Statistica Sinica **12**(2002), 801-821

DIRECT ESTIMATION IN AN ADDITIVE MODEL WHEN THE COMPONENTS ARE PROPORTIONAL

Lijian Yang

Michigan State University

Abstract: Component functions of an additive model can be estimated at univariate rate of convergence, by such methods as backfitting, marginal integration, etc. An alternative direct method is developed when the components are proportional. This new direct local polynomial estimator requires as little computing as a univariate estimator, less than the integration method by a factor of the sample size. Combination with one-step backfitting yields an improved estimator with univariate rate of convergence and "oracle" efficiency, and retains comparable computational efficiency. Monte-Carlo results indicate good performance of both estimators, which work much better than the integration method. The direct method is applied to a GARCH type model, illustrated by an analysis of the daily returns of Deutsche Mark/British Pound (DEM/GBP).

Key words and phrases: Coefficient parameter, DEM/GBP daily returns, efficient estimator, equivalent kernel, local polynomial, nonparametric GARCH, rule-of-thumb bandwidth.

1. Introduction

The volatility of foreign exchange daily returns depends on many past variables. This persistence is commonly modeled by exponentially decaying dependence, for example the GARCH(p, q) model in Bollerslev (1986). A GARCH$(1, 1)$ with mean zero describes a time series $\{Y_t\}_{t=0}^{\infty}$ as

$$Y_t = \sigma_t \xi_t, \quad \sigma_t^2 = w + \beta Y_{t-1}^2 + \gamma \sigma_{t-1}^2, \quad w \geq 0, \quad \beta, \gamma > 0, \quad \beta + \gamma < 1, \tag{1.1}$$

where the ξ_t's are i.i.d. with $E\xi_t = E\xi_t^3 = 0$, $E\xi_t^2 = 1$, $E\xi_t^4 = m_4 \in (0, \infty)$. Equivalently, the conditional volatility $\sigma_t^2 = \mathrm{Var}\,(Y_t | Y_{t-1}, Y_{t-2}, \ldots)$ is $\sigma_t^2 = g(Y_{t-1}) + \gamma g(Y_{t-2}) + \gamma^2 g(Y_{t-3}) + \cdots$, $g(y) = w + \beta y^2$.

The quadratic form of $g(y)$, however, is inadequate for certain data. To remove this restriction, Hafner (1998) proposed an alternative nonparametric model

$$\sigma_t^2 = g(Y_{t-1}) + \gamma g(Y_{t-2}) + \gamma^2 g(Y_{t-3}) + \cdots, \tag{1.2}$$

where g is any smooth function. Model (1.2) is a compromise between nonparametric flexibility and GARCH simplicity. An iterative procedure was developed in Hafner (1998) to estimate γ and g. Theoretical properties of these estimators are unavailable, due to an infinite number of explanatory variables. A manageable finite model is

$$\sigma_t^2 = g(X_1) + \gamma g(X_2) + \cdots + \gamma^{d-1} g(X_d) \tag{1.3}$$

for some sufficiently large integer d, where one denotes $X_i = Y_{t-i}$, $i = 1, \ldots, d$. Note that for $Z_t = Y_t^2$, $Z_t = m(\mathbf{X}_t) + \sigma(\mathbf{X}_t)\epsilon_t$, $m(\mathbf{X}_t) = \sum_{\alpha=1}^{d} \gamma^{\alpha-1} g(X_\alpha)$, $\sigma(\mathbf{X}_t) = m(\mathbf{X}_t)\sqrt{m_4 - 1}$, where $\epsilon_t = \left(\xi_t^2 - 1\right)/\sqrt{m_4 - 1}$ satisfies $E\epsilon_t = 0$, $\mathrm{Var}\ (\epsilon_t) = 1$.

The model (1.3) is a special case of the proportional additive model (PAM)

$$Y = m(\mathbf{X}) + \sigma(\mathbf{X})\epsilon, \tag{1.4}$$

where Y is a scalar dependent variable, $\mathbf{X} = (X_1, X_2, \ldots, X_d)$ a vector of explanatory variables (here one no longer has $X_i = Y_{t-i}$), ϵ is independent of $\mathbf{X}$ with $E\epsilon = 0, Var(\epsilon) = 1$, and the regression function is

$$m(\mathbf{X}) = c_1(\boldsymbol{\gamma}) g(X_1) + c_2(\boldsymbol{\gamma}) g(X_2) + \cdots + c_d(\boldsymbol{\gamma}) g(X_d). \tag{1.5}$$

Here $c_1, \ldots, c_d$ are known coefficients functions, and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_r)$ is a vector of parameters. For identifiability, we set c_1 equal to 1 throughout.

Model (1.5) is an additive model. In a regression model of the general form (1.4), without the restriction (1.5), $m(\cdot)$ can be estimated at the rate $O\left(n^{-(p+1)/(2p+d+2)}\right)$, where n is the sample size and $m(\cdot)$ has $p+1$ Lipschitz continuous derivatives. This rate is improved to the univariate $O\left(n^{-(p+1)/(2p+3)}\right)$ if the function $m(\cdot)$ is additive:

$$m(\mathbf{X}) = c + \sum_{\beta=1}^{d} g_\beta(X_\beta). \tag{1.6}$$

Properties of (1.6) were studied by Stone (1985). A backfitting algorithm was proposed by Hastie and Tibshirani (1990) for estimating the functions $\{g_\beta(\cdot)\}_{\beta=1}^{d}$ and $m(\cdot)$, and theoretical properties were established by Opsomer and Ruppert (1997) in the case $d = 2$. Marginal integration was proposed by Tjøstheim and Auestad (1994), Linton and Nielsen (1995) Masry and Tjøstheim (1996) and Yang, Härdle and Nielsen (1999), and has univariate rate of convergence. Carroll, Härdle and Mammen (1999) proposed to estimate $g(\cdot)$ and γ in (1.3) via integration or backfitting. In high dimensions ($d > 2$), however, the integration method behaves poorly when the data is sparse, while the backfitting method has no asymptotic distribution theory.

I present a direct local polynomial estimator of the function $g(\cdot)$ in (1.5) that takes advantage of the proportional structure. An improved estimator based on direct estimation and one-step backfitting achieves the optimal convergence rate. If the parameters $\gamma_1, \ldots, \gamma_r$ are unknown, they are estimated by minimizing the prediction error of the direct estimator. The proposed methods require less computing than the integration method by a factor $O(n)$, and perform better in practice, as seen in the Monte-Carlo study.

In Section 2, I describe the technical setting and propose direct and rate optimal estimators when the parameter vector $\boldsymbol{\gamma}$ is known. Section 3 proposes a $\sqrt{n}$-convergent estimator of $\boldsymbol{\gamma}$. Implementation is detailed in Section 4. Section 5 presents simulation examples of PAM and the nonparametric GARCH model, and the GARCH fit to the daily returns of Deutsche Mark/British Pound exchange rates. Proofs are in the Appendix.

2. Estimation When Parameters are Known

Suppose for now that the parameters $\gamma_1, \ldots, \gamma_r$ are known. Let $(\mathbf{X}_i, Y_i)$, $i = 1, 2, \ldots, n$ be a sample following (1.4) with $m(\cdot)$ as in (1.5). Recall that c_1 is 1. The observations are assumed to be i.i.d., but the method also works if the $\mathbf{X}_i$'s are geometrically β-mixing and strictly stationary.

For any $x \in A$, with set A defined in Assumption A4, one has a special Taylor expansion

$$m(\mathbf{z}) \approx g(x) \sum_{\alpha=1}^{d} c_\alpha(\boldsymbol{\gamma}) + g'(x) \sum_{\alpha=1}^{d} c_\alpha(\boldsymbol{\gamma})(z_\alpha - x) + \cdots + \frac{g^{(p)}(x)}{p!} \sum_{\alpha=1}^{d} c_\alpha(\boldsymbol{\gamma})(z_\alpha - x)^p, \quad (2.1)$$

which suggests that one estimate quantities $g(x), g'(x), \ldots, g^{(p)}(x)$ by regressing Y_i's on terms $\sum_{\alpha=1}^{d} c_\alpha(\boldsymbol{\gamma})(X_{i\alpha} - x)^\lambda, \lambda = 0, 1, \ldots,$ with d-dimensional kernel weights.

To implement the idea, define derivative and function estimators

$$\widehat{g^{(\lambda)}}(x) = \lambda! h^{-\lambda} E_\lambda' \left(Z_\gamma' W Z_\gamma\right)^{-1} Z_\gamma' W \mathbf{Y}, \quad \widehat{g}(x) = E_0' \left(Z_\gamma' W Z_\gamma\right)^{-1} Z_\gamma' W \mathbf{Y}, \quad (2.2)$$

where

$$Z_\gamma = \left\{ \sum_{\alpha=1}^{d} c_\alpha(\boldsymbol{\gamma}) \left(\frac{X_{i\alpha} - x}{h} \right)^\lambda \right\}_{1 \le i \le n, 0 \le \lambda \le p}, \qquad W = \operatorname{diag} \left\{ \frac{1}{n} K_h(\mathbf{X}_i - \mathbf{x}) \right\}_{i=1}^{n},$$

$\mathbf{Y} = (Y_i)_{n \times 1}$, E_λ is a $(p+1)-$ vector with the $(\lambda+1)-$ element being 1 and 0 elsewhere, $p > 0$ is an integer, $\mathbf{x} = (x, \ldots, x)$, K is a kernel function, $h > 0$ the bandwidth. In the following, I denote $K_h(u) = K(u/h)/h$ for any function K, and $K_{\lambda,\gamma}^*(\mathbf{u}) = \sum_{\alpha=1}^{d} \sum_{\lambda'=0}^{p} c_\alpha(\boldsymbol{\gamma}) s_{\lambda\lambda'}(\boldsymbol{\gamma}) u_\alpha^{\lambda'} K(\mathbf{u})$ with $\{s_{st}(\boldsymbol{\gamma})\}_{0 \le s,t \le p} = S_\gamma^{-1}$, where S_γ is defined in (A.3).

Theorem 1. *Under assumptions* A1-A4 *and* A6, *for any fixed* $x \in A$ *and* $\lambda \geq 0$ *such that* $p-\lambda$ *is odd, as* $nh^{2\lambda+d} \to \infty, h \to 0$, *the estimator* $\widehat{g^{(\lambda)}}(x)$ *defined by* (2.2) *satisfies* $\sqrt{nh^{2\lambda+d}}\left\{\widehat{g^{(\lambda)}}(x) - g^{(\lambda)}(x) - h^{p+1-\lambda}b_\lambda(x)\right\} \xrightarrow{D} N\{0, v_\lambda(x)\}$, *where*

$$b_\lambda(x) = \lambda!\Lambda_{\lambda,p+1,\gamma}g^{(p+1)}(x)/(p+1)!,\ v_\lambda(x) = (\lambda!)^2 \left\|K^*_{\lambda,\gamma}\right\|_2^2 \sigma^2(\mathbf{x})\varphi^{-1}(\mathbf{x}), \quad (2.3)$$

$\varphi(\cdot)$ *is the design density of* $\mathbf{X}$, *and* $\Lambda_{\lambda,p+1,\gamma}$ *is defined in* (A.4). *In particular, for odd* p, *and* $nh^d \to \infty, h \to 0, \sqrt{nh^d}\{\widehat{g}(x) - g(x) - h^{p+1}b_0(x)\} \xrightarrow{D} N\{0, v_0(x)\}$.

