The London School of Economics and Political Science

Statistical Inference on Linear and Partly Linear Regression with Spatial Dependence: Parametric and Nonparametric Approaches

Supachoke Thawornkaiwong

A thesis submitted to the Department of Economics of the London School of Economics for the degree of Doctor of Philosophy, London, August 2012

Declaration

I certify that the thesis I have presented for examination for the Ph.D. degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorisation does not, to the best of my belief, infringe the rights of any third party.

I declare that my thesis consists of approximately 70000 words.

Statement of conjoint work

I confirm that Chapter 3 was jointly co-authored with Professor Peter M. Robinson and I contributed 50% of this work.

Abstract

The typical assumption made in regression analysis with cross-sectional data is that of independent observations. However, this assumption can be questionable in some economic applications where spatial dependence of observations may arise, for example, from local shocks in an economy, interaction among economic agents and spillovers.

The main focus of this thesis is on regression models under three different models of spatial dependence. First, a multivariate linear regression model with the disturbances following the Spatial Autoregressive process is considered. It is shown that the Gaussian pseudo-maximum likelihood estimate of the regression and the spatial autoregressive parameters can be root-n-consistent under strong spatial dependence or explosive variances, given that they are not too strong, without making restrictive assumptions on the parameter space. To achieve efficiency improvement, adaptive estimation, in the sense of Stein (1956), is also discussed where the unknown score function is nonparametrically estimated by power series estimation. A large section is devoted to an extension of power series estimation for random variables with unbounded supports.

Second, linear and semiparametric partly linear regression models with the disturbances following a generalized linear process for triangular arrays proposed by Robinson (2011) are considered. It is shown that instrumental variables estimates of the unknown slope parameters can be root-n-consistent even under some strong spatial dependence. A simple nonparametric estimate of the asymptotic variance matrix of the slope parameters is proposed. An empirical illustration of the estimation technique is also conducted.

Finally, linear regression where the random variables follow a marked point process is considered. The focus is on a family of random signed measures, constructed from the marked point process, that are second-order stationary and their spectral properties are discussed. Asymptotic normality of the least squares estimate of the regression parameters are derived from the associated random signed measures under mixing assumptions. Nonparametric estimation of the asymptotic variance matrix of the slope parameters is discussed where an algorithm to obtain a positive definite estimate, with faster rates of convergence than the traditional ones, is proposed.

Acknowledgement

I would like to thank Professor Peter M. Robinson for his endless support, kindness and particularly his patience. I have learnt a great deal of econometrics and learnt to try to think more thoroughly from his detailed and substantive comments. He has been an immense source of good examples, inspiration, encouragement and support throughout my Ph.D. To him, I owe my deepest acknowledgements.

I would like to thank many individuals in the Econometrics group at the LSE during 2004 - 2008, in particular Professor Javier Hidalgo, Professor Oliver Linton, Dr. Marcia Schafgans and Dr. Myung Hwan Seo for lively discussions and support.

I also would like to thank many individuals who previously taught me mathematics, statistics, econometrics and economics, particularly Dr. Javier Hualde who gave an excellent introductory course in econometrics. His impressive teaching has played an important role in shaping my academic path towards research in econometrics.

An endless support form Mark Wilbor, the MRes-PhD in Economics Programme Manager, is appreciated.

I would like to thank the Bank of Thailand for the scholarship during the first year of my doctoral study and the LSE for the scholarship during the 2nd, 3rd, and 4th years of my study.

Chapter 2 of the thesis was written from May through early August 2012. The author will not be able to complete this chapter without the kind permission from many executives at the Bank of Thailand to take a leave from mid June to mid August.

Throughout this period, I am grateful to my family for their moral support. The support and love from Dr. Thanyaporn Chankrajang, my fiancee, has endlessly given me strength and energy to complete this thesis.

Contents

List of Tables and Figures

1 Introduction

Modern econometrics can, to some extent, be regarded as a branch of mathematical statistics aimed at providing statistical tools for economic analysis. Traditionally, cross-sectional data were analysed in microeconomic studies whereas time series data were employed in the macroeconomic counterpart. However, this distinction is no longer prevailing. There has been a rather significant movement among macroeconomists to collect and analyse cross-sectional data in order to understand macroeconomic behaviours. Household expenditure surveys have played a crucial role in helping macroeconomists understand consumption and saving behaviours. Surveys of consumer finances have also become popular for empirical analysis of asset pricing. Investment and R&D data at firm levels have improved macroeconomists' understanding of investment and R&D decisions, which play a role in short-term economic fluctuations and are widely accepted as being vital for economic growth. Cross-sectional data are currently playing key roles in other areas of studies such as unemployment and credit markets too.

There are many reasons explaining the popularity of cross-sectional data in macroeconomic analysis. Given that most macroeconomic theories are currently based on microeconomic foundations, which focus on decisions of economic agents in an economy, it is vital to check at the right level, e.g. households or firms, whether such theories are valid. Moreover, cross-sectional data are particularly useful for policy evaluation such as effects of minimum wages and monetary policy. If one were to rely on aggregate data, one would have to analyse only a few data points whereas the micro-level data can give a great deal of information.

The reader may be thinking of panel data and consider them as being different from the cross-secional one. However, given that most panel data used in economic analysis have much shorter time span compared with the number of cross-sectional observations, this type of panel data can be regarded, from the theoretical point of view, as cross-sectional data with higher dimensions. Hence the theories developed for cross-sectional data will be applicable to panel data (over a short time span) too. A serious discussion of panel data with large cross-sectional observations over a long period of time requires a proper theoretical foundation for spatio-temporal dependence, which is beyond the scope of this thesis.

Independence of observations is traditionally assumed when analysing cross-sectional data. However, this typical assumption can be questionable. A shutdown of a factory will affect many households' income in a given neighbourhood. A natural disaster or a contagious disease can substantially lead to a reduction of output of a large region of a country. Spillovers and externalities may carry some impacts of a certain economic shock to other communities outside the one where the shock takes place. Trade can indirectly induce interdependence in activities of economic agents. Many economic theories also suggest dependence of economic variables across space. A change of one player's strategy can result in a change of a Nash equilibrium. Risk averse agents will make insurance contracts allowing them to smooth idiosyncratic shocks and this implies dependence in consumption across individuals. In this thesis we call dependence across cross-sectional observations, spatial dependence. This kind of dependence does not necessarily arise from a physical space. It arise from some other economic spaces where an economic distance may be different from the physical one.

There are two main strategies in econometric literature aimed at modelling spatial dependence. The first line of research is based on the idea that a family of random variables exhibiting spatial dependence can be represented as a linear process with independent innovations. The most popular parametric model on this line of research is the Spatial Autoregressive (SAR) model. Recently Robinson (2011) proposed a generalized linear process for a triangular array of random variables and showed that a broad class of spatial processes can be represented by such a generalized linear process. It should be stressed that the generalized linear process of Robinson (2011) is a nonparametric model. The advantage of this modelling strategy is that many well-established results from linear time series can be extended.

The other line of research is to assume that the data is, to some degree, second-order stationary. Conley (1999) considered irregularly spaced data in $\mathbb{R}^2$. He assumed that the data is a marked point process where the marks and the ground process are independent. Moreover, he assumed that the marks are stationary random fields and the ground process is a hard-core process. The assumption that the ground process is a hard-core process allows researchers to regard irregularly spaced spatial data on $\mathbb{R}^2$ as a random field on the lattice $\mathbb{Z}^d$, where $\mathbb{Z}^d$ is the Cartesian d-product of the space of integers $\mathbb{Z}$. Even though Conley (1999) was able to show analytical tractability of his model, his assumptions on the hard-core process is restrictive and result in computationally intensive calculation.

In this thesis, we investigate both lines of researches. In Chapter 2, we investigate a multivariate linear regression model with the disturbances following a multivariate SAR model. The parametric set-up of the SAR model allows us to employ likelihood based inference. We first consider the Gaussian pseudo-maximum likelihood estimate of the unknowns. We show that under mild regularity conditions, such estimate can be root-n-consistent. Our regularity conditions are quite different from the ones in the existing literature. First, we do not impose excessive restriction on the parameter space. Second, we show and stress analytical tractability and flexibility of the spectral norm compared with the $\|\cdot\|_1$ and $\|\cdot\|_\infty$ norms commonly employed in the literature. Employing a different technique for proof of consistency of the estimate, we can avoid row or column normalization. We also allow the SAR process to exhibit long-range dependence or explosive variances while the existing literature focuses on short-range dependence and bounded variances. The Gaussian pseudo-maximum likelihood estimate will lose its efficiency if the innovations of the SAR process are not normally distributed. This leads us to consider efficiency improvement of the Gaussian pseudo-maximum likelihood estimate by nonparametrically estimating the unknown score function of the distribution of the innovations. This "adaptive" estimate of the slope parameters of the regression is asymptotically as efficient as the one obtained from the maximum likelihood estimation when the density function is known. Our nonparametric estimate of the unknown score functions is a power series nonparametric estimate. In order to allow the number of approximating functions to increase faster than the ones in the literature, we employ properties of orthonormal polynomials in our proof. We also extend some results in power series literature to allow for random variables with unbounded support.

In Chapter 3, we consider linear and partly linear regression models where the disturbances follow a generalized linear process in Robinson (2011). Central limit theorems are developed for instrumental variables estimates of linear and semiparametric partly linear regression models. We also show that the estimate of the slope parameters in the linear part of the partly linear model can be root-n-consistent similar to the case for independent data. We discuss estimation of the variance matrix, including estimates that are robust to disturbance heteroscedasticity and/or dependence. A Monte Carlo study of finite-sample performance is included. In an empirical example, the estimates and robust and non-robust standard errors are computed from Indian regional data, following tests for spatial correlation in disturbances, and nonparametric regression fitting. Some final comments discuss modifications and extensions.

In Chapter 4, we consider a certain class of a marked point process which can give a good representation of cross-sectional data exhibiting spatial dependence. This interpretation offers a nonparametric approach in capturing spatial dependence. Under some assumptions, a linear functional of the marked point process forms a second-order stationary random (signed) measure on the state space $\mathbb{R}^d$ and its spectral properties can be developed. We then consider a linear regression model from this marked point process. The asymptotic normality of the least squares estimate of the slope parameters of the model is derived based on laws of large numbers and central limit theorems for random (signed) measures. Estimation of spectral density of the random signed measure and the asymptotic variance matrix of the least squares estimate are discussed. Finally, we propose an algorithm which can be employed to obtain a positive definite estimate of an unknown positive definite matrix. Our algorithm can be applied to general estimation of unknown positive definite matrices. One advantage of this algorithm is that it can achieve faster rates of mean-square convergence of the estimate compared with other conventional positive semidefinite estimates commonly employed in the literature.

2 Likelihood Based Inference on Multivariate Regression with Spatial Autoregressive Disturbances

2.1 Introduction

In this chapter, we consider a parametric model employed to capture spatial dependence. The most popular parametric model in econometric literature is the Spatial Autoregressive (SAR) model introduced by Cliff and Ord (1973) and popularised by Anselin (1988). The simplest SAR model for a triangular array of random variables $\{u_{i,n},\ 1 \leq i \leq n,\ n \geq 1\}$ is of the form

$$u_{i,n} = \rho_0 \sum_{j=1}^{n} w_{ij,n} u_{j,n} + \varepsilon_i, \tag{2.1}$$

where ρ_0 is the spatial autoregressive parameter, $w_{ij,n}$ are nonstochastic weights and $\{\varepsilon_i\}$ is a sequence of uncorrelated random variables with zero mean and constant variance σ_0^2. For simplicity, the subscript n will be omitted from the presentation. The model in (2.1) can be re-written in a matrix form as

$$u = \rho_0 W u + \varepsilon, \tag{2.2}$$

where $u = (u_1,\ ...,\ u_n)'$, W is the $n \times n$ matrix whose (i,j)-th element is w_{ij}, $\varepsilon = (\varepsilon_1,\ ...,\ \varepsilon_n)'$ and the prime $'$ denotes transposition. The model in (2.2) can be re-written as

$$Su = (I_n - \rho_0 W)\, u = \varepsilon,$$

where $S = I_n - \rho_0 W$. If ρ_0 takes a value such that S is invertible, then the model implies that

$$u = S^{-1}\varepsilon \ \text{ and } \ Var\,(u) = \sigma_0^2\,(S'S)^{-1}.$$

Unlike in time series analysis, it may not be obvious for practitioners how the weights w_{ij} should be chosen. One natural choice of the weights is to rely on "economic" distances of each pair of observations. In this case, the weights should have inverse relationships with distances to reflect falling-off dependence as distances increase. However, the row-normalisation restriction, i.e.

$$w_{ij} = \frac{f\,(d_{ij})}{\sum_{j=1}^{n} f\,(d_{ij})},$$

where d_{ij} is a distance between the i-th and j-th observations and f is a chosen decreasing function, is sometimes imposed. This restriction can be a drawback since the model in (2.1) may lose its economic appeal. Further normalisation may be imposed to make the matrix W uniformly bounded in both row and column sums to satisfy certain theoretical assumptions in the existing literature. Moreover, the parameter space of the unknown ρ_0 is usually restricted to ensure that $S\,(\rho)^{-1}$ is uniformly bounded in both row and column sums for all ρ in the parameter space, where $S\,(\rho) = I_n - \rho W$. See, for example, Kelejian and Prucha (1998) and Lee (2004).

Practitioners may find restrictions on both the parameter space of ρ_0 and on W too restrictive. In many applications, practitioners may prefer a symmetric matrix reflecting distances between economic agents or observations, i.e. $w_{ij} = f(d_{ij})$, where d_{ij} is the distance between the i-th and j-th obsevations, as a natural choice of the weighting matrix. In this case, row or column normalisation will be restrictive. Moreover, when a chosen function f is known up to an unknown scale, the unknown scale can be absorbed by the spatial autoregressive parameter ρ_0. This implies that a further restriction commonly imposed on the parameter space of ρ_0 will become restrictive.

In this paper, we also show that the assumptions imposed in the existing literature do not cover two important scenarios, namely explosive variances of some observations and long range dependence.

In this paper, we show that these restrictions are unnecessary to obtain a root-n-consistent estimate of the unknown autoregressive parameter. Instead of considering a simple univariate SAR model, we consider a multivariate linear regression model with SAR disturbances. We show that with Gaussian pseudo-maximum likelihood estimation, we can obtain a root-n-consistent estimate of the unknown parameters under long-range spatial dependence or explosive variances.

When the innovations ε_i are i.i.d., we also show how to obtain efficiency improvement over the Gaussian pseudo-maximum likelihood estimate by nonparametrically approximate the unknown score function of the distribution of the innovations. Our efficient estimate is based on series approximation of the unknown score function suggested by Beran (1976) for finite dimensional cases. This estimate is computationally simple and the issue of selecting the trimming parameter can be avoided. In order to nonparametrically estimate the unknown score function in a general infinite-dimensional space, one has to allow the number of approximating functions to increase to infinity at an appropriate rate. Newey (1988) extended Beran's technique to obtain an adaptive estimate of the slope parameters of a linear regression model with i.i.d. data but the number of approximating functions has to go to infinity at a rate that is slower than logarithm of the sample size. Robinson (2005), considering efficient estimation of time series regression with fractional disturbances, showed that the condition in Newey (1988) can be relaxed and allow the number of approximating functions to increase at the rates slightly faster than that in Newey (1988).

In this paper, we show that in order to obtain an efficiency improvement of an estimate of the slope parameter in a multivariate linear regression with SAR disturbances, the number of approximating functions can indeed increase with the sample size at a polynomial rate. The proof relies on results from power series approximation literature. Unlike other papers in the literature, we show that in order to allow the number of approximating power functions to grow at the rate that is proportional to a fractional power of the sample size, we do not need to make a restrictive assumption that the density function of the disturbances must have bounded support. The result in this chapter should be applicable to other semiparametric models in econometrics, where the power series approximation is employed to estimate the nonparametric part of a model.

A simple univariate SAR model and its multivariate extension are discussed in Section 2. We show that the spectral norm can be more flexible than other norms such as the maximum column sum and maximum row sum norms commonly employed in the literature. We also discuss how to relax the condition on uniform boundedness of row and column sums to possibly allow for long-range dependence or explosive variances. Some further analytically tractable results can be obtained when W is symmetric. In section 3, we discuss consistency and asymptotic normality of the Gaussian pseudo-maximum likelihood estimate of a multivariate linear regression model with multivariate SAR disturbances. In section 4, we extend some results in nonparametric series approximation to allow for random variables with unbounded support. Finally, efficiency improvement of the slope parameter in the multivariate linear regression model is discussed. Proofs and technical lemmas are left in the Appendices.

2.2 Spatial Autoregressive Model

In this section, we discuss a spectral norm and show its analytical tractability for the SAR model. First, we introduce some notations. Let A be an $n \times n$ matrix and a_{ij} denote its (i,j)-th element. Define $\overline{\lambda}(A)$ and $\underline{\lambda}(A)$ as the largest and smallest eigenvalues of A, respectively. Define $\|A\|_1 = \max_{1 \le j \le n} \sum_{i=1}^n |a_{ij}|$, $\|A\|_\infty = \max_{1 \le i \le n} \sum_{j=1}^n |a_{ij}|$, and $\|A\| = \left[\overline{\lambda}(A'A)\right]^{1/2}$. Let $\rho(A) = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } A\}$ and $|A|$ be the determinant of A. In this chapter, a square matrix A of order n is positive definite (p.d.) if A is symmetric and for any $x \in \mathbb{R}^n$ such that $x \neq 0$, $x'Ax > 0$. Similarly, a square matrix A of order n is positive semidefinite (p.s.d.) if A is symmetric and $x'Ax \ge 0$ for any $x \in \mathbb{R}^n$.

2.2.1 Univariate Spatial Autoregressive Model

Consider a univariate SAR model in (2.2). As mentioned in the previous section, $Var(u) = \sigma_0^2 (S'S)^{-1}$. The most common assumption in the literature, as in Kelejian and Prucha (1998) and Lee (2004), is

Assumption A1 $\|S^{-1}\|_1 + \|S^{-1}\|_\infty$ *is bounded uniformly in* n*.*

Since it can be shown that both $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are matrix norms as defined in Horn (1985), one implication of this condition is that $\|Var(u)\|_1 + \|Var(u)\|_\infty$ is bounded uniformly in n. Now consider the spectral norm. One advantage of the spectral norm can be seen directly from Lemma A1 that for any square matrix A, $\|A\| = \|A'\|$. In order to get an analogous result on a bound for $Var(u)$, by employing the spectral norm, we need to make the following condition.

Assumption A2 $\underline{\lambda}(S'S)$ *is bounded away from zero uniformly in* n*.*

Under Assumption A2, $S'S$ is p.d. and hence $\overline{\lambda}\left((S'S)^{-1}\right) = \{\underline{\lambda}(S'S)\}^{-1}$. Because $\|A\| = \|A'\|$, Assumption A2 is equivalent to the condition that $\left\|S^{-1}\right\|$ is bounded uniformly in n. As $(S'S)^{-1}$ is p.d., by Lemma A3, $\left\|(S'S)^{-1}\right\| = \overline{\lambda}\left((S'S)^{-1}\right)$. Therefore, Assumption A2 is also equivalent to the condition that $\left\|(S'S)^{-1}\right\|$ and $\|Var(u)\|$ are bounded uniformly in n. As $Var(u)$ is symmetric, by Lemma A3,

$$\|Var(u)\| \leq \|Var(u)\|_1 \text{ and } \|Var(u)\| \leq \|Var(u)\|_\infty .$$

Therefore Assumption A1 implies Assumption A2. If variance matrices are of primary concern, the following theorem shows that the spectral norm is strictly weaker than $\|\cdot\|_1$ and $\|\cdot\|_\infty$ norms.

Theorem A *Consider any family $\{A_n\}$ of positive definite matrices of order n. (i) $\|A_n\| \leq \|A_n\|_1$ and $\|A_n\| \leq \|A_n\|_\infty$. (ii) There exists a family $\{A_n\}$ such that $\|A_n\|$ are uniformly bounded in n but $\|A_n\|_1$ and $\|A_n\|_\infty$ are not.*

Even though Assumption A2 is weaker than Assumption A1 that is commonly employed in the literature, it may be too strong. With reference to time series literature, consider a covariance stationary process $\{v_t\}$. One possible nonparametric definition for long-range dependece of $\{v_t\}$ is that

$$Var\left(n^{-1/2}\sum_{t=1}^{n} v_t\right) \to \infty, \quad \text{as } n \to \infty. \tag{2.3}$$

Let $z = n^{-1/2}(1, \ ..., \ 1)'$ be a vector in $\mathbb{R}^n$. Since $n^{-1/2}\sum_{t=1}^{n} v_t = z'v$ where $v = (v_1, \ ..., \ v_n)'$, it follows that $Var\left(n^{-1/2}\sum_{t=1}^{n} v_t\right) = z'Var(v)\,z$, where $\|z\| = 1$. Since $z'Var(v)\,z \leq \|Var(v)\|$, then $\|Var(v)\| \to \infty$, as $n \to \infty$ given that $\{v_t\}$ has long-range dependence. Alternatively, suppose $\{v_t\}$ has absolutely continuous spectral distribution and let f be the density function. It is common to say that $\{v_t\}$ has short memory if

$$0 < f(\lambda) < \infty, \text{ for all } \lambda \in [-\pi, \ \pi). \tag{2.4}$$

Note that this definition excludes seasonal long memory. If $\{v_t\}$ has short memory as defined in (2.4), then one observation in section 5.2 (b) in Grenander and Szego (1984) implies that $\|Var(v)\|$ is bounded uniformly in n. These two results suggest that uniform boundedness of $\|Var(v)\|$ can be a rather sensible description of time series exhibiting short-range dependence.

Given these results one may be tempted to say that a SAR process u has long range dependece if

$$\|Var(u)\| \to \infty \quad \text{as } n \to \infty. \tag{2.5}$$

It is clear that Assumption A2 does not allow for this possibility. This may be a serious drawback of Assumption A2 since some spatial data may be subject to long-range dependence. Farifield Smith (1938) mentioned the problem of (2.3) when analysing agricultural spatial data. However, this definition may be misleading since it is possible that (2.5) holds when $Var(u_i) \to \infty$ as $n \to \infty$ for some i but $Cov(u_i,\ u_j)$ becomes arbitrarily small sufficiently fast as the distance between the i-th and j-th observations increases. In other words, condition (2.5) may arise not from long-range dependence but from explosive variances of some observations. Neverthess, this explosive behaviour of the variances may arise naturally in some economic applications. With reference to economic geography literature, agglomeration of economic activities in a certain location may be a norm rather than an exception to benefit from economies of scales. See Fujita, Krugman and Venables (2001) for a reference. Hence concentration of economic activities may be sources of explosive variances.

As $S = I_n - \rho_0 W$, Lemma A4 implies that λ is an eigenvalue of S if and only if $\lambda = 1 - \rho_0 \omega$ where ω is an eigenvalue of W. If $\rho_0 = 0$, then $S = I_n$ and we have a trivial case. Suppose $\rho_0 \neq 0$, then S is invertible, i.e. $\lambda \neq 0$, as long as none of the eigenvalues of W is equal to $1/\rho_0$. Non-singularity of S can be regarded as an identification restriction so that each u_i can be written uniquely as a linear combination of ε_j. It is generally difficult to have much information about $\|Var(u)\|$ since it depends on the unknown ρ_0 and the relationship between $Var(u)$ and ρ_0 may be highly nonlinear. However, if W is symmetric, for example when economic distances are employed to construct W without row or column normalization, many analytically tractable results can be obtained.

If W is symmetric, then S and S^{-1} are symmetric. Lemma A5 also implies that $\|Var(u)\| = \sigma_0^2 \max\left\{(1-\rho_0\omega)^{-2} : \omega \text{ is an eigenvalue of } W\right\}$. Even though S is assumed to be non-singular, $\|Var(u)\|$ can becomes arbitrarily large if at least one of the eigenvalues of W gets arbitrarily close to $1/\rho_0$ as $n \to \infty$. The rate at which $\|Var(u)\|$ becomes explosive depends on the rate at which one of the eigenvalues of W gets arbitrarily close to $1/\rho_0$. In other words, for a given ρ_0, the explosive behaviour of $\|Var(u)\|$ depends on the characteristic values of the weight matrix W.

Compared with the symmetric case, $\|Var(u)\|$ loses its analytical tractability when W is not symmetric. The complexity of $\|Var(u)\|$ when W is not symmetric can be illustrated from the truncated first-order autoregressive model. Let $x_1 = \varepsilon_1$ and $x_t = \rho x_{t-1} + \varepsilon_t,\ t \geq 2$, where $\{\varepsilon_t\}$ is a white noise process. Then it can be shown that $x = (x_1,\ ...,\ x_n)'$ can be represented as in (2.2) with $w_{ij} = \delta_{i,j+1}$, where δ_{ij} is the Kronecker's delta. It follows that the resulting weight matrix W is a lower shift matrix and is also nilpotent. Hence, every eigenvalue of W is equal to zero and, by Lemma A4, every eigenvalue of S is equal to unity regardless of the value of ρ_0. This implies that S is always invertible and every eigenvalue of S^{-1} is equal to unity regardless of the value of ρ_0. However, when $\rho_0 = 1$, i.e. $\{x_t\}$ is a truncated random walk process, $\|Var(x)\|$ becomes explosive at a fairly fast rate. This example illustrates the difficulty in determining the behaviour of $\|Var(u)\|$ particularly when W is not symmetric.

The complexity of $\|Var(u)\|$ when W is not symmetric arises from the result shown in

Lemmas A1 and A10 in the Appendix that $\|Var(u)\|$ is proportional to

$$\left\|(S'S)^{-1}\right\| = \overline{\lambda}\left(S^{-1}\left(S^{-1}\right)'\right) \geq \rho\left(S^{-1}\right)^2.$$

Lemma A2 shows that this inequality becomes an equality when W is symmetric. When W is not symmetric, one can at least conclude that if $\rho\left(S^{-1}\right)$ is explosive, then $\left\|(S'S)^{-1}\right\|$ is also explosive. One can only infer that the rate at which $\left\|(S'S)^{-1}\right\|$ becomes explosive is at least as fast as the rate of $\rho\left(S^{-1}\right)^2$. However, a sharper rate at which $\left\|(S'S)^{-1}\right\|$ becomes explosive may not be easy to conclude. The truncated random walk process is a good example showing this complexity.

Define $G = WS^{-1}$. It will be clear later that G arises naturally from the Gaussian psuedo-maximum likelihood estimation of the unknown ρ_0. If $\rho_0 = 0$, then $G = W$. It follows form the proof of Lemma A6 that if $\rho_0 \neq 0$,

$$G = \rho_0^{-1}\left(S^{-1} - I_n\right). \tag{2.6}$$

This equality shows that when there is spatial dependence, i.e. $\rho_0 \neq 0$, the matrix $\rho_0 G$ is essentially the matrix S^{-1}. Recall that $u = S^{-1}\varepsilon$. It follows that $u_i = \sum_{j=1}^n b_{ij}\varepsilon_j$ where b_{ij} is the (i,j)-th element of S^{-1}. Following (2.6), the i-th element of $\rho_0 G\varepsilon$ is $\sum_{j=1}^n b_{ij}\varepsilon_j - \varepsilon_i$. That is $\rho_0 G\varepsilon$ is essentially the same as u. Moreover, Lemma A6 indicates that

$$\left\|S^{-1}\right\| - 1 \leq \|\rho_0 G\| = \left\|S^{-1}\right\| + 1.$$

Hence $\|G\| \sim |\rho_0|^{-1}\left\|S^{-1}\right\|$ as $\left\|S^{-1}\right\| \to \infty$, where '$\sim$' indicates that the ratio of left and right sides tends to 1. Note that $\left\|S^{-1}\right\|$ can get arbitrarily large only if $\rho_0 \neq 0$.

2.2.2 Multivariate Spatial Autoregressive Model

In a multivariate case, we have n observations and g equations. The univariate SAR model in (2.1) can be generalized to

$$u_{it} = \rho_{0t}\sum_{j=1}^n w_{ijt}u_{jt} + \varepsilon_{it}, \quad i = 1,\ ...,\ n,\ t = 1,\ ...,\ g, \tag{2.7}$$

where the index i is associated with the i-th observation and the index t is associated with the t-th equation. In a matrix form, this can be written as

$$u_{\cdot t} = \rho_{0t}W_t u_{\cdot t} + \varepsilon_{\cdot t}, \quad t = 1,\ ...,\ g, \tag{2.8}$$

where $u_{\cdot t} = (u_{1t},\ ...,\ u_{nt})'$, $\varepsilon_{\cdot t} = (\varepsilon_{1t},\ ...,\ \varepsilon_{nt})'$ and W_t is the matrix whose (i,j)-th element is w_{ijt}. This specification assumes no direct cross-equation effects but cross-equation dependence arises from dependence structure of $(\varepsilon_{i1},\ \varepsilon_{i2},\ ...,\ \varepsilon_{ig})'$. For square matrices $A_1,\ ...,\ A_n$, not necessarily of the same order, define $diag(A_1,\ ...,\ A_n)$ as the block-diagonal matrix whose diagonal blocks are $A_1,\ ...,\ A_n$, respectively. Then, the specification in (2.8) can be re-written as

$$u = diag\left(\rho_{01}W_1,\ ...,\ \rho_{0g}W_g\right)u + \varepsilon,$$

where $u = \left(u'_{\cdot 1},\ ...,\ u'_{\cdot g}\right)'$ and $\varepsilon = \left(\varepsilon'_{\cdot 1},\ ...,\ \varepsilon'_{\cdot g}\right)'$. Define $S_t = I_n - \rho_{0t}W_t$, then

$$Su = \varepsilon,$$

where $S = diag\left(S_1,\ ...,\ S_g\right)$.

Assumption B1 *For $t = 1,\ ...,\ g,\ W_t$ are $n \times n$ matrices of nonstochastic weights w_{ijt} and S_t are non-singular for all $n \geq 1$.*

Let $\mathbf{1}$ be the indicator function.

Assumption B2 *Let $\varepsilon_{i\cdot} = \left(\varepsilon_{i1},\ ...,\ \varepsilon_{ig}\right)'$. (i) $\mathbb{E}\left(\varepsilon_{i\cdot}\right) = 0$ for all i. (ii) $\mathbb{E}\left(\varepsilon_{i\cdot}\varepsilon'_{j\cdot}\right) = \Sigma_0 \mathbf{1}\left(i = j\right)$ for all $i,\ j$, where Σ_0 is p.d..*

Under Assumptions B1 and B2,

$$u = S^{-1}\varepsilon,$$

where $S^{-1} = diag\left(S_1^{-1},\ ...,\ S_g^{-1}\right)$,

$$Var\left(\varepsilon\right) = \Sigma_0 \otimes I_n,$$

where $\otimes$ is the Kronecker product, and

$$Var\left(u\right) = S^{-1}\left(\Sigma_0 \otimes I_n\right)\left(S^{-1}\right)' = \left\{S'\left(\Sigma_0^{-1} \otimes I_n\right)S\right\}^{-1}.$$

It follows from Lemmas A1 and A7 that

$$\left\|Var\left(u\right)\right\| \leq \left\|\left(\Sigma_0 \otimes I_n\right)\right\| \left\|S^{-1}\right\|^2 = \left\|\Sigma_0\right\| \left\|S^{-1}\right\|^2.$$

If (2.5) holds, then it must be the case that

$$\left\|S^{-1}\right\| \to \infty \text{ as } n \to \infty.$$

Hence $\left\|S^{-1}\right\|$ is the source of an explosive behaviour of $\left\|Var\left(u\right)\right\|$. Lemma A8 implies that $\left\|S^{-1}\right\| = \max_{1 \leq t \leq g}\left\|S_t^{-1}\right\|$. This simple relationship suggests that the results previously established for a univariate process can be applicable to a multivariate process. Similarly, we can define $G = diag\left\{G_1,\ ...,\ G_g\right\}$ where $G_t = W_tS_t^{-1}$. Lemma A8 implies that $\left\|G\right\| = \max_{1 \leq t \leq g}\left\|G_t\right\|$ and hence the results for a univariate case can be applied.

2.3 Multivariate Linear Regression

In this paper we consider a multivariate linear regression model

$$y_{it} = x_{it}'\beta_0 + u_{it}, \quad i = 1, \ ..., \ n, \ t = 1, \ ..., \ g,$$

where y_{it} are scalar random variables, x_{it} are $\mathbb{R}^K$-valued random variables, β_0 are unknown vectors in $\mathbb{R}^K$ and the disturbances u_{it} follow a multivariate SAR process as defined in (2.7). Then, for each $t = 1, \ ..., \ g$, we have

$$y_{\cdot t} = X_t\beta_0 + u_{\cdot t},$$

where $y_{\cdot t} = (y_{1t}, \ ..., \ y_{nt})'$, $X_t = (x_{1t}, \ ..., \ x_{nt})'$ and $u_{\cdot t} = (u_{1t}, \ ..., \ u_{nt})'$. This can be written as

$$y = X\beta_0 + u,$$

where $y = \left(y_{\cdot 1}', \ ..., \ y_{\cdot g}'\right)', \ X = \left(X_1', \ ..., \ X_g'\right)'$ and $u = \left(u_{\cdot 1}', \ ..., \ u_{\cdot g}'\right)'$.

Under the assumption that $\{\varepsilon_{i\cdot}\}$, as defined in the previous section, is a sequence of independent vectors of jointly normally distributed random variables with zero mean and $Var(\varepsilon_{i\cdot}) = \Sigma_0$, and the assumption that the regressors x_{it} and ε_{js} are independent for all $i, \ j, \ t$ and s, and that the distributions of x_{it} do not depend on $\beta_0, \ \Sigma_0$ and $\rho_{01}, \ ..., \ \rho_{0g}$, the log-likelihood function of y for the maximum likelihood estimation of the unknown parameters is

$$\begin{aligned} l_n(\beta, \ \Sigma, \ \rho) \quad = \quad & -\frac{ng}{2}\log 2\pi + \frac{1}{2}\log\left|S(\rho)' S(\rho)\right| - \frac{1}{2}\log|\Sigma \otimes I_n| \\ & -\frac{1}{2}u'(\beta) S'(\rho)\left(\Sigma^{-1} \otimes I_n\right) S(\rho) u(\beta), \end{aligned}$$

where $\rho = \left(\rho_1, \ ..., \ \rho_g\right)', \ S_t(\rho) = I_n - \rho_t W_t, \ S(\rho) = diag(S_1(\rho), \ ..., \ S_g(\rho))$, and $u(\beta) = y - X\beta$.

If the normality assumption does not hold, we can employ this log-likelihood function to construct a loss function

$$\begin{aligned} Q_n(\beta, \ \Sigma, \ \rho) \quad = \quad & -\frac{1}{2g}\log\left|\Sigma^{-1}\right| - \frac{1}{2ng}\sum_{t=1}^{g}\log\left|S_t(\rho)' S_t(\rho)\right| \\ & +\frac{1}{2ng}u(\beta)' S(\rho)'\left(\Sigma^{-1} \otimes I_n\right) S(\rho) u(\beta). \end{aligned} \tag{2.9}$$

Let $\widehat{\beta}, \ \widehat{\Sigma}$ and $\widehat{\rho}$ be the minimizer of this loss function. Given Σ and ρ, the minimizer

$$\widehat{\beta}(\Sigma, \ \rho) = \left\{X' S(\rho)'\left(\Sigma^{-1} \otimes I_n\right) S(\rho) X\right\}^{-1} X' S(\rho)'\left(\Sigma^{-1} \otimes I_n\right) S(\rho) y.$$

If we ignore symmetry of Σ, then, applying the relationship that

$$tr(ABCD) = vec(C)'(D \otimes B') vec(A'),$$

it follows that

$$\widehat{\Sigma}(\rho) = \underset{\Sigma \in \mathbb{R}^{g\times g}}{\arg\min}\ \frac{1}{2g}\log|\Sigma| + \frac{1}{2ng} tr\left\{\Sigma^{-1}\left(\sum_{i=1}^{n} \varepsilon_{i\cdot}\left(\widehat{\beta},\ \rho\right)\varepsilon_{i\cdot}\left(\widehat{\beta},\ \rho\right)'\right)\right\}, \tag{2.10}$$

where

$$(\varepsilon_{1t}(\beta,\ \rho),\ ...,\ \varepsilon_{nt}(\beta,\ \rho))' = \varepsilon_{\cdot t}(\beta,\ \rho) = S_t(\rho)\, u_t(\beta)$$

and $\varepsilon_{i\cdot}(\beta,\ \rho) = (\varepsilon_{i1}(\beta,\ \rho),\ ...,\ \varepsilon_{ig}(\beta,\ \rho))'$. For each i, $\varepsilon_{i\cdot}\left(\widehat{\beta},\ \rho\right)\varepsilon_{i\cdot}\left(\widehat{\beta},\ \rho\right)'$ is p.s.d. regardless of the values of ρ. As n increases it is more likely that $\sum_{i=1}^{n}\varepsilon_{i\cdot}\left(\widehat{\beta},\ \rho\right)\varepsilon_{i\cdot}\left(\widehat{\beta},\ \rho\right)'$ is p.d. and hence singular. Then we can apply Lemma 3.2.2 in Anderson (2003) to show that

$$\widehat{\Sigma}(\rho) = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_{i\cdot}\left(\widehat{\beta},\ \rho\right)\varepsilon_{i\cdot}\left(\widehat{\beta},\ \rho\right)'. \tag{2.11}$$

Note that there is a typographical error in the statement of Lemma 3.2.2. There should be "$-$" in front of $N\log|G|$. Moreover, the same Lemma implies that the minimum value of the objective function in (2.10) is

$$C - \frac{1}{2g}\log\left|\widehat{\Sigma}(\rho)\right|,$$

where C is a constant. Hence

$$\widehat{\rho} = \underset{\rho\in\mathbb{R}^{g}}{\arg\min}\frac{1}{2g}\log\left|\widehat{\Sigma}(\rho)\right| - \frac{1}{2ng}\sum_{t=1}^{g}\log\left|S_t(\rho)' S_t(\rho)\right|.$$

Now we give a formal statement for the minimisation problem. Following Abadir and Magnus (2005), for any symmetric matrix A of order g, define the half-vec of A, $vech(A)$, as the $g(g+1)/2\times 1$ vector that is obtained from $vec(A)$ by eliminating all supradiagonal elements of A. Let $\theta = \left(\theta_1',\ \theta_2',\ \theta_3'\right)'$ where $\theta_1 = \beta,\ \theta_2 = vech\left(\Sigma^{-1}\right)$ and $\theta_3 = \rho = \left(\rho_1,\ ...,\ \rho_g\right)'$.

Let

$$\widehat{\theta} = \underset{\theta\in\Theta}{\arg\min} Q_n(\theta),$$

where $Q_n(\theta)$ is the right side of (2.9), $\Theta = \Theta_1\times\Theta_2\times\Theta_3$, $\Theta_1\subset\mathbb{R}^K$, $\Theta_2\subset\mathbb{R}^{g(g+1)/2}$ and $\Theta_3\subset\mathbb{R}^g$. Since there may be some $\rho\in\Theta_3$ such that $S(\rho)$ is singular and hence $\left|S(\rho)' S(\rho)\right| = 0$, we need to define $\log(x) = -\infty$ for $x = 0$. Note that $\widehat{\theta}$ can be interpreted as a Gaussian psuedo-maximum likelihood estimate as in Lee (2004). The consideration of Σ^{-1} rather than Σ substantially simplifies our proofs. It is important to note that we take $\theta_2 = vech\left(\Sigma^{-1}\right)$ rather than $vec\left(\Sigma^{-1}\right)$ to ensure that the asymptotic covariance matrix of $\widehat{\theta}_2$, where $\widehat{\theta}_2$ is a sub-vector of $\widehat{\theta}$ associated with θ_2, is non-singular. The consideration of $vech\left(\Sigma^{-1}\right)$ implicitly assume that Σ is symmetric. We can make this assumption without loss of generality since (2.11) shows that it does not matter whether the assumption of symmetry is imposed. We also stress the importance of the expression $\log\left|S_t(\rho)' S_t(\rho)\right|$ since it is not generally true that $|S_t(\rho)|\geq 0$ but it is always the case that $\left|S_t(\rho)' S_t(\rho)\right|\geq 0$. Lee and Yu (2010) considered Gaussian pseudo-maximum likelihood estimation of a panel data model that is essentially a multivariate model. However, Lee and Yu (2010) assumed that $\Sigma = I_g$, $W_t = W$ and $\rho_t = \rho$ for all $t = 1, \ ..., \ g$. This assumption essentially simplifies a multivariate SAR model to a univariate one.

2.3.1 Consistency

To show consistency of $\widehat{\theta}$, we make the following assumptions. Let $\mathbb{N}$ be the set of natural numbers.

Assumption B3 *$\{\varepsilon_{i\cdot}\}$ is a sequence of independent $\mathbb{R}^g$-valued random variables such that (i) $\mathbb{E}(\varepsilon_{i\cdot}) = 0$ for all i in $\mathbb{N}$. (ii) $\mathbb{E}(\varepsilon_{i\cdot}\varepsilon_{i\cdot}') = \Sigma_0$ for all i in $\mathbb{N}$, where Σ_0 is a positive definite matrix. (iii) There is a finite constant C such that*

$$\max_{1\leq t\leq g} \max_{i\geq 1} \mathbb{E}\varepsilon_{it}^4 \leq C.$$

Assumption B4 *Θ is a compact set. In addition, Θ_2 is a subspace of $\mathbb{R}^{g(g+1)/2}$ such that Σ is positive definite for all $vech(\Sigma) \in \Theta_2$.*

For any positive definite matrix A, let $A^{1/2}$ be the square root matrix of A. Define

$$H(\theta_2, \ \theta_3) = \left(\Sigma_0^{1/2} \otimes I_n\right)\left(S^{-1}\right)' S(\rho)' \left(\Sigma^{-1} \otimes I_n\right) S(\rho) S^{-1} \left(\Sigma_0^{1/2} \otimes I_n\right).$$

By definition, $H(\theta_2, \ \theta_3)$ is positive semidefinite for all $\theta_2 \in \Theta_2$ and all $\theta_3 \in \Theta_3$. The matrix G_t defined in the previous section can be written as $G = WS^{-1}$, where $W = diag\{W_1, \ ..., \ W_g\}$.

Assumption B5 *Let $\lambda_1, \ ..., \ \lambda_{ng}$ be eigenvalues of $H(\theta_2, \ \theta_3)$. For any $\delta > 0$, there exists $\eta > 0$ such that for some N,*

$$\inf_{\|\tau-\tau_0\|\geq\delta} \left\{\frac{1}{ng}\sum_{i=1}^{ng}(\lambda_i - \log\lambda_i - 1)\right\} \geq \eta,$$

for all $n \geq N$, where $\tau = \left(\theta_2', \ \theta_3'\right)' \in \Theta_2 \times \Theta_3$.

Assumption B6 *(i) $\{x_{it}\}$ and $\{\varepsilon_{it}\}$ are independent. (ii) Let $X^* = SX$ and hence $X_t^* = S_tX_t$. As $n \to \infty$,*

$$\frac{1}{ng}\begin{pmatrix} X_s^{*\prime} \\ X_s^{*\prime}G_s' \end{pmatrix}\begin{pmatrix} X_t^* & G_tX_t^* \end{pmatrix} \to_p \begin{pmatrix} Q_{11}^{st} & Q_{12}^{st} \\ Q_{21}^{st} & Q_{22}^{st} \end{pmatrix},$$

and

$$\frac{1}{ng}\begin{pmatrix} (X^*)'\left(\Sigma_0^{-1}\otimes I_n\right)X^* & (X^*)'\left(\Sigma_0^{-1}\otimes I_n\right)GX^* \\ (X^*)'G'\left(\Sigma_0^{-1}\otimes I_n\right)X^* & (X^*)'G'\left(\Sigma_0^{-1}\otimes I_n\right)GX^* \end{pmatrix} \to_p \begin{pmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{pmatrix},$$

where O_{11} is p.d..

Let v_{ij} be the (i,j)-th element of the $G'G$, $(x_{it}^*)'$ be the i-th row of X_t^*, $(x_{it}^{**})'$ be the i-th row of $X_t^{**} = G_t X_t^*$ and u_{it}^* be the t-th element of $u_{\cdot t}^* = G_t u_{\cdot t}$.

Assumption B7 *As $n \to \infty$,*

$$\sum_{i=1}^{ng}\sum_{j=1}^{ng} v_{ij}^2 = o\left(n^2\right)$$

and for any $s,\ t = 1,\ ...,\ g$,

$$\begin{aligned} \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left(x_{is}^*\left(x_{jt}^*\right)'\right) Cov\left(u_{is}^*,\ u_{jt}^*\right) &= o\left(n^2\right), \\ \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left(x_{is}^{**}\left(x_{jt}^{**}\right)'\right) Cov\left(u_{is}^*,\ u_{jt}^*\right) &= o\left(n^2\right). \end{aligned}$$

Theorem B *Under Assumptions B1 and B3-B7, $\widehat{\theta} \to_p \theta_0$.*

Assumption B4 on Θ_2 may appear to be quite restrictive. However, (2.11) shows that without the assumption on positive definiteness of Σ, an unconstrained optimizer for θ_2 gives $\widehat{\Sigma}$ that is always positive semidefinite and usually positive definite in finite samples. Hence Assumption B4 is not really a practical issue.

Assumption B5 is quite common in multivariate analysis. Consider a function $f : \mathbb{R}^+ \to \mathbb{R}$ where $f(x) = x - \log x - 1$. We have that $f(x) > 0$ for all $x > 0$ except $x = 1$ and $f(1) = 0$. Hence $\frac{1}{ng}\sum_{i=1}^{ng}(\lambda_i - \log\lambda_i - 1) = 0$ if and only if $\lambda_i = 1$ for $i = 1,\ ...,\ n$. This is equivalent to $H(\theta_2,\ \theta_3) = I_n$. Note that $H(\theta_{02},\ \theta_{03}) = I_n$. Assumption B5 essentially states that when τ is sufficiently different from τ_0, $H(\theta_2,\ \theta_3)$ is sufficiently different from $H(\theta_{02},\ \theta_{03})$. There is one technical issue with Assumption B5. There may be some $\rho = \theta_3 \in \Theta_3$ such that $S(\rho)$ is singular. In this case, there is $\lambda_i = 0$, where λ_i is an eigenvalue of $H(\theta_2,\ \theta_3)$, and hence $\log\lambda_i$ is not generally well defined. However, we employ the rule stated earlier that we define $\log x = -\infty$ if $x = 0$. It is worth noting that Assumption B5 is for asymptotic identification of θ_2 and θ_3.

Assumption B6 is similar to the one made in Lee (2004). Since we consider a somewhat different linear model from the one considered in Lee (2004), it turns out that only positive definiteness of O_{11} plays a role in asymptotic identification of θ_1.

As can be seen in the proof of Theorem B, we can easily avoid considering the term $S(\rho)^{-1}$. Therefore we do not need to impose any assumption on Θ_3 so that $S(\rho)$ is invertible for any $\rho \in \Theta_3$ as in the literature. With reference to (2.9) any value of ρ making $S(\rho)$ singular cannot be a minimizer of $Q_n(\theta)$ since we set $\log x = -\infty$ for $x = 0$. Hence these values of ρ will be automatically removed from the "effective" parameter space when doing an optimization problem.

The most important point to be noted here is that we do not assume any bound on $\left\|S(\rho)^{-1}\right\|$ uniformly in Θ_3 as in the literature. The discussion at the end of Section 2 implies that Assumption B7 allows $\left\|S^{-1}\right\|$ and hence $\|Var(u)\|$ to be explosive but at an appropriate rate. Since $\left\|S^{-1}\right\|$ and G depends only on ρ_0, our assumptions are imposed on the true value ρ_0 but not the whole parameter space of ρ_0. As discussed at the end of Section 2, G is essentially S^{-1}. In a univariate case, since $Var(u) = \sigma_0^2\left(\left(S^{-1}\right)' S^{-1}\right)$, in the presence of spatial dependence, $G'G$ is essentially $Var(u)$. Hence Assumption B7 allows the variance matrix of u to be explosive but the rate at which it becomes explosive cannot be too fast. Moreover, in the presence of spatial dependence, i.e. $\rho_{0t} \neq 0$, $G_t = \rho_{0t}^{-1}\left(S_t^{-1} - I_n\right)$. Then

$$X_t^{**} = G_t X_t^* = \rho_{0t}^{-1} X_t - \rho_{0t}^{-1} X_t^* \text{ and } u_{\cdot t}^* = G_t \varepsilon_{\cdot t} = \rho_{0t}^{-1} u_{\cdot t} - \rho_{0t}^{-1} \varepsilon_{\cdot t}.$$

Hence the latter part of Assumption B7 is equivalent to

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}\left(x_{is}\left(x_{jt}\right)'\right) Cov\left(u_{is},\ u_{jt}\right) &= o\left(n^2\right), \\
\sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}\left(x_{is}^*\left(x_{jt}^*\right)'\right) Cov\left(u_{is},\ u_{jt}\right) &= o\left(n^2\right).
\end{aligned}$$

This is the limit on the joint explosive behaviour of the regressors and disturbances.

2.3.2 Asymptotic Normality

First we introduce notations employed in this part. For $t = 1, \ ..., \ g$, recall that X_t^* is defined as $S_t X_t$. Let $\left(x_{it}^*\right)'$ be the i-th row of X_t^*. For $s, \ t = 1, \ ..., \ g$, let σ_0^{st} and σ^{st} be the (s,t)-th element of Σ_0^{-1} and Σ^{-1}, respectively. Similarly, let σ_{0st} and σ_{st} be the (s,t)-th element of Σ_0 and Σ, respectively. Define $u_{\cdot t}(\beta) = y_{\cdot t} - X_t\beta$ and $\sigma_{st}(\theta) = u_{\cdot s}(\beta)' S_s(\rho)' S_t(\rho) u_{\cdot t}(\beta)$. Let $\Sigma(\theta)$ be the square matrix whose (s,t)-th element is $\sigma_{st}(\theta)$. For $t = 1, \ ..., \ g$, let g_{ijt} be the (i,j)-th element of G_t. For any value of ρ such that $S_t(\rho)$ is non-singular, define $G_t(\rho) = W_t S_t^{-1}(\rho)$ and $G(\rho) = diag\left(G_1(\rho), \ ..., \ G_g(\rho)\right)$. Following Abadir and Magnus (2005), for any symmetric matrix A of order n, define the duplication matrix D_n as the $n^2 \times n(n+1)/2$ matrix such that $D_n vech(A) = vec(A)$. Define δ_{st} as the Kronecker's delta.

The First Derivatives Consider θ in a neighbourhood of θ_0 such that $S_t(\rho)$ are non-singular for $t = 1, \ ..., \ g$. Recall that we impose the assumption that Σ and Σ^{-1} are symmetric. With reference to Lemmas C1 and C2, the first derivatives are

$$\frac{\partial Q_n(\theta)}{\partial \theta_1} = -\frac{1}{ng} X' S(\rho)' \left(\Sigma^{-1} \otimes I_n\right) S(\rho) u(\beta), \tag{2.12}$$

for $s \geq t$,

$$\frac{\partial Q_n(\theta)}{\partial \sigma^{st}} = -\frac{2-\delta_{st}}{2g}\sigma_{st} + \frac{2-\delta_{st}}{2ng}\sigma_{st}(\theta), \tag{2.13}$$

and for $t = 1, \ ..., \ g$,

$$\frac{\partial Q_n(\theta)}{\partial \theta_{3t}} = \frac{1}{ng} tr\left\{G_t(\rho)\right\} - \frac{1}{ng}\sum_{s=1}^{g} \sigma^{st} u_{\cdot s}(\beta)' S_s(\rho)' W_t u_{\cdot t}(\beta), \tag{2.14}$$

where θ_{3t} is the t-th element of θ_3. Note that (2.13) can be written in a matrix form as

$$\frac{\partial Q_n(\theta)}{\partial \theta_2} = -\frac{1}{2g} D_g' vec(\Sigma) + \frac{1}{2ng} D_g' vec(\Sigma(\theta)), \tag{2.15}$$

where D_g is the duplication matrix.

It follows from (2.12), (2.15) and (2.14) that

$$\frac{\partial Q_n(\theta_0)}{\partial \theta_1} = -\frac{1}{ng}(X^*)' \left(\Sigma_0^{-1} \otimes I_n\right) \varepsilon,$$

$$\frac{\partial Q_n(\theta_0)}{\partial \theta_2} = -\frac{1}{2g} D_g' vec(\Sigma_0) + \frac{1}{2ng} D_g' vec \begin{pmatrix} \varepsilon_{\cdot 1}'\varepsilon_{\cdot 1} & \cdots & \varepsilon_{\cdot 1}'\varepsilon_{\cdot g} \\ \vdots & \ddots & \vdots \\ \varepsilon_{\cdot g}'\varepsilon_{\cdot 1} & \cdots & \varepsilon_{\cdot g}'\varepsilon_{\cdot g} \end{pmatrix},$$

and for $t = 1, \ ..., \ g$,

$$\frac{\partial Q_n(\theta_0)}{\partial \theta_{3t}} = \frac{1}{ng} tr(G_t) - \frac{1}{ng}\sum_{s=1}^{g} \sigma_0^{st} \varepsilon_{\cdot s}' G_t \varepsilon_{\cdot t}.$$

Let

$$\begin{aligned} \Omega_n &= ng\mathbb{E}\left(\frac{\partial Q_n(\theta_0)}{\partial \theta}\frac{\partial Q_n(\theta_0)}{\partial \theta'}\bigg| X\right) \\ &= \begin{pmatrix} \Omega_{11,n} & \Omega_{12,n} & \Omega_{13,n} \\ \Omega_{21,n} & \Omega_{22,n} & \Omega_{23,n} \\ \Omega_{31,n} & \Omega_{32,n} & \Omega_{33,n} \end{pmatrix}, \end{aligned}$$

where $\Omega_{ij,n} = \mathbb{E}\left(\frac{\partial Q_n(\theta_0)}{\partial \theta_i}\frac{\partial Q_n(\theta_0)}{\partial \theta_j'}\Big| X\right)$.

By Assumption B6,

$$\Omega_{11,n} = \frac{1}{ng}(X^*)' \left(\Sigma_0^{-1} \otimes I_n\right) X^* \to_p O_{11},$$

where O_{11} is p.d..

By Lemma C3, each column of $\Omega_{12,n}$ is a multiple of

$$\frac{1}{ng}\sum_{s=1}^{g}\sum_{t=1}^{g}\sigma_0^{st}\sum_{i=1}^{n}x_{is}^{*}\mathbb{E}\left\{\varepsilon_{it}\left(\varepsilon_{iu}\varepsilon_{iv}-\sigma_{0uv}\right)\right\}.$$

Lemma C4 implies that the τ-th column of $\Omega_{13,n}$ is

$$\frac{1}{ng}\sum_{s=1}^{g}\sum_{t=1}^{g}\sum_{u=1}^{g}\sigma_0^{st}\sigma_0^{u\tau}\sum_{i=1}^{n}x_{is}^{*}g_{ii\tau}\mathbb{E}\left(\varepsilon_{it}\varepsilon_{iu}\varepsilon_{i\tau}\right).$$

By Lemma C5, each element of $\Omega_{22,n}$ is a multiple of

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\varepsilon_{is}\varepsilon_{it}\varepsilon_{iu}\varepsilon_{iv}-\sigma_{0st}\sigma_{0uv}\right).$$

Lemma C6 implies that each element of $\Omega_{23,n}$ is a multiple of

$$\frac{1}{n}\sum_{u=1}^{g}\sigma_0^{u\tau}\sum_{i=1}^{n}g_{ii\tau}\mathbb{E}\left(\varepsilon_{is}\varepsilon_{it}\varepsilon_{iu}\varepsilon_{i\tau}\right)-\frac{1}{n}tr\left(G_\tau\right)\sigma_{0st}.$$

Finally, by Lemma C7, the (τ,t)-th element of $\Omega_{33,n}$ is

$$\frac{1}{ng}\sum_{u=1}^{g}\sum_{s=1}^{g}\sigma_0^{u\tau}\sigma_0^{st}\sum_{i=1}^{n}g_{ii\tau}g_{iit}\left\{\mathbb{E}\left(\varepsilon_{iu}\varepsilon_{i\tau}\varepsilon_{is}\varepsilon_{it}\right)-\sigma_{0u\tau}\sigma_{0st}-\sigma_{0us}\sigma_{0\tau t}-\sigma_{0ut}\sigma_{0\tau s}\right\}$$
$$+\frac{1}{ng}tr\left(G_\tau' G_t\right)\sum_{u=1}^{g}\sum_{s=1}^{g}\sigma_0^{u\tau}\sigma_0^{st}\sigma_{0us}\sigma_{0\tau t}+\frac{1}{ng}tr\left(G_\tau G_t\right)\sum_{u=1}^{g}\sum_{s=1}^{g}\sigma_0^{u\tau}\sigma_0^{st}\sigma_{0ut}\sigma_{0\tau s}. \quad (2.16)$$

If $\mathbb{E}\left(\varepsilon_{iu}\varepsilon_{i\tau}\varepsilon_{is}\varepsilon_{it}\right)=\mathbb{E}\left(\varepsilon_{ju}\varepsilon_{j\tau}\varepsilon_{js}\varepsilon_{jt}\right)$ for all $i\neq j$, then the first term in (2.16) becomes

$$\left(\frac{1}{ng}\sum_{i=1}^{n}g_{ii\tau}g_{iit}\right)\sum_{u=1}^{g}\sum_{s=1}^{g}\sigma_0^{u\tau}\sigma_0^{st}\kappa\left(u,\tau,s,t\right),$$

where κ is the fourth cumulant. Under mild assumptions, it can be shown that all submatrices of Ω_n are convergent in probability. Hence, we make the following assumption.

Assumption C1 *As $n\to\infty$, $\Omega_n\to_p\Omega$ where Ω is a positive definite matrix.*

The Second Derivatives Now consider the second derivatives of $Q_n(\theta)$. With reference to Lemmas C8, C9 and C10,

$$\begin{aligned}
\frac{\partial^2 Q_n(\theta)}{\partial\theta_1\partial\theta_1'} &= \frac{1}{ng}X'S(\rho)'\left(\Sigma^{-1}\otimes I_n\right)S(\rho)X, \\
\frac{\partial^2 Q_n(\theta)}{\partial\theta_2\partial\theta_2'} &= \frac{1}{2g}D_g'(\Sigma\otimes\Sigma)D_g \\
\frac{\partial^2 Q_n(\theta)}{\partial\theta_1\partial\rho_t} &= \frac{1}{ng}\sum_{s=1}^{g}\left\{\sigma^{st}X_s'S_s(\rho)'W_t u_{\cdot t}(\beta)+\sigma^{ts}X_t'W_t'S_s(\rho)u_{\cdot s}(\beta)\right\}, \\
\frac{\partial^2 Q_n(\theta)}{\partial\rho_s\partial\rho_t} &= \frac{1}{ng}tr\left\{G_t(\rho)^2\right\}\delta_{st}+\frac{1}{ng}\sigma^{st}u_{\cdot s}(\beta)'W_s'W_t u_{\cdot t}(\beta),
\end{aligned}$$

for $s\geq t$,

$$\begin{aligned}
\frac{\partial^2 Q_n(\theta)}{\partial\theta_1\partial\sigma^{st}} &= -\frac{1}{ng}\left\{X_s'S_s(\rho)'S_t(\rho)u_{\cdot t}(\beta)+X_t'S_t(\rho)'S_s(\rho)u_{\cdot s}(\beta)(1-\delta_{st})\right\}, \\
\frac{\partial^2 Q_n(\theta)}{\partial\sigma^{tt}\partial\rho_t} &= -\frac{1}{ng}u_{\cdot t}(\beta)'W_t'S_t(\rho)u_{\cdot t}(\beta),
\end{aligned}$$

and for $s>t$,

$$\frac{\partial^2 Q_n(\theta)}{\partial\sigma^{st}\partial\rho_\tau} = -\frac{1}{ng}\left\{u_{\cdot\tau}(\beta)'W_\tau'S_t(\rho)u_{\cdot t}(\beta)\delta_{\tau t}+u_{\cdot s}(\beta)'S_s(\rho)'W_\tau u_{\cdot\tau}(\beta)\delta_{\tau s}\right\},$$

Assumption C2 *For* $s,\ t=1,\ ...,\ g,\ \lim_{n\to\infty}n^{-1}tr\{G_t\},\ \lim_{n\to\infty}n^{-1}tr\left\{G_t^2\right\}$ *and* $\lim_{n\to\infty}n^{-1}tr\{G_s'G_t\}$ *exist.*

Assumption C3 *For* $t=1,\ ...,\ g,$ *suppose* $\overline{\rho}-\rho_0=o_p(1)$ *as* $n\to\infty$, *then*

$$n^{-1}tr\left\{G_t(\overline{\rho})^2\right\}-n^{-1}tr\left\{G_t^2\right\}=o_p(1).$$

Under Assumptions B1, B3-B7, and C2-C3, Lemmas C11-C15, if $\overline{\theta}-\theta_0=o_p(1)$, then

$$\frac{\partial^2 Q_n(\overline{\theta})}{\partial\theta\partial\theta'}\to_p\begin{pmatrix}O_{11} & 0 & 0\\ 0 & (2g)^{-1}D_g'(\Sigma_0\otimes\Sigma_0)D_g & E_2'\\ 0 & E_2 & E_1\end{pmatrix}=E, \tag{2.17}$$

where

$$E_1=\lim_{n\to\infty}\frac{1}{ng}\left\{\begin{pmatrix}tr\left(G_1^2\right) & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & tr\left(G_g^2\right)\end{pmatrix}+\begin{pmatrix}\sigma_{011}\sigma_0^{11}tr\left(G_1'G_1\right) & \cdots & \sigma_{01g}\sigma_0^{1g}tr\left(G_1'G_g\right)\\ \vdots & \ddots & \vdots\\ \sigma_{0g1}\sigma_0^{g1}tr\left(G_g'G_1\right) & \cdots & \sigma_{0gg}\sigma_0^{gg}tr\left(G_g'G_g\right)\end{pmatrix}\right\}, \tag{2.18}$$

and E_2 are $g \times (g+1)\,g/2$ matrix whose elements correspond to: for $s > t$, $\partial^2 Q_n\left(\overline{\theta}\right)/\partial\sigma^{st}\partial\rho_\tau$ correspond to

$$-\lim_{n\to\infty}\left[\frac{\sigma_{\tau t}}{ng}tr\left(G_\tau\right)\delta_{\tau t}-\frac{\sigma_{\tau s}}{ng}tr\left(G_\tau\right)\delta_{\tau s}\right]$$

and $\partial^2 Q_n\left(\overline{\theta}\right)/\partial\sigma^{tt}\partial\rho_\tau$ correspond to

$$-\lim_{n\to\infty}\left[\frac{\sigma_{tt}}{ng}tr\left(G_t\right)\delta_{\tau t}\right].$$

By Assumptions B3 and B6, O_{11} and Σ_0 are p.d.. Exercise 11.34 (a) in Abadir and Magnus (2005) implies that $D_g'\left(\Sigma_0\otimes\Sigma_0\right)D_g$ is p.d.. Since the second term in (2.18) is the limit of

$$(ng)^{-1}\left[diag\left\{W_1u_{\cdot 1}\left(\overline{\beta}\right),\ ...,\ W_gu_{\cdot g}\left(\overline{\beta}\right)\right\}\right]'\left(\overline{\Sigma}^{-1}\otimes I_n\right)diag\left\{W_1u_{\cdot 1}\left(\overline{\beta}\right),\ ...,\ W_gu_{\cdot g}\left(\overline{\beta}\right)\right\}$$

that is p.s.d., Lemma B9 implies that it must be p.s.d.. The assumption $\lim_{n\to\infty}n^{-1}tr\left(G_t^2\right)>0$ for all $t=1,\ ...,\ g$, implies that E_1 is p.d.. To avoid the complication of showing that $E_1-E_2\left\{(2g)^{-1}D_g'\left(\Sigma_0\otimes\Sigma_0\right)D_g\right\}^{-1}E_2'$ is p.d., so that E is p.d., we make the following assumption.

Assumption C4 *The matrix E defined in (2.17) is positive definite.*

Assumption C5 *θ_0 is an interior point of Θ.*

Assumption C6 *Let g_{ijt} be the (i,j)-th element of G_t. As $n\to\infty$,*

$$\begin{aligned}\sum_{i=1}^n\left(\sum_{j=1}^n g_{ijt}^2\right)^2 &= o\left(n^2\right),\\ \max_{1\le t\le g}\max_{1\le j\le n}\sum_{i=1}^n\left|g_{ijt}\right|+\max_{1\le t\le g}\max_{1\le i\le n}\sum_{j=1}^n\left|g_{ijt}\right| &= o\left(n^{1/2}\right).\end{aligned}$$

Assumption C7 *Recall that $\left(x_{it}^*\right)'$ is the i-th row of X_t^*. There exists $\delta>0$ such that (i) there is a finite constant C such that*

$$\max_{1\le t\le g}\max_{i\ge 1}\mathbb{E}\left|\varepsilon_{it}\right|^{4+\delta}\le C;$$

(ii) as $n\to\infty$,

$$n^{-1}\sum_{t=1}^g\sum_{i=1}^n\left\|x_{it}^*\right\|^{2+\delta}=O_p\left(1\right);$$

and (iii)

$$n^{-1}\sum_{t=1}^{g}\sum_{i=1}^{n}\left|g_{iit}\right|^{2+\delta}=O\left(1\right).$$

Theorem C *Under Assumptions B1, B3-B7 and C1-C7, as $n\to\infty$,*

$$\sqrt{ng}\left(\widehat{\theta}-\theta_0\right)\to_d N\left(0,\ E^{-1}\Omega E^{-1}\right).$$

Remark Assumption C3 is imposed so that we can avoid making arbitrary assumptions on W. For a square matrix A, let $\lambda_i\left(A\right)$ be an eigenvalue of A. It can be shown that

$$n^{-1}tr\left\{G_t\left(\overline{\rho}\right)^2\right\}-n^{-1}tr\left\{G_t^2\right\}=n^{-1}\sum_{i=1}^{n}\left\{\lambda_i\left(G_t\left(\overline{\rho}\right)\right)\right\}^2-\left\{\lambda_i\left(G_t\right)\right\}^2.$$

Note that if A is an invertible matrix, then λ^{-1} is an eigenvalue of A^{-1} if and only if λ is an eigenvalue of A. By Lemmas A4 and A6, for $\rho_0\neq 0$ and $\overline{\rho}$ sufficiently near ρ_0,

$$\lambda_i\left(G_t\left(\overline{\rho}\right)\right)=\frac{1}{\overline{\rho}}\left(\left(1-\overline{\rho}\omega_i\right)^{-1}-1\right)=\frac{\omega_i}{1-\overline{\rho}\omega_i},$$

where ω_i are eigenvalues of W. Similarly

$$\lambda_i\left(G_t\right)=\frac{\omega_i}{1-\rho_0\omega_i}.$$

The convergence in Assumption C3 depends on the behaviour of ω_i, particularly on how many of ω_i and how fast ω_i get close to $1/\rho_0$ as $n\to\infty$. Clearly, if $\left\|S^{-1}\right\|$ is uniformly bounded as commonly assumed in the literature, it is rather straight forward to show that Assumption C3 holds. For the case where $\rho_0=0$, it is trivial since there is no spatial dependence.

Remark For a univariate case, it is very simple to show that Assumption C4 holds under more primitive assumptions. If $g=1$, then

$$E=\begin{pmatrix} O_{11} & 0 & 0\\ 0 & \frac{1}{2}\sigma_0^4 & -\frac{\sigma_0^2}{n}tr\left(G\right)\\ 0 & -\frac{\sigma_0^2}{n}tr\left(G\right) & \frac{1}{n}tr\left(G^2\right)+\frac{1}{n}tr\left(G'G\right)\end{pmatrix}.$$

Assuming that $\lim_{n\to\infty}n^{-1}tr\left(G'G\right)>0$, then a necessary and sufficient condition for E to be p.d. is that

$$\frac{1}{n}tr\left(G^2\right)+\frac{1}{n}tr\left(G'G\right)-2\left\{\frac{1}{n}tr\left(G\right)\right\}^2>0.$$

By Schur's inequality, $tr\left(G^2\right)\leq tr\left(G'G\right)$, a sufficient condition for this to hold is that

$\left\{n^{-1}tr\left(G\right)\right\}^2 < n^{-1}tr\left(G^2\right)$. Note that, the Cauchy's inequality,

$$\left\{n^{-1}tr\left(G\right)\right\}^2 = n^{-2}\left(\sum_{i=1}^{n}\lambda_i\right)^2 \leq n^{-1}\sum_{i=1}^{n}\lambda_i^2 = n^{-1}tr\left(G^2\right),$$

where λ_i are eigenvalues of G. The equality holds if and only if $\lambda_1 = \lambda_2 = \cdots = \lambda_n$.

Remark Since $tr\left(G_t'G_t\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} g_{ijt}^2$, Assumption C2 implies that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ijt}^2 = O\left(n\right).$$

This observation makes Assumption C6 analogous to a more familiar assumption in time series literature that

$$\max_{1\leq t\leq g}\max_{1\leq j\leq n}\sum_{i=1}^{n}\left|g_{ijt}\right| + \max_{1\leq t\leq g}\max_{1\leq i\leq n}\sum_{j=1}^{n}\left|g_{ijt}\right| = o\left(\left(\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ijt}^2\right)^{1/2}\right).$$

Recall again that, in the presence of spatial dependence, $G_t = \rho_{0t}^{-1}\left(S_t^{-1} - I_n\right)$. Hence G_t is essentially $\rho_{0t}^{-1}S_t^{-1}$ and elements of G_t directly controls the degree of spatial dependence of $u_{\cdot t}$ since $Var\left(u_{\cdot t}\right)$ is essentially $G_t'G_t$. Finally, Assumption C7 is for the Lyapunov condition.

