Nonparametric Trending Regression with Cross-Sectional Dependence

Peter M. Robinson*

London School of Economics

January 25, 2010

Abstract

Panel data, whose series length T is large but whose cross-section size N need not be, are assumed to have a common time trend. The time trend is of unknown form, the model includes additive, unknown, individual-specific components, and we allow for spatial or other cross-sectional dependence and/or heteroscedasticity. A simple smoothed nonparametric trend estimate is shown to be dominated by an estimate which exploits the availability of cross-sectional data. Asymptotically optimal choices of bandwidth are justified for both estimates. Feasible optimal bandwidths, and feasible optimal trend estimates, are asymptotically justified, the finite sample performance of the latter being examined in a Monte Carlo study. A number of potential extensions are discussed.

JEL Classifications: C13; C14; C23

Keywords: Panel data; nonparametric time trend; cross-sectional dependence; generalized least squares; optimal bandwidth

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

1. INTRODUCTION

Much econometric modelling of nonstationary time series employs deterministic trending functions that are polynomial, indeed frequently linear. However the penalties of mis-specifying parametric functions are well appreciated, and nonparametric modelling is increasingly widely accepted, at least in samples of reasonable size. The inability of polynomials to satisfactorily globally approximate general functions of time deters study of polynomial functions whose order increases slowly with sample size, and rather leads one to consider the possibility of a smooth trend mapped into the unit interval and approximated by a smoothed kernel regression. For example, Starica and Granger (2005) employed this approach in modelling series of stock prices. There is a huge literature on such fixed-design nonparametric regression, principally in the setting of a single time series.

Here we are concerned with panel data, where N series of length T have a common, nonparametric, time trend but also additive, fixed, individual effects, for which we have to correct before being able to form a trend estimate. We assume an asymptotic framework in which T is large, but not necessarily N, so that the cross-sectional mean at a given time point is not necessarily consistent for the trend, hence the recourse to smoothed nonparametric regression. A major feature of the paper is concern for possible cross-sectional correlation and/or heteroscedasticity. These influence the asymptotic variance of our trend estimate, and thence also the mean squared error and consequent optimal rules for bandwidth choice. The availability of cross-sectional data enables us to propose a trend estimate, based on the generalized least squares principle, that reduces the asymptotic variance. This estimate, along with its asymptotic variance (and that of the original trend estimate), depends on the cross-sectional covariance matrix. In general this is not wholly known, and possibly not known at all. Using residuals from the fitted trend, we consistently estimate its elements, so as to obtain a feasible improved trend estimate, and a consistent estimate of its variance, as well as feasible optimal bandwidths that are asymptotically equivalent to the infeasible versions. These results are valid with N remaining fixed as T increases, and they continue to hold if N is also allowed to increase, in which case there is a faster rate of convergence, and in this latter situation our results hold irrespective of whether or not the covariance matrix is finitely parameterized.

Section 2 describes the basic model. In Section 3 we present a simple trend estimate and its mean squared error properties. Improved estimation is discussed in Section 4. In Section 5 optimal bandwidths are reported. Section 6 suggests estimates of the cross-sectional covariance matrix, with asymptotic justifiication for their insertion in the optimal bandwidths and improved trend estimates. Section 7 suggests some directions for further research. Proof details may be found in two appendices.

2. PANEL DATA NONPARAMETRIC MODEL

We observe y_{it}, $i = 1, ..., N$, $t = 1, ..., T$, generated by

$$y_{it} = \alpha_i + \beta_t + x_{it}, \tag{2.1}$$

where the α_i and β_t are unknown constants, and the x_{it} are unobservable zero-mean random variables, uncorrelated and homoscedastic across time, but possibly correlated and heteroscedastic over the cross section. Thus we impose

Assumption 1 *For all* i, t,

$$E(x_{it}) = 0; \tag{2.2}$$

for all i, j, t *there exist finite constants* ω_{ij} *such that*

$$E\left(x_{it}x_{jt}\right) = \omega_{ij}; \tag{2.3}$$

and for all i, j, t, u,

$$E\left(x_{it}x_{ju}\right) = 0, \quad t \neq u. \tag{2.4}$$

Our focus is on estimating the time trend. Superficially this is represented in (2.1) by β_t, but the α_i and β_t are identified only up to location shift. To resolve this problem we initially impose the (arbitrary) restriction

$$\sum_{i=1}^{N} \alpha_i = 0. \tag{2.5}$$

An immediate consequence of (2.5) is the relationship

$$\bar{y}_{At} = \beta_t + \bar{x}_{At}, \tag{2.6}$$

defining the cross-sectional means

$$\bar{x}_{At} = N^{-1} \sum_{i=1}^{N} x_{it}, \quad \bar{y}_{At} = N^{-1} \sum_{i=1}^{N} y_{it}. \tag{2.7}$$

We use the A subscript to denote averaging, in (2.7) with respect to i and subsequently also with respect to t. In view of (2.2) an obvious estimate of β_t is thus

$$\beta_t^* = \bar{y}_{At}. \tag{2.8}$$

This can be a good estimate if N is large. For any fixed t, it is trivially seen that β_t^* is mean-square consistent for β_t if

$$\lim_{N \to \infty} \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \omega_{ij} = 0. \tag{2.9}$$

Condition (2.9) is trivially satisfied if the x_{it} are uncorrelated over i, but more generally if cross-sectional dependence is limited by the condition

$$\overline{\lim_{N \to \infty}} \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \omega_{ij} < \infty. \tag{2.10}$$

We mention (2.10) because it is analogous to a common weak dependence condition for time series, indeed it would correspond to an extension of the latter condition to stationary spatial lattice processes. Also (2.10) is slightly weaker than Condition C.3

of Bai and Ng (2002), in a different panel data context. The more general condition (2.9) permits a type of cross-sectional long range dependence.

While it can also cover certain models including additive factors, (2.9) does not, however, hold for factor models in which the x_{it}, for all i, are influenced by the same factor or factors. For this reason, and because we do not wish to model the β_t in terms of finitely many parameters (as would be the case for a polynomial trend, say), we use nonparametric smoothing across time which requires T to be large, but not necessarily N. In order to achieve consistent estimates we assume the existence of a function $\beta(\tau)$, $0 \leq \tau \leq 1$, that is suitably smooth, such that

$$\beta_t = \beta(t/T), \quad t = 1, ..., T. \tag{2.11}$$

The T-dependent argument in (2.11) enables information to be borrowed so as to permit consistent estimation of $\beta(\tau)$ for any fixed τ as $T \to \infty$. Thus β_t, and hence in turn y_{it}, should be regarded as triangular arrays, i.e. $\beta_t = \beta_{tT}$, $y_{it} = y_{itT}$, but for ease of notation we suppress reference to the T-subscript. Also, though our work is relevant in part to the case of a fixed N (as $T \to \infty$), we also allow $N \to \infty$ (slowly, as $T \to \infty$). In the latter circumstances, the α_i should also be regarded as triangular arrays, $\alpha_i = \alpha_{iN}$, in view of the restriction (2.5), and this would imply dependence of the y_{it} on N also, though again we suppress reference to an N subscript. In terms of practical applications, on the other hand, one envisages data for which T is much larger than N, one example being many frequent time series observations on a relatively modest number of stock prices.

Smoothed nonparametric estimation has been considered previously in a panel data setting. For example, Ruckstuhl, Welsh and Carroll (2000) considered a model in which β_t is replaced by a nonparametric function of a stochastic explanatory variable, which can vary across both i and t, the α_i are stochastic, and independent and homoscedastic across i, t, and N is fixed. Thus there is an explicit factor structure built in, with no other source of cross-sectional dependence (and no heteroscedasticity). We discuss a factor structure for x_{it} but as a special case only, and here as elsewhere we consider the possibility that N, like T, diverges. Also, Hart and Wehrly (1986) considered the special case of (2.1) with no individual-specific effects, i.e. $\alpha_i \equiv 0$, and with no cross-sectional correlation or heteroscedasticity. A deterministic nonparametric trend also features in the panel data partly linear semiparametric regression models considered by Severini and Stanisvalis (1994), Moyeed and Diggle (1994), for example.

