NON-PARAMETRIC REGRESSION UNDER LOCATION SHIFTS

BY

Peter C. B. Phillips and Liangjun Su

COWLES FOUNDATION PAPER NO. 1346

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

2012

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

The Econometrics Journal

Econometrics Journal (2011), volume **14**, pp. 457–486.
doi: 10.1111/j.1368-423X.2011.00344.x

Non-parametric regression under location shifts

PETER C. B. PHILLIPS[†,‡,§,¶] AND LIANGJUN SU[§§]

[†]*Cowles Foundation, P.O. Box 208281, Yale University, New Haven, CT 06520-8281, USA.*
[‡]*Department of Economics, Owen G. Glenn Building, The University of Auckland, Private Bag 92019, Auckland 1142, New Zealand.*
[§]*School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903.*
[¶]*Department of Economics, School of Social Sciences, University of Southampton, Southampton SO17 1BJ, UK.*
E-mail: peter.phillips@yale.edu

[§§]*School of Economics, Singapore Management University, 90 Stamford Road, Singapore 178903.*
E-mail: ljsu@smu.edu.sg

First version received: June 2009; final version accepted: February 2011

Summary Recent work by Wang and Phillips (2009b, 2011) has shown that ill-posed inverse problems do not arise in non-stationary non-parametric regression and there is no need for non-parametric instrumental variable estimation. Instead, simple Nadaraya–Watson non-parametric estimation of a cointegrating regression equation is consistent irrespective of the endogeneity in the regressor. The present paper shows that some closely related results apply in the case of structural non-parametric regression with independent data when there are continuous location shifts in the regressor. Some interesting cases are discovered where non-parametric regression is consistent, whereas parametric regression is inconsistent even when the true regression functional form is known and used in regression. This appears to be a paradox, as knowing the true functional form should not in general be detrimental in regression. The paradox arises because additional correct information is not necessarily advantageous when information is incomplete. In this case, endogeneity in the regressor introduces bias when the true functional form is known, but interestingly does not do so in local non-parametric regression. We propose two new consistent estimators for the parametric regression, which address the endogeneity in the regressor by means of spatial bounding and bias correction using non-parametric estimation.

Keywords: *Bias-correction, Endogeneity, Kernel regression, Location shift, Non-parametric IV, Non-stationarity, Paradox, Spatial $\mathcal{L}_2$ regression, Structural estimation.*

1. INTRODUCTION AND MOTIVATION

Much recent interest in non-parametric 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

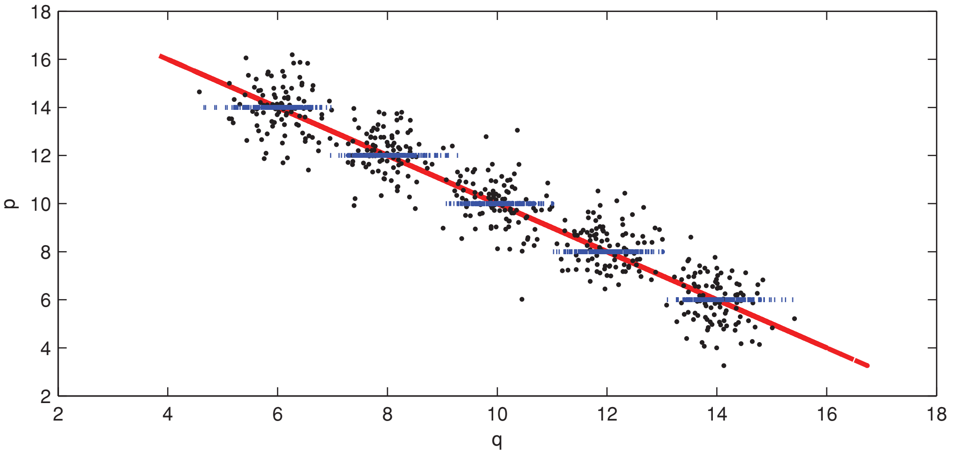

Figure 1. Demand and supply curves for five locations.

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 inverse problems. Research on this type of problem in econometrics has been under way 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.

To facilitate the use of standard functional analysis methods, it has become conventional in econometric treatments to restrict the 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 (such as unbounded densities) that are not innocuous and they can omit factors that are relevant in microeconometric modelling. 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 non-parametric regression, completely removing the ill-posed inverse problem and facilitating consistent estimation by conventional non-parametric 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 supply shifters are precisely the ones that are described in textbook treatments of identification

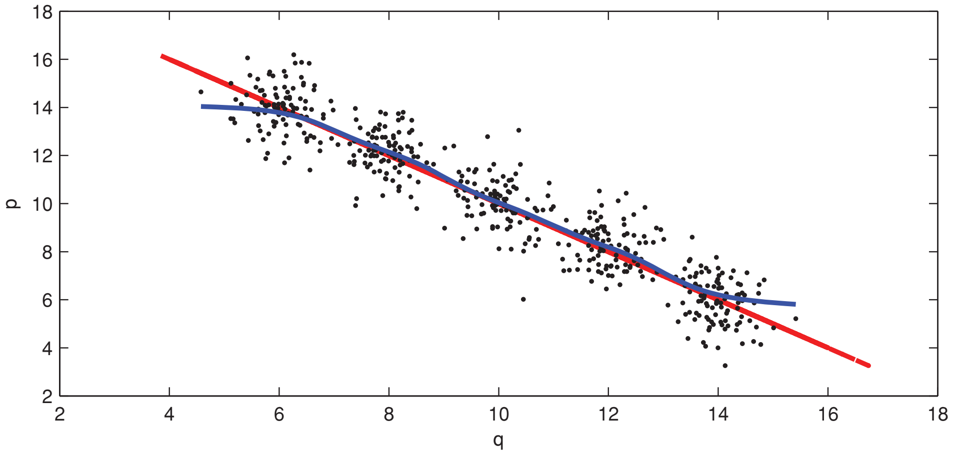

Figure 2. Nadaraya–Watson non-parametric regression demand curve estimate.

in simultaneous equations models of supply and demand. As the supply curve moves around it traces out a sequence of equilibria at each location. The key idea is illustrated in Figure 1,which provides a sample plot of data generated from the following location-specific linear Marshallian stochastic demand/supply system:[1]

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

$$\text{Supply:} \quad p_i = c + \sum_{\alpha=1}^{K} \mu_\alpha 1\{i \in A_\alpha\} + u_{pi}, \tag{1.2}$$

$$u_i = \{\theta u_{pi} + \epsilon_{qi}\}/(1+\theta^2)^{1/2}, \tag{1.3}$$

where $1\{\cdot\}$ is the indicator function, α signifies a particular market location, the A_α's are disjoint clusters of individuals associated with a mean price $c + \mu_\alpha$, the errors $(u_{pi}, \epsilon_{qi}) \equiv \text{i.i.d.}N(0, I_2)$, p_i and q_i are prices and quantities, respectively, and a, b, c and θ are parameters. In general, the locational equilibrium at location α is disturbed by errors and the data tend to cluster around each locational equilibrium point. Figure 1 illustrates this locational clustering phenomenon with a typical data set for $K = 5$ locations corresponding to $\mu_\alpha \in \{-4, -2, 0, 2, 4\}$ with $\theta = 2$, $a = 20$, $b = -1$, $c = 10$ and $M = 100$ observations for each α. Along the demand curve we observe clusters of points around price levels $\{6, 8, 10, 12, 14\}$ 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. Figure 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 bandwidth chosen by the Silverman rule of thumb. As is apparent, within the interior support of the sample data for (q_i, p_i) the data trace out the demand curve closely and the regression line is very well

[1] Alternatively, as a referee suggested, the model can be presented as $q_{\alpha k} = a + bp_{\alpha k} + u_{\alpha k}$, $p_{\alpha k} = c + \mu_\alpha + u_{p,\alpha k}$, $u_{\alpha k} = \{\theta u_{p,\alpha k} + \epsilon_{q,\alpha k}\}/(1+\theta^2)^{1/2}$, where α is a market index and k is a product index. Here we have K markets.

fitted by the non-parametric curve irrespective of the endogeneity. 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 first central finding of this paper. In spite of endogeneity in a regressor and provided the regressor has enough variation, simple non-parametric 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 shifts serve the role of a form of cross-section non-stationarity and in this sense the resulting consistency of non-parametric regression is analogous to that achieved in the case of a (near) unit root regressor in a time series setting. 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, 2011) showed that non-parametric 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.

As Phillips and Su (2009) remark, continuous location shifts are idealized in a cross-section setting because practical empirical examples with cross-section data where continuous location shifts may occur are likely to be uncommon in economics. However, as the above illustration shows, even a small number of discrete location shifts is sufficient to give non-parametric regression a significant advantage in dealing with endogeneity over a subset of the support. More importantly, a natural and realistic application of our results is a structural non-parametric panel data model:

$$y_{it} = g(X_{it}) + u_{it}, \quad X_{it} = \mu_{xi} + u_{xit}, \quad E(u_{it} \mid X_{it}) \neq 0,$$

where $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$, u_{it} and u_{xit} are error terms, and the individual effects μ_{xi} are assumed to be independent over i and uniformly distributed over an interval that is possibly expanding as N passes to infinity. Note that T may or may not pass to infinity. If the regressor X_{it} takes on different mean values in different but large numbers of subpopulations (as in a stratified random sample), the above model appears realistic and provides a natural mechanism for identifying structural non-parametric elements. In this case, the dispersion of the 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.

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 non-parametric IV literature. In the present context, we show that the non-parametric local level and local linear estimators can both achieve the conventional $\sqrt{nh}$ convergence rate of non-parametric 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. This outcome is paradoxical, as local level regression is consistent whereas linear parametric regression that uses the true functional form and global information is inconsistent. Clearly, using all information is costly here. The reason, of course, is the endogeneity in the regressor. Intriguingly, however, partial information usage of the type employed in kernel regression avoids most of the problems associated with parametric regression.

This turns out to be the second central finding of this paper. It makes clear that local non-parametric regression has robustness advantages beyond robustness to specific functional form, for which it is most commonly celebrated. As shown here, non-parametric regression may also display a robustness to endogeneity in a regression by concentrating attention on local information and attenuating tail information that may be more heavily subjected to endogeneity effects.

Intuitively, any regression approach like conventional parametric regression that uses global information can be subject to distortionary effects from outlying observations. Such behaviour is very well known in statistics. Anscombe (1960) coined the term 'outlier' and suggested trimming techniques to attentuate the effects of outliers in regression based on the insurance analogy of 'protection and premium' to guard against unwanted effects. In the present context, the reason for the outlier effect is that at the limits of the domain of definition, the observations are more affected by the endogeneity of the regressor, so bias arising from the ends of the domain can dominate a global regression and result in inconsistency. By contrast, in non-parametric regression mainly local information is used in estimation so the endogeneity effects in the tail can be well controlled and first-order bias, at least, can be eliminated in local regression. In effect, by concentrating attention on the cluster of observations around each point, non-parametric regression localizes attention and removes outlier effects. This heuristic suggests that to recover the true regression line by a parametric method, a natural approach is to modify the regression by removing the effects of tail observations. The idea is comparable to that of trimming or Winsorizing the data, on which there is a large literature in statistics stemming largely from Anscombe's (1960) study. In the present context, to use Anscombe's analogy, the idea is to provide protection (against the possible effects of endogeneity) by paying a premium in terms of losing some observations. Kernel regression accomplishes this task by using data that is effectively in the locality of each individual regression point, thereby sacrificing an (asymptotically larger) infinity of observations to achieve a local regression fit and, incidentally in the process, protection from the effects of endogeneity.

This paper is structured as follows. Section 2 lays out a formal model and presents the main finding on the consistency and limiting normal distribution of non-parametric regression in a structural model with continuous location shifts, followed by a series of remarks. Section 3 studies the asymptotic properties of the OLS estimator for the coefficients in a linear model with an endogenous regressor that is generated via location shifts. We demonstrate that the OLS estimator is inconsistent in the case of compact support and has slower convergence rate than the non-parametric one in the case of infinite support. For this reason, we propose two new consistent estimators, a spatial $\mathcal{L}_2$ estimator and a bias-corrected OLS estimator. Section 4 concludes. Proofs of the main results are given in the Appendix.

