)) = \eta_s^{-2} \tau_{suu'}(x_s) n^{-1} h^{-1} \{1 + o(1)\}$  where

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

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

## 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 the American Statistical Association*, 95, 888-902.
- Cai, Z., Fan, J. and Yao, Q. (2000), “Functional-Coefficient Regression Models for Nonlinear Times Series”, *Journal of the American Statistical Association*, 95, 941-956.
- Chen, R. and Liu, L. M. (1993a), “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.
- 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.

- 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: 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 and Hall.
- Hastie, T. J. and Tibshirani, R. J. (1993), “Varying-Coefficient Models”, *Journal of the Royal Statistical Society Series 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.
- 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*, 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”, *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.
- Masry, E. and Tjøstheim, D. (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.
- 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 and 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)”, *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.

Table 1: MISEs of the estimators  $\hat{f}_1$ ,  $\hat{f}_2$  and  $\hat{f}_3$  for the i.i.d. case.

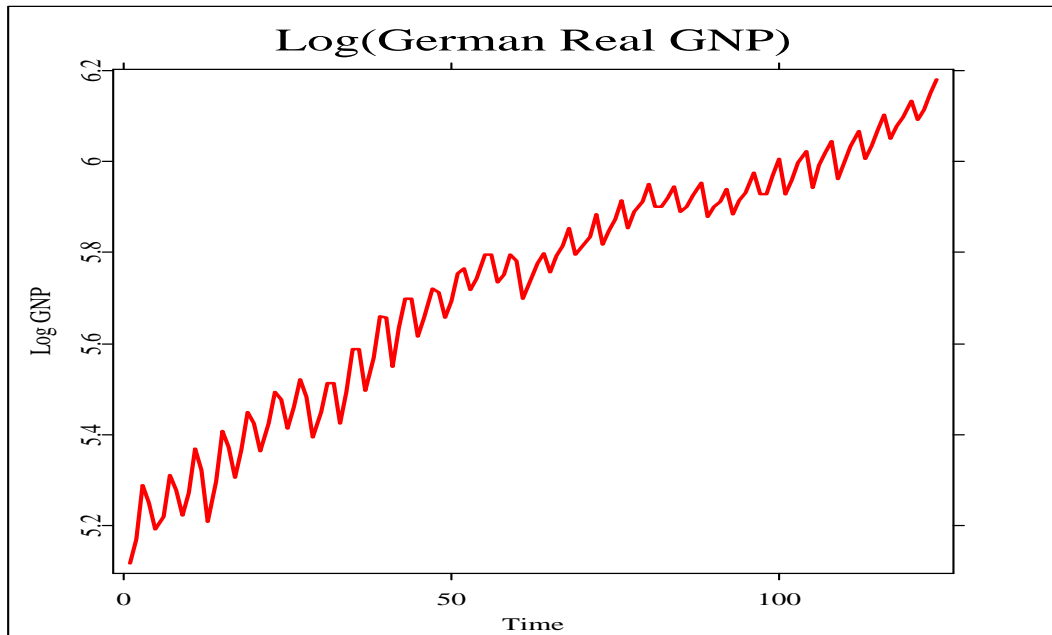
	$f_1$	$f_2$	$f_3$
n=50	0.0559	0.1144	0.1336
n=100	0.0300	0.0515	0.0617
n=250	0.0108	0.0223	0.0225

Table 2: Proportions among the 100 replications of rejecting the null hypotheses  $H_{s0}$ ,  $s = 1, 2, 3$ , at the significant level 0.05 for the i.i.d. case.

