Econometric Theory, **16**, 2000, 465–501. Printed in the United States of America.

NONPARAMETRIC ESTIMATION OF ADDITIVE NONLINEAR ARX TIME SERIES: LOCAL LINEAR FITTING AND PROJECTIONS

ZONGWU CAI
University of North Carolina at Charlotte

ELIAS MASRY
University of California at San Diego

We consider the estimation and identification of the components (endogenous and exogenous) of additive nonlinear ARX time series models. We employ a local polynomial fitting scheme coupled with projections. We establish the weak consistency (with rates) and the asymptotic normality of the projection estimates of the additive components. Expressions for the asymptotic bias and variance are given.

1. INTRODUCTION

Let $\{X_l, Y_l\}_{l=-\infty}^{\infty}$ be jointly stationary discrete-time processes. Among the nonlinear time series models popular in the econometrics literature is the bivariate ARX model:

$$Y_l = \tilde{g}_1(Y_{l-\bar{q}}, \ldots, Y_{l-1}) + \tilde{g}_2(X_{l-\bar{p}}, \ldots, X_l) + e_l, \tag{1.1}$$

$$X_l = \tilde{g}_3(X_{l-\bar{p}}, \ldots, X_{l-1}) + \varepsilon_l, \tag{1.2}$$

where $\{e_l\}$ and $\{\varepsilon_l\}$ are independent series each consisting of zero mean independent and identically distributed (i.i.d.) variables with finite variances σ_e^2 and σ_ε^2, respectively. Sufficient conditions for $\{X_l, Y_l\}$ defined in (1.1) and (1.2) to be stationary are given in Lemma 3.1 in Masry and Tjøstheim (1997). The variables $\{X_l\}$ and $\{Y_l\}$ are called exogenous and endogenous, respectively, and the estimation of the relationship between $\{X_l\}$ and $\{Y_l\}$ is of importance in econometric modeling.

Because of the "curse of dimensionality," reliable estimates of $\tilde{g}_1(\cdot)$, $\tilde{g}_2(\cdot)$, and $\tilde{g}_3(\cdot)$, using a moderate sample size n, are difficult to obtain nonparametri-

The work of E. Masry was supported by the National Science Foundation under grant DMS-97-03876. Z. Cai did much of his work on the paper during his visit at U.C. San Diego. He is grateful for the hospitality and the partial support offered by E. Masry. We thank the editor and the two referees for their insightful comments, which led to a substantial improvement of the paper. Address correspondence to: Zongwu Cai, Department of Mathematics, University of North Carolina, Charlotte, NC 28223, USA; e-mail: zcai@uncc.edu.

© 2000 Cambridge University Press 0266-4666/00 $9.50

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

cally even for relatively small values of $\bar{p}$ and $\bar{q}$. Following Masry and Tjøstheim (1997), we may consider, therefore, an ARX model with only a few significant lags. Let $0 \le i_1 < \cdots < i_{q+1}$ and $0 \le j_1 < \cdots < j_p$ be integers with $j_p \le i_{q+1}$ and assume

$$Y_{l+i_{q+1}} = g_1(Y_{l+i_1},\ldots,Y_{l+i_q}) + g_2(X_{l+j_1},\ldots,X_{l+j_p}) + e_{l+i_{q+1}}, \tag{1.3}$$

$$X_{l+j_p} = g_3(X_{l+j_1},\ldots,X_{l+j_{p-1}}) + \varepsilon_{l+j_p}. \tag{1.4}$$

Here i_q and j_p may be large, but p and q can be small. The values of $i_1,\ldots,i_{q+1}$ and $j_1,\ldots,j_p$ are assumed to be known; the procedures for selecting these lags were discussed in Tjøstheim and Auestad (1994b) in the univariate case.

Estimates of the g-functions given in (1.3) and (1.4) can be used to obtain nonparametric forecasts; i.e., estimate of the regression function $m(\cdot)$,

$$m(x_1,\ldots,x_p;y_1,\ldots,y_q) = E\{Y_{i_{q+1}} | X_{j_1} = x_1,\ldots,X_{j_p} = x_p;\, Y_{i_1} = y_1,\ldots,Y_{i_q} = y_q\}. \tag{1.5}$$

The interest in nonlinear ARX time series and regression models has been increasing in econometrics and also in related fields. Both parametric and nonparametric modeling have been considered. We refer to Chen and Tsay (1993) for nonparametric additive modeling, Friedman (1991) and Lewis and Stevens (1991) for MARS modeling, and Granger and Teräsvirta (1993) for parametric modeling. Whereas extensive and rigorous theory has been established (see, e.g., Pötscher and Prucha, 1991a, 1991b) in the parametric case, less has been achieved for the nonparametric methods. For additive models, rigorous results are mainly restricted to the case of independent components through the basic paper of Stone (1985); additional references, using the projection method, are given subsequently.

Because the function $g_3(\cdot)$ can be estimated straightforwardly by kernel-type estimation and was treated by Masry and Tjøstheim (1995), we put the emphasis on the estimation of $g_1(\cdot)$ and $g_2(\cdot)$ here. Our approach to the additive modeling in general and to the additive nonlinear ARX model in particular will be through local linear fitting coupled with the projection method. We note that the regression function $m(\cdot)$ defined in (1.5) can only identify the sum

$$m(x_1,\ldots,x_p;y_1,\ldots,y_q) = g_1(y_1,\ldots,y_q) + g_2(x_1,\ldots,x_p)$$

of the functions $g_1(\cdot)$ and $g_2(\cdot)$ in the ARX system. Define the projection as follows:

$$\begin{aligned} P_Y(y_1,\ldots,y_q) &= E\{m(X_{j_1},\ldots,X_{j_p};y_1,\ldots,y_q)\} \\ &= g_1(y_1,\ldots,y_q) + \int_{\Re^p} g_2(x_1,\ldots,x_p)\, dF(x_1,\ldots,x_p), \end{aligned} \tag{1.6}$$

where $F(\cdot)$ is the joint distribution function of $X_{j_1},\ldots,X_{j_p}$. We are able to identify $g_1(\cdot)$ up to an additive constant. The function $g_2(\cdot)$ can be retrieved likewise by projecting on the X-variables.

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

Projections were introduced in Auestad and Tjøstheim (1991), and they were more systematically explored in Tjøstheim and Auestad (1994a), in both cases for a univariate additive model with the purpose of identifying the functional structure of the components. The projection idea has been formulated independently by Linton and Nielsen (1995) under the name "marginal integration." For application of the projection method to additive regression models in an i.i.d. setting see Linton and Nielsen (1995), Linton and Härdle (1996), Fan, Härdle, and Mammen (1998), and Linton (1997). Masry and Tjøstheim (1997) considered the estimation of the additive components $g_1(\cdot)$ and $g_2(\cdot)$ using the Nadaraya–Watson approach for estimating $m(x_1,\ldots,x_p;y_1,\ldots,y_q)$ and the projection method for estimating the components $g_1(\cdot)$ and $g_2(\cdot)$. They established weak consistency and asymptotic normality for the projection estimates under a precise set of regularity conditions.

The goal of this paper is to extend the work of Masry and Tjøstheim (1997) by using local linear fitting plus the projection method to estimate the additive components $g_1(\cdot)$ and $g_2(\cdot)$ of the regression function $m(x_1,\ldots,x_p;y_1,\ldots,y_q)$.

Local linear regression estimation, and more generally local polynomial fitting, was introduced originally by Stone (1977) and studied by Cleveland (1979), Fan (1992, 1993), Ruppert and Wand (1994), Masry (1996a, 1996b), Masry and Fan (1997), and many others. See the book of Fan and Gijbels (1996) for additional references. Local polynomial fitting has significant advantages over Nadaraya–Watson regression estimates. It reduces the bias (Fan, 1992), and it adapts automatically to the boundary of design points (see Fan and Gijbels, 1996). Using a minimax argument, Fan (1993) showed that within the class of linear estimators that includes kernel and spline estimates, the local linear estimators achieve the best possible constant and rates of convergence.

Our formulation is not actually limited to the ARX model (1.3) and (1.4) but deals with the general additive regression model. We proceed as follows: Let $\{X_l,Y_l\}$ be jointly stationary processes. For integers $0 \le i_1 < \cdots < i_{q+1}$ and $0 \le j_1 < \cdots < j_p$, define the regression function

$$m(x_1,\ldots,x_p;y_1,\ldots,y_q) = E\{\phi(Y_{iq+1})|X_{j_1}=x_1,\ldots,X_{j_p}=x_p;\, Y_{i_1}=y_1,\ldots,Y_{i_q}=y_q\}, \tag{1.7}$$

where $\phi(\cdot)$ is an arbitrary measurable function on the real line and it is assumed that $E|\phi(Y_{i_{q+1}})| < \infty$. The introduction of $\phi(\cdot)$ allows us to estimate conditional distributions ($\phi(Y) = I(Y \le u)$) and conditional moments ($\phi(Y) = Y^r$). We assume that the regression function $m(x_1,\ldots,x_p;y_1,\ldots,y_q)$ has the additive decomposition

$$m(x_1,\ldots,x_p;y_1,\ldots,y_q) = g_1(y_1,\ldots,y_q) + g_2(x_1,\ldots,x_p). \tag{1.8}$$

Such a decomposition holds, for example, for the ARX model (1.1) and (1.2). Using local linear fitting coupled with the projection method, we establish weak consistency (with rates) and asymptotic normality for the projection estimates of $g_1(\cdot)$ and $g_2(\cdot)$. Explicit asymptotic expressions for the bias and variance of

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

the projection estimates are given. The results hold under a set of regularity conditions on the processes $\{X_l, Y_l\}$. In particular, it is assumed that $\{X_l, Y_l\}$ are strongly mixing with appropriate algebraic decay for their mixing coefficient. The results can be applied, in particular, to the ARX model (1.3) and (1.4). We would like to emphasize that, under the usual set of regularity conditions, the ARX model (1.3) and (1.4) is Markovian in nature and thus one can exploit this property in any direct analysis of the ARX system (1.3) and (1.4). On the other hand, the general additive nonlinear regression model (1.7) and (1.8), with which we are concerned, does not assume any Markovian structure, leading to a considerably more complex analysis.

The organization of the paper is as follows: In Section 2, we define the projection estimates. Preliminary results are given in Section 3. The weak consistency with rates is presented in Section 4, and the asymptotic normality is established in Section 5. Section 6 provides a discussion of the results of the paper. The derivations of the propositions and theorems stated in Sections 3, 4, and 5 are presented in Section 7. The Appendix contains the proofs of certain crucial lemmas needed in Section 4.

2. ESTIMATION PROCEDURE

Let

$$\underline{X}_l = (X_{l+j_1}, \ldots, X_{l+j_p}); \qquad \underline{Y}_l = (Y_{l+i_1}, \ldots, Y_{l+i_q}); \qquad \underline{Z}_l = (\underline{X}_l, \underline{Y}_l)$$

and

$$\underline{x} = (x_1, \ldots, x_p); \qquad \underline{y} = (y_1, \ldots, y_q); \qquad \underline{z} = (\underline{x}, \underline{y}).$$

Then the regression function (1.7) can be written as

$$m(\underline{x}, \underline{y}) = E\{\phi(Y_{l+i_{q+1}}) | \underline{X}_l = \underline{x}, \underline{Y}_l = \underline{y}\}. \tag{2.1}$$

We assume throughout the paper that the second order partial derivatives of $m(\underline{s})$ exist and are continuous at the point $\underline{z}$. We can approximate $m(\underline{s})$ locally by a multivariate polynomial at point $\underline{z}$ as follows:

$$m(\underline{s}) \approx \beta_0 + (\underline{s} - \underline{z})\underline{\beta}_1, \tag{2.2}$$

where $\beta_0 = m(\underline{z})$, $\underline{\beta}_1^T = \partial m(\underline{s})/\partial \underline{s}|_{\underline{s}=\underline{z}}$, and $\underline{\beta}_1^T$ denotes the transpose of $\underline{\beta}_1$ ($\underline{\beta}_1$ is a column vector). Note that $\underline{\beta}^T = (\beta_0, \underline{\beta}_1^T)$ depends on $\underline{z}$. Let $K(\underline{u})$ be a kernel function on $\Re^d$ with $d = p + q$ and $h = h_n$ be a bandwidth parameter. Given the observations $\{X_l, Y_l\}_{l=0}^n$, we consider the multivariate weighted least squares:

$$\sum_{l=0}^{n_1} \{\phi(Y_{l+i_{q+1}}) - \beta_0 - (\underline{Z}_l - \underline{z})\underline{\beta}_1\}^2 K_h(\underline{Z}_l - \underline{z}), \tag{2.3}$$

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

where $n_1 = n - 1 - i_{q+1}$, assumed to be positive, and $K_h(\underline{z}) = K(\underline{z}/h)/h^d$. Minimizing (2.3) with respect to β_0 and $\underline{\beta}_1$, one obtains estimates of β_0 and $\underline{\beta}_1$, respectively. The minimization of (2.3) leads to

$$\hat{\underline{\beta}}(\underline{z}) = \begin{pmatrix} \hat{\beta}_0(\underline{z}) \\ \hat{\underline{\beta}}_1(\underline{z}) \end{pmatrix} = \mathbf{Q}\mathbf{S}_n^{-1}(\underline{z})\underline{t}_n(\underline{z}), \tag{2.4}$$

where

$$\mathbf{Q} = \operatorname{diag}\{1, h_n^{-1}, \ldots, h_n^{-1}\}, \qquad \mathbf{S}_n = \mathbf{S}_n(\underline{z}) = \begin{pmatrix} s_{n,0}(\underline{z}) & \underline{s}_{n,1}^T(\underline{z}) \\ \underline{s}_{n,1}(\underline{z}) & \mathbf{S}_{n,2}(\underline{z}) \end{pmatrix},$$

and

$$\underline{t}_n(\underline{z}) = \begin{pmatrix} t_{n,0}(\underline{z}) \\ \underline{t}_{n,1}(\underline{z}) \end{pmatrix}$$

with

$$t_{n,0}(\underline{z}) = \frac{1}{n_1+1}\sum_{l=0}^{n_1} \phi(Y_{l+i_{q+1}})K_h(\underline{Z}_l - \underline{z}), \tag{2.5}$$

$$\underline{t}_{n,1}(\underline{z}) = \frac{1}{n_1+1}\sum_{l=0}^{n_1} \left(\frac{\underline{Z}_l - \underline{z}}{h}\right)^T \phi(Y_{l+i_{q+1}})K_h(\underline{Z}_l - \underline{z}), \tag{2.6}$$

$$s_{n,0}(\underline{z}) = \frac{1}{n_1+1}\sum_{l=0}^{n_1} K_h(\underline{Z}_1 - \underline{z}), \tag{2.7}$$

$$\underline{s}_{n,1}(\underline{z}) = \frac{1}{n_1+1}\sum_{l=0}^{n_1} \left(\frac{\underline{Z}_l - \underline{z}}{h}\right)^T K_h(\underline{Z}_l - \underline{z}), \tag{2.8}$$

and

$$\mathbf{S}_{n,2}(\underline{z}) = \frac{1}{n_1+1}\sum_{l=0}^{n_1} \left(\frac{\underline{Z}_l - \underline{z}}{h}\right)^T \left(\frac{\underline{Z}_l - \underline{z}}{h}\right) K_h(\underline{Z}_l - \underline{z}). \tag{2.9}$$

Therefore, our local linear estimator of $m(\underline{z})$ is

$$\hat{m}(\underline{z}) = \hat{\beta}_0(\underline{z}) = \underline{e}_1^T\hat{\underline{\beta}}(\underline{z}) = \underline{e}_1^T\mathbf{Q}\mathbf{S}_n^{-1}(\underline{z})\underline{t}_n(\underline{z}), \tag{2.10}$$

where $\underline{e}_1 = (1,0,\ldots,0)^T$.

