

# Inference on power law spatial trends

PETER M. ROBINSON

*Department of Economics, London School of Economics, London WC2A 2AE, UK.  
E-mail: p.m.robinson@lse.ac.uk*

Power law or generalized polynomial regressions with unknown real-valued exponents and coefficients, and weakly dependent errors, are considered for observations over time, space or space–time. Consistency and asymptotic normality of nonlinear least-squares estimates of the parameters are established. The joint limit distribution is singular, but can be used as a basis for inference on either exponents or coefficients. We discuss issues of implementation, efficiency, potential for improved estimation and possibilities of extension to more general or alternative trending models to allow for irregularly spaced data or heteroscedastic errors; though it focusses on a particular model to fix ideas, the paper can be viewed as offering machinery useful in developing inference for a variety of models in which power law trends are a component. Indeed, the paper also makes a contribution that is potentially relevant to many other statistical models: Our problem is one of many in which consistency of a vector of parameter estimates (which converge at different rates) cannot be established by the usual techniques for coping with implicitly-defined extremum estimates, but requires a more delicate treatment; we present a generic consistency result.

*Keywords:* asymptotic normality; consistency; correlation; generalized polynomial; lattice; power law

## 1. Introduction

Polynomial-in-time regression is one of the longest-established tools of time series analysis (see Jones [9]). In much empirical work, especially when stochastic trends, such as unit roots, are also involved, only a linear trend is countenanced, or merely a constant intercept. On the other hand, classical methods can test polynomial order when observations are equally spaced in time. With independent and identically distributed (i.i.d.) normal errors, a particularly elegant way of achieving this, with finite sample validity, results from an orthogonal polynomial representation – the covariance matrix of the least-squares estimate (LSE) is diagonalized, and contributions to the  $F$  statistic from individual regressors are i.i.d. (see Section 3.2.2 of Anderson [1]). Asymptotic theory is valid under much wider conditions on the errors; indeed from Section 7.4 of Grenander and Rosenblatt [5], the LSE is asymptotically efficient (in the Gauss–Markov sense) when the (possibly non-Gaussian) errors are covariance stationary with spectral density bounded and bounded away from zero at zero frequency, as with short memory processes. Polynomial models have also been extended to spatial lattice data (see Section 3.4 of Cressie [2]).

Polynomials are nevertheless restrictive. The Weierstrass theorem justifies their uniform approximation of any continuous function over a compact interval, but seems less practically relevant the longer the data set. Nonparametric smoothing may be unreliable in a series of moderate length, when instead richer parametric models than polynomials might be considered. One class that advantageously nests polynomials, which has received little theoretical attention, consists of “generalized polynomial” or “power law” models. With equally spaced time series observations

$y_u, u = 1, \dots, N$ , consider

$$y_u = \sum_{j=1}^p \beta_j u^{\theta_j} + x_u, \tag{1.1}$$

where the  $\theta_j$  and  $\beta_j$  are real valued and all can be unknown,  $\theta_j > -1/2$  for all  $j$ , and the zero-mean unobservable process  $x_u$  is covariance stationary with short memory. For  $\theta_j < -1/2$ ,  $\beta_j$  would not be estimable (whether  $\theta_j$  were known or unknown) because the corresponding signal is drowned by the noise. For  $\theta_j = -1/2$ ,  $\beta_j$  is estimable but we omit this possibility because our central limit theorem requires  $\theta_j$  to lie in the interior of a compact set. Polynomials, such as when  $\theta_j = j - 1$  for all  $j$ , are nested; indeed this is a hypothesis that might be tested within (1.1).

We consider the nonlinear least-squares estimate (NLSE) of the  $\theta_j, \beta_j$  in (1.1) and, more generally, of exponents and coefficients in an extended model defined on a lattice, applying to spatial and spatio-temporal data, where our provision, for example, for weaker trends than linear ones and for decaying trends seems practically useful. Unlike the LSE when exponents are known, the NLSE cannot be expressed in closed form and requires numerical optimization. Correspondingly, asymptotic theory, with sample size  $N$  increasing, is needed to justify rules of statistical inference even when errors are Gaussian. We establish consistency and asymptotic normality for the NLSE of exponent and coefficient estimates, achieving also an analogous efficiency bound to that described above. As with other implicitly defined estimates, asymptotic distribution theory makes use (in application of the mean value theorem) of an initial consistency proof. Many such proofs (see Jennrich [8], Malinvaud [12]) require regressors to be non-trending, whence under suitable additional conditions all parameter estimates are  $N^{1/2}$ -consistent. For the NLSE of (1.1), Wu [21] significantly relaxed this requirement but nevertheless appears to heavily restrict the diversity of trends. The discussion after Assumptions A and A' of Wu [21] indicates that they reduce in (1.1) with known  $\theta_j$  to the assumption  $\max_j \theta_j < \frac{1}{2} + 2 \min_j \theta_j$ , and no weaker requirement suffices in the case of unknown  $\theta_j$ . Example 4 of Wu [21] addressed the latter case but with  $p = 1$  only (and for  $\theta_1 \in (-\frac{1}{2}, 0]$ ) when the inequality is trivially satisfied. In general, more elaborate techniques seem required to establish consistency in (1.1). Moreover, Wu [21] established consistency with no rate, whereas we find that a slow rate of convergence in the  $\theta_j$  estimates is required before asymptotic normality is established. Wu [21] also established asymptotic normality of the NLSE in a quite general setting, but under the assumption that all parameter estimates converge at the same rate. This is not the case with (1.1); indeed all rates of  $\theta_j, \beta_j$  estimates turn out to differ. For implicitly defined extremum estimates such variation is typically associated with difficulty in the initial consistency proof due to the objective function not converging uniformly to a function that is uniquely optimized over the whole parameter space. Consistency proofs here have tended to be geared to the case at hand (see e.g. Giraitis, Hidalgo and Robinson [4], Nagaraj and Fuller [13], Nielsen [14], Robinson [15], Sun and Phillips [17]). Our consistency proof employs a generic result (presented and proved in Appendix A to avoid interrupting the flow) that seems likely to apply to a quite general class of estimates (not just the NLSE) of a variety of models. Our asymptotic distribution theory of estimates for (1.1) and its extension presents some other unusual features.

The following section presents the model, regularity conditions and three theorems describing asymptotic statistical properties. The main details of their proofs appear in Appendix B. These use a series of propositions, stated and proved in Appendix C, and relying in turn also on a

series of lemmas, in Appendix D. A Monte Carlo study of finite sample performance appears in Section 3, while Section 4 discusses aspects of the theoretical results and their implementation, with possible extensions.

## 2. Estimation of spatial lattice regression model

Let the integer  $d \geq 1$  represent the dimension on which data are observed, where  $d = 1$  for time series (as in (1.1)) and  $d \geq 2$  for spatial or spatio-temporal data. Generalize  $u$  to the  $d$ -dimensional multi-index  $u = (u_1, u_2, \dots, u_d)'$ . Denoting  $\mathbb{Z}_+ = \{j: j = 0, 1, \dots\}$ , generalize (1.1) to

$$y_u = \sum_{i=1}^d \sum_{j=1}^{p_i} \beta_{ij} u_i^{\theta_{ij}} + x_u = f(u; \theta)' \beta + x_u, \quad u \in \mathbb{Z}_+^d, \tag{2.1}$$

where  $x_u$  is described subsequently and  $\beta = (\beta'_1, \dots, \beta'_d)'$ ,  $\beta_i = (\beta_{i1}, \dots, \beta_{ip_i})'$ ,  $\theta = (\theta'_1, \dots, \theta'_d)'$ ,  $\theta_i = (\theta_{i1}, \dots, \theta_{ip_i})'$ ,  $f(u; \theta) = (f_1(u_1; \theta_1)', \dots, f_d(u_d; \theta_d)')'$ ,  $f_i(u_i; \theta_i) = (u_i^{\theta_{i1}}, \dots, u_i^{\theta_{ip_i}})'$ , for  $i = 1, \dots, d$ . Defining  $p = p_1 + \dots + p_d$ , the  $p \times 1$  vectors  $\beta$  and  $\theta$  are supposed unknown. Any  $f_i(u_i; \theta_i)$  might be absent from  $f(u; \theta)$  when corresponding  $\theta_i$  and  $\beta_i$  are void; we proceed as if corresponding  $p_i$  and sums over  $j = 1, \dots, p_i$  are zero, avoiding indicator functions to describe such circumstances.

Our consistency proof confines the NLSE of  $\theta$  to a compact set. Prescribe an (arbitrarily small) positive  $\delta$ , and for each  $i = 1, \dots, d$ , prescribe  $\underline{\Delta}_i, \bar{\Delta}_i$  such that  $-1/2 < \underline{\Delta}_i < \bar{\Delta}_i < \infty$ , and define

$$\Theta_i = \{h_1, \dots, h_{p_i}: h_1 \geq \underline{\Delta}_i; h_j - h_{j-1} \geq \delta, j = 2, \dots, p_i; h_{p_i} \leq \bar{\Delta}_i\} \tag{2.2}$$

and  $\Theta = \prod_{i=1}^d \Theta_i$ . We introduce two assumptions that imply identifiability of  $\theta$  and  $\beta$ .

**Assumption 1.**  $\theta \in \Theta$ .

**Assumption 2.**  $\theta_{ij} = 0$  for at most one  $(i, j)$ ;  $\beta_{ij} \neq 0$  for all  $(i, j)$ .

Assumption 1 implies

$$-1/2 < \theta_{i1} < \dots < \theta_{ip_i} < \infty, \quad i = 1, \dots, d. \tag{2.3}$$

The ordering in (2.3) is arbitrary, and distinctness of the  $\theta_{ij}$  across  $j$  along with the first part of Assumption 2 identifies  $\beta$ ; note that  $u_i^0 = 1$  for all  $i$  and that we allow an intercept but do not require one. The second part of Assumption 2 identifies  $\theta$ .

Given  $N = \prod_{i=1}^d n_i$  observations on  $y_u, u \in \mathbb{N} = \mathbb{N}_1 \times \dots \times \mathbb{N}_d$  and  $\mathbb{N}_i = (1, \dots, n_i)$ , define the NLSE of  $\beta, \theta$  by  $(\hat{\beta}, \hat{\theta}) = \arg \min_{b \in \mathbb{R}^p, h \in \Theta} Q(b, h)$ , where  $Q(b, h) = \sum_{u \in \mathbb{N}} \{y_u - b' f(u; h)\}^2$ . Asymptotic theory requires further assumptions. Let  $\mathbb{Z} = \{j: j = 0, \pm 1, \dots\}$ .

**Assumption 3.**  $x_u, u \in \mathbb{Z}^d$ , is covariance stationary with zero mean, and its autocovariance function,  $\gamma_u = \text{cov}(x_t, x_{t+u})$ , for the multi-index  $t = (t_1, \dots, t_d)'$ , satisfies  $\sum_{u \in \mathbb{Z}^d} |\gamma_u| < \infty$ .

Our parameter estimates make no attempt to correct for this possible nonparametric weak dependence of the  $x_u$  (permitted also in Assumption 5), and Cressie [2], page 25, stresses the importance of mean function specification relative to error specification. However, the NLSE turns out to be not only consistency-robust to spatial correlation but also asymptotically Gauss–Markov efficient.

The next assumption, of increase with algebraic rate of observations in all dimensions, is capable of generalization but is employed for simplicity.

**Assumption 4.**  $n_i \sim B_i N^{b_i}$ ,  $i = 1, \dots, d$ , as  $N \rightarrow \infty$ , where  $B_i > 0$ ,  $b_i > 0$ ,  $i = 1, \dots, d$ ,  $\prod_{i=1}^d B_i = \sum_{i=1}^d b_i = 1$ .

Define  $\zeta_{ij} = b_i \theta_{ij}$  and, with no loss of generality, identify dimension  $i = 1$  such that

$$\zeta_{11} = \min_{1 \leq i \leq d} \{\zeta_{i1}\}, \tag{2.4}$$

where, if two or more  $i$  satisfy (2.4), an arbitrary choice is made. Note that  $\zeta_{11} + \frac{1}{2} > 0$  is implied by  $\theta_{11} + \frac{1}{2} > 0$ .

**Theorem 1.** *Let Assumptions 1–4 hold. Then for  $j = 1, \dots, p_i$ ,  $i = 1, \dots, d$ , as  $N \rightarrow \infty$ ,*

$$\hat{\theta}_{ij} - \theta_{ij} = O_p(N^{\chi - \zeta_{ij} - 1/2}) \tag{2.5}$$

for any  $\chi > 0$ .

The proof is in Appendix B. As is common with initial consistency proofs, a sharp rate (corresponding to  $\chi = 0$  in (2.5)) is not delivered (smoothness conditions, in particular, are not exploited). Theorem 1 is used in the proof of our central limit theorem (CLT), for which we also need consistency, with a rate, for  $\beta$ . We state this result without the proof, which is a relatively straightforward application of Theorem 1, techniques used in its proof, Theorem 3 below and routine manipulations.

**Theorem 2.** *Let Assumptions 1–4 hold. Then, for  $j = 1, \dots, p_i$ ,  $i = 1, \dots, d$ ,*

$$\hat{\beta}_{ij} = \beta_{ij} + O_p((\log N) N^{\chi - \zeta_{ij} - 1/2}), \quad \text{as } N \rightarrow \infty.$$

The relative rates for the  $\hat{\theta}_{ij}$  and  $\hat{\beta}_{ij}$  in Theorems 1 and 2 are matched by relative rates that feature in our CLT. For this we introduce first

**Assumption 5.**  $x_u = \sum_{v \in \mathbb{Z}^d} \xi_v \varepsilon_{u-v}$ ,  $\sum_{v \in \mathbb{Z}^d} |\xi_v| < \infty$ ,  $u \in \mathbb{Z}^d$ , where  $v$  is the multi-index  $v = (v_1, \dots, v_d)'$ ,  $\{\varepsilon_u, u \in \mathbb{Z}^d\}$  are independent random variables with zero mean and unit variance,  $\{\varepsilon_u^2, u \in \mathbb{Z}^d\}$  are uniformly integrable and  $\sum_{v \in \mathbb{Z}^d} \xi_v \neq 0$ .

Assumption 5 implies Assumption 3, and both imply the existence and boundedness of the spectral density  $F(\lambda) = (2\pi)^{-1} |\sum_{v \in \mathbb{Z}^d} \xi_v e^{iv \cdot \lambda}|^2$  of  $x_u$ , where  $\lambda$  is the multi-index  $\lambda =$

$(\lambda_1, \dots, \lambda_d)'$ , while Assumption 5 also implies  $F(0) > 0$ . Stationary invertible autoregressive moving averages are among time series processes covered by Assumption 5, as are spatial generalizations of these (see e.g. Hallin, Lu and Tran [6], Robinson and Vidal Sanz [16], Tjøstheim [18, 19], Yao and Brockwell [24]). Mixing conditions, such as ones employed in a spatial context by Gao, Lu and Tjøstheim [3], Hallin, Lu and Yu [7], and Lu, Lundervold, Tjøstheim and Yao [11], provide an alternative route for establishing a CLT, but are not strictly weaker or stronger than Assumption 5, which we prefer here because  $x_u$ , unlike processes considered in the latter references, is involved only linearly.