Performance of the proposed function estimator is measured by $\sum_{\alpha=1}^d E\int_A \{\widehat{g}(x) - g(x)\}^2 \varphi_\alpha(x)dx$, where $\varphi_\alpha(\cdot)$ is the marginal density of X_α.

Corollary 1. *The global optimal bandwidth for estimating the function* $g(\cdot)$ *is*

$$h_{opt} = \left[\frac{d\{(p+1)!\}^2 \left\|K^*_{0,\gamma}\right\|_2^2 \sum_{\alpha=1}^d \int_A \sigma^2(\mathbf{x})\varphi^{-1}(\mathbf{x})\varphi_\alpha(x)dx}{2n(p+1)(\Lambda_{0,p+1,\gamma})^2 \sum_{\alpha=1}^d \int_A \{g^{(p+1)}(x)\}^2 \varphi_\alpha(x)dx}\right]^{1/(2p+d+2)}. \quad (2.4)$$

which yields mean integrated squared error of $\widehat{g}(x)$ *of order* $n^{-2(p+1)/(2p+2+d)}$.

The suboptimal rate of convergence in Theorem 1 can be improved by a one-step backfitting procedure. Let

$$Y_{i1} = Y_i - \{c_2(\boldsymbol{\gamma})g(X_{i2}) + \cdots + c_d(\boldsymbol{\gamma})g(X_{id})\}, \quad (2.5)$$

and define the "oracle" smoother

$$g_1(x) = E_0'(Z_1'W_1Z_1)^{-1} Z_1'W_1\mathbf{Y}_1, \quad (2.6)$$

with

$$Z_1 = \left\{\left(\frac{X_{i1}-x}{h_1}\right)^\lambda\right\}_{1\leq i\leq n, 0\leq\lambda\leq p},\ W_1 = \mathrm{diag}\left\{\frac{1}{n}K_{h_1}(X_{i1}-x)\right\}_{i=1}^n,\ \mathbf{Y}_1 = (Y_{i1})_{n\times 1},$$

and $h_1 = Cn^{-1/(2p+3)}, C > 0$. Define next the sample analogs of (2.5) and (2.6):

$$\widetilde{Y}_{i1} = Y_i - \{c_2(\boldsymbol{\gamma})\widehat{g}(X_{i2}) + \cdots + c_d(\boldsymbol{\gamma})\widehat{g}(X_{id})\}, \quad (2.7)$$

$$\widehat{g}_1(x) = E_0'(Z_1'W_1Z_1)^{-1} Z_1'W_1\widetilde{\mathbf{Y}}_1, \quad \widetilde{\mathbf{Y}}_1 = (\widetilde{Y}_{i1})_{n\times 1}, \quad (2.8)$$

in which the estimator $\widehat{g}(\cdot)$ is as defined in (2.2).

Theorem 2. *Under assumptions* A1-A4 *and* A6, *for any fixed* $x \in A$ *and odd* p, *as* $nh^d \to \infty, h/h_1 \to 0, h_1 = Cn^{-1/(2p+3)}, C > 0$, *the estimator* $\widehat{g}_1(x)$ *defined by* (2.8) *satisfies*

$$\sqrt{nh_1}\left\{\widehat{g}_1(x) - g(x) - h_1^{p+1}b_1^*(x)\right\} \xrightarrow{D} N\{0, v_1^*(x)\}, \quad (2.9)$$

$$b_1^*(x) = \Lambda_{0,p+1} g^{(p+1)}(x)/(p+1)!,$$

$$v_1^*(x) = \|K_0^*\|_2^2 \varphi_1^{-2}(x) \int \sigma^2(x, u_2, \ldots, u_d)\varphi(x, u_2, \ldots, u_d) du_2 \cdots du_d, \quad (2.10)$$

and $\Lambda_{0,p+1}$, K_0^* *are univariate constants* $\Lambda_{0,p+1,\gamma}$, $K_{0,\gamma}^*$. *In particular, if* $\sigma(\mathbf{x}) \equiv \sigma_0 > 0$, $v_1^*(x) = \|K_0^*\|_2^2 \sigma_0^2/\varphi_1(x)$, $\widehat{g}_1(x)$ *has the same asymptotic distribution as the local polynomial estimator for the univariate model* $Y_{i1} = g(X_{i1}) + \sigma_0\varepsilon_i$.

While the "oracle" estimator $\widehat{g}_1(x)$ enjoys the optimal convergence rate, its computing time is roughly d times that for $\widehat{g}(x)$, but much less than the integration estimator. The proof of Theorem 2 relies on the facts that the bias caused by substituting the $\widehat{g}(X_{i\alpha})$ for $g(X_{i\alpha})$ is negligible (of order $O\left(h^{p+1}\right) = o\left(h_1^{p+1}\right)$), and that the sum of the noises of these substitutions is of order $O\left(\sqrt{n^{-1}h_1^{-1}n^{-1}h^{-d} + n^{-1}}\right) = o\left(\sqrt{n^{-1}h_1^{-1}}\right)$.

3. Estimating the Parameters

If the parameter $\boldsymbol{\gamma}$ is unknown, I present a procedure that estimates $\boldsymbol{\gamma}$ at the usual $\sqrt{n}$-rate. Assume that $\boldsymbol{\gamma} \in$ the interior of $\boldsymbol{\Gamma}$, a compact subset of R^r.

For each value $\boldsymbol{\gamma}' \in \boldsymbol{\Gamma}$, define for any $x \in A$ the estimator

$$\widehat{g}_{\gamma'}(x) = E_0'\left(Z_{\gamma'}'WZ_{\gamma'}\right)^{-1} Z_{\gamma'}'W\mathbf{Y}, \ Z_{\gamma'} = \left\{\sum_{\alpha=1}^{d} c_\alpha(\boldsymbol{\gamma}')\left(\frac{X_{i\alpha}-x}{h}\right)^\lambda\right\}_{1\le i\le n, 0\le\lambda\le p}, \quad (3.1)$$

and define the function

$$L(\boldsymbol{\gamma}') = \frac{1}{n}\sum_{i=1}^{n}\left\{Y_i - \widehat{g}_{\gamma'}(X_{i1}) - c_2(\boldsymbol{\gamma}')\widehat{g}_{\gamma'}(X_{i2}) - \cdots - c_d(\gamma')\widehat{g}_{\gamma'}(X_{id})\right\}^2 \pi(\mathbf{X}_i), \quad (3.2)$$

where $\pi(\cdot)$ is a nonnegative and continuous weight function whose compact support is contained in A. Let $\widehat{\boldsymbol{\gamma}}$ be the minimizer of the function $L(\boldsymbol{\gamma}')$, i.e.,

$$\widehat{\boldsymbol{\gamma}} = \arg\min_{\gamma' \in \boldsymbol{\Gamma}} L(\boldsymbol{\gamma}'). \quad (3.3)$$

Theorem 3. *Under assumptions* A1-A6, *if* $h^{p+1} + n^{-1}h^{-d} = o(1/\sqrt{n})$, *then as* $n \to \infty$, *the* $\widehat{\boldsymbol{\gamma}}$ *defined by* (3.3) *satisfies*

$$\sqrt{n}\left(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\right) \to N(\mathbf{0}, \Sigma_\gamma) \quad (3.4)$$

for some positive definite matrix Σ_γ.

4. Implementation

The estimators defined in the previous sections are implemented with a simple rule-of-thumb (ROT) for the bandwidth h_{opt} given by (2.4). If the $\sqrt{n}$-convergent $\widehat{\boldsymbol{\gamma}}$ has been obtained or if $\boldsymbol{\gamma}$ is known, the ROT bandwidth is

$$\widehat{h}_{opt}=\left[\frac{d\left\{(p+1)!\right\}^2\left\|K^*_{0,\widehat{\boldsymbol{\gamma}}}\right\|_2^2\widetilde{\sigma}^2\sum_{\alpha=1}^{d}\sum_{i=1}^{n}\widetilde{\varphi}^{-1}(\mathbf{X}_{i\alpha})1_A(X_{i\alpha})}{2n(p+1)\left(\Lambda_{0,p+1,\widehat{\boldsymbol{\gamma}}}\right)^2\sum_{\alpha=1}^{d}\sum_{i=1}^{n}\left\{\widetilde{g}^{(p+1)}(X_{i\alpha})\right\}^2 1_A(X_{i\alpha})}\right]^{1/(2p+d+2)}, \quad (4.1)$$

where $\widetilde{\varphi}^{-1}(\mathbf{X}_{i\alpha})$ are kernel density estimates at points $\mathbf{X}_{i\alpha}=(X_{i\alpha},\ldots,X_{i\alpha})$, $i=1,\ldots,n$, $\alpha=1,\ldots,d$, using Silverman's kernel density estimation ROT bandwidth (Silverman (1986), p.86-87). The $\widetilde{g}^{(p+1)}$ and $\widetilde{\sigma}^2$ are obtained as in Fan and Gijbels (1996, equation (4.3), p.111) by an ordinary $p+2$ degree polynomial regression of the Y_i's on the $\mathbf{X}_i$'s. Constants $\Lambda_{0,p+1,\widehat{\boldsymbol{\gamma}}}$ and $\left\|K^*_{0,\widehat{\boldsymbol{\gamma}}}\right\|_2^2$ are calculated exactly.

If one needs to estimate $\boldsymbol{\gamma}$ first, one sets $d'=\min(d,3)$ and denotes by p' an odd positive integer such that $p-d/2>p'>d/2-1$. For any given $\boldsymbol{\gamma}'\in\boldsymbol{\Gamma}$, define

$$\widehat{h}_{\gamma'}=\left[\frac{\left\{(p'+1)!\right\}^2\left\|K^*_{0,\gamma'}\right\|_2^2\widetilde{\sigma}^2\sum_{\alpha=1}^{d'}\sum_{i=1}^{n}\widetilde{\varphi}^{-1}(\mathbf{X}_{i\alpha})1_A(X_{i\alpha})}{2(p'+1)\left(\Lambda_{0,p'+1,\gamma'}\right)^2\sum_{\alpha=1}^{d'}\sum_{i=1}^{n}\left\{\widetilde{g}^{(p'+1)}(X_{i\alpha})\right\}^2 1_A(X_{i\alpha})}\right]^{1/(2p'+d+2)}, \quad (4.2)$$

where $\widetilde{g}^{(p'+1)}$, $\widetilde{\varphi}^{-1}$and $\widetilde{\sigma}^2$ are estimated as in (4.1). This bandwidth $\widehat{h}_{\gamma'}$ satisfies the order requirement $h^{p+1}+n^{-1}h^{-d}=o(1/\sqrt{n})$ of Theorem 3 and is suitable for the $\sqrt{n}$ consistent estimation of $\boldsymbol{\gamma}$.