We now employ the projection technique to estimate $g_1(\underline{y})$, essentially up to a constant factor (see (1.6)). To this end, let D_1 be a compact subset of $\Re^p$ and D_2 be a compact subset of $\Re^q$ and let $D = D_1 \times D_2$. Define the weighting function $w(\cdot,\cdot)$ to be continuous on $\Re^d$ such that $w(\underline{x},\underline{y}) \leq 1$ and

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

$$w(\underline{x},\underline{y}) = \begin{cases} 1 & \text{for } (\underline{x},\underline{y}) \in D_0 \\ 0 & \text{for } (\underline{x},\underline{y}) \notin D, \end{cases}$$

where $D_0 = D_{1,0} \times D_{2,0}$, and $D_{i,0}$ is a subset of D_i such that volume(D_i) − volume$(D_{i,0}) < \varepsilon$ for some $\varepsilon > 0$ for $i = 1,2$. The purpose of introducing a weighting scheme here is to make estimates efficient and to screen out extreme observations (see, e.g., Tjøstheim and Auestad, 1994a; Masry and Tjøstheim, 1997). Define the projection

$$P_{Y,w}(\underline{y}) = E\{m(\underline{X}_0,\underline{y})w(\underline{X}_0,\underline{y})\}. \tag{2.11}$$

We have

$$P_{Y,w}(\underline{y}) = g_1(\underline{y})E\{w(\underline{X}_0,\underline{y})\} + E\{g_2(\underline{X}_0)w(\underline{X}_0,\underline{y})\}.$$

Because

$$E\{w(\underline{X}_0,\underline{y})\} = \int_{D_{1,0}} f_{\underline{X}_0}(\underline{u})\,d\underline{u} + \int_{D_1 - D_{1,0}} w(\underline{u},\underline{y})\,f_{\underline{X}_0}(\underline{u})\,d\underline{u},$$

the first term on the right-hand side approaches one as $D_{1,0}$ becomes large, and the second term is small for small ε. Thus, we can identify $g_1(\underline{y})$ for $\underline{y} \in D_2$ up to a multiplicative and additive constant, and the multiplicative constant will be close to 1 when the support D of $w(\cdot,\cdot)$ is taken to be large enough.

We now estimate $P_{Y,w}(\underline{y})$ by

$$\hat{P}_{Y,w}(\underline{y}) = \frac{1}{n_2+1}\sum_{l=0}^{n_2} \hat{m}(\underline{X}_l,\underline{y})w(\underline{X}_l,\underline{y}), \tag{2.12}$$

where $n_2 = n - 1 - j_p$ and $\hat{m}(\underline{x},\underline{y})$ is given by (2.10). Similarly, we take the projection of the function $g_2(\underline{x})$ as

$$P_{X,w}(\underline{x}) = E\{m(\underline{x},\underline{Y}_0)w(\underline{x},\underline{Y}_0)\} = g_2(\underline{x})E\{w(\underline{x},\underline{Y}_0)\} + E\{g_1(\underline{Y}_0)w(\underline{x},\underline{Y}_0)\},$$

and we estimate $P_{X,w}(\underline{x})$ by

$$\hat{P}_{X,w}(\underline{x}) = \frac{1}{n_3+1}\sum_{l=0}^{n_3} \hat{m}(\underline{x},\underline{Y}_l)w(\underline{x},\underline{Y}_l),$$

where $n_3 = n - 1 - i_q$. In the sequel we focus our attention on $P_{Y,w}(\cdot)$ and $\hat{P}_{Y,w}(\cdot)$; similar arguments apply to $P_{X,w}(\cdot)$ and $\hat{P}_{X,w}(\cdot)$. Henceforth, we drop the subscript w from $P_{Y,w}(\cdot)$ and $\hat{P}_{Y,w}(\cdot)$.

3. PRELIMINARIES

Our goal in this section is to obtain a centered expression for the estimation error $\hat{P}_Y(\underline{y}) - P_Y(\underline{y})$ and an expression for the bias term. This requires an appropriate Taylor series expansion and also uniform convergence in probability of $\mathbf{S}_n(\underline{x},\underline{y})$. Details are given in Section 7. Here we only state the conditions and the result.

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

Let $F(\underline{x})$ be the distribution function of $\underline{X}_0$ and $\hat{F}(\underline{x})$ be the corresponding empirical distribution function $\hat{F}(\underline{x}) = (n_2 + 1)^{-1} \sum_{l=0}^{n_2} I(\underline{X}_l \leq \underline{x})$. We center the vector $\underline{t}_n(\underline{z})$ of (2.4) by

$$t_{n,0}^*(\underline{z}) = \frac{1}{n_1 + 1} \sum_{l=0}^{n_1} \delta_l K_h(\underline{Z}_l - \underline{z}) \tag{3.1}$$

and

$$\underline{t}_{n,1}^*(\underline{z}) = \frac{1}{n_1 + 1} \sum_{l=0}^{n_1} \left(\frac{\underline{Z}_l - \underline{z}}{h} \right)^T \delta_l K_h(\underline{Z}_l - \underline{z}), \tag{3.2}$$

where

$$\delta_l = \phi(Y_{l+i_{q+1}}) - m(\underline{Z}_l). \tag{3.3}$$

The following conditions are needed in the proof of Proposition 1.

Condition 1.

(i) The kernel $K(\cdot)$ is symmetric with $\int_{\Re^d} K(\underline{u})\, d\underline{u} = 1$, is bounded with compact support (say, $K(\underline{u}) = 0$ for $\|\underline{u}\| > 1$), and is factorable $K(\underline{u}) = K_1(\underline{u}')K_2(\underline{u}'')$, $\underline{u}' \in \Re^p$, $\underline{u}'' \in \Re^q$.
(ii) $E\{|\phi(Y_0)|\}^\nu < \infty$ for some $\nu > 2$.
(iii) The probability densities $f(\underline{u},\underline{v})$ and $f_{\underline{X}_0}(\underline{u})$ are continuous on D and D_1, respectively.
(iv) The conditional density

$$f_{(\underline{X}_0, \underline{Y}_0, \underline{X}_l, \underline{Y}_l)|(Y_{i_{q+1}}, Y_{l+i_{q+1}})} \leq A_3 < \infty$$

for all $l \geq 1$.
(v) The conditional density

$$f_{(\underline{X}_0, \underline{Y}_0)|Y_{i_{q+1}}} \leq A_4 < \infty.$$

(vi) The processes $\{X_l, Y_l\}$ are strongly mixing with $\sum_{l=1}^{\infty} l^a \{\alpha(l)\}^{1-2/\nu} < \infty$ for some $\nu > 2$ and $a > 1 - 2/\nu$.

Remark 1. We remark that $\underline{X}_0$ and $\underline{X}_l$ may overlap when $l \leq j_p - j_1$. The situation is similar for $\underline{Y}_0$ and $\underline{Y}_l$ when $l \leq i_q - i_1$. In Condition 1(iv) the joint density is meant to be that of the distinct random variables in the set $(\underline{X}_0, \underline{Y}_0, \underline{X}_l, \underline{Y}_l)$.

Condition 2. Let D be the compact support of the weight function $w(\underline{x}, \underline{y})$. Then

$$\inf_{(\underline{x}, \underline{y}) \in D} f(\underline{x}, \underline{y}) = A_2 > 0$$

and $m(\underline{x}, \underline{y})$ has continuous second-order partial derivatives on D.

Condition 3. The kernel $K_1(\cdot)$ has an integrable Fourier transform.

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

PROPOSITION 1. *Under Conditions* 1–3, *and* $nh_n^{p+q} \to \infty$ *and* $nh_n^{2p} \to \infty$, *we have*

$$\hat{P}_Y(\underline{y}) - P_Y(\underline{y}) - \mathrm{BIAS}(\underline{y}) = J_1(\underline{y}) + o_p((nh_n^q)^{-1/2}) + o_p(h_n^2) + O_p(n^{-1/2}), \tag{3.4}$$

where

$$J_1(\underline{y}) = \int_{\Re^p} \frac{w(\underline{x}, \underline{y})}{f(\underline{x}, \underline{y})}\, t^*_{n,0}(\underline{x}, \underline{y})\, d\hat{F}(\underline{x}), \tag{3.5}$$

$$\mathrm{BIAS}(\underline{y}) = \frac{h_n^2}{2} \int_{\Re^p} w(\underline{u}, \underline{y}) \mathrm{tr}(\mathbf{M}_{1,1} \mathbf{V}(\underline{u}, \underline{y}))\, dF(\underline{u}), \tag{3.6}$$

the moment matrix $\mathbf{M}$ *is given by*

$$\mathbf{M} = \begin{pmatrix} 1 & \underline{0}^T \\ \underline{0} & \displaystyle\int_{\Re^d} \underline{u}^T \underline{u} K(\underline{u})\, d\underline{u} \end{pmatrix} = \begin{pmatrix} 1 & \underline{0}^T \\ \underline{0} & \mathbf{M}_{1,1} \end{pmatrix}, \tag{3.7}$$

and $\mathbf{V} = \mathbf{V}(\underline{z}) = (\partial^2 m(\underline{z})/\partial \underline{z}^T \partial \underline{z})_{d \times d}$ *is the Hessian matrix of* $m(\underline{z})$.

It is conjectured that the additional assumption on the bandwidth ($nh_n^{2p} \to \infty$) made in Proposition 1 is not necessary but we were not able to dispense with it (see Remark A in the Appendix).

We remark that by (3.6) the bias term, represented by $\mathrm{BIAS}(\underline{y})$, is of order h_n^2 and is proportional to the integrals (with respect to $\underline{x}$) of the second-order partial derivatives $\mathbf{V}(\underline{x}, \underline{y})$ of the regression function $m(\underline{x}, \underline{y})$. Also note that the first term on the right-hand side of (3.6) is centered via $t^*_{n,0}(\underline{x}, \underline{y})$.

4. WEAK CONSISTENCY AND RATES

In this section we obtain the second-order properties of the term $J_1(\underline{y})$ on the right-hand side of (3.4) that lead to establishing the weak consistency of $\hat{P}_Y(\underline{y})$.

We first decompose $J_1(\underline{y})$ as the sum of two integrals:

$$\begin{aligned} J_1(\underline{y}) &= \int_{\Re^p} \frac{w(\underline{x}, \underline{y})}{f(\underline{x}, \underline{y})}\, t^*_{n,0}(\underline{x}, \underline{y})\, dF(\underline{x}) + \int_{\Re^p} \frac{w(\underline{x}, \underline{y})}{f(\underline{x}, \underline{y})}\, t^*_{n,0}(\underline{x}, \underline{y})\, d\{\hat{F}(\underline{x}) - F(\underline{x})\} \\ &\equiv J_{1,1}(\underline{y}) + J_{1,2}(\underline{y}). \end{aligned} \tag{4.1}$$

Define

$$H(\underline{x}, \underline{y}) = \frac{w(\underline{x}, \underline{y})}{f(\underline{x}, \underline{y})}\, f_{\underline{X}_0}(\underline{x}) \tag{4.2}$$

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

and assume that

$$K(\underline{u}) = K_1(\underline{u}')K_2(\underline{u}''), \qquad \underline{u}' \in \Re^p, \qquad \underline{u}'' \in \Re^q,$$

so that $K_h(\underline{u}) = K_{1,h}(\underline{u}')K_{2,h}(\underline{u}'')$ with $K_{1,h}(\underline{u}') = K_1(\underline{u}'/h)/h^p$ and $K_{2,h}(\underline{u}'') = K_2(\underline{u}''/h)/h^q$. Then, substituting $t^*_{n,0}(\underline{z})$ into $J_{1,1}(\underline{y})$, we obtain

$$J_{1,1}(\underline{y}) = \frac{1}{n_1+1}\sum_{l=0}^{n_1}\delta_l K_{2,h}(\underline{Y}_l - \underline{y})\left\{\int_{\Re^p} H(\underline{x},\underline{y})K_{1,h}(\underline{X}_l - \underline{x})\,d\underline{x}\right\}.$$

Let

$$H^*_n(\underline{u},\underline{y}) = \int_{\Re^p} H(\underline{x},\underline{y})K_{1,h}(\underline{u}-\underline{x})\,d\underline{x}. \tag{4.3}$$

Then

$$\begin{aligned} J_{1,1}(\underline{y}) &= \frac{1}{n_1+1}\sum_{l=0}^{n_1}\delta_l H(\underline{X}_l,\underline{y})K_{2,h}(\underline{Y}_l - \underline{y}) \\ &\quad + \frac{1}{n_1+1}\sum_{l=0}^{n_1}\delta_l\{H^*_n(\underline{X}_l,\underline{y}) - H(\underline{X}_l,\underline{y})\}K_{2,h}(\underline{Y}_l - \underline{y}) \\ &\equiv G_n(\underline{y}) + G'_n(\underline{y}). \end{aligned} \tag{4.4}$$

Hence

$$J_1(\underline{y}) = G_n(\underline{y}) + G'_n(\underline{y}) + J_{1,2}(\underline{y}). \tag{4.5}$$

We show that the contribution of $G'_n(\underline{y})$ and $J_{1,2}(\underline{y})$ is negligible relative to $G_n(\underline{y})$. The proofs of the following two lemmas are quite involved and are relegated to the Appendix.

LEMMA 1. *Under Conditions* 1 *and* 2, *and* $nh_n^{p+q} \to \infty$,

$$G'_n(\underline{y}) = o_p((nh_n^q)^{-1/2}).$$

LEMMA 2. *Under Conditions* 1–3 *and* $nh_n^{p+q} \to \infty$ *and* $nh_n^{2p} \to \infty$,

$$J_{1,2}(\underline{y}) = o_p((nh_n^q)^{-1/2})$$

at continuity points of the function $a^2(\underline{v},\underline{y})$ *defined in* (4.11) (*subsequently*) *as a function of* $\underline{v}$.

By (3.4), (4.5), and Lemmas 1 and 2, we have the following proposition.

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

PROPOSITION 2. *Under Conditions* 1–3 *and* $nh_n^{p+q} \to \infty$ *and* $nh_n^{2p} \to \infty$,

$$\hat{P}_Y(\underline{y}) - P_Y(\underline{y}) - \mathrm{BIAS}(\underline{y}) = G_n(\underline{y}) + o_p(h_n^2) + o_p((nh_n^q)^{-1/2}). \tag{4.6}$$

We now proceed to obtain the asymptotic variance of $G_n(\underline{y})$. By (4.4),

$$G_n(\underline{y}) = \frac{1}{n_1+1}\sum_{l=0}^{n_1} \pi_{n,l} \tag{4.7}$$

with

$$\pi_{n,l} = \pi_{n,l}(\underline{y}) = \delta_l H(\underline{X}_l, \underline{y}) K_{2,h}(\underline{Y}_l - \underline{y}). \tag{4.8}$$

Then,

$$(n_1+1)\mathrm{Var}(G_n(\underline{y})) = \mathrm{Var}(\pi_{n,0}) + 2\sum_{l=1}^{n_1}\left(1 - \frac{l}{n_1+1}\right)\mathrm{Cov}(\pi_{n,0}, \pi_{n,l}). \tag{4.9}$$

Define

$$\sigma^2(\underline{u}, \underline{v}) = \mathrm{Var}\{\phi(Y_{i_{q+1}}) | \underline{X}_0 = \underline{u}, \underline{Y}_0 = \underline{v}\}, \tag{4.10}$$

$$a^2(\underline{v}, \underline{y}) = \int_{\Re^p} H^2(\underline{u}, \underline{y}) \sigma^2(\underline{u}, \underline{v}) f(\underline{u}, \underline{v})\, d\underline{u}, \tag{4.11}$$

and

$$\|K_2\|_2^2 = \int_{\Re^p} \{K(\underline{u})\}^2\, d\underline{u}. \tag{4.12}$$

THEOREM 1. *Under Conditions* 1 *and* 2 *and* $nh_n^q \to \infty$, *we have*

(a)

$$h_n^q \mathrm{Var}(\pi_{n,0}) \to \|K_2\|_2^2 a^2(\underline{y}, \underline{y})$$

at continuity points of $a^2(\underline{v}, \underline{y})$ *as a function of* $\underline{v}$.