Let  $I_r$  be the  $r$ -rowed identity matrix,  $\otimes$  denote the Kronecker product, and introduce  $p \times p$  matrices  $D = N^{1/2} \text{diag}\{n_1^{\theta_{11}}, \dots, n_1^{\theta_{1p_1}}, \dots, n_d^{\theta_{d1}}, \dots, n_d^{\theta_{dp_d}}\}$ ,  $L(s) = \text{diag}\{L_1(s_1), \dots, L_d(s_d)\}$ , where  $L_i(s_i) = (\log s_i)I_{p_i}$ , and  $(2p \times 2p)$  matrices  $D_+ = I_2 \otimes D$  and  $L_+ = \text{diag}\{I_p, L(n)\}$ . Define  $\alpha = (\theta', \beta')'$ ,  $\hat{\alpha} = (\hat{\theta}', \hat{\beta}')'$ . Denote by  $\mathfrak{N}_r(a, A)$  an  $r$ -dimensional normal vector with mean vector  $a$  and (possibly singular) covariance matrix  $A$ . Appendix B defines the  $p \times p$  matrix  $\Upsilon$  and  $p \times 2p$  matrix  $B$  and proves:

**Theorem 3.** *Let Assumptions 1, 2 and 5 hold. Then as  $N \rightarrow \infty$ ,*

$$D_+L_+^{-1}(\hat{\alpha} - \alpha) \rightarrow_d \mathfrak{N}_{2p}(0, 2\pi F(0)B'\Upsilon^{-1}B).$$

### 3. Finite sample properties

A small Monte Carlo study provides some information on finite sample performance. Issues of concern, given unknown  $\theta$ , are bias and variability of the NLSE and accuracy of large sample inference rules suggested by Theorem 3. We employed (2.1) with  $d = 2$ ,  $p_1 = p_2 = 1$ , picking 2  $(\theta_1, \theta_2) = (\theta_{11}, \theta_{21})$  combinations – (1, 1), (0.5, 2) – but throughout took  $\Theta_{i1} = [-0.45, 4]$ ,  $\beta_i = \beta_{i1} = 1$ ,  $i = 1, 2$ . We varied  $N$  absolutely and also the relative  $n_1, n_2$ , taking  $n_1, n_2 = (8, 12)$ , (10, 10), (11, 20), (15, 15).

Our first experiment took the  $x_u$  to be i.i.d.  $\mathfrak{N}_1(0, 1)$  variables. Tables 1 and 2 report, for the respective parameter combinations, bias (BIAS), mean squared error (MSE), and empirical size at 5% (SIZE5) and 1% (SIZE1) for the NLSE  $\hat{\theta}_i, \hat{\beta}_i$ , and also  $\tilde{\beta}_i$ , the LSE of  $\beta_i$  that correctly assumes  $\theta$ , for  $i = 1, 2$ , across 1000 replications. The sizes were proportions of significant estimates, using normal critical values scaled by estimated standard deviations which, in the case of the  $\hat{\theta}_i, \hat{\beta}_i$ , were computed on the basis of Theorem 3 with current parameter estimates replacing true values of  $\theta, \beta$ , and  $2\pi F(0)$  replaced by the sum of squared residuals divided by  $N$  (so the spatial independence of the  $x_u$  was treated as known, as it was also in the conventional scaling used for the  $\tilde{\beta}_i$ ).

The tables reveal a definite inferiority of the NLSE relative to the LSE, but unsurprisingly, as the LSE is exactly unbiased, more efficient and yields exact critical regions. Though the NLSE-based tests on  $\beta$  are nearly always over-sized, this phenomenon diminishes with increased  $N$ , and overall the discrepancy between the performances of the two classes of the  $\beta$  estimate does not seem very serious. There is also a predominate over-sizing of the tests on  $\theta$ , but again this falls as  $N$  increases, and, in Table 2 in particular, it is often modest. There is a tendency for the NLSE to over-estimate, but for  $\beta$  biases only exceed 2% of the parameter value when  $n_i = 8$

**Table 1.**  $\theta_1 = 1, \theta_2 = 1, \beta_1 = 1, \beta_2 = 1, \sigma^2 = 1, x_u$  i.i.d.

| $n_1$ | $n_2$ |       | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\beta}_1$ | $\tilde{\beta}_1$ | $\hat{\beta}_2$ | $\tilde{\beta}_2$ |
|-------|-------|-------|------------------|------------------|-----------------|-------------------|-----------------|-------------------|
| 8     | 12    | BIAS  | 0.008            | 0.007            | 0.024           | 0.000             | 0.017           | 0.000             |
|       |       | MSE   | 0.016            | 0.007            | 0.080           | 0.001             | 0.051           | 0.000             |
|       |       | SIZE5 | 0.100            | 0.125            | 0.151           | 0.048             | 0.166           | 0.055             |
|       |       | SIZE1 | 0.044            | 0.048            | 0.075           | 0.010             | 0.084           | 0.010             |
| 10    | 10    | BIAS  | 0.005            | 0.009            | 0.016           | -0.001            | 0.009           | 0.002             |
|       |       | MSE   | 0.010            | 0.009            | 0.060           | 0.006             | 0.063           | 0.007             |
|       |       | SIZE5 | 0.132            | 0.132            | 0.180           | 0.053             | 0.186           | 0.051             |
|       |       | SIZE1 | 0.055            | 0.050            | 0.084           | 0.015             | 0.090           | 0.011             |
| 11    | 20    | BIAS  | -0.002           | 0.002            | 0.016           | 0.000             | -0.007          | 0.000             |
|       |       | MSE   | 0.003            | 0.001            | 0.022           | 0.000             | 0.010           | 0.000             |
|       |       | SIZE5 | 0.086            | 0.104            | 0.115           | 0.039             | 0.120           | 0.051             |
|       |       | SIZE1 | 0.030            | 0.039            | 0.051           | 0.005             | 0.049           | 0.012             |
| 15    | 15    | BIAS  | 0.003            | 0.002            | 0.006           | 0.000             | -0.001          | 0.000             |
|       |       | MSE   | 0.002            | 0.002            | 0.013           | 0.000             | 0.013           | 0.000             |
|       |       | SIZE5 | 0.074            | 0.075            | 0.108           | 0.043             | 0.103           | 0.039             |
|       |       | SIZE1 | 0.024            | 0.022            | 0.033           | 0.010             | 0.037           | 0.010             |

and  $n_i = 12$ . For  $\theta$  they never reach 1%, while overall they mostly fall with increasing  $N$ , as does the MSE. In Table 2, the results are not in line with what the rates in Theorem 3 suggest, because the fall in MSE is greater for  $\hat{\theta}_2$  and  $\hat{\beta}_2$  than for  $\hat{\theta}_1$  and  $\hat{\beta}_1$ , despite the fact that  $\theta_1 = 2$  and  $\theta_2 = \frac{1}{2}$ . Nevertheless, it is not clear to what extent one would expect asymptotic theory to predict comparisons at this level of refinement in such sample sizes. Note that the Monte Carlo results are also difficult to judge relative to the theory because the various  $n_i$  did not result from fixing the  $b_i$  and  $B_i$  and then increasing  $n$ , but were chosen with a view to representing some variability in  $n$ , and relative to  $n_1$  and  $n_2$ . In addition, the convergence rates of  $\hat{\theta}_i$  and  $\hat{\beta}_i$  do not only depend on  $n_i$ , but on the overall  $n$ . Other results are more closely in line with the asymptotic theory. This is the case in Table 1 where, with  $\theta_1 = \theta_2 = 1$ , the above MSE ratios are sometimes greater for  $\hat{\theta}_2$  and/or  $\hat{\beta}_2$  and sometimes less. It is also the case in Table 2 for the LSE  $\tilde{\beta}_i$ , as elsewhere, that comparisons are sometimes difficult as a number of MSEs are zero to 3, and even to 4 (unreported here), decimal places.

Next we considered the effect of dependence, employing three different models for  $x_u$ , again with  $d = 2$ . All models entailed weak dependence, with varying spans, but in the first dependence was negative, so that the spectral density at zero was small, whereas in the other two it was positive, producing a peaked spectral density. In the following,  $\varepsilon_u \sim$  i.i.d.  $\mathfrak{N}_1(0, 1)$ .

1. Multiple direction MA(1):

$$x_u = \varepsilon_u - 0.12 \sum_{j=-1}^1 \sum_{\substack{k=-1 \\ (j,k) \neq 0}}^1 \varepsilon_{u_1+j, u_2+k}, \quad u_i = 1, \dots, n_i, i = 1, 2. \quad (3.1)$$

**Table 2.**  $\theta_1 = 2, \theta_2 = 1/2, \beta_1 = 1, \beta_2 = 1, \sigma^2 = 1, x_u$  i.i.d.

| $n_1$ | $n_2$ |       | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\beta}_1$ | $\tilde{\beta}_1$ | $\hat{\beta}_2$ | $\tilde{\beta}_2$ |
|-------|-------|-------|------------------|------------------|-----------------|-------------------|-----------------|-------------------|
| 8     | 12    | BIAS  | 0.008            | 0.001            | 0.024           | 0.003             | -0.002          | -0.000            |
|       |       | MSE   | 0.014            | 0.001            | 0.071           | 0.005             | 0.001           | 0.000             |
|       |       | SIZE5 | 0.063            | 0.060            | 0.087           | 0.077             | 0.053           | 0.090             |
|       |       | SIZE1 | 0.028            | 0.012            | 0.038           | 0.029             | 0.014           | 0.034             |
| 10    | 10    | BIAS  | 0.008            | 0.000            | 0.020           | 0.004             | 0.000           | -0.000            |
|       |       | MSE   | 0.013            | 0.003            | 0.074           | 0.004             | 0.001           | 0.000             |
|       |       | SIZE5 | 0.069            | 0.057            | 0.101           | 0.058             | 0.065           | 0.039             |
|       |       | SIZE1 | 0.033            | 0.013            | 0.047           | 0.015             | 0.017           | 0.009             |
| 11    | 20    | BIAS  | 0.005            | -0.000           | -0.001          | -0.002            | 0.000           | 0.000             |
|       |       | MSE   | 0.005            | 0.000            | 0.028           | 0.002             | 0.000           | 0.000             |
|       |       | SIZE5 | 0.052            | 0.054            | 0.069           | 0.030             | 0.059           | 0.041             |
|       |       | SIZE1 | 0.017            | 0.012            | 0.017           | 0.012             | 0.011           | 0.006             |
| 15    | 15    | BIAS  | 0.002            | 0.001            | 0.004           | 0.001             | 0.004           | 0.000             |
|       |       | MSE   | 0.004            | 0.001            | 0.025           | 0.001             | 0.000           | 0.000             |
|       |       | SIZE5 | 0.058            | 0.044            | 0.070           | 0.081             | 0.043           | 0.055             |
|       |       | SIZE1 | 0.018            | 0.011            | 0.019           | 0.019             | 0.010           | 0.020             |

2. Multilateral MA(4), no interactions:

$$x_u = \varepsilon_u + \sum_{\substack{j=-4 \\ j \neq 0}}^4 a_{|j|}(\varepsilon_{u_1+j, u_2} + \varepsilon_{u_1, u_2+j}), \quad u_i = 1, \dots, n_i, i = 1, 2 \quad (3.2)$$

for  $a_1 = 0.14, a_2 = 0.12, a_3 = 0.1, a_4 = 0.08$ .

3. Bilateral MA(9), on diagonal:

$$x_u = \varepsilon_u + \sum_{\substack{j=-9 \\ j \neq 0}}^9 (0.95)^{|j|} \varepsilon_{u_1+j, u_2+j}, \quad u_i = 1, \dots, n_i, i = 1, 2. \quad (3.3)$$

For the same parameter values as before, bias and MSE of the LSE and NLSE are presented in Tables 3–8, with Tables 3 and 4 referring to (3.1), Tables 5 and 6 to (3.2), and Tables 7 and 8 to (3.3). As before the LSE  $\tilde{\beta}_1, \tilde{\beta}_2$  are exactly unbiased, as the Monte Carlo results tend to illustrate. However, perhaps surprisingly, the dependent model (3.3) produces some very large biases in the NLSE  $\hat{\beta}_1$ , though not so much in  $\hat{\beta}_2, \hat{\theta}_1, \hat{\theta}_2$ . For the other dependence models the NLSE biases are not necessarily greater than under independence. The MSE magnitudes are not directly comparable to those of Tables 1 and 2, because scales were not calibrated, but a similar overall picture emerges: the NLSE of  $\beta$  often has much greater MSE than the LSE, but this falls with increasing  $N$ , as does that of the NLSE of  $\theta$ . In Tables 4, 6 and 8, where  $\theta_1 = 2, \theta_2 = \frac{1}{2}$ ,

**Table 3.**  $\theta_1 = 1, \theta_2 = 1, \beta_1 = 1, \beta_2 = 1, \sigma^2 = 1, x_u = (3.1)$

| $n_1$ | $n_2$ |      | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\beta}_1$ | $\tilde{\beta}_1$ | $\hat{\beta}_2$ | $\tilde{\beta}_2$ |
|-------|-------|------|------------------|------------------|-----------------|-------------------|-----------------|-------------------|
| 8     | 12    | BIAS | 0.005            | 0.003            | 0.006           | 0.000             | 0.002           | -0.000            |
|       |       | MSE  | 0.006            | 0.003            | 0.031           | 0.000             | 0.021           | 0.000             |
| 10    | 10    | BIAS | 0.005            | 0.001            | 0.001           | 0.000             | 0.008           | -0.000            |
|       |       | MSE  | 0.003            | 0.003            | 0.023           | 0.000             | 0.023           | 0.000             |
| 11    | 20    | BIAS | 0.001            | 0.0001           | 0.001           | -0.000            | 0.001           | 0.000             |
|       |       | MSE  | 0.001            | 0.000            | 0.006           | 0.000             | 0.003           | 0.000             |
| 15    | 15    | BIAS | 0.002            | -0.001           | -0.003          | -0.000            | 0.005           | 0.000             |
|       |       | MSE  | 0.000            | 0.000            | 0.004           | 0.000             | 0.004           | 0.000             |

the same somewhat surprising feature as noted in Table 2 appears, with  $\hat{\theta}_1$  and  $\hat{\beta}_1$  improving less than  $\hat{\theta}_2$  and  $\hat{\beta}_2$  with increasing  $n$ , and the only additional point to add to our previous discussion is that convergence is often expected to be slowed by dependence.

### 4. Final comments

1. For known  $\theta$ , long-established techniques (see [1], Section 2.6) give  $D(\hat{\beta}(\theta) - \beta) \rightarrow_d \mathfrak{N}_p(0, 2\pi F(0)\Phi^{-1})$  (where  $\Phi$  is defined near the start of Appendix B below), so ignorance of  $\theta$  incurs not only efficiency loss, but slightly slower convergence. Theorem 3 also implies a singularity in the limit distribution, whose covariance matrix has rank  $p$  only. This is due to bias in  $\hat{\beta}$ , which on expansion is seen to have a term linear in  $\hat{\theta} - \theta$  that dominates the contribution from  $\sum_{u \in \mathbb{N}} f(u; \theta)x_u$ . Nevertheless, Theorem 3 does provide separate inference on  $\beta$  (moreover, one can conduct joint inference that does not cover both  $\theta_{ij}$  and  $\beta_{ij}$  for any  $(i, j)$ ), though, given Assumption 1, we cannot test zero restrictions on  $\beta$ . In our setting,  $\beta$  may be of less initial

**Table 4.**  $\theta_1 = 2, \theta_2 = 1/2, \beta_1 = 1, \beta_2 = 1, \sigma^2 = 1, x_u = (3.1)$

| $n_1$ | $n_2$ |      | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\beta}_1$ | $\tilde{\beta}_1$ | $\hat{\beta}_2$ | $\tilde{\beta}_2$ |
|-------|-------|------|------------------|------------------|-----------------|-------------------|-----------------|-------------------|
| 8     | 12    | BIAS | 0.003            | 0.000            | 0.003           | -0.000            | -0.000          | 0.000             |
|       |       | MSE  | 0.003            | 0.000            | 0.017           | 0.001             | 0.000           | 0.000             |
| 10    | 10    | BIAS | -0.003           | 0.000            | 0.014           | -0.001            | -0.001          | 0.000             |
|       |       | MSE  | 0.003            | 0.000            | 0.018           | 0.001             | 0.000           | 0.000             |
| 11    | 20    | BIAS | -0.000           | 0.000            | 0.003           | 0.000             | -0.000          | -0.000            |
|       |       | MSE  | 0.001            | 0.000            | 0.004           | 0.000             | 0.000           | 0.000             |
| 15    | 15    | BIAS | -0.001           | 0.000            | 0.005           | 0.001             | -0.000          | -0.000            |
|       |       | MSE  | 0.000            | 0.000            | 0.004           | 0.000             | 0.000           | 0.000             |



**Table 5.**  $\theta_1 = 1, \theta_2 = 1, \beta_1 = 1, \beta_2 = 1, \sigma^2 = 1, x_u = (3.2)$

| $n_1$ | $n_2$ |      | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\beta}_1$ | $\tilde{\beta}_1$ | $\hat{\beta}_2$ | $\tilde{\beta}_2$ |
|-------|-------|------|------------------|------------------|-----------------|-------------------|-----------------|-------------------|
| 8     | 12    | BIAS | 0.032            | 0.020            | 0.053           | -0.001            | 0.035           | 0.000             |
|       |       | MSE  | 0.050            | 0.026            | 0.249           | 0.004             | 0.169           | 0.002             |
| 10    | 10    | BIAS | 0.029            | 0.020            | 0.017           | -0.005            | 0.047           | 0.003             |
|       |       | MSE  | 0.031            | 0.031            | 0.181           | 0.003             | 0.177           | 0.003             |
| 11    | 20    | BIAS | 0.010            | 0.003            | 0.017           | -0.001            | 0.015           | 0.001             |
|       |       | MSE  | 0.013            | 0.004            | 0.091           | 0.001             | 0.045           | 0.000             |
| 15    | 15    | BIAS | 0.008            | 0.007            | 0.006           | -0.001            | 0.014           | 0.000             |
|       |       | MSE  | 0.007            | 0.008            | 0.059           | 0.000             | 0.060           | 0.001             |

interest than  $\theta$ , and Theorem 3 allows inference on  $\theta$  with  $\hat{\theta}$  converging slightly faster than  $\hat{\beta}$ , and at what appears to be the optimal rate for this problem.

