ANNALS OF ECONOMICS AND FINANCE **1**, 337–357 (2000)

Estimating Semiparametric Econometrics Models by Local Linear Method: With An Application to Gross-Country Growth. *

Qi Li

Department of Economics,
Texas A&M University
College Station, TX 77843, U.S.A.

and

Jeffrey Wooldridge

Department of Economics,
Michigan State University
East Lansing, MI 48824-1038, U.S.A.

It is well established that local linear method dominates the conventional local constant method in estimating nonparametric regression models by kernel method. In this paper we consider the problem of estimating semiparametric econometric models by local linear method. We provide a simple proof of establishing the joint asymptotic normality of the local linear estimator. We then show that our results can be used to easily derive the asymptotic distributions of local linear estimators for several semiparametric econometric models. An empirical application of using a semiparametric local linear estimator to cross country growth data is examined.

Key Words: Local linear estimator; Asymptotic normality; Partially linear model; Smoothing coefficient model.
JEL Classification Numbers: C14, C21.

* Li's research was supported by Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, Ontario Premier's Research Excellence Awards, and Bush Program in the Economics of Public Policy.

1529-7373/2000

1. INTRODUCTION

There is a rich literature in estimating an unknown regression function and its derivatives using local linear method with kernel weight, see Fan (1992, 1993), Fan and Gijbels (1995), Ruppert and Wand (1994), among others. Local linear estimator has many attractive properties including its minimax efficiencies (e.g., Fan (1992)).

For most regression smoothers the rate of convergence at the boundary points is slower than at the points in the interior because there are less observations in the boundary region. In the literature this is referred to the boundary effects. Earlier methods of correcting the boundary effects are mainly focused on using various kinds of boundary correction kernels. However, there are many problems with the use of boundary kernels. For example, one needs to know the exact boundary in order to use a boundary correction kernel. Also it is difficult to construct boundary correction kernels for multivariate regression case.

The local linear regression estimator adjusts for the boundary automatically and it is nearly the best boundary correction method. It has nearly 100% asymptotic minimax efficient among the class of linear smoothers, both in the interior and at the boundary (Fan and Gijbels (1995,p.91)). In contrast a local constant kernel estimator has zero minimax efficiency compared to a local linear kernel estimator, this is because the local constant estimator has a large bias especially near the boundary region. Although asymptotically the length of the boundary region shrinks to zero, in finite sample applications, say with a sample size of a few hundreds, the boundary region can be quite large. As Fan and Gijbels (1995, p.69) put it: 'If the bandwidth h is chosen to be 25% of the data range, then the boundary region contains for about 50% of the whole data. In higher dimensions these figures are even more striking.'

We share the enthusiasm of Fan and Gijbels (1995) on local linear method and advocate the use of local linear methods in estimating semiparametric models. In order to derive the asymptotic distributions of local linear estimators for various semiparametric models, we first consider the case of a nonparametric regression model and provide a simple way of establishing the asymptotic normality of a local linear estimator. We then show that our results can be easily generalized to some semiparametric models. Specifically we will consider a semiparametric partially linear model and a semiparametric smooth coefficient model. We use a semiparametric partially linear specification to study cross country growth rates, our empirical study shows that, in the boundary region (which is about 40% of the data range), the local linear method give quite different estimation result compared with that obtained from a local constant estimator. This suggest that the local constant estimator suffers from substantial finite sample bias in this application.

The paper is organized as follows. In section 2.1 we provided a simple proof of establishing the asymptotic normality of the local linear estimator. Section 2.2 discusses the problem of estimating the unknown function in a partially linear model by the local linear method. Section 2.3 proposes to estimate a semiparametric smooth coefficient model by a local linear method. Section 3 uses a partially linear specification to examine a cross country growth data.

2. LOCAL LINEAR ESTIMATOR AND ITS ASYMPTOTIC DISTRIBUTION

In section 2.1 we first consider the case of a nonparametric regression model and derive the asymptotic distribution of a local linear estimator. We then show that the result can be easily generalized to a semiparametric partially linear model in section 2.2.

2.1. A Nonparametric Regression Model

Consider a nonparametric regression model:

$$y_i = g(x_i) + \epsilon_i, \quad (i = 1, ..., n) \tag{1}$$

where x_i is of dimension d. We are interested in estimating the unknown function $g(x)$ as well as the derivative of $g(x)$: $g^{(1)}(x) \stackrel{def}{=} \partial g(x)/\partial x$ ($g^{(1)}(.)$ is a $d \times 1$ vector). Define $\delta(x) = (g(x), (g^{(1)}(x))')'$. $\delta(x)$ is a $(d+1) \times 1$ vector function, its first component is $g(x)$ and the remaining d components are the first derivatives of $g(x)$. It is well established that the local linear method can be used to estimate the unknown function $g(x)$ as well as its derivative $g^{(1)}(x)$.

Below we discuss the point-wise estimation of $\delta(x) = (g(x), g^{(1)}(x)')'$ for some $x \in R^d$ with $f(x) > 0$ ($f(.)$ is the density function x_i).

Take a Taylor series expansion of $g(x_i)$ at x, we get $g(x_i) = g(x) + (x_i - x)g^{(1)}(x) + (x_i - x)'g^{(2)}(x)(x_i - x)/2 + R_m(x_i, x)$, where $g^{(2)}(x)$ is the $d \times d$ matrix of second order derivatives of $g(x)$ and $R_m(x_i, x)$ is the remainder term in the Taylor expansion. We rewrite (1) as

$$\begin{aligned} y_i &= g(x) + (x_i - x)'g^{(1)}(x) + (x_i - x)'g^{(2)}(x)(x_i - x)/2 + R_m(x_i, x) + \epsilon_i \\ &= (1, (x_i - x)')\delta(x) + (x_i - x)'g^{(2)}(x)(x_i - x)/2 + R_m(x_i, x) + \epsilon_i. \end{aligned} \tag{2}$$

A local linear kernel estimator of $\delta(x)$ is obtained by a (kernel) weighted regression of Y_i on $(1, (x_i - x)')$.[1]

$$\hat{\delta}(x) = \begin{pmatrix} \hat{g}(x) \\ \hat{g}^{(1)}(x) \end{pmatrix} = \left[\sum_i K_{i,x} \begin{pmatrix} 1, & (x_i - x)' \\ x_i - x, & (x_i - x)(x_i - x)' \end{pmatrix} \right]^{-1} \cdot \sum_i K_{i,x} \begin{pmatrix} 1 \\ x_i - x \end{pmatrix} y_i, \tag{3}$$

where $K_{i,x} = K((x_i - x)/h)$ is the kernel function and $h = h_n$ is a smoothing parameter.

Under some regularity conditions including $h \to 0$ and $nh^d \to \infty$ as $n \to \infty$, one can show that $\hat{\delta}(x)$ is a consistent estimator of $\delta(x)$. To establish the asymptoic normality result of $\hat{\delta}(x)$ given in (3), the following stronger assumptions will be used.

(A1): (x_i, y_i) are i.i.d. as (x_1, y_1). Both x_1 and y_1 have finite 4th moments. $g(x)$ and $f(x)$ are both differentiable up to third order. For $\theta = g$ or $\theta = f$, denote by $\theta^{(1)}(x)$ and $\theta^{(2)}(x)$ the gradient (first derivative) vector and the Hessian (second derivatives) matrix of $\theta(.)$, respecitvely. Then $\theta(x)$, $\theta^{(1)}(x)$ and $\theta^{(2)}(x)$ are all bounded by some functions that have finite second moment. $\sigma^2(x) = E(\epsilon_1^2 | x_1 = x)$ is continuous in x.

(A2): $K(.) : \mathcal{R}^d \to \mathcal{R}$ is a bounded symmetric non-negative function, $\int K(u)uu'du = c_k I_d$, $\int K^2(u)du = d_k$ and $\int K^2(u)uu'du = \nu_k I_d$, where I_d is an identity matrix of dimension d, c_k, d_k and ν_k are all finite positive constants.

