NONPARAMETRIC STRUCTURAL ESTIMATION VIA CONTINUOUS LOCATION SHIFTS IN AN ENDOGENOUS REGRESSOR

By

Peter C. B. Phillips and Liangjun Su

June 2009

COWLES FOUNDATION DISCUSSION PAPER NO. 1702

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

Nonparametric Structural Estimation via Continuous Location Shifts in an Endogenous Regressor *

Peter C. B. Phillips
Yale University, University of Auckland
University of York & Singapore Management University

Liangjun Su
School of Economics
Singapore Management University

May 12, 2009

Abstract

Recent work by Wang and Phillips (2009b, c) has shown that ill posed inverse problems do not arise in nonstationary nonparametric regression and there is no need for nonparametric instrumental variable estimation. Instead, simple Nadaraya Watson nonparametric estimation of a (possibly nonlinear) cointegrating regression equation is consistent with a limiting (mixed) normal distribution irrespective of the endogeneity in the regressor, near integration as well as integration in the regressor, and serial dependence in the regression equation. The present paper shows that some closely related results apply in the case of structural nonparametric regression with independent data when there are continuous location shifts in the regressor. In such cases, location shifts serve as an instrumental variable in tracing out the regression line similar to the random wandering nature of the regressor in a cointegrating regression. Asymptotic theory is given for local level and local linear nonparametric estimators, links with nonstationary cointegrating regression theory and nonparametric IV regression are explored, and extensions to the stationary strong mixing case are given. In contrast to standard nonparametric limit theory, local level and local linear estimators have identical limit distributions, so the local linear approach has no apparent advantage in the present context. Some interesting cases are discovered, which appear to be new in the literature, where nonparametric estimation is consistent whereas parametric regression is inconsistent even when the true (parametric)

*Phillips acknowledges partial support from NSF Grant #SES 06-47086. Address correspondence to Peter C. B. Phillips, Cowles Foundation for Research in Economics, Yale University, Box 208281, New Haven, Connecticut USA 06520-8281; e-mail: peter.phillips@yale.edu. Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903; e-mail: ljsu@smu.edu.sg.

regression function is known. The methods are further applied to establish a limit theory for nonparametric estimation of structural panel data models with endogenous regressors and individual effects. Some simulation evidence is reported.

JEL classification: C13, C14.
Keywords: Fixed effects, Kernel regression, Location shift, Mixing, Nonparametric IV, Nonstationarity, Panel model, Structural estimation.

1 Introduction

Much recent interest in nonparametric regression has focused on the effects of endogeneity in the regressor. Attention has primarily been microeconometric and so has naturally concentrated on a framework in which the observed data are independently distributed. Allowing for endogeneity is particularly important in practical applications where there are unobserved characteristics (such as inherent ability) that influence both the regressors and the equation errors. In such contexts, the unknown regression function is not recoverable as a conditional expectation but is submerged in a functional integral equation whose solution raises difficulties that fall under the category of ill-posed inversion problems. Research on this type of problem in econometrics has been underway for about a decade, following on from a much longer literature in mathematics, numerical analysis, image processing, and statistics. Methods of "regularizing" the inversion problem have become popular in the theoretical development of the subject and both kernel and series-based approaches have been considered. There is now a wide literature on the subject, including early work by Newey, Powell and Vella (1999), and Newey and Powell (2003), overviews by Blundell and Powell (2003), Florens (2003), Carrasco, Florens and Renault (2007), Chen (2007), and contributions by Hall and Horowitz (2005), Darolles, Florens and Renault, (2009) and many others. In addition to microeconometric contexts, the field of potential applications has recently widened to include asset pricing in financial econometrics (Simoni, 2009).

To facilitate the use of standard functional analysis methods, it has become conventional in econometric treatments to restrict the system variables to have bounded support and bounded densities. These restrictions appear innocent given that (distribution function) transformations can mechanically transform the support of all variables to the $[0, 1]$ interval. However, these transformations can induce subtle changes in the system that are not so innocuous and omit factors that are generally relevant in microeconometric modeling. For example, in applied microeconometrics, adding more data generally means adding more parameters to estimate and introducing more variation to be explained. Indeed, those very characteristics have driven much of the research on robust estimation and the treatment of individual fixed effects.

Some of these characteristics manifest in an important way in the setting explored in the present paper. In particular, we show that the variation induced by locational shifts in the regressor can have an enormous impact on the potential capabilities of nonparametric regression, completely removing the ill posed inverse problem and facilitating consistent estimation by conventional nonparametric regression techniques.

Fixed location effects may arise in various economic contexts. For example, markets for a product may be differentiated by location with different supply functions where the supply curve is influenced by factors that shift the supply curve around according to location. Such shifts in a supply curve are precisely the ones that are described in textbook treatments of identification in simultaneous equations models of supply and demand. As the supply curve moves around it traces out a sequence of equilibria associated with the Marshallian cross at each location.

The key idea is illustrated in Fig. 1, which provides a sample plot of data generated from the following location specific linear Marshallian stochastic demand/supply system

$$\text{Demand:} \qquad q_i = a + bp_i + u_i, \tag{1.1}$$

$$\text{Supply:} \qquad p_i = c + \mu_\alpha 1\{i \in \alpha\} + u_{pi}, \;\; \alpha = 1, 2, \cdots, K, \tag{1.2}$$

$$u_i = \{\theta u_{pi} + \epsilon_{qi}\} / \left(1 + \theta^2\right)^{1/2}$$

In this system, the errors $(u_{pi}, \epsilon_{qi}) \equiv iidN(0, I_2)$ and α signifies a particular market location in which, if there were no disturbances, the product price would be $p = c + \mu_\alpha$ and demand would be $q = a + b(c + \mu_\alpha)$. In general, the locational equilibrium at location α is disturbed by errors and the data tend to cluster around each locational equilibrium point. The nature and orientation of the clusters depends on the joint distribution of the equation errors (u_i, u_{pi}), which in turn depends on the endogeneity parameter θ, corresponding to which we have the correlation $\rho = \text{corr}(u_i, u_{pi}) = \theta / \left(1 + \theta^2\right)^{1/2}$. When $\theta = 0$, there is no endogeneity ($\rho = 0$), and when $\theta = 2$, there is substantial endogeneity ($\rho = 0.89$). Fig. 1 illustrates this locational clustering phenomena with a typical data set for $K = 5$ locations corresponding to $\mu_\alpha \in \{-4, -2, 0, 2, 4\}$ with $\theta = 2$, $c = 5$, $a = 10$, $b = -1$, and $M = 100$ observations for each α. Along the demand curve we observe clusters of points around price levels $\{1, 3, 5, 7, 9\}$ corresponding to each of the market locations. As the location α shifts, the data display a tendency to trace out the demand curve, just as in textbook discussions.

Fig. 2 displays the Nadaraya Watson local level estimate of the (demand) curve using all $n = M \cdot K = 500$ observations, a Gaussian kernel and a mechanical $n^{-1/5}$ bandwidth rule. As is apparent, within the support of the sample data for (q_i, p_i) the regression line is very well fitted by the nonparametric curve irrespective of the endogeneity. In this region, the data trace out the demand curve and it is this line that is well approximated by the regression. But at the extremes of the support the endogeneity in the data is more clearly manifest and the regression line tracks out the data, following the supply correspondence in those two regions.

This simple example illustrates the central finding of the present paper. In spite of endogeneity in a regressor and provided the regressor has enough variation, simple nonparametric regression may be consistent even with independent data, at least within the interior of the support. The simplest manner in which sufficient variation may be attained is for the regressors to sustain location shifts by way of external fixed effects, as in the illustration. These location

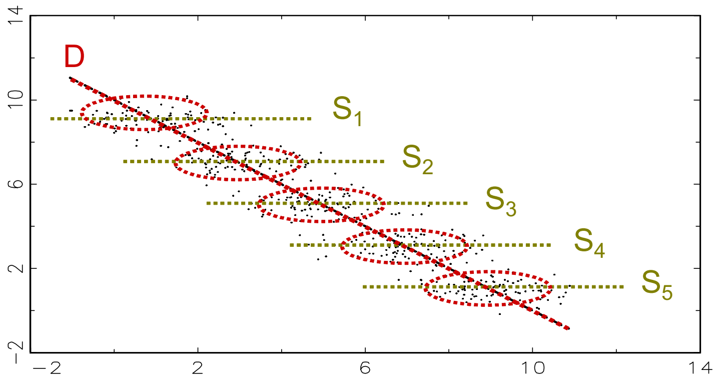

Figure 1: Demand and supply curves (for 5 locations) with sample data generated by (1.1) and (1.2).

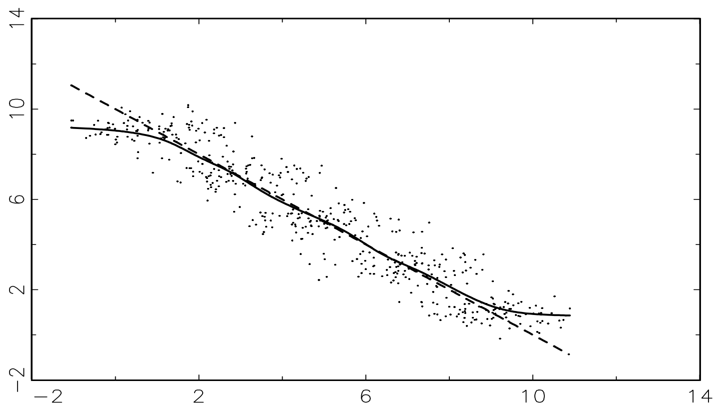

Figure 2: Nadaraya Watson nonparametric regression estimate of the demand curve (1.1) using sample data taken over all locations.

shifts serve the role of a form of nonstationarity and in this sense the resulting consistency of nonparametric regression is analogous to that achieved in the case of a (near) unit root regressor. In that case, the regressor is recurrent and visits every point in the space an infinite number of times, which is broadly speaking analogous to a regressor which undergoes continuous shifts in location as data accumulate, so that the variation of the regressor continues to increase with the sample size. Wang and Phillips (2009b, c) showed that nonparametric regression with unit root or near integrated regressors is consistent, but with a reduced rate of convergence compared with the conventional $\sqrt{nh}$ rate for sample size n and bandwidth h. Correspondingly, the present paper shows that within the support and with a reduced rate of convergence, nonparametric regression is consistent at all points where the location shifts are continuous in the limit as $n \to \infty$.

The case where the distribution of the regressor has compact support is also investigated. As mentioned above, this case has been extensively studied in the nonparametric IV literature. In the present context, we show that the nonparametric local level and local linear estimators can both achieve the conventional $\sqrt{nh}$ convergence rate of nonparametric estimation in models with exogenous regressors. However, parametric estimation turns out to be inconsistent, even though the true form of the (parametric) regression function is known and estimated. To the authors' knowledge, this is the first example of this type to be found in the literature.

The paper is organized as follows. Section 2 lays out a formal model and presents the main finding on the consistency and limiting normal distribution of nonparametric regression in a structural model with continuous location shifts. A series of remarks on the main theorem then discuss the significance, operational import, and various extensions of the result. Differences with standard limit theory are explained, including the elimination of the linear bias term in the limit distribution of the local level estimator, which gives local level estimation the same limit theory as local linear estimation when there are location shifts. The discussion makes clear that a central requirement in achieving consistency and avoiding the use of external instrumentation is the continuity in the limit of the location shifts, although as the above illustration shows even a small number of shifts can achieve a great deal of bias reduction in practice. The discussion also explores the links with recent findings on consistent estimation in nonparametric cointegrating regressions (Wang and Phillips, 2009b, c) and contrasts the present findings with those of the recent nonparametric IV literature. Particular consideration is given to whether the regressor has compact support, as is frequently assumed in the latter literature, or infinite support, as in the nonstationary time series literature. The case of compact support is especially interesting because of the differences between parametric and nonparametric regression. Even when the true functional form is known, nonlinear parametric regression is generally inconsistent because of endogeneity. In the case of infinite support, even if parametric regression is consistent, its convergence rate can be arbitrarily slow in relation to nonparametric estimation, which appears to be the first example of this kind in the literature.

Continuous location shifts are somewhat idealized in a cross section setting and are more likely to be realistic in a panel context where individual fixed effects play a significant role.

Section 3 of the paper therefore looks at extensions of the (cross section) location model to the structural panel data case and gives a corresponding limit theory for panel nonparametric regression. To the authors' knowledge, this is the first treatment of nonparametric estimation in structural panel data models with endogenous regressors. Section 4 reports some simulation evidence and Section 5 concludes. Proofs of the main results are given in Section 6, which also includes extensions of the limit theory in the paper to the stationary, strong mixing case.

2 Limit theory under location shifts

2.1 Models with infinite support

This section introduces a nonparametric regression model in which there are continuous location shifts in the regressor whose support is infinite. Observations $\{(y_t, X_t), 1 \le t \le n\}$ are generated as follows:

$$\begin{aligned} y_t &= g(X_t) + u_t,\ E(u_t|X_t) \neq 0 && (2.1)\\ X_t &= \mu_\alpha 1\{t \in A_\alpha\} + u_{xt}, && (2.2)\\ \mu_\alpha &= \frac{\alpha}{m^\lambda},\ \alpha = -m+1, \cdots, m,\ \lambda \in (0,1), && (2.3)\\ M &= \#\{t \in A_\alpha\}. \end{aligned}$$

We use the t subscript notation here (in spite of the possible cross section setting) because it fits most conveniently with the panel model extension that is explored in the next section of the paper. The mechanism allows for M observations in the vicinity of each locality and may easily be generalized to allow M to depend on α, although that extension involves only minor complications but more notational complexity so it will not be pursued here. The total observation count is then $n = 2mM$. Throughout the paper we require $m \to \infty$, but M can either be finite or pass to ∞ and this dual possibility is denoted by writing $(M, m) \to \infty$.

We start by considering the local constant (Nadaraya-Watson) kernel estimator of $g(x)$

$$\widehat{g}(x) = \frac{\frac{1}{n}\sum_{t=1}^{n} y_t K_h(X_t - x)}{\frac{1}{n}\sum_{t=1}^{n} K_h(X_t - x)}, \tag{2.4}$$

where $K(\cdot)$ is a kernel function, $K_h(\cdot) = h^{-1}K(\cdot/h)$ and $h \equiv h(M, m)$ is a bandwidth parameter. We make the following assumptions.

Assumption A1. (u_t, u_{xt}), $t = 1, \cdots, n$, *are independent and identically distributed (iid).*

Assumption A2. $E(u_t) = 0$, $E(u_t^2) = \sigma^2$, *and* $E|u_t|^{2+\delta} < \infty$ *for some* $\delta > 0$.

Assumption A3. *(i) The probability density function (p.d.f.)* $f(\cdot,\cdot)$ *of* (u_t, u_{xt}) *exists.* $f(\cdot,\cdot)$ *has second order partial derivative* $f_2''(u, u_x)$ *with respect to* u_x *such that* $f_2''(u, u_x)$ *is continuous in* u_x *and* $\int\int |u f_2''(u, u_x)|\, du du_x < \infty$*. The marginal p.d.f. of* u_{xt}, $f_{u_x}(\cdot)$, *has second order continuous derivatives such that* $\int_{-\infty}^{\infty} |f_{u_x}'(p)|\, dp < \infty$, *and* $\int_{-\infty}^{\infty} |f_{u_x}''(p)|\, dp < \infty$.

(ii) There exists a function $C_f(\cdot)$ such that for any sequence $c_n \to \infty$, we have $|\int_{-\infty}^{-c_n} f(u,u_x) du_x + \int_{c_n}^{\infty} f(u,u_x) du_x| \le c_n^{-\upsilon} C_f(u)$ *for some* $\upsilon > 1$ and $\int_{-\infty}^{\infty} |u| C_f(u) du < \infty$.

(iii) $f_{u_x}(p_n) = O(|p_n|^{-\upsilon-1})$ *as* $|p_n| \to \infty$.

Assumption A4. *For given x, $g(x)$ has a continuous, bounded second derivative in a small neighborhood of x.*

Assumption A5. *The kernel function $K(x)$ is a uniformly bounded symmetric p.d.f. such that $\int x^4 K(x) dx < \infty$.*

Assumption A6. *As $(M,m) \to \infty$, $Mm^{\lambda}h \to \infty$, $Mm^{\lambda}h^5 \to c \in [0,\infty)$, and $Mm^{\lambda}h(m^{-4\lambda} + m^{-2\upsilon(1-\lambda)}) \to 0$.*

The *iid* condition in Assumption A1 can be relaxed. For example, we can allow the process $\{(u_t, u_{xt}), t \ge 1\}$ to be strictly stationary and strong mixing with mixing coefficients that decay to zero at certain rates. This extension is provided in the Appendix, in order to keep the assumptions and development in the main text as simple as possible. Alternatively, it is possible to allow for unit root or near unit root behavior in u_{xt}, as in Wang and Phillips (2009b, c). Assumption A2 is weak. The zero-mean condition is needed for the identification of the nonparametric function $g(\cdot)$. But we do not impose the exogeneity condition $E(u_t|X_t) = 0$ a.s., nor do we assume conditional homoskedasticity. Assumption A3 imposes some smoothness and tail conditions on the joint and marginal p.d.f.'s. The tail condition on $f_{u_x}(\cdot)$ is equivalent to requiring that the first moment of u_{xt} exists. Assumption A4 imposes a standard smoothness condition on the regression function $g(\cdot)$. Assumption A5 is also standard, although the symmetry of the kernel function facilitates analysis and simplifies notation. Assumption A6 imposes conditions on the bandwidth and its association with M, m^{λ}, and υ.