2.4 Nonparametric Series Estimation

Before discussing how to obtain efficiency improvement, we first discuss nonparametric series estimation. The reason is that in order to obtain an estimate that is adaptive in the sense of Stein (1956), one needs to nonparametrically estimate the unknown score function. Our choice of nonparametric series estimation over the kernel estimation is based on the advantage that no trimming is required.

Consider a nonparametric model

$$\mathbb{E}\left(\left. y_i\right| x_i\right) = h\left(x_i\right), \quad i = 1,\ ...,\ n,$$

where x_i are $\mathbb{R}^g$-valued random variables. The main interest in series estimation literature is to nonparametrically approximate the unknown function h by a linear combination of approximating functions $p_1,\ ...,\ p_L$, $\widehat{h} = \sum_{l=1}^{L} c_l p_l$. The effectiveness of such approximation depends on the choices of approximating functions and coefficients c_l. For a given a family approximating functions $\{p_l\}$, the simplest way to choose the coefficients is to perform least squares regression of y_i on $p_l\left(x_i\right)$. To explicitly illustrate the idea, let

$$p^L\left(x\right) = \left(p_1\left(x\right),\ ...,\ p_L\left(x\right)\right)'$$

be a vector of approximating functions. The least squares approximation of h by approximating functions $p_1,\ ...,\ p_L$ at a point x is

$$\widehat{h}_L(x) = p^L(x)' \widehat{\gamma}_L, \tag{2.19}$$

where

$$\widehat{\gamma}_L = (P'P)^{-1} P'y, \quad P = \left(p^L(x_1),\ ...,\ p^L(x_n)\right)' \ \text{ and } \ y = (y_1,\ ...,\ y_n)'. \tag{2.20}$$

The main interest in the literature focuses on precision of such approximation with some families of approximating functions such as trigonometric functions, polynomials and regression splines, when the number of approximating functions L is allowed to become arbitrarily large as the sample size n increases. The precision is commonly evaluated from the mean-square and uniform convergence perspective where the rate of convergence is often of a main interest.

For clarity of the discussion, we introduce two assumptions.

Assumption D1 $\left\{(x_i',\ y_i)'\right\}$ *is an i.i.d. sequence of* $\mathbb{R}^{K+1}$*-valued random variables.*

Assumption D2 $\mathbb{E}\left\{h(x)^2\right\} < \infty$, *where x has the same distribution as x_i.*

For some semiparametric models including the one to be discussed in this chapter, the mean-square convergence of the type

$$\int \left[h(x) - \widehat{h}_L(x)\right]^2 \, dF_X(x) \to 0 \text{ as } L \to \infty, \tag{2.21}$$

where F_X is the distribution function of x, is sufficient to show required asymptotic properties as long as the first-order asymptotic is concerned. Our choice of a family of approximating functions is the polynomial type. Practitioners, particularly those from economic background, may find this choice of approximating functions natural and intuitive. With polynomials, the expression in (2.19) can be interpreted as a Taylor approximation of an unknown function h. Moreover, Newey (1988) showed that a series estimate employing polynomials arises naturally from GMM estimation.

Let $\mathbb{N}_0$ denote the set of nonnegative integers. A multi-index is denoted by $\lambda = (\lambda_1,\ ...,\ \lambda_g)' \in \mathbb{N}_0^g$ with norm $\|\lambda\|_1 = \sum_{t=1}^g |\lambda_t|$. For $\lambda \in \mathbb{N}_0^g$ and $x = (x_1,\ ...,\ x_g)' \in \mathbb{R}^g$, a monomial in variables $x_1,\ ...,\ x_g$ is a product

$$x^\lambda = x_1^{\lambda_1} ... \ x_g^{\lambda_g}.$$

The number $\|\lambda\|_1$ is the total degree of x^λ. A polynomial $p : \mathbb{R}^g \to \mathbb{R}$ in g variables is a linear combination of monomials

$$p(x) = \sum_\lambda c_\lambda x^\lambda,$$

where $c_\lambda \in \mathbb{R}$. The degree of a polynomial is defined as the highest total degree of its monomials. Denote the collection of polynomials in g variables by Π^g.

Let f_X be the probability density function of x and $\mathcal{X}$ be the support of x, i.e. $f_X(y) > 0$ for all $y \in \mathcal{X}$. Newey (1997) showed that under the assumptions that $\mathcal{X}$ is the Euclidean product of compact intervals on which f_X is bounded away from zero, and that h is continuously differentiable of order s on $\mathcal{X}$ (2.21) is $O\left(L/n + L^{-2s/g}\right)$ as $n \to \infty$. However, the assumption on f_X is restrictive since his results are not applicable to most well-known random variables such as the normal, Student's t, exponential and chi-squared random variables.

The main objective of this section is to relax Newey (1997)'s distributional assumption to allow for unbounded $\mathcal{X}$. When $\mathcal{X}$ is unbounded, in some circumstances a sharp result as in Newey (1997) may not be achievable since $|h(x)|$ may become arbitrarily large as $\|x\|$ goes to infinity. As mentioned earlier, in many semiparametric applications, including the one in this section, the convergence of the type (2.21) without the knowledge of the rate of convergence is sufficient as long as first-order asymptotic is concerned. Hence, we first discuss how to obtain (2.21) in a general case and later discuss how to be more precise about the rate of convergence.

The possibility for (2.21) to hold arises from the following result extended from the theorem for a univariate case in Freud (1971) that was employed in Newey (1988) and Robinson (2005, 2010). Let $\mathcal{M} = \mathcal{M}(\mathbb{R}^g)$ denote the set of nonnegative Borel measures on $\mathbb{R}^g$ having moments of all orders, i.e. if $\mu \in \mathcal{M}$, then

$$\int_{\mathbb{R}^g} \left|x^\lambda\right| \, d\mu(x) < \infty \quad \text{for all } \lambda \in \mathbb{N}_0^g. \tag{2.22}$$

Denote the class of square integrable functions with respect to a measure μ by $L^2(\mu)$, i.e. $f \in L^2(\mu)$ if and only if $\int_{\mathbb{R}^g} |f|^2 \, d\mu < \infty$, . Theorem 3.1.18 in Dunkl and Xu (2001) states that if $\mu \in \mathcal{M}$ satisfies

$$\int_{\mathbb{R}^g} \exp(c\|x\|) \, d\mu(x) < \infty, \tag{2.23}$$

where $\|\cdot\|$ is the Euclidean norm, for some constant $c > 0$, then the space of polynomial Π^g is dense in $L^2(\mu)$.

The distribution of x can generally be an issue for series estimation when polynomials are employed. First, higher order moments of x may not exist if the distribution of x has fat tails. Hence, (2.21) is not well defined. In addition, in order to employ the approximation result from Dunkl and Xu (2001), it is required that the moment of x must exist for all order. This problem can be overcome by employing polynomials in $\xi = T(x)$ rather than polynomial in x where $T : \mathbb{R}^g \to \mathbb{R}^g$ is a one-to-one bounded transformation such that $\|\xi\|$ is bounded. When $\widehat{h}_L(x)$ is replaced by $\widehat{h}_L(T(x))$, it follows that the mean-square criterion in (2.21) becomes

$$\int \left[h\left(T^{-1}(\xi)\right) - \widehat{h}_L(\xi)\right]^2 \, dF_\xi(\xi) \to 0 \text{ as } L \to \infty, \tag{2.24}$$

where $\xi = T(x)$, $\widehat{h}_L(\xi)$ is a polynomial in ξ and F_ξ is the distribution of ξ. Clearly the composite function $h \circ T^{-1}$ is in $L^2(F_\xi)$ and F_ξ satisfies (2.22) and (2.23) when $\|\xi\|$ is bounded.

Hence we employ approximating functions

$$p_l(T(x)) = T(x)^{\lambda(l)},$$

where $\{\lambda(l)\}$ is a sequence of distinct multi-indices. It is crucial to assume that the sequence $\{\lambda(l)\}_{l=1}^{\infty}$ of distinct multi-indices must include all distinct multi-indices. Moreover, it is assumed that the sequence $\{\lambda(l)\}$ is ordered so that $\|\lambda(l)\|_1 = \sum_{t=1}^{g} |\lambda_t(l)|$ is monotonically increasing.

Another problem arises from the fact that when the regression in (2.20) is employed to choose the coefficients c_λ for the polynomials, multi-colinearity of the approximating functions $p_l(T(x_i))$ may become an issue for a wide class of distribution, particularly as $L \to \infty$. This is precisely the problem faced by Newey (1988) and Robinson (2005, 2010). Under a general distributional assumption, Newey (1988) had to assume that $L \log(L) / \log(n) \to 0$ as $n \to \infty$. Robinson (2005) relaxed this slow rate of L slightly. This problem of multi-colinearity is particularly serious for x with high dimensions since one cannot employ many approximating functions as restricted by the rate of growth of L.

One effective way to get around with the multi-colinearity problem was proposed in Cox (1988). Cox (1988) pointed out that when polynomials are employed as approximating functions, $\widehat{h}_L(\xi)$ computed from polynomials in ξ will be numerically the same as that from orthonormal polynomials in ξ, with respect to some weight functions, of order corresponding to components of $\lambda(l)$. Newey (1997) employed this advantage to show that when the support $\mathcal{X}$ of x is bounded and f_X is bounded away from zero on $\mathcal{X}$, the appropriate orthonormal polynomials which could get rid of multi-colinearity is the Jacobi polynomials with respect to the uniform weight.

It turns out that a certain class of transformations T can play a crucial role in allowing for unbounded $\mathcal{X}$. Before discussing an appropriate class of transformations T, we first discuss an analogous result to that of Newey (1997).

Assumption D3 *There exists a bounded and one-one transformation $T : \mathbb{R}^g \to \mathbb{R}^g$ such that $\xi = T(x)$ where the support of ξ is the Cartesian product of bounded open intervals $\Pi_{t=1}^{g}(a_t,\ b_t)$ on which the probability density function of ξ is bounded away from zero almost everywhere, i.e. there is a constant C such that $f_\xi(\xi) \geq C > 0$ for all $\xi \in \Pi_{t=1}^{g}(a_t,\ b_t)$ except for ξ in a null set, where f_ξ is the density function of ξ.*

Assumption D4 *$Var(y_i|\, x_i)$ is bounded.*

Assumption D5 *As $n \to \infty$, $L^3/n \to 0$.*

Theorem D1 *Under Assumptions D1-D5, as $n \to \infty$,*

$$\int \left[h(x) - \widehat{h}_L(T(x))\right]^2 \ dF_X(x) = o(1). \tag{2.25}$$

Let $h_1 = h \circ T^{-1}$. Suppose it is known that h_1 is continuously differentiable of order v on the support $\Pi_{t=1}^{g}(a_t,\ b_t)$ of ξ. Suppose further that the following assumption holds.

Assumption D6 *It is possible to extend $h_1 : \Pi_{t=1}^{g}(a_t,\ b_t) \to \mathbb{R}$ to $h_2 : \Pi_{t=1}^{g}[a_t,\ b_t] \to \mathbb{R}$ where h_2 is also continuously differentiable of order v on $\Pi_{t=1}^{g}[a_t,\ b_t]$.*

Then we can be more precise regarding the rate of converge.

Theorem D2 *Under Assumptions D1-D6, as $n \to \infty$,*

$$\int \left[h(x) - \widehat{h}_L(T(x))\right]^2 \ dF_X(x) = O_p\left(L/n + L^{-2v/g}\right). \tag{2.26}$$

Theorem D2 is essentially the same as the first part of Theorem 4 in Newey (1997). Assumption D6 is the same as Assumption 9 in Newey (1997) so that we do not have to rely on the approximating result in L^2 space from Dunkl and Xu (2001). One necessary condition for Assumption D6 to hold is that $h : \mathbb{R}^g \to \mathbb{R}$ is a continuously differentiable of order v on $\mathbb{R}^g$ and that h must be bounded. Then under some conditions, it is possible to extend h_1 to satisfy Assumption D6. Hints for sufficient conditions for Assumption D6 may be seen from the discussion of the transformation T later in this section. Obviously, T must be smooth enough for h_2 to be smooth. An example of unknown function satisfying Assumption D6 may arise from applications where it is required to estimate an unknown distribution function of vectors of random variables such as in the semiparametric index models.

As Theorems D1 and D2 allow us to extend the result in Newey (1997) to more applications, we have not shown an existence of a transformation T satisfying Assumption D3. Actually, this is the most difficult part in this section and is our main contribution. It should be hinted from Assumption D3 that we only need to show that the required condition holds except for a null set. However, for simplicity of the proof, we make the following assumptions.

First, we introduce some notations applicable only for this section. For a function $f : A \to B$, define $f(A) = \{f(a) \in B : a \in A\}$ as the image of A under f. Let $\mathcal{X}$ denote the support of x and $\mathcal{X}_t$ denote the support of x_t. That is $\mathcal{X} = \{x \in \mathbb{R}^g : f_X(x) > 0\}$ and $\mathcal{X}_t = \{x_t \in \mathbb{R} : f_t(x_t) > 0\}$ where f_t is the marginal density function of x_t.

Assumption D7 *(i) The probability density function f_X of x is continuous on $\mathcal{X}$. (ii) $\mathcal{X}$ is the Euclidean product of unbounded open intervals.*

Assumption D7 is applicable for many families of multivariate random variables such as the normal, Student's t and exponential families. Lemma D3 (i) also shows that Assumption D7 (ii) is not too strong an additional condition from Assumption D7 (i). Again for simplicity of the proof, we restrict a transformation $T : \mathbb{R}^g \to \mathbb{R}^g$ to be of the form

$$T(x) = \xi = (m_1(x_1), \; ..., \; m_g(x_g))', \tag{2.27}$$

where $x = (x_1, \; ..., \; x_g)$ and m_t are functions $m_t : \mathbb{R} \to \mathbb{R}$.

Assumption D8 *Functions $m_t : \mathbb{R} \to \mathbb{R}$ are (i) strictly increasing; (ii) continuously differentiable and*

$$\frac{d}{du} m_t(u) > 0 \text{ for all } u \in \mathcal{X}_t;$$

and (iii) for all $u \in \mathbb{R}$, $|m_t(u)| \leq C$ for some finite constant C.

Note that Assumption D8 (i) can be replaced by a strictly monotonic function. However if m_t are decreasing, then $-m_t$ are increasing. Hence, there is no loss of generality. It follows from Lemma D3 that under Assumptions D7 (i) and D8, for any $\xi \in T(\mathcal{X})$,

$$f_\xi(\xi) = f_X\left(T^{-1}(\xi)\right) \prod_{t=1}^{g} \left[m_t'\left(m_t^{-1}(\xi_t)\right)\right]^{-1}, \tag{2.28}$$

and f_ξ is continuous on $T(\mathcal{X})$, where $m_t'(u) = dm_t(u)/du$. In order to see the significance of a right choice of transformations m_t and hence T, we restrict our intuitive discussion to a univariate case. As $f_X(x) > 0$ for all $x \in \mathbb{R}$, it follows that $\lim_{x \to \infty} f_X(x) = 0 = \lim_{x \to -\infty} f_X(x)$. To ensure that f_ξ satisfies Assumption D3, one difficulty may arise from the fact that f_X may converge to zero at a very fast rate as in the case of the Gaussian random variable. With respect to (2.28), the role of the transformation m is to make f_ξ to have fatter tails. That is a right choice of a function m has to move enough proportion of mass of the density f_X so that f_ξ have enough mass at the tails in order to satisfy Assumption D3. It turns out that the following family of transformations will do the job. In order to avoid making the proof excessively lengthy, we first restrict Assumption D7 to the following one.

Assumption D9 *Assumption D7 holds with $\mathcal{X} = \mathbb{R}^g$.*

We will later briefly show how to extend the result for $\mathcal{X}$ in the form

$$\mathcal{X} = \prod_{t=1}^{n} I_t,$$

where I_t can be any combination of I_t of the forms $(-\infty,\ b)$, $(b,\ \infty)$ and $(-\infty,\ \infty)$ where b is a finite real number, so that only Assumption D7 (i) holds precisely.

Definition D1 *A function $m : \mathbb{R} \to \mathbb{R}$ is in a class $\mathcal{E}$ if*

$$m(u) = -\frac{1}{1 + \exp(g(u))}, \tag{2.29}$$

where $g : \mathbb{R} \to \mathbb{R}$ is a continuously differentiable function such that its derivatives are strictly positive and

$$\lim_{u \to -\infty} g(u) = -\infty \quad \text{and} \quad \lim_{u \to \infty} g(u) = \infty.$$

Given the class $\mathcal{E}$, Theorem D3 gives a hint for a proper choice of function g such that the transformation T satisfies Assumption D3.

Theorem D3 *Suppose a transformation T has the form (2.27) where m_t are in the class $\mathcal{E}$ and Assumption D9 holds. Suppose there are functions $q_t : \mathbb{R} \to \mathbb{R}$ such that, for $t = 1,\ ...,\ g$,*

$$\lim_{x_t \to \infty} \frac{f_X(x)}{\Pi_{s=1}^{g} \exp(q_s(x_s))} \geq c_{1t}, \quad \lim_{x_t \to -\infty} \frac{f_X(x)}{\Pi_{s=1}^{g} \exp(q_s(x_s))} \geq c_{2t}, \tag{2.30}$$

where $c_{1t},\ c_{2t} > 0$ and can be infinite, for all $x_{-t} = (x_1,\ ...,\ x_{t-1},\ x_{t+1},\ ...,\ x_g)'$ in $\mathbb{R}^{g-1}$, and for all $t = 1,\ ...,\ g$,

$$\lim_{x_t \to \infty} \left[g(x_t) + q_t(x_t) - \log(g'(x_t)) \right] = \infty, \tag{2.31}$$

$$\lim_{x_t \to -\infty} \left[g(x_t) - q_t(x_t) + \log(g'(x_t)) \right] = -\infty, \tag{2.32}$$

where g is a function in (2.29). Then the transformation T satisfies Assumption D3.

Remark Theorem D3 gives sufficient conditions on f_X and proper choices of $g(u)$ so that the transformation T will satisfy Assumption D3. The complexity of conditions in Theorem D3 is mainly designed to turn a multivariate problem into a univariate one. It is much easier to find a limit of a function with one variable than with several variables. To appreciate the usefulness of Theorem D3, we consider one example. Suppose x is a multivariate normal random variables with zero mean and a p.d. covariance matrix Σ. It follows that

$$f_X(x) = C \exp\left(-\frac{1}{2} x' \Sigma^{-1} x\right), \quad \text{for some finite constant } C. \tag{2.33}$$

If we choose $g(u) = u^3 + u$, then $g'(u) = 3u^2 + 1 > 0$. It can be verified that this choice of g make m be in the class $\mathcal{E}$. Given f_X in (2.33), we can choose $q_t(x_t) = -c_0 x_t^2/2$ where $c_0 = \left\|\Sigma^{-1}\right\|$. It follows that

$$f_X(x) \geq C \exp\left(-\frac{1}{2} \|x\|^2 \left\|\Sigma^{-1}\right\|\right) = C \exp\left(-\frac{c_0}{2} \sum_{t=1}^{g} x_t^2\right).$$

Hence for all $x \in \mathbb{R}^g$,

$$\frac{f_X(x)}{\Pi_{s=1}^g \exp(q_s(x_s))} \geq C \prod_{s=1}^{g} \frac{\exp\left(-c_0 x_s^2/2\right)}{\exp\left(-c_0 x_s^2/2\right)} = C > 0.$$

Therefore, for all $t = 1, \ ..., \ g$,

$$\lim_{x_t \to -\infty} \frac{f_X(x)}{\Pi_{s=1}^g \exp(q_s(x_s))}, \lim_{x_t \to \infty} \frac{f_X(x)}{\Pi_{s=1}^g \exp(q_s(x_s))} \geq C > 0,$$

for all $x_{-t} \in \mathbb{R}^{g-1}$. Hence condition (2.30) holds. Now for each t,

$$\lim_{x_t \to \infty} \left[g(x_t) + q_t(x_t) - \log(g'(x_t))\right] = \lim_{x_t \to \infty} \left[x_t^3 + x_t - c_0 x_t^2/2 - \log\left(3x_t^2 + 1\right)\right] = \infty$$

and

$$\lim_{x_t \to -\infty} \left[g(x_t) - q_t(x_t) + \log(g'(x_t))\right] = \lim_{x_t \to -\infty} \left[x_t^3 + x_t + c_0 x_t^2/2 + \log\left(3x_t^2 + 1\right)\right] = -\infty.$$

Hence conditions (2.31) and (2.32) hold. Therefore the function

$$m(u) = -\frac{1}{1 + \exp(u^3 + u)} \tag{2.34}$$

can give the transformation T such that Assumption D3 holds for the multivariate normal distribution. It is easy to see, particularly from (2.30) that if f_X has fatter tails than the multivariate normal distribution, we can apply the same choices of q_t and g so that all conditions in Theorem D3 hold. Hence we can state the following result where the proof is omitted.

Corollary D *Suppose x is a random variable such that its support is $\mathbb{R}^g$ and $f_X(x) > 0$ for all $x \in \mathbb{R}^g$. Suppose that f_X is continuous and its tails approach zero at most as fast as that of the multivariate normal distribution. Then the transformation constructed from (2.34) will make Assumption D3 holds.*

Remark If the tails of f_X approach zero at a much faster rate than that of the normal distribution, then a choice of $g(u) = \exp(u)$ can be employed. This choice of function g moves much more mass of f_ξ towards the boundary of $T(\mathcal{X})$.

Remark Now we consider other forms of $\mathcal{X}$. As noted earlier, our assumptions and Lemmas D5, D6 essentially turn a multivariate problem into a univariate one. To avoid making many repetitive steps as for the proof of Theorem D3, we simply consider x where x is a real-valued random variable. An extension to multivariate x can be seen from all Lemmas associated with the proof of Theorem D3. Without loss of generality, suppose it is known that the support of x is $\mathcal{X} = (0,\ \infty)$. For more general cases namely $(-\infty,\ b)$ and $(b,\ \infty)$, one can always first take a linear transformation to get $\mathcal{X} = (0,\ \infty)$. Examples of x satisfying this assumption are the exponential distribution and other distributions taking only positive values. It is obvious that the choice of $g(u) = u^3 + u$ will not make $\xi = m(x)$ satisfy Assumption D3 since $\lim_{\xi \to -1/2} f_\xi(\xi) = 0$ where $T(\mathcal{X}) = (-1/2,\ 0)$.

Suppose that the right tail of f_X decreases to zero at a rate slower than that of the normal distribution and, for some constant k, the left tail approaches zero at the rate x^k as $x \to 0$. Then one can choose

$$g(u) = u^{k'},$$

where k' is the smallest odd number such that $k' > \min\{k+1,\ 3\}$, as a choice for m. Under regularity conditions as outlined above, it follows that

$$\begin{aligned} f_\xi(\xi) &= f_X\left(m^{-1}(\xi)\right)\left[m'\left(m^{-1}(\xi)\right)\right]^{-1} \\ &= \frac{f_X(x)}{\exp(g(x))\, k' x^{k'-1}} \left[1 + \exp(g(x))\right]^2, \end{aligned}$$

where $x = m^{-1}(\xi)$. Certainly $T(\mathcal{X}) = (m(0),\ 0)$ and $m(0)$ is the left boundary of $T(\mathcal{X})$. The difference between this choice of g and the previous one is that there is a number x_0 in $\mathbb{R}$ such that $g'(x_0) = 0$. In this case, $x_0 = 0$. Applying the steps shown for the normal example, it can be shown that $\lim_{\xi \to 0} f_\xi(\xi) = \infty$. Similarly, $\lim_{\xi \to m(0)} f_\xi(\xi) = \infty$ since $x^{k'-1}/x^k \to 0$ as $x \to 0$. As we have shown that f_ξ is continuous and its limits go to infinity as ξ approaches the boundary. Hence it follows that f_ξ is bounded away from zero on it support, i.e. Assumption D3 holds. As a consequence, for $\mathcal{X} = \Pi_{t=1}^{g} I_t$ where I_t are unbounded open intervals, we can choose the right function m_t to match the behaviour of f_X on each I_t so that Assumption D3 holds.

Remark In reality, practitioners may not have full knowledge of the support $\mathcal{X}$. One question we have to discuss is whether the wrong kind of transformation will have any significant effect on our result. First suppose that $\mathcal{X}$ is $\mathbb{R}$ but we employ the transformation for $\mathcal{X} = (0,\ \infty)$ as discussed above. It turns out that this mistake will not have a serious impact since Assumption D3 still holds. Recall that this type of transformation makes $\lim_{\xi \to T(0)} f_\xi(\xi) = \infty$ without changing the limits of f_ξ as $\xi \to -1$ and $\xi \to 0$. Moreover, $T(\mathcal{X})$ becomes $(-1,\ T(0)) \cup (T(0),\ 0)$. As f_ξ is still continuous on $T(\mathcal{X})$ and all boundary points are such that their limits approach infinity. Hence, f_ξ is bounded away from zero for every $\xi \in T(\mathcal{X})$. Then we can set $f_\xi(T(0)) = 0$ but Assumption D3 still holds since it only requires that f_ξ is bounded away from zero almost everywhere.

Now suppose $\mathcal{X} = \Pi_{t=1}^{g}[a_t,\ b_t]$, where $a_t,\ b_t$ are finite numbers and f_X is bounded away from zero for all x in $\mathcal{X}$, i.e. f_X satisfies the assumption in Newey (1997). If we employ a transformation for $\mathcal{X} = \mathbb{R}^g$ as for the case of the multivariate normal distribution, then this wrong kind of transformation still makes Assumption D3 hold. The reason is that f_ξ will still be bounded away from zero on the new support of the form $\Pi_{t=1}^{g}[m(a_t),\ m(b_t)]$. Similarly, one can argue that the transformation for $\mathcal{X}$ of the type $(0,\ \infty)$ will not affect Assumption D3 either. Finally, it is worth noting that if f_X behaves in such a way that there is no transformation T such that Assumption D3 holds, then the convergence in Theorem D1 still holds. The only effect is that the number of approximating functions employed in the series estimation has to grow very slowly due to the fact that the matrix $P'P$ employed in the approximation is near singular. This can be seen from Newey (1988).

2.5 Efficiency Improvement

In section 3, we see that the Gaussian pseudo-maximum likelihood estimate can be root-n-consistent. If precision of an estimate is a concern, particularly when the sample size is small, a maximum likelihood estimate can be employed. If the density of the innovations $\varepsilon_{i\cdot}$ are correctly specified, then, under mild assumptions, a maximum likelihood estimate will be more efficient than the estimate in the previous section. However, if the density is misspecified, the maximum likelihood estimate may become inconsistent.

Stone (1975) showed that, for a simple location model with independent and identically distributed (i.i.d.) data and symmetric distribution, in the absence of a complete knowledge of the distribution of the data, there exists an estimate that is asymptotically as efficient as the maximum likelihood estimate when the density function is known. Not knowing the density function of the data, Stone (1975) constructed his asymptotically efficient estimate from a nonparametric estimate of the unknown density function. Stone's estimate is adaptive in the sense of Stein (1956). That is the unknown location parameter can be estimated as well asymptotically not knowing the density function as knowing it.

Bickel (1982) and Newey (1988) extended Stone's result and showed how to obtain an adaptive estimate of the slope parameter in a linear regression model where the disturbances are i.i.d.. Many authors showed that adaptive estimates can be obtained even without i.i.d. data. For a linear regression model, Steigerwald (1992) considered the case when the disturbances follow an ARMA process while Robinson (2005) allowed the disturbances to follow a fractional process. More recently Robinson (2010) considered adaptive estimation of the slope and the spatial autoregressive parameters in a univariate SAR model.

Now consider a multivariate regression with SAR disturbances. Unlike the specification in the previous section, we now make a clear distinction between intercept and slope parameters. The model becomes

$$y_{it} = x_{it}'\beta_0 + u_{it}, \quad i = 1,\ ...,\ n,\ t = 1,\ ...,\ g, \tag{2.35}$$

where

$$u_{it} = \rho_{0t} \sum_{j=1}^{n} w_{ijt} u_{jt} + \alpha_{0t} + \varepsilon_{it}. \tag{2.36}$$

It is worth noting that this model is slightly different from the one considered earlier. The switch from the previous specification is a result of some complexity from the rate of convergence of the intercepts in the previous model. It is well-known from long memory time series that the rate of convergence of the intercept is slower than the slope parameters in the presence of long memory of the disturbances. As mentioned earlier, to allow for long-range dependence in u_{it}, the regressors x_{it} should be interpreted as mean-corrected random variables. With this new specification, the parameters α_{0t} play a role of location parameters. We still expect that the result established in the previous section should hold for this specification too. The reason for treating the intercept and slope parameters separately is based on the fact that, compared with β_0, α_{0t} can be adaptively estimable under relatively stronger assumptions. See, for example, Bickel (1982). Now we introduce some definitions and assumptions.

Definition E1 *Let $[a,\ b]$ be a closed interval. A function $f : [a,\ b] \to \mathbb{R}$ is absolutely continuous if for every $\varepsilon > 0$ there exists $\delta > 0$ such that if*

$$a \leq x_1 < y_1 \leq x_2 < y_2 \leq \ \dots\ \leq x_m < y_m \leq b$$

and

$$\sum_{i=1}^{m} (y_i - x_i) < \delta,$$

then

$$\sum_{i=1}^{m} |f(y_i) - f(x_i)| < \varepsilon.$$

For a function $f : \mathbb{R} \to \mathbb{R}$, we say that f is absolutely continuous if for every $x \in \mathbb{R}$, there exists $m > 0$ such that the restriction of f on the closed interval $[-m,\ m]$ is absolutely continuous.

Definition E2 *For a function $f : \mathbb{R}^g \to \mathbb{R}$, define a function $f_{x_{-t}} : \mathbb{R} \to \mathbb{R}$ by*

$$f_{x_{-t}}(x_t) = f(x_1,\ \dots,\ x_{t-1},\ x_t,\ x_{t+1},\ \dots,\ x_g)$$

where $x_{-t} = (x_1,\ \dots,\ x_{t-1},\ x_{t+1},\ \dots,\ x_g) \in \mathbb{R}^g$. A function $f : \mathbb{R}^g \to \mathbb{R}$ is in the class $\mathcal{AC}(\mathbb{R}^g)$ if functions $f_{x_{-t}}(x_t)$ are absolutely continuous almost everywhere (for all $x_{-t} \in \mathbb{R}^{g-1}$ except x_{-t} in a null set) for $t = 1,\ \dots,\ g$. Define $\partial f_{x_{-t}}(x_t)/\partial x_t$ as a derivative of $f_{x_{-t}}$ and

$$\frac{\partial f}{\partial x} = \left(\frac{\partial f_{x_{-1}}}{\partial x_1},\ \dots,\ \frac{\partial f_{x_{-g}}}{\partial x_g}\right)'.$$

Fix t and x_{-t}. If $f_{x_{-t}}$ is absolute continuous, then it is differentiable almost everywhere. Hence the derivative defined in Definition E1 is well-defined almost everywhere. A sufficient condition for a density $f : \mathbb{R}^g \to \mathbb{R}$ to be in the class $\mathcal{AC}(\mathbb{R}^g)$ is that f is continuously differentiable. Continuous differentiability of f implies that $\partial f / \partial e_i$ are continuous for all $i = 1, \ ..., \ g$. Hence $f_{e_{-i}}$ are absolutely continuous everywhere for $i = 1, \ ..., \ g$. A weaker sufficient condition such as a Lipschitz condition can be employed to check for absolute continuity too. Recall that $\varepsilon_{i\cdot} = (\varepsilon_{i1}, \ ..., \ \varepsilon_{ig})'$.

Assumption E1 *(i)* $\{\varepsilon_{i\cdot}\}$ *is an independent and identically distributed sequence of* $\mathbb{R}^g$*-valued random variables with the joint density function* f*. (ii) The function* f *is in the class* $\mathcal{AC}(\mathbb{R}^g)$ *with partial derivative* $\partial f / \partial e$*. For* e *such that* $f(e) > 0$*, let*

$$\psi(e) = -\frac{1}{f(e)} \frac{\partial f(e)}{\partial e}.$$

(iii)

$$\mathbb{E}\left\{\psi(\varepsilon_{i\cdot})\, \psi(\varepsilon_{i\cdot})'\right\} = \mathcal{L},$$

and $\mathcal{L}$ *is a finite and positive definite matrix.*

Define $\alpha_0 = (\alpha_{01}, \ ..., \ \alpha_{0g})'$ and $\theta_{04} = \left(\alpha_0', \ \beta_0', \ \rho_0'\right)'$. Similarly define $\alpha = (\alpha_1, \ ..., \ \alpha_g)'$, $\theta_4 = \left(\alpha', \ \beta', \ \rho'\right)'$,

$$\varepsilon_{it}(\theta_4) = (y_{it} - x_{it}'\beta) - \rho_t \sum_{j=1}^{n} w_{ijt}\left(y_{jt} - x_{jt}'\beta\right) - \alpha_t \tag{2.37}$$

and $\varepsilon_{i\cdot}(\theta_4) = (\varepsilon_{i1}(\theta_4), \ ..., \ \varepsilon_{ig}(\theta_4))'$.

Assumption E2 $\{x_{it}\}$ *and* $\{\varepsilon_{it}\}$ *are independent and the joint distribution function of* $\{x_{it}\}$ *does not depend on* θ_4*.*

In order to express the likelihood function of the data in a tractable form, we state one useful result. Consider any random variables $X_1, \ ..., \ X_n$ with the joint probability density function f. Let T be an $\mathbb{R}^n$-valued function such that $T(X_1, \ ..., \ X_n) = Y_1, \ ..., \ Y_n$ where $Y_1, \ ..., \ Y_n$ are just a re-arrangement of $X_1, \ ..., \ X_n$. It follows that the Jacobian matrix of T is a product of elementary matrices. Since these elementary matrices represent row-switching transformations, the modulus of the Jacobian determinant is unity. Hence $Y_1, \ ..., \ Y_n$ has the density function

$$g(y_1, \ ..., \ y_n) = f\left(T^{-1}(y_1, \ ..., \ y_n)\right), \tag{2.38}$$

where T^{-1} is the inverse of T.

Suppose the joint density function of $vec(X)$ is f_X. To discuss possibility of adaptive estimation, we first consider an arbitrary parametric submodel corresponding to a parameterization of the joint density of $vec(X)$ as $f_X(x,\ \eta_1)$ and of $\varepsilon_{i\cdot}$ as $f(e,\ \eta_2)$, where $f_X(x) = f_X(x,\ \eta_{01})$ and $f(e) = f(e,\ \eta_{02})$, for some η_{01} and η_{02}. It follows from Assumptions E1 - E2 and (2.38) that the log-likelihood of the sample is

$$l_n(\theta_4) = \log f_X(vec(X),\ \eta_1) + \frac{1}{2}\sum_{t=1}^{g}\log\left|S_t(\rho)' S_t(\rho)\right| + \sum_{i=1}^{n}\log f(\varepsilon_{i\cdot}(\theta_4),\ \eta_2), \quad (2.39)$$

where $S_t(\rho)$ are defined as in the previous section.

Following Stein (1956), θ_{04} is adaptively estimable if θ_{04} can be estimated as asymptotically efficient not knowing η_{01} and η_{02} as knowing η_{01} and η_{02}. A necessary condition for θ_{04} to be adaptively estimable is that the information matrix of θ_4 and $(\eta_1,\ \eta_2)'$ is block-diagonal. Under the Gaussian assumption, it can be seen from Theorem C that a necessary condition for ρ_0 to be adaptively estimable is that $\lim_{n\to\infty} n^{-1}tr(G_t) = 0$ for $t = 1,\ ...,\ g$. Since the required condition does not generally hold, particularly with our assumption that $\left\|S^{-1}\right\| \to \infty$ as $n \to \infty$, it follows that ρ_0 is generally not apatively estimable.

It should be noted that condition (3.13) in Robinson (2010) makes the information matrix of the spatial autoregressive parameter and the variance of the Gaussian innovations block-diagonal. Hence Stein's necessary condition for adaptive estimation is satisfied. With further assumptions, he show how to obtain adaptive estimates for both slope and spatial autoregressive parameters. However, Robinson's sufficient conditions for adaptive estimation are not of our interest since they depends on row normalization and the restriction imposed on the parameter space of the unknown spatial autoregressive parameter.

As ρ_0 is not generally adaptively estimable, for simplicity, we only focus on adaptive estimation of β_0. Now we discuss adaptive estimation of β_0. First, we introduce some notations. From (2.35), it follows that

$$y_{i\cdot} = X_{i\cdot}\beta_0 + u_{i\cdot},$$

where $y_{i\cdot} = (y_{i1},\ ...,\ y_{ig})'$, $X_{i\cdot} = (x_{i1},\ ...,\ x_{ig})'$ and $u_{i\cdot} = (u_{i1},\ ...,\ u_{ig})'$. Define

$$x_{it}^* = x_{it} - \rho_{0t}\sum_{j=1}^{n} w_{ijt}x_{jt}, \quad x_{it}^*(\rho_t) = x_{it} - \rho_t\sum_{j=1}^{n} w_{ijt}x_{jt}, \quad (2.40)$$

$$X_{i\cdot}^* = \left(x_{i1}^*,\ ...,\ x_{ig}^*\right)',\ X_{i\cdot}^*(\rho) = \left(x_{i1}^*(\rho_1),\ ...,\ x_{ig}^*(\rho_g)\right)', \quad (2.41)$$

and

$$\overline{X}_{\cdot}^*(\rho) = n^{-1}\sum_{i=1}^{n} X_{i\cdot}^*(\rho).$$

Let $\Lambda = diag\left(\rho_1,\ ...,\ \rho_g\right)$ and $W_{ij} = diag(w_{ij1},\ ...,\ w_{ijg})$. With this notation, it follows from (2.37) that

$$\varepsilon_{i\cdot}(\theta_4) = (y_{i\cdot} - X_{i\cdot}\beta) - \Lambda\sum_{j=1}^{n} W_{ij}(y_{j\cdot} - X_{j\cdot}\beta) - \alpha. \quad (2.42)$$

Suppose the density function of $\varepsilon_{i\cdot}$ and ρ_0 are known. Suppose there are initial estimates $\widetilde{\alpha}$ and $\widetilde{\beta}$ of α_0 and β_0. Then to avoid non-linear optimization, one can employ the linearized maximum likelihood estimate $\widehat{\beta}$ of β_0 of the form

$$\widehat{\beta}=\widetilde{\beta}+\left(\sum_{i=1}^{n}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)'\widehat{\mathcal{L}}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)\right)^{-1}\left(\sum_{i=1}^{n}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)'\psi\left(\varepsilon_{i\cdot}\left(\widetilde{\alpha},\ \widetilde{\beta},\ \rho_0\right)\right)\right),$$

where

$$\widehat{\mathcal{L}}=n^{-1}\sum_{i=1}^{n}\psi\left(\varepsilon_{i\cdot}\left(\widetilde{\alpha},\ \widetilde{\beta},\ \rho_0\right)\right)\psi\left(\varepsilon_{i\cdot}\left(\widetilde{\alpha},\ \widetilde{\beta},\ \rho_0\right)\right)'$$

and $\overline{X}_{\cdot}^{*}=n^{-1}\sum_{i=1}^{n}X_{i\cdot}^{*}$. Under weak regularity conditions, it can be shown that this estimate will be efficient in the Cramer-Rao sense.

If we no longer assume that the density function and ρ_0 are known, we can estimate ρ_0 as in Section 3 but we have to nonparametrically estimate the unknown score function ψ. In this chapter, we employ a series estimate developed earlier. Our adaptive estimate of β_0 is

$$\widehat{\beta}=\widetilde{\beta}+\left(\sum_{i=1}^{n}\left[X_{i\cdot}^{*}\left(\widetilde{\rho}\right)-\overline{X}_{\cdot}^{*}\left(\widetilde{\rho}\right)\right]'\widehat{\mathcal{L}}_L\left[X_{i\cdot}^{*}\left(\widetilde{\rho}\right)-\overline{X}_{\cdot}^{*}\left(\widetilde{\rho}\right)\right]\right)^{-1}\left(\sum_{i=1}^{n}\left[X_{i\cdot}^{*}\left(\widetilde{\rho}\right)-\overline{X}_{\cdot}^{*}\left(\widetilde{\rho}\right)\right]'\widehat{\psi}_L\left(\widetilde{\varepsilon}_{i\cdot}\right)\right),$$

where $\widehat{\mathcal{L}}=n^{-1}\sum_{i=1}^{n}\widehat{\psi}_L\left(\widetilde{\varepsilon}_{i\cdot}\right)\widehat{\psi}_L\left(\widetilde{\varepsilon}_{i\cdot}\right)'$, $\widetilde{\varepsilon}_{i\cdot}=\varepsilon_{i\cdot}\left(\widetilde{\theta}_4\right)$, $\widetilde{\theta}_4=\left(\widetilde{\alpha}',\ \widetilde{\beta}',\ \widetilde{\rho}'\right)'$ and $\widehat{\psi}_L\left(\widetilde{\varepsilon}_{i\cdot}\right)$ is a nonparametric estimate of $\psi\left(\widetilde{\varepsilon}_{i\cdot}\right)$ which will be discussed in details below.

As in the previous section, for any v in $\mathbb{R}^g$ and a multi-index $\lambda\left(l\right)$ in $\mathbb{N}_0^g$, define

$$p_l\left(v\right)=v^{\lambda(l)}\ \text{ and }\ p^L\left(v\right)=\left(p_1\left(v\right),\ ...,\ p_L\left(v\right)\right)'.$$

Let T be a one-one transformation described in the previous section. Define

$$\pi_t^L\left(u\right)=\left(\frac{\partial p_1\left(T\left(u\right)\right)}{\partial u_t},\ ...,\ \frac{\partial p_L\left(T\left(u\right)\right)}{\partial u_t}\right)'.$$

Let $\psi_t\left(e\right)$ be the $t-$th element of $\psi\left(e\right)$. Then, for $t=1,\ ...,\ g$, our nonparametric estimate of $\psi_t\left(\varepsilon_{i\cdot}\right)$ is

$$\widehat{\psi}_{tL}\left(\widetilde{\varepsilon}_{i\cdot}\right)=p^L\left(T\left(\widetilde{\varepsilon}_{i\cdot}\right)\right)'\left(\sum_{j=1}^{n}p^L\left(T\left(\widetilde{\varepsilon}_{j\cdot}\right)\right)p^L\left(T\left(\widetilde{\varepsilon}_{j\cdot}\right)\right)'\right)^{-1}\left(\sum_{j=1}^{n}\pi_t^L\left(\widetilde{\varepsilon}_{j\cdot}\right)\right), \tag{2.43}$$

where $\widetilde{\varepsilon}_{i\cdot}$ are as described above, and

$$\widehat{\psi}_L\left(\widetilde{\varepsilon}_{i\cdot}\right)=\left(\widehat{\psi}_{1L}\left(\widetilde{\varepsilon}_{i\cdot}\right),\ ...,\ \widehat{\psi}_{gL}\left(\widetilde{\varepsilon}_{i\cdot}\right)\right)'.$$

Assumption E3 *The transformation T is of the form (2.27) where $m_1=\cdots=m_g=m$, the function m is in the class E and the function g is well chosen so that Assumption D3* *holds.*

Assumption E4 *A function m in the class $\mathcal{E}$ is such that*

$$\sup_{u\in\mathbb{R}}\left|\frac{\partial}{\partial u}m\left(u\right)\right|+\sup_{u\in\mathbb{R}}\left|\frac{\partial^{2}}{\partial u^{2}}m\left(u\right)\right|<\infty$$

Assumption E5 *As $n\to\infty$, (i)*

$$n^{-1}\sum_{i=1}^{n}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)^{\prime}\mathcal{L}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)\to_{p}V,$$

where V is a positive definite matrix; (ii)

$$n^{-1/2}\max_{1\leq i\leq n}\left\|X_{i\cdot}^{*}\right\|=o_{p}\left(1\right);$$

(iii)

$$n^{-1}\sum_{i=1}^{n}\mathbb{E}\left\|X_{i\cdot}^{*}\right\|^{2}=O\left(1\right),\ n^{-1}\sum_{i=1}^{n}\mathbb{E}\left\|X_{i\cdot}^{*}\right\|^{3}=O\left(n^{\kappa_{1}}\right),\ n^{-1}\sum_{i=1}^{n}\mathbb{E}\left\|X_{i\cdot}^{*}\right\|^{4}=O\left(n^{\kappa_{2}}\right);$$

(iv)

$$n^{-1}\sum_{i=1}^{n}\mathbb{E}\left\|\sum_{j=1}^{n}W_{ij}X_{j\cdot}\right\|=O_{p}\left(n^{\zeta_{1}}\right),\ n^{-1}\sum_{i=1}^{n}\mathbb{E}\left\|\sum_{j=1}^{n}W_{ij}X_{j\cdot}\right\|^{2}=O_{p}\left(n^{\zeta_{2}}\right),$$

where $\zeta_1<1/2$ and $\zeta_2<1$; (v)

$$n^{-1}\sum_{i=1}^{n}\left\|\sum_{j=1}^{n}W_{ij}u_{j\cdot}\right\|=O_{p}\left(1\right)\ \text{and}\ n^{-1}\sum_{i=1}^{n}\left\|\sum_{j=1}^{n}W_{ij}u_{j\cdot}\right\|^{2}=O_{p}\left(1\right);$$

(vi)

$$n^{-1}\sum_{i=1}^{n}\left\|\sum_{j=1}^{n}W_{ij}X_{j\cdot}\right\|\left\|X_{i\cdot}^{*}\right\|^{2}=o_{p}\left(n^{1/2}\right),\ n^{-1}\sum_{i=1}^{n}\left\|\sum_{j=1}^{n}W_{ij}X_{j\cdot}\right\|\left\|X_{i\cdot}^{*}\right\|^{2}=o_{p}\left(n^{1/2}\right),$$

$$n^{-1}\sum_{i=1}^{n}\left\|X_{i\cdot}^{*}\right\|\left\|\sum_{j=1}^{n}W_{ij}X_{j\cdot}\right\|=O_{p}\left(n^{\zeta_{4}}\right);$$

(vii)

$$n^{-1}\sum_{i=1}^{n}\left\|\sum_{j=1}^{n}W_{ij}X_{j\cdot}\right\|\left\|\sum_{j=1}^{n}W_{ij}u_{j\cdot}\right\| = O_p\left(n^{\zeta_5}\right),\ n^{-1}\sum_{i=1}^{n}\left\|\sum_{j=1}^{n}W_{ij}u_{j\cdot}\right\|\left\|X_{i\cdot}^{*}\right\|^2 = O_p\left(n^{\zeta_3}\right),$$

$$n^{-1}\sum_{i=1}^{n}\left\|X_{i}^{*}\right\|\left\|\sum_{j=1}^{n}W_{ij}X_{j\cdot}\right\|\left\|\sum_{k=1}^{n}W_{ik}u_{k\cdot}\right\| = O_p\left(n^{\zeta_6}\right),\ n^{-1}\sum_{i=1}^{n}\left\|X_{i\cdot}^{*}\right\|\left\|\sum_{j=1}^{n}W_{ij}u_{j\cdot}\right\|^2 = O_p\left(n^{\zeta_7}\right).$$

Let

$$R_i = diag\left\{\sum_{j=1}^{n}w_{ij1}u_{j1},\ ...,\ \sum_{j=1}^{n}w_{ijg}u_{jg}\right\}, \tag{2.44}$$

$$\mathcal{D}_{1n} = \sum_{i=1}^{n}\mathbb{E}\left\|X_{i\cdot}^{*}\right\|^2 \text{ and } \mathcal{D}_{2n} = \sum_{i=1}^{n}\mathbb{E}\left\|h\left(\varepsilon_i\right)R_i\right\|^2$$

Assumption E6 *For a function $h:\mathbb{R}^g \to \mathcal{M}_g$, where $\mathcal{M}_g$ is the set of all real $g \times g$ matrices, let*

$$\mathcal{D}_{1n} = \sum_{i=1}^{n}\left\|X_{i\cdot}^{*}\right\|^2 \text{ and } \mathcal{D}_{2n} = \sum_{i=1}^{n}\left\|h\left(\varepsilon_i\right)R_i\right\|^2.$$

Assume that

$$\mathcal{D}_{1n}^{-1/2}\mathcal{D}_{2n}^{-1/2}\sum_{i=1}^{n}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)' h\left(\varepsilon_{i\cdot}\right)R_i = O_p\left(n^{\vartheta}\right),$$

where $\vartheta < 0$.

Assumption E7 *The sequence $\{\lambda(l)\}_{l=1}^{\infty}$ includes all distinct multi-indices. The sequence is ordered so that $\|\lambda(l)\|_1$ is monotonically increasing.*

Assumption E8 *As $n\to\infty$, $\widetilde{\theta}_4 - \theta_{04} = O_p\left(n^{-1/2}\right)$.*

Assumption E9 *For $\kappa_1,\ \kappa_2,\ \zeta_i,\ i = 1,\ ...,\ 7$, in Assumption E5, and ϑ in Assumption E6, as $n\to\infty$,*

$$L^{-1} + n^{-1}L^{16} + n^{-1/2+\max\{\kappa_1,\ \zeta_3\}}L^5 + n^{-1+\max\{\kappa_2,\ 2\zeta_4+2\zeta_5\}}L^6 + n^{-1+2\zeta_1}L$$
$$+n^{-(1-\zeta_2)}L + n^{-1+\max\{\zeta_6,\ \zeta_7\}}L^5 + n^{\vartheta}L^3$$

is $o(1)$.

Theorem E *Under Assumptions E1-E9,*

$$n^{-1/2}\left(\widehat{\beta}-\beta_0\right)\to_d N\left(0,\ V^{-1}\right).$$

It can be shown that the choice of functions g discussed in the previous section makes m satisfy Assumption E4. The complication of Assumption E5 arises from our attempt to accommodate some explosive behaviours of the regressors and disturbances. Recall that $X_{i\cdot}^* = X_{i\cdot} - \Lambda_0 \sum_{j=1}^n W_{ij} X_{j\cdot}$. For example, if we assume that as $n\to\infty$,

$$n^{-1}\sum_{i=1}^n \|X_{i\cdot}\| + n^{-1}\sum_{i=1}^n \|X_{i\cdot}^*\| = O_p(1), \tag{2.45}$$

then it follows that $n^{-1}\sum_{i=1}^n \left\|\sum_{j=1}^n W_{ij} X_{i\cdot}\right\| = O_p(1)$ provided that Λ_0 is nonsingular. Note that if Λ_0 is singular, it follows that at least one component of $u_{i\cdot}$ are actually independent, leading to a trivial case. Other terms in Assumptions E5 can be substantially simplified in a similar way if stronger conditions analogous to (2.45) are imposed. Assumption E6 is the most unique assumption for our model. It involves both the necessary condition for orthogonality so that β_0 can be adaptively estimated, and the strength of spatial dependence. The normalized sum in Assumption E6 is essentially an estimate of the covariance of $X_{i\cdot}^*$ and $h(\varepsilon_{i\cdot}) R_i$. By independence of $\{x_{it}\}$ and $\{\varepsilon_{it}\}$, this normalized sum should tend to zero in probability. The parameter ϑ determines the rate of convergence of this normalized sum to its population counterpart that is zero. Under some regularity conditions, it can be shown that under short-range dependence of both $X_{i\cdot}^*$ and $u_{i\cdot}$, $\vartheta = -1$. Under this and other stronger assumptions analogous to (2.45), Assumption E9 can be simplified so that it only requires

$$L\to\infty \text{ and } n^{-1}L^{16}\to 0$$

as $n\to\infty$. The slower rate of increase of L, compared with the ones discussed in the previous section, arises from the fact that our nonparametric estimate of the score function relies on integration-by-parts. Differentiation of the approximating polynomial functions is the source of this slow rate.

2.6 Final Comments

In this chapter we discuss estimation of a multivariate linear regression with spatial autoregressive disturbances. We show that the typical assumptions on the degree of spatial dependence and on the parameter space can be substantially relaxed. We illustrate the usefulness of the spectral norm over the row or column sum norms in defining an explosive behaviour of a spatial autoregressive process due to long-range dependence and explosive variances. This explosive behaviour may be a norm rather than an exception in cross-sectional data with spatial dependence. We also show that the pseudo-maximum likelihood estimate can be root-n-consistent in the presence of this explosive behaviour.

There are many possible extensions to the results established in this chapter. Our multivariate set-up should be readily applicable to linear panel data models. An extension to simultaneous equations models may require additional steps to deal with endogeneity within the models. It is also interesting to allow explosive behaviours of the disturbances in limited dependent variable models. Another possible extension is to allow spillover effects from regressors in the model.

Motivated by the success of the Autoregressive Fractionally Integrated Moving Average (ARFIMA) models in time series analysis, it is very interesting to investigate a possibility of a modelling strategy that can separate long-range spatial dependence from short-range spatial dependence. In the ARFIMA models, the memory parameter reflects long-run dynamics of the process whereas the ARMA parameters capture the short-run dynamics. In the first-order Autoregressive (AR(1)) model, i.e. $X_t = \rho X_{t-1} + \varepsilon_t$, $\{X_t\}$ is stationary if and only if $|\rho| < 1$ and there is an abrupt change when $|\rho| = 1$. This unsmooth behaviour of the variances and autocovariances is a reason for popularity of the ARFIMA model. However, in the first-order SAR model considered in this chapter, we show, in case of a symmetric weighting matrix, that the smoothness of the transition from completely stable variances and covariances to explosive ones is controlled by the rate at which one of the eigenvalues of the weighting matrix approaches the inverse of the spatial autoregressive parameter. As a result, it is not clear whether a modelling strategy that can directly separate short and long range dependence is needed. Nevertheless, it is important to investigate complexity of the dependence structure arising from higher-order spatial autoregressive models.

The second half of the chapter is devoted to discussion of efficiency improvement of the pseudo-maximum likelihood estimate by nonparametric estimates of the unknown score function of the distribution of the innovations in the model. The nonparametric power series estimation is employed to estimate the unknown score function. We stress the importance of a transformation prior to the nonparametric estimation especially when the distribution of the innovations has unbounded support. It is interesting to see sensitivity and relative performance of different choices of transformation in finite samples from Monte Carlo simulations.

Appendix 2.1: Proofs of Theorems

In the proofs, if not specified, C denotes a finite constant.

Proof of Theorem A By Lemma A3, (i) follows. Now consider a family $\{A_n\}$ such that all elements are uniformly bounded in n. Let

$$A_n = \begin{pmatrix} a_{11,n} & a'_{1n} \\ a_{1n} & B_n \end{pmatrix},$$

where $a_{11,n}$ is the $(1,1)$-th element of A_n, $a_{1n} \in \mathbb{R}^{n-1}$ is the first column of A_n with $a_{11,n}$ removed. Let $a_{j1,n}$ be the j-th element of a_{1n}. Suppose the family $\{a_{1n}\}$ of $\mathbb{R}^{n-1}$-vectors is such that $\sum_{j=1}^{n-1} |a_{j1n}|$ form an unbounded sequence but $\sum_{j=1}^{n-1} |a_{j1n}|^2$ form a bounded sequence. Then $\|A_n\|_1$ and $\|A_n\|_\infty$ are not uniformly bounded in n. Suppose for simplicity that $\{B_n\}$ is a family of positive definite matrices of order $n-1$ such that $\|B_n\|_1$ and $\|B_n\|_\infty$ are uniformly bounded in n. Note that if A_n is positive definite, then so is B_n. As A_n is positive definite,

$$\|A_n\| = \overline{\lambda}(A_n) = \sup_{\|x\|=1} x' A_n x.$$

Let $x = (x_1,\ x_2')' \in \mathbb{R}^n$ where $x_2 \in \mathbb{R}^{n-1}$. Then

$$\begin{aligned} \|A_n\| &= \sup_{\|x\|=1} \left(a_{11,n} x_1^2 + 2x_1 a_{1n}' x_2 + x_2' B_n x_2 \right) \\ &\leq |a_{11,n}| + 2 \sup_{\|x_2\|=1} a_{1n}' x_2 + \sup_{\|x_2\|=1} x_2' B_n x_2 \end{aligned} \tag{2.46}$$

By our assumption $|a_{11,n}|$ is uniformly bounded. As $\|B_n\|_1$ is uniformly bounded in n, and B_n is symmetric, Lemma A3 implies that,

$$\sup_{\|x_2\|=1} x_2' B_n x_2 = \|B_n\| \leq \|B_n\|_1 .$$

Hence the third term in (2.46) is uniformly bounded in n. Finally, Lemma A9 and the assumption that $\sum_{j=1}^{n-1} |a_{j1,n}|^2$ are uniformly bounded imply that the second term in (2.46) is uniformly bounded in n. Thus, $\|A_n\|$ is uniformly bounded in n.

Proof of Theorem B First we introduce some notations. Define $\Lambda_0 = diag\left(\rho_{01} I_n,\ ...,\ \rho_{0g} I_n\right)$ and $\Lambda = diag\left(\rho_1 I_n,\ ...,\ \rho_g I_n\right)$. Consider any p.d. matrix Σ and $\theta_3 \in \Theta_3$. Let σ^{st} be the (s,t)-th element of Σ^{-1}. Then

$$\begin{aligned} & n^{-1} (X^*)' G' (\Lambda_0 - \Lambda)' \left(\Sigma^{-1} \otimes I_n\right) (\Lambda_0 - \Lambda) G X^* \\ = \; & n^{-1} \sum_{s=1}^{g} \sum_{t=1}^{g} \sigma^{st} (\rho_{0s} - \rho_s)(\rho_{0t} - \rho_t)(X_s^*)' G_s' G_t X_t^* . \end{aligned}$$

Assumption B6 implies that this matrix converges in probability to a finite matrix. Similarly, it can be shown that, under Assumption B6,

$$\frac{1}{n} \begin{pmatrix} (X^*)' \left(\Sigma^{-1} \otimes I_n\right) X^* & (X^*)' \left(\Sigma^{-1} \otimes I_n\right) (\Lambda_0 - \Lambda) G X^* \\ (X^*)' G' (\Lambda_0 - \Lambda)' \left(\Sigma^{-1} \otimes I_n\right) X^* & (X^*)' G' (\Lambda_0 - \Lambda)' \left(\Sigma^{-1} \otimes I_n\right) (\Lambda_0 - \Lambda) G X^* \end{pmatrix} \tag{2.47}$$

converges in probability to a finite matrix. Denote its limit by $M_1(\theta_2,\ \theta_3)$. Since Σ is p.d., (2.47) can be written as $n^{-1} A A'$ where

$$A' = \left(\left(\Sigma^{-1/2} \otimes I_n\right) X^* \quad \left(\Sigma^{-1/2} \otimes I_n\right) (\Lambda_0 - \Lambda) G X^* \right).$$

Hence Lemma B9 implies that $M_1(\theta_2,\ \theta_3)$ is p.s.d.. Define

$$M_2(\theta_2,\ \theta_3) = \begin{pmatrix} I_K & I_K \end{pmatrix} M_1(\theta_2,\ \theta_3) \begin{pmatrix} I_K \\ I_K \end{pmatrix}. \tag{2.48}$$

Then $M_2(\theta_2,\ \theta_3)$ is p.s.d..

Our proof follows a standard procedure to show consistency of an extremum estimate. The loss function can be re-written as

$$Q_n(\theta) = -\frac{1}{2ng}\log\left|\Sigma^{-1}\otimes I_n\right| - \frac{1}{2ng}\log\left|S(\rho)' S(\rho)\right| + \frac{1}{2ng}u(\beta)' S(\rho)'\left(\Sigma^{-1}\otimes I_n\right)S(\rho)u(\beta).$$

Therefore

$$Q_n(\theta_0) = -\frac{1}{2ng}\log\left|\Sigma_0^{-1}\otimes I_n\right| - \frac{1}{2ng}\log\left|S'S\right| + \frac{1}{2ng}\varepsilon'\left(\Sigma_0^{-1}\otimes I_n\right)\varepsilon.$$

It follows that

$$\begin{aligned} & Q_n(\theta) - Q_n(\theta_0) \\ = & -\frac{1}{2ng}\log\left|H(\theta_2,\ \theta_3)\right| - \frac{1}{2ng}\varepsilon'\left(\Sigma_0^{-1}\otimes I_n\right)\varepsilon \\ & + \frac{1}{2ng}\left\{u + X(\beta_0-\beta)\right\}' S(\rho)'\left(\Sigma^{-1}\otimes I_n\right)S(\rho)\left\{u + X(\beta_0-\beta)\right\} \\ = & \ s_n(\theta) - t_n(\theta), \end{aligned}$$

where

$$\begin{aligned} s_n(\theta) = & \frac{1}{2ng}tr\left\{H(\theta_2,\ \theta_3)\right\} - \frac{1}{2ng}\log\left|H(\theta_2,\ \theta_3)\right| - \frac{1}{2} \\ & + \frac{1}{2g}(\beta_0-\beta)' M_2(\theta_2,\ \theta_3)(\beta_0-\beta), \end{aligned}$$

and

$$\begin{aligned} -t_n(\theta) = & \frac{1}{2ng}u' S(\rho)'\left(\Sigma^{-1}\otimes I_n\right)S(\rho)u - \frac{1}{2ng}tr\left\{H(\theta_2,\ \theta_3)\right\} \\ & + \frac{1}{ng}(\beta_0-\beta)' X' S(\rho)'\left(\Sigma^{-1}\otimes I_n\right)S(\rho)u \\ & + \frac{1}{2g}(\beta_0-\beta)'\left\{\frac{1}{n}X' S(\rho)'\left(\Sigma^{-1}\otimes I_n\right)S(\rho)X - M_2(\theta_2,\ \theta_3)\right\}(\beta_0-\beta) \\ & + \frac{1}{2} - \frac{1}{2ng}\varepsilon'\left(\Sigma_0^{-1}\otimes I_n\right)\varepsilon. \end{aligned}$$

By Lemma B0, to prove consistency we need to show that for any $\delta > 0$, there exists $\eta > 0$ such that for some N,

$$\inf_{\|\theta-\theta_0\|\geq\delta} s_n(\theta) \geq \eta \text{ for all } n \geq N, \tag{2.49}$$

and as $n \to \infty$

$$\sup_{\theta \in \Theta} |t_n(\theta)| \xrightarrow{p} 0. \tag{2.50}$$

Note that Lemma B0 is a slight modification of the standard theorem for consistency of an extremum estimate.

First we show positivity of $s_n(\theta)$. For any $\theta_2 \in \Theta_2$ and $\theta_3 \in \Theta_3$, $M_2(\theta_2,\ \theta_3)$ is p.s.d. as outlined above. Therefore for any $\theta \in \Theta$, $(\beta_0 - \beta)' M_2(\theta_2,\ \theta_3)(\beta_0 - \beta) \geq 0$. Recall that a function $f : \mathbb{R}^+ \to \mathbb{R}$ where $f(x) = x - \log x - 1$ is always positive, i.e. $f(x) \geq 0$. Our extension for $x = 0$ also gives $f(0) = \infty$. Since

$$\begin{aligned} & (ng)^{-1} \{tr[H(\theta_2,\ \theta_3)] - \log|H(\theta_2,\ \theta_3)| - ng\} \\ = \quad & (ng)^{-1} \sum_{i=1}^{ng} (\lambda_i - \log \lambda_i - 1), \end{aligned}$$

where λ_i are eigenvalues of $H(\theta_2,\ \theta_3)$, it follows that

$$\frac{1}{2ng} tr\{H(\theta_2,\ \theta_3)\} - \frac{1}{2ng} \log|H(\theta_2,\ \theta_3)| - \frac{1}{2} \geq 0$$

for all n and $\theta \in \Theta$. Hence $s_n(\theta) \geq 0$ for all n and $\theta \in \Theta$.

Let $\tau = (\theta_2,\ \theta_3) \in \Theta_2 \times \Theta_3$. Assumption B5 implies that for any $\delta > 0$, there exists $\eta > 0$ such that for some N,

$$\begin{aligned} & \inf_{\|\tau - \tau_0\| \geq \delta} \left\{ \frac{1}{ng} tr\{H(\theta_2,\ \theta_3)\} - \frac{1}{ng} \log|H(\theta_2,\ \theta_3)| - 1 \right\} \\ = \quad & \inf_{\|\tau - \tau_0\| \geq \delta} \left\{ \frac{1}{ng} \sum_{i=1}^{ng} (\lambda_i - \log \lambda_i - 1) \right\} \geq \eta \end{aligned}$$

for all $n \geq N$. Since this holds true for all $\beta \in \Theta_1$, to show (2.49) it suffices to show that when $\tau = \tau_0$, i.e. $(\theta_2,\ \theta_3)' = (\theta_{02},\ \theta_{03})'$

$$\inf_{\|\beta - \beta_0\| \geq \delta} \left((\beta_0 - \beta)' M_2(\theta_{02},\ \theta_{03})(\beta_0 - \beta) \right) \geq \eta.$$

This is indeed the case due to Lemma B7. Hence, (2.49) holds. With compactness of Θ_1, Lemmas B2, B4-B6 imply that (2.50) holds.

Proof of Theorem C Under Assumptions B1, B3-B7, the Gaussian-pseudo maximum likelihood estimate is consistent, i.e. $\widehat{\theta} - \theta_0 = o_p(1)$. With Assumptions C4 and C5, to prove Theorem C, it suffices to show that

$$\sqrt{ng} \frac{\partial Q_n(\theta_0)}{\partial \theta} \to_d N(0,\ \Omega).$$

Define

$$a_n = \sqrt{ng} \frac{\partial Q_n(\theta_0)}{\partial \theta},$$

and $a_{i,n} = \sqrt{ng} \frac{\partial Q_n(\theta_0)}{\partial \theta_i},\ i = 1,\ 2,\ 3$. Then $\Omega_n = \mathbb{E}(a_n a_n' | X)$. Assumption C1 implies that

$\Omega_n \to_p \Omega$ where Ω is p.d.. Hence Ω_n will be p.d. with probability approaching 1. For any $\lambda \in \mathbb{R}^{K+g(g+1)/2+g}$ such that $\|\lambda\| = 1$, we need to show that

$$\left(\lambda'\Omega_n\lambda\right)^{-1/2} \lambda' a_n \to_d N\left(0,\ 1\right). \tag{2.51}$$

Suppose $\lambda = \left(\lambda_1',\ \lambda_2',\ \lambda_3'\right)'$ where $\lambda_1 \in \mathbb{R}^K,\ \lambda_2 \in \mathbb{R}^{g(g+1)/2}$ and $\lambda_3 \in \mathbb{R}^g$. Then

$$\lambda' a_n = \lambda_1' a_{1,n} + \lambda_2' a_{2,n} + \lambda_3' a_{3,n}.$$

It follows that

$$\lambda_1' a_{1,n} = -\left(ng\right)^{-1/2} \sum_{i=1}^{n} \sum_{s=1}^{g} \sum_{t=1}^{g} \sigma_0^{st} \lambda_1' x_{is}^{*} \varepsilon_{it}, \tag{2.52}$$

where $x_{is}^{*\prime}$ is the i-th row of X_s^*. Each element of $a_{2,n}$ is of the form

$$\left(4ng\right)^{-1/2} \left(2 - \delta_{st}\right) \sum_{i=1}^{n} \left(\varepsilon_{is}\varepsilon_{it} - \sigma_0^{st}\right). \tag{2.53}$$

By symmetry of Σ_0^{-1}, each element of $a_{3,n}$ is of the form

$$\left(ng\right)^{-1/2} \sum_{i=1}^{n} g_{iit} - \left(ng\right)^{-1/2} \sum_{s=1}^{g} \sigma_0^{st} \left(\sum_{i=1}^{n} \sum_{j=1}^{n} g_{ijt}\varepsilon_{is}\varepsilon_{jt} \right),$$

where g_{ijt} is the (i, j)-th element of G_t. This can be rewritten as

$$-\left(ng\right)^{-1/2} \sum_{i=1}^{n} \left\{ \left(\sum_{s=1}^{g} \sigma_0^{st} g_{iit}\varepsilon_{is}\varepsilon_{it} - g_{iit} \right) + \sum_{j \neq i} \sum_{s=1}^{g} \sigma_0^{st} g_{ijt}\varepsilon_{is}\varepsilon_{jt} \right\}. \tag{2.54}$$

Since $\sum_{s=1}^{g} \sigma_0^{st}\sigma_{0st} = 1$, the expectation of the sum in the parentheses is 0.