(b)

$$h_n^q \sum_{l=1}^{n_1} |\mathrm{Cov}(\pi_{n,0}, \pi_{n,l})| = o(1), \quad \text{as } n \to \infty.$$

(c)

$$nh_n^q \mathrm{Var}(G_n(\underline{y})) \to \|K_2\|_2^2 a^2(\underline{y}, \underline{y})$$

at continuity points of $a^2(\underline{v}, \underline{y})$ *as a function of* $\underline{v}$.

The proof of Theorem 1 is given in Section 7. We now state the weak consistency of the projection estimate $\hat{P}_Y(\underline{y})$ in the following theorem; its proof is given in Section 7.

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

THEOREM 2. *Under Conditions* 1–3, *and the bandwidth* h_n *satisfying* $nh_n^{p+q} \to \infty$, $nh_n^{2p} \to \infty$, $nh_n^{q+4} = O(\log n)$, *we have*

(a)

$$\left(\frac{nh_n^q}{\log n}\right)^{1/2} (\hat{P}_Y(\underline{y}) - P_Y(\underline{y}) - \mathrm{BIAS}(\underline{y})) \xrightarrow{P} 0, \quad \text{as } n \to \infty.$$

(b) *If, in addition,* $nh_n^{q+4} = o(\log n)$, *then*

$$\left(\frac{nh_n^q}{\log n}\right)^{1/2} (\hat{P}_Y(\underline{y}) - P_Y(\underline{y})) \xrightarrow{P} 0, \quad \text{as } n \to \infty.$$

Remark 2. Note that the rate of the convergence in Theorem 2 is the expected one given that $\underline{y} \in \Re^q$. The requirement that $nh_n^{p+q} \to \infty$ arises from the nature of the projection method, which utilizes the regression estimate of $m(\underline{x}, \underline{y})$. This condition was also required in Masry and Tjøstheim (1997) where a Nadaraya–Watson regression estimator was employed. The requirement that $nh_n^{2p} \to \infty$ was already discussed following the statement of Proposition 1.

5. ASYMPTOTIC NORMALITY

In this section, we establish the asymptotic normality of the projection estimate $\hat{P}_Y(\underline{y})$. We need the following condition on the mixing coefficient $\alpha(l)$.

Condition 4. Let $h_n \to 0$ such that $nh_n^q \to \infty$ as $n \to \infty$. Assume that there is a sequence $\{v_n\}$ of positive integers satisfying $v_n \to \infty$ and $v_n = o\left(\sqrt{nh_n^q}\right)$ such that $(n/h_n^q)^{1/2}\alpha(v_n) \to 0$.

Before we proceed with the statement of the asymptotic normality result (Theorem 3, which follows), we discuss a technical continuity requirement that is needed in its proof. It is due to the presence of the arbitrary transformation $\phi(\cdot)$, which requires a truncation argument to be employed. Specifically, for any $L > 0$, let

$$\tau_L(y) = yI(|y| \le L), \tag{5.1}$$

where L is a fixed truncation point. Put

$$m_L(\underline{x}, \underline{y}) = E\{\tau_L(\phi(Y_{i_{q+1}}))|\underline{X}_0 = \underline{x}, \underline{Y}_0 = \underline{y}\}, \tag{5.2}$$

$$\sigma_L^2(\underline{x}, \underline{y}) = \mathrm{Var}[\tau_L(\phi(Y_{i_{q+1}}))|\underline{X}_0 = \underline{x}, \underline{Y}_0 = \underline{y}], \tag{5.3}$$

$$\tilde{\sigma}_L^2(\underline{x}, \underline{y}) = \mathrm{Var}[\phi(Y_{i_{q+1}})I\{|\phi(Y_{i_{q+1}})| > L\}|\underline{X}_0 = \underline{x}, \underline{Y}_0 = \underline{y}], \tag{5.4}$$

$$a_L^2(\underline{v}, \underline{y}) = \int_{\Re^p} H^2(\underline{u}, \underline{y})\sigma_L^2(\underline{u}, \underline{v}) f(\underline{u}, \underline{v})\, d\underline{u}, \tag{5.5}$$

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

and

$$\tilde{a}_L^2(\underline{v},\underline{y}) = \int_{\Re^p} H^2(\underline{u},\underline{y})\,\tilde{\sigma}_L^2(\underline{u},\underline{v})\,f(\underline{u},\underline{v})\,d\underline{u}. \tag{5.6}$$

As in Theorem 1, where the quadratic-mean convergence holds at the continuity point $\underline{v} = \underline{y}$ of the function $a^2(\underline{v},\underline{y})$ (as a result of Bochner's lemma), the quadratic-mean convergence of the corresponding truncated variables requires that the functions $a_L^2(\underline{v},\underline{y})$ and $\tilde{a}_L^2(\underline{v},\underline{y})$ be continuous at the point $\underline{v} = \underline{y}$ for every $L > 0$. We state this technical continuity requirement as Condition 5 subsequently. Note that in view of the relationships (5.5) and (5.6), a sufficient condition is that the functions $m_L(\underline{u},\underline{v})$, $\sigma^2(\underline{u},\underline{v})$, and $\tilde{\sigma}^2(\underline{u},\underline{v})$, are continuous on $D_1(\underline{y}) \equiv D_1 \times \{\underline{y}\}$ for every $L > 0$ (see Lemma A in the Appendix for details).

Condition 5. For every $L > 0$, the functions $a_L^2(\underline{v},\underline{y})$ and $\tilde{a}_L^2(\underline{v},\underline{y})$ are continuous at the point $\underline{v} = \underline{y}$.

Condition 6. The bandwidth parameter h_n satisfies $h_n \to 0$, $nh_n^{p+q} \to \infty$, $nh_n^{q+4} = O(1)$, $nh_n^{2p} \to \infty$, as $n \to \infty$.

THEOREM 3. *Under Conditions* 1–6, *we have*, *as* $n \to \infty$,

$$(nh_n^q)^{1/2}(\hat{P}_Y(\underline{y}) - P_Y(\underline{y}) - \mathrm{BIAS}(\underline{y})) \xrightarrow{\mathcal{L}} N(0, \|K_2\|_2^2\, a^2(\underline{y},\underline{y}))$$

at continuity points $\underline{v} = \underline{y}$ *of* $a^2(\underline{v},\underline{y})$, *where the asymptotic variance* $a^2(\underline{y},\underline{y})$ *is given in* (4.11).

Remark 3. We provide a sufficient condition for the mixing coefficient $\alpha(l)$ to satisfy Conditions 1(vi) and 4. Suppose that $h_n = An^{-\theta}(0 < \theta < 1/q, A > 0)$, $v_n = (nh_n^q/\log n)^{1/2}$, and $\alpha(l) = O(l^{-c})$ for some $c > 0$. Then Condition 1(vi) is satisfied for $c > 2(1 - 1/\nu)/(1 - 2/\nu)$ and Condition 4 is satisfied if $c > (1 + \theta q)/(1 - \theta q)$. Hence both conditions are satisfied if

$$\alpha(l) = O(l^{-c}), \qquad c > \max\left\{\frac{1+\theta q}{1-\theta q}, \frac{2(1-1/\nu)}{1-2/\nu}\right\}.$$

Note that this is a trade-off between the order ν of the moment of $\phi(Y_0)$ and the rate of decay of the mixing coefficient; the larger the order ν, the weaker is the decay rate of $\alpha(l)$.

Remark 4. Theorem 3 shows that the projection estimate $\hat{P}_y(\underline{y})$ has the following asymptotic expressions for its bias and variance (of the asymptotic distribution)

$$\mathrm{bias}(\hat{P}_Y(\underline{y})) \approx \frac{h_n^2}{2}\int_{\Re^p} w(\underline{u},\underline{y})\,\mathrm{tr}(\mathbf{M}_{1,1}\mathbf{V}(\underline{u},\underline{y}))\,f_{\underline{X}_0}(\underline{u})\,d\underline{u} \equiv h_n^2\,\Gamma(\underline{y})$$

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

and

$$\mathrm{Var}(\hat{P}_Y(\underline{y})) \approx (nh_n^q)^{-1} \|K_2\|_2^2 a^2(\underline{y}, \underline{y})$$

with

$$a^2(\underline{y}, \underline{y}) = \int_{\Re^p} \frac{w^2(\underline{u}, \underline{y})}{f(\underline{u}, \underline{y})} f_{\underline{X}_0}^2(\underline{u}) \sigma^2(\underline{u}, \underline{y})\, d\underline{u}.$$

The asymptotically optimal bandwidth is then given by

$$h_n = \left(\frac{q\|K_2\|_2^2 a^2(\underline{y}, \underline{y})}{4\Gamma^2(\underline{y})} \right)^{1/(q+4)} \cdot \frac{1}{n^{1/(q+4)}},$$

so that the corresponding rate of the "mean-square convergence" is $O(n^{-4/(q+4)})$. Note however that Condition 6 imposes certain constraints on the bandwidth parameter h_n: in particular, the optimal bandwidth satisfies Condition 6 only if $p < \min(4,(q + 4)/2)$.

The issue of selecting the bandwidth in a data-driven fashion for our projection estimates remains open. We note that for local linear regression estimation with i.i.d. data, data-driven bandwidth selection was recently considered by Fan and Gijbels (1995) and Ruppert, Sheather, and Wand (1995). In the context of this paper, where the underlying processes $\{X_l, Y_l\}$ are strongly mixing, the problem is quite complex analytically.

6. DISCUSSION

We have employed local linear fitting along with the projection method to obtain the estimates of the components of additive nonlinear regression models of the form (1.8). The main results of the paper are Theorem 2, which establishes rates of convergence (in probability) for the projection estimates, and Theorem 3, which establishes the asymptotic normality of the projection estimates. Asymptotic expressions for the bias and variance (of the asymptotic distribution) are also given.

The results of the paper can be applied to the ARX model (1.3) and (1.4). We first note that in general the ARX model (1.3) and (1.4) need not be stationary without constraints on the growth of the functions $g_1(\underline{y})$ and $g_2(\underline{x})$ at infinity (Tjøstheim, 1990). Under Assumption 3.3 in Masry and Tjøstheim (1997) (reproduced in the Appendix as Condition A), it is shown in Lemma 3.1 of Masry and Tjøstheim (1997) that the ARX model (1.3) and (1.4) is stationary and strongly mixing with geometric decay, $\alpha(l) = e^{-al}$, $a > 0$. It then follows that the assumptions imposed on $\alpha(l)$ in Conditions 1(vi) and 4 are automatically satisfied for the ARX system (1.3) and (1.4) under Condition A. Assuming that the other technical conditions of this paper are satisfied, the weak consistency (Theorem 2) and the asymptotic normality (Theorem 3) results for

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

the general additive nonlinear regression model (1.7) and (1.8) hold, in particular, for the ARX model (1.3) and (1.4).

As was mentioned in the introduction, the key ideas of this paper are similar to those used in Masry and Tjøstheim (1997), except that we employ local linear fitting instead of local constant fitting. It is of interest to compare the two procedures for estimating the additive components although it is well known that the local linear fitting reduces the bias in the standard nonparametric regression context (see Fan and Gijbels, 1996, p. 16). In view of Theorem 3 and Corollary 4.2 in Masry and Tjøstheim (1997), it follows immediately that both procedures have the same asymptotic variance for the additive components. To see the gain for the bias using local linear fitting, we first derive the asymptotic bias for the Nadaraya–Watson estimator under Condition 7 (which follows), which is stated here as the following lemma; its proof is given in the Appendix. Note that the "random bias" for the Nadaraya–Watson estimator, given in Corollary 4.2 of Masry and Tjøstheim (1997), takes the form

$$A_n(\underline{y}) = \int_{\Re^p} w(\underline{x}, \underline{y}) C_n(\underline{x}, \underline{y})\, d\hat{F}(\underline{x}), \tag{6.1}$$

where

$$C_n(\underline{z}) = \frac{E\{t_{n,0}(\underline{z})\} - m(\underline{z}) f(\underline{z}) - m(\underline{z})[E\{s_{n,0}(\underline{z})\} - f(\underline{z})]}{E\{s_{n,0}(\underline{z})\}}. \tag{6.2}$$

Condition 7. $f(\underline{x}, \underline{y})$ has continuous second-order partial derivatives on D.

LEMMA 3. *Under Condition* 7, *the "random bias"* $A_n(\underline{y})$ *of* (6.1) *has the following asymptotic expression*:

$$A_n(\underline{y}) = \mathrm{BIAS}(\underline{y}) + h_n^2 \int_{\Re^p} \frac{w(\underline{x}, \underline{y}) f'(\underline{x}, \underline{y}) \mathbf{M}_{1,1} \{m'(\underline{x}, \underline{y})\}^T}{f(\underline{x}, \underline{y})} f(\underline{x})\, d\underline{x} + o_p(h_n^2), \tag{6.3}$$

where $\mathrm{BIAS}(\underline{y})$ *is given in* (3.6) *and* $f'(\underline{z})$ *and* $m'(\underline{z})$ *are vectors representing the partial derivatives of* $f(\underline{z})$ *and* $m(\underline{z})$, *respectively*.

Note that the first term $\mathrm{BIAS}(\underline{y})$ on the right-hand side of (6.3) is the asymptotic bias of the local linear estimator in the additive models. It follows from Lemma 3 that the Nadaraya–Watson estimator suffers from large bias, particularly in regions where the derivatives of the regression function $m(\underline{z})$ or the underlying density function $f(\underline{z})$ are large as a result of the second term on the right-hand side of (6.3). It can have a large bias even when the true regression function $m(\underline{z})$ is linear. The bias can also be large when $f'(\underline{z})/f(\underline{z})$ is large.

In this paper we assumed that the bandwidth parameter h_n is identical in all directions, i.e., $K_h(\underline{u}) = h^{-d} K(\underline{u}/h)$. It may be desirable to have distinct smoothing in different directions. This can be accomplished by replacing the

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

kernel $K_h(\underline{u}) = h^{-d}K(\underline{u}/h)$ by $|\mathbf{H}|^{-1}K(\mathbf{H}^{-1}\underline{u})$ where $\mathbf{H}$ is a symmetric positive definite bandwidth matrix and $|\mathbf{H}|$ is its determinant as was done in Robinson (1983) and Ruppert and Wand (1994). In particular, $\mathbf{H}$ can be diagonal $\mathbf{H} = \text{diag}\{h_1, \ldots, h_d\}$. The analysis will go through with the obvious modifications such as $n|\mathbf{H}| \to \infty$ replacing $nh^d \to \infty$.

