

Statistical Inference on Regression with Spatial Dependence

Peter M Robinson* and Supachoke Thawornkaiwong
London School of Economics

April 29, 2010

Abstract

Central limit theorems are developed for instrumental variables estimates of linear and semi-parametric partly linear regression models for spatial data. General forms of spatial dependence and heterogeneity in explanatory variables and unobservable disturbances are permitted. We discuss estimation of the variance matrix, including estimates that are robust to disturbance heteroscedasticity and/or dependence. A Monte Carlo study of finite-sample performance is included. In an empirical example, the estimates and robust and non-robust standard errors are computed from Indian regional data, following tests for spatial correlation in disturbances, and nonparametric regression fitting. Some final comments discuss modifications and extensions.

JEL Classifications: C13; C14; C21

Keywords: Linear regression; Partly linear regression; Nonparametric regression; Spatial data; Instrumental variables; Asymptotic normality; Variance estimation.

*Corresponding author. Tel.: +44-20-7955-7516; fax: +44-20-7955-6592. *E-mail address:* p.m.robinson@lse.ac.uk

1 Introduction

The linear regression model, with estimation by ordinary least squares (LS) or instrumental variables (IV), is still a very popular statistical tool in empirical economic investigation. Often, however, the linearity seems an arbitrary restriction, while no specific nonlinear-in-parameters model is supported by economic theory. On the other hand, smoothed nonparametric regression encounters the curse of dimensionality unless very few explanatory variables are relevant or a huge sample is available. As a result, semiparametric models, such as partly linear regression, have been employed. For example, Robinson (1988) proposed estimates of the coefficients of the linear component of a partly linear regression and showed that they can compete with estimates of purely parametric models by converging at parametric rate and being asymptotically normal, in the setting of arbitrarily many stochastic explanatory variables in both the parametric and nonparametric parts. He assumed that observations are independent and identically distributed (i.i.d.). This is often questionable in economic applications, in particular, spatial dependence may arise from local shocks in an economy and interaction among economic agents, due for example to spill-overs, competition and externalities; Conley (1999) discussed in detail sources of spatial dependence, from both theoretical and empirical perspectives. The setting of the present paper is motivated by spatial dependence in general, but also covers, as a special case, time dependence, whose implications have already been widely studied in the parametric regression context, and to a much more limited extent (e.g. Fan and Li (1999)) in the partly linear context, but on the other hand our conditions also cover time dependence in panel data or spatio-temporal data settings.

Spatial dependence can arise in many forms of data, for example (equally-spaced) data observed on a regular lattice of two or more dimensions, data observed with irregular spacing on a geographic space, data for which only pairwise "economic distances" are available, and cross-sectional data that are feared to be dependent but for which no distance measures are postulated. Asymptotic statistical properties of estimates, such as of LS and IV estimates of linear regression, and estimates for the partly linear model, have not yet been developed under conditions that satisfactorily cover these possibilities. In an important class of cases, unobservable disturbances are i.i.d., and here the asymptotic distribution is expected, under suitable regularity conditions, to be unaffected, leaving intact rules of large sample inference. In other cases, disturbances will be mutually independent but conditionally or unconditionally heteroskedastic, where the asymptotic variance matrix is affected, so standard t-tests and interval estimates are invalidated, and Gauss-Markov efficiency properties (in case of LS regression estimates), or the achievement of a semiparametric efficiency bound (in case of Robinson's (1988) estimates of partly linear regression) are lost. The same is true when, on the other hand, homoskedasticity in disturbances is retained but independence is lost, and *a fortiori* when disturbances are both heteroskedastic and dependent. A desirable solution would entail correcting for whichever problem is present, using generalized least squares (GLS) ideas, as has been frequently done in dealing with heteroskedasticity, and also with time series dependence, and occasionally even with both problems simultaneously (see Hidalgo (1992)). It is relatively easy to see how to construct GLS estimates when dependence can be accurately parametrically modelled, but matters become more complicated in the more modern approach where disturbance correlation is treated as nonparametric, and certainly more consideration has to be given to the possible structure of dependence, reflecting the particular nature of the data, than in simple point estimates which ignore the problem. Moreover, if we begin from a situation in which correlation between regressors and disturbances is also feared, leading to use of instrumental variables, efficiency improvements are still harder to achieve.

In the setting of random design nonparametric regression, Robinson (2007b) proposed a triangular array structure which he justified as a possible representation for a broad class of spatial configurations, and presented conditions for consistency and asymptotic normality of Nadarya-Watson estimates. Disturbances were assumed to satisfy a kind of linear process, possibly allowing also for conditional or unconditional heteroscedasticity, and restrictions on dependence of regressors were expressed in terms of conditions on joint and marginal probability density functions, again also permitting some hetero-

generality. It was argued that these kinds of conditions might be suited to a wide range of spatial data.

We employ similar conditions here, in order to establish asymptotic normality of IV (and thus also LS) estimates of a linear regression (see the following section), and of (density-weighted IV) estimates of a partly linear model (see Section 3), allowing in both cases for spatial dependence in regressors and disturbances. Proofs of these results are left to three appendices, the first presenting the main steps, the second a sequence of propositions, and the third, technical lemmas. Section 4 discusses estimation of relevant large sample covariance matrices, some of which allow for disturbance heteroscedasticity and/or dependence, and thus provide robust inference, with the proof of a theorem contained in the fourth appendix. In an empirical study in Section 6, we develop the regression analysis of Banerjee and Iyer (2005) of the effect of systems for collecting land revenue instituted during British rule in India on present-day economic performance, after first finding evidence of spatial correlation of disturbances and carrying out nonparametric regression fitting. Sections 5 and 6 also include some discussion of the issue of bandwidth choice in partly linear regression. Section 7 discusses related aspects and possible modifications and extensions of our methods and theory.

2 Linear Regression

Given n observations on the p -dimensional column vector random variable X_{1in} and scalar random variable Y_{in} , we consider the linear regression

$$Y_{in} = \beta' X_{1in} + U_{in}, \quad 1 \leq i \leq n, \quad n = 1, 2, \dots, \quad (1)$$

where the p -dimensional column vector β is unknown, the prime denotes transposition, and the U_{in} are unobservable scalar disturbances. It is possible that X_{1in} includes an intercept. For spatial data there is generally no natural ordering, but an arbitrary one is employed in (1). The triangular array formulation, indicated by the n subscript, is used because some re-ordering may be natural when n increases, as discussed by Robinson (2007b), for example when observation points form a lattice in two or more dimensions. It is also essential when a variable is believed to be generated by a model such as a spatial autoregression (SAR) with row-normalized weight matrix. However, to avoid complicated notation we will mostly suppress reference to the n subscript in what follows, so in particular we write $U_i = U_{in}$, $X_{1i} = X_{1in}$, $Y_i = Y_{in}$, though from time to time we take the opportunity to remind the reader of the underlying potential dependence on n of various quantities.

Consider the IV estimate $\tilde{\beta} = \tilde{\beta}_n$ of β , given by

$$\tilde{\beta} = \left(\sum_{i=1}^n X_{2i} X_{1i}' \right)^{-1} \sum_{i=1}^n X_{2i} Y_i,$$

assuming we observe also the p -dimensional column vector random variable $X_{2i} = X_{2in}$ and the inverse exists. As usual X_{1i} and X_{2i} may overlap and $X_{2i} = X_{1i}$ is possible, when $\tilde{\beta}$ becomes LS, but IV estimation is as usual motivated by the fear of correlation between one or more elements of X_{1i} and U_i , and the hope of orthogonality between X_{2i} and U_i , and correlation between X_{1i} and X_{2i} .

We introduce the following assumptions, where the norm $\|A\|$ of a rectangular matrix A is defined as the square root of the trace of $A'A$, and C denotes a generic, finite constant, independent of n .

Assumption A1 (1) holds where

$$U_i = U_{in} = \sum_{k=1}^{\infty} b_{ik} \varepsilon_k, \quad 1 \leq i \leq n, \quad n = 1, 2, \dots, \quad (2)$$

where ε_k , $k = 1, 2, \dots$, are independent scalar random variables with zero mean and unit variance, and the scalar weights $b_{ik} = b_{ikn}$ satisfy

$$\sum_{k=1}^{\infty} b_{ik}^2 \leq C, \quad 1 \leq i \leq n, \quad n = 1, 2, \dots \quad (3)$$

Assumption A2 As $n \rightarrow \infty$,

$$\begin{aligned} \sum_{i=1}^n \mathbb{E} \|X_{2i}\|^2 &= O(n), \\ n^{-1} \sum_{i=1}^n X_{2i} X'_{1i} &\rightarrow_p \Phi, \end{aligned}$$

where Φ is a constant non-singular matrix.

Assumption A3 Denoting by \mathbb{N} the set of positive integers,

$$\lim_{\delta \rightarrow \infty} \sup_{k \in \mathbb{N}} \mathbb{E} \{ \varepsilon_k^2 1(|\varepsilon_k| > \delta) \} = 0.$$

We abbreviate the triangular array or sequence $\{b_i = b_{in}; 1 \leq i \leq n, n \geq 1\}$ to $\{b_i\}$.

Assumption A4 $\{X_{2i}\}$ and $\{\varepsilon_i\}$ are independent, and as $n \rightarrow \infty$,

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^{\infty} b_{ik} b_{jk} X_{2i} X'_{2j} \rightarrow_p \Sigma,$$

where Σ is positive definite (p.d.) and

$$n^{-1/2} \sup_{k \in \mathbb{N}} \left\| \sum_{i=1}^n X_{2i} b_{ik} \right\| \rightarrow_p 0. \quad (4)$$

Theorem A Under Assumptions A1-A4, as $n \rightarrow \infty$,

$$\sqrt{n} (\tilde{\beta} - \beta) \rightarrow_d \mathbf{N}(0, \Phi^{-1} \Sigma \Phi^{-1}).$$

Robinson (2007b) gave detailed motivation for using (2) and (3) to derive central limit theorems in the presence of spatial correlation and heterogeneity. Most basically, they imply that $\max_{1 \leq i \leq n} E(U_i^2) \leq C$. They also extend the kind of linear process used when the U_i form a stationary time series, and $b_{ij} = b_{i-j}$. The more general ij subscript conveys possible heterogeneity as well as correlation, and this and the suppressed n subscript on b_{ij} are required to cover models such as the SAR (which is nonstationary). In the SAR model for U_i the b_{ij} eventually vanish, for all i ($b_{ij} = 0$ for $j > n$), and (3) is satisfied under standard conditions, but it also covers infinite-order dependence, familiar from time series and lattice autoregressive and autoregressive moving average models. In these, the b_{ij} are absolutely summable, but (3) covers also possible "long memory". However, the extent to which this is possible depends also on the dependence within $\{X_{2i}\}$. As noted in the time series case by Robinson and

Hidalgo (1997), root- n -consistency is only possible if the collective memory in U_i and X_{2i} is sufficiently weak. In particular if X_{2i} includes an intercept, the first limit in Assumption A4 (which merely asserts convergence of the covariance matrix of $n^{-1/2} \sum_{i=1}^n X_{2i} U_i$) rules out long memory in U_i . However if (1) is reformulated in terms of mean-corrected observables long memory in U_i might be permitted in a corresponding central limit theorem for slope parameter estimates based on Assumption A1, cf Robinson and Hidalgo (1997). Independence of innovations (in Assumption A1) is standardly assumed both in models of SAR type and in lattice extensions of linear time series models; the martingale difference assumptions of time series models are hard to extend as there is no natural ordering to our data. Independence of $\{X_{2i}\}$ and $\{\varepsilon_i\}$ is a strong assumption and would be capable of some relaxation, but at a cost because our decoupling of conditions on disturbances and explanatory variables, here and even more so with respect to the partly linear model of the following section, has advantages, as discussed in Robinson (2007b). Assumption (4) is the required version of the asymptotic-negligibility condition to satisfy a Lindeberg condition. Note that if the U_i are uncorrelated, as implied when $b_{ik} = 0$ for $i \neq k$, (4) reduces to $n^{-\frac{1}{2}} \max_{1 \leq i \leq n} \|X_{2i}\| \rightarrow_p 0$, which, given the standard Assumption A2, is implied by the more familiar-looking condition $\max_{1 \leq i \leq n} \|X_{2i}\| / \left(\sum_{i=1}^n \|X_{2i}\|^2 \right)^{1/2} \rightarrow_p 0$. But the same conclusion results also under fairly general dependence in U_i . In particular this is the case if $\sum_{i=1}^n |b_{ik}| \leq C$ for all k , as is true if $|b_{ik}| \leq C |b_{i-k}|$ where $\sum_{i=-\infty}^{\infty} |b_i| < \infty$, to connect with weakly dependent stationary time series, or under an analogous condition relating to lattice processes. It is also the case with SAR models under normalization conditions. However, (4) is also true under more general dependence conditions, in particular if X_{2i} is uniformly bounded in probability it is only required that $\sup_{k \in \mathbb{N}} \sum_{i=1}^n |b_{ik}| = o\left(n^{\frac{1}{2}}\right)$, which for stationary time series and lattice data would permit long memory in U_{in} . Assumption A3 is just a standard uniform integrability requirement, avoiding identity of distribution.

3 Partly Linear Regression

Consider now the partly linear regression

$$Y_i = \beta' X_{1i} + \theta(Z_i) + U_i, \quad 1 \leq i \leq n, \quad (5)$$

where to extend the previous definitions $Z_i = Z_{in}$ is a q -dimensional observable column vector random variable, and θ is an unknown, nonparametric, function. As discussed by Robinson (1988), for identifiability X_{1i} cannot include an intercept and X_{1i} , Z_i cannot overlap.

We again focus on estimating β . As in Robinson (1988), we employ Nadaraya-Watson nonparametric regression estimation in estimating a transformed version of (5). Letting $k : \mathbb{R} \rightarrow \mathbb{R}$ be an even function, consider a product kernel $K : \mathbb{R}^q \rightarrow \mathbb{R}$ such that

$$K(z) = \prod_{t=1}^q k(z_t),$$

where z_t is the t -th element of z . For a positive scalar bandwidth sequence $a = a_n$, tending to zero as $n \rightarrow \infty$, denote

$$K_{ij} = K_{ijn} = K\left(\frac{Z_j - Z_i}{a}\right).$$

For a column vector triangular array $\{A_i = A_{in}\}$, define

$$A_i^* = A_{in}^* = \frac{1}{na^q} \sum_{j=1}^n (A_i - A_j) K_{ij},$$

and with $\{B_i = B_{in}\}$ also a column vector triangular array, define

$$S_{AB} = \frac{1}{n} \sum_{i=1}^n A_i^* B_i^{*'}.$$

Our semiparametric IV estimate of β is

$$\hat{\beta} = S_{X_2 X_1}^{-1} S_{X_2 Y},$$

assuming existence of the inverse. This is a density-weighted (as in Fan and Li (1999)) IV version of the estimate of Robinson (1998). For independent and homoskedastic U_i , Chamberlain (1992) showed that the latter estimate achieves a semiparametric efficiency bound. However, with spatial dependence in $\{U_i\}$, this property is lost, and without suitable spatial dependence structure, GLS-type estimation is ruled out. Because neither the estimate in Robinson (1988) nor the density-weighted version is efficient, and the former need not in general be the more efficient of the two, the latter may be preferable since the trimming in Robinson (1988) can thereby be avoided. However as in that reference, we still need to sufficiently reduce bias so as to obtain root- n -consistency in the presence of an arbitrarily high dimension of the vector Z_i , and this is achieved by employing a kernel k of suitably high order, and a corresponding degree of smoothness in the functions to be estimated. To describe these features we introduce the following definitions.

Definition 1 \mathcal{K}_l , $l \geq 1$, is the class of bounded and even functions $k : \mathbb{R} \rightarrow \mathbb{R}$ such that

$$\begin{aligned} \int_{\mathbb{R}} u^i k(u) du &= \delta_{i0}, \quad i = 0, \dots, l-1, \\ k(u) &= O\left(\left(1 + |u|^\zeta\right)^{-1}\right), \end{aligned}$$

as $|u| \rightarrow \infty$, where δ_{ij} is the Kronecker delta and $\zeta > \max(l+1, 2q)$.

Definition 2 A function $g : \mathbb{R}^q \rightarrow \mathbb{R}$ is in the class \mathcal{G}_μ^α , $\alpha > 0$, $\mu > 0$ (with respect to the triangular array $\{Z_i\}$) if: (i) g is $(m-1)$ -times partially differentiable, for $m-1 \leq \mu \leq m$; (ii) for some $\rho > 0$, $\sup_{y \in B(z, \rho)} |g(y) - g(z) - Q(y, z)| / \|y - z\|^\mu \leq h(z)$ for all z , where $B(z, \rho) = \{y : 0 < \|y - z\| < \rho\}$; $Q = 0$ when $m = 1$; (iii) Q is a $(m-1)$ -th degree homogeneous polynomial in $y - z$ with coefficients the partial derivatives of g at z of orders 1 through $m-1$ when $m > 1$; and (iv) $g(z)$, its partial derivatives of order $m-1$ and less, and $h(z)$, have average α th moments (averaged over Z_i , $1 \leq i \leq n$) that are uniformly bounded for all sufficiently large n .

We introduce the following assumptions.

Assumption B1 Assumption A1 holds with (1) replaced by (5).

Assumption B2 $\{\varepsilon_i\}$ is independent of $\{X_{2i}, Z_i\}$ and Assumption A3 holds.

Assumption B3 The following probability densities exist and have unbounded support: $f_i = f_{in}$, the density function of Z_i ; $f_{ij} = f_{ijn}$, the joint density function of Z_i and Z_j ; $f_{ijk} = f_{ijkn}$, the joint density function of Z_i , Z_j , and Z_k ; and $f_{ijkl} = f_{ijkln}$, the joint density function of Z_i , Z_j , Z_k and Z_l .

Assumption B4 For all $n \geq 1$ and $1 \leq i \leq n$,

$$X_{ti} = \xi_t(Z_i) + V_{ti}, \quad t = 1, 2,$$

where $V_{ti} = V_{tin}$ are p -dimensional column vector random variables such that for $t = 1, 2$,

$$\mathbb{E}(V_{ti} | Z_1, \dots, Z_n) = 0$$

and there exist functions $\sigma_t : R^q \times R^q \rightarrow R$ such that

$$\mathbb{E}(V_{ti}' V_{tj} | \{Z_1, \dots, Z_n\}) = \sigma_t(Z_i, Z_j) \gamma_{ij}^{(t)},$$

where $\gamma_{ij}^{(t)} = \gamma_{ijn}^{(t)} = \mathbb{E}(V_{ti}' V_{tj})$.