2. LIMIT THEORY UNDER LOCATION SHIFTS

2.1. Models with infinite support

This section introduces a non-parametric regression model in which there are continuous location shifts in the regressor whose support is infinite. Observations $\{(y_i, X_i), 1 \le i \le n\}$ are generated as follows:

$$y_i = g(X_i) + u_i, \quad E(u_i|X_i) \neq 0, \tag{2.1}$$

$$X_i = \sum_{\alpha=-m}^{m} \mu_\alpha 1\{i \in A_\alpha\} + u_{xi}, \tag{2.2}$$

$$\mu_\alpha = \frac{\alpha L_n}{2m}, \quad \alpha = -m, -m+1, \ldots, m, \tag{2.3}$$

where $\{A_\alpha\}$ are disjoint clusters of individuals associated with locations $\{\mu_\alpha\}$ for the endogenous regressor X, and both $m \equiv m_n$ and L_n depend on the sample size n. Let $M \equiv M_n = \sum_{i=1}^n 1\{i \in A_\alpha\}$. The mechanism allows for M observations in the vicinity of each locality and may be generalized to allow M to depend on α, although that extension involves more notational complexity so it will not be pursued here. The total observation count is then $n = (2m+1)M$. Throughout the paper we require that as $n \to \infty$, $m \to \infty$, but M can either be finite or pass to ∞. In the case where X has infinite support we require $L_n \to \infty$ as $n \to \infty$. Otherwise, L_n can either be finite or pass to ∞.

We start by considering the local level kernel estimator of $g(x)$:

$$\widehat{g}(x) = \frac{\frac{1}{n}\sum_{i=1}^n y_i K_h(X_i - x)}{\frac{1}{n}\sum_{i=1}^n K_h(X_i - x)}, \tag{2.4}$$

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

ASSUMPTION A1. *(u_i, u_{xi}), $i = 1, \ldots, n$, are independent and identically distributed (i.i.d.).*

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

ASSUMPTION A3. *(a) The probability density function (p.d.f.) $f(\cdot, \cdot)$ of (u_i, u_{xi}) 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 $\int\int \left|u f_2''(u, u_x)\right| du du_x < \infty$. The marginal p.d.f. of u_{xi}, $f_{u_x}(\cdot)$, has second-order continuous derivatives such that $\int_{-\infty}^{\infty} \left|f_{u_x}'(p)\right| dp < \infty$, and $\int_{-\infty}^{\infty} \left|f_{u_x}''(p)\right| dp < \infty$. (b) There exists a function $C_f(\cdot)$ such that for any sequence $p_n \to \infty$, we have $|\int_{-\infty}^{-p_n} f(u, u_x) du_x + \int_{p_n}^{\infty} f(u, u_x) du_x| \le p_n^{-\upsilon} C_f(u)$ for some $\upsilon > 1$ and $\int_{-\infty}^{\infty} |u| C_f(u) du < \infty$. (c) $f_{u_x}(p_n) = O(|p_n|^{-\upsilon-1})$ as $|p_n| \to \infty$.*

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

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

ASSUMPTION A6. *As $n \to \infty$, $m \to \infty$, $L_n \to \infty$, $h \to 0$, $nh/L_n \to \infty$, $nh^5/L_n \to c \in [0, \infty)$, and $nh(L_n^3 m^{-4} + L_n^{-2\upsilon-1}) \to 0$.*

The i.i.d condition in A1 can be relaxed. For example, we can allow the process $\{(u_i, u_{xi}), i \ge 1\}$ to be strictly stationary and strong mixing with mixing coefficients that decay to zero at certain rates. A2 is weak. The zero-mean condition is needed for the identification of the non-parametric function $g(\cdot)$. But we do not impose the exogeneity condition $E(u_i|X_i) = 0$ a.s., nor do we assume conditional homoscedasticity. 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_{xi} exists. A4 imposes a standard smoothness condition on the regression function $g(\cdot)$. A5 is also standard, although the symmetry of the kernel function facilitates analysis and simplifies notation. A6 imposes conditions on h and its association with n, m, L_n and υ.

We now state the first main result.

THEOREM 2.1. *Under Assumptions A1–A6, we have*

$$\sqrt{nh/L_n}\left[\widehat{g}(x) - g(x) - \tfrac{1}{2}h^2\mu_2(K)\,g''(x)\right] \to_d N\left[0, \sigma^2\nu_2(K)\right], \tag{2.5}$$

where $\mu_2(K) = \int x^2K(x)dx$, *and* $\nu_2(K) = \int K(x)^2dx$. *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 but exogeneity, and where the regressor is integrated.

REMARK 2.1 (Comparison with the exogenous regressor case). When the regressor X_i is exogenous, stationary and has p.d.f. $f(x)$, the local constant estimator has the following limit behaviour:

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

where $\sigma^2(x) \equiv E\left(u_i^2|X_i = x\right)$. 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 analogue 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 $f_{u_x}(\cdot)$ 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 $o(h^2)$ 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_i appears in (2.6) but not in (2.5). The reason is that the denominator in the definition of $\widehat{g}(x)$, $\frac{1}{n}\sum_{i=1}^{n} K_h(X_i - x)$, does not converge in probability to a density. Instead, as shown in the Appendix, $\frac{L_n}{n}\sum_{i=1}^{n} K_h(X_i - x) \to_p 1$, and this result arises because the location shifts in X_i imply that the averaging operator $(L_n/n)\sum_{i=1}^{n}$ ensures that the density of u_{xi} 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_i appears in (2.5), whereas the conditional variance $\sigma^2(x)$ of u_i appears in (2.6). The reason is that in the bias–variance decomposition of $\widehat{g}(x) - g(x)$, the variance term $\frac{\sqrt{h}}{\sqrt{n}}\sum_{i=1}^{n} u_iK_h(X_i - x)$ 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_i leads to further averaging in the central limit theory and we have

$$\frac{\sqrt{hL_n}}{\sqrt{n}}\sum_{i=1}^{n}\{u_iK_h(X_i-x)-E[u_iK_h(X_i-x)]\}\to_d N[0,\sigma^2\nu_2(K)].$$

REMARK 2.2 (Local linear non-parametric 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):

$$\widetilde{g}(x)=\sum_{i=1}^{n}w_iY_i\Big/\sum_{i=1}^{n}w_i,$$

where $w_i=K_h(X_i-x)\{S_{n2}-(X_i-x)S_{n1}\}$ and $S_{nj}=\sum_{i=1}^{n}(X_i-x)^jK_h(X_i-x)$ 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

$$\sqrt{nh/L_n}[\widetilde{g}(x)-g(x)-\tfrac{1}{2}h^2\mu_2(K)g''(x)]\ \to_d N[0,\sigma^2\nu_2(K)]. \tag{2.7}$$

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 non-stationary non-parametric cointegrating regression, Phillips and Wang (2011) 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 non-stationary and location shift regressions.

REMARK 2.3 (Comparison with non-parametric IV regression). Consider the general non-parametric instrumental variable (IV) regression:

$$y_i=g(X_i)+u_i,\quad E(u_i\mid X_i)\neq 0,\quad E(u_i\mid W_i)=0, \tag{2.8}$$

where $(y_i,X_i,W_i)_{i=1}^n$ are observed and W_i is used as an instrument for X_i. Observe that $E(y_i|W_i)=E\{g(X_i)|W_i\}=\int g(x)\frac{f_{xw}(x,W_i)}{f_w(W_i)}dx$, 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. In order for W_i to be a valid IV for X_i, we usually require that W_i 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$(X_i,W_i)\neq 0$. To complete the specification of the model (2.8), we add a reduced-form equation for the endogenous regressor X_i. Let $m(W_i)\equiv E(X_i|W_i)$ and $u_{xi}\equiv X_i-E(X_i|W_i)$, so that we have

$$X_i=m(W_i)+u_{xi}. \tag{2.9}$$

This reduced-form equation helps in identifying the structural curve $g(\cdot)$ in (2.8) provided W_i is observable and the systematic component $m(W_i)$ of X_i 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 (2.8) has a triangular array structure so that the systematic component has the form $m(W_{in})$. If the variance of $m(W_{in})$ *expands* as the sample size n increases, then the signal in the regressor X_i correspondingly increases, relative to the variance of the stationary error u_{xi} in (2.9), the variance of the structural equation error u_i in (2.8) and the covariance $\text{cov}(u_{xi}, u_i)$. 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 $m(W_{in})$ may be regarded as carrying location-specific information that corresponds to a potentially new location in the system (2.1)–(2.3). Interestingly in this case, one does not even need to observe W_{in} in order to identify and consistently estimate the true structural regression curve $g(\cdot)$. It is sufficient for this leverage from W_{in} to be present in the regressor X_i. In effect, X_i then has an array structure itself and it is this 'non-stationarity' in the regressor that opens up the possibility of consistent estimation by direct non-parametric regression.

REMARK 2.4 (Link with the non-stationary non-parametric cointegrating regression). Recently Wang and Phillips (2009b, 2011) studied structural models of non-parametric cointegration and developed a limit theory for kernel cointegrating regression. It was shown that no ill-posed inverse problem arises in non-parametric models with non-stationary 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 non-linear structural model of cointegration:

$$y_i = g\left(X_i\right) + u_i, \quad i = 1, 2, \ldots, n,$$

where u_i is a zero mean stationary process, X_i is a non-stationary $I(1)$ regressor, and $g(\cdot)$ is an unknown smooth function to be estimated. Wang and Phillips (2009b, 2011) show that under some regularity conditions, the local constant estimate $\widehat{g}(x)$ of $g(x)$ has the following limiting mixed normal distribution:

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

where σ_u is a constant that depends on the kernel and the parameters underlying the process $\{X_i, u_i\}$, and $\mathcal{N}$ 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 non-parametric 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 behaviour ensures that $\text{Var}(X_i)$ expands as n increases and that data accumulate steadily over the entire support. In effect, continuous location shifts provide a form of recurrence in X_i, corresponding to the capacity of a non-stationary random wandering process to visit all points in the space an infinite number of times. Second, as mentioned in Remark 2.2, both estimates have the same asymptotic distribution and bias as the corresponding local linear estimates. Third, both estimates have a slower convergence rate than non-parametric 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 non-parametric 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 nh/L_n which gives the convergence rate $\sqrt{nh/L_n}$. Fourth, for both types of estimates, the conditional variance $\sigma^2(x)$ does not play a role in the asymptotics. Instead, it is the unconditional variance that really matters.

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 n 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 non-parametric local constant and local linear estimates is reduced in the case of continuous location shifts in an endogenous regressor. Second, the sample variance of the endogenous regressor needs to *expand* as n increases. A natural question is whether these two conditions are vital to identification and consistent non-parametric estimation.

It turns out the expansion of regressor location over the whole real line is unnecessary when u_{xi} is compactly supported. As indicated earlier, the non-parametric IV literature frequently assumes compact support for the endogenous regressor in order to use standard functional analysis. We therefore consider a similar case where the location shifts occur in a compact set. The model is as specified in (2.1)–(2.3), but we now assume that the $2m + 1$ locations are spaced over a fixed compact interval $[-L/2, L/2]$, i.e. $L_n = L$ does not pass to infinity as $n \to \infty$. As before, since we do not restrict u_{xi} to have zero mean, it is not restrictive to require that the $2m + 1$ locations are spaced over a compact interval that is symmetric around 0. We continue to assume that $M = \sum_{i=1}^{n} 1\{i \in A_\alpha\}$ so that $n = (2m + 1)M$. The pair (u_i, u_{xi}) continues to satisfy Assumptions A1, A2 and A3(i).

To proceed, we make the following assumptions.

ASSUMPTION A7. *The error term u_{xi} has compact support, i.e. $u_{xi} \in [\underline{u}, \overline{u}]$ 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 $n \to \infty, m \to \infty, h \to 0, nh \to \infty, nh^5 \to c \in [0, \infty)$ and $Mh/m^3 \to 0$.*

Note that under A7, the tail conditions in Assumptions A3(b)–(c) 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.