2. If independence of the  $x_u$  is not assumed, the limiting covariance matrix in Theorem 3 can be consistently estimated (under additional conditions) by replacing  $F(0)$  by a parametric or smoothed nonparametric estimate based on NLSE residuals.

3. The form of the limiting covariance matrix in Theorem 3, with dependence simply reflected in the scale factor  $2\pi F(0)$ , suggests that a generalized NLSE, which corrects parametrically or nonparametrically for correlation in  $x_u$ , affords no efficiency improvement (cf. Section 7.4 of Grenander and Rosenblatt [5]).

4. On the other hand, our estimates are not Fisher efficient for non-Gaussian  $x_u$ . Departures from Gaussianity might be detected by, for example, nonparametric probability density estimation based on NLSE residuals; Hallin, Lu and Tran [6] studied density estimation for linear lattice processes. More efficient parameter estimates could be obtained by  $M$ -estimation using a correctly parameterized  $\varepsilon_u$  distribution, or adapting semi-parametrically to a nonparametric one, in either case employing parametric  $\{\xi_v\}$  or approximating them via a long autoregression. The extra proof details would be far from trivial, but convergence rates should be unaffected, with the limiting covariance matrix of Theorem 3 simply shrunk by a scalar factor.

**Table 6.**  $\theta_1 = 2, \theta_2 = 1/2, \beta_1 = 1, \beta_2 = 1, \sigma^2 = 1, x_u = (3.2)$

| $n_1$ | $n_2$ |      | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\beta}_1$ | $\tilde{\beta}_1$ | $\hat{\beta}_2$ | $\tilde{\beta}_2$ |
|-------|-------|------|------------------|------------------|-----------------|-------------------|-----------------|-------------------|
| 8     | 12    | BIAS | 0.064            | 0.001            | 0.048           | 0.005             | -0.001          | -0.000            |
|       |       | MSE  | 0.115            | 0.000            | 0.272           | 0.024             | 0.003           | 0.000             |
| 10    | 10    | BIAS | 0.067            | -0.001           | 0.023           | -0.002            | 0.005           | 0.000             |
|       |       | MSE  | 0.111            | 0.001            | 0.267           | 0.019             | 0.005           | 0.000             |
| 11    | 20    | BIAS | 0.019            | 0.000            | 0.035           | 0.000             | -0.001          | 0.000             |
|       |       | MSE  | 0.027            | 0.000            | 0.151           | 0.009             | 0.000           | 0.000             |
| 15    | 15    | BIAS | 0.008            | 0.000            | 0.046           | -0.002            | -0.001          | 0.000             |
|       |       | MSE  | 0.020            | 0.000            | 0.143           | 0.007             | 0.001           | 0.000             |

**Table 7.**  $\theta_1 = 1, \theta_2 = 1, \beta_1 = 1, \beta_2 = 1, \sigma^2 = 1, x_u = (3.3)$

| $n_1$ | $n_2$ |      | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\beta}_1$ | $\tilde{\beta}_1$ | $\hat{\beta}_2$ | $\tilde{\beta}_2$ |
|-------|-------|------|------------------|------------------|-----------------|-------------------|-----------------|-------------------|
| 8     | 12    | BIAS | 0.074            | 0.096            | 0.154           | 0.008             | 0.091           | -0.004            |
|       |       | MSE  | 0.129            | 0.157            | 0.738           | 0.048             | 0.549           | 0.024             |
| 10    | 10    | BIAS | 0.041            | 0.069            | 0.105           | -0.008            | 0.050           | 0.008             |
|       |       | MSE  | 0.080            | 0.097            | 0.455           | 0.033             | 0.371           | 0.032             |
| 11    | 20    | BIAS | 0.016            | 0.036            | 0.134           | 0.0010            | 0.017           | -0.000            |
|       |       | MSE  | 0.043            | 0.032            | 0.462           | 0.014             | 0.232           | 0.005             |
| 15    | 15    | BIAS | 0.013            | 0.024            | 0.061           | -0.003            | 0.028           | 0.002             |
|       |       | MSE  | 0.026            | 0.026            | 0.214           | 0.009             | 0.182           | 0.009             |

5. Another extension allows long or negative memory, in  $x_u$ , bearing in mind results of Yajima [22] for (1.1) with known integer  $\theta_i$ , and Yajima and Matsuda [23]; this would affect all convergence rates by the same scalar factor, the efficiency property in Comment 3 would be lost, and negative  $\theta_{ij}$ , and corresponding  $\beta_{ij}$  may not be estimable.

6. In an alternative formulation to (1.1),  $u^{\theta_j}$  is replaced by  $(u/N)^{\theta_j}$ , confining the regression to the unit interval, and (2.1) can be analogously modified. Consistency is then much easier to prove, all exponent estimates being  $\sqrt{N}$ -consistent. A similar device is employed in fixed-design nonparametric regression, but unlike there it is not essential in order to achieve consistency in our parametric setting, where we find it aesthetically unattractive given that  $x_u$  is defined on an increasing domain.

7. The results are straightforwardly extended to allow some  $\theta_{ij}$  in (2.1) to be known; for example, to specify an intercept by  $\theta_{11} = 0$ , though the norming factor and limit covariance matrix in Theorem 3 are affected.

8. Our notation suggests constant spacing between observations across all  $d$  dimensions, but allowing the interval of observation to vary with dimension affects each  $\beta_{ij}$  by a factor depending also on the corresponding  $\theta_{ij}$ , but not the  $\theta_{ij}$  themselves.

**Table 8.**  $\theta_1 = 2, \theta_2 = 1/2, \beta_1 = 1, \beta_2 = 1, \sigma^2 = 1, x_u = (3.3)$

| $n_1$ | $n_2$ |      | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\beta}_1$ | $\tilde{\beta}_1$ | $\hat{\beta}_2$ | $\tilde{\beta}_2$ |
|-------|-------|------|------------------|------------------|-----------------|-------------------|-----------------|-------------------|
| 8     | 12    | BIAS | 0.063            | -0.000           | 0.100           | 0.014             | 0.009           | -0.000            |
|       |       | MSE  | 0.518            | 0.003            | 1.217           | 0.291             | 0.019           | 0.000             |
| 10    | 10    | BIAS | 0.098            | -0.000           | 0.118           | 0.009             | 0.008           | -0.000            |
|       |       | MSE  | 0.512            | 0.003            | 0.912           | 0.222             | 0.016           | 0.000             |
| 11    | 20    | BIAS | -0.037           | -0.002           | -0.007          | -0.001            | 0.008           | 0.000             |
|       |       | MSE  | 0.275            | 0.000            | 1.059           | 0.128             | 0.004           | 0.000             |
| 15    | 15    | BIAS | 0.054            | 0.000            | 0.128           | -0.001            | 0.001           | 0.000             |
|       |       | MSE  | 0.226            | 0.000            | 0.616           | 0.086             | 0.003           | 0.000             |

9. Irregular spacing of observations, either due to missing data from an otherwise regular lattice, or with observations occurring anywhere on  $\mathbb{R}^d$ , can also be considered. In both of these settings asymptotic theory requires a degree of regularity in the observation locations, ruling out situations where observations become too sparse, for example. Given this, the extension is relatively simple with independent  $x_u$ . Under dependence, asymptotic variance formulae will be complicated by the irregular spacing and the efficiency property of Comment 3 will be lost. In addition, different kinds of assumptions from ours on the errors  $x_u$  may be needed. In the case of missing data from an otherwise regular lattice, our Assumptions 3 (for consistency) and 5 (for asymptotic normality) should still suffice. But for observations anywhere on  $\mathbb{R}^d$  it would be appropriate to consider an underlying continuous process. Then, for consistency, a suitable ergodicity property would be needed, whereas for asymptotic normality leading possibilities that can entail weak dependence analogous to that of Assumption 5 include suitable linear functionals of Brownian motion and mixing conditions.

10. A Bayesian treatment would be worthwhile, with suitable priors placed on the exponents and possibly also the coefficients.

11. When  $d \geq 2$  a more realistic model than (2.1) might allow interaction terms, that is, products of powers of  $u_i$  and  $u_k$ ,  $i \neq k$ . Our proof methods are extendable, but from a practical perspective the curse of dimensionality threatens and the issue of parsimonious specification, already posed by (2.1), becomes more pressing. A penalized procedure could be used.

12. Modified model classes might provide an alternative route to parsimony; for example, one might take  $p_i = 1$  with  $\beta_{i1}u_{i1}^{\theta_{i1}}$  replaced by  $\beta_{i1}(u_{i1} + \phi_{i1})^{\theta_{i1}}$  for known or unknown  $\phi_{i1}$  (cf. Example 3 of Wu [21]). Trigonometric factors might also be incorporated (cf. Section 7.5 of Grenander and Rosenblatt [5]).

13. For alternative classes of trending model (for example, involving wavelets), asymptotic estimation theory might be handled by similar techniques.

14. An alternative practically relevant modelling of the  $x_u$  treats them as heteroscedastic but possibly independent. Broadly similar proof techniques would provide corresponding results to ours, but the NLSE is less efficient than a suitably weighted estimate.

15. Though we have focussed on (1.1) and (2.1) to fix ideas, our methods and theory can be developed to cover models that incorporate power law trends along with other explanatory variables, both stochastic and non-stochastic, such as extensions of the nonparametric and semi-parametric spatial regressions considered by Gao, Lu and Tjøstheim [3] and Lu, Lundervold, Tjøstheim and Yao [11], and so the paper can be viewed as introducing machinery relevant to a wide variety of settings.

## Appendix A: Generic consistency theorem

We present a consistency theorem for a general, implicitly defined extremum estimate under unprimitive conditions that will be checked in the paper's setting and seem capable of checking in a number of others. As this appendix is self-contained, there seems no risk of confusion in employing notations that are similar to those elsewhere in the paper but can have slightly different meanings. We estimate the  $p \times 1$  vector parameter  $\theta$ , with elements  $\theta_i$ ,  $i = 1, \dots, p$ , by  $\hat{\theta} = \arg \min_{h \in \Theta} R(h)$ , where  $R(h) : \mathbb{R}^p \rightarrow \mathbb{R}$  depends on sample size  $N$  and  $\Theta \subset \mathbb{R}^p$  is a fixed

compact set. For positive scalars  $C_{iw}, i = 1, \dots, p, w = 1, 2, \dots$ , depending on  $N$  and such that  $C_{iw} \leq C_{i,w+1}, i = 1, \dots, p$ , define  $C_w = (C_{1w}, \dots, C_{pw})'$ , and

$$\begin{aligned} \mathcal{N}_i(C_{iw}) &= \{h_i: |h_i - \theta_i| < C_{iw}\}, & \mathcal{N}(C_w) &= \prod_{i=1}^p \mathcal{N}_i(C_{iw}), \\ \bar{\mathcal{N}}(C_w) &= \Theta \setminus \mathcal{N}(C_w), & \mathcal{S}_w &= \bar{\mathcal{N}}(C_w) \cap \mathcal{N}(C_{w+1}). \end{aligned} \tag{A.1}$$

**Theorem A.** Assume:

- (i)  $\Theta \subset \mathcal{N}(C_{W+1})$  for a finite integer  $W$  and  $N$  sufficiently large;
- (ii) There exist positive  $s_1, \dots, s_W$  and  $U(h), V(h)$  such that  $R(h) = R(\theta) + U(h) + V(h)$  and  $s_1 < \dots < s_W$ , and as  $N \rightarrow \infty, s_1 \rightarrow \infty$  and

$$P\left(\inf_{h \in \mathcal{S}_w} \frac{U(h)}{s_w} > \eta\right) \rightarrow 1, \quad \text{some } \eta > 0, \tag{A.2}$$

$$\sup_{h \in \mathcal{S}_w} \frac{|V(h)|}{s_w} = o_p(1). \tag{A.3}$$

Then

$$\hat{\theta} = \theta + \mathcal{O}_p(C_1), \quad \text{as } N \rightarrow \infty,$$

where  $\mathcal{O}_p(C_1)$  is a  $p \times 1$  vector with  $i$ th element  $\mathcal{O}_p(C_{i1})$ .

**Proof.** We show that  $P(\hat{\theta} \in \bar{\mathcal{N}}(C_1)) \rightarrow 0$  as  $N \rightarrow \infty$ . By a standard kind of argument

$$P(\hat{\theta} \in \bar{\mathcal{N}}(C_1)) \leq P\left(\inf_{h \in \bar{\mathcal{N}}(C_1)} \{R(h) - R(\theta)\} \leq 0\right).$$

Under (i),  $\bar{\mathcal{N}}(C_1) \subset \bar{\mathcal{N}}(C_1) \cap \mathcal{N}(C_{W+1}) = \bigcup_{w=1}^W \mathcal{S}_w$ . Thus the last probability is bounded by

$$\sum_{w=1}^W P\left(\inf_{h \in \mathcal{S}_w} \left\{ \frac{R(h) - R(\theta)}{s_w} \right\} \leq 0\right) \leq \sum_{w=1}^W P\left(\sup_{h \in \mathcal{S}_w} \frac{|V(h)|}{s_w} \geq \inf_{h \in \mathcal{S}_w} \frac{U(h)}{s_w}\right),$$

which is bounded by

$$\sum_{w=1}^W \left\{ P\left(\sup_{h \in \mathcal{S}_w} \frac{|V(h)|}{s_w} > \eta\right) + P\left(\inf_{h \in \mathcal{S}_w} \frac{U(h)}{s_w} \leq \eta\right) \right\}, \tag{A.4}$$

which tends to zero on applying (A.2) and (A.3). □

Three comments are relevant. (1) In the setting of the rest of the paper,  $U$  can be chosen non-stochastic but this is not possible in the context of such stochastic trends as unit roots, where the more general (A.2) is useful. (2) An almost sure convergence version of Theorem A is possible

under suitably strengthened versions of (A.2) and (A.3). (3) By comparison with our decomposition of  $\tilde{\mathcal{N}}(C_1)$  into  $\mathcal{S}_1, \dots, \mathcal{S}_W$ , van de Geer [20] (see pages 69, 70) employed a “peeling device” to obtain an exponential inequality for  $\sup_{g \in \mathcal{G}} \{|Z_N(g)|/\tau(g)\}$ , where  $Z_N(g)$  is a stochastic process,  $\tau(g)$  is a non-negative function and the set  $\mathcal{G}$  is “peeled off” as  $\bigcup_{j=1}^J \mathcal{G}_j$ , where  $\mathcal{G}_j = \{g \in \mathcal{G}: m_{j-1} \leq \tau(g) < m_j\}$ , for an increasing sequence  $\{m_j\}$ , and  $J$  need not be finite. Thus  $\sup_{g \in \mathcal{G}_j} \{|Z_N(g)|/\tau(g)\} \leq \{\sup_{g \in \mathcal{G}, \tau(g) < m_j} |Z_N(g)|\}/m_{j-1}$  and only the supremum of the numerator of the original statistic need be approximated. There is no denominator there like  $\tau(g)$  in our problem, and our decomposition of  $\tilde{\mathcal{N}}(C_1)$  is designed to suitably balance  $U(h)$  and  $V(h)$  on each  $\mathcal{S}_w$  to enable choices of the  $s_w$  that make all  $W$  summands in (A.4) small.