(A3): As $n \to \infty$, $nh^{d+2} \to \infty$ and $nh^{d+6} \to 0$.

The following Theorem establishes asymptotic normality of $\hat{\delta}(x)$.

THEOREM 2.1. *Define* $D(n) = \begin{pmatrix} (nh^d)^{1/2}, & 0 \\ 0, & (nh^{d+2})^{1/2} I_d \end{pmatrix}$, $\mu_k(x) = (1/2) c_k\, tr\{g^{(2)}(x)\}$. *Then under* **A1 - A3** *and assume that* $f(x) > 0$, *we have*

$$D(n)(\hat{\delta}(x) - \delta(x) - \begin{pmatrix} h^2 \mu_k(x) \\ 0 \end{pmatrix}) \to N(0, \Sigma_x) \text{ in distribution,}$$

where $\Sigma_x = \begin{pmatrix} d_k \sigma^2(x)/f(x), & 0 \\ 0, & \nu_k \sigma^2(x) I_d / (c_k^2 f(x)) \end{pmatrix}$.

[1] A similar idea is also used in Robinson (1989), who considered the problem of estimating a time varying parameter model: $y_t = x_t' \beta_t + \epsilon_t$, where $\beta_t = \beta(t)$ is a smooth but unknown function of t. Robinson (1989) proposed estimating $\beta(\tau)$ by (for $\tau \in [0,1]$) $\hat{\beta}(\tau) = (\sum_t k_{t,\tau} x_t x_t')^{-1} \sum_t k_{t,\tau} x_t y_t$, where $k_{t,\tau} = k(\frac{T\tau - t}{Th})$ is the kernel function and h is the smoothing parameter.

The proof of Theorem 2.1 is given in the Appendix A. The asymptotic normality of $\hat{g}(x)$ (the first component of $\hat{\delta}(x)$) was derived in Fan, Heckman and Wand (1995). Here we give the joint asymptotic distribution of $\hat{g}(x)$ and $\hat{g}^{(1)}(x)$. Also our proof differs from that of Fan, et al (1995). In particular we use a simple trick to handle the numerator and the denominator of $\hat{\delta}(x)$ separately which greatly simplifies the proofs. This method of proof enables us to generalize the result of Theorem 2.1 to some semiparametric models in a straightforward way.

The asymptotic distribution of $\hat{g}(x)$ is obtained as a corollary to Theorem 2.1.

COROLLARY 2.1. *Under the same conditions as in Theorem 2.1, we have*

$(nh^d)^{1/2}(\hat{g}(x) - g(x) - h^2\mu_k(x)) \rightarrow N(0, d_k\sigma^2(x)/f(x))$ *in distribution.*

Remark 2.1. Corollary 2.1 shows that the leading terms of bias and variance of $\hat{g}(x)$ are $h^2\mu_k(x)$ and $d_k\sigma^2(x)/[f(x)nh^d]$, respectively. It is well known that the local linear kernel estimator has the same variance as above but with a larger bias term of: $h^2\{\mu_k(x) + c_k[2(g^{(1)}(x))'f^{(1)}(x)]/f(x)\}$. Therefore, a local linear estimator has zero minimax efficiency compared with a local linear estimator. In the remaining part of this paper we will show similar results hold true for several semiparametric regression models.

2.2. A Partially Linear Model

A semiparametric partially linear model is given by (e.g., Engle, et al (1986), Robinson (1988) and Stock (1989))

$$y_i = z_i'\gamma + g(x_i) + \epsilon_i \quad (i = 1, ..., n) \tag{4}$$

where z_i is of dimension $q \times 1$, γ is a $q \times 1$ unknown parameter and x_i is $d \times 1$. It is well established that γ can be $\sqrt{n}$-consistently estimated by some nonparametric method. In this subsection we will use the kernel estimation method and show that the results of section 2.1 can be easily generalized to deliver the asymptotic distributions of $g(x)$ and $g^{(1)}(x)$ based on (4) by the local linear method.

Following Robinson (1998), we estimate $\hat{\gamma}$ by

$$\hat{\gamma} = S_{z-\hat{z}}^{-1} S_{z-\hat{z},y-\hat{y}}, \tag{5}$$

where for scalar or column-vector sequences A_i and B_i, $S_{A,B} = n^{-1}\sum_i A_i B_i' I_i$ and $S_A = S_{A,A}$. $I_i = I(\hat{f}_i > b)$, $b = b_n$ is a trimming parameter and $I(.)$

is the usual indicator function, i.e., $I_i = 1$ if $\hat{f}_i > b$ and $I_i = 0$ if $\hat{f}_i \le b$. $\hat{z}_i$ and $\hat{y}_i$ are the leave-one-out kernel estimators of $E(z_i|x_i)$ and $E(y_i|x_i)$, respectively, and $\hat{f}_i = \hat{f}(z_i)$ is the kernel estimator of the density function $f(z_i)$.

$$\hat{z}_i = \hat{E}(z_i|x_i) = \frac{1}{nh^d}\sum_j z_j K_{ij}/\hat{f}_i, \tag{6}$$

$$\hat{y}_i = \hat{E}(y_i|x_i) = \frac{1}{nh^d}\sum_j y_j K_{ij}/\hat{f}_i, \tag{7}$$

and

$$\hat{f}_i = \frac{1}{nh^d}\sum_j K_{ij}. \tag{8}$$

Lemma 2.1. *Under the same conditions as in Robinson (1988), and allow the error to be conditional heteroskedastic, $\sigma^2(x_i,z_i) = E(u_i^2|x_i,z_i)$. Then we have $\sqrt{n}(\hat{\gamma}-\gamma)\to N(0,\Sigma)$ in distribution, where $\Sigma = \Phi^{-1}\Psi\Phi^{-1}$, $\Phi = E(v_i v_i')$, $\Psi = E[\sigma^2(x_i,z_i)v_i v_i']$ and $v_i = z_i - E(z_i|x_i)$.*

After obtaining $\hat{\gamma}$, we subtract $z_i'\hat{\gamma}$ from both sides of (4) which gives

$$\begin{aligned} y_i - z_i'\hat{\gamma} &= g(x_i) + z_i'(\gamma-\hat{\gamma}) + \epsilon_i \\ &\equiv g(x_i) + u_i, \end{aligned} \tag{9}$$

where $u_i = z_i'(\gamma-\hat{\gamma}) + \epsilon_i$ acting as an error term in (9). Now one can estimate the unknown function $g(x)$ and its derivative $g^{(1)}(x)$ by the local linear method as given in equation (3) with the new dependent variable $y_i - z_i'\hat{\gamma}$ replacing y_i.

$$\begin{aligned} \hat{\delta}(x) &= \left[\sum_i K(\frac{x_i-x}{h})\begin{pmatrix}1\\ x_i-x\end{pmatrix}(1,(x_i-x)')\right]^{-1} \\ &\cdot \sum_i K(\frac{x_i-x}{h})\begin{pmatrix}1\\ x_i-x\end{pmatrix}(y_i - z_i'\hat{\gamma}). \end{aligned} \tag{10}$$

By the results of Theorem 2.1, we immediately have the following results.

Corollary 2.2. *Under the conditions of Theorem 2.1 and lemma 2.1, we have*

$$D(n)(\hat{\delta}(x) - \delta(x) - \begin{pmatrix}h^2\mu_k(x)\\ 0\end{pmatrix})\to N(0,\Sigma_x) \text{ in distribution,}$$

where $\hat{\delta}(x)$ is defined in (10), $D(n)$, $\mu_k(x)$ and Σ_x are the same as defined in Theorem 2.1.