We now state the first main result.

Theorem 2.1 *Under Assumptions A1-A6, we have*

$$\sqrt{Mm^{\lambda}h}\left\{\widehat{g}(x) - g(x) - \frac{1}{2}h^2\mu_2(K)g''(x)\right\} \to_d N\left(0, \sigma^2\nu_2(K)\right), \tag{2.5}$$

where $\mu_2(K) = \int x^2 K(x) dx$, and $\nu_2(K) = \int K(x)^2 dx$. If in addition $c = 0$ in Assumption A6, then the bias term in the braces vanishes.

The following remarks discuss this theorem in connection with limit theory in other cases, including cases where there are no location shifts, where there is exogeneity, and where the regressor is integrated. We discuss extensions to the panel data case in the next section.

Remark 1. (**Comparison with the exogenous regressor case**) When the regressor X_t is exogenous, stationary and has p.d.f. $f(x)$, the local constant estimator has the following limit behavior

$$\sqrt{nh}\left(\widehat{g}(x) - g(x) - h^2\mu_2(K)\left\{g'(x)f'(x)/f(x) + \frac{1}{2}g''(x)\right\}\right) \to_d N\left(0, \sigma^2(x)\nu_2(K)/f(x)\right), \tag{2.6}$$

where $\sigma^2(x) = E\left(u_t^2|X_t = x\right)$. While this result is obviously similar to the limit theory in (2.5) there are some important distinctions. The most important difference is that the expression for the bias in (2.5) involves only a single term, which corresponds to the second bias term in (2.6). The analog of the first bias term in (2.6) in the present setting is

$$h^2\mu_2(K)\,g'(x)\int_{-\infty}^{\infty} f'_{u_x}(p)\,dp,$$

which is 0 as the marginal p.d.f. $f_{u_x}(\cdot)$ vanishes at infinity. This explains why there is no linear term in the bias function appearing in (2.5), in contrast to the limit theory for local level estimation with stationary regressors. In the present context, endogeneity does not prevent consistency and makes a smaller $o\left(h^2\right)$ contribution to the bias so it does not figure in the limit theory (2.5). A second important distinction is that the marginal density of X_t appears in (2.6) but not in (2.5). The reason is that the denominator in the definition of $\widehat{g}(x)$

$$\frac{1}{n}\sum_{t=1}^{n} K_h(X_t - x)$$

does not converge in probability to a density. Instead, as shown in the appendix,

$$\frac{1}{Mm^\lambda}\sum_{t=1}^{n} K_h(X_t - x) \to_p 1,$$

and this result arises because the location shifts in X_t imply that the averaging operator $m^{-\lambda}M^{-1}\sum_{t=1}^{n}$ ensures that the density of u_{xt} is averaged over the whole domain leading to $\int f_{u_x}(x)\,dx = 1$. A third important distinction is that the (unconditional) variance of u_t appears in (2.5) whereas the conditional variance $\sigma^2(x)$ of u_t appears in (2.6). The reason is that in the bias-variance decomposition of $\widehat{g}(x) - g(x)$, the variance term

$$\frac{\sqrt{h}}{\sqrt{n}}\sum_{t=1}^{n} u_t K_h(X_t - x)$$

does not converge weakly to a normal distribution with mean zero and variance $\sigma^2(x)\,\nu_2(K)$ $/f(x)$. Instead, the presence of location shifts in X_t leads to further averaging in the central limit theory and we have

$$\frac{\sqrt{h}}{\sqrt{Mm^\lambda}}\sum_{t=1}^{n}\left\{u_t K_h(X_t - x) - E\left[u_t K_h(X_t - x)\right]\right\} \to_d N\left(0, \sigma^2\nu_2(K)\right).$$

Remark 2. (**Extension to the stationary strong mixing case**) It is well known that in the exogenous regressor case, the limit theory for the nonparametric estimator given in (2.6) is unaffected by the presence of weakly dependent data. The result is commonly referred to as working independence. Various forms of dependence have been considered in the literature and results are overviewed in Li and Racine (2007). In the present case of an endogenous regressor, the limit theory given in Theorem 2.1 is also unaffected by weak dependence, so working independence continues to apply in the location shift model. In particular, under some additional regularity conditions, the limit distribution (2.5) holds under stationary and strong mixing errors (u_t, u_{xt}). The analysis is given in Section 6.4. It is shown there that, just as in the independent case, the two terms in the decomposition

$$\widehat{g}(x) - g(x) = \frac{\frac{1}{Mm^{\lambda}} \sum_{t=1}^{n} \{g(X_t) - g(x)\} K_h(X_t - x)}{\frac{1}{Mm^{\lambda}} \sum_{t=1}^{n} K_h(X_t - x)} + \frac{\frac{1}{Mm^{\lambda}} \sum_{t=1}^{n} u_t K_h(X_t - x)}{\frac{1}{Mm^{\lambda}} \sum_{t=1}^{n} K_h(X_t - x)}$$

may be analyzed in a similar way, showing that the asymptotic effect of weak dependence on both the bias and the limit distribution terms is negligible. Interestingly, the limit results do not depend on the location shifts occuring sequentially in the time dimension, so we may have $t, t' \in A_\alpha$ even for distant pairs $\{t, t'\}$.

Remark 3. (**Local linear nonparametric estimation**) A popular choice in place of the estimator (2.4) in practical work is the local linear estimator (e.g., Fan and Gijbels, 1996)

$$\widetilde{g}(x) = \frac{\sum_{t=1}^{n} w_t Y_t}{\sum_{t=1}^{n} w_t},$$

where $w_t = K_h(X_t - x)\{S_{n2} - (X_t - x) S_{n1}\}$ and $S_{nj} = \sum_{t=1}^{n} (X_t - x)^j K_h(X_t - x)$ for $j = 1, 2$. Following the same lines as the proof of Theorem 2.1 and under the same conditions A1-A6 we find that

$$\sqrt{Mm^{\lambda}h}\left(\widetilde{g}(x) - g(x) - \frac{1}{2}h^2 \mu_2(K) g''(x)\right) \to_d N\left(0, \sigma^2 \nu_2(K)\right). \tag{2.7}$$

Thus, $\widetilde{g}(x)$ is consistent and has the same limit distribution and bias as the local level estimator (2.5). In the study of nonstationary nonparametric cointegrating regression, Phillips and Wang (2009c) found that the local linear and local level estimators also share the same asymptotic distribution and bias. So, local linear regression has no advantage over local level regression in terms of bias reduction in both nonstationary and location shift regressions.

Remark 4. (**Bandwidth choice**) Defining the limiting bias function as $B(x) = \frac{1}{2}\mu_2(K)$ $g''(x)$, Theorem 2.1 implies that the leading term in the asymptotic mean squared error (MSE) of $\widehat{g}(x)$ is given by

$$MSE(\widehat{g}(x)) = h^4 B(x)^2 + \frac{\sigma^2 \nu_2(K)}{Mm^{\lambda}h}.$$

Given x, the optimal bandwidth in terms of minimizing $MSE(\widehat{g}(x))$ is

$$h^*(x) = \left[\frac{\sigma^2 \nu_2(K)}{4B(x)^2}\right]^{1/5} \frac{1}{(Mm^{\lambda})^{1/5}}.$$

To estimate the whole regression function $g(\cdot)$, one can choose a single h to minimize the integrated mean square error $\int MSE(\widehat{g}(x)) w(x) dx$, where $w(\cdot)$ is a weight function that serves to avoid division by zero and to perform trimming in areas of sparse support. Then the optimal bandwidth is given by

$$h^* = \left[\frac{\sigma^2 \nu_2(K) \int w(x)\,dx}{4 \int B(x)^2 w(x)\,dx}\right]^{1/5} \frac{1}{(Mm^\lambda)^{1/5}}.$$

The optimal rate $\left(Mm^\lambda\right)^{-1/5}$ is larger than the optimal rate $n^{-1/5}$ for the conventional stationary (exogenous) regressor case. For example, if $\lambda = 0.5$, then $\left(Mm^\lambda\right)^{-1/5} = n^{-1/5} m^{1/10}$. Nevertheless, λ is unobserved in practice, at least without further specification involving measurable variables that determine location shifts. So, in general, we cannot use the plug-in principle to obtain an estimate of h^* or $h^*(x)$. Instead, we can use least squares cross-validation (LSCV) to choose the smoothing parameters. That is, choose h to minimize the following LSCV criterion

$$CV(h) = n^{-1} \sum_{t=1}^{n} \left(Y_t - \widehat{g}_{-t}(X_t)\right)^2 w(X_t), \tag{2.8}$$

where $\widehat{g}_{-t}(X_t)$ is the leave-one-out local constant estimate of $g(X_t)$ and $w(\cdot)$ is some given weight function. We conjecture that the cross-validated bandwidth will converge to the optimal rate $\left(Mm^\lambda\right)^{-1/5}$ at some rate, but do not pursue this study here.

Remark 5. (**Comparison with nonparametric IV regression**) Consider the general nonparametric instrumental variable (IV) regression

$$y_t = g(X_t) + u_t,\ E(u_t|X_t) \neq 0,\ E(u_t|W_t) = 0, \tag{2.9}$$

where $(y_t, X_t, W_t)_{t=1}^n$ are observed and W_t is used as an instrument for X_t. Observe that

$$E(y_t|W_t) = E\{g(X_t)|W_t\} = \int g(x) \frac{f_{xw}(x, W_t)}{f_w(W_t)} dx,$$

where $f_{xw}(\cdot,\cdot)$ and $f_w(\cdot)$ are the joint and marginal p.d.f.'s. An estimate of $g(\cdot)$ may be obtained by various functional inversion techniques. However, the inversion of the associated integral operator equation is typically ill-posed because inversion involves an operator that is not bounded or continuous. Using some standard regularization methods in functional analysis, Hall and Horowitz (2005) suggest two nonparametric approaches to consistently estimate $g(\cdot)$, derive convergence rates and show these are rate optimal.

In order for W_t to be a valid IV for X_t, we usually require that W_t be observed and that the association between the variables be strong enough for successful inversion (or operator inversion in the functional case) of the estimating equations. In the linear case it is usually sufficient to require that corr$(X_t, W_t) \neq 0$. To complete the specification of the model (2.9), we add a reduced form equation for the endogenous regressor X_t. Let $m(W_t) = E(X_t|W_t)$ and $u_{xt} = X_t - E(X_t|W_t)$, so that we have

$$X_t = m(W_t) + u_{xt}. \tag{2.10}$$

This reduced form equation helps in identifying the structural curve $g(\cdot)$ in (2.9) provided W_t is observable and the systematic component $m(W_t)$ of X_t provides sufficient leverage in estimation.

To make the link between this model and the location shift system (2.1) - (2.3) explicit, suppose the instrument variable in (2.9) has a triangular array structure so that the systematic component has the form $m(W_{tn})$. If the variance of $m(W_{tn})$ *expands* as the sample size n increases, then the signal in the regressor X_t correspondingly increases, relative to the variance of the stationary error u_{xt} in (2.10), the variance of the structural equation error u_t in (2.9) and the covariance $\text{cov}(u_{xt}, u_t)$. The possibility of consistent estimation then opens up even in the face of endogeneity in the regressor, a feature that is already well known in the analysis of simple linear parametric models (e.g. Hamilton, 1994, p. 234). As n increases, each distinct value of $m(W_{tn})$ may be regarded as carrying location specific information that corresponds to a potentially new location in the continuous location-shift system (2.1) - (2.3). Interestingly in this case, one does not even need to observe W_{tn} in order to identify and consistently estimate the true structural regression curve $g(\cdot)$ asymptotically. It is sufficient for this leverage from W_{tn} to be present in the regressor X_t. In effect, X_t then has an array structure itself and it is this "nonstationarity" in the regressor that opens up the possibility of consistent estimation by direct nonparametric regression.

Remark 6. (**Link with the nonstationary nonparametric cointegrating regression**) Recently Wang and Phillips (2009b, 2009c) studied structural models of nonparametric cointegration and developed a limit theory for kernel cointegrating regression. It was shown that no ill-posed inverse problem arises in nonparametric models with nonstationary endogenous regressors and identification does not require the existence of observable instrumental variables that are orthogonal to the structural equation errors. The location-shift model shares these features in much the same spirit. To see the analogy between the two types of models, consider the following nonlinear structural model of cointegration

$$y_t = g(X_t) + u_t, \ t = 1, 2, \cdots, n,$$

where u_t is a zero mean stationary process, X_t is a nonstationary $I(1)$ regressor, and $g(\cdot)$ is an unknown smooth function to be estimated. Wang and Phillips (2009b, 2009c) show that under some regularity conditions, the local constant estimate $\widehat{g}(x)$ of $g(x)$ has the following limiting mixed normal distribution:

$$\left(nh^2\right)^{1/4}\left(\widehat{g}(x) - g(x) - \frac{h^2\mu_2(K)}{2}g''(x)\right) \to_d \frac{\sigma_u \; N}{L^{1/2}(1,0)}$$

where σ_u is a constant that depends on the kernel and the parameters underlying the process $\{X_t, u_t\}$, and N is a standard normal variate independent of the local time process $L(t,0)$ of the Brownian motion associated with the limit of the standardized process $n^{-1/2}X_{[n\cdot]}$. The kernel estimates in the current paper are similar to the nonparametric cointegrating estimates of Wang and Phillips (2009b) in at least four aspects. First, both estimates deal with endogeneity in the regressor without using instrumental variables. Instead, identification occurs through the expansion of the variance of the regressors: either continuous location shifts or unit-root behavior ensures that $\mathrm{Var}(X_t)$ expands as the sample size increases and that data accumulates steadily over the entire support. In effect, continuous location shifts provide a form of recurrence in X_t, corresponding to the capacity of a nonstationary random wandering process to visit all points in the space an infinite number of times. This type of recurrent behavior in X_t ensures that the structural regression curve is effectively traced out continuously by the data and is correspondingly amenable to nonparametric regression, just as Fig. 2 suggests. Second, as mentioned in Remark 3, both estimates have the same asymptotic distribution and bias as the corresponding local linear estimates. Third, both estimates have a slower convergence rate than nonparametric regression with a stationary regressor, namely, $(nh)^{1/2}$. As noticed by Wang and Phillips (2009a), in the unit-root case, the amount of time spent by the process around any particular spatial point is of order $\sqrt{n}$ rather than n so that the corresponding convergence rate in nonparametric cointegrating regression is $\sqrt{\sqrt{n}h} = \left(nh^2\right)^{1/4}$. In the case of continuous location shifts, the number of effective observations at each location is of order $Mm^{\lambda}h$ which gives the convergence rate $\sqrt{Mm^{\lambda}h}$. Fourth, for both types of estimates, the conditional variance $\sigma^2(x) = E\left(u_t^2|X_t = x\right)$ does not play a role in the asymptotics. Instead, it is the unconditional variance that really matters. This result indicates that the width of the pointwise confidence bands for either type of estimate should remain largely the same across evaluation points in areas with abundant data.

2.2 Models with compact support

The formulation of the location-generating mechanism in (2.2) - (2.3) requires that the regressor locations eventually cover the whole real line as the sample size grows in order to accommodate the infinite support of the random elements in the structural model. This condition facilitates identification and the development of the asymptotic theory, but it also creates two problems. First, the usual convergence rate of the nonparametric local constant and local linear estimates is reduced in the case of continuous location shifts in an endogenous regressor because some portion of the data is used up to achieve identification by tracing out the curve through shifts in location. Second, the sample variance of the endogenous regressor needs to *expand* as the sample size n increases. A natural question is whether these two conditions are vital to identification and consistent nonparametric estimation.

It turns out the expansion of regressor location over the whole real line is unnecessary when u_{xt} is compactly supported. As indicated earlier, the nonparametric IV literature frequently assumes compact support for the endogenous regressor (e.g., Hall and Horowitz, 2005) in order to use standard functional analysis. We therefore consider a similar case where the location shifts occur in a compact set. The model we consider is as follows:

$$\begin{aligned} y_t &= g(X_t) + u_t,\ E(u_t|X_t) \neq 0 & (2.11)\\ X_t &= \mu_\alpha 1\{t \in A_\alpha\} + u_{xt}, & (2.12)\\ \mu_\alpha &= -\frac{L}{2} + \frac{L(\alpha-1)}{2m-1},\ \alpha = 1,\cdots,2m, & (2.13)\\ M &= \#\{t \in A_\alpha\},\ n = 2mM, \end{aligned}$$

where $t = 1, 2, \cdots, n$, the pair (u_t, u_{xt}) continues to satisfy Assumptions A1, A2, and A3(i), and L is a fixed positive number. As before, since we do not restrict u_{xt} to have zero mean, it is not restrictive to require that the $2m$ locations are spaced over a compact interval $[-L/2, L/2]$ that is symmetric around 0. Let $d_m = (2m-1)/L$. Then the distance between two contiguous locations is given by d_m^{-1}.

To proceed, we make the following assumptions.