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

THEOREM 2.2. *Suppose Assumptions A1–A2, A3(a), A4, A5 and A7–A9 hold. Then*

$$\sqrt{nh/L}\left\{\widehat{g}(x) - g(x) - h^2\mu_2(K)\left[g'(x)\int_{\underline{u}}^{\overline{u}} f'_{u_x}(p)\,dp + \frac{1}{2}g''(x)\right]\right\} \to_d N\left[0, \sigma^2\nu_2(K)\right]. \qquad (2.10)$$

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. The same is also true for the local linear estimate. This fast rate contrasts with the much slower rates achievable for non-parametric IV estimation without the advantage of location shifts. In comparison with the result in Theorem 2.1, we observe that the bias function in (2.10) 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 section reveals a further advantage of local smoothing techniques in dealing with issues of endogeneity.

3. INCONSISTENT PARAMETRIC REGRESSION UNDER LOCATION SHIFTS

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) is known to be of the parametric form $g(x) = g(x, \theta)$, say, where $g(\cdot, \cdot)$ is known up to the finite dimensional parameter, θ, then $g(\cdot)$ can be estimated by direct parametric estimation of θ. However, under the conditions given in Theorem 2.2 and assuming that the form of $g(x) = g(x, \theta)$ is known, the non-linear least squares (NLS) estimate $\widehat{\theta}$ of θ is generally inconsistent because of endogeneity in the regressor. To see this, we focus on the simple linear model

$$y_i = \beta_0 + \beta_1 X_i + u_i, \quad E(u_i \mid X_i) \neq 0, \quad E(u_i) = 0, \tag{3.1}$$

where the endogenous regressor X_i is generated according to (2.2)–(2.3). Let $\widehat{\beta}_0$ and $\widehat{\beta}_1$ denote the ordinary least squares (OLS) estimators of β_0 and β_1, respectively. Below we first demonstrate the inconsistency of these OLS estimators and then propose two methods for consistent estimation of (β_0, β_1).

3.1. *Inconsistency of* $(\widehat{\beta}_0, \widehat{\beta}_1)$

To study the asymptotic properties of the OLS estimator $(\widehat{\beta}_0, \widehat{\beta}_1)$, we maintain Assumptions A1–A2 and add the following assumption.

ASSUMPTION A2*. *$E(u_{xi}) = \mu_x$ and* $\mathrm{Var}(u_{xi}) = \sigma_x^2 < \infty$.

Note that the above assumption is automatically satisfied under A7 for the case of regressors with compact support. Under Assumptions A1, A2 and A2*, we first derive the probability limit of the OLS estimator $\widehat{\beta}_1$ and show that it is inconsistent for β_1 when L_n is fixed. The probability limit of $\widehat{\beta}_0$ follows straightforwardly. Write

$$\widehat{\beta}_1 = \beta_1 + \frac{n^{-1} \sum_{i=1}^{n} (X_i - \overline{X}) u_i}{n^{-1} \sum_{i=1}^{n} (X_i - \overline{X})^2}, \tag{3.2}$$

where $\overline{X} \equiv n^{-1} \sum_{i=1}^{n} X_i$. First, by the definition of $\{\mu_\alpha\}$, the weak law of large numbers (WLLN) and the Chebyshev inequality, we have

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \\
&= \frac{1}{2m+1}\sum_{\alpha=-m}^{m}\frac{1}{M}\sum_{i\in A_\alpha}\left[\frac{L_n\alpha}{2m} + (u_{xi} - \overline{u}_x)\right]^2 \\
&= \frac{L_n^2}{4m^2(2m+1)}\sum_{\alpha=-m}^{m}\alpha^2 + \frac{1}{n}\sum_{i=1}^{n}(u_{xi} - \overline{u}_x)^2 + \frac{L_n}{m(2m+1)}\sum_{\alpha=-m}^{m}\frac{\alpha}{M}\sum_{i\in A_\alpha}(u_{xi} - \overline{u}_x) \\
&= \frac{L_n^2}{2m^2(2m+1)}\sum_{\alpha=1}^{m}\alpha^2 + \sigma_x^2 + o_P(1) \\
&= \frac{L_n^2}{12}[1 + o(1)] + \sigma_x^2 + o_P(1), \qquad (3.3)
\end{aligned}$$

where $\overline{u}_x \equiv n^{-1}\sum_{i=1}^{n} u_{xi}$, and the third line follows provided $n/L_n^2 \to \infty$ as $n \to \infty$.[2]

Next

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})u_i \\
&= \frac{1}{2m+1}\sum_{\alpha=-m}^{m}\frac{1}{M}\sum_{i\in A_\alpha}\left(\frac{L_n\alpha}{2m} + u_{xi} - \overline{u}_x\right)u_i \\
&= \frac{L_n}{2m(2m+1)}\sum_{\alpha=-m}^{m}\frac{1}{M}\sum_{i\in A_\alpha}\alpha u_i + \frac{1}{n}\sum_{i=1}^{n}u_{xi}u_i - \frac{\overline{u}_x}{n}\sum_{i=1}^{n}u_i \\
&= \frac{L_n}{2m(2m+1)}\sum_{\alpha=-m}^{m}\frac{1}{M}\sum_{i\in A_\alpha}\alpha u_i + E(u_{xi}u_i) + o_P(1).
\end{aligned}$$

Consider the first term in the last expression. Let $T_{1n} \equiv \frac{L_n}{2m(2m+1)}\sum_{\alpha=-m}^{m}\frac{1}{M}\sum_{i\in A_\alpha}\alpha u_i$. Then $E(T_{1n}) = 0$ as $E(u_i) = 0$, and

$$\begin{aligned}
\mathrm{Var}(T_{1n}) &= \frac{L_n^2}{4m^2(2m+1)^2}\mathrm{Var}\left(\sum_{\alpha=-m}^{m}\frac{1}{M}\sum_{i\in A_\alpha}\alpha u_i\right) \\
&= \frac{L_n^2\sigma^2}{4Mm^2(2m+1)^2}\sum_{\alpha=-m}^{m}\alpha^2 \\
&= O\left(\frac{L_n^2}{n}\right) = o(1).
\end{aligned}$$

[2] Letting $T_n \equiv \frac{L_n}{m(2m+1)}\sum_{\alpha=-m}^{m}\frac{\alpha}{M}\sum_{i\in A_\alpha}(u_{xi} - \mu_x) = \frac{L_n}{m(2m+1)}\sum_{\alpha=-m}^{m}\frac{\alpha}{M}\sum_{i\in A_\alpha}(u_{xi} - \overline{u}_x)$, then $E(T_n) = 0$ and $\mathrm{Var}(T_n) = \frac{L_n^2}{Mm^2(2m+1)^2}\sum_{\alpha=-m}^{m}\alpha^2\sigma_x^2 = O\left(\frac{L_n^2}{Mm}\right) = O\left(\frac{L_n^2}{n}\right) = o(1)$.

Hence $T_{1n} = o_P(1)$ and $\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})u_i = E(u_{xi}u_i) + o_P(1)$. This, together with (3.2) and (3.3), implies

$$\widehat{\beta}_1 = \beta_1 + \frac{E(u_{xi}u_i) + o_P(1)}{\frac{L_n^2}{12}[1 + o(1)] + \sigma_x^2 + o_P(1)} = \beta_1 + \frac{E(u_{xi}u_i)}{\frac{L_n^2}{12} + \sigma_x^2}[1 + o_P(1)], \tag{3.4}$$

thereby giving the following result.

LEMMA 3.1. *Suppose Assumptions A1, A2 and A2 * hold and $n/L_n^2 \to \infty$. Then,*

$$\widehat{\beta}_1 = \beta_1 + \frac{E(u_{xi}u_i)}{\frac{L_n^2}{12} + \sigma_x^2}[1 + o_P(1)].$$

REMARK 3.1 (Inconsistency of parametric regression). We explore the implications of the above lemma in the two cases where $L_n = L$ is fixed and $L_n \to \infty$ as $n \to \infty$. When $L_n = L$ is fixed, $\widehat{\beta}_1$ has the probability limit $\beta_1 + E(u_{xi}u_i)/\left(\frac{L^2}{12} + \sigma_x^2\right)$ and is inconsistent unless $E(u_{xi}u_i) = 0$, viz. X_i is exogenous. For $\widehat{\beta}_0$, we have

$$\widehat{\beta}_0 - \beta_0 = -\overline{X}(\widehat{\beta}_1 - \beta_1) + n^{-1}\sum_{i=1}^{n} u_i = -\mu_x \frac{E(u_{xi}u_i)}{\frac{L^2}{12} + \sigma_x^2} + o_P(1), \tag{3.5}$$

so $\widehat{\beta}_0$ is inconsistent for β_0 unless either $\mu_x = 0$ or $E(u_{xi}u_i) = 0$. Hence, the parametric estimator $\widehat{g}^p(x) \equiv \widehat{\beta}_0 + \widehat{\beta}_1 x$ is inconsistent for $g(x) \equiv \beta_0 + \beta_1 x$ at all points except $x = \mu_x$. By contrast, according to Theorem 2.2, the non-parametric estimator is consistent for all x satisfying certain domain restrictions. When $L_n \to \infty$, the OLS estimator $\widehat{\beta}_1$ is consistent for β_1 as $\widehat{\beta}_1 = \beta_1 + o_P(1)$, due to the strengthening signal in the regressor as $L_n \to \infty$. This result, together with (3.5), implies that $\widehat{\beta}_0$ is consistent for β_0, and so the parametric regression estimator of $g(x) = \beta_0 + \beta_1 x$ is also consistent. However, if L_n diverges to infinity slowly like $L_n = \log n$, the estimation bias may disappear at a very slow rate.

To find the limit distribution of $(\widehat{\beta}_0, \widehat{\beta}_1)$, we add the following assumption.

ASSUMPTION A2**. $E[|u_i u_{xi}|^{2+\delta}] < \infty$ *for some* $\delta > 0$.

Let $\theta \equiv (\beta_0, \beta_1)'$ and $\widehat{\theta} \equiv (\widehat{\beta}_0, \widehat{\beta}_1)'$. Let $\underline{X}_i \equiv (1, X_i)'$, $\mathbf{X} \equiv (\underline{X}_1, \ldots, \underline{X}_n)'$, $\mathbf{u} \equiv (u_1, \ldots, u_n)'$, and $\mathbf{y} \equiv (y_1, \ldots, y_n)'$. Define $D_n \equiv \text{diag}(1, L_n)$:

$$\Gamma \equiv \lim_{n\to\infty} \begin{bmatrix} 1 & \frac{\mu_x}{L_n} \\ \frac{\mu_x}{L_n} & \frac{1}{12} + \frac{E\left(u_{xi}^2\right)}{L_n^2} \end{bmatrix} \quad \text{and} \quad \Omega \equiv \lim_{n\to\infty} \begin{bmatrix} \sigma^2 & \frac{E\left(u_{xi}u_i^2\right)}{L_n} \\ \frac{E\left(u_{xi}u_i^2\right)}{L_n} & \frac{\sigma^2}{12} + \frac{\text{Var}(u_{xi}u_i)}{L_n^2} \end{bmatrix}. \tag{3.6}$$

After centring, the limiting distribution of $\widehat{\theta}$ is given in the following theorem.

THEOREM 3.1. *Suppose Assumptions A1, A2, A2 * and A2 ** hold and* $n/L_n^2 \to \infty$. *Then*

$$\sqrt{n} D_n[\widehat{\theta} - \theta - (\mathbf{X}'\mathbf{X})^{-1} E(\mathbf{X}'\mathbf{u})] \to_d N(0, \Gamma^{-1}\Omega\Gamma^{-1}). \tag{3.7}$$

REMARK 3.2 (Different convergence rates for the intercept and slope estimates). Straightforward calculations show that

$$(\mathbf{X}'\mathbf{X})^{-1} E(\mathbf{X}\mathbf{u}) = \frac{E(u_{xi}u_i)}{n^{-1}\sum_{i=1}^n (X_i - \overline{X})^2} \begin{pmatrix} -\overline{X} \\ 1 \end{pmatrix}.$$

An immediate implication of Theorem 3.1 is therefore that

$$\sqrt{n}\left[\widehat{\beta}_0 - \beta_0 + \frac{\overline{X}\, E\,(u_{x1}u_1)}{n^{-1}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2}\right] \to_d N\,(0, \omega_{11})\,, \tag{3.8}$$

$$L_n\sqrt{n}\left[\widehat{\beta}_1 - \beta_1 - \frac{E\,(u_{x1}u_1)}{n^{-1}\sum_{i=1}^n (X_i - \overline{X})^2}\right] \to_d N\,(0, \omega_{22})\,, \tag{3.9}$$

where ω_{11} and ω_{22} are the (1,1) and (2,2) elements of $\Gamma^{-1}\Omega\Gamma^{-1}$, respectively. If $L_n \to \infty$ as $n \to \infty$, then $\Gamma^{-1}\Omega\Gamma^{-1}$ is diagonal, implying that the intercept estimator, after centring, is asymptotically independent of the slope estimator, and the slope coefficient estimator has a faster convergence rate, due to the stronger signal in the regressor. In contrast, in the fixed L_n case, the two estimators are asymptotically dependent, have the same rate of convergence, and the range of location shifts contributes to the asymptotic variance formula in (3.7) in a complicated way.

