Journal of Business & Economic Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/ubes20

Data-Driven Bandwidth Selection for Nonstationary Semiparametric Models

Yiguo Sun[a] & Qi Li[b]
[a] Department of Economics, University of Guelph, Guelph, ON, Canada N1G 2W1
[b] Department of Economics, Texas A&M University, College Station, TX 77843-4228
Published online: 24 Jan 2012.

Data-Driven Bandwidth Selection for Nonstationary Semiparametric Models

Yiguo SUN
Department of Economics, University of Guelph, Guelph, ON, Canada N1G 2W1 (yisun@uoguelph.ca)

Qi LI
Department of Economics, Texas A&M University, College Station, TX 77843-4228 (qi@econmail.tamu.edu)

This article extends the asymptotic results of the traditional least squares cross-validatory (CV) bandwidth selection method to semiparametric regression models with nonstationary data. Two main findings are that (a) the CV-selected bandwidth is stochastic even asymptotically and (b) the selected bandwidth based on the local constant method converges to 0 at a different speed than that based on the local linear method. Both findings are in sharp contrast to existing results when working with weakly dependent or independent data. Monte Carlo simulations confirm our theoretical results and show that the automatic data-driven method works well.

KEY WORDS: Integrated time series; Local constant; Local linear; Semiparametric varying-coefficient model.

1. INTRODUCTION

It is well known that the capability of selecting the optimal smoothing parameter is crucial to the application of nonparametric estimation techniques. For independent and weakly dependent data, plug-in methods and (generalized) cross-validation (CV) methods have been incorporated into many popular software packages, and the corresponding asymptotic theory has been well developed (e.g., Härdle and Marron 1985). However, to the best of our knowledge, no asymptotic analysis exists for studying the performance of bandwidth selection methods when integrated time series data are involved, even though nonparametric/semiparametric estimation of regression models with integrated processes has recently attracted much attention among econometricians and statisticians (see, e.g., Juhl 2005; Karlsen, Myklebust, and Tjostheim 2007; Cai, Li, and Park 2009; Phillips 2009; Wang and Phillips 2009a, 2009b; Xiao 2009). This article aims to fill this gap.

We focus on a particular type of semiparametric model, the semiparametric varying-coefficient model with integrated regressors. Cai, Li, and Park (2009) and Xiao (2009) derived the asymptotic properties of kernel estimators for this model, but used ad hoc methods for selecting the smoothing parameter. In this article we suggest using the least squares CV (LS–CV) method to choose the smoothing parameter and examine the asymptotic properties of this data-driven method–selected smoothing parameter. We consider the following semiparametric model:

$$Y_t = X_t^T \beta(Z_t) + u_t, \qquad 1 \le t \le n, \tag{1.1}$$

where Y_t, Z_t, and u_t are scalars, Z_t and u_t are stationary variables with $E(u_t|Z_t) = 0$ for all t, X_t is a $p \times 1$ vector containing some nonstationary components, and $\beta(Z_t)$ is a $p \times 1$ vector of unspecified smooth functions. In situations where all of the variables are weakly dependent or iid, model (1.1) was considered by Chen and Tsay (1993), Cai, Fan, and Yao (2000), and Zhou and Liang (2009), among others. When $Z_t = t$, Equation (1.1) was considered by Robinson (1989) and Cai (2007). Model (1.1) adds extra flexibility to a linear cointegrating regression model by allowing the cointegrating vector to be smooth functions of some stationary variables. If some or all of the coefficient functions are constant, then model (1.1) becomes a partially linear or linear cointegrating regression model. Thus our asymptotic result also covers the popular partially linear model; see Section 2.4 for a more detailed discussion of this.

To estimate the unknown coefficient curve $\beta(\cdot)$ in model (1.1), we apply both the local constant and the local linear regression approaches. The local constant (LC) estimator of $\beta(z)$ is given by

$$\hat{\beta}(z) = \left[\sum_{t=1}^n X_t X_t^T K_{h,tz}\right]^{-1} \sum_{t=1}^n X_t Y_t K_{h,tz}, \tag{1.2}$$

where $K_{h,tz} = h^{-1}K(\frac{Z_t - z}{h})$, $K(\cdot)$ is a kernel function, and h is the smoothing parameter. The local linear estimator of $\beta(z)$, along with the estimator of its derivative function $\beta^{(1)}(z) = d\beta(z)/dz$, is given by

$$\begin{pmatrix} \hat{\beta}(z) \\ \hat{\beta}^{(1)}(z) \end{pmatrix} = \left[\sum_{t=1}^n K_{h,tz} \begin{pmatrix} 1 & Z_t - z \\ Z_t - z & (Z_t - z)^2 \end{pmatrix} \otimes (X_t X_t^T)\right]^{-1} \times \sum_{t=1}^n K_{h,tz} \begin{pmatrix} X_t \\ X_t(Z_t - z) \end{pmatrix} Y_t, \tag{1.3}$$

where "⊗" denotes the Kronecker product.

In this article we are interested in studying the asymptotic properties of the LS–CV–selected bandwidth for both the local constant and the local linear estimator defined earlier when X_t contains integrated covariates. We show that the CV-selected bandwidth is stochastic even asymptotically, and that the CV-selected bandwidth for the local constant estimator and the local

Journal of Business & Economic Statistics
October 2011, Vol. 29, No. 4
DOI: 10.1198/jbes.2011.09159

linear estimator converge to 0 at different speeds. Both findings are in sharp contrast to the existing results obtained for independent and weakly dependent data cases. We show that the local linear estimation method has smaller estimated mean squared error than the local constant estimation method. We also show that the asymptotic properties of the CV-selected bandwidth remain unchanged when the $I(1)$ regressor is correlated with the error term u_t, provided that the correlation between the $I(1)$ regressor and the error term u_t is not overly persistent [see Equation (2.7) for the detailed condition].

To simplify notation/proofs without affecting the essence of our results, we give our theories and proofs for scalar cases; that is, X_t is a scalar $I(1)$ variable in Section 2, and $X_t = (X_{t1}, X_{t2})^T$ is a 2×1 vector, with X_{t1} an $I(0)$ variable and X_{t2} an $I(1)$ variable in Section 3. Allowing higher dimensions in X_t will only make the mathematical representation of the asymptotic results more complicated without providing any additional insight into the problem.

The rest of the article is organized as follows. Section 2 describes the cross-validation method and derives asymptotic results when X_t in model (1.1) is an $I(1)$ variable. Section 3 provides asymptotic analysis when $X_t = (X_{t1}, X_{t2})^T$, where X_{t1} is an $I(0)$ variable and X_{t2} is an $I(1)$ variable. Section 4 presents Monte Carlo simulations. We relegate all mathematical proofs to two appendices.

2. CROSS–VALIDATION METHOD WITH AN $I(1)$ COVARIATE

2.1 The CV Function and Regularity Conditions

Let $\hat{h}$ be the data-driven bandwidth selected to minimize the following LS–CV function:

$$CV(h) = \frac{1}{n^2}\sum_{t=1}^{n}[Y_t - X_t^T\hat{\beta}_{-t}(Z_t)]^2 M(Z_t), \quad (2.1)$$

where $0 \le M(\cdot) \le 1$ is a nonnegative weight function that trims out observations near the boundary of the support of Z_t and $\hat{\beta}_{-t}(Z_t)$ is the leave-one-out local constant or local linear estimator of $\beta(Z_t)$ defined in Section 1. Apparently, scaling the CV function by n^{-2} rather than the conventional choice of n^{-1} is introduced purely for theoretical reasons, and this does not affect the value of the selected bandwidth minimizing (2.1). With a scale of n^{-2}, $CV(h)$ asymptotically has the same order as $\int(\hat{\beta}(z) - \beta(z))^2 M(z)\,dz$, because $\sup_{1\le t\le n}\|X_t\|^2/n = O_p(1)$. Thus, $CV(h)$ with the scale of n^{-2}, can be roughly viewed as a weighted version of the average squared error of the nonparametric estimator $\hat{\beta}(\cdot)$; we provide a more detailed discussion of this point later.

To simplify notation, we write $\beta_t = \beta(Z_t)$, $\hat{\beta}_{-t} = \hat{\beta}_{-t}(Z_t)$, and $M_t = M(Z_t)$ for all t. Substituting (1.1) into (2.1) gives

$$CV(h) = \frac{1}{n^2}\sum_t [X_t^T(\beta_t - \hat{\beta}_{-t})]^2 M_t + \frac{2}{n^2}\sum_t u_t X_t^T(\beta_t - \hat{\beta}_{-t})M_t + \frac{1}{n^2}\sum_t u_t^2 M_t, \quad (2.2)$$

where the last term does not depend on h. Thus minimizing $CV(h)$ over h is equivalent to minimizing $CV_0(h)$, where $CV_0(h)$ consists of the first two terms of $CV(h)$,

$$\begin{aligned} CV_0(h) &\overset{\text{def}}{=} n^{-2}\sum_t [X_t^T(\beta_t - \hat{\beta}_{-t})]^2 M_t \\ &\quad + 2n^{-2}\sum_t u_t X_t^T(\beta_t - \hat{\beta}_{-t})M_t \\ &= CV_{0,1} + 2CV_{0,2}, \end{aligned} \quad (2.3)$$

where the definitions of $CV_{0,1}$ and $CV_{0,2}$ should be apparent. In Appendix A we show that $CV_{0,1}$ is the leading term of $CV_0(h)$. With $CV_{0,1} = n^{-1}\sum_t[\frac{X_t^T}{\sqrt{n}}(\beta_t - \hat{\beta}_{-t})]^2 M_t$, and by $\sup_t |X_t|/\sqrt{n} = O_p(1)$, we would expect $CV_{0,1}$ to have the same order as $\int(\beta(z) - \hat{\beta}(z))^2 M(z)\,dz$.

We make the following assumptions:

A1 $\{Z_t\}$ is a strictly β-mixing stationary sequence of size $-(2+\delta')/\delta'$ for some $0 < \delta' < \delta < 1$, and $E|Z_t|^{2+\delta} < M < \infty$. Define $\mathcal{M} = \{z \in R : M(z) > 0\}$ [the support of the weight function $M(\cdot)$]. We require that $\mathcal{M}$ be a compact subset of $\mathcal{R}$. Let $f(z)$ be the pdf of Z_t. Then $f(z)$ has bounded derivatives (uniformly in $z \in \mathcal{M}$) up to the fourth order, and $\inf_{z\in\mathcal{M}} f(z) > 0$.