3. SIMPLE TREND ESTIMATION

We introduce a kernel function $k(u)$, $-\infty < u < \infty$, satisfying

$$\int_{-\infty}^{\infty} k(u)du = 1, \tag{3.1}$$

and a positive scalar bandwidth $h = h_T$. Then with the abbreviation

$$k_{t\tau} = k\left(\frac{T\tau - t}{Th}\right), \tag{3.2}$$

define the estimate

$$\tilde{\beta}(\tau) = \sum_{t=1}^{T} k_{t\tau}\bar{y}_{At} / \sum_{t=1}^{T} k_{t\tau}, \quad \tau \in (0,1). \tag{3.3}$$

Important measures of goodness of nonparametric estimates, which lead to optimal choices of bandwidth h, are mean squared error, i.e.

$$MSE\left\{\tilde{\beta}(\tau)\right\} = E\left\{\tilde{\beta}(\tau) - \beta(\tau)\right\}^2, \tag{3.4}$$

and mean integrated squared error, i.e.

$$MISE\left\{\tilde{\beta}\right\} = \int_0^1 E\left\{\tilde{\beta}(u) - \beta(u)\right\}^2 du. \tag{3.5}$$

To approximate these we require conditions on β, k and h.

Assumption 2 *$\beta(\tau)$ is twice continuously differentiable, $0 \leq \tau \leq 1$.*

Assumption 3 *$k(u)$ satisfies (3.1) and is even, non-negative, and twice boundedly differentiable at all but possibly finitely many points, $k(u)$ and its derivative $k'(u) = (d/du)k(u)$ satisfying*

$$k(u) = O\left((1+u^4)^{-1}\right), \quad k'(u) = O\left((1+|u|^c)^{-1}\right), \ \textit{some } c > 2. \tag{3.6}$$

Assumption 4 *As $T \to \infty$*

$$h + (Th)^{-1} \to 0. \tag{3.7}$$

The smoothness condition on β in Assumption 2 could be relaxed at cost of inferior optimal rates of convergence, or on the other hand strengthened, which would lead to better rates if Assumption 3 were modified to permit higher order kernels; in any case the non-negativity condition on k is imposed only to simplify proofs. Assumption 4 is as usual a minimal condition for consistent estimation.

Define the $N \times N$ cross-sectional disturbance covariance matrix Ω, having $(i,j)-th$ element ω_{ij} (see Assumption 1), define by ℓ the $N \times 1$ vector of 1's, and also define

$$\kappa = \int_{-\infty}^{\infty} k^2(u)du, \quad \chi = \int_{-\infty}^{\infty} u^2 k(u)du, \tag{3.8}$$

$$\zeta(\tau) = \beta''(\tau)^2, \ \ \xi = \int_0^1 \beta''(u)^2 du, \tag{3.9}$$

where $\beta''(u) = \left(d^2/du^2\right)\beta(u)$.

Assumption 5 $\zeta(\tau) > 0$, $\xi > 0$.

Theorem 1 *Let Assumptions 1-5 hold. Then as* $T \to \infty$

$$MSE\left\{\tilde{\beta}(\tau)\right\} \sim \frac{\kappa}{Th}\frac{\ell'\Omega\ell}{N^2} + \frac{\zeta(\tau)\chi^2 h^4}{4}, \quad (3.10)$$

$$MISE\left\{\tilde{\beta}\right\} \sim \frac{\kappa}{Th}\frac{\ell'\Omega\ell}{N^2} + \frac{\xi\chi^2 h^4}{4}. \quad (3.11)$$

Theorem 1 contains nothing new in itself, for fixed $N, \ell'\Omega\ell/N^2$ being just the variance of the dependent variable $\bar{y}_{At}$ in the nonparametric regression. Versions of this result were given long ago, e.g. by Benedetti (1977), though conditions employed in Theorem 1 are essentially taken from Robinson (1997), and no proof is required. Hart and Wehrly (1986) also considered nonparametric regression of cross-sectional means in a panel data setting, but without allowing for individual-specific effects. However, the availability of cross-sectional data offers the possibility of improved estimation, in terms of variance-reduction, as explored in the following section.

4. IMPROVED TREND ESTIMATION

Improved estimation of the trend requires it to be identified in a different way from that in Section 2, in particular to shift its location. Consider the representation

$$y_{it} = \alpha_i^{(w)} + \beta_t^{(w)} + x_{it}, \quad (4.1)$$

where the bracketed superscript w represents a vector $w = (w_1, ..., w_n)'$ of weights, such that

$$w'\ell = 1, \quad (4.2)$$

$$w'\alpha^{(w)} = 0, \quad (4.3)$$

where ℓ is a $N \times 1$ vector of 1's. This represents a generalisation of (2.1), (2.5), in which $w = (1/N, ..., 1/N)'$. It is convenient to write (4.1), for $i = 1, ..., N$, in

N-dimensional column vector form as

$$y_{\cdot t} = \alpha^{(w)} + \beta_t^{(w)} \ell + x_{\cdot t}, \tag{4.4}$$

where

$$\alpha^{(w)} = \begin{bmatrix} \alpha_1^{(w)} \\ \vdots \\ \alpha_N^{(w)} \end{bmatrix}, \; x_{\cdot t} = \begin{bmatrix} x_{1t} \\ \vdots \\ x_{Nt} \end{bmatrix}, \; y_{\cdot t} = \begin{bmatrix} y_{1t} \\ \vdots \\ y_{Nt} \end{bmatrix}. \tag{4.5}$$

For given w consider a smooth function $\beta^{(w)}(\tau)$ such that $\beta_t^{(w)} = \beta^{(w)}(t/T)$. For any two weight vectors w_a, w_b we have

$$\alpha^{(w_a)} + \beta^{(w_a)}(\tau)\ell = \alpha^{(w_b)} + \beta(\tau)^{(w_b)}\ell. \tag{4.6}$$

Thus

$$\beta^{(w_b)}(\tau) = \beta^{(w_a)}(\tau) + w_b' \alpha^{(w_a)}, \tag{4.7}$$

$$\begin{aligned} \alpha^{(w_b)} &= \alpha^{(w_a)} + \left(\beta^{(w_a)}(\tau) - \beta^{(w_b)}(\tau)\right)\ell \\ &= \left(I_N - \ell w_b'\right)\alpha^{(w_a)}, \end{aligned} \tag{4.8}$$

where I_N is the $N \times N$ identity matrix.

The location shift in $\beta_t^{(w_b)}$ relative to $\beta_t^{(w_a)}$ is thus $+w_b'\alpha^{(w_a)}$ (and that in $\alpha_i^{(w_b)}$ relative to $\alpha_i^{(w_a)}$ is $-w_b'\alpha^{(w_a)}$, for each i). However, the time trend is scale- and shape-invariant to choice of w. Moreover, defining the estimate

$$\widetilde{\beta}^{(w)}(\tau) = \sum_{t=1}^{T} k_{t\tau} w' y_{\cdot t} \Big/ \sum_{t=1}^{T} k_{t\tau}, \tag{4.9}$$

we have, with β as defined in Section 1,

$$\begin{aligned} E\left\{\widetilde{\beta}^{(w)}(\tau)\right\} - \beta^{(w)}(\tau) &= \sum_{t=1}^{T} \left\{\beta^{(w)}(t/T) - \beta^{(w)}(\tau)\right\} k_{t\tau} \Big/ \sum_{t=1}^{T} k_{t\tau} \\ &= \sum_{t=1}^{T} \left\{\beta(t/T) - \beta(\tau)\right\} k_{t\tau} \Big/ \sum_{t=1}^{T} k_{t\tau} \end{aligned} \tag{4.10}$$

for any w, so bias is invariant to w. Thus,

$$MSE\left\{\tilde{\beta}^{(w)}(\tau)\right\} = E\left\{\tilde{\beta}^{(w)}(\tau) - \beta^{(w)}(\tau)\right\}^2 \tag{4.11}$$

and

$$MISE\left\{\tilde{\beta}^{(w)}\right\} = \int_0^1 E\left\{\tilde{\beta}^{(w)}(u) - \beta^{(w)}(u)\right\}^2 du \tag{4.12}$$

are affected by w only through the variance. This is approximated by the leading term on the right sides of (4.13) and (4.14) in Theorem 2 below, which requires the additional

Assumption 6 Ω *is non-singular.*

Theorem 2 *Let Assumptions 1-6 hold. Then as* $T \to \infty$,

$$MSE\left\{\tilde{\beta}^{(w)}(\tau)\right\} \sim \kappa\frac{(w'\Omega w)}{Th} + \frac{h^4\chi^2}{4}\zeta(\tau), \tag{4.13}$$

$$MISE\left\{\tilde{\beta}^{(w)}\right\} \sim \kappa\frac{(w'\Omega w)}{Th} + \frac{h^4\chi^2}{4}\xi. \tag{4.14}$$

Again no proof is needed. The bias contributions in Theorems 1 and 2 are identical as discussed above, or alternatively due to β and ρ having identical second derivatives. The right hand sides of (4.13) and (4.14) are minimized, subject to (4.2), by $w = w^*$, where

$$w^* = \frac{\Omega^{-1}\ell}{\ell'\Omega^{-1}\ell}, \tag{4.15}$$

For notational ease denote $\rho(\tau) = \beta^{(w^*)}(\tau) = \beta(\tau) + (\ell'\Omega^{-1}\alpha)/(\ell'\Omega^{-1}\ell)$, $\tilde{\rho}(\tau) = \tilde{\beta}^{(w^*)}(\tau)$. Disregarding the vertical shift, $\tilde{\rho}(\tau)$ can be said to be at least as good an estimate as $\tilde{\beta}(\tau)$.

