The London School of Economics and Political Science

Essays on Spatial Autoregressive Models with Increasingly Many Parameters

Abhimanyu Gupta

A thesis submitted to the Department of Economics of the London School of Economics and Political Science for the degree of Doctor of Philosophy

05/07/2013

Declaration

I certify that the thesis I have presented for examination for the PhD 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 $65,045$ words.

Abhimanyu Gupta

Abstract

Much cross-sectional data in econometrics is blighted by dependence across units. A solution to this problem is the use of spatial models that allow for an explicit form of dependence across space. This thesis studies problems related to spatial models with increasingly many parameters. A large proportion of the thesis concentrates on Spatial Autoregressive (SAR) models with increasing dimension. Such models are frequently used to model spatial correlation, especially in settings where the data are irregularly spaced.

Chapter 1 provides an introduction and background material for the thesis. Chapter 2 develops consistency and asymptotic normality of least squares and instrumental variables (IV) estimates for the parameters of a higher-order spatial autoregressive (SAR) model with regressors. The order of the SAR model and the number of regressors are allowed to approach infinity with sample size, and the permissible rate of growth of the dimension of the parameter space relative to sample size is studied.

An alternative to least squares or IV is to use the Gaussian pseudo maximum likelihood estimate (PMLE), studied in Chapter 3. However, this is plagued by finite-sample problems due to the implicit definition of the estimate, these being exacerbated by the increasing dimension of the parameter space. A computationally simple Newton-type step is used to obtain estimates with the same asymptotic properties as those of the PMLE.

Chapters 4 and 5 of the thesis deal with spatial models on an equally spaced, d-dimensional lattice. We study the covariance structure of stationary random fields defined on d-dimensional lattices in detail and use the analysis to extend many results from time series. Our main theorem concerns autoregressive spectral density estimation. Stationary random fields on a regularly spaced lattice have an infinite autoregressive representation if they are also purely non-deterministic. We use truncated versions of the AR representation to estimate the spectral density and establish uniform consistency of the proposed spectral density estimate.

To my parents

Acknowledgements

I am greatly indebted to my supervisor, Professor Peter Robinson, for his invaluable support and guidance.

The econometrics faculty at LSE were extremely supportive and helpful. I am grateful to Javier Hidalgo, Tatiana Komarova, Oliver Linton, Taisuke Otsu, Marcia Schafgans and Myung Hwan Seo for all the help and feedback over the years

I would also like to thank fellow researchers at the LSE, in particular the participants at the work in progress seminars, for valuable feedback on my work. I had many enjoyable discussions with Abhisek Banerjee, Svetlana Bryzgalova, Ziad Daoud, Jungyoon Lee, Francesca Rossi and Sorawoot Tang Srisuma. Sometimes we actually talked about econometrics.

Outside of the econometrics group, particular thanks are due to my long-suffering office-mates Wenya Cheng and Attakrit Leckcivilize, who stoically accepted the presence of a theoretical econometrician in their midst. Kitjawat Tacharoen, Daniel Vernazza and Mazhar Waseem ensured that there were always lighter moments.

I owe a lot to my family and friends outside LSE for their support over the past few years. In particular, none of this would have been possible without the love and support of my parents, Swati and my sister. Ishan Bhaskar, Dhananjai Jain, Udit Khanna, Priya Mukherjee, Abhishek Parikh and Shruti Sinha always provided the laughs when they were needed most, and the perspective when I took myself too seriously.

Contents

List of Tables

List of Figures

1 Introduction

In this chapter we provide the necessary background to evaluate the contribution of this thesis to the spatial econometrics and spatial statistics literature. Section 1.1 summarizes the key properties of spatial data and discusses solutions for issues that arise in the analysis of such data. Section 1.2 introduces spatial autoregressions for irregularly-spaced data, while Section 1.3 does the same for data on a regularly-spaced lattice. These sections also provide motivation for the analysis of higher-order autoregressions. Section 1.4 summarizes the literature on models with increasingly many parameters, while Section 1.5 outlines the contribution of this thesis to the literature. Finally, Section 1.6 introduces some notation and definitions that will be used throughout the thesis.

1.1 Issues in the analysis of spatial data

Correlation in cross-sectional data poses considerable challenges to econometricians and statisticians, complicating both modelling and statistical inference. In econometrics, a substantial literature collectively known as Spatial Econometrics has analysed the problems caused by correlation between observations at different points in space. This goes back as far as the early work by Moran (1950), and key waypoints in the journey of the literature have been the contributions by Cliff and Ord (1973) and Cressie (1993). Survey articles outlining recent developments in spatial econometrics include Robinson (2008) and Anselin (2010). A feature of the spatial econometrics literature is its focus on spatial data recorded at irregularly-spaced points. This is reflective of typical datasets available in economic applications. Due to the irregularity of the spacing and the ambiguity about the process generating the locations of observations, fairly strong assumptions are necessary to capture spatial correlation parsimoniously. In this thesis, we will concentrate on a class of assumptions that give rise to an 'autoregressive' model.

On the other hand, much of the spatial statistics literature has focused on data recorded on a regularly-spaced d-dimensional lattice, where $d > 1$. Typically the distance between observations is fixed within dimensions, but may vary between dimensions. This structure may lead the reader to anticipate the potential extension of asymptotic theory for time series. This is complicated by the fact that while the variate of a time series is influenced only by past values, for spatial processes the dependence extends in all directions. In a seminal contribution Whittle (1954) showed that, in general, multilateral models on lattices have a unilateral moving average representation on a 'half-plane', thereby extending the familiar Wold decomposition for time series. There are limits to the use of such a representation if interest is in the coefficients in the original multilateral model, as the coefficients in the unilateral representation may not have a closed-form expression in terms of the original ones even with seemingly simple multilateral models. However, as we show in Chapters 4 and 5, such unilateral representations can be extremely useful if our interest lies in prediction and spectral density estimation. Another complication is the bias in covariance estimates due to the 'edge-effect', noted by Guyon (1982) who proposed an incorrectly centred version of the covariance estimates to eliminate this effect. The edge-effect worsens with increasing d. Solutions to the edge-effect are also explored in Dahlhaus and Künsch (1987) and Robinson and Vidal Sanz (2006).

1.2 Spatial autoregressions for irregularly-spaced data

The reader may wonder if the theory of irregularly-spaced time series can be extended to the case of irregularly-spaced spatial data, just as we discussed the extension of the theory of regularly-spaced time series to many dimensions above. Robinson (1977) showed that some cases of irregularly-spaced time series can be described by an underlying continuous time process where spacing is generated by a point process. When the continuous time process is a first-order stochastic differential equation with constant coefficients and driven by white noise, consistent and asymptotically normal estimates of the unknown parameters can be obtained from an approximated Gaussian log-likelihood. This can be extended to situations when the data are recorded at irregularly-spaced geographical locations, but even then leads to complications in estimation and inference.

Besides such complications, 'space' in economic applications need not refer to geographic space. In fact the notion of economic distance encompasses many more possibilities (e.g. differences in income of economic agents), of which geographic distance is but one, and this notion of distance determines the spatial correlation between observations. In spatial econometrics, the economic distance between two economic agents (also called units) i and j is defined as the distance between two vectors of characteristics v_i and v_j. Note that we identified units with their location. This distance may be defined in a number of ways, without any geographical interpretation (see e.g. Conley and Ligon (2002), Conley and Dupor (2003)). If there is no geographical interpretation of the distance, any hope of extending the theory of irregularly-spaced time series is extinguished.

Instead, a commonly used framework for describing such data is the spatial autoregressive model, introduced by Cliff and Ord (1973, 1981). Given a sample of size n, the problem of irregular-spacing and location is circumvented by the introduction of an $n \times n$ spatial weights matrix, denoted W_n, which is chosen by the practitioner according to the particularities of the problem under consideration. Typically, the elements

$w_{ij,n}$ of W_n are inversely related to some measure of economic distance. This distance need not be geographic distance, as discussed above. The $w_{ij,n}$ may be binary, for instance taking the value 1 when two units are contiguous according to some definition of contiguity, and 0 otherwise. The SAR model can also be combined with explanatory variables to give rise to the mixed regressive SAR (MRSAR) specification, and multiple weight matrices may be included to cover spatial correlation arising from a variety of sources or from higher orders of spatial contiguity. A caveat is that adding more weight matrices can lead to circularity in dependence (see Blommestein (1985)), so care must be taken to guard against such redundancies to avoid identification problems.