In spite of the $O\left(\sqrt{n}\right)$ convergence rate for the intercept estimator and the $O\left(L_n\sqrt{n}\right)$ convergence rate for the slope estimator, the result in (3.7) does not seem useful for inferential purposes because the bias term in (3.7) does not appear to be estimable at the required $\sqrt{n}/L_n^2$-rate (for the intercept parameter) or $\sqrt{n}/L_n$-rate (for the slope parameter) to be eliminated (recall from (3.3) that $n^{-1}\sum_{i=1}^n (X_i - \overline{X})^2 = O_P(L_n^2)$). It is worth mentioning that the residuals $\{\widehat{u}_i\}$ from the OLS regression are useless in constructing a consistent estimate of $E(X_i u_i)$ because of the orthogonality of $\widehat{u}_i$ and X_i. One may consider consistent estimation of $E(X_i u_i)$ by estimating the model first via the local level (or local linear) non-parametric method at interior points and then obtaining the non-parametric residuals from the structural equation. Unfortunately, this approach usually requires uniform consistency of the local level (or local linear) estimate over the whole support of the regressor, which seems extremely difficult here given the fact that the variance of the regressor is expanding as the sample size increases or that the non-parametric estimates are only consistent at points within a subset of the interior of the support of the regressor for the fixed L_n case. Instead, the next subsection proposes two alternative methods to achieve consistent estimation of (β_0, β_1) by direct use of the non-parametric level estimate in a parametric regression.

3.2. Spatial $\mathcal{L}_2$ and bias-corrected OLS estimation

We now propose two methods for consistent estimation of (β_0, β_1). The first method is a spatial $\mathcal{L}_2$ regression and the second method involves bias-corrected OLS estimation. The $\mathcal{L}_2$ method treats the linear regression function as unknown, estimates it non-parametrically by $\widehat{g}(x)$, and then regresses $\widehat{g}(x)$ on $(1, x)$ to estimate the unknown parameter (β_0, β_1) by minimizing a spatial $\mathcal{L}_2$ criterion function, using a continuum of pseudo-observations on $(x, \widehat{g}(x))$ where x is restricted to be bounded away from the two tails. We prove that the resulting $\mathcal{L}_2$ estimator of β_0 can be $\sqrt{n}$-consistent and that of β_1 can be super-consistent. We show that the OLS bias terms can be corrected by using the residuals from the $\mathcal{L}_2$ regression, and the bias-corrected OLS estimators can achieve the same consistency rates as the $\mathcal{L}_2$ estimators.

3.2.1. Spatial $\mathcal{L}_2$ regression. Noting that the local level estimate $\widehat{g}(x)$ is consistent for $g(x) = \beta_0 + \beta_1 x$ in the interior of the regressor support, we propose to estimate the unknown parameter $\theta \equiv (\beta_0, \beta_1)'$ by minimizing the following (spatial) $\mathcal{L}_2$ criterion:

$$S_n(\beta_0, \beta_1) \equiv \int_{a_n}^{b_n} [\widehat{g}(x) - \beta_0 - \beta_1 x]^2 \widehat{f}(x) dx, \tag{3.10}$$

where a_n and b_n are integration limits that serve to truncate observations in the two tails, $\widehat{f}(x) \equiv N^{-1} \sum_{i=1}^{n} K_h(x - X_i)$ is a pseudo-estimate of the 'density' of X_i, and $N \equiv n/L_n$ signifies the effective number of observations used in the non-parametric estimation, which does not need to be observed in practice for implementation. In the case where $L_n \to \infty$ as $n \to \infty$, we allow but do not require that the integration limits a_n and b_n pass to (positive or negative) infinity: if they pass to infinity, we assume for simplicity that they do so at the same rate c_n and write $a_n = c_n a$ and $b_n = c_n b$, where $c_n \to \infty$ and $c_n = O(L_n)$ due to the endogeneity bias of the non-parametric estimators in the tails; if they do not pass to infinity, we can simply write a_n and b_n as a and b by taking $c_n = 1$. In the case where $L_n = L$ is fixed as $n \to \infty$, we can only consider fixed integration limits and take $c_n = 1$. Note that $\widehat{f}(x)$ is a weight function that serves to avoid division by zero and to perform trimming in areas of sparse support, and $[a_n, b_n]$ defines a (possibly expanding) compact set on which the non-parametric estimates $\widehat{g}(x)$ are used in the estimation of (β_0, β_1). Clearly, (3.10) provides a trimming operation implicitly via the local nature of the estimate $\widehat{g}(x)$ and explicitly via the use of the (truncated) domain $[a_n, b_n]$.

The minimizer of (3.10) $\widetilde{\theta} \equiv (\widetilde{\beta}_0, \widetilde{\beta}_1)'$ is given by

$$\widetilde{\theta} = Q_n^{-1} \begin{bmatrix} \int_{a_n}^{b_n} \widehat{g}(x) \widehat{f}(x) dx \\ \int_{a_n}^{b_n} x \widehat{g}(x) \widehat{f}(x) dx \end{bmatrix},$$

where

$$Q_n = \begin{bmatrix} \int_{a_n}^{b_n} \widehat{f}(x) dx & \int_{a_n}^{b_n} x \widehat{f}(x) dx \\ \int_{a_n}^{b_n} x \widehat{f}(x) dx & \int_{a_n}^{b_n} x^2 \widehat{f}(x) dx \end{bmatrix}.$$

To develop the limit theory, we make the following assumptions.

ASSUMPTION A3*. *Either one of the following conditions holds: (a) The error term* u_{xi} *has infinite support such that Assumption A.3(b) holds,* $L_n^{-v} = O(h^2)$, *and* $f_{u_x}(L_n) = O(h^2)$ *as* $L_n \to \infty$. *As* $n \to \infty$, $L_n \to \infty$. *(b) The error term* u_{xi} *has compact support, i.e.* $u_{xi} \in [\underline{u}, \overline{u}]$ *a.s. for some finite numbers* $\underline{u}$ *and* $\overline{u}$. L_n *is either fixed or tends to* ∞ *as* $n \to \infty$. *If* $L_n = L$ *is fixed,* L *is sufficiently large that* $x \in (-L/2 + \overline{u}, L/2 + \underline{u})$ *for all* $x \in [a, b]$.

ASSUMPTION A6*. *As* $n \to \infty$, $n/L_n^2 \to \infty$, $nh^4 c_n/L_n \to 0$, $nm^{-4} c_n L_n^3 \to 0$, $(n/L_n)^{-\delta/2} L_n \to 0$, *and* c_n *either passes to infinity or takes value 1 such that* $c_n = O(L_n)$.

A3* is modified from A3(b)–(c) for the infinite support case and A7–A8 for the compact support case. A6* is adapted from A6 by taking into account the known linear regression function and imposing an undersmoothing bandwidth sequence ($nh^4 c_n/L_n \to 0$). The next-to-last requirement in A6* is needed to verify the Liapounov condition.

THEOREM 3.2. *Suppose Assumptions A1, A2, A2 *, A2**, A3(a), A3*, A5 and A6* hold. Then*

$$\sqrt{n/L_n} C_n (\widetilde{\theta} - \theta) \to_d N\left(0, \sigma^2 Q^{-1}\right), \tag{3.11}$$

where

$$C_n \equiv \begin{pmatrix} \sqrt{c_n} & 0 \\ 0 & \sqrt{c_n^3} \end{pmatrix} \quad \text{and} \quad Q \equiv \begin{pmatrix} b - a & \frac{b^2 - a^2}{2} \\ \frac{b^2 - a^2}{2} & \frac{b^3 - a^3}{3} \end{pmatrix}.$$

REMARK 3.3 (Consistency of $\widetilde{\beta}_0$ and $\widetilde{\beta}_1$). Despite the non-parametric convergence rate of the regressand $\widehat{g}(x)$, Theorem 3.2 indicates that the $\mathcal{L}_2$ estimate $\widetilde{\beta}_0$ is $\sqrt{n}$-consistent, whereas the $\mathcal{L}_2$ estimate $\widetilde{\beta}_1$ is $\sqrt{nL_n}$-consistent provided one takes $c_n = L_n$. Both estimates achieve the parametric $\sqrt{n}$-rate of consistency for the case of fixed L_n, and $\widetilde{\beta}_1$ is super-consistent in the case where L_n passes to infinity as $n \to \infty$. To obtain these results in Theorem 3.2, we have employed two devices. First, undersmoothing is required to eliminate the $O(h^2)$ bias terms from the first-stage non-parametric regression estimate of $g(x)$. This is standard in the non-parametric or semiparametric literature when the first-stage non-parametric estimates are used in a second-stage parametric or non-parametric estimation. Second, to reduce the variation of the non-parametric estimates, we have used integration in the $\mathcal{L}_2$ regression. The smoothing operation of integration helps to produce the (nearly) parametric convergence rate of $\widetilde{\theta}$ despite the slow non-parametric convergence rate of $\widehat{g}(x)$. The mechanism is analogous to that of average marginal effect or derivative estimation.

Cristóbal et al. (1987) defined a class of linear regression estimators by minimizing the following $\mathcal{L}_2$ criterion that is similar to our spatial regression criterion:

$$\overline{S}_n(\beta_0, \beta_1) \equiv \int (g_n(x) - \beta_0 - \beta_1 x)^2 \, d\Omega_n(x),$$

where $g_n(x)$ is a non-parametric estimate of $g(x) \equiv E(y_i|X_i = x)$ and Ω_n is a weight measure that typically has an infinite support. They demonstrated that this class of estimators includes the ordinary and generalized ridge regression estimators as special cases and they provided asymptotic properties of these estimators for the classical i.i.d. design. In contrast, our model has endogenous regressors that involve a complicated location shift pattern, and we intentionally use a compactly supported weight function ($\widehat{f}(x)1\{a_n \le x \le b_n\}$), rather than infinite support, in order to provide protection against the bias effects of endogeneity in the tails.

3.2.2. Bias-corrected OLS estimation. We now propose a bias-correction procedure for the OLS estimator of $\theta = (\beta_0, \beta_1)'$ in the linear structural equation (3.1). Let $\widetilde{u}_i \equiv y_i - \widetilde{\beta}_0 - \widetilde{\beta}_1 X_i$ and $\widetilde{c} \equiv n^{-1}\sum_{i=1}^n \widetilde{u}_i X_i$. We define bias-corrected OLS estimators of β_0 and β_1, respectively, as

$$\widehat{\beta}_{0c} \equiv \widehat{\beta}_0 + \frac{\overline{X}\,\widetilde{c}}{n^{-1}\sum_{i=1}^n (X_i - \overline{X})^2} \quad \text{and} \quad \widehat{\beta}_{1c} \equiv \widehat{\beta}_1 - \frac{\widetilde{c}}{n^{-1}\sum_{i=1}^n (X_i - \overline{X})^2}.$$

Let $\widehat{\theta}_c \equiv (\widehat{\beta}_{0c}, \widehat{\beta}_{1c})'$. The following theorem establishes the consistency and asymptotic normality of $\widehat{\theta}_c$.