Let $\mathcal{F}_0$ be the trivial σ-field and $\mathcal{F}_i = \sigma\left(\varepsilon_{1\cdot},\ ...,\ \varepsilon_{i\cdot}\right)$ be the σ-field generated by $\varepsilon_{1\cdot},\ ...,\ \varepsilon_{i\cdot}$, where $\varepsilon_{i\cdot} = \left(\varepsilon_{i1},\ ...,\ \varepsilon_{ig}\right)'$. Conditional on X, following (2.52), (2.53) and (2.54), there exist random variables b_{in} such that

$$\left(\lambda'\Omega_n\lambda\right)^{-1/2} \lambda' a_n = \sum_{i=1}^{n} b_{in},$$

where $\{b_{in},\ 1 \le i \le n\}$ is a martingale difference sequence for each n, i.e. conditional on X, $\mathbb{E}\left(b_{in}\middle|\mathcal{F}_{i-1}\right) = 0$. Then, by Theorem 2 of Scott (1973), (2.51) holds if conditional on X, as $n \to \infty$,

$$\sum_{i=1}^{n} \mathbb{E}\left(b_{in}^2\middle|\mathcal{F}_{i-1}\right) \to_p 1 \tag{2.55}$$

and for any $\varepsilon > 0$,

$$\sum_{i=1}^{n} \mathbb{E}\left\{b_{in}^2 1\left(|b_{in}| \ge \varepsilon\right)\right\} \to_p 0. \tag{2.56}$$

Define

$$z_{i,n} = \left(\lambda'\Omega_n\lambda\right)^{1/2} b_{in}.$$

Then, conditional on X, $\lambda'\Omega_n\lambda = \sum_{i=1}^n \mathbb{E}\left(z_{i,n}\right)^2$ and a sufficient condition for (2.55) is

$$\sum_{i=1}^n \left\{\mathbb{E}\left(z_{i,n}^2\middle|\mathcal{F}_{i-1}\right) - \mathbb{E}\left(z_{i,n}^2\right)\right\} \to_p 0, \tag{2.57}$$

because, by Assumption C1, $\lambda'\Omega_n\lambda \to_p \lambda'\Omega\lambda > 0$. Let

$$z_{i,n} = z_{i1,n} + z_{i2,n} + z_{i3,n},$$

where $z_{i1,n}$, $z_{i2,n}$ and $z_{i3,n}$ correspond to (2.52), (2.53) and (2.54), respectively. It is clear from (2.52) and (2.53) that, conditional on X, $\mathbb{E}\left\{\left(z_{i1,n} + z_{i2,n}\right)^2\middle|\mathcal{F}_{i-1}\right\} = \mathbb{E}\left\{\left(z_{i1,n} + z_{i2,n}\right)^2\right\}$ for all i. Hence, conditional on X,

$$\sum_{i=1}^n \left\{\mathbb{E}\left(\left(z_{i1,n} + z_{i2,n}\right)^2\middle|\mathcal{F}_{i-1}\right) - \mathbb{E}\left(z_{i1,n} + z_{i2,n}\right)^2\right\} \to_p 0.$$

Hence for (2.57) to hold, it suffices to show that, conditional on X,

$$\sum_{i=1}^n \left[\mathbb{E}\left(z_{i3,n}^2\middle|\mathcal{F}_{i-1}\right) - \mathbb{E}\left(z_{i3,n}^2\right)\right] \to_p 0.$$

and

$$\sum_{i=1}^n \left\{\mathbb{E}\left[\left(z_{i1,n} + z_{i2,n}\right) z_{i3,n}\middle|\mathcal{F}_{i-1}\right] - \mathbb{E}\left[\left(z_{i1,n} + z_{i2,n}\right) z_{i3,n}\right]\right\} \to_p 0.$$

Lemma C16-C18 imply that these conditions hold. Hence (2.55) holds. To show that (2.56) holds, it suffices to show the Lyapunov condition that, conditional on X, there is $\delta > 0$ such that

$$\sum_{i=1}^n \mathbb{E}\left|b_{i,n}\right|^{2+\delta} \to_p 0. \tag{2.58}$$

A sufficient condition for (2.58) is that, conditional on X, there is $\delta > 0$ such that

$$\sum_{i=1}^n \mathbb{E}\left(\left|z_{1i,n}\right|^{2+\delta} + \left|z_{2i,n}\right|^{2+\delta} + \left|z_{3i,n}\right|^{2+\delta}\right) \to_p 0.$$

Assumption C7 implies that this is the case for $z_{1i,n}$ and $z_{2i,n}$. With reference to the first part of the proof of Lemma C16, to show that $\sum_{i=1}^n \mathbb{E}\left|z_{3i,n}\right|^{2+\delta} = o_p(1)$, we only need to show that $\sum_{i=1}^n \mathbb{E}\left|d_{ist}\right|^{2+\delta} = o(1)$, where

$$d_{ist} = g_{iit}\left(\varepsilon_{is}\varepsilon_{it} - \sigma_{0st}\right) + \varepsilon_{is}\left(\sum_{j<i} g_{ijt}\varepsilon_{jt} + \sum_{j>i} g_{ijt}\varepsilon_{jt}\right).$$

Then the derivation from (A.24) to (A.26) in Robinson (2008) and Assumptions C6 and C7 imply that $\sum_{i=1}^n \mathbb{E}\left|d_{ist}\right|^{2+\delta} = o(1)$. Hence (2.58) holds.

Proof of Theorem D1 Our proof is essentially the same as the proofs of Theorems 1 and 4 in Newey (1997). So the repetitive steps will be omitted. Let $h_1 = h \circ T^{-1}$. As discussed in (2.24), the left side of (2.25) becomes

$$\int \left[h_1(\xi) - \widehat{h}_L(\xi)\right]^2 \ f_\xi(\xi) \ d(\xi),$$

where f_ξ is the probability density function of ξ,

$$\widehat{h}_L(\xi) = p^L(\xi)' (P'P)^{-1} P'y,$$

$$P = \left(p^L(\xi_1),\ ...,\ p^L(\xi_n)\right)' \quad \text{and} \quad y = (y_1,\ ...,\ y_n)'.$$

The difference of our proof to that of Theorem 1 in Newey (1997) are from equations (A.2) and (A.3) of Newey (1997). We proceed as in the proof of Theorem 4 in Newey (1997) by showing that Assumptions 1 and 2 in Newey (1997) holds and point out that the precise rate of convergence as indicated by Assumption 3 in Newey (1997) can be replaced by the approximating result in Dunkl and Xu (2001). The approximating result in Dunkl and Xu (2001) can replace the precise result from Newey (1997)'s Assumption 3 in equations (A.2) and (A.3).

First, Assumptions D1 and D4 implies that Assumption 1 in Newey (1997) holds. Note that $Var(y_i|\, x_i) = Var(y_i|\, \xi_i)$ since the σ-field generated by x_i and ξ_i are the same under our one-one restriction on T.

Next, we employ the observation made by Cox (1988), in a univariate case, that $\widehat{h}_L(\xi)$ is numerically invariant if we replace $p_l(\xi)$ by orthonormal polynomials with the corresponding order $\lambda(l)$. In our multivariate case, this replacement is valid since the sequence $\{\lambda(l)\}$ is assumed to be ordered. Hence $\widehat{h}_L(\xi)$ can be written as

$$\widehat{h}_L(\xi) = p_*^L(\xi)' (P_*'P_*)^{-1} P_*'y,$$

where

$$p_*^L(\xi) = (p_1^*(\xi),\ ...,\ p_L^*(\xi))', \quad P_* = \left(p_*^L(\xi_1),\ ...,\ p_*^L(\xi_n)\right)',$$

$$p_l^*(\xi) = z_{\lambda_1(l)}(\xi_1)\ ...\ z_{\lambda_g(\lambda)}\left(\xi_g\right),$$

$$z_{\lambda_t}(\xi_t) = \left(\frac{2\lambda_t + 1}{b_t - a_t}\right)^{1/2} P_{\lambda_t}^{(0,\ 0)}\left(\frac{2(\xi_t - a_t)}{b_t - a_t} - 1\right), \tag{2.59}$$

where $P_{\lambda_t}^{(0,\ 0)}$ are univariate Jacobi polynomials on $[-1,\ 1]$ with degree λ_l and parameter $(0,\ 0)$ (see Abramowitz and Stegun (1964, p. 775 eqn. 22.3.1) and Andrews (1991 eqn. 3.12)). That is to transform univariate polynomials in $\xi_t \in (a_t,\ b_t)$ to orthonormal polynomials with respect to the uniform weight on $[a_t,\ b_t]$. Hence, for $\xi \in \Pi_{t=1}^g (a_t,\ b_t)$ standard polynomials in ξ are replaced by Jacobi polynomials of the same degree that is orthonormal with respect to the uniform density on $\Pi_{t=1}^g(-1,\ 1)$. The usefulness of this transformation can be seen from Lemma D2 that $\underline{\lambda}\left(\mathbb{E}\left[p_*^L(\xi)\, p_*^L(\xi)'\right]\right) \geq C$ for all $L \geq 1$. Then we can follow the proof in Andrews (1991) to verify that other requirements in Assumption 2 of Newey

(1997) hold. Note that the condition $L^3/n = o(1)$ is needed to verify that Assumption 2 of Newey (1997) hold for polynomials. See also the proof of Theorem 4 in Newey (1997) as a reference.

Finally, we need to replace β_K in Assumption 3 of Newey (1997) by d_L such that

$$\int \left[h_1(\xi) - d_L' p_*^L(\xi)\right]^2 \; dF_\xi(\xi) \to 0 \;\text{ as }\; L \to \infty, \tag{2.60}$$

from the result in Dunkl and Xu (2001). Note that since $\mathbb{E}\left[h_1(\xi)^2\right] < \infty$ and ξ are bounded, conditions (2.22) and (2.23) of Dunkl and Xu (2001) hold for polynomials in ξ with respect to its distribution function. Hence, by Theorem 3.1.18 in Dunkl and Xu (2001), there is a triangular array $\left\{d_L \in \mathbb{R}^L : L \geq 1\right\}$ such that (2.60) holds for polynomials in ξ. Since there is a one-one correspondence between polynomials in ξ and orthonormal polynomials in ξ, there is a triangular array $\{d_L : L \geq 1\}$ for orthonormal polynomials too. The replacement should be taken everywhere β appears in the proof of Theorem 1 of Newey (1997) by our δ_L where the sup norm should be replaced by the L^2 norm in the sense of Dunkl and Xu (2001) and the Markov's inequality can be applied. Then it follows that the conclusion of the theorem holds.

Proof of Theorem D2 It has been shown in the proof of Theorem D1 that Assumptions 1 and 2 of Newey (1997) holds. To show that Theorem D2 holds, it remains to show that Assumptions 3 of Newey (1997) holds. Certainly, Assumption D6 is analogous to Assumption 9 of Newey (1997) implying that Assumption 3 of Newey (1997) holds. Hence, the result follows directly from the proof of Theorem 4 in Newey (1997).

Proof of Theorem D3 Suppose a transformation T has the form (2.27) where m_t are in the class $\mathcal{E}$ and Assumption D9 holds, it follows from Lemma D4 that Assumption D8 holds and under Assumption D9 $T(\mathcal{X}) = (-1,\ 0)^g$. Hence T is bounded. By Lemma D3, T is one-one and with Assumption D9,

$$f_\xi(\xi) = f_X\left(T^{-1}(\xi)\right) \prod_{t=1}^{g} \left[m_t'\left(m_t^{-1}(\xi_t)\right)\right]^{-1},$$

where f_ξ is continuous on $T(\mathcal{X})$ and $f_\xi(\xi) > 0$ for all ξ in $T(\mathcal{X})$. To show that Theorem D3 holds, it suffices to show that there is a constant $C > 0$ such that $f_\xi(\xi) \geq C$ for all ξ in $T(\mathcal{X})$. To achieve this, we employ both Lemmas D5 and D6.

First fix $t = 1,\ \ldots,\ g$, and $\xi_{-t} \in (-1,\ 0)^{g-1}$. Consider

$$\begin{aligned} \lim_{\xi_t \to -1} f_\xi(\xi) &= \lim_{\xi_t \to -1} f_X\left(T^{-1}(\xi)\right) \prod_{s=1}^{g} \left[m_s'\left(m_s^{-1}(\xi_s)\right)\right]^{-1} \\ &= \lim_{x_t \to -\infty} \frac{f_X(x)}{\Pi_{s=1}^{g} m_s'(x_s)}, \end{aligned}$$

where $x = T^{-1}(\xi)$. By condition (2.30), there are functions q_s such that

$$\lim_{x_t \to -\infty} \frac{f_X(x)}{\Pi_{s=1}^{g} m_s'(x_s)} = \lim_{x \to -\infty} \left\{ \frac{f_X(x)}{\Pi_{s=1}^{g} \exp(q_s(x_s))} \left[\Pi_{s=1}^{g} \frac{\exp(q_s(x_s))}{m_s'(x_s)} \right] \right\}.$$

Since

$$m_t'(x_t) = \frac{\exp(g(x_t))\, g'(x_t)}{\left[1 + \exp(g(x_t))\right]^2},$$

it follows that

$$\frac{\exp(q_t(x_t))}{m_t'(x_t)} = \frac{\left[1 + \exp(g(x_t))\right]^2}{\exp\left[g(x_t) - q_t(x_t) + \log(g'(x_t))\right]}.$$

Since $g(x_t) \to -\infty$ as $x_t \to -\infty$,

$$\lim_{x_t \to -\infty} \left[1 + \exp(g(x_t))\right]^2 = 1.$$

Under condition (2.32),

$$\lim_{x \to -\infty} \exp\left[g(x_t) - q_t(x_t) + \log(g'(x_t))\right] = 0,$$

and thus

$$\lim_{x_t \to -\infty} \frac{\exp(q_t(x_t))}{m_t'(x_t)} = \infty.$$

Hence under condition (2.30),

$$\lim_{\xi_t \to -1} f_\xi(\xi) = \lim_{x_t \to -\infty} \frac{f_X(x)}{\Pi_{s=1}^{g} m_s'(x_s)} = \infty.$$

Now consider

$$\begin{aligned} \lim_{\xi_t \to 0} f_\xi(\xi) &= \lim_{x_t \to \infty} \frac{f_X(x)}{\Pi_{s=1}^{g} m_s'(x_s)} \\ &= \lim_{x \to \infty} \left\{ \frac{f_X(x)}{\Pi_{s=1}^{g} \exp(q_s(x_s))} \left[\Pi_{s=1}^{g} \frac{\exp(q_s(x_s))}{m_s'(x_s)} \right] \right\} \end{aligned}$$

Since

$$m_t'(x_t) = \frac{\exp(g(x_t))\, g'(x_t)}{\left[1 + \exp(g(x_t))\right]^2} > 0,$$

it follows that

$$\begin{aligned} \frac{\exp(q_t(x_t))}{m_t'(x_t)} &\geq \frac{\exp(g(x_t))^2}{\exp(g(x_t))\, g'(x_t)} \exp(q_t(x_t)) \\ &= \exp\left[g(x_t) + q_t(x_t) - \log(g'(x_t))\right]. \end{aligned}$$

Hence under conditions (2.30) and (2.31),

$$\lim_{\xi_t \to 0} f_\xi(\xi) = \lim_{x_t \to \infty} \frac{f_X(x)}{\Pi_{s=1}^{g} m_s'(x_s)} = \infty.$$

As this result holds for all $t = 1, \; ..., \; g$, and $\xi_{-t} \in (-1, \; 0)^{g-1}$, it follows from Lemma

D6 that for any y in the boundary of $T(\mathcal{X}) = (-1,\ 0)^g$,

$$\lim_{\xi \to y} f_\xi(\xi) = \infty.$$

Hence, as mentioned earlier that f_ξ is continuous on $T(\mathcal{X})$ and $f_\xi(\xi) > 0$ for all ξ in $T(\mathcal{X})$, by Lemma D5, there is a constant $C > 0$ such that $f_\xi(\xi) \geq C$ for all ξ in $T(\mathcal{X})$ as required.

Proof of Theorem E Our proof is quite different from those of Newey (1988) and Robinson (2005, 2010) for a number of reasons. First, we have to work with orthonormal polynomials. Second, we have to focus on obtaining relatively sharper results. Our proof is also different from that in Newey (1997) since we do not have an explicit regression form.

First, recall that for $v \in \mathbb{R}^g$ and $\lambda(l) \in \mathbb{N}_0^g$,

$$p_l(v) = v^{\lambda(l)}$$

is a monomial in $v_1,\ ...,\ v_g$ with total degree $\|\lambda(l)\|_1$ as described in the previous section. By (2.43) and Assumption E3, we only have to consider $v_t \in (a_t,\ b_t)$, where a_t and b_t are finite constants. By Assumption E7, the sequence $\{\lambda(l)\}$ is ordered. Then for any $v \in \Pi_{t=1}^g (a_t,\ b_t)$, there is a non-singular matrix B_{1L} of constants such that

$$p_*^L(v) = B_{1L} p^L(v),$$

where $p_*^L(v)$ is an $L \times 1$ vector of multivariate orthonormal Jocobi polynomials in v on $[-1,\ 1]^g$, with respect to the uniform weight, and with the corresponding order $\lambda(l)$. For $t = 1,\ ...,\ g$, since $\pi_t^L(u) = \frac{\partial}{\partial u_t} p^L(T(u))$, it follows that

$$B_{1L}\pi_t^L(u) = \frac{\partial}{\partial u_t} B_{1L} p^L(T(u)) = \frac{\partial}{\partial u_t} p_*^L(T(u)).$$

Define $\pi_{*t}^L(u) = \frac{\partial}{\partial u_t} p_*^L(T(u))$. Then $\widehat{\psi}_{tL}(\widetilde{\varepsilon}_{i\cdot})$ constructed from the standard multivariate polynomials is numerically the same as when Jacobi orthonormal polynomials are employed. That is

$$\widehat{\psi}_{tL}(\widetilde{\varepsilon}_{i\cdot}) = p_*^L(T(\widetilde{\varepsilon}_{i\cdot}))' \left(\sum_{j=1}^n p_*^L(T(\widetilde{\varepsilon}_{j\cdot})) p_*^L(T(\widetilde{\varepsilon}_{j\cdot}))' \right)^{-1} \left(\sum_{j=1}^n \pi_{*t}^L(\widetilde{\varepsilon}_{j\cdot}) \right).$$

The advantage of this approach is that it will substantially reduce multi-colinearity of the approximating functions. Under Assumption E3, Lemma D2 implies that there is a constant $C > 0$ such that for all $L \geq 1$, $\underline{\lambda}\big(\mathbb{E}\left[p_*^L(T(\varepsilon_{1\cdot})) p_*^L(T(\varepsilon_{1\cdot}))'\right]\big) \geq C$. Hence we can define

$$B_{2L} = \left\{\mathbb{E}\left[p_*^L(T(\varepsilon_{1\cdot})) p_*^L(T(\varepsilon_{1\cdot}))'\right]\right\}^{-1/2}$$

and B_{2L} is positive definite for all L. For $u,\ v \in \mathbb{R}^g$ and $t = 1,\ ...,\ g$, let

$$p_{**}^L(v) = B_{2L} p_*^L(v) \quad \text{and} \quad \pi_{**t}^L(u) = B_{2L}\pi_{*t}^L(u) = \frac{\partial}{\partial u_t} p_{**}^L(T(u)). \tag{2.61}$$

Then, it follows that

$$\widehat{\psi}_{tL}\left(\widetilde{\varepsilon}_{i\cdot}\right)=p_{**}^{L}\left(T\left(\widetilde{\varepsilon}_{i\cdot}\right)\right)'\left(\sum_{j=1}^{n}p_{**}^{L}\left(T\left(\widetilde{\varepsilon}_{j\cdot}\right)\right)p_{**}^{L}\left(T\left(\widetilde{\varepsilon}_{j\cdot}\right)\right)'\right)^{-1}\left(\sum_{j=1}^{n}\pi_{**t}^{L}\left(\widetilde{\varepsilon}_{j\cdot}\right)\right).$$

This step is employed in Newey (1997) to help increase the rate at which L can go to infinity. One advantage of this step is that

$$\mathbb{E}\left[p_{**}^{L}\left(T\left(\varepsilon_{1\cdot}\right)\right)p_{**}^{L}\left(T\left(\varepsilon_{1\cdot}\right)\right)'\right]=I_{L}, \tag{2.62}$$

where I_L is the identity matrix of order L, without seriously contaminating other terms since

$$\begin{aligned}\left\|B_{2L}\right\|^{2} &= \overline{\lambda}\left(\mathbb{E}\left[p_{*}^{L}\left(T\left(\varepsilon_{1\cdot}\right)\right)p_{*}^{L}\left(T\left(\varepsilon_{1\cdot}\right)\right)'\right]^{-1}\right)\\ &= \left\{\underline{\lambda}\left(\mathbb{E}\left[p_{*}^{L}\left(T\left(\varepsilon_{1\cdot}\right)\right)p_{*}^{L}\left(T\left(\varepsilon_{1\cdot}\right)\right)'\right]\right)\right\}^{-1}\\ &\leq 1/C\end{aligned}$$

uniformly in L. Hence $\|B_{2L}\|$ is bounded uniformly in L.

With this expression, define

$$\widetilde{\gamma}_{tL}=\left(\sum_{j=1}^{n}p_{**}^{L}\left(T\left(\widetilde{\varepsilon}_{j\cdot}\right)\right)p_{**}^{L}\left(T\left(\widetilde{\varepsilon}_{j\cdot}\right)\right)'\right)^{-1}\left(\sum_{j=1}^{n}\pi_{**t}^{L}\left(\widetilde{\varepsilon}_{j\cdot}\right)\right). \tag{2.63}$$

Hence

$$\widehat{\psi}_{tL}\left(\widetilde{\varepsilon}_{i\cdot}\right)=\left(\widetilde{\gamma}_{tL}\right)'p_{**}^{L}\left(T\left(\widetilde{\varepsilon}_{i\cdot}\right)\right). \tag{2.64}$$

Let

$$\widetilde{\Gamma}_{L}=\left(\widetilde{\gamma}_{1L},\ ...,\ \widetilde{\gamma}_{gL}\right)'. \tag{2.65}$$

Then

$$\widehat{\psi}_{L}\left(\widetilde{\varepsilon}_{i\cdot}\right)=\widetilde{\Gamma}_{L}\ p_{**}^{L}\left(T\left(\widetilde{\varepsilon}_{i\cdot}\right)\right). \tag{2.66}$$

Now we can start the standard procedure as in Newey (1988) and Robinson (2005, 2010). For $u\in\mathbb{R}^{g}$, define

$$\Pi_{**}^{L}\left(u\right)=\frac{\partial}{\partial u'}p_{**}^{L}\left(T\left(u\right)\right)=\left(\pi_{**1}^{L}\left(u\right),\ ...,\ \pi_{**g}^{L}\left(u\right)\right). \tag{2.67}$$

Define

$$R_{i}\left(\beta\right)=diag\left\{\sum_{j=1}^{n}w_{ij1}u_{j1}\left(\beta\right),\ ...,\ \sum_{j=1}^{n}w_{ijg}u_{jg}\left(\beta\right)\right\}, \tag{2.68}$$

where $u_{jt}\left(\beta\right)=y_{jt}-x_{jt}'\beta$. By the mean value theorem around θ_{04},

$$p_{**}^{L}\left(T\left(\widetilde{\varepsilon}_{i\cdot}\right)\right)=p_{**}^{L}\left(T\left(\varepsilon_{i\cdot}\right)\right)-\Pi_{**}^{L}\left(\overline{\varepsilon}_{i\cdot}\right)\left[\left(\widetilde{\alpha}-\alpha_{0}\right)+X_{i\cdot}^{*}\left(\overline{\rho}\right)\left(\widetilde{\beta}-\beta_{0}\right)+R_{i}\left(\overline{\beta}\right)\left(\widetilde{\rho}-\rho_{0}\right)\right], \tag{2.69}$$

where $\bar{\varepsilon}_{i\cdot} = \varepsilon_{i\cdot}\left(\bar{\theta}_4\right)$ for some $\bar{\theta}_4$ such that $\left\|\bar{\theta}_4 - \theta_{04}\right\| < \left\|\widetilde{\theta}_4 - \theta_{04}\right\|$.

Let $\widehat{V}_n = n^{-1}\sum_{i=1}^n \left[X_{i\cdot}^*\left(\widetilde{\rho}\right) - \overline{X}_\cdot^*\left(\widetilde{\rho}\right)\right]' \widehat{\mathcal{L}}_L \left[X_{i\cdot}^*\left(\widetilde{\rho}\right) - \overline{X}_\cdot^*\left(\widetilde{\rho}\right)\right]$. Then, by (2.66) and (2.69),

$$\begin{aligned} n^{1/2}\left(\widehat{\beta} - \beta_0\right) &= \widehat{V}_n^{-1} n^{-1/2} \sum_{i=1}^n \left[X_{i\cdot}^*\left(\widetilde{\rho}\right) - \overline{X}_\cdot^*\left(\widetilde{\rho}\right)\right]' \widetilde{\Gamma}_L\, p_{**}^L\left(T\left(\varepsilon_{i\cdot}\right)\right) \\ &\quad + \left[I_K - \widehat{V}_n^{-1} n^{-1} \sum_{i=1}^n \left[X_{i\cdot}^*\left(\widetilde{\rho}\right) - \overline{X}_\cdot^*\left(\widetilde{\rho}\right)\right]' \widetilde{\Gamma}_L \Pi_{**}^L\left(\bar{\varepsilon}_{i\cdot}\right) X_{i\cdot}^*\left(\bar{\rho}\right)\right] n^{1/2}\left(\widetilde{\beta} - \beta_0\right) \\ &\quad - \widehat{V}_n^{-1} n^{-1} \sum_{i=1}^n \left[X_{i\cdot}^*\left(\widetilde{\rho}\right) - \overline{X}_\cdot^*\left(\widetilde{\rho}\right)\right]' \widetilde{\Gamma}_L \Pi_{**}^L\left(\bar{\varepsilon}_{i\cdot}\right) \sqrt{n}\left(\widetilde{\alpha} - \alpha_0\right) \\ &\quad - \widehat{V}_n^{-1} n^{-1} \sum_{i=1}^n \left[X_{i\cdot}^*\left(\widetilde{\rho}\right) - \overline{X}_\cdot^*\left(\widetilde{\rho}\right)\right]' \widetilde{\Gamma}_L \Pi_{**}^L\left(\bar{\varepsilon}_{i\cdot}\right) R_i\left(\bar{\beta}\right) \sqrt{n}\left(\widetilde{\rho} - \rho_0\right). \end{aligned}$$

By Assumption E8, to proof Theorem E, it suffices to show that

$$\widehat{V}_n \to_p V, \tag{E.1}$$

$$n^{-1/2}\sum_{i=1}^n \left[X_{i\cdot}^*\left(\widetilde{\rho}\right) - \overline{X}_\cdot^*\left(\widetilde{\rho}\right)\right]' \widetilde{\Gamma}_L\, p_{**}^L\left(T\left(\varepsilon_{i\cdot}\right)\right) - n^{-1/2}\sum_{i=1}^n \left(X_{i\cdot}^* - \overline{X}_\cdot^*\right)' \psi\left(\varepsilon_{i\cdot}\right) \to_p 0. \tag{E.2}$$

$$n^{-1/2}\sum_{i=1}^n \left(X_{i\cdot}^* - \overline{X}_\cdot^*\right)' \psi\left(\varepsilon_{i\cdot}\right) \to_d N\left(0,\ V\right). \tag{E.3}$$

$$n^{-1}\sum_{i=1}^n \left(X_{i\cdot}^*\left(\widetilde{\rho}\right) - \overline{X}_\cdot^*\left(\widetilde{\rho}\right)\right)' \widetilde{\Gamma}_L \Pi_{**}^L\left(\bar{\varepsilon}_{i\cdot}\right) X_{i\cdot}^*\left(\bar{\rho}\right) - V \to_p 0, \tag{E.4}$$

$$n^{-1}\sum_{i=1}^n \left(X_{i\cdot}^*\left(\widetilde{\rho}\right) - \overline{X}_\cdot^*\left(\widetilde{\rho}\right)\right)' \widetilde{\Gamma}_L \Pi_{**}^L\left(\bar{\varepsilon}_{i\cdot}\right) R_i\left(\bar{\beta}\right) \to_p 0, \tag{E.5}$$

and

$$n^{-1}\sum_{i=1}^n \left(X_{i\cdot}^*\left(\widetilde{\rho}\right) - \overline{X}_\cdot^*\left(\widetilde{\rho}\right)\right)' \widetilde{\Gamma}_L \Pi_{**}^L\left(\bar{\varepsilon}_{i\cdot}\right) \to_p 0, \tag{E.6}$$

Hence Propositions E1 - E6 conclude the proof.

Appendix 2.2: Propositions for Proof of Theorem E

Proposition E1 *As $n \to \infty$,*

$$\widehat{\mathcal{L}}_L \to_p \mathcal{L} \quad \textit{and} \quad \widehat{V}_n \to_p V.$$

Proof. For $t = 1,\ \ldots,\ g$, let

$$\psi_{tL}\left(\varepsilon_{i\cdot}\right) = \gamma_{tL}'\, p_{**}^L\left(T\left(\varepsilon_{i\cdot}\right)\right), \tag{2.70}$$

where

$$\gamma_{tL} = \mathbb{E}\left[p_{**}^{L}\left(T\left(\varepsilon_{1\cdot}\right)\right)\psi_{t}\left(\varepsilon_{1\cdot}\right)\right] \tag{2.71}$$

Let

$$\psi_{L}\left(\varepsilon_{i\cdot}\right) = \left(\psi_{1L}\left(\varepsilon_{i\cdot}\right),\ ...,\ \psi_{gL}\left(\varepsilon_{i\cdot}\right)\right)'. \tag{2.72}$$

By Lemma E3, as $n \to \infty$,

$$\mathbb{E}\left[\psi_{L}\left(\varepsilon_{1\cdot}\right)\psi_{L}\left(\varepsilon_{1\cdot}\right)'\right] \to \mathcal{L}. \tag{2.73}$$

Hence, to show the first part of Proposition **??**, it suffices to show that, as $n \to \infty$,

$$\widehat{\mathcal{L}}_{L} \to_{p} \mathbb{E}\left[\psi_{L}\left(\varepsilon_{i\cdot}\right)\psi_{L}\left(\varepsilon_{i\cdot}\right)'\right]. \tag{2.74}$$

Let

$$\widetilde{I}_{L} = n^{-1}\sum_{i=1}^{n} p_{**}^{L}\left(T\left(\widetilde{\varepsilon}_{i\cdot}\right)\right)p_{**}^{L}\left(T\left(\widetilde{\varepsilon}_{i\cdot}\right)\right)'$$

Then, with reference to (2.63) and (2.64),

$$\widetilde{\gamma}_{tL} = \widetilde{I}_{L}^{-1}\left(n^{-1}\sum_{i=1}^{n}\pi_{**t}^{L}\left(\widetilde{\varepsilon}_{i\cdot}\right)\right) \quad \text{and} \quad \widehat{\psi}_{tL}\left(\widetilde{\varepsilon}_{i\cdot}\right) = \widetilde{\gamma}_{tL}' p_{**}^{L}\left(T\left(\widetilde{\varepsilon}_{i\cdot}\right)\right). \tag{2.75}$$

For $s,\ t = 1,\ ...,\ g$, by definition of $p_{**}^{L}\left(T\left(\varepsilon_{i\cdot}\right)\right)$,

$$\mathbb{E}\left[\psi_{sL}\left(\varepsilon_{1\cdot}\right)\psi_{tL}\left(\varepsilon_{1\cdot}\right)\right] = \gamma_{sL}'\mathbb{E}\ \left[p_{**}^{L}\left(T\left(\varepsilon_{1\cdot}\right)\right)p_{**}^{L}\left(T\left(\varepsilon_{1\cdot}\right)\right)'\right]\gamma_{tL} = \gamma_{sL}'\gamma_{tL},$$

and, by (2.75),

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}\widehat{\psi}_{sL}\left(\widetilde{\varepsilon}_{i\cdot}\right)\widehat{\psi}_{tL}\left(\widetilde{\varepsilon}_{i\cdot}\right) \\ = \quad & \widetilde{\gamma}_{sL}'\widetilde{I}_{L}\widetilde{\gamma}_{tL} = \left(n^{-1}\sum_{j=1}^{n}\pi_{**s}^{L}\left(\widetilde{\varepsilon}_{j\cdot}\right)\right)'\widetilde{I}_{L}^{-1}\left(n^{-1}\sum_{j=1}^{n}\pi_{**t}^{L}\left(\widetilde{\varepsilon}_{j\cdot}\right)\right). \end{aligned}$$

Hence sufficient conditions for (2.74) are that for $s,\ t = 1,\ ...,\ g$,

$$\begin{aligned} & n^{-1}\sum_{i=1}^{n}\widehat{\psi}_{sL}\left(\widetilde{\varepsilon}_{i\cdot}\right)\widehat{\psi}_{tL}\left(\widetilde{\varepsilon}_{i\cdot}\right) - \mathbb{E}\left[\psi_{sL}\left(\varepsilon_{1\cdot}\right)\psi_{tL}\left(\varepsilon_{1\cdot}\right)\right] \\ = \quad & \widetilde{\gamma}_{sL}'\widetilde{I}_{L}\widetilde{\gamma}_{tL} - \gamma_{sL}'\gamma_{tL} = o_{p}(1). \end{aligned} \tag{2.76}$$

For $t = 1,\ ...,\ g$, let

$$\widetilde{\phi}_{tL} = n^{-1}\sum_{j=1}^{n}\pi_{**t}^{L}\left(\widetilde{\varepsilon}_{j\cdot}\right). \tag{2.77}$$

Then the left side of (2.76) is

$$\widetilde{\phi}_{sL}'\widetilde{I}_{L}^{-1}\widetilde{\phi}_{tL} - \gamma_{sL}'\gamma_{tL}.$$

Lemma E9 and Assumption E9 imply that (2.76) holds, and thus, as $n \to \infty$, $\widehat{\mathcal{L}}_{L} - \mathcal{L} =$

$o_p(1)$.

Since

$$\widehat{V}_n = n^{-1}\sum_{i=1}^{n}\left(X_{i\cdot}^{*}(\rho)-\overline{X}_{\cdot}^{*}(\rho)\right)'\left(\widehat{\mathcal{L}}_L-\mathcal{L}\right)\left(X_{i\cdot}^{*}(\rho)-\overline{X}_{\cdot}^{*}(\rho)\right) + n^{-1}\sum_{i=1}^{n}\left(X_{i\cdot}^{*}(\rho)-\overline{X}_{\cdot}^{*}(\rho)\right)'\mathcal{L}\left(X_{i\cdot}^{*}(\rho)-\overline{X}_{\cdot}^{*}(\rho)\right). \quad (2.78)$$

By Lemma A1, the norm of the first term in (2.78) is bounded by

$$\left\|\widehat{\mathcal{L}}_L-\mathcal{L}\right\|\left(n^{-1}\sum_{i=1}^{n}\left\|X_{i\cdot}^{*}(\rho)-\overline{X}_{\cdot}^{*}(\rho)\right\|^2\right)=o_p(1)$$

by the previously established result and Lemma E6. The second term in (2.78) is

$$\begin{aligned}
&n^{-1}\sum_{i=1}^{n}\left[\left(X_{i\cdot}^{*}(\rho)-\overline{X}_{\cdot}^{*}(\rho)\right)-\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)\right]'\mathcal{L}\left[\left(X_{i\cdot}^{*}(\rho)-\overline{X}_{\cdot}^{*}(\rho)\right)-\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)\right]\\
&+n^{-1}\sum_{i=1}^{n}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)'\mathcal{L}\left[\left(X_{i\cdot}^{*}(\rho)-\overline{X}_{\cdot}^{*}(\rho)\right)-\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)\right]\\
&+n^{-1}\sum_{i=1}^{n}\left[\left(X_{i\cdot}^{*}(\rho)-\overline{X}_{\cdot}^{*}(\rho)\right)-\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)\right]'\mathcal{L}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)\\
&+n^{-1}\sum_{i=1}^{n}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)'\mathcal{L}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right) \qquad (2.79)
\end{aligned}$$

By Cauchy's inequality, the norm of the first three terms in (2.79) is bounded by

$$\begin{aligned}
&\|\mathcal{L}\|\left(n^{-1}\sum_{i=1}^{n}\left\|\left(X_{i\cdot}^{*}(\rho)-\overline{X}_{\cdot}^{*}(\rho)\right)-\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)\right\|^2\right)\\
&+\|\mathcal{L}\|\left(n^{-1}\sum_{i=1}^{n}\left\|\left(X_{i\cdot}^{*}(\rho)-\overline{X}_{\cdot}^{*}(\rho)\right)-\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)\right\|^2\right)^{1/2}\left(n^{-1}\sum_{i=1}^{n}\left\|X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right\|^2\right)^{1/2}\\
=\ & o_p(1),
\end{aligned}$$

by Lemmas E5 and E6 (ii). Hence by Assumption E5, $\widehat{V}_n \to_p V$ as required. ■

Proposition E2 *As* $n\to\infty$,

$$n^{-1/2}\sum_{i=1}^{n}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)'\psi(\varepsilon_{i\cdot})-n^{-1/2}\sum_{i=1}^{n}\left(X_{i\cdot}^{*}(\widetilde{\rho})-\overline{X}_{\cdot}^{*}(\widetilde{\rho})\right)'\widetilde{\Gamma}_L\ p_{**}^{L}\left(T(\varepsilon_{i\cdot})\right)=o_p(1).$$

Proof. Let $\widehat{\psi}_L(\varepsilon_{i\cdot})=\widetilde{\Gamma}_L\ p_{**}^{L}(T(\varepsilon_{i\cdot}))$ and $\psi_L(\varepsilon_{i\cdot})$ is defined as (2.66). The left side of the lemma is

$$+n^{-1/2}\sum_{i=1}^{n}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)'\left[\psi\left(\varepsilon_{i\cdot}\right)-\psi_{L}\left(\varepsilon_{i\cdot}\right)\right]+n^{-1/2}\sum_{i=1}^{n}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)'\left[\psi_{L}\left(\varepsilon_{i\cdot}\right)-\widehat{\psi}_{L}\left(\varepsilon_{i\cdot}\right)\right]$$
$$-n^{-1/2}\sum_{i=1}^{n}\left[\left(X_{i\cdot}^{*}\left(\widetilde{\rho}\right)-\overline{X}_{\cdot}^{*}\left(\widetilde{\rho}\right)\right)-\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)\right]'\widehat{\psi}_{L}\left(\varepsilon_{i\cdot}\right).$$

Lemmas E10 - E12 imply that each of these terms are $o_p(1)$. Hence the required result holds. ∎

Proposition E3 *As* $n\to\infty$,

$$n^{-1/2}\sum_{i=1}^{n}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)'\psi\left(\varepsilon_{i\cdot}\right)\to_{d}N\left(0,\ V\right). \tag{2.80}$$

Proof. Define $V_n=n^{-1}\sum_{i=1}^{n}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)'\mathcal{L}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)$. By Assumption E5, V_n is positive definite with probability approaching one. For any λ in $\mathbb{R}^K$ such that $\|\lambda\|=1$, define

$$c_i'=c_{in}'=n^{-1/2}\lambda' V_n^{-1/2}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)'.$$

To show (2.80), it suffices to show that as $n\to\infty$,

$$\sum_{i=1}^{n}c_i'\psi\left(\varepsilon_{i\cdot}\right)\to_{d}N\left(0,\ 1\right). \tag{2.81}$$

Our proof modifies the proof of Theorem 2 in Robinson and Hidalgo (1997). It suffices to show that (2.81) holds by showing that conditionally on $\{X_{i\cdot}^{*}\}$, (2.81) holds. Let $\mathcal{F}_0$ be the trivial σ-field and $\mathcal{F}_i=\sigma\left(\varepsilon_{1\cdot},\ ...,\ \varepsilon_{i\cdot}\right)$ be the σ-field generated by $\varepsilon_{1\cdot},\ ...,\ \varepsilon_{i\cdot}$. Conditional on $\{X_{i\cdot}^{*}\}$, for each $n\geq 1$, $\{c_i'\psi\left(\varepsilon_{i\cdot}\right),1\leq i\leq n\}$ is a martingale difference sequence, i.e. $\mathbb{E}\left\{c_i'\psi\left(\varepsilon_{i\cdot}\right)\middle|\mathcal{F}_{i-1}\right\}=0$. To show (2.81), following Scott (1973), it suffices to show that, conditional on $\{X_{i\cdot}^{*}\}$, as $n\to\infty$,

$$\sum_{i=1}^{n}\mathbb{E}\left\{\left[c_i'\psi\left(\varepsilon_{i\cdot}\right)\right]^2\middle|\mathcal{F}_{i-1}\right\}\to_{p}1, \tag{2.82}$$

and, for all $\eta>0$,

$$\mathbb{E}\left\{\sum_{i=1}^{n}\mathbb{E}\left\{\left[c_i'\psi\left(\varepsilon_{i}\right)\right]^2\mathbf{1}\left(\left|c_i'\psi\left(\varepsilon_{i}\right)\right|>\eta\right)\middle|\{X_{i\cdot}^{*}\}\right\}\right\}\to 0. \tag{2.83}$$

It follows from the way in which c_i are defined that (2.82) holds. For any $\delta>0$, under

Assumption E1, the left side of (2.83) is bounded by

$$\mathbb{E}\left\{\sum_{i=1}^{n}\|c_i\|^2\,\mathbb{E}\left[\|\psi(\varepsilon_{1\cdot})\|^2\,\mathbf{1}\left(\|\psi(\varepsilon_{1\cdot})\|>\eta/\delta\right)\right]\right\}+\mathbb{P}\left\{\max_{1\le i\le n}\|c_i\|>\delta\right\}$$
$$=\mathbb{E}\left[\|\psi(\varepsilon_{1\cdot})\|^2\,\mathbf{1}\left(\|\psi(\varepsilon_{1\cdot})\|>\eta/\delta\right)\right]\sum_{i=1}^{n}\mathbb{E}\|c_i\|^2+\mathbb{P}\left\{\max_{1\le i\le n}\|c_i\|>\delta\right\}. \qquad (2.84)$$

Since

$$\sum_{i=1}^{n}\mathbb{E}\|c_i\|^2=\lambda' V_n^{-1/2}\left[n^{-1}\sum_{i=1}^{n}\mathbb{E}\left(X_{i\cdot}^*-\overline{X}_{\cdot}^*\right)'\left(X_{i\cdot}^*-\overline{X}_{\cdot}^*\right)\right]V_n^{-1/2}\lambda,$$

Assumption E5 implies that $\sum_{i=1}^{n}\mathbb{E}\|c_i\|^2=O(1)$. Note that

$$\mathbb{E}\|\psi(\varepsilon_{1\cdot})\|^2=\mathbb{E}\left\{tr\left[\psi(\varepsilon_{i\cdot})'\psi(\varepsilon_{i\cdot})\right]\right\}=tr\left\{\mathbb{E}\left[\psi(\varepsilon_{i\cdot})\psi(\varepsilon_{i\cdot})'\right]\right\}=tr\mathcal{L}<\infty.$$

This and the fact that $\sum_{i=1}^{n}\mathbb{E}\|c_i\|^2=O(1)$ implies that the term on the right of (2.84) can be made arbitrarily small by choosing δ small enough, so it suffices for (2.83) to show that $\max_{1\le i\le n}\|c_i\|=o_p(1)$. Since $\left\|\overline{X}_{\cdot}^*\right\|\le\max_{1\le i\le n}\|X_{i\cdot}^*\|$, by Assumption E5

$$\max_{1\le i\le n}\|c_i\|\le\left\|V_n^{-1}\right\|\left(n^{-1/2}\max_{1\le i\le n}\|2X_{i\cdot}^*\|\right)=o_p(1).$$

■

Proposition E4 *As* $n\to\infty$,

$$n^{-1}\sum_{i=1}^{n}\left(X_{i\cdot}^*(\widetilde{\rho})-\overline{X}_{\cdot}^*(\widetilde{\rho})\right)'\widetilde{\Gamma}_L\Pi_{**}^L(\overline{\varepsilon}_{i\cdot})X_{i\cdot}^*(\overline{\rho})\to_p V.$$

Proof. It follows directly from Lemmas E13 - E16 that Proposition E4 holds. ■

Proposition E5 *As* $n\to\infty$,

$$n^{-1}\sum_{i=1}^{n}\left(X_{i\cdot}^*(\widetilde{\rho})-\overline{X}_{\cdot}^*(\widetilde{\rho})\right)'\widetilde{\Gamma}_L\Pi_{**}^L(\overline{\varepsilon}_{i\cdot})R_i\left(\overline{\beta}\right)\to_p 0$$

Proof. Proceed as in the proof of Proposition E4 to show that

$$n^{-1}\sum_{i=1}^{n}\left[\left(X_{i\cdot}^*(\widetilde{\rho})-\overline{X}_{\cdot}^*(\widetilde{\rho})\right)'\widetilde{\Gamma}_L\Pi_{**}^L(\overline{\varepsilon}_{i\cdot})R_i\left(\overline{\beta}\right)-\left(X_{i\cdot}^*-\overline{X}_{\cdot}^*\right)'\widetilde{\Gamma}_L\Pi_{**}^L(\varepsilon_{i\cdot})R_i\right]=o_p(1).$$

Then Lemma E17 concludes the proposition. ■

Proposition E6 *As $n \to \infty$,*

$$n^{-1} \sum_{i=1}^{n} \left(X_{i\cdot}^{*} \left(\widetilde{\rho} \right) - \overline{X}_{\cdot}^{*} \left(\widetilde{\rho} \right) \right)' \widetilde{\Gamma}_{L} \Pi_{**}^{L} \left(\overline{\varepsilon}_{i\cdot} \right) \to_{p} 0.$$

Proof. The proof is similar to the proof of Proposition E4 but simpler. ■

Appendix 2.3: Technical Lemmas for proofs of Theorems

Lemma A1 *For any matrix A, $\|A'\| = \|A\|$.*

Proof. Let A be an $m \times n$ matrix. Exercise 7.25 of Abadir and Magnus (2005) implies that for $\lambda \neq 0$, $|\lambda I_n - A'A| = \lambda^{n-m} |\lambda I_m - AA'|$. If $A'A$ only has zero eigenvalues, then AA' must also have only zero eigenvalues. Otherwise, this equality will lead to a contradiction. In this case $\|A'\| = \|A\|$. Suppose $A'A$ has a nonzero eigenvalue. Then this nonzero number is also an eigenvalue of AA'. Let $\mathcal{A}_1$, $\mathcal{A}_2$ be the sets of all nonzero eigenvalues of $A'A$ and AA', respectively. If $\lambda \in \mathcal{A}_1$, then the above equality implies that $\lambda \in \mathcal{A}_2$. The converse is also true, and hence $\mathcal{A}_1 = \mathcal{A}_2$. It follows that $0 \neq \|A\|^2 = \overline{\lambda}(A'A) = \overline{\lambda}(AA') = \|A'\|^2$. Hence the required equality holds. ■

Lemma A2 *Let A be a symmetric matrix. Then λ is an eigenvalue of $A'A$ if and only if $\lambda = \omega^2$, where ω is an eigenvalue of A.*

Proof. Suppose $\lambda = \omega^2$ where ω is an eigenvalue of A. Then $0 = |A - \omega I_n| |A + \omega I_n| = |A'A - \lambda I_n|$. Hence λ is an eigenvalue of $A'A$. Conversely, suppose λ is an eigenvalue of $A'A$. Since $A'A$ is p.s.d., $\lambda \geq 0$. It follows that $0 = |A'A - \lambda I_n| = \left| A - \sqrt{\lambda} I_n \right| \left| A + \sqrt{\lambda} I_n \right|$. That is either $\sqrt{\lambda}$ is an eigenvalue of A or $-\sqrt{\lambda}$ is an eigenvalue of A. Hence $\lambda = \omega^2$ where ω is an eigenvalue of A. ■

Lemma A3 *Let A be a symmetric matrix. Then $\|A\| = \rho(A)$, and*

$$\|A\| \leq \|A\|_1 \text{ and } \|A\| \leq \|A\|_\infty .$$

Proof. By definition $\|A\|^2 = \overline{\lambda}(A'A)$. Since A is symmetric, Lemma A2 implies that there is an eigenvalue ω of A such that $\omega^2 = \overline{\lambda}(A'A)$. However, $|\omega|$ must be equal to $\rho(A)$, otherwise Lemma A2 will implies a contradiction. Hence $\|A\| = \rho(A)$. As it can be shown that $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are matrix norms as defined in Horn and Johnson (1985). Then by

Theorem 5.6.9 of Horn and Johnson (1985), $\|A\| = \rho(A) \leq \|A\|_1$. Similarly it follows that $\|A\| \leq \|A\|_\infty$. ■

Lemma A4 *Let A be a square matrix of order n, c_1, c_2 be constants and $B = c_1 I_n + c_2 A$, where I_n is the identity matrix of order n. Then λ is an eigenvalue of B if and only if $\lambda = c_1 + c_2\omega$ where ω is an eigenvalue of A.*

Proof. Suppose $\lambda = c_1 + c_2\omega$ where ω is an eigenvalue of A. It follows that

$$|B - \lambda I_n| = |c_2 A - c_2 \omega I_n| = (c_2)^n |A - \omega I_n| = 0.$$

Hence λ is an eigenvalue of B. Conversely, suppose that λ is an eigenvalue of B. If $c_2 = 0$, then it is trivial that $\lambda = c_1 + c_2\omega$ where ω is an eigenvalue of A. If $c_2 \neq 0$, then

$$0 = |B - \lambda I_n| = |c_2 A - (\lambda - c_1) I_n| = (c_2)^n \left| A - \frac{\lambda - c_1}{c_2} I_n \right|.$$

Hence, there is ω, an eigenvalue of A, such that $\omega = (\lambda - c_1)/c_2$. It follows that $\lambda = c_1 + c_2\omega$ where ω is an eigenvalue of A as required. ■

Lemma A5 *If W is symmetric and $S = I_n - \rho_0 W$ is invertible, then*

$$\left\| (S'S)^{-1} \right\| = \max \left\{ (1 - \rho_0 \omega)^{-2} : \omega \text{ is an eigenvalue of } W \right\}.$$

Proof. It follows from Lemma A4 that λ is an eigenvalue of S if and only if $\lambda = 1 - \rho_0\omega$, where ω is an eigenvalue of W. Invertibility of S implies that $1 - \rho_0\omega \neq 0$ for all eigenvalues ω of W, and that $S'S$ is p.d.. Suppose that W is symmetric. Then S is also symmetric and, by Lemmas A2 and A3,

$$\begin{aligned} \left\| (S'S)^{-1} \right\| &= \overline{\lambda} \left\{ (S'S)^{-1} \right\} = \left\{ \underline{\lambda} (S'S) \right\}^{-1} \\ &= \max \left\{ (1 - \rho_0 \omega)^{-2} : \omega \text{ is an eigenvalue of } W \right\}. \end{aligned}$$

■

Lemma A6 *Suppose $\rho_0 \neq 0$. (i) λ is an eigenvalue of $G = WS^{-1}$ if and only if $\lambda = \rho_0^{-1}(\omega - 1)$ where ω is an eigenvalue of S^{-1}. (ii) For any real number ρ_0,*

$$\left\| S^{-1} \right\| - 1 \leq \|\rho_0 G\| \leq \left\| S^{-1} \right\| + 1$$

Proof. From the definition of G, it follows that

$$I_n = SS^{-1} = (I_n - \rho_0 W) S^{-1} = S^{-1} - \rho_0 G$$

and, given that $\rho_0 \neq 0$,

$$G = \rho_0^{-1} \left(S^{-1} - I_n \right).$$

Lemma A4 implies that Lemma A6 (i) holds. As $\rho_0 G = S^{-1} - I_n$, by the property of a matrix norm,

$$\|\rho_0 G\| \leq \left\| S^{-1} \right\| + \|I_n\| = \left\| S^{-1} \right\| + 1.$$

Similarly, $S^{-1} = I_n + \rho_0 G$ and thus

$$\left\| S^{-1} \right\| \leq 1 + \|\rho_0 G\|.$$

■

Lemma A7 *Let A is a square matrix of order g. (i) λ is an eigenvalue of $A \otimes I_n$ if and only if λ is an eigenvalue of A. (ii) If A is p.d., then $A \otimes I_n$ is also p.d., and $\|A\| = \|A \otimes I_n\|$.*

Proof. Consider $|(A \otimes I_n) - \lambda I_{ng}| = |(A \otimes I_n) - \lambda (I_g \otimes I_n)| = |(A - \lambda I_g) \otimes I_n| = |A - \lambda I_g|^n$. Hence Lemma A7 (i) holds. Now suppose A is p.d., i.e. A is symmetric and its eigenvalues are all positive. Symmetry of A implies symmetry of $A \otimes I_n$. Lemma A7 (i) implies that all eigenvalues of $A \otimes I_n$ are also positive and hence $A \otimes I_n$ is p.d.. Moreover, by Lemma A3, $\|A\| = \overline{\lambda}(A) = \overline{\lambda}(A \otimes I_n) = \|A \otimes I_n\|$. ■

Lemma A8 *Let $A_1, \ ..., \ A_g$ be $n \times n$ matrices and $A = diag\{A_1, \ ..., \ A_g\}$. Then $\|A\| = \max_{1 \leq t \leq g} \|A_t\|$.*

Proof. Since $A'A = diag\left\{A_1'A_1, \ ..., \ A_g'A_g\right\}$, $|A'A - \lambda I_{ng}| = \Pi_{t=1}^{g} |A_t'A_t - \lambda I_n|$. This implies that λ is an eigenvalue of $A'A$ if and only if λ is an eigenvalue of $A_t'A_t$ for some $t = 1, \ ..., \ g$. Hence

$$\|A\|^2 = \overline{\lambda}(A'A) = \max_{1 \leq t \leq g} \overline{\lambda}(A_t'A_t) = \max_{1 \leq t \leq g} \|A_t\|^2.$$

This implies that Lemma A8 holds. ■

Lemma A9 *For any $a \in \mathbb{R}^n$,*

$$\sqrt{a'a} = \sup\{a'x : x \in \mathbb{R}^n, \ x'x \leq 1\}.$$

Proof. For a given a, let $f(x) = a'x$. Consider a maximization problem $Max \ f(x)$ subject to $x \in D = \{x \in \mathbb{R}^n : 1 - x'x \geq 0\}$. Since

$$\frac{\partial^2}{\partial x \partial x'} f(x) = 0 \ \text{ and } \ \frac{\partial^2}{\partial x \partial x'} (1 - x'x) = -2I_n,$$

both $f(x)$ and the constraint $1 - x'x$ are concave. Moreover, $1 - x_1'x_1 > 0$ for $x_1 = (1/2,\ 0,\ ...,\ 0)'$. Then it follows from the Kuhn-Tucker theorem that x^* is a solution to the maximization problem if and only if there is $\lambda^* \in \mathbb{R}$ such that

$$a - 2\lambda^* x^* = 0, \quad \lambda^* \geq 0 \quad \text{and} \quad \lambda^* \left\{1 - (x^*)' x^*\right\} = 0.$$

As $x^* = a(a'a)^{-1/2}$ and $\lambda^* = (a'a/4)^{1/2}$ satisfy the sufficient and necessary conditions of the Kuhn-Tucker theorem, it follows that $f(x^*) = (a'a)^{1/2}$. ∎

Lemma A10 *Let A be a square matrix. Then $\overline{\lambda}(A'A) \geq \rho(A)^2$.*

Proof. Let λ_1 be an eigenvalue of A such that $|\lambda_1| = \rho(A)$. Let x_1 be the corresponding eigenvector of λ_1. By symmetry of $A'A$,

$$\overline{\lambda}(A'A) \geq \frac{x_1'A'Ax_1}{x_1'x_1} = \frac{(\lambda_1 x_1)' \lambda_1 x_1}{x_1'x_1} = \lambda_1^2 = \rho(A)^2.$$

∎

Lemma B0 *Let*

$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\min} Q_n(\theta),$$

where Θ is a compact subset of R^p. If (i) $\theta_0 \in \Theta$, (ii)

$$Q_n(\theta) - Q_n(\theta_0) = s_n(\theta) - t_n(\theta),$$

where $s_n(\theta)$ is nonstochastic, (iii) for any $\varepsilon > 0$, there exists $\eta > 0$ such that for some N,

$$\inf_{\|\theta - \theta_0\| \geq \varepsilon} s_n(\theta) \geq \eta$$

for all $n \geq N$, and (iv) $\sup_{\theta \in \Theta} |t_n(\theta)| \xrightarrow{p} 0$ as $n \to \infty$, then

$$\hat{\theta} \xrightarrow{p} \theta_0$$

as $n \to \infty$.

Proof. For any $\varepsilon > 0$, let $\mathcal{N} = \{\theta : \|\theta - \theta_0\| < \varepsilon\}$ and $\mathcal{N}^c = \Theta \backslash \mathcal{N}$. For $n \geq N$,

$$\begin{aligned}
\mathbb{P}\left\{\left\|\hat{\theta} - \theta_0\right\| \geq \varepsilon\right\} &\leq \mathbb{P}\left\{\inf_{\mathcal{N}^c} [Q_n(\theta) - Q_n(\theta_0)] \leq 0\right\} \\
&= \mathbb{P}\left\{\inf_{\mathcal{N}^c} [s_n(\theta) - t_n(\theta)] \leq 0\right\} \\
&\leq \mathbb{P}\left\{\inf_{\mathcal{N}^c} s_n(\theta) - \sup_{\mathcal{N}^c} |t_n(\theta)| \leq 0\right\} \\
&\leq \mathbb{P}\left\{\inf_{\mathcal{N}^c} s_n(\theta) \leq \sup_{\theta \in \Theta} |t_n(\theta)|\right\} \\
&\leq \mathbb{P}\left\{\eta \leq \sup_{\theta \in \Theta} |t_n(\theta)|\right\}.
\end{aligned}$$

Hence $\mathbb{P}\left\{\left\|\hat{\theta} - \theta_0\right\| \geq \varepsilon\right\} \to 0$ as $n \to \infty$. ■

Lemma B1 *Let* $x = \left(x_1', \ ..., \ x_g'\right)', \ y = \left(y_1', \ ..., \ y_g'\right)'$ *where* $x_t, \ y_t \in \mathbb{R}^n, \ t = 1, \ ..., \ g$. *Let* σ^{st} *be the* (s,t)*-th element of* Σ^{-1}. *Then*

$$x' \left(S^{-1}\right)' S(\rho)' \left(\Sigma^{-1} \otimes I_n\right) S(\rho) S^{-1} y = \sum_{s=1}^{g} \sum_{t=1}^{g} \sigma^{st} x_s' \left(S_s^{-1}\right)' S_s(\rho)' S_t(\rho) S_t^{-1} y_t, \quad (2.85)$$

where

$$\begin{aligned}
& \left(S_s^{-1}\right)' S_s(\rho)' S_t(\rho) S_t^{-1} \\
= \ & I_n - (\rho_s - \rho_{0s}) G_s' - (\rho_t - \rho_{0t}) G_t + (\rho_s - \rho_{0s})(\rho_t - \rho_{0t}) G_s' G_t. \qquad (2.86)
\end{aligned}$$

Proof. The (s,t)-th submatrix of $\left(S^{-1}\right)' S(\rho)' \left(\Sigma^{-1} \otimes I_n\right) S(\rho) S^{-1}$ is $\sigma^{st} \left(S_s^{-1}\right)' S_s(\rho)' S_t(\rho) S_t^{-1}$. Hence (2.85) follows. Recall that $G_t = W_t S_t^{-1}$. Then,

$$S_t(\rho) S_t^{-1} = \{S_t - (\rho_t - \rho_{0t}) W_t\} S_t^{-1} = I_n - (\rho_t - \rho_{0t}) G_t,$$

and hence (2.86) follows. ■

Lemma B2 *Let* σ_{0st} *be the* (s,t)*-th element of* Σ_0. *For any family of nonstochastic* $n \times n$ *matrices* $\{A_n\}_{n \geq 1}$ *such that, as* $n \to \infty$, $\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ijn}^2 = o\left(n^2\right)$, *where* a_{ijn} *is the* (i,j)*-th element of* A_n,

$$\frac{1}{n} \varepsilon_{\cdot s}' A_n \varepsilon_{\cdot t} - \frac{1}{n} tr(\sigma_{0st} A_n) = o_p(1), \quad s, \ t = 1, \ ..., \ g.$$

Proof. Consider

$$\frac{1}{n} \varepsilon_{\cdot s}' A_n \varepsilon_{\cdot t} - \frac{1}{n} tr(\sigma_{0st} A_n) = \frac{1}{n} \sum_{i=1}^{n} a_{iin} (\varepsilon_{is} \varepsilon_{it} - \sigma_{0st}) + \frac{1}{n} \sum\sum_{i \neq j} a_{ijn} \varepsilon_{is} \varepsilon_{jt}. \quad (2.87)$$

For the first term in (2.87), by Assumption B3,

$$\begin{aligned}\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}a_{iin}\left(\varepsilon_{is}\varepsilon_{it}-\sigma_{0st}\right)\right]^{2} &\leq 2\left\{\max_{1\leq t\leq g}\max_{i\geq 1}\mathbb{E}\left(\varepsilon_{it}^{4}\right)+\sigma_{0st}^{2}\right\}\frac{1}{n^{2}}\sum_{i=1}^{n}a_{iin}^{2} \\ &\leq Cn^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ijn}^{2}=o\left(1\right).\end{aligned}$$

Similarly, the mean square of the second term in (2.87) is

$$\begin{aligned}\mathbb{E}\left(\frac{1}{n}\sum\sum_{i\neq j}a_{ijn}\varepsilon_{is}\varepsilon_{jt}\right)^{2} &= \frac{1}{n^{2}}\left\{\sum\sum_{i\neq j}a_{ijn}^{2}\mathbb{E}\varepsilon_{is}^{2}\mathbb{E}\varepsilon_{jt}^{2}+\sum\sum_{i\neq j}a_{ijn}a_{jin}\mathbb{E}\left(\varepsilon_{is}\varepsilon_{it}\right)\mathbb{E}\left(\varepsilon_{js}\varepsilon_{jt}\right)\right\} \\ &\leq Cn^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ijn}^{2}+Cn^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|a_{ijn}a_{jin}\right|. \end{aligned} \tag{2.88}$$

The first term in (2.88) is $o\left(1\right)$ .By Cauchy's inequality, the second sum in (2.88) is bounded by

$$\sum_{i=1}^{n}\left(\sum_{j=1}^{n}a_{ijn}^{2}\right)^{1/2}\left(\sum_{j=1}^{n}a_{jin}^{2}\right)^{1/2}\leq\left(\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ijn}^{2}\right)^{1/2}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}a_{jin}^{2}\right)^{1/2}=o\left(n^{2}\right).$$

Therefore the second term in (2.88) is $o\left(1\right)$, and hence $n^{-1}\varepsilon_{\cdot s}'A_{n}\varepsilon_{\cdot t}-n^{-1}tr\left(\sigma_{0st}A_{n}\right)=o_{p}\left(1\right)$ as $n\to\infty$. ■

Lemma B3 *Consider any two independent families of* $\mathbb{R}^{n}$*-valued random variables* $\{x_{n}\}_{n\geq 1}$ *and* $\{u_{n}\}_{n\geq 1}$*, where* $\mathbb{E}u_{n}=0$ *for all* $n\geq 1$*. For each* $n\geq 1$*, let* x_{in} *and* u_{in} *be the i-th elements of* x_{n} *and* u_{n}*, respectively. As* $n\to\infty$*, if* $\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left(x_{in}x_{jn}\right)Cov\left(u_{in},\ u_{jn}\right)=o\left(n^{-2}\right)$*, then*

$$n^{-1}x_{n}'u_{n}=o_{p}\left(1\right).$$

Proof. By independence of x_{in} and u_{in}, and the fact that $\mathbb{E}\left(u_{n}\right)=0$,

$$\mathbb{E}\left(x_{n}'u_{n}\right)^{2}=\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left(x_{in}x_{jn}\right)Cov\left(u_{in},\ u_{jn}\right).$$

Hence the required result holds. ■

Lemma B4 *As* $n\to\infty$,

$$\sup_{\tau\in\Theta_{2}\times\Theta_{3}}\left|\frac{1}{2ng}u'S\left(\rho\right)'\left(\Sigma^{-1}\otimes I_{n}\right)S\left(\rho\right)u-\frac{1}{2ng}tr\left\{H\left(\theta_{2},\ \theta_{3}\right)\right\}\right|=o_{p}\left(1\right) \tag{2.89}$$

.

Proof. Let σ_{0st} and σ^{st} be the (s,t)-th element of Σ_0 and Σ, respectively. It follows, also from symmetry, that

$$\begin{aligned} tr\left\{H\left(\theta_2,\ \theta_3\right)\right\} &= tr\left\{\left(S^{-1}\right)' S\left(\rho\right)'\left(\Sigma^{-1}\otimes I_n\right) S\left(\rho\right) S^{-1}\left(\Sigma_0\otimes I_n\right)\right\} \\ &= \sum_{s=1}^{g}\sum_{t=1}^{g}\sigma_{0st}\sigma^{st} tr\left\{\left(S_s^{-1}\right)' S_s\left(\rho\right)' S_t\left(\rho\right) S_t^{-1}\right\}. \end{aligned}$$

The difference inside the modulus in (2.89) is

$$\begin{aligned} &\frac{1}{2ng}\varepsilon'\left(S^{-1}\right)' S\left(\rho\right)'\left(\Sigma^{-1}\otimes I_n\right) S\left(\rho\right) S^{-1}\varepsilon - \frac{1}{2ng} tr\left\{H\left(\theta_2,\ \theta_3\right)\right\} \\ = \ &\frac{1}{2ng}\sum_{s=1}^{g}\sum_{t=1}^{g}\sigma^{st}\left[\varepsilon_{\cdot s}'\left(S_s^{-1}\right)' S_s\left(\rho\right)' S_t\left(\rho\right) S_t^{-1}\varepsilon_{\cdot t} - \sigma_{0st} tr\left\{\left(S_s^{-1}\right)' S_s\left(\rho\right)' S_t\left(\rho\right) S_t^{-1}\right\}\right], \end{aligned}$$

where the last equality follows from Lemma B1. For fixed s and t, Lemma B1 implies that

$$\begin{aligned} &n^{-1}\varepsilon_{\cdot s}'\left(S_s^{-1}\right)' S_s\left(\rho\right)' S_t\left(\rho\right) S_t^{-1}\varepsilon_{\cdot t} - n^{-1}\sigma_{0st} tr\left\{\left(S_s^{-1}\right)' S_s\left(\rho\right)' S_t\left(\rho\right) S_t^{-1}\right\} \\ = \ &\left\{n^{-1}\varepsilon_{\cdot s}'\varepsilon_{\cdot t} - n^{-1} tr\left(\sigma_{0st} I_n\right)\right\} - \left(\rho_s - \rho_{0s}\right)\left\{n^{-1}\varepsilon_{\cdot s}' G_s'\varepsilon_{\cdot t} - n^{-1} tr\left(\sigma_{0st} G_s'\right)\right\} \\ &- \left(\rho_t - \rho_{0t}\right)\left\{n^{-1}\varepsilon_{\cdot s}' G_t\varepsilon_{\cdot t}' - n^{-1} tr\left(\sigma_{0st} G_t\right)\right\} \\ &+ \left(\rho_s - \rho_{0s}\right)\left(\rho_t - \rho_{0t}\right)\left\{n^{-1}\varepsilon_{\cdot s}' G_s' G_t\varepsilon_{\cdot t}' - n^{-1} tr\left(\sigma_{0st} G_s' G_t\right)\right\}. \end{aligned}$$

Assumption B7 and Lemma B2 imply that term in the last curly brackets is $o_p(1)$. Let g_{ijt} be the (i,j)-th element of G_t. Note that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ijt}^2 = tr\left(G_t' G_t\right) = \sum_{i=1}^{n}\lambda_i\left(G_t' G_t\right),$$

where $\lambda_i\left(G_t' G_t\right)$ are eigenvalues of $G_t' G_t$. If all eigenvalues of $G_t' G_t$ is bounded uniformly in n, then $\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ijt}^2 = O(n) = o\left(n^2\right)$. Hence Lemma B2 implies that the other terms in curly brackets are also $o_p(1)$. If some of the eigenvalues of $G_t' G_t$ approaches infinity as n increases, then $\sum_{i=1}^{n}\lambda_i\left(G_t' G_t\right)$ will be dominated by

$$\sum_{i=1}^{n}\lambda_i\left(G_t' G_t\right)^2 = tr\left(G_t' G_t G_t' G_t\right) \le \sum_{i=1}^{ng}\sum_{j=1}^{ng} v_{ij}^2 = o\left(n^2\right).$$

Hence Lemma B2 implies that the other terms in curly brackets are also $o_p(1)$. Hence, Assumption B7 and Lemma B2 imply that all terms in curly brackets are $o_p(1)$. Positive definiteness of Σ_0, compactness of Θ_3 and the required property of Θ_2 in Assumption B4 imply that Lemma B4 holds. ■

Lemma B5 *Let x_k be the k-th column of X. As $n \to \infty$,*

$$\sup_{\tau\in\Theta_2\times\Theta_3}\left|\frac{1}{ng} x_k' S\left(\rho\right)'\left(\Sigma^{-1}\otimes I_n\right) S\left(\rho\right) u\right| = o_p(1).$$

Proof. Let σ^{st} be the (s,t)-th element of Σ^{-1}. Then the term in the absolute sign is $(ng)^{-1}(x_k^*)'\left(S^{-1}\right)'S(\rho)'\left(\Sigma^{-1}\otimes I_n\right)S(\rho)S^{-1}\varepsilon$, where x_k^* is the k-th column of X^*. Let $x_k^*=\left(\left(x_{k,1}^*\right)',\ ...,\ \left(x_{k,g}^*\right)'\right)'$, where $x_{k,t}^*$ are $\mathbb{R}^n$ random vectors. It follows from Lemma B1 that

$$\begin{aligned}
& (ng)^{-1}(x_k^*)'\left(S^{-1}\right)'S(\rho)'\left(\Sigma^{-1}\otimes I_n\right)S(\rho)S^{-1}\varepsilon \\
= \ & (ng)^{-1}\sum_{s=1}^{g}\sum_{t=1}^{g}\sigma^{st}\left(x_{k,s}^*\right)'\left(S_s^{-1}\right)'S_s(\rho)'S_t(\rho)S_t^{-1}\varepsilon_{\cdot t}.
\end{aligned} \qquad (2.90)$$

Lemma B1 implies that

$$\begin{aligned}
& n^{-1}\left(x_{k,s}^*\right)'\left(S_s^{-1}\right)'S_s(\rho)'S_t(\rho)S_t^{-1}\varepsilon_{\cdot t} \\
= \ & \left\{n^{-1}\left(x_{k,s}^*\right)'\varepsilon_{\cdot t}\right\}-(\rho_s-\rho_{0s})\left\{n^{-1}\left(x_{k,s}^*\right)'G_s'\varepsilon_{\cdot t}\right\}-(\rho_t-\rho_{0t})\left\{n^{-1}\left(x_{k,s}^*\right)'G_t\varepsilon_{\cdot t}\right\} \\
& +(\rho_s-\rho_{0s})(\rho_t-\rho_{0t})\left\{n^{-1}\left(x_{k,s}^*\right)'G_s'G_t\varepsilon_{\cdot t}\right\}.
\end{aligned}$$

Assumption B7 and Lemma B3 can be employed to show that the terms in curly brackets are all $o_p(1)$. With reference to (2.90), the property of Θ_2 and compactness of Θ_3 in Assumption B4 imply that Lemma B5 holds. ■

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

$$\sup_{\tau\in\Theta_2\times\Theta_3}\left\|\frac{1}{n}X'S(\rho)'\left(\Sigma^{-1}\otimes I_n\right)S(\rho)X-M_2(\theta_2,\ \theta_3)\right\|=o_p(1).$$

Proof. Let x_k^* be the k-th column of X^* and σ^{st} be the (s,t)-th element of Σ^{-1}. Proceeding as in the proof of Lemma B1, it follows that

$$\begin{aligned}
& n^{-1}X'S(\rho)'\left(\Sigma^{-1}\otimes I_n\right)S(\rho)X \\
= \ & n^{-1}(X^*)'\left(S^{-1}\right)'S(\rho)'\left(\Sigma^{-1}\otimes I_n\right)S(\rho)S^{-1}X^* \\
= \ & n^{-1}\sum_{s=1}^{g}\sum_{t=1}^{g}\sigma^{st}(X_s^*)'\left(S_s^{-1}\right)'S_s(\rho)'S_t(\rho)S_t^{-1}X_t^*.
\end{aligned}$$

Employing (2.86), it follows that

$$\begin{aligned}
& n^{-1}(X_s^*)'\left(S_s^{-1}\right)'S_s(\rho)'S_t(\rho)S_t^{-1}X_t^* \\
= \ & n^{-1}(X_s^*)'X_t^*+(\rho_{0s}-\rho_s)\left\{n^{-1}(X_s^*)'G_s'X_t^*\right\}+(\rho_{0t}-\rho_t)\left\{n^{-1}(X_s^*)'G_tX_t^*\right\} \\
& +(\rho_s-\rho_{0s})(\rho_t-\rho_{0t})\left\{n^{-1}(X_s^*)'G_s'G_tX_t^*\right\}.
\end{aligned}$$

Similarly, it can be shown, under Assumption B6, that, for $M_2(\theta_2,\ \theta_3)$ defined in (2.48),

$$M_2(\theta_2,\ \theta_3)=\sum_{s=1}^{g}\sum_{t=1}^{g}\sigma^{st}A_{st},$$

where

$$g^{-1}A_{st} = Q_{11}^{st} + (\rho_{0s} - \rho_s)\, Q_{21}^{st} + (\rho_{0t} - \rho_t)\, Q_{12}^{st} + (\rho_s - \rho_{0s})\,(\rho_t - \rho_{0t})\, Q_{22}^{st}.$$

Then Assumptions B4 and B6 imply that Lemma B6 holds. ■

Lemma B7 *For any* $\delta > 0$, *there exists* $\eta > 0$ *such that*

$$\inf_{\|\beta-\beta_0\|\geq\delta} (\beta - \beta_0)' M_2 (\theta_{02},\ \theta_{03})\,(\beta - \beta_0) \geq \eta.$$

Proof. Assumption B6 implies that $M_2(\theta_{02},\ \theta_{03})$ is a p.d. matrix O_{11}. Then

$$\begin{aligned} & \inf_{\|\beta-\beta_0\|\geq\delta} \left((\beta - \beta_0)' M_2 (\theta_{02},\ \theta_{03})\,(\beta - \beta_0)\right) \\ \geq\ & \left(\inf_{\|\beta-\beta_0\|\geq\delta} (\beta - \beta_0)'(\beta - \beta_0)\right)\left(\inf_{\|\beta-\beta_0\|\geq\delta} \frac{(\beta - \beta_0)' M_2 (\theta_{02},\ \theta_{03})\,(\beta - \beta_0)}{(\beta - \beta_0)'(\beta - \beta_0)}\right) \\ \geq\ & \delta^2 \underline{\lambda}(\Omega_{11}) > 0. \end{aligned}$$

■

Lemma B8 *If a sequence of non-negative random variables* $\{X_n\}$ *converges in probability to a constant* c, *then* $c \geq 0$.