Assumption A7. *The error term u_{xt} has compact support, i.e., $u_{xt} \in [\underline{u}, \overline{u}]$ a.s. for some finite numbers $\underline{u}$ and $\overline{u}$.*

Assumption A8. *For given x, L is sufficiently large such that $x \in (-L/2 + \overline{u},\ L/2 + \underline{u})$.*

Assumption A9. *As $(M, m) \to \infty$, $Md_mh \to \infty$, $Md_mh^5 \to c \in [0, \infty)$, and $Mhd_m^{-3} \to 0$.*

Note that under A7, the tail conditions in Assumptions A3(ii)-(iii) are redundant. A8 requires that $L > \overline{u} - \underline{u}$. Intuitively, it implies that the larger L is, the greater the portion of the true regression curve that can be identified and consistently estimated. A9 parallels A6 with m^λ and υ replaced by d_m and ∞, respectively.

The following theorem establishes the consistency and asymptotic normality of $\widehat{g}(x)$.

Theorem 2.2 *Suppose Assumptions A1-A2, A3(i), A4, A5 and A7-A9 hold. Then*

$$\sqrt{Md_mh}\left(\widehat{g}(x) - g(x) - h^2\mu_2(K)\left\{g'(x)\int_{\underline{u}}^{\overline{u}} f'_{u_x}(p)\,dp + \frac{1}{2}g''(x)\right\}\right) \to_d N\left(0, \sigma^2\nu_2(K)\right). \quad (2.14)$$

Theorem 2.2 indicates that the local constant estimate in the case of continuous location shifts can achieve the usual $\sqrt{nh}$-rate of consistency, since $Md_m \sim 2mM/L = O(n)$ for fixed L. The same is also true for the local linear estimate. This fast rate contrasts with the much slower rates achievable for nonparametric IV estimation without the advantage of location shifts. In comparison with the result in Theorem 2.1, we observe that the bias function in (2.14) contains the linear term, $g'(x)\int_{\underline{u}}^{\overline{u}} f'_{u_x}(p)\,dp$, which vanishes if $f_{u_x}(\underline{u}) = f_{u_x}(\overline{u})$, as in the infinite support case. If the last condition holds, then the local constant and local linear estimates share the same asymptotic distribution and bias as well. Otherwise, they only share the same asymptotic distribution after bias correction. The following remark reveals a further advantage of local smoothing techniques in dealing with issues of endogeneity.

Remark 7. (**Inconsistency of parametric regression**) If the regression function is parametric and if its functional form is known, parametric estimation becomes possible. In particular, if $g(\cdot)$ in (2.1) or (2.11) is known to be of the parametric form $g(x) = g(x, \beta)$, say, where $g(\cdot, \cdot)$ is known up to the finite dimensional parameter, β, then $g(\cdot)$ can be estimated by direct parametric estimation of β. However, under the conditions given in Theorem 2.2 and assuming that the form of $g(x) = g(x, \beta)$ is known, the nonlinear least squares (NLS) estimate $\widehat{\beta}$ of β is generally inconsistent because of endogeneity in the regressor. To see this, consider the simple case where the model has the linear form

$$\begin{aligned} y_t &= \beta_0 + \beta_1 X_t + u_t, \\ u_t &= \sigma(\epsilon_{yt} + \theta u_{xt}) / \left(1 + \theta^2\right)^{1/2}, \end{aligned}$$

where $\sigma > 0$, ϵ_{yt} and u_{xt} are each *iid* with mean zero and variance 1 and are mutually independent. That is, $g(x) = \beta_0 + \beta_1 x$ in (2.11), and $\text{cov}(u_t, u_{xt}) = \sigma\theta / \left(1 + \theta^2\right)^{1/2}$. Then the least squares (OLS) estimate $\widehat{\beta}_1$ of β_1 has the following limit

$$\widehat{\beta}_1 = \beta_1 + \frac{n^{-1}\sum_{t=1}^n \left(X_t - \overline{X}\right) u_t}{n^{-1}\sum_{t=1}^n \left(X_t - \overline{X}\right)^2} \to_p \beta_1 + \frac{\sigma\theta}{\sigma_X^2 \left(1 + \theta^2\right)^{1/2}}, \tag{2.15}$$

where $\overline{X} \equiv n^{-1}\sum_{t=1}^n X_t$, and

$$\sigma_X^2 \equiv p \lim_{n\to\infty} n^{-1} \sum_{t=1}^n \left(X_t - \overline{X}\right)^2 = E\left(u_{xt}^2\right) + \frac{L^2}{12}.$$

That is, $\widehat{\beta}_1$ is inconsistent for β_1 unless $\theta = 0$, viz., when X_t is exogenous. On the other hand, under A2 and since $E(u_{xt}) = 0$, the OLS estimate $\widehat{\beta}_0$ of β_0 is still consistent. These results imply that the parametric estimate $\widehat{g}^p(x) \equiv \widehat{\beta}_0 + \widehat{\beta}_1 x$ is inconsistent for $g(x) = \beta_0 + \beta_1 x$ at all points except $x = 0$. By contrast, according to Theorem 2.2, the nonparametric estimator is consistent for all $x \in (-L/2 + \overline{u},\ L/2 + \underline{u})$.

It is worth mentioning that if X_t has infinite second moment, then the limit (2.15) reveals that parametric regression may be consistent, due to the stronger signal in the regressor. Correspondingly, under the conditions of Theorem 2.1, which produce additional location variation in X_t, the NLS estimator may be consistent. However, the estimation bias may disappear at a very slow rate in this case, depending on the nature of the location-generating mechanism, and may not be eliminated for inferential purposes. For example, let $g(x) = \beta_0 + \beta_1 x$ in (2.1), $X_t = \mu_\alpha 1\{t \in A_\alpha\} + u_{xt}$, and suppose that the $2m$ values of μ_α are equally spaced locations on the interval $(-L_n/2, L_n/2)$, where L_n is an increasing slowly varying function at infinity, such as $L_n = \log\log n$, and both $m, M \to \infty$ as $n = 2mM \to \infty$. By direct calculation the OLS signal in this case is $n^{-1}\sum_{t=1}^n \left(X_t - \overline{X}\right)^2 = \frac{L_n^2}{12}\{1 + o_P(1)\}$ and the OLS estimate $\widehat{\beta}_1$ of β_1 has bias of $O\left(L_n^{-2}\right)$. More specifically,

$$\widehat{\beta}_1 = \beta_1 + \frac{12\sigma\theta}{L_n^2 \left(1 + \theta^2\right)^{1/2}} \{1 + o_P(1)\},$$

and, after centering, the limit distribution is given by

$$\frac{L_n\sqrt{n}}{12}\left\{\widehat{\beta}_1-\beta_1-\frac{\sigma\theta}{n^{-1}\sum_{t=1}^{n}\left(X_t-\overline{X}\right)^2\left(1+\theta^2\right)^{1/2}}\right\}\to_d N\left(0,\frac{\sigma^2}{12}\right). \tag{2.16}$$

In spite of the $O\left(L_n\sqrt{n}\right)$ convergence rate, (2.16) is not useable for inferential purposes because the bias term in (2.16) does not appear to be estimable at the required rate to be eliminated. In a separate paper, we demonstrate that by using nonparametric local constant estimation and spatial L_2 regression, the bias of the OLS estimator can be corrected up to order $O(n^{-1/2}L_n^{5/2})$, yielding an $(n/L_n)^{1/2}$-rate of consistent bias-corrected OLS estimator. Further, noting that m/L_n plays the role of m^λ in (2.3) and A6, and by following the line of proof in Theorem 2.1, it can be shown that the local constant and local linear estimates are $(nh/L_n)^{1/2}$ convergent. When the true regression function $g\left(x\right)$ is linear, the bias term in (2.5) vanishes for all choices of bandwidth. In this case, the optimal bandwidth h for local constant estimation does not diminish to zero as $n\to\infty$. For fixed bandwidth, the local constant estimate is $(n/L_n)^{1/2}$-consistent and performs as well as the bias-corrected OLS estimator in terms of convergence rate. When $g\left(x\right)$ is nonlinear and an optimal bandwidth $h\sim\left(L_n^2/n\right)^{1/5}$ is selected, the nonparametric estimates are $(n/L_n)^{2/5}$ consistent. Intuitively, the consistency of the parametric estimate relies upon the fact that $L_n\to\infty$, so that the sample variance of $\{X_t, 1\leq t\leq n\}$ tends to infinity with n. This signal expansion identifies the true regression line in the limit. But if the rate at which the support of the $2m$ locations expands is slow, as in the above example, the bias in the parametric estimate is large and hard to eliminate. On the other hand, the nonparametric estimates draw support from the divergence of both M and m, converge to the true regression line at the rate $(nh/L_n)^{1/2}$, and may be used for inference as usual.

3 Application to structural panel models with individual effects

As remarked earlier, model (2.1) - (2.3) is somewhat idealized because practical empirical examples with cross section data where continuous location shifts may occur are likely to be uncommon in economics. However, panel models that involve individual effects covering a large and continuously distributed population present some interesting and potentially realistic applications of our results. To include such cases, we therefore consider the following nonparametric panel data model

$$y_{it} = g\left(X_{it}\right)+u_{it},\ E\left(u_{it}|X_{it}\right)\neq 0, \tag{3.1}$$

$$X_{it} = \mu_{xi}+u_{xit}, \tag{3.2}$$

where $i=1,2,\cdots,N$, $t=1,2,\cdots,T$, and the individual effects μ_{xi} are assumed to be independent over i and uniformly distributed (U) over an interval of expanding length $L=N^{1-\lambda}$ for some $\lambda\in(0,1)$. The precise conditions on the μ_{xi} and the errors u_{it} and u_{xit} are as follows.

Assumption A10. *(u_{it}) are iid $\left(0,\sigma^2\right)$ with $E\left(u_{it}^4\right)<\infty$, and (u_{xit}) are iid(μ_x,σ_x^2) over i and t.*

Assumption A11. *(μ_{xi}) are iid$U(-L/2,L/2)$ independent of the process $\{(u_{it},u_{xit})\}$ and with $L=N^{1-\lambda}$ for some $\lambda\in(0,1)$.*

Since our focus of attention is on the impact of the individual effects, we concentrate on the case where the component errors have an *iid* structure. However, we expect that our limit results will hold with minor modifications for cases where the equation errors are stationary, as in the analysis of Section 6.4. Also, related limit theory may be expected for integrated and near integrated panel regressors, following the treatment of Wang and Phillips (2009b, c). But we do not pursue those generalizations here. Instead, we work under A10 and, following our earlier regression approach, we consider the pooled local level nonparametric estimate

$$\widehat{g}_{panel}(x)=\frac{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}y_{it}K_h(X_{it}-x)}{\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}K_h(X_{it}-x)}.$$

The following theorem shows consistency and gives the asymptotic distribution of $\widehat{g}_{panel}(x)$.

Theorem 3.1 *Let panel observations (y_{it},X_{it}) be generated according to (3.1) - (3.2) with individual effects and errors satisfying A10 and A11. Suppose that the joint p.d.f. $f(\cdot,\cdot)$ of (u_{it},u_{xit}) and the marginal p.d.f. $f_{u_x}(\cdot)$ of u_{xit} satisfy the conditions in A3. Suppose further that $g(\cdot)$ and $K(\cdot)$ satisfy A4 and A5, respectively. Suppose also that as $N\to\infty$, T is either fixed or tends to ∞, that $TN^{\lambda}h\to\infty$, $TN^{\lambda}h^5\to c'\in[0,\infty)$, $TN^{\lambda-2v(1-\lambda)}h\to0$ and $Th\to0$. Then we have*

$$\sqrt{TN^{\lambda}h}\left\{\widehat{g}_{panel}(x)-g(x)-\frac{1}{2}h^2\mu_2(K)g''(x)\right\}\to_d N\left(0,\sigma^2\nu_2(K)\right), \tag{3.3}$$

where $\mu_2(K)$ and $\nu_2(K)$ are defined in Theorem 2.1.

Note that we have strengthened the moment condition on the structural error term to facilitate the verification of the Liapounov condition. In addition, the bandwidth condition is strengthened to take into account the dependence of X_{it} across t. Here, we require that $Th\to0$ in order to verify that the bias due to the presence of endogeneity is of order $o\left(h^2\right)$.

As in Section 2.2, one can also consider the case where u_{xit} has compact support $[\underline{u},\overline{u}]$ and the individual effects μ_{xi} are spread over a fixed compact interval. In this case, L in A11 is a large fixed constant. Under the condition that $x\in(-L/2+\overline{u},\,L/2+\underline{u})$, one can show that the result in Theorem 3.1 continues to hold with N^{λ} replaced by $(N-1)/L$ and the addition of the linear bias term $h^2\mu_2(K)g'(x)\int_{\underline{u}}^{\overline{u}}f'_{u_x}(p)\,dp$. Consequently, the convergence rate of the pooled local level nonparametric estimate is comparable to that of nonparametric regression with a stationary regressor on a compact support.

The model (3.1) - (3.2) may be extended to allow for additional individual effects on y_{it} that do not operate through the endogenous regressor X_{it}. To allow for this possibility, the equation error u_{it} in (3.1) may be specified in error component form as follows

$$u_{it} = \mu_{yi} + v_{it},\ E\left(v_{it}|X_{it}\right) \neq 0,\ E\left(\mu_{yi}|X_{it}\right) \neq 0.$$

Assume that the μ_{yi} are $iid(0, \sigma^2_{\mu_y})$ across i and the v_{it} are $iid\left(0, \sigma^2_v\right)$ across both i and t, and both have finite fourth moments and are independent of μ_{xi}. Then under similar conditions as those stated in Theorem 3.1, the result of the theorem continues to hold.

This limit theory reveals that panel models with individual effects covering a large population provide a natural mechanism for identifying structural nonparametric elements. The dispersion of individual effects in the population introduces a form of leverage similar to that of continuous location shifts in the cross section case, thereby helping to trace out functional form and enable consistent estimation even in the presence of endogenous regressors.

4 Simulations

This section reports a small Monte Carlo experiment to evaluate the finite sample performance of the nonparametric local level and local linear estimators when the regressor exhibits continuous location shifts. Data is generated according to the following data generating process (DGP)

$$\begin{aligned} y_t &= g\left(X_t\right) + u_t, \\ X_t &= \mu_\alpha 1\left\{t \in A_\alpha\right\} + u_{xt},\ \mu_\alpha = \frac{\alpha}{m^\lambda},\ \alpha = -m+1, \cdots, m, \\ u_t &= \sigma\left(\epsilon_{yt} + \theta u_{xt}\right) / \left(1+\theta^2\right)^{1/2},\ \sigma = S_g, \end{aligned}$$

where ϵ_{yt} are $iid\ N\left(0,1\right)$, u_{xt} are $iid\ \chi^2\left(1\right)$ variates normalized to have mean 0 and variance 1, and S_g is the sample standard deviation of $g\left(X_t\right)$. By construction, the signal to noise ratio is maintained to be 1 throughout these simulations in order to enhance comparability across experiments. We consider the two regression functions

DGP 1: $\quad g\left(x\right) = \beta_0 + \beta_1 x,\ \ \beta_0 = 10,\ \beta_1 = -1,$

DGP 2: $\quad g\left(x\right) = 1/(1 - 2\beta_0 \sin\left(x/\beta_1\right) + \beta_0^2),\ \beta_0 = 0.5,\ \beta_1 = 10.$

Simulations are performed for $\theta = 0.32$ (weak endogeneity, corr(u_t, u_{xt})= 0.3) and $\theta = 2.07$ (strong endogeneity, corr(u_t, u_{xt})= 0.9), location shift parameter $m = 0, 1, 2, ..., 50$, and for the sample sizes n = 200, 800 and 3200. We generate the location shift points μ_α as $2m$ evenly spaced points between $[-\sqrt{m}, \sqrt{m}]$ unless otherwise specified, so that $\lambda = 0.5$, $m = 0$ corresponds to a single location and $m = 1$ to two locations. To save space, findings are reported only for a selection of these cases and are summarized graphically in what follows.

We consider three estimators. The first is the parametric estimator, which is based on the presumption that the functional form of the true DGP is known. A linear regression is run in the case of DGP 1, and for DGP 2 a nonlinear least squares (NLS) regression yields estimates

$(\widehat{\beta}_0, \widehat{\beta}_1)$ from which the parametric estimate $\widehat{g}^p(x)$ of $g(x)$ at x is constructed. It is worth mentioning that the NLS criterion function is highly nonconvex and different starting values can result in very different estimates in numerical optimization. In the simulation, the starting values in the numerical iteration are set at the true values, which gives the most favorable estimates of (β_0, β_1). The other two estimators are the local level and local linear estimators based on a Gaussian kernel and with bandwidth chosen according to the Silverman rule of thumb $h_{rot} = s_X n^{-1/5}$. We also tried the LSCV technique with h_{cv} minimizing the criterion (2.8), choosing the simple weight function $w(X_t) = 1\{(|X_t - \overline{X}| \leq 1.5 s_X\}$, where $\overline{X}$ and s_X denote the sample mean and standard deviation of X_t, respectively. Results based on the LSCV were found to be similar to those based on h_{rot} but are much more costly in terms of computation time, so only the results for h_{rot} are reported.