THEOREM 3.3. *Suppose Assumptions A1, A2, A2*, A2**, A3(a), A5 and A6*. (a) If Assumption A3*(a) holds or A3*(b) holds with $L_n \to \infty$ as $n \to \infty$, then $\sqrt{n/L_n}C_n(\widehat{\theta}_c - \theta) \to_d N(0, \Psi)$, where*

$$\Psi = \sigma^2 \lim_{n\to\infty} \begin{pmatrix} c_n^{-2}\mu_x^2 q^{22} + \dfrac{c_n}{L_n} & -c_n^{-1}\mu_x q^{22} + \dfrac{c_n}{L_n}\gamma \\ -c_n^{-1}\mu_x q^{22} + \dfrac{c_n}{L_n}\gamma & q^{22} \end{pmatrix},$$

$\gamma = q^{21}(b-a) + \frac{1}{2}q^{22}\left(b^2 - a^2\right)$, and q^{ij} is the (i,j) element of Q^{-1} for $i,\ j = 1, 2$. *(b) If Assumption A3*(b) holds with $L_n = L$ fixed, then $\sqrt{n/L}(\widehat{\theta}_c - \theta) \to_d N(0, \Psi)$, where*

$$\Psi \equiv B^{-1}\Upsilon B^{-1}, \quad \Upsilon \equiv \sigma^2 \begin{pmatrix} L^{-1} & L^{-1} \\ L^{-1} & c'Q^{-1}c \end{pmatrix}, \quad B \equiv \begin{pmatrix} 1 & \mu_x \\ \mu_x & \dfrac{L^2}{12} + E(u_{xi}^2) \end{pmatrix},$$

and $c \equiv (\mu_x, c_x)'$ with $c_x \equiv \frac{L^2}{12} + E(u_{xi}^2)$.

REMARK 3.4 (Consistency and extension). A similar remark to that after Theorem 3.2 holds for $\widehat{\beta}_{0c}$ and $\widehat{\beta}_{1c}$ in part (i) of the above theorem. In either of the two most important special cases, $c_n = L_n$ and $c_n = 1$, the formula for Ψ can be greatly simplified. Even though we only focus on the linear structural equation model as specified in (3.1), it is straightforward to extend our theory to the general non-linear structural equation model. Suppose $\{y_i\}$ is generated according to

$$y_i = g(X_i, \theta) + u_i, \quad E(u_i \mid X_i) \neq 0, \tag{3.12}$$

where $g(\cdot, \theta)$ is known up to the finite dimensional parameter θ, and the endogenous regressor X_i satisfies (2.2) and (2.3). It is straightforward to show that the NLS estimator $\widehat{\theta}$ of θ is inconsistent. As before, the non-parametric local level estimate $\widehat{g}(x)$ of $g(x) = g(x, \theta)$ is consistent for a large portion of the domain of the regressor. Then, the unknown parameter θ can be estimated by minimizing the following (spatial) $\mathcal{L}_2$ criterion $S_n(\theta) \equiv \int_{a_n}^{b_n} (\widehat{g}(x) - g(x, \theta))^2 \widehat{f}(x) dx$, just as before. Following the proof of Theorem 3.2, we can establish the consistency of the $\mathcal{L}_2$ estimator under some regularity conditions. A bias-corrected NLS estimator can also be constructed. The details are similar to the linear case and are omitted.

4. 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 non-parametric regression in both cases. Importantly, there is no need for non-parametric IV estimation or the complications of functional inversion and regularization. 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 non-parametric cointegrating regression where the limit theory also involves only a single $O(h^2)$ bias term (Wang and Phillips, 2009b).

We also explore a paradox where the greater use of correct prior information on a model can be detrimental in regression. The explanation for this paradox is that even when we use additional correct information about the specification of a model, that information may still not be complete and, in consequence, may distort regression results. In the example studied here, the correct additional information used is considerable and is the full functional form specification of the model. Nevertheless, the omitted information (endogeneity) that makes the model specification incomplete is very important and leads to inconsistency in parametric regression. In such situations, it is very interesting that partial information can be successful where more complete information fails. In applied statistics, it has long been known that controlling for outliers in regression can help to achieve robustness. We show that non-parametric kernel regression naturally utilizes this mechanism to great advantage in structural regression. More specifically, kernel regression has a considerable additional advantage beyond its usually touted advantage of robustness to (unknown) functional form. Local non-parametric regression also provides robustness to endogeneity in the regressor when there are systematic influences that assure identification, such as location shifts or non-stationarity in the data. It is also possible to obtain consistent estimation by parametric methods in such cases. In particular, spatial $\mathcal{L}_2$ regression is shown to successfully remove endogeneity bias and inconsistency by bounding the domain of the regression. This approach is analogous to the treatment of outliers—it provides protection against possible effects of endogeneity in parametric regression by paying a premium through the loss of tail information in the data.

ACKNOWLEDGMENTS

We thank the referees and editor for helpful comments. Phillips acknowledges partial support from NSF Grants SES 06-47086 and SES 09-56687.

REFERENCES

Anscombe, F. J. (1960). Rejection of outliers. *Technometrics 2*, 123–47.

Cristóbal, J. A. C., P. F. Roca and W. G. Manteiga (1987). A class of linear regression parameter estimators constructed by nonparametric estimation. *Annals of Statistics 15*, 603–09.

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

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

Phillips, P. C. B. and L. Su (2009). Nonparametric structural estimation via continuous location shifts in an endogenous regressor. Working paper, Dept. of Economics, Yale University.

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

Wang, Q. and P. C. B. Phillips (2009a). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory 25*, 710–38.

Wang, Q. and P. C. B. Phillips (2009b). Structural nonparametric cointegrating regression. *Econometrica 77*, 1901–48.

Wang, Q. and P. C. B. Phillips (2011). Asymptotic theory for zero energy density estimation with nonparametric regression applications. *Econometric Theory 27*, 236–69.

APPENDIX A: TECHNICAL RESULTS AND PROOFS

We first state two lemmas that are used in the proof of Theorem 2.1. The proofs of these and later lemmas in the Appendix are given in supplementary material that is available online at http: //www.mysmu.edu/faculty/ljsu/Publications/Location˙supplement.pdf.

LEMMA A.1. *Let $f_u(\cdot)$ be the p.d.f. of $\{u_i\}$. Suppose A2–A3 hold. Then,*

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

LEMMA A.2. *Let $\Theta_{jn} \equiv \frac{L_n}{n} \sum_{i=1}^{n} (X_i - x)^j K_h(X_i - x)$ for $j = 0, 1, 2$. Suppose A1–A6 hold. Then (a) $E\Theta_{0n} = 1 + o(1)$, (b) $E\Theta_{1n} = o(h^2)$, (c) $E\Theta_{2n} = h^2\mu_2(K) + o(h^2)$ and (d) $\mathrm{Var}(\Theta_{jn}) = (h^{2j-1}L_n/n)\int z^{2j}K^2(z)dz + o(h^{2j-1}L_n/n)$ for $j = 0, 1, 2$.*

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

$$\widehat{g}(x) - g(x) = \frac{\frac{L_n}{n} \sum_{i=1}^{n} \{g(X_i) - g(x)\} K_h(X_i - x)}{\frac{L_n}{n} \sum_{i=1}^{n} K_h(X_i - x)} + \frac{\frac{L_n}{n} \sum_{i=1}^{n} u_i K_h(X_i - x)}{\frac{L_n}{n} \sum_{i=1}^{n} K_h(X_i - x)}. \tag{A.1}$$

By Lemmas A.2(a) and (d), the Chebyshev inequality, and Assumption A6,

$$\Theta_{0n} = \frac{L_n}{n}\sum_{i=1}^{n} K_h(X_i - x) = 1 + O_P\left[(nh/L_n)^{-1/2}\right] = 1 + o_P(1). \tag{A.2}$$

By the second-order Taylor expansion, $\frac{L_n}{n}\sum_{i=1}^{n}[g(X_i) - g(x)] K_h(X_i - x) = g'(x)\Theta_{1n} + \frac{g''(x)}{2}\Theta_{2n} + R_n(x)$, where $R_n(x) = \frac{L_n}{2n}\sum_{i=1}^{n}\left[g''\left(X_i^*\right) - g''(x)\right](X_i - x)^2 K_h(X_i - x)$, and X_i^* lies between X_i and x. Following the proof of Lemma A.2, it is easy to show that $R_n(x) = o_P(h^2)$. Then by Lemma A.2(b)–(d), the Chebyshev inequality, and Assumption A6:

$$\frac{L_n}{n}\sum_{i=1}^{n}[g(X_i) - g(x)] K_h(X_i - x) = h^2\mu_2(K)\frac{g''(x)}{2} + o_P(h^2). \tag{A.3}$$

Let $\Theta_{3n} \equiv \sqrt{L_n h/(2mM)}\sum_{i=1}^{n} u_i K_h(X_i - x)$. We show that $\Theta_{3n} - E\Theta_{3n} \to_d (N0, \sigma^2\, \nu_2(K))$. Write $\Theta_{3n} = \sum_{i=1}^{n} Z_i$, where $Z_i \equiv \sqrt{L_n h/(2mM)} u_i K_h(X_i - x)$. Let $\beta_n \equiv 2m/L_n$. By a change of variables, the Fubini theorem, Lemma A.1 and under A2 and A6:

$$\begin{aligned} E\Theta_{3n} &= \frac{\sqrt{Mh}}{\sqrt{2m/L_n}}\sum_{\alpha=-m}^{m} E\left[u_i K_h(\mu_\alpha + u_{xi} - x)\right] \\ &= \frac{\sqrt{Mh\beta_n}}{\beta_n}\sum_{\alpha=-m}^{m} uK(z) f(u, x - \mu_\alpha + hz)\, dz du \\ &= \sqrt{Mh\beta_n} K(z)\left[u\frac{1}{\beta_n}\sum_{\alpha=-m}^{m} f(u, x - \mu_\alpha + hz)\, du\right] dz \\ &= \sqrt{Mh\beta_n} K(z)\left[O\left(\beta_n^{-2} + L_n^{-\upsilon}\right) + uf_u(u)\, du\right] dz \\ &= \sqrt{nh/L_n}\, O\left(L_n^2/m^2 + L_n^{-\upsilon}\right) = o(1), \end{aligned} \tag{A.4}$$

where we use the fact that the result in Lemma A.1 also holds uniformly in a small neighbourhood of x. Similarly, by (A.4) and Assumptions A1 and A2,

$$\begin{aligned} \mathrm{Var}(\Theta_{3n}) &= \frac{h}{\beta_n}\sum_{\alpha=-m}^{m} E\left[u_i^2 K_h^2(\mu_\alpha + u_{xi} - x)\right] - o(1) \\ &= \frac{1}{\beta_n}\sum_{\alpha=-m}^{m}\int u^2 K(z)^2 f(u, x - \mu_\alpha + hz)\, dz du + o(1) \\ &= \nu_2(K)\int u^2 \int f(u, x - p)\, dp du + o(1) = \nu_2(K)\sigma^2 + o(1). \end{aligned}$$

To show the asymptotic normality of $\Theta_{3n} - E\Theta_{3n}$, by the above variance calculation and the independence of $\{u_i, u_{xi}\}$ across i, it suffices to check the Liapounov condition. Let $\overline{Z}_i \equiv Z_i - E(Z_i)$. Then by the C_r

and Jensen inequalities and Assumptions A2 and A4–A6,

$$\begin{aligned}
\sum_{i=1}^{n} E\left|\overline{Z}_i\right|^{2+\delta} &\leq 2^{2+\delta}\left(\frac{L_n h}{2mM}\right)^{1+\delta/2}\sum_{i=1}^{n} E\left|u_i K_h(X_i - x)\right|^{2+\delta} \\
&= 2^{2+\delta}\left(\frac{2mMh}{L_n}\right)^{-\delta/2}\frac{L_n}{2m}\sum_{\alpha=-m}^{m}\int |u|^{2+\delta} K(z)^{2+\delta} f(u, x-\mu_\alpha + hz)\, dz du \\
&\approx \left(\frac{2mMh}{L_n}\right)^{-\delta/2} 2^{2+\delta}\int K(z)^{2+\delta}\, dz \int |u|^{2+\delta}\frac{L_n}{2m}\sum_{\alpha=-m}^{m} f(u, x-\mu_\alpha)\, du \\
&= \left(\frac{nh}{L_n}\right)^{-\delta/2} 2^{2+\delta}\int K(z)^{2+\delta}\, dz\left[E|u|^{2+\delta} + o(1)\right] \to 0.
\end{aligned}$$

By the Liapounov CLT, $\Theta_{3n} - E(\Theta_{3n}) \to_d N[0, \sigma^2 \nu_2(K)]$. Combining this with (A.1)–(A.4) and noting that $n/(2mM) \to 1$ as $n \to \infty$, we obtain (2.5) and the proof is complete. □

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(a), A7 and A8, the result in Lemma A.1 changes to

$$\int u\left[L(2m)^{-1}\sum_{\alpha=-m}^{m} f(u, x-\mu_\alpha) - f_u(u)\right] du = O\left(m^{-2}\right). \tag{A.5}$$

To see this, noticing that $f\left(u, x - \frac{L}{2}\right) = f\left(u, x + \frac{(m+1)L}{2m}\right) = 0$ since there is zero density outside the support by A7 and A8, we can write $\frac{L}{2m}\sum_{\alpha=-m}^{m} f(u, x-\mu_\alpha) - f_u(u)$ as

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

Since $x + \frac{L}{2} > \overline{u}$ and $x - \frac{L}{2} < \underline{u}$, we have $I_{2m}(u) = \int_{x-L/2}^{x+L/2} f(u, u_x)\, du_x - \int_{\underline{u}}^{\overline{u}} f(u, u_x)\, du_x = 0$ by A7 and A8. By the Newton–Cotes formula for the trapezoidal rule of approximation (e.g. Stoer and Bulirsch, 1993, p. 162) and Assumption A3(a):

$$\begin{aligned}
(2m/L)^2\int_{-\infty}^{\infty} u I_{1m}(u)\, du &= \int_{-\infty}^{\infty}\frac{u}{12}\frac{L}{2m}\sum_{\alpha=-m}^{m} f_2''(u, x-p_\alpha)\, du \\
&\to \frac{1}{12}\int_{-\infty}^{\infty} u\int_{-L/2}^{L/2} f_2''(u, x-p)\, dp du \\
&= \frac{1}{12}\int_{-\infty}^{\infty}\int_{\underline{u}}^{\overline{u}} u f_2''(u, p)\, dp du,
\end{aligned}$$

where p_α lies between μ_α and $\mu_{\alpha+1}$. Thus (A.5) holds.