It is feasible to compute a plug-in (PI) bandwidth selector for h_{opt} similar to Yang and Tschernig (1999). The PI bandwidth is expected to improve on the ROT bandwidth, but is also more computationally intensive.

5. Monte-Carlo and Examples

I begin by applying the direct and oracle estimators to simulated examples of data $(\mathbf{X}_i,Y_i)$, $i=1,2,\ldots,n$, drawn from model (1.4), where the regression function $m(\cdot)$ is given by (1.5). Motivated by (1.3), the setup is: $\boldsymbol{\Gamma}=[0.5,0.9]$, $c_\alpha(\boldsymbol{\gamma})=\boldsymbol{\gamma}^{\alpha-1}$, $\alpha=1,\ldots,d$. Following common practice, I set $\sigma^2(\cdot)\equiv\sigma^2$. The design variable $\mathbf{X}\sim U\left(A^d\right)$, where the estimation set $A=[-0.5,0.5]$ and the weight function $\pi=1_{A^d}$. The component function $g(x)$ is defined on A and $\boldsymbol{\gamma}$ =0.7. Each experiment is carried out 100 times. Although many candidates of $g(x)$ are used in the experiments, I report only on two.

Example 1. The function $g(x) = 2 + \sin(5\pi x)$.

Example 2. The function $g(x) = 2 - 48(x + 0.5) + 218(x + 0.5)^2 - 315(x + 0.5)^3 + 145(x + 0.5)^4$.

The results of experiments are summarized in Figures 1, 2, and 3. For each example, the 100 estimates (dashed lines) are overlaid with the true function (thick solid line).

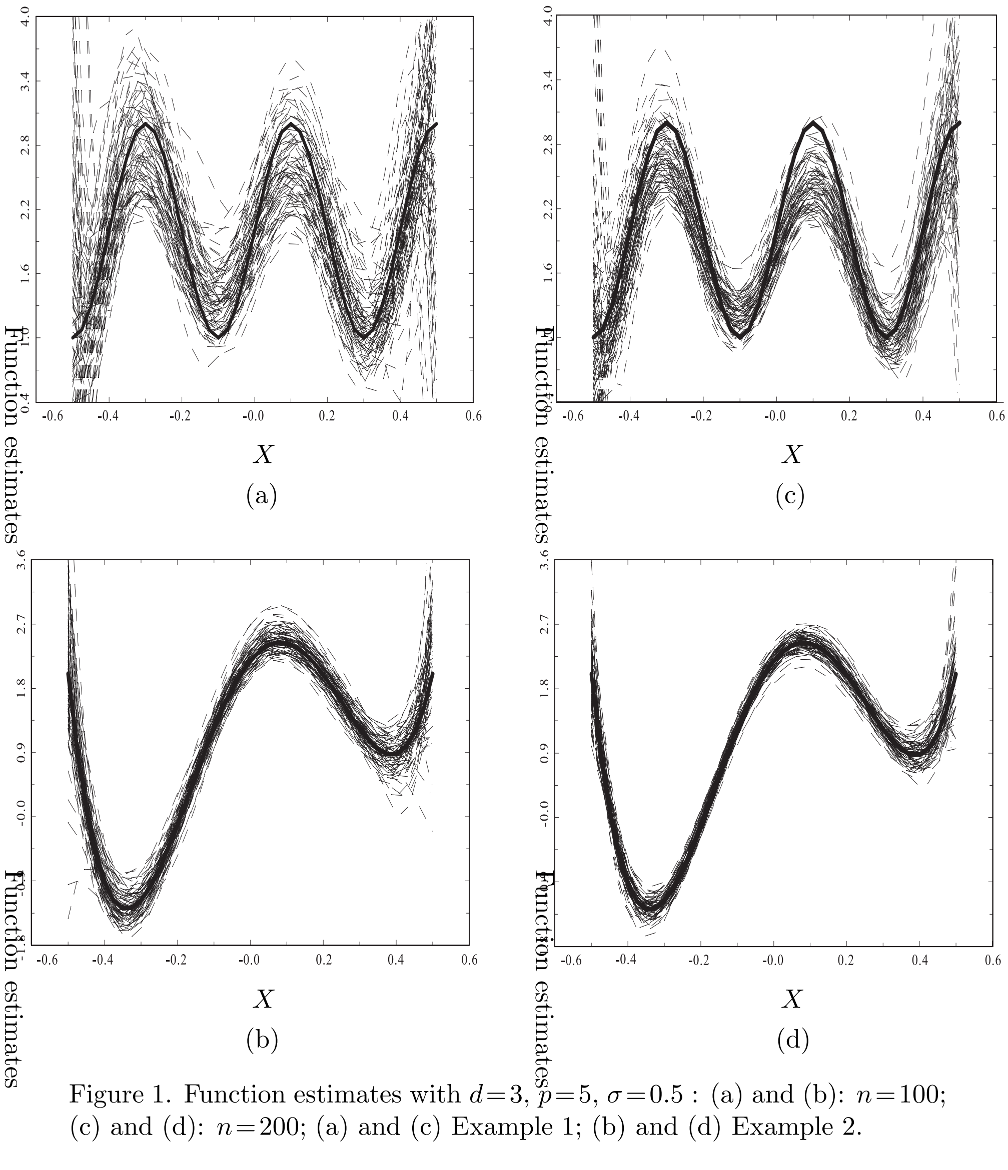

Figure 1. Function estimates with $d=3$, $p=5$, $\sigma=0.5$: (a) and (b): $n=100$; (c) and (d): $n=200$; (a) and (c) Example 1; (b) and (d) Example 2.

Figures 1 shows that for Examples 1 and 2, the function estimates hit their targets well, and when the sample size n is increased from 100 to 200, there is noticeable improvement. To verify the efficiency property (2.9) of the oracle estimator $\widehat{g}_1$ in (2.8) against the univariate estimator g_1, Figure 2 summarizes the results of comparing $\widehat{g}_1$ and g_1 for Example 1. Contrasting plots (a) and (b), and plots (c) and (d) indicates that the oracle estimator $\widehat{g}_1$ is nearly as accurate as g_1. The empirical efficiency of $\widehat{g}_1$ against g_1 is 74.2% for $n = 100$ and 90% for $n = 200$, consistent both with the plots and the conclusion of Theorem 2.

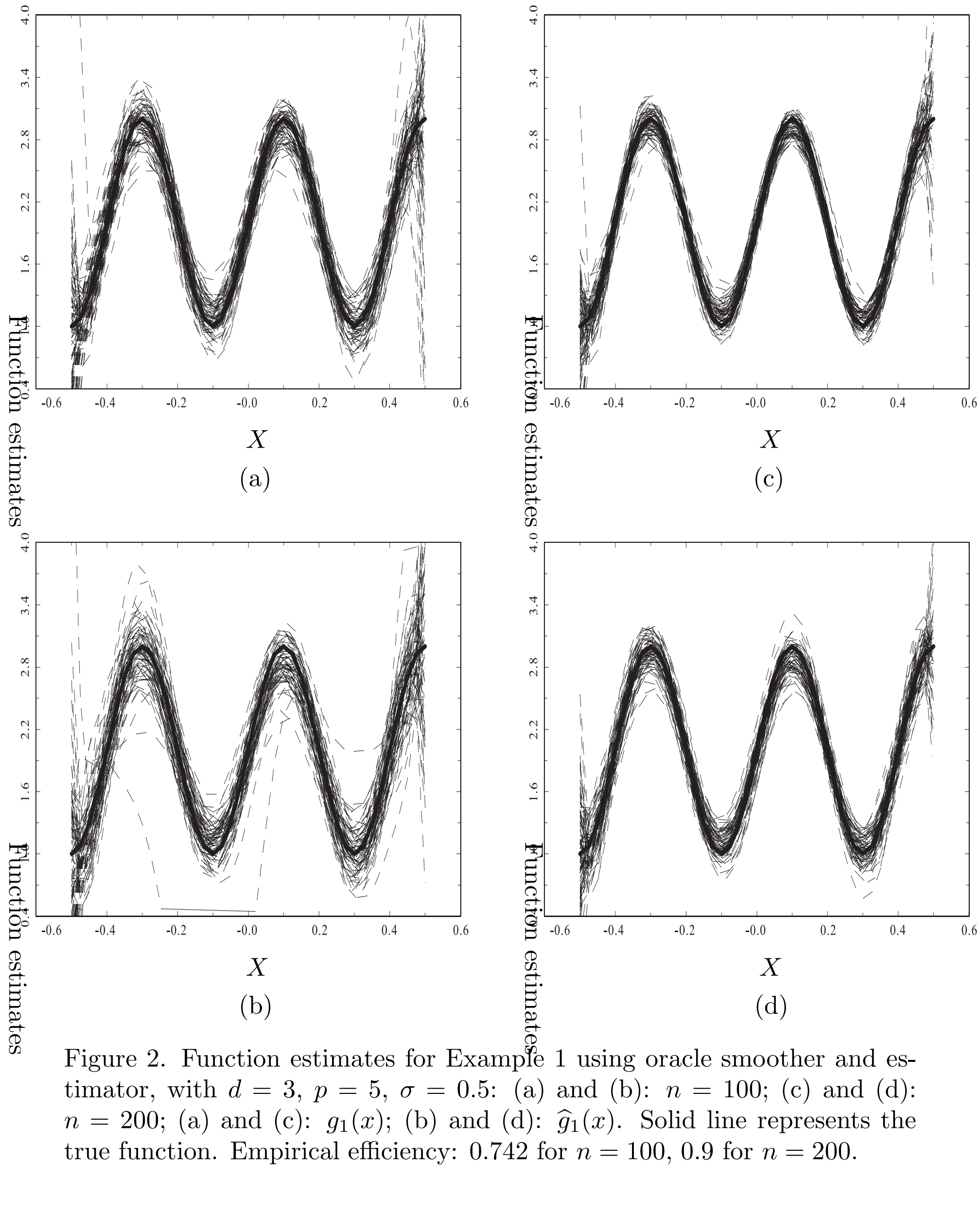

Figure 2. Function estimates for Example 1 using oracle smoother and estimator, with $d = 3$, $p = 5$, $\sigma = 0.5$: (a) and (b): $n = 100$; (c) and (d): $n = 200$; (a) and (c): $g_1(x)$; (b) and (d): $\widehat{g}_1(x)$. Solid line represents the true function. Empirical efficiency: 0.742 for $n = 100$, 0.9 for $n = 200$.