Define

$$\bar{f}(z) = \bar{f}_n(z) = \frac{1}{n} \sum_{i=1}^n f_i(z), \quad \bar{f}_i = \bar{f}(Z_i),$$

and

$$\gamma_{ij}^{(U)} = \gamma_{ijn}^{(U)} = E(U_i U_j).$$

Assumption B5 As $n \rightarrow \infty$,

$$n^{-1} \sum_{i=1}^n V_{2i} V_{1i}' \bar{f}_i^2 \rightarrow_p \Psi,$$

where Ψ is a constant non-singular matrix and

$$\max_{1 \leq i, j \leq n} |\gamma_{ij}^{(t)}| \leq C, \quad t = 1, 2,$$

$$\sum_{i,j=1}^n |\gamma_{ij}^{(1)}| = o(n^2), \quad \sum_{i,j=1}^n \left\{ |\gamma_{ij}^{(2)}| + |\gamma_{ij}^{(U)}| \right\} = o(n^{3/2}), \quad \text{as } n \rightarrow \infty.$$

Introduce the notation

$$\sum_{i_1, \dots, i_s}^n = \sum_{i_1=1}^n \sum_{i_2 \neq i_1}^n \cdots \sum_{i_s \neq i_1, \dots, i_s \neq i_{s-1}}^n.$$

Also introduce the dependence measures

$$\begin{aligned} F_{j:i}(z_2; z_1) &= f_{ij}(z_1, z_2) - f_i(z_1) f_j(z_2), \\ F_{jk:i}(z_2, z_3; z_1) &= f_{ijk}(z_1, z_2, z_3) - f_i(z_1) f_{jk}(z_2, z_3), \\ F_{ij:k:l}(z_1, z_2; z_3; z_4) &= f_{ijkl}(z_1, z_2, z_3, z_4) - f_{ij}(z_1, z_2) f_k(z_3) f_l(z_4). \end{aligned}$$

Assumption B6 For some $\varepsilon > 0$, $\{Z_i\}$ satisfies the following conditions as $n \rightarrow \infty$:

(i) denoting $B = B_n = \{z : \bar{f}(z) > 0\}$, $\mathcal{N}(z) = \{z_1 : \|z_1 - z\| < \varepsilon\}$,

$$\begin{aligned} \sup_{z_1 \in B} \sup_{z_2 \in \mathcal{N}(z_1)} \left\{ \frac{1}{\bar{f}(z_1)} \sum_{i,j}^n |F_{j:i}(z_2; z_1)| \right\} &= o(n^{3/2}), \\ \sup_{z_1 \in B} \sup_{z_2, z_3 \in \mathcal{N}(z_1)} \left\{ \frac{1}{\bar{f}(z_1)} \sum_{i,j,k}^n |F_{jk:i}(z_2, z_3; z_1)| \right\} &= o(n^{5/2}); \end{aligned}$$

(ii)

$$\begin{aligned} \sup_{z_1, z_2 \in \mathbb{R}^q} \sup_{z_3 \in \mathcal{N}(z_1) \cup \mathcal{N}(z_2)} \sum_{i,j,k}^n \left| \gamma_{ij}^{(U)} \gamma_{ij}^{(2)} F_{ij:k}(z_1, z_2; z_3) \right| &= o(n^2), \\ \sup_{z_1, z_2 \in \mathbb{R}^q} \sup_{z_3 \in \mathcal{N}(z_1), z_4 \in \mathcal{N}(z_2)} \left| \sum_{i,j,k,l}^n \gamma_{ij} F_{ij:kl}(z_1, z_2; z_3; z_4) \right| &= o(n^3), \end{aligned}$$

for $\gamma_{ij} = \gamma_{ij}^{(2)}$, $\gamma_{ij}^{(U)}$ and the product $\gamma_{ij}^{(2)} \gamma_{ij}^{(U)}$.

Assumption B7 For all sufficiently large n , $\bar{f} \in \mathcal{G}_\lambda^\infty$ for some $\lambda > 0$, and, for distinct $i, j, k, l \in [1, n]$,

$$\overline{\lim}_{n \rightarrow \infty} \left\{ \max_i \sup f_i(z) + \max_{i,j} \sup f_{ij}(z_1, z_2) + \max_{i,j,k} \sup f_{ijk}(z_1, z_2, z_3) + \max_{i,j,k,l} \sup f_{ijkl}(z_1, z_2, z_3, z_4) \right\} < \infty,$$

where the suprema are over all real values of the function arguments.

Introduce a scalar function $G(z)$ such that

$$\sum_{i=1}^n \mathbb{E} \{G^4(Z_i)\} = O(n), \quad \text{as } n \rightarrow \infty.$$

Assumption B8 For $t = 1, 2$, $\xi_t \in \mathcal{G}_\mu^4$ for some $\mu > 0$ and there exist $\varepsilon > 0$ such that for any $z \in \mathbb{R}^q$

$$\sup_{0 < \|u\| < \varepsilon} \frac{|\xi_t(z) - \xi_t(z+u)|}{\|u\|} \leq G(z).$$

Assumption B9 $\theta \in \mathcal{G}_\nu^4$ for some $\nu > 0$, and there exist $\varepsilon > 0$ such that for any $z \in \mathbb{R}^q$

$$\sup_{0 < \|u\| < \varepsilon} \frac{|\theta(z) - \theta(z+u)|}{\|u\|} \leq G(z).$$

Assumption B10 For $t = 1, 2$, as $n \rightarrow \infty$,

$$\int \sigma_t(z, z)^2 \bar{f}(z) dz + \int \sigma_2(z_1, z_2)^2 \bar{f}(z_1) \bar{f}(z_2) dz_1 dz_2 = O(1),$$

$$\max_{1 \leq i, j \leq n} \mathbb{E} |\sigma_2(Z_i, Z_j)| = O\left(n^{1/2}\right),$$

and there exist $\varepsilon > 0$ and functions $G_t(z_1, z_2)$ such that for any $z_1, z_2 \in R^q$,

$$\sup_{0 < \|(u, v)\| < \varepsilon} \frac{|\sigma_t(z_1, z_2) - \sigma_t(z_1 + u, z_2 + v)|}{\|(u, v)\|} \leq G_t(z_1, z_2),$$

where as $n \rightarrow \infty$

$$\int G_t(z, z) \bar{f}(z) dz + \int G_t(z_1, z_2) \bar{f}(z_1) \bar{f}(z_2) dz_1 dz_2 = O(1).$$

Assumption B11 As $n \rightarrow \infty$,

$$n^{-1/2} \sup_{j \in \mathbb{N}} \sum_{i=1}^n \|V_{2i}\| \bar{f}_i^2 |b_{ij}| \rightarrow_p 0$$

and

$$\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^{\infty} b_{ik} b_{jk} \bar{f}_i^2 \bar{f}_j^2 V_{2i} V_{2j}' \rightarrow_p \Omega,$$

where Ω is a constant p.d. matrix.

Assumption B12 For the same λ, μ, ν as in Assumptions B7 - B9, $k \in \mathcal{K}_{\max(l+m-1, l+r-1)}$ for integers l, m, r such that $l-1 < \lambda \leq l$, $m-1 < \mu \leq m$, $r-1 < \nu \leq r$.

Assumption B13 For the same λ, μ, ν as in Assumptions B7 - B9, as $n \rightarrow \infty$,

$$a + n^{-1/2} a^{-q} + n^{1/2} a^{\zeta-2q} + n^{1/2} (a^{2\mu} + a^{2\nu} + a^{2\lambda}) \rightarrow 0.$$

Theorem B Under Assumptions B1-B12, as $n \rightarrow \infty$,

$$\sqrt{n} (\hat{\beta} - \beta) \rightarrow_d \mathbf{N}(0, \Psi^{-1} \Omega \Psi^{-1'}).$$

To a substantial degree, the assumptions are a mixture or modification of ones in Robinson (1988, 2007b). In his i.i.d. data setting, Robinson (1988) was able to relax Assumption B4 to $\mathbb{E}(V_{ti}|Z_1, \dots, Z_n) = 0$ a.s., $t = 1, 2$, but for our potentially spatially dependent setting we have been unable to avoid more structure. Though Assumption B4 does allow for some conditional heteroscedasticity it is nevertheless strong, especially when $p > 1$, but we prefer to avoid milder but more complicated assumptions. Assumption B5 places an upper bound on the spatial dependence in U_i and V_{2i} that covers long memory. Assumption B6, as in the nonparametric regression setting of Robinson (2007b), constitutes an asymptotic independence assumption on Z_i ; part (ii) of it also involves the $\gamma_{ij}^{(U)}$ and $\gamma_{ij}^{(2)}$. It is difficult to check in general, but this is possible at least under Gaussianity: as noted in Robinson (2007b), a similar

(slightly stronger) condition was checked by Castellana and Leadbetter (1986), in the stationary scalar Gaussian time series case: there exists $\varepsilon > 0$ such that for $\mathcal{N}(z) = \{z_1 \in \mathbb{R} : |z - z_1| < \varepsilon\}$,

$$\sup_{z_1 \in \mathbb{R}} \sup_{z_2, z_3 \in \mathcal{N}(z_1)} \sum_{i,j,k}^n \left| \frac{F_{jk:i}(z_2, z_3; z_1)}{f(z_1)} \right| \leq Cn \sum_{i=1}^n \sum_{j=1}^n |Cov(Z_i, Z_{i+j})|.$$

In this setting at least, Assumption B6 allows $\{Z_i\}$ to have long memory. With respect to finding alternative sufficient conditions, there is always a difficulty, in either the spatial or time series contexts, in characterizing useful, coherent, joint, non-Gaussian, densities. To place matters in further perspective, mixing conditions would provide an alternative to B6, but though there has been a good deal of discussion of conditions for these with respect to time series, relatively little seems to be known in a spatial context, especially given the rather wide range of spatial configurations that we try to allow for.

4 Variance Estimation

For statistical inference the limiting covariance matrices in Theorems A and B must be consistently estimated. To focus particularly on the Theorem A, Assumption A2 gives a consistent estimate, $\widehat{\Phi}$, of Φ . Assuming no correlation in the U_i , Σ can be estimated by

$$\widehat{\Sigma}_1 = \widehat{\Sigma}_{1n} = \frac{1}{n} \sum_{i=1}^n X_{2i} X_{2i}' \widetilde{U}_i^2,$$

where

$$\widetilde{U}_i = \widetilde{U}_{in} = Y_i - \widetilde{\beta}' X_{1i},$$

so $\widehat{\Sigma}_1$ is a standard heteroscedasticity-robust estimate in the style of Eicker (1967). Assuming also homoscedasticity we have of course the estimate

$$\widehat{\Sigma}_2 = \widehat{\Sigma}_{2n} = \widetilde{\sigma}^2 \frac{1}{n} \sum_{i=1}^n X_{2i} X_{2i}',$$

where $\widetilde{\sigma}_n^2 = \widetilde{\sigma}^2 = (n-p)^{-1} \sum_{i=1}^n \widetilde{U}_i^2$. Consistency of $\widehat{\Sigma}_1$ and $\widehat{\Sigma}_2$ follows under mild additional conditions.

Estimation of Σ can be considerably more problematic when there is correlation in the U_i . Given a parametric model for U_i , such as a SAR or, with lattice data, a lattice extension of a stationary time series model such as an autoregressive moving average, matters are relatively straightforward. When U_i is not parametrically modelled, lattice data permit relatively straightforward extension of the heteroscedasticity-and-autocorrelation-consistent (HAC) variance estimates proposed for time series data, which are essentially smoothed nonparametric estimates of the spectral density matrix of a stationary process at zero frequency (though the edge-effect must be taken account of). For non-lattice data there is a fundamental difficulty of autocovariance estimation, for example when data are irregularly-spaced there are typically insufficient pairs of observations available to reliably estimate the autocovariance for a given lag using standard formulae. This problem is present with irregularly-spaced time series data, and the kernel smoothing method suggested there by Masry (1983), to estimate autocovariances at integer lags, can be extended to two or more dimensions, with the autocovariance estimates then straightforwardly inserted in a higher-dimensional HAC formula. This approach is based on stationarity, but as in the time series case it can doubtless be shown to be consistency-robust to a degree of heterogeneity. As an alternative way in which the problem can be transformed to one for a

stationary random field on a lattice, Conley (1999) modelled locations by a point process, dividing the sampling region into rectangular cells such that for each cell, there can be at most a single observation.

On the other hand an estimate which potentially covers both nonparametric dependence and heterogeneity is of form

$$\widehat{\Sigma}_3 = \widehat{\Sigma}_{3n} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n X_{2i} X_{2j}' \widetilde{U}_i \widetilde{U}_j' w_{ij}, \quad (6)$$

where the $w_{ij} = w_{ijn}$ form an array of weights, as in Kelejian and Prucha (2007). In their proof of consistency, they stress SAR-type U_i , but the property holds much more generally under Assumption A1. The quadratic-form estimate (6) reduces to a familiar HAC form if the w_{ij} are of the kernel form $w_{i-j} = w_{i-j,n}$, involving a bandwidth, but Kelejian and Prucha (2007) take $w_{ij} = w(d_{ij}/d)$, where the function $w(x)$ is suitably normalized and vanishes for $x > 1$, $d_{ij} = d_{ijn}$ is a known, positive (economic) distance between locations i and j , and $d = d_n \geq \max_{i,j} d_{ij}$ is regarded as increasing without bound with n . An alternative choice of w_{ij} is based on knowledge of observed locations $s_i^* \in \mathbb{R}^r$, for dimension $r \geq 1$, $i = 1, \dots, n$. Let s_i be a $r \times 1$ vector such that if s_{ik} and s_{ik}^* are the k -th elements of s_i and s_i^* , so s_{ik} is the smallest integer such that $s_{ik} \geq s_{ik}^*$. We can regard s_i as discretized locations on a rectangular grid. Define

$$w(s_i - s_j, m) = \prod_{k=1}^r h\{(s_{ik} - s_{jk})/m_k\},$$

where h is a real-valued function and $m_k = m_{kn}$ are non-negative integers forming a truncation vector $m = (m_1, \dots, m_r)$. Set $w_{ij} = w^*(s_i^*, s_j^*) = w(s_i - s_j, m)$.

With respect to variance estimation in Theorem B, Assumption B5 supplies a consistent estimate, $\widehat{\Psi}$, of Ψ , while to echo remarks of the previous section, after defining $\widehat{U}_{1i}^* = \widehat{U}_{1in}^* = Y_i^* - \widehat{\beta}' X_{1i}^*$, under regularity conditions a consistent estimate of Ω is

$$\widehat{\Omega}_1 = \widehat{\Omega}_{1n} = \frac{1}{n} \sum_{i=1}^n X_{2i}^* X_{2i}^{*'} \widehat{U}_{1i}^{*2},$$

when the U_i are independent, and

$$\widehat{\Omega}_2 = \widehat{\Omega}_{2n} = \widehat{\sigma}^2 \frac{1}{n} \sum_{i=1}^n X_{2i}^* X_{2i}^{*'},$$

when they are also homoscedastic, where $\widehat{\sigma}_n^2 = \widehat{\sigma}^2 = (n-p)^{-1} \sum_{i=1}^n \widehat{U}_{1i}^{*2}$. For dependent U_i one can use (cf (6))

$$\widehat{\Omega}_3 = \widehat{\Omega}_{3n} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n X_{2i}^* X_{2j}^{*'} \widehat{U}_{1i}^* \widehat{U}_{1j}^{*'} w_{ij}. \quad (7)$$

In order to provide some reasonably comprehensible theoretical justification, let us consider the infeasible estimate

$$\widetilde{\Sigma}_3 = \widetilde{\Sigma}_{3n} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n X_{2i} X_{2j}' U_i U_j' w_{ij}, \quad (8)$$

which becomes $\widehat{\Sigma}_3$ with U_i replaced by \widetilde{U}_i , and $\widehat{\Omega}_3$ with U_i , X_{2i} replaced by \widehat{U}_{1i}^* , X_{2i}^* respectively. For any $\lambda \in \mathbb{R}^p$,

$$\lambda' \widetilde{\Sigma}_3 \lambda = \frac{1}{n} \sum_{s \in \mathbb{L}} \sum_{t \in \mathbb{L}} v_s v_t w(s-t, m),$$

where $v_t = \sum_{i=1}^n \lambda' X_{2i} U_i \mathbf{1}(s_i = t)$ and $\mathbf{1}$ is the indicator function. This can be written as

$$\sum_{u \in \mathbb{L}^*} w(u, m) c_u,$$

where $\mathbb{L}^* = \{s - t : s \in \mathbb{L}, t \in \mathbb{L}\}$, $c_u = n^{-1} \sum_{\tau(u)} v_t v_{t+u}$, and $\tau(u) = \{t : t \in \mathbb{L}, t + u \in \mathbb{L}\}$, where we assume that $s_i \in \prod_{j=1}^r \{1, \dots, n_j\} = \mathbb{L}$ for all i , where \mathbb{L} is the smallest rectangular grid containing all s_i . If h is either the modified Bartlett window or the Parzen window, then $\lambda' \tilde{\Sigma}_3 \lambda \geq 0$ (see Robinson, 2007a), and hence $\tilde{\Sigma}_3$ is non-negative definite. We establish conditions for approximating

$$\Sigma_n = n^{-1} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}(X_{2i} X'_{2j}) \mathbb{E}(U_i U_j)$$

by $\tilde{\Sigma}_3$.

Assumption C1 *The kernel h is a real, even function such that $|h(u)| \leq 1$; $h(u) = 0$ if $|u| > 1$; and $\lim_{u \rightarrow 0} (1 - h(u)) / |u|^q = h_q$ for some $q > 0$ and $0 < h_q < \infty$.*

Assumption C2 *As $n \rightarrow \infty$,*

(i)

$$m_k \rightarrow \infty, \quad n_k \rightarrow \infty, \quad k = 1, \dots, r;$$

(ii)

$$\frac{m_k}{n_k} \rightarrow 0, \quad k = 1, \dots, r;$$

and there exist $0 < c_1 < c_2 < \infty$ such that

$$c_1 \prod_{k=1}^r n_k \leq n \leq c_2 \prod_{k=1}^r n_k$$

for sufficiently large n .

Define

$$S_n(u) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}(X_{2i} X'_{2j}) \mathbb{E}(U_i U_j) \mathbf{1}(s_i - s_j = u).$$

Assumption C3 *There exists a family of $p \times p$ matrices $\{G_u : u \in \mathbb{Z}^r\}$, where \mathbb{Z}^r is the r -Cartesian product of the set of integers, such that the absolute value of each element of $S_n(u)$ is bounded by the corresponding element of G_u , for all $u \in \mathbb{L}^*$, $n \in \mathbb{N}$, and $\sum_{u \in \mathbb{Z}^r} \sum_{k=1}^r |u_k|^q G_u$ is a finite matrix.*

Assumption C4 *Let x_{ti} be the t -th element of X_i . For all $t, s = 1, \dots, p$, as $n \rightarrow \infty$*

$$\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n |\kappa(x_{ti} U_i, x_{sj} U_j, x_{tk} U_k, x_{sl} U_l)| = O(n),$$

where κ is the cumulant function.

Define

$$\begin{aligned} S_{1ts,n}(u, v, u_1) &= n^{-1} \sum_{u, v, u_1} \mathbb{E}(x_{ti}x_{tk}U_iU_k) \mathbb{E}(x_{sj}x_{sl}U_jU_l), \\ S_{2ts,n}(u, v, u_1) &= n^{-1} \sum_{u, v, u_1} \mathbb{E}(x_{ti}x_{sk}U_iU_k) \mathbb{E}(x_{tj}x_{rl}U_jU_l), \end{aligned}$$

where the summation is over all i, j, k and l such that $s_i - s_j = u$, $s_k - s_l = v$ and $s_i - s_k = u_1$.

Assumption C5 There exist numbers $\{\gamma_{u,v} : u, v \in \mathbb{Z}^r\}$ such that $|S_{1ts,n}(u, v, u_1) + S_{2ts,n}(u, v, u_1)| \leq \gamma_{u_1, u_1+v-u}$ for all $t, s = 1, \dots, p$ and $u, v, u_1 \in \mathbb{L}^*$, $n \in \mathbb{N}$, and

$$\sum_{u \in \mathbb{Z}^r} \sum_{v \in \mathbb{Z}^r} \gamma_{u,v} < \infty.$$

Theorem C As $n \rightarrow \infty$, under Assumptions C1, C2 (i) and C3

$$\mathbb{E}(\tilde{\Sigma}_3 - \Sigma_n) = O\left(\sum_{k=1}^r m_k^{-q}\right);$$

and under Assumptions C2, C4 and C5,

$$\text{Var}(\tilde{\Sigma}_3) = O\left(n^{-1} \prod_{k=1}^r m_k\right).$$

Sharper results can be obtained if stronger assumptions are imposed. For example, if as $n \rightarrow \infty$, $S_n(u) \rightarrow S(u)$ for all u , for a well-defined function $S(u)$, the asymptotic bias can be made more precise. This assumption is similar to the definition of asymptotic stationarity of irregularly spaced time series in Parzen (1963). The same can be said for the variance if another type of asymptotic stationarity is introduced (see the proof of Theorem C). Under such assumptions, the asymptotic mean squared error can be used as a criterion for choosing a truncation vector, and a data-dependent plug-in procedure then employed.