Next, one can show that the results in Lemmas A.2(a) and (c)–(d) hold with L_n replaced by L. The result in Lemma A.2(b) now becomes

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

where the dominant term vanishes if and only if $f_{u_x}(\overline{u})=f_{u_x}\left(\underline{u}\right)$. As to Θ_{3n} defined in the proof of Theorem 2.1, we have that for sufficiently large n

$$
\begin{aligned}
E\Theta_{3n} &= \frac{\sqrt{2Mmh/L}}{2m/L}\sum_{\alpha=-m}^{m} E\left[u_i K_h(\mu_\alpha+u_{xi}-x)\right] \\
&= \sqrt{2Mmh/L}K(z)\left[u\frac{1}{2m/L}\sum_{\alpha=-m}^{m} f(u,x-\mu_\alpha+hz)\,du\right]dz \\
&= \sqrt{2Mmh/L}K(z)\left[O\left(m^{-2}\right)+uf_u(u)\,du\right]dz \\
&= \sqrt{nh}\,O\left(m^{-2}\right)=o(1),
\end{aligned}
$$

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 L_n replaced by L everywhere. In addition, the Liapounov condition is also satisfied. This completes the proof of the theorem. □

Proof of Theorem 3.1: Noting that $\sqrt{n}(\widehat{\theta}-\theta)=\left(n^{-1}\mathbf{X}'\mathbf{X}\right)^{-1}n^{-1/2}\mathbf{X}'\mathbf{u}$, we have

$$
\sqrt{n}D_n\left[\widehat{\theta}-\theta-\left(\mathbf{X}'\mathbf{X}\right)^{-1}E\left(\mathbf{X}'\mathbf{u}\right)\right]=\left(D_n^{-1}n^{-1}\mathbf{X}'\mathbf{X}D_n^{-1}\right)^{-1}D_n^{-1}n^{-1/2}\left[\mathbf{X}'\mathbf{u}-E\left(\mathbf{X}'\mathbf{u}\right)\right].
$$

We prove the theorem by showing that (a) $\Gamma_n\equiv D_n^{-1}n^{-1}\mathbf{X}'\mathbf{X}D_n^{-1}\to_p\Gamma$, and (b) $A_n\equiv D_n^{-1}n^{-1/2}\left(\mathbf{Xu}-E\left(\mathbf{Xu}\right)\right)\to_d N(0,\Omega)$, where Γ and Ω are defined in (3.6).

First, by the fact that $\overline{X}=\overline{u}_x\to_p\mu_x$, and (3.3), we have $L_n^{-2}n^{-1}\sum_{i=1}^n X_i^2=L_n^{-2}n^{-1}\sum_{i=1}^n(X_i-\overline{X})^2+L_n^{-2}\overline{X}^2=\frac{1}{12}+\frac{\sigma_x^2}{L_n^2}+\frac{\mu_x^2}{L_n^2}+o_P(1)=\frac{1}{12}+\frac{E\left(u_{x1}^2\right)}{L_n^2}+o_P(1)$. Thus (a) follows. To show (b), by the Cramér–Wold device it suffices to show that for any $\omega=(\omega_1,\omega_2)'$ with $\|\omega\|=1$, we have

$$
\omega' A_n=n^{-1/2}\sum_{i=1}^{n}\left\{\omega_1 u_i+\omega_2 L_n^{-1}\left[X_i u_i-E\left(X_i u_i\right)\right]\right\}\to_d N\left(0,\omega'\Omega\omega\right). \tag{A.6}
$$

By construction, $E(\omega' A_n) = 0$. We now calculate the asymptotic variance of $\omega' A_n$:

$$\begin{aligned}
\operatorname{Var}(\omega' A_n) &= n^{-1} \sum_{i=1}^{n} \operatorname{Var}\left\{\omega_1 u_i + \omega_2 L_n^{-1}\left[X_i u_i - E\left(X_i u_i\right)\right]\right\} \\
&= \omega_1^2 \sigma^2 + 2\omega_1\omega_2 L_n^{-1} n^{-1} \sum_{i=1}^{n} E\left(X_i u_i^2\right) + \omega_2^2 L_n^{-2} n^{-1} \sum_{i=1}^{n} \operatorname{Var}\left(X_i u_i\right), \\
&\to \omega' \Omega \omega,
\end{aligned}$$

because $L_n^{-1} n^{-1} \sum_{i=1}^{n} E\left(X_i u_i^2\right) = L_n^{-1} E\left(u_{xi} u_i^2\right)$ and

$$\begin{aligned}
\frac{1}{n L_n^2} \sum_{i=1}^{n} \operatorname{Var}\left(X_i u_i\right) &= \frac{1}{(2m+1) L_n^2} \sum_{\alpha=-m}^{m} \frac{1}{M} \sum_{i \in A_\alpha} \operatorname{Var}\left[\left(\frac{L_n \alpha}{2m} + u_{xi}\right) u_i\right] \\
&= \frac{1}{(2m+1) L_n^2} \sum_{\alpha=-m}^{m} \frac{1}{M} \\
&\quad \times \sum_{i \in A_\alpha} \left\{\frac{L_n^2 \alpha^2}{4m^2} \sigma^2 + \operatorname{Var}(u_{xi} u_i) + \frac{L_n \alpha}{m} E\left[u_{xi} u_i^2\right]\right\} \\
&= \frac{\sigma^2}{(2m+1) 2m^2} \frac{m\,(m+1)\,(2m+1)}{6} + \frac{\operatorname{Var}(u_{xi} u_i)}{L_n^2} \\
&\to \frac{\sigma^2}{12} + \lim_{n\to\infty} \frac{\operatorname{Var}(u_{xi} u_i)}{L_n^2}.
\end{aligned}$$

Let $\xi_{in} = n^{-1/2}\left\{\omega_1 u_i + \omega_2 L_n^{-1}\left[X_i u_i - E\left(X_i u_i\right)\right]\right\}$. By the C_r inequality,

$$\begin{aligned}
\sum_{i=1}^{n} E\left|\xi_{in}\right|^{2+\delta} &= \frac{1}{n^{1+\delta/2}} \sum_{i=1}^{n} E\left|\omega_1 u_i + \omega_2 L_n^{-1}\left[X_i u_i - E\left(X_i u_i\right)\right]\right|^{2+\delta} \\
&\le \frac{2^{1+\delta}}{n^{1+\delta/2}} \sum_{i=1}^{n} \left(E\left|\omega_1 u_i\right|^{2+\delta} + \left|\omega_2 L_n^{-1}\right|^{2+\delta} E\left\|X_i u_i - E\left(X_i u_i\right)\right\|^{2+\delta}\right) \\
&\le \frac{2^{1+\delta}\left|\omega_1\right|^{2+\delta}}{n^{1+\delta/2}} \sum_{i=1}^{n} E\left|u_i\right|^{2+\delta} + \left|\omega_2 L_n^{-1}\right|^{2+\delta} \frac{2^{3+2\delta}}{n^{1+\delta/2}} \sum_{i=1}^{n} E\left\|X_i u_i\right\|^{2+\delta} \\
&= O(n^{-\delta/2}) = o(1).
\end{aligned}$$

Then (A.6) follows by the Liapounov CLT. □

For notational simplicity, let $K_{ix} \equiv K_h(X_i - x)$. Recall $b_n = c_n b$, $a_n = c_n a$, and in the special case where $L_n = L$ is fixed, $c_n = 1$. We first state two further lemmas whose proofs are available in the supplementary material for this paper cited earlier.

LEMMA A.3. *Let* $\Phi_{jn} \equiv \int_{a_n}^{b_n} x^j \widehat{f}(x) dx$ *for* $j = 0, 1, 2$. *Suppose the conditions in Theorem 3.2 hold. Then for* $j = 0, 1, 2$, *(a)* $E\Phi_{jn} = (b_n^{j+1} - a_n^{j+1})/(j+1) + O(h^2 c_n^{j+1})$, *(b)* $\operatorname{Var}\left(\Phi_{jn}\right) = O(L_n c_n^{2j+1}/n)$ *and (c)* $c_n^{-(j+1)} \Phi_{jn} = \left(b^{j+1} - a^{j+1}\right)/(j+1) + o(1)$.

LEMMA A.4. *Let* $\Xi_{jn} \equiv \int_{a_n}^{b_n} N^{-1/2} \sum_{i=1}^{n} x^j (X_i - x) K_{ix} dx$ *for* $j = 0, 1$. *Then* $\Xi_{jn} = O_P(h^2 \sqrt{n/L_n} c_n^{j+1} + h c_n^{j+1/2})$.

Proof of Theorem 3.2: To prove Theorem 3.2, noticing that if $g(x) = \beta_0 + \beta_1 x$, then we can write $\int_{a_n}^{b_n} \widehat{g}(x) \widehat{f}(x) dx$ as

$$\begin{aligned}
&\int_{a_n}^{b_n} N^{-1} \sum_{i=1}^{n} K_h(X_i - x) y_i dx \\
&= \int_{a_n}^{b_n} N^{-1} \sum_{i=1}^{n} K_{ix} \left[\beta_0 + \beta_1 x + \beta_1 (X_i - x) + u_i\right] dx \\
&= \beta_0 \int_{a_n}^{b_n} \widehat{f}(x) dx + \beta_1 \int_{a_n}^{b_n} x \widehat{f}(x) dx + \beta_1 \int_{a_n}^{b_n} N^{-1} \sum_{i=1}^{n} (X_i - x) K_{ix} dx \\
&\quad + \int_{a_n}^{b_n} N^{-1} \sum_{i=1}^{n} K_{ix} u_i dx,
\end{aligned}$$

and similarly $\int_{a_n}^{b_n} x \widehat{g}(x) \widehat{f}(x) dx = \beta_0 \int_{a_n}^{b_n} x \widehat{f}(x) dx + \beta_1 \int_{a_n}^{b_n} x^2 \widehat{f}(x) dx + \beta_1 \int_{a_n}^{b_n} N^{-1} \sum_{i=1}^{n} x(X_i - x) K_{ix} dx + \int_{a_n}^{b_n} N^{-1} \sum_{i=1}^{n} x K_{ix} u_i dx$. It follows that

$$\begin{aligned}
\sqrt{N} C_n \left(\widetilde{\theta} - \theta\right) &= \beta_1 \left(C_n^{-1} Q_n C_n^{-1}\right)^{-1} C_n^{-1} \begin{bmatrix} \int_{a_n}^{b_n} N^{-1/2} \sum_{i=1}^{n} (X_i - x) K_{ix} dx \\ \int_{a_n}^{b_n} N^{-1/2} \sum_{i=1}^{n} x (X_i - x) K_{ix} dx \end{bmatrix} \\
&\quad + \left(C_n^{-1} Q_n C_n^{-1}\right)^{-1} C_n^{-1} \begin{pmatrix} \int_{a_n}^{b_n} N^{-1/2} \sum_{i=1}^{n} K_{ix} u_i dx \\ \int_{a_n}^{b_n} N^{-1/2} \sum_{i=1}^{n} x K_{ix} u_i dx \end{pmatrix}. \qquad \text{(A.7)}
\end{aligned}$$

By Lemma A.3,

$$C_n^{-1} Q_n C_n^{-1} \to_p \begin{pmatrix} b - a & \dfrac{b^2 - a^2}{2} \\ \dfrac{b^2 - a^2}{2} & \dfrac{b^3 - a^3}{3} \end{pmatrix} = Q, \qquad \text{(A.8)}$$

where $\det(Q) = (b - a)^4/12 > 0$ as $b \neq a$. This, together with Lemma A.4, implies that the first term in (A.7) is $O_P(h^2 \sqrt{n/L_n} c_n^{1/2} + h) = o_P(1)$ by Assumption A6*. We are left to show that the second term in (A.7) is asymptotically $N(0, \sigma^2 Q^{-1})$.