A2 $\beta(z)$ is not a linear function and is continuously differentiable up to the fourth order over $z \in \mathcal{M}$.

A3 Let $\mathcal{F}_{nt} = \sigma(X_{i+1}, Z_{i+1}, u_i : i \le t)$ be the smallest sigma field containing the past history of $\{(X_{i+1}, Z_{i+1}, u_i)\}_{i=0}^t$. Here $(u_t, \mathcal{F}_{nt})$ is a martingale difference sequence with $E(u_t^2|\mathcal{F}_{n,t-1}) = \sigma_u^2 < \infty$ for all t and $\sup_t E(|u_t|^q|\mathcal{F}_{n,t-1}) < \infty$ for some $q > 2$. In addition, the error terms u_t are independent of $\{(X_t, Z_t)\}_{t=1}^n$.

A4 The kernel function $K(\cdot)$ is a symmetric (around 0) bounded probability density function on the interval $[-1, 1]$.

A5 $nh^3(\ln n)^2 \to 0$, $h[\ln(n)]^4 \to 0$, and $n^{1-\epsilon}h \to \infty$ for some (arbitrarily) small $\epsilon > 0$, as $n \to \infty$.

A6 Let $v_t = \Delta X_t = X_t - X_{t-1}$ and $\eta_t(z) = e_t(z) - E(e_t(z))$, where $e_t(z) = (\beta(Z_t) - \beta(z))K_{h,tz}$ and $K_{h,tz} = h^{-1} \times K((Z_t - z)/h)$. The partial sums of the vector process $(v_t, \eta_t, u_t K_{h,tz})$ follow a multivariate invariance principle,

$$\begin{bmatrix} B_{n,x}(r) \\ B_{n,\beta,z}(r) \\ B_{n,u,z}(r) \end{bmatrix} = \begin{bmatrix} n^{-1/2}\sum_{t=1}^{[nr]} v_t \\ (nh)^{-1/2}\sum_{t=1}^{[nr]} \eta_t(z) \\ \sqrt{h/n}\sum_{t=1}^{[nr]} u_t K_{h,tz} \end{bmatrix} \Longrightarrow \begin{bmatrix} B_x(r) \\ B_{\beta,z}(r) \\ B_{u,z}(r) \end{bmatrix} \overset{\text{def}}{=} \text{BM}(0, \Omega), \quad (2.4)$$

where $\text{BM}(0, \Omega)$ denotes a Brownian motion with mean 0 and finite nonsingular variance–covariance matrix Ω. Here $[a]$ is the integer part of a and $r \in [0, 1]$.

A6* On a suitable probability space, there exists a vector Brownian process, $\text{BM}(0, \Omega)$, with mean 0 and finite nonsingular variance–covariance matrix such that

$$\text{(i)}\quad \sup_{r\in[0,1]} |B_{n,x}(r) - B_x(r)| = o_p(1),$$

$$\sup_{1\le t\le n} |X_t| = O\left(\sqrt{n \ln\ln n}\right)$$

almost surely; (2.5)

$$\text{(ii)}\quad \sup_{z\in\mathcal{M}} \sup_{r\in[0,1]} \left\| \begin{bmatrix} B_{n,\beta,z}(r) \\ B_{n,u,z}(r) \end{bmatrix} - \begin{bmatrix} B_{\beta,z}(r) \\ B_{u,z}(r) \end{bmatrix} \right\| = o_p(1). \quad (2.6)$$

Assumptions A1 and A2 impose a β-mixing weak dependence condition on Z_t and some moments and smoothness conditions on $f(z)$ and $\beta(z)$. Assumption A3 assumes that u_t is a martingale difference process independent of $\{(X_t, Z_t)\}_{t=1}^n$, and this assumption significantly simplifies our proofs. Later we discuss how to relax Assumption A3 in two ways: (a) allowing for u_t to be a stationary mixing process in Assumption A3′, which relaxes the martingale difference assumption, and (b) allowing u_t to be correlated with (X_t, Z_t) in Assumption A3″, which removes the independence assumption. Assumptions A4 and A5 impose mild conditions on the kernel function and the bandwidth h. The kernel function with a compact support is not essential and can be removed at the cost of a lengthy proof. Assumption A5 implies that $nh/[\ln(n)]^d \to \infty$ for any constant $d > 0$.

In Assumption A6, the weak convergence of the vector of the partial sums holds under some standard regularity conditions, such as strong mixing of $\{(v_t, Z_t, u_t)\}$ with some moment conditions. Assumption A6* imposes strong convergence results on the partial sums. Similar conditions were used by Wang and Phillips (2009a). Note that Equation (2.6) in Assumption A6* requires a uniform convergence result of $(B_{n,\beta,z}(r), B_{n,u,z}(r))$ over $z \in \mathcal{M}$. Equation (2.5) is a result of the functional law of the iterated logarithm (see Rio 1995).

Under Assumptions A1–A3, $B_x(r)$, $B_{\beta,z}(r)$, and $B_{u,z}(r)$ are independent of one another because the asymptotic covariances between each pair of the partial sums $B_{nx}(r)$, $B_{n,u,z}(\cdot)$, and $B_{n,\beta,z}(\cdot)$ are 0, with the variances of $B_{\beta,z}(r)$ and $B_{u,z}(r)$ given by $\sigma_1^2(z) = \lim_{n\to\infty} \text{var}(\sum_{t=1}^{[nr]} (nh)^{-1/2} \eta_{1,nt}(z)) = \nu_2(K) \times [\beta^{(1)}(z)]^2 f(z)$ and $\sigma_2^2(z) = \lim_{n\to\infty} \text{var}(\sqrt{h/n} \sum_{t=1}^{[nr]} u_t K_{h,tz}) = \sigma_u^2 \nu_0(K) f(z)$, respectively, with $\nu_j(K) = \int u^j K^2(u)\, du$. In Section 2.3, we show that the independence between $B_x(r)$ and $B_{u,z}(r)$ also holds under the "weaker" Assumption A3″. Write $W_\beta(r) = B_{\beta,z}(r)/\sigma_1(z)$ and $W_u(r) = B_{u,z}(r)/\sigma_2(z)$. Then $(W_\beta(r), W_u(r))$ is a bivariate standard Brownian motion vector independent of the stochastic process $B_x(r)$.

The foregoing assumptions exclude the cases where the error term u_t is serially correlated and the integrated variables are endogenous. These assumptions are imposed to simplify the proofs and can be replaced by the weaker Assumption A3′, or by Assumption A3″.

A3′ Same as Assumption A3 with this exception: Instead of assuming that u_t is a martingale difference process, u_t is a strictly stationary α-mixing process with mean 0, variance $\sigma_u^2 < \infty$, and mixing coefficients $\alpha(\tau) \equiv \alpha_\tau$ satisfying $\alpha_\tau = O(\tau^{-p})$ for some $p > \delta_1/(\delta_1 - 2)$ and $\delta_1 > 2$. Also, $E(|u_t|^{\delta_1})$ is finite.

The following (endogeneity) assumption is motived by Saikkonen's (1991) idea. Assumption 2 of Wang and Phillips (2009b) is in a similar spirit but more general than ours, because their assumption 2 allows a potential nonlinear relation between u_t and v_s.

A3″ The following representation of u_t allows X_t to be endogenous:

$$u_t = \sum_{j=-k_0}^{k_0} \alpha_j v_{t-j} + \varepsilon_t, \quad (2.7)$$

where $\{\varepsilon_t\}_{t=1}^n$ is independent of $\{(v_t, Z_t)\}_{t=1}^n$ and k_0 is a nonnegative integer. In addition, $\omega_t = (\varepsilon_t, v_t)^T$ is a strictly stationary β-mixing process with mean 0 and mixing coefficients $\beta(\tau) \equiv \beta_\tau$ satisfying $\beta_\tau = O(\tau^{-q})$ for some $q > 2(1+\delta')/\delta'$ and $0 < \delta' < \delta < 1$, and ω_t has a finite fourth moment. In addition, $\lim \text{var}(n^{-1/2} \times \sum_{t=1}^n \omega_t)$ is a finite, positive definite matrix. Moreover, $E(u_t^2|z_t = z)$ is continuously differentiable up to the second order over $z \in \mathcal{M}$.

Here we mainly use Assumption A3 (along with other assumptions) to prove the main results of the article. In supplemental Appendix B (available from the authors on request), we briefly discuss how the proofs can be modified so that our results remain valid even when Assumption A3 is replaced by Assumption A3′ or by Assumption A3″.

2.2 Main Results

In this section we present the asymptotic results of $\hat{h}$ minimizing (2.1) where the leave-one-out estimator can be either the local constant or the local linear estimator. The results are given in Theorem 2.1 for the local constant estimation case and in Theorem 2.2 for the local linear estimation case.

Theorem 2.1. Let $\hat{h}_{lc}$ denote the cross-validation–selected bandwidth based on the local constant estimation method. Under Assumptions A1–A5 and A6*, and assuming that $\beta(z)$ is not a constant function, we have

$$\text{(i)}\quad CV(h) - \frac{1}{n^2}\sum_{t=1}^n u_t^2 M(z_t) - CV_{lc,L}(h) = o_p(h/n + (n^2 h)^{-1}), \quad (2.8)$$

where

$$CV_{lc,L}(h) = \frac{h}{n}\nu_2(K) B_{x,(2)}^{-1} \left[\int M(z)\left(\beta^{(1)}(z)\right)^2 dz\right] \zeta_{\beta,2}^2 + \frac{1}{n^2 h}\nu_0(K)\sigma_u^2 \int M(z)\, dz\, B_{x,(2)}^{-1} \zeta_{u,1}^2, \quad (2.9)$$

$B_{x,(2)} = \int_0^1 B_x^2(r)\, dr$, $\zeta_{i,j} = \int_0^1 B_x^j(r)\, dW_i(r)$ for $i = \beta, u$ and $j = 1, 2$;

$$\text{(ii)}\quad \sqrt{n}\hat{h}_{lc} - \sigma_u \sqrt{\frac{\zeta_{u,1}^2 \nu_0(K) \int M(z)\, dz}{\zeta_{\beta,2}^2 \nu_2(K) \int M(z)[\beta^{(1)}(z)]^2\, dz}} \xrightarrow{p} 0. \quad (2.10)$$

The proof of Theorem 2.1 is given in Appendix A. Theorem 2.1(i) states that, apart from a term ($n^{-2}\sum_t u_t^2 M_t$) that does not depend on h, $CV_{lc,L}(h)$ is the leading term of $CV(h)$. This leading term consists of two parts: the $O_p(h/n)$ term corresponding to the leading bias term, and the $(n^2h)^{-1}$ term from the leading variance term. The selected bandwidth balances these two terms, and we obtain $\hat{h}_{lc} = O_p(n^{-1/2})$, as stated in Theorem 2.1(ii).

