Econometric Theory, **18**, 2002, 625–645. Printed in the United States of America.
DOI: 10.1017.S0266466602183034

SEMIPARAMETRIC ESTIMATION OF PARTIALLY LINEAR MODELS FOR DEPENDENT DATA WITH GENERATED REGRESSORS

QI LI
Texas A&M University

JEFFREY M. WOOLDRIDGE
Michigan State University

In this paper we consider the problem of estimating a semiparametric partially linear model for dependent data with generated regressors. This type of model comes naturally from various econometric models such as a semiparametric rational expectation model when the surprise term enters the model nonparametrically, or a semiparametric type-3 Tobit model when the error distributions are of unknown forms, or a semiparametric error correction model. Using the nonparametric kernel method and under primitive conditions, we show that the $\sqrt{n}$-consistent estimation results of the finite-dimensional parameter in a partially linear model can be generalized to the case of generated regressors with weakly dependent data. The regularity conditions we use are quite weak, and they are similar to those used in Robinson (1988, *Econometrica* 56, 931–954) for independent and observed data.

1. INTRODUCTION

Since Engle, Granger, Rice, and Weiss (1986), Robinson (1988), and Stock (1989), the partially linear model has attracted much attention among econometricians. See Ahn and Powell (1993), Ai and McFadden (1997), Andrews (1991, 1994), Chamberlain (1992), Donald and Newey (1994), Linton (1995), Fan and Li (1999a), Powell (1987), to mention only a few. For empirical applications of partially linear models in economics, see Engle et al. (1986), Stock (1989), and Stoker (1992), and Kniesner and Li (2002), among others. Most of the preceding works deal with the case with observed regressors, with the exceptions of Powell (1987), Ahn and Powell (1993), and Ai and McFadden (1997). Ahn and Powell (1993) and Powell (1987) consider the case in which the gen-

Helpful comments from two referees and Joel Horowitz greatly improved the paper. Li's research is supported by the Social Sciences and Humanities Research Council of Canada, the Natural Sciences Engineering Research Council of Canada, the Bush program in economics of public policy, and Private Enterprise Research Center, Texas A&M University. Address correspondence to: Qi Li, Department of Economics, Texas A&M University, College Station, TX 77845, USA.

© 2002 Cambridge University Press 0266-4666/02 $9.50

https://doi.org/10.1017/S0266466602183034 Published online by Cambridge University Press

erated regressors appear in the nonparametric part of the model, whereas Ai and McFadden (1997) consider the case in which the generated regressors appear in some known (possibly nonlinear) functions. Powell (1987), Ahn and Powell (1993), and Ai and McFadden (1997) all consider the independent data case. Under some high-level assumptions, Andrews (1994) establishes the $\sqrt{n}$-normality results of estimators from a wide class of semiparametric models. In this paper, using mild and primitive conditions, we generalize the $\sqrt{n}$-consistent estimation result of Robinson (1988) to time-series models with generated regressors.

There are many examples of econometric models that contain generated regressors. See Pagan (1984) for a general review on this issue and for the consequences of generated regressors on the estimation of parametric models. For time-series data, consider the following rational expectation model:

$$S_t = Z_t'\alpha + \eta_t, \quad \text{with } E(\eta_t|Z_t) = 0 \tag{1}$$

and

$$Y_t = X_t'\gamma + (S_t - E(S_t|Z_t))\beta + u_t, = X_t'\gamma + \eta_t\beta + u_t. \tag{2}$$

In (2), only the surprise term η_t, rather than S_t, has an impact on Y_t. However, the surprise term η_t may not enter the model linearly. One can let η_t enter the model nonparametrically (and additively). This results in a partially linear model with a generated regressor:

$$Y_t = X_t'\gamma + m(S_t - E(S_t|Z_t)) + u_t = X_t'\gamma + m(\eta_t) + u_t, \tag{3}$$

where $m(\cdot)$ is a smooth unknown function.

Another example in which generated regressors enter a regression model is an error correction model (ECM) (e.g., Granger, 1981; Engle and Granger, 1987). Engle and Granger (1987) consider the case in which the error term from a cointegrated system enters the short-run dynamic equation linearly. If one relaxes this linearity assumption and allows the error term to enter the short-run equation nonlinearly and nonparametrically, then one obtains a partially linear model with generated regressors. Using U.S. interest rate data, Bachmeier and Li (2002) find that a semiparametric ECM has smaller average squared prediction errors than that obtained from a linear ECM. This provides empirical evidence that a semiparametric ECM better describes the U.S. term structure data than the popular linear ECM.

A third example is a semiparametric type-3 Tobit model (usually with independent data). When the error distributions in a standard type-3 Tobit model are of unknown forms, it can be written as a partially linear model with the estimated errors from a selection equation entering the main equation nonparametrically, resulting in a partially linear model with a generated regressor in the nonparametric part. See Wooldridge (1998) for a detailed discussion on this.

https://doi.org/10.1017/S0266466602183034 Published online by Cambridge University Press

The paper is organized as follows. In Section 2 we show that the slope coefficient in a time-series semiparametric partially linear model with generated regressors can be $\sqrt{n}$-consistently estimated using the approach of Robinson (1988). Section 3 presents generalizations to the cases of a nonlinear regression model for η_t, a semiparametric type-3 Tobit model, and a semiparametric ECM. Section 4 discusses the semiparametric efficient bound for the independent data case. All the proofs appear in the two Appendixes.

2. THE MODEL AND THE MAIN RESULTS

Let $\{\mathcal{W}_t\} \equiv \{Y_t, X_t', S_t, Z_t'\}$ be a stationary, absolutely regular process. Namely, as $\tau \to \infty$,

$$\beta_\tau = \sup_{s \in \mathcal{N}} E\left[\sup_{A \in \mathcal{M}_{s+\tau}^{\infty}} \{|P(A|\mathcal{M}_{-\infty}^{s}(\mathcal{W})) - P(A)|\} \right] \to 0,$$

where $\mathcal{M}_s^t(\mathcal{W})$ denotes $\sigma(\mathcal{W}_s, \ldots, \mathcal{W}_t)$, the sigma algebra generated by $(\mathcal{W}_s, \ldots, \mathcal{W}_t)$, for $s \leq t$.

We consider the following partially linear model:

$$Y_t = X_t'\gamma + m(\eta_t) + u_t, \qquad E(u_t|X_t, Z_t, \eta_t) = 0, \qquad t = 1,2,\ldots \tag{4}$$

and

$$\eta_t = S_t - Z_t'\alpha, \qquad E(\eta_t|X_t, Z_t) = 0, \qquad t = 1,2,\ldots, \tag{5}$$

where X_t is of dimension $p \times 1$, Z_t is $q \times 1$, Y_t and S_t are scalars, γ and α are vectors of unknown parameters, and $m(\cdot)$ is an unknown smooth function. We will show in Section 3 that models (4) and (5) can be easily modified to cover the cases of a nonlinear regression model and a Tobit-3 model.

Following Robinson (1988) and Powell (1987), we will use the nonparametric kernel method to estimate γ. To avoid the random denominator problem in the kernel estimation, one can either estimate a density-weighted relationship (e.g., Powell, Stock, and Stoker, 1989) or introduce a trimming indicator function to trim out small values of estimated density function (e.g., Robinson, 1988). In this paper we will consider the density-weighted estimator in great detail and also present a conjecture on the result of using a trimming indicator function. To motivate our density-weighted estimator of γ, note that from (4), we have

$$(Y_t - E(Y_t|\eta_t))f_t = [(X_t - E(X_t|\eta_t))f_t]'\gamma + u_t f_t, \tag{6}$$

where $f_t = f(\eta_t)$ is the density function of η_t. Our semiparametric estimator of γ is the least squares estimator of γ based on (6) but with f_t, $E(Y_t|\eta_t)$, and $E(X_t|\eta_t)$ replaced by their kernel estimators: $\hat{f}_t$, $\hat{Y}_t$, and $\hat{X}_t$, respectively, where

https://doi.org/10.1017/S0266466602183034 Published online by Cambridge University Press

$$\hat{f}_t = \frac{1}{nh} \sum_{s=1, s\neq t}^{n} \hat{K}_{st}, \tag{7}$$

$$\hat{Y}_t = \hat{E}(Y_t|\eta_t) = \frac{1}{nh} \sum_{s=1, s\neq t}^{n} Y_s \hat{K}_{st}/\hat{f}_t, \tag{8}$$

$$\hat{X}_t = \hat{E}(X_t|\eta_t) = \frac{1}{nh} \sum_{s=1, s\neq t}^{n} X_s \hat{K}_{st}/\hat{f}_t, \tag{9}$$

where $\hat{K}_{st} = K((\hat{\eta}_s - \hat{\eta}_t)/h)$ is the kernel function, $h = h_n$ is the smoothing parameter, $\hat{\eta}_t = S_t - Z_t'\hat{\alpha}$, and $\hat{\alpha}$ is a $\sqrt{n}$-consistent estimator of α. See condition (C2) for the definition of $\hat{\alpha}$.

Our proposed estimator of γ is the least squares regression of $(Y_t - \hat{Y}_t)\hat{f}_t$ on $(X_t - \hat{X}_t)\hat{f}_t$:

$$\begin{aligned}\hat{\gamma} &= S^{-1}_{(X-\hat{X})\hat{f}} S_{(X-\hat{X})\hat{f},(Y-\hat{Y})\hat{f}} \\ &= \gamma + S^{-1}_{(X-\hat{X})\hat{f}} S_{(X-\hat{X})\hat{f},(m-\hat{m})\hat{f}+U\hat{f}-\hat{U}\hat{f}},\end{aligned} \tag{10}$$

where we used, as in Robinson (1988), the notation that for scalar or column-vector sequences A_t and B_t, $S_{A\hat{f},B\hat{f}} = (1/n)\sum_{t=1}^{n} A_t \hat{f}_t B_t' \hat{f}_t$ and $S_{A\hat{f}} = S_{A\hat{f},A\hat{f}}$. Also $\hat{m}(\eta_t)$ and $\hat{u}_t$ are defined as in the right-hand side of (8) with Y_s replaced by $m(\eta_s)$ and u_s, respectively.

To establish the $\sqrt{n}$-consistency and asymptotic normality of $\hat{\gamma}$ given in (10), we first need a definition and some assumptions. The following definition is from Robinson (1988).

DEFINITION 2.1. *$\mathcal{G}_\mu^\alpha$, $\alpha > 0$, $\mu > 0$, is the class of functions g: $R^d \to R$ satisfying the following conditions: g is $(l-1)$-times partially differentiable, for $l - 1 \leq \mu \leq l$; for some $\rho > 0$, $\sup_{y\in\phi_{z\rho}} |g(y) - g(z) - Q_g(y,z)|/|y - z|^\mu \leq G_g(z)$ for all z, where $\phi_{z\rho} = \{y : |y - z| < \rho\}$; $Q_g = 0$ when $l = 1$; Q_g is a $(l-1)$th degree homogeneous polynomial in $y - z$ with coefficients the partial derivatives of g at z of orders 1 through $l - 1$ when $l > 1$; and $g(z)$, its partial derivatives of order $l - 1$ and less, and $G_g(z)$ have finite αth moments.*

Let $\xi = \xi(\eta) = E(X|\eta)$ and $V = X - \xi$. Also $\|\cdot\|$ denotes the Euclidean norm. The following conditions will be used to establish the $\sqrt{n}$-consistency of $\hat{\gamma}$.

(C1)

(i) $\{\mathcal{W}_t\} = \{Y_t, X_t', S_t, Z_t'\}$, $t = 1,2,\ldots$, is a strictly stationary absolutely regular process with coefficient β_τ, satisfying $\beta_\tau^{\delta/(2+\delta)} = O(\tau^{-2+\rho})$ for some $1 > \rho > 0$, $\frac{1}{2} \geq \delta > 0$. In addition, $\{Y_t, X_t', S_t, Z_t'\}$ have the same marginal distribution as $\{Y, X', S, Z'\} = \{\mathcal{W}\}$.