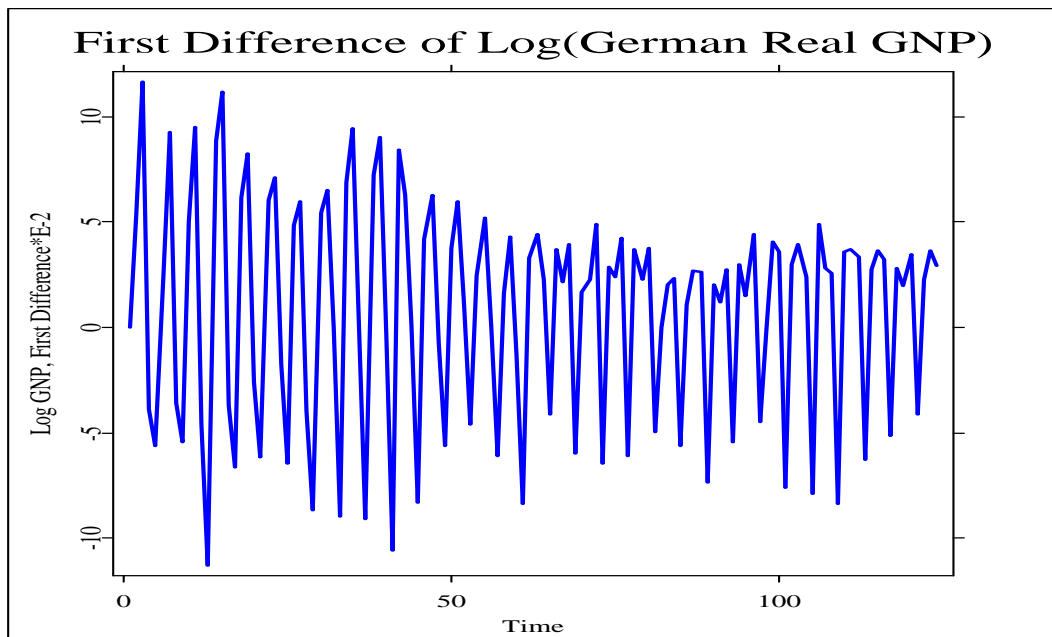
	$H_{10}$	$H_{20}$	$H_{30}$
n=50	0.94	0.85	0.02
n=100	1	1	0.04
n=250	1	1	0.08

Table 3: Average squared residuals (ASR) and average squared prediction errors (ASPE) obtained from fitting twelve VCAR models with the German real GNP data. Each model is identified by the four digits which indicate the order in which the lagged variables  $Y_{t-1}$ ,  $Y_{t-2}$ ,  $Y_{t-3}$  and  $Y_{t-4}$  enter the VCAR model. For example, the model ‘1234’ means  $Y_t = f_1(Y_{t-1})Y_{t-2} + f_2(Y_{t-3})Y_{t-4} + (\text{noise})$ . The partially linear VCAR model at the bottom is  $Y_t = f_1Y_{t-2} + f_2(Y_{t-3})Y_{t-4} + (\text{noise})$ .

Model	ASR	ASPE
1234	<b>0.00021</b>	0.00011
1243	0.00040	0.00019
1324	0.00025	0.00013
1342	0.00039	0.00016
1423	0.00026	0.00014
1432	0.00024	<b>0.00009</b>
2134	0.00023	0.00017
2143	0.00051	0.00037
2341	0.00049	0.00032
2431	0.00024	0.00015
3142	0.00041	0.00023
3241	0.00038	0.00017
Linear AR	0.00059	0.00041
Partially Linear VCAR	0.00032	0.00024



(a)



(b)

Figure 1: Plots of the West German real GNP quarterly data from 1960:1 to 1990:4. Panel (a) shows  $\log(\text{GNP})$  over time, and (b) depicts the first difference of  $\log(\text{GNP})$ .