A fundamental difference between the result presented in Theorem 2.1 and previously reported results (when dealing with independent or weakly dependent data) is that the "optimal" bandwidth is stochastic even asymptotically. More specifically, let the CV-selected bandwidth be $\hat{h}_{lc} = \hat{c}n^{-\alpha}$. With weakly dependent or independent data, it is well known that $\alpha = 1/5$ and $\hat{c} \xrightarrow{p} c_{opt}$, where $c_{opt} > 0$ is a nonstochastic (optimal) constant so that $\hat{h}_{lc}/h_{opt} \xrightarrow{p} 1$, and $h_{opt} = c_{opt}n^{-1/5}$ is the nonstochastic benchmark (optimal) bandwidth (see Härdle, Hall, and Marron 1992). In contrast, when X_t is an $I(1)$ process, Theorem 2.1 states that $\alpha = 1/2$, and that $\hat{c}$ does not converge to a nonstochastic constant, but instead $\hat{c}$ has a well-defined nondegenerate limiting distribution. Simulation results in Section 4 confirm the theoretical results of Theorem 2.1. There we show that, as the sample size n increases, $\hat{h}_{lc}$ shrinks to 0, whereas $\hat{c}$ has a stable nondegenerate distribution. The next theorem describes the asymptotic behavior for the LL–CV–selected bandwidth.

Theorem 2.2. Let $\hat{h}_{ll}$ denote the CV-selected bandwidth based on the local linear estimation method. Under Assumptions A1–A5 and A6*, we have

$$\text{(i)}\quad CV(h) - n^{-2}\sum_{t=1}^{n} u_t^2 M_t - CV_{ll,L}(h) = o_p(h^4 + (n^2h)^{-1}), \tag{2.11}$$

where

$$CV_{ll,L}(h) = (1/4)h^4 B_{x,(2)}\kappa_2^2 E\big[\big(\beta_t^{(2)}\big)^2 M_t\big] + (n^2h)^{-1}\nu_0(K)\sigma_u^2 B_{x,(2)}^{-1}\zeta_{u,1}^2 \int M(z)\,dz, \tag{2.12}$$

$\kappa_2 = \int v^2 K(v)\,dv$, and $\beta_t^{(2)} = d^2\beta(z)/dz^2|_{z=Z_t}$;

$$\text{(ii)}\quad n^{2/5}\hat{h}_{ll} - \left(\frac{4\sigma_u^2\nu_0(K)\int M(z)\,dz\,\zeta_{u,1}^2}{B_{x,(2)}^2\kappa_2^2 E[(\beta_t^{(2)})^2 M_t]}\right)^{1/5} \xrightarrow{p} 0. \tag{2.13}$$

The proof of Theorem 2.2 is given in supplemental Appendix B, where we show that $CV_{ll,L}(h) = O_p(h^4 + (n^2h)^{-1})$. The $O_p(h^4)$ term corresponds to the leading bias term, and the $O_p(n^2h)^{-1}$ term is the leading variance term. The "optimal" h balancing the two terms has an order of $n^{-2/5}$. We explain why the leading bias term from the LL estimation method differs from that obtained from the LC estimation method in Section 2.3.

Theorems 2.1 and 2.2 imply that $CV_{lc,L}(\hat{h}_{lc}) = O_p(n^{-3/2})$ and $CV_{ll,L}(\hat{h}_{ll}) = O_p(n^{-8/5})$, respectively. Thus the LC–CV method gives stochastically a larger average squared error than that obtained from the LL–CV method, indicating that the LL–CV method dominates the LC–CV method. This is in sharp contrast to the existing results obtained for independent or weakly dependent data, because it is well known that for independent or weakly dependent data cases, the CV functions for the local constant and the local linear methods have the same rate of convergence.

Note that the result of Theorem 2.1 requires that $\beta(z)$ be a nonconstant function, and that Theorem 2.2 assumes that $\beta(z)$ is nonlinear in z. Here we briefly comment on what happens if these assumptions are violated. First, if $\Pr\{\beta(Z_t) = c\} = 1$ for some constant c, then the true model reduces to a linear cointegration model. Ideally, one would like to select a sufficiently large h in this case, because when $h = +\infty$, $\hat{\beta}(z)$ becomes the least squares estimator of the constant parameter c. However, it can be shown that neither $\hat{h}_{lc}$ nor $\hat{h}_{ll}$ will converge ∞ in this case. Moreover, h will not converge to 0, so our Theorems 2.1 and 2.2 do not cover the case where $\beta(\cdot)$ is a constant function. We conjecture that the CV-selected bandwidth has a tendency to take large positive values but will not diverge to infinity even as $n \to \infty$. Simulations reported in Section 4 support our conjecture. The asymptotic behavior of the CV-selected bandwidth when the true regression model is linear (β is a constant) is quite complex, and it is beyond our present capabilities to derive the asymptotic distribution of the CV-selected bandwidth in this case.

Next, if $\Pr\{\beta(Z_t) = a + bZ_t\} = 1$, or $\beta(z)$ is a linear function in z, then it can be shown that in this case, Theorem 2.1 still holds true so that $\hat{h}_{lc} = O_p(n^{-1/2})$ ($\hat{h}_{lc}$ still converges to 0), whereas $\hat{h}_{ll}$ converges to neither 0 nor ∞. In Section 4 we use simulations to investigate the behavior of $\hat{h}_{cv}$ when $\beta(z) = a$ and $\beta(z) = a + bz$. We also examine a spurious regression case where $\beta(z) \equiv 0$ and u_t is an $I(1)$ process. The theoretical investigation of $\hat{h}_{cv}$ under a spurious regression model is quite complicated and is beyond the scope of this article.

2.3 Endogenous Regressor Case

For a linear cointegration model, it is well known that when X_t and u_t are correlated, the ordinary least squares (OLS) estimator of the cointegrating coefficient is still n-consistent but has an additional bias term of order n^{-1} (see Phillips and Hansen 1990; Phillips 1995). Recently, Wang and Phillips (2009b, theorem 3.2) considered a nonparametric cointegration model and showed that the asymptotic analysis remains unchanged even when X_t is correlated with u_t, provided that the correlation is not overly persistent. However, Wang and Phillips' framework is quite different from the semiparametric model that we consider here, because in their model, the nonstationary variable enters the model nonparametrically, whereas in our semiparametric model, the nonparametric component Z_t is a stationary variable. To the best of our knowledge, in the framework of a semiparametric varying-coefficient model with $I(1)$ regressors, the endogeneity issue has not been addressed. A priori, it is not clear whether an endogenous regressor will lead to a nonnegligible bias term. In this section we show that the asymptotic distribution of $\hat{\beta}(z)$ remains unchanged when X_t is correlated with u_t under some standard regularity conditions. For expositional simplicity, we consider only the local constant estimator in this section; a similar result can be shown to hold true for the local linear estimator.

Let $K_{h,tz} = h^{-1}K((Z_t - z)/h)$. Then the local constant estimator of $\beta(z)$ is given by

$$\hat{\beta}(z) = \beta(z) + \left(\frac{1}{n^2}\sum_{t=1}^{n} X_t^2 K_{h,tz}\right)^{-1} \times \left\{\frac{1}{n^2}\sum_{t=1}^{n} X_t^2\big(\beta(Z_t) - \beta(z)\big)K_{h,tz} + \frac{1}{n^2}\sum_{t=1}^{n} X_t u_t K_{h,tz}\right\} \equiv \beta(z) + A_{1n}^{-1}(A_{2n} + A_{3n}), \tag{2.14}$$

where the definitions of A_{jn} should be apparent ($j = 1, 2, 3$). Allowing X_t and u_t to be correlated may affect the asymptotic behavior only of A_{3n}, because other terms do not depend on u_t. We show that the asymptotic behavior of A_{3n} remains unchanged when Assumption A3 is replaced by Assumption A3$''$. Therefore, our result remains the same even when X_t is correlated with u_t.

Using $X_t = \sum_{s=1}^{t} v_t$, we have $A_{3n} = n^{-2}\sum_{t=2}^{n} X_{t-1}u_t K_{h,tz} + n^{-2}\sum_{t=1}^{n} v_t u_t K_{h,tz}$, where the stochastic property of v_t is described in Assumption A3$''$. If u_t is a martingale difference sequence as defined in Assumption A3, or if X_t is strictly exogenous with $E(u_t|Z_t) = 0$, we have $E(A_{3n}) = 0$. However, allowing X_t to be endogenous and u_t to be serially correlated generally leads to $E(A_{3n}) \neq 0$. Assumption A3$''$ ensures that $\sup_{z\in\mathcal{M}} n^{-1}\sum_{t=1}^{n} |v_t u_t| K_{h,tz} = O_p(1)$, and applying Assumption A6 and theorem 4.1 of De Jong and Davidson (2000) gives

$$n\sqrt{h}A_{3n} = \sum_{t=1}^{n} \frac{X_t}{\sqrt{n}}\frac{u_t}{\sqrt{n}}\sqrt{h}K_{h,tz} \xrightarrow{d} \int_0^1 B_x(r)\, dB_{u,z}(r) + \Lambda, \tag{2.15}$$

where $\Lambda = \lim_{n\to\infty}\Lambda_n$ with $\Lambda_n \equiv E(\sqrt{h}/n\sum_{t=1}^{n} X_{t-1}u_t K_{h,tz})$.

We show $\Lambda = 0$ below. Let $E(\cdot|z) = E(\cdot|Z_t = z)$. We assume that both $E(v_t^2|z)$ and $E(u_t^2|z)$ are bounded by a function (of z) that has a finite second moment. By Assumptions A1 and A3$''$, $\{(Z_t, v_t, u_t)\}_{t=1}^{n}$ is a β-mixing process. Thus, applying lemma 1 of Yoshihara (1976), we obtain

$$|\Lambda_n| \le \frac{\sqrt{h}}{n}\sum_{t=1}^{n}\sum_{s=1}^{t-1} |E(v_s u_t K_{h,tz})| \le M\frac{\sqrt{h}}{n}\sum_{t=1}^{n}\sum_{s=1}^{t-1} h^{-\delta/(1+\delta)}|t-s|^{-q\delta/(1+\delta)} = O\big(\sqrt{h}h^{-\delta/(1+\delta)}\big(1 - n^{1-q\delta/(1+\delta)}\big)\big) = o(1). \tag{2.16}$$

Hence, $\Lambda_n = o(1)$ and this implies that $\Lambda = \lim_{n\to\infty}\Lambda_n = 0$.