Proof. Given the convergence, for any $\delta > 0$, $\lim_{n\to\infty} \mathbb{P}\{|X_n - c| > \delta\} = 0$. Suppose $c < 0$. Since $X_n \geq 0$ and $c < 0$, $|X_n - c| \geq |c|$. Hence $\mathbb{P}\{|X_n - c| > |c|/2\} = 1$ for all n. This leads to a contradiction if we set $\delta = |c|/2$. Therefore, $c \geq 0$. ■

Lemma B9 *Let* $\{A_n\}$ *be a sequence of p.s.d. matrices of order* K. *If* $A_n \to_p A$, *as* $n \to \infty$, *then* A *is also p.s.d..*

Proof. Clearly A must be symmetric. For any $y \in \mathbb{R}^K$ such that $\|y\| = 1$, $y'A_n y \to_p y'Ay$. Let $t_n = y'A_n y$. Then $\{t_n\}$ is a sequence of non-negative real numbers and hence Lemma B8 implies that $t = y'Ay \geq 0$. Since this hold for any $y \in \mathbb{R}^K$, A must be p.s.d.. ■

Lemma C1 *For* $\tau = 1,\ \ldots,\ g$,

$$\frac{\partial}{\partial\rho_\tau} u(\beta)' S(\rho)' \left(\Sigma^{-1} \otimes I_n\right) S(\rho)\, u(\beta) = -2\sum_{s=1}^{g} \sigma^{s\tau} u_{\cdot s}(\beta)' S_s(\rho)' W_\tau u_{\cdot\tau}(\beta).$$

Proof. The left side is

$$
\begin{aligned}
& \frac{\partial}{\partial \rho_\tau} \sum_{s=1}^{g} \sum_{t=1}^{g} \sigma^{st} u_{\cdot s}(\beta)' S_s(\rho)' S_t(\rho) u_{\cdot t}(\beta) \\
= \quad & -\sum_{s=1}^{g} \sigma^{s\tau} u_{\cdot s}(\beta)' S_s(\rho)' W_\tau u_{\cdot \tau}(\beta)(1-\delta_{s\tau}) - \sum_{s=1}^{g} \sigma^{\tau s} u_{\cdot \tau}(\beta)' W_\tau' S_s(\rho) u_{\cdot s}(\beta)(1-\delta_{\tau s}) \\
& + \sigma^{\tau\tau} u_{\cdot \tau}(\beta)' \left[-W_\tau' (I_n - \rho_\tau W_\tau) - (I_n - \rho_\tau W_\tau)' W_\tau \right] u_{\cdot \tau}(\beta) \\
= \quad & -\sum_{s=1}^{g} \sigma^{s\tau} u_{\cdot s}(\beta)' S_s(\rho)' W_\tau u_{\cdot \tau}(\beta) - \sum_{s=1}^{g} \sigma^{\tau s} u_{\cdot \tau}(\beta)' W_\tau' S_s(\rho) u_{\cdot s}(\beta) \\
= \quad & -2 \sum_{s=1}^{g} \sigma^{s\tau} u_{\cdot s}(\beta)' S_s(\rho)' W_\tau u_{\cdot \tau}(\beta),
\end{aligned}
$$

by symmetry of Σ^{-1}. ■

Lemma C2 *For ρ in a neighbourhood of ρ_0 such that $S_t(\rho)$ are non-singular for all $t = 1, \ ..., \ g$,*

$$
\frac{\partial}{\partial \rho_t} \sum_{s=1}^{g} \log \left| S_s(\rho)' S_s(\rho) \right| = -2tr \left\{ G_t(\rho) \right\}.
$$

Proof. The left side is

$$
\begin{aligned}
\frac{\partial}{\partial \rho_t} \log \left| S_t(\rho)' S_t(\rho) \right| \quad &= \quad tr \left\{ \left[S_t(\rho)' S_t(\rho) \right]^{-1} \left[-W_t' (I_n - \rho_t W_t) - (I_n - \rho_t W_t)' W_t \right] \right\} \\
&= \quad -tr \left\{ S_t(\rho)^{-1} \left(S_t(\rho)^{-1} \right)' \left[W_t' S_t(\rho) + S_t(\rho)' W_t \right] \right\} \\
&= \quad -tr \left[\left(S_t(\rho)^{-1} \right)' W_t' \right] - tr \left[S_t(\rho)^{-1} W_t \right] \\
&= \quad -2tr \left[G_t(\rho) \right].
\end{aligned}
$$

■

Lemma C3 *For any $u, \ v = 1, \ ..., \ g$,*

$$
\begin{aligned}
& \mathbb{E} \left\{ \frac{1}{ng} (X^*)' \left(\Sigma_0^{-1} \otimes I_n \right) \varepsilon \left(\varepsilon_{\cdot u}' \varepsilon_{\cdot v} - n\sigma_{0uv} \right) \bigg| X \right\} \qquad (2.91) \\
= \quad & \frac{1}{ng} \sum_{s=1}^{g} \sum_{t=1}^{g} \sigma_0^{st} \sum_{i=1}^{n} x_{is}^* \mathbb{E} \left\{ \varepsilon_{it} \left(\varepsilon_{iu} \varepsilon_{iv} - \sigma_{0uv} \right) \right\}.
\end{aligned}
$$

Proof. The left side is

$$
\begin{aligned}
& \frac{1}{ng} \sum_{s=1}^{g} \sum_{t=1}^{g} \sigma_0^{st} (X_s^*)' \mathbb{E} \left[\varepsilon_{\cdot t} \left(\varepsilon_{\cdot u}' \varepsilon_{\cdot v} - n\sigma_{0uv} \right) \right] \\
= \quad & \frac{1}{ng} \sum_{s=1}^{g} \sum_{t=1}^{g} \sigma_0^{st} \sum_{i=1}^{n} x_{is}^* \mathbb{E} \left[\varepsilon_{it} \left(\varepsilon_{\cdot u}' \varepsilon_{\cdot v} - n\sigma_{0uv} \right) \right],
\end{aligned}
$$

where $(x_{is}^*)'$ is the i-th row of X_s^*. Since

$$\mathbb{E}\left[\varepsilon_{it}\left(\varepsilon'_{\cdot u}\varepsilon_{\cdot v}-n\sigma_{0uv}\right)\right]=\mathbb{E}\left[\varepsilon_{it}\sum_{j=1}^{n}\left(\varepsilon_{ju}\varepsilon_{jv}-\sigma_{0uv}\right)\right]=\mathbb{E}\left[\varepsilon_{it}\left(\varepsilon_{iu}\varepsilon_{iv}-\sigma_{0uv}\right)\right],$$

(2.91) holds. ■

Lemma C4 *For* $\tau=1,\ \ldots,\ g,$

$$\begin{aligned}&\mathbb{E}\left\{\frac{1}{ng}\left(X^*\right)'\left(\Sigma_0^{-1}\otimes I_n\right)\varepsilon\left[\sum_{u=1}^{g}\sigma_0^{u\tau}\varepsilon'_{\cdot u}G_\tau\varepsilon_{\cdot\tau}-tr\left(G_\tau\right)\right]\middle|X\right\}\\ =\ &\frac{1}{ng}\sum_{s=1}^{g}\sum_{t=1}^{g}\sum_{u=1}^{g}\sigma_0^{st}\sigma_0^{u\tau}\sum_{i=1}^{n}x_{is}^*g_{ii\tau}\mathbb{E}\left(\varepsilon_{it}\varepsilon_{iu}\varepsilon_{i\tau}\right)\end{aligned}\tag{2.92}$$

Proof. The conditional expectation in (2.92) is

$$\frac{1}{ng}\sum_{s=1}^{g}\sum_{t=1}^{g}\sum_{i=1}^{n}\sigma_0^{st}x_{is}^*\mathbb{E}\left\{\varepsilon_{it}\left[\sum_{u=1}^{g}\sigma_0^{u\tau}\varepsilon'_{\cdot u}G_\tau\varepsilon_{\cdot\tau}-tr\left(G_\tau\right)\right]\right\}.$$

Since

$$\begin{aligned}\mathbb{E}\left\{\varepsilon_{it}\sum_{u=1}^{g}\left[\sigma_0^{u\tau}\varepsilon'_{\cdot u}G_\tau\varepsilon_{\cdot\tau}-tr\left(G_\tau\right)\right]\right\}&=\sum_{u=1}^{g}\sigma_0^{u\tau}\sum_{j=1}^{n}\sum_{k=1}^{n}g_{jk\tau}\mathbb{E}\left(\varepsilon_{it}\varepsilon_{ju}\varepsilon_{k\tau}\right)\\&=\sum_{u=1}^{g}\sigma_0^{u\tau}g_{ii\tau}\mathbb{E}\left(\varepsilon_{it}\varepsilon_{iu}\varepsilon_{i\tau}\right),\end{aligned}$$

(2.92) follows. ■

Lemma C5 *For* $s,\ t,\ u,\ v=1,\ \ldots,\ g,$

$$n^{-1}\mathbb{E}\left(\varepsilon'_{\cdot s}\varepsilon_{\cdot t}\varepsilon'_{\cdot u}\varepsilon_{\cdot v}\right)-n\sigma_{0st}\sigma_{0uv}=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\varepsilon_{is}\varepsilon_{it}\varepsilon_{iu}\varepsilon_{iv}-\sigma_{0st}\sigma_{0uv}\right).$$

Proof. The left side is

$$\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left(\varepsilon_{is}\varepsilon_{it}\varepsilon_{ju}\varepsilon_{jv}-\sigma_{0st}\sigma_{0uv}\right)=\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left(\varepsilon_{is}\varepsilon_{it}\varepsilon_{iu}\varepsilon_{iv}-\sigma_{0st}\sigma_{0uv}\right).$$

■

Lemma C6 *For any* $s,\ t,\ \tau = 1,\ ...,\ g,$

$$\begin{aligned} & \frac{1}{n}\mathbb{E}\left\{\left(\varepsilon_{\cdot s}'\varepsilon_{\cdot t} - n\sigma_{0st}\right)\sum_{u=1}^{g}\sigma_0^{u\tau}\varepsilon_{\cdot u}'G_\tau\varepsilon_{\cdot\tau}\right\} \\ = & \frac{1}{n}\sum_{u=1}^{g}\sigma_0^{u\tau}\sum_{i=1}^{n}g_{ii\tau}\mathbb{E}\left(\varepsilon_{is}\varepsilon_{it}\varepsilon_{iu}\varepsilon_{i\tau}\right) - \frac{1}{n}tr\left(G_\tau\right)\sigma_{0st}. \end{aligned} \tag{2.93}$$

Proof. The expectation in (2.93) is

$$\begin{aligned} & \frac{1}{n}\sum_{u=1}^{g}\sigma_0^{u\tau}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}g_{jk\tau}\mathbb{E}\left\{\left(\varepsilon_{is}\varepsilon_{it} - \sigma_{0st}\right)\varepsilon_{ju}\varepsilon_{k\tau}\right\} \\ = & \frac{1}{n}\sum_{u=1}^{g}\sigma_0^{u\tau}\sum_{i=1}^{n}g_{ii\tau}\mathbb{E}\left\{\left(\varepsilon_{is}\varepsilon_{it} - \sigma_{0st}\right)\varepsilon_{iu}\varepsilon_{i\tau}\right\} \\ = & \frac{1}{n}\sum_{u=1}^{g}\sigma_0^{u\tau}\sum_{i=1}^{n}g_{ii\tau}\mathbb{E}\left(\varepsilon_{is}\varepsilon_{it}\varepsilon_{iu}\varepsilon_{i\tau}\right) - \frac{1}{n}tr\left(G_\tau\right)\sigma_{0st}\sum_{u=1}^{g}\sigma_0^{u\tau}\sigma_{0u\tau}. \end{aligned}$$

The fact that $\sum_{u=1}^{g}\sigma_0^{u\tau}\sigma_{0u\tau} = 1$ implies (2.93). ■

Lemma C7 *For* $\tau,\ t = 1,\ ...,\ g,$

$$\begin{aligned} & \frac{1}{ng}\mathbb{E}\left(\sum_{u=1}^{g}\sum_{s=1}^{g}\sigma_0^{u\tau}\sigma_0^{st}\varepsilon_{\cdot u}'G_\tau\varepsilon_{\cdot\tau}\varepsilon_{\cdot s}'G_t\varepsilon_{\cdot t}\right) - \frac{1}{ng}tr\left(G_t\right)\mathbb{E}\left(\sum_{u=1}^{g}\sigma_0^{u\tau}\varepsilon_{\cdot u}'G_\tau\varepsilon_{\cdot\tau}\right) \\ = & \frac{1}{ng}\sum_{u=1}^{g}\sum_{s=1}^{g}\sigma_0^{u\tau}\sigma_0^{st}\sum_{i=1}^{n}g_{ii\tau}g_{iit}\left\{\mathbb{E}\left(\varepsilon_{iu}\varepsilon_{i\tau}\varepsilon_{is}\varepsilon_{it}\right) - \sigma_{0u\tau}\sigma_{0st} - \sigma_{0us}\sigma_{0\tau t} - \sigma_{0ut}\sigma_{0\tau s}\right\} \\ & + \frac{1}{ng}tr\left(G_\tau'G_t\right)\sum_{u=1}^{g}\sum_{s=1}^{g}\sigma_0^{u\tau}\sigma_0^{st}\sigma_{0us}\sigma_{0\tau t} + \frac{1}{ng}tr\left(G_\tau G_t\right)\sum_{u=1}^{g}\sum_{s=1}^{g}\sigma_0^{u\tau}\sigma_0^{st}\sigma_{0ut}\sigma_{0\tau s}. \end{aligned}$$

Proof. Employing the fact that $\sum_{s=1}^{g}\sigma_0^{st}\sigma_{0st}=1$, the first term of the left side is

$$
\begin{aligned}
& \frac{1}{ng}\sum_{u=1}^{g}\sum_{s=1}^{g}\sigma_0^{u\tau}\sigma_0^{st}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}g_{ij\tau}g_{klt}\mathbb{E}\left(\varepsilon_{iu}\varepsilon_{j\tau}\varepsilon_{ks}\varepsilon_{lt}\right) \\
= & \frac{1}{ng}\sum_{u=1}^{g}\sum_{s=1}^{g}\sigma_0^{u\tau}\sigma_0^{st}\sum_{i=1}^{n}g_{ii\tau}g_{iit}\mathbb{E}\left(\varepsilon_{iu}\varepsilon_{i\tau}\varepsilon_{is}\varepsilon_{it}\right) \\
& +\frac{1}{ng}\sum_{u=1}^{g}\sigma_0^{u\tau}\sigma_{0u\tau}\sum_{s=1}^{g}\sigma_0^{st}\sigma_{0st}\sum\sum_{i\neq k}g_{ii\tau}g_{kkt} \\
& +\frac{1}{ng}\sum_{u=1}^{g}\sum_{s=1}^{g}\sigma_0^{u\tau}\sigma_0^{st}\sigma_{0us}\sigma_{0\tau t}\sum\sum_{i\neq j}g_{ij\tau}g_{ijt} \\
& +\frac{1}{ng}\sum_{u=1}^{g}\sum_{s=1}^{g}\sigma_0^{u\tau}\sigma_0^{st}\sigma_{0ut}\sigma_{0\tau s}\sum\sum_{i\neq j}g_{ij\tau}g_{jit} \\
= & \frac{1}{ng}\sum_{u=1}^{g}\sum_{s=1}^{g}\sigma_0^{u\tau}\sigma_0^{st}\sum_{i=1}^{n}g_{ii\tau}g_{iit}\left\{\mathbb{E}\left(\varepsilon_{iu}\varepsilon_{i\tau}\varepsilon_{is}\varepsilon_{it}\right)-\sigma_{0u\tau}\sigma_{0st}-\sigma_{0us}\sigma_{0\tau t}-\sigma_{0ut}\sigma_{0\tau s}\right\} \\
& +\frac{1}{ng}tr\left(G_\tau\right)tr\left(G_t\right)+\frac{1}{ng}tr\left(G_\tau'G_t\right)\sum_{u=1}^{g}\sum_{s=1}^{g}\sigma_0^{u\tau}\sigma_0^{st}\sigma_{0us}\sigma_{0\tau t} \\
& +\frac{1}{ng}tr\left(G_\tau G_t\right)\sum_{u=1}^{g}\sum_{s=1}^{g}\sigma_0^{u\tau}\sigma_0^{st}\sigma_{0ut}\sigma_{0\tau s}.
\end{aligned}
$$

The second term is just $(ng)^{-1}\,tr\left(G_\tau\right)tr\left(G_t\right)$. Hence the required result holds. ■

Lemma C8 *For* $\tau=1,\ \ldots,\ g,$

$$
\begin{aligned}
& \frac{\partial}{\partial\rho_\tau}X'S\left(\rho\right)'\left(\Sigma^{-1}\otimes I_n\right)S\left(\rho\right)u\left(\beta\right) \\
= & -\sum_{s=1}^{g}\left\{\sigma^{s\tau}X_s'S_s\left(\rho\right)'W_\tau u_{\cdot\tau}\left(\beta\right)+\sigma^{\tau s}X_\tau'W_\tau'S_s\left(\rho\right)u_{\cdot s}\left(\beta\right)\right\}.
\end{aligned}
$$

Proof. Proceed as in the proof of Lemma C1. ■

Lemma C9 *For values of* ρ_t *such that* $S_t\left(\rho\right)$ *is invertible,*

$$
\frac{\partial}{\partial\rho_t}tr\left\{W_tS_t\left(\rho\right)^{-1}\right\}=tr\left(G_t\left(\rho\right)^2\right).
$$

Proof. By the result from Exercise 13.22 (a) in Abadir and Magnus (2005),

$$\begin{aligned}
\frac{\partial}{\partial \rho_t} tr\left\{W_t S_t(\rho)^{-1}\right\} &= \frac{\partial tr\left\{W_t S_t(\rho)^{-1}\right\}}{\partial vec\left(S_t(\rho)\right)'}\frac{\partial vec\left(S_t(\rho)\right)}{\partial \rho_t} \\
&= \left[vec\left\{\left(S_t(\rho)^{-1} W_t S_t(\rho)^{-1}\right)'\right\}\right]' vec\left(W_t\right) \\
&= \left[vec\left\{\left(S_t(\rho)^{-1} W_t S_t(\rho)^{-1}\right)'\right\}\right]' \left(I_n \otimes I_n\right) vec\left(W_t\right) \\
&= tr\left\{W_t'\left(S_t(\rho)^{-1} W_t S_t(\rho)^{-1}\right)'\right\} \\
&= tr\left\{S_t(\rho)^{-1} W_t S_t(\rho)^{-1} W_t\right\} \\
&= tr\left\{G_t(\rho)^2\right\}.
\end{aligned}$$

■

Lemma C10

$$\frac{\partial}{\partial\left\{vech\left(\Sigma^{-1}\right)\right\}'} D_g' vec(\Sigma) = -D_g'\left(\Sigma \otimes \Sigma\right) D_g.$$

Proof. Recall that $D_g vech\left(\Sigma^{-1}\right) = vec\left(\Sigma^{-1}\right)$. Employing the fact that,

$$\frac{\partial vec\left(A^{-1}\right)}{\partial\left[vec(A)\right]'} = -\left(\left(A^{-1}\right)' \otimes A^{-1}\right),$$

it follows that

$$\begin{aligned}
\frac{\partial\left\{D_g' vec(\Sigma)\right\}}{\partial\left\{vech\left(\Sigma^{-1}\right)\right\}'} &= \frac{\partial\left\{D_g' vec(\Sigma)\right\}}{\partial\left\{vec(\Sigma)\right\}'}\frac{\partial\left\{vec(\Sigma)\right\}}{\partial\left\{vec\left(\Sigma^{-1}\right)\right\}'}\frac{\partial\left\{vec\left(\Sigma^{-1}\right)\right\}}{\partial\left\{vech\left(\Sigma^{-1}\right)\right\}'} \\
&= -D_g'\left(\Sigma \otimes \Sigma\right) D_g.
\end{aligned}$$

■

Lemma C11 *If* $\overline{\theta} \rightarrow_p \theta_0$, *then, as* $n \rightarrow \infty$,

$$(ng)^{-1} X' S(\overline{\rho})' \left(\overline{\Sigma}^{-1} \otimes I_n\right) S(\overline{\rho}) X \rightarrow_p O_{11}. \tag{2.94}$$

Proof. The left side is

$$\sum_{s=1}^{g}\sum_{t=1}^{g} \overline{\sigma}^{st}\left\{(ng)^{-1} X_s' S_s(\overline{\rho})' S_t(\overline{\rho}) X_t\right\}. \tag{2.95}$$

Employing Lemma B1, Assumption B6 and consistency of $\overline{\rho}$, as $n \to \infty$,

$$\begin{aligned} & (ng)^{-1} (X_s^*)' \left(S_s^{-1}\right)' S_s(\overline{\rho})' S_t(\overline{\rho}) S_t^{-1} X_t^* - Q_{11}^{st} \\ = & \left\{(ng)^{-1} (X_s^*)' X_t^*\right\} - (\overline{\rho}_s - \rho_{0s}) \left\{(ng)^{-1} (X_s^*)' G_s' X_t^*\right\} - (\overline{\rho}_t - \rho_{0t}) \left\{(ng)^{-1} (X_s^*)' G_t X_t^*\right\} \\ & + (\overline{\rho}_s - \rho_{0s}) (\overline{\rho}_t - \rho_{0t}) \left\{(ng)^{-1} (X_s^*)' G_s' G_t X_t^*\right\} - Q_{11}^{st} \\ = & o_p(1). \end{aligned}$$

With consistency of $\overline{\Sigma}^{-1}$, (2.95) converges in probability to $\sum_{s=1}^{g} \sum_{t=1}^{g} \sigma_0^{st} Q_{11}^{st} = O_{11}$. ■

Lemma C12 *If $\overline{\theta} \to_p \theta_0$, then, as $n \to \infty$,*

$$n^{-1} X_s' S_s(\overline{\rho})' W_t u_{\cdot t}\left(\overline{\beta}\right) = o_p(1) \tag{2.96}$$

and

$$n^{-1} X_t' W_t' S_s(\overline{\rho}) u_{\cdot s}\left(\overline{\beta}\right) = o_p(1). \tag{2.97}$$

Proof. Employing the fact that $S_t(\overline{\rho}) = S_t + (\rho_{0t} - \overline{\rho}_t) W_t$, $W_t = G_t S_t$ and $u_{\cdot t}\left(\overline{\beta}\right) = u_{\cdot t} - X_t\left(\overline{\beta} - \beta_0\right)$, the left side of (2.96) becomes

$$\begin{aligned} & n^{-1} (X_s^*)' \{I_n + (\rho_{0s} - \overline{\rho}_s) G_s'\} G_t \left\{\varepsilon_{\cdot t} - X_t^* \left(\overline{\beta} - \beta_0\right)\right\} \\ = & \left\{n^{-1} (X_s^*)' G_t \varepsilon_{\cdot t}\right\} + (\rho_{0s} - \overline{\rho}_s) \left\{n^{-1} (X_s^*)' G_s' G_t \varepsilon_{\cdot t}\right\} \\ & - \left[n^{-1} (X_s^*)' G_t X_t^*\right] \left(\overline{\beta} - \beta_0\right) - (\rho_{0s} - \overline{\rho}_s) \left[n^{-1} (X_s^*)' G_s' G_t X_t^*\right] \left(\overline{\beta} - \beta_0\right). \end{aligned}$$

Under Assumption B7, Lemma B3 implies that the terms in the curly brackets are $o_p(1)$. Assumption B6 implies that the terms in the square brackets are $O_p(1)$. This with consistency of $\overline{\theta}$ imply (2.96). Employing this technique, it follows that (2.97) holds. ■

Lemma C13 *If $\overline{\theta} \to_p \theta_0$, then, as $n \to \infty$,*

$$n^{-1} X_s' S_s(\overline{\rho})' S_t(\overline{\rho}) u_{\cdot t}\left(\overline{\beta}\right) = o_p(1). \tag{2.98}$$

Proof. Since $u_{\cdot t}\left(\overline{\beta}\right) = u_{\cdot t} - X_t\left(\overline{\beta} - \beta_0\right)$, the left side of (2.98) is

$$n^{-1} (X_s^*)' \left(S_s^{-1}\right)' S_s(\overline{\rho})' S_t(\overline{\rho}) S_t^{-1} \left\{\varepsilon_{\cdot t} - X_t^* \left(\overline{\beta} - \beta_0\right)\right\}.$$

By Lemma B1, this becomes

$$\begin{aligned} & \left\{n^{-1} (X_s^*)' \varepsilon_{\cdot t}\right\} - (\overline{\rho}_s - \rho_{0s}) \left\{n^{-1} (X_s^*)' G_s' \varepsilon_{\cdot t}\right\} - (\overline{\rho}_t - \rho_{0t}) \left\{n^{-1} (X_s^*)' G_t \varepsilon_{\cdot t}\right\} \\ & + (\overline{\rho}_s - \rho_{0s}) (\overline{\rho}_t - \rho_{0t}) \left\{n^{-1} (X_s^*)' G_s' G_t \varepsilon_{\cdot t}\right\} - \left[n^{-1} (X_s^*)' X_t^*\right] \left(\overline{\beta} - \beta_0\right) \\ & + (\overline{\rho}_s - \rho_{0s}) \left[n^{-1} (X_s^*)' G_s' X_t^*\right] \left(\overline{\beta} - \beta_0\right) + (\overline{\rho}_t - \rho_{0t}) \left[n^{-1} (X_s^*)' G_t X_t^*\right] \left(\overline{\beta} - \beta_0\right) \\ & - (\overline{\rho}_s - \rho_{0s}) (\overline{\rho}_t - \rho_{0t}) \left[n^{-1} (X_s^*)' G_s' G_t X_t^*\right] \left(\overline{\beta} - \beta_0\right). \end{aligned}$$

Under Assumption B7, Lemma B3 implies that the terms in the curly brackets are $o_p(1)$. Assumption B6 implies that the terms in the square brackets are $O_p(1)$. This with consistency of $\overline{\theta}$ imply (2.98). ■

Lemma C14 *If $\overline{\theta} \to_p \theta_0$, then, as $n \to \infty$,*

$$n^{-1}u_{\cdot s}\left(\overline{\beta}\right)' W_s' W_t u_{\cdot t}\left(\overline{\beta}\right) - n^{-1}\sigma_{0st} tr\left(G_s' G_t\right) = o_p(1). \tag{2.99}$$

Proof. Note that $u_{\cdot t}\left(\overline{\beta}\right) = u_{\cdot t} - X_t\left(\overline{\beta} - \beta_0\right)$. Applying the fact that $W_t = G_t S_t$, it follows that the left side of (2.99) is

$$\begin{aligned}
&\left\{n^{-1}\varepsilon_{\cdot s}' G_s' G_t \varepsilon_{\cdot t} - n^{-1}\sigma_{0st} tr\left(G_s' G_t\right)\right\} \\
&-\left(\overline{\beta} - \beta_0\right)' \left\{n^{-1}\left(X_s^*\right)' G_s' G_t \varepsilon_{\cdot t}\right\} - \left\{n^{-1}\varepsilon_{\cdot s}' G_s' G_t X_t^*\right\}\left(\overline{\beta} - \beta_0\right) \\
&+\left(\overline{\beta} - \beta_0\right)' \left[n^{-1}\left(X_s^*\right)' G_s' G_t X_t^*\right]\left(\overline{\beta} - \beta_0\right).
\end{aligned}$$

With Assumption B7, Lemmas B2 and B3 imply that the terms in the curly brackets are $o_p(1)$. Assumption B6 implies that the term in the square brackets is $O_p(1)$. This with consistency of $\overline{\beta}$ imply that (2.99) holds. ■

Lemma C15 *If $\overline{\theta} \to_p \theta_0$, then, as $n \to \infty$,*

$$n^{-1}u_{\cdot s}\left(\overline{\beta}\right)' S_s(\rho)' W_t u_{\cdot t}\left(\overline{\beta}\right) - n^{-1}\sigma_{0st} tr\left(G_t\right) = o_p(1).$$

Proof. The proof is similar to the proofs of Lemmas C12 and C14. ■

Lemma C16 *As $n \to \infty$,*

$$\sum_{i=1}^{n}\left\{\mathbb{E}\left(\left. z_{i3,n}^2 \right| \mathcal{F}_{i-1}\right) - \mathbb{E}\left(z_{i3,n}^2\right)\right\} \to_p 0. \tag{2.100}$$

Proof. Note that $z_{i3,n} = \sum_{t=1}^{g} \lambda_{3t} c_{it,n}$, where

$$c_{it,n} = -\sum_{s=1}^{g} \sigma_0^{st}(ng)^{-1/2}\left\{g_{iit}\left(\varepsilon_{is}\varepsilon_{it} - \sigma_{0st}\right) + \varepsilon_{is}\left(\sum_{j<i} g_{ijt}\varepsilon_{jt} + \sum_{j>i} g_{ijt}\varepsilon_{jt}\right)\right\}.$$

It follows that

$$\begin{aligned}
&\sum_{i=1}^{n}\left\{\mathbb{E}\left(\left. c_{it,n}^2 \right| \mathcal{F}_{i-1}\right) - \mathbb{E}\left(c_{it,n}^2\right)\right\} \qquad (2.101)\\
=\;& \sum_{s=1}^{g}\sum_{s_1=1}^{g} \sigma_0^{st}\sigma_0^{s_1 t}(ng)^{-1}\left[\sum_{i=1}^{n}\left\{\mathbb{E}\left(\left. d_{ist} d_{is_1 t} \right| \mathcal{F}_{i-1}\right) - \mathbb{E}\left(d_{ist,n} d_{is_1 t,n}\right)\right\}\right],
\end{aligned}$$

where

$$d_{ist} = g_{iit}\left(\varepsilon_{is}\varepsilon_{it} - \sigma_{0st}\right) + \varepsilon_{is}\left(\sum_{j<i} g_{ijt}\varepsilon_{jt} + \sum_{j>i} g_{ijt}\varepsilon_{jt}\right).$$

For simplicity, consider the case when $s = s_1$,

$$\begin{aligned} & \frac{1}{ng}\sum_{i=1}^{n}\left\{\mathbb{E}\left(d_{ist}^2\middle|\mathcal{F}_{i-1}\right) - \mathbb{E}\left(d_{ist}^2\right)\right\} \qquad (2.102) \\ = & \frac{\sigma_{0ss}}{ng}\sum_{i=1}^{n}\sum_{j\neq i} g_{ijt,n}^2\left(\varepsilon_{jt}^2 - \sigma_{0tt}\right) + \frac{2}{ng}\sum_{i=1}^{n} g_{iit,n}\mathbb{E}\left(\varepsilon_{is}^2\varepsilon_{it}\right)\sum_{j\neq i} g_{ijt,n}\varepsilon_{jt} \\ & +\frac{\sigma_{0ss}}{ng}\sum_{i=1}^{n}\sum_{j<i}\sum_{k<i} g_{ijt}g_{ikt}\varepsilon_{jt}\varepsilon_{kt}\left(1-\delta_{jk}\right) \\ = & e_{1n} + e_{2n} + e_{3n}. \end{aligned}$$

Since e_{1n} can be re-written as $\frac{\sigma_{0ss}}{ng}\sum_{j=1}^{n}\sum_{i>j} g_{ijt}^2\left(\varepsilon_{jt}^2 - \sigma_{0tt}\right)$, by Assumption B3,

$$\begin{aligned} \mathbb{E}\left(e_{1n}\right)^2 & \leq C\max_{1\leq j\leq n}\mathbb{E}\left(\varepsilon_{jt}^2 - \sigma_{0tt}\right)^2 n^{-2}\sum_{j=1}^{n}\left(\sum_{i>j} g_{ijt}^2\right)^2 \\ & \leq Cn^{-2}\sum_{j=1}^{n}\left(\sum_{i=1}^{n} g_{ijt}^2\right)^2. \end{aligned}$$

Hence, Assumption C6 implies that $e_{1n} = o_p(1)$. e_{2n} can be re-written as

$$\frac{2}{ng}\sum_{j=1}^{n}\sum_{i>j} g_{iit}\mathbb{E}\left(\varepsilon_{is}^2\varepsilon_{it}\right) g_{ijt}\varepsilon_{jt}.$$

By the Cauchy-Schwarz inequality

$$\begin{aligned} \mathbb{E}\left(e_{2n}^2\right) & \leq C\left\{\max_{1\leq i\leq n}\mathbb{E}\left(\varepsilon_{is}^2\varepsilon_{it}\right)\right\}^2 n^{-2}\sum_{j=1}^{n}\mathbb{E}\left(\varepsilon_{jt}^2\right)\left(\sum_{i>j} g_{iit}g_{ijt}\right)^2 \\ & \leq C\max_{1\leq i\leq n}\mathbb{E}\left(\varepsilon_{is}^4\right)\left\{\max_{1\leq i\leq n}\mathbb{E}\left(\varepsilon_{it}^2\right)\right\}^2 n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\left|g_{iit}g_{kkt}g_{ijt}g_{kjt}\right| \\ & \leq Cn^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\left|g_{ijt}g_{kjt}\right|\left(g_{iit}^2 + g_{kkt}^2\right) \\ & \leq Cn^{-2}\left(\max_{1\leq j\leq n}\sum_{k=1}^{n}\left|g_{kjt}\right|\right)\left(\max_{1\leq i\leq n}\sum_{j=1}^{n}\left|g_{ijt}\right|\right)\sum_{i=1}^{n} g_{iit}^2 \\ & \leq Cn^{-2}\left(\max_{1\leq j\leq n}\sum_{k=1}^{n}\left|g_{kjt}\right|\right)\left(\max_{1\leq i\leq n}\sum_{j=1}^{n}\left|g_{ijt}\right|\right)\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ijt}^2. \end{aligned}$$

Since $\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ijt}^2 = tr\left(G_t'G_t\right)$, Assumption C2 implies that $\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ijt}^2 = O(n)$. Hence, by Assumption C6, $e_{2n} = o_p(1)$. Finally, similar to the above derivation, following

(A.22) in Robinson (2008), it can be shown that

$$\mathbb{E}\left(e_{3n}^2\right) \leq Cn^{-2}\left(\max_{1\leq i\leq n}\sum_{j=1}^{n}|g_{ijt}|\right)\left(\max_{1\leq j\leq n}\sum_{i=1}^{n}|g_{ijt}|\right)\sum_{i=1}^{n}\sum_{j=1}^{n}g_{ijt}^2.$$

Hence by Assumptions C2 and C6, $e_{3n} = o_p(1)$. Therefore (2.102) is $o_p(1)$. Applying the derivation similar to the one shown above to other terms, it can be shown that (2.101) holds and for $u \neq t$,

$$\sum_{i=1}^{n}\left\{\mathbb{E}\left(\left. c_{it,n}c_{iu,n}\right|\mathcal{F}_{i-1}\right) - \mathbb{E}\left(c_{it,n}c_{iu,n}\right)\right\} = o_p(1).$$

Hence (2.100) holds. ■

Lemma C17 *As $n \to \infty$,*

$$\sum_{i=1}^{n}\left[\mathbb{E}\left\{\left. z_{i1,n}z_{i3,n}\right|\mathcal{F}_{i-1}\right\} - \mathbb{E}\left\{z_{i1,n}z_{i3,n}\right\}\right] \to_p 0. \quad (2.103)$$

Proof. Recall that $z_{1,n} = -(ng)^{-1/2}\sum_{s=1}^{g}\sum_{t=1}^{g}\sigma_0^{st}\lambda_1' x_{is}^*\varepsilon_{it}$, and $z_{i3,n} = \sum_{t=1}^{g}\lambda_{3t}c_{it,n}$ where $c_{it,n}$ is defined in the proof of Lemma C16. It suffices to show (2.103) by showing that for all $t = 1, \ ..., \ g$,

$$\sum_{i=1}^{n}\left[\mathbb{E}\left\{\left. z_{i1,n}c_{it,n}\right|\mathcal{F}_{i-1}\right\} - \mathbb{E}\left\{z_{i1,n}c_{it,n}\right\}\right] \to_p 0. \quad (2.104)$$

Analogous to the proof in Lemma C16, we have to consider all possible cross-product terms. However, we will only give one example to demonstrate how to show the rest. To consider $z_{i1,n}c_{it,n}$, it is essentially to consider

$$z_{i13,n} = (ng)^{-1}\lambda_1' x_{is}^*\varepsilon_{it}\left\{g_{iit}\left(\varepsilon_{iu}\varepsilon_{iv} - \sigma_{0uv}\right) + \varepsilon_{iu}\sum_{j\neq i}g_{ijv,n}\varepsilon_{jv}\right\}.$$

Then, conditional on X,

$$\begin{aligned}\sum_{i=1}^{n}\left\{\mathbb{E}\left(\left. z_{i13,n}\right|\mathcal{F}_{i-1}\right) - \mathbb{E}\left(z_{i13,n}\right)\right\} &= (ng)^{-1}\sum_{i=1}^{n}\lambda_1' x_{is}^*\mathbb{E}\left(\varepsilon_{it}\varepsilon_{iu}\right)\sum_{j<i}g_{ijv}\varepsilon_{jv} \\ &= 2(ng)^{-1}\sum_{j=1}^{n}\sum_{i>j}\lambda_1' x_{is}^*\mathbb{E}\left(\varepsilon_{it}\varepsilon_{iu}\right)g_{ijv}\varepsilon_{jv}.\end{aligned}$$

Similar to the proof in Lemma C16, by Assumption B3, the mean square of this is bounded by

$$
\begin{aligned}
& Cn^{-2}\sum_{j=1}^{n}\left(\sum_{i>j}\lambda_1' x_{is}^{*} g_{ijv}\right)^{2} \\
\leq\ & Cn^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\left|\left(\lambda_1' x_{is}^{*}\right)\left(\lambda_1' x_{ks}^{*}\right) g_{ijv} g_{kjv}\right| \\
\leq\ & Cn^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\left|g_{ijv} g_{kjv}\right|\left(\left|\lambda_1' x_{is}^{*}\right|^{2}+\left|\lambda_1' x_{ks}^{*}\right|^{2}\right) \\
\leq\ & Cn^{-2}\left(\max_{1\leq j\leq n}\sum_{k=1}^{n}\left|g_{kjv}\right|\right)\left(\max_{1\leq i\leq n}\sum_{j=1}^{n}\left|g_{ijv}\right|\right)\lambda_1'\left(\sum_{i=1}^{n} x_{is}^{*} x_{is}^{*\prime}\right)\lambda_1 .
\end{aligned}
$$

By Assumptions B6 and C6, $\sum_{i=1}^{n}\left\{\mathbb{E}\left(\left.z_{i13,n}\right|\mathcal{F}_{i-1}\right)-\mathbb{E}\left(z_{i13,n}\right)\right\}=o_p\left(1\right).$ ■

Lemma C18 *As $n\to\infty$,*

$$\sum_{i=1}^{n}\left[\mathbb{E}\left\{\left.z_{i2,n}z_{i3,n}\right|\mathcal{F}_{i-1}\right\}-\mathbb{E}\left\{z_{i2,n}z_{i3,n}\right\}\right]\rightarrow_p 0.$$

Proof. Similar to the proof of Lemma C16. ■

Lemma D1 *Consider functions $w:\mathbb{R}^g\to\mathbb{R}$ and $z_i:\mathbb{R}^g\to\mathbb{R}$, $i=1,\ ...,\ L$ such that $w\left(x\right)\geq 0$ for all $x\in\mathbb{R}^g$ and $\int_{\mathcal{A}} z_i^2\left(x\right)w\left(x\right)\ dx<\infty$ for all $i=1,...,L$, where $\mathcal{A}\subset\mathbb{R}^g$. Let $Z_L\left(x\right)=\left(z_1\left(x\right),...,z_L\left(x\right)\right)'$, then $\int_{\mathcal{A}} Z_L\left(x\right)Z_L\left(x\right)'w\left(x\right)\ dx$ is a finite and p.s.d. matrix.*

Proof. Let Ω denote the matrix of interest and $\omega_{ij}=\int_{\mathcal{A}} z_i\left(x\right)z_j\left(x\right)w\left(x\right)\ dx$ be its $\left(i,j\right)$-th element. By Schwarz's inequality, $\omega_{ij}<\infty$ for all $i,\ j=1,\ ...,\ L$. Ω is also symmetric. Now for any $y\in\mathbb{R}^g$,

$$
\begin{aligned}
y'\Omega y &= \sum_{i=1}^{L}\sum_{j=1}^{L} y_i y_j\omega_{ij} \\
&= \int_{\mathcal{A}}\sum_{i=1}^{L}\sum_{j=1}^{L} y_i y_j z_i\left(x\right)z_j\left(x\right)w\left(x\right)\ dx \\
&= \int_{\mathcal{A}}\left[\sum_{i=1}^{L} y_i z_i\left(x\right)\right]^2 w\left(x\right)\ dx\geq 0.
\end{aligned}
$$

Hence Ω is p.s.d.. ■

Lemma D2 *Under Assumption D3, there is a constant $C>0$ such that $\underline{\lambda}\left(\mathbb{E}\left[p_*^L\left(\xi\right)p_*^L\left(\xi\right)'\right]\right)\geq C$ for all $L\geq 1$.*

Proof. Let $\mathcal{A} = \Pi_{t=1}^{g}(a_t,\ b_t)$, $\overline{\mathcal{A}} = \Pi_{t=1}^{g}[a_t,\ b_t]$ and $\mathcal{A}_0$ be the subset of all ξ such that $f_\xi(\xi) \geq C > 0$, where C is a positive integer in Assumption D3. Then $\mathcal{A}_0 \subset \mathcal{A} \subset \overline{\mathcal{A}}$ and $\overline{\mathcal{A}} \cap \mathcal{A}_0^c$ is a null set with respect to the Lebesgue measure. It follows that

$$\begin{aligned} \mathbb{E}\left[p_*^L(\xi)\, p_*^L(\xi)'\right] &= \int_{\mathcal{A}} C p_*^L(\xi)\, p_*^L(\xi)'\ d\xi + \int_{\mathcal{A}} p_*^L(\xi)\, p_*^L(\xi)'\left[f_\xi(\xi) - C\right]\ d\xi \\ &= \int_{\overline{\mathcal{A}}} C p_*^L(\xi)\, p_*^L(\xi)'\ d\xi + \int_{\mathcal{A}_0} p_*^L(\xi)\, p_*^L(\xi)'\left[f_\xi(\xi) - C\right]\ d\xi \quad (2.105) \end{aligned}$$

Since ξ is bounded, both integrals in (2.105) are finite and Lemma D1 is applicable. Since $f(\xi) - C \geq 0$ for all ξ in $\mathcal{A}_0$ it follows from Lemma D1 that the last term in (2.105) is a p.s.d. matrix. Employing the fact that, for any symmetric matrices A and B, $\underline{\lambda}(A+B) \geq \underline{\lambda}(A) + \underline{\lambda}(B)$, we have

$$\underline{\lambda}\left(\mathbb{E}\left[p_*^L(\xi)\, p_*^L(\xi)'\right]\right) \geq \underline{\lambda}\left(\int_{\overline{\mathcal{A}}} C p_*^L(\xi)\, p_*^L(\xi)'\ d\xi\right)$$

since the other term is p.s.d.. As $p_*^L(\xi)$, defined in (2.59), is a vector of multivariate orthonormal polynomials with respect to the uniform weight over $\Pi_{t=1}^{g}[a_t,\ b_t]$,

$$\int_{\overline{\mathcal{A}}} C p_*^L(\xi)\, p_*^L(\xi)'\ d\xi = C I_L,$$

where I_L is the identity matrix of order L (see Abramowitz and Stegun (1964) and Andrews (1991)). Hence $\underline{\lambda}\left(\mathbb{E}\left[p_*^L(\xi)\, p_*^L(\xi)'\right]\right) \geq C > 0$ for all $L \geq 1$. ■

Lemma D3 *(i) Under Assumption D7 (i), if $x_t \notin \mathcal{X}_t$, then $x \notin \mathcal{X}$ for all x such that its t-th element is x_t. (ii) Under Assumption D8, the transformation T in (2.27) is one-one and continuously differentiable. Let $T(\mathcal{X}) = \{T(x) : x \in \mathcal{X}\}$ be the support of ξ, f_ξ be the probability density of ξ and $m_t'(u) = \frac{d}{du} m_t(u)$. (iii) Under Assumptions D7 (i) and D8, for any $\xi \in T(\mathcal{X})$,*

$$f_\xi(\xi) = f_X\left(T^{-1}(\xi)\right) \prod_{t=1}^{g}\left[m_t'\left(m_t^{-1}(\xi_t)\right)\right]^{-1}, \qquad (2.106)$$

and f_ξ is continuous on $T(\mathcal{X})$. (iv) As f_X is positive on $\mathcal{X}$, i.e. $f_X(x) > 0$ for all $x \in \mathcal{X}$, then $f_\xi(\xi) > 0$ for all $\xi \in T(\mathcal{X})$.

Proof. (i) Suppose $x_t \notin \mathcal{X}_t$, i.e. $f_t(x_t) = 0$. Suppose that there is $y \in \mathcal{X}$, i.e. $f_X(y) > 0$, such that its t-th element is x_t. Under Assumption D7 (i), f_X is continuous at y and hence there is $\delta > 0$ such that if $\|x - y\| < \delta$, then $|f_X(x) - f_X(y)| < f_X(y)/2$. That is $f_X(x) > f_X(y)/2 > 0$ for all x such that $\|x - y\| < \delta$. This leads to a contradiction since it follows that $f_t(x_t) = \int_{\mathbb{R}^{g-1}} f_X(x)\ dx_{-t} > 0$, where $x_{-t} = (x_1,\ \dots,\ x_{t-1},\ x_{t+1},\ \dots,\ x_g)$. Hence $x \notin \mathcal{X}$ for all x whose t-th element is x_t.

(ii) If $T(x) = T(x')$, then $m_t(x_t) = m_t(x_t')$ for all $t = 1,\ \dots,\ g$. Under Assumption D8 (i), m_t is one-one and hence $x_t = x_t'$ for all t, i.e. $x = x'$. Thus T is one-one. Under Assumption D8 (ii), T is also continuously differentiable since all its partial derivatives are continuously differentiable.

(iii) Restrict our attention to $\mathcal{X}$, T^{-1} is a function $T^{-1}: T(\mathcal{X}) \to \mathcal{X}$ of the form $T^{-1}(\xi) = \left(m_1^{-1}(\xi_1), ..., m_g^{-1}(\xi_g)\right)'$. Under Assumption D8 (ii), for all $\xi_t \in m_t(\mathcal{X}_t)$,

$$\frac{d}{d\xi_t} m_t^{-1}(\xi_t) = \left[m_t'\left(m_t^{-1}(\xi_t)\right)\right]^{-1} > 0 \tag{2.107}$$

and $m_t^{-1}(\xi_t)$ is continuously differentiable on $m_t(\mathcal{X}_t)$. Therefore T^{-1} is continuously differentiable on $T(\mathcal{X})$, and, for any $\xi \in T(\mathcal{X})$,

$$f_\xi(\xi) = f_X\left(T^{-1}(\xi)\right) \prod_{t=1}^{g} \frac{d}{d\xi_t} m_t^{-1}(\xi_t), \tag{2.108}$$

since the Jacobian matrix of T^{-1} is a diagonal matrix with positive diagonal elements. (2.106) follows directly from (2.107) and (2.108). Continuity of f_ξ follows from continuity of f_X, T^{-1} and $\frac{d}{d\xi_t} m_t^{-1}(\xi_t)$.

(iv) If $f_X(x) > 0$ for all $x \in \mathcal{X}$, then (2.106) and the fact that $m_t'(x_t) > 0$ for all $x_t \in \mathcal{X}_t$ imply that $f_\xi(\xi) > 0$ for all $\xi \in T(\mathcal{X})$. ■

Lemma D4 *Let m be a function in the class $\mathcal{E}$ and g' be the derivative of g. Then (i) m satisfies Assumption D8 and (ii) $m(\mathbb{R}) = \{m(u) : u \in \mathbb{R}\}$ is an open interval $(a, b) = (-1, 0)$.*

Proof. (i) As g' is continuous and strictly positive, m is continuously differentiable and $dm/du > 0$ for all u in $\mathbb{R}$. Hence m is strictly increasing. Moreover,

$$\lim_{u\to-\infty} m(u) = -1 \text{ and } \lim_{u\to\infty} m(u) = 0.$$

Hence $-1 < m(u) < 0$ for all u in $\mathbb{R}$. Thus, m satisfies Assumption D8.

(ii) It also follows that $b = 0 = \sup\{m(x) : x \in \mathbb{R}\}$ and $a = -1 = \inf\{m(x) : x \in \mathbb{R}\}$. Since m is strictly increasing, it follows that $a, b \notin m(\mathbb{R})$. For any natural number n, there are $x_{1n} < x_{2n}$ in $\mathbb{R}$ such that $m(x_{1n}) - a < 1/n$ and $b - m(x_{2n}) < 1/n$. Since m is strictly increasing, we can select sequences $\{x_{1n}\}$ and $\{x_{2n}\}$ so that the first sequence is decreasing and the second one is increasing. Moreover, $\lim_{n\to\infty} x_{1n} = -\infty$, $\lim_{n\to\infty} x_{2n} = \infty$, $\lim_{n\to\infty} m(x_{1n}) = a$ and $\lim_{n\to\infty} m(x_{2n}) = b$. As m is continuous and strictly increasing, $m([x_{1n}, x_{2n}]) = [m(x_{1n}), m(x_{2n})]$. Then

$$\begin{aligned} m(\mathbb{R}) &= m\left(\cup_{n=1}^{\infty}[x_{1n}, x_{2n}]\right) = \cup_{n=1}^{\infty} m([x_{1n}, x_{2n}]) \\ &= \cup_{n=1}^{\infty}[m(x_{1n}), m(x_{2n})] = (a, b) = (-1, 0). \end{aligned}$$

■

Lemma D5 *Let $\mathcal{A} = \Pi_{t=1}^{g}(a_t, b_t)$, where a_t, b_t are finite real numbers for all $t = 1, ..., g$. Suppose $f : \mathcal{A} \to \mathbb{R}$ is continuous on $\mathcal{A}$ and $f(x) > 0$ for all x in $\mathcal{A}$. Suppose there is a* *finite constant $C > 0$ such that for each y in the boundary of $\mathcal{A}$,*

$$\lim_{x \to y} f(x) \geq C,$$

where the limits can be infinite. Then there is a finite constant $C_1 > 0$ such that $f(x) \geq C_1$ for all x in $\mathcal{A}$.

Proof. Let $B = \{x \in \mathcal{A} : f(x) \leq C/2\}$. If B is empty, then $f(x) > C/2$ for all x, and thus the conclusion holds. Now suppose that B is non-empty. As $\mathcal{A}$ is bounded, B must be bounded. Our aim is to show that B is also a closed subset of $\mathbb{R}^g$. First suppose that B has no limit points. Then B is closed.

Next, suppose that B has at least one limit point. Let x_0 be a limit point of B, i.e. for any $\varepsilon > 0$ there is a point $x_\varepsilon \neq x_0$ in B such that $\|x_\varepsilon - x_0\| < \varepsilon$. As $\mathcal{A}$ is an open subset of $\mathbb{R}^g$, it follows that either $x_0 \in \mathcal{A}$ or x_0 is in the boundary of $\mathcal{A}$. Suppose x_0 is in the boundary of $\mathcal{A}$. Since $\lim_{x \to x_0} f(x) \geq C$, there is $\delta > 0$ such that, for all x in $\mathcal{A}$, if $\|x - x_0\| < \delta$, then $f(x) > C/2$. However, for a given $\delta > 0$, there is x_δ in B such that $\|x_0 - x_\delta\| < \delta$ and $f(x_\delta) \leq C/2$. This leads to a contradiction. Hence x_0 must be in $\mathcal{A}$. Since f is continuous on $\mathcal{A}$, it follows that $f(x_0) \leq C/2$. Otherwise, there will be a contradiction. Therefore $x_0 \in B$ and B is closed.

Since B is a closed and bounded subset of $\mathbb{R}^g$, B is a compact set. Following Weierstrass's Theorem in optimization theory, due to continuity of f on B which is also non-empty, there is a point x^* in B such that $f(x^*) \leq f(x)$ for all x in B. As it is assumed that $f(x^*) > 0$, it follows that $f(x) \geq \min\{f(x^*),\ C/2\} > 0$ for all x in $\mathcal{A}$. ■

Lemma D6 *Let $\mathcal{A} = \Pi_{t=1}^{g}(a_t,\ b_t)$, where a_t, b_t are finite real numbers for all $t = 1,\ ...,\ g$. Suppose $f : \mathcal{A} \to \mathbb{R}$ is a function, such that for each $t = 1,\ ...,\ g$,*

$$\lim_{x_t \to a_t} f(x) = \infty, \quad \lim_{x_t \to b_t} f(x) = \infty, \tag{2.109}$$

for all $x_{-t} = (x_1,\ ...,\ x_{t-1},\ x_{t+1},\ ...,\ x_g)$ in $\Pi_{s=1}^{t-1}(a_s,\ b_s) \times \Pi_{s=t+1}^{g}(a_s,\ b_s)$. Then, for any y in the boundary of $\mathcal{A}$,

$$\lim_{x \to y} f(x) = \infty.$$

Proof. For $t = 1,\ ...,\ g$, let $\mathcal{A}_t = \Pi_{s=1}^{t-1}(a_s,\ b_s) \times \Pi_{s=t+1}^{g}(a_s,\ b_s)$. Condition (2.109) implies that for any $C > 0$, there is $\delta_t > 0$ such that if $a_t < x_t < a_t + \delta_t$ or $b_t - \delta_t < x_t < b_t$, then $f(x) > C$ for all x_{-t} in $\mathcal{A}_t$. Now let y be a point in the boundary of $\mathcal{A}$, i.e. there is at least one element of y, say y_t, such that $y_t \notin (a_t,\ b_t)$. For all x in $\mathcal{A}$, if $0 < \|x - y\| < \delta_t$, then $0 < |x_t - y_t| < \delta_t$ and hence $f(x) > C$ for all x_{-t} in $\mathcal{A}_{-t}$, particularly for all x_{-t} such that $\|x_{-t} - y_{-t}\| < \delta_t$. Thus for any $C > 0$, there is $\delta > 0$ such that for all x in $\mathcal{A}$, if $0 < \|x - y\| < \delta$, then $f(x) > C$, where y is in the boundary of $\mathcal{A}$. That is $\lim_{x \to y} f(x) = \infty$ for any y in the boundary of $\mathcal{A}$. ■

Lemma E1 *For a function h in the class $\mathcal{AC}(\mathbb{R}^g)$ and a vector of random variables e satisfying Assumption D1 such that*

$$\mathbb{E}\left|h(e)\right| + \mathbb{E}\left|\frac{\partial}{\partial e} h(e)\right| + \mathbb{E}\left|h(e)\psi(e)\right| < \infty, \tag{2.110}$$

it follows that

$$\mathbb{E}\left\{\frac{\partial h(e)}{\partial e}\right\} = \mathbb{E}\left\{h(e)\psi(e)\right\}. \tag{2.111}$$

Proof. Consider $t = 1, ..., g$. Let $\psi_t(e)$ be the t-element of a vector function $\psi(e)$. It follows from (2.110) and Fubini's theorem that

$$\int_{\mathbb{R}} \left\{\left|\frac{\partial h(e)}{\partial e_t}\right| + |h(e)|\right\} f(e)\ de_t + \int_{\mathbb{R}} |h(e)\psi_t(e)|\ de_t < \infty, \tag{2.112}$$

for all $e_{-t} \in A \subset \mathbb{R}^{g-1}$, where A^c, the complement of A, is a null set. Since h and f are in the class $\mathcal{AC}(\mathbb{R}^g)$, $h_{e_{-t}}$ and $f_{e_{-t}}$ are absolutely continuous for every $e_{-t} \in B$, where B^c is a null set. Consider a fixed $e_{-t} \in A \cap B$. Hence $h_{e_{-t}}$ and $f_{e_{-t}}$ are absolutely continuous for all $e_{-t} \in A \cap B$ and $(A \cap B)^c$ is a null subset of $\mathbb{R}^{g-1}$. For each natural number m, $h_{e_{-t}}$ and $f_{e_{-t}}$ are differentiable for all $e_t \in [-m, m]$ except in C_m where C_m is a null set. Hence $h_{e_{-t}}$ and $f_{e_{-t}}$ are differentiable for all $e_t \in \mathbb{R}$ except in $C = \cup_{m=1}^{\infty} C_m$ where C is a null set. Let $\mathbf{1}_{[-m,\ m]}(e_t) = 1$ if $e_t \in [-m, m]$ and zero otherwise. It follows from the Integration by Parts theorem from Chapter 16.F in Jones (2001) that for all natural number m,

$$\begin{aligned} &\int_{\mathbb{R}} \left(\frac{\partial h(e)}{\partial e_t}\right) f(e)\, \mathbf{1}_{[-m,\ m]}(e_t)\ de_t \\ = \quad & h_{e_{-t}}(m) f_{e_{-t}}(m) - h_{e_{-t}}(-m) f_{e_{-t}}(-m) - \int_{\mathbb{R}} h(e) \left(\frac{\partial f(e)}{\partial e_t}\right) \mathbf{1}_{[-m,\ m]}(e_t)\ de_t\ . \end{aligned}$$

Letting $m \to \infty$, it follows from (2.112), dominated convergence and the fact that $h_{e_{-t}}$ and $f_{e_{-t}}$ are continuous that

$$\int_{\mathbb{R}} \left(\frac{\partial h(e)}{\partial e_t}\right) f(e)\ de_t = -\int_{\mathbb{R}} h(e) \left(\frac{\partial f(e)}{\partial e_t}\right)\ de_t.$$

for all $e_{-t} \in A \cap B$ where $(A \cap B)^c$ is a null set. It follows from Fubini's theorem that

$$\int_{\mathbb{R}^g} \left(\frac{\partial h(e)}{\partial e_t}\right) f(e)\ de = -\int_{\mathbb{R}^g} h(e) \left(\frac{\partial f(e)}{\partial e_t}\right)\ de.$$

Hence, for $t = 1, ..., g$,

$$\mathbb{E}\left\{\frac{\partial h(e)}{\partial e_t}\right\} = \mathbb{E}\left\{h(e)\psi_t(e)\right\},$$

and (2.111) follows. ■

Lemma E2 *Suppose $h : \mathbb{R}^g \to \mathbb{R}$ is a function such that $\mathbb{E}\left[h(\varepsilon_{1\cdot})^2\right] < \infty$. Let*

$$c_L = \mathbb{E}\left[p_{**}^L(T(\varepsilon_{1\cdot}))\, h(\varepsilon_{1\cdot})\right] \quad \textit{and} \quad h_L(\varepsilon_{1\cdot}) = c_L'\ p_{**}^L(T(\varepsilon_{1\cdot})),$$

where p_{**}^L is defined as in the proof of Theorem E. Then

$$\lim_{L\to\infty} \mathbb{E}\left[h(\varepsilon_{1\cdot}) - h_L(\varepsilon_{1\cdot})\right]^2 = 0.$$

Proof. From (2.62), if follows that

$$c_L = \left\{\mathbb{E}\left[p_{**}^L(T(\varepsilon_{1\cdot}))\, p_{**}^L(T(\varepsilon_{1\cdot}))'\right]\right\}^{-1} \mathbb{E}\left[p_{**}^L(T(\varepsilon_{1\cdot}))\, h(\varepsilon_{1\cdot})\right].$$

Hence

$$c_L = \arg\min_{c\in\mathbb{R}^L} \mathbb{E}\left[h(\varepsilon_{1\cdot}) - c' p_{**}^L(T(\varepsilon_{1\cdot}))\right]^2. \tag{2.113}$$

Let $e = T(\varepsilon_{1\cdot})$, where T is the transformation satisfying Assumption E3. As $\mathbb{E}\left[h(\varepsilon_{1\cdot})^2\right] < \infty$, $\mathbb{E}\left[h_1(e)\right]^2 < \infty$ and

$$\begin{aligned}
\mathbb{E}\left[h(\varepsilon_{1\cdot}) - h_L(\varepsilon_{1\cdot})\right]^2 &= \mathbb{E}\left[h\left(T^{-1}(e)\right) - c_L'\ p_{**}^L(e)\right]^2 \\
&= \mathbb{E}\left[h_1(e) - c_L'\ p_{**}^L(e)\right]^2,
\end{aligned}$$

where $h_1 = h \circ T^{-1}$. By Assumption E3, there exists a finite constant C such that $\sup_{u\in\mathbb{R}^g} \|T(u)\| < C$. Let F_E be the distribution function of e. Then conditions (2.22) and (2.23) hold with respect to F_E. Since elements of the vector $p_{**}^L(e)$ of transformed orthonormal polynomials and of the vector $p^L(e)$ of ordinary polynomials span the same space, and under Assumption E7, the sequence $\{\lambda(l)\}$ is ordered, by Theorem 3.1.18 of Dunkl and Xu (2001), there is a triangular array $\left\{d_L \in \mathbb{R}^L;\ L \geq 1\right\}$ such that

$$\lim_{L\to\infty} \mathbb{E}\left[h_1(e) - d_L' p_{**}^L(e)\right]^2 = 0.$$

By (2.113), for each $L \geq 1$,

$$\mathbb{E}\left[h_1(e) - c_L' p_{**}^L(e)\right]^2 \leq \mathbb{E}\left[h_1(e) - d_L' p_{**}^L(e)\right]^2.$$

Hence the required result holds. ■

Lemma E3 *For $s,\ t = 1,\ ...,\ g$, as $n \to \infty$, (i)*

$$\mathbb{E}\left[\psi_{tL}(\varepsilon_{1\cdot})\right] \to \mathbb{E}\left[\psi_t(\varepsilon_{1\cdot})\right];$$

(ii)

$$\mathbb{E}\left[\psi_{sL}(\varepsilon_{1\cdot})\,\psi_{tL}(\varepsilon_{1\cdot})\right] \to \mathbb{E}\left[\psi_s(\varepsilon_{1\cdot})\,\psi_t(\varepsilon_{1\cdot})\right];$$

(iii)

$$\|\gamma_{tL}\| = O(1) \quad \textit{and} \quad \left\|\mathbb{E}\left[p_{**}^L(\varepsilon_{1\cdot})\right]\right\| = O(1).$$

Proof. Fix $s,\ t$ in $\{1,\ ...,\ g\}$. Recall that $\psi_{tL}(\varepsilon_{i\cdot}) = \gamma_{tL}'\ p_{**}^L(T(\varepsilon_{i\cdot}))$, where $\gamma_{tL} = \mathbb{E}\left[p_{**}^L(T(\varepsilon_{1\cdot}))\,\psi_t(\varepsilon_{1\cdot})\right]$. By Assumption E1, $\mathbb{E}\left[\psi_t(\varepsilon_{1\cdot})^2\right] < \infty$. Therefore, Lemma E2 and Assumption E9 imply that $\mathbb{E}\left[\psi_t\left(\varepsilon_{1\cdot}\right)-\psi_{tL}\left(\varepsilon_{1\cdot}\right)\right]^2 = o(1)$ as $n \to \infty$. Hence, both Lemma E3 (i) and (ii) are immediate consequences of Proposition 2.7.1 in Brockwell and Davis (1991).

As $\mathbb{E}\left[p_{**}^L\left(T\left(\varepsilon_{1\cdot}\right)\right)p_{**}^L\left(T\left(\varepsilon_{1\cdot}\right)\right)'\right] = I_L$,

$$\left\|\gamma_{tL}\right\|^2 = \gamma_{tL}'\mathbb{E}\left[p_{**}^L\left(T\left(\varepsilon_{1\cdot}\right)\right)p_{**}^L\left(T\left(\varepsilon_{1\cdot}\right)\right)'\right]\gamma_{tL} = \mathbb{E}\left[\psi_{tL}\left(\varepsilon_{1\cdot}\right)^2\right] = O(1),$$

by Lemma E3 (ii). Hence the first result of (iii) holds. Finally, if we replace ψ_t by a constant function 1, then the second part of (iii) holds. ■

Lemma E4 *For* $t = 1,\ \ldots,\ g,$ *as* $n \to \infty$,

$$\sup_{u\in\mathbb{R}^g}\left\|p_{**}^L\left(T(u)\right)\right\| = O(L),\ \sup_{u\in\mathbb{R}^g}\left\|\pi_{**t}^L(u)\right\| = O\left(L^3\right),\ \sup_{u\in\mathbb{R}^g}\left\|\Pi_{2t}^L(u)\right\| = O\left(L^5\right),$$

where $\Pi_{2t}^L(u) = \frac{\partial}{\partial u/}\pi_{**t}^L(u)$.