For an $n \times 1$ vector of observations y_n, an $n \times k$ matrix of regressors X_n and $n \times n$ weight matrices W_{in}, $i = 1, \ldots, p$, it is assumed that there exist scalars $\lambda_1, \lambda_2, \ldots, \lambda_p$ and a $k \times 1$ vector β such that

$$y_n = \sum_{i=1}^{p} \lambda_i W_{in} y_n + X_n \beta + U_n \tag{1.2.1}$$

where U_n is an $n \times 1$ vector of disturbances. In this thesis we will refer to the MRSAR model as simply SAR and the SAR model without regressors as the pure SAR.

Typically diagonal elements of W_n are normalised to zero (see Assumption 2 and its discussion below). Another normalisation that weight matrices are frequently subjected to is row-normalisation, which ensures that each row of the normalised W_{in} sums to 1. In this case, taking $p = 1$ for illustrative purposes, the (i, j)-th element of W_{1n} is

$$w_{ij,n} = \frac{d_{ij,n}}{\sum_{h=1}^{n} d_{ih,n}} \tag{1.2.2}$$

where $d_{ij,n}$ is some measure of distance between observations at locations i and j. This provides motivation for allowing the $w_{ij,n}$ to depend on n, even if the $d_{ij,n}$ do not, implying that the y_n should be treated as triangular arrays as reflected in the subscripting with n. Kelejian and Prucha (2010) observed that if the weight matrices are subjected to a normalisation that is a function of sample size, the autoregressive parameters corresponding to the normalised weight matrices in the transformed model are dependent on n even if the original ones were not. It is clear that row-normalisation is an example of such a normalisation. The regressor matrix X_n may also contain spatial lags, and so it is attractive to allow both the autoregressive and regression parameters to vary with n. It is possible that $d_{ij,n} \neq d_{ji,n}$ so that spatial interactions are allowed to be asymmetric. See e.g. Arbia (2006) for a recent review of spatial autoregressions.

Several estimation methods have been considered for (1.2.1), the theory being generally presented for $p = 1$ i.e. the model

$$y_n = \lambda W_n y_n + X_n \beta + U_n. \tag{1.2.3}$$

The presence of spatially lagged y_n on the right side causes endogeneity problems, leading to ordinary least squares (OLS) estimation being summarily dismissed in much of the early spatial econometrics literature. However Lee (2002) showed that under additional conditions on the $w_{ij,n}$ OLS estimation can be consistent and asymptotically efficient. In particular, let h_n be a sequence that is bounded away from zero uniformly in n, and let primes indicate transposition. Lee (2002) proved that the OLS estimate of $(\lambda, \beta')'$ in (1.2.3) is consistent if $h_n \to \infty$ and the $w_{ij,n}$ are defined as in (1.2.2) with the $d_{ij,n}$ satisfying

$$c < \frac{\sum_{h=1}^{n} d_{ih,n}}{h_n},$$

where c is a generic, arbitrarily small but positive constant that is independent of n. If additionally $n^{\frac{1}{2}}/h_n \to 0$ as $n \to \infty$, then the OLS estimates are also asymptotically normal.

The instrumental variables (IV) estimate of Kelejian and Prucha (1998) is $n^{\frac{1}{2}}$-consistent (and also applicable to a version of (1.2.1) that allows for spatially correlated disturbances) under less restrictive conditions than the least squares estimate, since the introduction of h_n is not required, but is not efficient. On the other hand, it is computationally simpler than the generalized method of moments (GMM) estimate of Kelejian and Prucha (1999) and the (Gaussian) pseudo maximum likelihood estimate (PMLE) studied by Lee (2004), these being implicitly defined. The latter is obtained by maximising a Gaussian likelihood even when the disturbances are not actually Gaussian. If Gaussianity obtains, then the PMLE becomes the Maximum Likelihood Estimate (MLE) and is efficient in the Cramér-Rao sense. In fact, the asymptotic variance of the OLS estimates coincides with that of the MLE. Robinson (2010) developed asymptotic theory for efficient estimation of a semiparametric version of (1.2.1). Lee (2003) has also provided the optimal instruments for the IV estimator of Kelejian and Prucha (1998). For general, but fixed p, Lee and Liu (2010) justify an efficient GMM estimate.

The regressors X_n play a key role in estimation, with IV and OLS estimation possible only in their presence. The presence of even one non-intercept regressor can identify the spatial component of the model, as the regressor creates the correct deflation in the OLS and IV estimates. Without this deflation the deviation of the estimate from the true value converges to a non-degenerate distribution. As a result, the pure SAR model

$$y_n = \sum_{i=1}^{p} \lambda_i W_{in} y_n + U_n$$

cannot be estimated using a closed-form estimate in general. One implication of this observation is that IV and OLS estimates cannot be used to test the null hypothesis $\beta = 0$ in (1.2.1).

In Chapters 2 and 3, the spatial lag order p in (1.2.1) and the number of regressors k are allowed to increase slowly with n, as opposed to being fixed. This has attractions in that it allows for a richer model with increasing data. However, we now demonstrate by means of an example that such an asymptotic regime can arise quite naturally from applications.

A specification for the weight matrix that is frequently used for illustrative and simulation purposes is that used in Case (1991, 1992). In her scenario data are recorded in r districts, each of which contains m farmers, implying $n = mr$. It is assumed that farmers within each district impact each other equally and that there is inter-district independence between farmers so that we have

$$W_{in} = diag\left[0, \ldots, \underbrace{B_m}_{i^{th}\text{diagonal block}}, \ldots, 0\right]. \tag{1.2.4}$$

with

$$B_m = \frac{1}{m-1}\left(l_m l_m' - I_m\right) \tag{1.2.5}$$

where l_m is the m-dimensional vector of ones $(1, \ldots, 1)'$ and I_m the m-dimensional identity matrix.

With such a natural partitioning of the data, it is likely that the SAR parameters are unequal across districts. The true values may vary according to the properties of districts e.g. geographic or demographic differences to mention just two. Consider the model

$$y_n = \sum_{i=1}^{r} \lambda_i W_{in} y_n + X_n\beta + U_n, \tag{1.2.6}$$

contrasted with currently available theory, discussed above, that typically considers the specification (1.2.1) with $p = 1$ and

$$W_n = diag\left[B_m, \ldots, B_m\right]. \tag{1.2.7}$$

If we allow $n \to \infty$ with both $m \to \infty$ and $r \to \infty$ then the number of λ_is increase with n at rate r so that it is quite natural to consider an 'increasing-order' version of (1.2.1) where $p \to \infty$ as $n \to \infty$. In fact, as we demonstrate in Chapter 2, applications may even imply that both $p \to \infty$ and $k \to \infty$ as $n \to \infty$. As a result we introduce such a model and study various problems related to it in Chapters 2 and 3.

1.3 Spatial autoregressions for regularly-spaced lattice data

As mentioned before, the extension of time-series theory to even regularly-spaced lattices is not straightforward. We present a summary of the problems using examples from Whittle (1954). We first illustrate dependence from many directions by means of a simple bilateral model in one dimension ($d = 1$), and demonstrate that this can be converted into a unilateral model. Denoting observations by x_t and errors by ϵ_t, a simple bilateral autoregression in one dimension is

$$x_t = \alpha x_{t-1} + \beta x_{t+1} + \epsilon_t. \tag{1.3.1}$$

The estimation of this model by minimizing over α and β the usual least squares objective function

$$U(\alpha, \beta) = \sum_t (x_t - \alpha x_{t-1} - \beta x_{t+1})^2$$

leads to nonsensical results. This is due to the omission of the Jacobean of the transformation from ϵ_t to x_t, which is not unity for (1.3.1). The correct objective function is in fact $k(\alpha, \beta)U(\alpha, \beta)$ with

$$\log k(\alpha, \beta) = -\frac{1}{2\pi} \int_0^{2\pi} \log \left(\alpha e^{\imath\omega} - 1 + \beta e^{-\imath\omega}\right) \left(\alpha e^{-\imath\omega} - 1 + \beta e^{\imath\omega}\right) d\omega.$$

Evaluating the integral yields the objective function

$$\left\{1 + (1 - 4\alpha\beta)^{\frac{1}{2}}\right\}^{-2} \sum_t (x_t - \alpha x_{t-1} - \beta x_{t+1})^2 . \tag{1.3.2}$$

In fact (1.3.1) can be given a unilateral representation which generates the same autocorrelation function. Let a and b^{-1} be the roots of the polynomial $\alpha - z + \beta z^2$ and define A and B by comparing coefficients in

$$(z - a)(z - b) = z^2 + Az + B. \tag{1.3.3}$$

Then the AR(2) process

$$x_t + Ax_{t-1} + Bx_{t-2} = \epsilon_t$$

generates the same autocorrelations as (1.3.1). Transformation to A and B reveals that (1.3.2) is proportional to

$$\sum_t (x_t + Ax_{t-1} + Bx_{t+1})^2 .$$

Thus we have replaced (1.3.1) with a unilateral model, the parameters of which can be estimated by least-squares and used to solve for estimates of α and β via the relation

(1.3.3).