We have

$$w^{*\prime}\Omega w^* = \left(\ell'\Omega^{-1}\ell\right)^{-1}. \tag{4.16}$$

Thus the asymptotic variance, and MSE and MISE, of $\tilde{\beta}(\tau)$ equal those of $\tilde{\rho}(\tau)$ when

$$\frac{\ell'\Omega\ell}{N^2} = \ell'\Omega^{-1}\ell. \tag{4.17}$$

This is true if Ω has an eigenvector ℓ. One such situation of interest arises when the cross-sectional dependence in x_{it} is described by a spatial autoregressive model with row- and column-normalized weight matrix: in particular,

$$x_{\cdot t} = \lambda W x_{\cdot t} + \varepsilon_{\cdot t}, \tag{4.18}$$

where $\varepsilon_{\cdot t}$ is an $N \times 1$ vector of zero-mean uncorrelated, homoscedastic random variables (with variance constant also across T), the scalar $\lambda \in (-1, 1)$, and W is an $N \times N$ matrix with zero diagonal elements and satisfying $W\ell = W'\ell = \ell$. In practice elements of W are a measure of inverse economic distance between the corresponding elements of $x_{\cdot t}$. Thus W is often, though not always, symmetric, in which case if one of the latter equalities holds, so do both. This is the case in the familiar "farmers-districts" setting in which W is block-diagonal with symmetric blocks, including spatial correlation across farmers within a district, but not across districts. However, if W is non-symmetric, and only row-normalized, $\tilde{\rho}(\tau)$ is better than $\tilde{\beta}(\tau)$. As another example, suppose x_{it} has factor structure such that

$$\Omega = aI_N + bb' \tag{4.19}$$

where the scalar $a > 0$ and b is an $N \times 1$ non-null vector. Then

$$\frac{\ell'\Omega\ell}{N^2} = \frac{a}{N} + \frac{(\ell'b)^2}{N^2}, \tag{4.20}$$

$$(\ell'\Omega^{-1}\ell)^{-1} = a\left\{N - \frac{(\ell'b)^2}{a + b'b}\right\}^{-1}, \tag{4.21}$$

and (4.17) holds if and only if b is proportional to ℓ, i.e. if all factor weights are identical. Another simple example in which (4.17) holds is when there is no cross-sectional dependence in x_{it} but not all variances ω_{ii} are identical.

In terms of the impact on (4.13) and (4.14) of N increasing in the asymptotic theory, note that $\ell'\Omega^{-1}\ell \geq N/\|\Omega\|$, denoting by $\|A\|$ the square root of the greatest eigenvalue of $A'A$. Thus if $\|\Omega\|$ increases more slowly than N (it cannot increase faster) the variance components of (4.13) and (4.14) decrease faster than if N were fixed, indeed if $\|\Omega\|$ remains bounded the variance rate is $(NTh)^{-1}$, and the latter property holds even in the factor case (4.19), where $\|\Omega\|$ increases at rate N but, as (4.21) indicates, $(\ell'\Omega^{-1}\ell)^{-1} = O(N^{-1})$ if

$$\overline{\lim}_{N\to\infty}(\ell'b)^2/\{N(a+b'b)\} < 1, \tag{4.22}$$

where the expression whose upper limit is taken is clearly less than 1 for all fixed N (by the Cauchy inequality), but converges to 1 as $N \to \infty$ if, say, b is proportional to ℓ; essentially the requirement for (6.17) to hold for this factor model is one of sufficient variability in the factor weights.

5. OPTIMAL BANDWIDTH CHOICE

A key question in implementing either $\tilde{\beta}$ or $\tilde{\rho}$ is the choice of bandwidth h. Choices that are optimal in the sense of minimizing asymptotic MSE or MISE are conventional. The following theorem differs only from well known results in indicating the dependence of the optimal choices on Ω and N, and so again no proof is given.

Theorem 3 *Let Assumptions 1-5 hold. The h minimizing asymptotic MSE and MISE of $\tilde{\beta}$ are respectively*

$$h_{\beta,MSE}(\tau) = \left\{\kappa\frac{\ell'\Omega\ell}{TN^2}/\chi^2\zeta(\tau)\right\}^{1/5}, \tag{5.1}$$

$$h_{\beta,MISE} = \left\{\kappa\frac{\ell'\Omega\ell}{TN^2}/\chi^2\xi\right\}^{1/5}. \tag{5.2}$$

Let Assumption 6 also hold. The h minimizing asymptotic MSE and MISE of $\tilde{\rho}$ are respectively

$$h_{\rho,MSE}(\tau) = \left\{ \kappa \frac{(\ell'\Omega^{-1}\ell)^{-1}}{T} / \chi^2 \zeta(\tau) \right\}^{1/5}, \qquad (5.3)$$

$$h_{\rho,MISE} = \left\{ \kappa \frac{(\ell'\Omega^{-1}\ell)^{-1}}{T} / \chi^2 \xi \right\}^{1/5}. \qquad (5.4)$$

For fixed N, these optimal h all have the conventional $T^{-1/5}$ rate. If N is regarded as increasing also then unless (2.9) does not hold the rates of (5.1) and (5.2), and thus (5.3) and (5.4) (in view of (4.15)), are of smaller order. In any case (4.15) implies generally smaller optimal bandwidths for $\tilde{\rho}$ than $\tilde{\beta}$.

6. FEASIBLE OPTIMAL BANDWIDTH CHOICE AND TREND ESTIMATION

In practice the optimal bandwidths of the previous section cannot be computed. The constants κ and χ are trivially calculated, but $\zeta(\tau)$ and ξ are unknown. Discussion of their estimation can be found in the nonparametric smoothing literature, see e.g. Gasser, Kneip and Kohler (1991), and there is nothing about our setting to require additional treatment here, apart from the improved estimation possible by averaging over the cross section. More notable is the need to approximate the partly or wholly unknown Ω. This arises also if, instead of employing a plug-in proxy to the optimal bandwidths of the previous section, some automatic method such as cross-validation is employed. Estimation of Ω is also required in order to form a feasible version of the optimal trend estimate $\tilde{\rho}(\tau)$

In the simplest realistic situation, of no cross-sectional heteroscedasticity or dependence, $\Omega = \omega I_N$, with ω unknown, in which case $\tilde{\rho}(\tau) \equiv \tilde{\beta}(\tau)$. More generally we have a parametric structure permitting dependence and/or heteroscedasticity, such as in the examples of the previous section. Of course if N remains fixed as T increases, Ω is by definition parametric, so there is no theoretical loss in regarding Ω as an unrestricted positive definite $N \times N$ matrix. If then N is allowed to increase this corresponds to a nonparametric treatment of Ω. In all cases Ω is estimated by means of residuals.

It is natural to base the residuals on the originally parameterized model (2.1), but there is a choice of residuals depending on whether or not N is regarded as increasing with T, and, if it does, on the properties of the Ω sequence as N increases. Define the temporal means

$$\bar{x}_{iA} = T^{-1} \sum_{t=1}^{T} x_{it}, \ \ \bar{y}_{iA} = T^{-1} \sum_{t=1}^{T} y_{it}, \tag{6.1}$$

and the overall means

$$\bar{x}_{AA} = N^{-1} \sum_{i=1}^{N} \bar{x}_{iA}, \ \ \bar{y}_{AA} = N^{-1} \sum_{i=1}^{N} \bar{y}_{iA}. \tag{6.2}$$

Then we have, using (2.5),

$$y_{it} - \bar{y}_{iA} + \bar{y}_{AA} = \beta_t + x_{it} - \bar{x}_{iA} + \bar{x}_{AA}. \tag{6.3}$$

Employing also (2.6) gives

$$y_{it} - \bar{y}_{iA} - \bar{y}_{At} + \bar{y}_{AA} = x_{it} - \bar{x}_{iA} - \bar{x}_{At} + \bar{x}_{AA}. \tag{6.4}$$

This suggests the residual

$$\tilde{x}_{it} = y_{it} - \bar{y}_{iA} - \bar{y}_{At} + \bar{y}_{AA}. \tag{6.5}$$

Clearly $\bar{x}_{iA} = O_p(T^{-\frac{1}{2}})$ uniformly in i and $\bar{x}_{AA} = O_p(T^{-\frac{1}{2}})$, helping to justify (6.5). However, it is necessary also that $\bar{x}_{At} = o_p(1)$ and this requires $N \to \infty$, and, as discussed previously, will not hold even then in models such as (4.19) due to the strength of the cross-sectional dependence. Moreover, a satisfactory convergence rate may be required in order to show that the optimal bandwidth, and $\tilde{\rho}(\tau)$, are suitably closely approximated.