Proof. Let $v = T(u)$. Let $p_l^*(v)$ be the l-th element of the vector $p_*^L(v)$ of the Jacobi orthonormal polynomial of order $\left\|\lambda(L)\right\|_1$, with respect to the uniform weight as described in the previous section. It follows from equation (3.14) or (A40) in Andrews (1991) that there is a finite constant C such that $\sup_{u\in\mathbb{R}^g}\left|p_l^*(v)\right| \leq Cl^{1/2}$. Hence, with respect to (2.61),

$$\sup_{u\in\mathbb{R}^g}\left\|p_{**}^L\left(T(u)\right)\right\| \leq \left\|B_{2L}\right\|\sup_{u\in\mathbb{R}^g}\left\|p_*^L\left(T(u)\right)\right\| = O(L), \tag{2.114}$$

since $\left\|B_{2L}\right\|$ is uniformly bounded in L.

Now let $T(\mathbb{R}^g)$ be the image of $\mathbb{R}$ under T. By the choice of the transformation T and by Assumption E4, there is a finite constant C such that

$$\begin{aligned}\sup_{u\in\mathbb{R}^g}\left|\frac{\partial}{\partial u_t}p_l^*\left(T(u)\right)\right| &= \sup_{u\in\mathbb{R}^g}\left|\frac{\partial p_l^*(v)}{\partial v_t}\frac{\partial m(u_t)}{\partial u_t}\right| \\ &\leq \sup_{v\in T(\mathbb{R}^g)}\left|\frac{\partial p_l^*(v)}{\partial v_t}\right|\sup_{u_t\in\mathbb{R}}\left|\frac{\partial m(u_t)}{\partial u_t}\right| \\ &\leq Cl^{5/2},\end{aligned}$$

where the last inequality also follow from equation (A.44) in Andrews (1991). Recall that $\pi_{**t}^L(u) = B_{2L}\frac{\partial}{\partial u_t}p_*^L\left(T(u)\right)$. Applying the steps in (2.114), it follows that

$$\sup_{u\in\mathbb{R}^g}\left\|\pi_{**t}^L(u)\right\| = O\left(L^3\right).$$

Let δ_{st} be the Kronecker's delta. Similarly, by equation (A.44) in Andrews (1991) and

Assumption E4,

$$
\begin{aligned}
& \sup_{u\in\mathbb{R}^g}\left|\frac{\partial^2}{\partial u_s \partial u_t} p_l^*\left(T\left(u\right)\right)\right| \\
= & \sup_{u\in\mathbb{R}^g}\left|\frac{\partial^2 p_l^*\left(v\right)}{\partial v_s \partial v_t}\frac{\partial m\left(u_t\right)}{\partial u_t}\frac{\partial m\left(u_s\right)}{\partial u_s}+\frac{\partial p_l^*\left(v\right)}{\partial v_s}\frac{\partial^2 m\left(u_s\right)}{\partial u_s^2}\delta_{st}\right| \\
\leq & \sup_{v\in T(\mathbb{R}^g)}\left|\frac{\partial^2 p_l^*\left(v\right)}{\partial v_s \partial v_t}\right|\left(\sup_{u_t\in\mathbb{R}}\left|\frac{\partial m\left(u_t\right)}{\partial u_t}\right|\right)^2+\sup_{v\in T(\mathbb{R}^g)}\left|\frac{\partial p_l^*\left(v\right)}{\partial v_t}\right|\sup_{u_t\in\mathbb{R}}\left|\frac{\partial^2 m\left(u_t\right)}{\partial u_t^2}\right| \\
\leq & Cl^{9/2}.
\end{aligned}
$$

With the definition of Π_{2t}^L given in the lemma, it follows that the last required result holds. ■

Lemma E5 *Under Assumption E5 (iii),*

$$
n^{-1}\sum_{i=1}^{n}\mathbb{E}\left[\left(X_{i\cdot}^*-\overline{X}_{\cdot}^*\right)'\left(X_{i\cdot}^*-\overline{X}_{\cdot}^*\right)\right] = O\left(1\right), \tag{2.115}
$$

$$
n^{-1}\sum_{i=1}^{n}\mathbb{E}\left\|X_{i\cdot}^*\right\| = O\left(1\right), \tag{2.116}
$$

$$
\max_{1\leq t\leq g} n^{-1}\sum_{t=1}^{g}\sum_{i=1}^{n}\mathbb{E}\left\|x_{it}^*-\overline{x}_{\cdot t}^*\right\|^2 = O\left(1\right). \tag{2.117}
$$

Proof. The left side of (2.115) is

$$
n^{-1}\sum_{i=1}^{n}\mathbb{E}\left[\left(X_{i\cdot}^*\right)' X_{i\cdot}^*\right]-\mathbb{E}\left[\left(\overline{X}_{\cdot}^*\right)'\overline{X}_{\cdot}^*\right].
$$

By Assumption E5 (iii), the first term is $O\left(1\right).$ The norm of the second term is bounded by

$$
\begin{aligned}
n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left\|\left(X_{i\cdot}^*\right)' X_{j\cdot}^*\right\| & \leq n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left[\left\|X_{i\cdot}^*\right\|\left\|X_{j\cdot}^*\right\|\right] \\
& \leq n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\mathbb{E}\left\|X_{i\cdot}^*\right\|^2\mathbb{E}\left\|X_{j\cdot}^*\right\|^2\right]^{1/2} \\
& \leq n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\mathbb{E}\left\|X_{i\cdot}^*\right\|^2\mathbb{E}\left\|X_{j\cdot}^*\right\|^2+1\right] \\
& = 1+\left(n^{-1}\sum_{i=1}^{n}\mathbb{E}\left\|X_{i\cdot}^*\right\|^2\right)^2 \\
& = O\left(1\right),
\end{aligned}
$$

by Assumption E (iii), where the second inequality follows from Schwarz's inequality.

The left side of (2.116) is bounded by

$$n^{-1}\sum_{i=1}^{n}\left(\mathbb{E}\left\|X_{i\cdot}^{*}\right\|^{2}+1\right)=O\left(1\right),$$

by Assumption E5 (iii).

The left side of (2.117) is just trace of the left side of (2.115). Hence, (2.115) implies that (2.117) holds. ■

Lemma E6 *Suppose* $\overline{\beta}-\beta_0=O_p\left(n^{-1/2}\right)$ *and* $\overline{\rho}-\rho_0=O_p\left(n^{-1/2}\right)$ *as* $n\to\infty$. *Then (i)*

$$n^{-1}\sum_{i=1}^{n}\left\|X_{i\cdot}^{*}\left(\overline{\rho}\right)\right\|=O_p\left(1\right); \tag{2.118}$$

(ii)

$$n^{-1}\sum_{i=1}^{n}\left\|\left[X_{i\cdot}^{*}\left(\overline{\rho}\right)-\overline{X}_{\cdot}^{*}\left(\overline{\rho}\right)\right]-\left[X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right]\right\|^{2}=o_p\left(1\right); \tag{2.119}$$

(iii)

$$n^{-1}\sum_{i=1}^{n}\left\|X_{i\cdot}^{*}\left(\overline{\rho}\right)-\overline{X}_{\cdot}^{*}\left(\overline{\rho}\right)\right\|^{2}=O_p\left(1\right); \tag{2.120}$$

and (iv)

$$n^{-1}\sum_{i=1}^{n}\left\|R_i\left(\overline{\beta}\right)\right\|=O_p\left(1\right). \tag{2.121}$$

Proof. By definition of $X_{i\cdot}^{*}\left(\rho\right)$ and $X_{i\cdot}^{*}$ in (2.41),

$$X_{i\cdot}^{*}=X_{i\cdot}-\Lambda_0\sum_{j=1}^{n}W_{ij}X_{j\cdot},$$

where $\Lambda_0=diag\left\{\rho_{01},\ ...,\ \rho_{0g}\right\}$. Similarly,

$$X_{i\cdot}^{*}\left(\rho\right)=X_{i\cdot}-\Lambda\sum_{j=1}^{n}W_{ij}X_{j\cdot},$$

where $\Lambda=diag\left\{\rho_1,\ ...,\ \rho_g\right\}$. Let $\overline{\Lambda}=diag\left\{\overline{\rho}_1,\ ...,\ \overline{\rho}_g\right\}$. (i) Then, by Assumptions E5 (iv) and E8,

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}\left\|X_{i\cdot}^{*}\left(\overline{\rho}\right)-X_{i\cdot}^{*}\right\| &\leq \left\|\Lambda_0-\overline{\Lambda}\right\|\left(n^{-1}\sum_{i=1}^{n}\left\|\sum_{j=1}^{n}W_{ij}X_{j\cdot}\right\|\right) \\ &= o_p\left(1\right). \end{aligned}$$

Hence, with (2.116) in Lemma E5, the left side of (2.118) is bounded by

$$n^{-1}\sum_{i=1}^{n}\left\|X_{i\cdot}^{*}\left(\overline{\rho}\right)-X_{i\cdot}^{*}\right\|+n^{-1}\sum_{i=1}^{n}\left\|X_{i\cdot}^{*}\right\|=O_p\left(1\right).$$

(ii) The left side of (2.119) is bounded by

$$2n^{-1}\sum_{i=1}^{n}\left\|X_{i\cdot}^{*}\left(\overline{\rho}\right)-X_{i\cdot}^{*}\right\|^{2}+2\left\|\overline{X}_{\cdot}^{*}\left(\overline{\rho}\right)-\overline{X}_{\cdot}^{*}\right\|^{2}. \tag{2.122}$$

The first term in (2.122) is bounded by

$$2\left\|\overline{\Lambda}-\Lambda_{0}\right\|^{2}\left(n^{-1}\sum_{i=1}^{n}\left\|\sum_{j=1}^{n}W_{ij}X_{j\cdot}\right\|^{2}\right)=o_{p}\left(1\right),$$

by Assumption E5 (iv). The second term in (2.122) is bounded by

$$2\left[\left\|\overline{\Lambda}-\Lambda_{0}\right\|\left(n^{-1}\sum_{i=1}^{n}\left\|\sum_{j=1}^{n}W_{ij}X_{j\cdot}\right\|\right)\right]^{2}=o_{p}\left(1\right),$$

by Assumption E5 (iv). Hence the required result holds.

(iii) The left side of (2.120) is bounded by

$$2n^{-1}\sum_{i=1}^{n}\left\|\left[X_{i\cdot}^{*}\left(\overline{\rho}\right)-\overline{X}_{\cdot}^{*}\left(\overline{\rho}\right)\right]-\left[X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right]\right\|^{2}+2n^{-1}\sum_{i=1}^{n}\left\|X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right\|^{2}.$$

By (2.119), the first term is $o_p(1)$. By (2.115), the second term is $O_p(1)$. Hence the required result holds.

(iv) With (2.68), the (t,t)- element of the diagonal matrix $R_i\left(\overline{\beta}\right)$ is

$$R_{i}\left(\overline{\beta}\right)=\sum_{j=1}^{n}w_{ijt}u_{jt}\left(\overline{\beta}\right)=\sum_{j=1}^{n}w_{ijt}\left[x_{jt}'\left(\beta_{0}-\overline{\beta}\right)+u_{jt}\right].$$

Hence, by Assumption E5 (iv),

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}\left\|R_{i}\left(\overline{\beta}\right)\right\| &\leq C\left\|\overline{\beta}-\beta_{0}\right\|n^{-1}\sum_{i=1}^{n}\left\|\sum_{j=1}^{n}W_{ij}X_{j\cdot}\right\|+Cn^{-1}\sum_{i=1}^{n}\left\|\sum_{j=1}^{n}W_{ij}u_{j\cdot}\right\| \\ &= O_{p}\left(1\right). \end{aligned}$$

■

Lemma E7 *For $t=1,\ ...,\ g,$ as $n\to\infty$,*

$$\left\|\widetilde{\phi}_{tL}-\gamma_{tL}\right\|=O_{p}\left(n^{-1/2}L^{5}\right).$$

Proof. Fix t in $\{1,\ ...,\ g\}$. As in Lemma E4, define

$$\Pi_{2t}^{L}\left(u\right)=\frac{\partial}{\partial u'}\pi_{**t}^{L}\left(u\right).$$

By the mean value theorem around θ_{04},

$$\pi_{**t}^{L}\left(\widetilde{\varepsilon}_{i\cdot}\right)=\pi_{**t}^{L}\left(\varepsilon_{i\cdot}\right)-\Pi_{2t}^{L}\left(\overline{\varepsilon}_{i\cdot}\right)\left[\left(\widetilde{\alpha}-\alpha_{0}\right)+X_{i\cdot}^{*}\left(\overline{\rho}\right)\left(\widetilde{\beta}-\beta_{0}\right)+R_{i}\left(\overline{\beta}\right)\left(\widetilde{\rho}-\rho_{0}\right)\right],$$

where $\left\|\overline{\theta}_4-\theta_{04}\right\|<\left\|\widetilde{\theta}_4-\theta_{04}\right\|$, $\overline{\varepsilon}_{i\cdot}=\varepsilon_{i\cdot}\left(\overline{\theta}_4\right)$ and $R_i\left(\beta\right)$ is as defined in (2.68). By Lemmas E4 and E6,

$$\begin{aligned}n^{-1}\left\|\sum_{i=1}^{n}\left[\pi_{**t}^{L}\left(\widetilde{\varepsilon}_{i\cdot}\right)-\pi_{**t}^{L}\left(\varepsilon_{i\cdot}\right)\right]\right\| &\leq \max_{1\leq t\leq g}\sup_{u\in\mathbb{R}^{g}}\left\|\Pi_{2t}^{L}\left(u\right)\right\|\left[\left\|\widetilde{\rho}-\rho_{0}\right\|n^{-1}\sum_{i=1}^{n}\left\|R_{i}\left(\overline{\beta}\right)\right\|\right.\\ &\quad\left.+\left\|\widetilde{\alpha}-\alpha_{0}\right\|+\left\|\widetilde{\beta}-\beta_{0}\right\|n^{-1}\sum_{1=1}^{n}\left\|X_{i\cdot}^{*}\left(\overline{\rho}\right)\right\|\right]\\ &= O_{p}\left(n^{-1/2}L^{5}\right).\end{aligned}$$

By Lemma E4

$$\begin{aligned}\mathbb{E}\left\|n^{-1}\sum_{i=1}^{n}\left[\pi_{**t}^{L}\left(\varepsilon_{i\cdot}\right)-\mathbb{E}\left(\pi_{**t}^{L}\left(\varepsilon_{1\cdot}\right)\right)\right]\right\|^{2} &= n^{-2}\sum_{i=1}^{n}\mathbb{E}\left\|\pi_{**t}^{L}\left(\varepsilon_{i\cdot}\right)-\mathbb{E}\left(\pi_{**t}^{L}\left(\varepsilon_{1\cdot}\right)\right)\right\|^{2}\\ &\leq n^{-2}\sum_{i=1}^{n}\mathbb{E}\left\|\pi_{**t}^{L}\left(\varepsilon_{1\cdot}\right)\right\|^{2}\\ &\leq n^{-1}\left(\max_{1\leq t\leq g}\sup_{u\in\mathbb{R}^{g}}\left\|\pi_{**t}^{L}\left(u\right)\right\|^{2}\right)\\ &= O\left(n^{-1}L^{6}\right).\end{aligned}$$

By (2.71) and Lemma E1,

$$\gamma_{tL}=\mathbb{E}\left[p_{**}^{L}\left(T\left(\varepsilon_{1\cdot}\right)\right)\psi_{t}\left(\varepsilon_{1\cdot}\right)\right]=\mathbb{E}\left[\frac{\partial}{\partial\varepsilon_{1t}}p_{**}^{L}\left(T\left(\varepsilon_{1\cdot}\right)\right)\right]=\mathbb{E}\left[\pi_{**t}^{L}\left(\varepsilon_{1\cdot}\right)\right]. \tag{2.123}$$

Hence, with reference to (2.77) and the above result,

$$\begin{aligned}&\left\|\widetilde{\phi}_{tL}-\gamma_{tL}\right\|\\ \leq\ &\left\|n^{-1}\sum_{i=1}^{n}\left[\pi_{**t}^{L}\left(\widetilde{\varepsilon}_{i\cdot}\right)-\pi_{**t}^{L}\left(\varepsilon_{i\cdot}\right)\right]\right\|+\left\|n^{-1}\sum_{i=1}^{n}\left[\pi_{**t}^{L}\left(\varepsilon_{i\cdot}\right)-\mathbb{E}\left(\pi_{**t}^{L}\left(\varepsilon_{1\cdot}\right)\right)\right]\right\|\\ =\ &O_{p}\left(n^{-1/2}L^{5}+n^{-1/2}L^{3}\right)=O_{p}\left(n^{-1/2}L^{5}\right).\end{aligned}$$

■

Lemma E8 *As* $n\to\infty$,

$$\left\|\widetilde{I}_{L}-I_{L}\right\|=O_{p}\left(n^{-1/2}L^{3}\right)=o_{p}\left(1\right)..$$

Proof. The first part of the proof follows the proof of Theorem 1 in Newey (1997). Let

$p_l^{**}(T(\varepsilon_{i\cdot}))$ be the l-th element of $p_{**}^L(T(\varepsilon_{i\cdot}))$ and

$$\mathcal{I}_L = n^{-1}\sum_{i=1}^{n} p_{**}^L(T(\varepsilon_{i\cdot}))\, p_{**}^L(T(\varepsilon_{i\cdot}))'.$$

Let δ_{kl} be the Kronecker's delta and recall that $p_l^{**}(v)$ is the l-th element of $p_{**}^L(v)$ and $\mathbb{E}\left[p_{**}^L(T(\varepsilon_{1\cdot}))\, p_{**}^L(T(\varepsilon_{1\cdot}))'\right] = I_L$. Then

$$\begin{aligned}
& \mathbb{E}\left\|\mathcal{I}_L - I_L\right\|^2 \\
\leq\ & \mathbb{E}\left\{tr\left(\mathcal{I}_L - I_L\right)'\left(\mathcal{I}_L - I_L\right)\right\} \\
=\ & \mathbb{E}\sum_{k=1}^{L}\sum_{l=1}^{L}\left\{n^{-1}\sum_{i=1}^{n}\left[p_k^{**}(T(\varepsilon_{i\cdot}))\, p_l^{**}(T(\varepsilon_{i\cdot})) - \delta_{kl}\right]\right\}^2 \\
\leq\ & n^{-1}\sum_{k=1}^{L}\sum_{l=1}^{L}\mathbb{E}\left\{\left[p_k^{**}(T(\varepsilon_{1\cdot}))\right]^2\left[p_l^{**}(T(\varepsilon_{1\cdot}))\right]^2\right\} \\
=\ & n^{-1}\mathbb{E}\left\{\sum_{k=1}^{L}\left[p_k^{**}(T(\varepsilon_{1\cdot}))\right]^2\sum_{l=1}^{L}\left[p_l^{**}(T(\varepsilon_{1\cdot}))\right]^2\right\} \\
\leq\ & n^{-1}\left(\sup_{u\in\mathbb{R}^g}\left\|p_{**}^L(T(u))\right\|\right)^2 tr\left\{\mathbb{E}\left[p_{**}^L(T(\varepsilon_{1\cdot}))\, p_{**}^L(T(\varepsilon_{1\cdot}))'\right]\right\}. \qquad (2.124)
\end{aligned}$$

As the expectation in (2.124) is I_L, it follows from Lemma E4 that

$$\mathbb{E}\left\|\mathcal{I}_L - I_L\right\|^2 = O\left(n^{-1}L^3\right). \qquad (2.125)$$

Now

$$\begin{aligned}
\left\|\widetilde{I}_L - \mathcal{I}_L\right\| \leq\ & \left\|n^{-1}\sum_{i=1}^{n}\left[p_{**}^L(T(\widetilde{\varepsilon}_{i\cdot})) - p_{**}^L(T(\varepsilon_{i\cdot}))\right] p_{**}^L(T(\widetilde{\varepsilon}_{i\cdot}))'\right\| \\
& + \left\|n^{-1}\sum_{i=1}^{n} p_{**}^L(T(\varepsilon_{i\cdot}))\left[p_{**}^L(T(\widetilde{\varepsilon}_{i\cdot})) - p_{**}^L(T(\varepsilon_{i\cdot}))\right]'\right\| \\
\leq\ & 2\sup_{u\in\mathbb{R}^g}\left\|p_{**}^L(T(u))\right\|\left[n^{-1}\sum_{i=1}^{n}\left\|p_{**}^L(T(\widetilde{\varepsilon}_{i\cdot})) - p_{**}^L(T(\varepsilon_{i\cdot}))\right\|\right]. \qquad (2.126)
\end{aligned}$$

By the mean value theorem in (2.69), the term in the square brackets in (2.126) is bounded by

$$\sup_{u\in\mathbb{R}^g}\left\|\Pi_{**}^L(u)\right\|\left[\left\|\widetilde{\alpha} - \alpha_0\right\| + \left\|\widetilde{\beta} - \beta_0\right\| n^{-1}\sum_{i=1}^{n}\left\|X_{i\cdot}^*(\overline{\rho})\right\| + \left\|\widetilde{\rho} - \rho_0\right\| n^{-1}\sum_{i=1}^{n}\left\|R_i\left(\overline{\beta}\right)\right\|\right].$$

By Assumption E8, Lemmas E4 and E6, this term is $O_p\left(n^{-1/2}L^3\right)$. Hence, this result, (2.125) and Assumption E9 imply that

$$\left\|\widetilde{I}_L - I_L\right\| \leq \left\|\widetilde{I}_L - \mathcal{I}_L\right\| + \left\|\mathcal{I}_L - I_L\right\| = O_p\left(n^{-1/2}L^3\right) = o_p(1).$$

■

Lemma E9 *For* $t = 1, \ ..., \ g,$ *as* $n \to \infty,$

$$\|\widetilde{\gamma}_{tL} - \gamma_{tL}\| = O_p\left(n^{-1/2}L^5\right) \quad \textit{and} \quad \widetilde{\phi}'_{sL}\widetilde{I}_L^{-1}\widetilde{\phi}_{tL} - \gamma'_{sL}\gamma_{tL} = O_p\left(n^{-1}L^{10}\right).$$

Proof. Fix t in $\{1, \ ..., \ g\}$. With reference to (2.75) and (2.77)

$$\begin{aligned} \|\widetilde{\gamma}_{tL} - \gamma_{iL}\| &= \left\|\widetilde{I}_L^{-1}\widetilde{\phi}_{tL} - \gamma_{tL}\right\| \\ &\leq \left\|\left(\widetilde{I}_L^{-1} - I_L\right)\widetilde{\phi}_{tL}\right\| + \left\|\widetilde{\phi}_{tL} - \gamma_{tL}\right\| \\ &\leq \left\|\widetilde{I}_L^{-1}\right\| \left\|\left(\widetilde{I}_L - I_L\right)\right\| \left\|\widetilde{\phi}_{tL}\right\| + \left\|\widetilde{\phi}_{tL} - \gamma_{tL}\right\|. \end{aligned} \tag{2.127}$$

The following result on $\left\|\widetilde{I}_L^{-1}\right\|$ is based on the observation made in Newey (1997). Let $\rho(A)$ be the spectral radius of a matrix A as defined in Section 2. Based on the fact that for a symmetric matrix A, $\underline{\lambda}(A) = \inf_{\|x\|=1} x'Ax$, it can be shown that for symmetric matrices A and B, $\underline{\lambda}(A+B) \geq \underline{\lambda}(A) + \underline{\lambda}(B)$. Employing this property with the fact that $\underline{\lambda}(-A) = -\overline{\lambda}(A)$, it follows that for symmetric matrices A and B,

$$\underline{\lambda}(A) \geq \underline{\lambda}(B) - \overline{\lambda}(A-B) \geq \underline{\lambda}(B) - \rho(A-B).$$

As $A - B$ is symmetric, Lemma A3 implies that

$$\|A - B\| = \rho(A-B) \geq \underline{\lambda}(B) - \underline{\lambda}(A).$$

Similarly, it can be shown that $\|A - B\| \geq \underline{\lambda}(A) - \underline{\lambda}(B)$. Hence $\|A - B\| \geq |\underline{\lambda}(A) - \underline{\lambda}(B)|$. By Lemma E8,

$$\left|\underline{\lambda}\left(\widetilde{I}_L\right) - 1\right| = \left|\underline{\lambda}\left(\widetilde{I}_L\right) - \underline{\lambda}(I_L)\right| \leq \left\|\widetilde{I}_L - I_L\right\| = o_p(1).$$

Hence $\underline{\lambda}\left(\widetilde{I}_L\right) - 1 = o_p(1)$ and

$$\left\|\widetilde{I}_L^{-1}\right\| = \left[\underline{\lambda}\left(\widetilde{I}_L\right)\right]^{-1} \to_p 1, \text{ i.e. } \left\|\widetilde{I}_L^{-1}\right\| = O_p(1). \tag{2.128}$$

With Lemmas E3 (iii), E7 and Assumption E9,

$$\left\|\widetilde{\phi}_{tL}\right\| \leq \left\|\widetilde{\phi}_{tL} - \gamma_{tL}\right\| + \|\gamma_{tL}\| = O_p(1) \tag{2.129}$$

Hence Lemmas E3 (iii), E7, E8 and (2.129) imply that

$$\|\widetilde{\gamma}_{tL} - \gamma_{iL}\| = O_p\left(n^{-1/2}L^5\right).$$

Now fix $s,\ t$ in $\{1,\ ...,\ g\}$. Since

$$\begin{aligned}\left|\widetilde{\phi}'_{sL}\widetilde{I}_L^{-1}\widetilde{\phi}_{tL}-\gamma'_{sL}\gamma_{tL}\right| &\leq \left|\left(\widetilde{\phi}_{sL}-\gamma_{sL}\right)\widetilde{I}_L^{-1}\left(\widetilde{\phi}_{tL}-\gamma_{tL}\right)\right|+\left|\gamma'_{sL}\widetilde{I}_L^{-1}\left(\widetilde{\phi}_{tL}-\gamma_{tL}\right)\right| \\ &\quad +\left|\left(\widetilde{\phi}_{sL}-\gamma_{sL}\right)'\widetilde{I}_L^{-1}\gamma_{tL}\right|+\left|\gamma'_{sL}\left(\widetilde{I}_L^{-1}-I_L\right)\gamma_{tL}\right| \\ &\leq \left\|\widetilde{\phi}_{sL}-\gamma_{sL}\right\|\left\|\widetilde{I}_L^{-1}\right\|\left\|\widetilde{\phi}_{tL}-\gamma_{tL}\right\|+\left\|\gamma_{sL}\right\|\left\|\widetilde{I}_L^{-1}\right\|\left\|\widetilde{\phi}_{tL}-\gamma_{tL}\right\| \\ &\quad +\left\|\widetilde{\phi}_{sL}-\gamma_{sL}\right\|\left\|\widetilde{I}_L^{-1}\right\|\left\|\gamma_{tL}\right\|+\left|\gamma'_{sL}\left(\widetilde{I}_L^{-1}-I_L\right)\gamma_{tL}\right|,\end{aligned}$$

Lemmas E7 - E8 and (2.128) imply that

$$\left|\widetilde{\phi}'_{sL}\widetilde{I}_L^{-1}\widetilde{\phi}_{tL}-\gamma'_{sL}\gamma_{tL}\right|=O_p\left(n^{-1}L^{10}\right).$$

■

Lemma E10 *As* $n\to\infty$,

$$n^{-1/2}\sum_{i=1}^n\left(X_{i\cdot}^*-\overline{X}_{\cdot}^*\right)'\left[\psi\left(\varepsilon_{i\cdot}\right)-\psi_L\left(\varepsilon_{i\cdot}\right)\right]=o_p\left(1\right).$$

Proof. Define x_{it}^* as in (2.40) and $\overline{x}_{\cdot t}^*=n^{-1}\sum_{i=1}^n x_{it}^*$. As usual, define ψ_t as the t-th element of ψ and similarly for ψ_{tL}. Then the left side of the lemma becomes

$$\begin{aligned}& n^{-1/2}\sum_{i=1}^n\sum_{t=1}^g\left(x_{it}^*-\overline{x}_{\cdot t}^*\right)\left[\psi_t\left(\varepsilon_{i\cdot}\right)-\psi_{tL}\left(\varepsilon_{i\cdot}\right)\right] \\ =\ & n^{-1/2}\sum_{i=1}^n\sum_{t=1}^g\left(x_{it}^*-\overline{x}_{\cdot t}^*\right)\left[\psi_t\left(\varepsilon_{i\cdot}\right)-\psi_{tL}\left(\varepsilon_{i\cdot}\right)-\mathbb{E}\left(\psi_{tL}\left(\varepsilon_{i\cdot}\right)\right)\right].\end{aligned}$$

Therefore, by Assumption E2 and the fact that $\mathbb{E}\left[\psi_t\left(\varepsilon_{1\cdot}\right)\right]=0$ for $t=1,\ ...,\ g$,

$$\begin{aligned}& \mathbb{E}\left\|n^{-1/2}\sum_{i=1}^n\left(x_{it}^*-\overline{x}_{\cdot t}^*\right)\left[\psi_t\left(\varepsilon_{i\cdot}\right)-\psi_{tL}\left(\varepsilon_{i\cdot}\right)-\mathbb{E}\left(\psi_{tL}\left(\varepsilon_{i\cdot}\right)\right)\right]\right\|^2 \\ =\ & \mathbb{E}\left[\psi_t\left(\varepsilon_{1\cdot}\right)-\psi_{tL}\left(\varepsilon_{1\cdot}\right)-\mathbb{E}\left(\psi_{tL}\left(\varepsilon_{1\cdot}\right)\right)\right]^2\left\{n^{-1}\sum_{i=1}^n\mathbb{E}\left\|x_{it}^*-\overline{x}_{\cdot t}^*\right\|^2\right\}.\end{aligned}$$

By Lemma E5, the term in the curly brackets is $O\left(1\right)$. Since $\mathbb{E}\left(\psi_t\left(\varepsilon_{1\cdot}\right)\right)=0$, by Lemmas E2 and E3,

$$\begin{aligned}\mathbb{E}\left[\psi_t\left(\varepsilon_{1\cdot}\right)-\psi_{tL}\left(\varepsilon_{1\cdot}\right)-\mathbb{E}\left(\psi_{tL}\left(\varepsilon_{1\cdot}\right)\right)\right]^2 &\leq 2\left\{\mathbb{E}\left[\psi_t\left(\varepsilon_{1\cdot}\right)-\psi_{tL}\left(\varepsilon_{1\cdot}\right)\right]^2+\left[\mathbb{E}\left(\psi_{tL}\left(\varepsilon_{1\cdot}\right)\right)\right]^2\right\} \\ &= o\left(1\right).\end{aligned}$$

Hence the required result holds. ■

Lemma E11 *As* $n \to \infty$,

$$n^{-1/2} \sum_{i=1}^{n} \left(X_{i\cdot}^* - \overline{X}_{\cdot}^* \right)' \left[\psi_L \left(\varepsilon_{i\cdot} \right) - \widehat{\psi}_L \left(\varepsilon_{i\cdot} \right) \right] = O_p \left(n^{-1/2} L^{11/2} \right).$$

Proof. With $\widehat{\psi}_L (\varepsilon_{i\cdot})$ defined in Proposition E2 and $\overline{x}_{\cdot t}^*$ as defined in the proof of Lemma E10, the left side of the lemma is

$$\begin{aligned} & n^{-1/2} \sum_{i=1}^{n} \sum_{t=1}^{g} \left(x_{it}^* - \overline{x}_{\cdot t}^* \right) p_{**}^L \left(T \left(\varepsilon_{i\cdot} \right) \right)' \left(\gamma_{tL} - \widetilde{\gamma}_{tL} \right) \\ = & \sum_{t=1}^{g} \left[n^{-1/2} \sum_{i=1}^{n} \left(x_{it}^* - \overline{x}_{\cdot t}^* \right) \left[p_{**}^L \left(T \left(\varepsilon_{i\cdot} \right) \right) - \mathbb{E} p_{**}^L \left(T \left(\varepsilon_{1\cdot} \right) \right) \right]' \right] \left(\gamma_{tL} - \widetilde{\gamma}_{tL} \right). \end{aligned}$$

The norm of the second moment of the term in the square brackets is bounded by

$$\begin{aligned} & n^{-1} \sum_{i=1}^{n} \mathbb{E} \left\| x_{it}^* - \overline{x}_{\cdot t}^* \right\|^2 \mathbb{E} \left\| p_{**}^L \left(T \left(\varepsilon_{1\cdot} \right) \right) - \mathbb{E} p_{**}^L \left(T \left(\varepsilon_{1\cdot} \right) \right) \right\|^2 \\ \leq \; & 2 \left[\mathbb{E} \left\| p_{**}^L \left(T \left(\varepsilon_{1\cdot} \right) \right) \right\|^2 + \left\| \mathbb{E} p_{**}^L \left(T \left(\varepsilon_{1\cdot} \right) \right) \right\|^2 \right] n^{-1} \sum_{i=1}^{n} \mathbb{E} \left\| x_{it}^* - \overline{x}_{\cdot t}^* \right\|^2 \\ = \; & O_p \left(L \right), \end{aligned}$$

by Lemmas E3 (iii) and E5 and the fact that

$$\mathbb{E} \left\| p_{**}^L \left(T \left(\varepsilon_{1\cdot} \right) \right) \right\|^2 = tr \left\{ \mathbb{E} \left[p_{**}^L \left(T \left(\varepsilon_{1\cdot} \right) \right) p_{**}^L \left(T \left(\varepsilon_{1\cdot} \right) \right)' \right] \right\} = tr \left(I_L \right) = L.$$

Hence, this and Lemma E9 imply that the required result holds. ■

Lemma E12 *As* $n \to \infty$,

$$n^{-1/2} \sum_{i=1}^{n} \left[\left(X_{i\cdot}^* \left(\widetilde{\rho} \right) - \overline{X}_{\cdot}^* \left(\widetilde{\rho} \right) \right) - \left(X_{i\cdot}^* - \overline{X}_{\cdot}^* \right) \right]' \widehat{\psi}_L \left(\varepsilon_{i\cdot} \right) = o_p \left(1 \right).$$

Proof. The left side of the lemma is

$$\begin{aligned} & n^{-1/2} \sum_{i=1}^{n} \left[\left(X_{i\cdot}^* \left(\widetilde{\rho} \right) - \overline{X}_{\cdot}^* \left(\widetilde{\rho} \right) \right) - \left(X_{i\cdot}^* - \overline{X}_{\cdot}^* \right) \right]' \widetilde{\Gamma}_L \left\{ p_{**}^L \left(T \left(\varepsilon_{i\cdot} \right) \right) - \mathbb{E} \left[p_{**}^L \left(T \left(\varepsilon_{1\cdot} \right) \right) \right] \right\} \\ = \; & n^{-1/2} \sum_{i=1}^{n} \left[X_{i\cdot}^* \left(\widetilde{\rho} \right) - X_{i\cdot}^* \right]' \widetilde{\Gamma}_L \left\{ p_{**}^L \left(T \left(\varepsilon_{i\cdot} \right) \right) - \mathbb{E} \left[p_{**}^L \left(T \left(\varepsilon_{1\cdot} \right) \right) \right] \right\} \\ & - n^{-1/2} \sum_{i=1}^{n} \left[\overline{X}_{\cdot}^* \left(\widetilde{\rho} \right) - \overline{X}_{\cdot}^* \right]' \widetilde{\Gamma}_L \left\{ p_{**}^L \left(T \left(\varepsilon_{i\cdot} \right) \right) - \mathbb{E} \left[p_{**}^L \left(T \left(\varepsilon_{1\cdot} \right) \right) \right] \right\} \qquad (2.130) \end{aligned}$$

The first term on the right of (2.130) is

$$
\begin{aligned}
& n^{-1/2}\sum_{i=1}^{n}\sum_{t=1}^{g}\left[x_{it}^{*}(\widetilde{\rho})-x_{it}^{*}\right]\left\{p_{**}^{L}(T(\varepsilon_{i\cdot}))-\mathbb{E}\left[p_{**}^{L}(T(\varepsilon_{1\cdot}))\right]\right\}'\widetilde{\gamma}_{tL} \\
= & \sum_{t=1}^{g}(\rho_{0t}-\widetilde{\rho}_{t})\left[n^{-1/2}\sum_{i=1}^{n}\left(\sum_{j=1}^{n}w_{ijt}x_{jt}\right)\left[p_{**}^{L}(T(\varepsilon_{i\cdot}))-\mathbb{E}p_{**}^{L}(T(\varepsilon_{1\cdot}))\right]'\right]\widetilde{\gamma}_{tL}
\end{aligned} \tag{2.131}
$$

The norm of the second moment of the term in the square brackets on the right side of (2.131) is bounded by

$$
\left(n^{-1}\sum_{i=1}^{n}\mathbb{E}\left\|\sum_{j=1}^{n}w_{ijt}x_{jt}\right\|^{2}\right)\mathbb{E}\left\|p_{**}^{L}(T(\varepsilon_{1\cdot}))-\mathbb{E}p_{**}^{L}(T(\varepsilon_{1\cdot}))\right\|^{2}=O\left(n^{\zeta_2}L\right),
$$

by the Assumption E5 (iv) and the step employed in the proof of Lemma E11. From Lemmas E3 (iii), E9 and Assumption E9, for $t=1,\ ...,\ g$,

$$
\|\widetilde{\gamma}_{tL}\|\leq\|\widetilde{\gamma}_{tL}-\gamma_{tL}\|+\|\gamma_{tL}\|=O_p(1). \tag{2.132}
$$

With (2.131), the first term on the right of (2.130) is $O_p\left(n^{-(1-\zeta_2)/2}L^{1/2}\right)=o_p(1)$ by Assumption E9.

Now consider

$$
\widetilde{\Gamma}_L\left(n^{-1/2}\sum_{i=1}^{n}\left\{p_{**}^{L}(T(\varepsilon_{i\cdot}))-\mathbb{E}\left[p_{**}^{L}(T(\varepsilon_{1\cdot}))\right]\right\}\right). \tag{2.133}
$$

The norm of the second moment of the term in the parentheses is bounded by

$$
\mathbb{E}\left\|p_{**}^{L}(T(\varepsilon_{1\cdot}))-\mathbb{E}p_{**}^{L}(T(\varepsilon_{1\cdot}))\right\|^{2}=O(L).
$$

By (2.132), $\left\|\widetilde{\Gamma}_L\right\|=O_p(1)$. Hence the term in (2.133) is $O_p\left(L^{1/2}\right)$. As in the proof of Lemma E6 (ii), $\left\|\overline{X}_{\cdot}^{*}(\widetilde{\rho})-\overline{X}_{\cdot}^{*}\right\|^{2}$ is bounded by

$$
\left\|\widetilde{\Lambda}-\Lambda_0\right\|^{2}\left(n^{-1}\sum_{i=1}^{n}\left\|\sum_{j=1}^{n}W_{ij}X_{j\cdot}\right\|\right)^{2}=O_p\left(n^{-1+2\zeta_1}\right),
$$

by Assumptions E5 (iv) and E8. Hence, the second term on the right of (2.130) is $O_p\left(n^{-1/2+\zeta_1}L^{1/2}\right)=o_p(1)$ by Assumption E9, and the required result holds. ■

Lemma E13 *As* $n\to\infty$,

$$
n^{-1}\sum_{i=1}^{n}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)'\widetilde{\Gamma}_L\left[\mathbb{E}\Pi_{**}^{L}(\varepsilon_{1\cdot})\right]X_{i\cdot}^{*}\rightarrow_p V.
$$

Proof. The left side of the lemma is

$$n^{-1}\sum_{i=1}^{n}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)'\left[\widetilde{\Gamma}_{L}\mathbb{E}\Pi_{**}^{L}\left(\varepsilon_{1\cdot}\right)-\mathcal{L}\right]\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)+n^{-1}\sum_{i=1}^{n}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)'\mathcal{L}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right).$$

By Assumption E5 (i), it suffices to show that the first term is $o_p(1)$. The first term is

$$\sum_{s=1}^{g}\sum_{t=1}^{g}\left\{\widetilde{\gamma}_{sL}'\mathbb{E}\left[\pi_{**t}^{L}\left(\varepsilon_{1\cdot}\right)\right]-\mathcal{L}_{st}\right\}\left(n^{-1}\sum_{i=1}^{n}\left(x_{is}^{*}-\overline{x}_{\cdot s}^{*}\right)\left(x_{it}^{*}-\overline{x}_{\cdot t}^{*}\right)'\right), \tag{2.134}$$

where $\mathcal{L}_{st}$ is the (s,t)-th element of $\mathcal{L}$. By (2.123) in the proof of Lemma E7, the term in the curly brackets is

$$\left(\widetilde{\gamma}_{sL}-\gamma_{sL}\right)'\gamma_{tL}+\left(\gamma_{sL}'\gamma_{tL}-\mathcal{L}_{st}\right)=o_p(1).$$

The reason follows from Assumption E9, Lemmas E3 (iii), E9 and the result from the proof of Proposition E1 that $\gamma_{sL}'\gamma_{tL}-\mathcal{L}_{st}=o(1)$. The term in the parentheses in (2.134) is $O_p(1)$ by Lemma E5. Hence the required result holds. ■

Lemma E14 *As* $n\to\infty$,

$$n^{-1}\sum_{i=1}^{n}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)'\widetilde{\Gamma}_{L}\left[\Pi_{**}^{L}\left(\varepsilon_{i\cdot}\right)-\mathbb{E}\Pi_{**}^{L}\left(\varepsilon_{1\cdot}\right)\right]X_{i\cdot}^{*}=o_p(1).$$

Proof. The left side is

$$\sum_{s=1}^{g}\sum_{t=1}^{g}\left\{n^{-1}\sum_{i=1}^{n}\widetilde{\gamma}_{sL}'\left[\pi_{**t}^{L}\left(\varepsilon_{i\cdot}\right)-\mathbb{E}\pi_{**t}^{L}\left(\varepsilon_{1\cdot}\right)\right]\left(x_{is}^{*}-\overline{x}_{\cdot s}^{*}\right)\left(x_{it}^{*}\right)'\right\}.$$

It suffices to show that each term in the curly brackets are $o_p(1)$. The term in the curly brackets is

$$\begin{aligned}&n^{-1}\sum_{i=1}^{n}\left(\widetilde{\gamma}_{sL}-\gamma_{sL}\right)'\left[\pi_{**t}^{L}\left(\varepsilon_{i\cdot}\right)-\mathbb{E}\pi_{**t}^{L}\left(\varepsilon_{1\cdot}\right)\right]\left(x_{is}^{*}-\overline{x}_{\cdot s}^{*}\right)\left(x_{it}^{*}\right)'\\&+n^{-1}\sum_{i=1}^{n}\gamma_{sL}'\left[\pi_{**t}^{L}\left(\varepsilon_{i\cdot}\right)-\mathbb{E}\pi_{**t}^{L}\left(\varepsilon_{1\cdot}\right)\right]\left(x_{is}^{*}-\overline{x}_{\cdot s}^{*}\right)\left(x_{it}^{*}\right)'.\end{aligned} \tag{2.135}$$

The norm of the first term in (2.135) is bounded by

$$C\left\|\widetilde{\gamma}_{sL}-\gamma_{sL}\right\|\sup_{u\in\mathbb{R}^{g}}\left\|\pi_{**t}^{L}(u)\right\|n^{-1}\sum_{i=1}^{n}\left\|\left(x_{is}^{*}-\overline{x}_{\cdot s}^{*}\right)\left(x_{it}^{*}\right)'\right\|=O_p\left(n^{-1/2}L^{8}\right)=o_p(1),$$

by Lemmas E4, E9, the steps similar to the proof of Lemma E5, and Assumption E9.

The norm of the second moment of the term in the curly brackets is bounded by

$$\left\|\gamma_{sL}\right\|^{2}\left[\sup_{u\in\mathbb{R}^{g}}\left\|\pi_{**t}^{L}(u)\right\|^{2}\right]n^{-2}\left(\sum_{i=1}^{n}\mathbb{E}\left\|\left(x_{is}^{*}-\overline{x}_{\cdot s}^{*}\right)\right\|^{2}\left\|x_{it}^{*}\right\|^{2}\right)=O\left(n^{-1+\kappa}L^{6}\right)=o(1),$$

by Lemmas E3, E4, and Assumption E5 (iii) and E9. Note that, by Assumption E5 (iii) and the steps in the proof of Lemma E5, it can be shown that the sum in the parentheses is $O\left(n^{1+\kappa}\right)$. ■

Lemma E15 *As* $n \to \infty$,

$$n^{-1}\sum_{i=1}^{n}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)'\widetilde{\Gamma}_{L}\left[\Pi_{**}^{L}\left(\overline{\varepsilon}_{i\cdot}\right)-\Pi_{**}^{L}\left(\varepsilon_{i\cdot}\right)\right]X_{i\cdot}^{*}=o_{p}\left(1\right).$$

Proof. The left side can written as

$$\sum_{s=1}^{g}\sum_{t=1}^{g}\left(n^{-1}\sum_{i=1}^{n}\widetilde{\gamma}_{sL}'\left[\pi_{**t}^{L}\left(\overline{\varepsilon}_{i\cdot}\right)-\pi_{**t}^{L}\left(\varepsilon_{i\cdot}\right)\right]\left(x_{is}^{*}-\overline{x}_{\cdot s}^{*}\right)\left(x_{it}^{*}\right)'\right).$$

Note that

$$\|\widetilde{\gamma}_{sL}\|\leq\|\widetilde{\gamma}_{sL}-\gamma_{sL}\|+\|\gamma_{sL}\|=O_{p}\left(1\right),$$

by Lemmas E3 (iii), E9 and Assumption E9. As in the proof of Lemma E7, by the mean value theorem around θ_{04}, it suffices to consider

$$\left(\sup_{u\in\mathbb{R}^{g}}\left\|\Pi_{2t}^{L}\left(u\right)\right\|\right)\left[\left\|\overline{\beta}-\beta_{0}\right\|n^{-1}\sum_{i=1}^{n}\left\|X_{i\cdot}^{*}\left(\breve{\rho}\right)\right\|\left\|x_{is}^{*}\right\|\left\|x_{it}^{*}\right\|\right.$$
$$\left.\left\|\overline{\alpha}-\alpha_{0}\right\|n^{-1}\sum_{i=1}^{n}\left\|x_{is}^{*}\right\|\left\|x_{it}^{*}\right\|+\left\|\overline{\rho}-\rho_{0}\right\|n^{-1}\sum_{i=1}^{n}\left\|R_{i}\left(\breve{\beta}\right)\right\|\left\|x_{is}^{*}\right\|\left\|x_{it}^{*}\right\|\right], \quad (2.136)$$

where $\left\|\breve{\beta}-\beta_{0}\right\|<\left\|\widetilde{\beta}-\beta_{0}\right\|$ and $\left\|\breve{\rho}-\rho_{0}\right\|<\left\|\widetilde{\rho}-\rho_{0}\right\|$. By Assumptions E5 and E8, the second term in the square brackets in (2.136) is $O_{p}\left(n^{-1/2}\right)$. Employing the steps in the proof of Lemma E6 (i) and (iv), it can be shown that under Assumption E5 (vi) and E8,

$$n^{-1}\sum_{i=1}^{n}\left[\left\|X_{i\cdot}^{*}\left(\breve{\rho}\right)\right\|+\left\|R_{i}\left(\breve{\beta}\right)\right\|\right]\left\|x_{is}^{*}\right\|\left\|x_{it}^{*}\right\|=O_{p}\left(n^{-1/2+\max\{\kappa_{1},\,\zeta_{3}\}}\right).$$

Hence Lemma E4 and Assumption E9 imply that the required result holds. ■

Lemma E16 *As* $n \to \infty$,

$$n^{-1}\sum_{i=1}^{n}\left(X_{i\cdot}^{*}\left(\widetilde{\rho}\right)-\overline{X}_{\cdot}^{*}\left(\widetilde{\rho}\right)\right)'\widetilde{\Gamma}_{L}\Pi_{**}^{L}\left(\overline{\varepsilon}_{i\cdot}\right)X_{i\cdot}^{*}\left(\overline{\rho}\right)-n^{-1}\sum_{i=1}^{n}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)'\widetilde{\Gamma}_{L}\Pi_{**}^{L}\left(\overline{\varepsilon}_{i\cdot}\right)X_{i\cdot}^{*}=o_{p}\left(1\right).$$

Proof. The left side is

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}\left[\left(X_{i\cdot}^{*}\left(\widetilde{\rho}\right)-\overline{X}_{\cdot}^{*}\left(\widetilde{\rho}\right)\right)-\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)\right]^{\prime}\widetilde{\Gamma}_{L}\Pi_{**}^{L}\left(\overline{\varepsilon}_{i\cdot}\right)\left[X_{i\cdot}^{*}\left(\overline{\rho}\right)-X_{i\cdot}^{*}\right] \\
& +n^{-1}\sum_{i=1}^{n}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)^{\prime}\widetilde{\Gamma}_{L}\Pi_{**}^{L}\left(\overline{\varepsilon}_{i\cdot}\right)\left[X_{i\cdot}^{*}\left(\overline{\rho}\right)-X_{i\cdot}^{*}\right] \\
& +n^{-1}\sum_{i=1}^{n}\left[\left(X_{i\cdot}^{*}\left(\widetilde{\rho}\right)-\overline{X}_{\cdot}^{*}\left(\widetilde{\rho}\right)\right)-\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)\right]^{\prime}\widetilde{\Gamma}_{L}\Pi_{**}^{L}\left(\overline{\varepsilon}_{i\cdot}\right)X_{i\cdot}^{*}. \qquad (2.137)
\end{aligned}
$$

Employing various steps in the proof of Lemma E6, the fact that $\left\|\widetilde{\Gamma}_{L}\right\| = O_p(1)$, Lemma E4 and Assumption E9, every term in (2.137) can be shown to be $o_p(1)$. ■

Lemma E17 *As $n \to \infty$,*

$$
n^{-1}\sum_{i=1}^{n}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)^{\prime}\widetilde{\Gamma}_{L}\Pi_{**}^{L}\left(\varepsilon_{i\cdot}\right)R_{i}=o_{p}\left(1\right).
$$

Proof. The left side of the lemma is

$$
n^{-1}\sum_{i=1}^{n}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)^{\prime}\left(\widetilde{\Gamma}_{L}-\Gamma_{L}\right)\Pi_{**}^{L}\left(\varepsilon_{i\cdot}\right)R_{i}+n^{-1}\sum_{i=1}^{n}\left(X_{i\cdot}^{*}-\overline{X}_{\cdot}^{*}\right)^{\prime}\Gamma_{L}\Pi_{**}^{L}\left(\varepsilon_{i\cdot}\right)R_{i}. \qquad (2.138)
$$

As in the proof of Lemma E15, by Assumptions E5 and E9, the The first term in (2.138) is $O_p\left(n^{-1/2}L^{8}\right) = o_p(1)$. By Lemmas E3, E4 and Assumption E5,

$$
\begin{aligned}
\sum_{i=1}^{n}\left\|\Gamma_{L}\Pi_{**}^{L}\left(\varepsilon_{i\cdot}\right)R_{i}\right\|^{2} &\leq \left(\sup_{u\in\mathbb{R}^{g}}\left\|\Pi_{**}^{L}\left(u\right)\right\|\right)^{2}\left\|\Gamma_{L}\right\|^{2}\sum_{i=1}^{n}\left\|\sum_{j=1}^{n}W_{ij}u_{j\cdot}\right\|^{2} \\
&= O\left(L^{6}\right).
\end{aligned}
$$

Hence Assumptions E6 and E9 imply that the second term in (2.138) is $O_p\left(n^{\vartheta}L^{3}\right) = o_p(1)$.
■

3 Statistical Inference on Regression with Spatial Dependence

3.1 Introduction

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

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

In the setting of random design nonparametric regression, Robinson (2011) proposed a triangular array structure which he justified as a possible representation for a broad class of spatial configurations, and presented conditions for consistency and asymptotic normality of Nadarya-Watson estimates. Disturbances were assumed to satisfy a kind of linear process, possibly allowing also for conditional or unconditional heteroscedasticity, and restrictions on dependence of regressors were expressed in terms of conditions on joint and marginal probability density functions, again also permitting some heterogeneity. It was argued that these kinds of conditions might be suited to a wide range of spatial data.

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

3.2 Linear Regression

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

$$Y_{in} = \beta' X_{1in} + U_{in}, \quad 1 \leq i \leq n, \quad n = 1, 2, ..., \tag{3.1}$$

where the p-dimensional column vector β is unknown, the prime denotes transposition, and the U_{in} are unobservable scalar disturbances. It is possible that X_{1in} includes an intercept.

For spatial data there is generally no natural ordering, but an arbitrary one is employed in *(3.1)*. The triangular array formulation, indicated by the n subscript, is used because some re-ordering may be natural when n increases, as discussed by Robinson (2011), for example when observation points form a lattice in two or more dimensions. It is also essential when a variable is believed to be generated by a model such as a spatial autoregression (SAR) with row-normalized weight matrix. However, to avoid complicated notation we will mostly suppress reference to the n subscript in what follows, so in particular we write $U_i = U_{in}$, $X_{1i} = X_{1in}$, $Y_i = Y_{in}$, though from time to time we take the opportunity to remind the reader of the underlying potential dependence on n of various quantities.

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

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

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

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

Assumption A1 *(3.1) holds where*

$$U_i = U_{in} = \sum_{k=1}^{\infty} b_{ik} \varepsilon_k, \quad 1 \leq i \leq n, \ n = 1, 2, ..., \tag{3.2}$$

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

$$\sum_{k=1}^{\infty} b_{ik}^2 \leq C, \quad 1 \leq i \leq n, \quad n = 1, 2, ... \tag{3.3}$$

Assumption A2 *As $n \to \infty$,*

$$\sum_{i=1}^{n} \mathbb{E} \left\| X_{2i} \right\|^2 = O(n),$$

$$n^{-1} \sum_{i=1}^{n} X_{2i} X_{1i}' \to_p \Phi,$$

where Φ is a constant non-singular matrix.

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

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

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

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

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

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

$$n^{-1/2} \sup_{k \in \mathbb{N}} \left\| \sum_{i=1}^{n} X_{2i} b_{ik} \right\| \to_p 0. \tag{3.4}$$

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

$$\sqrt{n}\left(\tilde{\beta} - \beta\right) \to_d \mathsf{N}\left(0, \Phi^{-1}\Sigma\Phi^{-1\prime}\right).$$

Robinson (2011) gave detailed motivation for using (3.2) and (3.3) to derive central limit theorems in the presence of spatial correlation and heterogeneity. Most basically, they imply that $\max_{1 \le i \le n} E(U_i^2) \le C$. They also extend the kind of linear process used when the U_i form a stationary time series, and $b_{ij} = b_{i-j}$. The more general ij subscript conveys possible heterogeneity as well as correlation, and this and the suppressed n subscript on b_{ij} are required to cover models such as the SAR (which is nonstationary). In the SAR model for U_i the b_{ij} eventually vanish, for all i ($b_{ij} = 0$ for $j > n$), and (3.3) is satisfied under standard conditions, but it also covers infinite-order dependence, familiar from time series and lattice autoregressive and autoregressive moving average models. In these, the b_{ij} are absolutely summable, but (3.3) covers also possible "long memory". However, the extent to which this is possible depends also on the dependence within $\{X_{2i}\}$. As noted in the time series case by Robinson and Hidalgo (1997), root-$n-$consistency is only possible if the collective memory in U_i and X_{2i} is sufficiently weak. In particular if X_{2i} includes an intercept, the first limit in Assumption A4 (which merely asserts convergence of the covariance matrix of $n^{-1/2}\sum_{i=1}^{n} X_{2i}U_i$) rules out long memory in U_i. However if *(3.1)* is reformulated in terms of mean-corrected observables long memory in U_i might be permitted in a corresponding central limit theorem for slope parameter estimates based on Assumption A1, cf Robinson and Hidalgo (1997). Independence of innovations (in Assumption A1) is standardly assumed both in models of SAR type and in lattice extensions of linear time series models; the martingale difference assumptions of time series models are hard to extend as there is no natural ordering to our data. Independence of $\{X_{2i}\}$ and $\{\varepsilon_i\}$ is a strong assumption and would be capable of some relaxation, but at a cost because our decoupling of conditions on disturbances and explanatory variables, here and even more so with respect to the partly linear model of the following section, has advantages, as discussed in Robinson (2011). Assumption (3.4) is the required version of the asymptotic-negligibility condition to satisfy a Lindeberg condition. Note that if the U_i are uncorrelated, as implied when $b_{ik} = 0$ for $i \neq k$, (3.4) reduces to $n^{-\frac{1}{2}} \max_{1 \leq i \leq n} \|X_{2i}\| \rightarrow_p 0$, which, given the standard Assumption A2, is implied by the more familiar-looking condition $\max_{1 \leq i \leq n} \|X_{2i}\| / \left(\Sigma_{i=1}^n \|X_{2i}\|^2\right)^{1/2} \rightarrow_p 0$. But the same conclusion results also under fairly general dependence in U_i. In particular this is the case if $\Sigma_{i=1}^n |b_{ik}| \leq C$ for all k, as is true if $|b_{ik}| \leq C\,|b_{i-k}|$ where $\Sigma_{i=-\infty}^{\infty} |b_i| < \infty$, to connect with weakly dependent stationary time series, or under an analogous condition relating to lattice processes. It is also the case with SAR models under normalization conditions. However, (3.4) is also true under more general dependence conditions, in particular if X_{2i} is uniformly bounded in probability it is only required that $\sup_{k \in \mathbb{N}} \Sigma_{i=1}^n |b_{ik}| = o\left(n^{\frac{1}{2}}\right)$, which for stationary time series and lattice data would permit long memory in U_{in}. Assumption A3 is just a standard uniform integrability requirement, avoiding identity of distribution.

3.3 Partly Linear Regression

Consider now the partly linear regression

$$Y_i = \beta' X_{1i} + \theta\left(Z_i\right) + U_i, \quad 1 \leq i \leq n, \tag{3.5}$$

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

We again focus on estimating β. As in Robinson (1988), we employ Nadaraya-Watson nonparametric regression estimation in estimating a transformed version of *(3.5)*. Letting $k : \mathbb{R} \rightarrow \mathbb{R}$ be an even function, consider a product kernel $K : \mathbb{R}^q \rightarrow \mathbb{R}$ such that

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

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

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

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

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

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

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

Our semiparametric IV estimate of β is

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

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

Definition 3.1 *$\mathcal{K}_l,\ l \geq 1$, is the class of bounded and even functions $k : \mathbb{R} \to \mathbb{R}$ such that*

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

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

Definition 3.2 *A function $g : \mathbb{R}^q \to \mathbb{R}$ is in the class $\mathcal{G}_\mu^\alpha,\ \alpha > 0,\ \mu > 0$ (with respect to the triangular array $\{Z_i\}$) if: (i) g is $(m-1)$-times partially differentiable, for $m-1 \leq \mu \leq m$; (ii) for some $\rho > 0$,*

$$\sup_{y \in B(z,\rho)} |g(y) - g(z) - Q(y,z)| \,/\, \|y - z\|^\mu \leq h(z) \text{ for all } z,$$

where $B(z,\rho) = \{y : 0 < \|y - z\| < \rho\}$; $Q = 0$ when $m = 1$; (iii) Q is a $(m-1)$-th degree homogeneous polynomial in $y - z$ with coefficients the partial derivatives of g at z of orders 1 *through $m-1$ when $m>1$; and (iv) $g(z)$, its partial derivatives of order $m-1$ and less, and $h(z)$, have average αth moments (averaged over Z_i, $1 \le i \le n$) that are uniformly bounded for all sufficiently large n.*

We introduce the following assumptions.

Assumption B1 *Assumption A1 holds with (3.1) replaced by (3.5).*

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

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

Assumption B4 *For all $n \ge 1$ and $1 \le i \le n$,*

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

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

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

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

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

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

Define

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

and

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

Assumption B5 *As* $n \to \infty$,

$$n^{-1} \sum_{i=1}^{n} V_{2i} V_{1i}' \overline{f_i}^2 \to_p \Psi,$$

where Ψ *is a constant non-singular matrix and*

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

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

Introduce the notation

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

Also introduce the dependence measures

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

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

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

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

(ii)

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

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

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

$$\begin{aligned} & \overline{\lim_{n\to\infty}} \left\{ \max_i \sup f_i(z) + \max_{i,j} \sup f_{ij}(z_1, z_2) \right. \\ & \left. + \max_{i,j,k} \sup f_{ijk}(z_1, z_2, z_3) + \max_{i,j,k,l} \sup f_{ijkl}(z_1, z_2, z_3, z_4) \right\} \\ < \; & \infty, \end{aligned}$$

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

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

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

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

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

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

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

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

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

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

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

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

where as $n \to \infty$

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

Assumption B11 *As $n \to \infty$,*

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

and

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

where Ω is a constant p.d. matrix.

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

Assumption B13 *For the same $\lambda,\ \mu,\ \nu$ as in Assumptions B7 - B9, as $n \to \infty$,*

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

Theorem B *Under Assumptions B1-B13, as $n \to \infty$,*

$$\sqrt{n} \left(\hat{\beta} - \beta \right) \to_d \mathsf{N} \left(0, \Psi^{-1} \Omega \Psi^{-1\prime} \right).$$

To a substantial degree, the assumptions are a mixture or modification of ones in Robinson (1988, 2011). In his i.i.d. data setting, Robinson (1988) was able to relax Assumption B4 to $\mathbb{E}\left(V_{ti}|Z_1, ..., Z_n\right) = 0$ a.s., $t = 1, 2$, but for our potentially spatially dependent setting we have been unable to avoid more structure. Though Assumption B4 does allow for some conditional heteroscedasticity it is nevertheless strong, especially when $p > 1$, but we prefer to avoid milder but more complicated assumptions. Assumption B5 places an upper bound on the spatial dependence in U_i and V_{2i} that covers long memory. Assumption B6, as in the nonparametric regression setting of Robinson (2011), constitutes an asymptotic independence assumption on Z_i; part (ii) of it also involves the $\gamma_{ij}^{(U)}$ and $\gamma_{ij}^{(2)}$. It is difficult to check in general, but this is possible at least under Gaussianity: as noted in Robinson (2011), a similar (slightly stronger) condition was checked by Castellana and Leadbetter (1986), in the stationary scalar Gaussian time series case: there exists $\varepsilon > 0$ such that for $\mathcal{N}(z) = \{z_1 \in \mathbb{R} : |z - z_1| < \varepsilon\}$,

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

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

3.4 Variance Estimation

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

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

where

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

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

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

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

Estimation of Σ can be considerably more problematic when there is correlation in the U_i. Given a parametric model for U_i, such as a SAR or, with lattice data, a lattice extension of a stationary time series model such as an autoregressive moving average, matters are relatively straightforward. When U_i is not parametrically modelled, lattice data permit relatively straightforward extension of the heteroscedasticity-and-autocorrelation-consistent (HAC) variance estimates proposed for time series data, which are essentially smoothed nonparametric estimates of the spectral density matrix of a stationary process at zero frequency (though the edge-effect must be taken account of). For non-lattice data there is a fundamental difficulty of autocovariance estimation, for example when data are irregularly-spaced there are typically insufficient pairs of observations available to reliably estimate the autocovariance for a given lag using standard formulae. This problem is present with irregularly-spaced time series data, and the kernel smoothing method suggested there by Masry (1983), to estimate autocovariances at integer lags, can be extended to two or more dimensions, with the autocovariance estimates then straightforwardly inserted in a higher-dimensional HAC formula. This approach is based on stationarity, but as in the time series case it can doubtless be shown to be consistency-robust to a degree of heterogeneity. As an alternative way in which the problem can be transformed to one for a stationary random field on a lattice, Conley (1999) modelled locations by a point process, dividing the sampling region into rectangular cells such that for each cell, there can be at most a single observation.

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

$$\widehat{\Sigma}_3 = \widehat{\Sigma}_{3n} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} X_{2i}X_{2j}'\tilde{U}_i\tilde{U}_j w_{ij}, \tag{3.6}$$

where the $w_{ij} = w_{ijn}$ form an array of weights, as in Kelejian and Prucha (2007). In their proof of consistency, they stress SAR-type U_i, but the property holds much more generally under Assumption A1. The quadratic-form estimate (3.6) reduces to a familiar HAC form if the w_{ij} are of the kernel form $w_{i-j} = w_{i-j,n}$, involving a bandwidth, but Kelejian and Prucha (2007) take $w_{ij} = w\left(d_{ij}/d\right)$, where the function $w\left(x\right)$ is suitably normalized and vanishes for $x > 1$, $d_{ij} = d_{ijn}$ is a known, positive (economic) distance between locations i and j, and $d = d_n \geq \max_{i,j} d_{ij}$ is regarded as increasing without bound with n. An alternative choice of w_{ij} is based on knowledge of observed locations $s_i^* \in \mathbb{R}^r$, for dimension $r \geq 1$, $i = 1, ..., n$. Let s_i be a $r \times 1$ vector such that if s_{ik} and s_{ik}^* are the k-th elements of s_i and s_i^*, so s_{ik} is the smallest integer such that $s_{ik} \geq s_{ik}^*$. We can regard s_i as discretized locations on a rectangular grid. Define

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

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

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

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

when the U_i are independent, and

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

when they are also homoscedastic, where $\widehat{\sigma}_n^2 = \widehat{\sigma}^2 = \left(n-p\right)^{-1}\sum_{i=1}^{n}\widehat{U}_{1i}^{*2}$. For dependent U_i one can use (cf (3.6))

$$\widehat{\Omega}_3 = \widehat{\Omega}_{3n} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} X_{2i}^* X_{2j}^{*\prime}\widehat{U}_{1i}^*\widehat{U}_{1j}^* w_{ij}. \tag{3.7}$$

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

$$\widetilde{\Sigma}_3 = \widetilde{\Sigma}_{3n} = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} X_{2i} X_{2j}' U_i U_j w_{ij}, \tag{3.8}$$

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

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

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

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

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

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

by $\widetilde{\Sigma}_3$.

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

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

(i)

$$m_k \to \infty, \quad n_k \to \infty, \quad k = 1, ..., r;$$

(ii)

$$\frac{m_k}{n_k} \to 0, \quad k = 1, ..., r;$$

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

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

for sufficiently large n.

Define

$$S_n(u) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\left(X_{2i}X_{2j}'\right)\mathbb{E}\left(U_iU_j\right)\mathbf{1}\left(s_i - s_j = u\right).$$

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

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

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

where κ is the cumulant function.

Define

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

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

Assumption C5 *There exist numbers $\left\{\gamma_{u,v} : u, v \in \mathbb{Z}^r\right\}$ such that*

$$\left|S_{1ts,n}(u,v,u_1) + S_{2ts,n}(u,v,u_1)\right| \leq \gamma_{u_1,u_1+v-u}$$

for all $t, s = 1, ..., p$ and $u,\ v,\ u_1 \in \mathbb{L}^*,\ n \in \mathbb{N}$, and

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

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

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

and under Assumptions C2, C4 and C5,

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

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

3.5 Monte Carlo Study of Finite-Sample Performance

We examine first, for the linear regression *(3.1)* with $p = 1$, the size of 2-sided t -tests based on the LS version of $\widetilde{\beta}$ and the estimates $\widehat{\Sigma}_1$, $\widehat{\Sigma}_2$ and the second approach to forming $\widehat{\Sigma}_3$ described in the previous section. The locations $s_1, ..., s_n$ of the observations were generated by a random draw from the uniform distribution over $\left[0, 4n^{1/2}\right] \times \left[0, 4n^{1/2}\right]$. Given these (and keeping them fixed across replications), the U_i were generated as normal variables with mean zero and covariances $Cov\left(U_i, U_j\right) = \rho_U^{\|s_i - s_j\|}$, for prescribed $\rho_U \in (0, 1)$. Likewise the X_i $(= X_{1i} = X_{2i})$ were generated as scalar normal variables with mean unity and covariances $Cov\left(X_i, X_j\right) = \rho_X^{\|s_i - s_j\|}$, for prescribed $\rho_X \in (0, 1)$ (and independently of the U_i). We took $\beta = 1$, $(\rho_X, \rho_U) = (0.2, 0.3)$ and $(0.4, 0.5)$, $n = 100$ and 169, and generated 1000 replications. Table 1 reports empirical sizes of $t-$tests with nominal sizes $\alpha = 0.01$, 0.05 and 0.1 using $\widehat{\Sigma}_1$, denoted in the "m" column by H, $\widetilde{\Sigma}_2$, denoted there by C, and $\widetilde{\Sigma}_3$, for various values of m in the truncation vector (m, m), and using the Parzen kernel for h. There is some over-sizing, which diminishes with increasing n. The over-sizing is particularly acute with respect to the inappropriate variance estimates C and H, with the (heteroscedasticity-robust) H doing worse than the classical C. For $\widetilde{\Sigma}_3$ there is stability across m (though when we tried m outside the range used in Table 1 we found greater sensitivity.

Table 1 about here

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

Table 2 about here

We now turn to the semiparametric partly linear model *(3.5)*, and use the LS version of $\widehat{\beta}$.