(ii) $E(u_t|X_t, Z_t, \eta_t, \mathcal{M}_{-\infty}^{t-1}(\mathcal{W})) = 0$ and $E(\eta_t|Z_t, \mathcal{M}_{-\infty}^{t-1}(\mathcal{W})) = 0$.

https://doi.org/10.1017/S0266466602183034 Published online by Cambridge University Press

(iii) X_t has finite $4(1+\delta)$th moment, and Z_t has finite $8(1+\delta)$th moment, $f(\cdot) \in \mathcal{G}_\nu^\infty$, $f_w(\cdot) \in \mathcal{G}_\nu^\infty$, where $f_w(\cdot)$ is the joint density function of $W_t = (\eta_t, X_t, Z_t)$, $m(\cdot) \in \mathcal{G}_\nu^4$, and $\xi(\cdot) \in \mathcal{G}_\nu^4$, where $\nu \geq 2$ is an integer. Also f, f_w, m, and ξ all satisfy some (global) Lipschitz-type condition: $|g(y) - g(z)| \leq D_g(z)\|y - z\|$ for all y and z, where $D_g(z)$ has finite fourth moments and $g = f$, or f_w, or m, or ξ.

(iv) Let $U = Y - X'\gamma - m(\eta)$. Here $E(|U|^{4(1+\delta)}) < \infty$, $E(\|V\|^{4(1+\delta)}) < \infty$, and $E(|\eta|^{4(1+\delta)}) < \infty$. Further, both $\sigma_u^2(x,w) = E(U^2|X = x, \eta = w)$ and $\sigma_\eta^2(z) = E(\eta^2|Z = z)$ belong to $\mathcal{G}_1^2$.

(C2) $n^{1/2}(\hat{\alpha} - \alpha) = n^{-1/2}\sum_{t=1}^n r_t\,\eta_t + o_p(1) \to N(0, \Sigma_{22})$ in distribution, where $r_t = r(Z_t)$ and $\Sigma_{22} = E[\eta_t^2\, r_t\, r_t']$.

(C3)

(i) $K: \mathcal{R} \to \mathcal{R}$, is a nonnegative and bounded function, satisfying $\int K(u)du = 1$, $\int K(u)udu = 0$, and $\int K(u)u^2\,du > 0$ and finite.

(ii) $K(\cdot)$ has second-order continuous derivatives; $K(\cdot)$ and its derivatives (up to second order) are all bounded. Moreover $K(\cdot)$ satisfies a (global) Lipschitz-type condition: $|K(u+v) - K(u) - K^{(1)}(u)v| \leq \frac{1}{2}H(u)v^2$ for all $u, v \in \mathcal{R}$, where $K^{(1)}(u) = \partial K(u)/\partial u$; $H(\cdot)$ is a bounded function. To simplify the notation, we also assume $K(u) \leq H(u)$ and $K^{(1)}(u) \leq H(u)$.

(C4) As $n \to \infty$, $nh^4 \to \infty$, $nh^8 \to 0$.

Condition (C1)(i) states that the underlying process is absolutely regular with some mild restrictions on the decay rate of the mixing coefficient β_τ. Condition (C1)(ii) is a martingale difference assumption for the error terms. Conditions (C1)(iii) and (iv) are some mild smoothness and moments conditions. Condition (C2) assumes that $\hat{\alpha}$ is a $\sqrt{n}$-consistent estimator of α. It could be a least squares estimator when (4) and (5) are standard regression models. In this case we have $r_t = E[Z_t Z_t']^{-1} Z_t$. See Section 3 for extensions to the case of a nonlinear regression model and a type-3 Tobit model, where $\hat{\alpha}$ could be a nonlinear least squares estimator or a censored least absolute deviation (LAD) estimator. Condition (C3)(i) assumes that $K(\cdot)$ is a nonnegative second-order kernel function, and Condition (C3)(ii) further imposes some smoothness conditions on $K(\cdot)$. Condition (C4) imposes some restrictions on the smoothing parameter h; it comes from the requirement that the mean square error of the nonparametric kernel estimate is of smaller order than $n^{-1/2}$ and the second term Taylor expansion of $\hat{K}_{st}$ is of smaller order than $n^{-1/2}$; i.e., from $O(h^4 + (nh)^{-1}) = o(n^{-1/2})$ and $[(\hat{\alpha} - \alpha)'(\hat{\alpha} - \alpha)/h^2] = O_p((nh^2)^{-1}) = o_p(n^{-1/2})$, we get $nh^4 \to 0$ and $nh^8 \to \infty$. Condition (C4) can be relaxed to $nh^{2+\epsilon} \to \infty$ and $nh^8 \to 0$ for some arbitrarily small $\epsilon > 0$, at the cost of strengthening some of the moments conditions in (C1) and a more tedious proof.

Compared with the conditions in Robinson (1988), we impose slightly stronger moment and smoothness conditions on the regression function, the density function, and the kernel function. This is because in the proof we use Taylor

https://doi.org/10.1017/S0266466602183034 Published online by Cambridge University Press

expansions to deal with the generated regressor $\hat{\eta}_t$. To prove that various terms in the Taylor expansion are of the desired orders, we need slightly stronger conditions than in the case with observable and independent data (e.g., Robinson, 1988).

The next theorem gives the main result of this paper.

THEOREM 2.1. *Under (C1)–(C4), we have* $\begin{pmatrix} \sqrt{n}(\hat{\gamma} - \gamma) \\ \sqrt{n}(\hat{\alpha} - \alpha) \end{pmatrix} \to N(0, \Sigma)$ *in distribution, provided* Φ *and* $E(\eta_t^2 Z_t Z_t')$ *are positive definite, where* $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}$, $\Sigma_{11} = \Phi^{-1}[\Omega_1 + \Psi\Sigma_{22}\Psi']\Phi^{-1}$, $\Phi = E[(X_t - \xi_t)'(X_t - \xi_t)f_t^2]$, $\Sigma_{12} = \Phi^{-1}\Psi\Sigma_{22}$ *(Σ_{22} is defined in condition (C2)), with* $\Psi = E\{V_t f^2(\eta_t) m^{(1)}(\eta_1) \times [Z_t - E(Z_t|\eta_t)]\}$, $\Omega_1 = E[(X_t - \xi_t)'(X_t - \xi_t)U_t^2 f_t^4]$, and $m^{(1)}(\eta_1) = [\partial m(\eta)/\partial\eta]|_{\eta=\eta_1}$.

Moreover, Σ can be consistently estimated by $\hat{\Sigma}_{11} = (\hat{\Phi})^{-1}[\hat{\Omega}_1 + \hat{\Psi}\hat{\Sigma}_{22}\hat{\Psi}'] \times (\hat{\Phi})^{-1}$, $\hat{\Sigma}_{12} = (\hat{\Phi})^{-1}\hat{\Psi}\hat{\Sigma}_{22}$, *where* $\hat{\Phi} = (1/n)\sum_t (X_t - \hat{X}_t)'\hat{f}_t(X_t - \hat{X}_t)\hat{f}_t$, $\hat{\Omega}_1 = (1/n)\sum_i (X_i - \hat{X}_i)'\hat{f}_t(X_t - \hat{X}_t)\hat{f}_t(u_{nt}\hat{f}_t)^2$, $\hat{\Psi} = (1/n^2h^2)\sum\sum_{s\neq t}(X_t - \hat{X}_t) \times \hat{f}_t(m_{ns} - m_{nt})\hat{K}_{st}^{(1)}(Z_s - Z_t)'$, $m_{nt} = \hat{Y}_t - \hat{X}_t\hat{\gamma}$, $\hat{K}_{st}^{(1)} = K^{(1)}((\hat{\eta}_s - \hat{\eta}_t)/h)$, *and* $u_{nt}\hat{f}_t = (Y_t - \hat{Y}_t)\hat{f}_t - (X_t - \hat{X}_t)'\hat{\gamma}\hat{f}_t$ *is a consistent estimate for the density weighted error $u_t f_t$. When $\hat{\alpha}$ is the least squares estimator of α, we have* $r_t = \{E[Z_t Z_t']\}^{-1} Z_t u_t$ *and* $\Sigma_{22} = \{E[Z_t Z_t']\}^{-1} E[u_t^2 Z_t Z_t'] = \{E[Z_t Z_t']\}^{-1}$ *and a consistent estimator of Σ_{22} is given by* $\hat{\Sigma}_{22} = \{(1/n)\sum_t Z_t Z_t'\}^{-1} \times \{(1/n)\sum_t \hat{\eta}_t^2 Z_t Z_t'\}\{(1/n)\sum_t Z_t Z_t'\}^{-1}$.

The proof of Theorem 2.1 is given in Appendix A. Comparing the result of Theorem 2.1 to the case that η_t is observable (Robinson, 1988), we see that the asymptotic variance Σ_{11} contains an extra term $\Phi^{-1}\Psi\Sigma_{22}\Psi'\Phi^{-1}$. Hence, similar to the results of Powell (1987) and Ahn and Powell (1987), $\hat{\gamma}$ given in (10) is asymptotically less efficient compared to the estimator of Robinson (1988) when η_t is observable. This is similar to the case of a parametric regression model with generated regressors. We know that, in general, generated regressors will only affect the asymptotic variance but not the $\sqrt{n}$-consistency rate of the estimators (see Pagan, 1984).

It is worth pointing out that the asymptotic variance of $\hat{\gamma}$ is the same whether the data are weakly dependent or independent and identically distributed (i.i.d.) because of the martingale difference assumption (C1)(ii). Without this assumption, Theorem 2.1 still holds with a more complicated asymptotic variance. See Andrews (1994) for further discussion on this and for references on conditional heteroskedasticity and autocorrelation consistent covariance matrix estimation.

One can also use a trimming function to deal with the random denominator problem in kernel estimation. Following Robinson (1988), one can estimate γ by

$$\bar{\gamma} = S_{X-\hat{X}}^{-1} S_{X-\hat{X},Y-\hat{Y}}, \tag{11}$$

https://doi.org/10.1017/S0266466602183034 Published online by Cambridge University Press

where $S_{A,B} = n^{-1}\sum_t A_t B_t \mathbf{1}(\hat{f}_t \geq b)$, $S_A = S_{A,A}$, $\mathbf{1}(\hat{f}_t \geq b)$ is the usual indicator function, which takes value 1 if $\hat{f}_t \geq b$ and zero otherwise, and $b = b(n) > 0$, $b \to 0$ as $n \to \infty$. Condition (C4) needs to be replaced by the following condition.

(C4)′ As $n \to \infty$, $b \to 0$, $nh^4 b^{-2} \to \infty$, $nh^8 b^{-2} \to 0$.

We conjecture that the following result holds true for the trimming-based estimator $\bar{\gamma}$.

CONJECTURE 2.1. *Under Conditions (C1)–(C3) and (C4)′,*

$$\begin{pmatrix} \sqrt{\mathrm{n}}(\bar{\gamma} - \gamma) \\ \sqrt{n}(\hat{\alpha} - \alpha) \end{pmatrix} \to N(0, \Omega) \quad \text{in distribution,}$$

where Ω *is obtained from* Σ *by replacing* f_t *by* 1.

A formal proof of the preceding conjecture seems quite challenging and tedious as a result of the correlations of the trimming indicator function and various U-statistics contained in the expansion of $(\bar{\gamma} - \gamma)$. We leave the proof of Conjecture 2.1 for future research.