In contrast, matters are substantially more complicated in two dimensions. To illustrate this we first explain what is meant by a unilateral model in two dimensions. Suppose that x_{st} is an observable variate with each subscript being integral and denoting location in the respective dimension, and ϵ_{st} be the unobservable error. We call an autoregression of the variate x_{st} unilateral if it can be expressed as an autoregression of x_{st} on x_{su} and x_{vw} with $u > t$, $v > s$ and w unrestricted (also see Wiener (1949)). In the lattice of Figure 1.1, this means that the observation at the cross be must expressible in terms of of the observations at the black dots. The diagram motivates the use of the term 'half-plane'. Such a representation ensures a Jacobean that does not depend on the parameters, and implies that the parameters of the unilateral scheme are estimable by least-squares. The idea is easily extended to $d > 2$. We also illustrate the case of 'quarter-plane' dependence, a special case of half-plane dependence, by the region bounded by the dashed lines.

The definition of a half-plane or quarter-plane is clearly not unique but we will adopt the description of the previous paragraph and Figure 1.1 as convention (without loss of generality) in this thesis.

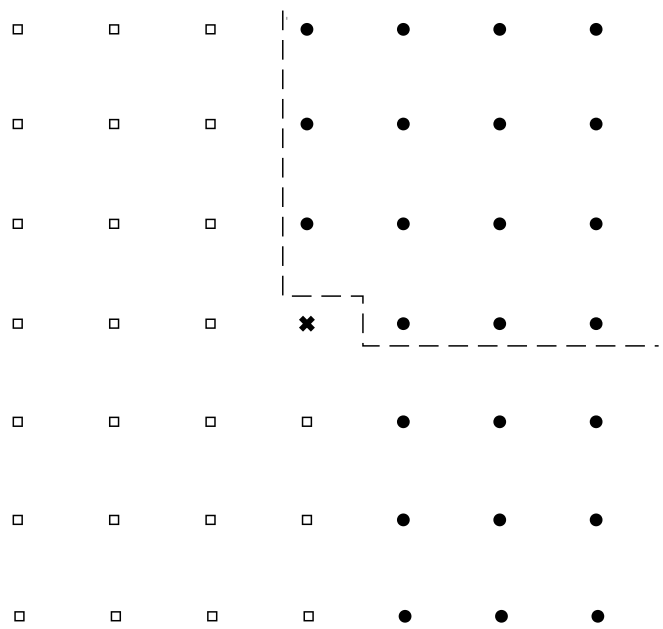

Figure 1.1: Half-plane and quarter-plane representations of two-dimensional lattice processes

Unfortunately the recovery of the parameters of the original scheme is not as straightforward as with the bilateral, $d = 1$ model (1.3.1). Indeed, it may even be impossible. The seemingly straightforward bilateral, $d = 2$ model

$$x_{st} = \alpha\left(x_{s+1,t} + x_{s-1,t} + x_{s,t+1} + x_{s,t-1}\right) + \epsilon_{st}$$

has a unilateral representation with coefficients that are expressible in no simpler form than elliptical integrals, hence yielding no closed-form. Matters are complicated further because unilateral representations of finite autoregressions may be infinite. Indeed, the finite autoregression

$$(1+\beta^2)x_{st} = \beta\left(x_{s+1,t} + x_{s,t+1} + x_{s,t-1}\right) + \epsilon_{st}$$

has an infinite unilateral representation given by

$$x_{st} = 2\beta x_{s,t+1} - \beta^2 x_{s,t+2} - \beta^2 x_{s+1,t+1} + \beta\left(1-\beta^2\right)\sum_{j=0}^{\infty}\beta^j x_{s+1,t-j} + \epsilon'_{st}, \tag{1.3.4}$$

where ϵ'_{st} is a white noise error term. Whittle (1954) proposes an approximation to the Gaussian likelihood, now called the Whittle likelihood, that permits estimation of the parameters in multilateral models.

On the other hand, the unilateral representation is extremely useful if our interest is in prediction purposes, or in spectral density estimation. The spectral density of the process x_{st} may estimated through least-squares estimation of the unilateral autoregressive representation. Autoregressive spectral estimation is well-established in time series, with roots in the contribution of Mann and Wald (1943). The advantages of autoregressive spectral estimation for time series were listed in Parzen (1969). These are enumerated in Chapter 5. The work of Akaike (1969) and Kromer (1970) established the techniques for this approach to estimating the spectrum with time series data. For spatial processes, this has been studied in a vast signal processing literature. Tjøstheim (1981) considers an autoregression defined unilaterally on a quarter plane and finds some evidence that autoregressive spectral estimation is superior to conventional spectral analysis methods. McClellan (1982) reviews seven different types of spectral estimates, the autoregressive estimator being one of them. Wester, Tummala, and Therrien (1990) propose iterative techniques to optimise computation of autoregressive estimates in both the half-plane and quarter-plane case.

However, both the time series and spatial literature mentioned in the preceding paragraph has assumes that the true model is a finite unilateral autoregression and it is rare that such an assumption can be justified, especially in view of representations such as (1.3.4). While it may be argued that truncated versions of (1.3.4) such as

$$x_{st} = 2\beta x_{s,t+1} - \beta^2 x_{s,t+2} - \beta^2 x_{s+1,t+1} + \beta\left(1-\beta^2\right)\sum_{j=0}^{k}\beta^j x_{s+1,t-j} + \epsilon'_{st}$$

need to be employed in practice, it is desirable to let $k \to \infty$ as the sample size increases to overcome the bias caused by using a finite autoregression. Even for time series, finite autoregressions may not capture the data generating process (DGP) and, for the regularly-spaced time series case, Berk (1974) provides results on the consistency and asymptotic normality of spectral density estimates with the order of the autoregression allowed to diverge with sample size. This approach has added appeal because any stationary, purely non-deterministic (in the linear prediction sense) time series has an infinite moving-average representation which, under invertibility conditions, yields an infinite autoregressive representation.

In Chapters 4 and 5, we extend Berk's consistency result to lattice processes. While we have already discussed that any multilateral process has a (possibly infinite) unilateral representation, Helson and Lowdenslager (1958, 1961) showed that even more generally all stationary, purely non-deterministic (in the linear prediction sense) spatial processes have a half-plane, infinite, moving-average representation. Again under invertibility conditions we can use this to write down an infinite autoregressive half-plane representation that is estimable by least-squares. As a result, there is strong motivation for an extension of the result of Berk (1974). We have seen that even for processes that already have a multilateral representation, the corresponding unilateral scheme may be infinite. This provides even greater reason to study the estimation of unilateral spatial autoregressions with diverging order in all dimensions. Extension is not straightforward, with complications arising due to the structure of the covariance matrix and the edge-effect.

1.4 Increasingly many parameters

While the preceding sections have demonstrated that the need for theory on models with increasingly many parameters arises quite naturally, this section summarises some of the key contributions in the general increasing parameter literature. This has been concentrated mostly in the statistical journals, even though the econometric implications are immediate. One of the earliest references to increasingly many parameters is Neyman and Scott (1948) who document the problem of incidental parameters potentially rendering maximum likelihood estimates inconsistent. In the regression context, the analysis of models with increasing dimension may be traced to Huber (1973). He considers the multiple regression model

$$y = X\beta + \epsilon \tag{1.4.1}$$

when the dimension of β is allowed to diverge with sample size. Let x_i denote the i-th column of X' and y_i the i-th element of the n-dimensional vector y. The M-estimate of β, denoted β_M, is the vector that solves

$$\sum_{i=1}^{n} x_i \psi\left(y_i - x_i'\beta\right) = 0,$$

where $\psi : \mathbb{R} \to \mathbb{R}$ is a given function. The least squares estimate is a particular case with $\psi(x) = x$. The asymptotic properties of β_M are then studied. This problem is considered further in Yohai and Maronna (1979) and Ringland (1983). The former showed that if the dimension of β is p, with $p \to \infty$, then $p^2/n \to 0$ is sufficient for

$$\|\beta_M - \beta\| \xrightarrow{p} 0, \tag{1.4.2}$$

where we employ Euclidean norm (see Section 1.6), and $p^{\frac{5}{2}}/n \to 0$ is sufficient for

$$a'\left(\beta_M - \beta\right) \xrightarrow{d} N(0,1), \tag{1.4.3}$$

where a is some appropriately bounded vector in $\mathbb{R}^p$. Portnoy (1984, 1985) improves the conditions for (1.4.2) and (1.4.3) to $p \log p/n \to 0$ and $(p \log p)^{\frac{3}{2}}/n$ respectively.