## Appendix B: Definitions and proofs of theorems

To define  $\Upsilon$ , introduce first, for  $i = 1, \dots, d$ , the  $p_i \times 1$  vector  $\phi_i(g_i)$  with  $j$ th element  $(g_{ij} + 1)^{-1}$  and the  $p_i \times p_i$  matrix  $\Phi_i(g_i, h_i)$  with  $(j, k)$ th element  $(g_{ij} + h_{ik} + 1)^{-1}$  for  $g_i = (g_{i1}, \dots, g_{ip_i})'$ ,  $h_i = (h_{i1}, \dots, h_{ip_i})'$ , where  $g_{ij}, h_{ij} > -1/2$  for all  $i, j$ . For  $g = (g'_1, \dots, g'_d)'$ ,  $h = (h'_1, \dots, h'_d)'$ , introduce the  $p \times p$  matrix  $\Phi(g, h)$  with  $(i, j)$ th  $p_i \times p_j$  block  $\Phi_i(g_i, h_i)$  when  $i = j$  and  $\phi_i(g_i)\phi_j(h_j)'$  when  $i \neq j$ . Denote  $\Phi = \Phi(\theta, \theta)$ . Writing  $\phi_i = \phi_i(\theta_i)$ ,  $\Phi_i = \Phi_i(\theta_i, \theta_i)$ , define  $p \times p$  matrices  $\Phi_+, \Phi_{++}$  with  $(i, j)$ th  $p_i \times p_j$  block  $\Phi_i \circ \Phi_i, 2\Phi_i \circ \Phi_i \circ \Phi_i$  when  $i = j$  and  $\phi_i(\phi_j \circ \phi_j)'$ ,  $(\phi_i \circ \phi_i)(\phi_j \circ \phi_j)'$  when  $i \neq j$ , where “ $\circ$ ” denotes the Hadamard product. Put  $\Upsilon = \Phi_{++} - \Phi'_+ \Phi^{-1} \Phi_+$ . Define  $B = (\beta_\Delta^{-1}, -I_p)$ , where  $\beta_\Delta$  is the  $p \times p$  diagonal matrix such that  $\beta_\Delta 1_p = \beta$  and  $1_p$  is the  $p \times 1$  vector of 1’s.

**Proof of Theorem 1.** We have  $\hat{\theta} = \arg \min_{h \in \Theta} R(h)$ ,  $\hat{\beta} = \hat{\beta}(\hat{\theta})$ , where

$$R(h) = Q(\hat{\beta}(h), h), \quad \hat{\beta}(h) = M(h, h)^{-1} \{M(h, \theta)\beta + m(h)\}$$

for  $M(g, h) = \sum_{u \in \mathbb{N}} f(u; g)f(u; h)'$ ,  $m(h) = \sum_{u \in \mathbb{N}} f(u; h)x_u$ . The subsequent proof implies that after suitable norming  $M(h, h)$  is well conditioned for relevant  $h$  and large  $N$ . In Theorem A, take  $U(h) = \beta' D\Psi(h)D\beta$ ,  $V(h) = V_1(h) - \{V_2(h) - V_2(\theta)\} - \{V_3(h) - V_3(\theta)\}$ , for  $V_1(h) = \beta' \{P(h) - D\Psi(h)D\}\beta$ ,  $V_2(h) = 2m(h)'M(h, h)^{-1}M(h, \theta)\beta$ ,  $V_3(h) = m(h)'M(h, h)^{-1}m(h)$ , with  $\Psi(h) = \Phi(\theta, \theta) - \Phi(\theta, h)\Phi(h, h)^{-1}\Phi(h, \theta)$ ,  $P(h) = M(\theta, \theta) - M(\theta, h)M(h, h)^{-1}M(h, \theta)$ . Define, for  $j = 1, \dots, p_i$ ,  $i = 1, \dots, d$ , and a finite  $W$ , positive scalars  $C_{ijw}$ ,  $w = 1, \dots, W$ , such that  $C_{ijw} \leq C_{ij, w+1}$  for each such  $w$ . Define

$$C_w = (C_{11w}, \dots, C_{1p_1w}, \dots, C_{d1w}, \dots, C_{dp_dw}), \quad w = 1, \dots, W + 1. \tag{B.1}$$

Define neighbourhoods  $\mathcal{N}_{ij}(C_{ijw}) = \{h_{ij}: |h_{ij} - \theta_{ij}| < C_{ijw}\}$ ,  $j = 1, \dots, p_i$ ,  $i = 1, \dots, d$ ,  $w = 1, \dots, W + 1$ . Finally, define for  $w = 1, \dots, W + 1$ ,

$$\mathcal{N}(C_w) = \prod_{i=1}^d \prod_{j=1}^{p_i} \mathcal{N}_{ij}(C_{ijw}), \tag{B.2}$$

and then  $\tilde{\mathcal{N}}(C_w)$ ,  $\mathcal{S}_w$  as in (A.1). Take  $C_{ij1} = N^{\chi - \zeta_{ij} - 1/2} \sim B_i^{\theta_{ij}} N^{\chi - 1/2} n_i^{-\theta_{ij}}$ ,  $j = 1, \dots, p_i$ ,  $i = 1, \dots, d$ , so we need to show that  $P(\hat{\theta} \in \tilde{\mathcal{N}}(C_1)) \rightarrow 0$  as  $N \rightarrow \infty$ . We check (i) and (ii) of

Theorem A, where (A.2) reduces to the requirement  $\inf_{h \in \mathcal{S}_w} U(h)/s_w > \eta$  for large enough  $N$  and  $\eta$  as in (A.2). From (B.1) and (B.2),

$$\mathcal{S}_w \subset \Theta \cap \mathcal{T}_w,$$

where

$$\mathcal{T}_w = \bigcup_{i=1}^d \bigcup_{j=1}^{p_i} \{h_{ij}: |h_{ij} - \theta_{ij}| \geq C_{ijw}; h_{kl}: h_{kl} \in (-1/2, \infty), \text{ all } (k, l) \neq (i, j)\}.$$

It follows from Proposition 1 that

$$\inf_{h \in \mathcal{S}_w} U(h) \geq \eta^* N \min_{i,j} \beta_{ij}^2 \sum_{i=1}^d \sum_{j=1}^{p_i} n_i^{2\theta_{ij}} C_{ijw}^2 \geq \frac{\eta}{p} \sum_{i=1}^d \sum_{j=1}^{p_i} N^{1+2\zeta_{ij}} C_{ijw}^2.$$

Thus (A.2) is satisfied when

$$\sum_{i=1}^d \sum_{j=1}^{p_i} N^{1+2\zeta_{ij}} C_{ijw}^2 \geq ps_w. \tag{B.3}$$

Next, (A.3) is implied if

$$\sup_{h \in \mathcal{S}_w} |V_1(h)| = o(s_w), \tag{B.4}$$

$$\sup_{h \in \mathcal{S}_w} |V_2(h) - V_2(\theta)| = o_p(s_w), \tag{B.5}$$

$$\sup_{h \in \mathcal{S}_w} |V_3(h)| = o_p(s_w), \tag{B.6}$$

as  $N \rightarrow \infty$ . Note that in (B.5) we are considering the difference  $V_2(h) - V_2(\theta)$  for  $h$  suitably close to  $\theta$  and this closeness is important in obtaining the desired result, whereas in the usual kind of consistency proof, for standard, non-mixed rate settings, one more simply shows the convergence to zero in probability of a suitably normalized  $V_2(h)$ , uniformly in  $h \in \Theta$ . Now (B.6) follows from Proposition 4, while (B.4) and (B.5) follow from Propositions 2 and 3, respectively, if

$$\sum_{i=1}^d \sum_{j=1}^{p_i} N^{1+2\zeta_{ij}-\delta^*} C_{ij,w+1}^2 = o(s_w), \tag{B.7}$$

where  $\delta^* = \min[\min_{1 \leq i \leq d} \{b_i/2 + \min(b_i \Delta_i, 0)\}, 2\chi]$  (implying  $\delta^* > 0$ ) and

$$\sum_{i=1}^d \sum_{j=1}^{p_i} N^{1/2+\zeta_{ij}+\varepsilon} C_{ij,w+1} = o(s_w) \tag{B.8}$$

for some  $\varepsilon > 0$ .

It remains to show that we can choose  $W$  and the  $s_w, C_{ijw}$ , to satisfy (i) of Theorem A and (B.3), (B.7) and (B.8). Now (B.3) holds for  $w = 1$  if  $s_1 = N^{2\chi}$ , and for  $w > 1$  if

$$s_w = s_1 N^{(w-1)\delta^*/2} = N^{2\chi+(w-1)\delta^*/2},$$

$$C_{ijw} = C_{ij1} N^{(w-1)\delta^*/4} = N^{\chi-\zeta_{ij}-1/2+(w-1)\delta^*/4}, \quad j = 1, \dots, p_i, i = 1, \dots, d.$$

Since

$$N^{1+2\zeta_{ij}} C_{ij1}^2 = s_1, \quad N^{1+2\zeta_{ij}-\delta^*} C_{ij,w+1}^2 = s_1 N^{(w/2-1)\delta^*} = s_w N^{-\delta^*/2}$$

for all  $i, j$ , (B.7) is satisfied. For all  $i, j$ ,

$$N^{1/2+\zeta_{ij}+\varepsilon} C_{ij,w+1} = N^{\chi+\varepsilon+w\delta^*/4} = s_w N^{\varepsilon-\chi+\delta^*/4+(1-w)\delta^*/4} = o(s_w),$$

on taking  $\varepsilon < \chi - \delta^*/4$ , to satisfy (B.8). Finally, for all  $i, j$ , though  $C_{ij1} \rightarrow 0$  as  $N \rightarrow \infty$  (no matter how small  $\delta^*$  or how large  $\zeta_{ij}$ ), we have  $C_{ijw} \rightarrow \infty$  as  $N \rightarrow \infty$  for large enough  $w$ , so there is a finite  $W$  to satisfy (i) of Theorem A. □

**Proof of Theorem 2.** Omitted. □

**Proof of Theorem 3.** Put  $a = (h', b)'$ ,  $Q(a) = Q(h, b)$  and define  $Q^{(1)}(a) = (\partial/\partial a)Q(a)$ ,  $Q^{(2)}(a) = (\partial/\partial a')Q^{(1)}(a)$ . We have

$$L_+ Q^{(1)}(a) = -2 \sum_{u \in \mathbb{N}} \{y_u - b' f(u; h)\} H(u; h, b),$$

where  $H(u; h, b) = [(L(u) f(u; h) \circ b)', (L f(u; h))']'$  with  $L = L(n)$  and  $L_+ Q^{(2)}(a) L_+ = \sum_{i=1}^3 Q_i^{(2)}(a)$ , with

$$Q_1^{(2)}(a) = 2 \sum_{u \in \mathbb{N}} H(u; h, b) H(u; h, b)',$$

$$Q_2^{(2)}(a) = 2 \sum_{u \in \mathbb{N}} \{b' f(u; h) - \beta' f(u; \theta)\} J(u; h, b),$$

$$Q_3^{(2)}(a) = -2 \sum_{u \in \mathbb{N}} x_u J(u; h, b),$$

in which  $J(u; h, b)$  is the  $2p \times 2p$  symmetric matrix with  $(i, j)$ th  $p \times p$  block  $L(u) f_{\Delta}(u; h) L(u) b_{\Delta}$  for  $i = j = 1$ ,  $L(u) f_{\Delta}(u; h) L$  for  $i = 1, j = 2$  and 0 for  $i = j = 2$ ,  $b_{\Delta}, f_{\Delta}(u, h)$  being the  $p \times p$  diagonal matrices such that  $b = b_{\Delta} 1_p, f(u; h) = f_{\Delta}(u; h) 1_p$ .

By the mean value theorem

$$D_+ L_+^{-1} (\hat{\alpha} - \alpha) = (D_+^{-1} L_+ \tilde{Q}^{(2)} L_+ D_+^{-1})^{-1} D_+^{-1} L_+ Q^{(1)}(\alpha), \tag{B.9}$$

where  $\tilde{Q}^{(2)}$  is formed from  $Q^{(2)}(a)$  by evaluating its  $i$ th row at  $a = \bar{\alpha}_{(i)}$ , where  $\|\bar{\alpha}_{(i)} - \alpha\| \leq \|\hat{\alpha} - \alpha\|$ ,  $i = 1, \dots, 2p$ . By Proposition 5 (B.9) is

$$\{D_+^{-1}L_+Q^{(2)}(\alpha)L_+D_+^{-1} + O_p(\log N)^{-2}\}^{-1}D_+^{-1}L_+Q^{(1)}(\alpha).$$

Let  $B_\Delta = \text{diag}(\beta_\Delta^{-1}, -I_p)$  and  $\Gamma$  be the  $2p \times 2p$  matrix with  $p \times p$  blocks  $\Gamma_{11} = 0$ ,  $\Gamma_{21} = \Gamma'_{12} = L^{-1}\Lambda$ ,  $\Gamma_{22} = -L^{-1}\Lambda - \Lambda'L^{-1}$ , with  $\Lambda = \Phi^{-1}\Phi_+\Upsilon^{-1}$ . Noting Proposition 6 and the representations

$$\begin{aligned} BD_+^{-1}L_+Q^{(1)}(\alpha) &= 2N^{-1/2} \sum_{u \in \mathbb{N}} \{L(u) - L\}D^{-1}f(u; \theta)x_u, \\ B_\Delta\Gamma B_\Delta D_+^{-1}L_+Q^{(1)}(\alpha) &= -2N^{-1/2} \sum_{u \in \mathbb{N}} [(\beta_\Delta^{-1}\Lambda)'], (L^{-1}\Lambda\{L(u) - L\} - \Lambda')' \\ &\quad \times D^{-1}f(u; \theta)x_u, \end{aligned}$$

we obtain from (B.9)

$$\begin{aligned} D_+L_+^{-1}(\hat{\alpha} - \alpha) &= -N^{-1/2}B \sum_{u \in \mathbb{N}} [\Upsilon^{-1}\{L(u) - L\} + \Lambda']D^{-1}f(u; \theta)x_u \\ &\quad - N^{-1/2} \sum_{u \in \mathbb{N}} [0, (L^{-1}\Lambda\{L(u) - L\}D^{-1}f(u; \theta)x_u)'] \\ &\quad + O_p((\log N)^{-2})N^{-1/2} \sum_{u \in \mathbb{N}} ((\beta_\Delta L(u))', L)'D^{-1}f(u; \theta)x_u. \end{aligned}$$

The last two terms are  $O_p((\log N)^{-1})$  by application of Lemmas 15 and 10, respectively. The proof is completed by applying Proposition 7 to the first term.  $\square$

### Appendix C: Propositions

**Proposition 1.** For all  $C_w$  given by (B.1) such that  $C_{ijw} > 0$ ,  $j = 1, \dots, p_i$ ,  $i = 1, \dots, d$ , there exists  $\eta^* > 0$  such that, for all  $\theta \in \Theta$

$$\inf_{h \in \mathcal{N}(C_w)} U(h) \geq \eta^* N \sum_{i=1}^d \sum_{j=1}^{p_i} \beta_{ij}^2 n_i^{2\theta_{ij}} C_{ijw}^2.$$

**Proof.** Non-singularity of  $\Phi(h, h)$  for  $h \in \Theta$ , and

$$\sup_{\Theta} \|\Phi(h, h)^{-1}\| \leq K, \tag{C.1}$$

where  $K$  throughout denotes a finite, positive generic constant, follow from Lemmas 2 and 3, numerators of elements of the inverse being bounded and denominators bounded away from



zero. Now  $\Psi(h) = [(I_p, 0)\Xi(h)^{-1}(I_p, 0)']^{-1}$ , where the  $2p \times 2p$  matrix  $\Xi(h)$  has  $(i, j)$ th  $p \times p$  submatrix  $\Phi(\theta 1(i=1) + h 1(i=2), \theta 1(i=1) + h 1(i=2))$ ,  $1(\cdot)$  denoting the indicator function and  $\Xi(h)^{-1}$  existing on  $\tilde{\mathcal{N}}(C_w)$  as implied below. Introduce the  $2p \times 2p$  orthogonal permutation matrix  $\Pi$  defined by  $\Pi(1_2 \otimes a) = ((1'_2 \otimes a'_1), \dots, (1'_2 \otimes a'_d))'$ , for any  $p \times 1$  vector  $a$  with  $i$ th  $p_i \times 1$  subvector  $a_i$ . Then  $\Pi \Xi(h) \Pi'$  has the form of  $T$  in Lemma 2 or 3.