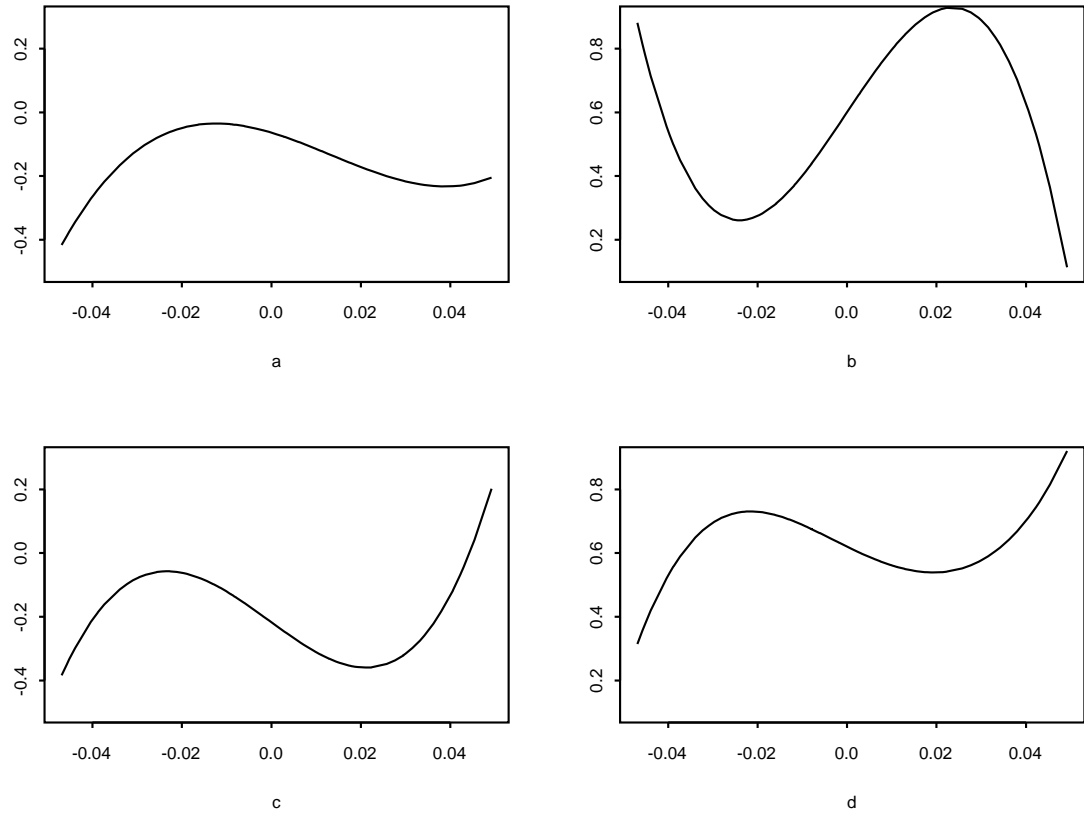


Figure 2: Estimated functions under the models (20) and (21). Panels (a) and (b) depict  $\hat{f}_1$  and  $\hat{f}_2$ , respectively, for the model (20), while (c) and (d) are for the model (21).