We have already mentioned the contribution of Berk (1974) to the time series increasing parameter literature. He proves that for an autoregression of order k with $k \to \infty$ as $n \to \infty$, least-squares estimates of autoregression coefficients are consistent and asymptotically normal in the sense of (1.4.2) and (1.4.3) if $k^2/n \to 0$ and the resulting spectral density estimate is consistent and asymptotically normal if $k^3/n \to 0$. Robinson (1979) establishes similar conditions for truncated approximations to systems with infinite distributed lags, but allows these conditions to vary with the strength of the assumptions on existence of moments for the errors. In Robinson (2003) simultaneous equation models with increasingly many equations are considered, which is equivalent to studying increasingly many coefficients on the endogenous and exogenous variables. It is shown that if the number of exogenous variables, m, is allowed to increase with n, then $m^2/n \to 0$ is sufficient for asymptotic normality of instrumental variable (IV) estimates of the parameters of a single equation nested in a system with increasingly many equations.

Another econometric example where the problem of increasingly many parameters arises is the panel data fixed-effects model, a spatial version of which is considered in Chapter 2. While we do not dwell on this in detail at this juncture, it should be mentioned that for such models Moreira (2009) has suggested a method based on using group actions and invariants (see also Eaton (1989)) to construct an objective function that is a function of a parameter of fixed dimension. The disadvantage of this approach is that the incidental parameters are treated purely as nuisance parameters and not actually estimated. In contexts such as the setting of Case (1991, 1992) discussed in Section 1.2, the incidental parameters are actually of interest and indeed tests of equality between them can be extremely useful in applied work.

1.5 Contributions of this thesis

This thesis makes several contributions. We list these by chapter. In Chapter 2, consistency and asymptotic normality of IV and OLS estimates in a SAR model with increasing autoregressive order and increasingly many regressors is considered. Permissible rates of growth of the parameter space relative to sample size are derived. This is more complicated than the model (1.4.1), due to the presence of spatially lagged y_n. In addition an empirical example illustrates a prescription for applied work: if the model design implies heterogeneity in spatial units then the spatial parameters should also reflect this.

Chapter 3 studies pseudo maximum likelihood estimates for the model considered in Chapter 2. The problem is challenging as it involves an implicitly defined estimate of a parameter of increasing dimension. A Monte Carlo study reveals that even the MLE suffers from finite-sample identification problems. Motivated by this, we also propose closed-form estimates obtained from a Newton-type step commencing from the IV and OLS estimates of Chapter 2. These are shown to have the same asymptotic distribution as the PMLE and their finite-sample properties are studied in a Monte Carlo experiment.

Chapters 4 and 5 concentrate on autoregressions defined on a regularly-spaced d-dimensional lattice. In particularly we focus on half-plane representations, which can be estimated by least-squares. Unlike in the time-series, unilateral representations of stationary processes do not yield a Toeplitz covariance matrix. Chapter 4 demonstrates that for spatial processes the covariance matrix may be nested inside a matrix which is block-Toeplitz with Toeplitz-blocks with $d-1$ levels of nesting. This contribution is important because the resulting analysis of eigenvalues can result in the kind of neat, unified asymptotic theory for spatial processes that Hannan (1973) derived for time series. This theory was derived by approaching the problem from the spectral domain. Indeed, in stationary time series the covariance matrix turns out to be approximately diagonalizable by a unitary matrix due to its Toeplitz structure, hence the favourable outcome of an elegant theory. Given that unilateral representations also result in a parameter-free Jacobean term there seems to be some scope for analogous results for spatial processes.

In Chapter 5 of this thesis, we exploit the structure derived in Chapter 4 to propose an autoregressive spectral density estimate for a stationary lattice process and prove that this is uniformly consistent under conditions that restrict the rate of growth of the autoregressive order in all dimensions. This is an important result due to the advantages of autoregressive spectral estimation listed in Section 1.3, and also because of the problems caused by the edge-effect in kernel-based spectral density estimation.

1.6 Some notation and definitions

We introduce some notation and definitions. These will be used throughout the thesis.

1. $1(\cdot)$ denotes the indicator function i.e.

$$1(x \in A) = \begin{cases} 1 & \text{if } x \in A; \\ 0 & \text{if } x \notin A. \end{cases}$$

2. For a generic $p \times p$ matrix A with real eigenvalues, the largest and smallest eigenvalues are denoted $\overline{\eta}(A)$ and $\underline{\eta}(A)$ respectively.

3. $\|\cdot\|$ denotes spectral norm i.e. for a generic real $p \times q$ matrix B,

$$\|B\| = \left\{\overline{\eta}(B'B)\right\}^{\frac{1}{2}}.$$

For vectors b we define Euclidean norm as $(b'b)^{\frac{1}{2}}$, so that spectral norm and Euclidean norm coincide for vectors.

4. For a generic real $p \times q$ matrix $B = [b_{ij}]$ we define

$$\|B\|_R = \max_{i=1,\ldots,p} \sum_{j=1}^{q} |b_{ij}|$$

and

$$\|B\|_C = \max_{j=1,\ldots,q} \sum_{i=1}^{p} |b_{ij}|,$$

which are the maximum absolute row-sum and column-sum norms respectively. If some W_{in} is row-normalized as in Section 1.2, then this implies that $\|W_{in}\|_R = 1$ if also W_{in} has non-negative elements.

5. $\|\cdot\|_F$ denotes the Frobenius norm i.e. i.e. for a generic real $p \times q$ matrix B

$$\|B\|_F = \left\{\sum_{i=1}^{p}\sum_{j=1}^{q} b_{ij}^2\right\}^{\frac{1}{2}}.$$

6. Throughout the thesis, C will denote a generic, arbitrarily large and positive constant that is independent of sample size, while c will denote a generic, arbitrarily small and positive constant that is independent of sample size.

7. *Consistency*: In this thesis, consistency of a parameter of increasing dimension is taken to mean consistency in Euclidean norm i.e. by the statement "θ^e is a consistent estimate of θ" we mean

$$\|\theta^e - \theta\| \xrightarrow{p} 0.$$

   Similarly, if we say that a matrix B of increasing dimensions can be consistently estimated by B^e we mean that

$$\|B^e - B\| \xrightarrow{p} 0.$$

2 IV and OLS estimation of higher-order SAR models

2.1 Introduction

In this chapter a version of (1.2.1) is considered where $p,\ k \to \infty$ as $n \to \infty$. This allows for more flexible modelling, in accordance with the idea that more parameters may be estimated as we increase the sample size, and explicitly permits asymptotic regimes prevalent in applied situations, as we illustrate later. Increasingly many parameters have been extensively studied in multiple regression, for instance by Huber (1973) and in a series of papers by Portnoy (1984, 1985). Berk (1974) and Robinson (2003) also studied problems with increasingly many parameters in time series autoregressions and simultaneous equations systems respectively. This literature has been discussed in Chapter 1.

In the next section, we introduce and discuss our model and also introduce some basic assumptions. Conditions and theorems for the consistency and asymptotic normality of least squares and instrumental variable (IV) estimates are presented in Section 2.3. In Section 2.4, we consider applications while Section 2.5 provides an empirical example. The proofs of the theorems and the sequences of lemmas that they rely on are left to appendices.

2.2 Model and basic assumptions

Given the existence of vectors $\lambda_{(n)} = (\lambda_{1n}, \ldots, \lambda_{p_n n})'$ and $\beta_{(n)} = (\beta_{1n}, \ldots, \beta_{k_n n})'$, where $'$ indicates transposition, we wish to model the $n \times 1$ observable vector $y_n = (y_{1n}, \ldots, y_{nn})'$ by the specification

$$y_n = \sum_{i=1}^{p_n} \lambda_{in} W_{in} y_n + X_n \beta_{(n)} + U_n \tag{2.2.1}$$

with $p_n \to \infty$ as $n \to \infty$, X_n an $n \times k_n$ matrix of constants with $k_n \to \infty$ as $n \to \infty$ and $U_n = (u_1, \ldots, u_n)'$ a vector of unobservable disturbances. We may rewrite (2.2.1) as

$$S_n y_n = X_n \beta_{(n)} + U_n \tag{2.2.2}$$

where $S_n = I_n - \sum_{i=1}^{p_n} \lambda_{in} W_{in}$ or equivalently $y_n = R_n \lambda_{(n)} + X_n \beta_{(n)} + U_n$ with $R_n = (W_{1n} y_n, \ldots, W_{p_n n} y_n)$. Note that in contrast to (1.2.1), in (2.2.1) we also allow the individual $\lambda_{(n)}$ and $\beta_{(n)}$ elements to vary with n as discussed in the previous section.