In the Lemma 2 situation, where no  $\theta_{ij}$  is zero and no  $h_{ij}$  is zero on  $\tilde{\mathcal{N}}(C_w)$ , we have  $r_i = 2p_i$ ,  $r = 2p$ , and  $v_{ik} = \theta_{ik}$ ,  $k = 1, \dots, p_i$ ,  $v_{ik} = \theta_{i, k-p_i}$ ,  $k = p_i + 1, \dots, 2p_i$ . Denoting  $E_i(h) = \text{diag}\{\theta_{i1} - h_{i1}, \dots, \theta_{ip_i} - h_{ip_i}\}$  and  $e_i(h) = \text{diag}\{E_i(h), -E_i(h)\}$ ,  $e(h) = \text{diag}\{e_1(h), \dots, e_d(h)\}$ , inspection of the results of Lemma 2 indicates that we may write  $(\Pi \Xi(h) \Pi')^{-1} = e(h)^{-1} G e(h)^{-1}$ , where the  $p \times p$  matrix  $G$  is non-singular and bounded on  $\tilde{\mathcal{N}}(C_w)$ . Then

$$\Psi(h) = (I_p, 0) \Pi' e(h)^{-1} G e(h)^{-1} \Pi (I_p, 0)' = E(h) \tilde{G}^{-1} E(h),$$

where  $E(h) = \text{diag}\{E_1(h), \dots, E_d(h)\}$ ,  $\tilde{G} = (I_p, 0) \Pi' G \Pi (I_p, 0)'$ . Thus  $U(h) = \beta' D E(h) \times \tilde{G}^{-1} E(h) D \beta \geq \beta' D^2 E(h)^2 \beta / \text{tr}(\tilde{G})$ , whence the result follows by boundedness of  $\tilde{G}$  and  $\inf_{h_{ij} \in \tilde{\mathcal{N}}(C_{ijw})} (\theta_{ij} - h_{ij})^2 = C_{ijw}^2$ .

The details in the Lemma 3 setting, in which either  $\theta_{ij} = 0$  for one  $(i, j)$ , or  $h_{ij}$  can be zero on  $\tilde{\mathcal{N}}(C_w)$  for one  $(i, j)$ , are too similar to warrant inclusion. □

**Proposition 2.**

$$\sup_{h \in \tilde{\mathcal{N}}(C_w)} |V_1(h)| \leq K \sum_{i=1}^d \sum_{j=1}^{p_i} N^{1+2\zeta_{ij}-\delta^*} C_{ijw}^2. \tag{C.2}$$

**Proof.** Define  $D(h) = N \text{diag}\{n_1^{h_{11}}, \dots, n_1^{h_{1p_1}}, \dots, n_d^{h_{d1}}, \dots, n_d^{h_{dp_d}}\}$ , so  $D = D(\theta)$ , and  $\tilde{M}(g, h) = D(g)^{-1} M(g, h) D(h)^{-1}$ , also  $F_1(h) = \tilde{M}(\theta, \theta) - \tilde{M}(\theta, h) - \tilde{M}(h, \theta) + \tilde{M}(h, h)$ ,  $F_2(h) = \{\tilde{M}(\theta, h) - \tilde{M}(h, h)\} \tilde{M}(h, h)^{-1} \{\tilde{M}(h, \theta) - \tilde{M}(h, h)\}$ , so we have the identity  $D^{-1} P(h) D^{-1} = F_1(h) - F_2(h)$ . Likewise,  $\Psi(h) = \Psi_1(h) - \Psi_2(h)$ , where

$$\begin{aligned} \Psi_1(h) &= \Phi(\theta, \theta) - \Phi(\theta, h) - \Phi(h, \theta) + \Phi(h, h), \\ \Psi_2(h) &= \{\Phi(\theta, h) - \Phi(h, h)\} \Phi(h, h)^{-1} \{\Phi(h, \theta) - \Phi(h, h)\}. \end{aligned}$$

Thus  $V_1(h) = V_{11}(h) - V_{12}(h)$ , where  $V_{1i}(h) = \beta' D \{F_i(h) - \Psi_i(h)\} D \beta$ ,  $i = 1, 2$ . Now  $V_{1i}(h)$  is bounded by

$$\begin{aligned} & \left| KN \sum_{i=1}^d \sum_{j=1}^{p_i} \sum_{\ell=1}^{p_i} n_i^{\theta_{ij} + \theta_{i\ell}} \left| \frac{1}{n_i} \sum_{u_i=1}^{n_i} v_{ij}(u_i/n_i) v_{i\ell}(u_i/n_i) - \int_0^1 v_{ij}(x) v_{i\ell}(x) dx \right| \right| \\ & + KN \sum_{i=1}^d \sum_{j=1}^{p_i} \sum_{k=1}^d \sum_{\ell=1}^{p_k} n_i^{\theta_{ij}} n_k^{\theta_{k\ell}} |\tilde{v}_{ij} \tilde{v}_{k\ell} - \bar{v}_{ij} \bar{v}_{k\ell}|, \end{aligned}$$

where  $v_{ij}(x) = v(x; \theta_{ij}, h_{ij})$  with  $v$  defined just before Lemma 6,  $\bar{v}_{ij} = \int_0^1 v_{ij}(x) dx$ ,  $\tilde{v}_{ij} = n_i^{-1} \sum_{u_i=1}^{n_i} v_{ij}(u_i/n_i)$ . From Lemma 8 the first modulus is bounded by

$$K|h_{ij} - \theta_{ij}||h_{i\ell} - \theta_{i\ell}|(\log n_i)^2/n_i^{1+\min(2\Delta_i, 0)} \leq KN^{\delta^*}$$

because  $\log n_i \leq \log N$ ,  $n_i^{1+\min(2\Delta_i, 0)} = (B_i N^{b_i})^{1+\min(2\Delta_i, 0)} \geq N^{2\delta^*}/K$ . The second modulus is bounded by

$$|\tilde{v}_{ij} - \bar{v}_{ij}||\tilde{v}_{k\ell}| + |\tilde{v}_{k\ell} - \bar{v}_{k\ell}||\bar{v}_{ij}|. \tag{C.3}$$

Now

$$|\tilde{v}_{k\ell}| \leq \left\{ \frac{1}{n_k} \sum_{u_k=1}^{n_k} v_{k\ell}(u_k/n_k)^2 \right\}^{1/2}, \quad |\bar{v}_{ij}| \leq \left\{ \int_0^1 v_{ij}(x)^2 dx \right\}^{1/2},$$

so from Lemmas 6, 7 and 8, (C.3) is bounded by

$$K \left\{ \frac{(\log n_i)^2}{n_i^{1+\min(\Delta_i, 0)}} + \frac{(\log n_k)^2}{n_k^{1+\min(\Delta_k, 0)}} \right\} |h_{ij} - \theta_{ij}||h_{k\ell} - \theta_{k\ell}|,$$

and the expression in braces is bounded by  $N^{-\delta^*}$ . Thus by elementary inequalities,  $\sup_{h \in \mathcal{N}(C_w)} |V_{li}(h)|$  has the bound (C.2). Next,  $F_2(h) - \Psi_2(h)$  is

$$\begin{aligned} & \{\tilde{M}(\theta, h) - \tilde{M}(h, h) - \Phi(\theta, h) + \Phi(h, h)\} \tilde{M}(h, h)^{-1} \{\tilde{M}(h, \theta) - \tilde{M}(h, h)\} \\ & + \{\Phi(\theta, h) - \Phi(h, h)\} \{\tilde{M}(h, h)^{-1} - \Phi(h, h)^{-1}\} \{\tilde{M}(h, \theta) - \tilde{M}(h, h)\} \\ & + \{\Phi(\theta, h) - \Phi(h, h)\} \Phi(h, h)^{-1} \{\tilde{M}(h, \theta) - \tilde{M}(h, h) - \Phi(h, \theta) + \Phi(h, h)\}. \end{aligned} \tag{C.4}$$

The final factor times  $D\beta$  has a norm bounded by

$$\begin{aligned} & KN^{1/2} \sum_{i=1}^d \sum_{j=1}^{p_i} \left| \sum_{\ell=1}^{p_i} n_i^{\theta_{i\ell}} \left\{ \frac{1}{n_i} \sum_{u_i=1}^{n_i} v_{ij}(u_i/n_i)(u_i/n_i)^{h_{i\ell}} - \int_0^1 v_{ij}(x)x^{h_{i\ell}} dx \right\} \right| \\ & + KN^{1/2} \sum_{i=1}^d \sum_{j=1}^{p_i} \left| \sum_{k=1}^d \sum_{\ell=1}^{p_k} n_k^{\theta_{k\ell}} \left\{ \tilde{v}_{ij} \frac{1}{n_k} \sum_{u_k=1}^{n_k} (u_k/n_k)^{h_{k\ell}} - \bar{v}_{ij} \int_0^1 x^{h_{k\ell}} dx \right\} \right|. \end{aligned} \tag{C.5}$$

The first term in braces is

$$\frac{1}{n_i} \sum_{u_i=1}^{n_i} v \left( \frac{u_i}{n_i}; \theta_{ij} + h_{i\ell}, h_{ij} + h_{i\ell} \right) - \int_0^1 v(x; \theta_{ij} + h_{i\ell}, h_{ij} + h_{i\ell}) dx.$$

By Lemma 8, this is bounded by

$$K|\theta_{ij} + h_{i\ell} - h_{ij} - h_{i\ell}|N^{-\delta^*} \leq K|\theta_{ij} - h_{ij}|N^{-\delta^*}.$$

After rearrangement as before, and application also of Lemma 8, the second term in braces in (C.5) has the same bound. Thus (C.5) is bounded over  $h \in \mathcal{N}(C_w)$  by  $K \times \sum_{i=1}^d \sum_{j=1}^{p_i} N^{1/2+\zeta_{ij}-\delta^*} C_{ijw}$ . On the other hand, using Lemma 6,

$$\|\tilde{\beta}' D\{\Phi(\theta, h) - \Phi(h, h)\}\| \leq K \sum_{i=1}^d \sum_{j=1}^{n_i} N^{1/2+\zeta_{ij}} C_{ijw} \tag{C.6}$$

uniformly in  $h \in \mathcal{N}(C_w)$ . Using (C.1), the contribution to  $V_{2i}(h)$  has the bound in (C.2). To deal with the contributions from the other two terms in (C.4), standard manipulations indicate that it suffices to show that

$$\sup_{h \in \mathcal{N}(C_w)} \|\{\tilde{M}(h, \theta) - \tilde{M}(h, h)\} D\beta\| \leq K \sum_{i=1}^d \sum_{j=1}^{n_i} N^{1/2+\zeta_{ij}+\delta^*} C_{ijw}, \tag{C.7}$$

$$\sup_{h \in \Theta} \|\tilde{M}(h, h) - \Phi(h, h)\| \leq K N^{-2\delta^*}. \tag{C.8}$$

Since the elements of  $\tilde{M}(h, \theta) - \tilde{M}(h, h)$  are of form  $n_i^{-1} \sum_{u \in \mathbb{N}} (u_i/n_i)^{h_{ij}} v_{k\ell}(u_k/n_k)$ , for  $i = k$  or  $i \neq k$ , (C.7) follows much as before, using Lemmas 4 and 7. Finally, (C.8) is an easy consequence of Lemma 5.  $\square$

**Proposition 3.** For any  $\varepsilon > 0$

$$\sup_{h \in \mathcal{N}(C_w)} |V_2(h) - V_2(\theta)| \leq K \sum_{i=1}^d \sum_{j=1}^{n_i} N^{1/2+\zeta_{ij}+\varepsilon} C_{ijw}. \tag{C.9}$$

**Proof.** We can write  $V_2(h) - V_2(\theta)$  as

$$\begin{aligned} & 2\{m(h) - m(\theta)\}'\beta + 2m(h)'M(h, h)^{-1}\{M(h, \theta) - M(h, h)\}\beta \\ &= 2\{\tilde{m}(h) - \tilde{m}(\theta)\}'D\beta \\ &+ 2\tilde{m}(h)'\tilde{M}(h, h)^{-1}\{\tilde{M}(h, \theta) - \tilde{M}(h, h)\}D\beta, \end{aligned} \tag{C.10}$$

where  $\tilde{m}(h) = D(h)^{-1}m(h)$ . Now

$$E \left\{ \sup_{h \in \Theta} \|\tilde{m}(h)\| \right\} \leq K \tag{C.11}$$

immediately from Lemma 10. From the proof of Proposition 2, the last term of (C.10) thus has the bound (C.9). Next,

$$|\{\tilde{m}(h) - \tilde{m}(\theta)\}'D\beta| \leq K \sum_{i=1}^d \sum_{j=1}^{n_i} N^{\zeta_{ij}} \left| \sum_{u \in \mathbb{N}} v_{ij}(u_i/n_i)x_u \right|$$

and by Lemma 11 its supremum over  $\mathcal{N}(C_w)$  has the bound (C.9).  $\square$

**Proposition 4.**

$$\sup_{h \in \Theta} |V_3(h)| \leq K.$$

**Proof.** Writing  $V_3(h) = \tilde{m}(h)' \tilde{M}(h, h)^{-1} \tilde{m}(h)$ , the result follows from (C.1), (C.8) and (C.11).  $\square$

**Proposition 5.** As  $N \rightarrow \infty$ ,

$$D_+^{-1} L \{ \tilde{Q}^{(2)} - Q^{(2)}(\alpha) \} L D_+^{-1} = O_p((\log N)^{-2}).$$