Therefore, the bias term due to the correlation between X_t and u_t is asymptotically negligible, which differs from the linear regression model case. Replacing $K_{h,tz}$ by 1 in (2.14), we obtain the OLS estimator of β. It is easy to see that for a linear cointegration model, the bias term is $\Lambda_0 \stackrel{\text{def}}{=} E(n^{-1}\sum_{t=1}^{n} X_t u_t) \neq 0$. In fact, it is easy to see that if both v_s and u_t are iid series but with $E(v_t u_t) \neq 0$, we have $\Lambda_0 = E(v_t u_t) \neq 0$. For our semiparametric model, even if $E(v_t u_t) \neq 0$, we have $\Lambda = 0$, because our bias term has an additional factor $\sqrt{h}K_{h,tz}$ and $E[\sqrt{h}K_{h,tz}] = O(\sqrt{h}) = o(1)$.

We use the foregoing decomposition to provide an intuitive explanation of Theorem 2.1 as to why $CV_{lc,L}(h) = O_p(h/n + (n^2h)^{-1})$. Applying Hansen's (1992) theorem 3.3 yields

$$A_{1n} = n^{-2}\sum_{t=1}^{n} X_t^2 K_{h,tz} = n^{-2}\sum_{t=1}^{n} X_t^2 E(K_{h,tz}) + o_p(1) \xrightarrow{d} f(z)\int_0^1 B_x^2(r)\, dr = O_e(1).$$

Here the notation $O_e(a_n)$ means an exact probability order of $O_p(a_n)$, but is not $o_p(a_n)$.

Next consider A_{2n}. By adding/subtracting terms, we rewrite $A_{2n} = A_{2n,1} + A_{2n,2}$, where

$$A_{2n,1} = \frac{1}{n^2}\sum_{t=1}^{n} X_t^2 E\big[(\beta(z_t) - \beta(z))K_{h,tz}\big] = c_{1n}[h^2 \operatorname{Bias}(z) + O(h^4)] \tag{2.17}$$

with $c_{1n} = n^{-2}\sum_{t=1}^{n} X_t^2$, $\operatorname{Bias}(z) = (1/2)\kappa_2[f(z)\beta^{(2)}(z) + f^{(1)}(z)\beta^{(1)}(z)]$, and $\kappa_2 = \int K(v)v^2\, dv$. By Assumption A6 and the independence between $B_x(\cdot)$ and $W_\beta(\cdot)$, we have

$$\begin{aligned} A_{2n,2} &= \sqrt{\frac{h}{n}}\sum_t \frac{X_{t-1}^2}{n}\frac{\eta_t(z)}{\sqrt{nh}} + \frac{\sqrt{h}}{n}\sum_t \frac{X_{t-1}}{\sqrt{n}}\frac{v_t\eta_t(z)}{\sqrt{nh}} + n^{-2}\sum_t v_t^2\eta_t(z) \\ &= \sqrt{\frac{h}{n}}\sum_t \frac{X_{t-1}^2}{n}\frac{\eta_t(z)}{\sqrt{nh}} + O_p\big(n^{-1}\sqrt{h} + n^{-1}h^2\big) \\ &= \sqrt{\frac{h}{n}}\bigg[f(z)\big(\beta^{(1)}(z)\big)^2\nu_2(K) \times \int_0^1 B_x(r)^2\, dW_\beta(r) + o_p(1)\bigg], \end{aligned} \tag{2.18}$$

where $\eta_t(z) = e_t(z) - E(e_t(z))$ and $e_t(z) = (\beta(Z_t) - \beta(z))K_{h,tz}$.

Equation (2.17) shows that the order of $A_{2n,1}$ is related to $E[e_t(z)]$, the mean part of $e_t(z)$, whereas Equation (2.18) tells us that the order of $A_{2n,2}$ is determined by the variance part of $e_t(z)$ as $\eta_t(z)$ has mean 0. With weakly dependent data, it is well known that the mean part of $e_t(z)$ dominates the variance part of $e_t(z)$. However, if the local constant estimation method is used to analyze integrated time series, then the variance part of $e_t(z)$ dominates the mean part of $e_t(z)$. To see this, we square each term to get rid of the square root expressions, which gives $A_{2n,1}^2 = O_p(h^4)$, $A_{2n,2}^2 = O_p(h/n)$, and $A_{3n}^2 = O_p((n^2h)^{-1})$. These results imply that $A_{2n,2}^2 + A_{3n}^2 = O_p(h/n + (n^2h)^{-1})$ has a stochastic order larger than that of $A_{2n,1}^2 = O_p(h^4)$.

The foregoing analysis also provides an explanation as to why the leading term of the CV function, $CV_{lc,L}(h)$, contains two terms of orders $O_p(h/n)$ and $O_p((n^2h)^{-1})$, which are related to the variances of $e_t(z)$ and $u_tK_{h,tz}$, respectively. The $O_p(h^4)$ term corresponding to the squared mean of $e_t(z)$ is asymptotically negligible. However, if the local linear estimation method is used instead of the local constant method, $\beta(Z_t) - \beta(Z_s)$ must be replaced by $R(Z_t, Z_s) \stackrel{\text{def}}{=} \beta(Z_t) - \beta(Z_s) -$

$\beta^{(1)}(Z_t)(Z_t - Z_s)$. In this case the two leading terms, $O_p(h^4)$ and $O_p((n^2h)^{-1})$ [of $CV_{ll,L}(h)$], correspond to the (squared) mean of $R(Z_t, Z_s)K_{h,tz}$ and the variance of $u_tK_{h,tz}$, whereas the term related to the variance of $R(Z_t, Z_s)K_{h,tz}$, which has order $O_p(h^4/n)$ (this term differs from the LC case), is asymptotically negligible. Therefore, the reason for the different convergence rates of $CV_{lc}(h)$ and $CV_{ll}(h)$ is that, unlike in the weakly dependent data case, the LC and the LL estimator have different stochastic orders in A_{n2}.

It can be shown that the results of Theorems 2.1 and 2.2 hold true when X_t and u_t are correlated, but the proofs will be lengthy and tedious and will provide no additional insight into the problem. Thus to save space, we do not pursue proofs for Theorems 2.1 and 2.2 for the endogenous X_t case. In Section 4 we report Monte Carlo simulations that allow X_t to be correlated with u_t. The simulation results reported there support the foregoing theoretical analysis and show that the estimation results are virtually unaffected whether or not X_t and u_t are correlated.

2.4 A Partially Linear Varying-Coefficient Model

In this section we consider the CV selection of the bandwidth h when estimating the following partially linear varying-coefficient model:

$$Y_t = S_t^T\gamma + X_t\beta(Z_t) + u_t, \tag{2.19}$$

where S_t is a $(d-1)\times 1$ vector of $I(1)$ variables, and γ is a $(d-1)\times 1$ vector of constant parameters. For expositional simplicity, we assume that X_t is a scalar $I(1)$ variable. Note that the foregoing partially linear model is a special case of a general varying-coefficient model $Y_t = S_t^T\alpha(Z_t) + X_t\beta(Z_t) + u_t$, where the restriction that $\alpha(z) \equiv \gamma$, a vector of constant parameters, is imposed.

We propose using a profile least squares approach to estimate γ. First, we treat γ as if it were known and rewrite (2.19) as $Y_t - S_t^T\gamma = X_t\beta(Z_t) + u_t$. Then the local constant estimator of $\beta(Z_t)$ is given by

$$\tilde\beta(Z_t) = \left(\sum_s X_s^2 K_{h,ts}\right)^{-1} \sum_s X_s(Y_s - S_s^T\gamma)K_{h,ts} \equiv A_{2t} - A_{1t}^T\gamma, \tag{2.20}$$

where $A_{1t} = (\sum_s X_s^2 K_{h,ts})^{-1}\sum_s X_sS_sK_{h,ts}$ and $A_{2t} = (\sum_s X_s^2 \times K_{h,ts})^{-1}\sum_s X_sY_sK_{h,ts}$. Note that $\tilde\beta(Z_t)$ defined in (2.20) is infeasible, because it depends on the unknown parameter γ. We provide a feasible estimator for $\beta(Z_t)$ in (2.23). Replacing $\beta(Z_t)$ by $\tilde\beta(Z_t)$ in (2.19) and rearranging terms, we obtain

$$Y_t - X_tA_{2t} = (S_t - X_tA_{1t})^T\gamma + \epsilon_t, \tag{2.21}$$

where $\epsilon_t \equiv Y_t - X_tA_{2t} - (S_t - X_tA_{1t})^T\gamma$. Applying the OLS method to model (2.21) leads to

$$\hat\gamma = \left[\sum_t (S_t - X_tA_{1t})(S_t - X_tA_{1t})^T\right]^{-1} \times \sum_t (S_t - X_tA_{1t})(Y_t - X_tA_{2t}). \tag{2.22}$$

Replacing γ with $\hat\gamma$ in (2.20), we obtain a feasible leave-one-out estimator of $\beta(Z_t)$ given by

$$\hat\beta_{-t}(Z_t) = \left(\sum_{s\neq t} X_s^2 K_{h,st}\right)^{-1}\sum_{s\neq t} X_s(Y_s - S_s^T\hat\gamma)K_{h,st}. \tag{2.23}$$

It can be shown that $S_t^T(\gamma - \hat\gamma)$ has a stochastic order smaller than $X_t(\beta_t - \hat\beta_{-t})$. Thus the leading term of $CV_{\hat\gamma}(h) = n^{-2} \times \sum_{t=1}^n [Y_t - S_t^T\hat\gamma - X_t\hat\beta_{-t}(Z_t)]^2M_t$ is given by $CV_\gamma(h) = n^{-2} \times \sum_{t=1}^n [Y_t - S_t^T\gamma - X_t\hat\beta_{-t}(Z_t)]^2M_t$. Obviously, $CV_\gamma(h)$ is the same as $CV(h)$ defined in (2.1). This is because if we define $\tilde Y_t = Y_t - S_t^T\gamma$, then model (2.19) can be written as $\tilde Y_t = X_t\beta(Z_t) + u_t$, which is identical to model (1.1). Thus the results of Theorem 2.1 remain valid for the partially linear model (2.19).