The model (2.2.1) cannot be considered a particular case of the models considered in the statistical literature surveyed in Chapter 1, due to the generation of y_n by a SAR model. Although Portnoy (1984, 1985) allowed his model to have stochastic regressors, these were not generated using a spatial process. In fact, in some sufficient conditions they were taken to be i.i.d.

Recently there has also been some interest in the estimation of spatial weight matrices, as opposed to assuming that they are exogenously chosen, see e.g. Bhattacharjee and Jensen-Butler (2013). A potential extension of the model considered in this chapter is to spatial weight matrix estimation, where each unit is influenced by a number of neighbours that increases slowly with sample size. In this case the quantities of interest are the elements of the weight matrices themselves, but these may be treated as linearly occurring parameters using suitable decompositions of the weight matrix/matrices, or a partitioning of the spatial domain.

We now introduce some basic assumptions.

Assumption 1. $U_n = (u_1, \ldots, u_n)'$ has iid elements with zero mean and finite variance σ^2.

Assumption 2. For $i = 1, \ldots, p_n$, the elements of W_{in} are uniformly $\mathscr{O}(1/h_n)$, where h_n is some positive sequence which may be bounded or divergent. If it is bounded, then it must also be bounded away from 0. The diagonal elements of each W_{in} are zero. We additionally assume that $n/h_n \to \infty$ as $n \to \infty$.

Different h_{in} sequences for each of the W_{in} may be used. However for least squares estimation, even for fixed p, Lee (2002) demonstrated that consistency requires divergence so that $\min_{i=1,\ldots,p_n} h_{in} \to \infty$ must be assumed and Assumption 2 entails no loss of generality. He also provides a detailed discussion of this assumption. In IV estimation, any mixture of bounded and divergent h_{in} sequences may be employed. However boundedness away from zero is crucial as even consistency of the error variance estimate based on IV residuals may fail if this does not hold. Indeed, an interpretation of h_n is that it is the number of neighbours of a unit and it is rather odd to allow this to go to zero as the sample size increases. The diagonal elements being zero implies that a unit is not regarded as its own neighbour.

Assumption 3. S_n is non-singular for sufficiently large n.

This assumption ensures that (2.2.2) has a solution for y_n. If the W_{in} happen to be block diagonal with a single non-zero block such that $\|W_{in}\|_R \leq 1$ for $i = 1, \ldots, d$, then we prove in Appendix 2.D that a sufficient condition for S_n to be non-singular is that $|\lambda_{in}| < 1$ for $i = 1, \ldots, d$. Such a situation is discussed in Section 2.4.1.

Assumption 4. $\left\|S_n^{-1}\right\|_R$, $\left\|S_n^{-1}\right\|_C$, $\|W_{in}\|_R$ and $\|W_{in}\|_C$ are uniformly bounded in n and i for all $i = 1, \ldots, p_n$ and sufficiently large n.

This assumption has its provenance in Kelejian and Prucha (1998). The parts pertaining to S_n^{-1} ensure that the spatial correlation is curtailed to a manageable degree because the covariance matrix of y_n is $\sigma^2 S_n^{-1} S_n'^{-1}$. The assumptions on the W_{in} are satisfied trivially if one unit is assumed to be a 'neighbour' of only a finite number of other units, and is also satisfied if a unit is a neighbour of infinitely many units as long as the $w_{ij,n}$ decline fast enough. The latter is natural if the $w_{ij,n}$ are decreasing functions of some measure of distance between units.

Assumption 5. The elements of X_n are constants and are uniformly bounded in n, in absolute value, for all sufficiently large n.

The assumption of non-stochastic regressors has been fairly standard in the theoretical spatial econometrics literature dealing with OLS estimation and the PMLE, see e.g. Lee (2002) and Lee (2004). In Kelejian and Prucha (1999) all expectations are to be read as conditional on the realisations of the explanatory variables, and so the regressors are treated as fixed in their theory. Assumption 5 is certainly strong, but we opt for it as the main purpose of this chapter is to study the implications of the increasing order of the SAR model. A similar discussion applies to Assumption 6 in the next section.

2.3 Consistency and asymptotic normality

2.3.1 IV estimation

Because of the endogeneity of the $W_{in}y_n$, $i = 1, \ldots, p_n$, IV estimation has been employed for estimation of SAR models. Let Z_n be an $n \times r_n$ matrix of instruments, with $r_n \geq p_n$ for all n and introduce

Assumption 6. The elements of Z_n are constants and are uniformly bounded in absolute value.

For the model (1.2.1) with $p = 1$, Kelejian and Prucha (1998) noted that $W_n \mathbb{E}(y_n)$ can be written as an infinite linear combination of the columns of the matrices

$$X_n, W_n X_n, W_n^2 X_n, \ldots,$$

assuming the existence of a convergent power (Neumann) series for $(I_n - \lambda W_n)^{-1}$. The existence of such a series is guaranteed if $\|\lambda W_n\| < 1$. It was suggested that the instrument matrix be constructed from linearly independent subsets of the columns of

$$X_n, W_n X_n, W_n^2 X_n, \ldots, W_n^q X_n,$$

where in principle $q \to \infty$ as $n \to \infty$ but $q = 2$ was regarded as sufficient from Monte Carlo experiments. Our theory allows the number of instruments to increase with sample size and provides a new result for the case when p_n is fixed while r_n is allowed to diverge with n. For the specification (2.2.1), we will have

$$\begin{aligned} \mathbb{E}(y_n) &= \left(I_n - \sum_{i=1}^{p_n} \lambda_{in} W_{in}\right)^{-1} X_n \beta_{(n)} \\ &= \left[\sum_{k=0}^{\infty} \left(\sum_{i=1}^{p_n} \lambda_{in} W_{in}\right)^k\right] X_n \beta_{(n)}, \end{aligned} \tag{2.3.1}$$

assuming that the power series is well-defined, so that instruments may be constructed as subsets of the linearly independent columns of

$$X_n, W_{1n} X_n, W_{1n}^2 X_n, \ldots, W_{2n} X_n, W_{2n}^2 X_n, \ldots, W_{p_n n} X_n, W_{p_n n}^2 X_n, \ldots \tag{2.3.2}$$

Columns of X_n pre-multiplied by cross-products of the W_{in} may also be employed in view of (2.3.1). Of course, other choices of instruments from outside the model are available to the practitioner depending on the problem under consideration.

We now provide sufficient conditions for the power series in (2.3.1) to be well-defined. A sufficient condition is

$$\left\| \sum_{i=1}^{p_n} \lambda_{in} W_{in} \right\| < 1, \tag{2.3.3}$$

for which either

$$\left(\max_{i=1,\ldots,p_n;\, n \geq 1} |\lambda_{in}| \right) \left\| \sum_{i=1}^{p_n} W_{in} \right\| < 1, \tag{2.3.4}$$

or

$$\left(\max_{i=1,\ldots,p_n;\, n \geq 1} \|W_{in}\| \right) \sum_{i=1}^{p_n} |\lambda_{in}| < 1 \tag{2.3.5}$$

suffices. When the W_{in} take the form (1.2.4), then $\sum_{i=1}^{p_n} W_{in} = W_n$ as given in (1.2.7). B_m as defined in (1.2.5) has one eigenvalue equal to 1 and also $-1/(m-1)$ as an eigenvalue with multiplicity $m-1$. Hence $\|W_n\| = \|W_{in}\| = \|B_m\| = 1$, $i = 1, \ldots, p_n$, and $\max_{i=1,\ldots,p_n;\, n \geq 1} |\lambda_{in}| < 1$ is sufficient for the power series to be valid, by (2.3.4). See also Proposition 2.7 in Appendix 2.D for an equivalent result. The condition from (2.3.5) is much stronger in this setting, requiring that $\sum_{i=1}^{p_n} |\lambda_{in}| < 1$.