After obtaining a $\sqrt{n}$-consistent estimator of γ, one can estimate $m(\eta)$ based on

$$Y_t - X_t'\hat{\gamma} = m(\eta_t) + u_t + X_t'(\gamma - \hat{\gamma}) = m(\eta_t) + \epsilon_t, \tag{12}$$

where $\epsilon_t = u_t + X_t'(\gamma - \hat{\gamma})$. A standard kernel estimator of $m(\eta)$ is given by

$$\tilde{m}(\eta) = \frac{\sum_s (Y_s - X_s'\hat{\gamma}) K\left(\dfrac{\hat{\eta}_s - \eta}{h}\right)}{\sum_s K\left(\dfrac{\hat{\eta}_s - \eta}{h}\right)}. \tag{13}$$

Using the facts that $\hat{\gamma} - \gamma = O_p(n^{-1/2})$ and $\hat{\alpha} - \alpha = O_p(n^{-1/2})$, it is easy to show that $\tilde{m}(\eta)$ has the same asymptotic distribution as an infeasible estimator $m_n(\eta)$ based on true values of γ and α:

$$m_n(\eta) = \frac{\sum_s (Y_s - X_s'\gamma) K\left(\dfrac{\eta_s - \eta}{h}\right)}{\sum_s K\left(\dfrac{\eta_s - \eta}{h}\right)}. \tag{14}$$

The asymptotic distribution of $m_n(\eta)$ is well established in the literature (e.g., Robinson, 1983) and therefore will not be presented here.

https://doi.org/10.1017/S0266466602183034 Published online by Cambridge University Press

3. GENERALIZATIONS

In this section we discuss three different generalizations of models (4) and (5).

Case 1. A Nonlinear Regression Model (5). In Section 2 we only consider the case that (5) is a linear regression model. In that case one can estimate α by the least squares method, which gives $\sqrt{n}(\hat{\alpha} - \alpha) = n^{-1/2}\sum_{t=1}^{n} r_t \eta_t + o_p(1)$ with $r_t = \{E[Z_t Z_t']\}^{-1} Z_t$ and $\Sigma_{22} = E[r_t r_t' \eta_t^2]$.

We now show that one can easily generalize model (5) to the case of a nonlinear regression model:

$$\eta_t = S_t - g(Z_t, \alpha), \tag{15}$$

where $g(.,.)$ is a known function with α being a vector of unknown parameters. One can estimate α by the nonlinear least squares method, and it is well established that $\sqrt{n}(\hat{\alpha} - \alpha) = n^{-1/2}\sum_{t=1}^{n} r_t \eta_t + o_p(1)$, where $r_t = \{E[\nabla g_t \nabla g_t']\}^{-1} \nabla g_t$, $\nabla g_t = [\partial g(Z_t, v)/\partial v]|_{v=\alpha}$. Also, in this case the definition of Ψ in Theorem 2.1 needs to be modified by replacing Z_t in Ψ by ∇g_t. With these changes, the conclusion of Theorem 2.1 holds true when η_t is determined by (15).

Case 2. A Type-3 Tobit Model. Theorem 2.1 includes the independent data as a special case ($\beta_\tau = 0$ for $\tau \geq 1$). Hence, it covers the case of a semiparametric type-3 Tobit model as discussed in Wooldridge (1998)[1] provided that one makes the following modifications. In this case, equation (5) corresponds to a selection equation, and one needs to replace the original equation (5) by

$$S_t = \max\{0, Z_t'\alpha + \eta_t\}. \tag{16}$$

When the distribution of η_t is unknown, one can $\sqrt{n}$-consistently estimate α by the censored LAD estimator of Powell (1984), denoting it by $\hat{\alpha}$.

Y_t is still determined by equation (4) but we observe Y_t only when $S_t > 0$ (a type-3 Tobit model). By assuming that X_t is independent with (η_t, U_t), we have $E(Y_t | S_t > 0, X_t, \eta_t) = (X_t'\gamma + m(\eta_t) | S_t > 0)$, where $m(\eta_t) = E(U_t | X_t, \eta_t) = E(u_t | \eta_t)$ has an unknown functional form because the joint distribution of (η, u) is unknown. Defining an indicator function $\mathbf{1}_t = 1$ if $S_t > 0$ and $\mathbf{1}_t = 0$ if $S_t \leq 0$, we have

$$Y_t \mathbf{1}_t = (X_t'\gamma + m(\eta_t) + v_t)\mathbf{1}_t, \tag{17}$$

where $E(v_t | X_t, \eta_t, S_t > 0) = 0$.

We see that if we only use data with $S_t > 0$ ($\mathbf{1}_t = 1$), then model (17) is the same as (4). Hence, we can estimate γ by $\hat{\gamma}$ as given in (10) with n being the sample size corresponding to $S_t > 0$, where the argument in the kernel function is $\hat{\eta}_t = S_t - Z_t'\hat{\alpha}$ and $\hat{\alpha}$ is the censored LAD estimator of α given in Powell (1984). Therefore, the asymptotic distribution of $\hat{\gamma}$ based on model (17) is the same as that of equation (10) with the following changes.

https://doi.org/10.1017/S0266466602183034 Published online by Cambridge University Press

(i) We only use data with $S_t > 0$.

(ii) The definition of Σ needs to be modified by multiplying the indicator function. By noting that $\mathbf{1}_t^m = \mathbf{1}_t$ for any positive integer m, it is easy to see that Σ in this case can be obtained from the Σ given in Theorem 2.1 with f_t replaced by $f_t\mathbf{1}_t$. For example Φ becomes $\Phi = E[(X_t - \xi_t)(X_t - \xi_t)' f_t^2 \mathbf{1}_t]$, and Ψ, and Ω_1 all need to be modified by multiplying the indicator function $\mathbf{1}_t$.

(iii) $\hat{\alpha}$ is the censored LAD estimator given in Powell (1984), Σ_{22} is the asymptotic variance of $\sqrt{n}(\hat{\alpha} - \alpha)$, and $\hat{\Sigma}_{22}$ is a consistent estimator of Σ_{22} (see Powell, 1984). With these changes, Theorem 2.1 provides an asymptotic theory for the semiparametric type-3 Tobit estimator $(\hat{\gamma}, \hat{\alpha})$.

The simulation study in Sheu (2000) shows that the semiparametric estimator proposed by Wooldridge (1998) performs competitively with other existing semiparametric estimators for a type-3 Tobit model (e.g., Chen, 1997; Honore, Kyriazidou and Udry, 1997; Lee, 1994).

For the i.i.d. data case, Ahn and Powell (1993) consider a more general case where η_t is an error term from a nonparametric regression model. Although generalizing (5) to a nonparametric regression model is an interesting research topic, the technicality involved is nontrivial because we consider dependent data in this paper.

One can test the linearity of model (5) using nonparametric estimation techniques. For weakly dependent data, a kernel-based test for a parametric functional form is given by Fan and Li (1999b) (see also Bierens and Ploberger, 1997). It is also possible to develop a consistent test for testing model (2) versus model (4), i.e., testing for linearity in a time-series partially linear model with *generated regressors*. However, the regularity conditions for such a test will be somewhat complicated (as a result of the presence of the generated regressors) and are likely to be more restrictive than the ones used in Theorem 2.1. We will not pursue this problem in this paper.

Case 3. A Semiparametric Error Correction Model. We now show that the result of Theorem 2.1 can also cover the case of a semiparametric ECM (with $\Sigma_{22} = 0$). To avoid using too much notation, we will use the same notation as in Section 2. Now equation (4) is a short-run dynamic equation (in first difference form), where Y_t, X_t, η_t, and u_t are all $I(0)$ variables. We treat equation (5) as a cointegration equation:

$$\eta_t = S_t - Z_t'\alpha,$$

where S_t and Z_t are $I(1)$ variables and η_t is a $I(0)$ error term. We can estimate α by the least squares method $(\hat{\alpha})$, but now $\hat{\alpha}$ is superconsistent (i.e., $\hat{\alpha} - \alpha = O_p(n^{-1})$), and this will imply that $\Sigma_{22} = 0$ (because $\sqrt{n}(\hat{\alpha} - \alpha) = o_p(1)$). The semiparametric estimator of γ is the same as given by equation (10). But now its asymptotic variance has a simple form because $\hat{\alpha}$ is superconsistent. It can be shown that

$$\sqrt{n}(\hat{\gamma} - \gamma) \to N(0, \Phi^{-1}\Omega_1\Phi^{-1}) \quad \text{in distribution,} \tag{18}$$

https://doi.org/10.1017/S0266466602183034 Published online by Cambridge University Press

provided Φ is positive definite, where Φ and Ω_1 are defined as in Theorem 2.1. A proof of this result can be obtained from the authors upon request. Expression (18) shows that $\hat{\gamma}$ has the same asymptotic distribution as in the case when η_t is observable (e.g., Robinson, 1988). This is similar to the parametric regression model case because we know that for a parametric ECM, the residual (the generated regressor) from a cointegrated equation does not affect the asymptotic distribution of the estimated parameter in a short-run dynamic equation.

4. SEMIPARAMETRIC EFFICIENCY BOUND FOR INDEPENDENT DATA

In this section we discuss the (local) semiparametric efficient bound for models (4) and (5) under the assumption of independent data. To derive the (local) semiparametric lower bound we assume that the errors are conditional homoskedastic, $\sigma_u^2 = E(u_t|W_t) = E(u_t^2)$ and $\sigma_\eta^2 = E(\eta_t^2|W_t) = E(\eta_t^2)$, and that u_t and η_t are both normally distributed. We will follow the approach of Newey (1990). The discussion here will be informal and can readily be made rigorous following the approach of Newey (1990). See Begun et al. (1983) on a general treatment on efficient bounds of semiparametric models and see also the related work of Rilstone (1993), Pakes and Olley (1995), and Ai and Chen (1999). The semiparametric multi-index model considered by Ichimura and Lee (1991) is also similar to the model we discuss here. Let $g(\eta, \delta)$ be a parametric submodel such that $g(\eta, \delta_0) = m(\eta)$ for some δ_0, where δ is a $k \times 1$ vector of nuisance parameters ($k \geq q$). The parameters of such a parametric submodel are $(\gamma', \alpha', \delta')$. The parameter of interest is $\theta = (\gamma', \alpha')'$. For each parametric submodel there is a vector-valued score function l_ψ, which one can partition with respect to the parameter of interest θ and the nuisance parameter δ so that $l_\psi' = (l_\theta', l_\delta')$. The semiparametric bound can be interpreted as the supremum of the covariance matrices over the parametric submodels. Define the tangent set, $\mathcal{J}$, for the nuisance function as the mean square closure of all $k \times 1$ linear combinations of l_δ. Let $\mathcal{P}[l_\theta|\mathcal{J}]$ denote the projection of l_θ onto the space $\mathcal{J}$ and define $l_\theta^* = l_\theta - \mathcal{P}[l_\theta|\mathcal{J}]$. Then the semiparametric lower bound for the model is $\mathcal{V}_\theta = \{E[l_\theta^* l_\theta^{*'}]\}^{-1}$.