Proof. It is easy to show that the contribution of the extra term $z_i'(\gamma-\hat{\gamma})$ is asymptotically negligible since $\hat{\gamma}-\gamma = O_p(n^{-1/2})$ and $D(n)n^{-1/2} = o(1)$. The proof of corollary 2.2 follows from the results of Theorem 2.1. Note that corollary 2.2 implies that $(nh^d)^{1/2}(\hat{g}(x)-g(x)-h^2\mu_k(x)) \to N(0, d_k\sigma^2(x)/f(x))$ in distribution, where $\hat{g}(x)$ is the first component of $\hat{\delta}(x)$ in (10). ∎

2.3. A Semiparametric Smooth Coefficient Model

In this ssubection we show that the earlier results can be used to easily derive the asymptotic distributions of local linear estimators for other semiparametric models. We will consider the semiparametric smooth coefficient model as discussed in Li, Huang and Fu (1997). A semiparametric smooth coefficient model has the following form:

$$y_i = z_i'\alpha(x_i) + \epsilon_i, \tag{11}$$

where z_i is of dimension $q \times 1$ and x_i is of dimension $d \times 1$, $\alpha(x)$ is an unspecified (unknown) smooth function. Li, et al (1997) motivated model (11) via a production frontier model, where y_i is firm i's output, z_i is an input vector and x_i is some other variables that may affect firm's production efficiency. To give a specific example, let us assume that x_i is firm's ratio of R&D to output and $z_i' = (1, z_{1i}, z_{2i})$, where z_{1i} is labor input and z_{2i} is capital input. Then model (11) says that not only labor and capital, but also *R&D* expenditure will affect firm's output.

Model (11) contains a linear regression model as a special case. To see this, let $\alpha(x) = (\alpha_0(x), \alpha_1, \alpha_2)' \equiv (\gamma_0 + \gamma_1 x, \alpha_1, \alpha_2)'$, then model (11) becomes:

$$y_i = \gamma_0 + \gamma_1 x_i + \alpha_1 z_{1i} + \alpha_2 z_{2i} + \epsilon_i. \tag{12}$$

However, (12) is restrictive because it assumes that firm's *R&D* (x) only affects the intercept: $\alpha_0(x) = \gamma_0 + \gamma_1 x$, but not the slope coefficient α_1 and α_2. In practice different levels of *R&D* ratio is likely to cause different marginal productivities of labor and capital. The semiparametric model (11) allows this possibility with flexible functional forms for the marginal productivities because the smooth coefficient function $\alpha(x) = (\alpha_0(x), \alpha_1(x), \alpha_2(x))'$ is not specified to have a known functional form.

For notational simplicity, we will only consider the case that x_i is a scalar ($d = 1$) in this subsection. Li, et al (1997) suggested to estimate the unknown smooth coefficient function $\alpha(x)$ by a local least squares method on regression y_i on z_i:

$$\tilde{\alpha}(x) = \left[\sum_i K(\frac{x_i - x}{h}) z_i' z_i\right]^{-1} \sum_i K(\frac{x_i - x}{h}) z_i' y_i. \tag{13}$$

The above estimator $\tilde{\alpha}(x)$ may look like a local linear estimator, but in fact it is a local constant estimator because $\tilde{\alpha}(x)$ is the solution of the following minimization problem:

$$\tilde{\alpha}(x) = argmin_a \sum_i [y_i - z_i' a]^2 K(\frac{x_i - x}{h}). \tag{14}$$

From (14) it is clear that the unknown function $\alpha(x_i)$ is treated locally as a constant $\alpha(x_i) \approx a$. The kernel weight function $K((x_i - x)/h)$ ensures that only local observations (x_i close to x) are used.

One can also use a local linear method to estimate $\alpha(x)$ and its derivative $\alpha^{(1)}(x)$. The idea is similar to that of section 2.1, we replace the local constant a in (14) by a local linear approximation: $a + b(x_i - x)$. Then the local linear estimators of $\alpha(x)$ and $\alpha^{(1)}(x)$ are the solutions of a and b in the following minimization problem.

$$\min_{a,b} \sum_i \{y_i - z_i'[a + b(x_i - x)]\}^2 K(\frac{x_i - x}{h}). \tag{15}$$

Define a $(2q) \times 1$ vector variable $\mathcal{Z}_i = (z_i', z_i'(x_i - x))'$ and denotes $\beta(x) = (\alpha(x)', (\alpha^{(1)}(x))')'$. Then from (15) we get the local linear estimator of $\hat{\beta}(x)$ given by

$$\hat{\beta}(x) = \begin{pmatrix} \hat{\alpha}(x) \\ \hat{\alpha}^{(1)}(x) \end{pmatrix} = \{\sum_i K(\frac{x_i - x}{h}) \mathcal{Z}_i \mathcal{Z}_i'\}^{-1} \sum_i K(\frac{x_i - x}{h}) \mathcal{Z}_i y_i. \tag{16}$$

The next Theorem establishes the asymptotic distribution of $\hat{\beta}(x)$ as defined in (16).

THEOREM 2.2. *Let $\beta(x)$ satisfy the same conditions as $\delta(x)$ given Theorem 2.1. The joint distribution of (x_i, z_i) satisfies the same conditions as in Li et al (1997). Define $\mathcal{D}(n) = \begin{pmatrix} (nh)^{1/2} I_q, & 0, \\ 0, & (nh^3)^{1/2} I_q \end{pmatrix}$. Also assume that $f_x(x) > 0$ ($f_x(.)$ is the marginal density of x_i). Then we have*

$$\mathcal{D}(n)(\hat{\beta}(x) - \beta(x) - \begin{pmatrix} h^2 \eta_k(x) \\ 0 \end{pmatrix}) \to N(0, \Omega_x) \text{ in distribution, where}$$

$$\Omega_x = \begin{pmatrix} d_k m^{-1}(x)\xi(x)m^{-1}(x)/f_x(x), & 0 \\ 0, & \nu_k m^{-1}(x)\xi(x)m^{-1}(x)/(c_k^2 f_x(x)) \end{pmatrix},$$

$\eta_k(x) = (1/2)c_k E(z_i|x_i = x)f(x)g^{(2)}(x)$, $m(x) = E[z_i z_i'|x_i = x]$, $\xi(x) = E[z_i z_i'\sigma^2(x_i, z_i)|x_i = x]$, c_k, d_k and ν_k are the same as defined in assumption **(A2)**.

The proof of Theorem 2.2 is given in the Appendix B. The asymptotic distribution of the local linear estimator for the smooth coefficient $\alpha(x)$ is obtained as a corollary of Theorem 2.2.

COROLLARY 2.3. *Under the same conditions as in Theorem 2.2, we have*

$$(nh)^{1/2}(\hat{\alpha}(x) - \alpha(x) - h^2\eta_k(x)) \to N(0, d_k m^{-1}(x)\xi(x)m^{-1}(x)/f_x(x))$$

in distribution, where $\hat{\alpha}(x)$ is the first p elements of $\beta(x)$ given in (16), $\eta_k(x)$, $m(x)$ and $\xi(x)$ are the same as defined in Theorem 2.2.

Remark 2.2. Note that the leading terms of bias and variance of $\hat{\alpha}(x)$ are $\eta_k(x)h^2$ and $c_k m^{-1}(x)\xi(x)m^{-1}(x)/f_x(x)$, respectively. It can be shown that while the local constant estimator $\tilde{\alpha}(x)$ (see (13)) has the same variance as that of $\hat{\alpha}(x)$, the bias of $\tilde{\alpha}(x)$ is

$$h^2\{\eta_k(x) + E(z_i|x_i = x)[2(g^{(1)}(x))' f_x^{(1)}(x)]/f_x(x)\}.$$

Hence, $\tilde{\alpha}(x)$ has a larger bias than that of $\hat{\alpha}(x)$. The local constant estimator $\tilde{\alpha}(x)$ of (13) has zero minimax efficiency compared with the local linear estimator $\hat{\alpha}(x)$. This is similar to the case of a standard nonparametric regression model as discussed in Remark 2.1.