Denoting $\theta_{(n)} = \left(\lambda'_{(n)}, \beta'_{(n)}\right)$, define the IV estimate of $\theta_{(n)}$ as

$$\hat{\theta}_{(n)} = \hat{Q}_n^{-1} \hat{K}'_n J_n^{-1} \hat{k}_n, \tag{2.3.6}$$

with

$$\hat{Q}_n = \hat{K}'_n J_n^{-1} \hat{K}_n$$

where

$$\hat{K}_n = \frac{1}{n} \begin{bmatrix} Z'_n \\ X'_n \end{bmatrix} [R_n, X_n],\ \hat{k}_n = \frac{1}{n} \begin{bmatrix} Z'_n \\ X'_n \end{bmatrix} y_n,\ J_n = \frac{1}{n} \begin{bmatrix} Z'_n \\ X'_n \end{bmatrix} [Z_n, X_n].$$

This implies that

$$\hat{\theta}_{(n)} - \theta_{(n)} = \hat{Q}_n^{-1} \hat{K}'_n J_n^{-1} q_n,$$

where

$$q_n = \frac{1}{n} \begin{bmatrix} Z'_n \\ X'_n \end{bmatrix} U_n.$$

Since (2.2.2) and Assumption 3 imply that $y_n = S_n^{-1} X_n \beta_{(n)} + S_n^{-1} U_n$ we can write $R_n = A_n + B_n$ where

$$A_n = (G_{1n} X_n \beta_{(n)}, \ldots, G_{p_n n} X_n \beta_{(n)}),\ B_n = (G_{1n} U_n, \ldots, G_{p_n n} U_n),$$

and $G_{in} = W_{in} S_n^{-1}$ for $i = 1, \ldots, p_n$. Also define

$$K_n = \frac{1}{n} \begin{bmatrix} Z'_n \\ X'_n \end{bmatrix} [A_n, X_n],\ Q_n = K'_n J_n^{-1} K_n,\ L_n = \frac{1}{n} \begin{bmatrix} A'_n \\ X'_n \end{bmatrix} [A_n, X_n].$$

Note that J_n and L_n are symmetric matrices.

Introduce the following assumptions.

Assumption 7. $\overline{\lim}_{n\to\infty} \overline{\eta}(J_n) < \infty$ and $\underline{\lim}_{n\to\infty} \underline{\eta}(K'_n K_n) > 0$.

Assumption 8. $\underline{\lim}_{n\to\infty} \underline{\eta}(J_n) > 0$ and $\overline{\lim}_{n\to\infty} \overline{\eta}(K'_n K_n) < \infty$.

These are asymptotic non-multicollinearity and finiteness conditions, which can to some extent be checked as we discuss in the next sub-section.

Lemma 2.1. *Under Assumptions 7 and 8 respectively*

(i) $\underline{\lim}_{n\to\infty} \underline{\eta}\,(Q_n) > 0$.

(ii) $\overline{\lim}_{n\to\infty} \overline{\eta}\,(Q_n) < \infty$.

For just-identified (i.e. IV) estimation, we have $p_n = r_n$ implying that $\hat{K}_n$ and K_n are square matrices so that $\hat{\theta}_{(n)} = \hat{K}_n^{-1}\hat{k}_n$ and $Q_n^{-1} = K_n^{-1} J_n K_n'^{-1}$.

Theorem 2.1. *Let Assumptions 1-7 hold. Suppose also that*

$$\frac{1}{p_n} + \frac{1}{r_n} + \frac{1}{k_n} + \frac{p_n\left(r_n + k_n\right)}{n} \to 0 \text{ as } n \to \infty. \tag{2.3.7}$$

Then

$$\left\| \hat{\theta}_{(n)} - \theta_{(n)} \right\| \xrightarrow{p} 0.$$

Condition (2.3.7) details the restrictions on the rate of growth of the number of instruments and regressors, and implies a restriction on the rate of growth of the parameter space because $p_n \leq r_n$. Slightly weakened conditions yield the same result for the just identified case $p_n = r_n$.

Corollary 2.2. *Suppose $p_n = r_n$. Let Assumptions 1-6 hold. Suppose also that*

$$\varliminf_{n\to\infty} \underline{\eta}(K_n' K_n) > 0$$

and

$$\frac{1}{p_n} + \frac{1}{k_n} + \frac{p_n\left(p_n + k_n\right)}{n} \to 0 \text{ as } n \to \infty.. \tag{2.3.8}$$

Then

$$\left\| \hat{\theta}_{(n)} - \theta_{(n)} \right\| \xrightarrow{p} 0.$$

The error variance may be estimated using the natural estimate

$$\hat{\sigma}^2_{(n)} = \frac{1}{n}\left(y_n - \left(R_n, X_n\right)\hat{\theta}_{(n)}\right)'\left(y_n - \left(R_n, X_n\right)\hat{\theta}_{(n)}\right). \tag{2.3.9}$$

Assumption 9. $\varlimsup_{n\to\infty} \overline{\eta}(L_n) < \infty$.

Theorem 2.2. *Let Assumptions 1-7 and 9 hold. Suppose also that*

$$\frac{1}{p_n} + \frac{1}{r_n} + \frac{1}{k_n} + \frac{\left(p_n + k_n\right)\left(r_n + k_n\right)}{n} \to 0 \text{ as } n \to \infty. \tag{2.3.10}$$

Then

$$\hat{\sigma}^2_{(n)} \xrightarrow{p} \sigma^2.$$

A similar theorem hold in the just identified case $p_n = r_n$ but we omit the statement for brevity. Here the requirement that h_n be bounded away from zero if it is bounded is crucial (see (2.B.8)), with consistency possibly failing otherwise. We can also record a central limit theorem for finitely many arbitrary linear combinations of $\hat{\theta}_{(n)} - \theta_{(n)}$ under stronger conditions which restrict the growth of p_n and r_n relative to n further.

Theorem 2.3. *Let Assumptions 1-9 hold. Suppose also that*

$$\frac{1}{p_n} + \frac{1}{k_n} + \frac{1}{r_n} + \frac{p_n\left(r_n^2 + k_n^2\right)}{n} + \frac{k_n\left(r_n + k_n\right)}{n} \to 0 \text{ as } n \to \infty. \tag{2.3.11}$$

Then, for any $s \times (p_n + k_n)$ matrix of constants Ψ_n with full row-rank,

$$\frac{n^{\frac{1}{2}}}{(p_n + k_n)^{\frac{1}{2}}} \Psi_n \left(\hat{\theta}_{(n)} - \theta_{(n)}\right) \xrightarrow{d} N\left(0, \lim_{n\to\infty} \frac{\sigma^2}{p_n + k_n} \Psi_n Q_n^{-1} \Psi_n'\right),$$

where the asymptotic covariance matrix exists, and is positive definite, by Lemma 2.1. It may be consistently estimated by

$$\frac{\hat{\sigma}_{(n)}^2}{p_n + k_n} \Psi_n \hat{Q}_n^{-1} \Psi_n'.$$

Corollary 2.3. *Suppose $p_n = r_n$. Let Assumptions 1-6, 8 and 9 hold. Suppose also that*

$$\varliminf_{n\to\infty} \underline{\eta}(K_n' K_n) > 0$$

and

$$\frac{1}{p_n} + \frac{1}{k_n} + \frac{p_n^3}{n} + \frac{p_n k_n^2}{n} \to 0 \text{ as } n \to \infty. \tag{2.3.12}$$

Then, for any $s \times (p_n + k_n)$ matrix of constants Ψ_n with full row-rank,

$$\frac{n^{\frac{1}{2}}}{(p_n + k_n)^{\frac{1}{2}}} \Psi_n \left(\hat{\theta}_{(n)} - \theta_{(n)}\right) \xrightarrow{d} N\left(0, \lim_{n\to\infty} \frac{\sigma^2}{p_n + k_n} \Psi_n K_n^{-1} J_n K_n'^{-1} \Psi_n'\right),$$

where the asymptotic covariance matrix exists, and is positive definite, by Assumptions 7 and 8. It may be consistently estimated by

$$\frac{\hat{\sigma}_{(n)}^2}{p_n + k_n} \Psi_n \hat{K}_n^{-1} J_n \hat{K}_n'^{-1} \Psi_n'.$$

Note that in Theorem 2.3 the condition $p_n r_n^2/n \to 0$ implies $p_n k_n^2/n \to 0$ as long as $k_n = \mathscr{O}(r_n)$ i.e. the number of instruments and regressors increase at the same rate. In particular if r_n is fixed (implying that p_n is fixed), $k_n = \mathscr{O}(r_n)$ is not satisfied unless k_n is also fixed. Similarly $r_n k_n/n \to 0$ implies $k_n^2/n \to 0$ if $k_n = \mathscr{O}(r_n)$.