As performance measures, we report absolute bias (Bias), and root mean squared error (Rmse) for the three types of estimates on a grid of G values of x using $R = 10,000$ replications. Since different estimators have different boundary properties, we distinguish between interior and boundary points. For interior points, we choose $G = 101$ grid points which are equally spaced on the interval $[-m^{1/3}, m^{1/3}]$, whereas for boundary points we choose $G = 100$ grid points which are equally spaced on $[-m^{1/2}, -m^{1/3}] \cup [m^{1/3}, m^{1/2}]$. When $m = 0$ and 1, we only consider interior points of evaluation. When $m = 0$, few observations are smaller than -0.8 and we evaluate the estimates over $[-0.8, 1.5]$.

Figs. 3 and 4 plot samples of observations (X_t, y_t) based on DGPs 1 and 2 respectively for various choices of m and under both weak and strong endogeneity. The top panel plots in both Figs. 3 and 4 show that for either weak or strong endogeneity, the data do not cluster around the true regression curve, suggesting that the true curves are not identified without external information in this case. Interestingly, as we allow the regressor X_t to exhibit location shifts, the shifts begin to track out and identify the regression curve, as suggested in the original illustration of Fig. 1. Even with only $2m = 4$ locations shifts, the regression curves can be well identified in areas of abundant observations. As the number of location shifts increase, the data can identify a greater portion of the true regression curves. With $2m = 100$ location shifts, we find that almost the whole curve is identified, the only exception occurring in the far right tail where the data is sparse and outliers dominate the regression.

Figs. 3 and 4 show the effects of numerous small location shifts on the capacity to identify and consistently estimate unknown structural functions. In practice, there may be fewer location shifts and these may be more spread out over the domain of the function, as in the original discussion of the demand and supply example of Figs. 1 and 2. This type of scenario is investigated further in Figs. 7 and 8, which demonstrate that as few as 5-8 locations with about 50-100 observations around each location may be sufficient to provide good estimates of the true regression function in practice.

Figs 5 and 6 show the relative performance of the three estimates $\widehat{g}^p$, $\widehat{g}$, and $\widetilde{g}$ of g for DGPs 1 and 2, respectively. In the case of weak endogeneity, the bias of $\widehat{g}^p$ is not very severe in comparison with the average magnitude of $g(x)$. Despite this, the local constant and local

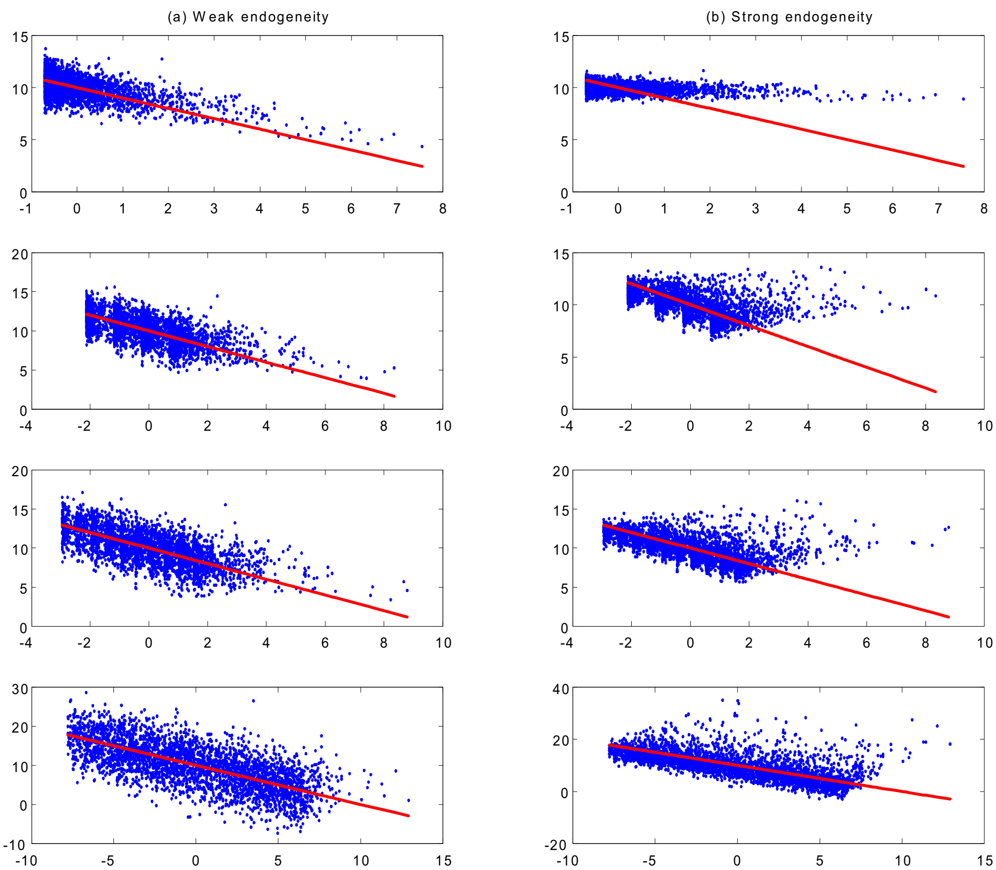

Figure 3: Sample plots of generated data (X_t, y_t) for DGP 1 with continuous location shifts ($n = 3200$). Horizontal axis: X_t, vertical axis: y_t. The solid line is the true regression curve. The four panels from the top to the bottom correspond to $m = 0$, 2, 5, and 50, respectively.

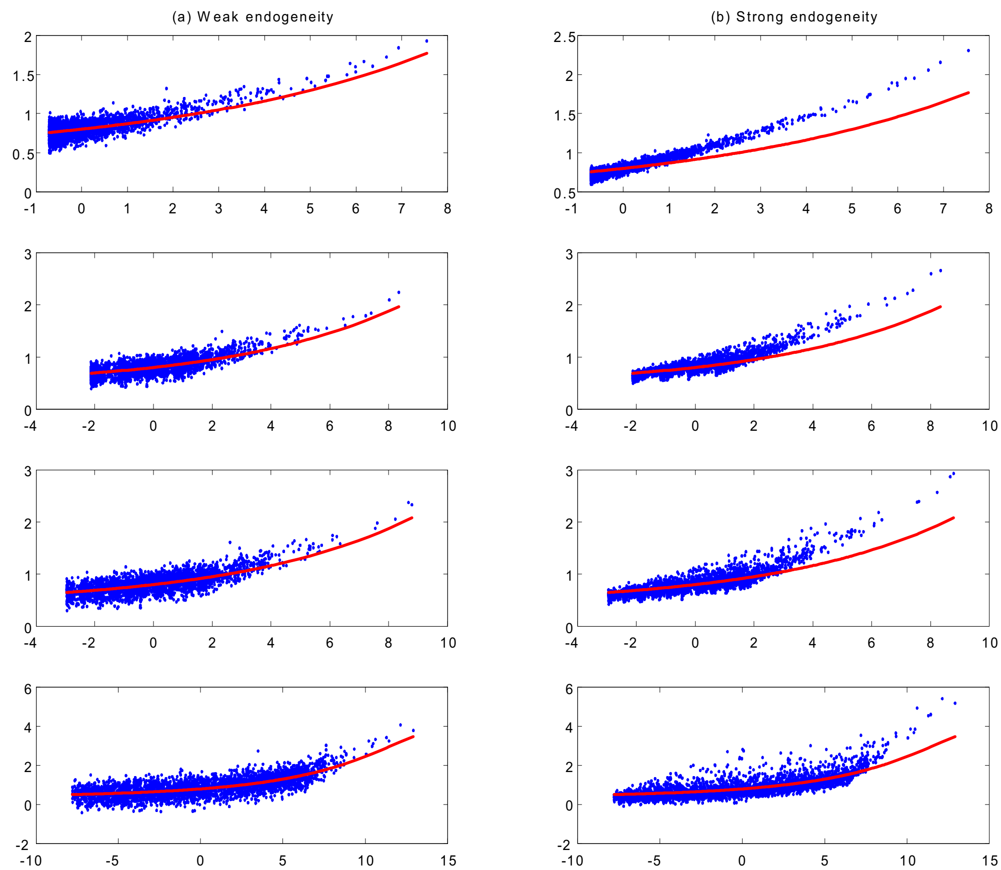

Figure 4: Sample plots of generated data (X_t, y_t) for DGP 2 with continuous location shifts $(n = 3200)$. Horizontal axis: X_t, vertical axis: y_t. The solid line is the true regression curve. The four panels from the top to the bottom correspond to $m = 0$, 2, 5, and 50, respectively.

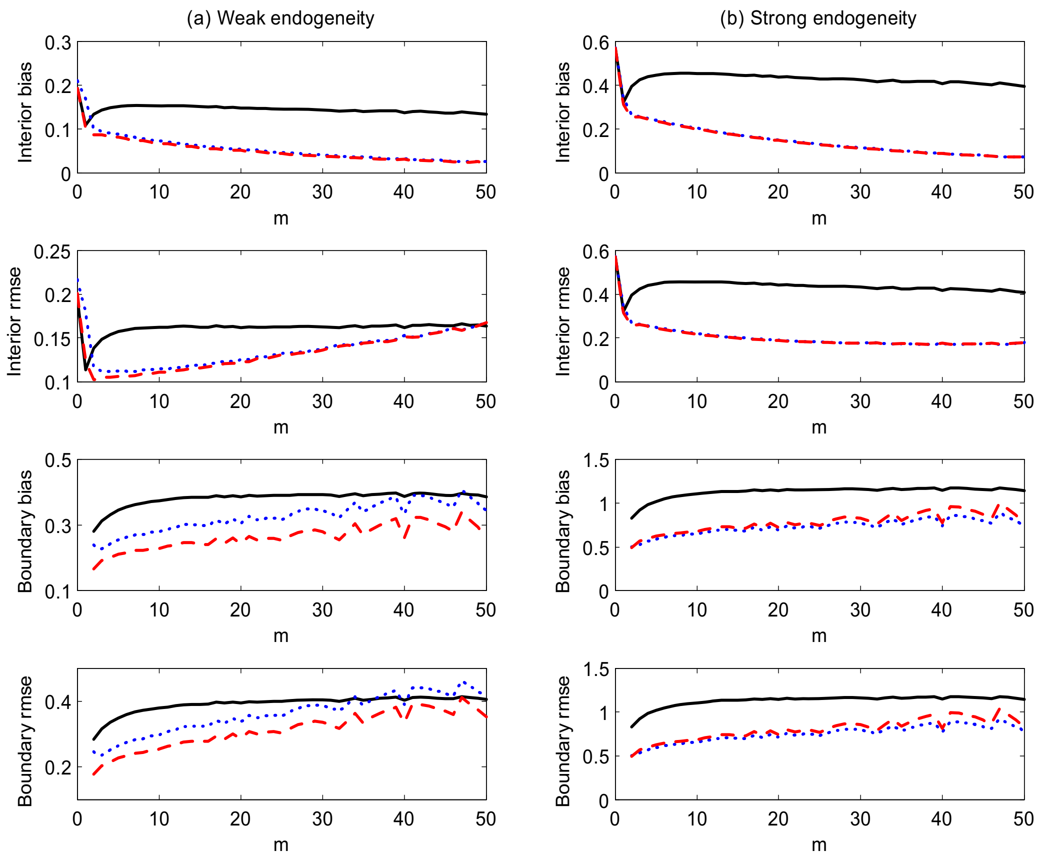

Figure 5: Interior and boundary bias and rmse for DGP 1 with continuous location shifts ($n = 3200$). Parametric estimate (solid line), local constant estimate (dotted line), local linear estimate (dashed line). $\mathrm{Corr}(u_t, u_{xt}) = 0.3$ and 0.9 for weak and strong endogeneity cases, respectively.

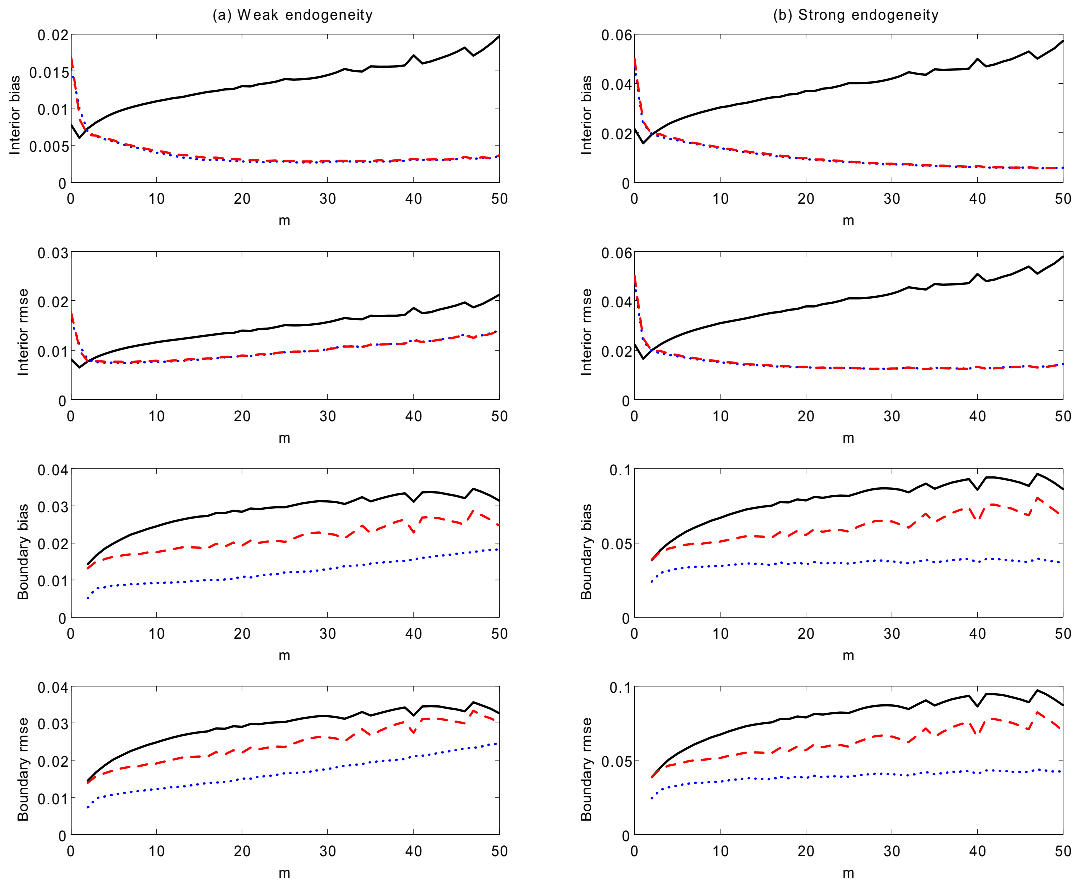

Figure 6: Interior and boundary bias and rmse for DGP 2 with continuous location shifts ($n = 3200$). Parametric estimate (solid line), local constant estimate (dotted line), local linear estimate (dashed line). $\text{Corr}(u_t, u_{xt}) = 0.3$ and 0.9 for weak and strong endogeneity cases, respectively.

linear estimates achieve significant bias reduction for both interior and boundary points and with as small as $2m = 4$ locations. Due to the high variance of the kernel estimates, the reduction in Rmse is not as large as it is in bias. In the case of strong endogeneity, both the local constant and local linear estimates achieve substantially more bias reduction than in the case of weak endogeneity. In terms of Rmse, these kernel estimates dominate the parametric estimates almost uniformly in the number of location shifts.

The behavior of the local constant and local linear estimates is not completely predicted by either Theorem 2.1 or Theorem 2.2. For interior points, the local linear estimates behave similarly to the local constant estimates, as predicted by Theorem 2.1, but they substantially differ at the extremities of the range. Perhaps unexpectedly (given the usual bias reduction capability of the local linear estimator at the boundary), the local linear estimator exhibits larger bias and root mean squared error over the local level estimator close to the boundary for DGP 2. Near the boundary, the local linear estimator appears to be more susceptible to the endogeneity bias of the parametric estimator.

Even though it is not reported here, the performance of these estimates was also investigated for small and intermediate sample sizes ($n = 200$, 800) and for a full range of the location shift parameter m. The superior performance of the kernel estimates over the parametric estimates holds uniformly for both small and intermediate sample sizes when the endogeneity is strong. In the case of weak endogeneity, both nonparametric estimates achieve significant bias reduction at interior points, but higher Rmse than the parametric estimates for small sample sizes. But as the sample size increases, the gains in bias reduction for the nonparametric estimates exceed the loss in variance inflation so that they still exhibit lower Rmse in general for both interior and boundary points.

In an additional experiment, Fig. 7 shows results for Monte Carlo approximations to the mean of the three estimates of $g(x)$ in DGP2 for the case of a single value $m = 4$, producing $2m = 8$ locations equally spaced over the interval $[-6, 6]$. Fig. 8 shows Monte Carlo approximations to the parametric and local level estimates of $g(x)$ in DGP 2 together with 95% pointwise "estimation bands".[1,2] Apparently, both the local constant and local linear estimates outperform the parametric estimates over much of the range of the data. The two kernel estimates behave equally well in the middle range of the data and differ slightly from each other at points close to the boundary. Because the shape parameter (β_1) in DGP 2 cannot be estimated accurately by parametric methods, the 95% estimation bands for the parametric estimates of $g(x)$ do not contain the true regression curves over a wide range of the regressor. By contrast, the 95% estimation bands for the local constant estimates contain the true regression curves most of the time and miss the target only in the left tail. Similar results were found for DGP1

[1] As in Hall and Horowitz (2005) and Wang and Phillips (2009b), for each grid point x_i the estimation bands contain 95% of the 10,000 simulated values of $\widehat{g}^p(x_i)$ or $\widehat{g}(x_i)$.