3. SEMIPARAMETRIC ANALYSIS OF CROSS COUNTRY GROWTH

In this section we use a semiparametric partially linear specification to study the effects of initial output level on cross country growth rates. We are interested in whether a semiparametric specification supports the so called ‘convergence’ hypothesis. ‘Convergence’ means that the speeds of the logarithm of per-capita output of different countries (rich or poor) tend to some steady-state value from different initial conditions.

The traditional approach of cross country growth analysis usually based on a Solow type growth model and assuming a Cobb-Douglas production function that imposes a common technology which leads to a log-linear relationship between changes in per-capita output and a set of variables that includes physical and human capital, initial conditions and population changes, see for example Durlauf and Johnson (1995), Lee, Pesaran and Smith (1997) for an exposition.

A linear regression model imposes strong functional form restrictions. Recently a number of researchers attempts to introduce certain nonlinearities in the multivariate regression framework in hoping to obtain more reliable estimation results, see Durlauf and Johnson (1995), Hansen (1996) and Liu and Stengos (1998).

We use the following partially linear specification.

$$y_i = z_i'\beta + g(x_i) + u_i, \tag{17}$$

where z_i is of dimension 7×1 containing growth of population (POP), the ratio of investment to GDP (INV), human capital which is measured as the enrollment rate in secondary school (SEC), all the three variables are in logarithms, and the four 0-1 dummy are: 1960's dummy, 1970's dummy, Latin American country dummy (LAAM) and Africa country dummy. x_i is the logarithm of initial value of GDP (1960 value). Our data is obtained from King and Levine (1993), which is a pooled cross-country data and averaged over the 1960's, 1970's and 1980's, the total observation is $n = 313$. We are mainly interested in finding whether the initial conditions affect DGP linearly or not. Therefore, we choose GDP60 as the nonparametric component in (17).

Our formulation is similar in spirit to that of Durlauf and Johnson (1995), Hansen (1996), and Liu and Stengos (1998) which allows possible nonlinearities and hence may lead to different equilibria growth paths for different groups of countries. Liu and Stengos (1998) used a similar but smaller data set (with 258 observations) and they also chose to use the schooling variable as a nonparametric component and estimated an additive partially model. While a formal test of testing a partially linear model versus an additive partially linear is not available in the literature, Liu and Stengos' estimation result shows that the straight line of the least squares estimate of schooling lies inside the 95% confidence bands of a semiparametric estimate of schooling effects, this suggests that the schooling variable enters the model linearly.

For comparison we also estimated a bench-mark linear regression model:

$$y_i = z_i'\beta + x_i\gamma + u_i, \tag{18}$$

The estimation results for the parametric components ($\beta's$ and γ) are given in Table 1. For the parametric components we observe that semiparametric and OLS methods give similar estimation results. For the OLS regression, the coefficient of $ln(GDP60)$ is negative, but it is not statistically significant. This result can be viewed as weak evidence in supporting the convergence hypothesis based on a parametric linear model.

Table 1: Cross Country Growth Regression

	OLS	(t-ratio)	Semi.	(t-ratio)
Constant	0.0329	(1.40)		
ln(POP)	-0.0137	(-1.25)	-0.0089	(-0.94)
ln(INV)	0.0227	(5.50)	0.0216	(5.90)
ln(SEC)	-.0004	(-0.20)	-0.0012	(-0.65)
Dummy 60	0.0249	(7.61)	0.0243	(7.25)
Dummy 70	0.0182	(6.30)	0.0196	(6.71)
LAAM	-0.0168	(-5.46)	-0.0206	(-6.16)
Africa	-0.0193	(-4.06)	-0.0171	(-4.19)
ln(GDP60)	-0.0022	(-0.91)		

The estimated function of $g(x)$ based on the local linear method (denoted it $\hat{g}_L(x)$) and the local constant method (denoted it $\hat{g}_C(x)$) are given in figure 1. We used a standard normal kernel and the smoothing parameter is chosen via the method of cross validation. From figure 1 one can clearly see that both the local linear and local constant estimators of $g(x)$ suggest that the initial GDP affects the output nonlinearly. This is in agreement with the findings of Durlauf and Johnson (1995), Lee, et al (1997), Bianchi (1997), Hansen (1996), and Liu and Stengos (1998). The result of figure 1 appears to support the view of clustering and stratification of growth patterns over time, which in sharp contrast to the convergence hypothesis.

Figure 1 also shows that $\hat{g}_L(x)$ differs from $\hat{g}_C(x)$ in the boundary region which is about 40% of the total region. As we mention in the introduction section, in finite sample applications, the boundary region can easily be as large as half of the total region. We observe that the nonlinearity of the local linear estimator $\hat{g}_L(x)$ is more pronounced than that of the local constant estimator $\hat{g}_C(x)$. We know that a local constant estimator has large bias than that of a local linear estimator especially in the boundary region. Thus our result suggests that the local constant estimator $\hat{g}_C(x)$ does not give a reliable estimation result in the boundary region due to its substantial bias.

Figure1:

In a convincing Monte Carlo study, Racine (1997) also demonstrates that local constant estimators have substantial bias in the boundary region in finite samples. Thus our empirical result is consistent the earlier theoretical results (e.g., Fan (1992, 1993)) and simulation evidence (Racine (1997)).

Our estimation result on cross country grow is in agreement with those of Durlauf and Johnson (1995), Hansen (1996), and Liu and Stengos (1998). By allowing possible nonlinearities in the regression model we observe different equilibria growth paths for countries with different initial wealth.

4. CONCLUSION

In this paper we provide a simple proof of establishing the asymptotic normality of a local linear kernel estimator. We show that our method of proof can be used to easily derive the asymptotic distributions of local linear estimators for semiparametric econometric models. An empirical example of cross country growth analysis also demonstrates that a local constant estimator can lead to severe bias in the boundary region which can be easily as large as half of the total region in finite sample applications. Therefore we advocate the use of local linear methods for estimating semiparametric econometric models.

APPENDIX A

Proof of Theorem 2.1

To simplify the proof, we first re-write (3) in an equivalent form. Insert the identity matrix $I_{d+1} = G_n^{-1} G_n$ in the middle of (3), where $G_n = \begin{pmatrix} h^2, & 0 \\ 0, & I_d \end{pmatrix}$. We get

$$\begin{aligned} \hat{\delta}(x) &= [\sum_i K(\frac{x_i - x}{h}) G_n \begin{pmatrix} 1 \\ x_i - x \end{pmatrix} (1, (x_i - x)')]^{-1} \\ &\cdot \sum_i K(\frac{x_i - x}{h}) G_n \begin{pmatrix} 1 \\ x_i - x \end{pmatrix} y_i \\ &= [\sum_i K(\frac{x_i - x}{h}) \begin{pmatrix} h^2 \\ x_i - x \end{pmatrix} (1, (x_i - x)')]^{-1} \\ &\cdot \sum_i K(\frac{x_i - x}{h}) \begin{pmatrix} h^2 \\ x_i - x \end{pmatrix} y_i. \end{aligned} \tag{A.1}$$

The advantage of using (A.1) in the proof is that

$$\frac{1}{nh^{d+2}} \sum_i K(\frac{x_i - x}{h}) \begin{pmatrix} h^2 \\ x_i - x \end{pmatrix} (h^2, (x_i - x)')$$

converges in probability to a non-singular matrix. Hence, we can analyze the denominator and numerator of (A.1) separately and thus greatly simplify the derivations. On the other hand, it is easy to show that for any choices of non-stochastic sequences of $c(n)$,

$$c(n)\sum_i K(\frac{x_i-x}{h})\begin{pmatrix}1\\ x_i-x\end{pmatrix}(1,(x_i-x)')$$

will not converge to a non-singular matrix.