5 Monte Carlo Study of Finite-Sample Performance

We examine first, for the linear regression (1) with $p = 1$, the size of 2-sided t -tests based on the LS version of $\tilde{\beta}$ and the estimates $\hat{\Sigma}_1, \hat{\Sigma}_2$ and the second approach to forming $\hat{\Sigma}_3$ described in the previous section. The locations s_1, \dots, s_n of the observations were generated by a random draw from the uniform distribution over $[0, 4n^{1/2}] \times [0, 4n^{1/2}]$. Given these (and keeping them fixed across replications), the U_i were generated as normal variables with mean zero and covariances $\text{Cov}(U_i, U_j) = \rho_U^{\|s_i - s_j\|}$, for prescribed $\rho_U \in (0, 1)$. Likewise the X_i ($= X_{1i} = X_{2i}$) were generated as scalar normal variables with

mean unity and covariances $Cov(X_i, X_j) = \rho_X^{\|s_i - s_j\|}$, for prescribed $\rho_X \in (0, 1)$ (and independently of the U_i). We took $\beta = 1$, $(\rho_X, \rho_U) = (0.2, 0.3)$ and $(0.4, 0.5)$, $n = 100$ and 169 , and generated 1000 replications. Table 1 reports empirical sizes of t -tests with nominal sizes $\alpha = 0.01, 0.05$ and 0.1 using $\tilde{\Sigma}_1$, denoted in the "m" column by H, $\tilde{\Sigma}_2$, denoted there by C, and $\tilde{\Sigma}_3$, for various values of m in the truncation vector (m, m) , and using the Parzen kernel for h . There is some over-sizing, which diminishes with increasing n . The over-sizing is particularly acute with respect to the inappropriate variance estimates C and H, with the (heteroscedasticity-robust) H doing worse than the classical C. For $\tilde{\Sigma}_3$ there is stability across m (though when we tried m outside the range used in Table 1 we found greater sensitivity).

Table 1 about here

Power was investigated in the same setting, against the incorrect null hypothesis that $\beta = 0.8$, but with $U_i \sim NID(0, 1)$, $X_i \sim NID(1, 1)$. Monte Carlo powers are displayed in Table 2. The main findings are that choice of variance estimate here makes little difference, and that power increases quite significantly with the rather modest increase in n . The experiment was repeated with the incorrect null hypothesis $\beta = 0.5$, when all powers were perfect.

Table 2 about here

We now turn to the semiparametric partly linear model (5), and use the LS version of $\hat{\beta}$. This depends on a bandwidth a . In general one expects less sensitivity to bandwidth choice in semiparametric than in nonparametric estimation. Moreover, the problem with trying to use a data-dependent bandwidth, especially in a relatively complicated, semiparametric, situation like this, is not so much the computational effort as that one is then at the mercy of a mechanical procedure that is itself rather arbitrarily selected. Even in the semiparametric literature often optimal bandwidths originally devised for purely nonparametric estimation are used, but clearly their relevance to the semiparametric model is unclear. Alternatively one can develop some procedure based on the semiparametric model itself. Our view here is that if the goal is statistical inference based on the central limit theorem, rather than using, say, minimum-mean-squared error or cross-validation procedures, it is more appropriate to choose a bandwidth that minimizes the error in the normal approximation. Nishiyama and Robinson (2000) achieved this for semiparametric averaged derivatives but even that case is complicated and in the current one, if feasible, it would be more so. Moreover, they assumed independence of observations, which would clearly be inappropriate here given the paper's overall focus. Even weak disturbance correlation would affect this optimal bandwidth (unlike in the pure nonparametric setting), let alone the strong correlation which we allow for. Another point to bear in mind is that our asymptotic theory, like the bulk of the nonparametric and semiparametric literature, assumes a data-free bandwidth. In any case some experience over the years suggests that unless an "optimal" bandwidth is available and well-motivated it may be desirable to employ a range of bandwidths, which also allows one to assess sensitivity, and this was done in the following experiment (though cross-validation was tried in the empirical study of the following section).

In (5) we took $p = 1$, $q = 2$ and $X_i = 1 + Z_{1i} + Z_{2i} + V_i$, $\theta(Z_i) = Z_{1i}^2 + Z_{2i}^2$, where the Z_{1i} , Z_{2i} , V_i were generated as normal variables with mean zero and such that $Cov\{X_i, X_j\} = \rho_X^{\|s_i - s_j\|}$, the U_i as normal with mean zero and $Cov\{U_i, U_j\} = \rho_U^{\|s_i - s_j\|}$, and $\{Z_{1i}\}$, $\{Z_{2i}\}$, $\{V_i\}$ and $\{U_i\}$ were independent. We again took $\beta = 1$, $(\rho_X, \rho_U) = (0.2, 0.3)$ and $(0.4, 0.5)$, $n = 100$ and 169 , and generated 1000 replications. We employed $a = 1.0, 1.2$ and 1.4 . We used two different kernels k , namely $k_2(z) = \phi(z)$ and $k_4(z) = (3 - z^2)\phi(z)$, where ϕ is the standard Gaussian density; k_2 and k_4 are respectively second- and fourth-order kernels, and are thus not of high enough order to satisfy the conditions of Theorem B, but this strategy was adopted due to the imprecision likely to be caused by a high order kernel in the relatively modest sample sizes.

There is interest in the effect on bias (BI) and standard deviation (SD) of the point estimate $\hat{\beta}$ of the choice of kernel and bandwidth. The results for k_2 were as follows. With $(\rho_X, \rho_U) =$

(0.2, 0.3), BI(SD) was, for $a = 1.0, 1.2, 1.4$, respectively .0062(.1200), .0059(.1184), .0057(.1187) when $n = 100$, and .0047(.0872), .0037(.0852), .0026(.0849) when $n = 169$; with $(\rho_X, \rho_U) = (0.4, 0.5)$, BI(SD) was .0052(.1260), .0048(.1259), .0047(.1281) when $n = 100$, and .0045(.0909), .0035(.0894), .0024(.0897). The results for k_4 were as follows. With $(\rho_X, \rho_U) = (0.2, 0.3)$, BI(SD) was .0063(.1245), .0060(.1224), .0059(.1214) when $n = 100$, and .0053(.0910), .0045(.0886), .0035(.0872) with $n = 169$; with $(\rho_X, \rho_U) = (0.4, 0.5)$, BI(SD) was .0052(.1300), .0050(.1291), .0048(.1295) when $n = 100$, and .0050(.0945), .0043(.0925), .0033(.0915) when $n = 169$. Both BI and SD fall with increasing n . There is no clear pattern discernable from changing (ρ_X, ρ_U) . The fact that k_2 on average produces lower BI than k_4 is due to the fact that the same bandwidths were used for both, whereas k_4 demands a larger bandwidth than k_2 . Nevertheless, k_2 still produces a lower SD.

Tables 3 and 4 about here

From the same replications t -ratios were computed for each choice of kernel and bandwidth, and using $\widehat{\Omega}_1$, denoted by H, $\widehat{\Omega}_2$, denoted by C, and $\widehat{\Omega}_3$, which employed the Parzen kernel and m in the truncation vectors (m, m) . Empirical sizes using k_2 and k_4 are displayed in Tables 3 and 4 respectively. There is clearly some sensitivity to choice of a , with sometimes a monotone change, and sometimes a peak or trough, observed on increasing it, though the discrepancies do not seem huge. Use of the C or H estimates tends to produce marked over-sizing when $(\rho_X, \rho_U) = (0.4, 0.5)$, but the correlation-robust tests are quite stable across m . Generally, performance deteriorates with greater spatial correlation, but it also improves with increasing n , and when $n = 169$ it is surprisingly better than for the parametric linear model (1). Comparing Tables 3 and 4, k_2 generally fares better than k_4 , possibly due to the relative BI and SD behaviour reported above.

Finally Table 5 displays empirical powers, against the incorrect null hypothesis that $\beta = 0.7$, in the previous setting but with $U_i, V_i, Z_{1i}, Z_{2i} \sim NID(0, 1)$. Powers mostly increase somewhat with a and markedly with n , but tend to be stable across the variance estimates, with the larger powers for C possibly due to over-sizing. In another experiment using the incorrect null hypothesis that $\beta = 0.5$, perfect powers were observed throughout.

Table 5 about here

6 Empirical Illustration

The present section develops an empirical analysis of Banerjee and Iyer (2005), which employed linear regression modelling and estimation to study the influence of different systems for collecting land revenue in India, instituted during British colonial rule, on present-day economic performance. In a threefold classification of these systems, in a given area revenue was collected either through the local landlord, or through the village, or from the individual cultivator. Banerjee and Iyer (2005) used district-level data, and calculated the proportion of "non-landlord" areas within a district (in the 1870's or 1880's); in some cases this could not be done accurately and a proportion of 0 or 1 was assigned. This non-landlord proportion, denoted NL, was the explanatory variable of chief interest in Banerjee and Iyer's (2005) study: on the basis of economic theory and empirical evidence, agricultural investment and yields are positively related to NL, and income/wealth inequality are negatively related to it. Their data on measures of economic performance and productivity, used as dependent variables, consisted of a panel (annually, over the period 1956 through 1987 and across some 271 districts in 13 major states). As well as carrying out LS regressions (correcting also for various control variables), because of concerns about endogeneity (non-landlord areas are inherently more productive), Banerjee and Iyer (2005) also used IV estimation with a dummy, which we denote C0, for whether or not a district was conquered between 1820 and 1856 as instrument for NL. Districts are intrinsically of irregular size and shape, and are thus intrinsically geographically irregularly-spaced, and moreover the lack of data for some states

produces huge spatial gaps, as Figure 1 of Banerjee and Iyer (2005) indicates. However, they did not explore the possibility of spatial or serial correlation, and employed standard inference rules based on uncorrelated and homoskedastic disturbances, and nor did they explore semiparametric modelling.

We consider the possibility of spatial correlation of disturbances, and its affect on inference, as well as the use of partly linear, and also pure nonparametric, regression. To maintain focus and prevent matters becoming over-complicated, we employ data from only one year, 1984; incidentally, Banerjee and Iyer's (2005) model was static, with time-invariant slope parameters. Employing data from near the end of the period also takes account of the "Green Revolution" (see e.g. Munshi (2004)), which started in the early 1960s to combat famine in certain Indian states, and was later extended throughout the country; as Banerjee and Iyer's (2005) aim was to study effects of local institutions, later periods in the sample could provide better regression fits.

We first tested for spatial correlation of the disturbances in some of Banerjee and Iyer's (2005) regression models, employing LS and IV residuals in members of the class of tests proposed by Robinson (2008). These tests include a number of previously-proposed ones as special cases, and can be designed to have a Lagrange multiplier interpretation with respect to certain spatially correlated alternatives to the null of uncorrelatedness, for example against a SAR alternative, when the test statistic depends on the chosen spatial weight matrix or matrices. For certain choices, several members of this class of statistics, including ones with finite-sample corrections, were computed, for the four regressions with proportion of irrigated land (IL), fertilizer use (FU), $\log(\text{yield } 15 \text{ crops})$ (L15), and $\log(\text{rice yield})$ (LR) as dependent variable Y . For the most part the tests rejected, suggesting possible spatial correlation in disturbances (though as always some other source of misspecification could be the cause). The detailed results can be obtained from the authors on request.

We next carried out some simple Nadaraya-Watson nonparametric regression fits, of each of the same four Y on NL. Under similar assumptions to ours, Robinson (2007b) showed consistency and asymptotic normality of this estimate. Though his conditions require the explanatory variable to be continuous, whereas as previously noted NL has a mixed distribution, nevertheless the exercise may be helpful in reflecting nonlinearity and hinting at its form. Figures 1-4 contain scatter plots for the four dependent variables and nonparametric regression fits using a Gaussian kernel with bandwidth 0.3. This choice was the smallest one that did not give very unsmooth curves, and much larger ones appeared to oversmooth, indeed NL takes values in $[0,1]$. In any case the purpose of the nonparametric regression is only exploratory, to hint at possible structure. The Figures suggest in each case a mode, and possibly a mild secondary one, and thus evidence of nonlinearity, contrary to the modelling of Banerjee and Iyer (2005).

Figures 1-4 about here

Our parametric and semiparametric regression models included (unlike in Banerjee and Iyer (2005)) the square (NL2) of NL as a regressor (as well as NL itself), as just suggested by the nonparametric fitting. We also replaced two of Banerjee and Iyer's (2005) explanatory variables by proxies which may be more appropriate. For their panel data set, mean annual rainfall was constructed over 1931-1960, but rainfall records from several decades earlier than 1984, the only year which we analyze, may not be relevant, especially for agricultural yields. We used instead a precipitation variable (PRE) constructed by Mitchell and Jones (2005), based on a method which they argued offers some improvement over existing ones in the climatology literature: their dataset included 6 monthly climate elements over a 0.5° grid, over which variation is small, and we used longitude and latitude of district headquarters to obtain a weighted average at surrounding grids for 1984, district headquarters tending to be in areas of high population density which themselves tend to be relatively fertile. Second, Banerjee and Iyer (2005) included latitude (but not longitude) as an explanatory variable, but latitude behaves like a linear trend in a time series regression, and thus affects the rate of convergence of estimates, in a way determined by the scatter of district headquarters. We replaced latitude by annual temperature (TEM), which varies considerably across India and is more likely to influence agricultural yields and hence investment

decisions. As an additional modification, we discarded Thanjavur district because it appears to have serious measurement error: it is the only district having IL exceeding unity, and FU in Thanjavur was 79.44 in 1981, rose to 301.18 in 1982, and has remained high since, whereas average FU excluding Thanjavur in 1984 was only 61.15.

IV estimation in the presence of the additional, NL-dependent, regressor NL2, requires an additional instrument. The one selected, denoted C1, takes the value unity if a district was acquired between 1820 and 1856, and otherwise its value is determined by the cause of acquisition: 0.1 for "lapse", 0.3 for "misrule", 0.5 for conques, 0.7 for "grant", and 0.8 for "ceded". The ordering is based on a likely strategy for security of the British administration, the higher value for "ceded" to "grant" due to the latter being more common at the beginning of the British colonisation when landlord land-revenue systems predominated. C1 can be considered as a finer version of C0, and should likewise be uncorrelated with omitted districts' characteristics which determine 1984 investment and productivity; both are one-off historical events. On the other hand C0 and C1 are not highly correlated but are both highly correlated with NL. We used C0 and C1 as instruments for NL2 and NL respectively, C1 having relatively higher sample correlation with NL.

In (1) we took $Y = \text{IL, FU, L15 and LR}$, as above ($n = 164, 164, 165$ and 165 respectively), with

$$X_1 = (1, \text{NL, NL2, DBC, CD, BSD, RSD, ASD, ALT, PRE, TEM})'$$

where DBC = date district came under British control, CD=coastal dummy, BSD=black soil dummy, RSD=red soil dummy, ASD=alluvial soil dummy, and ALT=altitude. We computed $\tilde{\beta}$ both with $X_2 = X_1$ (LS) and with

$$X_2 = (1, \text{C1, C0, DBC, CD, BSD, RSD, ASD, ALT, PRE, TEM})'$$

(IV). Standard errors (SEs) were computed using $\hat{\Phi}$, $\hat{\Sigma}_2$ and $\hat{\Sigma}_3$ as described in Sections 5 and 6, with for $m = 2, 4, 6$. Next, in (5) we took

$$X_1 = (\text{NL, NL2, DBC, CD, BSD, RSD, ASD})', \quad Z = (\text{ALT}^*, \text{PRE}^*, \text{TEM}^*)'$$

where $\text{ALT}^*, \text{PRE}^*, \text{TEM}^*$ are ALT, PRE, TEM normalized to have sample variances approximately 1 (in order to better justify use of a scalar bandwidth). This selection keeps NL, NL2 and DBC in the parametric part, these being the explanatory variables of most interest, along with the dummies, and puts into the nonparametric part control variables that can be taken to be continuous. We computed $\tilde{\beta}$ with Z as above, and both with $X_2 = X_1$ ("partly LS") and

$$X_2 = (\text{C1, C0, DBC, CD, BSD, RSD, ASD})'$$

("partly IV"). For choosing the bandwidth a we tried the partial LS cross-validation procedure (and an IV modification) of Gao (1988), justified by Gao and Yee (2000), though this does not quite fit with our density-weighted estimate $\tilde{\beta}$. (The elements of Z were previously normalized to have unit sample variance.) Unfortunately this tended to deliver data-dependent bandwidths that are far too large. There was a tendency for the cross-validation objective function to first decrease rapidly as a increases, then remain quite flat over a wide range before increasing. Thus we proceeded in a semi-automatic way, choosing two relatively small a that lie in the flat region of the cross-validation objective function, these bandwidths varying across the partly LS and IV estimates and across the same two kernels, k_2 and k_4 , as used in the previous section. SEs were computed using $\hat{\Psi}$, $\hat{\Omega}_2$ and $\hat{\Omega}_3$ as described in Sections 5 and 6, the latter being implemented in the same way as $\hat{\Sigma}_3$, and for $m = 2, 4, 6$; we justify these smallish values by the fact that the data locations of the Indian districts data fit within a 25 17 rectangle, where the units are latitude and longitude. The results are presented in Tables 6-9, for respectively irrigated land (IL), fertilizer use (FU), log(yield 15 crops) (L15), and log(rice yield) (LR)

as dependent variable, with point estimates in bold-face and SEs reported in parentheses beneath them (non-robust ones above the three robust ones).

Tables 6-9 about here

Considering first the parametric LS and IV estimates, sometimes marked differences between them are seen and neither estimate is statistically significant. In Tables 6, 7 and 8 none of the IV estimates on NL and NL2 is significant, but all the LS is significant, and in Table 9 NL is significant. This outcome also reflects the larger SEs for IV, which were anticipated. The signs of both LS and IV estimates of coefficients of NL and NL2 are mostly consistent with the inverted U-shape seen in Figures 1-4. Also in accordance with Banerjee and Iyer (2005), DBC was nearly always found to have a significantly negative effect; the exceptions were for the larger m , SEs tending to increase with m , a fairly general feature, though in most cases the variation did not affect the question of significance. Nor did the non-robust SEs often differ much from the robust ones. Turning to the semiparametric estimates, both the LS and IV versions of $\hat{\beta}$ tend to be in the same ball-park as LS (but not IV) $\tilde{\beta}$, at least where NL, NL2 and DBC are concerned, though in Table 8, where LS and IV are relatively close, there is a larger discrepancy for NL and NL2 with semiparametric IV exceeding in absolute value all the other estimates in case of NL and NL2. Again, using instruments tends to increase SE. There is some sensitivity to choice of bandwidth a and kernel k , though seldom enough to affect significance, keeping m fixed. With respect to kernel choice, k_4 does not necessarily produce larger SEs than k_2 , perhaps because of our simultaneous variation in bandwidth a . On the whole it could be said that Banerjee and Iyer's (2005) fully linear specifications are not contradicted by our results, except of course, and importantly, where our extra regressor NL is concerned, and the results here do strongly confirm the pattern found in our nonparametric regression fits.

7 Final Comments

We have developed asymptotic properties useful in statistical inference on regression coefficients in parametric and semiparametric partly linear models, in the context of a potentially wide range of spatial or spatio-temporal data. Consistent estimation of limiting covariance matrices is required, and we have also discussed this topic both when the disturbances are uncorrelated, and when they are spatially correlated. Finite-sample performance has been investigated in a simulation study, and the methods applied to an Indian regional data-set.

A number of related issues and extensions can be pursued.

1. As mentioned in the Introduction, mixing conditions represent an alternative class of dependence conditions, to replace our linear process assumption on disturbances and density-based assumptions on regressors. A recent econometric reference is Jenish and Prucha (2008), who develop the (regular lattice) mixing condition theory of Bolthausen (1982), establishing asymptotic normality (and laws of large numbers) for the sample mean of a scalar process observed on a possibly irregular lattice whose exogenous locations are separated by distances that are bounded away from zero. Analogous conditions can undoubtedly be developed for our more complicated statistics, dependent on multivariate data (with probably faster convergence of mixing rates required), and this kind of approach would enable a relaxation of our assumption of independence between regressors and observables. On the other hand, our conditions are potentially applicable beyond their irregular lattice context (in particular when observation locations are not known even approximately), and further discussion of the advantages and disadvantages of mixing conditions relative to ours can be found in Robinson (2007b). Another kind of condition that has been employed in the

spatial lattice context is based on "FKG inequalities" (see Newman, 1980), but it appears to be very restrictive.