This paper established the weak consistency and asymptotic normality of the components of additive nonparametric regression models of the form (1.8). We have not addressed the issue of establishing almost sure uniform convergence rates because of space limitations. However, in view of the results of Masry (1996b) for local polynomial regression, we expect that under appropriate set of regularity conditions, we have

$$\sup_{\underline{y}\in D_2} |\hat{P}_Y(\underline{y}) - P_Y(\underline{y}) - \text{BIAS}(\underline{y})| = O\left(\left(\frac{\ln n}{nh_n^q}\right)^{1/2}\right) \text{ almost surely}$$

and of course $\sup_{\underline{y}\in D_2}|\text{BIAS}(\underline{y})| = O(h_n^2)$. This issue will be pursued elsewhere.

As was noted in Remark 4, the problem of selecting the bandwidth h_n in date-driven fashion is quite complex under the general setting of this paper and remains open.

7. DERIVATIONS

In this section, we present the derivations of Proposition 1 and Theorems 1, 2, and 3.

We first note that by (2.10)–(2.12),

$$\hat{P}_Y(\underline{y}) = \int_{\Re^p} w(\underline{x}, \underline{y}) \underline{e}_1^T \hat{\underline{\beta}}(\underline{x}, \underline{y})\, d\hat{F}(\underline{x})$$

and

$$P_Y(\underline{y}) = \int_{\Re^p} w(\underline{x}, \underline{y}) \underline{e}_1^T \underline{\beta}(\underline{x}, \underline{y})\, dF(\underline{x}),$$

so that

$$\begin{aligned}\hat{P}_Y(\underline{y}) - P_Y(\underline{y}) &= \int_{\Re^p} w(\underline{x}, \underline{y}) \underline{e}_1^T \{\hat{\underline{\beta}}(\underline{x}, \underline{y}) - \underline{\beta}(\underline{x}, \underline{y})\}\, d\hat{F}(\underline{x}) \\ &\quad + \int_{\Re^p} w(\underline{x}, \underline{y}) m(\underline{x}, \underline{y})\, d\{\hat{F}(\underline{x}) - F(\underline{x})\} \\ &\equiv I_1 + I_2.\end{aligned}$$

The following lemma shows that I_2 is relatively negligible.

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

LEMMA 4. *Let the processes* $\{X_l, Y_l\}$ *be strongly mixing with mixing coefficient* $\alpha(l)$ *satisfying* $\sum_{l=1}^{\infty} \alpha(l) < \infty$. *Then,* $I_2 = O_p(n^{-1/2})$.

Proof. See Lemma 6.1 in Masry and Tjøstheim (1997). It follows that

$$\hat{P}_Y(\underline{y}) - P_Y(\underline{y}) = \int_{\Re^p} w(\underline{x}, \underline{y}) \underline{e}_1^T \{\hat{\underline{\beta}}(\underline{x}, \underline{y}) - \underline{\beta}(\underline{x}, \underline{y})\} \, d\hat{F}(\underline{x}) + O_p(n^{-1/2}). \quad \textbf{(7.1)}$$

To establish the convergence properties of the first term on the right-hand side of (7.1), we center the vector $\underline{t}_n(\underline{z})$ in (2.4) as in (3.1) and (3.2):

$$t^*_{n,0}(\underline{z}) = \frac{1}{n_1 + 1} \sum_{l=0}^{n_1} \delta_l K_h(\underline{Z}_l - \underline{z}) \quad \textbf{(7.2)}$$

and

$$\underline{t}^*_{n,1}(\underline{z}) = \frac{1}{n_1 + 1} \sum_{l=0}^{n_1} \left(\frac{\underline{Z}_l - \underline{z}}{h} \right)^T \delta_l K_h(\underline{Z}_l - \underline{z}). \quad \textbf{(7.3)}$$

Therefore,

$$t_{n,0}(\underline{z}) - t^*_{n,0}(\underline{z}) = \frac{1}{n_1 + 1} \sum_{l=0}^{n_1} m(\underline{Z}_l) K_h(\underline{Z}_l - \underline{z}) \quad \textbf{(7.4)}$$

and

$$\underline{t}_{n,1}(\underline{z}) - \underline{t}^*_{n,1}(\underline{z}) = \frac{1}{n_1 + 1} \sum_{l=0}^{n_1} \left(\frac{\underline{Z}_l - \underline{z}}{h} \right)^T m(\underline{Z}_l) K_h(\underline{Z}_l - \underline{z}). \quad \textbf{(7.5)}$$

Expanding $m(\underline{Z}_l)$ in a Taylor series around $\|\underline{Z}_l - \underline{z}\| \leq h$, because $m(\underline{z})$ has continuous second-order partial derivatives, we have

$$m(\underline{Z}_l) = \beta_0 + (\underline{Z}_l - \underline{z})\underline{\beta}_1 + \tfrac{1}{2}(\underline{Z}_l - \underline{z})\mathbf{V}(\underline{z})(\underline{Z}_l - \underline{z})^T + o_p(h_n^2), \quad \textbf{(7.6)}$$

where $\mathbf{V} = \mathbf{V}(\underline{z}) = (\partial^2 m(\underline{z})/\partial \underline{z}^T \partial \underline{z})_{d \times d}$. We rewrite the quadratic term as follows: Because $\mathbf{V}$ is symmetric, let $\text{vech}(\mathbf{V})$ be the $(d/2)(d+1)$ column vector representing the stacked up columns of $\mathbf{V}$ that are on and below the diagonal of $\mathbf{V}$. Then $\underline{a}^T \mathbf{V} \underline{a} = \text{vech}^T(2\mathbf{V} - \text{diag}(\mathbf{V}))\text{vech}(\underline{a}\underline{a}^T)$ for any vector $\underline{a}$. Substituting (7.6) in (7.2)–(7.5), we have

$$t_{n,0}(\underline{z}) - t^*_{n,0}(\underline{z}) = \beta_0(\underline{z}) s_{n,0}(\underline{z}) + h_n \underline{s}^T_{n,1}(\underline{z}) \beta_1(\underline{z}) + \frac{h_n^2}{2} \underline{b}^T_{n,1}(\underline{z}) \text{vech}(2\mathbf{V} - \text{diag}(\mathbf{V})) + o_p(h_n^2)$$

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

and

$$\underline{t}_{n,1}(\underline{z}) - \underline{t}^*_{n,1}(\underline{z}) = \beta_0(\underline{z})\underline{s}_{n,1}(\underline{z}) + h_n \mathbf{S}_{n,2}(\underline{z})\beta_1(\underline{z}) + \frac{h_n^2}{2}\mathbf{B}_{n,2}(\underline{z})\mathrm{vech}(2\mathbf{V} - \mathrm{diag}(\mathbf{V})) + o_p(h_n^2),$$

where

$$\underline{b}_{n,1}(\underline{z}) = \frac{1}{n_1+1}\sum_{l=0}^{n_1}\mathrm{vech}\left\{\left(\frac{\underline{Z}_l - \underline{z}}{h}\right)^T\left(\frac{\underline{Z}_l - \underline{z}}{h}\right)\right\}K_h(\underline{Z}_l - \underline{z})$$

and

$$\mathbf{B}_{n,2}(\underline{z}) = \frac{1}{n_1+1}\sum_{l=0}^{n_1}\left(\frac{\underline{Z}_l - \underline{z}}{h}\right)^T\mathrm{vech}^T\left\{\left(\frac{\underline{Z}_l - \underline{z}}{h}\right)^T\left(\frac{\underline{Z}_l - \underline{z}}{h}\right)\right\}K_h(\underline{Z}_l - \underline{z}).$$

Therefore,

$$\underline{t}_n(\underline{z}) - \underline{t}^*_n(\underline{z}) = \mathbf{S}_n(\underline{z})\mathbf{Q}^{-1}\underline{\beta}(\underline{z}) + \frac{h_n^2}{2}\mathbf{B}_n(\underline{z})\mathrm{vech}(2\mathbf{V} - \mathrm{diag}(\mathbf{V})) + o_p(h_n^2),$$

where

$$\mathbf{B}_n(\underline{z}) = \begin{pmatrix}\underline{b}^T_{n,1}(\underline{z})\\ \mathbf{B}_{n,2}(\underline{z})\end{pmatrix}. \tag{7.7}$$

By (2.4), we have

$$\underline{t}^*_n(\underline{z}) = \mathbf{S}_n(\underline{z})\mathbf{Q}^{-1}(\hat{\underline{\beta}}(\underline{z}) - \underline{\beta}(\underline{z})) - \frac{h_n^2}{2}\mathbf{B}_n(\underline{z})\mathrm{vech}(2\mathbf{V} - \mathrm{diag}(\mathbf{V})) + o_p(h_n^2),$$

so that

$$\hat{\underline{\beta}}(\underline{z}) - \underline{\beta}(\underline{z}) = \mathbf{Q}\mathbf{S}_n^{-1}(\underline{z})\underline{t}^*_n(\underline{z}) + \frac{h_n^2}{2}\mathbf{Q}\mathbf{S}_n^{-1}(\underline{z})\mathbf{B}_n(\underline{z})\mathrm{vech}(2\mathbf{V} - \mathrm{diag}(\mathbf{V})) + \mathbf{Q}\mathbf{S}_n^{-1}o_p(h_n^2). \tag{7.8}$$

Equation (7.8) is crucial to the analysis because it gives an expression for the estimation error $\hat{\underline{\beta}}(\underline{z}) - \underline{\beta}(\underline{z})$ in terms of a centered $\underline{t}^*_n(\underline{z})$ and a "bias" term. Substituting (7.8) into (7.1), we obtain

$$\begin{aligned}\hat{P}_Y(\underline{y}) - P_Y(\underline{y}) &= \int_{\Re^p} w(\underline{x},\underline{y})\underline{e}_1^T\mathbf{Q}\mathbf{S}_n^{-1}(\underline{x},\underline{y})\underline{t}^*_n(\underline{x},\underline{y})\,d\hat{F}(\underline{x})\\ &\quad + \frac{h_n^2}{2}\int_{\Re^p} w(\underline{x},\underline{y})\underline{e}_1^T\mathbf{Q}\mathbf{S}_n^{-1}(\underline{x},\underline{y})\mathbf{B}_n(\underline{x},\underline{y})\mathrm{vech}(2\mathbf{V} - \mathrm{diag}(\mathbf{V}))\,d\hat{F}(\underline{x})\\ &\quad + o_p(h_n^2)\int_{\Re^p} w(\underline{x},\underline{y})\underline{e}_1^T\mathbf{Q}\mathbf{S}_n^{-1}(\underline{x},\underline{y})\,d\hat{F}(\underline{x}) + O_p(n^{-1/2}).\end{aligned} \tag{7.9}$$

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

We now state the quadratic-mean convergence of the $(d+1)\times(d+1)$ stochastic matrix $\mathbf{S}_n(\underline{z})$ and of the $(d+1)\times(d/2)(d+1)$ stochastic matrix $\mathbf{B}_n(\underline{z})$.

LEMMA 5. *Under Condition* 1 *and* $nh_n^{p+q}\to\infty$, *we have*

$$\sup_{(\underline{x},\underline{y})\in\Re^{p+q}}|\mathbf{S}_n(\underline{x},\underline{y})-(\mathbf{M}f(\underline{x},\underline{y})|=o_p(1)$$

and

$$\sup_{(\underline{x},\underline{y})\in\Re^{p+q}}|\mathbf{B}_n(\underline{x},\underline{y})-(\mathbf{B}f(\underline{x},\underline{y})|=o_p(1),$$

where

$$\mathbf{M}_1=\begin{pmatrix}\displaystyle\int_{\Re^d}K(\underline{u})\,d\underline{u} & \displaystyle\int_{\Re^d}\underline{u}K(\underline{u})\,d\underline{u}\\ \displaystyle\int_{\Re^d}\underline{u}^TK(\underline{u})\,d\underline{u} & \displaystyle\int_{\Re^d}\underline{u}^T\underline{u}K(\underline{u})\,d\underline{u}\end{pmatrix}=\begin{pmatrix}m_{0,0} & \underline{m}_{1,0}^T\\ \underline{m}_{1,0} & \mathbf{M}_{1,1}\end{pmatrix} \tag{7.10}$$

and

$$\mathbf{B}=\begin{pmatrix}\operatorname{vech}^T\left(\displaystyle\int_{\Re^d}\underline{u}^T\underline{u}K(\underline{u})\,d\underline{u}\right)\\ \displaystyle\int_{\Re^d}\underline{u}^T\operatorname{vech}^T(\underline{u}^T\underline{u})K(\underline{u})\,d\underline{u}\end{pmatrix}. \tag{7.11}$$

Proof. See Proposition 1 and Theorem 1 in Masry (1996b).

Proof of Proposition 1. Observe that

$$\mathbf{S}_n^{-1}=\mathbf{S}^{-1}[\mathbf{I}+(\mathbf{S}-\mathbf{S}_n)\mathbf{S}_n^{-1}],$$

where $\mathbf{S}=\mathbf{M}_1f(\underline{x},\underline{y})$. By Lemma 5, $\mathbf{S}-\mathbf{S}_n\xrightarrow{P}0$ uniformly in $(\underline{x},\underline{y})\in\mathcal{R}^d$. By Lemma 5 and Condition 2, $\mathbf{S}_n^{-1}\xrightarrow{P}\mathbf{M}_1^{-1}/f(\underline{x},\underline{y})$ uniformly in $(\underline{x},\underline{y})\in D$ provided $\mathbf{M}_1^{-1}$ exists. It then follows that

$$\mathbf{S}_n^{-1}(\underline{x},\underline{y})=\mathbf{S}^{-1}(\underline{x},\underline{y})(1+o_p(1)), \tag{7.12}$$

where $o_p(1)$ is uniform in $(\underline{x},\underline{y})\in D$. Now substitute (7.12) into (7.9) to obtain

$$\begin{aligned}\hat{P}_Y(\underline{y})-P_Y(\underline{y})&\equiv\int_{\Re^p}w(\underline{x},\underline{y})\underline{e}_1^T\mathbf{Q}\mathbf{S}^{-1}(\underline{x},\underline{y})\underline{t}_n^*(\underline{x},\underline{y})\,d\hat{F}(\underline{x})\\&\quad+\frac{h_n^2}{2}\int_{\Re^p}w(\underline{x},\underline{y})\underline{e}_1^T\mathbf{Q}\mathbf{S}^{-1}(\underline{x},\underline{y})\mathbf{B}(\underline{x},\underline{y})f(\underline{x},\underline{y})\operatorname{vech}(2\mathbf{V}-\operatorname{diag}(\mathbf{V}))\,d\hat{F}(\underline{x})\\&\quad+o_p(h_n^2)\int_{\Re^p}w(\underline{x},\underline{y})\underline{e}_1^T\mathbf{Q}\mathbf{S}^{-1}(\underline{x},\underline{y})\,d\hat{F}(\underline{x})+O_p(n^{-1/2})+\Delta(\underline{y}),\end{aligned} \tag{7.13}$$

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

where the error term

$$\Delta(\underline{y}) = \int_{\Re^p} w(\underline{x},\underline{y})\underline{e}_1^T\mathbf{Q}[\mathbf{S}_n^{-1}(\underline{x},\underline{y}) - \mathbf{S}^{-1}(\underline{x},\underline{y})]\underline{t}_n^*(\underline{x},\underline{y})\, d\hat{F}(\underline{x})$$
$$+ \frac{h_n^2}{2}\int_{\Re^p} w(\underline{x},\underline{y})\underline{e}_1^T\mathbf{Q}[\mathbf{S}_n^{-1}(\underline{x},\underline{y})\mathbf{B}_n(\underline{x},\underline{y}) - \mathbf{S}^{-1}(\underline{x},\underline{y})\mathbf{B}(\underline{x},\underline{y})]f(\underline{x},\underline{y})$$
$$\times \operatorname{vech}(2\mathbf{V} - \operatorname{diag}(\mathbf{V}))\, d\hat{F}(\underline{x})$$
$$+ o_p(h_n^2)\int_{\Re^p} w(\underline{x},\underline{y})\underline{e}_1^T\mathbf{Q}[\mathbf{S}_n^{-1}(\underline{x},\underline{y}) - \mathbf{S}^{-1}(\underline{x},\underline{y})]\, d\hat{F}(\underline{x}). \qquad \textbf{(7.14)}$$