Substituting the Taylor expansion (2) into (A.1) and also multiply both the numerator and denominator of $(A.1)$ by $1/(nh^{d+2})$, we have

$$\begin{aligned}
\hat{\delta}(x) &= [\frac{1}{nh^{d+2}}\sum_i K_{i,x}\begin{pmatrix}h^2\\ x_i-x\end{pmatrix}(1,(x_i-x)')]^{-1}\\
&\quad\cdot\{\frac{1}{nh^{d+2}}\sum_i K_{i,x}\begin{pmatrix}h^2\\ x_i-x\end{pmatrix}[g(x_i)+\epsilon_i]\}\\
&= \delta(x)+[\frac{1}{nh^{d+2}}\sum_i K_{i,x}\begin{pmatrix}h^2, & h^2(x_i-x)'\\ (x_i-x), & (x_i-x)(x_i-x)'\end{pmatrix}]^{-1}\\
&\quad\times\{\frac{1}{nh^{d+2}}\sum_i K_{i,x}\begin{pmatrix}h^2\\ x_i-x\end{pmatrix}[(x_i-x)'g^{(2)}(x)(x_i-x)/2\\
&\quad+\epsilon_i+R_m(x,x_i)]\}\\
&\equiv \delta(x)+[\,A^{1,x}\,]^{-1}\{A^{2,x}+A^{3,x}\}+(s.o.), \qquad\qquad (A.2)
\end{aligned}$$

where

$$A^{1,x}=\frac{1}{nh^{d+2}}\sum_i K_{i,x}\begin{pmatrix}h^2, & h^2(x_i-x)'\\ (x_i-x), & (x_i-x)(x_i-x)'\end{pmatrix} \qquad (A.3)$$

$$A^{2,x}=\frac{1}{nh^{d+2}}\sum_i K_{i,x}\begin{pmatrix}h^2\\ x_i-x\end{pmatrix}(x_i-x)'g^{(2)}(x)(x_i-x)/2, \quad (A.4)$$

$$A^{3,x}=\frac{1}{nh^{d+2}}\sum_i K_{i,x}\begin{pmatrix}h^2\\ x_i-x\end{pmatrix}\epsilon_i, \qquad (A.5)$$

and the smaller order $(s.o.)$ term comes from $[A^{1,x}]^{-1}\frac{1}{nh^{d+2}}\sum_i K_{i,x}R_m(x,x_i)$, which has an order smaller than $[A^{1,x}]^{-1}A^{2,x}$.

Lemma C.1 of Appendix C shows that $A^{1,x}=M+o_p(1)$, where $M=\begin{pmatrix}f(x), & 0\\ c_k f^{(1)}(x), & c_k f(x)I_d\end{pmatrix}$. M is obviously non-singular because $det(M)=c_k(f(x))^{d+1}>0$.

Rewrite (A.2) as

$$D(n)(\hat{\delta}(x)-\delta(x))=D(n)[A^{1,x}]^{-1}[A^{2,x}+A^{3,x}]+(s.o.)$$

Define a diagonal matrix $R = diag(M^{-1}) = \begin{pmatrix} 1/f(x), & 0 \\ 0, & I_d/(c_k f(x)) \end{pmatrix}$ and $V = \begin{pmatrix} d_k\sigma^2(x)f(x), & 0 \\ 0, & \nu_k\sigma^2(x)f(x)I_d \end{pmatrix}$. Then Theorem 2.1 will be proved if we can show the showings:

(i) $D(n)[A^{1,x}]^{-1}[A^{2,x} + A^{3,x}] = D(n)M^{-1}[A^{2,x} + A^{3,x}] + o_p(1)$,

(ii) $D(n)M^{-1}[A^{2,x} + A^{3,x}] = R\,D(n)[A^{2,x} + A^{3,x}] + o_p(1)$,

(iii) $D(n)A^{2,x} = \begin{pmatrix} (nh^{d+4})^{1/2}\mu_k(x)f(x) \\ 0 \end{pmatrix} + o_p(1)$,

(iv) $D(n)A^{3,x} \to N(0, V)$ in distribution.

These are proved in lemmas A.1 - A.4 below.

Lemma A.1. $D(n)[A^{1,x}]^{-1}[A^{2,x}+A^{3,x}] = D(n)M^{-1}[A^{2,x}+A^{3,x}]+o_p(1)$

Proof. Writing $D(n)[A^{1,x}]^{-1}[A^{2,x} + A^{3,x}] = D(n)M^{-1}[A^{2,x} + A^{3,x}] + D(n)[(A^{1,x})^{-1} - M^{-1}][A^{2,x} + A^{3,x}]$. Therefore, it suffices to show that

$$D(n)[(A^{1,x})^{-1} - M^{-1}][A^{2,x} + A^{3,x}] = o_p(1). \tag{A.6}$$

Using the partitioned inverse, we get $M^{-1} = \begin{pmatrix} 1/f(x), & 0 \\ -f^{(1)}(x)/f^2(x), & I_d/(c_k f(x)) \end{pmatrix}$. From $A^{1,x} = M + o_p(1)$ (see lemma C.1 (v)) we know that

$$(A^{1,x})^{-1} = M^{-1} + o_p(1). \tag{A.7}$$

Let $J \overset{def}{=} (A^{1,x}_{11})^{-1} - M^{-1} \equiv \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22}I_d \end{pmatrix}$. (A.7) implies that $J_{ij} = o_p(1)$ for $i, j = 1, 2$. In order to prove (A.6), we need a sharp rate for J_{12}. Below we show that $J_{12} = O_p(h^2)$.

Define $F = (A^{1,x}_{22} - A^{1,x}_{21}(A^{1,x}_{11})^{-1}A^{1,x}_{12})^{-1}$. Using the partitioned inverse and the results of lemma C.1, we obtain

$$\begin{aligned}(A^{1,x})^{-1} &= \begin{pmatrix} (A^{1,x}_{11})^{-1}(I_d + A^{1,x}_{12}FA^{1,x}_{21}(A^{1,x}_{11})^{-1}) & -A^{1,x}_{11}A_{12}F \\ -FA^{1,x}_{21}(A^{1,x}_{11})^{-1} & F \end{pmatrix} \\ &= \begin{pmatrix} (1/f(x))I_d + o_p(1) & O_p(h^2) \\ -f^{(1)}(x)/(c_k f(x)) + o_p(1) & I_d/(c_k f(x)) + o_p(1) \end{pmatrix} \end{aligned} \tag{A.8}$$

because $A^{1,x}_{12} = O_p(h^2)$ by lemma C.1 (iii). (A.8) leads to

$$J = (A^{1,x}_{11})^{-1} - M^{-1} = \begin{pmatrix} o_p(1) & O_p(h^2) \\ o_p(1) & o_p(1)I_d \end{pmatrix}, \tag{A.9}$$

which proves that $J_{12} = O_p(h^2)$. Using (A.9) we immediately have $D(n)[(A^{1,x})^{-1} - M^{-1}][A^{2,x} + A^{3,x}] = \begin{pmatrix} (nh^d)^{1/2} & 0 \\ 0 & (nh^{d+2})^{1/2} I_d \end{pmatrix} \cdot \begin{pmatrix} o_p(1) & O_p(h^2) \\ o_p(1) & o_p(1) I_d \end{pmatrix} [\begin{pmatrix} A_1^{2,x} \\ A_2^{2,x} \end{pmatrix} + \begin{pmatrix} A_1^{3,x} \\ A_2^{3,x} \end{pmatrix}] = o_p(1)$, because $(nh^d)^{1/2} O_p(h^2) A_2^{2,x} = O_p((nh^d)^{1/2} h^4) = o_p(1)$ and $(nh^d)^{1/2} O_p(h^2) A_2^{3,x} = (nh^d)^{1/2} h^2 O_p((nh^{d+2})^{-1/2}) = O_p(h)$ by lemmas C.2 (ii) and C.3 (ii). This proves (A.6). ▮

LEMMA A.2. $D(n) M^{-1} [A^{2,x} + A^{3,x}] = R\, D(n) [A^{2,x} + A^{3,x}] + o_p(1)$.