Let $\omega \equiv (\omega_1, \omega_2)'$ be such that $\|\omega\| = 1$, and define

$$\begin{aligned}
\Theta_n &= \omega' C_n^{-1} \begin{pmatrix} \int_{a_n}^{b_n} N^{-1/2} \sum_{i=1}^{n} K_{ix} u_i dx \\ \int_{a_n}^{b_n} N^{-1/2} \sum_{i=1}^{n} x K_{ix} u_i dx \end{pmatrix} \\
&= N^{-1/2} \sum_{i=1}^{n} \int_{a_n}^{b_n} \left(c_n^{-1/2} \omega_1 + c_n^{-3/2} \omega_2 x\right) K_{ix} u_i dx.
\end{aligned}$$

We complete the proof by showing that $E(\Theta_n) = o(1)$ and

$$\Theta_n - E(\Theta_n) \to_d N\left(0, \sigma^2 \omega' Q \omega\right). \tag{A.9}$$

By a change of variables, the Fubini theorem, Lemma A.1 and Assumptions A2, A3(i) and A6*, we obtain

$$\begin{aligned}
E\Theta_n &= \sqrt{\frac{ML_n}{2m+1}} \sum_{\alpha=-m}^{m} \int_{a_n}^{b_n} E\left[\left(c_n^{-1/2}\omega_1 + c_n^{-3/2}\omega_2 x\right) K_h(\mu_\alpha + u_{x1} - x) u_1\right] dx \\
&= \sqrt{\frac{ML_n}{2m+1}} \sum_{\alpha=-m}^{m} \int_{a_n}^{b_n} \int\int \left(c_n^{-1/2}\omega_1 + c_n^{-3/2}\omega_2 x\right) uK(z) f(u, x - \mu_\alpha + hz)\, dz du dx \\
&= \sqrt{\frac{n}{L_n}} \int_{a_n}^{b_n} \left(c_n^{-1/2}\omega_1 + c_n^{-3/2}\omega_2 x\right) \int K(z) \\
&\quad \times \left[u \frac{L_n}{2m+1} \sum_{\alpha=-m}^{m} f(u, x - \mu_\alpha + hz)\, du\right] dz dx \\
&= \sqrt{N} \int_{a_n}^{b_n} \left(c_n^{-1/2}\omega_1 + c_n^{-3/2}\omega_2 x\right) \int K(z) \left\{O\left[(m/L_n)^{-2} + h^2\right] + uf(u)\, du\right\} dz dx \\
&= \sqrt{N} c_n^{1/2} O\left[(m/L_n)^{-2} + h^2\right] = o(1).
\end{aligned}$$

Next, letting $w_n(x) \equiv c_n^{-1/2}\omega_1 + c_n^{-3/2}\omega_2 x$, we have

$$\begin{aligned}
\operatorname{Var}(\Theta_n) &= N^{-1} \sum_{i=1}^{n} \operatorname{Var}\left[u_i \int_{a_n}^{b_n} \left(c_n^{-1/2}\omega_1 + c_n^{-3/2}\omega_2 x\right) K_{ix} dx\right] \\
&= \frac{L_n}{2m+1} \sum_{\alpha=-m}^{m} \int_{a_n}^{b_n} \int_{a_n}^{b_n} E[w_n(x) w_n(\widetilde{x}) u_1^2 K_h(\mu_\alpha + u_{x1} - x) \\
&\quad \times K_h(\mu_\alpha + u_{x1} - \widetilde{x})] dx d\widetilde{x} + o(1) \\
&= \frac{L_n}{2m+1} \sum_{\alpha=-m}^{m} \int_{a_n}^{b_n} \int_{a_n}^{b_n} w_n(x) w_n(\widetilde{x}) \int\int u^2 K_h(\mu_\alpha + u_x - x) K_h(\mu_\alpha + u_x - \widetilde{x}) \\
&\quad \times f(u, u_x)\, du_x du dx d\widetilde{x} + o(1) \\
&= \int_{a_n}^{b_n} \int_{a_n}^{b_n} w_n(x) w_n(\widetilde{x}) \int\int u^2 K(z) K\left(z + \frac{x - \widetilde{x}}{h}\right) \frac{L_n}{2m+1} \sum_{\alpha=-m}^{m} f(u, x - \mu_\alpha) \\
&\quad \times dz du dx d\widetilde{x} + o(1) \\
&= \sigma^2 \int_{a_n}^{b_n} \int_{a_n}^{b_n} w_n(x) w_n(\widetilde{x}) \int K(z) K\left(z + \frac{x - \widetilde{x}}{h}\right) dz\, dx d\widetilde{x} + o(1) \\
&= \sigma^2 \int_{a_n}^{b_n} \int_{(a_n - x)/h}^{(b_n - x)/h} w_n(x) w_n(x + hv) \int K(z) K(z - v)\, dz dv\, dx + o(1) \\
&= \sigma^2 \int_{a_n}^{b_n} \left(c_n^{-1/2}\omega_1 + c_n^{-3/2}\omega_2 x\right)^2 dx + o(1) \to \sigma^2 \omega' Q \omega,
\end{aligned}$$

where we have used the fact that $\frac{L_n}{2m} \sum_{\alpha=-m}^{m} f(u, x - \mu_\alpha) \to \int_{-\infty}^{\infty} f(u, x - p)\, dp = f(u)$. To show the asymptotic normality of $\Theta_n - E(\Theta_n)$, by the above variance calculation and the independence of (u_i, u_{xi}) across i, it suffices to check the Liapounov condition. Let $\overline{\xi}_{ni} = \xi_{ni} - E(\xi_{ni})$, where $\xi_{ni} = N^{-1/2} \int_{a_n}^{b_n} w_n(x) K_{ix} u_i dx$. Then by the C_r inequality,

$$\begin{aligned}\sum_{i=1}^{n} E \left| \overline{\xi}_{ni} \right|^{2+\delta} &\leq 2^{1+\delta} N^{-(1+\delta/2)} \sum_{i=1}^{n} E \left| u_i \int_{a_n}^{b_n} w_n(x) K_{ix} dx \right|^{2+\delta} \\ &\quad + 2^{1+\delta} N^{-(1+\delta/2)} \sum_{i=1}^{n} \left| E \left[u_i \int_{a_n}^{b_n} w_n(x) K_{ix} dx \right] \right|^{2+\delta} \\ &\equiv L_{n1} + L_{n2}, \text{ say.}\end{aligned}$$

First,

$$\begin{aligned}L_{n1} &= 2^{1+\delta} N^{-\delta/2} \frac{L_n}{2m+1} \sum_{\alpha=-m}^{m} \int \int \left| u \int_{a_n}^{b_n} w_n(x) K_h (\mu_\alpha + u_x - x) dx \right|^{2+\delta} f(u, u_x) du_x du \\ &\leq c_\delta N^{-\delta/2} \frac{L_n}{2m+1} \sum_{\alpha=-m}^{m} \int \int |u|^{2+\delta} \left| \int_{a_n}^{b_n} K_h (\mu_\alpha + u_x - x) dx \right|^{2+\delta} f(u, u_x) du_x du \\ &\leq c_\delta N^{-\delta/2} \frac{L_n}{2m+1} \sum_{\alpha=-m}^{m} \int \int |u|^{2+\delta} f(u, u_x) du_x du \\ &= c_\delta N^{-\delta/2} E |u|^{2+\delta} L_n \\ &= O(N^{-\delta/2} L_n) = o(1),\end{aligned}$$

where $c_\delta = 2^{1+\delta} \sup_{a_n \leq |x| \leq b_n} |w_n(x)|^{2+\delta} < \infty$. By the Jensen inequality, $L_{n2} \leq L_{n1} = o(1)$. Then by the Liapounov CLT (A.9) follows, and the proof is complete. □

Proof of Theorem 3.3: Let $B_n \equiv n^{-1}\mathbf{X}'\mathbf{X}$. Noting that

$$\frac{1}{n^{-1} \sum_{i=1}^{n} (X_i - \overline{X})^2} \begin{pmatrix} \overline{X}\, \widetilde{c} \\ -\widetilde{c} \end{pmatrix} = B_n^{-1} n^{-1} \sum_{i=1}^{n} \begin{pmatrix} 0 \\ -X_i \widetilde{u}_i \end{pmatrix},$$

we have

$$\begin{aligned}\sqrt{N} C_n \left(\widehat{\theta}_c - \theta \right) &= \sqrt{N} C_n \left(\widehat{\theta} - \theta \right) + C_n B_n^{-1} n^{-1} \sum_{i=1}^{n} \begin{pmatrix} 0 \\ -X_i \widetilde{u}_i \end{pmatrix} \\ &= \sqrt{N} C_n B_n^{-1} n^{-1} \sum_{i=1}^{n} \begin{bmatrix} u_i \\ \left(X_i, X_i^2 \right) (\widetilde{\theta} - \theta) \end{bmatrix}.\end{aligned}$$

Let $\omega \equiv (\omega_1, \omega_2)'$ with $\|\omega\| = 1$. Define $T_n \equiv \sqrt{N} \omega' C_n (\widehat{\theta}_c - \theta)$. By the Cramér–Wold device, it suffices to show that $T_n \to_d N(0, \omega' \Psi \omega)$. We prove this by distinguishing whether L_n is allowed to approach ∞ as $n \to \infty$.