2. As also mentioned in the Introduction, more efficient estimates than ours may be available. For example, by comparison with our simple IV estimate, when the number of available estimates exceeds the number of regressors a two-stage least squares (2SLS) estimate will be more efficient given disturbances that are both uncorrelated and homoscedastic. However, when either or both of these conditions are not met, 2SLS is not guaranteed to beat even a simple IV estimate. This drawback can be overcome by suitable GLS or generalized method-of-moment estimates, entailing either a parametric or nonparametric modelling of disturbance correlation or heteroscedasticity, but this would require further structure.
3. In the partly linear model (5), there may also be interest in estimating the nonparametric function $\theta(z)$. A simple estimate is

$$\hat{\theta}(z) = \sum_{i=1}^n (Y_i - \hat{\beta}' X_{1i}) K\left(\frac{z - Z_i}{h}\right) / \sum_{i=1}^n K\left(\frac{z - Z_i}{h}\right).$$

Under related conditions to ours, $\hat{\theta}(z)$ is likely to share the (simple, normally distributed) asymptotic properties of the infeasible estimate for which $Y_i - \beta' X_{1i}$ is replaced by $\theta(Z_i) + U_i$.

4. We have focussed on relatively simple models in this paper, but undoubtedly analogous conditions to ours can be employed in establishing, in a similarly general spatial context, asymptotic properties of estimates in more general parametric models (such as nonlinear regression and simultaneous equation models) and semiparametric models (such as those described in Robinson, 1988, Section 7).

Appendix 1: Proofs of Theorems A and B

Proof of Theorem A The proof modifies one in Robinson and Hidalgo (1997). Defining $r_n = n^{-1/2} \sum_{i=1}^n X_{2i} U_i$, by Assumption A2 it suffices to show that $r \rightarrow_d \mathbf{N}(0, \Omega)$. Now

$$r = n^{-1/2} \sum_{k=1}^{\infty} W_k \varepsilon_k,$$

where $W_k = W_{kN} = \sum_{i=1}^n X_{2i} b_{ik}$. By Lemma A1, there is a sequence $\{N = N_n\}$, increasing in n without bound, such that $r - r_{(N)} = o_p(1)$, where

$$r_{(N)} = n^{-1/2} \sum_{k=1}^N W_k \varepsilon_k.$$

Let $D = D_n = n^{-1} \sum_{k=1}^N W_k W_k'$. From the proof of Lemma A1, $\lim_{n \rightarrow \infty} \mathbb{E} \left(n^{-1} \sum_{k=N+1}^{\infty} \|W_k\|^2 \right) = 0$, so from Assumption A4, $D \rightarrow_p \Sigma$. For any $\lambda \in \mathbb{R}^p$ such that $\|\lambda\| = 1$, let $c_N = \lambda' D^{-\frac{1}{2}} r^{(N)}$ and $w_k = w_{kn} = n^{-1/2} \lambda' D^{-\frac{1}{2}} W_k$. Then $c_N = \sum_{k=1}^N w_k \varepsilon_k$, where by Assumption A4 $\{w_k \varepsilon_k, 1 \leq k \leq N\}$ is a martingale difference sequence for each $N \geq 1$. It suffices to show that conditional on $\{X_{2i}\}$, $\sum_{k=1}^N w_k \varepsilon_k \xrightarrow{d} \mathbf{N}(0, 1)$, which follows from Theorem 2 of Scott (1973) if, conditional on $\{X_{2i}\}$, as $n \rightarrow \infty$,

$$\mathbb{E} \left(\sum_{k=1}^N w_k^2 \varepsilon_k^2 \mid \varepsilon_j, j < k \right) \rightarrow_p 1, \quad (9)$$

and for all $\eta > 0$,

$$\mathbb{E} \left\{ \sum_{k=1}^N w_k^2 \mathbb{E} (\varepsilon_k^2 \mathbf{1}(|w_k \varepsilon_k| > \eta) | \{X_{2i}\}) \right\} \rightarrow 0. \quad (10)$$

The left side of (9) is $\lambda' D^{-\frac{1}{2}} \left(\frac{1}{n} \sum_{k=1}^N W_k W_k' \right) D^{-\frac{1}{2}} \lambda = 1$, so (9) holds. The left side of (10) is bounded by

$$\begin{aligned} & \mathbb{E} \left\{ \sum_{k=1}^N w_k^2 \mathbb{E} (\varepsilon_k^2 \mathbf{1}(|\varepsilon_k| > \eta/\delta)) \right\} + \mathbb{P} \left(\max_{1 \leq k \leq N} |w_k| > \delta \right) \\ & \leq \sup_{1 \leq k \leq N} \mathbb{E} (\varepsilon_k^2 \mathbf{1}(|\varepsilon_k| > \eta/\delta)) + \mathbb{P} \left(\max_{1 \leq k \leq N} |w_k| > \delta \right), \end{aligned} \quad (11)$$

for $\delta > 0$. By Assumption A3, the first term on the right can be made arbitrarily small by choosing δ small enough, so it suffices to show that $\max_{1 \leq k \leq N} |w_k| = o_p(1)$. By Assumptions A2, A3 and A4,

$$\max_{1 \leq k \leq N} |w_k| \leq n^{-1/2} \left\| D^{-\frac{1}{2}} \right\| \max_{1 \leq k \leq N} \left\| \sum_{i=1}^n X_{2i} b_{ik} \right\| = o_p(1).$$

Proof of Theorem B The proof modifies ones of Robinson (1988), Fan and Li (1999). We have

$$\hat{\beta} - \beta = S_{X_2 X_1}^{-1} (S_{X_2 \theta} + S_{X_2 U}),$$

where $S_{X_2 \theta}$ involves the array $\{\theta_i = \theta(Z_i)\}$. We show that $S_{X_2 X_1} \rightarrow_p \Psi$, $\sqrt{n} S_{X_2 \theta} \rightarrow_p 0$, $\sqrt{n} S_{X_2 U} \rightarrow_d \mathbf{N}(0, \Omega)$. With likewise $\xi_{ti} = \xi_t(Z_i)$, $t = 1, 2$, we have

$$S_{X_2 X_1} = S_{\xi_2 \xi_1} + S_{\xi_2 V_1} + S_{V_2 \xi_1} + S_{V_2 V_1}, \quad S_{X_2 \theta} = S_{\xi_2 \theta} + S_{V_2 \theta}, \quad S_{X_2 U} = S_{\xi_2 U} + S_{V_2 U}.$$

Applying the Cauchy inequality, i.e. $\mathbb{E} \|S_{AB}\| \leq (\mathbb{E} \|S_{AA}\| \mathbb{E} \|S_{BB}\|)^{1/2}$, and the propositions of the following appendix, the proof is completed by noting that $S_{\xi_2 \xi_1} \rightarrow_p 0$ (Propositions B2 and B3), $S_{\xi_2 V_1} \rightarrow_p 0$ (Proposition B4), $S_{V_2 \xi_1} \rightarrow_p 0$ (Proposition B5), $S_{V_2 V_1} \rightarrow_p \Psi$ (Proposition B6), $\sqrt{n} S_{\xi_2 \theta} \rightarrow_p 0$ (Propositions B1 and B2), $\sqrt{n} S_{V_2 \theta} \rightarrow_p 0$ (Proposition B7), $\sqrt{n} S_{\xi_2 U} \rightarrow_p 0$ (Proposition B8) and $\sqrt{n} S_{V_2 U} \rightarrow_d \mathbf{N}(0, \Sigma)$ (Proposition B9).

Appendix 2: Propositions for proofs of Theorems A and B

In this and the following appendix, it is frequently the case that a particular result requires an order bound for several quantities, but because these are often similarly handled details are not given for all, in order to conserve on space.

Define, for $1 \leq i \leq n$,

$$\hat{f}_i = \hat{f}_i(Z_i) = (na^q)^{-1} \sum_{j \neq i}^n K_{ij},$$

and for a triangular array $\{A_i\}$, $\bar{A}_i = (na^q)^{-1} \sum_{j \neq i}^n A_j K_{ij}$, so that $A_i^* = A_i \hat{f}_i - \bar{A}_i$ in the definition of S_{AB} .

Proposition B1 As $n \rightarrow \infty$,

$$\mathbb{E}(S_{\theta\theta}) = o(n^{-1/2}).$$

Proof. We have

$$\begin{aligned}\mathbb{E}(S_{\theta\theta}) &= \frac{1}{n^3 a^{2q}} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{k \neq i}^n \mathbb{E}\{(\theta_i - \theta_j)(\theta_i - \theta_k) K_{ij} K_{ik}\} \\ &= \frac{1}{n^3 a^{2q}} \left[\sum_{i,j}^n \mathbb{E}\{(\theta_i - \theta_j)^2 K_{ij}^2\} + \sum_{i,j,k}^n \mathbb{E}\{(\theta_i - \theta_j)(\theta_i - \theta_k) K_{ij} K_{ik}\} \right].\end{aligned}$$

The result follows from Lemmas B1, B2 in the following appendix, and Assumption B13. ■

Proposition B2 As $n \rightarrow \infty$,

$$\mathbb{E} \|S_{\xi_2 \xi_2}\| = o(n^{-1/2}).$$

Proof. Similar to that of Proposition B1. ■

Proposition B3 As $n \rightarrow \infty$,

$$\mathbb{E} \|S_{\xi_1 \xi_1}\| = o(1).$$

Proof. Similar to that of Proposition B1, except that the result is weaker because milder conditions are imposed on ξ_1 than on ξ_2 or θ . ■

Proposition B4 As $n \rightarrow \infty$,

$$S_{\xi_2 V_1} \rightarrow_p 0.$$

Proof. The left side is

$$n^{-1} \sum_{i=1}^n \left\{ \xi_{2i}^* V_{1i}' \bar{f}_i + \xi_{2i}^* V_{1i}' (\hat{f}_i - \bar{f}_i) - \xi_{2i}^* \bar{V}_{1i}' \right\}. \quad (12)$$

By Proposition B2, Lemmas B4 and B5, and the Cauchy inequality, the contributions from the last two summands in (12) are $o_p(1)$. Due to Assumptions B5, B7 and B10 for $t = 1, 2$,

$$\mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n \|V_{ti}\|^2 \bar{f}_i^2 \right) \leq \max_{1 \leq i \leq n} |\gamma_{ii}^{(t)}| \sup_{z \in \mathbb{R}^q} \bar{f}(z)^2 \int |\sigma_t(z, z)| \bar{f}(z) dz = O(1). \quad (13)$$

Proposition B2, (13) and the Cauchy inequality imply that the contribution from the first summand in (12) is also $o_p(1)$. ■

Proposition B5 As $n \rightarrow \infty$,

$$S_{\xi_1 V_2} \rightarrow_p 0.$$

Proof. Similar to that of Proposition B4. ■

Proposition B6 As $n \rightarrow \infty$,

$$S_{V_2 V_1} \rightarrow_p \Psi.$$

Proof. The left side is

$$n^{-1} \sum_{i=1}^n \left(V_{2i} V'_{1i} \hat{f}_i^2 - V_{2i} \bar{V}'_{1i} \hat{f}_i - \bar{V}_{2i} V'_{1i} \hat{f}_i + \bar{V}_{2i} \bar{V}'_{1i} \right). \quad (14)$$

For $t = 1, 2$,

$$n^{-1} \sum_{i=1}^n \|V_{ti}\|^2 \hat{f}_i^2 = n^{-1} \sum_{i=1}^n \|V_{ti}\|^2 \left\{ \bar{f}_i^2 + 2\bar{f}_i (\hat{f}_i - \bar{f}_i) + (\hat{f}_i - \bar{f}_i)^2 \right\}. \quad (15)$$

Lemma B4, (13) and the Cauchy inequality imply that the left side of (15) is $O_p(1)$. Hence with Lemma B5 and the Cauchy inequality, the contributions from the last three summands in (14) are $o_p(1)$. The contribution from the first summand in (14) is

$$n^{-1} \sum_{i=1}^n V_{2i} V'_{1i} \left\{ \bar{f}_i^2 + 2\bar{f}_i (\hat{f}_i - \bar{f}_i) + (\hat{f}_i - \bar{f}_i)^2 \right\}.$$

The proof is completed by applying Assumption B5, Lemma B4, (13) and the Cauchy inequality. ■

Proposition B7 As $n \rightarrow \infty$,

$$S_{V_2\theta} = o_p(n^{-1/2}).$$

Proof. The left side is

$$\frac{1}{n} \sum_{i=1}^n \left\{ V_{2i} \theta_i^* \bar{f}_i + V_{2i} \theta_i^* (\hat{f}_i - \bar{f}_i) - \bar{V}_{2i} \theta_i^* \right\}. \quad (16)$$

By Proposition B1, Lemmas B4 and B5, and the Cauchy inequality, the contribution from the last two summands are $o_p(n^{-1/2})$. The squared norm of the contribution from the first summand has expectation

$$n^{-2} \sum_{i=1}^n \gamma_{ii}^{(2)} \mathbb{E} \left(\sigma_2(Z_i, Z_i) \theta_i^{*2} \bar{f}_i^2 \right) + n^{-2} \sum_{i,j} \gamma_{ij}^{(2)} \mathbb{E} \left(\sigma_2(Z_i, Z_j) \theta_i^* \theta_j^* \bar{f}_i \bar{f}_j \right). \quad (17)$$

The first term in (17) is bounded by

$$\max_{1 \leq i \leq n} \left| \gamma_{ii}^{(2)} \right| n^{-2} \sum_{i=1}^n \mathbb{E} \left(\sigma_2(Z_i, Z_i) \theta_i^{*2} \bar{f}_i^2 \right) = o(n^{-1}),$$

by repeating the proof of Proposition B1. The second term in (17) is

$$\begin{aligned} & \frac{1}{n^4 a^{2q}} \mathbb{E} \left[\sum_{i,j,k,l} \gamma_{ij}^{(2)} \sigma_2(Z_i, Z_j) (\theta_i - \theta_k) (\theta_j - \theta_l) K_{ik} K_{jl} \bar{f}_i \bar{f}_j \right. \\ & \left. + \sum_{i,j,k} \left\{ (\theta_i - \theta_k) (\theta_j - \theta_k) K_{ik} K_{jk} + (\theta_i - \theta_k) (\theta_j - \theta_i) K_{ik} K_{ji} + (\theta_i - \theta_j) (\theta_j - \theta_k) K_{ij} K_{jk} \right\} \right. \\ & \left. \times \bar{f}_i \bar{f}_j \sigma_2(Z_i, Z_j) \right] - \frac{1}{n^4 a^{2q}} \sum_{i,j} \gamma_{ij}^{(2)} \mathbb{E} \left\{ \sigma_2(Z_i, Z_j) (\theta_i - \theta_j)^2 K_{ij}^2 \bar{f}_i \bar{f}_j \right\}. \quad (18) \end{aligned}$$

Lemma B6 and Assumption B13 imply that the contribution from the first term in square brackets is

$$o\left(n^{-1}a^2 + n^{-1/2}a^{2\min(\lambda+1,\nu)} + n^{-1/2}a^{\zeta-2q}\right) = o(n^{-1}).$$

The remaining contributions to (18) can likewise be shown to be $o(n^{-1})$. ■

Proposition B8 *As $n \rightarrow \infty$,*

$$S_{\xi_2 U} = o_p\left(n^{-1/2}\right).$$

Proof. Similar to that of Proposition B7. ■

Proposition B9 *As $n \rightarrow \infty$,*

$$n^{1/2}S_{V_2 U} \rightarrow_d \mathbf{N}(0, \Omega).$$

Proof. The left side is

$$n^{-1/2} \sum_{i=1}^n \left(V_{2i} U_i \hat{f}_i^2 - V_{2i} \bar{U}_i \hat{f}_i - \bar{V}_{2i} U_i \hat{f}_i + \bar{V}_{2i} \bar{U}_i \right). \quad (19)$$

By Lemma B5, the contribution from the last summand is $o_p(1)$. The contribution from the third summand in (19) is

$$n^{-1/2} \sum_{i=1}^n \left\{ \bar{V}_{2i} U_i \bar{f}_i + \bar{V}_i U_i \left(\hat{f}_i - \bar{f}_i \right) \right\} = o_p(1),$$

by Lemmas B4, B5 and B8 and the Cauchy inequality, and that from the second summand in (19) can similarly be shown to be $o_p(1)$. The contribution from the first summand in (19) is

$$n^{-1/2} \sum_{i=1}^n V_{2i} U_i \left\{ \bar{f}_i^2 + 2\bar{f}_i \left(\hat{f}_i - \bar{f}_i \right) + \left(\hat{f}_i - \bar{f}_i \right)^2 \right\}.$$

The proof is completed by applying Lemmas B4 and B10, and proceeding as in the proof of Lemma A1 and Theorem A. ■

Appendix 3 : Technical Lemmas for proofs of Theorems A and B

Lemma A1 *There exists an increasing sequence $N = N_n$ such that $N \rightarrow \infty$ as $n \rightarrow \infty$ and*

$$\lim_{n \rightarrow \infty} \mathbb{E} \left\| r_n - r_{(N)} \right\|^2 = 0.$$

Proof. By independence of the ε_k ,

$$\begin{aligned} \mathbb{E} \left\| r - r_{(N)} \right\|^2 &= \frac{1}{n} \sum_{k=N+1}^{\infty} \mathbb{E} \|W_k\|^2 \\ &= \frac{1}{n} \sum_{k=N+1}^{\infty} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E} (X'_{2i} X_{2j}) b_{ik} b_{jk} \\ &\leq \left(n \max_{1 \leq i \leq n} \sum_{k=N+1}^{\infty} b_{ik}^2 \right) \left\{ \frac{1}{n} \sum_{i=1}^n \left(\mathbb{E} \|X_{2i}\|^2 + 1 \right) \right\}^2. \end{aligned}$$

The result follows from Assumptions A1, A2 and Lemma C1. ■

We repeatedly use the following consequences of Definition 1:

$$\sup_{u \in \mathbb{R}^q} |K(u)| + \int \|u\| |K(u)| du + \int \|u\|^2 K^2(u) du < \infty; \quad \sup_{\|u\| \geq \delta/a} |K(u)| = O(a^\zeta) \text{ for all } \delta > 0.$$

We also introduce the abbreviations

$$\phi(z_1, z_2) = \theta(z_1) - \theta(z_2), \quad K(z_1, z_2) = K\left(\frac{z_2 - z_1}{a}\right).$$

Lemma B1 As $n \rightarrow \infty$,

$$n^{-3} \mathbb{E} \left\{ \sum_{i,j}^n (\theta_i - \theta_j)^2 K_{ij}^2 \right\} = o(a^{q+2} n^{-3/2}) + O(n^{-1} a^{q+2} + n^{-1} a^{2\zeta}).$$

Proof. The left side is

$$\begin{aligned} & \frac{1}{n^3} \int \phi(z_1, z_2)^2 K(z_1, z_2)^2 \sum_{i,j}^n f_{ij}(z_1, z_2) dz_1 dz_2 \leq \frac{1}{n} \left\{ \int \phi(z_1, z_2)^2 K(z_1, z_2)^2 \frac{1}{n^2} \sum_{i,j}^n F_{j:i}(z_2; z_1) dz_1 dz_2 \right. \\ & \left. + \int \phi(z_1, z_2)^2 K(z_1, z_2)^2 \bar{f}(z_1) \bar{f}(z_2) dz_1 dz_2 \right\}. \end{aligned} \quad (20)$$

Let

$$p(z, au) = \phi(z, z + au)^2 K^2(u) \frac{1}{n^2} \sum_{i,j}^n F_{j:i}(z + au; z).$$

The first integral in braces in (20) is

$$a^q \int_{\mathbb{R}^q} \int_{\mathbb{R}^q} p(z, au) dudz = a^q \left[\int_{\mathbb{R}^q} \int_{J_1(\varepsilon)} p(z, au) dudz + \int_{\mathbb{R}^q} \int_{J_2(\varepsilon)} p(z, au) dudz \right],$$

where

$$J_1(\varepsilon) = \{u : \|au\| < \varepsilon\}, \quad J_2(\varepsilon) = \{u : \|au\| \geq \varepsilon\}.$$