The log-likelihood function corresponding to the parametric submodel is

$$\mathcal{L}(\gamma, \alpha, \delta) = \text{Constant} - \frac{(Y - X'\gamma - g(S - Z'\alpha, \delta))^2}{2\sigma_u^2} - \frac{(S - Z'\alpha)^2}{2\sigma_\eta^2}. \tag{19}$$

The score function when evaluated at the true model is

$$\begin{pmatrix} l_\gamma \\ l_\alpha \\ l_\delta \end{pmatrix} = \begin{pmatrix} (1/\sigma_u^2) UX \\ (1/\sigma_u^2) U m^{(1)}(\eta) Z + (1/\sigma_\eta^2) \eta Z \\ -(1/\sigma_u^2) U [\partial g(\eta, \delta)/\partial \delta]|_{\delta = \delta_0} \end{pmatrix}, \tag{20}$$

https://doi.org/10.1017/S0266466602183034 Published online by Cambridge University Press

where we used $\partial g(\eta, \delta_0)/\partial \eta = m^{(1)}(\eta)$. Because $m(\eta)$ is not specified, we obtain the tangent set $\Lambda(.)$, the mean square closure of all $k \times 1$ linear combinations of l_δ, given by

$$\Lambda = \{U\Delta(\eta) : \Delta(\eta) \in L^2(R)\}. \tag{21}$$

Projections of l_γ, l_α on $\Lambda(.)$ are $(1/\sigma_u^2)UE(X|\eta)$ and $(1/\sigma_u^2)Um^{(1)}(\eta) \times E(Z|\eta) + (1/\sigma_\eta^2)\eta E(Z|\eta)$, respectively. We obtain the efficient score as ($\theta = (\gamma', \alpha')'$)

$$l_\theta^* \stackrel{\text{def}}{=} l_\theta - E(l_\theta|\Lambda(.)) = \begin{pmatrix} (1/\sigma_u^2)U(X - E(X|\eta)) \\ [(1/\sigma_u^2)Um^{(1)}(\eta) + (1/\sigma_\eta^2)\eta][Z - E(Z|\eta)] \end{pmatrix}. \tag{22}$$

Therefore, the semiparametric efficient bound for models (4) and (5) is

$$\mathcal{V}_\theta = \begin{pmatrix} \mathcal{V}_\gamma & \mathcal{V}_{\gamma,\alpha} \\ \mathcal{V}_{\alpha,\gamma} & \mathcal{V}_\alpha \end{pmatrix} = \{E[l_\theta^* l_\theta^{*'}]\}^{-1} \equiv \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}^{-1}, \tag{23}$$

where $A_{11} = \sigma_u^2 E[(X - \xi)(X - \xi)']$, $A_{12} = (1/\sigma_u^2)E[m^{(1)}(\eta)(X - \xi)(Z - \zeta)']$, $A_{21} = A_{12}'$, and $A_{22} = E[(1/\sigma_u^2)(m^{(1)}(\eta))^2 + (1/\sigma_\eta^2)](Z - \zeta)(Z - \zeta)']$ with $\xi = E(X|\eta)$ and $\zeta = E(Z|\eta)$.

Note that from (23) and using partitioned inverse, we have $\mathcal{V}_\gamma = A_{11}^{-1} + A_{11}^{-1}A_{12}F_2A_{12}'A_{11}^{-1}$, where $F_2 = (A_{22} - A_{12}'A_{11}^{-1}A_{12})^{-1}$. The expression $A_{11}^{-1} = \sigma_u^2\{E[(X - \xi)(X - \xi)']\}^{-1}$ is the semiparametric lower bound for γ when η is observable (or α is known). The second term in $\mathcal{V}_\gamma$ is thus the penalty for having to estimate α.

Comparing (23) with Σ given in Theorem 2.1 or Ω given in Conjecture 2.1, it can be shown that $\Sigma - \mathcal{V}_\theta$ and $\Omega - \mathcal{V}_\theta$ are both positive semidefinite. Therefore, our proposed estimators are not semiparametric efficient in the sense that the asymptotic variances Σ and Ω are larger than the semiparametric lower bound $\mathcal{V}_\theta$ for the model we consider. We leave the problem of semiparametric efficient estimation of models (4) and (5) for future research.

NOTE

1. Wooldridge (1998) did not provide asymptotic distribution theory of his proposed semiparametric estimator. Powell (1987) and Lee (1994) have considered different semiparametric estimators for a type-3 Tobit model.

REFERENCES

Ahn, H. & J.L. Powell (1993) Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics* 58, 3–29.

Ai, C. & X. Chen (1999) Efficient Estimation of Models with Conditional Moments Restrictions Containing Unknown Functions. Manuscript.

Ai, C. & D. McFadden (1997) Estimation of some partially specified nonlinear models. *Journal of Econometrics* 76, 1–37.

https://doi.org/10.1017/S0266466602183034 Published online by Cambridge University Press

Andrews, D.W.K. (1991) Asymptotic normality of series estimation for nonparametric and semiparametric regression models. *Econometrica* 59, 307–345.
Andrews, D.W.K. (1994) Asymptotic for semiparametric econometric models via stochastic equicontinuity. *Econometrica* 62, 43–72.
Bachmeier, L. & Q. Li (2002) Further evidence of nonlinearity in term structure. *Applied Economics Letters* 9, 151–153.
Begun, J.M., W. Hall, W.M. Huang, & J.A. Wellner (1983) Information and asymptotic efficiency in parametric-semiparametric models. *Annals of Statistics* 11, 432–452.
Bierens, H.J. & W. Ploberger (1997) Asymptotic theory of integrated conditional moment tests. *Econometrica* 65, 1129–1151.
Brown, B.M. (1971) Martingale central limit theorems. *Annals of Mathematics and Statistics* 42, 59–66.
Chamberlain, G. (1992) Efficiency bounds for semiparametric regression. *Econometrica* 60, 567–596.
Chen, S. (1997) Semiparametric estimation of Type-3 Tobit model. *Journal of Econometrics* 80, 1–34.
Denker, M. & G. Keller (1983) On *U*-statistics and v. Mises' statistics for weakly dependent processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 64, 505–522.
Donald, S.G. & W.K. Newey (1994) Series estimation of semilinear models. *Journal of Multivariate Analysis* 50, 30–40.
Engle, R.F. & C.W.J. Granger (1987) Co-integration and error correction: Representation, estimation, and testing. *Econometrica* 55, 251–276.
Engle, R.F., C.W.J. Granger, J. Rice, & A. Weiss (1986) Semiparametric estimates of the relationship between weather and electricity scale. *Journal of the American Statistical Association* 87, 310–320.
Fan, Y. & Q. Li (1999a) Root-n-consistent estimation of partially linear time-series models. *Journal of Nonparametric Statistics* 11, 251–269.
Fan, Y. & Q. Li (1999b) Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification testing. *Journal of Nonparametric Statistics* 10, 245–271.
Granger, C.W.J. (1981) Some properties of time series data and their use in econometric specification. *Journal of Econometrics* 25, 121–130.
Honore, B.E., E. Kyriazidou, & C. Udry (1997) Estimation of Type-3 Tobit models using symmetric trimming and pairwise comparisons. *Journal of Econometrics* 76, 107–128.
Ichimura, H. & L.F. Lee (1991) Semiparametric least squares estimation of multiple index models: Single equation estimation. In W.A. Barnett, J. Powell, & G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, pp. 3–49. Cambridge, UK: Cambridge University Press.
Kniesner, T. & Q. Li (2002) Nonlinearity in dynamic adjustment: Semiparametric estimation of panel labor supply. *Empirical Economics*, forthcoming.
Lee, L.F. (1994) Semiparametric two-stage estimation of sample selection models subject to Tobit-type selection variable. *Journal of Econometrics* 61, 305–344.
Linton, O. (1995) Second order approximation in the partially linear regression model. *Econometrica* 63, 1079–1112.
Newey, W. (1990) Semiparametric efficiency bounds. *Journal of Applied Econometrics* 5, 99–135.
Pagan, A. (1984) Econometric issues in the analysis of regressions with generated regressors. *International Economic Review* 25, 221–247.
Pakes, A. & S. Olley (1995) A limit theorem for a smooth class of semiparametric estimators. *Journal of Econometrics* 65, 295–332.
Powell, J.L. (1984) Least absolute deviations estimation for the censored regression model. *Journal of Econometrics* 25, 303–325.
Powell, J.L. (1987) Semiparametric Estimation of Bivariate Latent Variable Models. Discussion paper, Social System Research Institute, University of Wisconsin.

https://doi.org/10.1017/S0266466602183034 Published online by Cambridge University Press

Powell, J.L., J.H. Stock, & T.M. Stoker (1989) Semiparametric estimation of the index coefficients. *Econometrica* 57, 1043–1430.
Rilstone, P. (1993) Calculating the (local) semiparametric efficient bounds for the generated regressors problem. *Journal of Econometrics* 56, 357–370.
Robinson, P.M. (1983) Nonparametric estimators for time series. *Journal of Time Series Analysis* 4, 185–207.
Robinson, P.M. (1988) Root-N-consistent semiparametric regression. *Econometrica* 56, 931–954.
Sheu, S. (2000) Monte Carlo Study on Some Recent Type-3 Tobit Semiparametric Estimators. Manuscript, Texas A&M University.
Stock, J.H. (1989) Nonparametric policy analysis. *Journal of American Statistical Association* 84, 567–575.
Stoker, T.M. (1992) *Lectures on semiparametric econometrics*. Louvain-la-Neuve: CORE Foundation.
Wooldridge, J.M. (1998) Selection Corrections with a Censored Selection Rule. Manuscript, Michigan State University.
Yoshihara, K. (1976) Limiting behavior of *U*-statistics for stationary, absolutely regular processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 35, 237–252.

APPENDIX A: PROOF OF THE MAIN RESULTS

This Appendix proves Theorem 2.1. We write $g_t = g(Z_t)$ for $g = m$ or ξ, and $f_t = f(\eta_t)$. For $W_t = (u_t, \eta_t, Z_t)$ or (X_t, η_t, Z_t), we use $F(w_t)$, $F_{i_1 i_2}(w_{i_1}, w_{i_2})$, and $F_{i_1 i_2 i_3}(w_{i_1}, w_{i_2}, w_{i_3})$ to denote the distribution function of W_t and the joint distribution functions of (W_{i_1}, W_{i_2}) and $(W_{i_1}, W_{i_2}, W_{i_3})$, respectively. Also $\sum\sum_{s\neq t} = \sum_{t=1}^{n}\sum_{s=1, s\neq t}^{n}$ and $\sum\sum\sum_{t\neq s\neq r}$ denotes triple summation with three different indices. The terms C_1, C_2, and C denote generic positive constants. The tilde notation is exclusively used for independent process. Hence, $\{\widetilde{W}_t\}$ is an independent process having the same marginal distribution as $\{W_t\}$. For example, $\tilde{m}_t = m(\tilde{\eta}_t)$, $\tilde{K}_{st} = K((\tilde{\eta}_s - \tilde{\eta}_t)/h)$, and if $A = A(W_{t_1}, W_{t_2}, \ldots, W_{t_l})$, then $\tilde{A} = A(\widetilde{W}_{t_1}, \widetilde{W}_{t_2}, \ldots, \widetilde{W}_{t_l})$.

We estimate η_t by $\hat{\eta}_t = S_t - Z_t'\hat{\alpha} = \eta_t - Z_t'(\hat{\alpha} - \alpha)$. Hence,

$$\hat{\eta}_t - \hat{\eta}_s = (\eta_t - \hat{\eta}_s) + (Z_s - Z_t)'(\hat{\alpha} - \alpha). \tag{A.1}$$

For a Tobit-3 model, (A.1) also holds true provided that one only uses data with $S_t > 0$.

We give two useful formulas subsequently that will be used throughout this Appendix. Using a Taylor expansion argument, we have

$$\hat{K}_{st} - K_{st} = K_{st}^{(1)} \frac{(Z_s - Z_t)'(\hat{\alpha} - \alpha)}{h} + \frac{1}{2}\bar{K}_{st}^{(2)}\left(\frac{(Z_s - Z_t)'(\hat{\alpha} - \alpha)}{h}\right)^2, \tag{A.2}$$

where $K^{(l)}(v) = \partial^l K(v)/\partial v^l$ $(l = 1,2)$, $K_{st}^{(1)} = K^{(1)}((\eta_s - \eta_t)/h)$, $\bar{K}_{st}^{(2)} = K^{(2)}((\bar{\eta}_s - \bar{\eta}_t)/h)$, and $(\bar{\eta}_t, \bar{\eta}_s)$ is between the line segment of (η_t, η_s) and $(\hat{\eta}_t, \hat{\eta}_s)$. From (C3)(ii), we know that

https://doi.org/10.1017/S0266466602183034 Published online by Cambridge University Press

$$|K_{st}^{(i)}| \le H\left(\frac{\eta_s - \eta_t}{h}\right) \quad \text{for } i = 0,1, \quad \text{and} \quad |\bar{K}_{st}^{(2)}| \le H\left(\frac{\eta_s - \eta_t}{h}\right), \tag{A.3}$$

where $K_{st}^{(0)} = K_{st}$ and $H(\cdot)$ is the bounded function given in (C3)(ii). Both (A.1) and (A.2) will be used extensively in the proofs that follows. Also the notation of $A \sim B$ means that A and B have the same order. We first present some lemmas that will be used to prove Theorem 2.1.