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

In *(3.5)* we took $p = 1$, $q = 2$ and $X_i = 1 + Z_{1i} + Z_{2i} + V_i$, $\theta(Z_i) = Z_{1i}^2 + Z_{2i}^2$, where the Z_{1i}, Z_{2i}, V_i were generated as normal variables with mean zero and such that $Cov\{X_i, X_j\} = \rho_X^{\|s_i - s_j\|}$, the U_i as normal with mean zero and $Cov\{U_i, U_j\} = \rho_U^{\|s_i - s_j\|}$, and $\{Z_{1i}\}$, $\{Z_{2i}\}$, $\{V_i\}$ and $\{U_i\}$ were independent. We again took $\beta = 1$, $(\rho_X, \rho_U) = (0.2, 0.3)$ and $(0.4, 0.5)$, $n = 100$ and 169, and generated 1000 replications. We employed $a = 1.0$, 1.2 and 1.4. We used two different kernels k, namely $k_2(z) = \phi(z)$ and $k_4(z) = \left(3 - z^2\right)\phi(z)$, where ϕ is the standard Gaussian density; k_2 and k_4 are respectively second- and fourth-order kernels, and are thus not of high enough order to satisfy the conditions of Theorem B, but this strategy was adopted due to the imprecision likely to be caused by a high order kernel in the relatively modest sample sizes.

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

Tables 3 and 4 about here

From the same replications t-ratios were computed for each choice of kernel and bandwidth, and using $\widehat{\Omega}_1$, denoted by H, $\widehat{\Omega}_2$, denoted by C, and $\widehat{\Omega}_3$, which employed the Parzen kernel and m in the truncation vectors (m, m). Empirical sizes using k_2 and k_4 are displayed in Tables 3 and 4 respectively. There is clearly some sensitivity to choice of a, with sometimes a monotone change, and sometimes a peak or trough, observed on increasing it, though the discrepancies do not seem huge. Use of the C or H estimates tends to produce marked over-sizing when $(\rho_X, \rho_U) = (0.4, 0.5)$, but the correlation-robust tests are quite stable across m. Generally, performance deteriorates with greater spatial correlation, but it also improves with increasing n, and when $n = 169$ it is surprisingly better than for the parametric linear model *(3.1)*. Comparing Tables 3 and 4, k_2 generally fares better than k_4, possibly due to the relative BI and SD behaviour reported above.

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

Table 5 about here

3.6 Empirical Illustration

The present section develops an empirical analysis of Banerjee and Iyer (2005), which employed linear regression modelling and estimation to study the influence of different systems for collecting land revenue in India, instituted during British colonial rule, on present-day economic performance. In a threefold classification of these systems, in a given area revenue was collected either through the local landlord, or through the village, or from the individual cultivator. Banerjee and Iyer (2005) used district-level data, and calculated the proportion of "non-landlord" areas within a district (in the 1870's or 1880's); in some cases this could not be done accurately and a proportion of 0 or 1 was assigned. This non-landlord proportion, denoted NL, was the explanatory variable of chief interest in Banerjee and Iyer's (2005) study: on the basis of economic theory and empirical evidence, agricultural investment and yields are positively related to NL, and income/wealth inequality are negatively related to it. Their data on measures of economic performance and productivity, used as dependent variables, consisted of a panel (annually, over the period 1956 through 1987 and across some 271 districts in 13 major states). As well as carrying out LS regressions (correcting also for various control variables), because of concerns about endogeneity (non-landlord areas are inherently more productive), Banerjee and Iyer (2005) also used IV estimation with a dummy, which we denote C0, for whether or not a district was conquered between 1820 and 1856 as instrument for NL. Districts are intrinsically of irregular size and shape, and are thus intrinsically geographically irregularly-spaced, and moreover the lack of data for some states produces huge spatial gaps, as Figure 1 of Banerjee and Iyer (2005) indicates. However, they did not explore the possibility of spatial or serial correlation, and employed standard inference rules based on uncorrelated and homoskedastic disturbances, and nor did they explore semiparametric modelling.

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

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

We next carried out some simple Nadaraya-Watson nonparametric regression fits, of each of the same four Y on NL. Under similar assumptions to ours, Robinson (2011) showed consistency and asymptotic normality of this estimate. Though his conditions require the explanatory variable to be continuous, whereas as previously noted NL has a mixed distribution, nevertheless the exercise may be helpful in reflecting nonlinearity and hinting at its form. Figures 1-4 contain scatter plots for the four dependent variables and nonparametric regression fits using a Gaussian kernel with bandwidth 0.3. This choice was the smallest one that did not give very unsmooth curves, and much larger ones appeared to oversmooth, indeed NL takes values in [0,1]. In any case the purpose of the nonparametric regression is only exploratory, to hint at possible structure. The Figures suggest in each case a mode, and possibly a mild secondary one, and thus evidence of nonlinearity, contrary to the modelling of Banerjee and Iyer (2005).

Figures 1-4 about here

Our parametric and semiparametric regression models included (unlike in Banerjee and Iyer (2005)) the square (NL2) of NL as a regressor (as well as NL itself), as just suggested by the nonparametric fitting. We also replaced two of Banerjee and Iyer's (2005) explanatory variables by proxies which may be more appropriate. For their panel data set, mean annual rainfall was constructed over 1931-1960, but rainfall records from several decades earlier than 1984, the only year which we analyze, may not be relevant, especially for agricultural yields. We used instead a precipitation variable (PRE) constructed by Mitchell and Jones (2005), based on a method which they argued offers some improvement over existing ones in the climatology literature: their dataset included 6 monthly climate elements over a 0.5° grid, over which variation is small, and we used longitude and latitude of district headquarters to obtain a weighted average at surrounding grids for 1984, district headquarters tending to be in areas of high population density which themselves tend to be relatively fertile. Second, Banerjee and Iyer (2005) included latitude (but not longitude) as an explanatory variable, but latitude behaves like a linear trend in a time series regression, and thus affects the rate of convergence of estimates, in a way determined by the scatter of district headquarters. We replaced latitude by annual temperature (TEM), which varies considerably across India and is more likely to influence agricultural yields and hence investment decisions. As an additional modification, we discarded Thanjavur district because it appears to have serious measurement error: it is the only district having IL exceeding unity, and FU in Thanjavur was 79.44 in 1981, rose to 301.18 in 1982, and has remained high since, whereas average FU excluding Thanjavur in 1984 was only 61.15.

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

In *(3.1)* we took Y = IL, FU, L15 and LR, as above (n = 164, 164, 165 and 165 respectively), with

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

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

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

(IV). Standard errors (SEs) were computed using $\widehat{\Phi}$, $\widehat{\Sigma}_2$ and $\widehat{\Sigma}_3$ as described in Sections 5 and 6, with for $m = 2$, 4, 6. Next, in *(3.5)* we took

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

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

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

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

Tables 6-9 about here

Considering first the parametric LS and IV estimates, sometimes marked differences between them are seen and neither estimate is statistically significant. In Tables 6, 7 and 8 none of the IV estimates on NL and NL2 is significant, but all the LS is significant, and in Table 9 NL is significant. This outcome also reflects the larger SEs for IV, which were anticipated. The signs of both LS and IV estimates of coefficients of NL and NL2 are mostly consistent with the inverted U-shape seen in Figures 1-4. Also in accordance with Banerjee and Iyer (2005), DBC was nearly always found to have a significantly negative effect; the exceptions were for the larger m, SEs tending to increase with m, a fairly general feature, though in most cases the variation did not affect the question of significance. Nor did the non-robust SEs often differ much from the robust ones. Turning to the semiparametric estimates, both the LS and IV versions of $\widehat{\beta}$ tend to be in the same ball-park as LS (but not

IV) $\widetilde{\beta}$, at least where NL, NL2 and DBC are concerned, though in Table 8, where LS and IV are relatively close, there is a larger discrepancy for NL and NL2 with semiparametric IV exceeding in absolute value all the other estimates in case of NL and NL2. Again, using instruments tends to increase SE. There is some sensitivity to choice of bandwidth a and kernel k, though seldom enough to affect significance, keeping m fixed. With respect to kernel choice, k_4 does not necessarily produce larger SEs than k_2, perhaps because of our simultaneous variation in bandwidth a. On the whole it could be said that Banerjee and Iyer's (2005) fully linear specifications are not contradicted by our results, except of course, and importantly, where our extra regressor NL is concerned, and the results here do strongly confirm the pattern found in our nonparametric regression fits.

3.7 Final Comments

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

A number of related issues and extensions can be pursued.

1. As mentioned in the Introduction, mixing conditions represent an alternative class of dependence conditions, to replace our linear process assumption on disturbances and density-based assumptions on regressors. A recent econometric reference is Jenish and Prucha (2008), who develop the (regular lattice) mixing condition theory of Bolthausen (1982), establishing asymptotic normality (and laws of large numbers) for the sample mean of a scalar process observed on a possibly irregular lattice whose exogenous locations are separated by distances that are bounded away from zero. Analogous conditions can undoubtedly be developed for our more complicated statistics, dependent on multivariate data (with probably faster convergence of mixing rates required), and this kind of approach would enable a relaxation of our assumption of independence between regressors and observables. On the other hand, our conditions are potentially applicable beyond their irregular lattice context (in particular when observation locations are not known even approximately), and further discussion of the advantages and disadvantages of mixing conditions relative to ours can be found in Robinson (2011). Another kind of condition that has been employed in the spatial lattice context is based on "FKG inequalities" (see Newman, 1980), but it appears to be very restrictive.

2. As also mentioned in the Introduction, more efficient estimates than ours may be available. For example, by comparison with our simple IV estimate, when the number of available estimates exceeds the number of regressors a two-stage least squares (2SLS) estimate will be more efficient given disturbances that are both uncorrelated and

homoscedastic. However, when either or both of these conditions are not met, 2SLS is not guaranteed to beat even a simple IV estimate. This drawback can be overcome by suitable GLS or generalized method-of-moment estimates, entailing either a parametric or nonparametric modelling of disturbance correlation or heteroscedasticity, but this would require further structure.

3. In the partly linear model *(3.5)*, there may also be interest in estimating the nonparametric function $\theta(z)$. A simple estimate is

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

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

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

Appendix 3.1: Proofs of Theorems A and B

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

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

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

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

Let $D = D_n = n^{-1} \sum_{k=1}^{N} W_k W_k'$. From the proof of Lemma A1,

$$\lim_{n \to \infty} \mathbb{E} \left(n^{-1} \sum_{k=N+1}^{\infty} \|W_k\|^2 \right) = 0,$$

so from Assumption A4, $D \to_p \Sigma$. For any $\lambda \in \mathbb{R}^p$ such that $\|\lambda\| = 1$, let $c_N = \lambda' D^{-\frac{1}{2}} r^{(N)}$ and $w_k = w_{kn} = n^{-1/2} \lambda' D^{-\frac{1}{2}} W_k$. Then $c_N = \sum_{k=1}^{N} w_k \varepsilon_k$, where by Assumption A4 $\{w_k \varepsilon_k,\ 1 \le k \le N\}$ is a martingale difference sequence for each $N \ge 1$. It suffices to show that conditional on $\{X_{2i}\}$, $\sum_{k=1}^{N} w_k \varepsilon_k \xrightarrow{d} \mathsf{N}(0,1)$, which follows from Theorem 2 of Scott (1973) if, conditional on $\{X_{2i}\}$, as $n \to \infty$,

$$\mathbb{E}\left(\sum_{k=1}^{N} w_k^2 \varepsilon_k^2 \middle| \varepsilon_j, \ j < k \right) \to_p 1, \tag{3.9}$$

and for all $\eta > 0$,

$$\mathbb{E}\left\{ \sum_{k=1}^{N} w_k^2 \mathbb{E}\left(\varepsilon_k^2 1\left(|w_k \varepsilon_k| > \eta\right) \middle| \{X_{2i}\} \right) \right\} \to 0. \tag{3.10}$$

The left side of (3.9) is $\lambda' D^{-\frac{1}{2}} \left(\frac{1}{n} \sum_{k=1}^{N} W_k W_k' \right) D^{-\frac{1}{2}} \lambda = 1$, so (3.9) holds. The left side of (3.10) is bounded by

$$\begin{aligned} & \mathbb{E}\left\{ \sum_{k=1}^{N} w_k^2 \mathbb{E}\left(\varepsilon_k^2 1 \left|\varepsilon_k\right| > \eta/\delta \right) \right\} + \mathbb{P}\left(\max_{1 \le k \le N} |w_k| > \delta \right) \\ \le \quad & \sup_{1 \le k \le N} \mathbb{E}\left(\varepsilon_k^2 1 \left|\varepsilon_k\right| > \eta/\delta \right) + \mathbb{P}\left(\max_{1 \le k \le N} |w_k| > \delta \right), \end{aligned} \tag{3.11}$$

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

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

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

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

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

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

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

Appendix 3.2: Propositions for proofs of Theorems A and B

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

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

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

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

Proposition B1 *As $n \to \infty$,*

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

Proof. We have

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

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

Proposition B2 *As $n \to \infty$,*

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

Proof. Similar to that of Proposition B1. ■

Proposition B3 *As $n \to \infty$,*

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

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

Proposition B4 *As $n \to \infty$,*

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

Proof. The left side is

$$n^{-1}\sum_{i=1}^{n}\left\{\xi_{2i}^{*}V_{1i}'\overline{f}_i+\xi_{2i}^{*}V_{1i}'\left(\hat{f}_i-\overline{f}_i\right)-\xi_{2i}^{*}\overline{V}_{1i}'\right\}. \tag{3.12}$$

By Proposition B2, Lemmas B4 and B5, and the Cauchy inequality, the contributions from the last two summands in (3.12) are $o_p(1)$. Due to Assumptions B5, B7 and B10 for $t=1,2$,

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\|V_{ti}\|^2\overline{f}_i^{-2}\right)\leq\max_{1\leq i\leq n}\left|\gamma_{ii}^{(t)}\right|\sup_{z\in\mathbb{R}^q}\overline{f}(z)^2\int|\sigma_t(z,z)|\,\overline{f}(z)\,dz=O(1). \tag{3.13}$$

Proposition B2, (3.13) and the Cauchy inequality imply that the contribution from the first summand in (3.12) is also $o_p(1)$. ■

Proposition B5 *As* $n\to\infty$,

$$S_{\xi_1V_2}\to_p 0.$$

Proof. Similar to that of Proposition B4. ■

Proposition B6 *As* $n\to\infty$,

$$S_{V_2V_1}\to_p \Psi.$$

Proof. The left side is

$$n^{-1}\sum_{i=1}^{n}\left(V_{2i}V_{1i}'\hat{f}_i^2-V_{2i}\overline{V}_{1i}'\hat{f}_i-\overline{V}_{2i}V_{1i}'\hat{f}_i+\overline{V}_{2i}\overline{V}_{1i}'\right). \tag{3.14}$$

For $t=1,2$,

$$n^{-1}\sum_{i=1}^{n}\|V_{ti}\|^2\hat{f}_i^2=n^{-1}\sum_{i=1}^{n}\|V_{ti}\|^2\left\{\overline{f}_i^2+2\overline{f}_i\left(\hat{f}_i-\overline{f}_i\right)+\left(\hat{f}_i-\overline{f}_i\right)^2\right\}. \tag{3.15}$$

Lemma B4, (3.13) and the Cauchy inequality imply that the left side of (3.15) is $O_p(1)$. Hence with Lemma B5 and the Cauchy inequality, the contributions from the last three summands in (3.14) are $o_p(1)$. The contribution from the first summand in (3.14) is

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

The proof is completed by applying Assumption B5, Lemma B4, (3.13) and the Cauchy inequality. ■

Proposition B7 *As* $n\to\infty$,

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

Proof. The left side is

$$\frac{1}{n}\sum_{i=1}^{n}\left\{V_{2i}\theta_i^*\overline{f}_i+V_{2i}\theta_i^*\left(\hat{f}_i-\overline{f}_i\right)-\overline{V}_{2i}\theta_i^*\right\}. \tag{3.16}$$

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

$$n^{-2}\sum_{i=1}^{n}\gamma_{ii}^{(2)}\mathbb{E}\left(\sigma_2\left(Z_i,Z_i\right)\theta_i^{*2}\overline{f}_i^2\right)+n^{-2}\sum_{i,j}^{n}\gamma_{ij}^{(2)}\mathbb{E}\left(\sigma_2\left(Z_i,Z_j\right)\theta_i^*\theta_j^*\overline{f_i f_j}\right). \tag{3.17}$$

The first term in (3.17) is bounded by

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

by repeating the proof of Proposition B1. The second term in (3.17) is

$$\frac{1}{n^4a^{2q}}\mathbb{E}\left[\sum_{i,j,k,l}^{n}\gamma_{ij}^{(2)}\sigma_2\left(Z_i,Z_j\right)\left(\theta_i-\theta_k\right)\left(\theta_j-\theta_l\right)K_{ik}K_{jl}\overline{f}_i\overline{f_j}\right.$$
$$+\sum_{i,j,k}^{n}\left\{\left(\theta_i-\theta_k\right)\left(\theta_j-\theta_k\right)K_{ik}K_{jk}+\left(\theta_i-\theta_k\right)\left(\theta_j-\theta_i\right)K_{ik}K_{ji}+\left(\theta_i-\theta_j\right)\left(\theta_j-\theta_k\right)K_{ij}K_{jk}\right\}$$
$$\left.\times\overline{f}_i\overline{f_j}\sigma_2\left(Z_i,Z_j\right)\right]-\frac{1}{n^4a^{2q}}\sum_{i,j}^{n}\gamma_{ij}^{(2)}\mathbb{E}\left\{\sigma_2\left(Z_i,Z_j\right)\left(\theta_i-\theta_j\right)^2K_{ij}^2\overline{f}_i\overline{f_j}\right\}. \tag{3.18}$$

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

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

The remaining contributions to (3.18) can likewise be shown to be $o\left(n^{-1}\right)$. ■

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

$$S_{\xi_2U}=o_p\left(n^{-1/2}\right).$$

Proof. Similar to that of Proposition B7. ■

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

$$n^{1/2}S_{V_2U}\to_d \mathsf{N}\left(0,\Omega\right).$$

Proof. The left side is

$$n^{-1/2}\sum_{i=1}^{n}\left(V_{2i}U_i\hat{f}_i^2-V_{2i}\overline{U}_i\hat{f}_i-\overline{V}_{2i}U_i\hat{f}_i+\overline{V}_{2i}\overline{U}_i\right). \tag{3.19}$$

By Lemma B5, the contribution from the last summand is $o_p(1)$. The contribution from the third summand in (3.19) is

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

by Lemmas B4, B5 and B8 and the Cauchy inequality, and that from the second summand in (3.19) can similarly be shown to be $o_p(1)$. The contribution from the first summand in (3.19) is

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

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

Appendix 3.3 : Technical Lemmas for proofs of Theorems A and B

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

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

Proof. By independence of the ε_k,

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

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

We repeatedly use the following consequences of Definition 1:

$$\begin{aligned}\sup_{u\in\mathbb{R}^q}\left|K(u)\right|+\int\|u\|\left|K(u)\right|du+\int\|u\|^2K^2(u)\,du &< \infty;\\ \sup_{\|u\|\ge\delta/a}\left|K(u)\right| &= O\left(a^{\zeta}\right)\ \textit{for all }\delta>0.\end{aligned}$$

We also introduce the abbreviations

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

Lemma B1 *As* $n \to \infty$,

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

Proof. The left side is

$$\begin{aligned}
&\frac{1}{n^3}\int \phi(z_1,z_2)^2 K(z_1,z_2)^2 \sum_{i,j}^{n} f_{ij}(z_1,z_2)\,dz_1dz_2 \\
\leq\ &\frac{1}{n}\left\{\int \phi(z_1,z_2)^2 K(z_1,z_2)^2 \frac{1}{n^2}\sum_{i,j}^{n} F_{j:i}(z_2;z_1)\,dz_1dz_2\right. \\
&\left.+\int \phi(z_1,z_2)^2 K(z_1,z_2)^2\,\overline{f}(z_1)\,\overline{f}(z_2)\,dz_1dz_2\right\}. \qquad (3.20)
\end{aligned}$$

Let

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

The first integral in braces in (3.20) is

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

where

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

Let

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

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

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

Now

$$\begin{aligned}
& \int_{\mathbb{R}^q}\int_{J_2(\varepsilon)} |p(z,au)| \; dudz \\
\leq \; & \frac{1}{n^2a^q}\sup_{\|au\|\geq\varepsilon} K^2(u)\int_{\mathbb{R}^{2q}}\phi(z_1,z_2)^2\sum_{i,j}^{n}\left\{f_{ij}(z_1,z_2)+f_i(z_1)f_j(z_2)\right\}dz_1dz_2 \\
= \; & O\left(a^{2\zeta-q}\right),
\end{aligned}$$

because the double integral is

$$n^{-2}\sum_{i,j}^{n}\left\{\mathbb{E}\left(\theta_i-\theta_j\right)^2+\mathbb{E}\left(\theta_i^2\right)+\mathbb{E}\left(\theta_j^2\right)-2\mathbb{E}\left(\theta_i\right)\mathbb{E}\left(\theta_j\right)\right\}=O(1), \tag{3.21}$$

by Assumption B9. Hence the first integral in braces in (3.20) is $o\left(a^{q+2}n^{-1/2}\right)+O(a^{2\zeta})$. The second integral in braces in (3.20) is

$$a^q\int\phi(z,z+au)^2K^2(u)\overline{f}(z)\overline{f}(z+au)\,dzdu=a^q\left(\int_{\mathbb{R}^q}\int_{J_1(\varepsilon)}+\int_{\mathbb{R}^q}\int_{J_2(\varepsilon)}\right). \tag{3.22}$$

The first integral on the right in (3.22) is bounded by

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

and the second integral is bounded by

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

Hence the second integral in (3.20) is $O\left(a^{q+2}+a^{2\zeta}\right)$. ■

Lemma B2 *As* $n\to\infty$,

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

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

$$
\begin{aligned}
& \frac{1}{n^3}\sum_{i,j,k}^{n}\int s\left(z_1,z_2,z_3\right) f_{ijk}\left(z_1,z_2,z_3\right)\prod_{i=1}^{3} dz_i \\
= & \int s\left(z_1,z_2,z_3\right)\frac{1}{n^3}\sum_{i,j,k}^{n} F_{jk:i}(z_2,z_3;z_1)\prod_{i=1}^{3} dz_i \\
& +\int s\left(z_1,z_2,z_3\right)\frac{1}{n^3}\sum_{i,j,k}^{n} f_i\left(z_1\right) F_{k:j}\left(z_2;z_3\right)\prod_{i=1}^{3} dz_i \\
& +\int s\left(z_1,z_2,z_3\right)\frac{1}{n^3}\sum_{i,j,k}^{n}\left\{f_i\left(z_1\right) f_j\left(z_2\right) f_k\left(z_3\right)-\overline{f}\left(z_1\right)\overline{f}\left(z_2\right)\overline{f}\left(z_3\right)\right\}\prod_{i=1}^{3} dz_i \\
& +\frac{(n-1)(n-2)}{n^2}\int s\left(z_1,z_2,z_3\right)\overline{f}\left(z_1\right)\overline{f}\left(z_2\right)\overline{f}\left(z_3\right)\prod_{i=1}^{3} dz_i . \qquad (3.23)
\end{aligned}
$$

With the further abbreviation $p\left(z,u,v,a\right)=\phi\left(z,z+au\right)\phi\left(z,z+av\right)K\left(u\right)K\left(v\right)$, the first integral in (3.23) is

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

where

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

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

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

By similar reasoning to that in (3.21) in the proof of Lemma B1,

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

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

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

The first integral in (3.23) is thus $o\left(a^{2q+2}n^{-1/2}\right)+O(a^{\zeta})$. The second integral in (3.23) is

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

Now

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

Then proceeding as above, the second integral of (3.23) is $o(n^{-1/2}a^{2q+2})+O(a^{\zeta})$. Because

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

proceeding as in the last part of the proof of Lemma B1, using Assumption B7, the third integral of (3.23) is $O\left(n^{-1}a^{2q+2}+n^{-1}a^{\zeta}\right)$. Finally by Assumptions B7, B9 and B12, Lemma 5 of Robinson (1988) implies that the last integral of (3.23) is $O\left(a^{2\{q+\min(\nu,\lambda+1)\}}\right)$. ■

Lemma B3 *As $n\to\infty$,*

(i)

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

(ii)

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

Proof. Denoting

$$g(z_1; z_2, z_3) = \sigma_2(z_1, z_1)\left\{K(z_1, z_2) - a^q \overline{f}(z_1)\right\}\left\{K(z_1, z_3) - a^q \overline{f}(z_1)\right\},$$

the left side of (i) can be written

$$\begin{aligned}
&\frac{1}{n^3}\sum_{i,j,k}^{n}\int g(z_1; z_2, z_3) f_{ijk}(z_1, z_2, z_3)\prod_{i=1}^{3} dz_i \\
= \quad &\frac{1}{n^3}\sum_{i,j,k}^{n}\int g(z_1; z_2, z_3)\left[f_{ijk}(z_1, z_2, z_3) - \overline{f}(z_1)\overline{f}(z_2)\overline{f}(z_3)\right]\prod_{i=1}^{3} dz_i \\
&+\frac{(n-1)(n-2)}{n^2}\int g(z_1; z_2, z_3)\overline{f}(z_1)\overline{f}(z_2)\overline{f}(z_3)\prod_{i=1}^{3} dz_i. \qquad (3.24)
\end{aligned}$$

Writing $L(z_1; z_2, z_3) = \sigma_2(z_1, z_1) K(z_1, z_2) K(z_1, z_3)$, the first integral in (3.24) is

$$\begin{aligned}
&\frac{1}{n^3}\int L(z_1; z_2, z_3)\sum_{i,j,k}^{n}\left\{f_{ijk}(z_1, z_2, z_3) - \overline{f}(z_1)\overline{f}(z_2)\overline{f}(z_3)\right\}\prod_{i=1}^{3} dz_i \\
&-a^q\frac{1}{n^3}\int \sigma_2(z_1, z_1) K(z_1, z_3)\overline{f}(z_1)\sum_{i,j,k}^{n}\left\{f_{ik}(z_1, z_3) - \overline{f}(z_1)\overline{f}(z_3)\right\} dz_1 dz_3 \\
&-a^q\frac{1}{n^3}\int \sigma_2(z_1, z_1) K(z_1, z_2)\overline{f}(z_1)\sum_{i,j,k}^{n}\left\{f_{ij}(z_1, z_2) - \overline{f}(z_1)\overline{f}(z_2)\right\} dz_1 dz_2. \qquad (3.25)
\end{aligned}$$

The first term is

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

which, as in Lemma B2, is $o\left(n^{-1/2}a^{2q}\right) + O(a^{\zeta})$. The last two terms in (3.25) are bounded in absolute value by

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

which, by Assumption B6, can be shown to be $o\left(n^{-1/2}a^{2q}\right) + O\left(a^{\zeta+q}\right)$. Finally by Lemma 4 of Robinson (1988) and Assumption B4, the second integral in (3.24) is $O\left(a^{2(q+\lambda)}\right)$.

The left side of (ii) is bounded by

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

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

Lemma B4 *As* $n \to \infty$,

$$n^{-1} \sum_{i=1}^{n} \left(U_i^2 + \|V_{1i}\|^2 + \|V_{2i}\|^2 \right) \left(\hat{f}_i - \overline{f}_i \right)^2 = o_p \left(n^{-\frac{1}{2}} \right). \tag{3.26}$$

Proof. By Assumption B4, the expectation of the last contribution to (3.26) is

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

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

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

Lemma B3 and Assumption B13 imply that

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

The remainder of the proof is very similar. ■

Lemma B5 *As* $n \to \infty$,

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

Proof. The last expectation is

$$\begin{aligned} \mathbb{E}\left(\frac{1}{n^3 a^{2q}} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{k \neq i}^{n} V_{2j}' V_{2k} K_{ij} K_{ik} \right) \; = \; & \frac{1}{n^3 a^{2q}} \sum_{i,j,k}^{n} \gamma_{jk}^{(2)} \mathbb{E}\left(\sigma_2 \left(Z_j, Z_k \right) K_{ij} K_{ik} \right) \\ & + \frac{1}{n^3 a^{2q}} \sum_{i,j} \gamma_{jj}^{(2)} \mathbb{E}\left(\sigma_2 \left(Z_j, Z_j \right) K_{ij}^2 \right). \end{aligned} \tag{3.27}$$

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

$$\begin{aligned}
& \frac{1}{n^3 a^{2q}} \int l(z_1; z_2, z_3) \sum_{i,j,k}^{n} \gamma_{jk}^{(2)} F_{jk:i}(z_2, z_3; z_1) \prod_{i=1}^{3} dz_i \\
& + \frac{1}{n^3 a^{2q}} \int l(z_1; z_2, z_3) \sum_{i,j,k}^{n} \gamma_{jk}^{(2)} f_i(z_1) F_{k:j}(z_3; z_2) \prod_{i=1}^{3} dz_i \\
& + \frac{1}{n^3 a^{2q}} \int l(z_1; z_2, z_3) \sum_{i,j,k}^{n} \gamma_{jk}^{(2)} \left\{ f_i(z_1) f_j(z_2) f_k(z_3) - \overline{f}(z_1) \overline{f}(z_2) \overline{f}(z_3) \right\} \prod_{i=1}^{3} dz_i \\
& + \frac{1}{n^3 a^{2q}} \sum_{i,j,k}^{n} \gamma_{jk}^{(2)} \int l(z_1; z_2, z_3) \prod_{i=1}^{3} \left\{ \overline{f}(z_i) dz_i \right\}. \qquad (3.28)
\end{aligned}$$

The last term in (3.28) is bounded in absolute value by

$$\begin{aligned}
& \frac{1}{n^3 a^{2q}} \sum_{i,j,k}^{n} \left| \gamma_{jk}^{(2)} \right| \int \left| \sigma_2(z_2, z_3) - \sigma_2(z_1, z_1) \right| \left| K(z_1, z_2) K(z_1, z_3) \right| \prod_{i=1}^{3} \left\{ \overline{f}(z_i) dz_i \right\} \\
& + \frac{1}{n^3 a^{2q}} \sum_{i,j,k}^{n} \left| \gamma_{jk}^{(2)} \right| \int \left| \sigma_2(z_1, z_1) K(z_1, z_2) K(z_1, z_3) \right| \prod_{i=1}^{3} \left\{ \overline{f}(z_i) dz_i \right\}. \qquad (3.29)
\end{aligned}$$

Applying the last part of the proof of Lemma B1, Assumptions B7 and B10 imply that the integral of the first term in (3.29) is $O\left(a^{2q+1} + a^{\zeta}\right)$. Hence by Assumptions B5 and B13, the first term of (3.29) is $o\left(n^{-1/2}\right)$. The second term in (3.29) is bounded by

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

by Assumptions B5 and B7. For other terms in (3.28), apply the proof of Lemma B2. Altogether it is found that the first term of (3.27) is $o(n^{-1/2}) + O(a^{\zeta - 2q})$.

The second term of (3.27) is bounded by

$$\begin{aligned}
& \max_{1 \leq i \leq n} \left| \gamma_{ii}^{(2)} \right| \frac{1}{n^3 a^{2q}} \sum_{i,j}^{n} \mathbb{E} \left| \sigma_2(Z_j, Z_j) K_{ij}^2 \right| \\
\leq \; & \frac{C}{n^3 a^{2q}} \sum_{i,j}^{n} \left\{ \mathbb{E} \left| \sigma_2(Z_j, Z_j) - \sigma_2(Z_i, Z_i) \right| K_{ij}^2 + \mathbb{E} \left| \sigma_2(Z_i, Z_i) K_{ij}^2 \right| \right\}. \qquad (3.30)
\end{aligned}$$

Applying the proof of Lemma B1, (3.30) is $O\left(n^{-3/2} a^{-q}\right) + O\left(n^{-1} a^{-q} + n^{-1} a^{2\zeta - 2q}\right)$. This proves the last result. The others can be shown similarly. ∎

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

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

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

$$n^{-3}\int u\left(z_1,z_2,z_3,z_4\right)\overline{f}\left(z_1\right)\overline{f}\left(z_2\right)\sum_{i,j,k,l}^{n}\gamma_{ij}^{(2)}F_{ij:k:l}\left(z_1,z_2;z_3;z_4\right)\prod_{i=1}^{4}dz_i$$
$$+n^{-3}\int u\left(z_1,z_2,z_3,z_4\right)\overline{f}\left(z_1\right)\overline{f}\left(z_2\right)\sum_{i,j,k,l}^{n}\gamma_{ij}^{(2)}f_{ij}\left(z_1,z_2\right)\left\{f_k\left(z_3\right)f_l\left(z_4\right)-\overline{f}\left(z_3\right)\overline{f}\left(z_4\right)\right\}\prod_{i=1}^{4}dz_i$$
$$+\frac{1}{n^3}\sum_{i,j,k,l}^{n}\gamma_{ij}^{(2)}\int u\left(z_1,z_2,z_3,z_4\right)f_{ij}\left(z_1,z_2\right)\left\{\prod_{i=1}^{4}\overline{f}\left(z_i\right)dz_i\right\}. \tag{3.31}$$

As in Lemma B2, the first integral is $o\left(a^{2+2q}\right)+o(n^{1/2}a^{\zeta})$. Similarly, the second term in (3.31) can be shown to be of no greater order. The integral of the last term of (3.31) is bounded in absolute value by

$$\left\{\sup_{z_1,z_2}f_{ij}\left(z_1,z_2\right)\right\}\int_{\mathbb{R}^{2q}}\left|\int_{\mathbb{R}^q}\phi(z_1,z_3)K\left(z_1,z_3\right)\overline{f}\left(z_3\right)dz_3\right|$$
$$\left|\int_{\mathbb{R}^q}\phi(z_2,z_4)K\left(z_2,z_4\right)\overline{f}\left(z_4\right)dz_4\right|\sigma_2\left(z_1,z_2\right)\overline{f}\left(z_1\right)\overline{f}\left(z_2\right)dz_1dz_2$$
$$= O(a^{2\{q+\min(\lambda+1,\nu)\}})$$

by Lemma 4 of Robinson (1988), Assumptions B7 and the Cauchy inequality. Thus the last term in (3.31) is $o(n^{1/2}a^{2\{q+\min(\lambda+1,\nu)\}})$ by Assumption B5. ■

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

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

and

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

Proof. Writing

$$l_{ij:kl}\left(z_1,z_2;z_3,z_4\right)=K\left(z_1,z_3\right)K\left(z_2,z_4\right)\overline{f}\left(z_1\right)\overline{f}\left(z_2\right)f_{ijkl}\left(z_1,z_2,z_3,z_4\right),$$

$$
\begin{aligned}
& \mathbb{E}\left|\sigma_2\left(Z_k, Z_l\right) K_{ik} K_{jl} \overline{f}_i \overline{f}_j\right| \\
= & \int \left|\sigma_2\left(z_3, z_4\right) l_{ij:kl}\left(z_1, z_2; z_3, z_4\right)\right| \prod_{i=1}^{4} dz_i \\
\leq & \int \left\{\left|\sigma_2\left(z_3, z_4\right)-\sigma_2\left(z_1, z_2\right)\right|+\left|\sigma_2\left(z_1, z_2\right)\right|\right\}\left|l_{ij:kl}\left(z_1, z_2, z_3, z_4\right)\right| \prod_{i=1}^{4} dz_i . \quad (3.32)
\end{aligned}
$$

The second term in (3.32) is bounded by

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

By Assumption B7, it is uniformly $O\left(a^{2q}\right)$. Writing

$$
p\left(z_1, z_2, u, v, a\right)=\left|\sigma_2\left(z_1+a u, z_2+a v\right)-\sigma_2\left(z_1, z_2\right)\right||K(u) K(v)|,
$$

the first term in (3.32) is

$$
\begin{aligned}
& a^{2q} \int p\left(z_1, z_2, u, v, a\right) \overline{f}\left(z_1\right) \overline{f}\left(z_2\right) f_{ijkl}\left(z_1, z_2, z_1+a u, z_2+a v\right) dz_1 dz_2 du dv \\
= & \left(\int \int_{J_1(\varepsilon/2)}+\int \int_{J_2(\varepsilon/2)}+\int \int_{J_3(\varepsilon/2)}+\int \int_{J_4(\varepsilon/2)}\right), \quad (3.33)
\end{aligned}
$$

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

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

Assumption B7 and B10 imply that $\int \int_{J_2(\varepsilon)}$ is uniformly $O\left(n^{1/2} a^{\zeta-2q}\right)$. Similarly for the other terms in (3.33). The remaining terms of the lemma can be dealt with similarly. ■

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

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

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

$$
\begin{aligned}
& \mathbb{E}\left\|\frac{1}{n a^q} \sum_{i,j}^{n} U_i V_{2j} K_{ij} \overline{f}_i\right\|^2 \leq \frac{1}{n^2 a^{2q}} \max_{1 \leq i \leq n} \gamma_{ii}^{(U)} \mathbb{E}\left(\sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{k \neq i}^{n} V_{2j}^{\prime} V_{2k} K_{ij} K_{ik} \overline{f}_i^2\right) \\
& +\frac{1}{n^2 a^{2q}} \sum_{i,j}^{n} \sum_{k \neq i}^{n} \sum_{l \neq j}^{n} \gamma_{ij}^{(U)} \gamma_{kl}^{(2)} \mathbb{E}\left(\sigma_2\left(Z_k, Z_l\right) K_{ik} K_{jl} \overline{f}_i \overline{f}_j\right) . \quad (3.34)
\end{aligned}
$$

The first term in (3.34) is bounded in absolute value by

$$\frac{C}{n^2a^{2q}}\left\{\max_{1\le j\le n}\left|\gamma_{jj}^{(2)}\right|\sum_{i,j}^{n}\mathbb{E}\left|\sigma_2\left(Z_j,Z_j\right)K_{ij}^2\overline{f}_i^2\right|+\sum_{i,j,k}^{n}\left|\gamma_{jk}^{(2)}\right|\mathbb{E}\left|\sigma_2\left(Z_j,Z_k\right)K_{ij}K_{ik}\overline{f}_i^2\right|\right\}. \tag{3.35}$$

By Lemma B7 the double sum in (3.35) is $O\left(n^2a^q\right)$ and, with Assumption B5, the the triple sum in (3.35) is $o\left(n^{5/2}a^{2q}\right)$. Hence the first term in (3.34) is $O\left(a^{-q}\right)+o(n^{1/2})=o(n)$. The second term in (3.34) is bounded in absolute value by

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

By Lemma B7 and Assumption B5, the second term in (3.34) is $o\left(n+n^{1/2}+n^{-1/2}a^{-q}\right)=o(n)$. This proves the first result. The other can be shown similarly. ■

Lemma B9 *As $n\to\infty$,*

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

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

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

the left side is

$$\begin{aligned}
&\int\sigma_2\left(z_1,z_2\right)v(z_1;z_3)v(z_2;z_4)\overline{f}\left(z_1\right)\overline{f}\left(z_2\right)w\left(z_1,z_2;z_3;z_4\right)\prod_{i=1}^{4}dz_i\\
&+\frac{1}{n^3}\sum_{i,j,k,l}^{n}\gamma_{ij}^{(2)}\gamma_{ij}^{(U)}\int\sigma_2\left(z_1,z_2\right)v(z_1;z_3)v(z_2;z_4)f_{ij}\left(z_1,z_2\right)\prod_{i=1}^{4}\left\{\overline{f}\left(z_i\right)dz_i\right\}.
\end{aligned} \tag{3.36}$$

The first term in (3.36) is

$$\int \sigma_2(z_1,z_2)\,K(z_1,z_3)v(z_2;z_4)w(z_1,z_2;z_3;z_4)\,\overline{f}(z_1)\,\overline{f}(z_2)\prod_{i=1}^{4}dz_i$$
$$-a^q\int \sigma_2(z_1,z_2)\,K(z_2,z_4)w(z_1,z_2;z_3;z_4)\,\overline{f}^2(z_1)\,\overline{f}(z_2)\prod_{i=1}^{4}dz_i, \qquad (3.37)$$

because

$$\int_{\mathbb{R}^{2q}}\left\{f_{ijkl}(z_1,z_2,z_3,z_4)-f_{ij}(z_1,z_2)\,\overline{f}(z_3)\,\overline{f}(z_4)\right\}dz_3dz_4\equiv 0.$$

A leading term in (3.37) is

$$\int \sigma_2(z_1,z_2)\,K(z_1,z_3)\,K(z_2,z_4)\,\overline{f}(z_1)\,\overline{f}(z_2)\left[\frac{1}{n^3}\sum_{i,j,k,l}^{n}\gamma_{ij}^{(2)}\gamma_{ij}^{(U)}F_{ij:k:l}(z_1,z_2;z_3;z_4)\right.$$
$$\left.+\frac{1}{n^3}\sum_{i,j,k,l}^{n}\gamma_{ij}^{(2)}\gamma_{ij}^{(U)}f_{ij}(z_1,z_2)\left\{f_k(z_3)f_l(z_4)-\overline{f}(z_3)\,\overline{f}(z_4)\right\}\right]\prod_{i=1}^{4}dz_i. \qquad (3.38)$$

Similar to the proof of Lemma B2, the integral of the first sum in (3.38) can be shown to be $o\left(a^{2q}+n^{1/2}a^{\zeta}\right)$. Proceeding as in the proofs of Lemmas B6 and B3, remaining terms can be dealt with such that the first term in (3.36) is $o\left(a^{2q}+n^{1/2}a^{\zeta}\right)$. Proceeding as in the proof of Lemma B6, the second term in (3.36) is $o\left(n^{1/2}a^{2(\lambda+q)}\right)$ by Assumptions B5 and B7 and Lemma 4 of Robinson (1988). ■

Lemma B10 *As* $n\to\infty$,

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

Proof. The left side is

$$\mathbb{E}\sum_{i=1}^{n}\gamma_{ii}^{(2)}\gamma_{ii}^{(U)}\sigma_2(Z_i,Z_i)\left(\hat{f}_i-\overline{f}_i\right)^2\overline{f}_i^2$$
$$+\mathbb{E}\sum_{i,j}^{n}\gamma_{ij}^{(2)}\gamma_{ij}^{(U)}\sigma_2(Z_i,Z_j)\left(\hat{f}_i-\overline{f}_i\right)\left(\hat{f}_j-\overline{f}_j\right)\overline{f}_i\overline{f}_j. \qquad (3.39)$$

The first term in (3.39) is bounded by

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

by the proof of Lemma B4 and Assumptions B1, B5 and B7. The second term in (3.39) is

$$\begin{aligned}
&\frac{1}{n^2a^{2q}}\mathbb{E}\sum_{i,j}^{n}\sum_{k\neq i}^{n}\sum_{l\neq j}^{n}\gamma_{ij}^{(2)}\gamma_{ij}^{(U)}\sigma_2\left(Z_i,Z_j\right)M_{ik}M_{jl}\overline{f}_i\overline{f}_j \\
&-\frac{1}{n^2a^{q}}\mathbb{E}\sum_{i,j}^{n}\sum_{l\neq j}^{n}\gamma_{ij}^{(2)}\gamma_{ij}^{(U)}\sigma_2\left(Z_i,Z_j\right)M_{jl}\overline{f}_i^2\overline{f}_j \\
&-\frac{1}{n^2a^{q}}\mathbb{E}\sum_{i,j}^{n}\sum_{k\neq i}^{n}\gamma_{ij}^{(2)}\gamma_{ij}^{(U)}\sigma_2\left(Z_i,Z_j\right)M_{ik}\overline{f}_i\overline{f}_j^2 \\
&+\frac{1}{n^2}\mathbb{E}\sum_{i,j}^{n}\gamma_{ij}^{(2)}\gamma_{ij}^{(U)}\sigma_2\left(Z_i,Z_j\right)\overline{f}_i^2\overline{f}_j^2. \qquad (3.40)
\end{aligned}$$

By Assumptions B5 and B7, the last term in (3.40) is $o\left(n^{-1/2}\right)$. The absolute value of the second term in (3.40) is bounded by

$$\frac{1}{n^2a^q}\sum_{i,j}^{n}\sum_{l\neq j}^{n}\left|\gamma_{ij}^{(2)}\gamma_{ij}^{(U)}\right|\left\{\mathbb{E}\left|\sigma_2\left(Z_i,Z_j\right)K_{jl}\overline{f}_i^2\overline{f}_j\right|+a^q\mathbb{E}\left|\sigma_2\left(Z_i,Z_j\right)\overline{f}_i^2\overline{f}_j^2\right|\right\}. \qquad (3.41)$$

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

$$a^q\sup f_{ijl}\left(z_1,z_2,z_3\right)\sup\overline{f}\left(z\right)\int\left|\sigma_2\left(z_1,z_2\right)\right|\overline{f}\left(z_1\right)\overline{f}\left(z_2\right)\left|K\left(u\right)\right|dudz_1dz_2$$

which, by Assumption B7, is uniformly $O\left(a^q\right)$. Thus by Assumption B5, (3.41) is $o(n^{1/2})$. The same conclusion can be drawn for the third term in (3.40). The first term in (3.40) is

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

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

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

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

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

Appendix 4: Proof of Theorem C

Each element of

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

is bounded in absolute value by that of

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

Then proceed as Robinson (2007) and conclude that

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

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

$$\begin{aligned}
&\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\mathbb{E}\left(x_{ti}x_{tk}U_iU_k\right)\mathbb{E}\left(x_{sj}x_{sl}U_jU_l\right)w\left(s_i - s_j, m\right)w\left(s_k - s_l, m\right)\\
&+\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\mathbb{E}\left(x_{ti}x_{sl}U_iU_l\right)\mathbb{E}\left(x_{sj}x_{tk}U_jU_k\right)w\left(s_i - s_j, m\right)w\left(s_k - s_l, m\right) + O\left(n^{-1}\right).
\end{aligned}$$

The first term has modulus

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

The second term can be handled similarly.

Table 1

Linear regression *(3.1):* Empirical sizes of tests with size α

		$\rho_X = 0.2,\ \rho_U = 0.3$				$\rho_X = 0.4,\ \rho_U = 0.5$		
n	m	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	m	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
	C	.021	.058	.125	C	.037	.119	.185
	H	.027	.063	.138	H	.049	.123	.196
	2	.026	.058	.125	6	.029	.088	.154
100	4	.024	.052	.117	8	.029	.085	.152
	6	.022	.050	.115	10	.029	.084	.152
	8	.023	.052	.119	12	.027	.082	.153
	10	.024	0.56	.122	14	.027	.085	.151
	C	.013	.052	.106	C	.025	.084	.159
	H	.017	.056	.114	H	.028	.095	.163
	3	.016	.054	.109	6	.023	.069	.130
169	6	.013	.050	.104	9	.019	.067	.121
	9	.013	.049	.115	12	.019	.066	.120
	12	.014	.051	.118	15	.020	.067	.125
	15	.016	.061	.120	18	.020	.070	.131

Table 2

Linear regression *(3.1):* Empirical powers of tests with $\beta = 0.8$ and size α

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

Table 3

Partly linear regression *(3.5):* Empirical sizes of tests with size α using k_2

		$\rho_X = 0.2, \quad \rho_U = 0.3$								
		$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.10$		
n	m/a	1.0	1.2	1.4	1.0	1.2	1.4	1.0	1.2	1.4
	C	.012	.011	.009	.057	.048	.047	.111	.094	.087
	H	.013	.015	.015	.056	.053	.064	.109	.109	.107
	2	.015	.014	.014	.055	.050	.061	.109	.106	.106
100	4	.013	.016	.015	.053	.051	.060	.107	.106	.105
	6	.013	.016	.015	.053	.052	.060	.107	.110	.102
	8	.014	.016	.016	.058	.054	.063	.109	.110	.108
	12	.014	.015	.018	.061	.059	.067	.114	.113	.119
	C	.008	.004	.003	.052	.041	.030	.106	.090	.081
	H	.009	.006	.005	.045	.040	.039	.096	.088	.087
	3	.009	.006	.005	.045	.043	.040	.094	.088	.085
169	6	.010	.007	.008	.051	.043	.044	.091	.083	.083
	9	.012	.011	.009	.051	.046	.044	.087	.083	.084
	12	.014	.012	.011	.053	.050	.047	.093	.087	.089
	15	.013	.012	.011	.057	.051	.048	.103	.095	.090

		$\rho_X = 0.4, \quad \rho_U = 0.5$								
		$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.10$		
n	m/a	1.0	1.2	1.4	1.0	1.2	1.4	1.0	1.2	1.4
	C	.021	.018	.016	.069	.064	.063	.127	.123	.117
	H	.017	.019	.027	.076	.071	.073	.133	.143	.135
	6	.014	.014	.024	.066	.065	.070	.116	.125	.119
100	8	.013	.014	.022	.069	.068	.072	.117	.124	.120
	10	.014	.017	.024	.070	.067	.077	.121	.126	.127
	12	.016	.020	.026	.076	.077	.081	.124	.128	.128
	14	.018	.025	.029	.077	.084	.088	.127	.133	.134
	C	.011	.006	.004	.065	.049	.050	.124	.098	.085
	H	.010	.009	.010	.056	.054	.059	.104	.100	.102
	6	.010	.007	.009	.053	.053	.053	.099	.097	.099
169	9	.010	.009	.010	.056	.052	.055	.098	.095	.091
	12	.010	.013	.011	.056	.052	.054	.095	.090	.091
	15	.015	.016	.012	.056	.055	.053	.102	.098	.095
	18	.016	.018	.014	.061	.055	.055	.109	.109	.104

Table 4

Partly linear regression *(3.5):* Empirical sizes of tests with size α using k_4

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

Table 5

Partly linear regression *(3.5)*: Empirical powers of tests with $\beta = 0.7$ using k_2, k_4 at level α.

		k_2								
		$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.10$		
n	m/a	1.0	1.2	1.4	1.0	1.2	1.4	1.0	1.2	1.4
100	C	.536	.521	.485	.760	.744	.728	.830	.826	.826
	H	.519	.527	.535	.743	.744	.744	.817	.831	.842
	2	.515	.531	.534	.739	.750	.750	.817	.834	.846
	4	.511	.534	.537	.741	.752	.751	.818	.831	.844
	6	.511	.543	.541	.743	.757	.761	.819	.831	.845
	8	.521	.543	.556	.745	.757	.762	.823	.829	.841
	12	.530	.547	.559	.744	.754	.767	.827	.835	.844
169	C	.810	.794	.788	.929	.929	.918	.962	.960	.958
	H	.775	.795	.801	.917	.923	.928	.950	.957	.964
	3	.778	.796	.804	.914	.925	.927	.951	.956	.961
	6	.778	.798	.804	.916	.925	.926	.947	.956	.963
	9	.777	.807	.810	.910	.920	.927	.949	.958	.958
	12	.782	.808	.816	.913	.922	.929	.949	.958	.959
	15	.790	.815	.823	.914	.922	.927	.946	.959	.958

		k_4								
		$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.10$		
n	m/a	1.4	1.6	1.8	1.4	1.6	1.8	1.4	1.6	1.8
100	C	.546	.523	.499	.753	.737	.730	.825	.814	.812
	H	.508	.517	.523	.723	.737	.738	.797	.813	.820
	2	.503	.519	.524	.723	.736	.743	.797	.812	.821
	4	.508	.519	.523	.721	.735	.740	.796	.810	.821
	6	.501	.518	.535	.724	.735	.743	.799	.811	.821
	8	.508	.529	.538	.723	.742	.745	.802	.812	.823
	12	.518	.536	.548	.725	.743	.747	.804	.817	.827
169	C	.805	.791	.787	.924	.925	.918	.955	.956	.955
	H	.759	.774	.791	.903	.916	.918	.943	.952	.955
	3	.759	.773	.793	.903	.919	.922	.945	.952	.955
	6	.766	.774	.796	.902	.916	.923	.944	.952	.953
	9	.760	.779	.798	.897	.914	.922	.946	.949	.956
	12	.764	.784	.805	.900	.913	.921	.945	.950	.956
	15	.767	.792	.804	.902	.914	.924	.943	.951	.956

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

	LS	IV	Partly LS				Partly IV			
			k_2		k_4		k_2		k_4	
			a=2	a=2.5	a=1.4	a=1.9	a=1.7	a=2.2	a=1.4	a=1.9
NL	**.72**	**.37**	**.95**	**.96**	**.85**	**.90**	**.80**	**.81**	**.76**	**.78**
	(.21)	(.63)	(.18)	(.18)	(.16)	(.17)	(.41)	(.42)	(.40)	(.41)
m=2	(.23)	(.58)	(.22)	(.22)	(.22)	(.22)	(.43)	(.44)	(.42)	(.42)
m=4	(.25)	(.69)	(.24)	(.24)	(.24)	(.24)	(.52)	(.53)	(.51)	(.52)
m=6	(.27)	(.73)	(.25)	(.25)	(.25)	(.25)	(.54)	(.55)	(.53)	(.54)
NL2	**−.71**	**−.29**	**−.91**	**−.92**	**−.83**	**−.87**	**−.70**	**−.69**	**−.71**	**−.70**
	(.19)	(.61)	(.17)	(.17)	(.16)	(.16)	(.41)	(.42)	(.40)	(.41)
m=2	(.20)	(.60)	(.20)	(.20)	(.19)	(.19)	(.45)	(.46)	(.43)	(.44)
m=4	(.23)	(.74)	(.21)	(.21)	(.21)	(.21)	(.56)	(.58)	(.53)	(.55)
m=6	(.24)	(.79)	(.22)	(.22)	(.21)	(.21)	(.58)	(.60)	(.54)	(.57)
DBC$\times10^{-3}$	**−1.62**	**−1.90**	**−1.53**	**−1.40**	**−2.02**	**−1.77**	**−1.87**	**−1.80**	**−2.07**	**−1.94**
	(.66)	(.77)	(.63)	(.64)	(.57)	(.60)	(.70)	(.73)	(.62)	(.67)
m=2	(.74)	(.77)	(.72)	(.73)	(.70)	(.70)	(.68)	(.70)	(.68)	(.68)
m=4	(.85)	(.87)	(.82)	(.84)	(.80)	(.80)	(.77)	(.78)	(.77)	(.77)
m=6	(.83)	(.82)	(.80)	(.83)	(.75)	(.75)	(.72)	(.74)	(.70)	(.71)
CD$\times10^{-1}$	**.41**	**.38**	**−.60**	**−.65**	**−.39**	**−.49**	**−.69**	**−.80**	**−.43**	**−.60**
	(.57)	(.72)	(.51)	(.50)	(.54)	(.52)	(.55)	(.55)	(.58)	(.56)
m=2	(.56)	(.63)	(.54)	(.54)	(.51)	(.52)	(.56)	(.58)	(.53)	(.55)
m=4	(.52)	(.58)	(.59)	(.60)	(.55)	(.57)	(.64)	(.67)	(.58)	(.62)
m=6	(.48)	(.53)	(.63)	(.64)	(.57)	(.60)	(.69)	(.72)	(.59)	(.66)
BSD	**−.16**	**−.21**	**−.13**	**−.13**	**−.11**	**−.12**	**−.15**	**−.17**	**−.12**	**−.14**
	(.05)	(.07)	(.04)	(.05)	(.04)	(.04)	(.06)	(.06)	(.05)	(.05)
m=2	(.04)	(.06)	(.04)	(.04)	(.03)	(.04)	(.06)	(.06)	(.05)	(.05)
m=4	(.05)	(.09)	(.04)	(.04)	(.04)	(.04)	(.07)	(.08)	(.06)	(.07)
m=6	(.06)	(.10)	(.05)	(.05)	(.04)	(.04)	(.08)	(.09)	(.06)	(.07)
RSD$\times10^{-1}$	**.14**	**.25**	**−.44**	**−.49**	**−.23**	**−.33**	**−.37**	**−.43**	**−.20**	**−.31**
	(.48)	(.57)	(.44)	(.44)	(.42)	(.43)	(.46)	(.46)	(.46)	(.45)
m=2	(.51)	(.60)	(.51)	(.52)	(.47)	(.49)	(.53)	(.54)	(.53)	(.52)
m=4	(.40)	(.55)	(.56)	(.47)	(.42)	(.44)	(.50)	(.50)	(.51)	(.49)
m=6	(.34)	(.51)	(.43)	(.44)	(.40)	(.41)	(.47)	(.48)	(.50)	(.47)
ASD$\times10^{-1}$	**.62**	**.54**	**.81**	**.79**	**.84**	**.83**	**.75**	**.71**	**.82**	**.78**
	(.35)	(.38)	(.35)	(.35)	(.33)	(.34)	(.36)	(.37)	(.34)	(.35)
m=2	(.35)	(.36)	(.38)	(.38)	(.38)	(.38)	(.37)	(.38)	(.37)	(.37)
m=4	(.37)	(.37)	(.41)	(.42)	(.39)	(.40)	(.39)	(.40)	(.38)	(.39)
m=6	(.34)	(.34)	(.38)	(.39)	(.36)	(.36)	(.36)	(.37)	(.35)	(.35)

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

Table 7: $Y =$ Fertilizer use (FU)

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

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

Table 8: $Y =$ Log(yield 15 major crops) (L15)

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

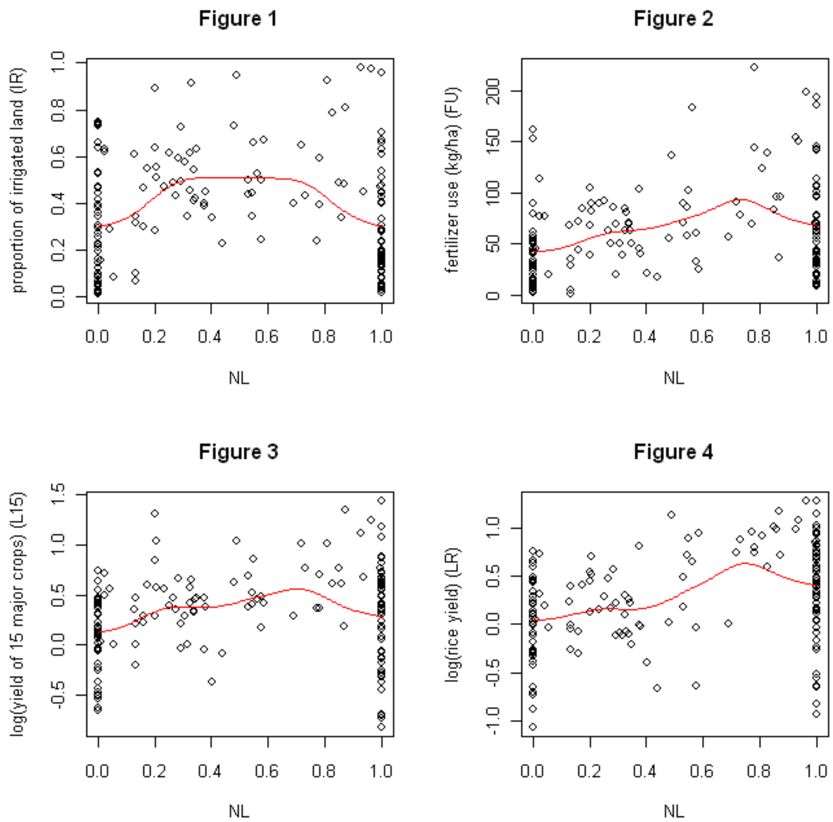
Slope estimates are in bold; SEs are in parentheses; with non-robust ones in the top row, and robust ones below computed using truncation vectors (m, m) where $m = 2$, 4 and 6 respectively; columns under Partial LS and Partial IV refer to choices of bandwidth a and kernel $(k_2,\ k_4)$.

Table 9: $Y =$ Log(rice yield) (LR)

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

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

Figures 1-4: Nonparametric regressions



4 Linear Analysis with Irregularly Spaced Data

4.1 Introduction

In this chapter, we consider linear regressions and some linear analysis with dependent data. The main focus is on dependence across (economic) space but the same principle could be applied to dependence through time. However, dependence through time may need a special treatment since some nice results can be established only when working through time by exploiting the fact that the real number system is an ordered field. See for example Robinson (1980). The main problem arising with spatial data is that (economic) locations are usually irregularly spaced. This makes statistical inference difficult or even analytically intractable (asymptotically) due to the fact that the covariance structure of the disturbances are unlikely to have the Toeplitz form. Various models have been proposed to overcome this difficulty. There seem to be two popular approaches in econometric literature and both can be regarded as extensions from classical time series analysis.

The most common methodology is to approximate dependence of the data by a linear process. The Spatial Autoregressive (SAR) model has been a popular parametric model in econometric literature. More recently Robinson (2011) proposed a generalized linear process for a triangular array of random variables nesting the SAR model as a special case. Even though this approach offers a close resemblance to most linear time series models, it often lacks stationarity and hence asymptotic covariances of some simple statistics may become intractable.

The other methodology is to maintain some form of stationarity. Conley (1999), for example, regarded an irregularly spaced data as a random sample of some underlying random fields on a lattice. There are a few problems related to this particular interpretation. First, the computation can be an issue. Conley assumed that the locations follow a hard-core point process, i.e. there are no pairs of locations whose distances are smaller than a particular positive number. Some computation involves dividing a subset of $\mathbb{R}^2$ into squares where there is at most one observation in each square. Consider an analysis of data on factory plants collected from several districts. Due to regulations, availability of infrastructure or economies of scale, their locations often cluster on a few small areas rather than scattering uniformly over the area of interest. In this case, a large proportion of squares will be empty, and an analysis based on dividing into squares as well as computation from those empty squares will be very computationally intensive.

Second, the assumption of a hard-core point process may be too strong. It prohibits locations of the observations from getting arbitrarily close to each other. This drawback makes the hard-core process inappropriate for many applications. In statistical analysis of locations, the Poisson process is often a popular choice for modelling town or village locations that could be useful for data related to applications in political economy or development economics. A Poisson cluster point process such as the Neyman-Scott process may be more suitable for modelling locations of industrial plants in applications related to economic geography, trade or innovation and growth. As mentioned above, due to regulations, infrastructure and, more importantly, economies of scale, economic activities, such as firm locations, tend to cluster around a few hubs. See Fujita, Krugman and Venables (2001).

Some underlying processes determining the locations of economic activities such as a cluster point process may be more appropriate than a hard-core point process.

The model considered in this chapter is very similar to the ones in Brillinger (1972), (1975) and (1986). There are two main differences. First, to derive asymptotic normality our proof is based upon mixing assumptions whereas Brillinger's asymptotic normality was derived using the method of moments requiring the data to have moments of all orders that could be too strong for economic data. Second, Brillinger assumed that the sampling process and the underlying random field are independent. Even though this assumption is important for a study of an underlying random field, it can be dropped as far as the unknown slope parameters of a linear regression are concerned.

Section 2 begins with a motivation for interpreting economic data as a realization of a marked point process and later we will show that many statistics of interest can be regarded as a random (signed) measure. To avoid making complicated assumptions, we mainly focus on the implication of second-order stationarity. Following Thornett (1979), there is a unique measure, we call it a spectral measure, related to a random (signed) measure exhibiting some kind of second-order stationarity. The relationship is similar to that of a spectral measure and a covariance function of covariance stationary time series. However, unlike in time series analysis, the spectral measure associated with a random (signed) measure is not totally finite. Hence, the standard technique involving inversion of a Fourier transform may not be sufficient to determine some properties of the spectral measure of the point process. Also in Section 2 we discuss existence of a continuous spectral density of a random (signed) measure. In Section 3, we discuss a law of large number and a central limit theorem for a random (signed) measure. There is one difficulty arising once one moves from time to space. The region, called the sampling region, from which the data are collected could have an irregular shape. This problem may be a norm rather than an exception. As a result, we need to allow the sampling region to have an arbitrary shape. The concept of van Hove convergence is employed and some discussion on van Hove convergence is provided in the same section. In Section 4 we discuss asymptotic distribution of the least squares estimate of a linear regression model. Section 5 is concerned with spectral estimation of the continuous spectral density of a random signed measure discussed in Section 2. In Section 6, somewhat unrelated to the other parts of the paper, we discuss how to transform an estimate of an unknown positive definite (p.d.) matrix into a positive definite estimate such that if the mean square error (MSE) of the original estimate is $O\left(n^{-\delta}\right)$, where $\delta > 0$, then for any $\varepsilon > 0$, the MSE of the transformed estimate can be $o\left(n^{-\delta+\varepsilon}\right)$. This result can be applied to an estimation of the covariance matrix in Section 4, as well as an estimation problem in other context such as an estimation of an optimal weighting matrix of the GMM objective function that is required to be at least positive semidefinite (p.s.d.). This would enable practitioners to employ smoothing with higher-order kennels, and later obtain a transformed estimate that is p.d. as well as achieving a rate of convergence, in mean square, arbitrarily closed to the original one. Proof of Theorems and technical lemmas can be found in the Appendices.

4.2 Models

Because economic experiments are rare, most economic data is observational. In empirical work, a practitioner usually has to select a particular span of time or an area, such as district and country, or both for panel data, to collect data. If variables of interest Z_i are observed at locations s_i, then the data is $\{(s_1, Z_1), ..., (s_N, Z_N)\}$. With regularly spaced time series, the data can be ordered through time and the locations can be treated as natural numbers so that the time index can represent the points at which Z_i are observed. As a result, the data can be treated as a sequence of random variables $\{Z_1,\ ...,\ Z_N\}$. When the data are collected across space, the number observations N and the locations s_i can potentially become random.

Theoretically one can suppose that there is a sequence of random variables $\{(s_i,\ Z_i)\}$ where s_i and Z_i are $\mathbb{R}^d$-valued and $\mathbb{R}^p$-valued random variables. If a practitioner chooses to collect data over an area A, then the observed data is $\{(s_i,\ Z_i) : s_i \in A\}$. It is assumed in this paper that a sampling area A is restricted to belong to the Borel σ-field of $\mathbb{R}^d$, denoted by $\mathcal{B}(\mathbb{R}^d)$ only. If $A \in \mathcal{B}(\mathbb{R}^d)$, A is said to be a Borel subset of $\mathbb{R}^d$, or simply a Borel set. In general s_i can be generalized to take values in a complete separable metric space but since the main focus of this paper is on weak stationarity we are only concerned with the Euclidean space $\mathbb{R}^d$.

Suppose that the underlying probability space is $(\Omega,\ \mathcal{F},\ \mathbb{P})$.

Assumption A1 *There exists an event $F_0 \in \mathcal{F}$ such that $\mathbb{P}(F_0) = 0$ and $\omega \notin F_0$ implies that for any bounded $A \in \mathcal{B}(\mathbb{R}^d)$, only a finite number of the elements of $\{s_i(\omega)\}$ lie in A.*

Define

$$N_g(A) = \sum_{i=1}^{\infty} \delta_{s_i}(A),$$

where δ_x is the Dirac measure, i.e. $\delta_x(A) = 1$ if $x \in A$ and zero otherwise. Under Assumption A1, by Proposition 9.1.X in Daley and Vere-Jones (2008), N_g is a point process. For any bounded $B \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^p)$ define

$$N(B) = \sum_{i=1}^{\infty} \delta_{(s_i,\ Z_i)}(B).$$

Then Assumption A1 and the Proposition in Daley and Vere-Jones imply that N is also a point process. For $\omega \notin F_0$ and for any bounded $B \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^p)$, there exists a bounded $A \in \mathcal{B}(\mathbb{R}^d)$ such that $B \subset A \times \mathbb{R}^p$, and there are only finite elements of $\{(s_i(\omega),\ Z_i(\omega))\}$ lying in $A \times \mathbb{R}^p$. As both N and N_g are point processes, N can be regarded as a marked point process where Z_i are the marks and N_g is called the ground process. Thanks to this result and the weakness of Assumption A1, observational economic data whose locations are random may be naturally regarded as a marked point process. Before continuing further discussion it is worth introducing some definitions, related to point processes, less common in econometric literature. Good sources of reference are Daley and Vere-Jones (2003, 2008)

4.2.1 Background on Point Processes

Let $\mathcal{S}$ be a complete separable metric space (c.s.m.s.) and $\mathcal{B}(\mathcal{S})$ be the σ-field of its Borel sets. A Borel measure μ on the c.s.m.s. $\mathcal{S}$ is boundedly finite if $\mu(A) < \infty$ for every bounded Borel set A. A boundedly finite integer-valued measure is called a counting measure. Let $\mathcal{N}_{\mathcal{S}}$ be the set of all counting measures; $\mathcal{N}^*_{\mathcal{S}}$ be the set of all simple counting measures, i.e. $N \in \mathcal{N}^*_{\mathcal{S}}$ if and only if $N \in \mathcal{N}_{\mathcal{S}}$ and $N(\{s\}) = 0$ or 1 for all $s \in \mathcal{S}$; and $\mathcal{N}^g_{\mathcal{S}\times\mathcal{K}}$ be the set of all boundedly finite counting measures defined on the product space $\mathcal{B}(\mathcal{S}\times\mathcal{K})$, where $\mathcal{K}$ is a c.s.m.s. of marks, such that the ground measure N_g defined by $N_g(A) \equiv N(A\times\mathcal{K})$, for all $A \in \mathcal{B}(\mathcal{S})$, is a boundedly finite simple counting measure, i.e. $N_g \in \mathcal{N}^*_{\mathcal{S}}$.