**Proof.** By elementary inequalities, the result follows if

$$D_+^{-1} L \{ \tilde{Q}^{(2)}(\bar{\alpha}) - Q^{(2)}(\alpha) \} L D_+^{-1} = O_p((\log N)^{-2})$$

for any  $\bar{\alpha}$  such that  $\|\bar{\alpha} - \alpha\| \leq \|\hat{\alpha} - \alpha\|$ . A typical element of  $Q_1^{(2)}(\bar{\alpha}) - Q_1^{(2)}(\alpha)$  is

$$\begin{aligned} & 2 \sum_{u \in \mathbb{N}} (\log u_i)^{\rho_1} (\log n_i)^{1-\rho_1} (\log u_k)^{\rho_2} (\log n_k)^{1-\rho_2} \\ & \times (\bar{\beta}_{ij}^{\rho_1} u_i^{\bar{\theta}_{ij}} u_k^{\bar{\theta}_{k\ell}} \bar{\beta}_{k\ell}^{\rho_2} - \beta_{ij}^{\rho_1} u_i^{\theta_{ij}} u_k^{\theta_{k\ell}} \beta_{k\ell}^{\rho_2}) \end{aligned} \tag{C.12}$$

for  $i = k$  and  $i \neq k$ , and  $\rho_1, \rho_2 = 0, 1$ . We need to show that (C.12) =  $O_p(N^{1+\zeta_{ij}+\zeta_{k\ell}}/(\log N)^2)$ . With  $\bar{\beta}_{ij}, \bar{\beta}_{k\ell}$  replaced by  $\beta_{ij}, \beta_{k\ell}$ , it is bounded by

$$K (\log N)^2 \frac{N}{n_i} \sum_{u_i=1}^{n_i} |u_i^{\bar{\theta}_{ij}+\bar{\theta}_{i\ell}} - u_i^{\theta_{ij}+\theta_{i\ell}}|, \quad i = k, \tag{C.13}$$

or by

$$K (\log N)^2 \frac{N}{n_i n_k} \sum_{u_i=1}^{n_i} \sum_{u_k=1}^{n_k} |u_i^{\bar{\theta}_{ij}} u_k^{\bar{\theta}_{k\ell}} - u_i^{\theta_{ij}} u_k^{\theta_{k\ell}}|, \quad i \neq k. \tag{C.14}$$

Note that, for example,

$$\sum_{u_i=1}^{n_i} u_i^{\bar{\theta}_{ij}} = O_p \left( n_i^{\bar{\theta}_{ij}} \sup_{h_{ij} \in \Theta_{ij}} \left| n_i^{-h_{ij}} \sum_{u_i=1}^{n_i} u_i^{h_{ij}} \right| \right) = O_p(n_i^{\theta_{ij}}),$$

since  $n_i^{\bar{\theta}_{ij}} = O_p(n_i^{\theta_{ij}} \exp(N^{\chi-\zeta_{ij}-1/2} \log N)) = O_p(n_i^{\theta_{ij}})$ , taking  $\chi < \zeta_{11} + \frac{1}{2}$ . Then from Theorem 2 and Lemma 12, (C.13) is

$$O_p((\log N)^3 N^{1+\zeta_{ij}+\zeta_{i\ell}} (N^{\chi-\zeta_{ij}-1/2} + N^{\chi-\zeta_{i\ell}-1/2})),$$

which is  $O_p(N^{1+\zeta_{ij}+\zeta_{i\ell}}/(\log N)^2)$  as desired, while using Lemma 4, (C.14) is

$$O_p((\log N)^3 N^{1+\zeta_{ij}+\zeta_{k\ell}} (N^{\chi-\zeta_{ij}-1/2} + N^{\chi-\zeta_{k\ell}-1/2})),$$

which is  $O_p(N^{1+\zeta_{ij}+\zeta_{k\ell}}/(\log N)^2)$  as desired. Using Theorem 3, it is readily seen that (C.12) =  $O_p(N^{1+\zeta_{ij}+\zeta_{k\ell}}/(\log N)^2)$ .

The only elements of  $Q_2^{(2)}(\bar{\alpha}) - Q_2^{(2)}(\alpha)$  that are not identically zero are the diagonal elements corresponding to the three non-null submatrices in  $J(u; h, b)$ , and are of form

$$2 \sum_{u \in \mathbb{N}} \{ \bar{\beta}' f(u; \bar{\theta}) - \beta' f(u; \theta) \} u_i^{\bar{\theta}_{ij}} \{ (\log u_i)^2 \bar{\beta}_{ij} \}^\rho (\log u_i \log n_i)^{1-\rho} \tag{C.15}$$

for  $\rho = 0, 1$ . We have to show this is  $O_p(N^{1+2\zeta_{ij}}/(\log N)^2)$ . After replacing  $\bar{\beta}_{ij}$  by  $\beta_{ij}$ , it is bounded by

$$K (\log N)^2 \sum_{k=1}^d \sum_{\ell=1}^{p_k} \left| \sum_{u \in \mathbb{N}} (u_k^{\bar{\theta}_{k\ell}} - u_k^{\theta_{k\ell}}) u_i^{\bar{\theta}_{ij}} \right|.$$

Proceeding much as before, this is

$$\begin{aligned} & O_p \left( (\log N)^2 \sum_{k=1}^d \sum_{\ell=1}^{p_k} N^{1+\zeta_{k\ell}+\zeta_{ij}} N^{\chi-\zeta_{k\ell}-1/2} \right) \\ & = O_p((\log N)^2 N^{1/2+\zeta_{ij}+\chi}) = O_p(N^{1+2\zeta_{ij}}/(\log N)^2). \end{aligned}$$

Again, the same bound holds for (C.15).

Finally,  $Q_3^{(2)}(\bar{\alpha}) - Q_3^{(2)}(\alpha)$  has non-zero elements at the same locations, and they are of form

$$-2(\log n_i)^{1-\rho} \sum_{u \in \mathbb{N}} x_u \{ (\bar{\beta}_{ij} \log u_i)^\rho u_i^{\bar{\theta}_{ij}} - (\beta_{ij} \log u_i)^\rho u_i^{\theta_{ij}} \} \tag{C.16}$$

for  $\rho = 0, 1$ , which again will be shown to be  $O_p(N^{1+2\zeta_{ij}}/(\log N)^2)$ . Replacing  $\bar{\beta}_{ij}$  by  $\beta_{ij}$  gives

$$\begin{aligned} & -2\beta_{ij}(\log n_i)^{1-\rho} \left\{ n_i^{\bar{\theta}_{ij}} \sum_{u \in \mathbb{N}} x_u (\log u_i)^\rho v(u_i/n_i; \bar{\theta}_{ij}, \theta_{ij}) \right. \\ & \quad \left. + (n^{\bar{\theta}_{ij}-\theta_{ij}} - 1) \sum_{u \in \mathbb{N}} x_u (\log u_i)^\rho u_i^{\theta_{ij}} \right\} \\ & = O_p((\log N)^2 N^{1/2+\zeta_{ij}+\chi}) = O_p(N^{1+2\zeta_{ij}}/(\log N)^2), \end{aligned}$$

applying Lemmas 10 and 11, and  $n^{\bar{\theta}_{ij}-\theta_{ij}} - 1 = O_p((\log N)|\bar{\theta}_{ij} - \theta_{ij}|)$ . We can show, as before, that (C.16) has the same bound. □

**Proposition 6.** As  $N \rightarrow \infty$ ,

$$D_+L^{-1}Q^{(2)}(\alpha)^{-1}L^{-1}D_+ = \frac{1}{2}B\Upsilon^{-1}B' + \frac{1}{2}B_\Delta\Gamma B_\Delta + O_p((\log N)^{-2}).$$

**Proof.** Clearly  $Q_2^{(2)}(\alpha) \equiv 0$ . A typical non-zero element of  $Q_3^{(2)}(\alpha)$  is

$$-2 \sum_{u \in \mathbb{N}} \{(\log u_i)^2 \beta_{ij}\}^\rho (\log u_i \log n_i)^{1-\rho} u_i^{\theta_{ij}} x_u$$

for  $\rho = 0, 1$ , and from Lemma 10 this is

$$O_p((\log N)^2 N^{1/2+\zeta_{ij}}) = O_p(N^{1+2\zeta_{ij}}/(\log N)^2)$$

as desired. From Lemmas 13 and 14,

$$D_+^{-1}Q_1^{(2)}D_+^{-1} = 2 \operatorname{diag}\{\beta_\Delta, I_p\}(A + O((\log N)^{-2})) \operatorname{diag}\{\beta_\Delta, I_p\},$$

where  $A$  has  $p \times p$  submatrices  $A_{ij}$  such that  $A_{11} = L\Phi L - L\Phi_+ - \Phi'_+L + \Phi_{++}$ ,  $A_{12} = A'_{21} = L\Phi L - \Phi'_+L$ ,  $A_{22} = L\Phi L$ . Thus  $A^{-1}$  has  $p \times p$  submatrices  $A^{ij}$  such that  $A^{11} = \Upsilon^{-1}$ ,  $A^{12} = A^{21'} = \Lambda'L^{-1} - \Upsilon^{-1}$ ,  $A^{22} = L^{-1}\Phi^{-1}(L^{-1} + (\Phi L - \Phi_+)\Upsilon^{-1}(L\Phi - \Phi_+)\Phi^{-1}L^{-1})$ . It follows that  $A^{-1} = (I_p, -I_p)\Upsilon^{-1}(I_p, -I_p)' + \Gamma$ . The proof is straightforwardly concluded.  $\square$

**Proposition 7.** As  $N \rightarrow \infty$ ,

$$N^{-1/2} \sum_{u \in \mathbb{N}} [\Upsilon^{-1}\{L(u) - L\} + \Lambda'] D^{-1} f(u; \theta) x_u \rightarrow_d \mathfrak{N}_p(0, 2\pi F(0)\Upsilon^{-1}).$$

**Proof.** Write  $x_u = x_{u1} + x_{u2}$  for  $x_{u1} = \sum_{v \in E_M} \xi_v \varepsilon_{u-v}$ ,  $x_{u2} = \sum_{v \in \bar{E}_M} \xi_v \varepsilon_{u-v}$ , with  $E_M = \{u: |u_i| \leq M, i = 1, \dots, d\}$ ,  $\bar{E}_M = \mathbb{Z}^d \setminus E_M$ , for positive integer  $M$ . For  $\eta > 0$ , choose  $M$  such that  $\sum_{v \in \bar{E}_M} |\xi_v| < \eta$ . Writing

$$g_u = [\Upsilon^{-1}\{L(u) - L\} + \Lambda'] D^{-1} f(u; \theta),$$

we have

$$\begin{aligned} E \left\| N^{-1/2} \sum_{u \in \mathbb{N}} g_u x_{u2} \right\|^2 &= N^{-1} \sum_{v \in \bar{E}_M} \sum_{w \in \bar{E}_M} \xi_v \xi_w \sum_{u, u-v+w \in \mathbb{N}} g'_u g_{u-v+w} \\ &\leq \left( \sum_{v \in \bar{E}_M} |\xi_v| \right)^2 N^{-1} \sum_{u \in \mathbb{N}} \|g_u\|^2, \end{aligned} \tag{C.17}$$

and

$$\frac{1}{N} \sum_{u \in \mathbb{N}} \|g_u\|^2 \leq \frac{K}{N} \sum_{u \in \mathbb{N}} \sum_{i=1}^d \sum_{j=1}^{p_i} [1 + \{\log(u_i/n_i)\}^2] (u_i/n_i)^{2\theta_{ij}} \leq K,$$

by Lemmas 13 and 14. Then (C.17)  $\leq K\eta^2$ . Next write  $N^{-1/2} \sum_{u \in \mathbb{N}} g_u x_{1u} = N^{-1/2} \sum_{w \in E'} \varepsilon_w \times \sum_{u \in E''} \xi_{u-w} g_u$ , where

$$E' = \{w: 1 - M \leq w_i \leq n_i + M, i = 1, \dots, d\},$$

$$E'' = \{u: \max(1, w_i - M) \leq u_i \leq \min(n_i, w_i + M), i = 1, \dots, d\}.$$

We may then apply a CLT, with  $N$  and thus  $N^* = \prod_{i=1}^d (n_i + 2M)$  increasing, for independent random variables whose squares are uniformly integrable. It remains to check two aspects. The first is the Lindeberg condition,

$$\frac{1}{N} \max_{w \in E'} \left\| \sum_{u \in E''} \xi_{u-w} g_u \right\|^2 \rightarrow 0, \quad \text{as } N \rightarrow \infty.$$

The left side is bounded by

$$\frac{K}{N} \max_{u \in \mathbb{N}} \|g_u\|^2 \leq \frac{K}{N} \sum_{i=1}^d \sum_{j=1}^{p_i} \max_{u_i} [\{\log(u_i/n_i)\}^2 + 1] (u_i/n_i)^{2\theta_{ij}} \rightarrow 0,$$

since, for some  $\eta > 0$ ,

$$(u_i/n_i)^{2\theta_{ij}} \leq 1(\theta_{ij} \geq 0) + n_i^{2|\theta_{ij}|} 1(\theta_{ij} < 0) \leq N^{1-\eta}, \quad |\log(u_i/n_i)| \leq K \log N.$$

The second aspect is the covariance structure:

$$E \left\{ N^{-1/2} \sum_{u \in \mathbb{N}} g_u x_{1u} \right\} \left\{ N^{-1/2} \sum_{u \in \mathbb{N}} g_u x_{1u} \right\}' = N^{-1} \sum_{v, w \in \bar{E}_M} \xi_v \xi_w \sum' g_u g_{u+w-v}, \quad (\text{C.18})$$

where the primed sum is over all  $u$  such that  $u, u + w - v \in \mathbb{N}$ . Since  $M$  is fixed and  $\|g_u\| \leq KN^{1-\eta}$ , for some  $\eta > 0$ , (C.18) differs by  $o(1)$ , as  $N \rightarrow \infty$ , from  $N^{-1} \sum_{v, w \in E_M} \xi_v \xi_w \times \sum_{u \in \mathbb{N}} g_u g_{u+w-v}$ . Using Lemma 16, this differs by  $o(1)$  from  $N^{-1} (\sum_{v \in E_M} \xi_v)^2 \sum_{u \in \mathbb{N}} g_u g'_u$ , which, by Lemmas 13 and 14 and straightforward calculation and elimination, equals

$$\begin{aligned} & \left( \sum_{v \in E_M} \xi_v \right)^2 \{ \Upsilon^{-1} \Phi_{++} \Upsilon^{-1} - \Upsilon^{-1} \Phi'_+ \Lambda - \Lambda' \Phi_+ \Upsilon^{-1} + \Lambda' \Phi \Lambda + O(1/\log N) \} \\ & = \left( \sum_{v \in E_M} \xi_v \right)^2 \{ \Upsilon^{-1} + O(1/\log N) \} \rightarrow \left( \sum_{v \in E_M} \xi_v \right)^2 \Upsilon^{-1} \end{aligned}$$

as  $N \rightarrow \infty$ , and the last displayed expression differs by  $O(\eta)$  from  $(\sum_{v \in \mathbb{Z}^d} \xi_v)^2 \Upsilon^{-1} = 2\pi F(0) \Upsilon^{-1}$ . □

### Appendix D: Technical lemmas

**Lemma 1.** Let  $T$  be an  $r \times r$  matrix, with  $(i, j)$ th  $r_i \times r_j$  block  $T_{ij}$ ,  $i, j = 1, \dots, d$ , where  $\sum_{i=1}^d r_i = r$ . Let  $t_i$  be a column vector such that  $T_{ij} = t_i t'_j$ ,  $i \neq j$ , and  $T_{ii} - t_i t'_i$  is positive definite,  $i, j = 1, \dots, d$ . Then  $T$  is non-singular, with  $(i, j)$ th  $r_i \times r_j$  submatrix

$$T_{ii}^{-1} + \frac{T_{ii}^{-1} t_i t'_i T_{ii}^{-1}}{1 - \tau_i} \sum_{\substack{s=1 \\ s \neq i}}^d \frac{\tau_s}{1 - \tau_s} / (1 + \sigma), \quad i = j, \tag{D.1}$$

and

$$\frac{-T_{ii}^{-1} t_i t'_j T_{jj}^{-1}}{(1 - \tau_i)(1 - \tau_j)} / (1 + \sigma), \quad i \neq j, \tag{D.2}$$

where  $\tau_i = t'_i T_{ii}^{-1} t_i$ ,  $\sigma = \sum_{i=1}^d \tau_i / (1 - \tau_i)$ .