The integration method is applied to Example 1 with optimal bandwidths. The 100 direct and integration estimates are plotted in (a) and (c) of Figure 3, respectively. One observes that the integration method has much larger errors than the direct method estimators. This confirms the original motivation of the direct method and, although the integration method has a faster rate of convergence, this advantage does not kick in for moderate sample sizes.

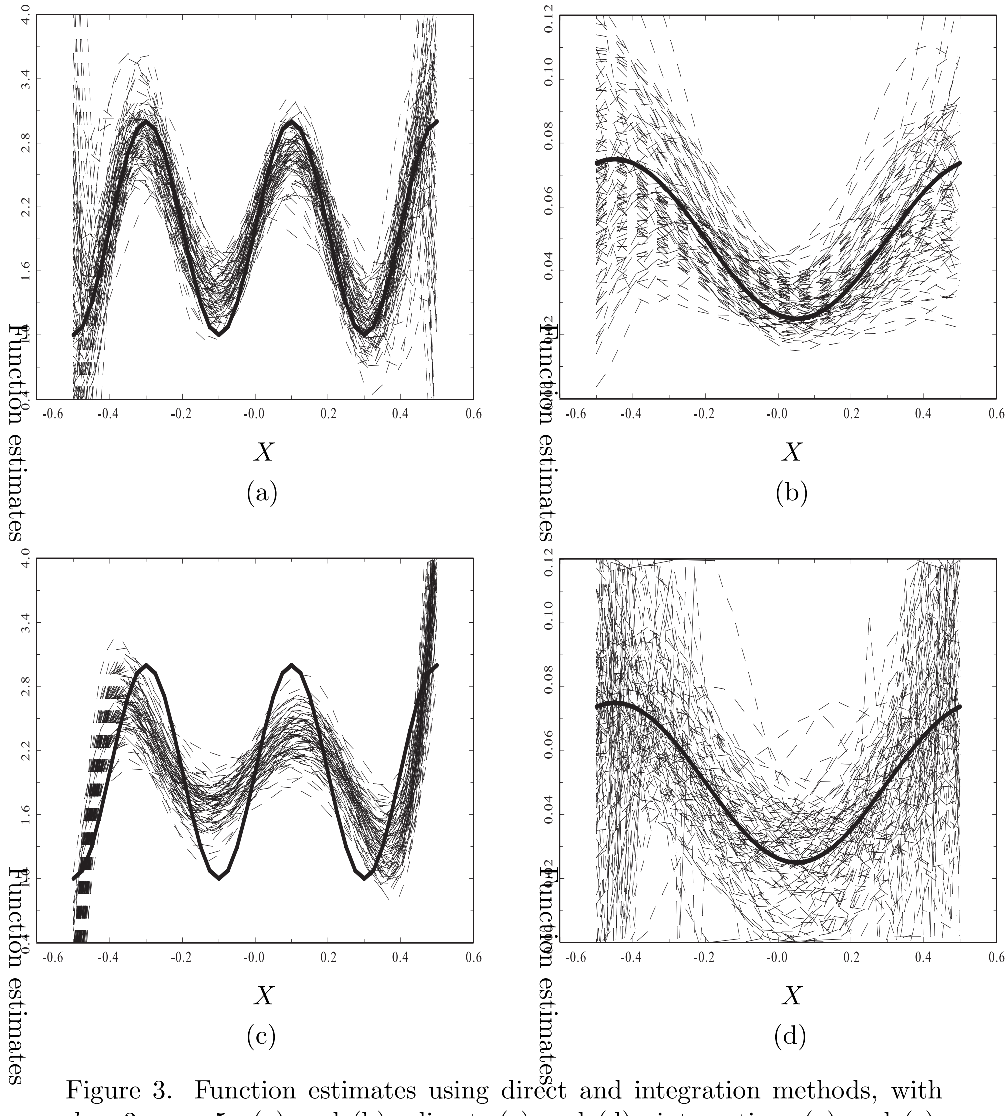

Figure 3. Function estimates using direct and integration methods, with $d = 3$, $p = 5$: (a) and (b): direct; (c) and (d): integration; (a) and (c): Example 1, $n = 100$, $\sigma = 0.5$, $\gamma = 0.7$; (b) and (d): Example 3, $n = 200$, $\gamma = 0.8$, MISE = 0.010303 for (d) and MISE = 0.058432 for (d). Solid line represents the true function.

A much more extensive study was done, but I have included only the most important results due to space restriction. In addition, the study found that having more variates (i.e., larger d) has little effect on function estimation precision, while the effect of noise σ^2 on function estimation is very significant. Similar phenomena hold for the parameter estimation. The function g can be estimated considerably more accurately when the parameter is known than when it is unknown and has to be estimated. Details of the full study can be downloaded from http://www.msu.edu/yangli/pam.pdf.

I now discuss model (1.3), the motivation for studying PAM. Under geometric ergodicity assumptions on $\{Y_t\}_{t=0}^{\infty}$, one can apply the estimation method of Section 4 to the data $\{(\mathbf{X}_t, Z_t)\}_{t=d}^{n}$. General methods of estimating volatility are found in Härdle and Tsybakov (1997), Härdle, Tsybakov and Yang (1998) and Fan and Yao (1998).

Example 3. For simulation, process $\{Y_t\}_{t=0}^{\infty}$ is generated according to (1.3) with $g(x) \equiv \{2 - \sin(2\pi x + 0.4\pi)\}/40$, $\gamma = 0.8$, $d = 3$. It is easy to verify geometric ergodicity of $\{Y_t\}_{t=0}^{\infty}$, see Doukhan (1994, Theorem 3, p.91). I use a total of 100 replications. For each replication, $n + 1000$ observations $\{Y_t\}_{t=-1000}^{n-1}$ are generated, with the last n taken as the data $\{Y_t\}_{t=0}^{n-1}$ to ensure stationarity. The function $g(x) \approx \left\{1 + 2\pi^2 (x - 0.05)^2\right\}/40$ around $x = 0.05$, resembling the asymmetric GARCH model of Engle and Ng (1993). I estimate $g(x)$ on $[-0.5, 0.5]$, where over 90% of the data lie. For the 100 replications of size $n = 200$, the integration method took 90 hours to finish in Gauss Windows 95, while the direct method took 2.5 hours. For larger n, the difference would grow at the rate of $O(n)$. Figure 3 (b) and (d) show estimates of $g(x)$ and the superior performance of the direct method over integration.

Example 4. A data example consists of the daily returns of Deutsche Mark/British Pound (DEM/GBP) from January 2, 1980 to October 30, 1992, with $n = 3212$. The data is trimmed at the 1.25 and 97.5 percentiles to reduce outlier influence. The first 1606 returns are shown in Figure 4(a). I fit Model (1.3) to these data with $d = 5$. The data is divided into 2 subsamples of $n = 1606$ observations each. The first subsample $\{Y_t\}_{t=0}^{1605}$ is used to construct the estimated function $\widehat{g}(\cdot)$ and parameter $\widehat{\gamma}$, and then the estimated volatility series and residuals $\widehat{\sigma}_t^2 = \sum_{\nu=1}^{d} \widehat{\gamma}^{\nu-1} \widehat{g}(Y_{t-\nu})$, $\widehat{\xi}_t = Y_t/\widehat{\sigma}_t = (\sigma_t/\widehat{\sigma}_t)\xi_t$, $t = d, \ldots, 3211$. The parameter estimate is $\widehat{\gamma} = 0.98$, while the estimated function $\widehat{g}(\cdot)$ is shown in Figure 4(c). The visual difference between $\widehat{g}(\cdot)$ (solid line) and its least squares quadratic fit (dotted line) strongly suggest that the nonparametric GARCH model (1.3) deviates from the asymmetric GARCH model of Engle and Ng (1993). The autocorrelation functions (ACFs) of $\{Y_t^2\}_{t=0}^{3211}$ and $\left\{\widehat{\xi}_t^2\right\}_{t=d}^{3211}$ are shown in Figure 4 (b) and (d) for both subsamples. The thick horizontal lines in (b) and (d) represent

0 and the 95% confidence bounds around 0. The ACFs of the squared returns (solid lines) die out slowly, showing strong positive correlation. The squared residuals show almost no correlation for the in-sample prediction, and very little

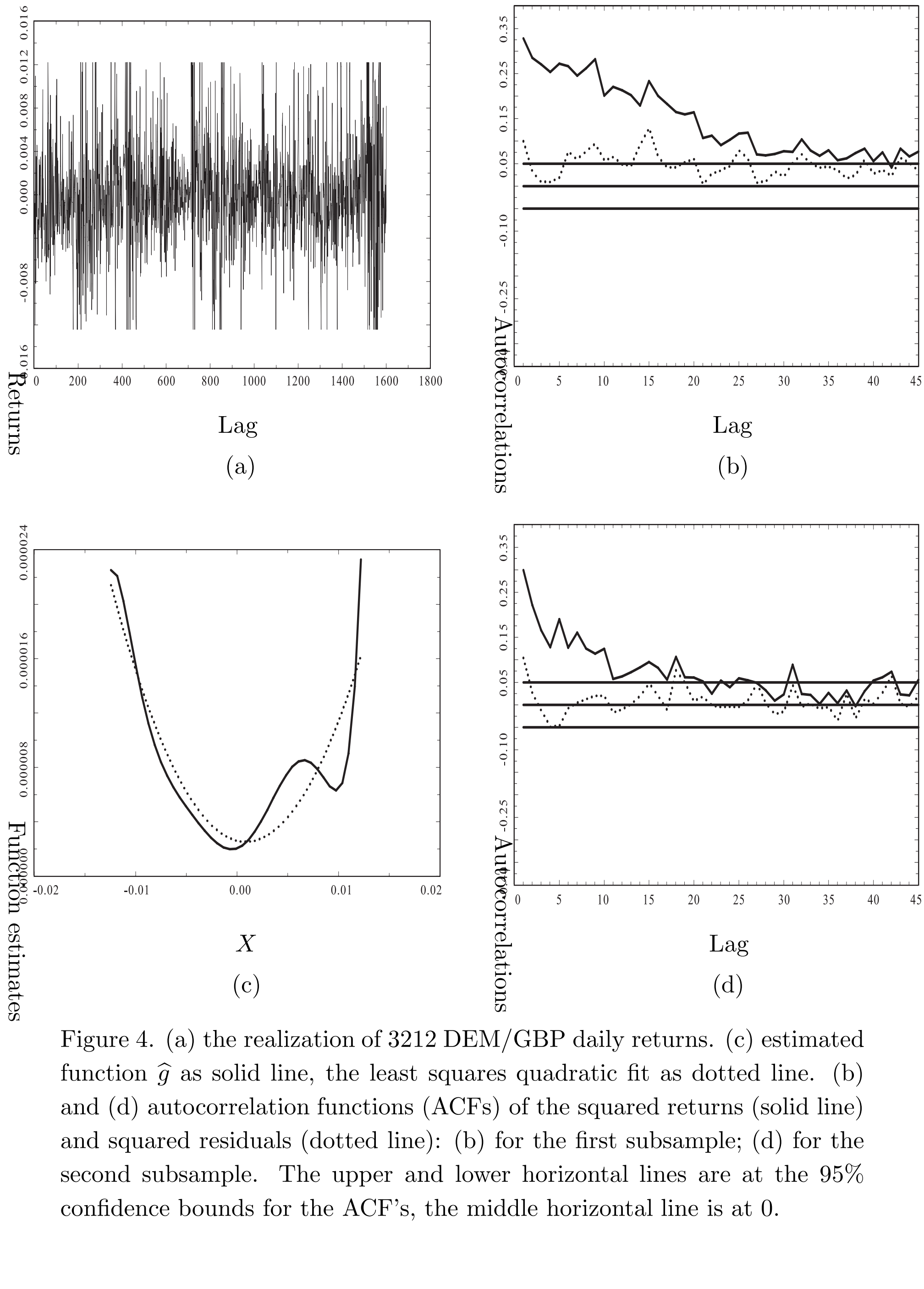

Figure 4. (a) the realization of 3212 DEM/GBP daily returns. (c) estimated function $\widehat{g}$ as solid line, the least squares quadratic fit as dotted line. (b) and (d) autocorrelation functions (ACFs) of the squared returns (solid line) and squared residuals (dotted line): (b) for the first subsample; (d) for the second subsample. The upper and lower horizontal lines are at the 95% confidence bounds for the ACF's, the middle horizontal line is at 0.