By assumption $\mathbf{M}_1 = \mathbf{M}$ and

$$\underline{e}_1^T\mathbf{Q}\mathbf{S}^{-1}(\underline{x},\underline{y}) = \frac{\underline{e}_1^T\mathbf{M}^{-1}}{f(\underline{x},\underline{y})} = \frac{\underline{e}_1^T}{f(\underline{x},\underline{y})}.$$

Thus,

$$\hat{P}_Y(\underline{y}) - P_Y(\underline{y}) = J_1(\underline{y}) + J_2(\underline{y}) + \Delta(\underline{y}) + o_p(h_n^2) + O_p(n^{-1/2}), \qquad \textbf{(7.15)}$$

where $J_1(\underline{y})$ is given in (3.5) and

$$J_2(\underline{y}) = \frac{h_n^2}{2}\int_{\Re^p} w(\underline{x},\underline{y})\operatorname{vech}^T(\mathbf{M}_{1,1})\operatorname{vech}(2\mathbf{V} - \operatorname{diag}(\mathbf{V}))\, d\hat{F}(\underline{x}). \qquad \textbf{(7.16)}$$

We consider the contribution of J_2. We note that $\operatorname{vech}^T(\mathbf{M}_{1,1})\operatorname{vech}(2\mathbf{V} - \operatorname{diag}(\mathbf{V})) = \operatorname{tr}(\mathbf{M}_{1,1}\mathbf{V}(\underline{x},\underline{y}))$. Then J_2 becomes

$$J_2 = \frac{h_n^2}{2}\frac{1}{n_2+1}\sum_{l=0}^{n_2} w(\underline{X}_l,\underline{y})\operatorname{tr}(\mathbf{M}_{1,1}\mathbf{V}(\underline{X}_l,\underline{y})). \qquad \textbf{(7.17)}$$

Because the summands in (7.17) are bounded random variables, then as in Lemma 4, we have

$$\frac{J_2}{h_n^2} \xrightarrow{P} \frac{1}{2}\int_{\Re^p} w(\underline{u},\underline{y})\operatorname{tr}(\mathbf{M}_{1,1}\mathbf{V}(\underline{u},\underline{y}))\, dF(\underline{u}). \qquad \textbf{(7.18)}$$

To complete the proof of the proposition, it suffices to show that $\Delta(\underline{y}) = o_p(h_n^2) + o_p((nh_n^q)^{-1/2})$. By (7.12), it is seen that the third term on the right-hand side of (7.14) is clearly $o_p(h_n^2)$; for the second term, observe that

$$\sup_{(\underline{x},\underline{y})\in D} |\mathbf{S}_n^{-1}(\underline{x},\underline{y})\mathbf{B}_n(\underline{x},\underline{y}) - \mathbf{S}^{-1}(\underline{x},\underline{y})\mathbf{B}(\underline{x},\underline{y})| = o_p(1)$$

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

component-wise and that the integrand is bounded on the compact domain of integration. Hence this term is also $o_p(h_n^2)$. It follows that

$$\Delta(\underline{y}) = o_p(1)\int_{\Re^p} w(\underline{x},\underline{y})|t^*_{n,0}(\underline{x},\underline{y})|\,d\hat{F}(\underline{x}) + o_p(h_n^2)$$
$$\equiv o_p(1)\Delta_1(\underline{y}) + o_p(h_n^2). \tag{7.19}$$

Under the assumptions in Proposition 1, it is shown in the Appendix, following the line of the analysis of the term $J_1(\underline{y})$ in Section 4, that

$$\Delta_1(\underline{y}) = O_p((nh_n^q)^{-1/2}). \tag{7.20}$$

This completes the proof of the proposition. ■

Proof of Theorem 1. Note that $E(\pi_{n,l}) = 0$. By (4.8), conditioning on $(\underline{X}_0, \underline{Y}_0)$, we have

$$h_n^q \operatorname{Var}(\pi_{n,0}) = h_n^q E\{H^2(\underline{X}_0,\underline{y})K_{2,h}^2(\underline{Y}_0 - \underline{y})\sigma^2(\underline{X}_0,\underline{Y}_0)\}$$
$$= \int_{\Re^q} \frac{1}{h_n^q} K_2^2\left(\frac{\underline{v}-\underline{y}}{h_n}\right) a^2(\underline{v},\underline{y})\,d\underline{v}.$$

By Bochner's lemma $h_n^q \operatorname{Var}(\pi_{n,0}) \to \|K_2\|_2^2 a^2(\underline{y},\underline{y})$ at continuity points of $a^2(\underline{v},\underline{y})$ as a function of $\underline{v}$ provided $a^2(\underline{v},\underline{y}) \in L_1(d\underline{v})$ which we proceed to show. It suffices to show that $E\{H^2(\underline{X}_0,\underline{y})\sigma^2(\underline{X}_0,\underline{Y}_0)\} < \infty$. By (4.2) and Conditions 1(iii) and 2

$$\sup_{\underline{x}\in\Re^p} H(\underline{x},\underline{y}) \le A_6 < \infty, \tag{7.21}$$

and $E\{\sigma^2(\underline{X}_0,\underline{Y}_0)\} < \infty$ because $E\{\phi^2(Y_0)\} < \infty$. For part (b), we decompose the sum into three terms as follows:

$$\sum_{l=1}^{n_1} = \sum_{l=1}^{m} + \sum_{l=m+1}^{c_n} + \sum_{l=c_n+1}^{n_1} \equiv J_7 + J_8 + J_9, \tag{7.22}$$

where $m = \max\{i_q - i_1, j_p - j_1\}$ and $c_n \to \infty$ such that $c_n h_n^q \to 0$ as $n \to \infty$. Clearly, there is an overlap of components in J_7 but not in J_8 or J_9. For J_7, let $\tilde{p}$ be the number of common elements in $(\underline{X}_0, \underline{X}_l)$ and $\tilde{q}$ be the number of common elements in $(\underline{Y}_0, \underline{Y}_l)$, $1 \le l \le m$. Then for $1 \le l \le m$, we have

$$\operatorname{Cov}(\pi_{n,0},\pi_{n,l}) = E[\delta_0\delta_l H(\underline{X}_0,\underline{y})K_{2,h}(\underline{Y}_0 - \underline{y})H(\underline{X}_l,\underline{y})K_{2,h}(\underline{Y}_l - \underline{y})].$$

Note that $H(\underline{u},\underline{y}) = 0$ for $\underline{u} \notin D_1$. Also K_2 has compact support. Because $m(\underline{u},\underline{v})$ is continuous, it follows that $m(\underline{u},\underline{v})$ is bounded over $\underline{u} \in D_1$ and $\{\underline{v} : \|\underline{v} - \underline{y}\| \le 1\}$. Let

$$A_7 = \sup_{\underline{u}\in D_1, \{\underline{v}:\|\underline{v}-\underline{y}\|\le 1\}} |m(\underline{u},\underline{v})|. \tag{7.23}$$

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

Conditioning on $(Y_{i_{q+1}}, Y_{l+i_{q+1}})$ and using Condition 1(iv), we obtain

$$|J_7| \le \text{const.}\, E[\{|\phi(Y_{i_{q+1}})| + A_7\}\{|\phi(Y_{l+i_{q+1}})| + A_7\}]$$
$$\times \int_{\Re^{2d-\bar{p}-\bar{q}}} H(\underline{u}', \underline{u}'', \underline{y}) H(\underline{u}'', \underline{u}''', \underline{y}) K_{2,h}(\underline{v}', \underline{v}'')$$
$$\times K_{2,h}(\underline{v}'', \underline{v}''')\, d\underline{u}'\, d\underline{u}''\, d\underline{u}'''\, d\underline{v}'\, d\underline{v}''\, d\underline{v}'''.$$

Here $\underline{u}'$ and $\underline{u}'''$ have dimensions $p - \tilde{p}$ and $\underline{u}''$ has dimensions $\tilde{p}$; $\underline{v}'$ and $\underline{v}'''$ have dimensions $q - \tilde{q}$ and $\underline{v}''$ has dimensions $\tilde{q}$. Therefore,

$$|J_7| \le \text{const.}\, h^{-\tilde{q}} E\{|\phi(Y_0)| + A_7\}^2 \int_{\Re^{2q-\bar{q}}} K_2(\underline{v}', \underline{v}'') K_2(\underline{v}'', \underline{v}''')\, d\underline{v}'\, d\underline{v}''\, d\underline{v}'''$$
$$\times \int_{\Re^{2p-\bar{p}}} H(\underline{u}', \underline{u}'', \underline{y}) H(\underline{u}'', \underline{u}''', \underline{y})\, d\underline{u}'\, d\underline{u}''\, d\underline{u}'''.$$

The integral over the K-functions is finite by Condition 1(i) and the integral over H-functions is finite because of the compact support of $w(\underline{x}, \underline{y})$ and (7.21). Therefore,

$$J_7 = O(h_n^{-\tilde{q}}).$$

Hence,

$$h_n^q J_7 = O(h_n^{q-\tilde{q}}) = o(1). \tag{7.24}$$

Next, we consider J_8 of (7.22). Here, there are no overlaps between $\underline{Z}_0$ and $\underline{Z}_l$. By conditioning on $(Y_{i_{q+1}}, Y_{l+i_{q+1}})$ and using Condition 1(iv) and (7.23),

$$|\text{Cov}(\pi_{n,0}, \pi_{n,l})|$$
$$\le E[(|\phi(Y_{i_{q+1}})| + A_7)(|\phi(Y_{l+i_{q+1}})| + A_7)$$
$$\times H(\underline{X}_0, \underline{y}) K_{2,h}(\underline{Y}_0 - \underline{y}) H(\underline{X}_l, \underline{y}) K_{2,h}(\underline{Y}_l - \underline{y})]$$
$$\le \text{const.}\, E[(|\phi(Y_{i_{q+1}})| + A_7)(|\phi(Y_{l+i_{q+1}})| + A_7)]$$
$$\times \int_{\Re^{2p+2q}} H(\underline{u}, \underline{y}) K_2(\underline{v}) H(\underline{u}', \underline{y}) K_2(\underline{v}')\, d\underline{u}\, d\underline{v}\, d\underline{u}'\, d\underline{v}'$$
$$\le \text{const.}\, E\{|\phi(Y_0)| + A_7\}^2 \left\{ \int_{\Re^p} H(\underline{u}, \underline{y})\, d\underline{u} \right\}^2 = O(1).$$

Thus,

$$J_8 = O(c_n - m) \quad \text{and} \quad h_n^q J_8 = O((c_n - m) h_n^q) = o(1) \tag{7.25}$$

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

as $n \to \infty$ by the choice of c_n. Finally, for J_9, it follows from Davydov's inequality (see, e.g., Hall and Heyde, 1980, Corollary A.2) that

$$|\mathrm{Cov}(\pi_{n,0}, \pi_{n,l})| \le 8\,[\alpha(\tilde{l})]^{1-2/\nu}\{E|\pi_{n,0}|^{\nu}\}^{2/\nu},$$

where $\tilde{l} = l + \min\{i_1, j_1\} - \max\{j_p, i_{q+1}\}$. Now, conditioning on $Y_{i_{q+1}}$, and using Condition 1(v) and (7.23), we have

$$E|\pi_{n,0}|^{\nu} \le A_4 E\{|\phi(Y_0)| + A_7\}^{\nu} \int_{\Re^p} H^{\nu}(\underline{u}, \underline{y})\, d\underline{u} \int_{\Re^q} K_{2,h}^{\nu}(\underline{v} - \underline{y})\, d\underline{v}$$

$$\le \text{const.} \int_{\Re^q} K_{2,h}^{\nu}(\underline{v} - \underline{y})\, d\underline{v}$$

by Condition 1(ii) and the compact support of $H(\underline{u}, \underline{y})$ with respect to $\underline{u}$. Therefore,

$$(E|\pi_{n,0}|^{\nu})^{2/\nu} \le \frac{\text{const.}}{h_n^{2q(1-1/\nu)}}.$$

Thus,

$$h_n^q |J_9| \le \frac{\text{const.}}{h_n^{q(1-2/\nu)}} \sum_{l=c_n+1}^{\infty} [\alpha(l)]^{1-2/\nu} \le \frac{\text{const.}}{h_n^{q(1-2/\nu)} c_n^a} \sum_{l=c_n+1}^{\infty} l^a [\alpha(l)]^{1-2/\nu}$$

and with $c_n = h_n^{-q(1-2/\nu)/a}$

$$h_n^q J_9 \to 0 \quad \text{as } n \to \infty \tag{7.26}$$

by Condition 1(vi). By (7.22) and (7.24)–(7.26), we have

$$h_n^q \sum_{l=0}^{n_1} |\mathrm{Cov}(\pi_{n,0}, \pi_{n,l})| = o(1),$$

which proves part (b). Part (c) follows from parts (a) and (b). ■

Proof of Theorem 2. By Theorem 1,

$$\left(\frac{nh_n^q}{\log n}\right)^{1/2} G_n(\underline{y}) \xrightarrow{P} 0, \quad \text{as } n \to \infty.$$

Part (a) follows from (4.6). Part (b) follows from part (a) and the fact that $\mathrm{BIAS}(\underline{x}, \underline{y}) = O(h_n^2)$ by (3.6). ■

Proof of Theorem 3. By (4.6) and Condition 6,

$$(nh_n^q)^{1/2}(\hat{P}_Y(\underline{y}) - P_Y(\underline{y}) - \mathrm{BIAS}(\underline{y})) = (nh_n^q)^{1/2} G_n(\underline{y}) + o_p(1),$$

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

so it suffices to show that

$$(nh_n^q)^{1/2} G_n(\underline{y}) \xrightarrow{\mathcal{L}} N(0, \|K_2\|_2^2 a^2(\underline{y}, \underline{y})).$$

Let

$$\Delta_{n,l} = h_n^{q/2} \pi_{n,l}, \tag{7.27}$$

where $\pi_{n,l}$ is defined in (4.8) and

$$U_n = \sum_{l=0}^{n-1} \Delta_{n,l}. \tag{7.28}$$

Then,

$$(nh_n^q)^{1/2} G_n(\underline{y}) = \left(\frac{n}{n_1 + 1} \right)^{1/2} \frac{1}{\sqrt{n_1 + 1}} U_{n - i_{q+1}}.$$

Hence it suffices to show that

$$\frac{1}{\sqrt{n}} U_n \xrightarrow{\mathcal{L}} N(0, \|K_2\|_2^2 a^2(\underline{y}, \underline{y})). \tag{7.29}$$

By Theorem 1, we have

$$\mathrm{Var}(\Delta_{n,0}) \to \|K_2\|_2^2 a^2(\underline{y}, \underline{y}), \quad \text{and} \quad \sum_{l=1}^{n-1} \mathrm{Cov}(\Delta_{n,0}, \Delta_{n,l}) = o(1). \tag{7.30}$$

Partition the set $\{0,1,\ldots,n-1\}$ into $2k_n + 1$ subsets with large blocks of size $u = u_n$ and small blocks of size $v = v_n$ where

$$k = k_n = \left\lfloor \frac{n}{u_n + v_n} \right\rfloor. \tag{7.31}$$

Define the random variables

$$\eta_j = \sum_{i=j(u+v)}^{j(u+v)+u-1} \Delta_{n,i}, \qquad 0 \le j \le k-1, \tag{7.32}$$

$$\xi_j = \sum_{i=j(u+v)-u}^{(j+1)(u+v)-1} \Delta_{n,i}, \qquad 0 \le j \le k-1, \tag{7.33}$$

and

$$\zeta_k = \sum_{k=k(u+v)}^{n-1} \Delta_{n,i}. \tag{7.34}$$

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

Write

$$U_n = \sum_{j=0}^{k-1} \eta_j + \sum_{j=0}^{k-1} \xi_j + \zeta_k = U_n' + U_n'' + U_n'''. \tag{7.35}$$

We show that, as $n \to \infty$,

$$\frac{1}{n} E(U_n'')^2 \to 0, \qquad \frac{1}{n} E(U_n''')^2 \to 0, \tag{7.36}$$

$$\left| E[\exp(itU_n')] - \prod_{j=0}^{k-1} E[\exp(it\eta_j)] \right| \to 0, \tag{7.37}$$

$$\frac{1}{n} \sum_{j=0}^{k-1} E(\eta_j^2) \to \|K_2\|_2^2 a^2(\underline{y}, \underline{y}), \tag{7.38}$$

and

$$\frac{1}{n} \sum_{j=0}^{k-1} E\{\eta_j^2 I(|\eta_j| > \varepsilon \|K_2\|_2 a(\underline{y}, \underline{y})\sqrt{n})\} \to 0 \tag{7.39}$$

for every $\varepsilon > 0$. Equation (7.36) implies that U_n'' and U_n''' are asymptotically negligible in probability; (7.37) shows that the summands $\{\eta_j\}$ in U_n' are asymptotically independent; and (7.38) and (7.39) are the standard Lindeberg–Feller conditions for asymptotic normality of U_n' for the independent setup.