The $n^{\frac{1}{2}}/(p_n + k_n)^{\frac{1}{2}}$-norming is needed to ensure a finite asymptotic covariance matrix, and implies a slower than $n^{\frac{1}{2}}$ rate of convergence due to the increasing parameter space dimension, while conditions (2.3.11) and (2.3.12) restrict the growth of the parameter space. Indeed, if only $n^{\frac{1}{2}}$-norming was employed the rows of Ψ_n would have to be assumed to have uniformly bounded norm which implies a similar normalisation as these rows have increasing dimension. The norming can change if the rows of Ψ_n contain many zero elements, indeed the number of non-zero elements can even be allowed to increase at a rate slower than the rate of increase of the parameters. In particular, Theorem 2.3 may be easily rewritten if the interest is in obtaining a central limit theorem for a fixed number of the parameters rather than an increasing number. Suppose without loss of generality that we are interested in, say, the first l elements of $\theta_{(n)}$. In this case we take Ψ_n to be the $1 \times (p_n + k_n)$ row vector with all elements after the l-th entry equal to zero. We then recover a $n^{\frac{1}{2}}$-consistency result.

Corollary 2.4.

(i) Let Assumptions 1-9 hold. Suppose also that (2.3.11) holds. Then

$$n^{\frac{1}{2}}\left(\hat{\theta}_{(n)} - \theta_{(n)}\right)_l \xrightarrow{d} N\left(0, \sigma^2\left(\lim_{n\to\infty} Q_n\right)_l^{-1}\right),$$

where $\left(\hat{\theta}_{(n)} - \theta_{(n)}\right)_l$ *denotes the first* l *elements of* $\hat{\theta}_{(n)} - \theta_{(n)}$ *while the top-left* $l \times l$ *block of* $\left(\lim_{n\to\infty} Q_n\right)^{-1}$ *is denoted* $\left(\lim_{n\to\infty} Q_n\right)_l^{-1}$*. The existence of the limit is guaranteed by Lemma* 2.1*.*

(ii) Suppose $p_n = r_n$*. Let Assumptions 1-6, 8 and 9 hold. Suppose also that*

$$\lim_{n\to\infty} \underline{\eta}(K_n' K_n) > 0$$

and (2.3.12) hold. Then

$$n^{\frac{1}{2}}\left(\hat{\theta}_{(n)} - \theta_{(n)}\right)_l \xrightarrow{d} N\left(0, \sigma^2\left[\left(\lim_{n\to\infty} K_n\right)^{-1} \lim_{n\to\infty} J_n \left(\lim_{n\to\infty} K_n'\right)^{-1}\right]_l\right),$$

where $\left[\left(\lim_{n\to\infty} K_n\right)^{-1} \lim_{n\to\infty} J_n \left(\lim_{n\to\infty} K_n'\right)^{-1}\right]_l$ *denotes the top-left* $l \times l$ *block of* $\left(\lim_{n\to\infty} K_n\right)^{-1} \lim_{n\to\infty} J_n \left(\lim_{n\to\infty} K_n'\right)^{-1}$*. The asymptotic covariance matrices are estimated as in Theorem 2.3. The existence of the limit is guaranteed by Assumptions 7 and 8.*

Corollary 2.4 indicates that the definition of simple t-statistics do not change from the fixed-dimension model (1.2.1) to (2.2.1).

2.3.2 Least squares estimation

Define the OLS estimate of $\theta_{(n)}$ as

$$\tilde{\theta}_{(n)} = \hat{L}_n^{-1} \hat{l}_n, \tag{2.3.13}$$

where

$$\hat{L}_n = \frac{1}{n}\begin{bmatrix} R_n' \\ X_n' \end{bmatrix}[R_n, X_n], \;\; \hat{l}_n = \frac{1}{n}\begin{bmatrix} R_n' \\ X_n' \end{bmatrix} y_n$$

so

$$\tilde{\theta}_{(n)} - \theta_{(n)} = \hat{L}_n^{-1} w_n,$$

with

$$w_n = \frac{1}{n}\begin{bmatrix} R_n' \\ X_n' \end{bmatrix} U_n.$$

Analogous to the IV case, we also have an asymptotic non-multicollinearity condition given by

Assumption 10. $\varliminf_{n\to\infty} \underline{\eta}(L_n) > 0$.

This can be checked under more primitive conditions. For instance, if X_n contains a column of ones (i.e. the model (2.2.1) has an intercept) and there exists a row-normalised W_{in} with equal off-diagonal elements (such as (1.2.4) defined below) then $W_{in}y_n$ is asymptotically collinear with the intercept. In this case Assumption 10 fails, and in fact so does $\varliminf_{n\to\infty} \underline{\eta}(K_n'K_n) > 0$. This problem is discussed further in Kelejian and Prucha (2002). A necessary condition for both Assumption 10 and $\varliminf_{n\to\infty} \underline{\eta}(K_n'K_n) > 0$ to hold is that, for all $i = 1, \ldots, p_n$, W_{in} are linearly independent for sufficiently large n, failing which some of the λ_{in} are not identified. It is clear that identification of the λ_{in} is particularly transparent when the W_{in} have a single non-zero block structure, a situation that will be discussed in detail in Section 2.4.

Theorem 2.4. *Let Assumptions 1-5 and 10 hold. Suppose also that*

$$\frac{1}{p_n} + \frac{1}{k_n} + \frac{p_n k_n^2 (p_n + k_n)}{n} + \frac{p_n}{h_n} \to 0 \text{ as } n \to \infty. \tag{2.3.14}$$

Then

$$\left\| \tilde{\theta}_{(n)} - \theta_{(n)} \right\| \xrightarrow{p} 0.$$

Lee (2002) demonstrated consistency of least-squares parameter estimates for the model (1.2.1), for $p = 1$, when $h_n \to \infty$. This condition ensures that the endogeneity problem discussed above vanishes asymptotically. Our condition (2.3.14) is suitably strengthened to also account for the increasing p_n and k_n. To obtain a central limit theorem, we additionally assume

Assumption 11. $\mathbb{E}\left(u_i^4\right) \le C$ for $i = 1, \ldots, n$.

While finite fourth order moments are not required for consistency, they are needed to prove asymptotic normality. The details are in Appendix 2.D, but briefly this is because

$$\left\| \frac{1}{n} \begin{bmatrix} B_n' \\ 0 \end{bmatrix} U_n \right\| = o_p(1)$$

under both second and fourth order moments, but if only second order moments are employed then the stochastic order of the last displayed expression is such that no normalisation factor is available to ensure a non-degenerate asymptotic distribution. We first introduce the error variance estimate

$$\tilde{\sigma}^2_{(n)} = \frac{1}{n}\left(y_n - (R_n, X_n)\,\tilde{\theta}_{(n)}\right)'\left(y_n - (R_n, X_n)\,\tilde{\theta}_{(n)}\right). \tag{2.3.15}$$

Theorem 2.5. *Let Assumptions 1-5 and 9-11 hold. Suppose also that*

$$\frac{1}{p_n} + \frac{1}{k_n} + \frac{p_n k_n^2 \left(p_n + k_n\right)}{n} + \frac{p_n}{h_n} \to 0 \text{ as } n \to \infty. \tag{2.3.16}$$

Then

$$\tilde{\sigma}^2_{(n)} \xrightarrow{p} \sigma^2.$$

Theorem 2.6. *Let Assumptions 1-5 and 9-11 hold. Suppose also that*

$$\frac{1}{p_n} + \frac{1}{k_n} + \frac{p_n^2 k_n^4 \left(p_n + k_n\right)}{n} + n^{\frac{1}{2}} \frac{p_n^{\frac{1}{2}}}{h_n} \to 0 \text{ as } n \to \infty. \tag{2.3.17}$$

Then, for any $s \times (p_n + k_n)$ matrix of constants Ψ_n with full row-rank,

$$\frac{n^{\frac{1}{2}}}{\left(p_n + k_n\right)^{\frac{1}{2}}} \Psi_n \left(\tilde{\theta}_{(n)} - \theta_{(n)}\right) \xrightarrow{d} N\left(0, \lim_{n\to\infty} \frac{\sigma^2}{p_n + k_n} \Psi_n L_n^{-1} \Psi_n'\right).$$

The asymptotic covariance matrix exists, and is positive definite, by Assumptions 9 and 10. It may be estimated consistently using

$$\frac{\tilde{\sigma}^2_{(n)}}{p_n + k_n} \Psi_n \hat{L}_n^{-1} \Psi_n'.$$

Corollary 2.5. *Let Assumptions 1-5 and 9-11 hold. Suppose also that (2.3.17) holds. Then*

$$n^{\frac{1}{2}} \left(\tilde{\theta}_{(n)} - \theta_{(n)}\right)_l \xrightarrow{d} N\left(0, \sigma^2 \left(\lim_{n\to\infty} L_n\right)_l^{-1}\right),$$

where $\left(\tilde{\theta}_{(n)} - \theta_{(n)}\right)_l$ denotes the first l elements of $\hat{\theta}_{(n)} - \theta_{(n)}$ while the top-left $l \times l$ block of $\left(\lim_{n\to\infty} L_n\right)^{-1}$ is denoted $\left(\lim_{n\to\infty} L_n\right)_l^{-1}$ and the asymptotic covariance matrix is estimated as in Theorem 2.6. The existence of the asymptotic covariance matrix is *guaranteed by Assumptions 9 and 10.*

2.4 Applications

SAR models have found widespread application in many situations where cross-sectional dependence has to be modelled for units observed with irregular spacing. A general attitude adopted by modellers is that the more data we have, the more parameters we can hope to estimate with reasonable precision. The asymptotic theory presented above takes this into consideration. While the allowance for the number of parameters to increase as $n \to \infty$ can be rather theoretical we show in this section that there at least two classes of SAR models where the need for such theory arises naturally. We also present an illustration of when this type of theory may be relevant, even though the model does not give rise to increasingly many parameters by its very design.