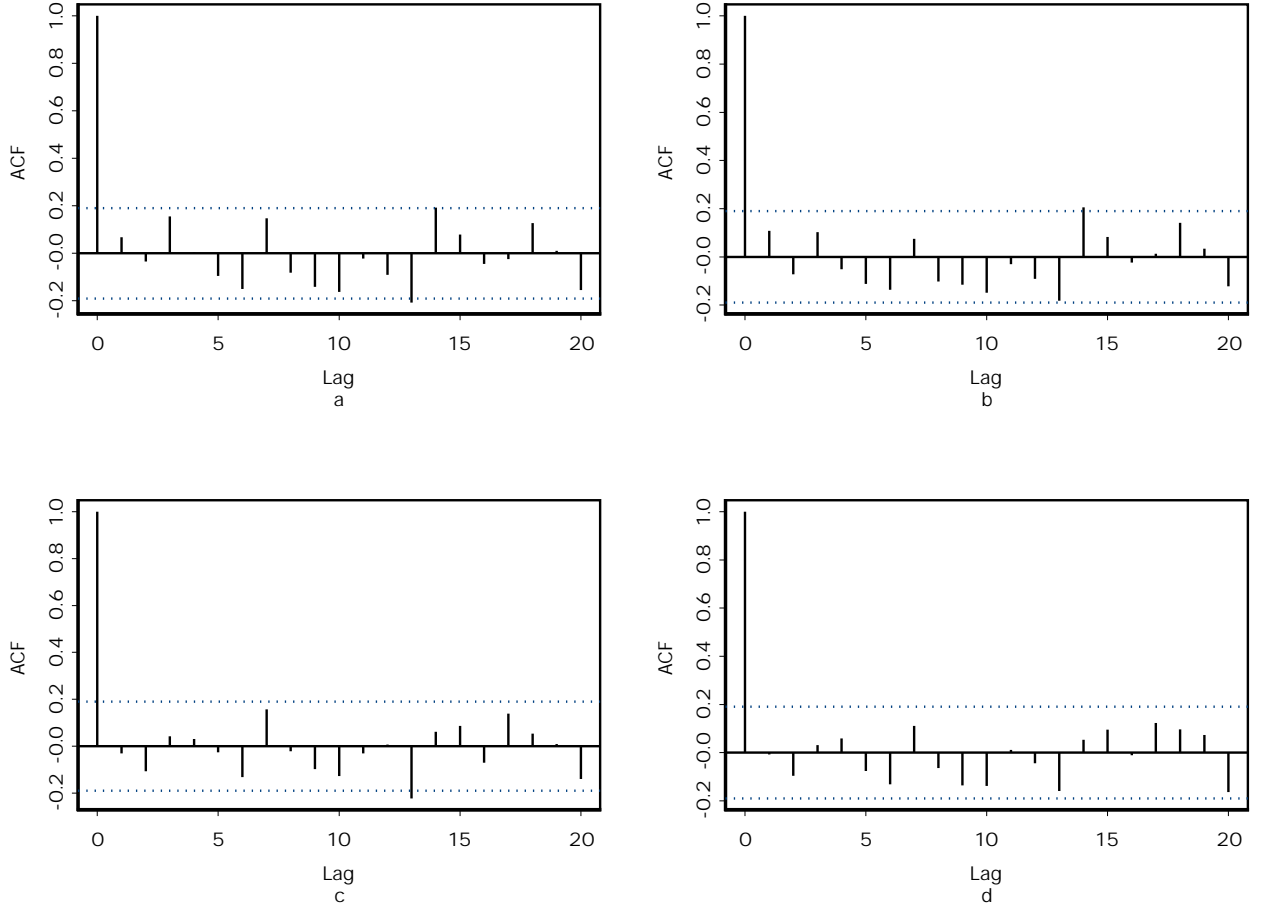
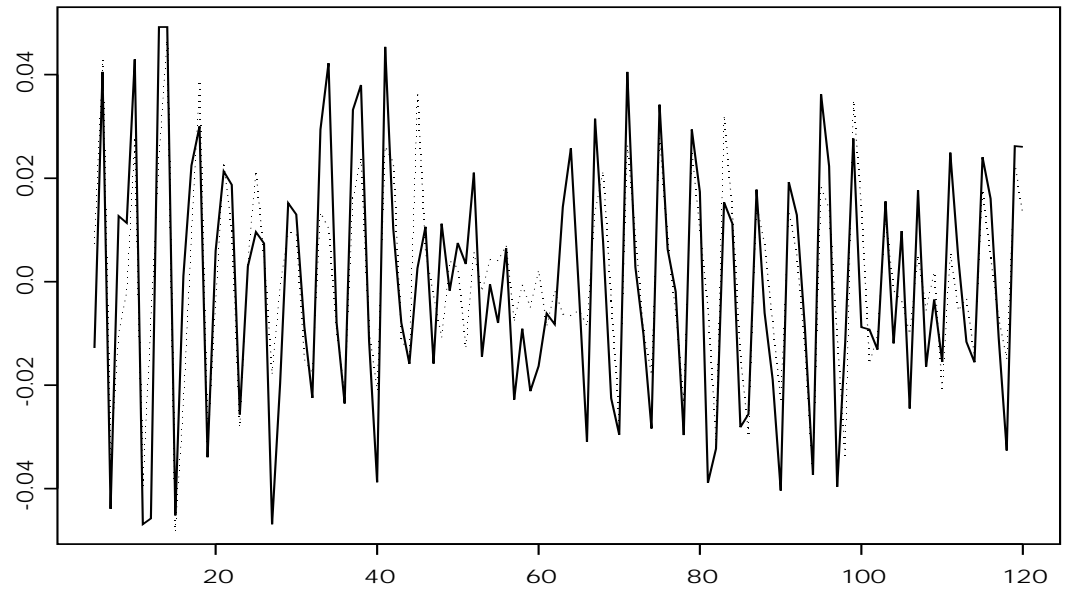
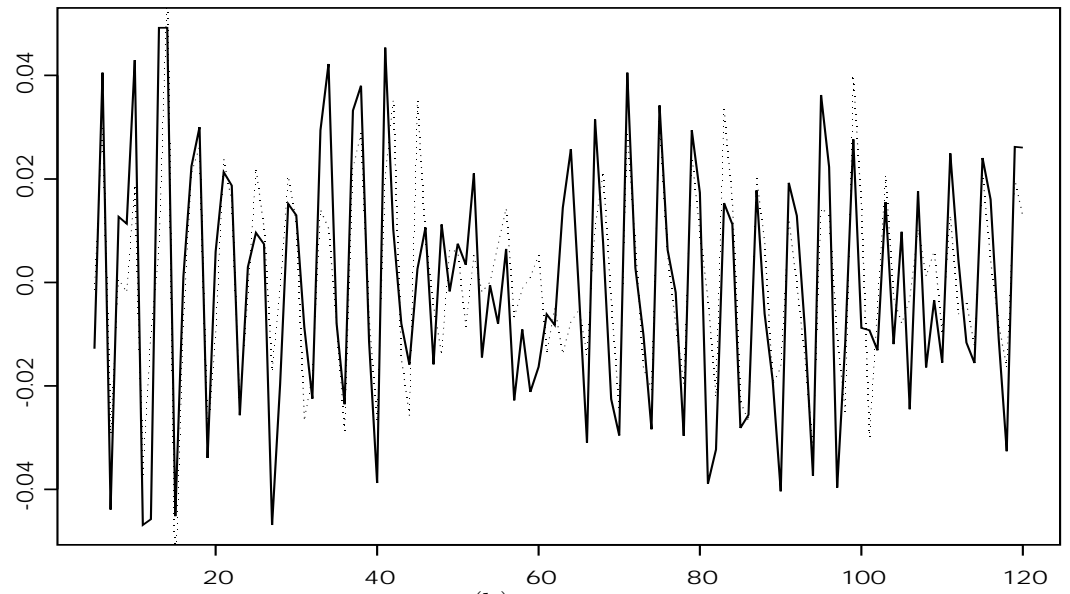