Proof. Note that R is a diagonal matrix with $R = diag(M^{-1})$, we have $D(n) M^{-1} [A^{2,x} + A^{3,x}] = R\, D(n) [A^{2,x} + A^{3,x}] + o_p(1)$ because the terms associated with the off-diagonal element of M^{-1} are all $o_p(1)$, i.e., $(nh^{d+2})^{1/2} A_1^{2,x} = \sqrt{nh^{d+2}} O_p(h^2) = O_p((nh^{d+6})^{-1/2}) = o_p(1)$ and $(nh^{d+2})^{1/2} A_1^{3,x} = \sqrt{nh^{d+2}} O_p((nh^d)^{-1/2}) = O_p(h) = o_p(1)$. ▮

LEMMA A.3. $D(n) A^{2,x} = \begin{pmatrix} (nh^{d+4})^{1/2} \mu_k(x) f(x) \\ 0 \end{pmatrix} + o_p(1)$.

Proof. By lemma C.2, we have $(nh^d)^{1/2} A_1^{2,x} = \sqrt{nh^{d+4}} [\mu_k(x) f(x) + o_p(1)]$ and $(nh^{d+2})^{1/2} A_2^{2,x} = \sqrt{nh^{d+2}} O_p(h^2) = O_p((nh^{d+6})^{1/2}) = o_p(1)$. Hence $D(n) A^{2,x} = \begin{pmatrix} (nh^d)^{1/2} A_1^{2,x} \\ (nh^{d+2})^{1/2} A_2^{2,x} \end{pmatrix} = \begin{pmatrix} (nh^{d+4})^{1/2} \mu_k(x) f(x) \\ 0 \end{pmatrix} + o_p(1)$. ▮

LEMMA A.4. *$D(n) A^{3,x} \to N(0, V)$ in distribution, where*

$$V = \begin{pmatrix} d_k f(x) \sigma^2(x), & 0 \\ 0, & \nu_k f(x) \sigma^2(x) I_d \end{pmatrix}.$$

Proof. By lemma C.3, $var((nh^d)^{1/2} A_1^{3,x}) = d_k f(x) \sigma^2(x) + o(1)$, $var((nh^{d+2})^{1/2} A_2^{3,x}) = \nu_k f(x) \sigma^2(x) I_d + o(1)$ and

$$cov((nh^d)^{1/2} A_1^{3,x}, (nh^{d+2})^{1/2} A_2^{3,x}) = o(1).$$

Hence, $var(D(n) A^{3,x}) = V + o(1)$. Also note that $A^{3,x}$ has mean zero.

Thus $D(n) A^{3,x} \to N(0, V)$ in distribution by a triangular-array (see Serfling (1980) page 32) central limit theorem. ▮

Proof of Theorem 2.1

Lemmas A.1 - A.4 imply that

$$\begin{aligned}
& D(n)(\hat{\delta}(x) - \delta(x) - \begin{pmatrix} h^2\mu_k(x) \\ 0 \end{pmatrix}) \\
= \ & RD(n)[A^{2,x} + A^{3,x}] - \begin{pmatrix} (nh^{d+4})^{1/2}\mu_k(x) \\ 0 \end{pmatrix} + o_p(1) \\
= \ & R\,D(n)A^{3,x} + o_p(1) \rightarrow R\,N(0,V) + o_p(1) \rightarrow N(0,\Sigma)
\end{aligned}$$

in distribution, where $\Sigma = RVR$ is the same Σ as given in Theorem 2.1. ▌

APPENDIX B

Proof of Theorem 2.2

Similar to the proof of Theorem 2.1, we first re-write (16) in an equivalently form by inserting an identity matrix $I_{2p} = \mathcal{G}_n\mathcal{G}_n^{-1}$, where $\mathcal{G}_n = \begin{pmatrix} h^2 I_q, & 0 \\ 0, & I_q \end{pmatrix}$. Using (11) and replacing $\alpha(x_i)$ by the Taylor expansion: $\alpha(x) + \alpha^{(1)}(x)(x_i - x) + \alpha^{(2)}(x)(x_i - x)^2/2 + R_m(x, x_i)$, we get

$$\begin{aligned}
\hat{\beta}(x) \ = \ & [\frac{1}{nh^3}\sum_i K_{i,x} \begin{pmatrix} h^2 z_i \\ z_i(x_i - x) \end{pmatrix} (z_i', z_i'(x_i - x))]^{-1} \\
\cdot \ & \{\frac{1}{nh^3}\sum_i K_{i,x} \begin{pmatrix} h^2 z_i \\ z_i(x_i - x) \end{pmatrix} y_i\} \\
= \ & \beta(x) + [\frac{1}{nh^3}\sum_i K_{i,x} \begin{pmatrix} h^2 z_i z_i', & h^2 z_i z_i'(x_i - x)^2 \\ z_i z_i'(x_i - x), & z_i z_i'(x_i - x)^2 \end{pmatrix}]^{-1} \\
\times \ & \{\frac{1}{nh^3}\sum_i K_{i,x} \begin{pmatrix} h^2 z_i \\ z_i(x_i - x) \end{pmatrix} [\alpha^{(2)}(x)(x_i - x)^2/2 + \epsilon_i + R_m(x, x_i)]\} \\
\equiv \ & \beta(x) + [\,B^{1,x}\,]^{-1}\{B^{2,x} + B^{3,x}\} + (s.o.), \qquad \text{(B.1)}
\end{aligned}$$

where

$$\begin{aligned}
B^{1,x} \ &= \ \frac{1}{nh^3}\sum_i K_{i,x} \begin{pmatrix} h^2 z_i' z_i, & h^2 z_i' z_i(x_i - x) \\ z_i' z_i(x_i - x), & z_i z_i'(x_i - x)^2 \end{pmatrix} \\
&\equiv \ \begin{pmatrix} B_{11}^{1,x} & B_{12}^{1,x} \\ B_{21}^{1,x} & B_{22}^{1,x} \end{pmatrix}, \qquad \text{(B.2)}
\end{aligned}$$

$B_{ij}^{1,x}$ are all of dimension $q \times q$ $(i, j = 1, 2)$,

$$B^{2,x} \ = \ \frac{1}{nh^3}\sum_i K_{i,x} \begin{pmatrix} h^2 z_i \\ z_i(x_i - x) \end{pmatrix} \alpha^{(2)}(x)(x_i - x)^2/2$$

$$\equiv \begin{pmatrix} B_1^{2,x} \\ B_2^{2,x} \end{pmatrix}, \tag{B.3}$$

$B_i^{2,x}$ are all of dimension $q \times 1$ ($i = 1, 2$), and

$$B^{3,x} = \frac{1}{nh^{d+2}} \sum_i K_{i,x} \begin{pmatrix} h^2 z_i \\ z_i(x_i - x) \end{pmatrix} \epsilon_i \equiv \begin{pmatrix} B_1^{3,x} \\ B_2^{3,x} \end{pmatrix}, \tag{B.4}$$

$B_i^{3,x}$ are all of dimension $q \times 1$ ($i = 1, 2$). The smaller order term ($s.o.$) on the right hand of (B.1) comes from $[B^{1,x}]^{-1} \frac{1}{nh^3} \sum_i K_{i,x} R_m(x, x_i)$, which has an order smaller than $[B^{1,x}]^{-1} B^{2,x}$.

Lemmas B.1 to B.3 below give the probability orders of $B^{j,x}$ for $j = 1, 2, 3$. The proof of Theorem 2.5 follows exactly the same arguments as in the proof of Theorem 2.1 except that one needs to cite lemmas B.1 to B.3 (rather than citing lemmas C.1 to C.3) in the proof.