Let

$$B = \{z : \bar{f}(z) > 0\}, \quad m(z_1, z_2) = n^{-2} \bar{f}(z_1)^{-1} \sum_{i,j}^n |F_{j:i}(z_2; z_1)|.$$

Note that $B^C \times \mathbb{R}^q$, where B^C is the complement of B , is a null set with respect to the probability measure of Z_i, Z_j for all $i \neq j$. Then by Assumptions B6 and B9,

$$\begin{aligned} \int_{\mathbb{R}^q} \int_{J_1(\varepsilon)} |p(z, au)| dudz & \leq \int_B \int_{J_1(\varepsilon)} \phi(z, z + au)^2 K^2(u) m(z, z + au) \bar{f}(z) dudz \\ & \leq a^2 \left(\sup_{z_1 \in B} \sup_{z_2 \in \mathcal{N}(z_1)} m(z_1, z_2) \right) \int G^2(z) \bar{f}(z) dz \int \|u\|^2 K^2(u) du \\ & = o(a^2 n^{-1/2}). \end{aligned}$$

Now

$$\begin{aligned} \int_{\mathbb{R}^q} \int_{J_2(\varepsilon)} |p(z, au)| \, dudz &\leq \frac{1}{n^2 a^q} \sup_{\|au\| \geq \varepsilon} K^2(u) \int_{\mathbb{R}^{2q}} \phi(z_1, z_2)^2 \sum_{i,j}^n \{f_{ij}(z_1, z_2) + f_i(z_1) f_j(z_2)\} dz_1 dz_2 \\ &= O(a^{2\zeta - q}), \end{aligned}$$

because the double integral is

$$n^{-2} \sum_{i,j}^n \left\{ \mathbb{E}(\theta_i - \theta_j)^2 + \mathbb{E}(\theta_i^2) + \mathbb{E}(\theta_j^2) - 2\mathbb{E}(\theta_i) \mathbb{E}(\theta_j) \right\} = O(1), \quad (21)$$

by Assumption B9. Hence the first integral in braces in (20) is $o(a^{q+2}n^{-1/2}) + O(a^{2\zeta})$. The second integral in braces in (20) is

$$a^q \int \phi(z, z+au)^2 K^2(u) \bar{f}(z) \bar{f}(z+au) \, dzdu = a^q \left(\int_{\mathbb{R}^q} \int_{J_1(\varepsilon)} + \int_{\mathbb{R}^q} \int_{J_2(\varepsilon)} \right). \quad (22)$$

The first integral on the right in (22) is bounded by

$$a^2 \left(\sup_{z \in \mathbb{R}^q} \bar{f}(z) \right) \int \|u\|^2 K^2(u) \, du \int G^2(z) \bar{f}(z) \, dz = O(a^2),$$

and the second integral is bounded by

$$a^{-q} \sup_{\|au\| \geq \varepsilon} K^2(u) \left[\frac{2}{n} \sum_{i=1}^n \mathbb{E}(\theta_i^2) + 2 \left(\frac{1}{n} \sum_{i=1}^n \mathbb{E}(\theta_i) \right)^2 \right] = O(a^{2\zeta - q}).$$

Hence the second integral in (20) is $O(a^{q+2} + a^{2\zeta})$. ■

Lemma B2 As $n \rightarrow \infty$,

$$n^{-3} \mathbb{E} \left\{ \sum_{i,j,k}^n (\theta_i - \theta_j)(\theta_i - \theta_k) K_{ij} K_{ik} \right\} = o(n^{-1/2} a^{2q+2}) + O(a^\zeta + a^{2\{q+\min(\nu, \lambda+1)\}}).$$

Proof. With the abbreviation $s(z_1, z_2, z_3) = \phi(z_1, z_2)\phi(z_1, z_3)K(z_1, z_2)K(z_1, z_3)$, the left side is

$$\begin{aligned} &\frac{1}{n^3} \sum_{i,j,k}^n \int s(z_1, z_2, z_3) f_{ijk}(z_1, z_2, z_3) \prod_{i=1}^3 dz_i = \int s(z_1, z_2, z_3) \frac{1}{n^3} \sum_{i,j,k}^n F_{jk:i}(z_2, z_3; z_1) \prod_{i=1}^3 dz_i \\ &+ \int s(z_1, z_2, z_3) \frac{1}{n^3} \sum_{i,j,k}^n f_i(z_1) F_{k:j}(z_2; z_3) \prod_{i=1}^3 dz_i \\ &+ \int s(z_1, z_2, z_3) \frac{1}{n^3} \sum_{i,j,k}^n \{f_i(z_1) f_j(z_2) f_k(z_3) - \bar{f}(z_1) \bar{f}(z_2) \bar{f}(z_3)\} \prod_{i=1}^3 dz_i \\ &+ \frac{(n-1)(n-2)}{n^2} \int s(z_1, z_2, z_3) \bar{f}(z_1) \bar{f}(z_2) \bar{f}(z_3) \prod_{i=1}^3 dz_i. \end{aligned} \quad (23)$$

With the further abbreviation $p(z, u, v, a) = \phi(z, z + au) \phi(z, z + av) K(u) K(v)$, the first integral in (23) is

$$\begin{aligned} & \frac{a^{2q}}{n^3} \int p(z, u, v, a) \sum_{i,j,k}^n F_{jk:i}(z + au; z + av; z) dz dudv \\ &= a^{2q} \left(\int_{\mathbb{R}^q} \int_{J_1(\varepsilon)} + \int_{\mathbb{R}^q} \int_{J_2(\varepsilon)} + \int_{\mathbb{R}^q} \int_{J_3(\varepsilon)} + \int_{\mathbb{R}^q} \int_{J_4(\varepsilon)} \right), \end{aligned}$$

where

$$\begin{aligned} J_1(\varepsilon) &= \{u, v : \|au\| < \varepsilon, \|av\| < \varepsilon\}, \quad J_2(\varepsilon) = \{u, v : \|au\| < \varepsilon, \|av\| \geq \varepsilon\}, \\ J_3(\varepsilon) &= \{u, v : \|au\| \geq \varepsilon, \|av\| < \varepsilon\}, \quad J_4(\varepsilon) = \{u, v : \|au\| \geq \varepsilon, \|av\| \geq \varepsilon\}. \end{aligned}$$

Let $B = \{z_1 : \bar{f}(z_1) > 0\}$ and $m(z_1, z_2, z_3) = n^{-3} \bar{f}(z_1)^{-1} \sum_{i,j,k}^n |F_{jk:i}(z_2, z_3; z_1)|$. Then by Assumption B6 the first integral is bounded by

$$\begin{aligned} & a^2 \int_B \sup_{z_2, z_3 \in \mathcal{N}(z_1)} m(z_1, z_2, z_3) G^2(z_1) \int_{J_1(\varepsilon)} |K(u) K(v)| \|u\| \|v\| \bar{f}(z_1) dudvdz_1 \\ & \leq a^2 \left(\sup_{z_1 \in B} \sup_{z_2, z_3 \in \mathcal{N}(z_1)} m(z_1, z_2, z_3) \right) \int G^2(z_1) \bar{f}(z_1) dz_1 \left(\int \|u\| |K(u)| du \right)^2 = o(a^2 n^{-1/2}). \end{aligned}$$

By similar reasoning to that in (21) in the proof of Lemma B1,

$$\begin{aligned} \left| a^{2q} \int_{\mathbb{R}^q} \int_{J_2(\varepsilon)} \right| & \leq n^{-3} \sup_{\|av\| \geq \varepsilon} |K(v)| \sup_u |K(u)| \int |\phi(z_1, z_2) \phi(z_1, z_3)| \sum_{i,j,k}^n \{f_{ijk}(z_1, z_2, z_3) + f_i(z_1) f_{jk}(z_2, z_3)\} \prod_{i=1}^3 dz_i \\ & = O(a^\zeta). \end{aligned}$$

The same result holds for $\left| a^{2q} \int_{\mathbb{R}^q} \int_{J_3(\varepsilon)} \right|$. Finally

$$\begin{aligned} \left| a^{2q} \int_{\mathbb{R}^q} \int_{J_4(\varepsilon)} \right| & \leq n^{-3} \sup_{\|au\| \geq \varepsilon} K(u)^2 \int |\phi(z_1, z_2) \phi(z_1, z_3)| \sum_{i,j,k}^n \{f_{ijk}(z_1, z_2, z_3) + f_i(z_1) f_{jk}(z_2, z_3)\} \prod_{i=1}^3 dz_i \\ & = O(a^{2\zeta}). \end{aligned}$$

The first integral in (23) is thus $o(a^{2q+2} n^{-1/2}) + O(a^\zeta)$. The second integral in (23) is

$$\begin{aligned} & n^{-3} a^{2q} \int p(z, u, v, a) \sum_{i,j,k}^n f_i(z) F_{k:j}(z + av; z + au) dudvdz \\ &= a^{2q} \left[\int_{\mathbb{R}^q} \int_{J_1(\varepsilon)} + \int_{\mathbb{R}^q} \int_{J_2(\varepsilon)} + \int_{\mathbb{R}^q} \int_{J_3(\varepsilon)} + \int_{\mathbb{R}^q} \int_{J_4(\varepsilon)} \right]. \end{aligned}$$

Now

$$\frac{1}{n} \sum_{i,j,k}^n |f_i(z_1) F_{k:j}(z_3; z_2)| \leq \bar{f}(z_1) \sum_{i,j}^n |F_{j:i}(z_3; z_2)|.$$

Then proceeding as above, the second integral of (23) is $o(n^{-1/2}a^{2q+2}) + O(a^\zeta)$. Because

$$\begin{aligned} & \sum_{i,j,k}^n \{f_i(z_1) f_j(z_2) f_k(z_3) - \bar{f}(z_1) \bar{f}(z_2) \bar{f}(z_3)\} \\ &= \frac{3n-2}{n^2} \sum_{i,j,k}^n f_i(z_1) f_j(z_2) f_k(z_3) - \frac{(n-1)(n-2)}{n^2} \left[\sum_{i=1}^n \sum_{j=1}^n f_i(z_1) f_i(z_2) f_j(z_3) \right. \\ & \quad \left. + \sum_{i,j}^n f_i(z_1) f_j(z_2) \{f_i(z_3) + f_j(z_3)\} \right], \end{aligned}$$

proceeding as in the last part of the proof of Lemma B1, using Assumption B7, the third integral of (23) is $O(n^{-1}a^{2q+2} + n^{-1}a^\zeta)$. Finally by Assumptions B7, B9 and B12, Lemma 5 of Robinson (1988) implies that the last integral of (23) is $O(a^{2\{q+\min(\nu, \lambda+1)\}})$. ■

Lemma B3 As $n \rightarrow \infty$,

(i)

$$n^{-3} \sum_{i,j,k}^n \mathbb{E} \{ \sigma_2(Z_i, Z_i) (K_{ij} - a^q \bar{f}_i) (K_{ik} - a^q \bar{f}_i) \} = o(n^{-1/2}a^{2q}) + O(a^\zeta + a^{2(q+\lambda)});$$

(ii)

$$n^{-2} \sum_{i,j}^n \mathbb{E} \{ \sigma_2(Z_i, Z_i) (K_{ij} - a^q \bar{f}_i)^2 \} = O(a^q).$$

Proof. Denoting $g(z_1; z_2, z_3) = \sigma_2(z_1, z_1) \{K(z_1, z_2) - a^q \bar{f}(z_1)\} \{K(z_1, z_3) - a^q \bar{f}(z_1)\}$, the left side of (i) can be written

$$\begin{aligned} \frac{1}{n^3} \sum_{i,j,k}^n \int g(z_1; z_2, z_3) f_{ijk}(z_1, z_2, z_3) \prod_{i=1}^3 dz_i &= \frac{1}{n^3} \sum_{i,j,k}^n \int g(z_1; z_2, z_3) [f_{ijk}(z_1, z_2, z_3) - \bar{f}(z_1) \bar{f}(z_2) \bar{f}(z_3)] \prod_{i=1}^3 dz_i \\ & \quad + \frac{(n-1)(n-2)}{n^2} \int g(z_1; z_2, z_3) \bar{f}(z_1) \bar{f}(z_2) \bar{f}(z_3) \prod_{i=1}^3 dz_i. \end{aligned} \quad (24)$$

Writing $L(z_1; z_2, z_3) = \sigma_2(z_1, z_1) K(z_1, z_2) K(z_1, z_3)$, the first integral in (24) is

$$\begin{aligned} & \frac{1}{n^3} \int L(z_1; z_2, z_3) \sum_{i,j,k}^n \{f_{ijk}(z_1, z_2, z_3) - \bar{f}(z_1) \bar{f}(z_2) \bar{f}(z_3)\} \prod_{i=1}^3 dz_i \\ & - a^q \frac{1}{n^3} \int \sigma_2(z_1, z_1) K(z_1, z_3) \bar{f}(z_1) \sum_{i,j,k}^n \{f_{ik}(z_1, z_3) - \bar{f}(z_1) \bar{f}(z_3)\} dz_1 dz_3 \\ & - a^q \frac{1}{n^3} \int \sigma_2(z_1, z_1) K(z_1, z_2) \bar{f}(z_1) \sum_{i,j,k}^n \{f_{ij}(z_1, z_2) - \bar{f}(z_1) \bar{f}(z_2)\} dz_1 dz_2. \end{aligned} \quad (25)$$

The first term is

$$\frac{1}{n^3} \int L(z_1; z_2, z_3) \sum_{i,j,k}^n \left[F_{jk:i}(z_2, z_3; z_1) + f_i(z_1) F_{k:j}(z_3; z_2) + \{f_i(z_1) f_j(z_2) f_k(z_3) - \bar{f}(z_1) \bar{f}(z_2) \bar{f}(z_3)\} \right] \prod_{i=1}^3 dz_i,$$

which, as in Lemma B2, is $o(n^{-1/2}a^{2q}) + O(a^\zeta)$. The last two terms in (25) are bounded in absolute value by

$$\frac{2a^q}{n^2} \left\{ \int \sigma_2(z_1, z_1) \bar{f}(z_1) |K(z_1, z_2)| \sum_{i,j}^n |f_{ij}(z_1, z_2) - \bar{f}(z_1) \bar{f}(z_2)| dz_1 dz_2 \right\},$$

which, by Assumption B6, can be shown to be $o(n^{-1/2}a^{2q}) + O(a^{\zeta+q})$. Finally by Lemma 4 of Robinson (1988) and Assumption B4, the second integral in (24) is $O(a^{2(q+\lambda)})$.

The left side of (ii) is bounded by

$$n^{-2} \int |g(z_1; z_2, z_2)| \sum_{i,j}^n |F_{j;i}(z_2; z_1)| dz_1 dz_2 + \int |g(z_1; z_2, z_2)| \bar{f}(z_1) \bar{f}(z_2) dz_1 dz_2.$$

To estimate the first integral complete the square and proceed as in Lemma B1. The second integral is dominated by $a^q \sup_z \bar{f}(z) \int \sigma_2(z, z) K^2(u) \bar{f}(z) dudz = O(a^q)$. ■

Lemma B4 As $n \rightarrow \infty$,

$$n^{-1} \sum_{i=1}^n \left(U_i^2 + \|V_{1i}\|^2 + \|V_{2i}\|^2 \right) \left(\hat{f}_i - \bar{f}_i \right)^2 = o_p \left(n^{-\frac{1}{2}} \right). \quad (26)$$

Proof. By Assumption B4, the expectation of the last contribution to (26) is

$$\begin{aligned} & \mathbb{E} \left\{ \frac{1}{n} \sum_{i=1}^n \gamma_{ii}^{(2)} \sigma_2(Z_i, Z_i) \left(\hat{f}_i - \bar{f}_i \right)^2 \right\} \\ & \leq \max_{1 \leq i \leq n} \left| \gamma_{ii}^{(2)} \right| \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left| \sigma_2(Z_i, Z_i) \left(\frac{1}{na^q} \sum_{j \neq i}^n M_{ij} - \frac{\bar{f}_i}{n} \right)^2 \right| \\ & \leq \frac{C}{n^3} \sum_{i=1}^n \mathbb{E} \left\{ |\sigma_2(Z_i, Z_i)| \left(\frac{1}{a^q} \sum_{j \neq i}^n M_{ij} \right)^2 + |\sigma_2(Z_i, Z_i)| \bar{f}_i^2 \right\}. \end{aligned}$$

where $M_{ij} = K_{ij} - a^q \bar{f}_i$. By Assumption B7 the contribution from the second term in brackets is $O(n^{-2})$. That from the first term is

$$\frac{C}{n^3 a^{2q}} \sum_{i,j,k}^n \mathbb{E} \{ |\sigma_2(Z_i, Z_i)| M_{ij} M_{ik} \} + \frac{C}{n^3 a^{2q}} \sum_{i,j}^n \mathbb{E} \{ |\sigma_2(Z_i, Z_i)| M_{ij}^2 \}.$$

Lemma B3 and Assumption B13 imply that

$$\mathbb{E} \left\{ \frac{1}{n} \sum_{i=1}^n \|V_{2i}\|^2 \left(\hat{f}_i - \bar{f}_i \right)^2 \right\} = o(n^{-1/2}) + O(a^{\zeta-2q} + a^{2(\lambda+1)} + n^{-1}a^{-q}) = o(n^{-1/2}).$$

The remainder of the proof is very similar. ■

Lemma B5 As $n \rightarrow \infty$,

$$\mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n \bar{U}_i^2 \right) = o(n^{-1/2}), \quad \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n \|\bar{V}_{1i}\|^2 \right) = o(1), \quad \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n \|\bar{V}_{2i}\|^2 \right) = o(n^{-1/2}).$$

Proof. The last expectation is

$$\begin{aligned} \mathbb{E} \left(\frac{1}{n^3 a^{2q}} \sum_{i=1}^n \sum_{j \neq i}^n \sum_{k \neq i}^n V_{2j}' V_{2k} K_{ij} K_{ik} \right) &= \frac{1}{n^3 a^{2q}} \sum_{i,j,k} \gamma_{jk}^{(2)} \mathbb{E} (\sigma_2(Z_j, Z_k) K_{ij} K_{ik}) \\ &\quad + \frac{1}{n^3 a^{2q}} \sum_{i,j} \gamma_{jj}^{(2)} \mathbb{E} (\sigma_2(Z_j, Z_j) K_{ij}^2). \end{aligned} \quad (27)$$

Denoting $l(z_1; z_2, z_3) = \sigma_2(z_2, z_3) K(z_1, z_2) K(z_1, z_3)$, the first term on the right is

$$\begin{aligned} &\frac{1}{n^3 a^{2q}} \int l(z_1; z_2, z_3) \sum_{i,j,k} \gamma_{jk}^{(2)} F_{jk;i}(z_2, z_3; z_1) \prod_{i=1}^3 dz_i + \frac{1}{n^3 a^{2q}} \int l(z_1; z_2, z_3) \sum_{i,j,k} \gamma_{jk}^{(2)} f_i(z_1) F_{k;j}(z_3; z_2) \prod_{i=1}^3 dz_i \\ &+ \frac{1}{n^3 a^{2q}} \int l(z_1; z_2, z_3) \sum_{i,j,k} \gamma_{jk}^{(2)} \{f_i(z_1) f_j(z_2) f_k(z_3) - \bar{f}(z_1) \bar{f}(z_2) \bar{f}(z_3)\} \prod_{i=1}^3 dz_i \\ &+ \frac{1}{n^3 a^{2q}} \sum_{i,j,k} \gamma_{jk}^{(2)} \int l(z_1; z_2, z_3) \prod_{i=1}^3 \{\bar{f}(z_i) dz_i\}. \end{aligned} \quad (28)$$

The last term in (28) is bounded in absolute value by

$$\begin{aligned} &\frac{1}{n^3 a^{2q}} \sum_{i,j,k} \left| \gamma_{jk}^{(2)} \right| \int |\sigma_2(z_2, z_3) - \sigma_2(z_1, z_1)| |K(z_1, z_2) K(z_1, z_3)| \prod_{i=1}^3 \{\bar{f}(z_i) dz_i\} \\ &+ \frac{1}{n^3 a^{2q}} \sum_{i,j,k} \left| \gamma_{jk}^{(2)} \right| \int |\sigma_2(z_1, z_1) K(z_1, z_2) K(z_1, z_3)| \prod_{i=1}^3 \{\bar{f}(z_i) dz_i\}. \end{aligned} \quad (29)$$

Applying the last part of the proof of Lemma B1, Assumptions B7 and B10 imply that the integral of the first term in (29) is $O(a^{2q+1} + a^\zeta)$. Hence by Assumptions B5 and B13, the first term of (29) is $o(n^{-1/2})$. The second term in (29) is bounded by

$$\begin{aligned} &\frac{1}{n^3} \sum_{i,j,k} \left| \gamma_{jk}^{(2)} \right| \int |K(u) K(v)| |\sigma_2(z, z)| \bar{f}(z) \bar{f}(z+au) \bar{f}(z+av) dudvdz \\ &\leq \frac{1}{n^2} \sum_{i,j} \left| \gamma_{jk}^{(2)} \right| \left(\sup_{z \in \mathbb{R}^q} \bar{f}(z) \right)^2 \left(\int |K(u)| du \right)^2 \int |\sigma_2(z, z)| \bar{f}(z) dz = o(n^{-1/2}) \end{aligned}$$

by Assumptions B5 and B7. For other terms in (28), apply the proof of Lemma B2. Altogether it is found that the first term of (27) is $o(n^{-1/2}) + O(a^\zeta - 2q)$.