[2] The local linear estimates and their estimation bands are not included in Fig. 8, as they largely coincide with results for the local level estimator except in the tail, where the local linear estimates tend to have a narrower estimation band than the local level estimates.

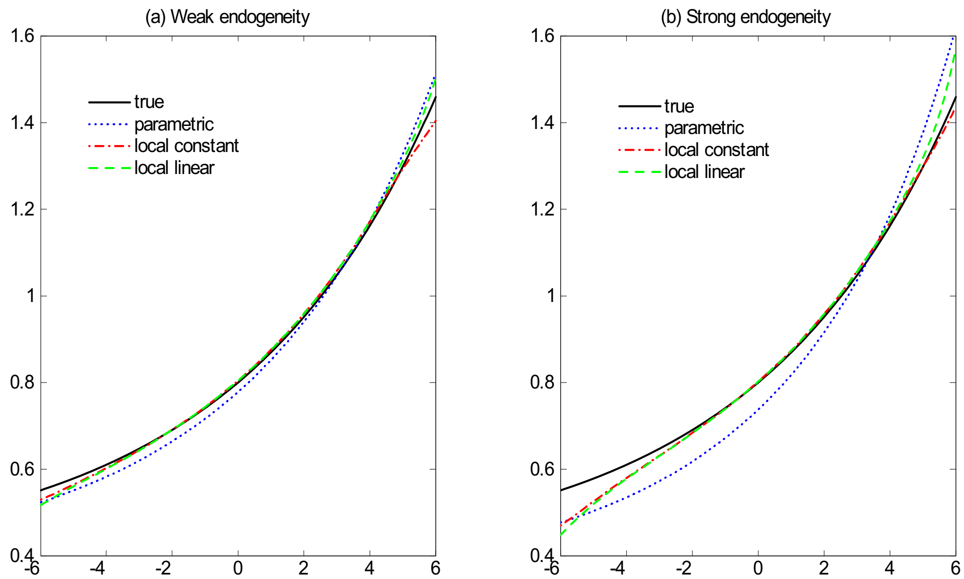

Figure 7: Graphs of various estimates for $g(x)$ in DGP 2 (solid line): $n = 800$, $m = 4$, $\mathrm{corr}(u_t, u_{xt}) = 0.3$ and 0.9 for the weak and stong endogeneity cases, respectively

and are not reported here to save space.

5 Concluding remarks

This paper shows that location shifts in a regressor can play an effective role in tracing out a regression curve in spite of endogeneity in the regressor. In part, these location shifts act as an instrument that moves the data along the curve, and in part they add variation to the regressor that enhances the signal/noise ratio. In both respects, such location shifts act in a manner analogous to the random wandering feature of unit root regressors in a cointegrating regression equation, thereby explaining the consistency of simple nonparametric regression in both cases.

Importantly, in all these cases there is no need for nonparametric IV estimation or the complications of functional inversion and regularization. Just as in textbook discussions of identification, regression curves may be identified and estimated by nonparametric methods provided the data embody a mechanism for tracing out the regression curve of interest. Our results also reveal the significance of the common assumption in the nonparametric IV literature of a compact support. In such cases, nonparametric regression can have an important advantage over parametric regression (even when the true form of the regression function is known) in terms of its consistency and a $\sqrt{nh}$ convergence rate in contrast to the inconsistency of

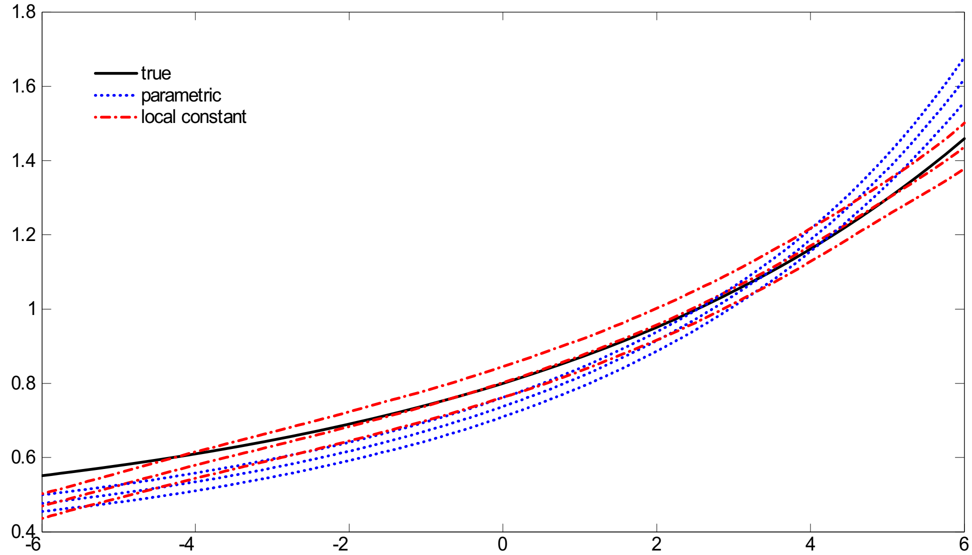

Figure 8: Graphs of the parametric and local constant estimates for $g(x)$ in DGP 2 with 95% estimation bands. $n = 800$, $m = 4$, $\text{corr}(u_t, u_{xt}) = 0.9$.

parametric regression.

As our main results show, location shifts remove the effects of endogeneity and ensure that the local linear and local constant estimates have the same asymptotic distribution and bias in the case of infinite support. This result is analogous to nonparametric cointegrating regression where the limit theory also involves only a single $O\left(h^2\right)$ bias term (Wang and Phillips, 2009b). Interestingly, simulations reveal that when there is strong endogeneity in the regression, local linear estimation is outperformed by local level regression near the boundary and appears to be more susceptible to endogeneity bias in this region of the sample space.

The main practical import of the results in the present paper is likely to be in panel data applications of the type studied in Theorem 3.1, where individual effects provide a natural mechanism for shifting around the regressor and increasing its variation. Structural panel data regression models of this type seem relevant in many contexts where there is an unknown but common regression function and the regressor is endogenous.

6 Appendix: technical results and proofs

6.1 Proof of Theorem 2.1

We first state and prove two lemmas that are used in the proof of Theorem 2.1.

Lemma 6.1 *Let $f_u(\cdot)$ be the p.d.f. of $\{u_t\}$. Suppose A2-A3 hold. Then*

$\int u \left[m^{-\lambda} \sum_{a=-m+1}^{m} f(u, x-\mu_\alpha) - f_u(u) \right] du = O\left(m^{-2\lambda} + m^{-v(1-\lambda)}\right).$

Proof. The trapezoidal rule approximates $\int_a^b g(p)\, dp$ by $\frac{1}{2}(b-a)(g(a)+g(b))$ and the approximation error is given by the Newton-Cotes formula (e.g., Stoer and Bulirsch (1993, p. 162))

$$\frac{1}{2}(b-a)(g(a)+g(b)) - \int_a^b g(p)\, dp = \frac{1}{12}(b-a)^3 g''(c), \tag{6.1}$$

where the second order derivative $g''(\cdot)$ of $g(\cdot)$ is continuous and $c \in (a,b)$. Write

$$\begin{aligned}
& \frac{1}{m^\lambda} \sum_{a=-m+1}^{m} f(u, x-\mu_\alpha) - f_u(u) \\
= & \left[\frac{1}{m^\lambda} \sum_{a=-m+1}^{m} f(u, x-\mu_\alpha) - \int_{(-m+1)/m^\lambda}^{m/m^\lambda} f(u, x-p)\, dp \right] \\
& + \int_{(-m+1)/m^\lambda}^{m/m^\lambda} f(u, x-p)\, dp - \int_{-\infty}^{\infty} f(u, x-p)\, dp \\
= & \left\{ \sum_{a=-m+1}^{m} \frac{1}{2} \left[f\left(u, x - \frac{\alpha-1}{m^\lambda}\right) + f\left(u, x - \frac{\alpha}{m^\lambda}\right) \right] \frac{1}{m^\lambda} - \int_{(-m+1)/m^\lambda}^{m/m^\lambda} f(u, x-p)\, dp \right\} \\
& + \left\{ \int_{(-m+1)/m^\lambda}^{m/m^\lambda} f(u, x-p)\, dp - \int_{-\infty}^{\infty} f(u, x-p)\, dp \right\} \\
& + \frac{1}{2m^\lambda} \left\{ f\left(u, x - \frac{m}{m^\lambda}\right) - f\left(u, x + \frac{m}{m^\lambda}\right) \right\} \\
\equiv & \; I_{1m}(u) + I_{2m}(u) + I_{3m}(u).
\end{aligned}$$

By (6.1) and Assumption A3(i)

$$\begin{aligned}
m^{2\lambda} \int_{-\infty}^{\infty} u I_{1m}(u)\, du &= \int_{-\infty}^{\infty} \frac{u}{12 m^\lambda} \sum_{a=-m+1}^{m} f_2''(u, x-p_\alpha)\, du \\
&\rightarrow \frac{1}{12} \int_{-\infty}^{\infty} u \int_{-\infty}^{\infty} f_2''(u, x-p)\, dp du = \frac{1}{12} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u f_2''(u, p)\, dp du,
\end{aligned}$$

where p_α lies between $(\alpha-1)/m^\lambda$ and α/m^λ. By Assumptions A2 and A3 (ii)-(iii)

$$\begin{aligned}
\left| \int_{-\infty}^{\infty} u I_{2m}(u)\, du \right| &\leq \int_{-\infty}^{\infty} |u| \left| \int_{(-m+1)/m^\lambda}^{m/m^\lambda} f(u, x-p)\, dp - \int_{-\infty}^{\infty} f(u, x-p)\, dp \right| du \\
&\leq (m-1)^{-v(1-\lambda)} \int_{-\infty}^{\infty} |u| C_f(u)\, du = O\left(m^{-v(1-\lambda)}\right),
\end{aligned}$$

and

$$\begin{aligned}
\left| \int_{-\infty}^{\infty} u I_{3m}(u)\, du \right| &\leq \frac{1}{2m^\lambda} \left\{ \int_{-\infty}^{\infty} |u| f\left(u | x - \frac{m}{m^\lambda}\right) du f_{u_x}\left(x - \frac{m}{m^\lambda}\right) \right. \\
&\quad \left. + \int_{-\infty}^{\infty} |u| f\left(u | x + \frac{m}{m^\lambda}\right) du f_{u_x}\left(x + \frac{m}{m^\lambda}\right) \right\} \\
&= O\left(m^{-(v+1)(1-\lambda)} m^{-\lambda}\right),
\end{aligned}$$

where $f(\cdot|\cdot)$ is the conditional p.d.f. of u_t given u_{xt}. This completes the proof. ■

Lemma 6.2 *Let* $\Theta_{jn} = \frac{1}{Mm^{\lambda}} \sum_{t=1}^{n} (X_t - x)^j K_h(X_t - x)$ *for* $j = 0, 1, 2$. *Suppose A1-A6 hold. Then*

(i) $E\Theta_{0n} = 1 + o(1)$,

(ii) $E\Theta_{1n} = o(h^2)$,

(iii) $E\Theta_{2n} = h^2 \mu_2(K) + o(h^2)$,

(iv) $Var(\Theta_{jn}) = h^{2j-1} (Mm^{\lambda})^{-1} \int z^{2j} K^2(z)\, dz + o(h^{2j-1} (Mm^{\lambda})^{-1})$ *for* $j = 0, 1, 2$.

Proof.

$$\begin{aligned} E\Theta_{jn} &= \frac{1}{m^{\lambda}} \sum_{a=-m+1}^{m} E\left[(\mu_\alpha + u_{xt} - x)^j K_h(\mu_\alpha + u_{xt} - x)\right] \\ &= \frac{1}{m^{\lambda}} \sum_{a=-m+1}^{m} \int z^j K(z) f_{u_x}(x - \mu_\alpha + hz)\, dz. \end{aligned}$$

If $j = 0$, we have

$$E\Theta_{0n} = \frac{1}{m^{\lambda}} \sum_{a=-m+1}^{m} f_{u_x}(x - \mu_\alpha) + \frac{h^2 \mu_2(K)}{2m^{\lambda}} \sum_{a=-m+1}^{m} f''_{u_x}(x - \mu_\alpha) + R_{0n},$$

where $R_{0n} = \frac{1}{2} h^2 \int z^2 K(z) \int_0^1 m^{-\lambda} \sum_{a=-m+1}^{m} \left[f''_{u_x}(x - \mu_\alpha + whz) - f''_{u_x}(x - \mu_\alpha)\right](1 - w)\, dwdz$ is the remainder term. Noting

$$\begin{aligned} \frac{1}{m^{\lambda}} \sum_{a=-m+1}^{m} f_{u_x}(x - \mu_\alpha) &\approx \int_{(-m+1)/m^{\lambda}}^{m/m^{\lambda}} f_{u_x}(x - p)\, dp \to \int_{-\infty}^{\infty} f_{u_x}(x - p)\, dp = 1, \\ \frac{1}{m^{\lambda}} \sum_{a=-m+1}^{m} \left|f''_{u_x}(x - \mu_\alpha)\right| &\approx \int_{(-m+1)/m^{\lambda}}^{m/m^{\lambda}} \left|f''_{u_x}(x - p)\right| dp \to \int_{-\infty}^{\infty} \left|f''_{u_x}(x - p)\right| dp < \infty, \end{aligned}$$

it follows that $R_{0n} = o(h^2)$ by the dominated convergence theorem. Thus, $E\Theta_{0n} = 1 + o(1)$. Similarly, for $j = 1, 2$, we have

$$E\Theta_{1n} = \frac{h^2 \mu_2(K)}{m^{\lambda}} \sum_{a=-m+1}^{m} f'_{u_x}(x - \mu_\alpha) + o(h^2) = h^2 \mu_2(K) \int_{-\infty}^{\infty} f'_{u_x}(x - p)\, dp + o(h^2) = o(h^2),$$

since $\int_{-\infty}^{\infty} f'_{u_x}(x - p)\, dp = 0$, and

$$E\Theta_{2n} = \frac{h^2 \mu_2(K)}{m^{\lambda}} \sum_{a=-m+1}^{m} f_{u_x}(x - \mu_\alpha) + o(h^4) = h^2 \mu_2(K) + o(h^2).$$

Now, by Jensen's inequality, a change of variables, and since (u_t, u_{xt}) is iid, we have

$$\begin{aligned}
\operatorname{Var}(\Theta_{jn}) &= \frac{1}{Mm^{2\lambda}} \sum_{a=-m+1}^{m} \operatorname{Var}\left((\mu_\alpha + u_x - x)^j K_h (\mu_\alpha + u_x - x) \right) \\
&\leq \frac{h^{2j}}{Mm^{2\lambda}h} \sum_{a=-m+1}^{m} \int z^{2j} K^2 (z) f_{u_x} (x - \mu_\alpha + hz) \, dz \\
&= \frac{h^{2j-1} \int z^{2j} K^2 (z) \, dz}{Mm^\lambda} \frac{1}{m^\lambda} \sum_{a=-m+1}^{m} f_{u_x} (x - \mu_\alpha) + O\left(\frac{h^{2j+1}}{Mm^\lambda} \right) \\
&= \frac{h^{2j-1} \int z^{2j} K^2 (z) \, dz}{Mm^\lambda} + o\left(\frac{h^{2j-1}}{Mm^\lambda} \right).
\end{aligned}$$

■

To prove Theorem 2.1, consider the usual bias-variance decomposition of $\widehat{g}(x) - g(x)$:

$$\widehat{g}(x) - g(x) = \frac{\frac{1}{Mm^\lambda} \sum_{t=1}^{n} \{ g(X_t) - g(x) \} K_h (X_t - x)}{\frac{1}{Mm^\lambda} \sum_{t=1}^{n} K_h (X_t - x)} + \frac{\frac{1}{Mm^\lambda} \sum_{t=1}^{n} u_t K_h (X_t - x)}{\frac{1}{Mm^\lambda} \sum_{t=1}^{n} K_h (X_t - x)}. \quad (6.2)$$

By Lemmas 6.2(i) and (iv), the Chebyshev inequality, and Assumption A6,

$$\Theta_{0n} = \frac{1}{Mm^\lambda} \sum_{t=1}^{n} K_h (X_t - x) = 1 + O_P\left((Mm^\lambda h)^{-1/2} \right) = 1 + o_P(1). \quad (6.3)$$