LEMMA B.1. *Define* $\mathcal{M} = \begin{pmatrix} f_x(x)m(x), & 0 \\ c_k m(x) f_x^{(1)}(x), & c_k f_x(x) m(x) \end{pmatrix}$, *where* $m(x) = E(z_i z_i' | x_i = x)$ *($m(x)$ is $q \times q$ positive definite matrix). Then*

$$B^{1,x} = \mathcal{M} + o_p(1).$$

Proof. It suffices to prove the followings:
(i) $B_{11}^{1,x} = f_x(x) m(x) + o_p(1)$,
(ii) $B_{21}^{1,x} = c_k m(x) f_x^{(1)}(x) + o_p(1)$,
(iii) $B_{12}^{1,x} = O_p(h^2) = o_p(1)$ and
(iv) $B_{22}^{1,x} = c_k f_x(x) m(x) + o_p(1)$.

Since the proofs are almost identical to the proof of lemma C.2, we will only provide a proof for (i) here. (ii)-(iv) can be proved similarly. Let $f(.,.)$ and $f(.|.)$ denote the joint density of (x_i, z_i) and the conditional density of z_i conditional on x_i, respectively. Then we have

$$\begin{aligned} E(B_{11}^{1,x}) &= h^{-1} E[z_1' z_1 K_{1x}] = h^{-1} \int\int f(x_1, z_1) z_1' z_1 K((x_1 - x)/h) dx_1 dz_1 \\ &= h^{-1} \int\int f_x(x_1) f(z_1|x_1) z_1' z_1 K((x_1 - x)/h) dx_1 dz_1 \\ &= \int\int f_x(x + au) f(z_1 | x_1 = x + au) z_1' z_1 K(u) du dz_1 \\ &= f_x(x) \int f(z_1 | x_1 = x) z_1' z_1 dz_1 + O(h) \\ &= f_x(x) E[z_1' z_1 | x_1 = x] + o(1) \equiv f_x(x) m(x) + o(1). \end{aligned}$$

Similarly one can easily show that $E[||B_{11}^{1,x} - E(B^{1,x})||^2] = o(1)$. Hence $B_{11}^{1,x} = f_x(x) m(x) + o_p(1)$. ∎

LEMMA B.2. *(i)* $B_1^{2,x} = h^2 c_k E(z_i|x_i = x) f(x)\, g^{(2)}(x)/2 + o_p(1)] \equiv h^2[\eta_k(x) + o_p(1)]$,
(ii) $B_2^{2,x} = O_p(h^2)$.

Proof. Mimic the proof of lemma C.2. ▮

LEMMA B.3. *Define* $\xi(x) = E[z_i z_i' \sigma^2(x_i, z_i)|x_i = x]$, *and*

$$\mathcal{V} = \begin{pmatrix} d_k f_x(x)\xi(x), & 0 \\ 0, & \nu_k f_x(x)\xi(x) \end{pmatrix}.$$

Then $\mathcal{D}(n)B^{3,x} \to N(0, \mathcal{V})$ *in distribution.*

Proof. Similar to the proof of lemma C.3, one can show the followings:
(i) $var((nh)^{1/2} B_1^{3,x}) = d_k f(x)\xi(x) + o(1)$,
(ii) $var((nh^3)^{1/2} B_2^{3,x}) = \nu_k f(x)\xi(x) + o(1)$ and
(iii) $cov((nh)^{1/2} B_1^{3,x}, (nh^3)^{1/2} B_2^{3,x}) = O(h) = o(1)$.
Then lemma B.3 follows by a triangular-array central limit theorem. Since the detailed proof is almost identical to the proof of lemma C.3, we only outline a proof for (i) below.

$$\begin{aligned}
& var((nh)^{1/2} B_1^{3,x}) \\
= \ & (nh)^{-1} \textstyle\sum_i E[z_i z_i' \epsilon_i^2 K_{ix}^2] \\
= \ & h^{-1} E[z_1 z_1' \sigma^2(x_1, z_1) K^2((x_1 - x)/h)] \\
= \ & h^{-1} \textstyle\int\int f(x_1, z_1) z_1 z_1' \sigma^2(x_1, z_1) K^2((x_1 - x)/h) dx_1 dz_1 \\
= \ & \textstyle\int\int f_x(x + au) f(z_1|x_1 = x + au) z_1 z_1' \sigma^2(x + au, z_1) K^2(u) du dz_1 \\
= \ & f_x(x)[\textstyle\int K^2(u) du][\int f(z_1|x_1 = x) z_1 z_1' \sigma^2(x, z_1) dz_1] + O(h) \\
= \ & d_k f_x(x) E[z_1 z_1' \sigma^2(x_1, z_1)|x_1 = x] + o(1) \\
= \ & d_k f_x(x)\xi(x) + o(1).
\end{aligned}$$

▮

Proof of Theorem 2.2

Define a diagonal matrix $\mathcal{R} = diag(\mathcal{M}^{-1}) = \begin{pmatrix} (f_x(x)m(x))^{-1}, & 0 \\ 0, & (c_k f_x(x)m(x))^{-1} \end{pmatrix}$. Using the results of Lemmas B.1 to B.3 and by the same arguments as in the proof of Theorem 2.1, one can show that

$$\begin{aligned}
& \mathcal{D}(n)(\hat{\beta}(x) - \beta(x) - \begin{pmatrix} h^2 \eta_k(x) \\ 0 \end{pmatrix}) \\
= \ & \mathcal{R}\mathcal{D}(n)[B^{2,x} + B^{3,x}] - \begin{pmatrix} (nh^5)^{1/2} \eta_k(x) \\ 0 \end{pmatrix} + o_p(1) \\
= \ & \mathcal{R}\,\mathcal{D}(n) B^{3,x} + o_p(1) \\
= \ & \mathcal{R}\, N(0, \mathcal{V}) + o_p(1) \to N(0, \Omega_x)
\end{aligned}$$

in distribution, where $\Omega_x = \mathcal{R}\mathcal{V}\mathcal{R}$ is the same Ω_x as given in Theorem 2.2 ($\mathcal{V}$ is defined in lemma B.3). ▮

APPENDIX C

Some Useful Lemmas

In this Appendix we provide some useful lemmas. Let $A^{1,x}$, $A^{2,x}$ and $A^{3,x}$ be the same as defined in the Appendix A, i.e., $A^{1,x} = \begin{pmatrix} A_{11}^{1,x}, & A_{12}^{1,x} \\ A_{21}^{1,x}, & A_{22}^{1,x} \end{pmatrix}$, where

$$\begin{aligned} A_{11}^{1,x} &= \tfrac{h^2}{nh^d} \textstyle\sum_i K(\tfrac{x_i - x}{h}), \\ A_{21}^{1,x} &= \tfrac{1}{nh^{d+2}} \textstyle\sum_i K(\tfrac{x_i - x}{h})(x_i - x), \\ A_{12}^{1,x} &= h^2 (A_{21}^{1,x})' \\ A_{22}^{1,x} &= \tfrac{1}{nh^{d+2}} \textstyle\sum_i K(\tfrac{x_i - x}{h})(x_i - x)(x_i - x)'. \end{aligned}$$

$A^{2,x} = \begin{pmatrix} A_1^{2,x} \\ A_2^{2,x} \end{pmatrix}$, where

$$\begin{aligned} A_1^{2,x} &= \tfrac{h^2}{nh^{d+2}} \textstyle\sum_i K(\tfrac{x_i - x}{h})(x_i - x)' g^{(2)}(x)(x_i - x)/2 \\ A_2^{2,x} &= \tfrac{1}{nh^{d+2}} \textstyle\sum_i K(\tfrac{x_i - x}{h})(x_i - x)(x_i - x)' g^{(2)}(x)(x_i - x)/2. \end{aligned}$$

$A^{3,x} = \begin{pmatrix} A_1^{3,x} \\ A_2^{3,x} \end{pmatrix}$, where

$$\begin{aligned} A_1^{3,x} &= \tfrac{1}{nh^d} \textstyle\sum_i K(\tfrac{x_i - x}{h})\epsilon_i \\ A_2^{3,x} &= \tfrac{1}{nh^{d+2}} \textstyle\sum_i K(\tfrac{x_i - x}{h})(x_i - x)\epsilon_i. \end{aligned}$$

LEMMA C.1. *Under the same conditions as in Theorem 2.1, we have*
(i) $A_{11}^{1,x} = f(x) + o_p(1)$,
(ii) $A_{21}^{1,x} = c_k f^{(1)}(x) + o_p(1)$,
(iii) $A_{12}^{1,x} = O_p(h^2)$,
(iv) $A_{22}^{1,x} = c_k f(x) I_d + o_p(1)$,
(v) $A^{1,x} = M + o_p(1)$, *where* $M = \begin{pmatrix} f(x), & 0 \\ c_k f^{(1)}(x), & c_k f(x) I_d \end{pmatrix}$.