Thus, instead of considering (6.5) further, we rely solely on T being large and in view of (6.3) consider instead the residual

$$\hat{x}_{it} = y_{it} - \bar{y}_{iA} - \tilde{\beta}_t + \bar{y}_{AA}, \tag{6.6}$$

where $\tilde{\beta}_t = \tilde{\beta}(t/T)$. Thus estimate ω_{ij} by

$$\hat{\omega}_{ij} = \frac{1}{T}\sum_{t=1}^{T} \hat{x}_{it}\hat{x}_{jt}. \tag{6.7}$$

We strengthen Assumption 4 to

Assumption 7 *As $T \to \infty$,*

$$h + \left(T^2 h^3\right)^{-1} \to 0. \tag{6.8}$$

Theorem 4 *Let Assumptions 1-3, 6 and 7 hold. Then as $T \to \infty$*

$$E\left|\hat{\omega}_{ij} - \omega_{ij}\right| = O\left(h^3 + \frac{h^{3/2}}{T^{\frac{1}{2}}} + \frac{1}{Th^{\frac{1}{2}}}\right), \tag{6.9}$$

uniformly in i and j.

The proof appears in Appendix A.

Now define $\hat{\Omega}$ to be the $N \times N$ matrix with (i,j)-th element $\hat{\omega}_{ij}$, and consider the replacement of Ω by $\hat{\Omega}$ in the optimal bandwidths (5.1)-(5.4). For completeness we also assume estimates of $\hat{\zeta}(\tau)$, $\hat{\xi}$ of $\zeta(\tau)$, ξ respectively. Notice that from Theorem 4

$$\left\|\hat{\Omega} - \Omega\right\| \leq \left\{\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\hat{\omega}_{ij} - \omega_{ij}\right)^2\right\}^{\frac{1}{2}} = O_p\left(Nh^3 + \frac{Nh^{3/2}}{T^{\frac{1}{2}}} + \frac{N}{Th^{\frac{1}{2}}}\right). \tag{6.10}$$

Assumption 8

$$\hat{\zeta}(\tau) - \zeta(\tau) = O_p\left(\|\Omega\|^{-1}\left\|\hat{\Omega} - \Omega\right\|\right), \ \hat{\xi} - \xi = O_p\left(\|\Omega\|^{-1}\left\|\hat{\Omega} - \Omega\right\|\right). \tag{6.11}$$

Assumption 8 is expressed in a somewhat unprimitive way, but the order in (6.10) is actually what is required to ensure that the effect of estimating bias is negligible, and as the bias issue is of very secondary importance here we do not go into details about the rates attainable by a particular estimate. However the rate would depend *inter alia* on the rate of decay of a bandwidth (not necessarily h) used in the nonparametric estimation of β'', or more particularly, in view of (6.1), on its rate relative to that of h. Note that if N remains fixed as T increases the rate in (6.11) must be $O_p\left(h^3 + h^{3/2}T^{-\frac{1}{2}} + \left(Th^{\frac{1}{2}}\right)^{-1}\right)$, whereas if N increases then the latter rate is still relevant in, for example, the factor case (4.19), but if $\|\Omega\|$ does not increase or increases more slowly than N, then (6.11) entails a milder restriction.

Define

$$\hat{h}_{\beta,MSE}(\tau) = \left\{\kappa\frac{\ell'\hat{\Omega}\ell}{T}/\chi^2\hat{\zeta}(\tau)\right\}^{1/5}, \tag{6.12}$$

$$\hat{h}_{\beta,MISE} = \left\{\kappa\frac{\ell'\hat{\Omega}\ell}{T}/\chi^2\hat{\xi}\right\}^{1/5}, \tag{6.13}$$

$$\hat{h}_{\rho,MSE}(\tau) = \left\{\kappa\frac{(\ell'\hat{\Omega}^{-1}\ell)^{-1}}{T}/\chi^2\hat{\zeta}(\tau)\right\}^{1/5}, \tag{6.14}$$

$$\hat{h}_{\rho,MISE} = \left\{\kappa\frac{(\ell'\hat{\Omega}^{-1}\ell)^{-1}}{T}/\chi^2\hat{\xi}\right\}^{1/5}. \tag{6.15}$$

Assumption 9 *If* $N \to \infty$ *as* $T \to \infty$ *then*

$$Nh^3 + \frac{Nh^{3/2}}{T^{\frac{1}{2}}} + \frac{N}{Th^{\frac{1}{2}}} \to 0. \tag{6.16}$$

Assumption 10 *If* $N \to \infty$ *as* $T \to \infty$ *then*

$$\left\|\Omega^{-1}\right\| + \frac{N\ell'\Omega^{-2}\ell}{(\ell'\Omega^{-1}\ell)^2} = O(1). \tag{6.17}$$

Boundedness of the first term on the left is equivalent to Ω having smallest eigenvalue that is bounded away from zero for sufficiently large N, as in Assumption 6. By the Cauchy inequality, the second term on the left of (6.17) is no less than 1, so it cannot converge to zero. On the other hand, since

$$\ell'\Omega^{-1}\ell \geq \max\left\{\frac{\ell'\Omega^{-2}\ell}{\|\Omega^{-1}\|}, \frac{N}{\|\Omega\|}\right\}, \tag{6.18}$$

it follows that if also the greatest eigenvalue of Ω is bounded, the second term on the left of (6.17) is also bounded. But the greatest eigenvalue of Ω can diverge with N, as in the factor example (4.19), whereas (6.17) may be held even in this case. To see this, note that $\ell'\Omega^{-2}\ell = a^{-2}\left\{N - (\ell'b)^2\left(2a + b'b\right)/(a + b'b)^2\right\} = O(N)$, so in view of (4.21) it is sufficient that (4.22) holds.

Theorem 5 *Let Assumptions 1-3 and 5-10 hold. Then as $T \to \infty$, and possibly $N \to \infty$ also,*

$$\frac{\hat{h}_{\beta,MSE}(\tau)}{h_{\beta,MSE}(\tau)}, \frac{\hat{h}_{\beta,MISE}}{h_{\beta,MISE}}, \frac{\hat{h}_{\rho,MSE}(\tau)}{h_{\rho,MSE}(\tau)}, \frac{\hat{h}_{\rho,MISE}}{h_{\rho,MISE}} \to_p 1. \tag{6.19}$$

The proof is in Appendix A.

We can also use $\hat{\Omega}$ to obtain a feasible version of $\tilde{\rho}(\tau)$, namely

$$\hat{\rho}(\tau) = \frac{\ell'\hat{\Omega}^{-1}}{\ell'\hat{\Omega}^{-1}\ell}\sum_{t=1}^{T} k_{t\tau}y_{\cdot t}/\sum_{t=1}^{T} k_{t\tau}. \tag{6.20}$$

Theorem 2 indicates that

$$\tilde{\rho}(\tau) - \rho(\tau) = O_e\left(\frac{(\ell'\Omega^{-1}\ell)^{-\frac{1}{2}}}{(Th)^{\frac{1}{2}}} + h^2\right), \tag{6.21}$$

where O_e denotes an exact stochastic order. It is of interest to show that $\hat{\rho}(\tau)$ achieves the same asymptotic MSE as $\tilde{\rho}(\tau)$, in other words that $\tilde{\rho}(\tau) - \tilde{\rho}(\tau)$ is of smaller order than (6.21). For this purpose we introduce:

Assumption 11 *If* $N \to \infty$ *as* $T \to \infty$ *then*

$$Nh + \frac{N}{Th^{\frac{1}{2}}} \to 0. \tag{6.22}$$

Theorem 6 *Let Assumptions 1-4 and 5-11 hold. Then as* $T \to \infty$*, and possibly* $N \to \infty$ *also,*

$$\hat{\rho}(\tau) - \tilde{\rho}(\tau) = o_p\left(\frac{(\ell'\Omega^{-1}\ell)^{-\frac{1}{2}}}{(Th)^{\frac{1}{2}}} + h^2 \right). \tag{6.23}$$

The proof is in Appendix A.

7. MONTE CARLO STUDY OF FINITE SAMPLE PERFORMANCE

As always when large sample asymptotic results are presented, the issue of finite-sample relevance arises. In the present case, one interesting question is the extent to which $\hat{\rho}(\tau)$ matches the efficiency of $\tilde{\rho}(\tau)$, and whether it is actually better than $\tilde{\beta}(\tau)$, given the sampling error in estimating Ω. We study this question by Monte Carlo simulations in the case where Ω has the factor structure (4.19).