A point process N on state space $\mathcal{S}$ is a measurable mapping $N : \Omega \to \mathcal{N}_{\mathcal{S}}$. A point process N is simple when $\mathbb{P}\{N \in \mathcal{N}^*_{\mathcal{S}}\} = 1$. A marked point process on $\mathcal{S}$ with marks in $\mathcal{K}$ is a point process N on $\mathcal{B}(\mathcal{S}\times\mathcal{K})$ for which $\mathbb{P}\left\{N \in \mathcal{N}^g_{\mathcal{N}\times\mathcal{K}}\right\} = 1$ where its ground process is $N_g(\cdot) \equiv N(\cdot\times\mathcal{K})$. In general we could regard a marked point process as a point process on the space $\mathcal{S}\times\mathcal{K}$ as described above or as a sequence of pairs of vectors of random variables $\{(s_i, \kappa_i)\}$ where it requires that for any bounded Borel sets $A \in \mathcal{B}(\mathcal{S})$, with probability one only a finite number of the element $\{s_i\}$ lie in A, and $s_i \neq s_j$ for all $i \neq j$.

For any bounded $A \in \mathcal{B}(\mathcal{S})$, $N(A)$ is a random variable (see Corollary 9.1.IX in Daley and Vere-Jones (2008)) and define $M(A) = \mathbb{E}[N(A)]$. Since $\mathcal{S}$ is a c.s.m.s., there exists a class of bounded Borel sets generating $\mathcal{B}(\mathcal{S})$. Let $\mathcal{R}_1$ be a ring generated by all bounded Borel sets. Suppose that $M(A) < \infty$ for all $A \in \mathcal{R}_1$. Then M is finitely additive on $\mathcal{R}_1$ since N is also finitely additive. For any increasing sequence of sets in $\mathcal{R}_1$, $\{A_n\}$, such that $\lim_n A_n = A \in \mathcal{R}_1$, by monotone convergence $\lim_{n\to\infty} M(A_n) = M(A)$. Hence M is countably additive on $\mathcal{R}_1$ and therefore M can be uniquely extended to a measure on $\mathcal{B}(\mathcal{S})$. We say that the first moment measure of N exists when $\mathbb{E}[N(A)] < \infty$ for all bounded Borel sets A.

Let $N_2(A_1\times A_2) = N(A_1)N(A_2)$ for any Borel sets A_1 and A_2. Then $N_2(\cdot)$ can be extended to a product measure on the product σ-field $\mathcal{B}(\mathcal{S}\times\mathcal{S})$. Then it can be shown that $N_2(\cdot)$ is also a point process on $\mathcal{S}\times\mathcal{S}$. Define $M_2(A) = \mathbb{E}[N_2(A)]$ for a Borel set A in $\mathcal{B}(\mathcal{S}\times\mathcal{S})$. We say that the second moment measure of N exists when $\mathbb{E}[N_2(A)] < \infty$ for all bounded Borel sets A. Let $f : \mathcal{S} \to \mathbb{R}$ be a Borel measurable function. The integral $\int_{\mathcal{X}} f(s)\, N(ds)$ is defined as the Lebesgue integral on a realization-by-realization basis.

From now on, let $N(\cdot)$ be a marked point process on the state space $\mathbb{R}^d \times \mathbb{R}^p$, where $d,\ p \in \mathbb{N}$, the set of natural numbers. For a measurable function $f : \mathbb{R}^p \to \mathbb{R}$ and a bounded Borel set $A \in \mathcal{B}\left(\mathbb{R}^d\right)$, the sum

$$\sum_{s_i\in A} f(z_i) = \int_{A\times\mathbb{R}^p} f(z)\, N(ds\times dz) = N_f(A)$$

can be shown to be a well defined random variable. If f is the indicator function of a bounded Borel subset of $\mathbb{R}^p$, then $N_f(A)$ is a random variable. For a nonnegative f, we can approximate it by an increasing sequence of simple functions $\{f_n\}$. For each n, $N_{f_n}(A)$ is a random variable, and by Monotone convergence, $N_f(A)$ is a random variable. This can be extended to a real-valued measurable function f in the usual way.

4.2.2 Second-Order Stationary Point Processes

In this chapter, we employ the definition of a second-order stationary point process from Daley and Vere-Jones (2003).

Definition A1 *A point process $N(\cdot)$ on the state space $\mathbb{R}^d$ is second-order stationary if its second moment measure exists and*

(i) for any bounded measurable function f of bounded support,

$$\int_{\mathbb{R}^d} f(s)\, M(ds) = \mu \int_{\mathbb{R}^d} f(s)\, ds, \tag{4.1}$$

where nonnegative constant μ is the mean density;

(ii) for any bounded measurable function f of bounded support,

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} f(s,t)\, M_2(ds \times dt) = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} f(x, x+u)\ dx\ \breve{M}_2(du), \tag{4.2}$$

where $\breve{M}_2(\cdot)$ is the reduced second moment measure.

If $N(\cdot)$ is a stationary point process on the state space $\mathbb{R}^d$ with a second moment measure, then the conditions in Definition A1 hold. For any subset A of $\mathbb{R}^d$ and $u \in \mathbb{R}^d$, define

$$T_u A = \{a + u : a \in A\}. \tag{4.3}$$

If a point process $N(\cdot)$ on $\mathbb{R}^d$ is second-order stationary with mean density μ, then for any bounded Borel sets $A,\ B$ and $u \in \mathbb{R}^d,\ \mathbb{E}[N(A)] = \mathbb{E}[N(T_u A)]$ and $\mathbb{E}[N(A)\, N(B)] = \mathbb{E}[N(T_u A)\, N(T_u B)]$.

For a second-order stationary point process, we define the reduced covariance (signed) measure in the following differential form

$$\breve{C}_2(du) = \breve{M}_2(du) - \mu^2 du.$$

It can be regarded as the reduced measure of the covariance measure where the covariance measure is defined in a similar fashion as the second moment measure. In fact, the covariance measure can be regarded as the second-moment measure of the random signed measure $\widetilde{N}(A) = N(A) - \mu\ell(A)$ defined on any Borel set A, where $\ell(\cdot)$ is the Lebesgue measure.

Definition A2 *We say that a measure is translation-bounded if for all $h > 0$ and $x \in \mathbb{R}^d$ there exists a finite constant K_h such that, for every ball $B_h(x) = \left\{y \in \mathbb{R}^d : \|x - y\| < h\right\}$,*

$$|\mu(B_h(x))| \le K_h,$$

where $\|\cdot\|$ is the usual Euclidean norm.

It follows from Proposition 8.2.I in Daley and Vere-Jones (2003) that if N is a second-order stationary point process, then there exists a symmetric, translation-bounded measure F on $\mathcal{B}\left(\mathbb{R}^d\right)$ such that for all ψ in the Schwartz space, denoted by $\mathcal{S}\left(\mathbb{R}^d\right)$,

$$\int_{\mathbb{R}^d} \psi(x)\ \breve{C}_2(dx) = \int_{\mathbb{R}^d} \widetilde{\psi}(\lambda)\ F(d\lambda), \tag{4.4}$$

where $\widetilde{\psi}(\lambda) = \int_{\mathbb{R}^d} e^{i\langle\lambda,u\rangle}\psi(u)\ du$ and $\langle\lambda,u\rangle = \sum_{i=1}^{d}\lambda_i u_i$ is the usual inner product on $\mathbb{R}^d$. Daley and Vere-Jones (2003) call the measure F the spectral measure. Since the Parseval identity holds for all $\psi \in \mathcal{S}\left(\mathbb{R}^d\right)$, it follows that the translation-bounded measure F is uniquely determined. Unlike in time series analysis, the spectral measure is not totally finite. It is just translation-bounded. Now we state our first result.

Theorem A1 *Suppose that a random measure ξ on the state space $\mathbb{R}^d$ is second-order stationary with reduced covariance measure $\breve{C}_2$ such that*

$$\int_{\mathbb{R}^d} \left|\breve{C}_2\right|(du) < \infty,$$

where the measures $\left|\breve{C}_2\right| = \breve{C}_2^+ + \breve{C}_2^-$, $\breve{C}_2^+$ and $\breve{C}_2^-$ are defined by the Hahn decomposition of the signed mesure $\breve{C}_2$, then its spectral measure F_ξ is absolutely continuous with non-negative and continuous Radon-Nikodym derivative

$$f_\xi(\lambda) = (2\pi)^{-d}\int_{\mathbb{R}^d} e^{-i\langle\lambda,u\rangle}\ \breve{C}_2(du).$$

Following Theorem A1, since F_ξ is absolutely continuous, its Radon-Nikodym derivative is unique only almost everywhere. Because f_ξ is continuous, it is the only continuous Radon-Nikodym derivative of F_ξ. Hence, we call f_ξ the spectral density of ξ.

4.2.3 Wide-Sense Second-Order Stationary Random Signed Measure

In most economic applications we have to deal with random signed measures rather than random measures. So we adopt the following definition for second-moment stationarity for a random signed measure from Thornett (1979). Let ℓ denote the Lebesgue measure. For a sequence of sets $\{A_n\}$ we say that it is increasing if $A_n \subset A_{n+1}$ for all $n \in \mathbb{N}$ and decreasing if $A_n \supset A_{n+1}$ for all $n \in \mathbb{N}$. For any complex number W, let $\overline{W}$ denote its complex conjugate.

Definition A3 *A wide-sense second-order stationary random signed measure on $\mathbb{R}^d$ is a jointly distributed family of real- or complex-valued random variables $\left\{W(A) : A \in \mathcal{B}\left(\mathbb{R}^d\right)\right\}$*

satisfying the following conditions, for any bounded $A,\ B \in \mathcal{B}\left(\mathbb{R}^d\right)$ *and any sequence of bounded borel sets* $\{A_n\}$,

(i) $\mathbb{E}\left|W(A)\right|^2 < \infty$;

(ii) *for some constant* μ, $\mathbb{E}\{W(A)\} = \mu\ell(A)$,

(iii) $\mathbb{E}\left\{W(T_uA)\overline{W(T_uB)}\right\} = \mathbb{E}\left\{W(A)\overline{W(B)}\right\}$ *for all* $u \in \mathbb{R}^d$;

(iv) $W(A\cup B) = W(A) + W(B)$ *for disjoint* $A,\ B$; *and*

(v) *if* $\{A_n\}$ *is decreasing and* $\lim_{n\to\infty}\ell(A_n) = 0$, *then* $\lim_{n\to\infty}\mathbb{E}\left\{W(A_n)^2\right\} = 0$.

The equality in (iii) is in the mean square sense.

Let $\mathcal{C}$ be the set of all Borel measures F such that (i) $\int_{\mathbb{R}^d}\left|\widetilde{\mathbf{1}}_A(\lambda)\right|^2\ F(d\lambda) < \infty$ for all bounded $A \in \mathcal{B}\left(\mathbb{R}^d\right)$, where $\mathbf{1}$ is the indicator function; (ii) if $\{A_n\}$ is a decreasing sequence of bounded Borel sets such that $\lim_{n\to\infty}\ell(A_n) = 0$, then $\lim_{n\to\infty}\int_{\mathbb{R}^d}\left|\widetilde{1}_{A_n}(\lambda)\right|^2\ F(d\lambda) = 0$. Thornett (1979) extended Bochner's theorem by showing the following result.

Proposition A *If* $\{W(A)\}$ *is wide-sense second-order stationary random signed measure on* $\mathbb{R}^d$ *with* $\mu = 0$, *then there is a unique measure* F_W *in* $\mathcal{C}$ *such that*

$$\mathbb{E}\left\{W(A)\overline{W(B)}\right\} = \int_{\mathbb{R}^d}\widetilde{\mathbf{1}}_A(\lambda)\,\overline{\widetilde{\mathbf{1}}}_B(\lambda)\ F_W(d\lambda) \quad \textit{for all bounded } A,\ B \in \mathcal{B}\left(\mathbb{R}^d\right).$$

The measure F_W *is called the spectral measure of* W.

In order to discuss spectral density of a random signed measure, we restrict ourselves to a certain class of random signed measures.

Assumption A2 W *is a real-valued wide-sense second-order stationary random signed measure on* $\mathbb{R}^d$ *such that for any bounded Borel sets* A *and* B, *there exists a signed measure* $\breve{C}_2$ *such that*

$$Cov\left(\int_{\mathbb{R}^d}\mathbf{1}_A(s)\ W(ds),\ \int_{\mathbb{R}^d}\mathbf{1}_B(s)\ W(ds)\right) = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathbf{1}_A(s)\,\mathbf{1}_B(s+u)\ ds\ \breve{C}_2(du).$$

Theorem A2 *Suppose that Assumption A2 holds and*

$$\int_{\mathbb{R}^d}\left|\breve{C}_2\right|(du) < \infty, \tag{4.5}$$

then the spectral measure F_W *of* W *is absolutely continuous with continuous Radon-Nikodym derivative, called spectral density,*

$$f_W(\lambda) = (2\pi)^{-d}\int_{\mathbb{R}^d}e^{-i\langle\lambda,u\rangle}\ \breve{C}_2(du).$$

Now consider the main object of interest

$$\zeta(A) = \sum_{s_i \in A} z_i,$$

where $\{(s_i, z_i)\}$ is a marked point process. This linear functional of the marked point process arises naturally in many statistics popular in econometric literature such as the least squares estimate of a linear regression. In economic applications it may be a sensible assumption to assume that given the locations $\{s_i\}$ the conditional expectation of marks z_i and z_j depends only on $s_i - s_j$ and conditional covariances vanish to zero as the distances increase. In addition, it may be sensible to assume that the ground process N_g is second-order stationary. As a result, we make the following assumptions.

Assumption A3 *Assumption A1 holds where $s_i \neq s_j$ for $i \neq j$, and the ground process N_g is second-order stationary.*

Assumption A4 *For all $i \in \mathbb{N}$, $\mathbb{E}(z_i | N_g) = 0$. In addition, there exists a measurable function γ_z such that $E(z_i z_j | N_g) = \gamma_z(s_i - s_j)$ for all $i,\ j \in \mathbb{N}$ and $\gamma_z(0) < \infty$, where $\mathbb{E}(\cdot | N_g)$ is a conditional expectation given the ground process.*

Theorem A3 *Under Assumptions A3 and A4, the random signed measure defined by $\zeta(A) = \sum_{s_i \in A} z_i$ for any $A \in \mathcal{B}(\mathbb{R}^d)$ is wide-sense second-order stationary. If*

$$\int_{\mathbb{R}^d} |\gamma_z(u)| \ \breve{M}_2(du) < \infty, \tag{4.6}$$

where $\breve{M}_2$ is the reduced second-order moment measure of the ground process, then the spectral measure F_ζ of ζ is absolutely continuous with continuous density

$$(2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i\langle\lambda,u\rangle} \gamma_z(u) \ \breve{M}_2(du).$$

Condition (4.6) in Theorem A3 is analogous to the one of time series with short memory. Suppose that z_i are $\mathbb{R}^p$-valued random variables and

$$\int_{\mathbb{R}^d} |\gamma_{rs}(u)| \ \breve{M}_2(du) < \infty, \quad r,\ s = 1,\ \cdots,\ p, \tag{4.7}$$

where $\gamma_{rs}(u)$ is the (r, s)-th element of the matrix $\gamma_z(u)$. Motivated by Theorem A3, we may regard $(2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i\langle\lambda,u\rangle} \gamma_z(u) \ \breve{M}_2(du)$ as the spectral density of the $\mathbb{R}^p$-valued random signed measure $\zeta = \sum_{s_i \in A} z_i$.

4.3 Asymptotic Properties of Random Signed Measures

We consider some asymptotic properties of a sequence of random variables $\{\xi(A_n)\}$ where ξ is a random signed measure and $\{A_n\}$ is a sequence of bounded Borel sets. These asymptotic properties are also applicable to non-stationary random signed measures. Typically in statistical literature, asymptotic properties of a random signed measure are restricted to a certain class of sequences of bounded Borel sets. The main attention in point processes is on sequences of convex Borel sets. For a given Borel set with Lebesgue measure of unity, Brillinger (1986) considered a sequence of this particular Borel set scaled by some indices where the indices go to infinity. One problem facing practitioners is that, in economic applications, typical assumptions on a sequence of sampling regions mentioned above tend not to hold. Empirical economists usually select a city or a county to collect data. The concept of the sampling region going to infinity can be more naturally interpreted as including other cities or counties in the sample. Therefore, we discuss a weak law of large numbers and a central limit theorem covering a sequence of arbitrary bounded Borel sets. However we still impose a weak assumption on the way in which these sampling regions get arbitrarily large.

Our asymptotic results are based on the following concept of strong mixing. First, we introduce some notation. For any subset A of $\mathbb{R}^d$, let $\delta(A)$ denote the diameter of A, i.e. $\delta(A) = \sup\{\|t-s\| : s,\ t \in A\}$, where $\|\cdot\|$ denotes the Euclidean norm. For any nonempty subsets A and B of $\mathbb{R}^d$, their distance is defined by $D(A,\ B) = \inf\{\|t-s\| : s \in A,\ t \in B\}$. It should be stressed that D is not a metric. Given a point process or a random signed measure ξ, for any $E \in \mathcal{B}(\mathbb{R}^d)$, let $\mathcal{F}_\xi(E)$ be the σ-field generated by the random variables $\xi(F)$ for all Borel sets F contained in E. For any $E_1,\ E_2 \in \mathcal{B}(\mathbb{R}^d)$, define

$$\alpha(E_1,\ E_2) = \sup_{A_1 \in \mathcal{F}_\xi(E_1),\ A_2 \in \mathcal{F}_\xi(E_2)} |\mathbb{P}(A_1A_2) - \mathbb{P}(A_1)\mathbb{P}(A_2)|. \tag{4.8}$$

4.3.1 Weak Law of Large Numbers

Let $a = (a_1,\ \cdots,\ a_d)' \in \mathbb{R}^d$ and $\Pi(a) = \{x \in \mathbb{R}^d : 0 < x_i \le a_i,\ i = 1,\ \cdots,\ d\}$. Let $\mathbb{Z}$ denote the set of integers and $\mathbb{Z}^d$ denote the Cartesian product $\Pi_{i=1}^d \mathbb{Z}$ of $\mathbb{Z}$. The translate of $\Pi(a)$ by the integral vector $ma = (m_1a_1,\ \cdots,\ m_da_d)$, where $m \in \mathbb{Z}^d$, is denoted by $\Pi_m = T_{ma}\Pi(a)$. The family of sets $\Pi_m,\ m \in \mathbb{Z}^d$, forms a partition of $\mathbb{R}^d$. For a subset $A \subset \mathbb{R}^d$, define $N_a^+(A)$ as the number of sets Π_m for which $A \cap \Pi_m \ne \emptyset$ and $N_a^-(A)$ as the number of Π_m such that $\Pi_m \subset A$.

Assumption B1 *There exist $C_1,\ C_2 < \infty$ such that, for $B \in \mathcal{B}(\mathbb{R}^d)$, if $\ell(B) < C_1$, then $\mathbb{E}|\xi(B)| \le C_2$, otherwise $\mathbb{E}|\xi(B)| \le C_2\ell(B)$.*

Assumption B2 *There is $a \in \mathbb{R}^d$ such that (i) the family $\{\xi(\Pi_m) : m \in \mathbb{Z}^d\}$ of random variables is uniformly integrable; (ii) as $n \to \infty$, $N_a^-(A_n) \to \infty$, $\ell(D_n)/\ell(A_n) \to 1$, where*

$D_n = \cup \{\Pi_m : \Pi_m \subset A_n\}$; *and (iii) letting* $\alpha(r) = \sup \alpha(\Pi_i,\ \Pi_j)$, *where the supremum is taken over all rectangles such that* $D(\Pi_i,\ \Pi_j) \geq r$, $\lim_{r\to\infty} \alpha(r) = 0$.

Theorem B1 *Under Assumptions A1 and A2, if* $\mathbb{E}\{\xi(\Pi_m)\} = 0$ *for all* $m \in \mathbb{Z}^d$, *then as* $n \to \infty$, $\ell(A_n)^{-1} \xi(A_n) \to_1 0$.

Consider a second-order stationary point process N. Assumption A1 holds for N since $\mathbb{E}N(A) = \mu\ell(A)$ for any bounded Borel set A, where μ is the mean density of N. Assumption A1 also holds for random signed measures having fixed atomic points, i.e. $\xi(x) \neq 0$ a.s. for some $x \in \mathbb{R}^d$. From the proof of Theorem B1, it can be seen that the notion of strong mixing is not necessary to show the weak law of large numbers. Assumptions on weak correlation among Π_m could have been imposed without affecting the conclusion.

4.3.2 Central Limit Theorem

In this chapter, we simply employ the central limit theorem proved in Bulinskii and Zhurbenko (1976) that can also be found in Zhurbenko (1986). Their proof of the central limit theorem is based on the Bernstein technique and the following definition of strong mixing. Define

$$\alpha^*(r,\ k) = \sup \alpha(E_1,\ E_2), \tag{4.9}$$

where the supremum is taken over all Borel sets E_1 and E_2 such that $D(E_1,\ E_2) \geq r$ and $\delta(E_i) \leq k,\ i = 1,\ 2$. Assumption B2 and the one in Bulinskii and Zhurbenko (1976) that is similar to our Assumption B2 can be difficult to verify. For example Assumption B2 requires that there exists $a \in \mathbb{R}^d$ such that $\{\xi(\Pi_m)\}$ is uniformly integrable and as n is sufficiently large the sampling regions A_n are essentially the union of those $\Pi_m \subset A_n$. In practice, practitioners may not know the exact value of a and hence may fail to verify that Assumption B2 holds. To avoid this difficulty, Bulinskii and Zhurbenkon (1976) considered a sequence of bounded Borel sets that converges to infinity in the sense of van Hove.

4.3.3 van Hove Convergence

Definition B1 *A sequence of sets* $\{A_n\}$ *converges to infinity in the sense of van Hove if for each* a,

$$\lim_{n\to\infty} N_a^-(A_n) = \infty \quad \textit{and} \quad \lim_{n\to\infty} N_a^-(A_n)/N_a^+(A_n) = 1,$$

where $N_a^+(A)$ *and* $N_a^-(A)$ *are defined as in the previous sub-section.*

Clearly if $\{A_n\}$ converges to infinity in the sense of van Hove, the problem mentioned at the end of the previous subsection is solved. Let $A_n = (s_{1n},\ t_{1n}] \times \cdots \times (s_{dn},\ t_{dn}]$ be a rectangle in $\mathbb{R}^d$. A sequence of rectangles $\{A_n\}$ such that $\lim_{n\to\infty} (t_{in} - s_{in}) = \infty,\ i = 1,\ \cdots,\ d$, converges to infinity in the sense of van Hove. However for a sequence of arbitrary shapes in $\mathbb{R}^d$, it may be much harder to verify if it converges to infinity in the sense of van Hove. Theorem B2 suggests how to check whether van Hove convergence holds.

For a set $A \in \mathcal{B}\left(\mathbb{R}^d\right)$, let ∂A be the boundary of A.

Theorem B2 *For any Π_m, if $A \cap \Pi_m \neq \emptyset$ and Π_m is not contained in A, then $\partial A \cap \Pi_m \neq \emptyset$.*

In most applications, the interest may be upon sampling regions that are subsets of $\mathbb{R}^2$. In this case, ∂A_n is the boundary of A_n and it is possible to find its length. To avoid making complicated assumption, we first focus on arbitrary shapes in $\mathbb{R}^d$ having no holes. The advantage of this restriction is that ∂A_n will be connected. To avoid confusion, we introduce some definitions that may not be often employed in econometrics. See, for example, Wilson (2008).

Definition B2 *A curve (or path) in a metric space $(X,\ d)$ is a continuous function $\gamma : [a,\ b] \to X$, for some real closed interval $[a,\ b]$.*

Definition B3 *For a curve $\gamma : [a,\ b] \to X$ on a metric space $(X,\ d)$, the length of γ, length(γ), is defined as*

$$\text{length}(\gamma) = \sup_P\ l_P,$$

where $l_P = \sum_{i=1}^m d\left(\gamma(t_i),\ \gamma(t_{i-1})\right)$ and $P = \{t_0,\ t_1,\ \cdots,\ t_m\}$ is a partition of the interval $[a,\ b]$.

Suppose that ∂A is the image of a closed curve γ, a continuous function $\gamma : [a,\ b] \to \mathbb{R}^2$ such that $\gamma(a) = \gamma(b)$. Then the perimeter of ∂A can be defined as length(γ).

Theorem B3 *Given a sequence of bounded Borel subsets $\{A_n\}$ of $\mathbb{R}^2$, suppose that for all n, ∂A_n are the image of closed curves γ_n, and as $n \to \infty$, $\ell(A_n) \to \infty$, length$(\gamma_n) \to \infty$, and length$(\gamma_n) = o\left(\ell(A_n)\right)$. Then $\{A_n\}$ converges to infinity in the sense of van Hove.*

Now if $\{A_n\}$ is a sequence of bounded Borel subsets of $\mathbb{R}^2$ such that there is $N_1 < \infty$ such that for all $n \in \mathbb{N}$, A_n have at most N_1 holes, and each hold has finite Lebesgue measure uniformly over n, then $\{A_n\}$ satisfying the condition in Theorem B3 is still a van Hove sequence. The condition in Theorem B3 may be relaxed further without affecting the result, but the proof may have to be on a case-by-case basis. It seems possible to extend this result to higher dimensions but the proof will be more complicated. It is worth noting that Lemma 5 is employed in the proof of Theorem B3 and it says that for a given curve with finite length, we can divide the curve into many sections with any required length. However, for example, in $\mathbb{R}^3$ ∂A is the surface of A. It is unclear how to divide the surface into smaller subsets of A with any desired area. Some further conditions may be needed.

4.4 Least Squares Estimation

Suppose that for a given marked point process $\{(s_i, (x_i, y_i))\}$ on $\mathbb{R}^d \times \mathbb{R}^{p+1}$ such that the marks x_i and y_i exhibits a linear relationship

$$y_i = \beta_0' x_i + \varepsilon_i,$$

where the p-dimensional column vector β_0 is unknown, the prime denotes transposition, and the ε_i are unobserved. Suppose we only observe a realization of the marked point process when the locations are in a bounded Borel set A. The least squares estimate (LSE) of β_0 constructed from the data is

$$\hat{\beta} = \left(\sum_{s_i \in A} x_i x_i' \right)^{-1} \sum_{s_i \in A} x_i y_i.$$

Let $z_i = x_i \varepsilon_i$ for all $i \in \mathbb{N}$. Define $\zeta(A) = \sum_{s_i \in A} z_i$ for all bounded Borel sets A.

Assumption C1 *Assumption A3 holds for the marked point process $\{(s_i, (x_i, y_i))\}$. There exists a measurable function σ_x such that $\sigma_x(s_i) = \mathbb{E}(x_i x_i' | N_g)$, where the modulus of all elements of $\sigma_x(s)$ are uniformly bounded for $s \in \mathbb{R}^d$. For any sequence $\{A_n\}$ of bounded Borel sets such that there exists a sequence of balls contained in A_n with radii r_n such that, $\lim_{n\to\infty} r_n = \infty$, $\ell(A_n)^{-1} \int_{A_n} \sigma_x(s)\ ds \to \Phi$ where Φ is positive definite. In addition, Assumption A4 holds for z_i where $\gamma(s_j - s_i)$ denotes $\mathbb{E}\left(z_i z_j' \middle| N_g \right)$, $\gamma(0)$ has bounded elements, and (4.7) holds and $\Sigma = \int_{\mathbb{R}^d} \gamma(u)\ \breve{M}_2(du)$ is positive definite.*

Assumption C2 *There exists $l^0 = \left(l_1^0, ..., l_d^0\right)' \in \mathbb{R}^d$ such that for some $\delta_1 > 0$ and some constant c, $\mathbb{E}\|\zeta(\Pi)\|^{2+\delta_1} \leq c$ for all rectangles Π whose j-th edges have length l_j^0, $j = 1, \cdots, d$. Let $\xi(A) = \sum_{s_i \in A} x_i x_i'$. There exists $l^{00} = \left(l_1^{00}, \cdots, l_d^{00}\right)' \in \mathbb{R}^d$ such that for some $\delta_2 > 0$ and some constant c, $\mathbb{E}|\xi_{rs}(\Pi_m)|^{1+\delta_2} \leq c$ for all rectangles $\Pi_m = T_{ml^{00}} \Pi\left(l^{00}\right)$, where $\xi_{rs}(\Pi)$ is the (r, s)-th element of $\xi(\Pi)$.*

Assumption C3 *The sequence of bounded Borel subsets* $\{A_n\}$ *is increasing and converges to infinity in the sense of van Hove. Moreover,* $\ell(A_n) \geq c[\delta(A_n)]^a$ *for some* $0 \leq a \leq d$ *and some constant* c*, where* $\delta(A_n)$ *is the diameter of* A_n*.*

Let $\alpha_1^*(r,\ k)$ be defined as in (4.8) and (4.9), where the the σ-field of interest is generated by the random (signed) measure $\zeta(\cdot)$ and $\alpha_2^*(r,\ k)$ be defined for the σ-field generated by the random (signed) measure $\xi(\cdot)$.

Assumption C4 *For some number* c *and* k_0,

$$\alpha_1^*(r,k) \leq c\frac{k^w}{r^{d+\varepsilon}} \text{ for } k > k_0, \text{ where } \varepsilon > \frac{2d}{\delta_1} \text{ and } w < \frac{a\delta_1}{2(1+\delta_1)}\left(\frac{\varepsilon}{d} - \frac{2}{\delta_1}\right)$$

and for $k = \delta(\Pi(a))$,

$$\lim_{r\to\infty} \alpha_2^*(r,\ k) = 0.$$

Theorem C1 *Under Assumptions C1-C4,*

$$[\ell(A_n)]^{1/2}\left(\hat{\beta} - \beta_0\right) \xrightarrow{d} N\left(0, \Phi^{-1}\Sigma\Phi^{-1}\right).$$

Assumption C2 requires weak dependence of N_g and of $\{z_i\}$ given the ground process N_g. Let $\breve{C}_2$ be the reduced covariance (signed) measure of N_g, i.e. $\breve{C}_2(du) = \breve{M}_2(du) - m^2\ell(du)$. Sufficient conditions for (4.7) are $\int_{\mathbb{R}^d} |\gamma_z(u)|\, du < \infty$ and $\int_{\mathbb{R}^d} |\gamma_z(u)| \left|\breve{C}_2\right|(du)$. Sufficient conditions for Assumption A3 are that for some finite constant c, (i) $\mathbb{E}(|z_i z_j z_k||\, N_g) < c$ for all $i,\ j,\ k \in \mathbb{N}$; and (ii) $\mathbb{E}[N_g(\Pi)]^3 < c$ uniformly for all Π described in Assumption A3. Condition (ii) holds if the ground process is third-order stationary.

In this chapter we allow dependence between the marks and the ground process. To cover the popular model in the statistical literature, we briefly outline how to show asymptotic normality of its LSE of $\widehat{\beta}$. If the marks and the ground process, assumed to be independent, arises from a random sampling of a second-order stationary random field, then the function γ_z may be regarded as the unconditional covariance function of the random field $\left\{z(s),\ s \in \mathbb{R}^d\right\}$. More direct assumptions on weak dependence of the sampling process $N_g(\cdot)$ and the random field can be given.

Suppose that the strong mixing assumption on dependence of the random field $\{z(s)\}$ and maximal correlation mixing assumption on the sampling process are given. Then some useful result on covariances of random variables generated by the sampling process and random field can be obtained. Let $A_1,\ A_2 \in \mathcal{B}\left(\mathbb{R}^d\right)$ such that $\delta(A_1),\ \delta(A_2) \leq k$ and

$D(A_1, A_2) \geq r$. Let V_i be a complex random variable measurable with respect to $\mathcal{F}_N(A_i)$, where $\mathcal{F}_N(A_i)$ is the σ-field as defined earlier, N is the marked point process $\{(s_i,\ z_i)\}$, and $\mathbb{E}|V_i|^{2+\delta} < \infty$. Following Politis and Sherman (2001),

$$Cov(V_1, V_2) = \mathbb{E}\left[Cov^{N_g}(V_1, V_2)\right] + Cov\left[\mathbb{E}(V_1|N_g), \mathbb{E}(V_2|N_g)\right],$$

where Cov^{N_g} denotes covariance conditional on the ground process N_g. Due to independence of N_g and $\{z(s)\}$, by Theorem 17.2.2 in Ibragimov and Linnik (1971), $Cov^{N_g}(V_1, V_2)$ can be bounded by a multiple of the strong mixing coefficient of the random field $\{z(s)\}$. Then this inequality remains valid after taking the expectation. Since $\mathbb{E}(V_i|\, N_g)$ is measurable with respect to the σ-field generated by $\{N_g(E) : E \subset A_i\}$, using the Jensen's inequality, $Cov\left[\mathbb{E}(V_1|N_g), \mathbb{E}(V_2|N_g)\right]$ can be bounded in absolute value by a multiple of the maximal correlation mixing coefficient. Then the proof based on the Bernstein's technique can rely upon this covariance bound.

It can also be shown that as $n \to \infty$, $N_g(A_n)/\ell(A_n) \xrightarrow{p} \mu$, where μ is the mean intensity of the ground process. Then it follows that $[N_g(A_n)]^{-1/2}\left(\widehat{\beta} - \beta_0\right) = O_p(1)$. One may wish to compare this result with the standard root-n consistency but it should be noted that the number of observations, $N_g(A_n)$, is now a random variable. Finally, as noted in the last paragraph of Section 2, Σ may be regarded as the spectral density function of the $\mathbb{R}^p$-valued random signed measure ζ at zero frequency. Hence, the asymptotic covariance matrix of $\widehat{\beta}$ is analogous to that of stationary time series.

4.5 Spectral Density Estimation

Motivated by the asymptotic covariance of the least square estimate of a linear regression, we mainly focus on estimating the spectral density function of ζ as defined in the previous section. We hope that our results may give some hint on how to consistently estimate spectral density of other wide-sense second-order stationary random signed measures. However further regularity conditions may be needed. If we consider a random signed measure W, then the variance of the spectral density estimate may not be defined since an integral with respect to W is only defined up to a mean square sense.

From now on, we will consider only ζ as defined in Theorem A3 but z_i are now $\mathbb{R}^p$-valued. We also assume that z_i are fully observed, if $s_i \in A$, to avoid further complexity from approximating z_i by the residuals from the least square estimate. It can be seen from the proof of Theorem 1 that a periodogram is an asymptotically unbiased estimate of spectral density. Let $d_{A_n}(\lambda) = \int_{A_n} e^{i\langle\lambda, s\rangle}\, \zeta(ds)$ be the finite Fourier transform and define the periodogram as

$$I_{A_n}(\lambda) = \frac{1}{(2\pi)^d\, \ell(A_n)} d_{A_n}(\lambda)\, \overline{d_{A_n}(\lambda)}'.$$

When the state space is $\mathbb{R}$, Brillinger (1972) showed that, as in stationary time series, the variance of a periodogram does vanish to zero as the length of the time span where data is observed goes to infinity.

Rather than employing an averaged periodogram, we instead restrict our attention to the following estimate similar to the one employed in Masry (1978)

$$\widehat{f}_\zeta(\lambda) = \frac{1}{(2\pi)^d \ell(B_n)} \sum_{s_j \in B_n} \sum_{s_k \in B_n} e^{-i\langle \lambda, s_k - s_j\rangle} w_n(s_k - s_j) z_j z_k',$$

where B_n is a subset of A_n. The reason for using B_n rather than A_n is due to the "edge effect" to be discussed later. The main reason for considering this estimate is that it is commonly employed in economic applications.

This estimate can be written as

$$\widehat{f}_\zeta(\lambda) = \frac{1}{(2\pi)^d \ell(B_n)} \int_{B_n \times \mathbb{R}} \int_{B_n \times \mathbb{R}} e^{-i\langle \lambda, t-s\rangle} w_n(t-s) z_1 z_2' \ N(ds \times dz_1) \ N(dt \times dz_2).$$

Certainly we need that w must be measurable. A product measure of N can be defined so that we can proceed as earlier to show that this integral is indeed a well-defined matrix of random variables.

Also we restrict our discussion on w_n such that

$$w_n(u) = \prod_{j=1}^d k(u_j / m_{jn}),$$

where the conditions on k are to be given later. We also assume that $\lim_{n\to\infty} m_{jn} \to \infty$, $j = 1, \cdots, d$.

For $\theta \in \mathbb{R}^d$, define

$$W_n(\theta) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{i\langle \theta, u\rangle} w_n(u) \ du.$$

If k is continuous and integrable, then

$$w_n(u) = \int_{\mathbb{R}^d} e^{-i\langle \theta, u\rangle} W_n(\theta) \ d\theta$$

Now fix n and hence B. Then

$$\begin{aligned} & \int_{\mathbb{R}^d} I_B(\theta) W_n(\theta - \lambda) \ d\theta \\ = \ & (2\pi)^{-d} \ell(B)^{-1} \sum_{s_j \in B} \sum_{s_k \in B} z_j z_k' e^{-i\langle \lambda, s_k - s_j\rangle} \left(\int_{\mathbb{R}^d} e^{-i\langle \theta - \lambda, s_k - s_j\rangle} W_n(\theta - \lambda) \ d\theta \right) \\ = \ & (2\pi)^{-d} \ell(B)^{-1} \sum_{s_j \in B} \sum_{s_k \in B} z_j z_k' e^{-i\langle \lambda, s_k - s_j\rangle} w_n(s_k - s_j) = \widehat{f}_\zeta(\lambda). \end{aligned}$$

Therefore positive semidefiniteness depends on the choice of W_n employed. If k is the modified Bartlett kernel, i.e.

$$k(u) = (1 - |u|) 1(|u| \leq 1),$$

then

$$W_n(\theta) = (2\pi)^{-d} \prod_{j=1}^d 2 \left(\frac{1 - \cos(\theta_j m_{jn})}{m_{jn} \theta_j^2} \right) \geq 0.$$

If k is the Parzen kernel, i.e.

$$\begin{aligned} k(u) &= 1 - 6u^2 + 6|u|^3, \quad |u| \le 1/2, \\ &= 2(1-|u|)^3, \quad 1/2 < |u| \le 1, \\ &= 0, \quad |u| \ge 1, \end{aligned}$$

then

$$W_n(\theta) = \prod_{j=1}^{d} \left[\frac{192}{m_{jn}^3 \theta_j^4} \{\sin(\theta_j m_{jn}/4)\}^4 \right] \ge 0.$$

For higher-order kernels, the weight functions $W_n(\theta)$ may not be nonnegative.

4.5.1 Bias

Similar to the notation in the previous section, let $\gamma_{rs}(u)$ denote the (r,s)-th element of the matrix $\gamma(u)$.

Assumption D1 *$k : \mathbb{R} \to \mathbb{R}$ is an even, Lebesgue integrable function such that $k(0) = 1$, $|k(u)| \le 1$, and for some $q > 0$,*

$$\lim_{u \to 0} \left\{ \frac{1 - k(u)}{|u|^q} \right\} = k_q,$$

where k_q is finite and strictly positive.

Assumption D2 *For $r, s = 1, \cdots, p$, $\sum_{j=1}^{d} \int_{\mathbb{R}^d} |u_j|^{\max\{q,1\}} |\gamma_{rs}(u)| \; \breve{M}_2(du) < \infty$.*

Assumption D3 *For each $n \in \mathbb{N}$, A_n has a subset B_n containing $R_n = \Pi_{i=1}^{d}[a_{in}, b_{in}]$, where $\lim_{n\to\infty}(b_{in} - a_{in}) = \infty$ for $i = 1, \cdots, d$. Moreover, $\{B_n\}$ is such that for some constant C,*

$$\frac{\ell(B_n \backslash T_{-u} B_n)}{\ell(B_n)} \le C \sum_{i=1}^{d} \frac{|u_i|}{b_{in} - a_{in}}.$$

Assumption D4 *For $j = 1, \cdots, d$, as $n \to \infty$, $m_{jn} \to \infty$*

Theorem D1 *Under Assumptions D1 - D4, as $n \to \infty$,*

$$(2\pi)^d \left\{ f_\zeta(\lambda) - \mathbb{E}\widehat{f}_\zeta(\lambda) \right\} = \alpha_{1n} + \alpha_{2n} + o\left(\sum_{j=1}^{d} \left\{ m_{jn}^{-q} + (b_{jn} - a_{jn})^{-1} \right\} \right),$$

where

$$\begin{aligned} \alpha_{1n} &= k_q \sum_{j=1}^{d} m_{jn}^{-q} \int_{\mathbb{R}^d} |u_j|^q \gamma(u) \ \breve{M}_2(du), \\ \alpha_{2n} &= O\left(\sum_{j=1}^{d} (b_{jn} - a_{jn})^{-1} \right). \end{aligned}$$

Robinson (2007) called α_{2n} the "edge effect" term. In the proof of Theorem D1, it can be seen that this term arises from

$$\int_{B_n'} \left\{ \frac{\ell(B_n \backslash T_{-u} B_n)}{\ell(B_n)} \right\} w_n(u) e^{-i\langle \lambda, u \rangle} \gamma(u) \ \breve{M}_2(du).$$

If $\{B_n\}$ is any van Hove sequence, then Lemma 4 only implies that $\ell(B_n)^{-1} \ell(B_n \backslash T_{-u} B_n)$ converges to zero. It is unclear how to determine the rate at which $\ell(B_n)^{-1} \ell(B_n \backslash T_{-u} B_n)$ converges to zero. As a result, we suggest using only a subset B_n of A_n exhibiting the property in Assumption D3 so that it becomes easier to determine a more precise bias from the edge effect term.

4.5.2 Variance

To avoid making complicated, despite being relatively weak, assumptions, we simply assume that the ground point process is 4-th order stationary. For any sets A_i, $i = 1, \cdots, k$, we denote its Cartesian product by $\Pi_{i=1}^k A_i$. Now for a point process N on the state space $\mathcal{S}$, define $N^k\left(\Pi_{i=1}^k A_i\right) = \Pi_{i=1}^k N(A_i)$, where A_i are Borel sets in $\mathcal{S}$. It follows that N^k can be extended to be a point process on $\mathcal{S}^k$. Define $M_k(A) = \mathbb{E}\left\{N^k(A)\right\}$. Again M_k can be extended to be a measure. If $M_k(A) < \infty$, for any bounded Borel set A in $\mathcal{S}^k$, then we say that M_k is the k-th moment measure of N. If N is a point process on $\mathbb{R}^d$ such that its k-th moment measure exists, and for each $j = 1, \cdots, k$, bounded Borel subsets $A_1, \cdots, A_j$ of $\mathbb{R}^d$, $u \in \mathbb{R}^d$,

$$M_j(T_u A_1 \times \cdots \times T_u A_j) = M_j(A_1 \times \cdots \times A_j),$$

then we say that the point process N is k-th order stationary. This is the generalization of the definition of second-order stationary in Section 2. The definition of higher-order cumulant measures can be generalized in a similar fashion. The main technical advantage of the k-th order stationarity is that there exist the reduced k-th order moment measure,

$\breve{M}_k$, and the reduced k-th order cumulant measure, $\breve{C}_k$, such that for any bounded function f of bounded support

$$\begin{aligned} & \int_{\mathcal{S}^k} f\left(s_1, \cdots, s_k\right) \; M_k\left(ds_1 \times \cdots \times ds_k\right) \\ = & \int_{\mathcal{S}^k} f\left(s, s+u_1, \cdots, s+u_{k-1}\right) \; ds \; \breve{M}_k\left(du_1 \times \cdots \times du_{k-1}\right), \end{aligned}$$

and similarly for $\breve{C}_k$ (see Proposition 12.6.III in Daley and Vere Jones (2008)).

Assumption D5 *The ground process is 4-th order stationary with the mean density* μ *and such that* $\int_{\mathbb{R}^{(j-1)d}} \left|\breve{C}_j\right| \left(\Pi_{i=1}^{j-1} ds_i\right) < \infty, \; j = 2, \; 3, \; 4.$

Suppose that $\kappa_{N_g}\left(z_{ir}, z_{js}, z_{kr}, z_{ls}\right)$ is the conditional fourth cumulant of $z_{ir}, \; z_{js}, \; z_{kr}$ and z_{ls} given the point process N_g.

Assumption D6 *For any* $i, \; j, \; k, \; l \in \mathbb{N}$, *and* $r, \; s = 1, \; \cdots, p$, *there exist functions* κ_{rs} *such that* $\kappa_{N_g}\left(z_{ir}, z_{js}, z_{kr}, z_{ls}\right) = \kappa_{rs}\left(s_i, s_j, s_k, s_l\right)$, *where* κ_{rs} *is bounded on any bound subset of* $\mathbb{R}^{4d}$. *In addition for all* $r, \; s = 1, \; \cdots, \; p$,

$$\sup_{x \in \mathbb{R}^d} \int_{\mathbb{R}^{3d}} \left|\kappa_{rs}\left(x, x+u_1, x+u_2, x+u_3\right)\right| \breve{M}_4\left(du_1 \times du_2 \times du_3\right) < \infty.$$

Assumption D7 *For* $r, \; s = 1, \; \cdots, \; p, \; \gamma_{rr}(0) < \infty$ *and* $\int_{\mathbb{R}^d} \left|\gamma_{rs}(u)\right| \; du < \infty$, *where* $\gamma_{rs}\left(s_j - s_i\right) = \mathbb{E}\left\{\left. z_{ri} z_{sj} \right| N_g\right\}$.

Theorem D2 *Suppose* A_n *contains* R_n *defined in Assumption D3. Under Assumptions D4-D7, for* $r, \; s = 1, \; \cdots, \; p$, *as* $n \to \infty$

$$\frac{\ell\left(A_n\right)}{m_n} (2\pi)^{2d} var\left(\widehat{f}_{rs}(\lambda)\right) \to \int_{\mathbb{R}^d} w(u)^2 \, du \left\{f_{rr}(\lambda) f_{ss}(\lambda) + \mathbf{1}(\lambda = 0) f_{rs}(\lambda)^2\right\}, \quad (4.10)$$

where $m_n = \Pi_{j=1}^d m_{jn}, \; w(u) = \Pi_{i=1}^d k\left(u_i\right).$

As discussed in Robinson (2007), without taking into account the contribution from the edge effect to the MSE of $\widehat{f}_{rs}(\lambda)$, this modified MSE can be minimized by choosing $m_{jn}^* = c_j \ell\left(B_n\right)^{1/(d+2q)}$ for some positive constants c_j. If, as $n \to \infty$, $\alpha_{2n} = o\left(\sum_{j=1}^d \left(m_{jn}^*\right)^{-q}\right)$, then the edge effect is dominated by the standard bias term. Therefore the optimal choice of

m_{jn} that minimizes the MSE of $\widehat{f}_{rs}(\lambda)$ is the same as m_{jn}^*. In this case, the MSE vanishes at the rate $\ell(B_n)^{-2q/(d+2q)}$. The usual curse of dimensionality is present in this optimal rate. If there is a rectangle Q_n such that it is a subset of A_n and $\ell(Q_n) \geq C\ell(A_n)$ for some positive constant C, then one can choose $B_n = Q_n$ so that the rate at which the MSE vanishes to zero will not be affected.

4.6 Positive Definite Estimate

In Section 4 we see that the asymptotic covariance matrix of the least squares estimate of the unknown slope parameter of a linear regression is a function of the spectral density at zero frequency of the random signed measure ζ defined by $\zeta(A) = \sum_{s_i \in A} x_i \varepsilon_i$. The results in Section 5 suggest how we can obtain a consistent estimate of the matrix of interest. Under weak dependence of the marked point process, higher-order kernels can be employed to reduce the bias and hence to achieve a better rate of convergence in the mean square sense. As can be seen also from Section 5, the curse of dimensionality affecting the rate of convergence makes higher-order kernels relatively more attractive. However, higher-order kernels will generally give an estimate that is not positive semidefinite. This can cause a problem if such estimate, obtained from higher-order kernels, are employed to construct Wald statistics.

Another application where the trade-off between positive semidefiniteness of an estimate and its rate of convergence is prominent is the GMM estimation with dependent data. A similar estimation problem arises naturally when one has to estimate the optimal weighting matrix. There the problem is more serious. An estimate of the optimal weighting matrix that is not positive semidefinite can make the nonlinear optimization more complicated. Some discussion can be found in Newey and West (1987).

As a result, we will discuss an algorithm which can be employed to convert an estimate that is not positive semidefinite into a positive definite one. This section is somewhat independent from the previous sections since the framework we consider is general enough to cover other estimation problems that are affected by the trade-off described above.

Before continuing the discussion some matrix notations are introduced. As in other chapters, a square matrix A if said to be positive definite if it is symmetric and $x'Ax > 0$ for any $x \neq 0$. For a matrix A, denote its (i, j)-th element by a_{ij}. For a square matrix A denote the largest and smallest eigenvalues of A by $\overline{\lambda}(A)$ and $\underline{\lambda}(A)$, respectively. For any $p \times p$ symmetric matrix A, let A_i be the leading principal submatrix of A determined by the first i rows and columns. From its definition associated with the quadratic form, it can be easily shown that a square matrix A of order p is positive definite if and only if A_i are positive definite for all $i = 1, ..., p$.

Algorithm

Suppose that we are given a $p \times p$ matrix B. For simplicity of notation, we assume that B is symmetric. Otherwise, we can take a transformation $(B + B')/2$ so that the transformed matrix is symmetric. Now let c and u be positive numbers. We can modify a symmetric matrix B to obtain a new matrix $T(B;c,u)=E$ that is positive definite using the following algorithm.

(i) If $b_{11}\geq c$, then set $e_{11}=b_{11}$. Otherwise, set $e_{11}=c$.

(ii) For any $i=2,...,p$, determine the leading principal submatrix E_{i-1} of E. Let e_{i-1} be an $(i-1)$-vector $e_{i-1}=(e_{1i},\ \cdots,\ e_{i-1,i})'$. For each $j=1,\ \cdots,\ i-1$, if $|b_{ji}|\leq u$, set $e_{ji}=b_{ji}$. Otherwise, set $e_{ji}=-u$ if $b_{ji}<0$, and set $e_{ji}=u$ if $b_{ji}>0$, where b_{ji} is the (j,i)-th element of B.

(iii) Set

$$E_i=\begin{pmatrix}E_{i-1} & e_{i-1}\\ e'_{i-1} & e_{ii}\end{pmatrix}.$$

If $b_{ii}\geq e'_{i-1}E_{i-1}^{-1}e_{i-1}+c$, then set $e_{ii}=b_{ii}$. Otherwise, set $e_{ii}=e'_{i-1}E_{i-1}^{-1}e_{i-1}+c$.

Unlike proofs of other theorems, the proof of Theorem 6.1 is presented here as it may help justify our algorithm.

Theorem E1 *For any $p\times p$ matrix B, the matrix $E=T(B;c,u)$ is positive definite.*

Proof. For simplicity of notation, we assume that B is already symmetric. The proof is based on the necessary and sufficient condition on a positive definite matrix given above and mathematical induction. Let $E=T(B;c,u)$. It is clear that the first leading principal submatrix E_1 is positive definite. Now suppose that for some $i=1,...,p-1$, E_i is positive definite. For any non-zero vector $x\in\mathbb{R}^{i+1}$, partition x such that $x=(x'_1,x_2)'$ where $x_1\in\mathbb{R}^i$ and x_2 is a real number. Then

$$\begin{aligned}x'E_{i+1}x &= x'_1E_ix_1+x_2e'_ix_1+x_2x'_1e_i+e_{i+1,i+1}x_2^2\\ &= \left(x_1+E_i^{-1}e_ix_2\right)'E_i\left(x_1+E_i^{-1}e_ix_2\right)+x_2^2\left(e_{i+1,i+1}-e'_iE_i^{-1}e_i\right). \quad (4.11)\end{aligned}$$

If $x_2=0$, then $x_1\neq 0$ and the first term in (4.11) is positive. If $x_2\neq 0$, x_1 can be chosen so that $x_1=-E_i^{-1}e_ix_2$. In this case, the necessary and sufficient condition for E_{i+1} to be positive definite is that $e_{i+1,i+1}>e'_iE_i^{-1}e_i$. By our definition of $e_{i+1,i+1}$, this is indeed the case. Hence, the required result holds. Note again that E_{i+1} is positive definite if and only if $e_{i+1,i+1}>e'_iE_i^{-1}e_i$. ■

This algorithm does two jobs. First, steps (i) and (iii) ensure that $e_{11}\geq c$ and $e_{i+1,i+1}\geq e'_{i+1}E_i^{-1}e_{i+1}+c,\ i=2,\ \cdots,\ p$, so that the matrix $E=T(B;c,u)$ is positive definite. Second, step (ii) sets an upper bound u for the absolute values of the off-diagonal elements of B. In practice, we can set u to be so large that none of the off-diagonal elements of B will be affected. From the proof of Theorem 6.1, a reason each leading principal submatrix $B_i,\ i=1,\ \cdots,\ p$, of B, is not positive definite is that the last inequality of the proof of Theorem 6.1 does not hold. Recall that positive definiteness of a matrix depends (necessarily and sufficiently) on positive definiteness of its leading principal submatrices. As a result, the algorithm proposed here seems to make a minimum alteration of the original matrix to make it positive definite. One may argue that rather than requiring $b_{ii} \geq e_i' E_{i-1}^{-1} e_i + c$, we simply need $b_{ii} > e_i' E_{i-1}^{-1} e_i$ without affecting positive definiteness of $T(B; c, u)$. This is indeed correct but there can be some undesirable consequences. First, if $b_{ii} - e_i' E_{i-1}^{-1} e_i$ is very small then the matrix E_i is near singularity. This can cause some serious computational problems. Second, the near singularity of E_i can lead to another theoretical consequence that some of its diagonal elements that is greater than $e_i' E_i^{-1} e_i$ may not have finite second moment. The latter reason also explains why we impose an upper bound u on the off-diagonal elements. Rather than choosing fixed values of u and c, it is possible to employ sequences $\{u_n\}$ and $\{c_n\}$ where as $n \to \infty$, $u_n \to \infty$ and $c_n \to 0$ without a significant impact on the rate of convergence of $T(B; c, u)$. Taking c_n and u_n into account, it is possible to determine a lower bound for the smallest eigenvalue of $T(B; c, u)$.

Theorem E2 *For given values of c and u, for $i = 1, \ldots, p$, let E_i be the leading principal submatrix of $E = T(B; c, u)$. Let $a_1 = c$ and $a_i = c\left(1 + (i-1) u^2 a_{i-1}^{-2}\right)^{-1}$, $i = 2, \cdots, p$. If $c \leq 1$ and $u \geq 1$, then*

$$\underline{\lambda}(E_i) \geq a_i, \quad i = 1, \cdots, p.$$

Theorem 6.2 indicates another advantage of our algorithm. The choices of u and c allow us to control the condition number of the matrix $T(B;\ c,\ u)$. If the actual interest is on the inverse of $T(B;\ c,\ u)$ rather $T(B;\ c,\ u)$ itself, then, from a computational point of view, c and u can be chosen to avoid the "ill conditioned" problem. Now consider an impact of our algorithm on the rate of convergence, in the mean square sense, of an original matrix.

Theorem E3 *Suppose that $\widehat{\Omega}$ is an estimate of an unknown matrix Ω_0 whose elements have the mean square error (MSE) of order $O(f_n)$. Let $\{c_n\}$ and $\{u_n\}$ be sequences of positive numbers employed in the algorithm mentioned above. Define $\tilde{\Omega} = T\left(\widehat{\Omega}; c_n, u_n\right)$ and let ω_{0ij}, $\widetilde{\omega}_{ij}$ be the (i,j)-th element of Ω_0 and $\tilde{\Omega}$ respectively. Suppose that as $n \to \infty$, $c_n = o(1)$ and $u_n \to \infty$. Then as $n \to \infty$, for $i,\ j = 1, \cdots, p$,*

$$\mathbb{E}\left(\widetilde{\omega}_{ij} - \omega_{0ij}\right)^2 = O\left(u_n^2 f_n\right), \ i \neq j,$$

and

$$\begin{aligned} \mathbb{E}\left(\widetilde{\omega}_{ii} - \omega_{0ii}\right)^2 &= O(f_n), \quad i = 1, \\ &= O(a_{in} f_n), \ i \geq 2, \end{aligned}$$

where $a_{2n} = c_n^{-2} u_n^6$ and $a_{in} = c_n^{-2^i+2} u_n^{2^i} a_{i-1n}$, $i > 2$.

Now go back to the trade-off problem. Suppose that there are two estimates $\widehat{\Omega}_1$ and $\widehat{\Omega}_2$ whose elements have the MSE of order $n^{-\delta_1}$ and $n^{-\delta_2}$ respectively, where $0 < \delta_1 < \delta_2$. It is often the case that the unknown matrix Ω_0 is positive definite, $\widehat{\Omega}_1$ is positive semidefinite but $\widehat{\Omega}_2$ is not positive semidefinite. If it is desirable for an estimate to be at least positive semidefinite, practitioners may normally choose $\widehat{\Omega}_1$ over $\widehat{\Omega}_2$ despite a faster rate of convergence of $\widehat{\Omega}_2$. Now if we set $c_n = O\left((\log n)^{-1}\right)$ and $u_n = O(\log n)$, then for any $\varepsilon > 0$, the MSE of elements of $T\left(\widehat{\Omega}_2; c_n, u_n\right)$ is $o\left(n^{-(\delta_2-\varepsilon)}\right)$. Therefore $T\left(\widehat{\Omega}_2; c_n, u_n\right)$ is both positive definite and converges faster than $\widehat{\Omega}_1$.

4.7 Final Comments

In this chapter, we propose an interpretation that irregularly-spaced cross-sectional data can naturally be regarded as a realization of a marked point process. We also show that linear functionals of a marked point process can be employed to construct a random (signed) measure appearing in many econometric applications including the least squares estimate of unknown slope parameters in a linear regression model. Under reasonably weak assumptions, including the presence of spatial dependence among observations, such a random (signed) measure is wide-sense second-order stationary and thus has a spectral measure analogous to a spectral measure of a second-order stationary time series. Based on mixing assumptions, we develop asymptotic properties of a random (signed) measure which can be applied in econometric applications. We show the asymptotic normality of the least squares estimate of regression coefficients and find that its asymptotic variance matrix is the spectral density at zero frequency of the associated random (signed) measure.

Even though the Toeplitz structure of the variance matrix of spatial observations is lost when the locations are irregularly spaced, the finding in this chapter shows that there is a close connection between an analysis of regularly spaced time series and of irregularly spaced spatial data. The finding suggests that many known spectral analysis of time series should be applicable to spatial data. For example, it should be possible to perform a nonparametric test for zero spatial correlation of the observations(the marks) by considering the shape of the spectral density of the associated random (signed) measure. An analysis of cross spectra my be useful in investigating interdependence or linkages among various cross-sectional variables. Moreover, our finding suggests that spatio-temporal dependence can be modelled in a unified framework via the use of spectral analysis. However, a success of an attempt to extend known results in spectral analysis of time series to the spatial context would depend on a success in dealing with the spectral measure of the spatial case that is not totally finite.

Asymptotic properties of a random (signed) measure, which are developed based on mixing assumptions, should be directly applicable to nonlinear estimation with spatial data. It should not be difficult to extend our results to GMM estimation. Concerning estimation of the spectral density of a random (signed) measure, a subset B_n of a sampling region A_n and its required properties are introduced to avoid bias from the edge effect. One possible approach to avoid employing a subset B_n and retain efficiency by employing A_n is to employ tapering in spectral density estimation as shown in Robinson (2007).

In the last section, rather independent from the other sections of the chapter, we discuss estimation of an unknown positive definite matrix. We propose an algorithm which can be employed to transform an estimate of the unknown matrix into a positive definite estimate. Despite an arbitrary decrease in the rate of mean square convergence of the estimate, this algorithm opens an opportunity for higher-order kernels to become more useful in many econometric applications such as estimation of asymptotic variance matrices or optimal weighting matrices in GMM estimation. Simulation results showing finite-sample properties of applications of this algorithm on estimates with higher-order kernels compared with standard estimates with modified Bartlett or Parzen kernels should be conducted. An improvement of precision of an estimate based on this algorithm will generate a challenging problem for both theorists and practitioners. For example, it is known that the weaker the time or spatial dependence is, the faster the rate of convergence of a spectral density estimate could be when higher-order kernels are employed. It is therefore crucial to get some information concerning the degree of time or spatial dependence so that an appropriate choice of kernel can be chosen. A data-dependent procedure which can reflect the degree of time or spatial dependence will be crucial to future development of spectral density or asymptotic variance estimation.

Appendix 4.1: Proof of Theorems

For the rest of this paper, for any subset $A,\ B$ of X, let $B \backslash A = \{x \in X : x \in B \text{ and } x \notin A\}$.

Proof of Theorem A1 By Lemma 2, f_ξ is a non-negative and continuous function. Define $G_\xi(A) = \int_A f_\xi(\lambda)\ d\lambda$ for any Borel set A. Then G_ξ is a measure that is absolutely continuous. For any $\psi \in \mathcal{S}(\mathbb{R}^d)$, $\psi(x) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i\langle \lambda, x\rangle} \widetilde{\psi}(\lambda)\ d\lambda$, by Fubini's theorem,

$$\begin{aligned}\int_{\mathbb{R}^d} \psi(x)\ \breve{C}_2(dx) &= \int_{\mathbb{R}^d} \widetilde{\psi}(\lambda) \left\{(2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i\langle \lambda, x\rangle} \breve{C}_2(dx)\right\} d\lambda \\ &= \int_{\mathbb{R}^d} \widetilde{\psi}(\lambda) f_\xi(\lambda)\ d\lambda \\ &= \int_{\mathbb{R}^d} \widetilde{\psi}(\lambda)\ G_\xi(d\lambda).\end{aligned}$$

Since the Parseval identity (4.4) holds for every ψ in the Schwartz space, $G_\xi = F_\xi$.

Proof of Theorem A2 By Assumption A2,for any bounded $A,\ B \in \mathcal{B}(\mathbb{R}^d)$

$$\begin{aligned}Cov(W(A),\ W(B)) &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{1}_B(s)\, \mathbf{1}_A(s+u)\ \breve{C}_2(du)\ ds \\ &= \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \mathbf{1}_B(s)\, \mathbf{1}_A(s+u)\ ds\right) \breve{C}_2(du). \qquad (4.12)\end{aligned}$$

By Lemma 3, the integral in the brackets is Lebesgue integrable and continuous in u. Moreover $\int_{\mathbb{R}^d} e^{i\langle\lambda,u\rangle}\left(\int_{\mathbb{R}^d} \mathbf{1}_B(s)\,\mathbf{1}_A(s+u)\ ds\right)\ du = \widetilde{\mathbf{1}}_A(\lambda)\,\overline{\widetilde{\mathbf{1}}}_B(\lambda)$. Since the Lebesgue measure is translation bounded, $\widetilde{\mathbf{1}}_A(\lambda)\,\overline{\widetilde{\mathbf{1}}}_B(\lambda)$ is Lebesgue integrable (see exercise 8.6.8 in Daley and Vere-Jones (2003) and employ Schwarz's inequality). Similar to the proof of Theorem A1, the Fourier inversion theorem of continuous and Lesbesgue integrable functions can be employed to show that the right of (4.12) is

$$(2\pi)^{-d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} e^{-i\langle\lambda,u\rangle}\widetilde{\mathbf{1}}_A(\lambda)\,\overline{\widetilde{\mathbf{1}}}_B(\lambda)\ d\lambda\ \breve{C}_2(du) = \int_{\mathbb{R}^d}\widetilde{\mathbf{1}}_A(\lambda)\,\overline{\widetilde{\mathbf{1}}}_B(\lambda)\,f_W(\lambda)\ d\lambda. \tag{4.13}$$

Proceeding as in the proof of Lemma 2, f_W is a non-negative continuous function. It remains to show that the measure defined by $F_W(A) = \int_A f_W(\lambda)\ d\lambda$ for any Borel set A, is in $\mathcal{C}$. First for any bounded Borel set A,

$$\begin{aligned}\int_{\mathbb{R}^d}\left|\widetilde{\mathbf{1}}_A(\lambda)\right|^2\ F_W(d\lambda) &= \int_{\mathbb{R}^d}\left|\widetilde{\mathbf{1}}_A(\lambda)\right|^2\ f_W(\lambda)\ d\lambda \\ &\le \left\{\sup_{\lambda\in\mathbb{R}^d}\ f_W(\lambda)\right\}\int_{\mathbb{R}^d}\left|\widetilde{\mathbf{1}}_A(\lambda)\right|^2\ d\lambda < \infty\end{aligned}$$

since, by (4.5), the function f_W is uniformly bounded and $\int_{\mathbb{R}^d}\left|\widetilde{\mathbf{1}}_A(\lambda)\right|^2\ d\lambda < \infty$ as the Lebesgue measure is translation bounded. It follows from (4.12) and (4.13) that $\mathbb{E}\{W(A)\}^2 = \int_{\mathbb{R}^d}\left|\widetilde{\mathbf{1}}_A(\lambda)\right|^2 f_W(\lambda)\ d\lambda$ for any bounded Borel set A. Consider a decreasing sequence $\{A_n\}$ of bounded Borel sets such that $\lim_{n\to\infty}\ell(A_n) = 0$. As the process W is wide-sense second-order stationary,

$$\lim_{n\to\infty}\int_{\mathbb{R}^d}\left|\widetilde{\mathbf{1}}_{A_n}(\lambda)\right|^2 f_W(\lambda)\ d\lambda = \lim_{n\to\infty}\mathbb{E}\left\{W(A_n)^2\right\} = 0.$$

Hence $F_W \in \mathcal{C}$.

Proof of Theorem A3 Throughout the proof of this theorem, sets A, B and $\{A_n\}$ denote bounded Borel sets and a sequence of bounded Borel sets in $\mathbb{R}^d$. Using iterated expectation, Assumption A4 implies that $\mathbb{E}\{\zeta(A)\} = 0$ for all A. For any A and B, by Assumptions A3 and A4,

$$\begin{aligned}Cov\{\zeta(A),\zeta(B)\} &= \mathbb{E}\left[\mathbb{E}\{\zeta(A)\,\zeta(B)|\,N_g\}\right] \\ &= \mathbb{E}\left\{\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} 1_A(s)\,1_B(t)\,\gamma_z(t-s)\ \ N_g(ds)\,N_g(dt)\right\} \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} 1_A(s)\,1_B(s+u)\,\gamma_z(u)\ \ ds\ \breve{M}_2(du) \end{aligned}\tag{4.14}$$

By Schwarz's inequality, $|\gamma_z(u)| \le \gamma_z(0)$ for all $u \in \mathbb{R}^d$. As A is bounded, $Var\{\zeta(A)\} \le |\gamma_z(0)|\,M_2(A\times A) < \infty$ where M_2 is the second moment measure of N_g. Define $B-A$ as in (4.27), then

$$Cov\{\zeta(A),\zeta(B)\} = \int_{B-A}\ell(A\cap T_{-u}B)\,\gamma_z(u)\ \ \breve{M}_2(du) \tag{4.15}$$

For any $v \in \mathbb{R}^d$, it it follows that

$$\begin{aligned} Cov\left\{\zeta\left(T_v A\right), \zeta\left(T_v B\right)\right\} &= \int_{T_v B - T_v A} \ell\left(T_v A \cap T_{-u} T_v B\right) \gamma_z(u) \ \breve{M}_2(du) \\ &= \int_{B-A} \ell\left\{T_v\left(A \cap T_{-u} B\right)\right\} \gamma_z(u) \ \breve{M}_2(du), \end{aligned}$$

as $T_v B - T_v A = B - A$. Since the Lebesgue measure is translation invariant,

$$Cov\left\{\zeta\left(T_v A\right), \zeta\left(T_v B\right)\right\} = Cov\left\{\zeta(A), \zeta(B)\right\},$$

for all $A,\ B \in \mathcal{B}\left(\mathbb{R}^d\right)$ and $v \in \mathbb{R}^d$. By definition of ζ, for disjoint A and B, $\zeta(A \cup B) = \zeta(A) + \zeta(B)$ a.s. and hence ζ is finitely additive in the mean square sense too. Define A' as in (4.27). Now consider a decreasing sequence such that $\lim_{n\to\infty} \ell(A_n) = 0$. It follows from (4.15) that for all $n \in \mathbb{N}$,

$$\begin{aligned} \mathbb{E}\left\{\zeta\left(A_n\right)^2\right\} &\leq \left|\gamma_z(0)\right| \int_{\mathbb{R}^d} 1_{A_n'}(u)\, \ell\left(A_n \cap T_{-u} A_n\right) \ \breve{M}_2(du) \\ &\leq \left|\gamma_z(0)\right| \int_{\mathbb{R}^d} 1_{A_1'}(u)\, \ell\left(A_1 \cap T_{-u} A_1\right) \ \breve{M}_2(du) \\ &= \left|\gamma_z(0)\right| M_2\left(A_1 \times A_1\right) < \infty. \end{aligned}$$

As $\ell\left(A_n \cap T_{-u} A_n\right) \leq \ell(A_n) \to 0$ as $n \to \infty$, it follows from dominated convergence that $\lim_{n\to\infty} \mathbb{E}\left\{\zeta\left(A_n\right)^2\right\} = 0$. Hence ζ is *wide-sense second-order stationary.*

Let $\gamma_z^+ = \max\{\gamma_z,\ 0\}$ and $\gamma_z^- = -\min\{\gamma_z,\ 0\}$. Define $\nu^+(A) = \int_A \gamma_z^+(u) \ \breve{M}_2(du)$ and $\nu^-(A) = \int_A \gamma_z^-(u) \ \breve{M}_2(du)$ for any Borel set A. By (4.6), ν^+ and ν^- are finite measures that are absolutely continuous with respect to the measure $\breve{M}_2$ where γ_z^+ and γ_z^- are their Radon-Nikodym derivatives. Then $\nu = \nu^+ - \nu^-$ is a signed measure. For bounded Borel sets A and B, using Fubini's theorem,

$$\begin{aligned} Cov\left(\int_{\mathbb{R}^d} \mathbf{1}_A(s) \ \zeta(ds), \int_{\mathbb{R}^d} \mathbf{1}_B(s) \ \zeta(ds)\right) &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{1}_A(s)\, \mathbf{1}_B(s+u)\, \gamma_z(u) \ ds \ \breve{M}_2(du) \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mathbf{1}_A(s)\, \mathbf{1}_B(s+u) \ ds \ \nu(du). \end{aligned}$$

Since ν^+ and ν^- are finite measures, $\int_{\mathbb{R}^d} |\nu|(du) < \infty$, where $|\nu| = \nu^+ + \nu^-$. Hence, by Theorem A2, the spectral measure of ζ is absolutely continuous with the continuous spectral density

$$f_\zeta(\lambda) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i\langle\lambda,u\rangle} \ \nu(du) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{-i\langle\lambda,u\rangle} \gamma_z(u) \ \breve{M}_2(du).$$

Proof of Theorem B1 For any finite set B, let $|B|$ be its cardinality. Consider $a \in \mathbb{R}^d$ satisfying Assumption B2. Assumption B2 (ii) implies that $\lim_{n\to\infty} \ell\left(A_n \backslash D_n\right)/\ell\left(A_n\right) = 0$. Since

$$\ell\left(A_n\right)^{-1}\left\{\xi\left(A_n\right) - \xi\left(D_n\right)\right\} = \ell\left(A_n\right)^{-1} \xi\left(A_n \backslash D_n\right),$$

Assumption B1 implies that $\ell(A_n)^{-1}\xi(A_n)-\ell(A_n)^{-1}\xi(D_n)\to_1 0$. Let $B_n=\{m:\Pi_m\subset A_n\}$. Then $\xi(D_n)=\sum_{m\in B_n}\xi(\Pi_m)$. The remaining part of the proof is just a slight modification of a standard proof of weak law of large numbers. For any $\varepsilon>0$,

$$\ell(A_n)^{-1}\xi(D_n)=\ell(A_n)^{-1}\sum_{m\in B_n}\{\xi'(\Pi_m)-\mathbb{E}\xi'(\Pi_m)\}+\ell(A_n)^{-1}\sum_{m\in B_n}\{\xi''(\Pi_m)-\mathbb{E}\xi''(\Pi_m)\},\tag{4.16}$$

where $\xi'(\Pi_m)=\xi(\Pi_m)1(|\xi(\Pi_m)|\le\delta)$, $\xi''(\Pi_m)=\xi(\Pi_m)1(|\xi(\Pi_m)|>\delta)$ and, by Assumption B2 (i), δ is chosen so that $\sup_{m\in\mathbb{Z}^d}\mathbb{E}|\xi''(\Pi_m)|<\varepsilon\ell(\Pi(a))/2$.