Proof. (i) $A_{11}^{1,x}$ is a kernel estimator of $f(x)$, $A_{11}^{1,x} = f(x) + o_p(1)$ is a well established result.
(ii)

$$\begin{aligned} E[A_{21}^{1,x}] &= h^{-(d+2)} E[K(\tfrac{x_1 - x}{h})(x_i - x)] = h^{-2}[\textstyle\int f(x + hv) K(v) hv dv] \\ &= [0 + (\textstyle\int K(v) v v' dv) f^{(1)}(x) + O(h)] = c_k f^{(1)}(x) + o(1). \end{aligned}$$

It is straightforward to show that $var(A_{21}^{1,x}) = O((nh^{d+2})^{-1}) = o(1)$. Hence $A_{21}^{1,x} = c_k f^{(1)}(x) + o_p(1)$.

(iii) $A_{12}^{1,x} = h^2 (A_{21}^{1,x})' = O_p(h^2)$ because $A_{21}^{1,x} = O_p(1)$ by (ii) above.

(iv) Similarly one can easily show that $A_{22}^{1,x} = f(x) \int K(v) v v' dv + o_p(1) = c_k f(x) I_d + o_p(1)$.

(v) follows directly from (i)-(iv). ▮

LEMMA C.2. *Under the same conditions as in Theorem 2.1, we have*
(i) $A_1^{2,x} = h^2[(1/2) c_k f(x) tr\{g^{(2)}(x)\} + o_p(1)] \equiv h^2[\eta_k(x) f(x) + o_p(1)]$,
(ii) $A_2^{2,x} = O_p(h^2)$.

Proof. (i) It is straightforward to show that

$$\begin{aligned} E(A_1^{2,x}/h^2) &= (1/2) h^{-(d+2)} E[K(\tfrac{x_1 - x}{h})(x_1 - x)' g^{(2)}(x)(x_1 - x)] \\ &= (1/2) c_k f(x) tr[g^{(2)}(x)] + o(1) \end{aligned}$$

and $var(A_1^{2,x}/h^2) = O((nh^p)^{-1})$. Hence, $A_1^{2,x}/h^2 = (1/2) c_k f(x) tr[g^{(2)}(x)] + o_p(1)$, which proves (i).

(ii)

$$\begin{aligned} E[A_2^{2,x}] &= \tfrac{1}{h^{d+2}} E[K(\tfrac{x_1 - x}{h})(x_1 - x)'(x_1 - x)' g^{(2)}(x)(x_1 - x)] \\ &= h[\int f(x + hv) K(v) v'(v' g^{(2)}(x) v) dv] = h[0 + O(h)] = O(h^2) \end{aligned}$$

and $var(A_2^{2,x}) = O(h^2 (nh^d)^{-1})$. Therefore, $A_2^{2,x} = O_p(h^2 + h(nh^d)^{-1/2}) = O_p(h^2 + h^2 (nh^{d+2})^{-1/2}) = O_p(h^2)$. ▮

LEMMA C.3. *Under the same conditions as in Theorem 2.1, we have*
(i) $Var((nh^d)^{1/2} A_1^{3,x}) = d_k f(x) \sigma^2(x) + o(1)$,
(ii) $Var((nh^{d+2})^{1/2} A_2^{3,x}) = \nu_k f(x) \sigma^2(x) I_d + o(1)$,
(iii) $cov((nh^d)^{1/2} A_1^{3,x}, (nh^{d+2})^{1/2} A_2^{3,x}) = O(h) = o(1)$.

Proof. (i) $(nh^d)^{1/2} A_1^{3,x}$ has mean zero and its variance is

$$E[nh^d (A_1^{3,x})^2] = \{f(x) \sigma^2(x) \int K^2(v) dv + o(1)\} = d_k f(x) \sigma^2(x) + o(1).$$

(ii) $(nh^{d+2})^{1/2} A_2^{3,x}$ has mean zero and its variance is

$$\begin{aligned} & E[K^2(\tfrac{x_1 - x}{h}) \sigma^2(x_1)(x_i - x)(x_i - x)'] \\ =& \int f(x + hv) \sigma^2(x + hv) K^2(v) v v' dv \\ =& [f(x) \sigma^2(x) \int K^2(v) v v' dv + o(1)] \\ =& \nu_k f(x) \sigma^2(x) I_d + o(1). \end{aligned}$$

(iii)

$$
\begin{aligned}
& cov((nh^d)^{1/2} A_1^{3,x}, (nh^{d+2})^{1/2} A_2^{3,x}) \\
= \;& nh^{d+1} E[A_1^{3,x} A_2^{3,x}] \\
= \;& h[\textstyle\int K^2(v) v v' dv] \partial(f(x)\sigma^2(x))/\partial x + o(h) = O(h) = o(1).
\end{aligned}
$$

$\blacksquare$

REFERENCES

Bianchi, M., 1997, Testing for convergence: Evidence from non-parametric multimodality tests. *Journal of Applied Econometrics* **12**, 393-409.

Durlauf, S. N. and P. A. Johnson, 1995, Multiple regimes and cross-country growth behavior. *Journal of Applied Econometrics* **10**, 365-384.

Engle JASA

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

Fan, J., 1993, Local linear regression smoothers and their minimax efficiencies. *The Annals of Statistics*, **21**, 196-216.

Fan, J. and I. Gijbels, 1995, Local Polynomial Modeling and Its Applications. Chapman & Hall.

Fan, J., N. E. Heckman, and M. P. Wand, 1995, Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *Journal of American Statistical Association.*

Hansen, B., 1996, Sample splitting and threshold estimation. *Mimeo.*

King, R. G. and R. Levine, 1993, Finance and growth: Schumpeter might be right. *Quarterly Journal of Economics* **CVIII**, 717-737.

Li, Q., C. Huang, and F. Fu, 1997, Semiparametric smooth coefficient model. *Mimeo.*

Liu, Z. and T. Stengos, 1998, Non-linearity in cross country growth regressions: A semiparametric approach. *Journal of Applied Econometrics*, forthcoming.

Lee, K., M. H. Pesaran, and R. Smith, 1997, Growth and convergence in a multi-country empirical stochastic Solow model. *Journal of Applied Econometrics* **12**, 357-392.

Racine, J., 1997, Bias-corrected kernel regression. *Mimeo.*

Robinson, P., 1988, Root-N-consistent semiparametric estimation. *Econometrica* **56**, 931-954.

Ruppert, D. and M. P. Wand, 1994, Multivariate weighted least squares regression. *Annals of Statistics* **22**, 1346-1370.

Serfling, R. J.,1980, *Approximation Theorem of Mathematical Statistics.* John Wiley & Sons, New York.

Stock, J. H.,1989, Nonparametric policy analysis. *Journal of American Statistical Association* **84**, 567-575.