We now prove (7.36)–(7.39). We first consider the choice of the large block size u_n. Condition 4 implies that there exist integers $\gamma_n \to \infty$ such that

$$\gamma_n v_n = o\left(\sqrt{nh_n^q}\right) \quad \text{and} \quad \gamma_n (n/h_n^q)^{1/2} \alpha(v_n) \to 0. \tag{7.40}$$

Now define the large block size u_n by $u_n = \lfloor (nh_n^q)^{1/2}/\gamma_n \rfloor$. Using (7.40) and simple algebra show that the following properties hold as $n \to \infty$,

$$v_n/u_n \to 0, \qquad u_n/n \to 0, \qquad u_n (nh_n^q)^{-1/2} \to 0, \tag{7.41}$$

$$(n/u_n)\alpha(v_n) \to 0. \tag{7.42}$$

We now establish (7.36).

$$E(U_n'')^2 = \sum_{j=0}^{k-1} \mathrm{Var}(\xi_j) + 2 \sum_{0 \le i < j \le k-1} \mathrm{Cov}(\xi_i, \xi_j) \equiv F_1 + F_2. \tag{7.43}$$

By stationarity and (7.30)

$$\begin{aligned} \mathrm{Var}(\xi_j) &= v_n \mathrm{Var}(\Delta_{n,0}) + 2v_n \sum_{i=1}^{v_n - 1} \left(1 - \frac{i}{v_n}\right) \mathrm{Cov}(\Delta_{n,0}, \Delta_{n,i}) \\ &= v_n \{\|K_2\|_2^2 a^2(\underline{y}, \underline{y}) + o(1)\}. \end{aligned} \tag{7.44}$$

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

Thus,

$$F_1 = k_n v_n \{\|K_2\|_2^2 a^2(\underline{y}, \underline{y}) + o(1)\} = o(n). \tag{7.45}$$

Next consider the term F_2 in (7.43). Write $r_j = j(u+v) + u$; we have

$$F_2 = \sum_{0 \le i \ne j \le k-1} \sum_{l_1=0}^{v-1} \sum_{l_2=0}^{v-1} \mathrm{Cov}(\Delta_{n, r_i + l_1}, \Delta_{n, r_j + l_2}), \tag{7.46}$$

but because $i \ne j$, $|r_j - r_i + l_1 - l_2| \ge u$ so that

$$|F_2| \le 2 \sum_{l_1=0}^{n-u-1} \sum_{l_2=l_1+u}^{n-1} |\mathrm{Cov}(\Delta_{n, l_1}, \Delta_{n, l_2})|.$$

By stationarity, (7.30), and $u_n \to \infty$,

$$|F_2| \le 2n \sum_{l=u}^{n-1} |\mathrm{Cov}(\Delta_{n,0}, \Delta_{n,l})| = o(n). \tag{7.47}$$

Hence by (7.43), (7.45), and (7.47), we have

$$\frac{1}{n} E(U_n'')^2 = o(1).$$

Using a similar argument, we find together with (7.30) and (7.41),

$$\frac{1}{n} E(U_n''')^2 \le \frac{n - k(u+v)}{n} \mathrm{Var}(\Delta_{n,0}) + 2 \sum_{l=1}^{n-1} |\mathrm{Cov}(\Delta_{n,0}, \Delta_{n,l})| = o(1). \tag{7.48}$$

To establish (7.37), we make use of the following lemma due to Volkonskii and Rozanov (1959).

LEMMA 6. *Let $V_1, \ldots, V_J$ be random variables measurable with respect to the σ-algebras $\mathcal{F}_{\mu_1}^{\rho_1}, \ldots, \mathcal{F}_{\mu_J}^{\rho_J}$, respectively, with $1 \le \mu_1 < \rho_1 < \mu_2 < \cdots < \rho_J \le n$, $\mu_{j+1} - \rho_j \ge \omega$, and $|V_j| \le 1$ for $j = 1, \ldots, J$. Then*

$$\left| E\left[\sum_{j=1}^{J} V_j\right] - \prod_{j=1}^{J} E[V_j] \right| \le 16(J-1)\alpha(\omega).$$

We note that by (4.8), (7.27), and (7.32) η_j is a function of the random variables $\{X_{j(u+v)+j_1}, \ldots, X_{j(u+v)+u+j_p-1}; Y_{j(u+v)+i_1}, \ldots, Y_{j(u+v)+u+i_{q+1}}-1\}$. Hence, we have

$$\left| E[\exp(itU_n')] - \prod_{j=0}^{k-1} E[\exp(it\eta_j)] \right| \le 16 k_n \alpha(\tilde{v}_n) \le \text{const.} \frac{n}{u_n} \alpha(v_n)$$

with $\tilde{v}_n = v_n + 1 \min(i_1, j_1) - \max(j_p, i_{q+1})$ and the right-hand side tends to zero by (7.42).

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

Next we establish (7.38). By stationarity and (7.44), with u_n replaced by v_n, we have

$$\mathrm{Var}(\eta_j) = \mathrm{Var}(\eta_0) = u_n\{\|K_2\|_2^2 a^2(\underline{y}, \underline{y}) + o(1)\},$$

so that

$$\frac{1}{n}\sum_{j=0}^{k-1} E(\eta_j)^2 = \frac{k_n u_n}{n}\{\|K_2\|_2^2 a^2(\underline{y}, \underline{y}) + o(1)\} \to \|K_2\|_2^2 a^2(\underline{y}, \underline{y}) \tag{7.49}$$

because $v_n/u_n \to 0$.

It remains to establish (7.39). We employ a truncation argument because $\phi(\cdot)$ is not necessarily a bounded function. Put

$$\Delta_{n,l}^L = h_n^{q/2}[\tau_L(\phi(Y_{l+i_{q+1}})) - m_L(\underline{Z}_l)]\, H(\underline{X}_l, \underline{y}) K_{2,h}(\underline{Y}_l - \underline{y}) \tag{7.50}$$

and

$$U_n^L = \sum_{l=0}^{n-1} \Delta_{n,l}^L, \qquad \tilde{U}_n^L = \sum_{l=0}^{n-1} (\Delta_{n,l} - \Delta_{n,l}^L). \tag{7.51}$$

Using the fact that $H(\underline{u}, \underline{y})$ is bounded (cf. (7.21)) and $K_2(\cdot)$ is bounded, we have

$$|\Delta_{n,l}^L| \le \frac{\text{const.}}{h_n^{q/2}}.$$

This implies by (7.30) that

$$n^{-1/2} \max_{1\le j\le k-1} |\eta_j^L| \le \frac{\text{const.}\, u_n}{\sqrt{n h_n^q}} \to 0$$

by (7.41). Therefore, the set $\{|\eta_j^L| > \varepsilon \|K_2\|_2 a_L(\underline{y}, \underline{y})\sqrt{n}\}$ becomes an empty set when n is sufficiently large and (7.39) holds for η_j^L. Consequently, we have established the following asymptotic normality:

$$\frac{1}{\sqrt{n}} U_n^L \xrightarrow{\mathcal{L}} N(0, \|K_2\|_2^2 a_L^2(\underline{y}, \underline{y})) \tag{7.52}$$

at continuity points $\underline{v} = \underline{y}$ of $a_L^2(\underline{v}, \underline{y})$. To complete the proof, namely, to establish (7.39) in the general case, it suffices to show that

$$\frac{1}{n}\mathrm{Var}(\tilde{U}_n^L) \to 0 \tag{7.53}$$

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

as first $n \to \infty$ and then $L \to \infty$. Observe that

$$
\begin{aligned}
&|E\exp(itn^{-1/2}U_n) - \exp(-\|K_2\|_2^2 a_L^2(\underline{y},\underline{y})t^2/2)| \\
&\quad \le |E\exp(itn^{-1/2}(U_n^L + \tilde{U}_n^L)) - \exp(-\|K_2\|_2^2 a_L^2(\underline{y},\underline{y})t^2/2)| \\
&\qquad + |\exp(-\|K_2\|_2^2 a_L^2(\underline{y},\underline{y})t^2/2) - \exp(-\|K_2\|_2^2 a^2(\underline{y},\underline{y})t^2/2)| \\
&\quad \le |E\exp(itn^{-1/2}U_n^L) - \exp(-\|K_2\|_2^2 a_L^2(\underline{y},\underline{y})t^2/2)| + E|\exp(itn^{-1/2}\tilde{U}_n^L) - 1| \\
&\qquad + |\exp(-\|K_2\|_2^2 a_L^2(\underline{y},\underline{y})t^2/2) - \exp(-\|K_2\|_2^2 a^2(\underline{y},\underline{y})t^2/2)|.
\end{aligned}
$$

As $n \to \infty$, the first term goes to zero by (7.52) for each $L > 0$; the second term converges to zero by (7.53), as first $n \to \infty$ and then $L \to \infty$; and the third term also goes to 0 as $L \to \infty$ by dominated convergence. Therefore, it remains to prove (7.53). Note that $\tilde{U}_n^L$ has the same structure as U_n except that $\phi(Y_{l+i_{q+1}})$ is replaced by $\phi(Y_{l+i_{q+1}})I\{\phi(Y_{l+i_{q+1}}) > L\}$. Hence, as in the proof of Theorem 1, using Condition 5, we have

$$\frac{1}{n}\operatorname{Var}(\tilde{U}_n^L) \to \|K_2\|_2^2 \tilde{a}_L^2(\underline{y},\underline{y})$$

as $n \to \infty$, where $\tilde{a}_L^2(\underline{v},\underline{y})$ is given in (5.6). It is clear that $\tilde{\sigma}_L^2(\underline{u},\underline{v})$ of (5.4) tends to zero as $L \to \infty$ for each $(\underline{u},\underline{v})$. Thus,

$$I_L(\underline{u},\underline{y}) = H^2(\underline{u},\underline{y})\tilde{\sigma}_L^2(\underline{u},\underline{y})f(\underline{u},\underline{y}) \to 0$$

as $L \to \infty$. Now

$$I_L(\underline{u},\underline{y}) \le H^2(\underline{u},\underline{y})E[\phi^2(Y_{i_{q+1}})|\underline{X}_0 = \underline{u}, \underline{Y}_0 = \underline{y}]f(\underline{u},\underline{y}),$$

and by (7.21),

$$I_L(\underline{u},\underline{y}) \le A_6^2 E[\phi^2(Y_{i_{q+1}})|\underline{X}_0 = \underline{u}, \underline{Y}_0 = \underline{y}]f(\underline{u},\underline{y})$$

and the right-hand side is integrable with respect to $\underline{u}$ and the integral is equal to

$$A_6^2 E[\phi^2(Y_{i_{q+1}})|\underline{Y}_0 = \underline{y}]f_{Y_0}(\underline{y}),$$

which is finite a.e.$(\underline{y})$. Hence, by dominated convergence, $\tilde{a}_L^2(\underline{y},\underline{y}) \to 0$ as $L \to \infty$. ■

REFERENCES

Auestad, B. & D. Tjøstheim (1991) Functional identification in nonlinear time series. In G.G. Roussas (ed.), *Nonparametric Functional Estimation and Related Topics* pp. 493–507. Amsterdam: Kluwer Academic.

Chen, R. & R. Tsay (1993) Nonlinear additive ARX models. *Journal of the American Statistical Association* 88, 310–320.

Cleveland, W.S. (1979) Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 74, 829–836.

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

Fan, J. (1992) Design-adaptive nonparametric regression. *Journal of the American Statistical Association* 87, 998–1004.

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

Fan, J. & I. Gijbels (1995) Data driven bandwidth selection in local polynomial fitting: Variable bandwidth spatial adaptation. *Journal of the Royal Statistical Society*, *Series B* 57, 371–394.

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

Fan, J., W. Härdle, & E. Mammen (1998) Direct estimation of low dimensional components in additive models. *Annals of Statistics* 26, 943–971.

Friedman, J. (1991) Multivariate adaptive regression splines (with discussion). *Annals of Statistics* 19, 1–141.

Granger, C.W.J. & T. Teräsvirta (1993) *Modeling Nonlinear Dynamic Relationships.* Oxford: Oxford University Press.

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

Lewis, P.A.W. & J.G. Stevens (1991) Nonlinear modeling of time series using multivariate adaptive splines (MARS). *Journal of the American Statistical Association* 86, 864–877.

Linton, O. (1997) Efficient estimation of additive nonparametric regression models. *Biometrika* 84, 469–473.

Linton, O. & W. Härdle (1996) Estimation of additive regression models with known links. *Biometrika* 83, 529–540.

Linton, O. & J.P. Nielsen (1995) A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* 82, 93–100.

Masry, E. (1996a) Multivariate regression estimation: Local polynomial fitting for time series. *Stochastic Processes and Their Applications* 65, 81–101.

Masry, E. (1996b) Multivariate local polynomial regression estimation for time series: Uniform strong consistency and rates. *Journal of Time Series Analysis* 17, 571–599.

Masry, E. & J. Fan (1997) Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics* 24, 165–179.

Masry, E. & D. Tjøstheim (1995) Nonparametric estimation and identification of nonlinear ARCH time series: Strong convergence and asymptotic normality. *Econometric Theory* 11, 258–289.

Masry, E. & D. Tjøstheim (1997) Additive nonlinear ARX time series and projection estimates. *Econometric Theory* 13, 214–252.

Pötscher, B.M. & I.R. Prucha (1991a) Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part I: Consistency and approximations concepts. *Econometric Review* 10, 125–216.

Pötscher, B.M. & I.R. Prucha (1991b) Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part II: Asymptotic normality. *Econometric Review* 10, 253–325.

Robinson, P.M. (1983) Nonparametric estimators for time series. *Journal of Time Series Analysis* 4, 185–297.