In (2.1), we thus take

$$x_{it} = b_i \eta_i + \sqrt{a} \varepsilon_{it}, \tag{7.1}$$

where the η_i and ε_{it} have mean zero and variance 1, and there is independence throughout the $\{\eta_i, i = 1, ..., \varepsilon_{it}; i, t = 1, 2, ...\}$, more particularly we take the η_i, ε_{it} to be normally distributed. We tried various values of a ($a = \frac{1}{2}, 1, 2$) and set the b_i factor weights by first generating $b_1, ..., b_N$ independently from a normal distribution, but then keeping them fixed across replications; in particular we considered the three choices $b \sim \mathcal{N}(0, I_N)$, $\mathcal{N}(0, 5I_N)$ and $N(0, 10I_N)$. With respect to the deterministic component of (2.1), we took $\beta(u) = (1+u^2)^{-1}$ throughout and fixed the individual effects α_i by first generating $\alpha_1, ..., \alpha_{N-1}$ independently from the standard normal distribution, then taking $\alpha_N = -\alpha_1 - ... - \alpha_{N-1}$ in order to satisfy (2.5), then keeping these α_i fixed across replications. The estimates $\tilde{\beta}(\tau)$, $\tilde{\rho}(\tau)$ and $\hat{\rho}(\tau)$ were computed at points $\tau = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$. For k we used the uniform kernel on $[-\frac{1}{2}, \frac{1}{2}]$, and the bandwidth values used were $h = 0.1, 0.5, 1$. Two (N, T) combinations employed, $(5, 100)$ and $(10, 500)$. Monte Carlo MSE, "$\widehat{MSE}$" in Tables 1-3, was computed in each case, based on 1000 replications. In Table 1 $Var\ b_i = 1$, in Table 2 $Var\ b_i = 5$, and in Table 3 $Var\ b_i = 10$.

(Tables 1-3 about here)

When $(N, T) = (5, 100)$ it was found that in every case $\hat{\rho}$ had larger $\widehat{MSE}$ than $\tilde{\rho}$ but smaller $\widehat{MSE}$ than $\tilde{\beta}$. However, when $(N, T) = (10, 500)$ $\hat{\rho}$ was always worse than $\tilde{\beta}$ when the b_i were generated from a distribution with variance 1, though when the latter variance was increased to 5 $\hat{\rho}$ was better in nearly half of the cases (13 out of 27), and when this variance was 10 $\hat{\rho}$ was worse than $\tilde{\beta}$ in only one case. It would appear that these results reflect the fact that the difference between variances of $\tilde{\rho}$ and $\hat{\rho}$, namely

$$\frac{\kappa}{Th} \left\{ \frac{\ell' \Omega \ell}{N^2} - \left(\ell' \Omega^{-1} \ell \right)^{-1} \right\}, \tag{7.2}$$

has to be large enough relative to the variability increase due to estimation of Ω, in order for $\hat{\rho}$ to beat $\tilde{\beta}$. In some sense the farther away Ω is from an identity matrix the better for $\hat{\rho}$'s relative performance, explaining why a large variance in generating the b_i helps, though for given such variance increasing N hurts $\hat{\rho}$ relatively (despite the helpful simultaneous increase of T).

The above description of the results is not informative of the extent to which $\hat{\rho}$ and $\tilde{\rho}$ are close, or to whether $\hat{\rho}$ beats or is beaten by $\tilde{\beta}$, and Tables 1-3 reveal these details. As a increases so does the variance of $\tilde{\rho}$, so again the potential for $\hat{\rho}$ to improve over $\tilde{\beta}$ is reduced, and the tables illustrate this. For $(N, T) = (5, 100)$, $\hat{\rho}$'s performance is often very roughly midway between those of $\tilde{\beta}$ and $\tilde{\rho}$. Also for $(N, T) = (5, 100)$, $\widehat{MSE}$ is often U-shaped in h in Tables 1 and 2, and also for $\tilde{\rho}$ and $\hat{\rho}$ in Table 3, but tends to be decreasing in h for $\tilde{\beta}$ in Table 3. The pattern of performance relative to h is less clear when $(N, T) = (10, 500)$.

In these simulations we have not pursued the discussion of optimal bandwidth choice (see Theorem 3) by incorporating feasible optimal bandwidths (as justified in Theorem 5). Some discussion of this issue was presented in Section 6, but a more explicit one is included in the following sequence of steps. Given use of a twice differenciable k, a feasible $\hat{h}_{\beta,MISE}$ (6.13) can be computed (or approximated, perhaps by cross-validation), and employed in place of h in the estimate $\tilde{\beta}(\tau)$ (3.3). Then the residuals (6.6) are calculated, followed by $\hat{\Omega}$, using (6.7). Finally, $\hat{h}_{\beta,MISE}$ (6.13) is computed and used in the estimate $\hat{\rho}(\tau)$ (6.20).

8. FURTHER DIRECTIONS FOR RESEARCH

1. Our asymptotic variance formulae for $\hat{\beta}(\tau)$ and $\hat{\rho}(\tau)$ appear also in central limit theorems, under some additional conditions, indeed one could develop joint central limit theorems for both $\tilde{\beta}(\tau)$ and $\tilde{\rho}(\tau)$ at finitely many, r, fixed frequencies $\tau_1, ..., \tau_r$, with asymptotic independence across the τ_i. When the bias is negligible relative to the standard deviation, the convergence rate will be $(Th)^{\frac{1}{2}}$ when N is fixed, and faster if N is allowed to increase with T. We could also develop a central limit theorem for $\hat{\rho}(\tau_i)$, $i = 1, ..., r$, giving the same limit distribution.

2. The assumption of regular spacing across t is easily relaxed with some regularity conditions on the spacings, since much of the fixed-design nonparametric regression estimation literature permits this.

3. Temporal correlation can also be introduced. For example, suppose

$$x_{\cdot t} = \sum_{j=0}^{\infty} A_j \varepsilon_{\cdot t-j}, \tag{8.1}$$

where the $\varepsilon_{\cdot t}$ are uncorrelated with covariance matrix I_N, and the $N \times N$ matrix A_j satisfies

$$\sum_{j=0}^{\infty} A_j z^j \neq 0, \quad |z| = 1, \tag{8.2}$$

$$\sum_{j=0}^{\infty} \|A_j\| < \infty. \tag{8.3}$$

Then

$$Var\left\{\tilde{\beta}(\tau)\right\} \sim \kappa \frac{\ell' f(0) \ell}{N^2 T h} \tag{8.4}$$

as $T \to \infty$, where

$$f(0) = \left(\sum_{j=0}^{\infty} A_j\right)\left(\sum_{j=0}^{\infty} A_j\right)', \tag{8.5}$$

which is proportional to the value of the spectral density matrix of $x_{\cdot t}$ at frequency zero. Likewise as $T \to \infty$

$$Var\left\{\widetilde{\rho}(\tau)\right\} \sim \kappa \frac{(\ell' f(0)^{-1} \ell)^{-1}}{Th}. \tag{8.6}$$

We can consistently estimate $f(0)$ by procedures of smoothed probability density estimation, and thence extend the results of Section 6.

4. We could proceed further by relaxing (8.3) to permit long memory, or on the other hand relax (8.2) to permit antipersistence. Relevant variance formula for univariate fixed-design regression with long memory or antipersistent disturbances can be found in Robinson (1997), whose results can be extended to our otherwise more general setting, one further issue arising being the possibility of varying memory parameters over the cross-section.