for out-of-sample prediction. It is significant that predicted volatilities based on the first subsample remove much of the dependence from the second subsample $\{Y_t\}_{t=1606}^{3211}$. This shows that the model fits well the intrinsic dynamics of DEM/GBP daily returns.

In summary, both the function and the parameter estimation work well given the moderate sample sizes and the high dimension d used in the simulation. The direct estimator substantially improves upon the integration method, contrary to asymptotic analysis. The oracle estimator performs nearly as well as its competing univariate estimator, just as Theorem 2 has concluded.

Acknowledgements

The research was supported in part by NSF grant DMS 9971186. The helpful comments of two anonymous referees are gratefully acknowledged. The stimulating comments of the Co-Editor and an associate editor led to the formulation of the oracle estimator.

Appendix

The following assumptions are used.

A1: *The kernel $K(\cdot)$ is a symmetric, compact supported and continuous probability density.*

A2: *The function $g(\cdot)$ has a Lipschitz continuous $(p+1)$-st derivative.*

A3: *The variance function $\sigma^2(\cdot)$ is Lipschitz continuous.*

A4: *The design variable $\mathbf{X}$ has density $\varphi(\cdot)$ and marginal densities $\varphi_\alpha(\cdot)$, $\alpha = 1, \ldots, d$, which are Lipschitz continuous and satisfy $\inf_{1\le\alpha\le d, x\in A} \varphi_\alpha(x) \ge c_A > 0$, where A is a compact subset of R with nonempty interior.*

A5: *The set $\boldsymbol{\gamma}$ is an r-dimensional submanifold of R^r with C^2 boundary, while the map $\boldsymbol{\gamma}' \to \{c_\alpha(\boldsymbol{\Gamma}')\}_{\alpha=1}^d$ is a C^2 diffeomorphism from $\boldsymbol{\Gamma}$ onto its image. There exists a constant $C > 0$ such that for a given compactly supported weight function $\pi(\cdot)$, the following condition holds:*

$$\int \left[\sum_{\alpha,\alpha'=1}^{d} \{c_{\alpha'}(\boldsymbol{\gamma}')c_\alpha(\boldsymbol{\gamma}'') - c_{\alpha'}(\boldsymbol{\gamma}'')c_\alpha(\boldsymbol{\gamma}')\} g(w_\alpha) \right]^2 \pi(\mathbf{w})d\mathbf{w} \ge C \left\|\boldsymbol{\gamma}'' - \boldsymbol{\Gamma}'\right\|^2, \quad \forall \boldsymbol{\gamma}'', \boldsymbol{\gamma}' \in \boldsymbol{\Gamma}. \tag{A.1}$$

A6: *There exists a constant $C > 0$ such that*

$$\left| \sum_{\alpha=1}^{d} c_\alpha(\boldsymbol{\gamma}') \right| \ge C, \boldsymbol{\gamma}' \in \boldsymbol{\Gamma}. \tag{A.2}$$

One should note that assumptions A1-A6 are satisfied in all our Monte-Carlo examples.

Denote $\mu_r(K) = \int u^r K(u)du$ and let $\{s_{st}(\boldsymbol{\gamma})\}_{0\le s,t\le p} = S_\gamma^{-1}$, where the matrix S_γ is defined as

$$S_{1,\gamma} = \sum_{\alpha=1}^{d} c_\alpha^2(\boldsymbol{\gamma})\left(\mu_{\lambda+\lambda'}(K)\right)_{\lambda,\lambda'=0}^{p} + \sum_{\alpha\neq\alpha'} c_\alpha(\boldsymbol{\gamma})c_{\alpha'}(\boldsymbol{\gamma})\left(\mu_\lambda(K)\right)_{\lambda=0}^{p}\left\{\left(\mu_\lambda(K)\right)_{\lambda=0}^{p}\right\}^T. \tag{A.3}$$

Note here by the definition of matrix S_γ, the multivariate equivalent kernel $K^*_{\lambda,\gamma}(\mathbf{u})$ satisfies

$$\sum_{\alpha'=1}^{d}\int K^*_{\lambda,\gamma}(\mathbf{u})c_{\alpha'}(\boldsymbol{\gamma})u_{\alpha'}^{\lambda''}d\mathbf{u} = \begin{cases} 1 & \lambda''=\lambda \\ 0 & 0\le\lambda''\le p, \quad \lambda''\neq\lambda \\ \Lambda_{\lambda,p+1,\gamma} & \lambda''=p+1, \end{cases} \tag{A.4}$$

where $\Lambda_{\lambda,p+1,\gamma}$ is a mixture of $(p+1)$-st moments depending on the parameters $\boldsymbol{\gamma} = \{\gamma_\alpha\}_{\alpha=1}^{q}$.

Lemma A.1. *As* $n\to\infty$,

$$Z'_\gamma W Z_\gamma = \varphi(\mathbf{x})S_\gamma\{I+o_p(1)\}, \quad \left(Z'_\gamma W Z_\gamma\right)^{-1} = \varphi(\mathbf{x})^{-1}S_\gamma^{-1}\{I+o_p(1)\}. \tag{A.5}$$

Proof. The (λ,λ')-th entry of $Z'_\gamma W Z_\gamma$ is of the form

$$\frac{1}{n}\sum_{i=1}^{n}K_h(\mathbf{X}_i-\mathbf{x})\left\{\sum_{\alpha=1}^{d}c_\alpha((\boldsymbol{\gamma})\left(\frac{X_{i\alpha}-x}{h}\right)^{\lambda}\right\}\left\{\sum_{\alpha=1}^{d}c_\alpha(\boldsymbol{\gamma})\left(\frac{X_{i\alpha}-x}{h}\right)^{\lambda'}\right\}$$

$$= \sum_{\alpha=1}^{d}\sum_{\alpha'=1}^{d}\frac{1}{n}\sum_{i=1}^{n}K_h(\mathbf{X}_i-\mathbf{x})c_\alpha(\boldsymbol{\gamma})\left(\frac{X_{i\alpha}-x}{h}\right)^{\lambda}c_{\alpha'}(\boldsymbol{\gamma})\left(\frac{X_{i\alpha'}-x}{h}\right)^{\lambda'}$$

$$= \sum_{\alpha=1}^{d}\sum_{\alpha'=1}^{d}\int K_h(\mathbf{w}-\mathbf{x})c_\alpha(\boldsymbol{\gamma})\left(\frac{w_\alpha-x}{h}\right)^{\lambda}c_{\alpha'}(\boldsymbol{\gamma})\left(\frac{w_{\alpha'}-x}{h}\right)^{\lambda'}\varphi(\mathbf{w})d\mathbf{w}\{1+o_p(1)\}.$$

By a change of variable $\mathbf{w} = \mathbf{x}+h\mathbf{u}$, this becomes

$$\sum_{\alpha=1}^{d}\sum_{\alpha'=1}^{d}\int K(\mathbf{u})c_\alpha(\boldsymbol{\gamma})c_{\alpha'}(\boldsymbol{\gamma})u_\alpha^{\lambda}u_{\alpha'}^{\lambda'}\varphi(\mathbf{x}+h\mathbf{u})d\mathbf{u}\{1+o_p(1)\}$$

$$= \varphi(\mathbf{x})\sum_{\alpha=1}^{d}\sum_{\alpha'=1}^{d}c_\alpha(\boldsymbol{\gamma})c_{\alpha'}(\boldsymbol{\gamma})\int K(\mathbf{u})u_\alpha^{\lambda}u_{\alpha'}^{\lambda'}d\mathbf{u}\{1+o_p(1)\}.$$

By direct calculation,

$$\int K(\mathbf{u})u_\alpha^{\lambda}u_{\alpha'}^{\lambda'}d\mathbf{u} = \begin{cases} \mu_\lambda(K)\mu_{\lambda'}(K) & \alpha\neq\alpha', \quad \text{both } \lambda \text{ and } \lambda' \text{ even} \\ 0 & \alpha\neq\alpha', \quad \lambda \text{ or } \lambda' \text{ odd} \\ 0 & \alpha=\alpha', \quad \lambda+\lambda'\text{odd} \\ \mu_{\lambda+\lambda'}(K) & \alpha=\alpha', \quad \lambda+\lambda' \text{ even,} \end{cases}$$

and so the (λ, λ')-th entry of $Z_\gamma' W Z_\gamma$ is

$$\begin{cases} \varphi(\mathbf{x})\left\{\sum_{\alpha=1}^{d} c_\alpha^2(\boldsymbol{\gamma})\mu_{\lambda+\lambda'}(K)\right. \\ \quad \left. + \sum_{\alpha\neq\alpha'} c_\alpha(\boldsymbol{\gamma})c_{\alpha'}(\boldsymbol{\gamma})\mu_\lambda(K)\mu_{\lambda'}(K)\right\} & \text{both } \lambda \text{ and } \lambda' \text{ even} \\ \qquad\qquad 0 & \lambda \text{ and } \lambda' \text{ one odd, one even} \\ \varphi(\mathbf{x})\left\{\sum_{\alpha=1}^{d} c_\alpha^2(\boldsymbol{\gamma})\mu_{\lambda+\lambda'}(K)\right\} & \text{both } \lambda \text{ and } \lambda' \text{ odd.} \end{cases}$$

Now (A.5) follows from the definitions of $S_{1,\gamma}$ and $S_{2,\gamma}$, see (A.3). The second equation follows from the first.