The second term of (27) is bounded by

$$\begin{aligned} &\max_{1 \leq i \leq n} \left| \gamma_{ii}^{(2)} \right| \frac{1}{n^3 a^{2q}} \sum_{i,j} \mathbb{E} |\sigma_2(Z_j, Z_j) K_{ij}^2| \\ &\leq \frac{C}{n^3 a^{2q}} \sum_{i,j} \{ \mathbb{E} |\sigma_2(Z_j, Z_j) - \sigma_2(Z_i, Z_i)| K_{ij}^2 + \mathbb{E} |\sigma_2(Z_i, Z_i) K_{ij}^2| \}. \end{aligned} \quad (30)$$

Applying the proof of Lemma B1, (30) is $O(n^{-3/2}a^{-q}) + O(n^{-1}a^{-q} + n^{-1}a^{2\zeta-2q})$. This proves the last result. The others can be shown similarly. ■

Lemma B6 *As $n \rightarrow \infty$,*

$$\frac{1}{n^3} \sum_{i,j,k,l} \gamma_{ij}^{(2)} \mathbb{E} \{ \sigma_2(Z_i, Z_j) (\theta_i - \theta_k) (\theta_j - \theta_l) K_{ik} K_{jl} \bar{f}_i \bar{f}_j \} = o(a^{2+2q} + n^{1/2}a^\zeta + n^{1/2}a^{2\{q+\min(\lambda+1, \nu)\}}).$$

Proof. Writing $u(z_1, z_2, z_3, z_4) = \sigma_2(z_1, z_2) \phi(z_1, z_3) \phi(z_2, z_4) K(z_1, z_3) K(z_2, z_4)$, the left side is

$$\begin{aligned} & n^{-3} \int u(z_1, z_2, z_3, z_4) \bar{f}(z_1) \bar{f}(z_2) \sum_{i,j,k,l} \gamma_{ij}^{(2)} F_{ij:k:l}(z_1, z_2; z_3, z_4) \prod_{i=1}^4 dz_i \\ & + n^{-3} \int u(z_1, z_2, z_3, z_4) \bar{f}(z_1) \bar{f}(z_2) \sum_{i,j,k,l} \gamma_{ij}^{(2)} f_{ij}(z_1, z_2) \{ f_k(z_3) f_l(z_4) - \bar{f}(z_3) \bar{f}(z_4) \} \prod_{i=1}^4 dz_i \\ & + \frac{1}{n^3} \sum_{i,j,k,l} \gamma_{ij}^{(2)} \int u(z_1, z_2, z_3, z_4) f_{ij}(z_1, z_2) \left\{ \prod_{i=1}^4 \bar{f}(z_i) dz_i \right\}. \end{aligned} \quad (31)$$

As in Lemma B2, the first integral is $o(a^{2+2q}) + o(n^{1/2}a^\zeta)$. Similarly, the second term in (31) can be shown to be of no greater order. The integral of the last term of (31) is bounded in absolute value by

$$\begin{aligned} & \left\{ \sup_{z_1, z_2} f_{ij}(z_1, z_2) \right\} \int_{\mathbb{R}^{2q}} \left| \int_{\mathbb{R}^q} \phi(z_1, z_3) K(z_1, z_3) \bar{f}(z_3) dz_3 \right| \\ & \left| \int_{\mathbb{R}^q} \phi(z_2, z_4) K(z_2, z_4) \bar{f}(z_4) dz_4 \right| \sigma_2(z_1, z_2) \bar{f}(z_1) \bar{f}(z_2) dz_1 dz_2 \\ & = O(a^{2\{q+\min(\lambda+1, \nu)\}}) \end{aligned}$$

by Lemma 4 of Robinson (1988), Assumptions B7 and the Cauchy inequality. Thus the last term in (31) is $o(n^{1/2}a^{2\{q+\min(\lambda+1, \nu)\}})$ by Assumption B5. ■

Lemma B7 *For distinct i, j, k and l , uniformly in $1 \leq i, j, k, l \leq n$, $n \geq 1$,*

$$\begin{aligned} & \mathbb{E} \left\{ \left| \sigma_2(Z_k, Z_l) K_{ik} K_{jl} \bar{f}_i \bar{f}_j \right| + \left| \sigma_2(Z_i, Z_k) K_{ik} K_{ij} \bar{f}_i \bar{f}_j \right| + \left| \sigma_2(Z_k, Z_k) K_{ik} K_{jk} \bar{f}_i \bar{f}_j \right| \right. \\ & \left. + \left| \sigma_2(Z_j, Z_k) K_{ij} K_{jk} \bar{f}_i \bar{f}_j \right| + \left| \sigma_2(Z_j, Z_k) K_{ij} K_{ik} \bar{f}_i^2 \right| \right\} \\ & = O(a^{2q}), \end{aligned}$$

and

$$\mathbb{E} \left\{ \left| \sigma_2(Z_i, Z_j) K_{ij}^2 \bar{f}_i \bar{f}_j \right| + \left| \sigma_2(Z_j, Z_j) K_{ij}^2 \bar{f}_i^2 \right| \right\} = O(a^q).$$

Proof. Writing $l_{ij:kl}(z_1, z_2; z_3, z_4) = K(z_1, z_3) K(z_2, z_4) \bar{f}(z_1) \bar{f}(z_2) f_{ijkl}(z_1, z_2, z_3, z_4)$

$$\begin{aligned} & \mathbb{E} \left| \sigma_2(Z_k, Z_l) K_{ik} K_{jl} \bar{f}_i \bar{f}_j \right| \\ & = \int |\sigma_2(z_3, z_4) l_{ij:kl}(z_1, z_2; z_3, z_4)| \prod_{i=1}^4 dz_i \\ & \leq \int \{ |\sigma_2(z_3, z_4) - \sigma_2(z_1, z_2)| + |\sigma_2(z_1, z_2)| \} |l_{ij:kl}(z_1, z_2, z_3, z_4)| \prod_{i=1}^4 dz_i. \end{aligned} \quad (32)$$

The second term in (32) is bounded by

$$a^{2q} \sup f_{ijkl}(z_1, z_2, z_3, z_4) \int |\sigma_2(z_1, z_2)| \bar{f}(z_1) \bar{f}(z_2) dz_1 dz_2 \left(\int |K(u)| du \right)^2.$$

By Assumption B7, it is uniformly $O(a^{2q})$. Writing $p(z_1, z_2, u, v, a) = |\sigma_2(z_1 + au, z_2 + av) - \sigma_2(z_1, z_2)| |K(u)K(v)|$, the first term in (32) is

$$\begin{aligned} & a^{2q} \int p(z_1, z_2, u, v, a) \bar{f}(z_1) \bar{f}(z_2) f_{ijkl}(z_1, z_2, z_1 + au, z_2 + av) dz_1 dz_2 dudv \\ &= \left(\int \int_{J_1(\varepsilon/2)} + \int \int_{J_2(\varepsilon/2)} + \int \int_{J_3(\varepsilon/2)} + \int \int_{J_4(\varepsilon/2)} \right), \end{aligned} \quad (33)$$

where $J_i(\varepsilon)$, $i = 1, \dots, 4$ are defined as in the proof of Lemma B2. By Assumptions B7 and B10, the first integral is uniformly $O(a)$. Since

$$\begin{aligned} \int \int_{J_2(\varepsilon/2)} &\leq \sup_{\|av\| \geq \varepsilon/2} |K(v)| \sup_u |K(u)| a^{-2q} \left\{ \sup \bar{f}(z)^2 \mathbb{E} |\sigma_2(Z_i, Z_j)| \right. \\ &\quad \left. + \sup f_{ij}(z_1, z_2) \int |\sigma_2(z_1, z_2)| \bar{f}(z_1) \bar{f}(z_2) dz_1 dz_2 \right\}, \end{aligned}$$

Assumption B7 and B10 imply that $\int \int_{J_2(\varepsilon)}$ is uniformly $O(n^{1/2} a^{\zeta-2q})$. Similarly for the other terms in (33). The remaining terms of the lemma can be dealt with similarly. ■

Lemma B8 As $n \rightarrow \infty$,

$$\sum_{i=1}^n \bar{V}_{2i} U_i \bar{f}_i = o_p(n^{1/2}), \quad \sum_{i=1}^n V_{2i} \bar{U}_i \bar{f}_i = o_p(n^{1/2}).$$

Proof. The expectation of the squared norm of the first sum is

$$\begin{aligned} \mathbb{E} \left\| \frac{1}{na^q} \sum_{i,j} U_i V_{2j} K_{ij} \bar{f}_i \right\|^2 &\leq \frac{1}{n^2 a^{2q}} \max_{1 \leq i \leq n} \gamma_{ii}^{(U)} \mathbb{E} \left(\sum_{i=1}^n \sum_{j \neq i} \sum_{k \neq i} V_{2j}' V_{2k} K_{ij} K_{ik} \bar{f}_i^2 \right) \\ &+ \frac{1}{n^2 a^{2q}} \sum_{i,j} \sum_{k \neq i} \sum_{l \neq j} \gamma_{ij}^{(U)} \gamma_{kl}^{(2)} \mathbb{E} (\sigma_2(Z_k, Z_l) K_{ik} K_{jl} \bar{f}_i \bar{f}_j). \end{aligned} \quad (34)$$

The first term in (34) is bounded in absolute value by

$$\frac{C}{n^2 a^{2q}} \left\{ \max_{1 \leq j \leq n} |\gamma_{jj}^{(2)}| \sum_{i,j} \mathbb{E} |\sigma_2(Z_j, Z_j) K_{ij}^2 \bar{f}_i^2| + \sum_{i,j,k} |\gamma_{jk}^{(2)}| \mathbb{E} |\sigma_2(Z_j, Z_k) K_{ij} K_{ik} \bar{f}_i^2| \right\}. \quad (35)$$

By Lemma B7 the double sum in (35) is $O(n^2 a^q)$ and, with Assumption B5, the the triple sum in (35) is $o(n^{5/2} a^{2q})$. Hence the first term in (34) is $O(a^{-q}) + o(n^{1/2}) = o(n)$. The second term in (34) is

bounded in absolute value by

$$\begin{aligned} & \frac{1}{n^2 a^{2q}} \left\{ \sum_{i,j,k,l}^n \left| \gamma_{ij}^{(U)} \gamma_{kl}^{(2)} \right| \mathbb{E} \left| \sigma_2(Z_k, Z_l) K_{ik} K_{jl} \bar{f}_i \bar{f}_j \right| + \sum_{i,j,k}^n \left| \gamma_{ij}^{(U)} \gamma_{ik}^{(2)} \right| \mathbb{E} \left| \sigma_2(Z_i, Z_k) K_{ik} K_{ij} \bar{f}_i \bar{f}_j \right| \right. \\ & + \sum_{i,j,k}^n \left| \gamma_{ij}^{(U)} \gamma_{kk}^{(2)} \right| \mathbb{E} \left| \sigma_2(Z_k, Z_k) K_{ik} K_{jk} \bar{f}_i \bar{f}_j \right| + \sum_{i,j}^n \left| \gamma_{ij}^{(U)} \gamma_{ij}^{(2)} \right| \mathbb{E} \left| \sigma_2(Z_i, Z_j) K_{ij}^2 \bar{f}_i \bar{f}_j \right| \\ & \left. + \sum_{i,j,k}^n \left| \gamma_{ij}^{(U)} \gamma_{jk}^{(2)} \right| \mathbb{E} \left| \sigma_2(Z_j, Z_k) K_{ij} K_{jk} \bar{f}_i \bar{f}_j \right| \right\}. \end{aligned}$$

By Lemma B7 and Assumption B5, the second term in (34) is $o(n + n^{1/2} + n^{-1/2} a^{-q}) = o(n)$. This proves the first result. The other can be shown similarly. ■

Lemma B9 As $n \rightarrow \infty$,

$$\frac{1}{n^3} \mathbb{E} \sum_{i,j,k,l}^n \gamma_{ij}^{(2)} \gamma_{ij}^{(U)} \sigma_2(Z_i, Z_j) (K_{ik} - a^q \bar{f}_i) (K_{jl} - a^q \bar{f}_j) \bar{f}_i \bar{f}_j = o\left(a^{2q} + n^{1/2} a^\zeta + n^{1/2} a^{2(\lambda+q)}\right).$$

Proof. Writing $v(z_1; z_2) = K(z_1, z_2) - a^q \bar{f}(z_1)$ and

$$w(z_1, z_2; z_3; z_4) = n^{-3} \sum_{i,j,k,l}^n \gamma_{ij}^{(2)} \gamma_{ij}^{(U)} \{f_{ijkl}(z_1, z_2, z_3, z_4) - f_{ij}(z_1, z_2) \bar{f}(z_3) \bar{f}(z_4)\},$$

the left side is

$$\begin{aligned} & \int \sigma_2(z_1, z_2) v(z_1; z_3) v(z_2; z_4) \bar{f}(z_1) \bar{f}(z_2) w(z_1, z_2; z_3; z_4) \prod_{i=1}^4 dz_i \\ & + \frac{1}{n^3} \sum_{i,j,k,l}^n \gamma_{ij}^{(2)} \gamma_{ij}^{(U)} \int \sigma_2(z_1, z_2) v(z_1; z_3) v(z_2; z_4) f_{ij}(z_1, z_2) \prod_{i=1}^4 \{\bar{f}(z_i) dz_i\}. \end{aligned} \quad (36)$$

The first term in (36) is

$$\begin{aligned} & \int \sigma_2(z_1, z_2) K(z_1, z_3) v(z_2; z_4) w(z_1, z_2; z_3; z_4) \bar{f}(z_1) \bar{f}(z_2) \prod_{i=1}^4 dz_i \\ & - a^q \int \sigma_2(z_1, z_2) K(z_2, z_4) w(z_1, z_2; z_3; z_4) \bar{f}^2(z_1) \bar{f}(z_2) \prod_{i=1}^4 dz_i, \end{aligned} \quad (37)$$

because

$$\int_{\mathbb{R}^{2q}} \{f_{ijkl}(z_1, z_2, z_3, z_4) - f_{ij}(z_1, z_2) \bar{f}(z_3) \bar{f}(z_4)\} dz_3 dz_4 \equiv 0.$$

A leading term in (37) is

$$\begin{aligned} & \int \sigma_2(z_1, z_2) K(z_1, z_3) K(z_2, z_4) \bar{f}(z_1) \bar{f}(z_2) \left[\frac{1}{n^3} \sum_{i,j,k,l}^n \gamma_{ij}^{(2)} \gamma_{ij}^{(U)} F_{ij:k:l}(z_1, z_2; z_3; z_4) \right. \\ & \left. + \frac{1}{n^3} \sum_{i,j,k,l}^n \gamma_{ij}^{(2)} \gamma_{ij}^{(U)} f_{ij}(z_1, z_2) \{f_k(z_3) f_l(z_4) - \bar{f}(z_3) \bar{f}(z_4)\} \right] \prod_{i=1}^4 dz_i. \end{aligned} \quad (38)$$

Similar to the proof of Lemma B2, the integral of the first sum in (38) can be shown to be $o(a^{2q} + n^{1/2}a^q)$. Proceeding as in the proofs of Lemmas B6 and B3, remaining terms can be dealt with such that the first term in (36) is $o(a^{2q} + n^{1/2}a^q)$. Proceeding as in the proof of Lemma B6, the second term in (36) is $o(n^{1/2}a^{2(\lambda+q)})$ by Assumptions B5 and B7 and Lemma 4 of Robinson (1988). ■

Lemma B10 As $n \rightarrow \infty$,

$$\mathbb{E} \left\| \sum_{i=1}^n V_{2i} U_i (\hat{f}_i - \bar{f}_i) \bar{f}_i \right\|^2 = o(n).$$

Proof. The left side is

$$\begin{aligned} & \mathbb{E} \sum_{i=1}^n \gamma_{ii}^{(2)} \gamma_{ii}^{(U)} \sigma_2(Z_i, Z_i) (\hat{f}_i - \bar{f}_i)^2 \bar{f}_i^2 \\ & + \mathbb{E} \sum_{i,j}^n \gamma_{ij}^{(2)} \gamma_{ij}^{(U)} \sigma_2(Z_i, Z_j) (\hat{f}_i - \bar{f}_i) (\hat{f}_j - \bar{f}_j) \bar{f}_i \bar{f}_j. \end{aligned} \quad (39)$$

The first term in (39) is bounded by

$$\max_{1 \leq i \leq n} \left| \gamma_{ii}^{(2)} \right| \max_{1 \leq i \leq n} \gamma_{ii}^{(U)} \sup_{z \in \mathbb{R}^q} \bar{f}(z)^2 \sum_{i=1}^n \mathbb{E} \left| \sigma_2(Z_i, Z_i) (\hat{f}_i - \bar{f}_i)^2 \right| = o(n^{1/2})$$

by the proof of Lemma B4 and Assumptions B1, B5 and B7. The second term in (39) is

$$\begin{aligned} & \frac{1}{n^2 a^{2q}} \mathbb{E} \sum_{i,j}^n \sum_{k \neq i}^n \sum_{l \neq j}^n \gamma_{ij}^{(2)} \gamma_{ij}^{(U)} \sigma_2(Z_i, Z_j) M_{ik} M_{jl} \bar{f}_i \bar{f}_j \\ & - \frac{1}{n^2 a^q} \mathbb{E} \sum_{i,j}^n \sum_{l \neq j}^n \gamma_{ij}^{(2)} \gamma_{ij}^{(U)} \sigma_2(Z_i, Z_j) M_{jl} \bar{f}_i^2 \bar{f}_j \\ & - \frac{1}{n^2 a^q} \mathbb{E} \sum_{i,j}^n \sum_{k \neq i}^n \gamma_{ij}^{(2)} \gamma_{ij}^{(U)} \sigma_2(Z_i, Z_j) M_{ik} \bar{f}_i \bar{f}_j^2 + \frac{1}{n^2} \mathbb{E} \sum_{i,j}^n \gamma_{ij}^{(2)} \gamma_{ij}^{(U)} \sigma_2(Z_i, Z_j) \bar{f}_i^2 \bar{f}_j^2. \end{aligned} \quad (40)$$

By Assumptions B5 and B7, the last term in (40) is $o(n^{-1/2})$. The absolute value of the second term in (40) is bounded by

$$\frac{1}{n^2 a^q} \sum_{i,j}^n \sum_{l \neq j}^n \left| \gamma_{ij}^{(2)} \gamma_{ij}^{(U)} \right| \left\{ \mathbb{E} \left| \sigma_2(Z_i, Z_j) K_{jl} \bar{f}_i \bar{f}_j \right| + a^q \mathbb{E} \left| \sigma_2(Z_i, Z_j) \bar{f}_i^2 \bar{f}_j^2 \right| \right\}. \quad (41)$$