Figure 3: Autocorrelations of standardized residuals  $\hat{\varepsilon}_t$ . Panels (a) and (b) are for the model (20) and depict the autocorrelations of  $|\hat{\varepsilon}_t|$  and  $\hat{\varepsilon}_t^2$ , respectively, while (c) and (d) are for the model (21). The dotted horizontal lines at levels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



(a)



(b)

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 the model (20), and (b) is for (21). Solid lines represent the predicted values  $\hat{Y}_t$ , while the dotted are for the observed values  $Y_t$ .

## SFB 649 Discussion Paper Series

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 001 "Nonparametric Risk Management with Generalized Hyperbolic Distributions" by Ying Chen, Wolfgang Härdle and Seok-Oh Jeong, January 2005.
- 002 "Selecting Comparables for the Valuation of the European Firms" by Ingolf Dittmann and Christian Weiner, February 2005.
- 003 "Competitive Risk Sharing Contracts with One-sided Commitment" by Dirk Krueger and Harald Uhlig, February 2005.
- 004 "Value-at-Risk Calculations with Time Varying Copulae" by Enzo Giacomini and Wolfgang Härdle, February 2005.
- 005 "An Optimal Stopping Problem in a Diffusion-type Model with Delay" by Pavel V. Gapeev and Markus Reiß, February 2005.
- 006 "Conditional and Dynamic Convex Risk Measures" by Kai Detlefsen and Giacomo Scandolo, February 2005.
- 007 "Implied Trinomial Trees" by Pavel Čížek and Karel Komorád, February 2005.
- 008 "Stable Distributions" by Szymon Borak, Wolfgang Härdle and Rafal Weron, February 2005.
- 009 "Predicting Bankruptcy with Support Vector Machines" by Wolfgang Härdle, Rouslan A. Moro and Dorothea Schäfer, February 2005.
- 010 "Working with the XQC" by Wolfgang Härdle and Heiko Lehmann, February 2005.
- 011 "FFT Based Option Pricing" by Szymon Borak, Kai Detlefsen and Wolfgang Härdle, February 2005.
- 012 "Common Functional Implied Volatility Analysis" by Michal Benko and Wolfgang Härdle, February 2005.
- 013 "Nonparametric Productivity Analysis" by Wolfgang Härdle and Seok-Oh Jeong, March 2005.
- 014 "Are Eastern European Countries Catching Up? Time Series Evidence for Czech Republic, Hungary, and Poland" by Ralf Brüggemann and Carsten Trenkler, March 2005.
- 015 "Robust Estimation of Dimension Reduction Space" by Pavel Čížek and Wolfgang Härdle, March 2005.
- 016 "Common Functional Component Modelling" by Alois Kneip and Michal Benko, March 2005.
- 017 "A Two State Model for Noise-induced Resonance in Bistable Systems with Delay" by Markus Fischer and Peter Imkeller, March 2005.
- 018 "Yxilon – a Modular Open-source Statistical Programming Language" by Sigbert Klinke, Uwe Ziegenhagen and Yuval Guri, March 2005.
- 019 "Arbitrage-free Smoothing of the Implied Volatility Surface" by Matthias R. Fengler, March 2005.
- 020 "A Dynamic Semiparametric Factor Model for Implied Volatility String Dynamics" by Matthias R. Fengler, Wolfgang Härdle and Enno Mammen, March 2005.
- 021 "Dynamics of State Price Densities" by Wolfgang Härdle and Zdeněk Hlávka, March 2005.
- 022 "DSFM fitting of Implied Volatility Surfaces" by Szymon Borak, Matthias R. Fengler and Wolfgang Härdle, March 2005.