Now write

$$\begin{aligned}
& \frac{1}{Mm^\lambda} \sum_{t=1}^{n} \{ g(X_t) - g(x) \} K_h (X_t - x) \\
= \; & \frac{g'(x)}{Mm^\lambda} \sum_{t=1}^{n} (X_t - x) K_h (X_t - x) + \frac{g''(x)}{2Mm^\lambda} \sum_{t=1}^{n} (X_t - x)^2 K_h (X_t - x) + R_n(x) \\
\equiv \; & g'(x) \Theta_{1n} + \frac{g''(x)}{2} \Theta_{2n} + R_n(x),
\end{aligned}$$

where $R_n(x) = \frac{1}{2Mm^\lambda} \sum_{t=1}^{n} [g''(X_t^*) - g''(x)] (X_t - x)^2 K_h (X_t - x)$, and X_t^* lies between X_t and x. Following the proof of Lemma 6.2, it is easy to show that $R_n(x) = o(h^2)$. Then by Lemma 6.2 (ii)-(iv), the Chebyshev inequality, and Assumption A6

$$\begin{aligned}
& \frac{1}{Mm^\lambda} \sum_{t=1}^{n} \{ g(X_t) - g(x) \} K_h (X_t - x) \\
= \; & h^2 \mu_2 (K) \frac{g''(x)}{2} + o_P(h^2) + O_P\left(\sqrt{\frac{h}{Mm^\lambda}} \right) \\
= \; & h^2 \mu_2 (K) \frac{g''(x)}{2} + o_P(h^2). \quad (6.4)
\end{aligned}$$

Let $\Theta_{3n} = \frac{\sqrt{h}}{\sqrt{Mm^\lambda}} \sum_{t=1}^{n} u_t K_h (X_t - x)$. We show that $\Theta_{3n} - E\Theta_{3n} \to_d N\left(0, \sigma^2 \nu_2 (K)\right)$. Write

$$\Theta_{3n} \equiv \frac{\sqrt{h}}{\sqrt{Mm^\lambda}} \sum_{t=1}^{n} u_t K_h (X_t - x) \equiv \sum_{t=1}^{n} Z_t,$$

where $Z_t = (\sqrt{h}/\sqrt{Mm^\lambda})u_t K_h(X_t - x)$. By a change of variables, the Fubini theorem, Lemma 6.1 and under A2 and A6

$$\begin{aligned}
E\Theta_{3n} &= \frac{\sqrt{Mm^\lambda h}}{m^\lambda}\sum_{a=-m+1}^{m} E\left[u_t K_h\left(\mu_\alpha + u_{xt} - x\right)\right] \\
&= \frac{\sqrt{Mm^\lambda h}}{m^\lambda}\sum_{a=-m+1}^{m}\int\int uK(z)\, f(u, x-\mu_\alpha + hz)\, dzdu \\
&= \sqrt{Mm^\lambda h}\int K(z)\left\{\int u\frac{1}{m^\lambda}\sum_{a=-m+1}^{m} f(u, x-\mu_\alpha+hz)\, du\right\} dz \\
&= \sqrt{Mm^\lambda h}\int K(z)\left\{O\left(m^{-2\lambda} + m^{-\upsilon(1-\lambda)}\right) + \int uf_u(u)\, du\right\} dz \\
&= \sqrt{Mm^\lambda h}O\left(m^{-2\lambda} + m^{-\upsilon(1-\lambda)}\right) = o(1), \qquad (6.5)
\end{aligned}$$

where we use the fact that the result in Lemma 6.1 also holds uniformly in a small neighborhood of x. Similarly, by (6.5) and Assumptions A1 and A2,

$$\begin{aligned}
\operatorname{Var}(\Theta_{3n}) &= \frac{h}{m^\lambda}\sum_{a=-m+1}^{m} E\left[u_t^2 K_h^2\left(\mu_\alpha + u_{xt} - x\right)\right] - o(1) \\
&= \frac{1}{m^\lambda}\sum_{a=-m+1}^{m}\int u^2 K(z)^2 f(u, x-\mu_\alpha+hz)\, dzdu + o(1) \\
&= \nu_2(K)\int u^2 \int f(u, x-p)\, dpdu + o(1) \\
&= \nu_2(K)\sigma^2 + o(1).
\end{aligned}$$

To show the asymptotic normality of $\Theta_{3n} - E\Theta_{3n}$, by the above variance calculation and the independence of $\{u_t, u_{xt}\}$ across t, it suffices to check the Liapounov condition. Let $\overline{Z}_t = Z_t - E[Z_t]$. Then by the C_r and Jensen inequalities and Assumptions A2 and A4-A6,

$$\begin{aligned}
\sum_{t=1}^{n} E\left|\overline{Z}_t\right|^{2+\delta} &\leq 2^{2+\delta}\left(\frac{h}{Mm^\lambda}\right)^{1+\delta/2}\sum_{t=1}^{n} E\left|u_t K_h(X_t - x)\right|^{2+\delta} \\
&= 2^{2+\delta}\left(Mm^\lambda h\right)^{-\delta/2}\frac{1}{m^\lambda}\sum_{a=-m+1}^{m}\int |u|^{2+\delta} K(z)^{2+\delta} f(u, x-\mu_\alpha+hz)\, dzdu \\
&\approx \left(Mm^\lambda h\right)^{-\delta/2} 2^{2+\delta}\int K(z)^{2+\delta} dz \int |u|^{2+\delta}\frac{1}{m^\lambda}\sum_{a=-m+1}^{m} f(u, x-\mu_\alpha)\, du \\
&= \left(Mm^\lambda h\right)^{-\delta/2} 2^{2+\delta}\int K(z)^{2+\delta} dz\left\{E|u|^{2+\delta} + o(1)\right\} \to 0
\end{aligned}$$

Then by the Liapounov CLT,

$$\Theta_{3n} - E(\Theta_{3n}) \to_d N\left(0, \sigma^2\nu_2(K)\right). \qquad (6.6)$$

Combining (6.2)-(6.6), we obtain (2.5) and the proof is complete.

6.2 Proof of Theorem 2.2

The proof is analogous to that of Theorem 2.1, so only the differences are sketched here. First, under A3(i), A7, and A8, the result in Lemma 6.1 changes to

$$\int u\left[d_m^{-1}\sum_{a=-m+1}^{m} f\left(u,x-\mu_\alpha\right)-f_u\left(u\right)\right]du=O\left(d_m^{-2}\right). \qquad (6.7)$$

To see this, noticing that $f\left(u,x-\frac{L}{2}\right)=f\left(u,x+\frac{L}{2}\right)=0$ since there is zero density outside the support by A7 and A8, we have

$$\begin{aligned}
&\frac{1}{d_m}\sum_{\alpha=1}^{2m} f\left(u,x-\mu_\alpha\right)-f_u\left(u\right)\\
=\ &\left\{\frac{1}{d_m}\sum_{a=1}^{2m}\frac{1}{2}\left[f\left(u,x+\frac{L}{2}-\frac{L\left(\alpha\text{-}1\right)}{2m\text{-}1}\right)+f\left(u,x+\frac{L}{2}-\frac{L\alpha}{2m\text{-}1}\right)\right]-\int_{-L/2}^{L/2} f\left(u,x-p\right)dp\right\}\\
&+\left\{\int_{-L/2}^{L/2} f\left(u,x-p\right)dp-\int_{\underline{u}}^{\overline{u}} f\left(u,u_x\right)du_x\right\}\\
\equiv\ & I_{1m}\left(u\right)+I_{2m}\left(u\right).
\end{aligned}$$

Since $x+\frac{L}{2}>\overline{u}$ and $x-\frac{L}{2}<\underline{u}$, we have $I_{2m}\left(u\right)=\int_{x-L/2}^{x+L/2} f\left(u,u_x\right)du_x-\int_{\underline{u}}^{\overline{u}} f\left(u,u_x\right)du_x=0$ by A7 and A8. By (6.1) and Assumption A3(i)

$$\begin{aligned}
\left(d_m\right)^2\int_{-\infty}^{\infty} uI_{1m}\left(u\right)du &= \int_{-\infty}^{\infty}\frac{u}{12d_m}\sum_{a=1}^{2m} f_2''\left(u,x-p_\alpha\right)du\\
&\to \frac{1}{12}\int_{-\infty}^{\infty} u\int_{-L/2}^{L/2} f_2''\left(u,x-p\right)dpdu=\frac{1}{12}\int_{-\infty}^{\infty}\int_{\underline{u}}^{\overline{u}} uf_2''\left(u,p\right)dpdu,
\end{aligned}$$

where p_α lies between μ_α and $\mu_{\alpha+1}$. Thus (6.7) holds.

Next, one can show that the results in Lemmas 6.2(i) and (iii)-(iv) hold with m^λ replaced by d_m. The result in Lemma 6.2(ii) now becomes

$$\begin{aligned}
E\Theta_{1n} &= \frac{h^2\mu_2\left(K\right)}{d_m}\sum_{a=-m+1}^{m} f_{u_x}'\left(x-\mu_\alpha\right)+o\left(h^2\right)=h^2\mu_2\left(K\right)\int_{-L/2}^{L/2} f_{u_x}'\left(x-p\right)dp+o\left(h^2\right)\\
&= h^2\mu_2\left(K\right)\int_{x-L/2}^{x+L/2} f_{u_x}'\left(p\right)dp+o\left(h^2\right)=h^2\mu_2\left(K\right)\left[f_{u_x}\left(\overline{u}\right)-f_{u_x}\left(\underline{u}\right)\right]+o\left(h^2\right),
\end{aligned}$$

where the dominant term vanishes if and only if $f_{u_x}\left(\overline{u}\right)=f_{u_x}\left(\underline{u}\right)$.

As to Θ_{3n} defined in the proof of Theorem 2.1, we have that for sufficiently large n

$$\begin{aligned}
E\Theta_{3n} &= \frac{\sqrt{Mm^{\lambda}h}}{d_m}\sum_{a=-m+1}^{m} E\left[u_t K_h\left(\mu_\alpha + u_{xt} - x\right)\right] \\
&= \sqrt{Mm^{\lambda}h}\int K(z)\left\{\int u\frac{1}{d_m}\sum_{a=-m+1}^{m} f\left(u, x-\mu_\alpha+hz\right)du\right\}dz \\
&= \sqrt{Mm^{\lambda}h}\int K(z)\left\{O\left(d_m^{-2}\right)+\int u f_u(u)\,du\right\}dz \\
&= \sqrt{Mm^{\lambda}h}O\left(d_m^{-2}\right) = o(1),
\end{aligned}$$

where we have used the fact if $x+\frac{L}{2} > \overline{u}$ and $x-\frac{L}{2} < \underline{u}$, then $x+hz+\frac{L}{2} \geq \overline{u}$ and $x+hz-\frac{L}{2} \leq \underline{u}$ for sufficiently large n and fixed z as $h \to 0$.

These results imply that the bias and variance calculations in the proof of Theorem 2.1 continue to hold with the change due to $E\Theta_{1n}$ and with m^{λ} replaced by d_m everywhere. In addition, the Liapounov condition is also satisfied. This completes the proof of the theorem.

6.3 Proof of Theorem 3.1

We first state and prove a lemma that is used in the proof of Theorem 3.1.

Lemma 6.3 *Let* $\Theta_{nj} = \frac{1}{TN^{\lambda}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it}-x\right)^j K_h\left(X_{it}-x\right)$ *for* $j = 0, 1, 2$. *Then*

(i) $E\Theta_{n0} = 1 + o(1)$,

(ii) $E\Theta_{n1} = o\left(h^2\right)$,

(iii) $E\Theta_{n2} = h^2\mu_2(K) + o\left(h^2\right)$,

(iv) $Var(\Theta_{nj}) = O\left(\frac{h^{2j}}{TN^{\lambda}h} + \frac{h^{2j}}{N^{\lambda}}\right)$ *for* $j = 0, 1, 2$.

Proof. Noting that $E\Theta_{nj} = \int_{\infty}^{\infty}\int_{-L}^{L} z^j K(z) f_{u_x}\left(x-\mu+hz\right)d\mu dz$, we have

$$\begin{aligned}
E\Theta_{n0} &\to \int_{\infty}^{\infty}\int_{\infty}^{\infty} K(z) f_{u_x}\left(x-\mu\right)d\mu dz = 1, \\
E\Theta_{n1} &= h^2\mu_2(K)\int_{-L}^{L} f'_{u_x}\left(x-\mu\right)d\mu + o\left(h^2\right) = h^2\mu_2(K)\int_{-\infty}^{\infty} f'_{u_x}(p)\,dp + o\left(h^2\right) = o\left(h^2\right), \\
E\Theta_{n2} &= h^2\mu_2(K)\int_{-L}^{L} f_{u_x}\left(x-\mu\right)d\mu + o\left(h^2\right) = h^2\mu_2(K) + o\left(h^2\right).
\end{aligned}$$

Now, by the Jensen inequality and a change of variables

$$\begin{aligned}
& \operatorname{Var}\left(\Theta_{nj}\right) \\
= & \frac{1}{T^2 N^{2\lambda}} \sum_{i=1}^{N} \operatorname{Var}\left(\sum_{t=1}^{T} \left(X_{it}-x\right)^j K_h\left(X_{it}-x\right)\right) \\
\leq & \frac{1}{T^2 N^{2\lambda}} \sum_{i=1}^{N} \sum_{t=1}^{T} E\left[\left(X_{it}-x\right)^{2j} K_h^2\left(X_{it}-x\right)\right] \\
& + \frac{1}{T^2 N^{2\lambda}} \sum_{i=1}^{N} \sum_{t=1}^{T} \sum_{s\neq t}^{T} E\left[\left(X_{it}-x\right)^j \left(X_{is}-x\right)^j K_h\left(X_{it}-x\right) K_h\left(X_{is}-x\right)\right] \\
= & \frac{h^{2j}}{TN^\lambda h} \int \int_{-L}^{L} z^{2j} K^2(z) f_{u_x}(x-\mu+hz)\, d\mu dz \\
& + \frac{(T-1) h^{2j}}{TN^\lambda} \int \int \int_{-L}^{L} z^j z'^j K(z) K\left(z'\right) f_{u_x}(x-\mu+hz) f_{u_x}\left(x-\mu+hz'\right) d\mu dz dz' \\
= & O\left(\frac{h^{2j}}{TN^\lambda h} + \frac{h^{2j}}{N^\lambda}\right).
\end{aligned}$$

■

As in the proof of Theorem 2.1, we consider the decomposition

$$\begin{aligned}
\widehat{g}(x) - g(x) = & \frac{\frac{1}{TN^\lambda} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{g\left(X_{it}\right) - g(x)\right\} K_h\left(X_{it}-x\right)}{\frac{1}{TN^\lambda} \sum_{i=1}^{N} \sum_{t=1}^{T} K_h\left(X_{it}-x\right)} \\
& + \frac{\frac{1}{TN^\lambda} \sum_{i=1}^{N} \sum_{t=1}^{T} u_{it} K_h\left(X_{it}-x\right)}{\frac{1}{TN^\lambda} \sum_{i=1}^{N} \sum_{t=1}^{T} K_h\left(X_{it}-x\right)}. \qquad (6.8)
\end{aligned}$$

Analogous to the proof of Theorem 2.1, we can apply Lemma 6.3, a Taylor expansion and the Chebyshev inequality to obtain

$$\frac{1}{TN^\lambda} \sum_{i=1}^{N} \sum_{t=1}^{T} K_h\left(X_{it}-x\right) = 1 + o_P(1), \qquad (6.9)$$

and

$$\frac{1}{TN^\lambda} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{g\left(X_{it}\right) - g(x)\right\} K_h\left(X_{it}-x\right) = h^2 \mu_2(K) \frac{g''(x)}{2} + o_P\left(h^2\right). \qquad (6.10)$$

Now, let $\Theta_{n3} = \frac{\sqrt{h}}{\sqrt{TN^\lambda}} \sum_{i=1}^{N} \sum_{t=1}^{T} z_{it}$ and $\overline{\Theta}_{n3} = \frac{\sqrt{h}}{\sqrt{TN^\lambda}} \sum_{i=1}^{N} \sum_{t=1}^{T} \overline{z}_{it}$, where $z_{it} = u_{it} K_h(X_{it} - x)$, and $\overline{z}_{it} = z_{it} - E\left(z_{it} | \mu_{xi}\right)$. We show that $\Theta_{n3} - \overline{\Theta}_{n3} = o_P(1)$ and $\overline{\Theta}_{n3} \to_d N\left(0, \sigma^2 \nu_2(K)\right)$. By the independence between μ_{xi} and $\left(u_{it}, u_{xit}\right)$, a change of variables, and the Fubini theorem, we have

$$\begin{aligned} E\left[\Theta_{n3}-\overline{\Theta}_{n3}\right] &= \frac{\sqrt{TN^{\lambda}h}}{N^{\lambda}} NE\left[u_{i1}K_h\left(\mu_{xi}+u_{xi1}-x\right)\right] \\ &= \sqrt{TN^{\lambda}h} N\int_{-L/2}^{L/2}\int\int uK\left(z\right)f\left(u,x-\mu+hz\right)dzdud\mu \\ &= \sqrt{TN^{\lambda}h}\int K\left(z\right)\left\{\int u\int_{-L/2}^{L/2} f\left(u,x-\mu+hz\right)d\mu du\right\}dz \\ &= \sqrt{TN^{\lambda}h} O\left(N^{-v(1-\lambda)}\right)=o\left(1\right), \end{aligned}$$

where we use the fact that

$$\begin{aligned} &\left|\int\int_{-L/2}^{L/2} uf\left(u,x-\mu+hz\right)d\mu du\right| \\ = &\left|\int\left\{\int_{-L/2}^{L/2} uf\left(u,x-\mu+hz\right)d\mu-\int_{-\infty}^{\infty} uf\left(u,x-\mu+hz\right)d\mu\right\}du\right| \\ \leq &\ O\left(L^{-v}\right)\int_{-\infty}^{\infty}\left|u\right|C_f\left(u\right)du=O\left(N^{-v(1-\lambda)}\right). \end{aligned}$$