Case 1. $L_n \to \infty$ as $n \to \infty$. By the proof of Theorem 3.2,

$$\begin{aligned}\sqrt{N}c_n^{3/2}(\widetilde{\beta}_1 - \beta_1) &= \left(q^{21}, q^{22}\right)\begin{pmatrix} c_n^{-1/2}\int_{a_n}^{b_n} N^{-1/2}\sum_{i=1}^{n} K_{ix}u_i dx \\ c_n^{-3/2}\int_{a_n}^{b_n} N^{-1/2}\sum_{i=1}^{n} xK_{ix}u_i dx \end{pmatrix} + o_P(1) \\ &= q^{21}c_n^{-1/2}\int_{a_n}^{b_n} N^{-1/2}\sum_{i=1}^{n} K_{ix}u_i dx \\ &\quad + q^{22}c_n^{-3/2}\int_{a_n}^{b_n} N^{-1/2}\sum_{i=1}^{n} xK_{ix}u_i dx + o_P(1),\end{aligned}$$

where for $i, j = 1, 2$, q^{ij} is the (i, j) element of Q^{-1}:

$$Q^{-1} = \frac{12}{(b-a)^4}\begin{pmatrix} \dfrac{b^3 - a^3}{3} & -\dfrac{b^2 - a^2}{2} \\ -\dfrac{b^2 - a^2}{2} & b - a \end{pmatrix} \equiv \begin{pmatrix} q^{11} & q^{12} \\ q^{21} & q^{22} \end{pmatrix}.$$

Noting that $\overline{X} = n^{-1}\sum_{i=1}^n u_{xi} \to_p \mu_x$, $L_n^{-2}n^{-1}\sum_{i=1}^n X_i^2 = L_n^{-2}n^{-1}\sum_{i=1}^n (X_i - \overline{X})^2 + L_n^{-2}\overline{X}^2 \to_p \frac{1}{12}$, we have $S_{xnc}^2/S_x^2 \to_p 1$ and $\overline{X}/S_x^2 \to_p 0$, where $S_x^2 = n^{-1}\sum_{i=1}^n (X_i - \overline{X})^2$ and $S_{xnc}^2 = n^{-1}\sum_{i=1}^n X_i^2$. Also note that $B_n^{-1} = \frac{1}{S_x^2}\begin{pmatrix} S_{xnc}^2 & -\overline{X} \\ -\overline{X} & 1 \end{pmatrix}$. It follows that

$$\begin{aligned} T_n &= \sqrt{N}n^{-1}\omega' C_n \sum_{i=1}^n \frac{1}{S_x^2}\begin{bmatrix} S_{xnc}^2 u_i - \overline{X}\left(X_i, X_i^2\right)(\widetilde{\theta} - \theta) \\ -\overline{X}u_i + \left(X_i, X_i^2\right)(\widetilde{\theta} - \theta) \end{bmatrix} \\ &= \sqrt{N}n^{-1}\sum_{i=1}^n \frac{1}{S_x^2}\left[\left(\omega_2 c_n^{3/2} - \omega_1 c_n^{1/2}\overline{X}\right)\left(X_i, X_i^2\right)(\widetilde{\theta} - \theta) + \left(\omega_1 c_n^{1/2} S_{xnc}^2 - \omega_2 c_n^{3/2}\overline{X}\right)u_i\right] \\ &= \frac{\omega_2 c_n^{3/2} - \omega_1 c_n^{1/2}\overline{X}}{S_x^2}\left(c_n^{-1/2}\overline{X},\ c_n^{-3/2}S_{xnc}^2\right)\sqrt{N}C_n(\widetilde{\theta} - \theta) \\ &\quad + L_n^{-1/2}\frac{1}{S_x^2}\left(\omega_1 c_n^{1/2}S_{xnc}^2 - \omega_2 c_n^{3/2}\overline{X}\right)n^{-1/2}\sum_{i=1}^n u_i \\ &= \left(0, \omega_2 - c_n^{-1}\omega_1\mu_x\right)\sqrt{N}C_n(\widetilde{\theta} - \theta) + \omega_1\left(c_n/L_n\right)^{1/2}n^{-1/2}\sum_{i=1}^n u_i + o_p(1) \\ &= \left(\omega_2 - c_n^{-1}\omega_1\mu_x\right)\sqrt{N}c_n^{3/2}(\widetilde{\beta}_1 - \beta_1) + \omega_1\left(c_n/L_n\right)^{1/2}n^{-1/2}\sum_{i=1}^n u_i + o_p(1) \\ &= \left(\omega_2 - c_n^{-1}\omega_1\mu_x\right) \end{aligned}$$

$$\times \left(q^{21} c_n^{-1/2} N^{-1/2} \sum_{i=1}^{n} \int_{a_n}^{b_n} K_{ix} u_i dx + q^{22} c_n^{-3/2} N^{-1/2} \sum_{i=1}^{n} \int_{a_n}^{b_n} x K_{ix} u_i dx \right)$$

$$+ \omega_1 (c_n/L_n)^{1/2} n^{-1/2} \sum_{i=1}^{n} u_i + o_p(1)$$

$$\to_d N\left[0, \sigma^2 \lim_{n\to\infty} \{\left(\omega_2 - c_n^{-1}\omega_1\mu_x\right)^2 q^{22} + \omega_1^2 (c_n/L_n) + 2\left(\omega_2 - c_n^{-1}\omega_1\mu_x\right)\omega_1 (c_n/L_n)\gamma\right]$$

$$= N\left[0, \sigma^2 \lim_{n\to\infty} \omega' \begin{pmatrix} c_n^{-2}\mu_x^2 q^{22} + \dfrac{c_n}{L_n} & -c_n^{-1}\mu_x q^{22} + \dfrac{c_n}{L_n}\gamma \\ -c_n^{-1}\mu_x q^{22} + \dfrac{c_n}{L_n}\gamma & q^{22} \end{pmatrix} \omega\right],$$

where $\gamma = q^{21}(b-a) + \frac{1}{2}q^{22}\left(b^2 - a^2\right)$.

Case 2. $L_n = L$ is fixed as $n \to \infty$. In this case, $c_n = 1$ and we can write a_n and b_n as a and b throughout. By the proof of Theorem 3.2 (see (A.7) and arguments thereafter),

$$\sqrt{N}\left(\widetilde{\theta} - \theta\right) = Q^{-1} N^{-1/2} \sum_{i=1}^{n} \begin{bmatrix} \int_a^b \left[K_{ix}u_i - E\left(K_{ix}u_i\right)\right] dx \\ \int_a^b x\left[K_{ix}u_i - E\left(K_{ix}u_i\right)\right] dx \end{bmatrix} + o_P(1)$$

$$\equiv \sum_{i=1}^{n} \begin{pmatrix} \xi_{i1} \\ \xi_{i2} \end{pmatrix} + o_P(1),$$

where

$$\xi_{i1} \equiv N^{-1/2} \left\{ q^{11} \int_a^b \left[K_{ix}u_i - E\left(K_{ix}u_i\right)\right] dx + q^{12} \int_a^b x\left[K_{ix}u_i - E\left(K_{ix}u_i\right)\right] dx \right\} \qquad \text{(A.10)}$$

and

$$\xi_{i2} \equiv N^{-1/2} \left\{ q^{21} \int_a^b \left[K_{ix}u_i - E\left(K_{ix}u_i\right)\right] dx + q^{22} \int_a^b x\left[K_{ix}u_i - E\left(K_{ix}u_i\right)\right] dx \right\}. \qquad \text{(A.11)}$$

Let

$$R_n \equiv \sqrt{N} n^{-1} \sum_{i=1}^{n} \begin{bmatrix} u_i \\ \left(X_i, X_i^2\right)\left(\widetilde{\theta} - \theta\right) \end{bmatrix}.$$

Then

$$\omega' R_n = \omega_2 \sqrt{N}\left(\widetilde{\beta}_1 - \beta_1\right)\left\{ n^{-1} \sum_{i=1}^{n} X_i^2 \right\} + \omega_2 \sqrt{N}\left(\widetilde{\beta}_0 - \beta_0\right)\left\{ n^{-1} \sum_{i=1}^{n} X_i \right\}$$

$$+ \omega_1 \sqrt{N} n^{-1} \sum_{i=1}^{n} u_i$$

$$= \omega_2 c_x \sqrt{N}\left(\widetilde{\beta}_1 - \beta_1\right) + \omega_2 \mu_x \sqrt{N}\left(\widetilde{\beta}_0 - \beta_0\right) + \omega_1 \sqrt{N} n^{-1} \sum_{i=1}^{n} u_i + o_P(1)$$

$$= \sum_{i=1}^{n} \left(\omega_2 c_x \xi_{i2} + \omega_2 \mu_x \xi_{i1} + \omega_1 \sqrt{N} n^{-1} u_i \right) + o_P(1)$$

$$\equiv \overline{R}_n + o_P(1),$$

where recall $c_x \equiv \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n} E\left(X_i^2\right) = \frac{L^2}{12} + E(u_{xi}^2)$. Let $c \equiv (\mu_x, c_x)'$. Note that $E\left(\overline{R}_n\right) = 0$, and

$$\begin{aligned}
\mathrm{Var}\left(\overline{R}_n\right) &= \sum_{i=1}^{n} \mathrm{Var}\left(\omega_2 c_x \xi_{i2} + \omega_2 \mu_x \xi_{i1} + \omega_1 \sqrt{N} n^{-1} u_i\right) \\
&= \omega_2^2 \sum_{i=1}^{n} \mathrm{Var}\left(c_x \xi_{i2} + \mu_x \xi_{i1}\right) + 2\omega_2\omega_1 \sum_{i=1}^{n} \mathrm{Cov}\left(c_x \xi_{i2} + \mu_x \xi_{i1}, \sqrt{N} n^{-1} u_i\right) \\
&\quad + \omega_1^2 \sum_{i=1}^{n} \mathrm{Var}\left(\sqrt{N} n^{-1} u_i\right) \\
&= \sigma^2 \omega_2^2 c' Q^{-1} c \\
&\quad + 2\omega_2\omega_1 \sigma^2 L^{-1} \left\{ q^{11} \int_a^b dx + q^{12} \int_a^b x dx + q^{21} \int_a^b dx + q^{22} \int_a^b x dx \right\} \\
&\quad + \omega_1^2 \sigma^2 L^{-1} + o(1) \\
&\to \sigma^2 \omega_2^2 c' Q^{-1} c + 2\omega_2\omega_1 \sigma^2 L^{-1} + \omega_1^2 \sigma^2 L^{-1} \\
&= \sigma^2 \omega' \begin{pmatrix} L^{-1} & L^{-1} \\ L^{-1} & c' Q^{-1} c \end{pmatrix} \omega,
\end{aligned}$$

where the third line follows from the definitions of ξ_{i1} and ξ_{i2}, Fubini, a change of variables, and the following two limits:

$$\begin{aligned}
&\sum_{i=1}^{n} E\left(\xi_{i1} \sqrt{N} n^{-1} u_i\right) \\
&= n^{-1} \sum_{i=1}^{n} E\left\{ \left[q^{11} \int_a^b \left[K_{ix} u_i - E\left(K_{ix} u_i\right)\right] dx + q^{12} \int_a^b x \left[K_{ix} u_i - E\left(K_{ix} u_i\right)\right] dx \right] u_i \right\} \\
&= q^{11} n^{-1} \sum_{i=1}^{n} E\left\{ u_i^2 \int_a^b K_{ix} dx \right\} + q^{12} n^{-1} \sum_{i=1}^{n} E\left\{ u_i^2 \int_a^b x K_{ix} dx \right\} \\
&= q^{11} (2m+1)^{-1} \sum_{\alpha=-m}^{m} \int_a^b \int\int u^2 K(z) f(u, x - u_\alpha + hz) dz du dx \\
&\quad + q^{12} (2m+1)^{-1} \sum_{\alpha=-m}^{m} \int_a^b x \int\int u^2 K(z) f(u, x - u_\alpha + hz) dz du dx \\
&= q^{11} L^{-1} \int_a^b \int u^2 \left\{ \frac{L}{2m} \sum_{\alpha=-m}^{m} f(u, x - u_\alpha) \right\} du dx \{1 + o(1)\} \\
&\quad + q^{12} L^{-1} \int_a^b x \int u^2 \left\{ \frac{L}{2m} \sum_{\alpha=-m}^{m} f(u, x - u_\alpha) \right\} dx \{1 + o(1)\} \\
&= \sigma^2 q^{11} L^{-1} \int_a^b dx \{1 + o(1)\} + \sigma^2 q^{12} L^{-1} \int_a^b x dx \{1 + o(1)\} \\
&\to \sigma^2 L^{-1} \left\{ q^{11} \int_a^b dx + q^{12} \int_a^b x dx \right\};
\end{aligned}$$

and, in a similar way,

$$\sum_{i=1}^{n} E\left(\xi_{i2}\sqrt{N}n^{-1}u_i\right) = n^{-1}\sum_{i=1}^{n} E\left[q^{21}u_i^2\int_a^b K_{ix}u_i dx + q^{22}u_i^2\int_a^b xK_{ix}dx\right]$$
$$\to \sigma^2 L^{-1}\left\{q^{21}\int_a^b dx + q^{22}\int_a^b x dx\right\}.$$

The Liapounov condition follows from the verification in the proof of Theorem 3.2, the fact that $\sqrt{N}n^{-1}\sum_{i=1}^{n} u_i$ also satisfies the Liapounov condition, and the C_r inequality. It follows that

$$R_n \to_d N\left[0, \sigma^2\begin{pmatrix} L^{-1} & L^{-1} \\ L^{-1} & c'Q^{-1}c \end{pmatrix}\right]$$

and

$$T_n \to_d N\left[0, \sigma^2\omega' B^{-1}\begin{pmatrix} L^{-1} & L^{-1} \\ L^{-1} & c'Q^{-1}c \end{pmatrix}B^{-1}\omega\right]. \qquad \square$$