The foregoing discussion is based on the local constant estimation method. The local linear estimation method could be used as well. In this case, $\tilde\beta(z)$ in (2.20) is replaced by the local linear estimator, which is also linear in γ. The remaining estimation steps are similar to the local constant estimation case discussed earlier. The asymptotic behavior of the LS–CV–selected h is the same as presented in Theorem 2.2.

3. X_t CONTAINS BOTH $I(0)$ AND $I(1)$ COMPONENTS

The results in Section 2 are obtained assuming that X_t contains only $I(1)$ components. This assumption simplifies our theoretical analysis and makes it easier to see why the LC and LL estimation methods lead to different rates of convergence. However, in practice, a cointegration model also may include an intercept term and some other stationary variables in addition to some $I(1)$ regressors. Therefore, in this section we investigate the asymptotic behavior of the LS–CV–selected bandwidth when X_t contains both $I(0)$ and $I(1)$ components. Specifically, we consider $X_t = (X_{1t}, X_{2t})^T$ with X_{1t} and X_{2t} being $I(0)$ and $I(1)$ variables, respectively. We replace Assumption A6* by the following assumption:

B Let $X_t = (X_{1t}, X_{2t})^T$, where X_{2t} satisfies Assumption A6* given in Section 2 and X_{1t} is a strictly stationary β-mixing sequence of size $-(2+\delta')/\delta'$ for some $0 < \delta' < \delta$ with $E(|X_{1t}|^{2+\delta}) < M < \infty$. In addition, both $E(X_{1t}|Z_t = z)$ and $E(X_{1t}^2|Z_t = z)$ are twice continuously differentiable over $z \in \mathcal{M}$.

Given that the local linear method has a smaller asymptotic MSE than the local constant method, we consider only the local linear estimation method in this section.

Theorem 3.1. Let $\hat h$ denote the CV-selected bandwidth via the local linear method. Under Assumptions A1–A5 and B, we have

$$n^{2/5}\hat h - c_{3n} \xrightarrow{p} 0, \tag{3.1}$$

where c_{3n} is a well-defined $O_e(1)$ random variable that is determined by (B1.2) and (B.13) in supplemental Appendix B.

Comparing the result of Theorem 3.1 with that of Cai, Li, and Park (2009), we see that $\hat h$ (defined in Theorem 3.1) is optimal for estimating $\beta_2(\cdot)$, the coefficient function associated with the integrated variable, because Cai, Li, and Park (2009) showed that the optimal h for estimating $\beta_1(\cdot)$ should have an order of $n^{-1/5}$. The reason that the CV method selects an optimal h for estimating $\beta_2(\cdot)$ is because $X_{2t}^T\beta_2(Z_t)$ is the dominant component of $X_{1t}^T\beta_1(Z_t) + X_{2t}^T\beta_2(Z_t)$. Thus, to minimize the CV function, one must choose an h that is optimal for estimating $\beta_2(\cdot)$, the coefficient function of the $I(1)$ covariate. In the next section, we examine a two-step estimation method as suggested by Cai, Li, and Park (2009) via Monte Carlo simulations, where after obtaining $\hat{\beta}_1(Z_t)$ and $\hat{\beta}_2(Z_t)$ in the first step, we use $Y_t - X_{2t}\hat{\beta}(Z_t)$ as the new dependent variable and estimate $\beta_1(Z_t)$ using a new CV-selected smoothing parameter in the second step. We show that this two-step method leads to improved estimation results for $\beta_1(z)$.

4. MONTE CARLO SIMULATIONS

4.1 Case (a): X_t Is $I(1)$

We consider the following data-generating process (DGP):

$$Y_t = X_t\beta(Z_t) + u_t, \qquad Z_t = 0.5Z_{t-1} + \eta_{3t}, \qquad \text{and}$$
$$u_t = (\eta_{2t} + \theta\eta_{1t})/\sqrt{1+\theta^2},$$

where $X_t = X_{t-1} + \eta_{1t}$, $(\eta_{1t}, \eta_{2t})^T$ is iid N$(0, I_2)$, η_{3t} is iid Uniform[0, 1], and $X_0 = 0$. The data-generating mechanism for u_t is the same as that of Wang and Phillips (2009b). It is easy to show that $\text{corr}(\Delta X_t, u_t) = \theta/\sqrt{1+\theta^2}$. We take two values for θ, $\theta = 0$ and $\theta = 0.2$, where the former case implies that X_t and u_t are independent of one another, whereas the latter case gives that $\text{corr}(\Delta X_t, u_t) = 0.1961$. Z_t is a stationary AR(1) process with the innovation η_{3t} taking values in [0, 1].

We consider four different DGPs and index them by DGP$_{i,j}$, specifically, $i = 1$ if $\theta = 0$, $i = 2$ if $\theta = 0.2$, $j = 1$ if $\beta(z) = 1 + z + 2z^2$ whose value increases as z increases, and $j = 2$ if $\beta(z) = \sin(3z)$, which is not monotone and has more curvature than the quadratic function. Comparing DGP$_{1,j}$ with DGP$_{2,j}$ for $j = 1$ and 2, we aim to show that X_t is allowed to be correlated with u_t as discussed in Section 2.3. Comparing DGP$_{i,1}$ with DGP$_{i,2}$ for $i = 1$ and 2, we examine how extra curvature of a coefficient function affects the finite-sample performance of our proposed CV method.

The sample sizes are $n = 100$, 200, and 400. The number of simulations is $m = 1000$. We report the square root of the average squared error, $RMSE = \sqrt{AMSE}$, and the mean absolute bias, $MABIAS$, where for the lth simulation, we define $AMSE_l = n^{-1}\sum_{t=1}^{n}(\beta(z_t) - \hat{\beta}_l(z_t))^2$ and $MABIAS_l = n^{-1}\sum_{t=1}^{n}|\beta(z_t) - \hat{\beta}_l(z_t)|$.

Figure 1 plots the kernel density functions of the CV-selected constant $\hat{c}$ and the bandwidth $\hat{h}$, where $\hat{h} = \hat{c}n^{-\alpha}$ with $\alpha = 1/2$ for the LC–CV method and $\alpha = 2/5$ for the LL–CV method. Figure 1 is obtained from DGP$_{2,1}$, where the dotted line represents $n = 100$, the dashed line represents $n = 200$, and the solid line represents $n = 400$. As predicted by Theorems 2.1 and 2.2, we see that the CV-selected bandwidth $\hat{h}$ becomes smaller as sample size increases, and that $\hat{c}$ does not converge to a constant as the estimated density function for $\hat{c}$ is rather stable for different sample sizes.

Table 1 reports the mean and standard deviation (over the 1000 replications) of the RMSE and the MABIAS. Several interesting patterns are observed. First, the LL–CV method has

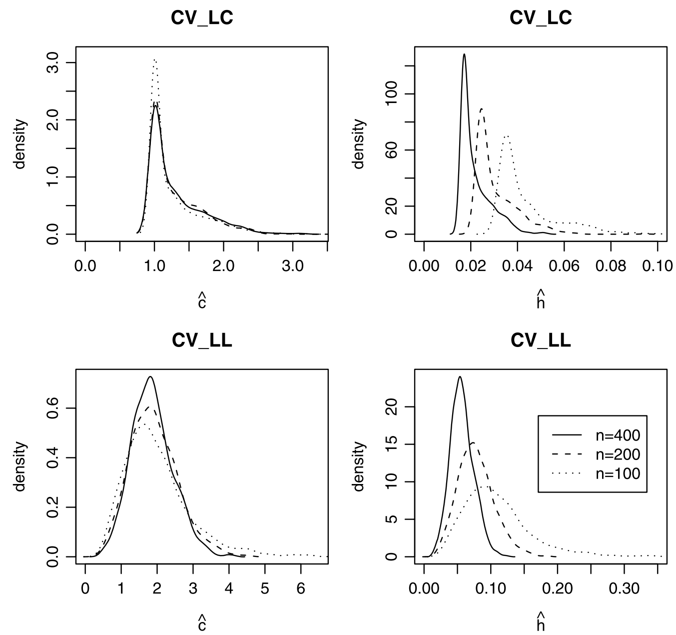

Figure 1. Kernel density estimate of CV-selected constant and bandwidth.

smaller (mean value of) RMSE and MABIAS compared with the LC–CV method. This is consistent with our theory, because the LL–CV method has smaller asymptotic MSE than the LC–CV method, as shown in Theorems 2.1 and 2.2. Second, the estimation efficiency gain of the LL–CV method over the LC–CV method is more pronounced for DGP$_{i1}$ than for DGP$_{i2}$ for $i = 1, 2$ when the unknown curve has more curvature (i.e., more nonlinearity). Third, comparing the results of DGP$_{1,j}$ with DGP$_{2,j}$ for $j = 1$ and 2 confirms our theoretical analysis presented in Section 2.3 that the CV method is valid even when u_t and X_t are contemporaneously correlated.

To show how the CV-selected bandwidth behaves when $\beta(\cdot)$ is constant, Table 2 reports the first quartile, median, mean, and third quartile of the CV-selected bandwidths, along with the RMSE and the MABIAS for the LC and LL kernel estimators, where $\beta(z) \equiv 1$. The results for $\theta = 0.2$ and $\theta = 0$ are very similar, so we only report the results for $\theta = 0$. In addition, the bandwidth has an upper bound of five times the interquartile range of Z_t (i.e., 9.45), because allowing the bandwidth to increase further will improve the CV value only beyond the sixth decimal point. In this case, with a much larger selected bandwidth, the LC–CV method gives smaller RMSE and MABIAS than the LL–CV method. We also see that the median value of the LC–CV–selected h is quite stable and does not seem to change as n increases. This result is consistent with our earlier finding that when $\beta(z)$ is a constant function, the CV method tends to choose a large value of h, but the CV-selected h will converge neither to ∞ nor to 0.