Note that

$$\begin{aligned} \operatorname{Var}\left(\Theta_{n3}-\overline{\Theta}_{n3}\right) &= \frac{Th}{N^{\lambda}}\sum_{i=1}^{N}\operatorname{Var}\left(E\left[u_{i1}K_h\left(\mu_{xi}+u_{xi1}-x\right)\left|\mu_{xi}\right.\right]\right) \\ &= \frac{Th}{N^{\lambda}}\sum_{i=1}^{N}\operatorname{Var}\left(\int\int uK\left(z\right)f\left(u,x-\mu_{xi}+hz\right)dzdu\right) \\ &\leq Th\int_{-L/2}^{L/2}\left(\int\int uK\left(z\right)f\left(u,x-\mu+hz\right)dzdu\right)^2 d\mu \\ &= Th\int_{-L/2}^{L/2}\left(\int uf\left(u,x-\mu\right)du\right)^2 d\mu+O\left(Th^5\right) \\ &= Th\int_{-\infty}^{\infty}\left(\int uf\left(u,x-\mu\right)du\right)^2 d\mu+O\left(Th\left(L^{-v}+h^4\right)\right)=o\left(1\right). \end{aligned}$$

It follows that

$$\Theta_{n3}-\overline{\Theta}_{n3}=o_P\left(1\right). \tag{6.11}$$

Similarly, since (u_{it}, u_{xit}) are *iid* and are independent of μ_{xi},

$$\begin{aligned}
\operatorname{Var}\left(\overline{\Theta}_{n3}\right) &= \frac{h}{TN^{\lambda}}\sum_{i=1}^{N}\sum_{t=1}^{T}\operatorname{Var}\left(\overline{z}_{it}\right) \\
&= \frac{h}{N^{\lambda}}\sum_{i=1}^{N} E\left[u_{i1}^2 K_h^2\left(\mu_{xi}+u_{xi1}-x\right)\right] + O\left(h\right) \\
&= \int_{-L/2}^{L/2}\int\int u^2 K\left(z\right)^2 f\left(u, x-\mu+hz\right) dzdud\mu + o\left(1\right) \\
&= \nu_2\left(K\right)\int u^2 \int f\left(u, x-\mu\right) d\mu du + o\left(1\right) = \nu_2\left(K\right)\sigma^2 + o\left(1\right).
\end{aligned}$$

To show the asymptotic normality of $\overline{\Theta}_{n3}$, by the above variance calculation and the independence of $\{\mu_{xi}, u_{it}, u_{xit}\}$ across i, it suffices to check the Liapounov condition. Let $Z_{Ni} = \frac{\sqrt{h}}{\sqrt{TN^{\lambda}}}\sum_{t=1}^{T}\overline{z}_{it}$. Noting that conditional on μ_{xi}, $\overline{z}_{it}$ are independent across t, by the law of iterated expectations and the C_r inequality we have

$$\begin{aligned}
\sum_{i=1}^{N} E\left|Z_{Ni}\right|^4 &= \left(\frac{h}{TN^{\lambda}}\right)^2 \sum_{i=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{t_3=1}^{T}\sum_{t_4=1}^{T} E\left\{E\left[\overline{z}_{it_1}\overline{z}_{it_2}\overline{z}_{it_3}\overline{z}_{it_4}|\mu_{xi}\right]\right\} \\
&= \frac{h^2}{T^2N^{2\lambda}}\sum_{i=1}^{N}\left\{\sum_{t=1}^{T} E\left[E\left(\overline{z}_{it}^4|\mu_{xi}\right)\right] + 6\sum_{t=1}^{T}\sum_{s\neq t}^{T} E\left[E\left(\overline{z}_{it}^2|\mu_{xi}\right)E\left(\overline{z}_{is}^2|\mu_{xi}\right)\right]\right\} \\
&\leq \frac{8}{TN^{\lambda}h}\int_{-L/2}^{L/2}\int\int u^4 K^4\left(z\right) f\left(u, x-\mu+hz\right) dzdud\mu \\
&\quad + \frac{24\left(T-1\right)}{TN^{\lambda}}\int_{-L/2}^{L/2}\left(\int\int u^2 K^2\left(z\right) f\left(u, x-\mu+hz\right) dzdu\right)^2 d\mu \\
&= O\left(\frac{1}{TN^{\lambda}h} + \frac{1}{N^{\lambda}}\right) = o\left(1\right).
\end{aligned}$$

Then by the Liapounov CLT,

$$\overline{\Theta}_{n3} \rightarrow_d N\left(0, \sigma^2\nu_2\left(K\right)\right). \tag{6.12}$$

Combining (6.8)-(6.12), we obtain (3.3) and the proof is complete.

6.4 The stationary, strong-mixing case

This appendix proves the result in Theorem 2.1 under strong mixing conditions on the process $\{(u_t, u_{xt}), t \geq 1\}$. Note that we do not restrict the location shifts to occur sequentially in the time dimension, which allows for $t, t' \in A_{\alpha}$ even if t and t' are distant observations. We use C to denote a generic large positive constant whose value varies across lines. In addition to A2-A5, the following assumptions are used in the proof that follows.

Assumption B1. (i) *The process* $\{(u_t, u_{xt}), t \geq 1\}$ *is a stationary strong mixing process with mixing coefficients* $\alpha(j)$ *such that* $\sum_{j=1}^{\infty}\alpha(j)^{\delta/(2+\delta)} < \infty$, *where* δ *is defined in A2.*

(ii) *There exist a positive number* a *and a sequence of positive numbers* $a_n \to \infty$ *such that* $a_n h \log m = o(1)$, $\sum_{j \geq a_n}^{\infty} j^a \alpha(j)^{\delta/(2+\delta)} = o(1)$, *and* $a_n^a h^{(1+2\delta)/(2+\delta)} \to (0, \infty]$.

Assumption B2. (i) *For any* $t = 1, 2, \cdots$, *there exists a continuous function* $\overline{\sigma}^2(\cdot)$ *such that* $E\left(u_t^2 | X_t\right) \leq \overline{\sigma}^2(X_t)$ *a.s., and* $\int \overline{\sigma}^2(p) f_{u_x}(p)\, dp < \infty$.

(ii) *For any* $t, s = 1, 2, \cdots$, *the joint density function* $f_{ts}(\cdot, \cdot)$ *of* (X_t, X_s) *exists.*

(iii) *For any* $t, s = 1, 2, \cdots$, *there exist continuous functions* $\overline{\sigma}^2(\cdot, \cdot)$ *and* $\overline{f}(\cdot, \cdot)$ *such that* $|E(u_t u_s | X_t, X_s)| \leq \overline{\sigma}^2(X_t, X_s)$ *a.s.,* $f_{ts}(X_t, X_s) \leq \overline{f}(X_t, X_s)$ *a.s., and* $max_{-m \leq \alpha \leq m}$ $\displaystyle\int \overline{\sigma}^2(p, x - \mu_\alpha)\, \overline{f}(p, x - \mu_\alpha)\, dp = O(\log m)$.

(iv) $\int \left\{ \int |u|^{2+\delta} f(u, u_x)\, du \right\}^{2/(2+\delta)} du_x < \infty$.

Assumption B3. *There exist a positive number* $b \in [1/2, 1)$ *and a sequence of positive integers* s_n *such that as* $(M, m) \to \infty$, $s_n \to \infty$, $s_n = o(M^b m^{(b+1)\lambda - 1} h^b)$, $n\alpha(s_n) / \left(Mm^\lambda h\right)^b = o(1)$, *and* $\left(Mm^\lambda h\right)^b h \to \infty$.

Assumption B4. *There exists* $\delta^* > 2 + \delta$ *such that* $E[|u_t|^{\delta^*} | X_t] = \sigma^{\delta^*}(X_t)$ *a.s. and* $sup_x \sigma^{\delta^*}(x) f_{u_x}(x) < \infty$.

Assumption B5. $\int |zK(z)|^{2+\delta} dz < \infty$.

Assumption B6. *As* $(M, m) \to \infty$, $Mm^\lambda h \to \infty$, $Mm^\lambda h^5 \to c \in [0, \infty)$, $Mm^\lambda h(m^{-4\lambda} + m^{-2v(1-\lambda)}) \to 0$, $Mm^\lambda h^{2(1+\delta)/(2+\delta)} \to \infty$, *and* $M^{(b-1)\delta/2} m^{1-\lambda+\lambda\delta(b-1)/2} \times h^{(2+\delta)/\delta^* - 1 + (b-1)\delta/2} = O(1)$.

Note that given A5, Assumption B5 is redundant if $\delta \leq 2$. Recall that $\Theta_{jn} = \frac{1}{Mm^\lambda} \sum_{t=1}^n (X_t - x)^j K_h(X_t - x)$ for $j = 0, 1, 2$. We first prove a lemma that is used in the main proof.

Lemma 6.4 *Var*$(\Theta_{jn}) = h^{2j-1}\left(Mm^\lambda\right)^{-1} \int z^{2j} K^2(z)\, dz + O(\left(Mm^\lambda h^{2(1+\delta)/(2+\delta)}\right)^{-1} h^{2j}) + o(\left(Mm^\lambda\right)^{-1} h^{2j-1})$ *for* $j = 0, 1, 2$.

Proof. Write $\text{Var}(\Theta_{jn}) = \frac{1}{m^{2\lambda} M^2} \sum_{t=1}^n \text{Var}((X_t - x)^j K_{tx}) + \frac{1}{m^{2\lambda} M^2} \sum_{t=1}^n \sum_{s=1, s \neq t}^n \text{Cov}((X_t - x)^j K_{tx}, (X_s - x)^j K_{sx}) \equiv I_{n1} + I_{n2}$. I_{n1} was studied in Lemma 6.2(iv). By the Davydov inequality (e.g., Hall and Heyde (1980, p. 278)) and Assumptions B1(i) and B5,

$$\begin{aligned}
I_{n2} &= \frac{1}{M^2 m^{2\lambda}} \sum_{t=1}^{n} \sum_{s=1, s\neq t}^{n} \operatorname{Cov}\Big((X_t - x)^j K_{tx}, (X_s - x)^j K_{sx} \Big) \\
&\leq \frac{8h^{2j}}{M^2 m^{2\lambda} h^{\frac{2(1+\delta)}{2+\delta}}} \sum_{a=-m+1}^{m} \sum_{a'=-m+1}^{m} \sum_{t\in A_\alpha} \sum_{s\in A_{\alpha'}} \left\{ \int |zK(z)|^{2+\delta} f_{u_x}(x - \mu_\alpha + hz)\, dz \right\}^{1/(2+\delta)} \\
&\quad \times \left\{ \int |zK(z)|^{2+\delta} f_{u_x}(x - \mu_{\alpha'} + hz)\, dz \right\}^{1/(2+\delta)} \alpha(|s-t|)^{\delta/(2+\delta)} \\
&\leq \frac{8\left[\nu_{2+\delta}(K)\right]^{2/(2+\delta)} h^{2j}}{M m^\lambda h^{2(1+\delta)/(2+\delta)}} \frac{1}{m^\lambda} \sum_{a=-m+1}^{m} f_{u_x}(x-\mu_\alpha)^{2/(2+\delta)} \sum_{s=1}^{\infty} \alpha(s)^{\delta/(2+\delta)} \{1 + o(1)\} \\
&\leq \frac{Ch^{2j}}{M m^\lambda h^{2(1+\delta)/(2+\delta)}} \int f_{u_x}(x-p)^{2/(2+\delta)}\, dp \{1 + o(1)\} = O\left(\frac{h^{2j}}{M m^\lambda h^{2(1+\delta)/(2+\delta)}} \right),
\end{aligned}$$

where $\nu_{2+\delta}(K) = \int |zK(z)|^{2+\delta} dz$. This together with Lemma 6.2(iv) finishes the proof of the lemma. ■

To prove the theorem under the strong mixing condition, we also use the bias-variance decomposition in (6.2), and analyze each of the quantities on the right hand side of (6.2). By Lemmas 6.2(i)-(iii) and 6.4, the Chebyshev inequality, and Assumption B6,

$$\frac{1}{Mm^\lambda} \sum_{t=1}^{n} K_h(X_t - x) = 1 + O_P((Mm^\lambda h)^{-1/2} + (Mm^\lambda h^{2(1+\delta)/(2+\delta)})^{-1/2}) = 1 + o_P(1), \quad (6.13)$$

and

$$\begin{aligned}
& \frac{1}{m^\lambda M} \sum_{t=1}^{n} \{g(X_t) - g(x)\} K_h(X_t - x) \\
&= \frac{h^2 \mu_2(K) g''(x)}{2} + O_P\left(\left(\frac{h}{Mm^\lambda} + \frac{Ch^2}{Mm^\lambda h^{2(1+\delta)/(2+\delta)}} \right)^{1/2} \right) + o_P\left(h^2\right) \\
&= \frac{h^2 \mu_2(K) g''(x)}{2} + o_P\left(h^2\right). \qquad (6.14)
\end{aligned}$$

Recall $\Theta_{3n} \equiv \frac{\sqrt{h}}{\sqrt{m^\lambda M}} \sum_{t=1}^{n} u_t K_h(X_t - x)$. The calculation of $E\Theta_{3n}$ in (6.5) continues to hold. It remains to show that

$$\Theta_{3n} - E\Theta_{3n} \to_d N\left(0, \sigma^2 \nu_2(K)\right). \qquad (6.15)$$

We first calculate the asymptotic variance of Θ_{3n} and then prove the asymptotic normality of $\Theta_{3n} - E\Theta_{3n}$. Write $\operatorname{Var}(\Theta_{3n}) = \frac{h}{m^{2\lambda}M^2} \sum_{t=1}^{n} \operatorname{Var}(u_t K_{tx}) + \frac{2h}{m^{2\lambda}M^2} \sum_{1\leq s<t\leq n} \operatorname{Cov}$

$(u_t K_{tx}, u_s K_{sx}) \equiv I_{n3} + 2I_{n4}$. As in the proof of Theorem 2.1 in the independent case,

$$\begin{aligned} I_{n3} &= \frac{h}{m^{\lambda}} \sum_{a=-m+1}^{m} \mathrm{Var}\left(u_t K_h\left(\mu_\alpha + u_{xt} - x\right)\right) \\ &= \frac{h}{m^{\lambda}} \sum_{a=-m+1}^{m} E\left(u_t^2 K_h^2\left(\mu_\alpha + u_{xt} - x\right)\right) + o\left(1\right) \\ &= \sigma^2 \int K^2\left(z\right) dz + o\left(1\right). \end{aligned}$$

To obtain an upper bound for I_{n4}, we split it in two as follows:

$$I_{n4} = \frac{h}{Mm^{\lambda}} \sum_{s=1}^{n-1} \sum_{t=s+1}^{s+a_n-1} \mathrm{Cov}\left(u_t K_{tx}, u_s K_{sx}\right) + \frac{h}{Mm^{\lambda}} \sum_{s=1}^{n-a_n} \sum_{t=s+a_n}^{n} \mathrm{Cov}\left(u_t K_{tx}, u_s K_{sx}\right) \equiv I_{n4a} + I_{n4b},$$

where a_n is specified in Assumption B1(ii). By Assumptions B1(ii) and B2, and the dominated convergence theorem

$$\begin{aligned} I_{n4a} &\leq \frac{h}{Mm^{\lambda}} \sum_{s=1}^{n-1} \sum_{t=s+1}^{s+a_n-1} \left|\mathrm{Cov}\left(u_t K_{tx}, u_s K_{sx}\right)\right| \\ &\leq \frac{Ch}{Mm^{\lambda}} \sum_{s=1}^{n-1} \sum_{t=s+1}^{s+a_n-1} \int\int \overline{\sigma}^2\left(x-\mu_\alpha+hz_t, x-\mu_{\alpha'}+hz_s\right) K\left(z_t\right) K\left(z_s\right) \\ &\quad \times \overline{f}\left(x-\mu_\alpha+hz_t, x-\mu_{\alpha'}+hz_s\right) dz_t dz_s 1\left\{t \in A_\alpha\right\} 1\left\{s \in A_{\alpha'}\right\} \\ &\leq \frac{Cha_n}{m^{\lambda}} \max_{\alpha'} \sum_{\alpha=-m+1}^{m} \int\int \overline{\sigma}^2\left(x-\mu_\alpha+hz_t, x-\mu_{\alpha'}+hz_s\right) K\left(z_t\right) K\left(z_s\right) \\ &\quad \times \overline{f}\left(x-\mu_\alpha+hz_t, x-\mu_{\alpha'}+hz_s\right) dz_t dz_s \\ &\to Cha_n \max_{\alpha'} \int \overline{\sigma}^2\left(p, x-\mu_{\alpha'}\right) \overline{f}\left(p, x-\mu_{\alpha'}\right) dp \\ &= O\left(ha_n \log m\right) = o\left(1\right). \end{aligned}$$