**Proof.** Let  $\tilde{T}$  be the  $r \times r$  matrix with diagonal blocks  $\tilde{T}_i = T_{ii} - \tau_i t'_i$ , and zeros elsewhere, so  $T = \tilde{T} + t t'$ , where  $t = (t'_1, \dots, t'_d)'$ . Now because  $\det\{T_{ii} - t_i t'_i\} = \det\{T_{ii}\}(1 - \tau_i)$ , it follows that  $\tau_i < 1$ , and

$$\tilde{T}_i^{-1} = T_{ii}^{-1} \{I_{r_i} + (1 - \tau_i)^{-1} t_i t'_i T_{ii}^{-1}\}, \quad i = 1, \dots, d. \tag{D.3}$$

Then  $\tilde{T}^{-1}$  is the  $r \times r$  matrix with diagonal blocks  $\tilde{T}_i^{-1}$ . Thus

$$T^{-1} = \tilde{T}^{-1} \{I_r - (1 + t' \tilde{T}^{-1} t)^{-1} t t' \tilde{T}^{-1}\}. \tag{D.4}$$

Now  $t'_i \tilde{T}_i^{-1} = (1 + \tau_i (1 - \tau_i)^{-1}) t'_i T_{ii}^{-1} = (1 - \tau_i)^{-1} t'_i T_{ii}^{-1}$ ,  $i = 1, \dots, d$ , and so  $t' \tilde{T}^{-1} = \{(1 - \tau_1)^{-1} t'_1 T_{11}^{-1}, \dots, (1 - \tau_d)^{-1} t'_d T_{dd}^{-1}\}$ , and thus  $t' \tilde{T}^{-1} t = \sigma$ . From (D.4), the  $(i, j)$ th  $r_i \times r_j$  submatrix of  $T^{-1}$ , for  $i \neq j$ , is  $-\tilde{T}_i^{-1} t_i t'_j \tilde{T}_j^{-1} / (1 + t' \tilde{T}^{-1} t)$ , which equals (D.2) on substituting (D.3), while for  $i = j$  it is

$$T_{ii}^{-1} + \frac{T_{ii}^{-1} t_i t'_i T_{ii}^{-1}}{1 - \tau_i} - \frac{T_{ii}^{-1} t_i t'_i T_{ii}^{-1}}{(1 - \tau_i)^2} / \{1 + \sigma\},$$

which equals (D.1) after straightforward algebra. □

**Lemma 2.** Let  $T_{ii}$  be a Cauchy matrix, having  $(j, k)$ th element  $(1 + v_{ij} + v_{ik})^{-1}$ , and let the  $j$ th element of  $t_i$  be  $(1 + v_{ij})^{-1}$ , where  $v_{ij} \in (-\frac{1}{2}, \infty) \setminus \{0\}$ , all  $i, j$  and  $v_{ij} \neq v_{ik}$ , for  $j \neq k$ . Then  $T$  as defined in Lemma 1 is non-singular, and its inverse  $T^{-1}$  has  $(i, j)$ th  $r_i \times r_j$  block with  $(k, \ell)$ th



element

$$(1 + 2v_{ik})(1 + 2v_{i\ell}) \prod_{\substack{m=1 \\ m \neq k}}^{r_i} \frac{1 + v_{ik} + v_{im}}{v_{ik} - v_{im}} \prod_{\substack{m=1 \\ m \neq \ell}}^{r_i} \frac{1 + v_{i\ell} + v_{im}}{v_{i\ell} - v_{im}}$$

$$\times \left[ \frac{1}{1 + v_{ik} + v_{i\ell}} - \left\{ \frac{1}{v_{ik}(1 + v_{ik})v_{i\ell}(1 + v_{i\ell})} \prod_{m=1}^{r_i} \left( \frac{1 + v_{im}}{v_{im}} \right)^2 \right\} \right] \tag{D.5}$$

$$\left/ \left\{ \sum_{s=1}^d \prod_{m=1}^{r_i} \left( \frac{1 + v_{sm}}{v_{sm}} \right)^2 + 1 - d \right\} \right., \quad i = j,$$

$$\frac{(1 + 2v_{ik})(1 + 2v_{j\ell})}{v_{ik}^2 v_{j\ell}^2} \prod_{\substack{m=1 \\ m \neq k}}^{r_i} \frac{(1 + v_{ik} + v_{im})(1 + v_{im})}{(v_{im} - v_{ik})v_{im}}$$

$$\times \prod_{\substack{m=1 \\ m \neq \ell}}^{r_i} \frac{(1 + v_{j\ell} + v_{jm})(1 + v_{jm})}{(v_{jm} - v_{j\ell})v_{jm}} \tag{D.6}$$

$$\left/ \left\{ \sum_{s=1}^d \prod_{m=1}^{r_i} \left( \frac{1 + v_{sm}}{v_{sm}} \right)^2 + 1 - d \right\} \right., \quad i \neq j.$$

**Proof.** From page 31 of Knuth [10],  $T_{ii}^{-1}$  has  $(k, \ell)$ th element

$$\prod_{m=1}^{r_i} (1 + v_{ik} + v_{im})(1 + v_{i\ell} + v_{im})$$

$$\left/ \left\{ (1 + v_{ik} + v_{i\ell}) \prod_{\substack{m=1 \\ m \neq \ell}}^{r_i} (v_{ik} - v_{im}) \prod_{\substack{m=1 \\ m \neq k}}^{r_i} (v_{i\ell} - v_{im}) \right\} \right.$$

For each  $i$  define the  $(r_i + 1) \times (r_i + 1)$  non-singular Cauchy matrix  $T_{ii}^+$  whose first  $r_i$  rows are  $(T_{ii}, t_i)$  and whose last row is  $(t_i', 1)$ . Thus, again from page 31 of Knuth [10], the  $(r_i + 1, r_i + 1)$ th element of its inverse is  $(1 - \tau_i)^{-1} = \prod_{\ell=1}^{r_i} (1 + v_{i\ell}^{-1})^2$ . Thus

$$1 + \sigma = \prod_{\ell=1}^{r_i} (1 + v_{i\ell}^{-1})^2 + 1 - d. \tag{D.7}$$

Also, the leading  $r_i \times 1$  subvector of the  $(r_i + 1)$ th column of  $T_{ii}^{+-1}$  is  $(1 - \tau_i)^{-1} T_{ii}^{-1} t_i$ , which has  $k$ th element

$$\begin{aligned} & (1 + v_{ik}) \prod_{m=1}^{r_i} (1 + v_{ik} + v_{im})(1 + v_{im}) \bigg/ \left\{ (1 + v_{ik}) v_{ik} \prod_{\substack{m=1 \\ m \neq i}}^{r_i} (v_{ik} - v_{im}) \prod_{m=1}^{r_i} (-v_{im}) \right\} \\ &= \frac{1 + 2v_{ik}}{v_{ik}^2} \prod_{\substack{m=1 \\ m \neq k}}^{r_i} \frac{(1 + v_{ik} + v_{im})(1 + v_{im})}{(v_{im} - v_{ik})v_{im}}. \end{aligned}$$

The proof is completed by substitution and rearrangement. □

**Lemma 3.** Let  $T^+$  be the  $(r + 1) \times (r + 1)$  matrix whose first  $r$  rows are  $(T, t)$  and whose last row is  $(t', 1)$ , with  $T$  and  $t$  defined as in Lemmas 1 and 2. Then

$$T^{+-1} = \begin{bmatrix} T^{-1}(I_r + t t' T^{-1}(1 - t' T^{-1} t)^{-1}) & -T^{-1} t (1 - t' T^{-1} t)^{-1} \\ -(1 - t' T^{-1} t)^{-1} t' T^{-1} & (1 - t' T^{-1} t)^{-1} \end{bmatrix},$$

where  $(1 - t' T^{-1} t)^{-1} = 1 + \sigma$ .

**Proof.** From (D.1) and (D.2)

$$t' T^{-1} t = \sum_{i=1}^d \left\{ \tau_i + \frac{\tau_i^2}{(1 + \sigma)(1 - \tau_i)} \left( \sigma - \frac{\tau_i}{1 - \tau_i} \right) \right\} - \frac{\sigma^2}{1 + \sigma} + \frac{1}{(1 + \sigma)} \sum_{i=1}^d \frac{\tau_i^2}{(1 - \tau_i)^2} = \frac{\sigma}{1 + \sigma}$$

after routine algebra. Thus  $1 - t' T^{-1} t = 1/(1 + \sigma)$ , and the proof is readily completed. □

**Lemma 4.** For  $a > -1$

$$\sup_{a \geq a} \left| \frac{1}{J} \sum_{j=1}^J \left( \frac{j}{J} \right)^a \right| \leq K.$$

**Proof.** The expression within the modulus is bounded by

$$\int_0^1 x^a dx + 1 = \frac{1}{a + 1} + 1 \leq \frac{1}{a + 1} + 1 \leq K. \quad \square$$

**Lemma 5.** For  $a > -1$ ,

$$\sup_{a \geq a} \left| \frac{1}{J} \sum_{j=1}^J \left( \frac{j}{J} \right)^a - \frac{1}{1 + a} \right| \leq \frac{K}{J^{1 + \min(a, 0)}}.$$

**Proof.** The expression within the modulus is

$$\frac{1}{J} \sum_{j=2}^{J-1} \int_{(j-1)/J}^{j/J} \left\{ \left( \frac{j}{J} \right)^a - x^a \right\} dx + \frac{1}{J^{a+1}} - \int_0^{1/J} x^a dx + \frac{1}{J} - \int_{1-1/J}^1 x^a dx. \tag{D.8}$$

Using the mean value theorem, the first term in (D.8) is bounded by

$$\frac{2a}{J} \sum_{j=1}^J \left( \frac{j}{J} \right)^{a-1} 1(a > 0) + \frac{|a|}{J^{a+1}} \sum_{j=1}^J j^{a-1} 1(a < 0) \leq \frac{2}{J} + \frac{K}{J^{a+1}}.$$

The last two integrals in (D.8) are bounded by

$$\frac{(1/J)^{a+1}}{a+1} + \frac{1}{a+1} \left\{ 1 - \left( 1 - \frac{1}{J} \right)^a \right\} \leq \frac{K}{J^{a+1}} + \frac{2}{J}. \quad \square$$

Define, for  $s \in [0, 1]$ ,  $v(s; a, b) = s^a - s^b$ .

**Lemma 6.** For  $\underline{a} > -\frac{1}{2}$

$$\sup_{a, b \geq \underline{a}} (a - b)^{-2} \int_0^1 v(x; a, b)^2 dx \leq K.$$

**Proof.** The integral is

$$\frac{1}{2a+1} - \frac{2}{a+b+1} + \frac{1}{2b+1} = \frac{2(a-b)^2}{(2a+1)(a+b+1)(2b+1)} \leq K(a-b)^2. \quad \square$$

**Lemma 7.** For  $\underline{a} > -\frac{1}{2}$ ,

$$\sup_{a, b \in [\underline{a}, \bar{a}]} \left\{ (a-b)^{-2} \sum_{j=1}^J v\left( \frac{j}{J}; a, b \right)^2 \right\} \leq K J (\log J)^2. \tag{D.9}$$

**Proof.** By the mean value theorem,

$$|v(s; a, b)| \leq s^c |\log s| |a - b|, \quad s \in (0, 1], \tag{D.10}$$

where  $|a - c| \leq |a - b|$ . Also, for such  $c$ ,

$$s^c \leq s^{\underline{a}}, \quad s \in (0, 1]. \tag{D.11}$$

Thus the quantity in braces in (D.9) is bounded by

$$K (\log J)^2 \sum_{j=1}^J \left( \frac{j}{J} \right)^{2\underline{a}} \leq K J (\log J)^2, \tag{D.12}$$

because  $\underline{a} > -\frac{1}{2}$ . □

**Lemma 8.** For  $-1 < \underline{a} < \bar{a} < \infty$ ,

$$\sup_{a,b \in [\underline{a}, \bar{a}]} |a - b|^{-1} \left| \frac{1}{J} \sum_{j=1}^J v\left(\frac{j}{J}; a, b\right) - \int_0^1 v(x; a, b) dx \right| \leq \frac{K(\log J)^2}{J^{1+\min(\underline{a}, 0)}}.$$

**Proof.** The expression within the modulus is

$$\sum_{j=2}^J \int_{(j-1)/J}^{j/J} \left\{ v\left(\frac{j}{J}; a, b\right) - v(x; a, b) \right\} dx + \frac{1}{J} v\left(\frac{1}{J}; a, b\right) - \int_0^{1/J} v(x; a, b) dx. \quad (\text{D.13})$$

From (D.10) and (D.11), the last integral is bounded by

$$K \int_0^{1/J} x^{\underline{a}} |\log x| dx |a - b| \leq K(\log J) J^{-\underline{a}-1} |a - b|,$$

and the same bound results for the penultimate term of (D.13). By the mean value theorem  $|v(s; a, b) - v(s - r; a, b)|$  is bounded by

$$|s^c \log s - (s - r)^c \log(s - r)| |a - b|, \quad 0 \leq r \leq 1/J, s \geq 2/J, \quad (\text{D.14})$$

where  $|a - c| \leq |a - b|$ , and the first modulus is bounded by

$$\begin{aligned} & \left\{ |s^c - (s - r)^c| |\log s| + |(s - r)^c \{\log s - \log(s - r)\}| \right. \\ & \leq s^c |\log s| \left\{ \left| 1 - \left(1 - \frac{r}{s}\right)^c \right| + \left(1 - \frac{r}{s}\right)^c \left| \log\left(1 - \frac{r}{s}\right) \right| \right\} \\ & \leq K \frac{s^{c-1}}{J} |\log s|. \end{aligned} \quad (\text{D.15})$$

Thus the first term of (D.13) is bounded by  $|a - b|$  times

$$\frac{K \log J}{J^2} \sum_{j=1}^J \left(\frac{j}{J}\right)^{\underline{a}-1} = O\left(\frac{\log J}{J^{\underline{a}+1}} 1(\underline{a} \leq 0) + \frac{(\log J)^2}{J} 1(\underline{a} = 0) + \frac{\log J}{J} 1(\underline{a} > 0)\right). \quad (\text{D.16}) \quad \square$$

**Lemma 9.** For  $\underline{a} > -\frac{1}{2}$ ,

$$\begin{aligned} & \sup_{\substack{a_j, b_i \in [\underline{a}, \bar{a}] \\ i=1,2}} \left\{ \prod_{i=1}^2 |a_i - b_i| \right\}^{-1} \left| \frac{1}{J} \sum_{j=1}^J \prod_{i=1}^2 v\left(\frac{j}{J}; a_i, b_i\right) - \int_0^1 \prod_{i=1}^2 v(x; a_i, b_i) dx \right| \\ & \leq \frac{K(\log J)^3}{J^{1+\min(2\underline{a}, 0)}}. \end{aligned}$$

**Proof.** The expression within the second modulus is

$$\begin{aligned} & \sum_{j=2}^J \int_{(j-1)/J}^{j/J} \left\{ \prod_{i=1}^2 v\left(\frac{j}{J}; a_i, b_i\right) - \prod_{i=1}^2 v(x; a_i, b_i) \right\} dx \\ & + J^{-1} \prod_{i=1}^2 v\left(\frac{1}{J}; a_i, b_i\right) - \int_0^{1/J} \prod_{i=1}^2 v(x; a_i, b_i) dx. \end{aligned} \tag{D.17}$$

Similarly to the proof of Lemma 7, the last term is bounded by

$$K \int_0^{1/J} x^{2a} (\log x)^2 dx \prod_{i=1}^2 |a_i - b_i| \leq \frac{K (\log J)^2}{J^{2a+1}} \prod_{i=1}^2 |a_i - b_i|.$$

The expression in braces in (D.17) can be written

$$\begin{aligned} & \left\{ v\left(\frac{j}{J}; a_1, b_1\right) - v(x; a_1, b_1) \right\} v\left(\frac{j}{J}; a_2, b_2\right) \\ & + v(x; a_1, b_1) \left\{ v\left(\frac{j}{J}; a_2, b_2\right) - v(x; a_2, b_2) \right\}. \end{aligned}$$