Table 3 reports the results when X_t and Y_t are two independent random-walk processes without drift (a spurious regression model). In this case the $CV(h)$ objective function is quite flat, and the RMSE and the MBIASE do not change much even when the h changes substantially, resulting a wide range of selected $\hat{h}_{cv}$, as shown in Table 3. Taken together, the results in Tables 2 and 3 show that an unusually large LC–CV–selected

Table 1. The LC–CV vs. the LL–CV method when X_t is an $I(1)$ variable

		RMSE		MABIAS		RMSE		MABIAS	
Method	n	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.
		DGP_{11}				DGP_{21}			
CV–LC	100	0.2828	0.0513	0.0213	0.0462	0.2864	0.0477	0.0224	0.0251
	200	0.2203	0.0392	0.0174	0.2197	0.2223	0.0364	0.0137	0.0872
	400	0.1701	0.0277	0.0073	0.0983	0.1714	0.0262	0.0043	0.0023
CV–LL	100	0.2229	0.0539	0.0114	0.0157	0.2325	0.0557	0.0148	0.0224
	200	0.1669	0.0391	0.0052	0.0105	0.1737	0.0411	0.0066	0.0061
	400	0.1249	0.0273	0.0022	0.0013	0.1296	0.0283	0.003	0.0014
		DGP_{12}				DGP_{22}			
CV–LC	100	0.2429	0.0508	0.0131	0.0159	0.2468	0.0504	0.0164	0.0241
	200	0.1861	0.0365	0.0058	0.0107	0.1888	0.0366	0.0071	0.0058
	400	0.1428	0.0263	0.0025	0.0025	0.1448	0.0263	0.0033	0.0022
CV–LL	100	0.2274	0.0524	0.0124	0.0156	0.2315	0.0525	0.0154	0.0109
	200	0.1704	0.0385	0.0055	0.0108	0.1735	0.0383	0.0069	0.0064
	400	0.1291	0.0272	0.0024	0.0018	0.1309	0.0271	0.0032	0.0017

$\hat{h}$ may indicate $\beta(z) = c$, a constant, or may suggest a spurious relationship between Y_t and X_t. Further diagnostics are needed to distinguish these two possibilities.

4.2 Case (b): X_t Contains Both $I(0)$ and $I(1)$ Variables

We consider the following DGP:

$$Y_t = X_{1t}\beta_1(Z_t) + X_{2t}\beta_2(Z_t) + u_t, \quad (4.1)$$

where $X_{2t} = X_{2,t-1} + \eta_{1t}$, $X_{1t} = 1 + 0.5X_{1,t-1} + \eta_{4t}$ with $\eta_{4t} \sim \text{iid}\,\mathrm{N}(0,1)$, u_t and Z_t are generated as in Section 4.1, $\beta_1(z) = 1 + z + 2z^2$, and $\beta_2(z) = \sin(3z)$. We use DGP_3 and DGP_4 to denote the cases where $\theta = 0$ and $\theta = 2$, respectively. For DGP_4, $\mathrm{corr}(\Delta X_{2t}, u_t) = 0.8944$, a rather high contemporaneous correlation.

Table 4 presents simulation results for two estimators: $\hat{\beta}(\cdot)$, the LL–CV estimator using the LL–CV–selected bandwidth $\hat{h}$, and $\tilde{\beta}_1(\cdot)$, the two-step LL–CV estimator for $\beta_1(\cdot)$ in which we construct a new dependent variable $Y_t - X_{2t}\hat{\beta}_2(Z_t)$ and reestimate $\beta_1(Z_t)$ in the second stage, where $\hat{\beta}_2(Z_t)$ is the first-stage estimator of $\beta_2(Z_t)$. We also selected a new bandwidth via the LL–CV method in the second-stage estimation. We observe the following. First, the coefficient curve $\beta_2(\cdot)$ associated with the $I(1)$ variable is estimated more accurately than the coefficient curve $\beta_1(\cdot)$ associated with the $I(0)$ variable. Second, the second-step estimation of $\beta_1(\cdot)$ by the LL–CV method performs slightly better than the one-step estimation result of $\beta_1(\cdot)$. Finally, Table 4 also indicates that we do not need the strictly exogenous assumption to validate the CV method.

APPENDIX A: PROOF OF THEOREM 2.1

We use the notation $A_n = B_n + (s.o.)$ to denote that $A_n = B_n +$ terms are of smaller order than B_n. We denote $\beta_t = \beta(z_t)$, $\beta_t^{(j)} = \beta^{(j)}(z_t) = d^j\beta(z)/dz^j|_{z=z_t}$, $\hat{\beta}_{-t} = \hat{\beta}_{-t}(z_t)$, $K_{h,ts} = h^{-1}K((z_t - z_s)/h)$, $f_t = f(z_t)$, $M_t = M(z_t)$, and $\sum_{s\neq t} = \sum_{s=1,s\neq t}^{n}$. Also, $\nu_j(K) = \int u^j K(u)^2\,du$, $\kappa_2 = \int u^2 K(u)\,du$, $c_{1n} = n^{-2}\sum_{t=1}^{n} x_t^2$, $c_{2n} = n^{-3}\sum_{t=1}^{n} x_t^4$, and $\mathcal{M} = \{z \in \mathcal{R} : M(z) > 0\}$.

Proof of Theorem 2.1

By (1.2), $\hat{\beta}_{-t} = (\sum_{s\neq t} x_s^2 K_{h,ts})^{-1}\sum_{s\neq t} x_s Y_s K_{h,ts}$. Replacing Y_s in $\hat{\beta}_{-t}$ by $Y_s = x_s\beta_s + u_s = x_s\beta_t + x_s(\beta_s - \beta_t) + u_s$, we obtain

$$\hat{\beta}_{-t} = \beta_t + \hat{A}_t^{-1}(\hat{B}_t + \hat{C}_t), \quad (A.1)$$

where $\hat{A}_t = n^{-2}\sum_{s\neq t} x_s^2 K_{h,ts}$, $\hat{B}_t = n^{-2}\sum_{s\neq t} x_s^2(\beta_s - \beta_t)K_{h,ts}$, and $\hat{C}_t = n^{-2}\sum_{s\neq t} x_s u_s K_{h,ts}$.

Note that for cases of independent or weakly dependent data, $\hat{A}_t^{-1}\hat{B}_t$ and $\hat{A}_t^{-1}\hat{C}_t$ correspond to the bias and variance terms, respectively. Therefore, for convenience we refer these two terms as bias and variance terms.

Table 2. The LC–CV vs. the LL–CV method when X_t is an $I(1)$ variable and $\beta(z) \equiv 1$

		$\hat{h}$				RMSE		MABIAS	
Method	n	$h_{0.25}$	$h_{0.5}$	h_{mean}	$h_{0.75}$	Mean	St. dev.	Mean	St. dev.
CV–LC	100	8.0816	8.3528	8.3508	8.6341	0.0166	0.0168	0.0167	0.0166
	200	8.3893	8.6057	8.5907	8.8187	0.008	0.0081	0.0082	0.0077
	400	8.6099	8.7869	8.7863	8.9654	0.0037	0.0043	0.0041	0.0039
CV–LL	100	8.0823	8.3528	8.3588	8.634	0.0271	0.0214	0.0234	0.0186
	200	8.3893	8.6057	8.5991	8.8187	0.0132	0.0104	0.0114	0.0088
	400	8.6099	8.7869	8.7863	8.9654	0.0066	0.0055	0.0059	0.0046

Table 3. The LC–CV vs. the LL–CV method when X_t and Y_t are independent random walk processes

		$\hat{h}$				RMSE		MABIAS	
Method	n	$h_{0.25}$	$h_{0.5}$	h_{mean}	$h_{0.75}$	Mean	St. dev.	Mean	St. dev.
CV–LC	100	0.1355	0.3146	2.6275	7.8475	0.7366	0.6379	0.7156	0.6332
	200	0.1347	0.2912	2.5119	7.8823	0.6990	0.6113	0.6863	0.6089
	400	0.1589	0.3265	2.5747	6.3165	0.7014	0.6320	0.6941	0.6307
CV–LL	100	0.2393	0.7662	3.8878	8.2403	0.7538	0.6425	0.7233	0.6335
	200	0.1943	0.5494	3.5212	8.4263	0.7157	0.6239	0.6923	0.6086
	400	0.2195	0.7223	3.6277	8.5907	0.7133	0.6443	0.6981	0.6344

Substituting (A.1) into $CV_{0,1}$ defined in (2.3), we obtain

$$\begin{aligned}CV_{0,1} &= n^{-2}\sum_t (x_t\hat{A}_t^{-1}\hat{B}_t)^2 M_t + n^{-2}\sum_t (x_t\hat{A}_t^{-1}\hat{C}_t)^2 M_t \\ &\quad + 2n^{-2}\sum_t x_t^2\hat{A}_t^{-2}\hat{B}_t\hat{C}_t M_t \\ &\equiv CV_1 + CV_2 + 2CV_3, \qquad \text{(A.2)}\end{aligned}$$

where the definitions of CV_j $(j = 1, 2, 3)$ should be apparent.

In Lemmas A.2–A.4 we show that

$$\begin{aligned}CV_1 &= (h/n)B_{x,(2)}^{-1}\nu_2(K)E\Big[M_t f_t^{-1}\big(\beta_t^{(1)}\big)^2\Big] \\ &\quad \times \left[\int_0^1 B_x^2(r)\,dW_\beta(r)\right]^2 + o_p(h/n + h^3), \qquad \text{(A.3)}\end{aligned}$$

$$\begin{aligned}CV_2 &= \frac{\nu_0(K)\sigma_u^2}{n^2 h}B_{x,(2)}^{-1}\left[\int_0^1 B_x(r)\,dW_u(r)\right]^2\int M(z)\,dz \\ &\quad + o_p((n^2h)^{-1}), \qquad \text{(A.4)}\end{aligned}$$

$$CV_3 = o_p(h/n), \qquad \text{(A.5)}$$

where $B_{x,(2)} = \int_0^1 B_x(r)^2\,dr$ and $B_x(r)$, $W_\beta(r)$, and $W_u(r)$ are as defined in Section 2.1. Under Assumptions A1–A3 and A6, $W_\beta(r)$ and $W_u(r)$ are standard Brownian motions independent of the stochastic process $B_x(r)$. Using similar arguments, it can be shown that [$CV_{0,2}$ is as defined in (2.3)],

$$CV_{0,2} = o_p(h/n) + O_p\big((n^2h^{1/2})^{-1}\big) = o_p(h/n). \qquad \text{(A.6)}$$

Combining (A.3)–(A.6), we see that the leading term of $CV_0(h)$ defined in (2.3) is given by

$$\begin{aligned}CV_{lc,L}(h) &= (h/n)\nu_2(K)B_{x,(2)}^{-1} \\ &\quad \times E\Big[M_t f_t^{-1}\big(\beta_t^{(1)}\big)^2\Big]\left[\int_0^1 B_x^2(r)\,dW_\beta(r)\right]^2 \\ &\quad + \frac{\nu_0(K)\sigma_u^2}{n^2h}B_{x,(2)}^{-1}\left[\int_0^1 B_x(r)\,dW_u(r)\right]^2 \\ &\quad \times \int M(z)\,dz. \qquad \text{(A.7)}\end{aligned}$$

Obviously, $CV_{lc,L}(h)$ is minimized at

$$\begin{aligned}h_0 &= \sigma_u n^{-1/2}\sqrt{\frac{\zeta_{u,1}^2\nu_0(K)\int M(z)\,dz}{\zeta_{\beta,2}^2\nu_2(K)\int M(z)(\beta^{(1)}(z))^2\,dz}} \\ &= O_e\big(n^{-1/2}\big), \qquad \text{(A.8)}\end{aligned}$$

where $\zeta_{ij} = \int_0^1 B_x^j(r)\,dW_i(r)$ for $i = \beta, u$ and $j = 1, 2$.