Before I proceed to prove Theorem 1, note by definition that

$$S_\gamma^{-1} Z_\gamma' W = \frac{1}{n\varphi(\mathbf{x})} \begin{Bmatrix} K_{0,\gamma,h}^*(\mathbf{X}_1-\mathbf{x}) & \cdots & K_{0,\gamma,h}^*(\mathbf{X}_n-\mathbf{x}) \\ \vdots & \ddots & \vdots \\ K_{p,\gamma,h}^*(\mathbf{X}_1-\mathbf{x}) & \cdots & K_{p,\gamma,h}^*(\mathbf{X}_n-\mathbf{x}) \end{Bmatrix}. \tag{A.6}$$

Proof of Theorem 1. By definition of the matrices, for a fixed λ, $E_\lambda'(Z_\gamma' W Z_\gamma)^{-1} Z_\gamma' W Z_\gamma E_\lambda = 1$, $E_\lambda' \left(Z_\gamma' W Z_\gamma\right)^{-1} Z_\gamma' W Z_\gamma E_{\lambda'} = 0$, $0 \leq \lambda' \leq p$, $\lambda' \neq \lambda$, so

$$\begin{aligned} \widehat{g^{(\lambda)}}(x) - g^{(\lambda)}(x) &= \lambda! h^{-\lambda} E_\lambda' (Z_\gamma' W Z_\gamma)^{-1} Z_\gamma' W \mathbf{Y} - g^{(\lambda)}(x) E_\lambda' (Z_\gamma' W Z_\gamma)^{-1} Z_\gamma' W Z_\gamma E_\lambda \\ &\quad - \sum_{\lambda'\geq 0, \lambda'\neq\lambda} \frac{\lambda!}{\lambda'!} g^{(\lambda')}(x) h^{\lambda'} E_\lambda' \left(Z_\gamma' W Z_\gamma\right)^{-1} Z_\gamma' W Z_\gamma E_{\lambda'} \\ &= \frac{\lambda!}{\varphi(\mathbf{x}) h^\lambda} E_\lambda' S_\gamma^{-1} Z_\gamma' W \left\{ \mathbf{Y} - \frac{g^{(\lambda)}(x) h^\lambda}{\lambda!} Z_\gamma E_\lambda - \sum_{\lambda'\geq 0, \lambda'\neq\lambda} \frac{1}{\lambda'!} g^{(\lambda')}(x) h^{\lambda'} Z_\gamma E_{\lambda'} \right\} \\ &\quad \times \{1 + o_p(1)\} = I_1 + I_2, \end{aligned} \tag{A.7}$$

$$I_1 = \frac{\lambda!}{n\varphi(\mathbf{x}) h^\lambda} \sum_{i=1}^{n} K_{\lambda,\gamma,h}^*(\mathbf{X}_i - \mathbf{x}) \times$$
$$\left[\sum_{\alpha=1}^{d} c_\alpha(\boldsymbol{\gamma}) \left\{ g(X_{i\alpha}) - g(x) - \sum_{\lambda'=1}^{p} \frac{1}{\lambda'!} g^{(\lambda')}(x) \left(X_{i\alpha} - x\right)^{\lambda'} \right\} \right] \{1 + o_p(1)\}, \tag{A.8}$$

$$I_2 = \frac{\lambda!}{n\varphi(\mathbf{x}) h^\lambda} \sum_{i=1}^{n} K_{\lambda,\gamma,h}^*(\mathbf{X}_i - \mathbf{x}) \sigma(\mathbf{X}_i) \varepsilon_i \{1 + o_p(1)\}, \tag{A.9}$$

by (A.5) and (A.6). Note next that

$$\frac{\lambda!}{n\varphi(\mathbf{x}) h^\lambda} \sum_{i=1}^{n} K_{\lambda,\gamma,h}^*(\mathbf{X}_i - \mathbf{x}) \left[\sum_{\alpha=1}^{d} c_\alpha(\boldsymbol{\gamma}) \left\{ g(X_{i\alpha}) - g(x) - \sum_{\lambda=1}^{p} \frac{1}{\lambda!} g^{(\lambda)}(x) \left(X_{i\alpha} - x\right)^\lambda \right\} \right]$$

$$= \frac{\lambda!}{n\varphi(\mathbf{x})h^\lambda} \int K^*_{\lambda,\gamma,h}(\mathbf{u}-\mathbf{x}) \left[\sum_{\alpha=1}^{d} c_\alpha(\boldsymbol{\gamma}) \left\{ g(u_\alpha) - g(x) - \sum_{\lambda=1}^{p} \frac{1}{\lambda!} g^{(\lambda)}(x)\,(u_\alpha - x)^\lambda \right\} \right]$$
$$\times \varphi(\mathbf{u}) d\mathbf{u} \left\{1 + o_p(1)\right\}.$$

By a change of variable $\mathbf{u} = \mathbf{x} + h\mathbf{v}$, this last becomes

$$\frac{\lambda!}{n\varphi(\mathbf{x})h^\lambda} \int K^*_{\lambda,\gamma}(\mathbf{v}) \left[\sum_{\alpha=1}^{d} c_\alpha(\boldsymbol{\gamma}) \left\{ g(x+hv_\alpha) - g(x) - \sum_{\lambda=1}^{p} \frac{1}{\lambda!} g^{(\lambda)}(x) h^\lambda v_\alpha^\lambda \right\} \right] \varphi(\mathbf{x}+h\mathbf{v}) d\mathbf{v}$$
$$= \frac{\lambda!}{\varphi(\mathbf{x})h^\lambda} \varphi(\mathbf{x}) \frac{h^{p+1}}{(p+1)!} g^{(p+1)}(x) \int K^*_{\lambda,\gamma}(\mathbf{v}) \left\{ \sum_{\alpha=1}^{d} c_\alpha(\boldsymbol{\gamma}) v_\alpha^{p+1} \right\} d\mathbf{v} \left\{1 + o_p(1)\right\},$$

which yields

$$I_1 = \frac{\lambda! \Lambda_{\lambda,p+1,\gamma} g^{(p+1)}(x)}{(p+1)!} h^{p+1-\lambda} + o_p(h^{p+1-\lambda}). \tag{A.10}$$

On the other hand, I_2 is asymptotically normal, with variance

$$\frac{(\lambda!)^2}{n\varphi^2(\mathbf{x})h^{2\lambda}} E\left\{ K^*_{\lambda,\gamma,h}(\mathbf{X}_1 - \mathbf{x})\sigma(\mathbf{X}_1)\varepsilon_1 \right\}^2 \left\{1 + o_p(1)\right\}$$
$$= \frac{(\lambda!)^2}{n\varphi^2(\mathbf{x})h^{2\lambda}} \int \left\{ K^*_{\lambda,,h}(\mathbf{u} - \mathbf{x})\sigma(\mathbf{u}) \right\}^2 \varphi(\mathbf{u}) d\mathbf{u} \left\{1 + o_p(1)\right\}.$$

By a change of variable $\mathbf{u} = \mathbf{x} + h\mathbf{v}$, this becomes

$$var(I_2) = \frac{(\lambda!)^2 \left\| K^*_{\lambda,\gamma} \right\|_2^2 \sigma^2(\mathbf{x})}{nh^{2\lambda+d}\varphi(\mathbf{x})} \left\{1 + o_p(1)\right\}. \tag{A.11}$$

Combining (A.10) and (A.11), I have finished the proof of the Theorem.

Proof of Theorem 2. In order to show (2.9), one shows

$$\sqrt{nh_1} \left\{ g_1(x) - g(x) - h_1^{p+1} b_1^*(x) \right\} \xrightarrow{D} N\left\{0, v_1^*(x)\right\}, \tag{A.12}$$

$$\sqrt{nh_1} \left\{ \widehat{g}_1(x) - g_1(x) \right\} \to 0, \tag{A.13}$$

respectively. To prove (A.12), one uses arguments similar to those used in the proof of Theorem 1 which easily yield the bias formula of $b_1^*(x)$ in (2.10). To derive the variance formula of $v_1^*(x)$, notice the noise term in $g_1(x) - g(x)$ is

$$\frac{1}{n\varphi_1(x)} \sum_{i=1}^{n} K^*_{0,h_1}(X_{i1} - x)\sigma(\mathbf{X}_i)\varepsilon_i \left\{1 + o_p(1)\right\},$$

whose variance is

$$\frac{1}{n\varphi_1^2(x)}EK_{0,h_1}^{*2}(X_{11}-x)\sigma^2(\mathbf{X}_1)\varepsilon_1^2\{1+o(1)\}$$
$$=\frac{1}{n\varphi_1^2(x)}\int K_{0,h_1}^{*2}(u_1-x)\sigma^2(\mathbf{u})\varphi(\mathbf{u})d\mathbf{u}\{1+o(1)\}$$
$$=\frac{1}{n\varphi_1^2(x)h_1}\int K_0^{*2}(v)\sigma^2(x+h_1v,z)\varphi(x+h_1v,z)dvdz\{1+o(1)\}$$
$$=\frac{v_1^*(x)}{nh_1}+o\left(\frac{1}{nh_1}\right)$$

according to the definition of $v_1^*(x)$ in (2.10).

To prove (A.13), note by definitions in (2.6), (2.5), (2.8) and (2.7),

$$\widehat{g}_1(x)-g_1(x)=E_0'\left(Z_1'W_1Z_1\right)^{-1}Z_1'W_1\left(\widetilde{\mathbf{Y}}_1-\mathbf{Y}_1\right)$$
$$=E_0'\left(Z_1'W_1Z_1\right)^{-1}Z_1'W_1\left[\sum_{\alpha=2}^{d}c_\alpha(\boldsymbol{\gamma})\{g(X_{i\alpha})-\widehat{g}(X_{i\alpha})\}\right]_{i=1}^{n}$$

which, according to the decomposition formula (A.7) and equations (A.10) and (A.9), equals B_e+V_e, where

$$B_e=\frac{1}{n^2\varphi_1(x)}\sum_{i=1}^{n}K_{0,h_1}^*(X_{i1}-x)\sum_{\alpha=2}^{d}c_\alpha(\boldsymbol{\gamma})\frac{\Lambda_{0,p+1,\boldsymbol{\gamma}}g^{(p+1)}(X_{i\alpha})}{(p+1)!}h^{p+1}\{1+o_p(1)\},\tag{A.14}$$

$$V_e=\frac{1}{n^2\varphi_1(x)}\sum_{i=1}^{n}\sum_{j=1}^{n}K_{0,h_1}^*(X_{i1}-x)\sum_{\alpha=2}^{d}c_\alpha(\boldsymbol{\gamma})\frac{K_{0,\boldsymbol{\gamma},h}^*(\mathbf{X}_j-\mathbf{X}_{i\alpha})}{\varphi(\mathbf{X}_{i\alpha})}\sigma(\mathbf{X}_j)\varepsilon_j\{1+o_p(1)\}.\tag{A.15}$$

Under the assumption that $h/h_1\to 0$, it is clear from (A.14) that the extra bias term $B_e=o_p\left(h_1^{p+1}\right)$.