Because $\xi'(\Pi_m)$ is $\mathcal{F}_\xi(\Pi_m)$-measurable, following the proof of Theorem 17.2.1 in Ibragimov and Linnik (1971), it can be shown that, for $m\neq m'$,

$$|Cov\{\xi'(\Pi_m),\ \xi'(\Pi_{m'})\}|\le 4\delta^2\alpha(\Pi_m,\ \Pi_{m'}).$$

Since $\alpha(\Pi_m,\ \Pi_{m'})\le\alpha(r)$ if $D(\Pi_m,\ \Pi_{m'})\ge r$, Assumption B2 (iii) implies that there is $r_0<\infty$ such that $\alpha(\Pi_m,\ \Pi_{m'})<\varepsilon/\left(8\ell(\Pi(a))^{-2}\delta^2\right)$ for all $m,\ m'$ such that $D(\Pi_m,\ \Pi_{m'})\ge r_0$. The second moment of the first sum in (4.16) is bounded by

$$\begin{aligned}&\ell(A_n)^{-2}\sum\sum_{m,m'\in B_n}|Cov\{\xi'(\Pi_m),\xi'(\Pi_{m'})\}|\\ \le\ &\ell(\Pi(a))^{-2}|B_n|^{-2}\left[\sum_{E_{1n}}4\delta^2\alpha(\Pi_m,\ \Pi_{m'})+\sum_{E_{2n}}|Cov\{\xi'(\Pi_m),\xi'(\Pi_{m'})\}|\right],\end{aligned}\tag{4.17}$$

where

$$\begin{aligned}E_{1n}&=\{(m,\ m'):m,\ m'\in B_n,\ D(\Pi_m,\ \Pi_{m'})\ge r_0\},\\ E_{2n}&=\{(m,\ m'):m,\ m'\in B_n,\ D(\Pi_m,\ \Pi_{m'})<r_0\}.\end{aligned}$$

The contribution from the first term in (4.17) is bounded by $\varepsilon/2$. Moreover, there is $C_{r_0}<\infty$ such that $|E_{2n}|<C_{r_0}|B_n|$. Hence the second term in (4.17) is bounded by $\ell(\Pi(a))^{-2}|B_n|^{-1}\delta^2C_{r_0}$. Assumptions B1 (i) and B2 (ii) implies that there is $N<\infty$ such that for all $n>N$, the second term in (4.17) is less than $\varepsilon/2$. Thus the first sum in (4.16) converges to zero in the second mean.

For the second sum in (4.16),

$$\begin{aligned}\mathbb{E}\left|\ell(A_n)^{-1}\sum_{m\in B_n}\{\xi''(\Pi_m)-\mathbb{E}\xi''(\Pi_m)\}\right|&\le\ell(A_n)^{-1}\left\{2\sum_{m\in B_n}\mathbb{E}|\xi''(\Pi_m)|\right\}\\ &\le\ell(A_n)^{-1}\left\{2|B_n|\sup_{m\in\mathbb{Z}^d}\mathbb{E}|\xi''(\Pi_m)|\right\}\\ &\le 2\ell(\Pi(a))^{-1}\sup_{m\in\mathbb{Z}^d}\mathbb{E}|\xi''(\Pi_m)|<\varepsilon.\end{aligned}$$

Hence the second sum in (4.16) converges to zero in the first mean.

Proof of Theorem B2 It follows that there is $a,\ b \in \Pi_m$ such that $a \in A$, and $b \in A^c$, where A^c denote the complement of A. Consider $c(t) = ta + (1-t)b$, where $t \in [0,\ 1]$. Let $C = \{c(t) : 0 \le t \le 1\}$. By definition of Π_m, it follows that $C \subset \Pi_m,\ C \cap A \ne \emptyset$ and $C \cap A^c \ne \emptyset$. For a subset A of a metric space X, let $int(A)$ be the interior of A. If either $a \notin int(A)$ or $b \notin int(A^c)$, then either $a \in \partial A$ or $b \in \partial A$. Hence, the required result holds. It remains to consider the case when $a \in int(A)$ and $b \in int(B)$.

Let $D = \{s \in [0,1] : c(t) \in int(A) \text{ for all } t \le s\}$. Clearly $c(0) = a \in int(A)$ so that $D \ne \emptyset$. In addition there is $\varepsilon_1 > 0$ such that $[0,\ \varepsilon_1] \subset D$. Similarly $c(1) = b \in int(A^c)$ so that 1 is an upper bound for D and there is $\varepsilon_2 > 0$ such that $c(t) \in int(A^c)$ for $1-\varepsilon_2 \le t \le 1$. Then $\sup D$ exists and let $\delta = \sup D$. It follows that $0 < \delta < 1$. Clearly $c(\delta) \notin int(A)$, otherwise there is a contradiction. Similarly $c(\delta) \notin int(A^c)$. Hence $c(\delta) \in C \subset \Pi_m$ and $c(\delta) \in \partial A$.

Proof of Theorem B3 Consider any $a = (a_1,\ a_2)' \in \mathbb{R}^2$ such that $a_i \ne 0,\ i = 1,\ 2$. Let $a_0 = \min\{a_1,\ a_2\}$. Let $B_{1n} = \{m : \Pi_m \subset int(A_n)\}$, $B_{2n} = \{m : \Pi_m \subset A_n\}$, $B_{3n} = \{m : \Pi_m \cap A_n \ne \emptyset\}$, and $B_{4n} = \left\{m : \Pi_m \cap \overline{A}_n \ne \emptyset\right\}$, where $int(A_n)$ and $\overline{A}_n$ are the interior and closure of A_n respectively. It follows that $B_{1n} \subset B_{2n} \subset B_{3n} \subset B_{4n}$. Hence $|B_{1n}| \le |B_{2n}| = N_a^-(A_n)$ and $N_a^+(A_n) = |B_{3n}| \le |B_{4n}|$. Then

$$N_a^+(A_n) - N_a^-(A_n) \le |B_{4n} \backslash B_{1n}|. \tag{4.18}$$

Since $B_{4n} \backslash B_{1n} = \left\{m : x \in \overline{A}_n \text{ and } y \notin int(A_n) \text{ for some } x,\ y \in \Pi_m\right\}$, $B_{4n} \backslash B_{1n} \subset B_{5n}$ where $B_{5n} = \{m : \Pi_m \cap \partial A_n \ne \emptyset\}$ by Assumption B2.

By Lemma 5, the function L, defined there, is continuous. For sufficiently large n, by the intermediate value theorem, there exists t such that $L(t) = a_0/2$. Take t_1 to be the supremum of such t. Similarly we can find t_r such that $L(t_r) = ra_0/2 \le \text{length}(\gamma_n)$. As L is nondecreasing, $t_r \le t_{r+1}$. If $2\text{length}(\gamma_n)/a_0 > \lfloor 2\text{length}(\gamma_n)/a_0 \rfloor$, where $\lfloor \cdot \rfloor$ is the floor function, set $R_n = \lfloor 2\text{length}(\gamma_n)/a_0 \rfloor + 1$. Then $\partial A_n = \cup_{r=1}^{R_n} \{\gamma([t_{r-1},\ t_r])\}$, where $\gamma([t_{r-1},\ t_r])$ is the image of γ over $[t_{r-1},\ t_r]$ with length at most $a_0/2$.

Note that in the Euclidean space, any curve joining two endpoints with the minimum length is a straight line segment. It can be shown, by considering every possible cases, that each $\gamma([t_{r-1},\ t_r])$ can be contained within 4 adjacent $\Pi_m s$, whose union is a $2a_1 \times 2a_2$ rectangle. Let B_{6n} be the union, over $r = 1,\ 2,\ \cdots,\ R_n$, of all such m. Then $B_{5n} \subset B_{6n}$ and $|B_{6n}| \le 4R_n$. Let $D_{1n} = \cup\{\Pi_m : m \in B_{1n}\}$, $D_{4n} = \cup\{\Pi_m : m \in B_{4n}\}$, and $D_{6n} = \cup\{\Pi_m : m \in B_{6n}\}$. Since $B_{4n} \backslash B_{1n} \subset B_{6n}$, $\ell(A_n) \le \ell(D_{4n}) \le \ell(D_{6n}) + \ell(D_{1n})$. Hence $\ell(A_n) \ge \ell(D_{1n}) \ge \ell(A_n) - 4R_n a_1 a_2$. Since as $n \to \infty$, $\text{length}(\gamma_n) = o(\ell(A_n))$, $\ell(D_{1n})/\ell(A_n) \to 1$. Hence $\lim_{n\to\infty} N_a^-(A_n) = \infty$. Since $N_a^+(A_n) \ge \ell(D_{1n})/(a_1 a_2)$ and $B_{4n} \backslash B_{1n} \subset B_{6n}$, (4.18) implies that $\lim_{n\to\infty} \{N^+(A_n) - N_a^-(A_n)\}/N_a^+ = 0$. Therefore $\{A_n\}$ converges to infinity in the sense of van Hove.

Proof of Theorem C Note that

$$[\ell(A_n)]^{1/2}\left(\hat{\beta}-\beta_0\right)=\left(\ell(A_n)^{-1}\sum_{s_i\in A_n}x_ix_i'\right)^{-1}\left(\ell(A_n)^{-1/2}\sum_{s_i\in A_n}z_i\right).$$

For $r=1,\ \cdots,\ p$, let $\sigma_r=\mathbb{E}\left(x_{ir}^2\middle|N_g\right)$. For any bounded Borel set A, by Assumption C1 and Schwarz's inequality,

$$\begin{aligned}\mathbb{E}\left|\xi_{rs}(A)\right| &\leq \mathbb{E}\left(\sum_{s_i\in A}\left|x_{ir}x_{is}\right|\right)\leq\mu\int_A\sigma_r(t)^{1/2}\sigma_s(t)^{1/2}\ dt\\ &\leq \mu\left(\int_A\sigma_r(t)\ dt\right)^{1/2}\left(\int_A\sigma_s(t)\ dt\right)^{1/2}\leq C\mu\ell(A),\end{aligned}$$

where μ is the mean density of N_g. Hence Assumption B1 holds for each element of ξ.

For l^{00} in Assumption C2, take $\Pi_m=T_{ma}\Pi(a)$ where $a=l^{00}$. Assumption C2 implies that the family $\{\xi_{rs}(\Pi_m)\}$ is uniformly integrable. Since $\alpha(r)\leq\alpha^*(r,k)$, where $k=\delta(\Pi(a))$, Assumption C4 implies that $\lim_{r\to\infty}\alpha(r)=0$. Hence these results and Assumption C3 implies that Assumption B2 holds. Thus Theorem B1, Assumption C1 and Lemma 4 imply that

$$\ell(A_n)^{-1}\xi(A_n)\to_p\Phi. \tag{4.19}$$

For any $\lambda\in\mathbb{R}^p$ such that $\|\lambda\|=1$, define

$$\zeta_\lambda(A_n)=\lambda'\zeta(A_n)=\lambda'\sum_{s_i\in A_n}z_i.$$

Assumption C1 implies that $\mathbb{E}\zeta_\lambda(A_n)=0$ for all $n\in\mathbb{N}$. Let $A_n^0=\cup\{\Pi_m'':\Pi_m''\subset A_n\}$, where $\Pi_m''=T_{ma'}\Pi(a'')$ and $a''=l^0$ in Assumption C2. Since $\{A_n\}$ converges to infinity in the sense of van Hove, $\lim_{n\to\infty}\left[\ell\left(A_n^0\right)/\ell(A_n)\right]=1$. By Lemma 6, as $n\to\infty$,

$$\frac{1}{\ell(A_n)}\mathbb{E}\left[\zeta_\lambda(A_n)-\zeta_\lambda\left(A_n^0\right)\right]^2=\frac{Var\left\{\zeta_\lambda\left(A_n\backslash A_n^0\right)\right\}}{\ell(A_n)}\to 0. \tag{4.20}$$

Hence

$$\ell(A_n)^{-1/2}\left\{\zeta_\lambda(A_n)-\zeta_\lambda\left(A_n^0\right)\right\}\to_p 0. \tag{4.21}$$

Since Lemma 6 and Assumptions C2-C4 imply that all conditions in Bulinskii and Zhurbenko (1976) are satisfied, it follows that

$$\zeta_\lambda\left(A_n^0\right)/Var\left(\zeta_\lambda\left(A_n^0\right)\right)^{1/2}\to_d N(0,1). \tag{4.22}$$

Now Lemma 6 and (4.20) and imply that

$$Var\left\{\zeta_\lambda\left(A_n^0\right)\right\}/\ell(A_n)\to\lambda'\Sigma\lambda. \tag{4.23}$$

Then (4.22) and (4.23) imply that

$$\ell(A_n)^{-1/2}\zeta_\lambda\left(A_n^0\right)\to_d N\left(0,\lambda'\Sigma\lambda\right).$$

Hence this with (4.19) and (4.21) conclude the proof.

Proof of Theorem D1 Let $f_{rs}(\lambda)$ be the (r,s)-th element of $f_\zeta(\lambda)$ and similarly for $\widehat{f}_\zeta(\lambda)$.

$$\begin{aligned}
& (2\pi)^d \left[f_{rs}(\lambda) - \mathbb{E}\left\{ \widehat{f}_{rs}(\lambda) \right\} \right] \\
= & \int_{\mathbb{R}^d} e^{-i\langle \lambda, u\rangle} \gamma_{rs}(u) \ \breve{M}_2(du) - \int_{\mathbb{R}^d} \ell(B_n)^{-1} \ell(B_n \cap T_{-u} B_n) e^{-i\langle\lambda,u\rangle} w_n(u) \gamma_{rs}(u) \ \breve{M}_2(du) \\
= & \int_{B_n'} \{1 - w_n(u)\} e^{-i\langle\lambda,u\rangle} \gamma_{rs}(u) \ \breve{M}_2(du) + \int_{(B_n')^c} e^{-i\langle\lambda,u\rangle} \gamma_{rs}(u) \ \breve{M}_2(du) \qquad (4.24) \\
& + \int_{B_n'} \left\{ 1 - \frac{\ell(B_n \cap T_{-u} B_n)}{\ell(B_n)} \right\} w_n(u) e^{-i\langle\lambda,u\rangle} \gamma_{rs}(u) \ \breve{M}_2(du).
\end{aligned}$$

Recall that $R_n' = \Pi_{i=1}^d [a_{in} - b_{in}, \ b_{in} - a_{in}]$. Note that for any $u \in \mathbb{R}^d$, $\mathbf{1}_{B_n'}(u) \geq \mathbf{1}_{R_n'}(u)$ and $\lim_{n\to\infty} \mathbf{1}_{R_n'}(u) = 1$. For the first term in (4.24), it can be proceeded similar to Robinson (2007), with the summation sign replaced by the integral sign, to show that the first term is the (r,s)-th element of $\alpha_{1n} + o(\alpha_{1n})$.

The modulus of the second term in (4.24) is bounded by

$$\begin{aligned}
\int_{(R_n')^c} |\gamma_{rs}(u)| \ \breve{M}_2(du) \quad & \leq \quad \sum_{j=1}^d (b_{jn} - a_{jn})^{-1} \int_{(R_n')^c} |u_j| \, |\gamma_{rs}(u)| \ \breve{M}_2(du) \\
& = \quad o\left(\sum_{j=1}^d (b_{jn} - a_{jn})^{-1} \right).
\end{aligned}$$

The modulus of last term in (4.24) is bounded by

$$C \int_{B_n'} \sum_{j=1}^d \frac{|u_j|}{b_{jn} - a_{jn}} |\gamma_{rs}(u)| \ \breve{M}_2(du) = O\left(\sum_{j=1}^d (b_{jn} - a_{jn})^{-1} \right).$$

Proof of Theorem D2 Let

$$\begin{aligned}
\eta_{rs}(s_1, s_2, s_3, s_4) &= \kappa_{rs}(s_1, s_2, s_3, s_4) + \gamma_{rr}(s_3 - s_1) \gamma_{ss}(s_4 - s_2) + \gamma_{rs}(s_4 - s_1) \gamma_{sr}(s_3 - s_2), \\
\phi(s_1, s_2, s_3, s_4) &= e^{-i\langle \lambda, s_2 - s_1 - s_4 + s_3\rangle} w_n(s_2 - s_1) w_n(s_4 - s_3).
\end{aligned}$$

The left side of (4.10) is

$$\frac{1}{\ell(A_n)\,m_n}\int_{B_n^4}\phi(s_1,s_2,s_3,s_4)\,\eta_{rs}(s_1,s_2,s_3,s_4)\,M_4\left(\Pi_{i=1}^d ds_i\right) \tag{4.25}$$
$$+\frac{1}{\ell(A_n)\,m_n}\left\{\int_{B_n^4}\phi(s_1,s_2,s_3,s_4)\,\gamma_{rs}(s_2-s_1)\,\gamma_{rs}(s_4-s_3)\,M_4\left(\Pi_{i=1}^d ds_i\right)-\right.$$
$$\left.\int_{B_n^4}\phi(s_1,s_2,s_3,s_4)\,\gamma_{rs}(s_2-s_1)\,\gamma_{rs}(s_4-s_3)\,M_2(ds_1\times ds_2)\,M_2(ds_3\times ds_4)\right\}.$$

Lemma 7 implies that the contribution from κ_{rs} to the first integral in (4.25) is $O(\ell(B_n))$.

In differential form, it follows that

$$\begin{aligned}
& M_4(dx_1\times\cdots\times dx_4)\\
= \; & C_4(dx_1\times\cdots\times dx_4)+M_1(dx_1)\,C_3(dx_2\times dx_3\times dx_4)\\
& +M_1(dx_2)\,C_3(dx_1\times dx_3\times dx_4)+M_1(dx_3)\,C_3(dx_1\times dx_2\times dx_4)\\
& +M_1(dx_4)\,C_3(dx_1\times dx_2\times dx_3)+C_2(dx_1\times dx_2)\,C_2(dx_3\times dx_4)\\
& +C_2(dx_1\times dx_3)\,C_2(dx_2\times dx_4)+C_2(dx_1\times dx_4)\,C_2(dx_2\times dx_3)\\
& +M_1(dx_1)\,M_1(dx_2)\,C_2(dx_3\times dx_4)+M_1(dx_1)\,M_1(dx_3)\,C_2(dx_2\times dx_4)\\
& +M_1(dx_1)\,M_1(dx_4)\,C_2(dx_2\times dx_3)+M_1(dx_2)\,M_1(dx_3)\,C_2(dx_1\times dx_4)\\
& +M_1(dx_2)\,M_1(dx_4)\,C_2(dx_1\times dx_3)+M_1(dx_3)\,M_1(dx_4)\,C_2(dx_1\times dx_2)\\
& +M_1(dx_1)\,M_1(dx_2)\,M_1(dx_3)\,M_1(dx_4),
\end{aligned}$$

where C_j are the j-th cumulant (signed) measure. Since N_g is also 1-st order moment stationary, $M_1(dx)=\mu\,dx$. The results and proofs from Lemmas 7-10 can be employed to show that most contribution, from the expansion above, to the first term in (4.25) that is associated with $\gamma_{rr}(s_3-s_1)\,\gamma_{ss}(s_4-s_2)$ is $O\left(m_n^{-1}\right)$.

The nontrivial contribution is from

$$\begin{aligned}
& M_2(dx_1\times dx_3)\,M_2(dx_2\times dx_4)\\
= \; & C_2(dx_1\times dx_3)\,C_2(dx_2\times dx_4)+\mu^2 C_2(dx_1\times dx_3)\,dx_2dx_4\\
& +\mu^2 dx_1dx_3C_2(dx_2\times dx_4)+\mu^4 dx_1dx_2dx_3dx_4.
\end{aligned}$$

Therefore Lemma 11 implies that the contribution to the first term in (4.25) from $\gamma_{rr}(s_3-s_1)\,\gamma_{ss}(s_4-s_2)$ is precisely $f_{rr}(\lambda)\,f_{ss}(\lambda)\int_{\mathbb{R}^d}w(u)^2\,du$.

The same reasoning and the standard step employed for time series can be employed to show that if $\lambda\neq 0$, the contribution from $\gamma_{rs}(s_4-s_1)\,\gamma_{rs}(s_3-s_2)$ is 0. However if $\lambda=0$, then the contribution is $f_{rs}(\lambda)^2\int_{\mathbb{R}^d}w(u)^2\,du$. Finally the contribution from the terms in brackets in (4.25) is also $O\left(m_n^{-1}\right)$.

Proof of Theorem E2 For $i=1$, $\underline{\lambda}(E_1)\geq c=a_1$. The remaining part of the proof employs the well-known result that for any symmetric $p\times p$ matrix A, $\overline{\lambda}(A)=$

$\sup_{x\in\mathbb{R}^p:\|x\|=1} x'Ax$ and $\underline{\lambda}(A) = \inf_{x\in\mathbb{R}^p:\|x\|=1} x'Ax$. Now consider $i \geq 2$,

$$E_i = \begin{pmatrix} E_{i-1} & e_{i-1} \\ e'_{i-1} & e_{ii} \end{pmatrix}.$$

For $x = (x_1', x_2) \in \mathbb{R}^i$ where x_2 is a scalar, from the proof of Theorem E1,

$$x'E_ix = \left(x_1 + E_{i-1}^{-1}e_{i-1}x_2\right)' E_{i-1}\left(x_1 + E_{i-1}^{-1}e_{i-1}x_2\right) + x_2^2\left(e_{ii} - e'_{i-1}E_{i-1}^{-1}e_{i-1}\right).$$

Suppose that $x \in \mathbb{R}^i$ such that $\|x\| = 1$. If $x_2 = 0$, then $\|x_1\| = 1$ and $x'E_ix = x_1'E_{i-1}x_1 \geq \underline{\lambda}(E_{i-1})$. Now if $x_2 \neq 0$, then $x'E_ix \geq x_2^2\left(e_{ii} - e'_{i-1}E_{i-1}^{-1}e_{i-1}\right)$ since x_1 can be chosen so that $x_1 = -E_{i-1}^{-1}e_{i-1}x_2$. As $x_1 = -E_{i-1}^{-1}e_{i-1}x_2$, it follows that, in order to have $\|x\| = 1$, $x_2^2 = \left(1 + e'_{i-1}E_{i-1}^{-1}E_{i-1}^{-1}e_{i-1}\right)^{-1}$. Hence

$$\underline{\lambda}(E_i) = \min\left\{\underline{\lambda}(E_{i-1}),\ \left(1 + e'_{i-1}E_{i-1}^{-1}E_{i-1}^{-1}e_{i-1}\right)^{-1}\left(e_{ii} - e'_{i-1}E_{i-1}^{-1}e_{i-1}\right)\right\}.$$

The results we have shown so far are independent of our algorithm. Now, under our algorithm, $e_{ii} - e'_{i-1}E_{i-1}^{-1}e_{i-1} \geq c$. Since E_{i-1} is positive definite,

$$\begin{aligned} e'_{i-1}E_{i-1}^{-1}E_{i-1}^{-1}e_{i-1} &\leq \|e_{i-1}\|^2\,\overline{\lambda}\left(E_{i-1}^{-1}E_{i-1}^{-1}\right) \leq (i-1)\,u^2\overline{\lambda}\left(E_{i-1}^{-1}\right)^2 \\ &= (i-1)\,u^2\underline{\lambda}(E_{i-1})^{-2}. \end{aligned}$$

Hence

$$\left(1 + e'_{i-1}E_{i-1}^{-1}E_{i-1}^{-1}e_{i-1}\right)^{-1}\left(e_{ii} - e'_{i-1}E_{i-1}^{-1}e_{i-1}\right) \geq c\left(1 + (i-1)\,u^2\underline{\lambda}(E_{i-1})^{-2}\right)^{-1}.$$

Since $c \leq 1$ and $u \geq 1$, by simple arithmetic,

$$\underline{\lambda}(E_{i-1}) \geq c\left(1 + (i-1)\,u^2\underline{\lambda}(E_{i-1})^{-2}\right)^{-1}.$$

Hence

$$\underline{\lambda}(E_i) \geq c\left(1 + (i-1)\,u^2\underline{\lambda}(E_{i-1})^{-2}\right)^{-1}.$$

Suppose that for $i \geq 2$, $\underline{\lambda}(E_{i-1}) \geq a_{i-1}$. Then $\underline{\lambda}(E_i) \geq c\left(1 + (i-1)\,u^2a_{i-1}^{-2}\right)^{-1} = a_i$ since the function $f(x) = \left(1 + cx^{-2}\right)^{-1}$, where $c > 0$, is increasing in x when $x > 0$. Hence the required result holds by mathematical induction.

Proof of Theorem E3 Let $\widehat{\omega}_{ij}$ be the (i,j)-th element of $\widehat{\Omega}$. For $i \neq j$

$$\begin{aligned} \widetilde{\omega}_{ij} - \omega_{0ij} &= \left(\widehat{\omega}_{ij} - \omega_{0ij}\right)\mathbf{1}\left(|\widehat{\omega}_{ij}| \leq u_n\right) + \left(u_n - \omega_{0ij}\right)\mathbf{1}\left(\widehat{\omega}_{ij} > u_n\right) \\ &\quad + \left(-u_n - \omega_{0ij}\right)\mathbf{1}\left(\widehat{\omega}_{ij} < -u_n\right). \end{aligned}$$

Clearly

$$\mathbb{E}\left\{\left(\widehat{\omega}_{ij} - \omega_{0ij}\right)\mathbf{1}\left(|\widehat{\omega}_{ij}| \leq u_n\right)\right\}^2 \leq \mathbb{E}\left(\widehat{\omega}_{ij} - \omega_{0ij}\right)^2 = O(f_n).$$

Now

$$\mathbb{E}\left\{(u_n-\omega_{0ij})\,\mathbf{1}\left(\widehat{\omega}_{ij}>u_n\right)\right\}^2=(u_n-\omega_{0ij})^2\,\mathbb{P}\left\{\widehat{\omega}_{ij}>u_n\right\}=O\left(u_n^2 f_n\right),$$

where the last equality follows from the following argument. As $\lim_{n\to\infty}u_n=\infty$, for sufficiently large n,

$$\begin{aligned}\mathbb{P}\left\{|\widehat{\omega}_{ij}|>u_n\right\} &\le \mathbb{P}\left\{|\widehat{\omega}_{ij}|>2\,|\omega_{0ij}|\right\}\le \mathbb{P}\left\{|\widehat{\omega}_{ij}-\omega_{0ij}|>|\omega_{0ij}|\right\}\\ &\le |\omega_{0ij}|^2\,\mathbb{E}\,|\widehat{\omega}_{ij}-\omega_{0ij}|^2=O\left(f_n\right),\end{aligned}$$

by Markov's inequality. The same result can be shown for $(-u_n-\omega_{0ij})\,\mathbf{1}\left(\widehat{\omega}_{ij}<-u_n\right)$. Hence, the first required result holds.

The rest of the proof is based on mathematical induction. Now

$$\widetilde{\omega}_{11}-\omega_{011}=(\widehat{\omega}_{11}-\omega_{011})\,\mathbf{1}\left(\widehat{\omega}_{11}\ge c_n\right)+(c_n-\omega_{011})\,\mathbf{1}\left(\widehat{\omega}_{11}<c_n\right).$$

Clearly the second moment of the first term is $O\left(f_n\right)$. By Lemma 13,

$$\mathbb{E}\left\{(c_n-\omega_{011})\,\mathbf{1}\left(\widehat{\omega}_{11}<c_n\right)\right\}^2=(c_{1n}-\omega_{011})^2\,\mathbb{E}\left\{\mathbf{1}\left(\widehat{\omega}_{11}<c_n\right)\right\}=O\left(f_n\right).$$

Then the required result holds for $i=1$.

For $i\ge 2$, suppose that the mean square error of each of the element of $\widetilde{\Omega}_{i-1}$, the leading principal submatrix of $\widetilde{\Omega}$, is $O\left(a_{i-1,n}f_n\right)$. Now

$$\widetilde{\omega}_{ii}-\omega_{0ii}=(\widehat{\omega}_{ii}-\omega_{0ii})\,\mathbf{1}\left(\widehat{\omega}_{ii}\ge \widetilde{t}_i+c_n\right)+\left(\widetilde{t}_i+c_n-\omega_{0ii}\right)\mathbf{1}\left(\widehat{\omega}_{ii}<\widetilde{t}_i+c_n\right)\tag{4.26}$$

where $\widetilde{t}_i=\widetilde{\omega}_{i-1}'\widetilde{\Omega}_{i-1}^{-1}\widetilde{\omega}_{i-1}$ and $\widetilde{\omega}_{i-1}=\left(\widetilde{\omega}_{1i},\ \cdots,\ \widetilde{\omega}_{i-1,i}\right)'$. Again the second moment of the first term is $O\left(f_n\right)$. Recall that $\widetilde{\Omega}_{i-1}$ is positive definite. If $\widetilde{\omega}_{i-1}\ne 0$,

$$\widetilde{t}_i=\widetilde{\omega}_{i-1}'\widetilde{\Omega}_{i-1}^{-1}\widetilde{\omega}_{i-1}\le \|\widetilde{\omega}_{i-1}\|^2\,\overline{\lambda}\left(\widetilde{\Omega}_{i-1}^{-1}\right)\le (i-1)\,u_n^2\left\{\underline{\lambda}\left(\widetilde{\Omega}_{i-1}\right)\right\}^{-1}.$$

Using the result from Theorem E2, it can be shown by induction that

$$\left\{\underline{\lambda}\left(\widetilde{\Omega}_i\right)\right\}^{-1}=O\left(c_n^{-2^i+1}u_n^{2^i-2}\right).$$

Hence

$$\widetilde{t}_i=O\left(c_n^{-2^{i-1}+1}u_n^{2^{i-1}}\right).$$

Let

$$E_i=\begin{pmatrix}\widetilde{\Omega}_{i-1} & \widetilde{\omega}_{i-1}\\ \widetilde{\omega}_{i-1}' & \widehat{\omega}_{ii}\end{pmatrix}.$$

From the proof of Theorem E2, it follows that

$$\underline{\lambda}\left(E_i\right)=\min\left\{\underline{\lambda}\left(\widetilde{\Omega}_{i-1}\right),\ \left(1+\widetilde{\omega}_{i-1}'\widetilde{\Omega}_{i-1}^{-1}\widetilde{\Omega}_{i-1}^{-1}\widetilde{\omega}_{i-1}\right)^{-1}\left(\widehat{\omega}_{ii}-\widetilde{\omega}_{i-1}'\widetilde{\Omega}_{i-1}^{-1}\widetilde{\omega}_{i-1}\right)\right\}.$$

Therefore $\underline{\lambda}\left(E_i\right)\le\widehat{\omega}_{ii}-\widetilde{t}_i$. Hence $\widehat{\omega}_{ii}<\widetilde{t}_i+c_n$ implies $\underline{\lambda}\left(E_i\right)<c_n$.

For $i=2$, due to the off-diagonal elements, the mean square error of each elements of

E_2 is $O\left(u_n^2 f_n\right)$. Hence, by Lemma 13,

$$\begin{aligned}
\mathbb{E}\left\{\mathbf{1}\left(\widehat{\omega}_{ii} < \widetilde{t}_i + c_n\right)^2\right\} &= \mathbb{P}\left\{\widehat{\omega}_{ii} < \widetilde{t}_i + c_n\right\} \\
&\leq \mathbb{P}\left\{\underline{\lambda}\left(E_i\right) < c_n\right\} \\
&= O\left(u_n^2 f_n\right).
\end{aligned}$$

Therefore the second moment of the second term in (4.26) is $O\left(c_n^{-2} u_n^6 f_n\right)$. Thus $\mathbb{E}\left(\widetilde{\omega}_{ii} - \omega_{0ii}\right)^2 = O\left(c_n^{-2} u_n^6 f_n\right)$.

For $i > 2$, the MSE of each elements of E_i is $O\left(a_{i-1,n} f_n\right)$. Therefore $\mathbb{E}\left\{\mathbf{1}\left(\widehat{\omega}_{ii} < \widetilde{t}_i + c_n\right)^2\right\} = O\left(a_{i-1,n} f_n\right)$ by Lemma 13. Hence

$$\mathbb{E}\left(\widetilde{\omega}_{ii} - \omega_{0ii}\right)^2 = O\left(c_n^{-2^i+2} u_n^{2^i} a_{i-1,n} f_n\right).$$

Appendix 4.2: Technical Lemmas for proofs of Theorems

For any subsets $A,\ B$ of $\mathbb{R}^d$, let

$$B - A = \{b - a : a \in A,\ b \in B\} \text{ and } A' = A - A. \tag{4.27}$$

Also let $\mathbf{1}_A$ be the indicator function such that $\mathbf{1}_A(s) = 1$ if $x \in A$ and 0 otherwise.

Lemma 1 *For any subsets $A,\ B$ of $\mathbb{R}^d$ and any $s,\ u \in \mathbb{R}^d$,*

$$\mathbf{1}_A(s)\,\mathbf{1}_B(s+u) = \mathbf{1}_{A \cap T_{-u} B}(s)\,\mathbf{1}_{B-A}(u),$$

where $T_u A$ is defined as in (4.3)

Proof. For s and u such that $s \in A$ and $s + u \in B$, there exists $t \in B$ such that $s + u = t$. Hence $u = t - s$ where $s \in A,\ t \in B$, i.e. $u \in B - A$. Since $s + u \in B$, there exists $t \in B$ such that $s + u = t$ and thus $s = t - u$, that is $s \in T_{-u}B$. Therefore $s \in A \cap T_{-u}B$. It should be noted that if $u \in B - A$, then $A \cap T_{-u}B$ is non-empty. If $u \in B - A$, i.e. $u = t - s$ for some $s \in A,\ t \in B$, then $s = t - u$ that is $s \in A$ and $s \in T_{-u}B$.

On the other hand, suppose that s and u are such that $s \in A \cap T_{-u}B$ and $u \in B - A$. Since $u \in B - A,\ A \cap T_{-u}B$ is nonempty. Since $s \in A \cap T_{-u}B$, it follows that $s \in A$ and $s + u \in B$. ■

Lemma 2 *Suppose that $\breve{C}_2$ is a reduced covariance measure such that $\int_{\mathbb{R}^d} \left|\breve{C}_2\right|(du) < \infty$. Define $f(\lambda) = \int_{\mathbb{R}^d} e^{-i\langle\lambda,u\rangle}\, \breve{C}_2(du)$ for all $\lambda \in \mathbb{R}^d$. Then f is a real-valued function that is nonnegative and continuous.*

Proof. Fix any $\lambda \in \mathbb{R}^d$. Let N be the point process whose reduced covariance measure is $\breve{C}_2$. Recall that ℓ denotes the Lebesgue measure. For any bounded Borel set A, the variance of the normalized finite Fourier transform $\ell(A)^{-1/2} \int_A e^{i\langle\lambda, s\rangle}\ N(ds)$ is

$$\begin{aligned} & \ell(A)^{-1} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} 1_A(s)\, 1_A(s+u)\, e^{-i\langle\lambda,u\rangle}\ \breve{C}_2(du)\ \ ds \\ = \ & \ell(A)^{-1} \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} 1_A(s)\, 1_A(s+u)\ ds \right)\ e^{-i\langle\lambda,u\rangle}\ \breve{C}_2(du) \\ = \ & \int_{\mathbb{R}^d} 1_{A'}(u) \left\{ \ell(A \cap T_{-u}A) / \ell(A) \right\}\ e^{-i\langle\lambda,u\rangle}\ \breve{C}_2(du), \end{aligned} \tag{4.28}$$

where the last equality follows from Lemma 1. Consider a sequence of bounded Borel sets $\{A_n\}$ such that for each $n,\ A_n$ is a rectangle $\Pi_{i=1}^d [a_{in},\ b_{in}]$. Suppose that as $n \to \infty$, $b_{in} - a_{in} \to \infty,\ i = 1,\ \cdots,\ d$, then the integral in (4.28) converges to $\int_{\mathbb{R}^d} e^{-i\langle\lambda,u\rangle}\ \breve{C}_2(du)$ by dominated convergence. Since the last integral is the limit of a sequence of non-negative real numbers, it is also real and non-negative.

To show continuity of f, it suffices to consider

$$g(h) = \int_{\mathbb{R}^d} \left\{ \cos(\langle\lambda + h,\ u\rangle) - \cos(\langle\lambda,\ u\rangle) \right\}\ \left|\breve{C}_2\right|(du).$$

Consider any sequence $\{h_n\}$ such that $\lim_{n\to\infty} h_n = 0$. Since

$$\int_{\mathbb{R}^d} \left|\cos(\langle\lambda + h_n,\ u\rangle) - \cos(\langle\lambda,\ u\rangle)\right|\ \left|\breve{C}_2\right|(du) \le 2 \int_{\mathbb{R}^d} \left|\breve{C}_2\right|(du) < \infty,$$

by dominated convergence, $\lim_{n\to\infty} g(h_n) = 0$. Hence f is continuous. ■

Lemma 3 *For any bounded Borel subsets $A,\ B$ of $\mathbb{R}^d$, let $g(u) = \int_{\mathbb{R}^d} \mathbf{1}_B(s)\, \mathbf{1}_A(s+u)\ ds$. Then g is continuous and Lebesgue integrable.*

Proof. By Fubini's Theorem,

$$\begin{aligned} \int_{\mathbb{R}^d} g(u)\ du &= \int_{\mathbb{R}^d} \mathbf{1}_B(s) \left(\int_{\mathbb{R}^d} \mathbf{1}_A(s+u)\ du \right)\ ds \\ &= \ell(B)\, \ell(A) < \infty. \end{aligned}$$

Hence g is Lebesgue integrable. For $\varepsilon > 0$, by continuity of translation of integrable functions, see Section D of Chapter 7 in Jones (2001), there exists $\delta > 0$ such that, for all $u' \in \mathbb{R}^d$, if $\|u - u'\| < \delta$, then

$$\int_{\mathbb{R}^d} \left|\mathbf{1}_A(s+u) - \mathbf{1}_A(s+u')\right| ds < \varepsilon.$$

For such $u' \in \mathbb{R}^d$, since

$$\left|\mathbf{1}_A(s+u) - \mathbf{1}_A(s+u')\right|^2 = \left|\mathbf{1}_A(s+u) - \mathbf{1}_A(s+u')\right|,$$

by Schwarz's inequality,

$$\begin{aligned}
|g(u)-g(u')| &\leq \int_{\mathbb{R}^d} \mathbf{1}_B(s)\,|\mathbf{1}_A(s+u)-\mathbf{1}_A(s+u')|\;ds \\
&\leq \left(\int_{\mathbb{R}^d} \mathbf{1}_B(s)^2\;ds\right)^{1/2}\left(\int_{\mathbb{R}^d} |\mathbf{1}_A(s+u)-\mathbf{1}_A(s+u')|^2\,ds\right)^{1/2} \\
&= \varepsilon^{1/2}\ell(B)^{1/2}.
\end{aligned}$$

Hence g is continuous in u. ■

Lemma 4 *If an increasing sequence of sets $\{A_n\}$ converges to infinity in the sense of van Hove, then*

(i) there exists a sequence of balls $\{B_n\}$ such that, for each n, $B_n \subset A_n$ and their radii $r_n \to \infty$ as $n \to \infty$;

(ii) $\lim_{n\to\infty} 1_{A'_n} = 1_{\mathbb{R}^d}$, where $A'_n = \{t_1 - t_2 : t_1,\ t_2 \in A_n\}$; and

(iii) for each $u \in \mathbb{R}^d$, $\lim_{n\to\infty}[\ell(A_n \cap (A_n - u))/\ell(A_n)] = 1$, where $A_n - u = \{t - u : t \in A_n\}$.

Proof. (i) Suppose that there is no such a sequence of balls. For each n let B_n be the biggest ball contained in A_n. Then there exists a finite constant c such that $r_n \leq c$ for all n. Let a be a vector in $\mathbb{R}^d$ defined as in the definition of van Hove convergence. Set $a = (4c, ..., 4c)'$. Since A_n converges to infinity in the sense of van Hove, for some large n there exists a rectangular parallelepiped contained in A_n such that the lengths of its edges are $4c$. As a ball with radius $2c$ can be contained in this rectangular parallelepiped, this leads to a contradiction. Hence the truth of statement (i) is proven.

(ii) For any $x \in \mathbb{R}^d$, due to part (i), for some sufficiently large N there exists a ball B_N contained in A_N with radius $r_N > \|x\|$. Let c_N be the centre of the ball. Then $x + c_N \in B_N$ and therefore $x \in A'_N$. This is also the case for all $n \geq N$.

(iii) Now fix $u \in \mathbb{R}^d$. Let $A_n(h)$ be the set of points with distance less than or equal to h to the boundary of A_n. Clearly $A_n \backslash A_n(2\|u\|) \subset A_n \cap (A_n - u)$. Choose $a \in \mathbb{R}^d$ that are associated with van Hove convergence so that $a_i \geq 3\|u\|$ for all $i = 1, ..., d$, then

$$\bigcup_{m:\Pi_m \subset A_n} \Pi_m \subset A_n \backslash A_n(2\|u\|) \subset A_n \cap (A_n - u).$$

Therefore $\lim_{n\to\infty}[\ell(A_n \cap (A_n - u))/\ell(A_n)] = 1$. ■

Lemma 5 *Consider a curve $\gamma : [a,\ b] \to X$ on a metric space $(X,\ d)$. Suppose that γ has a finite length, define a function $L : [a,\ b] \to \mathbb{R}$ by $L(x) = \text{length}\left(\gamma_{[a,x]}\right)$ where $\gamma_{[a,x]}$ is the restriction of γ to $[a,\ x]$. Then L is continuous on $[a,\ b]$ and non-decreasing.*

Proof. For $a \leq x \leq y \leq b$, let $\gamma_{[x,y]}$ be the restriction of γ to $[x, y]$. It follows that for $a \leq x \leq y \leq z \leq b$,

$$\text{length}\left(\gamma_{[x,z]}\right) = \text{length}\left(\gamma_{[x,y]}\right) + \text{length}\left(\gamma_{[y,z]}\right). \tag{4.29}$$

Note that the definition of length is analogous to that of total variation. The proof of the analogous additive property of total variation can be employed. Hence it follows that L is non-decreasing since length is, by definition, non-negative.

For any $\varepsilon > 0$, there is a partition $P = \{t_0,\ t_1,\ \cdots,\ t_m\}$ such that

$$L(b) - \sum_{i=1}^{m} d(\gamma(t_i),\ \gamma(t_{i-1})) < \varepsilon/2.$$

Since γ is continuous on $[a,\ b]$, it is also uniformly continuous on $[a,\ b]$. Hence there is $\delta > 0$ such that whenever $|t-s| < \delta$, then $d(\gamma(s),\ \gamma(t)) < \varepsilon/2(m+2)$. For any $x \in [a,\ b)$, take $\delta_1 = \min\{\delta,\ b-x\}$. By (4.29), for any $0 < h < \delta_1$,

$$L(b) = \text{length}\left(\gamma_{[a,x]}\right) + \text{length}\left(\gamma_{[x,x+h]}\right) + \text{length}\left(\gamma_{[x+h,b]}\right).$$

Notice that if P' is a refinement of P, i.e. $P \subset P'$, then $l_{P'} \geq l_P$. Let $P' = P \cup \{x,\ x+h\}$. Then $L(b) - l_{P'} < \varepsilon/2$ too. This implies that

$$\text{length}\left(\gamma_{[x,x+h]}\right) - \sum'' d(\gamma(s),\ \gamma(t)) < \varepsilon/2,$$

where the summation is over $s,\ t \in P' \cap [x,\ x+h]$. Using the properties mentioned above, it follows that

$$L(x+h) - L(x) = \text{length}\left(\gamma_{[x,x+h]}\right) < \varepsilon.$$

Similarly for $h < 0$ and $a < x \leq b$. Hence L is continuous on $[a,\ b]$. ■

Lemma 6 *For a sequence of bounded Borel sets* $\{B_n\}$,

$$Var(\zeta_\lambda(B_n)) = O(\ell(B_n)),$$

and if $\{B_n\}$ *satisfies Assumption C3, as* $n \to \infty$,

$$[\ell(B_n)]^{-1} Var[\zeta_\lambda(B_n)] \to \lambda' \left\{\int_{\mathbb{R}^d} \gamma(u)\ \breve{M}_2(du)\right\} \lambda. \tag{4.30}$$

Proof. Recall that $\mathbb{E}[\zeta_\lambda(B_n)] = 0$. Proceeding as in the proof of Theorem A3, it follows that

$$Var[\zeta_\lambda(B_n)] = \ell(B_n)\left[\lambda' \left\{\int_{\mathbb{R}^d} 1_{B'_n}(u) \frac{\ell(B_n \cap T_{-u}B_n)}{\ell(B_n)} \gamma(u)\ \breve{M}_2(du)\right\} \lambda\right]. \tag{4.31}$$

By definition $\gamma(u)$ is p.s.d. for all $u \in \mathbb{R}^d$. By Assumption C1, for all $n \geq 1$, the term in the square brackets in (4.31) is bounded in absolute value by $\int_{\mathbb{R}^d} \lambda' \gamma(u) \lambda\ \breve{M}_2(du) < \infty$. Hence $Var[\zeta(B_n)] = O(\ell(B_n))$ as $n \to \infty$.

Suppose that Assumption C3 holds for the sequence $\{B_n\}$. Fix $u \in \mathbb{R}^d$. From Lemma 4, the integrand in (4.31) converges to $\gamma_z(u)$ as $n \to \infty$. As element of the integrand is also bounded in absolute value by $|\gamma_{rs}(u)|$, which is integrable with respect to $\breve{M}_2$, Assumption C1 and dominated convergence imply (4.30). ■

Lemma 7 *There exists a finite constant C such that*

$$\left| \int_{A_n^4} \phi(s_1, \cdots, s_4) \kappa_{rs}(s_1, \cdots, s_4) M_4 \left(\Pi_{j=1}^4 ds_j \right) \right| \leq C\ell(A_n).$$

Proof. By Assumption D5, the left side is

$$\int_{\mathbb{R}^{4d}} \mathbf{1}_{A_n}(x) \mathbf{1}_{A_n}(x+u_1) \mathbf{1}_{A_n}(x+u_2) \mathbf{1}_{A_n}(x+u_3) \phi(x, x+u_1, x+u_2, x+u_3)$$
$$\kappa_{rs}(x, x+u_1, x+u_2, x+u_3) \; dx \; \breve{M}_4(du_1 \times du_2 \times du_3).$$

Its absolute value is bounded by

$$\sup_{x \in \mathbb{R}^d} \int_{\mathbb{R}^{3d}} |\kappa_{rs}(x, x+u_1, x+u_2, x+u_3)| \breve{M}_4 \left(\Pi_{j=1}^3 du_j \right) \int_{A_n} dx.$$

Assumption D6 implies that this is not greater than $C\ell(B_n)$ for some finite constant C. ■

Lemma 8 *As $n \to \infty$,*

$$\int_{B_n^4} \phi(s_1, s_2, s_3, s_4) \gamma_{rr}(s_3 - s_1) \gamma_{ss}(s_4 - s_2) C_3(ds_2 \times ds_3 \times ds_4) ds_1 = O(\ell(B_n)).$$

Proof. The left-side is

$$\int_{\mathbb{R}^{4d}} \mathbf{1}_{B_n}(x_1) \mathbf{1}_{B_n}(x_2) \mathbf{1}_{B_n}(x_2+u_1) \mathbf{1}_{B_n}(x_2+u_2) \phi(x_1, x_2+u_1, x_2, x_2+u_2)$$
$$\gamma_{rr}(x_2 - x_1) \gamma_{ss}(u_2 - u_1) \breve{C}_3(du_1 \times du_2) dx_1 dx_2.$$

Hence, by Assumptions D5 and D7, its modulus is bounded by

$$\begin{aligned} & \gamma_{ss}(0) \int_{\mathbb{R}^{2d}} \mathbf{1}_{B_n}(x_1) \mathbf{1}_{B_n}(x_2) |\gamma_{rr}(x_2 - x_1)| dx_1 dx_2 \int_{\mathbb{R}^{2d}} \left| \breve{C}_3 \right| (du_1 \times du_2) \\ \leq \; & C\ell(A_n) \int_{B_n'} \frac{\ell(B_n \cap T_{-u} B_n)}{\ell(B_n)} |\gamma_{rr}(u)| \; du = O(\ell(B_n)). \end{aligned}$$

■

Lemma 9 *As $n \to \infty$,*

$$\int_{B_n^4} \phi(s_1, s_2, s_3, s_4) \gamma_{rr}(s_3 - s_1) \gamma_{ss}(s_4 - s_2) C_2(dx_1 \times dx_2) C_2(dx_3 \times dx_4) = O(\ell(B_n)).$$

Proof. Proceeding as in the proof of the previous lemma, it can be shown that the modulus of the left side is bounded by

$$\gamma_{ss}(0)\left\{\int_{\mathbb{R}^d}\left|\breve{C}_2\right|(du_1)\right\}^2\int_{\mathbb{R}^{2d}}\mathbf{1}_{B_n}(x_1)\,\mathbf{1}_{B_n}(x_2)\left|\gamma_{rr}(x_2-x_1)\right|dx_1dx_2=O(\ell(B_n)).$$

■

Lemma 10 *As $n\to\infty$,*

$$\mu^2\int_{B_n^4}\phi(s_1,s_2,s_3,s_4)\,\gamma_{rr}(s_3-s_1)\,\gamma_{ss}(s_4-s_2)\,C_2(dx_3\times dx_4)\,dx_1dx_2=O(\ell(B_n)).$$

Proof. The modulus of the left side is bounded by

$$\begin{aligned}&\mu^2\int_{\mathbb{R}^d}\mathbf{1}_{B_n}(x_3)\int_{\mathbb{R}^d}\left(\int_{\mathbb{R}^d}\left|\gamma_{rr}(x_3-x_1)\right|dx_1\right)\left(\int_{\mathbb{R}^d}\left|\gamma_{ss}(x_3+u-x_2)\right|dx_2\right)\left|\breve{C}_2\right|(du)\,dx_3\\ =\;&\mu^2\left(\int_{\mathbb{R}^d}\left|\gamma_{rr}(x)\right|dx\right)\left(\int_{\mathbb{R}^d}\left|\gamma_{ss}(x)\right|dx\right)\left(\int_{\mathbb{R}^d}\left|\breve{C}_2\right|(du)\right)\ell(B_n).\end{aligned}$$

■

Lemma 11 *As $n\to\infty$,*

$$\begin{aligned}&\frac{1}{\ell(B_n)\,m_n}\int_{B_n^4}\phi(s_1,s_2,s_3,s_4)\,\gamma_{rr}(s_3-s_1)\,\gamma_{ss}(s_4-s_2)\,M_2(ds_1\times ds_3)\,M_2(ds_2\times ds_4)\\ \to\;&f_{\zeta,rr}(\lambda)\,f_{\zeta,ss}(\lambda)\int_{\mathbb{R}^d}w(u)^2\,du,\end{aligned}$$

where $w(u)=\Pi_{i=1}^d k(u_i)$.

Proof. Using Lemma 1, the left side is

$$\begin{aligned}&\frac{1}{\ell(B_n)\,m_n}\int_{\mathbb{R}^{4d}}\mathbf{1}_{B_n}(s_1)\,\mathbf{1}_{B_n}(s_2)\,\mathbf{1}_{B_n}(s_3)\,\mathbf{1}_{B_n}(s_4)\,e^{-i\langle\lambda,s_2-s_1-s_4+s_3\rangle}w_n(s_2-s_1)\\ &w_n(s_4-s_3)\,\gamma_{rr}(s_3-s_1)\,\gamma_{ss}(s_4-s_2)\,M_2(ds_1\times ds_3)\,M_2(ds_2\times ds_4)\\ =\;&\frac{1}{\ell(B_n)\,m_n}\int_{\mathbb{R}^{4d}}\mathbf{1}_{B_n\cap T_{-u}B_n}(s_1)\,\mathbf{1}_{B_n\cap T_{-v}B_n}(s_2)\,\mathbf{1}_{B_n'}(u)\,\mathbf{1}_{B_n'}(v)\,e^{-i\langle\lambda,u-v\rangle}w_n(s_2-s_1)\\ &w_n(s_2-s_1+v-u)\,\gamma_{rr}(u)\,\gamma_{ss}(v)\,ds_1ds_2\breve{M}_2(du)\,\breve{M}_2(dv)\\ =\;&\frac{1}{\ell(B_n)\,m_n}\int_{\mathbb{R}^{3d}}\mathbf{1}_{B_n'}(u)\,e^{-i\langle\lambda,u\rangle}\gamma_{rr}(u)\,\mathbf{1}_{B_n'}(v)\,e^{i\langle\lambda,v\rangle}\gamma_{ss}(v)\,w_n(u')\,w_n(u'+v-u)\\ &\left\{\int_{\mathbb{R}^d}\mathbf{1}_{B_n\cap T_{-u}B_n\cap T_{-u'}\{B_n\cap T_{-v}B_n\}}(s)\,ds\right\}\mathbf{1}_{\{B_n\cap T_{-v}B_n\}-\{B_n\cap T_{-u}B_n\}}(u')\,du'\breve{M}_2(du)\,\breve{M}_2(dv)\\ =\;&\int_{\mathbb{R}^{2d}}\mathbf{1}_{B_n'}(u)\,e^{-i\langle\lambda,u\rangle}\gamma_{rr}(u)\,\mathbf{1}_{B_n'}(v)\,e^{i\langle\lambda,v\rangle}\gamma_{ss}(v)\left\{\int_{\mathbb{R}^d}\frac{1}{m_n}w_n(u')\,w_n(u'+v-u)\right.\\ &\left.\frac{\ell(B_n\cap T_{-u}B_n\cap T_{-u'}B_n\cap T_{-v-u'}B_n)}{\ell(B_n)}\mathbf{1}_{\{B_n\cap T_{-v}B_n\}-\{B_n\cap T_{-u}B_n\}}(u')\,du'\right\}\breve{M}_2(du)\,\breve{M}_2(dv).\end{aligned}$$

To conclude the result, one can follow standard steps employed in time series which rely on dominated convergence. The proof of Lemma 4 can be extended to show that, under van Hove convergence, $\ell(B_n \cap T_{-u}B_n \cap T_{-u'}B_n \cap T_{-v-u'}B_n)/\ell(B_n) \to 1$. ■

Lemma 12 *Suppose that B_0 is an unknown $p \times p$ matrix and $\widehat{B}$ is an estimate such that $\mathbb{E}\left(\widehat{b}_{ij} - b_{0ij}\right)^2 = O(f_n)$ for all $i,\ j = 1, ..., p$. Suppose that x is a p-random vector such that $\|x\|^2 = 1$ a.s., then $\mathbb{E}\left(x'\widehat{B}x - x'B_0x\right)^2 = O(f_n)$.*

Proof. Let $x = (x_1,\ ...,\ x_p)'$. By Schwarz's inequality,

$$\begin{aligned}
\mathbb{E}\left(x'\widehat{B}x - x'B_0x\right)^2 &= \mathbb{E}\left\{\sum_{i=1}^{p}\sum_{j=1}^{p}\left(\widehat{b}_{ij} - b_{0ij}\right)x_ix_j\right\}^2 \\
&\leq \sum_{i=1}^{p}\sum_{j=1}^{p}\sum_{k=1}^{p}\sum_{l=1}^{p}\mathbb{E}\left|\widehat{b}_{ij} - b_{0ij}\right|\left|\widehat{b}_{kl} - b_{0kl}\right| \\
&= O(f_n).
\end{aligned}$$

■

Lemma 13 *Suppose that B_0 is an unknown $p \times p$ matrix and $\widehat{B}$ is an estimate such that $\mathbb{E}\left(\widehat{b}_{ij} - b_{0ij}\right)^2 = O(f_n)$ for all $i,\ j = 1, ..., p$. Let $\{c_n\}$ be a sequence of positive numbers such that $\lim_{n\to\infty} c_n = 0$. If B_0 is positive definite, then $\mathbb{E}\left\{\mathbf{1}\left(\underline{\lambda}\left(\widehat{B}\right) < c_n\right)\right\} = O(f_n)$, where $\mathbf{1}$ is the indicator function.*

Proof. For a random matrix $\widehat{B}$ it is possible to construct a measurable function h such that $h\left(\widehat{B}\right) = \widehat{x}$ where $\widehat{x}$ is a normalized eigenvector corresponding to $\underline{\lambda}\left(\widehat{B}\right)$. The construction is based on employing row operations on the matrix $\widehat{B} - \underline{\lambda}\left(\widehat{B}\right)I_p$, where I_p denotes the identity matrix of order p, to obtain a reduced row-echelon form. By the previous lemma $\mathbb{E}\left(\widehat{x}'\widehat{B}\widehat{x} - \widehat{x}'B_0\widehat{x}\right)^2 = O(f_n)$. Since $c_n \to 0$ as $n \to \infty$, there is $N < \infty$ such that $c_n < \underline{\lambda}(B_0)/2$ for all $n \geq N$. Hence for large enough n, using the fact that $\widehat{x}'B_0\widehat{x} \geq \underline{\lambda}(B_0)$, $\left|\widehat{x}'\widehat{B}\widehat{x} - \widehat{x}'B_0\widehat{x}\right| \leq \underline{\lambda}(B_0)/2$ implies $\widehat{x}'\widehat{B}\widehat{x} \geq \widehat{x}'B_0\widehat{x} - \underline{\lambda}(B_0)/2 \geq \underline{\lambda}(B_0)/2 \geq c_n$. Since $\widehat{x}'\widehat{B}\widehat{x} = \underline{\lambda}\left(\widehat{B}\right)$, $\left|\widehat{x}'\widehat{B}\widehat{x} - \widehat{x}'B_0\widehat{x}\right| \leq \underline{\lambda}(B_0)/2$ implies $\underline{\lambda}\left(\widehat{B}\right) \geq c_n$. Therefore, for sufficiently large n, using Markov's inequality,

$$\begin{aligned}
\mathbb{E}\left|\mathbf{1}\left(\underline{\lambda}\left(\widehat{B}\right) < c_n\right)\right| &= \mathbb{P}\left\{\underline{\lambda}\left(\widehat{B}\right) < c_n\right\} \\
&\leq \mathbb{P}\left\{\left|\widehat{x}'\widehat{B}\widehat{x} - \widehat{x}'B_0\widehat{x}\right| > \underline{\lambda}(B_0)/2\right\} \\
&\leq \{\underline{\lambda}(B_0)/2\}^2\,\mathbb{E}\left(\widehat{x}'\widehat{B}\widehat{x} - \widehat{x}'B_0\widehat{x}\right)^2 \\
&= O(f_n).
\end{aligned}$$

■

Reference

Abadir, K.M., Magnus J.R., 2005. Matrix algebra. Cambridge University Press, New York.

Abramowitz, M., Stegun I.A., Eds., 1964. Handbook of Mathematical Functions. Dover, New York.

Anderson, T.W., 2003. An introduction to multivariate statistical analysis. Wiley, New Jersey.

Andrews, D.W.K., 1991. Asymptotic normality of series estimators for nonparametric and semiparametric Models. Econometrica 59, 307-345.

Anselin, L., 1988. Spatial econometrics: methods and models. Kluwer Academic Publishers, The Netherlands.

Banerjee, A.V., Iyer L., 2005. History, institutions and economic performance: the legacy of colonial land tenure systems in India. American Economic Review 95, 1190-1213.

Beran, R., 1976. Adaptive estimates for autoregressive processes. Annals of the Institute of Statistical Mathematics 26, 77-89.

Bickel, P., 1982. On adaptive estimation. Annals of Statistics 10, 647-671.

Bolthausen, E., 1982. On the central limit theorem for stationary mixing random fields. Annals of Probability 10, 1047-1050.

Brillinger, D.R., 1972. The spectral analysis of stationary interval functions. In Proc. Sixth Berkeley Symposium. Vol. I, 483-513. University of California Press, Berkeley.

Brillinger, D.R., 1973. Estimation of the mean of a stationary time series by sampling. Journal of Applied Probability 10, 419-431.

Brillinger, D.R., 1986. Regression for randomly sampled spatial series: the trigonometric case. Journal of Applied Probability 23, 275-289.

Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods (Second Edition). Springer and Verlag, New York.

Bulinskii, A.V., Zhurbenko, I.G., 1976. A central limit theorem for additive random functions. Theory of Probability and Its Applications 4, 687-697.

Castellana, J.V., Leadbetter, M.R., 1986. On smoothed probability density estimation for stationary processes. Stochastic Processes and Their Applications 21, 179-193.

Chamberlain G., 1992. Efficiency bounds for semiparametric regression. Econometrica 60, 567-596.

Cliff, A.D., Ord, J.K., 1973. Spatial autocorrelation. Pion Ltd, London.

Conley, T.G., 1999. GMM estimation with cross sectional dependence. Journal of Econometrics 92, 1-45.

Cox, D.D., 1988. Approximation of least squares regression on nested subspaces. Annals of Statistics 16, 713-732.

Daley, D.J., Vere-Jones, D., 2003. An introduction to the theory of point processes volume I: Elementary theory and methods. Springer, New York.

Daley, D.J., Vere-Jones, D., 2008. An introduction to the theory of point processes volume II: General theory and structure. Springer, New York.

Dunkl, C.F., Xu, Y., 2001. Orthogonal Polynomials of Several Variables. Cambridge University Press, New York.

Eicker, F., 1967. Limit theorems for regression with unequal and dependent errors. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, University of California Press, Berkeley, 59–82.

Fairfield Smith, H., 1938. An empirical law describing heterogeneity in the yields of agricultural crops. Journal of Agricultural Science 28, 1-23.

Fan, Y., Li, Q., 1999. Root-n-consistent estimation of partially linear time series models. Journal of Nonparametric Statistics 11, 251-269.

Freud, G., 1971. Orthogonal Polynomials. Pergamon Press, Oxford.

Fujita, M., Krugman, P., Venables A.J., 2001. The spatial economy: cities, regions and international trade. MIT Press, Cambridge MA. and London.

Gao, J.T., 1998. Semiparametric regression smoothing of nonlinear time series. Scandinavian Journal of Statistics 25, 521-539.

Gao, J.T., Yee, T., 2000. Adaptive estimation in partially linear autoregressive models. Canadian Journal of Statistics 28, 571-586.

Grenander, U., Szego, G., 1984. Toeplitz forms and their applications. Chelsea Publishing Company, New York.

Guan, Y., Sherman, M., Calvin, J.A., 2004. A nonparametric test for spatial isotropy using subsampling, Journal of the American Statistical Association 467, 810-821.

Hidalgo, J., 1992. Adaptive estimation in time series regression models with heteroskedasticity of unknown form. Econometric Theory 8, 161-187.

Horn, R.A., Johnson C.R., 1985. Matrix Analysis. Cambridge University Press, New York.

Ibragimov, I.A., Linnik, Y.V., 1971. Independent and Stationary Sequences of Random Variables. Wolters-Noordhoff Publishing, Groningen.

Jenish, N., Prucha, I.R., 2009. Central limit theorems and uniform laws of large numbers for arrays of random fields. Journal of Econometrics 150, 86-98.

Jones, F., 2001. Lebesgue integration on Euclidean space. Jones and Bartlett Publishers International, Sudbury MA.

Kelejian, H.H., Prucha, I.R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. Journal of Real Estate Finance and Economics 17, 99-121.

Kelejian, H.H., Prucha, I.R., 2007. HAC estimation in a spatial framework. Journal of Econometrics 140, 131-154.

Lee, L.F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72, 1899-1925.

Lee, L.F., Yu, J., 2010. A spatial dynamic panel data model with both time and individual fixed effects. Econometric Theory 26, 564-597.

Masry, E., 1978. Poisson sampling and spectral estimation of continuous-time Processes, IEEE Transactions on Information Theory 34, 173-183.

Masry, E., 1983. Non-parametric covariance estimation from irregularly-spaced data. Advances in Applied Probability 15, 113-132.

Mitchell, T.D., Jones, P.D., 2005. An improved method of constructing a database of monthly climate observations and associated high-resolution grids. International Journal of Climatology 25, 693-712.

Munshi, K., 2004. Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution. Journal of Development Economics 73, 185-213.

Newey, W.K., 1988. Adaptive estimation of regression models via moment restrictions. Journal of Econometrics 38, 301-339.

Newey, W.K., 1997. Convergence rates and asymptotic normality for series estimators. Journal of Econometrics 79, 147-168.

Newey, W.K., West K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55, 703-708.

Newman, C.M., 1980. Normal fluctuations and the FKG inequalities. Communications in Mathematical Physics 74, 119-128.

Nishiyama, Y., Robinson, P.M., 2000. Edgeworth expansions for semiparametric averaged derivatives. Econometrica 68, 931-979.

Parzen, E., 1963. On spectral analysis with missing observations and amplitude modulation. Sankya 25, 180-189.

Politis, D.N., Sherman, M., 2001. Moment estimation for statistics from marked point processes. Journal of Royal Statistical Society Series B 63, 261-275.

Robinson, P.M., 1980. Continuous model fitting from discrete data. In Directions in Time Serie (D. R. Brillinger and G. C. Tiao, eds), 263-278, Institute of Mathematical Statistics.

Robinson, P.M., 1988. Root-n-consistent semiparametric regression. Econometrica 56, 931-954.

Robinson, P.M., 2005. Efficiency improvements in inference on stationary and nonstationary fractional time series. Annals of Statistics 33, 1800-1842.

Robinson, P.M., 2007. Nonparametric spectrum estimation for spatial data. Journal of Statistical Planning and Inference (Special Issue in honour of Madan L. Puri) 137, 1024-1034.

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

Robinson, P.M., 2010. Efficient estimation of the semiparametric spatial autoregressive model. Journal of Econometrics 157, 6-17.

Robinson, P.M., 2011. Asymptotic theory for nonparametric regression with spatial data. Journal of Econometrics 165, 5-19.

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

Ruelle, D., 1999. Statistical Mechanics: Rigorous Results. World Scientific Publishing Co Pte Ltd, Singapore.

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

Stein, C., 1956. Efficient nonparametric testing and estimation. Proceedings of the Third

Berkeley Symposium on Mathematical Statistics and Probability 1, 187-196. University of California Press.

Stone, C.J., 1975. Adaptive Maximum-Likelihood Estimators of a Location Parameter. Annals of Statistics 3, 267-284.

Thornett, M.L., 1979. A class of second-order stationary random measures. Stochastic Processes and their Applications 8, 323-334.

Wilson, P.M.H., 2008. Curved Spaces: From Classical Geometries to Elementary Differential Geometry. Cambridge University Press, UK.

Zhurbenko, I.G., 1986. The Spectral Analysis of Time Series. North-Holland, Amsterdam.