The foregoing result can be extended to $\sup_{h\in\mathcal{H}_n}|CV_0(h) - CV_{lc,L}(h)| = o_p(n^{-1/2})$, where $\mathcal{H}_n = (an^{-0.6}, bn^{-0.4})$ for some $a > 0$ and $b > 0$. This completes the proof of Theorem 2.1.

Lemma A.1. Under Assumptions A1, A2, A4, A5, and A6, we have

$$\sup_{z_t\in\mathcal{M}}|\hat{A}_t - \tilde{\mu}_1(z_t)| = O_p\left(\frac{(\ln n)^{1/2}}{(nh)^{1/4}}\right), \qquad \text{(A.9)}$$

where $\tilde{\mu}_1(z) = (n^{-2}\sum_t x_t^2)f(z) \equiv c_{1n}f(z)$.

Table 4. The LC–CV vs. the LL–CV method when X_{1t} is an $I(0)$ variable and X_{2t} is an $I(1)$ variable

		$\theta = 0$				$\theta = 2$			
		MABIAS		RMSE		MABIAS		RMSE	
Estimator	n	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.	Mean	St. dev.
$\hat{\beta}_2(z)$	100	0.0724	0.0355	0.1407	0.1231	0.0777	0.0417	0.1482	0.1181
	200	0.0395	0.0182	0.0869	0.0761	0.042	0.0211	0.0962	0.0969
	400	0.0216	0.0093	0.0567	0.0547	0.0225	0.0105	0.0601	0.0565
$\hat{\beta}_1(z)$	100	0.1577	0.0955	0.3313	0.5636	0.1616	0.0842	0.3153	0.3329
	200	0.1061	0.0563	0.2233	0.3771	0.1082	0.0487	0.2246	0.2297
	400	0.0744	0.0349	0.1734	0.2541	0.0745	0.0314	0.163	0.1622
$\tilde{\beta}_1(z)$	100	0.146	0.0685	0.244	0.1784	0.1543	0.0738	0.2647	0.2191
	200	0.1016	0.0457	0.1786	0.1243	0.1056	0.0441	0.1982	0.1486
	400	0.0717	0.0293	0.1337	0.0894	0.0737	0.0291	0.1518	0.1114

Proof. By (2.5), we have $\Delta_t = \hat{A}_t - \tilde{\mu}_1(z_t) = n^{-2}\sum_{s\neq t} x_s^2 \times e_{st} + O(n^{-1}\ln\ln n)$, where $e_{st} = K_{h,ts} - f(z_t)$ and $E|e_{st}| = O(1)$ for all $s \neq t$. We follow the proof technique of Hansen (1992, pp. 497–498). For any small $\lambda \in (0,1)$, let $N = [\lambda^{-1}]$ ($[\lambda^{-1}]$ denote the integer part of λ^{-1}), $s_k = [kn/N] + 1$, $s_k^* = s_{k+1} - 1$, $N^* = N - 1$, and $s_k^{**} = \min\{s_k^*, n\}$. Also denote $M_{n,s} = n^{-1}x_s^2$ for any s and $M_n(r) = M_{n,[nr]}$ for $r \in [0,1]$. By Assumption A1, and applying the continuous mapping theorem, we have $M_n(\cdot) \Rightarrow B_x^2(\cdot)$. We then have

$$\begin{aligned}
\left|n^{-2}\sum_{s\neq t} x_s^2 e_{st}\right|
&\leq \left|n^{-1}\sum_{k=0}^{N^*}\sum_{\substack{s=s_k\\ s\neq t}}^{s_k^{**}} M_{n,s_k} e_{st}\right| + \left|n^{-1}\sum_{k=0}^{N^*}\sum_{\substack{s=s_k\\ s\neq t}}^{s_k^{**}} (M_{ns} - M_{n,s_k}) e_{st}\right| \\
&\leq \sup_{r\in[0,1]} M_n(r) n^{-1}\sum_{k=0}^{N^*}\left|\sum_{\substack{s=s_k\\ s\neq t}}^{s_k^{**}} e_{st}\right| \\
&\quad + \sup_{|r-r'|\leq\lambda} |M_n(r) - M_n(r')| n^{-1}\sum_{s\neq t}|e_{st}| \\
&= O_p\left(h^2 + \sqrt{\frac{\ln(n\lambda)}{n\lambda h}}\right) + O_p\left(\sqrt{\lambda\ln\lambda^{-1}}\right) \\
&= O_p\left(\frac{(\ln n)^{1/2}}{(nh)^{1/4}}\right),
\end{aligned} \tag{A.10}$$

where the last equality follows by choosing $\lambda = (nh)^{-1/2}$, and we also use the following results: by Assumption A6, we have

$$\sup_{r\in[0,1]} M_n(r) = O_p(1), \tag{A.11}$$

$$\begin{aligned}
\sup_{|r-r'|\leq\lambda} |M_n(r) - M_n(r')| &= \sup_{|r-r'|\leq\lambda} |B_x^2(r) - B_x^2(r')| + (s.o.) \\
&= O_p\left(\sqrt{\lambda\ln(\lambda^{-1})}\right).
\end{aligned} \tag{A.12}$$

In addition, in deriving the third line of Equation (A.10) we use theorem 6 of Hansen (2008):

$$\begin{aligned}
\sup_{z_t\in\mathcal{M}} n^{-1}\sum_{k=0}^{N^*}\left|\sum_{s=s_k}^{s_k^{**}} e_{st}\right| &\leq \sup_{z_t\in\mathcal{M}}\sup_{s\neq t}\frac{1}{n\lambda}\left|\sum_{j=s}^{s+[n\lambda]} e_{jt}\right| \\
&= O_p\left(h^2 + \sqrt{\ln([n\lambda])/([n\lambda]h)}\right).
\end{aligned}$$

Lemma A.2. Under Assumptions A1, A2, A4, A5, and A6*, (A.3) holds true.

Proof. Define CV_1^0 by replacing $\hat{A}_t$ with $\tilde{\mu}_1(z_t) = c_{1n}f_t$ in CV_1; that is, $CV_1^0 = n^{-2}\sum_t [x_t\hat{B}_t/\tilde{\mu}_1(z_t)]^2 M_t = n^{-2}c_{1n}^{-2} \times \sum_t x_t^2\hat{B}_t^2 M_t f_t^{-2}$, where $\hat{B}_t = n^{-2}\sum_{s\neq t} x_s^2 e_{st}$ and $e_{st} = (\beta_s - \beta_t)K_{h,ts}$. Equation (A.9) implies that CV_1^0 is the leading term of CV_1.

We decompose $\hat{B}_t$ into two terms: $\hat{B}_t = n^{-2}\sum_{s\neq t} x_s^2 E(e_{st}|z_t) + n^{-2}\sum_{s\neq t} x_s^2[e_{st} - E(e_{st}|z_t)] = \omega_{1t} + \omega_{2t}$, where the definitions of ω_{1t} and ω_{2t} should be obvious. Thus,

$$\begin{aligned}
CV_1^0 &= n^{-2}c_{1n}^{-2}\sum_t x_t^2\omega_{1t}^2 M_t f_t^{-2} + 2n^{-2}c_{1n}^{-2}\sum_t x_t^2\omega_{1t}\omega_{2t}M_t f_t^{-2} \\
&\quad + n^{-2}c_{1n}^{-2}\sum_t x_t^2\omega_{2t}^2 M_t f_t^{-2} \\
&\equiv \Delta_{n1} + \Delta_{n2} + \Delta_{n3}.
\end{aligned} \tag{A.13}$$

We evaluate Δ_{nj} for $j = 1, 2, 3$ separately. First, we define $e_t(z) = (\beta(z_t) - \beta(z))K_h(z_t - z)$. Simple calculations show that $E(e_t(z)) = h^2 B(z) + O(h^4)$ holds uniformly over $z \in \mathcal{M}$, where $B(z) = (\kappa_2/2)[\beta^{(2)}(z) + \beta^{(1)}(z)f^{(1)}(z)/f(z)]$. This gives

$$\sup_{z_t\in\mathcal{M}} |\omega_{1t}| = O_p(h^2) \quad \text{and} \quad \Delta_{n1} = O_p(h^4). \tag{A.14}$$

Second, applying the same proof technique used in the proof of Lemma A.1, we can derive the stochastic order for ω_{2t}. Specifically, we have

$$\begin{aligned}
|\omega_{2t}| &\leq \left|n^{-1}\sum_{k=0}^{N^*}\sum_{\substack{s=s_k\\ s\neq t}}^{s_k^{**}} M_{n,s_k}[e_{st} - E(e_{st}|z_t)]\right| \\
&\quad + \left|n^{-1}\sum_{k=0}^{N^*}\sum_{\substack{s=s_k\\ s\neq t}}^{s_k^{**}} (M_{ns} - M_{n,s_k})[e_{st} - E(e_{st}|z_t)]\right| \\
&\leq \sup_{r\in[0,1]} M_n(r) n^{-1}\sum_{k=0}^{N^*}\left|\sum_{\substack{s=s_k\\ s\neq t}}^{s_k^{**}}[e_{st} - E(e_{st}|z_t)]\right| \\
&\quad + \sup_{|r-r'|\leq\lambda}|M_n(r) - M_n(r')| n^{-1}\sum_{s\neq t}\left|e_{st} - E(e_{st}|z_t)\right| \\
&= O_p\left(\sqrt{\frac{h\ln(n\lambda)}{n\lambda}}\right) + O_p\left(h\sqrt{\lambda\ln\lambda^{-1}}\right) \\
&= O_p\left(h(nh)^{-1/4}(\ln n)^{1/2}\right)
\end{aligned} \tag{A.15}$$

by choosing $\lambda = (nh)^{-1/2}$, where we again use (A.11) and (A.12). Also, applying theorem 2 of Hansen (2008) gives

$$\begin{aligned}
\sup_{z\in\mathcal{M}}\left|n^{-1}\sum_{t=1}^{n}\left(e_t(z) - E[e_t(z)]\right)\right| &= O_p\left(\sqrt{(h^2\ln n)/(nh)}\right) \\
&= O_p\left(\sqrt{(h\ln n)/n}\right).
\end{aligned} \tag{A.16}$$