The extra noise term V_e is comprised of sums of diagonal and off-diagonal terms:

$$V_{ed}=\frac{1}{n^2\varphi_1(x)}\sum_{i=1}^{n}K_{0,h_1}^*(X_{i1}-x)\sum_{\alpha=2}^{d}c_\alpha(\boldsymbol{\gamma})\frac{K_{0,\boldsymbol{\gamma},h}^*(\mathbf{X}_i-\mathbf{X}_{i\alpha})}{\varphi(\mathbf{X}_{i\alpha})}\sigma(\mathbf{X}_i)\varepsilon_i\{1+o_p(1)\},\tag{A.16}$$

$$V_{eod}=\frac{1}{n^2\varphi_1(x)}\sum_{i,j=1,i\neq j}^{n}K_{0,h_1}^*(X_{i1}-x)\sum_{\alpha=2}^{d}c_\alpha(\boldsymbol{\gamma})\frac{K_{0,\boldsymbol{\gamma},h}^*(\mathbf{X}_j-\mathbf{X}_{i\alpha})}{\varphi(\mathbf{X}_{i\alpha})}\sigma(\mathbf{X}_j)\varepsilon_j\{1+o_p(1)\}.\tag{A.17}$$

The diagonal sum V_{ed} is a mean zero sample mean whose variance is

$$
\begin{aligned}
EV_{ed}^2 &= \frac{1}{n^3\varphi_1^2(x)} E\left\{K_{0,h_1}^*(X_{11}-x)\sum_{\alpha=2}^{d} c_\alpha(\boldsymbol{\gamma})\frac{K_{0,\gamma,h}^*(\mathbf{X}_1-\mathbf{X}_{1\alpha})}{\varphi(\mathbf{X}_{1\alpha})}\sigma(\mathbf{X}_1)\right\}^2 \{1+o(1)\} \\
&= \frac{1}{n^3\varphi_1^2(x)}\int\left\{K_{0,h_1}^*(u_1-x)\sum_{\alpha=2}^{d} c_\alpha(\boldsymbol{\gamma})\frac{K_{0,\gamma,h}^*(\mathbf{u}-\mathbf{u}_\alpha)}{\varphi(\mathbf{u}_\alpha)}\sigma(\mathbf{u})\right\}^2 \varphi(\mathbf{u})d\mathbf{u}\,\{1+o(1)\}.
\end{aligned}
$$

After doing a change of variable $u_1 - x = h_1 v$, this becomes

$$
\begin{aligned}
&= \frac{1}{n^3 h_1\varphi_1^2(x)\varphi^2(\mathbf{x})}\int\left\{K_0^*(v)\sum_{\alpha=2}^{d} c_\alpha(\boldsymbol{\gamma})\frac{K_{0,\gamma,h}^*(\mathbf{u}-\mathbf{u}_\alpha)}{\varphi(\mathbf{u}_\alpha)}\sigma(\mathbf{u})\right\}^2 \varphi(x+h_1v,u_2,\ldots,u_d) \\
&\qquad\qquad dv du_2\cdots du_d\,\{1+o(1)\} \\
&= O\left(\frac{1}{nh_1 nh^d nh^d}\right) = o\left(\frac{1}{nh_1}\right),
\end{aligned}
$$

because $nh^d \to \infty$.

The variance of the off-diagonal sum V_{ed} is

$$
\begin{aligned}
EV_{eod}^2 &= \frac{\{1+o(1)\}}{n^2\varphi_1^2(x)\varphi^2(\mathbf{x})} E\left\{K_{0,h_1}^*(X_{11}-x)\sum_{\alpha=2}^{d} c_\alpha(\boldsymbol{\gamma})\frac{K_{0,\gamma,h}^*(\mathbf{X}_2-\mathbf{X}_{1\alpha})}{\varphi(\mathbf{X}_{1\alpha})}\sigma(\mathbf{X}_2)\varepsilon_2\right\}^2 \\
&\quad + \frac{\{1+o(1)\}}{n\varphi^2(x)\varphi^2(\mathbf{x})} \times E\Bigg[K_{0,h_1}^*(X_{11}-x)K_{0,h_1}^*(X_{21}-x) \\
&\quad \times\left\{\sum_{\alpha=2}^{d} c_\alpha(\boldsymbol{\gamma})\frac{K_{0,\gamma,h}^*(\mathbf{X}_3-\mathbf{X}_{1\alpha})}{\varphi(\mathbf{X}_{1\alpha})}\right\}\left\{\sum_{\alpha=2}^{d} c_\alpha(\boldsymbol{\gamma})\frac{K_{0,\gamma,h}^*(\mathbf{X}_3-\mathbf{X}_{2\alpha})}{\varphi(\mathbf{X}_{2\alpha})}\right\}\sigma^2(\mathbf{X}_3)\varepsilon_3^2\Bigg] \\
&= \frac{\{1+o(1)\}}{n^2\varphi_1^2(x)}\int E\left\{K_{0,h_1}^*(v_1-x)\sum_{\alpha=2}^{d} c_\alpha(\boldsymbol{\gamma})\frac{K_{0,\gamma,h}^*(\mathbf{u}-\mathbf{v}_\alpha)}{\varphi(\mathbf{v}_\alpha)}\sigma(\mathbf{u})\right\}^2 \varphi(\mathbf{u})d\mathbf{u}\varphi(\mathbf{v})d\mathbf{v} \\
&\quad + \frac{\{1+o(1)\}}{n\varphi_1^2(x)} \times \int K_{0,h_1}^*(u_1-x)K_{0,h_1}^*(v_1-x)\left\{\sum_{\alpha=2}^{d} c_\alpha(\boldsymbol{\gamma})\frac{K_{0,\gamma,h}^*(\mathbf{w}-\mathbf{u}_\alpha)}{\varphi(\mathbf{u}_\alpha)}\right\} \\
&\qquad \times\left\{\sum_{\alpha=2}^{d} c_\alpha(\boldsymbol{\gamma})\frac{K_{0,\gamma,h}^*(\mathbf{w}-\mathbf{v}_\alpha)}{\varphi(\mathbf{v}_\alpha)}\right\}\sigma^2(\mathbf{w})\varphi(\mathbf{u})d\mathbf{u}\varphi(\mathbf{v})d\mathbf{v}\varphi(\mathbf{w})d\mathbf{w}.
\end{aligned}
$$

After applying appropriate changes of variables, this is $O\left(n^{-1}h_1^{-1}n^{-1}h^{-d}+n^{-1}\right) = o\left(n^{-1}h_1^{-1}\right)$. Having shown that both V_{ed} in (A.16) and V_{ed} in (A.17) have variances of order $o\left(n^{-1}h_1^{-1}\right)$, it follows that V_e in (A.15) is of order $o\left(1/\sqrt{nh_1}\right)$. Now since both B_e and V_e are of order $o\left(1/\sqrt{nh_1}\right)$, (A.13) is proved, and hence (2.9).

Proof of Theorem 3. First note that $L(\boldsymbol{\gamma}')$ apparently allows the following decomposition:

$$L(\boldsymbol{\gamma}')=\frac{1}{n}\sum_{i=1}^{n}\left\{m(\mathbf{X}_i)-\widehat{g}_{\gamma'}(X_{i1})-c_2(\boldsymbol{\gamma}')\widehat{g}_{\gamma'}(X_{i2})-\cdots-c_d(\boldsymbol{\gamma}')\widehat{g}_{\gamma'}(X_{id})\right\}^2\pi(\mathbf{X}_i)$$
$$+\frac{2}{n}\sum_{i=1}^{n}\{m(\mathbf{X}_i)-\widehat{g}_{\gamma'}(X_{i1})-c_2(\boldsymbol{\gamma}')\widehat{g}_{\gamma'}(X_{i2})-\cdots-c_d(\boldsymbol{\gamma}')\widehat{g}_{\gamma'}(X_{id})\}\sigma(\mathbf{X}_i)\varepsilon_i\pi(\mathbf{X}_i)$$
$$+\frac{1}{n}\sum_{i=1}^{n}\sigma^2(\mathbf{X}_i)\varepsilon_i^2\pi(\mathbf{X}_i),$$

where the third term does not affect the minimization, and the second term is asymptotically normal with $\sqrt{n}$-rate uniformly for all $\boldsymbol{\gamma}'\in\Gamma$. So one is interested in minimizing the first term, which is equivalent to minimizing

$$\begin{aligned}L_1(\boldsymbol{\gamma}')&=\int\left\{m(\mathbf{w})-\widehat{g}_{\gamma'}(w_1)-c_2(\boldsymbol{\gamma}')\widehat{g}_{\gamma'}(w_2)-\cdots-c_d(\boldsymbol{\gamma}')\widehat{g}_{\gamma'}(w_d)\right\}^2\pi(\mathbf{w})d\mathbf{w}\\ &=\int\left[\left\{m(\mathbf{w})-\sum_{\alpha=1}^{d}c_\alpha(\boldsymbol{\gamma}')g(w_\alpha)\right\}+\sum_{\alpha=1}^{d}c_\alpha(\boldsymbol{\gamma}')\left\{g(w_\alpha)-\widehat{g}_{\gamma'}(w_\alpha)\right\}\right]^2\pi(\mathbf{w})d\mathbf{w}\\ &=\int\left[\sum_{\alpha=1}^{d}\left\{c_\alpha(\boldsymbol{\gamma})-c_\alpha(\boldsymbol{\gamma}')\right\}g(w_\alpha)+\sum_{\alpha=1}^{d}c_\alpha(\boldsymbol{\gamma}')\left\{g(w_\alpha)-\widehat{g}_{\gamma'}(w_\alpha)\right\}\right]^2\pi(\mathbf{w})d\mathbf{w}.\end{aligned}\tag{A.18}$$

In the following, note $h^{p+1}=o(1/\sqrt{n})$ and $nh^d=o(1/\sqrt{n})$, so one is free to collect any terms of order h^{p+1} or nh^d into $o(1/\sqrt{n})$.