LEMMA A.1. *$S_{\varepsilon(\hat{f}-f)} = o_p(n^{-1/2})$, where $\epsilon = U$ or V $(V = X - E(X|\eta))$.*

Proof. We will prove the result for a univariate random variable ε, which applies to U and each component of V. Because V is a finite-dimensional vector, it is sufficient to verify the result for each component of V. For a univariate random variable ϵ, we have

$$\begin{aligned} S_{\epsilon(\hat{f}-f)} &= \left(\frac{1}{n}\right)\sum_t \epsilon_t^2(\hat{f}_t - f_t)^2 \\ &= \frac{1}{n^3h^2}\left\{\sum\sum_{s\neq t}\epsilon_t^2(\hat{K}_{st} - hf_t)^2 + \sum\sum\sum_{r\neq s\neq t}\epsilon_t^2(\hat{K}_{st} - hf_t)(\hat{K}_{rt} - hf_t)\right\} \\ &\equiv \frac{1}{n^3h^2}\{A_{1n} + A_{2n}\}. \end{aligned}$$

By adding and subtracting terms in A_{1n}, we have

$$\begin{aligned} A_{1n} &= \sum\sum_{s\neq t}\epsilon_t^2[(K_{st} - hf_t)^2 + (\hat{K}_{st} - K_{st})^2 + 2(\hat{K}_{st} - K_{st})(K_{st} - hf_t)] \\ &\equiv A_{1n1} + A_{1n2} + 2A_{1n3}. \end{aligned}$$

Here $A_{1n1} = O_p(n^2h)$ by Lemma A.1 of Fan and Li (1999a). Using (A.1), we have $A_{1n2} = \sum_{i=1}^{2}\sum_{i'=1}^{2}(1/i!)(1/i'!)A_{1n2,ii'}[(\hat{\alpha} - \alpha)/h]^{i+i'}$, where $A_{1n2,ii'} = \sum\sum_{s\neq t}\epsilon_t^2 K_{st}^{(i)}K_{st}^{(i')}(Z_t - Z_s)^{i+i'}$.

By Lemma B.1 in Appendix B and recalling that the tilde notation denotes independent process, we have $E|A_{1n2,ii'}| \le \sum\sum_{s\neq t}E|\epsilon_t^2 K_{st}^{(i)}K_{st}^{(i')}(Z_t - Z_s)^{i+i'}| \le E|\tilde{A}_{1n2,ii'}| + \sum\sum_{s\neq t}M_n\beta_{|s-t|}^{\delta/(1+\delta)} = O(n^2h^2) + O(n) = O(n^2h^2)$ by Assumption (C1)(i) and the fact that $M_n \sim E|\epsilon_t^2 K_{st}^{(i)}K_{st}^{(i')}(Z_t - Z_s)^{i+i'}| \le \{E[|\epsilon_t|^{4(1+\delta)}]E[(Z_t - Z_s)|^{8(1+\delta)}]\}^{1/2} = O(1)$. Hence, $A_{1n2} = o_p(n^2h)$ because $(\hat{\alpha} - \alpha)/h = O_p((nh^2)^{-1/2}) = o_p(1)$. Also $A_{1n3} \le (A_{1n1} + A_{1n2})/2$. Thus we have shown that $A_{1n} = A_{1n1} + A_{1n2} + 2A_{1n3} = O_p(n^2h)$.

Next we consider A_{2n}. Again by adding and subtracting terms, we have

$$\begin{aligned} A_{2n} &= \sum\sum\sum_{r\neq s\neq t}\epsilon_t^2\{[(K_{st} - hf_t) + (\hat{K}_{st} - K_{st})][(K_{rt} - hf_t) + (\hat{K}_{rt} - K_{rt})]\} \\ &= \sum\sum\sum_{r\neq s\neq t}\epsilon_t^2\{[(K_{st} - hf_t)(K_{rt} - hf_t) + (\hat{K}_{st} - K_{st})(\hat{K}_{rt} - K_{rt}) \\ &\qquad + (K_{st} - hf_t)(\hat{K}_{rt} - K_{rt}) + (K_{rt} - hf_t)(\hat{K}_{st} - K_{st})]\} \\ &\equiv A_{2n1} + A_{2n2} + A_{2n3} + A_{2n4}. \end{aligned}$$

By Lemma A.1 of Fan and Li (1999a), we know $A_{2n1} = O_p(n^3h^{2(1+2)} + n^2h^{3/(2+2\delta)} + n) = O_p(n^3h^6) + O_p(n^2h)$ because $\delta < 2$.

https://doi.org/10.1017/S0266466602183034 Published online by Cambridge University Press

We will use Lemma B.2 to get an order for A_{2n2}. Using (A.1), $A_{2n2} = (1/i!i'!) \times \sum_{i=1}^{2}\sum_{i'=1}^{2} A_{2n2,ii'}[(\hat{\alpha} - \alpha)/h]^{i+i'}$, where $A_{2n2,ii'} = \sum\sum\sum_{r\neq s\neq t} \epsilon_t^2 K_{st}^{(i)} K_{rt}^{(i')}(Z_s - Z_t)^i(Z_r - Z_t)^{i'}$.

Using (A.2), it is easy to see that $E[|\tilde{A}_{2n2,ii'}|] \leq n^3 E[|\tilde{\epsilon}_1^2 \tilde{H}_{21}\tilde{H}_{31}(\tilde{Z}_2 - \tilde{Z}_1)^i(\tilde{Z}_3 - \tilde{Z}_1)^{i'}|] = n^3 E\{\sigma_\epsilon^2(\eta_1, Z_1)E_1[|\tilde{H}_{21}(\tilde{Z}_2 - \tilde{Z}_1)^i|]E_1[|\tilde{H}_{31}(\tilde{Z}_3 - \tilde{Z}_1)^{i'}|]\} = O(n^3h^2)$. Let $\phi(W_t, W_s, W_r) = \epsilon_t^2 K_{st}^{(i)} K_{rt}^{(i')}(Z_s - Z_t)^i(Z_r - Z_t)^{i'}$, where $W_t = (\epsilon_t, \eta_t, Z_t)$. The expression $\phi(.,.,.)$ is not symmetric in its arguments; we need to symmetrize it to use Lemma B.2 in Appendix B. However, it is obvious that the extra terms in the symmetrized version of ϕ have the same order as ϕ itself. Hence, we will apply Lemma B.2 to ϕ. Let M_{n12} and M_{n3} be defined as in Lemma B.2 with $\phi(.,.,.)$ defined previously. Then we have

$$M_{n12} \sim \iint |\epsilon_t^2 K_{st}^{(2)} K_{rt}^{(2)}(z_s - z_t)^2(z_r - z_t)^2|^{1+\delta} dF(w_r)dF(w_t, w_s)$$

$$\leq C \iint |\epsilon_t^2 (z_s - z_t)^2(z_r - z_t)^2|^{1+\delta} dF(w_r)dF(w_t, w_s)$$

$$\leq C\{E[|\epsilon_t|^{4(1+\delta)}]\}\{E[|(z_s - z_t)z_s|^{4(1+\delta)}]\}^{1/2} = O(1).$$

Similarly, it is easy to check that $M_{3n} < \infty$. Hence, by Lemma B.2, we know that $E|A_{2n2,ii'}| \leq E|\tilde{A}_{2n2,ii'}| + O(n^2 M_{n12}) + O(nM_{n3}) = O(n^3h^2) + O(n^2) + O(n) = O(n^3h^2)$. This gives that $A_{2n2} = O_p(n^3h^2)O_p((nh^2)^{-1})$ because $\hat{\delta} - \delta = O_p(n^{-1/2})$ and $i + i' \geq 2$.

Finally one can easily show that $A_{2n3} + A_{2n4}$ has an order no larger than $A_{2n1} + A_{2n2}$. Thus $A_{2n} = A_{2n1} + A_{2n2} + A_{2n3} + A_{2n4} \sim A_{2n1} + A_{2n2} = O_p(n^3h^6) + O_p(n^3h^2)O_p((nh^2)^{-1})$.

Summarizing the preceding discussion, we have $S_{\epsilon(\hat{f}-f)} = (n^3h^2)^{-1}[A_{1n} + A_{2n}] = (n^3h^2)^{-1}\{O_p(n^2h) + O_p(n^3h^6) + O_p(n^3h^2)O_p((nh^2)^{-1})\} = O_p((nh)^{-1}) + O_p(h^4) + O_p((nh^2)^{-1}) = o_p(n^{-1/2})$ by Condition (C4). ■

LEMMA A.2. *$S_{(\hat{g}-g)\hat{f}} = o_p(n^{-1/2})$, where $g = m$ or ξ.*

Proof. It suffices to consider the case that g is a univariate function.

$$S_{(\hat{g}-g)\hat{f}} = \frac{1}{n}\sum_t (\hat{g}_t - g_t)^2 \hat{f}_t^2$$

$$= (n^3h^2)^{-1}\left\{\sum\sum_{s\neq t}(g_s - g_t)^2 \hat{K}_{st}^2 + \sum\sum\sum_{r\neq s\neq t}(g_r - g_t)(g_s - g_t)\hat{K}_{st}\hat{K}_{rt}\right\}$$

$$= (n^3h^2)^{-1}\{B_{1n} + B_{2n}\}.$$

For $j = 1,2$, note that the major difference between B_{jn} defined earlier and A_{jn} (introduced in the proof of Lemma A.1) is that $\hat{K}_{st} - hf_t$ in A_{jn} is replaced by $(g_s - g_t)\hat{K}_{st}$. Hence, by following *exactly* the same steps as in the proof of Lemma A.1, but citing Lemma A.2 of Fan and Li (1999a) instead of Lemma A.1 of Fan and Li (1999a), one can show that $S_{(\hat{g}-g)\hat{f}} = o_p(n^{-1/2})$.

LEMMA A.3. *$S_{(\hat{\xi}-\xi)\hat{f},U\hat{f}} = o_p(n^{-1/2})$.*

Proof. We only need to consider the case that ξ is a univariate function. Here $S_{(\hat{\xi}-\xi)\hat{f},U\hat{f}} = S_{(\hat{\xi}-\xi)\hat{f},U(\hat{f}-f)} + S_{(\hat{\xi}-\xi)\hat{f},Uf}$. First, $S_{(\hat{\xi}-\xi)\hat{f},U(\hat{f}-f)} = o_p(n^{-1/2})$ by Lemmas

https://doi.org/10.1017/S0266466602183034 Published online by Cambridge University Press

A.1, A.2 and Cauchy inequality. The expression $S_{(\hat{\xi}-\xi)\hat{f},Uf} = (1/n^2h)\sum\sum_{s\neq t} u_t f_t(\xi_s - \xi_t)\hat{K}_{st} = (1/n^2h)\sum\sum_{s\neq t} u_t f_t\{(\xi_s - \xi_t)K_{st} + (\xi_s - \xi_t)[\hat{K}_{st} - K_{st}]\} = C_{1n} + C_{2n}$. Here $C_{1n} = O_p(n^{-1/2})$ by Lemma A.3 of Fan and Li (1999a).

Next we show $C_{2n} = o_p(n^{-1/2})$. Using equation (A.1), we have $C_{2n} = \sum_{i=1}^{2}(1/i!) \times C_{2n,i}[(\hat{\alpha} - \alpha)/h]^i$, where $C_{2n,i} = (n^2h)^{-1}\sum\sum_{s\neq t} u_t f_t(\xi_s - \xi_t)K_{st}^{(i)}(Z_s - Z_t)^i$. Comparing $C_{2n,1}$ with C_{1n}, we see that when $K_{st} = K((\eta_s - \eta_t)/h)$ in C_{1n} is replaced by $K_{st}^{(1)}(Z_s - Z_t) = K^{(1)}((\eta_s - \eta_t)/h)(Z_s - Z_t)$, we obtain $C_{2n,1}$. Obviously replacing K_{st} by $K_{st}^{(1)}$ makes no difference as far as the order of C_{1n} is concerned; also it is easy to see that adding an extra factor $Z_s - Z_t$ in C_{1n} will not change its order (because $E(u_t|X_t, Z_t, \eta_t) = 0$). Hence, by following the same steps as in the proof of $C_{1n} = o_p(n^{-1/2})$ (i.e., Fan and Li, 1999a, Lemma A.3), one can show that $C_{2n,1} = o_p(n^{-1/2})$.