By Assumptions B7 and B10, the last expectation is uniformly bounded, whereas the first is bounded by

$$a^q \sup_{f_{ijl}(z_1, z_2, z_3)} \sup \bar{f}(z) \int |\sigma_2(z_1, z_2)| \bar{f}(z_1) \bar{f}(z_2) |K(u)| du dz_1 dz_2$$

which, by Assumption B7, is uniformly $O(a^q)$. Thus by Assumption B5, (41) is $o(n^{1/2})$. The same conclusion can be drawn for the third term in (40). The first term in (40) is

$$\begin{aligned} & \frac{1}{n^2 a^{2q}} \mathbb{E} \sum_{i,j,k,l}^n \gamma_{ij}^{(2)} \gamma_{ij}^{(U)} \sigma_2(Z_i, Z_j) M_{ik} M_{jl} \bar{f}_i \bar{f}_j + \frac{1}{n^2 a^{2q}} \mathbb{E} \sum_{i,j}^n \gamma_{ij}^{(2)} \gamma_{ij}^{(U)} \sigma_2(Z_i, Z_j) M_{ij} M_{ji} \bar{f}_i \bar{f}_j \\ & + \frac{1}{n^2 a^{2q}} \mathbb{E} \sum_{i,j,k}^n \gamma_{ij}^{(2)} \gamma_{ij}^{(U)} \sigma_2(Z_i, Z_j) (M_{ik} M_{ji} + M_{ik} M_{jk} + M_{ij} M_{jk}) \bar{f}_i \bar{f}_j. \end{aligned}$$

Lemma B9 and Assumption B13 imply that the first term is $o(n)$. Other terms can likewise be shown to be $o(n)$. ■

Lemma C1 For all $1 \leq i \leq n$, $n \geq 1$, let $c_{ijn} \geq 0$ for all $j \geq 1$ and $\sum_{j=1}^{\infty} c_{ijn} < C$. Then for any $K < \infty$, there exists a sequence $\{N_n\}$ increasing in n without bound such that

$$n^K \max_{1 \leq i \leq n} \sum_{j=N_n+1}^{\infty} c_{ijn} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Proof. Fix $n \geq 1$ and $1 \leq i \leq n$. There exists M_{in} such that $\sum_{j=m+1}^{\infty} c_{ijn} < n^{-K-1}$ for all $m \geq M_{in}$. Let $M_n = \max_{1 \leq i \leq n} M_{in}$. Then for each $n \geq 1$, $\max_{1 \leq i \leq n} \sum_{j=m+1}^{\infty} c_{ijn} < n^{-K-1}$ for all $m \geq M_n$. Put $N_n = \max(N_{n-1}, M_n) + 1$. Then $n^K \max_{1 \leq i \leq n} \sum_{j=N_n+1}^{\infty} c_{ijn} < n^{-1} \rightarrow 0$ as $n \rightarrow \infty$. ■

Appendix 4: Proof of Theorem C

Each element of

$$\mathbb{E} \left(\Sigma_n - \tilde{\Sigma}_3 \right) = \sum_{u \in \mathbb{L}^*} S_n(u) \{1 - w(u, m)\}$$

is bounded in absolute value by that of

$$\sum_{u \in \mathbb{L}^*} G_u \{1 - w(u, m)\}.$$

Then proceed as Robinson (2007a) and conclude that

$$\mathbb{E} \left(\Sigma_n - \tilde{\Sigma}_3 \right) = O \left(k_q \sum_{k=1}^r m_k^{-q} \sum_{u \in \mathbb{Z}^r} |u_k|^q G_u \right).$$

The variance of the (t, s) -th element of $\tilde{\Sigma}_3$ is, by Assumption C4,

$$\begin{aligned} & \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \mathbb{E}(x_{ti} x_{tk} U_i U_k) \mathbb{E}(x_{sj} x_{sl} U_j U_l) w(s_i - s_j, m) w(s_k - s_l, m) \\ & + \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n \mathbb{E}(x_{ti} x_{sl} U_i U_l) \mathbb{E}(x_{sj} x_{tk} U_j U_k) w(s_i - s_j, m) w(s_k - s_l, m) + O(n^{-1}). \end{aligned}$$

The first term has modulus

$$\begin{aligned} & \left| n^{-2} \sum_{u \in \mathbb{L}^*} \sum_{v \in \mathbb{L}^*} w(u, m) w(v, m) \sum_{u_1 \in \mathbb{L}^*} S_{1rs,n}(u, v, u_1) \right| \\ & \leq n^{-1} \sum_{u \in \mathbb{L}^*} \sum_{v \in \mathbb{L}^*} |w(u, m) w(v, m)| \sum_{u_1 \in \mathbb{L}^*} \gamma_{u_1, u_1+v-u} \\ & \leq n^{-1} \sum_{u_1 \in \mathbb{L}^*} \sum_{v_1 \in \mathbb{L}^{**}} \gamma_{u_1, u_1-v_1} \sum_{u \in \mathbb{L}^*} |w(u, m)| \\ & \leq cn^{-1} \prod_{k=1}^d m_k \sum_{u \in \mathbb{Z}^r} \sum_{v \in \mathbb{Z}^r} \gamma_{u,v}. \end{aligned}$$

The second term can be handled similarly.

Acknowledgement

This research was supported by ESRC Grant RES-062-23-0036. We are grateful for the constructive comments of two referees.

References

- Banerjee, A. V., and L. Iyer, 2005, History, institutions and economic performance: the legacy of colonial land tenure systems in India. *American Economic Review* 95, 1190-1213.
- Bolthausen, E., 1982, On the central limit theorem for stationary mixing random fields. *Annals of Probability* 10, 1047-1050.
- Castellana, J. V., and M. R. Leadbetter, 1986, On smoothed probability density estimation for stationary processes. *Stochastic Processes and Their Applications* 21, 179-193.
- Chamberlain G., 1992, Efficiency bounds for semiparametric regression. *Econometrica* 60, 567-596.
- Conley, T. G., 1999, GMM estimation with cross sectional dependence. *Journal of Econometrics* 92, 1-45.
- Eicker, F., 1967, Limit theorems for regression with unequal and dependent errors. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, University of California Press, Berkeley, 59-82.
- Fan, Y., and Q. Li, 1999, Root-n-consistent estimation of partially linear time series models. *Journal of Nonparametric Statistics* 11, 251-269.
- Gao, J.T., 1998, Semiparametric regression smoothing of nonlinear time series. *Scandinavian Journal of Statistics* 25, 521-539.
- Gao, J.T., and T. Yee, 2000, Adaptive estimation in partially linear autoregressive models. *Canadian Journal of Statistics* 28, 571-586.
- Hidalgo, J., 1992, Adaptive estimation in time series regression models with heteroskedasticity of unknown form. *Econometric Theory* 8, 161-187.
- Jenish, N., and I. R. Prucha, 2008, Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics*, forthcoming.
- Kelejian, H. H., and I. R. Prucha, 2007, HAC estimation in a spatial framework. *Journal of Econometrics* 140, 131-154.
- Masry, E., 1983, Non-parametric covariance estimation from irregularly-spaced data. *Advances in Applied Probability* 15, 113-132.
- Mitchell, T. D., and P. D. Jones, 2005, An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology* 25, 693-712.
- Munshi, K., 2004, Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution. *Journal of Development Economics* 73, 185-213.
- Newman, C.M., 1980, Normal fluctuations and the FKG inequalities. *Communications in Mathematical Physics* 74, 119-128.
- Nishiyama, Y., and P. M. Robinson, 2000, Edgeworth expansions for semiparametric averaged derivatives. *Econometrica* 68, 931-979.
- Parzen, E., 1963, On spectral analysis with missing observations and amplitude modulation. *Sankhya* 25, 180-189.
- Robinson, P. M., 1988, Root-n-consistent semiparametric regression. *Econometrica* 56, 931-954.
- Robinson, P. M., 2007a, Nonparametric spectrum estimation for spatial data. *Journal of Statistical Planning and Inference* 137, 1024-1034.

Robinson, P. M., 2007b, Nonparametric regression with spatial data. *Journal of Econometrics*, forthcoming.

Robinson, P. M., 2008, Correlation testing in time series, spatial and cross-sectional data, *Journal of Econometrics* 147, 5-16.

Robinson, P. M., and F. J. Hidalgo, 1997, Time series regression with long-range dependence. *Annals of Statistics* 25, 77-104.

Scott, D. J., 1973, Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach. *Advances in Applied Probability* 5, 119-137.

Table 1
 Linear regression (1): Empirical sizes of tests with size α

n	m	$\rho_X = 0.2, \rho_U = 0.3$			m	$\rho_X = 0.4, \rho_U = 0.5$		
		$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$		$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
100	C	.021	.058	.125	C	.037	.119	.185
	H	.027	.063	.138	H	.049	.123	.196
	2	.026	.058	.125	6	.029	.088	.154
	4	.024	.052	.117	8	.029	.085	.152
	6	.022	.050	.115	10	.029	.084	.152
	8	.023	.052	.119	12	.027	.082	.153
	10	.024	.056	.122	14	.027	.085	.151
169	C	.013	.052	.106	C	.025	.084	.159
	H	.017	.056	.114	H	.028	.095	.163
	3	.016	.054	.109	6	.023	.069	.130
	6	.013	.050	.104	9	.019	.067	.121
	9	.013	.049	.115	12	.019	.066	.120
	12	.014	.051	.118	15	.020	.067	.125
	15	.016	.061	.120	18	.020	.070	.131

Table 2
 Linear regression (1): Empirical powers of tests with $\beta = 0.8$ and size α

n	m	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	n	m	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
100	C	.605	.827	.902	169	C	.869	.962	.980
	H	.620	.838	.902		H	.877	.966	.983
	2	.618	.838	.901		3	.879	.964	.983
	4	.628	.838	.897		6	.876	.964	.981
	6	.637	.834	.900		9	.881	.963	.982
	8	.641	.834	.897		12	.881	.969	.983
	10	.641	.837	.900		15	.889	.970	.982
12	.655	.841	.904	18	.893	.971	.982		

Table 3
Partly linear regression (5): Empirical sizes of tests with size α using k_2

		$\rho_X = 0.2, \rho_U = 0.3$								
n	m/a	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.10$		
		1.0	1.2	1.4	1.0	1.2	1.4	1.0	1.2	1.4
100	C	.012	.011	.009	.057	.048	.047	.111	.094	.087
	H	.013	.015	.015	.056	.053	.064	.109	.109	.107
	2	.015	.014	.014	.055	.050	.061	.109	.106	.106
	4	.013	.016	.015	.053	.051	.060	.107	.106	.105
	6	.013	.016	.015	.053	.052	.060	.107	.110	.102
	8	.014	.016	.016	.058	.054	.063	.109	.110	.108
	12	.014	.015	.018	.061	.059	.067	.114	.113	.119
169	C	.008	.004	.003	.052	.041	.030	.106	.090	.081
	H	.009	.006	.005	.045	.040	.039	.096	.088	.087
	3	.009	.006	.005	.045	.043	.040	.094	.088	.085
	6	.010	.007	.008	.051	.043	.044	.091	.083	.083
	9	.012	.011	.009	.051	.046	.044	.087	.083	.084
	12	.014	.012	.011	.053	.050	.047	.093	.087	.089
	15	.013	.012	.011	.057	.051	.048	.103	.095	.090
		$\rho_X = 0.4, \rho_U = 0.5$								
n	m/a	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.10$		
		1.0	1.2	1.4	1.0	1.2	1.4	1.0	1.2	1.4
100	C	.021	.018	.016	.069	.064	.063	.127	.123	.117
	H	.017	.019	.027	.076	.071	.073	.133	.143	.135
	6	.014	.014	.024	.066	.065	.070	.116	.125	.119
	8	.013	.014	.022	.069	.068	.072	.117	.124	.120
	10	.014	.017	.024	.070	.067	.077	.121	.126	.127
	12	.016	.020	.026	.076	.077	.081	.124	.128	.128
	14	.018	.025	.029	.077	.084	.088	.127	.133	.134
169	C	.011	.006	.004	.065	.049	.050	.124	.098	.085
	H	.010	.009	.010	.056	.054	.059	.104	.100	.102
	6	.010	.007	.009	.053	.053	.053	.099	.097	.099
	9	.010	.009	.010	.056	.052	.055	.098	.095	.091
	12	.010	.013	.011	.056	.052	.054	.095	.090	.091
	15	.015	.016	.012	.056	.055	.053	.102	.098	.095
	18	.016	.018	.014	.061	.055	.055	.109	.109	.104

Table 4
Partly linear regression (5): Empirical sizes of tests with size α using k_4

$\rho_X = 0.2, \rho_U = 0.3$										
n	m/a	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.10$		
		1.4	1.6	1.8	1.4	1.6	1.8	1.4	1.6	1.8
100	C	.015	.011	.010	.065	.055	.051	.126	.109	.101
	H	.015	.016	.016	.066	.057	.065	.116	.117	.113
	2	.014	.015	.015	.063	.054	.058	.114	.113	.112
	4	.013	.017	.016	.063	.051	.058	.112	.114	.112
	6	.013	.016	.016	.066	.051	.055	.113	.116	.118
	8	.013	.016	.018	.067	.058	.060	.120	.115	.117
	12	.016	.018	.019	.070	.070	.065	.125	.120	.118
169	C	.015	.009	.005	.060	.050	.037	.119	.098	.090
	H	.011	.010	.007	.051	.044	.042	.097	.097	.099
	3	.010	.011	.010	.050	.045	.045	.099	.091	.093
	6	.014	.013	.010	.055	.047	.049	.096	.094	.090
	9	.014	.014	.012	.055	.048	.054	.102	.091	.088
	12	.015	.015	.012	.056	.055	.057	.106	.094	.092
	15	.017	.015	.012	.057	.056	.057	.109	.010	.102
$\rho_X = 0.4, \rho_U = 0.5$										
n	m/a	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.10$		
		1.4	1.6	1.8	1.4	1.6	1.8	1.4	1.6	1.8
100	C	.024	.018	.020	.084	.068	.067	.154	.138	.134
	H	.017	.020	.020	.074	.078	.080	.133	.145	.149
	6	.017	.014	.020	.068	.073	.072	.122	.133	.131
	8	.015	.013	.020	.071	.074	.076	.124	.132	.132
	10	.016	.014	.023	.079	.076	.081	.131	.133	.135
	12	.020	.018	.024	.081	.078	.082	.141	.135	.139
	14	.022	.022	.027	.081	.085	.091	.146	.140	.143
169	C	.016	.010	.008	.074	.058	.054	.133	.114	.099
	H	.012	.014	.012	.057	.056	.062	.111	.107	.111
	6	.011	.012	.010	.055	.052	.056	.109	.103	.104
	9	.011	.012	.011	.058	.054	.056	.107	.098	.102
	12	.011	.013	.015	.064	.057	.059	.104	.099	.103
	15	.015	.017	.017	.062	.062	.060	.109	.105	.110
	18	.016	.018	.018	.063	.065	.063	.117	.119	.113

Table 5
Partly linear regression (5): Empirical powers of tests with $\beta = 0.7$ using k_2, k_4 at level α .

		k_2								
n	m/a	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.10$		
		1.0	1.2	1.4	1.0	1.2	1.4	1.0	1.2	1.4
100	C	.536	.521	.485	.760	.744	.728	.830	.826	.826
	H	.519	.527	.535	.743	.744	.744	.817	.831	.842
	2	.515	.531	.534	.739	.750	.750	.817	.834	.846
	4	.511	.534	.537	.741	.752	.751	.818	.831	.844
	6	.511	.543	.541	.743	.757	.761	.819	.831	.845
	8	.521	.543	.556	.745	.757	.762	.823	.829	.841
	12	.530	.547	.559	.744	.754	.767	.827	.835	.844
169	C	.810	.794	.788	.929	.929	.918	.962	.960	.958
	H	.775	.795	.801	.917	.923	.928	.950	.957	.964
	3	.778	.796	.804	.914	.925	.927	.951	.956	.961
	6	.778	.798	.804	.916	.925	.926	.947	.956	.963
	9	.777	.807	.810	.910	.920	.927	.949	.958	.958
	12	.782	.808	.816	.913	.922	.929	.949	.958	.959
	15	.790	.815	.823	.914	.922	.927	.946	.959	.958
		k_4								
n	m/a	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.10$		
		1.4	1.6	1.8	1.4	1.6	1.8	1.4	1.6	1.8
100	C	.546	.523	.499	.753	.737	.730	.825	.814	.812
	H	.508	.517	.523	.723	.737	.738	.797	.813	.820
	2	.503	.519	.524	.723	.736	.743	.797	.812	.821
	4	.508	.519	.523	.721	.735	.740	.796	.810	.821
	6	.501	.518	.535	.724	.735	.743	.799	.811	.821
	8	.508	.529	.538	.723	.742	.745	.802	.812	.823
	12	.518	.536	.548	.725	.743	.747	.804	.817	.827
169	C	.805	.791	.787	.924	.925	.918	.955	.956	.955
	H	.759	.774	.791	.903	.916	.918	.943	.952	.955
	3	.759	.773	.793	.903	.919	.922	.945	.952	.955
	6	.766	.774	.796	.902	.916	.923	.944	.952	.953
	9	.760	.779	.798	.897	.914	.922	.946	.949	.956
	12	.764	.784	.805	.900	.913	.921	.945	.950	.956
	15	.767	.792	.804	.902	.914	.924	.943	.951	.956

Table 6: $Y =$ Proportion of irrigated land (IR)

	LS	IV	Partly LS				Partly IV			
			k_2		k_4		k_2		k_4	
			a=2	a=2.5	a=1.4	a=1.9	a=1.7	a=2.2	a=1.4	a=1.9
NL	.72 (.21)	.37 (.63)	.95 (.18)	.96 (.18)	.85 (.16)	.90 (.17)	.80 (.41)	.81 (.42)	.76 (.40)	.78 (.41)
m=2	(.23)	(.58)	(.22)	(.22)	(.22)	(.22)	(.43)	(.44)	(.42)	(.42)
m=4	(.25)	(.69)	(.24)	(.24)	(.24)	(.24)	(.52)	(.53)	(.51)	(.52)
m=6	(.27)	(.73)	(.25)	(.25)	(.25)	(.25)	(.54)	(.55)	(.53)	(.54)
NL2	-.71 (.19)	-.29 (.61)	-.91 (.17)	-.92 (.17)	-.83 (.16)	-.87 (.16)	-.70 (.41)	-.69 (.42)	-.71 (.40)	-.70 (.41)
m=2	(.20)	(.60)	(.20)	(.20)	(.19)	(.19)	(.45)	(.46)	(.43)	(.44)
m=4	(.23)	(.74)	(.21)	(.21)	(.21)	(.21)	(.56)	(.58)	(.53)	(.55)
m=6	(.24)	(.79)	(.22)	(.22)	(.21)	(.21)	(.58)	(.60)	(.54)	(.57)
DBC $\times 10^{-3}$	-1.62 (.66)	-1.90 (.77)	-1.53 (.63)	-1.40 (.64)	-2.02 (.57)	-1.77 (.60)	-1.87 (.70)	-1.80 (.73)	-2.07 (.62)	-1.94 (.67)
m=2	(.74)	(.77)	(.72)	(.73)	(.70)	(.70)	(.68)	(.70)	(.68)	(.68)
m=4	(.85)	(.87)	(.82)	(.84)	(.80)	(.80)	(.77)	(.78)	(.77)	(.77)
m=6	(.83)	(.82)	(.80)	(.83)	(.75)	(.75)	(.72)	(.74)	(.70)	(.71)
CD $\times 10^{-1}$.41 (.57)	.38 (.72)	-.60 (.51)	-.65 (.50)	-.39 (.54)	-.49 (.52)	-.69 (.55)	-.80 (.55)	-.43 (.58)	-.60 (.56)
m=2	(.56)	(.63)	(.54)	(.54)	(.51)	(.52)	(.56)	(.58)	(.53)	(.55)
m=4	(.52)	(.58)	(.59)	(.60)	(.55)	(.57)	(.64)	(.67)	(.58)	(.62)
m=6	(.48)	(.53)	(.63)	(.64)	(.57)	(.60)	(.69)	(.72)	(.59)	(.66)
BSD	-.16 (.05)	-.21 (.07)	-.13 (.04)	-.13 (.05)	-.11 (.04)	-.12 (.04)	-.15 (.06)	-.17 (.06)	-.12 (.05)	-.14 (.05)
m=2	(.04)	(.06)	(.04)	(.04)	(.03)	(.04)	(.06)	(.06)	(.05)	(.05)
m=4	(.05)	(.09)	(.04)	(.04)	(.04)	(.04)	(.07)	(.08)	(.06)	(.07)
m=6	(.06)	(.10)	(.05)	(.05)	(.04)	(.04)	(.08)	(.09)	(.06)	(.07)
RSD $\times 10^{-1}$.14 (.48)	.25 (.57)	-.44 (.44)	-.49 (.44)	-.23 (.42)	-.33 (.43)	-.37 (.46)	-.43 (.46)	-.20 (.46)	-.31 (.45)
m=2	(.51)	(.60)	(.51)	(.52)	(.47)	(.49)	(.53)	(.54)	(.53)	(.52)
m=4	(.40)	(.55)	(.56)	(.47)	(.42)	(.44)	(.50)	(.50)	(.51)	(.49)
m=6	(.34)	(.51)	(.43)	(.44)	(.40)	(.41)	(.47)	(.48)	(.50)	(.47)
ASD $\times 10^{-1}$.62 (.35)	.54 (.38)	.81 (.35)	.79 (.35)	.84 (.33)	.83 (.34)	.75 (.36)	.71 (.37)	.82 (.34)	.78 (.35)
m=2	(.35)	(.36)	(.38)	(.38)	(.38)	(.38)	(.37)	(.38)	(.37)	(.37)
m=4	(.37)	(.37)	(.41)	(.42)	(.39)	(.40)	(.39)	(.40)	(.38)	(.39)
m=6	(.34)	(.34)	(.38)	(.39)	(.36)	(.36)	(.36)	(.37)	(.35)	(.35)

Slope estimates are in bold; SEs are in parentheses; with non-robust ones in the top row, and robust ones below computed using truncation vectors (m, m) where $m = 2, 4$ and 6 respectively; columns under Partial LS and Partial IV refer to choices of bandwidth a and kernel (k_2, k_4) .