**SFB 649, Spandauer Straße 1, D-10178 Berlin**  
**<http://sfb649.wiwi.hu-berlin.de>**

This research was supported by the Deutsche  
Forschungsgemeinschaft through the SFB 649 "Economic Risk".



- 023 "Towards a Monthly Business Cycle Chronology for the Euro Area" by Emanuel Mönch and Harald Uhlig, April 2005.
- 024 "Modeling the FIBOR/EURIBOR Swap Term Structure: An Empirical Approach" by Oliver Blaskowitz, Helmut Herwartz and Gonzalo de Cadenas Santiago, April 2005.
- 025 "Duality Theory for Optimal Investments under Model Uncertainty" by Alexander Schied and Ching-Tang Wu, April 2005.
- 026 "Projection Pursuit For Exploratory Supervised Classification" by Eun-Kyung Lee, Dianne Cook, Sigbert Klinke and Thomas Lumley, May 2005.
- 027 "Money Demand and Macroeconomic Stability Revisited" by Andreas Schabert and Christian Stoltenberg, May 2005.
- 028 "A Market Basket Analysis Conducted with a Multivariate Logit Model" by Yasemin Boztuğ and Lutz Hildebrandt, May 2005.
- 029 "Utility Duality under Additional Information: Conditional Measures versus Filtration Enlargements" by Stefan Ankirchner, May 2005.
- 030 "The Shannon Information of Filtrations and the Additional Logarithmic Utility of Insiders" by Stefan Ankirchner, Steffen Dereich and Peter Imkeller, May 2005.
- 031 "Does Temporary Agency Work Provide a Stepping Stone to Regular Employment?" by Michael Kvasnicka, May 2005.
- 032 "Working Time as an Investment? – The Effects of Unpaid Overtime on Wages, Promotions and Layoffs" by Silke Anger, June 2005.
- 033 "Notes on an Endogenous Growth Model with two Capital Stocks II: The Stochastic Case" by Dirk Bethmann, June 2005.
- 034 "Skill Mismatch in Equilibrium Unemployment" by Ronald Bachmann, June 2005.
- 035 "Uncovered Interest Rate Parity and the Expectations Hypothesis of the Term Structure: Empirical Results for the U.S. and Europe" by Ralf Brüggemann and Helmut Lütkepohl, April 2005.
- 036 "Getting Used to Risks: Reference Dependence and Risk Inclusion" by Astrid Matthey, May 2005.
- 037 "New Evidence on the Puzzles. Results from Agnostic Identification on Monetary Policy and Exchange Rates." by Almuth Scholl and Harald Uhlig, July 2005.
- 038 "Discretisation of Stochastic Control Problems for Continuous Time Dynamics with Delay" by Markus Fischer and Markus Reiss, August 2005.
- 039 "What are the Effects of Fiscal Policy Shocks?" by Andrew Mountford and Harald Uhlig, July 2005.
- 040 "Optimal Sticky Prices under Rational Inattention" by Bartosz Maćkowiak and Mirko Wiederholt, July 2005.
- 041 "Fixed-Prize Tournaments versus First-Price Auctions in Innovation Contests" by Anja Schöttner, August 2005.
- 042 "Bank finance versus bond finance: what explains the differences between US and Europe?" by Fiorella De Fiore and Harald Uhlig, August 2005.
- 043 "On Local Times of Ranked Continuous Semimartingales; Application to Portfolio Generating Functions" by Raouf Ghomrasni, June 2005.
- 044 "A Software Framework for Data Based Analysis" by Markus Krätzig, August 2005.
- 045 "Labour Market Dynamics in Germany: Hirings, Separations, and Job-to-Job Transitions over the Business Cycle" by Ronald Bachmann, September 2005.

**SFB 649, Spandauer Straße 1, D-10178 Berlin**  
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche  
 Forschungsgemeinschaft through the SFB 649 "Economic Risk".



- 046 "Paternal Uncertainty and the Economics of Mating, Marriage, and Parental Investment in Children" by Dirk Bethmann and Michael Kvasnicka, September 2005.
- 047 "Estimation and Testing for Varying Coefficients in Additive Models with Marginal Integration " by Lijian Yang, Byeong U. Park, Lan Xue and Wolfgang Härdle, September 2005.

**SFB 649, Spandauer Straße 1, D-10178 Berlin**  
**<http://sfb649.wiwi.hu-berlin.de>**

This research was supported by the Deutsche  
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