Ruppert, D., S.J. Sheather, & M.P. Wand (1995) An effective bandwidth selection for local least squares regression. *Journal of the American Statistical Association* 90, 1257–1270.

Ruppert, D., & M.P. Wand (1994) Multivariate weighted least squares regression. *Annals of Statistics* 22, 1346–1370.

Stone, C.J. (1977) Consistent nonparametric regression. *Annals of Statistics* 5, 595–645.

Stone, C.J. (1985) Additive regression and other nonparametric models. *Annals of Statistics* 13, 685–705.

Tjøstheim, D. (1990) Non-linear time series and Markov chains. *Advances in Applied Probability* 22, 587–611.

Tjøstheim, D. & B. Auestad (1994a) Nonparametric identification of nonlinear time series: Projections. *Journal of the American Statistical Association* 89, 1398–1409.

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

Tjøstheim, D. & Auestad (1994b) Nonparametric identification of nonlinear time series: Selecting significant lags. *Journal of the American Statistical Association* 89, 1410–1419.
Volkonskii, V.A. & Yu.A. Rozanov (1959) Some limit theorems for random functions. I. *Theory of Probability and Its Applications* 4, 178–197.

APPENDIX

In this section, we first present the proofs of Lemmas 1, 2, and 3 and (7.20). Also, we state a set of sufficient conditions for the nonlinear ARX model (1.3) and (1.4) to be stationary. Finally, we provide a sufficient condition for Condition 5.

Proof of Lemma 1. Let

$$\Delta'_{n,l} = \Delta'_{n,l}(\underline{y}) = h_n^{q/2}\delta_l\{H_n^*(\underline{X}_l, \underline{y}) - H(\underline{X}_l, \underline{y})\}K_{2,h}(\underline{Y}_l - \underline{y}); \tag{A.1}$$

then $E(\Delta'_{n,l}) = 0$ and

$$h_n^{q/2}G'_n(\underline{y}) = \frac{1}{n_1+1}\sum_{l=0}^{n_1}\Delta'_{n,l}.$$

so that

$$(n_1+1)h_n^q \operatorname{Var}(G'_n(\underline{y})) = \operatorname{Var}(\Delta'_{n,0}) + 2\sum_{l=1}^{n_1}\left(1 - \frac{l}{n_1+1}\right)\operatorname{Cov}(\Delta'_{n,0}, \Delta'_{n,l}). \tag{A.2}$$

By (A.2), (4.3), and Jensen's inequality,

$$\operatorname{Var}(\Delta'_{n,0}) = h_n^q E[\delta_0^2\{H_n^*(\underline{X}_0, \underline{y}) - H(\underline{X}_0, \underline{y})\}^2 K_{2,h}^2(\underline{Y}_0 - \underline{y})]$$

$$\leq h_n^q \int_{\Re^p} E[\delta_0^2\{H(\underline{X}_0 - h_n\underline{u}', \underline{y}) - H(\underline{X}_0, \underline{y})\}^2 K_{2,h}^2(\underline{Y}_0 - \underline{y})] \times K_1(\underline{u}')\, d\underline{u}'.$$

Note that $H(\underline{u}, \underline{y}) = 0$ for $\underline{u} \notin D_1$ and $K_1(\cdot)$ and $K_2(\cdot)$ have compact support. Because $m(\underline{u},\underline{v})$ is continuous, it follows that $m(\underline{u},\underline{v})$ is bounded over $D_1^* \times D_2^*$ with $D_1^* = \{\underline{u} : \underline{u} - \underline{u}' \in D_1;\ \|\underline{u}'\| \leq 1\} \cup D_1$ and $D_2^* = \{\underline{v} : \|\underline{v} - \underline{y}\| \leq 1\}$. Let

$$A_{10} = \sup_{\substack{\underline{u}\in D_1^* \\ \underline{v}\in D_2^*}} |m(\underline{u},\underline{v})|. \tag{A.3}$$

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

Conditioning on $Y_{i_{q+1}}$ and using Condition 1(v) and (A.3), we obtain

$$\begin{aligned}\mathrm{Var}(\Delta'_{n,0}) &\le A_4 h_n^q E\{|\phi(Y_0)| + A_{10}\}^2 \int_{\Re^{2p+q}} \{H(\underline{u} - h_n\underline{u}', \underline{y}) - H(\underline{u}, \underline{y})\}^2 K_{2,h}^2(\underline{y} - \underline{v}) \\ &\quad \times K_1(\underline{u}')\, d\underline{u}\, d\underline{u}'\, d\underline{v} \\ &\le \text{const.} \int_{\Re^{2p}} \{H(\underline{u} - h_n\underline{u}', \underline{y}) - H(\underline{u}, \underline{y})\}^2 K_1(\underline{u}')\, d\underline{u}\, d\underline{u}' \\ &= \text{const.} \int_{\|\underline{u}'\| \le 1} \int_{D_1^*} K_1(\underline{u}')\{H(\underline{u} - h_n\underline{u}', \underline{y}) - H(\underline{u}, \underline{y})\}^2\, d\underline{u}\, d\underline{u}'.\end{aligned}$$

By Condition 1(iii), $H(\underline{u}, \underline{y})$ is continuous in $\underline{u} \in D_1$ for any fixed $\underline{y}$. Then, by (7.21) and dominated convergence, we have

$$\mathrm{Var}(\Delta'_{n,0}) = o(1). \tag{A.4}$$

We decompose the sum on the right-hand side of (A.2) into three terms

$$\sum_{l=1}^{n_1} = \sum_{l=1}^{m} + \sum_{l=m+1}^{c_n} + \sum_{l=c_n+1}^{n_1} \equiv I'_1 + I'_2 + I'_3, \tag{A.5}$$

where $m = \max\{i_q - i_1, j_p - j_1\}$ and $c_n \to \infty$ such that $c_n h_n^q \to 0$ as $n \to \infty$. Clearly, there is an overlap of components in I'_1 but not in I'_2 and I'_3. For I'_1, by the Cauchy-Schwartz inequality and (A.4), we have

$$|\mathrm{Cov}(\Delta'_{n,0}, \Delta'_{n,l})| \le \mathrm{Var}(\Delta'_{n,0}) = o(1),$$

so that

$$I'_1 = o(1). \tag{A.6}$$

For I'_2,

$$\begin{aligned}&|\mathrm{Cov}(\Delta'_{n,0}, \Delta'_{n,l})| \\ &\le h_n^q \int_{\Re^{2p}} E[\{|\phi(Y_{i_{q+1}})| + |m(\underline{Z}_0)|\}\{|\phi(Y_{l+i_{q+1}})| + |m(\underline{Z}_l)|\} \\ &\qquad \times \{H(\underline{X}_0 - \underline{u}', \underline{y}) + H(\underline{X}_0, \underline{y})\}\{H(\underline{X}_l - \underline{u}'', \underline{y}) + H(\underline{X}_l, \underline{y})\} \\ &\qquad \times K_{2,h}(\underline{Y}_0 - \underline{y})K_{2,h}(\underline{Y}_l - \underline{y})] \\ &\qquad \times K_{1,h}(\underline{u}')K_{1,h}(\underline{u}'')\, d\underline{u}'\, d\underline{u}''.\end{aligned}$$

Conditioning on $(Y_{i_{q+1}}, Y_{l+i_{q+1}})$ and using Condition 1(iv) and (A.3), we obtain

$$\begin{aligned}|\mathrm{Cov}(\Delta'_{n,0}, \Delta'_{n,l})| &\le A_3 h_n^q E\{|\phi(Y_0)| + A_{10}\}^2 \int_{\Re^{4p+2q}} \{H(\underline{u}_3 - \underline{u}_1, \underline{y}) + H(\underline{u}_3, \underline{y})\} \\ &\quad \times \{H(\underline{u}_4 - \underline{u}_2, \underline{y}) + H(\underline{u}_4, \underline{y})\}K_{2,h}(\underline{v}_1 - \underline{y})K_{2,h}(\underline{v}_2 - \underline{y})K_{1,h}(\underline{u}_1)K_{1,h}(\underline{u}_2) \\ &\quad \times d\underline{u}_1\, d\underline{u}_2\, d\underline{u}_3\, d\underline{u}_4\, d\underline{v}_1\, d\underline{v}_2 \le \text{const.}\, h_n^q.\end{aligned} \tag{A.7}$$

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

Therefore,

$$I_2' = O(c_n h_n^q) \to 0 \tag{A.8}$$

by the choice of c_n. For I_3', we use Davydov's inequality to obtain

$$|\mathrm{Cov}(\Delta_{n,0}', \Delta_{n,l}')| \le 8[\alpha(\tilde{l})]^{1-2/\nu}(E|\Delta_{n,0}'|^\nu)^{2/\nu},$$

where $\tilde{l} = l + \min\{i_1, j_1\} - \max\{j_p, i_{q+1}\}$. By the C_r inequality,

$$E|\Delta_{n,0}'|^\nu \le 2^{\nu-1}h_n^{\nu q/2}E[\{|\phi(Y_{i_{q+1}})| + |m(\underline{Z}_0)|\}^\nu K_{2,h}^\nu(\underline{Y}_0 - \underline{y})\{(H_n^*(\underline{X}_0, \underline{y}))^\nu + H^\nu(\underline{X}_0, \underline{y})\}].$$

By Jensen's inequality,

$$\{H_n^*(\underline{X}_0, \underline{y})\}^\nu \le \int_{\Re^p} H^\nu(\underline{X}_0 - h_n\underline{u}', \underline{y})K_1(\underline{u}')\,d\underline{u}'.$$

Then,

$$E|\Delta_{n,0}'|^\nu \le 2^{\nu-1}h_n^{\nu q/2}\int_{\Re^p} E[\{|\phi(Y_{i_{q+1}})| + |m(\underline{Z}_0)|\}^\nu K_{2,h}^\nu(\underline{Y}_0 - \underline{y}) \times \{H^\nu(\underline{X}_0 - h_n\underline{u}', \underline{y}) + H^\nu(\underline{X}_0, \underline{y})\}]K_1(\underline{u}')\,d\underline{u}'.$$

Using the same bounding as in (A.7), we have

$$E|\Delta_{n,0}'|^\nu = O(h_n^{q(1-\nu/2)}).$$

Now we follow the same bounding in the proof of (7.26) to obtain

$$|I_3'| \le \text{const.}\, h_n^{-q(1-2/\nu)}c_n^{-a}\sum_{l=c_n}^{\infty} l^a[\alpha(l)]^{1-2/\nu},$$

and with $c_n = h_n^{-q(1-2/\nu)/a}$

$$I_3' = o(1). \tag{A.9}$$

By (A.5), (A.6), (A.8), and (A.9),

$$\sum_{l=1}^{n_1}|\mathrm{Cov}(\Delta_{n,0}', \Delta_{n,l}')| = o(1),$$

which, in conjunction with (A.2) and (A.4), implies that

$$nh_n^q\,\mathrm{Var}(G_n'(\underline{y})) = o(1).$$

This completes the proof of the lemma. ■

Proof of Lemma 2. Substituting

$$K_1(\underline{u}) = \int_{\Re^p} e^{i\underline{\lambda}\cdot\underline{u}}\tilde{K}_1(\underline{\lambda})\,d\underline{\lambda}$$

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

into $t^*_{n,0}(\underline{x},\underline{y})$ of (2.1), we obtain

$$t^*_{n,0}(\underline{x},\underline{y}) = \int_{\Re^p} I_{11}(\underline{y},\underline{\lambda})\,\frac{1}{h_n^p}\,e^{i\underline{\lambda}\cdot\underline{x}/h_n}\widetilde{K}_1(\underline{\lambda})\,d\underline{\lambda}, \tag{A.10}$$

where

$$I_{11}(\underline{y},\underline{\lambda}) = \frac{1}{n_1+1}\sum_{l=0}^{n_1}\delta_l\,e^{-i\underline{\lambda}\cdot\underline{X}_l/h_n}K_{2,h}(\underline{Y}_l-\underline{y}).$$

Substituting (A.10) into $J_{1,2}(\underline{y})$ of (4.1), we have

$$J_{1,2}(\underline{y}) = \int_{\Re^p} I_{11}(\underline{y},\underline{\lambda})I_{12}(\underline{y},\underline{\lambda})\widetilde{K}_1(\underline{\lambda})\,d\underline{\lambda}, \tag{A.11}$$

where

$$I_{12}(\underline{y},\underline{\lambda}) = \int_{\Re^p}\frac{w(\underline{u},\underline{y})}{f(\underline{u},\underline{y})}\,\frac{1}{h_n^p}\,e^{i\underline{\lambda}\cdot\underline{u}/h_n}\,d\{\hat{F}(\underline{u})-F(\underline{u})\}.$$

Writing $e^{-i\underline{X}_l\cdot\underline{\lambda}/h_n}=\cos(\underline{X}_l\cdot\underline{\lambda}/h_n)-i\sin(\underline{X}_l\cdot\underline{\lambda}/h_n)$, the real and imaginary parts can be treated separately. The expression $I_{11}(\underline{y},\underline{\lambda})$ can be analyzed step by step in the manner of the proof of part (c) of Theorem 1 to obtain

$$\sup_{\underline{\lambda}\in\Re^p} E\{|I^2_{11}(\underline{y},\underline{\lambda})|\} = O((nh_n^q)^{-1}). \tag{A.12}$$

By (6.61) in Masry and Tjøstheim (1997),

$$\sup_{\underline{\lambda}\in\Re^p} E\{|I^2_{12}(\underline{y},\underline{\lambda})|\} = O((nh_n^{2p})^{-1}). \tag{A.13}$$

By the Cauchy–Schwartz inequality, (A.11)–(A.13), and Condition 3,

$$\begin{aligned} E|J_{1,2}(\underline{y})| &\le \sup_{\underline{\lambda}\in\Re^p}[E\{|I^2_{11}(\underline{y},\underline{\lambda})|\}E\{|I^2_{12}(\underline{y},\underline{\lambda})|\}]^{1/2}\int_{\Re^p}|\widetilde{K}_1(\underline{\lambda})|\,d\underline{\lambda} \\ &= O((n^2h_n^{2p+q})^{-1/2}) = o((nh_n^q)^{-1/2}). \end{aligned} \tag{A.14}$$

This completes the proof of the lemma. ■

Remark A. It is seen that under the condition $nh_n^{2p}\to\infty$, we have that $J_{1,2}$ can be neglected compared to G_n. Whereas $O_p((nh_n^q)^{-1/2})$ is the best rate attainable for G_n, it is not at all clear that the rate $O_p((n^2h_n^{2p+q})^{-1/2})$ obtained in (A.14) is the best rate attainable and $J_{1,2}$. Intuitively, if we look at the expressions for G_n and $J_{1,2}$ in (4.4) and (4.1), it seems obvious that $J_{1,2}$ should always be of lower order than G_n and that its rate should be $O_p((n^2h_n^{p+q})^{-1/2})$, and we conjecture that this rate is attainable.