Table 7: $Y = \text{Fertilizer use (FU)}$

	LS	IV	Partly LS				Partly IV			
			k_2		k_4		k_2		k_4	
			a=1.8	a=2.3	a=1.6	a=2.2	a=1.4	a=1.9	a=1.5	a=2.0
NL	115.90	30.51	115.58	119.07	111.94	114.15	104.47	88.28	117.28	100.84
	(42.28)	(135.11)	(31.37)	(32.69)	(28.48)	(30.80)	(75.57)	(79.25)	(71.92)	(76.32)
m=2	(36.40)	(118.17)	(32.40)	(33.28)	(32.63)	(32.20)	(74.91)	(74.19)	(78.04)	(74.43)
m=4	(40.00)	(121.92)	(36.84)	(37.94)	(37.17)	(36.55)	(82.85)	(80.17)	(87.87)	(81.98)
m=6	(42.56)	(123.52)	(38.23)	(39.77)	(37.91)	(37.73)	(84.12)	(80.50)	(90.01)	(83.10)
NL2	-82.02	32.98	-87.23	-88.74	-86.64	-86.52	-65.05	-42.94	-81.10	-60.93
	(39.57)	(130.50)	(29.66)	(30.95)	(26.88)	(29.09)	(73.40)	(77.75)	(69.32)	(74.24)
m=2	(34.80)	(118.89)	(30.59)	(31.52)	(30.58)	(30.37)	(78.54)	(78.56)	(80.64)	(78.22)
m=4	(36.28)	(127.68)	(33.09)	(34.27)	(33.08)	(32.77)	(87.99)	(86.88)	(91.52)	(87.41)
m=6	(36.53)	(132.60)	(32.42)	(34.00)	(32.08)	(31.94)	(89.33)	(87.87)	(93.42)	(88.65)
DBC	-.31	-.43	-.24	-.25	-.21	-.24	-.26	-.30	-.23	-.26
	(.14)	(.16)	(.11)	(.11)	(.10)	(.11)	(.11)	(.12)	(.10)	(.11)
m=2	(.14)	(.19)	(.10)	(.11)	(.10)	(.10)	(.11)	(.12)	(.11)	(.11)
m=4	(.17)	(.24)	(.12)	(.13)	(.11)	(.12)	(.13)	(.15)	(.11)	(.13)
m=6	(.18)	(.25)	(.12)	(.14)	(.10)	(.12)	(.12)	(.15)	(.10)	(.13)
CD	2.84	-1.59	-3.55	-3.89	-6.59	-3.61	-8.55	-7.59	-11.61	-7.87
	(11.59)	(15.51)	(10.43)	(10.11)	(11.43)	(10.59)	(11.74)	(11.32)	(12.35)	(11.63)
m=2	(16.10)	(17.12)	(12.13)	(12.11)	(12.20)	(12.18)	(12.13)	(12.08)	(12.58)	(12.17)
m=4	(17.53)	(17.80)	(12.93)	(13.35)	(11.14)	(12.70)	(11.15)	(12.32)	(10.02)	(11.42)
m=6	(18.61)	(18.68)	(14.10)	(14.72)	(11.43)	(13.74)	(11.27)	(13.19)	(9.10)	(11.65)
BSD	-9.26	-22.85	1.24	-1.09	5.64	2.22	-1.07	-6.87	3.32	-1.90
	(9.78)	(14.30)	(7.80)	(8.16)	(7.17)	(7.65)	(9.49)	(10.25)	(8.89)	(9.61)
m=2	(9.10)	(15.29)	(8.10)	(8.15)	(8.24)	(8.13)	(12.25)	(12.70)	(12.04)	(12.34)
m=4	(9.54)	(18.45)	(8.22)	(8.26)	(8.51)	(8.26)	(13.73)	(14.65)	(13.19)	(13.88)
m=6	(9.92)	(20.28)	(8.25)	(8.29)	(8.57)	(8.30)	(14.01)	(15.38)	(13.14)	(14.23)
RSD	3.19	5.13	6.23	4.02	11.08	7.24	8.60	6.18	10.91	8.20
	(9.76)	(12.18)	(8.07)	(8.24)	(7.86)	(8.00)	(8.95)	(8.92)	(9.09)	(8.91)
m=2	(12.00)	(12.78)	(10.07)	(9.87)	(11.10)	(10.22)	(11.85)	(10.89)	(13.26)	(11.67)
m=4	(12.68)	(13.76)	(11.53)	(11.24)	(12.67)	(11.72)	(13.85)	(12.72)	(15.39)	(13.66)
m=6	(13.55)	(15.04)	(12.49)	(12.15)	(13.81)	(12.70)	(15.27)	(14.06)	(16.97)	(15.07)
ASD	18.72	15.60	23.18	22.55	25.11	23.55	23.15	21.55	24.73	22.98
	(7.23)	(8.03)	(6.29)	(6.50)	(5.97)	(6.21)	(6.23)	(6.60)	(6.02)	(6.29)
m=2	(8.31)	(9.26)	(6.53)	(7.02)	(5.76)	(6.34)	(6.15)	(6.90)	(5.77)	(6.26)
m=4	(8.55)	(10.30)	(6.51)	(7.14)	(5.50)	(6.25)	(6.16)	(7.26)	(5.46)	(6.32)
m=6	(8.77)	(10.77)	(6.46)	(7.15)	(5.45)	(6.20)	(6.19)	(7.43)	(5.37)	(6.37)

Slope estimates are in bold; SEs are in parentheses; with non-robust ones in the top row, and robust ones below computed using truncation vectors (m, m) where $m = 2, 4$ and 6 respectively; columns under Partial LS and Partial IV refer to choices of bandwidth a and kernel (k_2, k_4) .

Table 8: $Y = \text{Log}(\text{yield 15 major crops})$ (L15)

	LS	IV	Partly LS				Partly IV			
			k_2		k_4		k_2		k_4	
			a=1.4	a=1.9	a=1.4	a=1.7	a=1.4	a=1.9	a=1.5	a=2.1
NL	1.71 (.35)	2.07 (1.03)	1.65 (.30)	1.54 (.31)	1.74 (.29)	1.69 (.30)	2.29 (.76)	1.81 (.76)	2.61 (.76)	2.15 (.77)
m=2	(.40)	(1.02)	(.38)	(.37)	(.40)	(.39)	(.88)	(.82)	(.94)	(.86)
m=4	(.44)	(1.16)	(.41)	(.40)	(.43)	(.42)	(.99)	(.95)	(1.02)	(.98)
m=6	(.34)	(1.19)	(.40)	(.39)	(.40)	(.40)	(1.03)	(1.00)	(1.05)	(1.03)
NL2	-1.41 (.33)	-1.66 (1.00)	-1.38 (.29)	-1.27 (.29)	-1.47 (.27)	-1.42 (.28)	-1.96 (.74)	-1.47 (.75)	-2.28 (.73)	-1.82 (.75)
m=2	(.37)	(1.02)	(.34)	(.33)	(.35)	(.34)	(.90)	(.85)	(.94)	(.89)
m=4	(.38)	(1.17)	(.35)	(.35)	(.36)	(.36)	(1.02)	(1.00)	(1.05)	(1.02)
m=6	(.37)	(1.20)	(.33)	(.32)	(.33)	(.33)	(1.07)	(1.04)	(1.09)	(1.07)
DBC $\times 10^{-3}$	-2.65 (1.12)	-3.06 (1.26)	-2.61 (1.04)	-2.73 (1.08)	-2.42 (1.00)	-2.55 (1.03)	-2.95 (1.12)	-3.04 (1.17)	-2.84 (1.09)	-3.01 (1.14)
m=2	(1.11)	(1.15)	(1.10)	(1.11)	(1.10)	(1.10)	(1.10)	(1.12)	(1.11)	(1.11)
m=4	(1.30)	(1.33)	(1.32)	(1.35)	(1.29)	(1.32)	(1.28)	(1.14)	(1.25)	(1.31)
m=6	(1.33)	(1.33)	(1.38)	(1.42)	(1.30)	(1.36)	(1.28)	(1.14)	(1.20)	(1.33)
CD $\times 10^{-1}$.51 (.95)	.06 (1.17)	-1.07 (1.10)	-7.73 (1.01)	-1.61 (1.25)	-1.27 (1.15)	-1.39 (1.18)	-1.03 (1.08)	-1.80 (1.30)	-1.29 (1.14)
m=2	(.92)	(1.22)	(1.16)	(1.05)	(1.30)	(1.21)	(1.34)	(1.21)	(1.47)	(1.30)
m=4	(.95)	(1.29)	(1.04)	(1.01)	(1.03)	(1.04)	(1.25)	(1.19)	(1.27)	(1.24)
m=6	(.95)	(1.26)	(1.01)	(1.01)	(.92)	(.99)	(1.16)	(1.14)	(1.13)	(1.16)
BSD $\times 10^{-1}$	-1.60 (.80)	-1.58 (1.09)	-1.29 (.75)	-1.58 (.77)	-1.01 (.74)	-1.17 (.74)	-.97 (.96)	-1.59 (.99)	-.54 (.94)	-1.16 (.97)
m=2	(.82)	(1.09)	(.85)	(.86)	(.84)	(.85)	(1.07)	(1.08)	(1.08)	(1.07)
m=4	(.84)	(1.13)	(.87)	(.88)	(.86)	(.87)	(1.08)	(1.12)	(1.05)	(1.09)
m=6	(.84)	(1.14)	(.84)	(.85)	(.83)	(.84)	(1.01)	(1.09)	(.96)	(1.04)
RSD $\times 10^{-1}$.22 (.80)	.01 (.93)	.21 (.80)	.23 (.79)	.25 (.82)	.21 (.81)	-.16 (.90)	.09 (.85)	-.36 (.95)	-.08 (.88)
m=2	(.74)	(.85)	(.86)	(.81)	(.96)	(.89)	(.97)	(.88)	(1.10)	(.93)
m=4	(.70)	(.87)	(.88)	(.82)	(1.00)	(.91)	(1.01)	(.93)	(1.13)	(.97)
m=6	(.68)	(.89)	(.91)	(.84)	(1.06)	(.95)	(1.06)	(.98)	(1.19)	(1.02)
ASD $\times 10^{-1}$	2.56 (.60)	2.44 (.62)	2.45 (.62)	2.65 (.62)	2.22 (.62)	2.37 (.62)	2.39 (.63)	2.58 (.63)	2.24 (.64)	2.45 (.63)
m=2	(.58)	(.58)	(.68)	(.67)	(.70)	(.69)	(.68)	(.66)	(.70)	(.67)
m=4	(.55)	(.56)	(.66)	(.68)	(.62)	(.65)	(.64)	(.66)	(.62)	(.65)
m=6	(.50)	(.51)	(.64)	(.66)	(.59)	(.62)	(.62)	(.65)	(.57)	(.63)

Slope estimates are in bold; SEs are in parentheses; with non-robust ones in the top row, and robust ones below computed using truncation vectors (m, m) where $m = 2, 4$ and 6 respectively; columns under Partial LS and Partial IV refer to choices of bandwidth a and kernel (k_2, k_4) .

Table 9: $Y = \text{Log}(\text{rice yield})$ (LR)

	LS	IV	Partly LS				Partly IV			
			k_2		k_4		k_2		k_4	
			a=1.5	a=2.0	a=1.3	a=1.5	a=0.9	a=1.3	a=1.3	a=1.6
NL	.99 (.43)	.35 (1.28)	1.25 (.38)	1.15 (.39)	1.43 (.37)	1.38 (.37)	1.12 (.96)	.93 (.96)	1.14 (.95)	1.03 (.96)
m=2	(.48)	(1.16)	(.54)	(.51)	(.59)	(.57)	(1.04)	(1.02)	(1.05)	(1.03)
m=4	(.56)	(1.37)	(.63)	(.61)	(.68)	(.66)	(1.19)	(1.19)	(1.20)	(1.21)
m=6	(.57)	(1.38)	(.63)	(.62)	(.67)	(.65)	(1.18)	(1.20)	(1.20)	(1.21)
NL2	-.53 (.40)	.08 (1.24)	-.85 (.36)	-.71 (.37)	-1.07 (.35)	-1.01 (.35)	-.94 (.92)	-.69 (.93)	-.96 (.91)	-.80 (.93)
m=2	(.42)	(1.22)	(.46)	(.44)	(.50)	(.41)	(1.09)	(1.09)	(1.09)	(1.10)
m=4	(.46)	(1.44)	(.51)	(.49)	(.55)	(.53)	(1.27)	(1.29)	(1.28)	(1.30)
m=6	(.45)	(1.44)	(.49)	(.48)	(.53)	(.51)	(1.27)	(1.29)	(1.29)	(1.31)
DBC $\times 10^{-3}$	-4.23 (1.39)	-4.08 (1.56)	-3.65 (1.32)	-3.81 (1.34)	-3.21 (1.27)	-3.38 (1.29)	-2.67 (1.36)	-3.01 (1.40)	-2.76 (1.36)	-2.95 (1.39)
m=2	(1.61)	(1.69)	(1.61)	(1.60)	(1.62)	(1.62)	(1.52)	(1.47)	(1.51)	(1.48)
m=4	(2.02)	(2.18)	(2.03)	(2.02)	(2.01)	(2.02)	(1.76)	(1.82)	(1.76)	(1.80)
m=6	(2.21)	(2.42)	(2.20)	(2.20)	(2.10)	(2.16)	(1.75)	(1.92)	(1.76)	(1.88)
CD $\times 10^{-1}$	1.33 (1.18)	1.73 (1.46)	-.05 (1.36)	.03 (1.24)	.05 (1.66)	.01 (1.55)	.82 (1.78)	.60 (1.53)	.70 (1.75)	.57 (1.53)
m=2	(1.12)	(1.38)	(1.48)	(1.29)	(1.99)	(1.80)	(2.23)	(1.81)	(2.25)	(1.95)
m=4	(1.26)	(1.52)	(1.55)	(1.42)	(1.82)	(1.72)	(2.09)	(1.80)	(2.11)	(1.90)
m=6	(1.34)	(1.58)	(1.64)	(1.56)	(1.69)	(1.68)	(1.95)	(1.81)	(1.96)	(1.86)
BSD $\times 10^{-1}$	-.39 (.99)	-.80 (1.35)	.73 (.95)	.43 (.96)	1.12 (.95)	1.01 (.95)	1.46 (1.17)	1.05 (1.20)	1.41 (1.17)	1.19 (1.19)
m=2	(1.08)	(1.32)	(1.17)	(1.17)	(1.17)	(1.17)	(1.32)	(1.34)	(1.32)	(1.33)
m=4	(1.19)	(1.51)	(1.29)	(1.29)	(1.28)	(1.28)	(1.47)	(1.51)	(1.47)	(1.50)
m=6	(1.21)	(1.54)	(1.25)	(1.33)	(1.30)	(1.31)	(1.43)	(1.47)	(1.43)	(1.54)
RSD $\times 10^{-1}$	2.30 (.99)	2.60 (1.15)	2.25 (1.01)	2.07 (.98)	2.61 (1.07)	2.49 (1.05)	2.90 (1.26)	2.64 (1.15)	2.89 (1.25)	2.71 (1.19)
m=2	(.98)	(1.11)	(1.10)	(1.00)	(1.33)	(1.25)	(1.56)	(1.30)	(1.54)	(1.38)
m=4	(.98)	(1.20)	(1.19)	(1.08)	(1.48)	(1.37)	(1.76)	(1.45)	(1.74)	(1.54)
m=6	(1.00)	(1.25)	(1.25)	(1.12)	(1.58)	(1.45)	(1.90)	(1.54)	(1.88)	(1.64)
ASD $\times 10^{-1}$	2.48 (.74)	2.52 (.77)	2.34 (.78)	2.57 (.77)	1.86 (.80)	2.03 (.79)	1.93 (.81)	2.33 (.80)	1.93 (.81)	2.20 (.80)
m=2	(.81)	(.83)	(1.03)	(.99)	(1.09)	(1.07)	(1.08)	(1.02)	(1.08)	(1.04)
m=4	(.83)	(.89)	(1.02)	(1.01)	(1.04)	(1.04)	(1.03)	(1.03)	(1.03)	(1.03)
m=6	(.75)	(.83)	(.90)	(.89)	(.90)	(.90)	(.90)	(.92)	(.90)	(.92)

Slope estimates are in bold; SEs are in parentheses; with non-robust ones in the top row, and robust ones below computed using truncation vectors (m, m) where $m = 2, 4$ and 6 respectively; columns under Partial LS and Partial IV refer to choices of bandwidth a and kernel (k_2, k_4) .

Nonparametric regressions

Figure 1

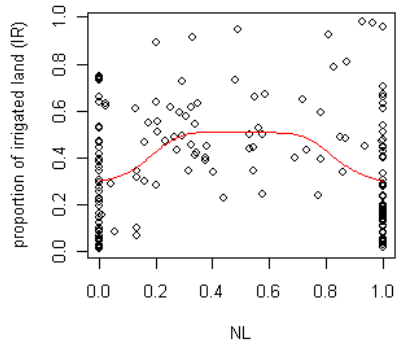


Figure 2

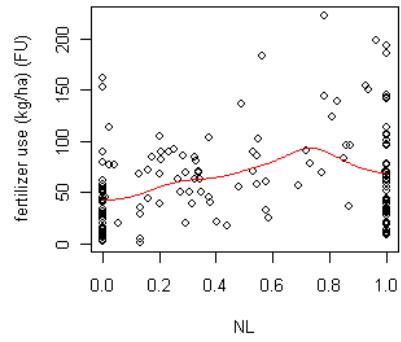


Figure 3

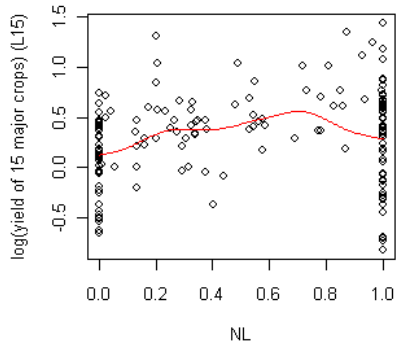


Figure 4