For I_{n4b}, by applying the Davydov inequality and Assumptions B1-B2, we have

$$\begin{aligned} I_{n4b} &\leq \frac{h}{Mm^{\lambda}} \sum_{s=1}^{n-a_n} \sum_{t=s+a_n}^{n} \left|\mathrm{Cov}\left(u_t K_{tx}, u_s K_{sx}\right)\right| \\ &\leq \frac{1}{Mm^{\lambda} h^{\frac{1+2\delta}{2+\delta}}} \sum_{a=-m+1}^{m} \sum_{a'=-m+1}^{m} \sum_{s \in A_\alpha} \sum_{t=s+a_n, t \in A_{\alpha'}} \left\{\int\int \left|uK\left(z\right)\right|^{2+\delta} f\left(u, x-\mu_\alpha+hz\right) dz du\right\}^{1/(2+\delta)} \\ &\quad \times \left\{\int\int \left|uK\left(z\right)\right|^{2+\delta} f\left(x-\mu_{\alpha'}+hz\right) dz du\right\}^{1/(2+\delta)} \alpha\left(t-s\right)^{\delta/(2+\delta)} \end{aligned}$$

$$
\begin{aligned}
&\leq \left\{\int K(z)^{2+\delta} dz\right\}^{2/(2+\delta)} h^{-\frac{1+2\delta}{2+\delta}} a_n^{-a} \frac{1}{m^\lambda} \sum_{a=-m+1}^{m} \left\{\int |u|^{2+\delta} f(u, x-\mu_\alpha) du\right\}^{2/(2+\delta)} \\
&\quad \times \sum_{k\geq a_n}^{n} k^a \alpha(k)^{\delta/(2+\delta)} \{1+o(1)\} \\
&= o\left(h^{-\frac{1+2\delta}{2+\delta}} a_n^{-a}\right) = o(1).
\end{aligned}
$$

Consequently, we have

$$
\operatorname{Var}(\Theta_{3n}) = \nu_2(K)\sigma^2 + o(1). \tag{6.16}
$$

To show the asymptotic normality of $\Theta_{3n} - E(\Theta_{3n})$, we apply the standard Doob large-block and small-block technique. We partition $\{1, 2, \cdots, n\}$ into $2q_n+1$ subsets with large-block of size $r = r_n$ and small-block of size $s = s_n$. Set $q = q_n = \lfloor n/(r_n + s_n) \rfloor$. Define the random variables, for $0 \leq j \leq q-1$

$$
\eta_j = \sum_{i=j(r+s)+1}^{j(r+s)+r} \xi_i, \ \varsigma_j = \sum_{i=j(r+s)+r+1}^{(j+1)(r+s)} \xi_i, \text{ and } \eta_q = \sum_{i=q(r+s)+1}^{n} \xi_i,
$$

where $\xi_i = \sqrt{nh/(Mm^\lambda)}\{u_i K_h(X_i - x) - E[u_i K_h(X_i - x)]\}$. Then

$$
\Theta_{3n} - E(\Theta_{3n}) = \frac{1}{\sqrt{n}} \left\{ \sum_{j=0}^{q-1} \eta_j + \sum_{j=0}^{q-1} \varsigma_j + \eta_q \right\} \equiv \frac{1}{\sqrt{n}} \{Q_{n1} + Q_{n2} + Q_{n3}\}.
$$

Assumption B3 implies that there exists a sequence of positive constants $\iota_n \to \infty$ such that

$$
\iota_n s_n = o(M^b m^{(b+1)\lambda - 1} h^b), \ \iota_n n \alpha(s_n)/(Mm^\lambda h)^b = o(1), \text{ and } \left(Mm^\lambda h\right)^b h/\iota_n \to \infty,
$$

where $b \in [1/2, 1)$. We will choose the large-block size $r_n = \lfloor \left(Mm^\lambda h\right)^b / \iota_n \rfloor$ and small block size s_n. By construction, we have, as $(M, m) \to \infty$,

$$
s_n/r_n \to 0, \ r_n/(Mm^\lambda h) \to 0, \ q_n s_n/(Mm^\lambda) \to 0, \ (n/r_n)\alpha(s_n) \to 0, \text{ and } a_n/r_n \to 0.
$$

We will show that, as $(M, m) \to \infty$ and $h \to 0$, (i) $\frac{1}{n}E[Q_{n2}]^2 \to 0$, (ii) $\frac{1}{n}E[Q_{n3}]^2 \to 0$, (iii) $|E\exp(itQ_{n1}) - \Pi_{j=0}^{q-1} E[\exp(it\eta_j)]| \to 0$, (iv) $\frac{1}{n}\sum_{j=0}^{q-1} E(\eta_j^2) \to \nu_2(K)\sigma^2$, and (v) $\frac{1}{n}\sum_{j=0}^{q-1} E|\eta_j|^2 \times 1\{|\eta_j| \geq \epsilon\nu_2(K)\sigma^2\sqrt{n}\} \to 0$ for any $\epsilon > 0$. We will prove each of these results in turn.

To show (i), write

$$
\frac{1}{n}E[Q_{n2}]^2 = \frac{1}{n}E\left[\sum_{j=0}^{q-1} \varsigma_j\right]^2 = \frac{1}{n}\sum_{j=0}^{q-1} \operatorname{Var}(\varsigma_j) + \frac{1}{n}\sum_{j=0}^{q-1} \sum_{j' \neq j, j'=0}^{q-1} \operatorname{Cov}(\varsigma_j, \varsigma_{j'}) \equiv I_{n5} + I_{n6}.
$$

$$
\begin{aligned}
I_{n5} &= \frac{h}{Mm^{\lambda}} \sum_{j=0}^{q-1} \sum_{t=j(r+s)+r+1}^{(j+1)(r+s)} \operatorname{Var}\left(u_t K_h\left(X_t - x\right)\right) \\
&\quad + \frac{h}{Mm^{\lambda}} \sum_{j=0}^{q-1} \sum_{t=j(r+s)+r+1}^{(j+1)(r+s)} \sum_{t'=j(r+s)+r+1, t' \neq t}^{(j+1)(r+s)} \operatorname{Cov}\left(u_t K_h\left(X_t - x\right), u_{t'} K_h\left(X_{t'} - x\right)\right) \\
&\equiv I_{n5}a + I_{n5b}.
\end{aligned}
$$

The first term is $O\left(q_n s_n / \left(Mm^{\lambda}\right)\right) = o(1)$. Analogous to the proof of I_{n4}, we can show the second term is $O\left(ha_n \log m\right) + o\left(h^{-(1+2\delta)/(2+\delta)} a_n^{-a}\right) = o(1)$. Hence $I_{n5} = o(1)$. Now let $l_j = j\left(r_n + s_n\right)$. Then $l_j - l_i \geq r_n$ for all $j > i$. Let $e_t = u_t K_h\left(X_t - x\right)$. It follows that

$$
\begin{aligned}
I_{n6} &= \frac{2h}{Mm^{\lambda}} \sum_{0 \leq j < j' \leq q-1} \sum_{t=l_j+r+1}^{l_{j+1}} \sum_{t'=l_{j'}+r+1}^{l_{j'+1}} \operatorname{Cov}\left(u_t K_h\left(X_t - x\right), u_{t'} K_h\left(X_{t'} - x\right)\right) \\
&\leq \frac{2h}{Mm^{\lambda}} \sum_{0 \leq i < j \leq q-1} \sum_{j_1=1}^{s_n} \sum_{j_2=1}^{s_n} \left|\operatorname{Cov}\left(e_{l_i + r_n + j_1}, e_{l_j + r_n + j_2}\right)\right| \\
&\leq \frac{2h}{Mm^{\lambda}} \sum_{j_1=1}^{n-r_n} \sum_{j_2=j_1+r_n}^{n} \left|\operatorname{Cov}\left(e_{j_1}, e_{j_2}\right)\right| \\
&\leq \frac{Cr_n^{-a}}{h^{(1+2\delta)/(2+\delta)}} \frac{1}{m^{\lambda}} \sum_{\alpha=-m+1}^{m} \left\{\int |u|^{2+\delta} f\left(u, x - \mu_{\alpha}\right) du\right\}^{2/(2+\delta)} \sum_{j=r_n}^{\infty} j^a \alpha(j)^{\delta/(2+\delta)} \\
&= o\left(r_n^{-a} h^{-(1+2\delta)/(2+\delta)}\right) = o(1),
\end{aligned}
$$

where we have used the fact that $\sum_{j=r_n}^{\infty} j^a \alpha(j)^{\delta/(2+\delta)} = o(1)$ and $a_n / r_n = o(1)$.

(ii) It is easy to see that

$$
\begin{aligned}
\frac{1}{n} E\left[Q_{n3}\right]^2 &= \frac{h}{Mm^{\lambda}} \sum_{i=q(r+s)+1}^{n} \sum_{j=q(r+s)+1}^{n} \operatorname{Cov}\left(u_i K_h\left(X_i - x\right), u_j K_h\left(X_j - x\right)\right) \\
&= O\left(\frac{n - q_n\left(r_n + s_n\right)}{Mm^{\lambda}}\right) = O\left(\frac{r_n + s_n}{Mm^{\lambda}}\right) = o(1).
\end{aligned}
$$

(iii) Noting that stationarity is not necessary for Lemma 1.1 of Volkonskii and Rozanov (1959) (see also Li and Racine (2007, p.571)), we can apply this lemma to obtain

$$
\left|E \exp\left(itQ_{n1}\right) - \Pi_{j=0}^{q-1} E\left[\exp\left(it\eta_j\right)\right]\right| \leq 16\left(q_n - 1\right) \alpha\left(s_n\right) \leq \frac{16n}{r_n} \alpha\left(s_n\right) = o(1).
$$

(iv) Noting that

$$
\operatorname{Var}(\Theta_{3n}) = \sum_{s=1}^{3} \operatorname{Var}\left(Q_{ns}\right) + 2 \sum_{1 \leq s < t \leq 3} \operatorname{Cov}\left(Q_{ns}, Q_{nt}\right),
$$

by (6.16), (i), (ii) and the Cauchy-Schwarz inequality, we can deduce $\frac{1}{n}\mathrm{Var}(Q_{n1}) = O(1)$, which further implies that

$$\frac{1}{n}\mathrm{Var}\left(Q_{n1}\right) = \frac{1}{n}\mathrm{Var}\left(\Theta_{3n}\right) + o\left(1\right) = \nu_2\left(K\right)\sigma^2 + o\left(1\right).$$

Next,

$$\frac{1}{n}\sum_{j=0}^{q-1} E\left(\eta_j^2\right) = \frac{1}{n}\mathrm{Var}\left(Q_{n1}\right) - \frac{2}{n}\sum_{0\leq j<j'\leq q-1}\sum_{t=j(r+s)+1}^{j(r+s)+r}\sum_{t'=j'(r+s)+1}^{j'(r+s)+r}\mathrm{Cov}\left(\xi_t,\xi_{t'}\right) + o\left(1\right).$$

Following the proof of I_{n4}, one shows that the second term in the last expression is $o(1)$. It follows that $\frac{1}{n}\sum_{j=0}^{q-1}E(\eta_j^2) = \nu_2(K)\sigma^2 + o(1)$.

(v) By the Hölder and Chebyshev inequalities, Theorem 4.1 of Shao and Yu (1996), and Assumption B4, we have

$$\begin{aligned}
& \frac{1}{n}\sum_{j=0}^{q-1} E\left|\eta_j\right|^2 1\left\{\left|\eta_j\right| \geq \epsilon\nu_2\left(K\right)\sigma^2\sqrt{n}\right\} \\
\leq\ & \frac{1}{n}\sum_{j=0}^{q-1}\left\{E\left|\eta_j\right|^{2+\delta}\right\}^{2/(2+\delta)} P\left(\left|\eta_j\right| \geq \epsilon\nu_2\left(K\right)\sigma^2\sqrt{n}\right)^{\delta/(2+\delta)} \leq Cn^{-(1+\delta/2)}\sum_{j=0}^{q-1}E\left|\eta_j\right|^{2+\delta} \\
\leq\ & Cn^{-(1+\delta/2)}r_n^{1+\delta/2}\sum_{j=0}^{q-1}\max_{j(r+s)+1\leq i\leq j(r+s)+r}\left\{E\left|\xi_i\right|^{\delta^*}\right\}^{(2+\delta)/\delta^*}.
\end{aligned}$$

By the C_r inequality and Assumption B4,

$$\begin{aligned}
& E\left[\left|\xi_i\right|^{\delta^*}|i\in A_\alpha\right] \\
=\ & 2^{\delta^*/2-1}\left(\frac{nh}{Mm^\lambda}\right)^{\delta^*/2} E\left[\left|u_iK_h\left(X_i-x\right)\right|^{\delta^*}|i\in A_\alpha\right] \\
=\ & 2^{\delta^*/2-1}\left(\frac{n}{Mm^\lambda}\right)^{\delta^*/2} h^{1-\delta^*/2}\int\left|uK\left(z\right)\right|^{\delta^*} f\left(u,x-\mu_\alpha+hz\right)dzdu \\
\leq\ & C\left(\frac{n}{Mm^\lambda}\right)^{\delta^*/2} h^{1-\delta^*/2}\int\left|u\right|^{\delta^*} f\left(u,x-\mu_\alpha\right)du = O\left(\left(\frac{n}{Mm^\lambda}\right)^{\delta^*/2} h^{1-\delta^*/2}\right).
\end{aligned}$$

It follows that

$$\begin{aligned}
& \frac{1}{n}\sum_{j=0}^{q-1} E\left|\eta_j\right|^2 1\left\{\left|\eta_j\right| \geq \epsilon\nu_2\left(K\right)\sigma^2\sqrt{n}\right\} \\
\leq\ & Cn^{-(1+\delta/2)}r_n^{1+\delta/2}\sum_{j=0}^{q-1}\left\{\left(\frac{n}{Mm^\lambda}\right)^{\delta^*/2} h^{1-\delta^*/2}\right\}^{(2+\delta)/\delta^*} \\
=\ & O\left(n^{-(1+\delta/2)}r_n^{1+\delta/2}q_n\left(\frac{n}{Mm^\lambda}\right)^{(2+\delta)/2} h^{(2+\delta)/\delta^*-(2+\delta)/2}\right) \\
=\ & O\left(r_n^{1+\delta/2}q_n\left(Mm^\lambda\right)^{-(2+\delta)/2} h^{(2+\delta)/\delta^*-(2+\delta)/2}\right) \\
=\ & O\left(\iota_n^{-\delta/2}M^{(b-1)\delta/2}m^{1-\lambda+\lambda\delta(b-1)/2}h^{(2+\delta)/\delta^*-1+(b-1)\delta/2}\right) = o\left(1\right).
\end{aligned}$$

References

Blundell, R. W. and J. L. Powell (2003). "Endogeneity in semiparametric and nonparametric regression models." In *Advances in Economics and Econometrics: Theory and Applications.* (M. Dewatripont, L. P. Hansen and S. J. Turnovsky, eds.), 2, 312 - 357. Cambridge University Press.

Carrasco, M., Florens, J-P., and E. Renault (2007). "Linear inverse problems in structural econometrics: Estimation based on spectral decomposition and regularization." *Handbook of Econometrics*, Vol. 6, Part 2, 5633 - 5751.

Chen, X. (2007). "Large sample sieve estimation of semi-nonparametric models." Ch. 76 in J. J. Heckman and E. E. Leamer (eds.) *Handbook of Econometrics,* Vol. 6B, North Holland.

Darolles, S., J-P Florens, E. Renault, (2009). "Nonparametric instrumental regression." Working paper, Toulouse School of Economics.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*, vol. 66 of Monographs on Statistics and Applied Probability. Chapman and Hall.

Florens, J.–P. (2003). "Inverse problems and structural econometrics: The example of instrumental variables." In *Advances in Economics and Econometrics: Theory and Applications.* (M. Dewatripont, L. P. Hansen and S. J. Turnovsky, eds.), 2, 284-311. Cambridge University Press.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Applications.* Academic Precess, San Diego.

Hall, P. and Horowitz, J. L. (2005). "Nonparametric methods for inference in the presence of instrumental variables." *Annals of Statistics* 33, 2904-2929.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton: Princeton University Press.

Li, Q. and Racine, J. S. (2007). *Nonparametric Econometrics: Theory and Practice.* Princeton University Press, Princeton and Oxford.

Newey, W. K. and J. L. Powell (2003). "Instrumental variable estimation of nonparametric models." *Econometrica* 71, 1565 - 1578.

Newey, W. K., J. L. Powell, and F. Vella (1999). "Nonparametric estimation of triangular simultaneous equations models." *Econometrica* 67, 565 - 604.

Shao, Q. and Yu, H. (1996). "Weak convergence for weighted empirical processes of dependent sequences." *Annals of Probability* 24, 2098-2127.

Simoni, A. (2009). "Bayesian nonparametric estimation of asset price functionals." Working paper, Toulouse School of Economics.

Stoer, J. and Bulirsch R. (1993). *Introduction to Numerical Analysis*. Springer-Verlag, New York.

Volkonskii, V. A. and Rozanov Y. A. (1959). "Some limit theorems for random functions I." *Theory of Probability and Its Applications* 4, 178-197.

Wang, Q. and Phillips, P. C. B. (2009a). "Asymptotic theory for local time density estimation and nonparametric cointegrating regression." Forthcoming in *Econometric Theory*.

Wang, Q. and Phillips, P. C. B. (2009b). "Structural nonparametric cointegrating regression." Forthcoming in *Econometrica*.

Wang, Q. and Phillips, P. C. B. (2009c). "Asymptotic theory for zero energy density estimation with nonparametric regression applications." Cowles Foundation Discussion Paper 1687. Yale University.