5. Unknown individual-specific multiplicative effects can also be incorporated. To extend (2.1),

$$y_{\cdot t} = \alpha + \gamma \beta_t + x_{\cdot t}, \tag{8.7}$$

where γ is an $N \times 1$ unknown vector. Since the scale, as well as location, of β_t is not fixed we need a normalization restriction on γ. Imposing

$$\ell'\gamma = N, \tag{8.8}$$

in addition to (2.5), we readily see that $\tilde{\beta}(\tau)$ retains the properties described in Theorem 1. However, improving on $\tilde{\beta}(\tau)$ along the lines discussed in Section 4 is more problematic. It was seen in Section 4 that changing w does not change the bias of $\tilde{\beta}^{(w)}(\tau)$ as an estimate of $\beta^{(w)}(\tau)$, and changes the asymptotic variance by a factor that depends only on w. But consider any $\alpha^{(w)}, \gamma^{(w)}, \beta^{(w)}(\tau)$, where

$$w'\alpha^{(w)} = 0, \;\; w'\gamma^{(w)} = 1 \tag{8.9}$$

and

$$\alpha + \gamma\beta(\tau) = \alpha^{(w)} + \gamma^{(w)}\beta^{(w)}(\tau). \tag{8.10}$$

We deduce that

$$\beta^{(w)}(\tau) = w'\alpha + (w'\gamma)\beta(\tau) \tag{8.11}$$

and thence

$$\alpha^{(w)} = \left(I_N - \frac{\gamma w'}{w'\gamma}\right)\alpha, \tag{8.12}$$

$$\gamma^{(w)} = \frac{\gamma}{w'\gamma}. \tag{8.13}$$

Then the bias

$$\begin{aligned} E\left\{\tilde{\beta}^{(w)}(\tau)\right\} - \beta^{(w)}(\tau) &= \sum_{t=1}^{T} k_{t\tau}\left\{\beta^{(w)}(t/T) - \beta^{(w)}(\tau)\right\} / \sum_{t=1}^{T} k_{t\tau} \\ &= (w'\gamma) BIAS\left\{\tilde{\beta}^{(w)}(\tau)\right\}, \end{aligned} \tag{8.14}$$

where $BIAS\left\{\tilde{\beta}^{(w)}(\tau)\right\}$ is (4.10). Thus, varying w changes the bias by a factor which reflects the unknown multiplicative effects, so the incorporation of these complicates the study of optimal trend estimation.

6. Analogous methods and theory could be developed for corresponding models in which the regressor in the nonparametric function is stochastic (and possibly multivariate), rather than deterministic. Robinson (2007) considers static stochastic-design nonparametric regression but without distinguishing between a time and cross-sectional dimension, and without individual effects, but his conditions on spatial dependence in regressors and disturbances are quite general and could be employed in more general settings, such as that just envisaged.

Appendix A: Proofs of Theorems

Proof of Theorem 4 We have

$$\hat{\omega}_{ij} - \omega_{ij} = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{x}_{it}\hat{x}_{jt} - x_{it}x_{jt}\right) + \frac{1}{T}\sum_{t=1}^{T}\left(x_{it}x_{jt} - \omega_{ij}\right). \tag{A.1}$$

The second term on the right has mean zero and variance bounded by CT^{-1}, where C throughout denotes a generic constant (so in the present case there is uniformity in i and j). From (6.3) and (6.6) we may write

$$\hat{x}_{it} = x_{it} + d_i + e_t, \tag{A.2}$$

where

$$d_i = \bar{x}_{AA} - \bar{x}_{iA},\ \ e_t = \beta_t - \tilde{\beta}_t. \tag{A.3}$$

Then

$$\begin{aligned} \hat{x}_{it}\hat{x}_{jt} - x_{it}x_{jt} &= \left(\hat{x}_{it} - x_{it}\right)\left(\hat{x}_{jt} - x_{jt}\right) + x_{it}\left(\hat{x}_{jt} - x_{jt}\right) + \left(\hat{x}_{it} - x_{it}\right)x_{jt} \\ &= \left(d_i + e_t\right)\left(d_j + e_t\right) + x_{it}\left(d_j + e_t\right) + \left(d_i + e_t\right)x_{jt}. \end{aligned} \tag{A.4}$$

The first term on the right of (A.4) is bounded in absolute value by $2\left(d_i^2 + d_j^2 + e_t^2\right)$. From Assumption 1, $E(d_i^2) \leq CT^{-1}$ uniformly in i. Next write

$$e_t = f_t/k_t - m_t/k_t, \tag{A.5}$$

where

$$f_t = (Th)^{-1} \sum_{s=1}^{T} k_{st}(\beta_t - \beta_s), \tag{A.6}$$

$$m_t = (Th)^{-1} \sum_{s=1}^{T} k_{st}\bar{x}_{As}, \tag{A.7}$$

$$k_t = (Th)^{-1} \sum_{s=1}^{T} k_{st}. \tag{A.8}$$

By Lemma 3, $|k_t|^{-1} \leq C$ uniformly in t and T. Next, m_t has mean zero and variance bounded by

$$C(Th)^{-2} \sum_{s=1}^{T} k_{st}^2 \leq C(Th)^{-1}, \tag{A.9}$$

applying Lemma 1. By the mean value theorem

$$\beta_s - \beta_t = \left(\frac{s-t}{T}\right)\beta_t' + \frac{1}{2}\left(\frac{s-t}{T}\right)^2 \beta_{st}'', \tag{A.10}$$

where $\beta_t' = \beta'(t/T)$ and $|\beta_{st}''| \leq C$. Thus

$$|f_t| \leq |\beta_t'| \left|(Th)^{-1} \sum_{s=1}^{T} \left(\frac{s-t}{T}\right) k_{st}\right| + (Th)^{-1} \sum_{s=1}^{T} \left(\frac{s-t}{T}\right)^2 |k_{st}\beta_{st}''|. \tag{A.11}$$

The last term is $O(h^2)$ by Lemma 1. The first factor of the first term on the right is uniformly bounded and the second factor may be written

$$h\left|(Th)^{-1} \sum_{s=1}^{T} \left(\frac{s-t}{Th}\right) k_{st} - \int_{-\infty}^{\infty} uk(u)du\right|. \tag{A.12}$$

From Lemma 2 this is bounded by

$$Ch\left\{\frac{1}{Th} + \frac{1}{T^2h^3} + \left(\frac{t}{Th}\right)^{-1} + \left(\frac{T-t}{Th}\right)^{-1}\right\} \tag{A.13}$$

for $Th \leq t \leq T - Th$. For other t we use the bound Ch from Lemma 1. It follows from Lemma 3 that f_t/k_t has the same bounds. Thus

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T}\left(\frac{f_t}{k_t}\right)^2 &\leq C\frac{Th}{T}h^2 + C\left(\frac{1}{T^2}+\frac{1}{T^4h^4}+\frac{h^2}{T}\sum_{t\geq Th}(t/Th)^{-2}\right)\\ &\leq C\left(h^3+\frac{1}{T^2}+\frac{1}{T^4h^4}+\frac{h^2.Th}{T}\right)\\ &\leq C\left(h^3+\frac{1}{T^2}+\frac{1}{T^4h^4}\right). \qquad (A.14)\end{aligned}$$

Thus, for large T,

$$E\left\{\frac{1}{T}\sum_{t=1}^{T}\left(d_i^2+d_j^2+e_t^2\right)\right\}\leq C\left(h^3+\frac{1}{Th}\right). \qquad (A.15)$$

Next, looking at the second term in (A.4) we have

$$\begin{aligned}E\left|d_j\frac{1}{T}\sum_{t=1}^{T}x_{it}\right| &\leq \left\{Ed_j^2\frac{1}{T^2}\sum_{t=1}^{T}E\left(x_{it}^2\right)\right\}^{\frac{1}{2}}\\ &\leq CT^{-1}. \qquad (A.16)\end{aligned}$$

Also

$$\begin{aligned}E\left|\frac{1}{T}\sum_{t=1}^{T}x_{it}f_t/k_t\right| &\leq \frac{C}{T}\left\{\sum_{t=1}^{T}\left(\frac{f_t}{k_t}\right)^2\right\}^{\frac{1}{2}}\\ &\leq \frac{C}{T^{\frac{1}{2}}}\left(h^{3/2}+\frac{1}{T}+\frac{1}{T^2h^2}\right). \qquad (A.17)\end{aligned}$$

Write

$$x_{it}m_t = \frac{k(0)}{Th}x_{it}\bar{x}_{At}+\frac{x_{it}}{Th}\sum_{\substack{s=1\\ s\neq t}}^{T}k_{st}\bar{x}_{As}. \qquad (A.18)$$

Thus

$$\begin{aligned}E\left|\sum_{t=1}^{T}x_{it}\frac{m_t}{k_t}\right| &\leq \frac{k(0)}{Th}E\left|\sum_{t=1}^{T}\frac{x_{it}\bar{x}_{At}}{k_t}\right|\\ &\quad +\frac{1}{Th}\left\{E\left(\sum_{t=1}^{T}\frac{x_{it}}{k_t}\sum_{\substack{s=1\\ s\neq t}}^{T}k_{st}\bar{x}_{As}\right)^2\right\}^{\frac{1}{2}}. \qquad (A.19)\end{aligned}$$

The first term on the right is bounded by $C/(T^{\frac{1}{2}}h)$. The second term is

$$\frac{1}{Th}\left\{E\sum_{t=1}^{T}\frac{x_{it}^2}{k_t^2}\left(\sum_{\substack{s=1\\s\neq t}}^{T}k_{st}\bar{x}_{As}\right)^2\right\}^{\frac{1}{2}}+\frac{1}{Th}E\left(\sum_{\substack{t=1}}^{T}\sum_{\substack{v=1\\v\neq t}}^{T}\frac{k_{tv}^2}{k_tk_v}x_{it}\bar{x}_{Av}x_{iv}\bar{x}_{At}\right)^{\frac{1}{2}}$$

$$\leq \frac{C}{Th}\left(\sum_{t=1}^{T}\sum_{s=1}^{T}k_{sv}^2\right)^{\frac{1}{2}}\leq C/h^{\frac{1}{2}} \tag{A.20}$$

by Lemma 1. It follows that

$$E\left|T^{-1}\sum_{t=1}^{T}x_{it}m_t/k_t\right|\leq C/(Th^{\frac{1}{2}}). \tag{A.21}$$

Thus,

$$\begin{aligned}E\left|\hat{\omega}_{ij}-\omega_{ij}\right| &\leq C\left(h^3+\frac{1}{Th}\right)\\ &\quad+\frac{C}{T^{\frac{1}{2}}}\left(h^{3/2}+\frac{1}{T}+\frac{1}{T^2h^2}\right)+\frac{C}{Th^{\frac{1}{2}}}\\ &\leq C\left(h^3+\frac{h^{3/2}}{T^{\frac{1}{2}}}+\frac{1}{Th}\right).\end{aligned} \tag{A.22}$$

The proof is completed.