Both terms are treated similarly; we consider only the first. From the bounds (D.14), (D.15) its first factor is bounded by  $(K/J)(j/J)^{a-1}(\log J)|a_1 - b_1|$ , and its second one by  $K(j/J)^a(\log J)|a_2 - b_2|$ . Thus its contribution is

$$O\left( (\log J)^2 J^{1+2a} \sum_{j=1}^J j^{2a-1} \right),$$

whence the result follows by an analogous calculation to (D.16). □

**Lemma 10.** For  $i = 1, \dots, d$  and  $-\frac{1}{2} < a < \bar{a} < \infty$ , and all  $q \geq 0$

$$E \left\{ \sup_{a \in [a, \bar{a}]} \left| N^{-1/2} \sum_{u \in \mathbb{N}} \left(\frac{u_i}{n_i}\right)^a (\log u_i)^q x_u \right| \right\} \leq K (\log N)^q. \tag{D.18}$$

**Proof.** By summation by parts

$$\begin{aligned} & \sum_{u_i=1}^{n_i} \left(\frac{u_i}{n_i}\right)^a (\log u_i)^q x_u \\ & = \sum_{u_i=1}^{n_i-1} \left\{ \left(\frac{u_i}{n_i}\right)^a - \left(\frac{u_i+1}{n_i}\right)^a \right\} \sum_{\ell=1}^{u_i} (\log \ell)^q x_{u_1, \dots, \ell, \dots, u_d} + \sum_{u_i=1}^{n_i} (\log u_i)^q x_u, \end{aligned}$$

where  $x_{u_1, \dots, \ell, \dots, u_d}$  is  $x_u$  with  $u_i$  replaced by  $\ell$ . Thus the expression in the modulus in (D.18) is

$$N^{-1/2} \sum_{u_i=1}^{n_i-1} \left(\frac{u_i}{n_i}\right)^a \{1 - (1 + u_i^{-1})^a\} H_i(u_i) + n^{-1/2} H_i(n_i), \tag{D.19}$$

where

$$H_i(s) = \sum_{\substack{u_k=1 \\ k=1, \dots, d \\ k \neq i}}^{n_k} \sum_{\ell=1}^s x_{u_1, \dots, \ell, \dots, u_d} (\log \ell)^q.$$

The factor in braces in (D.19) is bounded by  $|a|/u_i \leq K/u_i$ , whereas  $(u_i/n_i)^a \leq (u_i/n_i)^{\underline{a}}$ . Thus the left side of (D.18) is bounded by

$$\begin{aligned} &KN^{-1/2} \sum_{u_i=1}^{n_i-1} \left(\frac{u_i}{n_i}\right)^{\underline{a}} \frac{1}{u_i} E|H_i(u_i)| + n^{-1/2} E|H_i(n_i)| \\ &\leq K(\log n_i)^q n_i^{-1/2-\underline{a}} \sum_{u_i=1}^{n_i} u_i^{\underline{a}-1/2} + K(\log N)^q \leq K(\log N)^q, \end{aligned}$$

since  $\underline{a} > -\frac{1}{2}$  and

$$\begin{aligned} E H_i(s)^2 &= \sum_{\substack{u_k=1 \\ k=1, \dots, d \\ k \neq i}}^{n_k} \sum_{\ell=1}^s \sum_{\substack{v_k=1 \\ k=1, \dots, d \\ k \neq i}}^{n_k} \sum_{m=1}^s \gamma_{u_1-v_1, \dots, \ell-m, \dots, u_d-v_d} (\log \ell)^q (\log m)^q \\ &\leq \frac{KNs}{n_i} (\log s)^{2q} \sum_{u \in \mathbb{Z}^d} |\gamma_u| \leq \frac{KNs(\log s)^{2q}}{n_i}. \end{aligned} \quad \square$$

**Lemma 11.** For  $\underline{a} > -\frac{1}{2}$ ,

$$E \left\{ \sup_{a, b \in [a, \bar{a}]} |a - b|^{-1} \left| \sum_{u \in \mathbb{N}} v(u_i/n_i; a, b) x_u \right| \right\} \leq KN^{1/2} \log N. \tag{D.20}$$

**Proof.** By summation by parts,

$$\sum_{u_i=1}^{n_i} v(u_i/n_i; a, b) x_u = \sum_{u_i=1}^{n_i-1} \{v(u_i/n_i; a, b) - v((u_i + 1)/n_i; a, b)\} \sum_{\ell=1}^{u_i} x_{u_1, \dots, \ell, \dots, u_d}.$$

From (D.14) and (D.15), the expression in braces is bounded by

$$K \left( \frac{\log n_i}{n_i} \right) \left( \frac{u_i + 1}{n_i} \right)^{a-1} \leq \frac{K \log N}{u_i} \left( \frac{u_i}{n_i} \right)^a.$$

Thus the left side of (D.20) is bounded by  $K \log N \sum_{u_i=1}^{n_i-1} (u_i/n_i)^a u_i^{-1} E|H_i(u_i)|$ , which, from the proof of Lemma 10 (with  $q = 0$ ), has the desired bound.  $\square$

**Lemma 12.** *Let  $a > -\frac{1}{2}$  be a scalar and  $\tilde{a} = \tilde{a}_J$  be a sequence such that  $\tilde{a} - a = O_p(J^{-\eta})$  as  $J \rightarrow \infty$ , for some  $\eta > 0$ . Then for all  $q \geq 0$ ,*

$$J^{-1-a} \sum_{j=1}^J (\log j)^q |j^{\tilde{a}} - j^a| = O_p(J^{-\eta}), \quad \text{as } J \rightarrow \infty.$$

**Proof.** The left side is bounded by

$$\begin{aligned} \sum_{j=1}^J (\log j)^q \left( \frac{j}{J} \right)^a |j^{\tilde{a}-a} - 1| &\leq \frac{1}{J} \sum_{j=1}^J (\log j)^{q+1} \left( \frac{j}{J} \right)^a |\tilde{a} - a| \\ &\leq K J^{\eta/2} O_p(J^{-\eta}) \frac{1}{J} \sum_{j=1}^J \left( \frac{j}{J} \right)^a = O_p(J^{-\eta/2}). \end{aligned} \quad \square$$

**Lemma 13.** *For  $a > -\frac{1}{2}$ , there is an  $\eta > 0$  such that for all sufficiently large  $J$ ,*

$$\left| \frac{1}{J} \sum_{j=1}^J (\log j) \left( \frac{j}{J} \right)^a - \frac{\log J}{a+1} + \frac{1}{(a+1)^2} \right| \leq K J^{-\eta}. \tag{D.21}$$

**Proof.** The left side is bounded by

$$\frac{1}{J^{a+1}} \sum_{j=2}^J \int_{j-1}^j |(\log x)x^a - (\log j)j^a| dx + \left| \frac{1}{J^{a+1}} \int_0^1 (\log x)x^a dx \right|. \tag{D.22}$$

The first modulus is bounded by

$$\begin{aligned} |\log x||x^a - j^a| + |\log(x/j)|j^a &\leq K(\log j)\{(j-1)^{a-1} + j^{a-1}\} + j^{a-1} \\ &\leq K(\log j)j^{a-1} \end{aligned}$$

for  $x \in [j-1, j]$ ,  $j \geq 2$ . Thus the first term of (D.23) is  $O((\log J)J^{-a-1})$  for  $a < 0$ ,  $O((\log J)^2 J^{-1})$  for  $a = 0$ , and  $O((\log J)J^{-1})$  for  $a > 0$ . The last integral is  $O(J^{a-1})$ . Since  $a > -1$  there is an  $\eta > 0$  to satisfy (D.21).  $\square$

**Lemma 14.** For any  $a > -\frac{1}{2}$ , there is an  $\eta > 0$  such that for all sufficiently large  $J$ ,

$$\left| \frac{1}{J} \sum_{j=1}^J (\log j)^2 \left(\frac{j}{J}\right)^a - \frac{(\log J)^2}{a+1} + \frac{2 \log J}{(a+1)^2} - \frac{2}{(a+1)^3} \right| \leq J^{-\eta}.$$

**Proof.** The left side is bounded by

$$\frac{1}{J^{a+1}} \sum_{j=2}^J \int_{j-1}^j |(\log x)^2 x^a - (\log j)^2 j^a| dx + \left| \frac{1}{J^{a-1}} \int_0^1 (\log x)^2 x^a dx \right|.$$

The first integrand is bounded by

$$(\log x)^2 |x^a - j^a| + |\log(x/j)| |\log(xj)| j^a \leq K (\log j)^2 j^{a-1}$$

as in the proof of Lemma 13; the proof is completed in similar fashion. □

**Lemma 15.** For any  $a > -\frac{1}{2}$  and all sufficiently large  $N$ ,

$$E \left\{ N^{-1/2} \sum_{u \in \mathbb{N}} \log(u_i/n_i) (u_i/n_i)^a x_u \right\}^2 \leq K.$$

**Proof.** The left side is

$$\begin{aligned} N^{-1} \sum_{u,v \in \mathbb{N}} \sum_{i \in \mathbb{N}} \left(\frac{u_i}{n_i}\right)^a \left(\frac{v_i}{n_i}\right)^a \log\left(\frac{u_i}{n_i}\right) \log\left(\frac{v_i}{n_i}\right) \gamma_{u-v} \\ \leq N^{-1} \sum_{u \in \mathbb{N}} \sum_{i \in \mathbb{N}} \left(\frac{u_i}{n_i}\right)^{2a} \log^2\left(\frac{u_i}{n_i}\right) \sum_{v \in \mathbb{Z}^d} |\gamma_{u-v}| \leq K, \end{aligned}$$

by Assumption 3 and straightforward application of Lemmas 13 and 14. □

**Lemma 16.** For  $a_1, a_2 > \frac{1}{2}$ ,  $q_1, q_2 \geq 0$ , and any finite positive or negative integer  $M$ , there is an  $\eta > 0$  such that for all sufficiently large  $J$ ,

$$\begin{aligned} \frac{1}{J} \sum_{j=1}^J \left\{ \log^{q_1} \left(\frac{j}{J}\right) \right\} \left(\frac{j}{J}\right)^{a_1} \left\{ \log^{q_2} \left(\frac{j+M}{J}\right) \left(\frac{j+M}{J}\right)^{a_2} - \log^{q_2} \left(\frac{j}{J}\right) \left(\frac{j}{J}\right)^{a_2} \right\} \\ \leq |M| J^{-\eta}. \end{aligned} \tag{D.23}$$

**Proof.** We have

$$\left| \left(\frac{j+M}{J}\right)^{a_2} - \left(\frac{j}{J}\right)^{a_2} \right| \leq \frac{M}{j} \left(\frac{j}{J}\right)^{a_2},$$



$$\left| \log^{q_2} \left( \frac{j+M}{J} \right) - \log^{q_2} \left( \frac{j}{J} \right) \right| \leq \frac{M}{j}.$$

By elementary inequalities the left side of (D.23) is bounded by

$$\begin{aligned} & \frac{KM(\log J)^{q_1+q_2}}{J^{a_1+a_2+1}} \sum_{j=1}^J j^{a_1+a_2-1} \\ & \leq K|M|(\log J)^{q_1+q_2} \left\{ \frac{1(a_1+a_2 < 0)}{J^{a_1+a_2+1}} + \frac{1(a_1+a_2 = 0)}{J} \log J + \frac{1(a_1+a_2 > 0)}{J} \right\}, \end{aligned}$$

which is  $O(|M|J^{-\eta})$ . □

## Acknowledgements

I thank the Editor, Associate Editor and two referees for helpful comments. I am also grateful to Francesca Rossi for carrying out the Monte Carlo simulations. This research was supported by ESRC Grant RES-062-23-0036, Spanish Plan Nacional de I+D+I Grant SEJ2007-62908/ECON, and a Catedra de Excelencia at Universidad Carlos III, Madrid.

## References

- [1] Anderson, T.W. (1971). *The Statistical Analysis of Time Series*. New York: Wiley. [MR0283939](#)
- [2] Cressie, N.A.C. (1993). *Statistics for Spatial Data*. New York: Wiley. Revised reprint of the 1991 edition. [MR1239641](#)
- [3] Gao, J., Lu, Z. and Tjøstheim, D. (2006). Estimation in semiparametric spatial regression. *Ann. Statist.* **34** 1395–1435. [MR2278362](#)
- [4] Giraitis, L., Hidalgo, J. and Robinson, P.M. (2001). Gaussian estimation of parametric spectral density with unknown pole. *Ann. Statist.* **29** 987–1023. [MR1869236](#)
- [5] Grenander, U. and Rosenblatt, M. (1984). *Statistical Analysis of Stationary Time Series*, 2nd ed. New York: Chelsea. [MR0890514](#)
- [6] Hallin, M., Lu, Z. and Tran, L.T. (2001). Density estimation for spatial linear processes. *Bernoulli* **7** 657–668. [MR1849373](#)
- [7] Hallin, M., Lu, Z. and Yu, K. (2009). Local linear spatial quantile regression. *Bernoulli* **15** 659–686. [MR2555194](#)
- [8] Jennrich, R.I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* **40** 633–643. [MR0238419](#)
- [9] Jones, H.T. (1943). Fitting polynomial trends to seasonal data by the method of least squares. *J. Amer. Statist. Assoc.* **38** 453–465.
- [10] Knuth, D.E. (1968). *The Art of Computer Programming. Vol. 1: Fundamental Algorithms*. Reading, MA: Addison-Wesley.
- [11] Lu, Z., Lundervold, A., Tjøstheim, D. and Yao, Q. (2007). Exploring spatial nonlinearity using additive approximation. *Bernoulli* **13** 447–472. [MR2331259](#)
- [12] Malinvaud, E. (1970). The consistency of nonlinear regressions. *Ann. Math. Statist.* **41** 956–969. [MR0261754](#)

- [13] Nagaraj, N.K. and Fuller, W.A. (1991). Estimation of the parameters of linear time series models subject to nonlinear restrictions. *Ann. Statist.* **19** 1143–1154. [MR1126318](#)
- [14] Nielsen, M.Ø. (2007). Local Whittle analysis of stationary fractional cointegration and the implied-realized volatility relation. *J. Bus. Econom. Statist.* **25** 427–446. [MR2410027](#)
- [15] Robinson, P.M. (2008). Multiple local Whittle estimation in stationary systems. *Ann. Statist.* **36** 2508–2530. [MR2458196](#)
- [16] Robinson, P.M. and Vidal Sanz, J. (2006). Modified Whittle estimation of multilateral models on a lattice. *J. Multivariate Anal.* **97** 1090–1120. [MR2276150](#)
- [17] Sun, Y. and Phillips, P.C.B. (2003). Nonlinear log-periodogram regression for perturbed fractional processes. *J. Econometrics* **115** 355–389. [MR1984781](#)
- [18] Tjøstheim, D. (1978). Statistical spatial series modelling. *Adv. in Appl. Probab.* **10** 130–154. [MR0471224](#)
- [19] Tjøstheim, D. (1983). Statistical spatial series modelling. II. Some further results on unilateral lattice processes. *Adv. in Appl. Probab.* **15** 562–584. [MR0706617](#)
- [20] van de Geer, S.A. (2000). *Applications of Empirical Process Theory. Cambridge Series in Statistical and Probabilistic Mathematics* **6**. Cambridge: Cambridge Univ. Press. [MR1739079](#)
- [21] Wu, C.F. (1981). Asymptotic theory of nonlinear least squares estimation. *Ann. Statist.* **9** 501–513. [MR0615427](#)
- [22] Yajima, Y. (1988). On estimation of a regression model with long-memory stationary errors. *Ann. Statist.* **16** 791–807. [MR0947579](#)
- [23] Yajima, Y. and Matsuda, Y. (2008). Asymptotic properties of the LSE of a spatial regression in both weakly and strongly dependent stationary random fields. Preprint CIRJE-F-587. Faculty of Economics, Univ. Tokyo.
- [24] Yao, Q. and Brockwell, P.J. (2006). Gaussian maximum likelihood estimation for ARMA models II: Spatial processes. *Bernoulli* **12** 403–429.

*Received January 2010 and revised December 2010*