For general $L_1(\boldsymbol{\gamma}')$, start by calculating $\widehat{g}_{\gamma'}(\cdot)$. Using (A.5) of Lemma A.1, $\left(Z'_{\gamma'}WZ_{\gamma'}\right)^{-1}=\varphi(\mathbf{x})^{-1}S_{\gamma'}^{-1}\left\{I+o_p(1)\right\}$ and, by definition (3.1) of $\widehat{g}_{\gamma'}(x)$,

$$\begin{aligned}&\widehat{g}_{\gamma'}(x)-g(x)\\ &=E_0'\left(Z'_{\gamma'}WZ_{\gamma'}\right)^{-1}Z'_{\gamma'}W\mathbf{Y}-g(x)E_0'\left(Z'_{\gamma'}WZ_{\gamma'}\right)^{-1}Z'_{\gamma'}WZ_{\gamma'}E_0\\ &\quad-\sum_{\lambda=1}^{p}\frac{1}{\lambda!}g^{(\lambda)}(x)h^\lambda E_0'\left(Z'_{\gamma'}WZ_{\gamma'}\right)^{-1}Z'_{\gamma'}WZ_{\gamma'}E_\lambda\\ &=\frac{1}{\varphi(\mathbf{x})}E_0'S_{\gamma'}^{-1}Z'_{\gamma'}W\left\{\mathbf{Y}-g(x)Z_{\gamma'}E_0-\sum_{\lambda=1}^{p}\frac{1}{\lambda!}g^{(\lambda)}(x)h^\lambda Z_{\gamma'}E_\lambda\right\}\left\{1+o_p(1)\right\}\\ &=I_1+I_2,\end{aligned}$$

$$I_1=\frac{\{1+o_p(1)\}}{n\varphi(\mathbf{x})}\sum_{i=1}^{n}K^*_{0,\gamma',h}(\mathbf{X}_i-\mathbf{x})\sum_{\alpha=1}^{d}\left[c_\alpha(\boldsymbol{\gamma})g(X_{i\alpha})-c_\alpha(\boldsymbol{\gamma}')\times\left\{g(x)+\sum_{\lambda=1}^{p}\frac{1}{\lambda!}g^{(\lambda)}(x)\left(X_{i\alpha}-x\right)^\lambda\right\}\right],$$

$$I_2 = \frac{1}{n\varphi(\mathbf{x})}\sum_{i=1}^{n} K^*_{0,\gamma',h}(\mathbf{X}_i - \mathbf{x})\sigma(\mathbf{X}_i)\varepsilon_i \left\{1 + o_p(1)\right\},$$

by (A.6). The term I_1 is then calculated the same way as in the case when the correct parameter $\boldsymbol{\gamma}$ is used, successively

$$I_1 = \frac{\{1+o_p(1)\}}{n\varphi(\mathbf{x})}\sum_{i=1}^{n} K^*_{0,\gamma',h}(\mathbf{X}_i - \mathbf{x})\sum_{\alpha=1}^{d} c_\alpha(\boldsymbol{\gamma}')\left\{g(X_{i\alpha}) - g(x) - \sum_{\lambda=1}^{p}\frac{1}{\lambda!}g^{(\lambda)}(x)(X_{i\alpha}-x)^\lambda\right\}$$
$$+\frac{\{1+o_p(1)\}}{n\varphi(\mathbf{x})}\sum_{i=1}^{n} K^*_{0,\gamma',h}(\mathbf{X}_i - \mathbf{x})\sum_{\alpha=1}^{d}\left\{c_\alpha(\boldsymbol{\gamma}) - c_\alpha(\boldsymbol{\gamma}')\right\} g(X_{i\alpha}),$$

$$I_1 = \frac{\Lambda_{0,p+1,\gamma'}\, g^{(p+1)}(x)}{(p+1)!}h^{p+1} + o_p(h^{p+1})$$
$$+\frac{\{1+o_p(1)\}}{\varphi(\mathbf{x})}\int K^*_{0,\gamma',h}(\mathbf{w}-\mathbf{x})\sum_{\alpha=1}^{d}\left\{c_\alpha(\boldsymbol{\gamma}) - c_\alpha(\boldsymbol{\gamma}')\right\} g(w_\alpha)\varphi(\mathbf{w})d\mathbf{w}.$$
$$=\frac{\{1+o_p(1)\}}{\varphi(\mathbf{x})}\int K^*_{0,\gamma'}(\mathbf{u})\sum_{\alpha=1}^{d}\left\{c_\alpha(\boldsymbol{\gamma}) - c_\alpha(\boldsymbol{\gamma}')\right\} g(x+hu_\alpha)\varphi(\mathbf{x}+h\mathbf{u})d\mathbf{u} + o_p(\frac{1}{\sqrt{n}})$$
$$= g(x)\int K^*_{0,\gamma'}(\mathbf{u})\sum_{\alpha=1}^{d}\left\{c_\alpha(\boldsymbol{\gamma}) - c_\alpha(\boldsymbol{\gamma}')\right\} d\mathbf{u}\left\{1 + o_p(1)\right\} + o_p(\frac{1}{\sqrt{n}})$$
$$= g(x)\frac{\sum_{\alpha=1}^{d}\left\{c_\alpha(\boldsymbol{\gamma}) - c_\alpha(\boldsymbol{\gamma}')\right\}}{\sum_{\alpha=1}^{d} c_\alpha(\boldsymbol{\gamma}')}\left\{1 + o_p(1)\right\} + o_p(\frac{1}{\sqrt{n}}).$$

Meanwhile, it is easy to verify by *U-statistic* arguments that I_2^2, when integrated, is of order nh^d, and hence $o_p(1/\sqrt{n})$. Using these properties of $\widehat{g}_{\gamma'}(\cdot)$ and (A.18), $L_1(\boldsymbol{\gamma}')$ is

$$\int\Big[\sum_{\alpha=1}^{d}\left\{c_\alpha(\boldsymbol{\gamma}) - c_\alpha(\boldsymbol{\gamma}')\right\} g(w_\alpha) - \frac{\sum\limits_{\alpha'=1}^{d}\left\{c_{\alpha'}(\boldsymbol{\gamma}) - c_{\alpha'}(\boldsymbol{\gamma}')\right\}}{\sum\limits_{\alpha'=1}^{d} c_{\alpha'}(\boldsymbol{\gamma}')}\sum_{\alpha=1}^{d} c_\alpha(\boldsymbol{\gamma}')g(w_\alpha)\Big]^2\pi(\mathbf{w})$$
$$d\mathbf{w}\left\{1 + o_p(1)\right\}$$

$$=\left\{\sum_{\alpha'=1}^{d} c_{\alpha'}(\boldsymbol{\gamma}')\right\}^{-2}\int\left[\sum_{\alpha,\alpha'=1}^{d}\left\{c_{\alpha'}(\boldsymbol{\gamma}')c_\alpha(\boldsymbol{\gamma}) - c_{\alpha'}(\boldsymbol{\gamma})c_\alpha(\boldsymbol{\gamma}')\right\} g(w_\alpha)\right]^2\pi(\mathbf{w})d\mathbf{w}\{1+o_p(1)\}$$
$$\geq C'\left\|\boldsymbol{\gamma} - \boldsymbol{\gamma}'\right\|^2\left\{1 + o_p(1)\right\}$$

for all $\boldsymbol{\gamma}' \in \boldsymbol{\Gamma}$, based on (A.2) and (A.1). Now note that

$$\begin{aligned}
&L_1(\boldsymbol{\gamma})\\
&=\int \left[\sum_{\alpha=1}^{d} c_\alpha(\boldsymbol{\gamma}) \left\{g(w_\alpha) - \widehat{g}_\gamma(w_\alpha)\right\}\right]^2 \pi(\mathbf{w})d\mathbf{w}\\
&=\int\left[\sum_{\alpha=1}^{d}c_\alpha(\boldsymbol{\gamma})\left\{b(w_\alpha)h^{p+1}+\frac{1}{n\varphi(\mathbf{w}_\alpha)}\sum_{i=1}^{n}K^*_{0,\gamma,h}(\mathbf{X}_i-\mathbf{w}_\alpha)\sigma(\mathbf{X}_i)\varepsilon_i\right\}\right]^2\pi(\mathbf{w})d\mathbf{w}\{1+o_p(1)\}\\
&=o_p(h^{2p+2})+\int\left\{\sum_{\alpha=1}^{d} c_\alpha(\boldsymbol{\gamma})\frac{1}{n\varphi(\mathbf{w}_\alpha)}\sum_{i=1}^{n}K^*_{0,\gamma,h}(\mathbf{X}_i-\mathbf{w}_\alpha)\sigma(\mathbf{X}_i)\varepsilon_i\right\}^2\pi(\mathbf{w})d\mathbf{w}\\
&=O_p(1/n).
\end{aligned}$$

This implies $O_p(1/n) = L_1(\boldsymbol{\gamma}) \geq L_1(\widehat{\boldsymbol{\gamma}}) \geq C' \left\|\boldsymbol{\gamma} - \widehat{\boldsymbol{\gamma}}\right\|^2 \{1 + o_p(1)\}$, which entails that

$$\|\boldsymbol{\gamma} - \widehat{\boldsymbol{\gamma}}\| = O_p(1/\sqrt{n}). \tag{A.19}$$

One can then use (A.19) and proceed by routine differentiation of $L_1(\boldsymbol{\gamma}')$ to obtain the asymptotic normality (3.4) of $\widehat{\boldsymbol{\gamma}}$.

References

Bollerslev, T. P. (1986). Generalized autoregressive conditional heteroscedasticity. *J. Econometrics* **31**, 307-327.

Carroll, R., Härdle, W. and Mammen, E. (1999). Estimation in an additive model when the components are linked parametrically. *Discussion Paper* 01, SFB 373, Humboldt Universität zu Berlin, *Econometric Theory*, tentatively accepted.

Engle, R. F. and Ng, V. (1993). Measuring and testing the impact of news on volatility. *J. Finance* **48**, 1749-1778.

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

Fan, J. and Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **85**, 645-660.

Hafner, C. (1998). *Nonlinear Time Series Analysis with Applications to Foreign Exchange Rate Volatility.* Physica-Verlag, Heidelberg.

Härdle, W. and Tsybakov, A. B. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *J. Econometrics* **81**, 223-242.

Härdle, W., Tsybakov, A. B. and Yang, L. (1998). Nonparametric vector autoregression. *J. Statist. Plann. Inference* **68**, 221-245.

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

Linton, O. B. and Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82**, 93-101.

Masry, E. and Tjøstheim, D. (1996). Additive nonlinear ARX time series and projection estimates. *Econometric Theory* **13**, 214-252.

Opsomer, J. D. and Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *Ann. Statist.* **25**, 186-211.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, London.

Stone, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13**, 689-705.

Tjøstheim, D. and Auestad, B. H. (1994). Nonparametric identification of nonlinear time series: projections. *J. Amer. Statist. Assoc.* **89**, 398-1409.

Yang, L., Härdle, W. and Nielsen, J. P. (1999). Nonparametric autoregression with multiplicative volatility and additive mean. *J. Time Ser. Anal.* **20**, 579-604.

Yang, L. and Tschernig, R. (1999). Multivariate bandwidth selection for local linear regression. *J. Roy. Statist. Soc. Ser. B* **61**, 793-815.

Department of Statistics and Probability, Michigan State University, East Lansing, MI 48824, U.S.A.

E-mail: yang@stt.msu.edu

(Received September 2000; accepted November 2001)