Finally by Lemma B.1, we have $E|C_{2n,2}| \leq E|\tilde{C}_{2n,2}| + C(n^2h)^{-1}\sum\sum_{s\neq t}\beta_{|s-t|}^{\delta/(1+\delta)} = O(1) + O((nh)^{-1}) = O(1)$. Thus $C_{2n,2}[(\hat{\alpha} - \alpha)/h]^2 = O_p((nh^2)^{-1}) = o_p(n^{-1/2})$, which together with $C_{2n,1} = o_p(n^{-1/2})$ implies $C_{2n} = o_p(n^{-1/2})$.

Hence, $S_{(\hat{\xi}-\xi)\hat{f},U\hat{f}} = C_{1n} + C_{2n} = o_p(n^{-1/2})$. ■

LEMMA A.4.

(i) $S_{U\hat{f}} = o_p(n^{-1/2})$,
(ii) $S_{V\hat{f}} = o_p(n^{-1/2})$,
(iii) $S_{U\hat{f},V\hat{f}} = o_p(n^{-1/2})$,
(iv) $S_{V\hat{f},U\hat{f}} = o_p(n^{-1/2})$,
(v) $S_{V\hat{f},V\hat{f}} = o_p(n^{-1/2})$,
(vi) $S_{U\hat{f},V\hat{f}} = o_p(n^{-1/2})$.

Note that it suffices to consider the case that V_t is a scalar because we only need to obtain probability orders for Lemma A.4.

Proof of (i). $S_{U\hat{f}} = (1/n)\sum_t(\hat{u}_t\hat{f}_t)^2 = (n^3h^q)^{-1}\{\sum\sum_{s\neq t} u_s^2\hat{K}_{st}^2 + \sum\sum\sum_{r\neq s\neq t} u_s \times u_r\hat{K}_{st}\hat{K}_{rt}\} = (n^3h^2)^{-1}\{D_{1n} + D_{2n}\}$. Here $D_{1n} = \sum\sum_{\neq t} u_t^2\{K_{st}^2 + [\hat{K}_{st}^2 - K_{st}^2]\} = D_{1n1} + D_{1n2}$. Obviously $D_{1n1} = O_p(n^2h)$. Using equation (A.1), we have $D_{1n2} = \sum_{i=1}^{2}\sum_{i'=1}^{2}(1/s!)(1/s'!)\sum\sum_{s\neq t} U_t^2K_{st}^{(i)}K_{st}^{(i')}(Z_s - Z_t)^{i+i'}[(\hat{\alpha} - \alpha)/h]^{i+i'} = \sum_{i=1}^{2}\sum_{i'=1}^{2}(1/s!)(1/s'!)D_{1n2,ii'}[(\hat{\alpha} - \alpha)/h]^{i+i'}$, where $D_{1n2,ii'} = \sum\sum_{s\neq t} U_t^2K_{st}^iK_{st}^{(i')}(Z_s - Z_t)^{i+i'}$. The expression $E|D_{1n2,ii'}| \leq n^2\{E[U_t^4]E[|K_{st}^iK_{st}^{(i')}(Z_s - Z_t)^{i+i'}|^2]\}^{1/2} = O(n^2)$. Using $\hat{\alpha} - \alpha = O_p(n^{-1/2})$ and $i + i' \geq 2$, we get $D_{1n2} = O_p(n^2(nh^2)^{-1}) = o_p(n^2h)$. Hence, $D_{1n} = D_{1n1} + D_{1n2} = O_p(n^2h)$.

Next we consider D_{2n}; using equation (A.1), $D_{2n} = \sum\sum\sum_{r\neq s\neq t} u_su_r\{K_{st}K_{rt} + (\hat{K}_{st} - K_{st})K_{rt} + (\hat{K}_{rt} - K_{rt})\hat{K}_{st}\} = D_{2n1} + D_{2n2} + D_{2n3}$. The term $D_{2n1} = O_p(n^2h + n^2h^{2/(1+\delta)}) = O_p(n^2)$ by Lemma A.4(i) of Fan and Li (1999a). For D_{2n2}, using equation (A.1), we have $D_{2n2} = \sum_{i=1}^{2}(1/i!)\sum\sum\sum_{r\neq s\neq t} u_su_tK_{rt}K_{st}^{(i)}(Z_s - Z_t)^i[(\hat{\alpha} - \alpha)/h]^i \equiv \sum_{i=1}^{2}(1/i!)D_{2n2,i}[(\hat{\alpha} - \alpha)/h]^i$, where $D_{2n2,i} = \sum\sum\sum_{r\neq s\neq t} u_su_tK_{rt}K_{st}^{(i)}(Z_s - Z_t)^i[(\hat{\alpha} - \alpha)/h]^i$. Note that $D_{2n2,1}$ can be obtained by replacing K_{st} in D_{2n1} by $K_{st}^{(1)}(Z_s - Z_t)$. Then by the same arguments as in the proof of Lemma A.3, we know that $D_{2n2,1}$ has the same order as D_{2n1}. Thus $D_{2n2,1} = O_p(n^2)$.

For $D_{2n2,2}$, first note that $E[|\tilde{D}_{2n2,2}|] \leq n^3E[|\tilde{u}_1\tilde{u}_2\tilde{K}_{31}\tilde{K}_{21}(\tilde{Z}_2 - \tilde{Z}_1)^2|] = O(n^3h^2)$. Then by Lemma B.2 and equation (A.1), we have $E|D_{2n2,2}| \leq E|\tilde{D}_{2n2,2}| + C\sum\sum\sum_{r\neq s\neq t} M_n^{1/(1+\delta)}(\beta_{|s-t|}^{\delta/(1+\delta)} + \beta_{|s-r|}^{\delta/(1+\delta)}) = O(n^3h^2) + O(n^2) = O(n^3h^2)$ because $M_n \sim E|u_su_tK_{rt}H_{st}(Z_s - Z_t)^2|^{1+\delta} \leq C\{E|u_su_t|^{2(1+\delta)}\}^{1/2}\{E|(Z_s - Z_t)^2|^{2(1+\delta)}\}^{1/2} = O(1)$.

https://doi.org/10.1017/S0266466602183034 Published online by Cambridge University Press

Using $\hat{\alpha} - \alpha = O_p(n^{-1/2})$, we have $D_{2n2} = D_{2n2,1}O_p((nh^2)^{-1/2}) + D_{2n2,2}O_p((nh^2)^{-1}) = O_p(n^2)O_p((nh^2)^{-1/2}) + O_p(n^3h^2)O_p((nh^2)^{-1})$.

Similarly, one can show that D_{2n3} has an order no larger than D_{2n2}.

Hence, $S_{\hat{U}\hat{f}} = (n^3h^2)^{-1}\{D_{1n} + D_{2n}\} = (n^3h^2)^{-1}\{O_p(n^2) + O(n^3h^2)O_p((nh^2)^{-1})\} = O_p((nh^2)^{-1}) = o_p(n^{-1/2})$. ■

Proof of (ii). Follow the same proof as (i).

Proof of (iii). $S_{U\hat{f},\hat{V}\hat{f}} = S_{U(\hat{f}-f),\hat{V}\hat{f}} + S_{Uf,\hat{V}\hat{f}}$. Here $S_{U(\hat{f}-f),\hat{V}\hat{f}} = o_p(n^{-1/2})$ by Cauchy inequality, Lemma A.1, and (ii).

The expression $S_{Uf,\hat{V}\hat{f}} = (1/n^2h)\sum\sum_{s\neq t} u_t f_t V_s \hat{K}_{st} = (1/n^2h)\sum\sum_{s\neq t} u_t f_t V_s\{K_{st} + [\hat{K}_{st} - K_{st}]\} = F_{1n} + F_{2n}$. The term $F_{1n} = o_p(n^{-1/2})$ by Lemma A.2(iii) of Fan and Li (1999a). Using equation (A.1), we have $F_{2n} = \sum_{i=1}^{2}(1/i!)(n^2h)^{-1}\sum\sum_{s\neq t} K^{(i)} \times (Z_s - Z_t)^i[(\hat{\alpha} - \alpha)/h]^i = \sum_{i=1}^{2}(1/i!)F_{2n,i}[(\hat{\alpha} - \alpha)/h]^i$, where $F_{2n,i} = (n^2h)^{-1} \times \sum\sum_{s\neq t} K^{(i)}(Z_s - Z_t)^i$. Note that $F_{2n,1}$ can be obtained by replacing K_{st} in F_{1n} by $K_{st}^{(1)}(Z_s - Z_t)$; thus by exactly the same proof as $F_{1n} = o_p(n^{-1/2})$, one can show $F_{2n,1} = o_p(n^{-1/2})$. Finally using Lemma B.1 and equation (A.2), we have $E|F_{2n,2}| \leq (n^2h)^{-1}\sum\sum_{s\neq t} E|u_t f_t V_s H_{st}(Z_s - Z_t)| \leq h^{-1}E|\tilde{u}_1\tilde{f}_1\tilde{V}_2 H((\tilde{\eta}_2 - \tilde{\eta}_1)/h)(\tilde{Z}_2 - \tilde{Z}_a)| + C(n^2h)^{-1}\sum\sum_{s\neq t}\beta_{|s-t|}^{\delta/(1+\delta)} = O(1) + O((nh)^{-1}) = O(1)$. Hence, $F_{2n,2}[(\hat{\alpha} - \alpha)/h]^2 = O_p((nh^2)^{-1}) = o_p(n^{-1/2})$, and consequently, $F_{2n} = o_p(n^{-1/2})$. ■

Proof of (iv). Follow the same proof as (iii).

Proof of (v). Follow a proof similar to (iii).

Proof of (vi). By (i), (ii), and Cauchy inequality.

LEMMA A.5. $S_{V\hat{f}} = \Phi + o_p(1)$.

Proof. $S_{V\hat{f}} = S_{Vf} + S_{V(\hat{f}-f)} + 2S_{Vf,V(\hat{f}-f)} = S_{Vf} + o_p(1)$ by Cauchy inequality, Lemma A.1, and the fact that $S_{Vf} = O_p(1)$. Finally $S_{Vf} = \Phi + o_p(1)$ by Lemma A.6 of Fan and Li (1999a). ■

LEMMA A.6. $S_{V\hat{f},(\hat{m}-m)\hat{f}} = \Psi[\sqrt{n}(\hat{\alpha} - \alpha)] + o_p(1)$.

Proof. $S_{V\hat{f},(\hat{m}-m)\hat{f}} = S_{V(\hat{f}-f),(\hat{m}-m)\hat{f}} + S_{Vf,(\hat{m}-m)\hat{f}}$. Here $S_{V(\hat{f}-f),(\hat{m}-m)\hat{f}} \leq \{S_{V(\hat{f}-f)}S_{(\hat{m}-m)\hat{f}}\}^{1/2} = o_p(n^{-1/2})$ by Lemmas A.1 and A.2.

$S_{Vf,(\hat{m}-m)\hat{f}} = (1/n)\sum_t V_t f_t(\hat{m}_t - m_t)\hat{\tilde{f}}_t = (1/n^2h)\sum\sum_{s\neq t} V_t f_t(m_s - m_t)\hat{K}_{st} = (1/n^2h)\sum\sum_{s\neq t} V_t f_t(m_s - m_t)\{K_{st} + [\hat{K}_{st} - K_{st}]\} = \{G_{1n} + G_{2n}\}$.

The term $G_{1n} = o_p(n^{-1/2})$ by Lemma A.3 of Fan and Li (1999a).