Proof of Lemma 3. First, we compute the expectations $E\{s_{n,0}(\underline{z})\}$ and $E\{t_{n,0}(\underline{z})\}$. To this end, expanding $f(\underline{s})$ in a Taylor series around $\|\underline{s}-\underline{z}\|\le h$, by Condition 7, we have

$$f(\underline{s}) = f(\underline{z}) + (\underline{s}-\underline{z})f'(\underline{z}) + \tfrac{1}{2}(\underline{s}-\underline{z})f''(\underline{z})(\underline{s}-\underline{z})^T + o(h_n^2),$$

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

where $f''(\underline{z})$ is the Hessian matrix of $f(\underline{z})$, $o(h_n^2)$ is uniform in $\|\underline{s} - \underline{z}\| \leq h_n$, and $\underline{z} \in D_1(\underline{y})$. Then, we have, uniformly in $\underline{z} \in D_1(\underline{y})$,

$$E\{s_{n,0}(\underline{z})\} = f(\underline{z}) + \frac{h_n^2}{2} \operatorname{tr}\{\mathbf{M}_{1,1} f''(\underline{z})\} + o(h_n^2) \tag{A.15}$$

and

$$E\{t_{n,0}(\underline{z})\} = \psi(\underline{z}) + \frac{h_n^2}{2} \operatorname{tr}\{\mathbf{M}_{1,1} \psi''(\underline{z})\} + o(h_n^2), \tag{A.16}$$

where $\psi(\underline{z}) = m(\underline{z}) f(\underline{z})$ and $\psi''(\underline{z})$ is the Hessian matrix of $\psi(\underline{z})$. Substituting (A.15) and (A.16) into (6.2) gives, uniformly in $\underline{z} \in D_1(\underline{y})$,

$$E\{t_{n,0}(\underline{z})\} - m(\underline{z}) f(\underline{z}) - m(\underline{z})[E\{s_{n,0}(\underline{z})\} - f(\underline{z})] = \frac{h_n^2}{2} f(\underline{z}) C(\underline{z}) + o(h_n^2)$$

and

$$C_n(\underline{z}) = \frac{h_n^2}{2} C(\underline{z}) + o(h_n^2), \tag{A.17}$$

where

$$C(\underline{z}) = \operatorname{tr}\{\mathbf{M}_{1,1} \mathbf{V}(\underline{z})\} + \frac{2}{f(\underline{z})} f'(\underline{z}) \mathbf{M}_{1,1} \{m'(\underline{z})\}^T.$$

Then, substituting (A.17) into (6.1), we obtain

$$A_n(\underline{y}) = \frac{h_n^2}{2} \int_{D_1} w(\underline{x}, \underline{y}) C(\underline{x}, \underline{y}) \, d\hat{F}(\underline{x}) + o_p(h_n^2). \tag{A.18}$$

Because $w(\underline{z})C(\underline{z})$ is bounded in $D_1(\underline{y})$, we have as in Lemma 4

$$\int_{D_1} w(\underline{x}, \underline{y}) C(\underline{x}, \underline{y}) \, d\hat{F}(\underline{x}) = \int_{D_1} w(\underline{x}, \underline{y}) C(\underline{x}, \underline{y}) f(\underline{x}) \, d\underline{x} + o_p(1), \tag{A.19}$$

which, in conjunction with (A.18), implies that

$$\begin{aligned} A_n(\underline{y}) &= \frac{h_n^2}{2} \int_{\Re^p} w(\underline{x}, \underline{y}) C(\underline{x}, \underline{y}) f(\underline{x}) \, d\underline{x} + o_p(h_n^2) \\ &= \text{BIAS}(\underline{y}) + h_n^2 \int_{\Re^p} \frac{w(\underline{x}, \underline{y}) f'(\underline{x}, \underline{y}) \mathbf{M}_{1,1} \{m'(\underline{x}, \underline{y})\}^T}{f(\underline{x}, \underline{y})} f(\underline{x}) \, d\underline{x} + o_p(h_n^2). \end{aligned} \tag{A.20}$$

This concludes the proof of the lemma. ■

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

Proof of (7.20). We only present the outline of the basic steps for the proof of (7.20) because the proof is similar to that of the term $J_1(\underline{y})$ in Section 4. We rewrite $\Delta_1(\underline{y})$ as

$$\Delta_1(\underline{y}) = \int_{\Re^p} w(\underline{x}, \underline{y}) c_n(\underline{x}, \underline{y}) t^*_{n,0}(\underline{x}, \underline{y})\, d\hat{F}(\underline{x}), \tag{A.21}$$

where $c_n(\underline{x}, \underline{y}) = \mathrm{sgn}(t^*_{n,0}(\underline{x}, \underline{y}))$ so that $|c_n(\underline{x}, \underline{y})| = 1$. Similar to (4.1), we decompose $\Delta_1(\underline{y})$ into two parts as follows:

$$\Delta_1(\underline{y}) \equiv \Delta_{11}(\underline{y}) + \Delta_{12}(\underline{y}),$$

where

$$\Delta_{11}(\underline{y}) = \int_{\Re^p} w(\underline{x}, \underline{y}) c_n(\underline{x}, \underline{y}) t^*_{n,0}(\underline{x}, \underline{y})\, dF(\underline{x})$$

and

$$\Delta_{12}(\underline{y}) = \int_{\Re^p} w(\underline{x}, \underline{y}) c_n(\underline{x}, \underline{y}) t^*_{n,0}(\underline{x}, \underline{y})\, d[\hat{F}(\underline{x}) - F(\underline{x})].$$

Then, substituting $t^*_{n,0}(\underline{z})$ into $\Delta_{11}(\underline{y})$, we obtain

$$\Delta_{11}(\underline{y}) = \frac{1}{n_1 + 1} \sum_{l=0}^{n_1} \delta_l K_{2,h}(\underline{Y}_l - \underline{y}) \tilde{H}^*_n(\underline{X}_l, \underline{y}),$$

where δ_l is defined in (3.3), and with $\tilde{H}_n(\underline{x}, \underline{y}) = w(\underline{x}, \underline{y}) c_n(\underline{x}, \underline{y})$,

$$\tilde{H}^*_n(\underline{u}, \underline{y}) = \int_{\Re^p} \tilde{H}_n(\underline{x}, \underline{y}) K_{1,h}(\underline{u} - \underline{x})\, d\underline{x}.$$

Then,

$$\sup_{(\underline{x}, \underline{y}) \in D} |\tilde{H}_n(\underline{x}, \underline{y})| \le 1 \quad \text{and} \quad \sup_{(\underline{x}, \underline{y}) \in D} |\tilde{H}^*_n(\underline{x}, \underline{y})| \le 1. \tag{A.22}$$

Similar to (4.4),

$$\begin{aligned}\Delta_{11}(\underline{y}) &= \frac{1}{n_1 + 1} \sum_{l=0}^{n_1} \delta_l K_{2,h}(\underline{Y}_l - \underline{y}) \tilde{H}_n(\underline{X}_l, \underline{y}) \\ &\quad + \frac{1}{n_1 + 1} \sum_{l=0}^{n_1} \delta_l K_{2,h}(\underline{Y}_l - \underline{y}) \{\tilde{H}^*_n(\underline{X}_l, \underline{y}) - \tilde{H}_n(\underline{X}_l, \underline{y})\} \\ &\equiv \Delta'_{11}(\underline{y}) + \Delta''_{11}(\underline{y}).\end{aligned}$$

Proceeding as in the proof of Theorem 1 by bounding the second moment of $\Delta'_{11}(\underline{y})$ and using (A.22), we find

$$\Delta'_{11}(\underline{y}) = O_p((nh^q_n)^{-1/2}).$$

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

Now using the facts that $\tilde{H}_n(\underline{u}, \underline{y}) = 0$ if $\underline{u} \notin D_1$ and $K_1(\cdot)$ and $K_2(\cdot)$ have compact support and following the same steps as in the proof of Lemma 1, we obtain

$$\Delta''_{11}(\underline{y}) = O_p((nh^q)^{-1/2}).$$

Therefore, we have

$$\Delta_{11}(\underline{y}) = O_p((nh^q)^{-1/2}).$$

Finally, for $\Delta_{12}(\underline{y})$, we follow the same arguments used in the proof of Lemma 2. As in (A.11),

$$\Delta_{12}(\underline{y}) = \int_{\Re^p} I_{11}(\underline{y}, \underline{\lambda}) I^*_{12}(\underline{y}, \underline{\lambda}) \tilde{K}_1(\underline{\lambda})\, d\underline{\lambda},$$

where

$$I^*_{12}(\underline{y}, \underline{\lambda}) = \int_{\Re^p} \tilde{H}_n(\underline{u}, \underline{y}) \frac{1}{h_n^p} e^{i\underline{\lambda}\cdot\underline{u}/h_n}\, d\{\hat{F}(\underline{u}) - F(\underline{u})\}.$$

By (6.61) in Masry and Tjøstheim (1997),

$$\sup_{\underline{\lambda}\in\Re^p} E\{|I^*_{12}(\underline{y}, \underline{\lambda})|^2\} = O((nh_n^{2p})^{-1}),$$

which, in conjunction with (A.12) and the Cauchy–Schwartz inequality, implies that

$$\begin{aligned} E|\Delta_{12}(\underline{y})| &\le \sup_{\underline{\lambda}\in\Re^p} [E\{|I^2_{11}(\underline{y}, \underline{\lambda})|\} E\{|I^*_{12}(\underline{y}, \underline{\lambda})|^2\}]^{1/2} \int_{\Re^p} |\tilde{K}_1(\underline{\lambda})\, d\underline{\lambda} \\ &= O((n^2 h_n^{2p+q})^{-1/2}) = o((nh_n^q)^{-1/2}). \end{aligned}$$

The proof is complete. ■

Condition A. The nonlinear ARX model (1.3) and (1.4) satisfies the following.

(i) The functions $g_1(\cdot)$, $g_2(\cdot)$, and $g_3(\cdot)$ are nonperiodic and bounded on compact sets, and $g_2(\underline{x}) = O(\|\underline{x}\|^{\gamma_1})$ as $\|\underline{x}\| \to \infty$ for some real γ_1.

(ii) The i.i.d. random variables $\{e_t\}$ ad $\{\varepsilon_t\}$ have probability density functions that are positive on $\Re^1$ and such that $E(|\varepsilon_t|^{\max(1, \gamma_1+\gamma_2)}) < \infty$ for some $\gamma_2 > 0$.

(iii) There exist column vectors $\underline{a} \in \Re^q$ and $\underline{c} \in \Re^{p-1}$ (each of which may be the zero vector)such that

$$g_1(\underline{y}) = \underline{y}\underline{a} + 0(\|\underline{y}\|) \quad \text{and} \quad g_3(\underline{x}) = \underline{x}\underline{c} + o(\|\underline{x}\|)$$

as $\|\underline{y}\|$ and $\|\underline{x}\| \to \infty$. Moreover, let $a'_{i_r+i_1+1} = a_r$, $r = 1, \ldots, q$, and $a'_j = 0$ otherwise, and similarly $c'_{j_r-j_1+1} = c_r$, $r = 1, \ldots, p-1$, and $c'_j = 0$ otherwise. Then the $(i_{q+1} - i_1)$ square matrix $\mathbf{A}$ defined by 0 if $\underline{a} = \underline{0}$, and by

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & \dots & 0 & a'_1 \\ 1 & 0 & \dots & 0 & a'_2 \\ 0 & 1 & \dots & 0 & a'_3 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & a'_{i_{q+1}-i_1} \end{pmatrix}$$

otherwise, and the $(j_p - j_1)$-dimensional square matrix $\mathbf{C}$ defined by

$$\mathbf{C} = \begin{pmatrix} 0 & 0 & \dots & 0 & c'_1 \\ 1 & 0 & \dots & 0 & c'_2 \\ 0 & 1 & \dots & 0 & c'_3 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & c'_{j_p-j_1} \end{pmatrix}$$

satisfy $\rho(\mathbf{A}) < 1$ and $\rho(\mathbf{C}) < 1$, where ρ denotes the spectral radius.

Next, we provide a sufficient condition for Condition 5, stated as the following lemma, and its justification is also included.

LEMMA A. *For any fixed point $L > 0$, assume that the functions $\sigma^2(\underline{u},\underline{v})$, $\sigma_L^2(\underline{u},\underline{v})$, and $m_L(\underline{u},\underline{v})$ are continuous on $D_1(\underline{y})$ and suppose Condition* 1(iii) *holds. Then $a^2(\underline{v},\underline{y})$, $a_L^2(\underline{v},\underline{y})$, and $\tilde{a}_L^2(\underline{v},\underline{y})$, as functions of $\underline{v}$, are continuous at the point $\underline{v} = \underline{y}$.*

Proof. Because $\sigma^2(\underline{u},\underline{v})$ is continuous on $D_1(\underline{y})$, then it is continuous on $D_1 \times N(\underline{y})$, where $N(\underline{y})$ is a neighborhood of $\underline{y}$. Therefore, $\sigma^2(\underline{u},\underline{v})$ is bounded on $D_1 \times N(\underline{y})$. Let

$$B_1 = \sup_{\underline{u}\in D_1, \underline{v}\in N(\underline{y})} \sigma^2(\underline{u},\underline{v}). \tag{A.23}$$

For any small $\underline{\delta}$, Condition 1(iii) and (A.23) imply that

$$\sup_{\underline{u}\in D_1} \{\sigma^2(\underline{u},\underline{y}+\underline{\delta}) f(\underline{u},\underline{y}+\underline{\delta})\} = B_2 < \infty.$$

By the boundedness of $H(\underline{u},\underline{y})$ and dominated convergence we obtain

$$a^2(\underline{y}+\underline{\delta},\underline{y}) = \int_{D_1} H^2(\underline{u},\underline{y})\sigma^2(\underline{u},\underline{y}+\underline{\delta}) f(\underline{u},\underline{y}+\underline{\eta})\, d\underline{u} \to a^2(\underline{y},\underline{y}) \tag{A.24}$$

as $\|\underline{\delta}\| \to 0$, which implies that $a^2(\underline{v},\underline{y})$, as a function of $\underline{v}$, is continuous at the point $\underline{v} = \underline{y}$. Analogously, it can be shown that $a_L^2(\underline{v},\underline{y})$ is continuous at the point $\underline{v} = \underline{y}$ because $\sigma_L^2(\underline{u},\underline{v})$ is continuous on $D_1(\underline{y})$. Some algebraic computations yield

$$\tilde{\sigma}_L^2(\underline{x},\underline{y}) = \sigma^2(\underline{x},\underline{y}) - \sigma_L^2(\underline{x},\underline{y}) + 2m_L(\underline{x},\underline{y})(m(\underline{x},\underline{y}) - m_L(\underline{x},\underline{y})),$$

which implies that

$$\tilde{a}_L^2(\underline{v},\underline{y}) = a^2(\underline{v},\underline{y}) - a_L^2(\underline{v},\underline{y}) + 2I^*(\underline{v},\underline{y}), \tag{A.25}$$

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press

where

$$I^*(\underline{v},\underline{y}) = \int_{\Re^p} H^2(\underline{u},\underline{y}) m_L(\underline{u},\underline{v})\{m(\underline{u},\underline{v}) - m_L(\underline{u},\underline{v})\} f(\underline{u},\underline{v})\, d\underline{u}. \quad \textbf{(A.26)}$$

Next, we wish to show that $\tilde{a}^2(\underline{v},\underline{y})$ is continuous at the point $\underline{v} = \underline{y}$. By (A.25) it suffices to show that $I^*(\underline{v},\underline{y})$ is continuous at $\underline{v} = \underline{y}$. Equation (A.26) is equal to

$$I^*(\underline{v},\underline{y}) = \int_{D_1} H^2(\underline{u},\underline{y}) m_L(\underline{u},\underline{v})\{m(\underline{u},\underline{v}) - m_L(\underline{u},\underline{v})\} f(\underline{u},\underline{v})\, d\underline{u}.$$

Because $m(\underline{u},\underline{v})$ and $m_L(\underline{u},\underline{v})$ are continuous on $D_1(\underline{y})$, the result follows in the manner of the proof of (A.24). ■

https://doi.org/10.1017/S0266466600164011 Published online by Cambridge University Press