Proof of Theorem 5 It suffices to consider only the MISE-optimal bandwidths, the proofs for the MSE-optimal ones being identical. We have

$$\hat{h}_{\beta,MISE}-h_{\beta,MISE}=\left(\frac{\kappa}{\chi^2TN^2}\right)^{1/5}\left\{\left(\frac{\ell'\hat{\Omega}\ell}{\hat{\xi}}\right)^{1/5}-\left(\frac{\ell'\Omega\ell}{\xi}\right)\right\}^{1/5}. \tag{A.23}$$

By the mean value theorem, the last factor is bounded in absolute value by

$$\frac{1}{5}\left|\tilde{r}_1\right|\left|\ell'\left(\hat{\Omega}-\Omega\right)\ell\right|+\frac{1}{5}\left|\tilde{r}_2\right|\left|\hat{\xi}-\xi\right|, \tag{A.24}$$

where $\tilde{r}_1$ lies between $(\ell'\Omega\ell)^{-4/5}\xi^{-1/5}$ and $(\ell'\hat{\Omega}\ell)^{-4/5}\hat{\xi}^{-1/5}$ and $\tilde{r}_2$ lies between $(\ell'\Omega\ell)^{1/5}\xi^{-6/5}$

and $(\ell'\hat{\Omega}\ell)^{1/5}\hat{\xi}^{-6/5}$. Then straightforwardly we deduce that (A.24) is

$$\begin{aligned}
& \left((\ell'\Omega\ell)^{-4/5}\left\{N\left\|\hat{\Omega}-\Omega\right\|+(\ell'\Omega\ell)\left|\hat{\xi}-\xi\right|\right\}\right) \\
= \; & O_p\left(N^{1/5}\left(\left\|\hat{\Omega}-\Omega\right\|+\|\Omega\|\left|\hat{\xi}-\xi\right|\right)\right) \\
= \; & O_p\left(N^{1/5}\left\|\hat{\Omega}-\Omega\right\|\right) \qquad \text{(A.25)}
\end{aligned}$$

by Assumption 8. Thus (A.23) is

$$\begin{aligned}
& O_p\left(\left(\frac{\ell'\Omega\ell}{TN^2}\right)^{1/5}\left(\frac{N}{\ell'\Omega\ell}\right)^{1/5}\left\|\hat{\Omega}-\Omega\right\|\right) \\
= \; & O_p\left(\left(\frac{\ell'\Omega\ell}{TN^2}\right)^{1/5}\left\|\hat{\Omega}-\Omega\right\|\right) = o_p\left(h_{\beta,MSE}\right). \qquad \text{(A.26)}
\end{aligned}$$

Next consider

$$\hat{h}_{\rho,MSE}-h_{\rho,MSE}=\left(\frac{\kappa}{\chi^2T}\right)^{1/5}\left\{\frac{\left(\ell'\hat{\Omega}^{-1}\ell\right)^{-1/5}}{\hat{\xi}^{1/5}}-\frac{(\ell'\Omega^{-1}\ell)^{-1/5}}{\xi^{1/5}}\right\}. \qquad \text{(A.27)}$$

The last factor is bounded in absolute value by

$$\frac{1}{5}\left|\tilde{s}_1\right|\left|\left(\ell'\hat{\Omega}^{-1}\ell\right)^{-1}-\left(\ell'\Omega^{-1}\ell\right)^{-1}\right|+\frac{1}{5}\left|\tilde{s}_2\right|\left|\hat{\xi}-\xi\right|, \qquad \text{(A.28)}$$

where $\tilde{s}_1$ lies between $(\ell'\Omega^{-1}\ell)^{4/5}\xi^{-1/5}$ and $\left(\ell'\hat{\Omega}^{-1}\ell\right)^{4/5}\hat{\xi}^{-1/5}$ and $\tilde{s}_2$ lies between $(\ell'\Omega^{-1}\ell)^{-1/5}\xi^{-6/5}$ and $\left(\ell'\hat{\Omega}^{-1}\ell\right)^{-1/5}\hat{\xi}^{-6/5}$. Now

$$\begin{aligned}
\left|\left(\ell'\hat{\Omega}^{-1}\ell\right)^{-1}-\left(\ell'\Omega^{-1}\ell\right)^{-1}\right| \; & \leq \; \frac{\left|\ell'\left(\hat{\Omega}^{-1}-\Omega^{-1}\right)\ell\right|}{\left(\ell'\hat{\Omega}^{-1}\ell\right)(\ell'\Omega^{-1}\ell)} \\
& = \; \frac{\left|\ell'\hat{\Omega}^{-1}\left(\hat{\Omega}-\Omega\right)\Omega^{-1}\ell\right|}{\left(\ell'\hat{\Omega}^{-1}\ell\right)(\ell'\Omega^{-1}\ell)} \\
& = \; O_p\left(\frac{(\ell'\Omega^{-2}\ell)}{(\ell'\Omega^{-1}\ell)^2}\left\|\hat{\Omega}-\Omega\right\|\right) \\
& = \; O_p\left(\frac{\left\|\hat{\Omega}-\Omega\right\|}{N}\right). \qquad \text{(A.29)}
\end{aligned}$$

Thus (A.28) is

$$
\begin{aligned}
& O_p\left(\left(\ell'\Omega^{-1}\ell\right)^{4/5}\frac{\left\|\hat{\Omega}-\Omega\right\|}{N}+\left(\ell'\Omega^{-1}\ell\right)^{-1/5}\left|\hat{\xi}-\xi\right|\right) \\
= \; & O_p\left(\frac{\left(\ell'\Omega^{-1}\ell\right)^{4/5}}{N}\left(\left\|\hat{\Omega}-\Omega\right\|+\|\Omega\|\left|\hat{\xi}-\xi\right|\right)\right). \qquad \text{(A.30)}
\end{aligned}
$$

It follows from Assumption 8 that

$$
\begin{aligned}
\hat{h}_{\rho,MISE}-h_{\rho,MISE} &= O_p\left(\frac{\left(\ell'\Omega^{-1}\ell\right)^{4/5}}{NT^{1/5}}\left\|\hat{\Omega}-\Omega\right\|\right) \\
&= O_p\left(\left(T\ell'\Omega^{-1}\ell\right)^{-1/5}\left\|\hat{\Omega}-\Omega\right\|\right) \\
&= o_p\left(h_{\rho,MISE}\right). \qquad \text{(A.31)}
\end{aligned}
$$

The proof is completed.