Using equation (A.1), $G_{2n} = (n^2h)^{-1}\sum_{i=1}^{2}(1/i!)\sum\sum_{s\neq t} V_t f_t(m_s - m_t)K_{st}^{(i)}[(Z_s - Z_t)(\hat{\alpha} - \alpha)/h]^i = (n^2h)^{-1}\sum_{i=1}^{2}(1/i!)G_{2n,s}[(\hat{\alpha} - \alpha)/h]^i$, where $G_{2n,1} = (n^2h)^{-1}\sum\sum_{s\neq t} V_t f_t(m_s - m_t)K_{st}^{(1)}(Z_s - Z_t)'$ and $G_{2n,2} = (n^2h)^{-1}\sum\sum_{s\neq t} V_t f_t(m_s - m_t)K_{st}^{(2)}(Z_s - Z_t)^2$. Note that we deliberately define $G_{2n,1}$ in a way that allows Z_t to be a vector because we need the exact expression for $G_{2n,1}$, whereas for $G_{2n,2}$ we only consider the case of a scalar Z_t because we only need to get probability order for $G_{2n,2}$.

We will use Lemma B.3 to show that $G_{2n,1} = h[\Psi + o_p(1)]$. Instead of considering $G_{2n,1}$ directly, it is convenient to consider $hG_{2n,1}$.

Note that $(m_s - m_t)$, $K_{st}^{(1)}$, and $Z_s - Z_t$ are all antisymmetric. We have $hG_{2n,1} = n^{-2}\sum\sum_{s<t}[V_t f_t - V_s f_s](m_s - m_t)K_{st}^{(1)}(Z_s - Z_t)' = (2/n^2)\sum\sum_{s<t} H(W_s, W_t)$ where $H(W_s, W_t) = [V_t f_t - V_s f_s](m_s - m_t)K_{st}^{(1)}(Z_s - Z_t)'/2$ and $W_t = (X_t, Z_t, \eta_t)$ (because $V_t = X_t - E(X_t|\eta_t) \equiv V(X_t, \eta_t)$). Let ζ, $H_1(\cdot)$, s_δ, and R_n be defined in the same way as

https://doi.org/10.1017/S0266466602183034 Published online by Cambridge University Press

in Lemma B.3 with the preceding function $H(\cdot,\cdot)$. Then by Lemma B.3, we have $hG_{2n,1} = \zeta + (2/n)\sum_t (H(W_t) - \zeta) + R_n$, where $H_1(w) \equiv H_1(z,x,\eta)$ and

$$\begin{aligned}
H_1(z,x,\eta) &= (1/2)\int [vf(\eta) - v_s f(\eta_s)](m(\eta_s) - m(\eta))K^{(1)} \\
&\quad \times \left(\frac{\eta_s - \eta}{h}\right)(z_s - z)' f_w(\eta_s, x_s, z_s)\, d\eta_s\, dx_s\, dz_s \\
&= \frac{1}{2}h\int [vf(\eta) - (x_s - \xi(\eta + hu))f(\eta + hu)](m(\eta + hu) - m(\eta)) \\
&\quad \times K^{(1)}(u)(z_s - z)' f_w(\eta + hu, x_s, z_s)\, du dx_s\, dz_s \\
&= \frac{1}{2}h^2\int [vf(\eta) - (x_s - \xi(\eta + hu))f(\eta + hu)]m^{(1)}(\bar{\eta})u \\
&\quad \times K^{(1)}(u)(z_s - z)' f_w(\eta + hu, x_s, z_s)\, du dx_s\, dz_s, \qquad \textbf{(A.4)}
\end{aligned}$$

where $\bar{\eta}$ is between η and $\eta + hu$.

Subsequently, we first evaluate the order of $(1/n)\sum_t (H(W_t) - \zeta)$. Here $E\{[(1/n)\sum_t (H_1(W_t) - \zeta)]^2\} = (1/n^2)\{\sum_t E[(H_1(W_t) - \zeta)^2] + \sum\sum_{s\neq t} E[(H_1(W_s) - \zeta)(H_1(W_t) - \zeta)]\} \equiv n^{-2}\{J_{1n} + J_{2n}\}$.

Using equation (A.3), it is easy to see that $J_{1n} = nE[(H_1(W_1) - \zeta)^2] = O(nh^4)$.

Let $M_n = \max_{\{s\neq t\}}\{E[|(H_1(W_s) - \zeta)(H_1(W_t) - \zeta)|^{1+\delta}], E[|(H_1(\widetilde{W}_s) - \zeta)(H_1(\widetilde{W}_t) - \zeta)|^{1+\delta}]\}$. Then using equation (A.3) and $m(\eta + hu) - m(\eta) = m(\bar{\eta})hu$, it is straightforward to show that $M_n = O(h^{4(1+\delta)})$. Hence by Lemma B.1, it follows that $J_{2n} = \sum\sum_{s\neq t} E[(H_1(W_s) - \zeta)(H_1(W_t) - \zeta)] = \sum\sum_{s\neq t}\{E[(H_1(\widetilde{W}_s) - \zeta)(H_1(\widetilde{W}_t) - \zeta)] + CM_n^{1/(1+\delta)}\beta_{|s-t|}^{\delta/(1+\delta)}\} = 0 + O(nh^4)$.

Thus we have shown that $E\{[(1/n)\sum_t (H_1(W_t) - \zeta)]^2\} = n^{-2}\{J_{1n} + J_{2n}\} = n^{-2}\{O(nh^4) + O(nh^4)\} = O(h^4 n^{-1})$. Hence, $(1/n)\sum_t (H_1(W_t) - \zeta) = O_p(h^2 n^{-1/2}) = o_p(hn^{-1/2})$.

Next we consider the remainder term R_n. By Lemma B.3, $R_n = O(n^{-1+\epsilon/2}s_\delta)$, where $s_\delta \leq C_1 \sup_{s<t}\{E|V_t f_t(m_s - m_t)K_{st}^{(1)}(Z_s - Z_t)|^{2+\delta}\}^{1/(2+\delta)} = O(hh^{1/(2+\delta)})$.

Hence, $R_n = n^{-1+\epsilon/2}O(hh^{1/(2+\delta)}) = o_p(hn^{-1/2})$ for small enough ϵ.

Thus using Lemma B.3, we have obtained that $hG_{2n,1} = \zeta + o_p(hn^{-1/2})$. It remains to show $\zeta = h^2[\Psi + o_p(1)]$. Let $f_{\eta z}(\eta, z)$ denote the density function of (η_s, Z_s) and $f_z(z|\eta)$ denote the conditional density of z_s conditional on η_s. We have

$$\begin{aligned}
\zeta &= E[H(\widetilde{W}_t, \widetilde{W}_S)] \\
&= \iiint v_t f(\eta_t)(m(\eta_s) - m(\eta_t))K^{(1)}((\eta_s - \eta_t)/h)(z_s - z_t) f_{\eta z}(\eta_s, z_s) d\eta_s\, dz_s\, dF(w_t) \\
&= h\iiint v_t f(\eta_t)(m(\eta_t + hu) - m(\eta_t))K^{(1)}(u)(z_s - z_t) f_{\eta z}(\eta_t + hu, z_s)\, du dz_s\, dF(w_t) \\
&= h^2\left\{\left[\int uK^{(1)}(u)\, du\right]\left[\int v_t f(\eta_t)m^{(1)}(\eta_t)(z_s - z_t) f_{\eta z}(\eta_t, z_s)\, dz_s dF(w_t) + O(h)\right]\right\} \\
&= h^2\left\{\left[\int uK^{(1)}(u)\, du\right]\left[\int v_t f^2(\eta_t)m^{(1)}(\eta_t)\int (z_s - z_t) f_z(z_s|\eta_s = \eta_t)\, dz_s dF(w_t) + O(h)\right]\right\} \\
&= h^2\left[\int uK^{(1)}(u)\, du\right]\{E[V_1 f^2(\eta_1)m^{(1)}(\eta_1)[E(Z_2|\eta_2 = \eta_1) - Z_1)] + O(h)\} \\
&= h^2\{E[V_1 f^2(\eta_1)m^{(1)}(\eta_1)(Z_1 - E(Z_1|\eta_1))] + O(h)\} \equiv h^2\{\Psi + o(1)\},
\end{aligned}$$

https://doi.org/10.1017/S0266466602183034 Published online by Cambridge University Press

where we used $\int uK^{(1)}(u)\,du = -1$. Hence, $G_{2n,1} = h^{-1}[\zeta + o_p(n^{-1/2}h)] = h[\Psi + o_p(n^{-1/2})]$ and $G_{2n,1}[(\hat{\alpha} - \alpha)/h] = \Psi(\hat{\alpha} - \alpha) + o_p(n^{-1/2})$.

For $G_{2n,2}$, we have $E|G_{2n,2}| \leq (n^2h)^{-1}\sum\sum_{s\neq t} E|V_t f_t(m_s - m_t)H_{st}(Z_s - Z_t)^2| \leq h^{-1}E|\tilde{V}_1\tilde{f}_2(\tilde{m}_2 - \tilde{m}_1)\tilde{H}_{21}(\tilde{Z}_2 - \tilde{Z}_1)^2| + C(n^2h)^{-1}\sum\sum_{s\neq t}\beta_{|s-t|}^{\delta/(1+\delta)} = O(1 + (nh)^{-1}) = O(1)$. Hence, $G_{2n,2}[(\hat{\alpha} - \alpha)/h]^2 = O_p(1)O_p((nh^2)^{-1}) = o_p(n^{-1/2})$.

Summarizing the preceding discussion, we have shown that $S_{V\hat{f},(\hat{m}-m)\hat{f}} = \{G_{1n} + G_{2n}\} = \Psi(\hat{\alpha} - \alpha) + o_p(n^{-1/2})$. ■

LEMMA A.7. *$S_{(\hat{g}-g)\hat{f},\hat{\epsilon}\hat{f}} = o_p(n^{-1/2})$, where $g = m$ or ξ and $\epsilon = U$ or V.*

Proof. The proof is obtained, using Cauchy inequality, Lemma A.2, and Lemma A.4(i) and (ii).

LEMMA A.8. *$S_{V\hat{f},U\hat{f}} = S_{Vf,Uf} + o_p(n^{-1/2})$.*

Proof. $S_{V\hat{f},U\hat{f}} = S_{Vf,Uf} + S_{V(\hat{f}-f),U(\hat{f}-f)} + S_{Vf,U(\hat{f}-f)} + S_{V(\hat{f}-f),Uf} = S_{Vf,Uf} + L_{1n} + L_{2n} + L_{3n}$. Here $L_{2n} = o_p(n^{-1/2})$ by Cauchy inequality and Lemma A.1, and $L_{2n} = O_p(n^{-1/2})$ and $L_{3n} = O_p(n^{-1/2})$ by the same proof as that of Lemma A.2. ■

To prove Theorem 2.1 (i), we first verify the following results.

(i) $S_{(X-\hat{X})\hat{f}} = \Phi + o_p(1)$,
(ii) $\sqrt{n}S_{(X-\hat{X})\hat{f},(m-\hat{m})\hat{f}} = \Psi[\sqrt{n}(\hat{\alpha} - \alpha)] + o_p(1)$,
(iii) $\sqrt{n}S_{(X-\hat{X})\hat{f},\hat{U}\hat{f}} = o_p(1)$
(iv) $\sqrt{n}S_{(X-\hat{X})\hat{f},U\hat{f}} = \sqrt{n}S_{Vf,Uf} + o_p(1)$,
(v) $\sqrt{n}S_{Vf,Uf} + \Psi[\sqrt{n}(\hat{\alpha} - \alpha)] \to N(0, \Omega_1 + \Psi\Omega_2\Psi')$ in distribution.