Combining (A.15) and $|\omega_{1t}| = O_p(h^2)$ yields

$$\begin{aligned}
\sup_{z_t\in\mathcal{M}} |\hat{B}_t| &\leq \sup_{z_t\in\mathcal{M}} |\omega_1(z_t)| + \sup_{z_t\in\mathcal{M}} |\omega_2(z_t)| \\
&= O_p\left(h^2 + h(nh)^{-1/4}(\ln n)^{1/2}\right) \\
&= O_p\left(h(nh)^{-1/4}(\ln n)^{1/2}\right).
\end{aligned} \tag{A.17}$$

By $n^{-2}\sum_s x_s^2 = O_p(1)$, it is easy to see that

$$\begin{aligned}
\Delta_{n2} &= O_p(1)\sup_{z_t\in\mathcal{M}}|\omega_{1t}\omega_{2t}| = O_p\left(h^3(nh)^{-1/4}(\ln(n))^{1/2}\right) \\
&= o_p(h^3).
\end{aligned} \tag{A.18}$$

Finally, we consider $\Delta_{n3} = n^{-2}c_{1n}^{-2}\sum_t x_t^2 \omega_{2t}^2 M_t f_t^{-2}$. By Assumption A6*, (2.6) with $\sigma_1(z) = |\beta^{(1)}(z)|\sqrt{\nu_2(K)f(z)}$, we have $\frac{1}{\sigma_1(z)\sqrt{nh}}\sum_s \frac{x_s^2}{n}[e_s(z) - E(e_s(z))] - \int_0^1 B_x^2(r)\,dW_\beta(r) = o_p(1)$ uniformly over $z \in \mathcal{M}$. Also note that when $c_{1n} = n^{-2}\sum_{s=1}^n x_s^2 = B_{x,(2)} + o_p(1)$, where $B_{x,(2)} = \int_0^1 B_x(r)^2\,dr$, we have

$$\Delta_{n3} = (h/n)\nu_2(K)B_{x,(2)}^{-1}E\big[M_t f_t^{-1}\big(\beta_t^{(1)}\big)^2\big]\left[\int_0^1 B_x^2(r)\,dW_\beta(r)\right]^2 + o_p(h/n). \quad \text{(A.19)}$$

Combining (A.14), (A.18), and (A.19) completes the proof of Lemma A.2.

Lemma A.3. Under Assumptions A1–A6*, (A.4) holds true.

Proof. Lemma A.1 implies that the leading term of CV_2, denoted by CV_2^0, is obtained from CV_2 by replacing $\hat{A}_t$ with $\tilde{\mu}_1(z_t)$. That is, $CV_2^0 = n^{-2}\sum_t x_t^2\hat{C}_t^2/\tilde{\mu}_1^2(z_t)M_t = n^{-2}c_{1n}^{-2} \times \sum_t x_t^2\hat{C}_t^2 f_t^{-2}M_t$, where $\hat{C}_t = n^{-2}\sum_{s\neq t} x_s u_s K_{h,ts}$. By Assumption A6*, we have $n\sqrt{h}\hat{C}_t = \sigma_u\sqrt{\nu_0(K)f(z_t)}\int_0^1 B_x(r)\,dW_u(r) + o_p(1)$. Therefore, we obtain

$$n^2hCV_2^0 = \sigma_u^2\nu_0(K)c_{1n}^{-1}\left[\int_0^1 B_x(r)\,dW_u(r)\right]^2\int M(z)\,dz + o_p(1). \quad \text{(A.20)}$$

This completes the proof of Lemma A.3.

Lemma A.4. Under Assumptions A1–A6, (A.5) holds true.

Proof. By definition, $CV_3 = n^{-4}\sum_t x_t^2\hat{B}_t\hat{A}_t^{-2}M_t \times \sum_{s\neq t} x_s \times u_s K_{h,ts}$. Assumption A3 implies that $E(CV_3) = 0$ and $\{u_t\}_{t=1}^n$ are serially uncorrelated. Letting $\Delta_{3,n} = E(CV_3^2|\{x_t, z_t\}_{t=1}^n)$, we have $\Delta_{3,n} = n^{-8}\sigma_u^2\sum_t(x_t^2\hat{B}_t\hat{A}_t^{-2}M_t)^2\sum_{s\neq t}x_s^2K_{h,ts}^2 + n^{-8} \times \sigma_u^2\sum_t\sum_{t'\neq t}x_t^2\hat{B}_t\hat{A}_t^{-2}M_t x_{t'}^2\hat{B}_{t'}\hat{A}_{t'}^{-2}M_{t'}\sum_{s\neq t\neq t'}x_s^2K_{h,ts}K_{h,t's} = \Delta_{3,1n} + \Delta_{3,2n}$.

Lemma A.1 and (A.17) imply that $\sup_{z_t\in\mathcal{M}}|\hat{A}_t| = O_e(1)$ and that $\sup_{z_t\in\mathcal{M}}|\hat{B}_t| = O_p(h\delta_n)$, where $\delta_n = (nh)^{-1/4}(\ln(n))^{1/2}$. Applying the same technique used in the proof of Lemma A.1, we can show that $\sup_{z_t\in\mathcal{M}} n^{-2}\sum_{s\neq t}x_s^2K_{h,ts}^2 = O_p(h^{-1})$. Therefore, we have $\Delta_{3,1n} = O_p(h^2\delta_n^2)n^{-5}\sup_{z_t\in\mathcal{M}}\sum_{s\neq t}x_s^2K_{h,ts}^2 = O_p(n^{-3}h\delta_n^2)$ and $\Delta_{3,2n} = O_p(h^2\delta_n^2)n^{-8}\sum_s x_s^2\sum_{t\neq s}x_t^2K_{h,ts} \times M_t\sum_{t'\neq t\neq s}x_{t'}^2K_{h,t's}M_{t'} = O_p(n^{-2}h^2\delta_n^2)$ by Lemma A.1. Because $n^{-1} = o(h)$, $\Delta_{3,2n}$ asymptotically dominates $\Delta_{3,1n}$. Thus we have shown that $\mathrm{var}(CV_3|\{(x_t, z_t)\}_{t=1}^n) = O_p(n^{-2}h^2\delta_n^2)$. This implies that $CV_3 = O_p((h/n)\delta_n) = o_p(h/n)$ by Markov's inequality.

ACKNOWLEDGMENTS

The authors thank two referees, an associate editor, former coeditor Serena Ng, and coeditor Jonathan Wright for insightful comments that greatly improved the article. This research was partially supported by the SSHRC Fund of Canada and the National Science Foundation of China (project 70773005).

[Received June 2009. Revised December 2011.]

REFERENCES

Cai, Z. (2007), "Trending Time Varying Coefficient Time Series Models With Serially Correlated Errors," *Journal of Econometrics*, 136, 163–188. [541]

Cai, Z., Fan, J., and Yao, Q. (2000), "Functional Coefficient Regression Models for Nonlinear Time Series," *Journal of the American Statistical Association*, 95, 941–956. [541]

Cai, Z., Li, Q., and Park, J. Y. (2009), "Functional-Coefficient Models for Nonstationary Time Series Data," *Journal of Econometrics*, 148, 101–113. [541,546,547]

Chen, R., and Tsay, C. Y. (1993), "Functional Coefficient Autoregressive Models," *Journal of the American Statistical Association*, 88, 298–308. [541]

De Jong, R., and Davison, J. (2000), "The Functional Central Limit Theorem and Weak Convergence to Stochastic Integrals I: Weakly Dependent Processes," *Econometric Theory*, 16, 621–642. [545]

Hansen, B. E. (1992), "Convergence to Stochastic Integrals for Dependent Heterogeneous Processes," *Econometric Theory*, 8, 489–500. [544,545,550]

——— (2008), "Uniform Convergence Rates for Kernel Estimation With Dependent Data," *Econometric Theory*, 24, 726–748. [550]

Härdle, W., and Marron, J. S. (1985), "Optimal Bandwidth Selection in Nonparametric Regression Function Estimation," *The Annals of Statistics*, 13, 1465–1481. [541]

Juhl, T. (2005), "Functional Coefficient Models Under Unit Root Behavior," *Econometrics Journal*, 8, 197–213. [541]

Karlsen, H. A., Myklebust, T., and Tjostheim, D. (2007), "Nonparametric Estimation in a Nonlinear Cointegration Type Model," *The Annals of Statistics*, 35, 252–299. [541]

Phillips, P. C. B. (1995), "Fully Modified Least Squares and Vector Autoregression," *Econometrica*, 63, 1023–1078. [544]

——— (2009), "Local Limit Theory and Spurious Nonparametric Regression," *Econometric Theory*, 25, 1466–1497. [541]

Phillips, P. C. B., and Hansen, B. E. (1990), "Statistical Inference in Instrumental Variables Regressions With $I(1)$ Processes," *Review of Economic Studies*, 57, 99–125. [544]

Rio, E. (1995), "The Functional Law of the Iterated Logarithm for Stationary Strongly Mixing Sequences," *The Annals of Probability*, 23, 1188–1203. [543]

Robinson, P. M. (1989), "Nonparametric Estimation of Time-Varying Parameters," in *Statistical Analysis and Forecasting of Economic Structural Change*, ed. P. Hackl, Berlin: Springer-Verlag, pp. 253–264. [541]

Saikkonen, P. (1991), "Asymptotically Efficient Estimation of Cointegration Regression," *Econometric Theory*, 7, 1–21. [543]

Wang, Q., and Phillips, P. C. B. (2009a), "Asymptotic Theory for Local Time Density Estimation and Nonparametric Cointegration Regression," *Econometric Theory*, 25, 710–738. [541,543]

——— (2009b), "Structural Nonparametric Cointegrating Regression," *Econometrica*, 77, 1901–1948. [541,543,544,547]

Xiao, Z. (2009), "Functional Coefficient Cointegration Models," *Journal of Econometrics*, 152, 81–92. [541]

Yoshihara, K. (1976), "Limiting Behavior of U–Statistics for Stationary, Absolutely Regular Processes," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 35, 237–252. [545]

Zhou, Y., and Liang, H. (2009), "Statistical Inference for Semiparametric Varying-Coefficient Partially Linear Models With Error-Prone Linear Covariates," *The Annals of Statistics*, 37, 427–458. [541]