Proof of Theorem 6 We have

$$
\begin{aligned}
\hat{\rho}(\tau)-\tilde{\rho}(\tau) &= \left\{\frac{\ell'\hat{\Omega}^{-1}}{\ell'\hat{\Omega}^{-1}\ell}-\frac{\ell'\Omega^{-1}}{\ell'\Omega^{-1}\ell}\right\}\sum_{t=1}^{T}k_{t\tau}\left(\alpha+\beta_t\ell+x_{\cdot t}\right)/\sum_{t=1}^{T}k_{t\tau} \\
&= \left\{\frac{\ell'\hat{\Omega}^{-1}}{\ell'\hat{\Omega}^{-1}\ell}-\frac{\ell'\Omega^{-1}}{\ell'\Omega^{-1}\ell}\right\}\sum_{t=1}^{T}k_{t\tau}\left(\alpha+x_{\cdot t}\right)/\sum_{t=1}^{T}k_{t\tau}. \qquad \text{(A.32)}
\end{aligned}
$$

It clearly suffices to show that the factor in braces has norm $o_p\left(\left(\ell'\Omega^{-1}\ell\right)^{-\frac{1}{2}}\left(h^2+N^{-\frac{1}{2}}(Th)^{-\frac{1}{2}}\right)\right)$. This norm is bounded by

$$
\begin{aligned}
& \left\|\ell'\hat{\Omega}^{-1}\right\|\left|\left(\ell'\hat{\Omega}^{-1}\ell\right)^{-1}-\left(\ell'\Omega^{-1}\ell\right)^{-1}\right| \\
& +\left(\ell'\Omega^{-1}\ell\right)^{-1}\left\|\ell'(\hat{\Omega}^{-1}-\Omega^{-1})\right\|. \qquad \text{(A.33)}
\end{aligned}
$$

From the proof of Theorem 5,

$$
\left|\left(\ell'\hat{\Omega}^{-1}\ell\right)^{-1}-\left(\ell'\Omega^{-1}\ell\right)^{-1}\right| = O_p\left(\frac{\left\|\hat{\Omega}-\Omega\right\|}{N}\right), \qquad \text{(A.34)}
$$

so since $\left\|\ell'\hat{\Omega}^{-1}\right\| \leq N^{\frac{1}{2}}\left\|\hat{\Omega}^{-1}\right\|$ the first term is $O_p\left(N^{-\frac{1}{2}}\left\|\hat{\Omega}-\Omega\right\|\right)$. Also,

$$\left\|\ell'(\hat{\Omega}^{-1}-\Omega^{-1})\right\| = O_p\left(\left(\ell'\Omega^{-2}\ell\right)^{\frac{1}{2}}\left\|\hat{\Omega}-\Omega\right\|\right) \tag{A.35}$$

and thus from Assumption 10 the second term has the same bound. This bound is

$$\begin{aligned} & O_p\left(\left(\ell'\Omega^{-1}\ell\right)^{-\frac{1}{2}}\left\|\hat{\Omega}-\Omega\right\|\right) \\ \leq\ & O_p\left(\left(\ell'\Omega^{-1}\ell\right)^{-\frac{1}{2}} N\left(h^3+\frac{h^{3/2}}{T^{\frac{1}{2}}}+\frac{1}{Th^{\frac{1}{2}}}\right)\right). \end{aligned} \tag{A.36}$$

Now $Nh^3 = o(h^2)$ if $Nh \to 0$ and $Nh^{3/2}/T^{\frac{1}{2}} = o(h^2)$ if $N = o\left((Th)^{\frac{1}{2}}\right)$. Also, $N/(Th^{\frac{1}{2}}) = o\left((Th)^{-\frac{1}{2}}\right)$ if $N = o(T^{\frac{1}{2}})$ but this is implied by the last condition. The conclusion follows.

Appendix B: Technical Lemmas

Lemma 1 *For given $d \geq 0$ let the function $g(u)$ be such that for $c > d$, $|g(u)| \leq C\left(1+u^c\right)^{-1}$. Then*

$$\max_{1\leq t\in T}\left\{\frac{1}{Th}\sum_{s=1}^{T}\left|\frac{s-t}{Th}\right|^d\left|g\left(\frac{s-t}{Th}\right)\right|\right\} \leq C. \tag{B.1}$$

Proof. The expression in braces is bounded by

$$C(Th)^{-d-1}\sum_{|s-t|\leq U}|s-t|^d + C(Th)^{c-d-1}\sum_{|s-t|\geq U}|s-t|^{d-c}, \tag{B.2}$$

for any positive U. This is bounded by

$$C(Th)^{-d-1}U^{d+1} + C(Th)^{c-d-1}U^{d-c+1}. \tag{B.3}$$

Choosing $U \sim Th$ gives the result. ■

Lemma 2 *Let the function g be twice boundedly differentiable, and let*

$$g(u) = (1+|u|^c)^{-1} \tag{B.4}$$

for $c > 1$, and let $g(u)$ and its derivative $g'(u)$ be integrable. Then for $t \geq Th$, $t \leq T - Th$,

$$\left| \frac{1}{Th} \sum_{s=1}^{T} g\left(\frac{s-t}{Th}\right) - \int_{-\infty}^{\infty} g(u)du \right| \leq C\left(\frac{1}{Th} + \frac{1}{T^2h^3} + \left(\frac{t}{Th}\right)^{1-c} + \left(\frac{T-t}{Th}\right)^{1-c}\right). \tag{B.5}$$

Proof. The left side is

$$\sum_{s=1}^{T} \int_{(s-t-1)/Th}^{(s-t)/Th} \left\{ g\left(\frac{s-t}{Th}\right) - g(u) \right\} du \tag{B.6}$$

$$- \int_{(T-t)/Th}^{\infty} g(u)du - \int_{-\infty}^{-t/Th} g(u)du. \tag{B.7}$$

The first integrand is

$$g'\left(\frac{s-t}{Th}\right)\left(\frac{s-t}{Th} - u\right) + \frac{1}{2} g''_{st} \left(\frac{s-t}{Th} - u\right)^2, \tag{B.8}$$

where $|g''_{st}| \leq C$. Thus (B.6) is bounded in absolute value by

$$\frac{1}{(Th)^2} \sum_{s=1}^{T} \left| g'\left(\frac{s-t}{Th}\right) \right| + \frac{CT}{(Th)^3} \leq \frac{C}{Th} + \frac{C}{T^2h^3}, \tag{B.9}$$

using Lemma 1. On the other hand (B.7) is bounded in absolute value by

$$C\left(\frac{T-t}{Th}\right)^{1-c} + C\left(\frac{t}{Th}\right)^{1-c}. \tag{B.10}$$

■

Lemma 3 *For all sufficiently large T*

$$\min_{1 \leq t \leq T} k_t \geq 1/8. \tag{B.11}$$

Proof. Define

$$\kappa_t = \int_{-t/(Th)}^{(T-t)/(Th)} k(u)du. \tag{B.12}$$

Then

$$k_t - \kappa_t = \sum_{s=1}^{T} \int_{-t/(Th)}^{(T-t)/(Th)} \left\{ k\left(\frac{s-t}{Th}\right) - k(u) \right\} du. \tag{B.13}$$

The integrand is

$$k'\left(\frac{s-t}{Th}\right)\left(\frac{s-t}{Th} - u\right) + C_t\left(\frac{s-t}{Th} - u\right)^2, \tag{B.14}$$

where

$$\max_{1 \le t \le T} C_t < \infty. \tag{B.15}$$

Thus (B.13) is bounded in absolute value by

$$\begin{aligned} & (Th)^{-2}\sum_{s=1}^{T}\left|k'\left(\frac{s-t}{Th}\right)\right| + C_t T (Th)^{-3} \le C(Th)^{-1} + CT^{-2}h^3 \\ \le \;& 1/8, \end{aligned} \tag{B.16}$$

say, for large enough T, using Lemma 1 and Assumption 7. Now for large enough T,

$$\min_{1 \le t \le T} \kappa_t \ge \int_0^{1/(2h)} k(u)du \ge \frac{1}{4} \tag{B.17}$$

and thus

$$\min_{1 \le t \le T} k_t = \min_{1 \le t \le T} \kappa_t - \max_{1 \le t \le T} |k_t - \kappa_t| > 1/8. \tag{B.18}$$

■

Acknowledgement

This research was supported by ESRC Grant RES-062-23-0036. I am grateful for the helpful comments of a referee and Jungyoon Lee, and to the latter also for carrying out the simulations.

References

Bai, J., S. Ng, 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191-221.

Benedetti, J.K., 1977. On the nonparametric estimation of regression functions. Journal of the Royal Statisticl Society Series B39, 248-257.

Gasser, T., A. Kneip, W. Kohler, 1991. A flexible and fast method for automatic smoothing. Journal of the American Statistical Association 86, 647-652.

Hart, J.D., T.E. Wehrly, 1986. Kernel regression estimation using repeated measurements data. Journal of the American Statistical Association 81, 1080-1088.

Moyeed, K.A., P.J. Diggle, 1994. Rates of convergence in semi-parametric modelling of longitudinal data. Australian Journal of Statistics 36, 75-93.

Robinson, P.M., 1997. Large-sample inference for nonparametric regression with dependent errors. Annals of Statistics 25, 2059-2083.

Robinson, P.M., 2007. Asymptotic properties of nonparametric regression with spatial data. Preprint.

Ruckstuhl, A.F., A.H. Welsh, R.J. Carroll, 2000. Nonparametric function estimation of the relationship between two repeatedly measured variables. Statistica Sinica 10, 51-71.

Severini, T.A., J.G. Staniswalis, 1994. Quasi-likelihood estimation in semiparametric models. Journal of the American Statistical Association 89, 501-512.

Starica, C., C. Granger, 2005. Nonstationarities in stock returns. Review of Economics and Statistics 87, 503-522.