Proof of (i). $S_{(X-\hat{X})\hat{f}} = S_{V\hat{f}-\hat{V}\hat{f}+(\xi-\hat{\xi})\hat{f}} = S_{V\hat{f}} + S_{\hat{V}\hat{f}} + S_{(\xi-\hat{\xi})\hat{f}} + 2S_{(\xi-\hat{\xi})\hat{f},V\hat{f}} - 2S_{(\xi-\hat{\xi})\hat{f},\hat{V}\hat{f}} - 2S_{V\hat{f},\hat{V}\hat{f}} = S_{V\hat{f}} + o_p(1) = \Phi + o_p(1)$ by Lemma A.5, where the third equality follows from Lemmas A.2, A.4(ii) and (v), and A.7 and Cauchy inequality. ■

Proof of (ii). $\sqrt{n}S_{(X-\hat{X})\hat{f},(m-\hat{m})\hat{f}} = \sqrt{n}\{S_{(\xi-\hat{\xi})\hat{f},(m-\hat{m})\hat{f}} + S_{V\hat{f},(m-\hat{m})\hat{f}} - S_{\hat{V}\hat{f},(m-\hat{m})\hat{f}}\} = o_p(1) + \Psi[\sqrt{n}(\hat{\delta} - \delta)] + o_p(1)$ by Lemmas A.2, A.6, and A.7 and Cauchy inequality. ■

Proof of (iii). $\sqrt{n}S_{(X-\hat{X})\hat{f},\hat{U}\hat{f}} = \sqrt{n}\{S_{(\xi-\hat{\xi})\hat{f},\hat{U}\hat{f}} + S_{V\hat{f},\hat{U}\hat{f}} - S_{\hat{V}\hat{f},\hat{U}\hat{f}}\} = o_p(1)$ by Lemma A.4(iv) and (vi), and Lemma A.7. ■

Proof of (iv). $\sqrt{n}S_{(X-\hat{X})\hat{f},U\hat{f}} = \sqrt{n}S_{V\hat{f},U\hat{f}} + \{\sqrt{n}S_{(\xi-\hat{\xi})\hat{f},U\hat{f}} - \sqrt{n}S_{\hat{V}\hat{f},U\hat{f}}\} = \sqrt{n}S_{Vf,Uf} + o_p(1)$ by Lemmas A.3, A.4(iii), and A.8. ■

Proof of (v). $\sqrt{n}(\hat{\alpha} - \alpha) = (1/\sqrt{n})\sum_t r_t + o_p(1)$. Hence, $\sqrt{n}S_{Vf,Uf} + \Psi[\sqrt{n}(\hat{\alpha} - \alpha)] = (1/\sqrt{n})\sum_t[V_t u_t + \Psi r_t] + o_p(1) \equiv (1/\sqrt{n})\sum_t R_t + o_p(1)$, where $R_t = V_t u_t + \Psi r_t$. Obviously R_t is a martingale difference process, and $\mathrm{var}(R_t) = \mathrm{Var}(V_t u_t) + \Psi\mathrm{Var}(r_t)\Psi' + 2\mathrm{cov}(V_t u_t, \Psi r_t) = \mathrm{Var}(V_t u_t) + \Psi\mathrm{Var}(r_t)\Psi' = \Omega_1 + \Psi\Omega_2\Psi'$ because $\mathrm{cov}(V_t u_t, \Psi r_t) = E[V_t u_t \Psi r_t] = E[V_t \Psi r_t E(u_t|W_t)] = 0$, where $W_t = (X_t, Z_t, \eta_t)$. Hence, $\sqrt{n}S_{Vf,Uf} + \Psi[\sqrt{n}(\hat{\alpha} - \alpha)] = (1/\sqrt{n})\sum_t[V_t u_t + \Psi r_t] + o_p(1) \to N(0, \Omega_1 + \Psi\Omega_2\Psi)$ by the central limit theorem of Brown (1971) for martingale difference processes. ■

https://doi.org/10.1017/S0266466602183034 Published online by Cambridge University Press

Proof of Theorem 2.1. We consider $\sqrt{n}(\hat{\gamma} - \gamma)$. From (10), we have

$$\begin{aligned}\sqrt{n}(\hat{\gamma} - \gamma) &= S_{(X-\hat{X})\hat{f}}^{-1}\sqrt{n}\{S_{(X-\hat{X})\hat{f},(m-\hat{m})\hat{f}+U\hat{f}-\hat{U}\hat{f}}\} \\ &= S_{(X-\hat{X})\hat{f}}^{-1}\sqrt{n}\{S_{(X-\hat{X})\hat{f},(m-\hat{m})\hat{f}} + S_{(X-\hat{X})\hat{f},U\hat{f}} - S_{(X-\hat{X})\hat{f},\hat{U}\hat{f}}\} \\ &= [S_{Vf} + o_p(1)]^{-1}\{\sqrt{n}S_{Vf,Uf} + \Psi[\sqrt{n}(\hat{\alpha} - \alpha)] + o_p(1)\} \\ &\xrightarrow{d} \Phi^{-1}N(0, \Omega_1 + \Psi\Sigma_{22}\Psi') = N(0, \Sigma_{11}),\end{aligned}$$

by (I)–(V), where $\Sigma_{11} = \Phi^{-1}(\Omega_1 + \Psi\Sigma_{22}\Psi')\Phi^{-1}$.

Condition (C2) assumes that $\sqrt{n}(\hat{\alpha} - \alpha) = (1/\sqrt{n})\sum_{t=1}^{n} r_t u_t + o_p(1) \to N(0, \Sigma_{22})$ in distribution.

From Lemmas A.1–A.8, we know that $\operatorname{cov}(\sqrt{n}(\hat{\gamma} - \gamma), \sqrt{n}(\hat{\alpha} - \alpha)) = \operatorname{cov}(\Phi^{-1}\Psi\sqrt{n}(\hat{\alpha} - \alpha), \sqrt{n}(\hat{\alpha} - \alpha)) + o(1) = \Phi^{-1}\Psi\operatorname{var}(\sqrt{n}(\hat{\alpha} - \alpha)) = \Phi^{-1}\Psi\Sigma_{22} + o(1) = \Sigma_{12} + o(1)$. Theorem 2.1 follows by a martingale central limit theorem. ■

Proof of $\hat{\Sigma} = \Sigma + o_p(1)$. It follows from the proof of result (i) that $\hat{\Phi} = S_{(X-\hat{X})\hat{f}}$ is a consistent estimator of Φ. The expression $\hat{\Omega}_1 = \Omega_1 + o_p(1)$ follows a similar proof, and $\hat{\Sigma}_{22} = \Sigma_{22} + o_p(1)$ by standard law of large number arguments. Finally, it is easy to show that

$$\begin{aligned}\hat{\Psi} &= \frac{1}{n^2h^2}\sum\sum_{s\neq t}(X_t - \hat{X}_t)\hat{f}_t(m_{ns} - m_{nt})\hat{K}_{st}^{(1)}(Z_s - Z_t)' \\ &= \frac{1}{n^2h^2}\sum\sum_{s\neq t}V_t f_t(m_s - m_t)K_{st}^{(1)}(Z_s - Z_t)' + o_p(1) \\ &\equiv h^{-1}G_{2n,1} + o_p(1) \\ &= \Psi + o_p(1).\end{aligned}$$

In the preceding expression the second equality follows from the facts that $X_t - \hat{X}_t = V_t + o_p(1)$, $\hat{f}_t = f_t + o_p(1)$, $\hat{K}_{st}^{(1)} = K_{st}^{(1)} + o_p(1)$, and $m_{nt} = m_t + o_p(1)$; the last equality was proved in the proof of Lemma A.6. Hence, $\hat{\Sigma} = \Sigma + o_p(1)$. ■

The Nonlinear Regression Case and the Type 3-Tobit Model Case

For the nonlinear regression case of $\eta_t = S_t - g(Z_t, \alpha)$, $\hat{\Sigma}$ is basically the same as in the linear model case defined previously with Z_t replaced by $\nabla g(Z_t, \hat{\alpha})$.

In the case of a Tobit-3 model, where $\hat{\alpha}$ is the censored LAD estimator of α, the definition of Σ_{22} and a consistent estimator $\hat{\Sigma}_{22}$ can be found in Powell (1984).

APPENDIX B: SOME USEFUL LEMMAS

LEMMA B.1. *Let* $g_n(x_1, x_2)$ *be a Borel function such that* $\int_{R^{2r}}|g_n(x_1, x_2)|^{1+\delta}dF_{1,1+j}(x_1, x_2) \leq M_n$ *and* $\int_{R^{2r}}|g_n(x_1, x_2)|^{1+\delta}dF(x_1)dF(x_2) \leq M_n$ *for some* $\delta > 0$ *and* $M_n > 0$. *Then* $|\int\int g_n(x_1, x_2)dF_{1,1+j}(x_1, x_2) - \int\int_{R^{2p}} g_n(x_1, x_2)\,dF(x_1)dF(x_2)| \leq 4M_n^{1/(1+\delta)}[\beta(j)]^{\delta/(1+\delta)}$.

https://doi.org/10.1017/S0266466602183034 Published online by Cambridge University Press

Proof. This is Lemma 1 of Yoshihara (1976).

LEMMA B.2. *Let $\{W_t\}$ be a strictly stationary, absolutely regular process with coefficient β_m and take values in R^r. Let $\phi(.,.,.)$ be a symmetric Borel measurable function defined on $R^r \times R^r \times R^r$ such that*

$$M_{n3} = \max\left\{\int_{R^{3r}} |\phi(x_1,x_2,x_3)|^{1+\delta}\, dF_{i_1 i_2 i_3}(x_1,x_2,x_3),\right.$$
$$\left.\int_{R^{3r}} |\phi(x_1,x_2,x_3)|^{1+\delta} dF(x_1)\, dF_{i_2 i_3}(x_2,x_3)\right\}$$

and

$$M_{n12} = \max\left\{\int_{R^{3r}} |\phi(x_1,x_2,x_3)|^{1+\delta}\, dF(x_1) dF_{i_2 i_3}(x_2,x_3),\right.$$
$$\left.\int_{R^{3r}} |\phi(x_1,x_2,x_3)|^{1+\delta} \prod_{i=1}^{3} dF(x_i)\right\},$$

for all $1 \le i_1 < i_2 < i_3 \le n$. Then

$$E\left[\sum\sum \sum_{1\le i_1<i_2<i_3\le n} |\phi(W_{i_1},W_{i_2},W_{i_3})|\right]$$
$$= O(n^3 E[|\phi(\widetilde{W}_1,\widetilde{W}_2,\widetilde{W}_3)|]) + O(n^2 M_{n12}^{1/(1+\delta)}) + O(nM_{n3}^{1/(1+\delta)}),$$

where $\{\widetilde{W}_t\}$ is an independent process having the same marginal distribution as $\{W_t\}$.

Proof. See the proof of Lemma B.1 of Fan and Li (1999a).

LEMMA B.3. *Let $\{W_t\}$ be as in Lemma B.2 and let*

$$u_n = \binom{n}{k}^{-1} \sum_{1\le t_1<t_2<\cdots<t_k\le n} H(W_{t_1},\ldots,W_{t_k})$$

be a kth-order U-statistic. Define $H_1(w_1) = \int\ldots\int H(w_1,\ldots,w_m)\prod_{i=2}^{k} dF(w_i)$ and $R_n = u_n - \zeta - (k/n)\sum_{t=1}^{n}(H_1(W_t) - \zeta)$, where $\zeta = \int H(w_1,\ldots,w_k)\prod_{i=1}^{k} dF(w_i)$. Further assume that $\beta_n^{\delta/(2+\delta)} = O(n^{-2+\epsilon})$ for some $\delta,\epsilon > 0$. Then

$$R_n = O_p(n^{-1+\epsilon/2} s_\delta).$$

where $s_\delta = \sup_{1\le t_1<\cdots<t_k\le n}(E|H(W_{t_1},\ldots,W_{t_k})|^{2+\delta})^{1/(2+\delta)}$.

Proof. This is Proposition 2 of Denker and Keller (1983).

https://doi.org/10.1017/S0266466602183034 Published online by Cambridge University Press