2.4.1 Farmer-district type models

The setting of Case (1991, 1992) was discussed in Chapter 1 as a natural motivator for the work in this chapter. From an applied point of view a parsimonious model may be quite desirable, and so some districts can be allowed to have the same λ_is on the basis of some homogeneity which will vary with application. There are other reasons to allow for a slower increase of λ_is than with r. For instance consider the condition $p_n^3/n \to 0$ (we keep k_n fixed for simplicity). In this setting this translates into requiring that $r^2/m \to 0$. For finite samples an approximation to this would be that the ratio r^2/m be small, but this may not be reasonable if, say, $r = 10$ and $m = 100$. It would be natural then to allow a slower increase of the parameter space than r, and attempts can be made to combine λ_is to reduce the ratio r^2/m. Combinations can be made according to geography, demographics or other criteria based on the context.

2.4.2 Panel data SAR models with fixed effects

Consider a balanced spatial panel data set with N observations in each of T individual panels, so that the sample size is $n = NT$. Let $y_{t,N}$ be the $N \times 1$ vector of observations on the dependent variable for the t-th panel, where t may correspond to a time period or a more general spatial unit like a school, village or district. Also let $X_{t,N}$ and F_N be $N \times k_1$ and $N \times k_2$ matrices of regressors respectively. $X_{t,N}$ contains panel-varying regressors while F_N does not. Let W_{iN}, $i = 1, \dots, p$, be a set of spatial weight matrices and consider the model

$$y_{t,N} = l_N \alpha_t + X_{t,N}\beta + F_N \gamma_t + \sum_{i=1}^{p} \lambda_i W_{iN} y_{t,N} + U_{t,N}, \quad t = 1, \dots, T \tag{2.4.1}$$

where $U_{t,N}$ is the $N \times 1$ vector of disturbances for each panel, which we take to be formed of iid components. The α_t, $t = 1, \ldots, T$, are scalar parameters that control for fixed effects with respect to panels, the λ_i, $i = 1, \ldots, p$, are scalar spatial autoregressive parameters and β is a $k_1 \times 1$ panel-invariant parameter vector. On the other hand γ_t is a $k_2 \times 1$ parameter vector that varies over panels. For this reason, the variables in F_N may be thought of as controlling for 'quasi' fixed-effects. Denote $y_n = \left(y_{1,n}', \ldots, y_{T,n}'\right)'$, $X_n = \left(X_{1,n}', \ldots, X_{T,n}'\right)'$, $U_n = \left(U_{1,n}', \ldots, U_{T,n}'\right)'$, $\alpha = (\alpha_1, \ldots, \alpha_T)'$ and $\gamma = (\gamma_1, \ldots, \gamma_T)'$. We can then stack (2.4.1) to obtain

$$y_n = (I_T \otimes l_N)\,\alpha + X_n\beta + (I_T \otimes F_N)\,\gamma + \sum_{i=1}^{p} \lambda_i\,(I_T \otimes W_{iN})\,y_n + U_n. \tag{2.4.2}$$

This model is an extension of that considered in Kelejian, Prucha, and Yuzefovich (2006), and was employed by Yuzefovich (2003). The latter is used as the basis for the empirical example we consider below. The former noted that the above model is again subject to asymptotic multicollinearity between the 'constant' and spatial lags if any of the W_{iN} have equal elements. We allow both $T \to \infty$ and $N \to \infty$ for our asymptotic theory, while they only allowed the latter. This implies that the number of regression parameters in (2.4.2) increases asymptotically. Not only this, since the $I_T \otimes W_{iN}$ are block diagonal it would be natural to fear that spatial autoregressive parameters differ for each panel, or at least among subsets of the panels. To illustrate, suppose for the moment that $p = 1$. Allowing a separate spatial parameter for each panel implies the model

$$y_n = (I_T \otimes l_N)\,\alpha + X_n\beta + (I_T \otimes F_N)\,\gamma + \sum_{i=1}^{T} \lambda_i W_N^i y_n + U_n \tag{2.4.3}$$

where

$$W_N^i = diag\left[0, \ldots, \underbrace{W_N}_{i^{th}\text{diagonal block}}, \ldots, 0\right].$$

The model (2.4.3) has $k_1 + T(k_2 + 1)$ regression parameters and T spatial parameters, making it fit naturally into the asymptotic regime discussed in Section 2.3. As in Section 2.4.1 a point of concern may be that conditions such as $p_n^3 k_n^4$ diverging slower than n (needed for asymptotic normality of least squares estimation in Section 2.3.2) translate here into requiring that

$$\frac{T^6}{N} \to 0 \text{ as } N, T \to \infty. \tag{2.4.4}$$

In finite samples we would like the above ratio to be somewhat small, but this may be impossible to achieve. For even $T = 2$, $T^6 = 64$ which may not be small compared to N. A solution is to use a smaller number of spatial parameters in (2.4.3), thereby allowing the number of spatial parameters to increase more slowly with T. For example, if t represents monthly observations we may allow the spatial parameters to change on a quarterly basis so that we have $T/4$ spatial parameters, assuming that T is divisible by 4 for simplicity. Then we would need

$$\frac{T^6}{256N} \to 0 \text{ as } N, T \to \infty$$

as opposed to (2.4.4). The last two displayed conditions are asymptotically the same but in finite samples it is more likely that the last displayed ratio is small.

2.4.3 Another illustration

In Kolympiris, Kalaitzandonakes, and Miller (2011), the authors attempt to explain the level of venture capital funding (provided by venture capital firms (VCFs)) for dedicated biotechnology firms (DBFs) with a SAR model. In particular, the hypotheses are that the level of VC funding for a DBF increases with the number of VCFs located in close proximity to the DBF and with the number of other DBFs located in close proximity to the DBF. To model this, specification (1.2.1) is employed, with the dependent variable being defined as the natural logarithm of the amount invested by VCFs in each of the $n = 816$ observed DBFs. The spatial weight matrices are defined using a binary neighbourhood criterion and then row-normalised. In particular, three weight matrices are employed (i.e. $p = 3$) with each based on a 3 sequential 10-mile rings from the origin DBF. The set of DBFs situated less than 10 miles from the origin DBF are considered one set of neighbours, those situated 10.1-20 miles from the origin form the second set and the third set of neighbours is defined in the obvious way. Their model also has $k = 21$, including an intercept term. The asymptotic multicollinearity problem caused by an intercept that was discussed after Assumption 10 does not arise here because the weight matrices have unequal off diagonal elements in general. Because the number of neighbours may be taken to increase with sample size, least squares is used to estimate the model. The results indicate that that only the first spatial lag of y_n, corresponding to those DBFs situated less than 10 miles from the origin DBF, is significant. Our theory is relevant here, since if data on more DBFs were to become available it would be attractive to reduce the radius of the rings used in defining neighbours. As discussed earlier, more parsimonious specifications such as the original may still be attractive to the practitioner but various models can be employed and relevant statistical tests run to arrive at a more informed choice.

2.5 Empirical example: A spatial approach to estimating contagion

The purpose of this section is to provide a practical example where the theory we have presented may be useful, and the new approach that we have suggested may lead to different conclusions from the empirical evidence on hand. It is, however, not intended to be a detailed econometric study of the problem. Yuzefovich (2003) carries out a study of 'contagion' of financial crises using (2.4.2), improving upon the treatment of Hernández and Valdés (2001). He studies the aftermath of three financial crises viz. the Asian, Russian and Brazilian crises of July 1997, August 1998 and January 1999. The idea is to identify channels of contagion using weekly stock market returns as the dependent variable. In particular, it is proposed that the stock market return of a country in a given week is determined by a set of fixed effects, exogenous variables (common shocks) and also a weighted average of returns of other countries in the same time period.

Four kinds of spatial weight matrices are employed, so that we have $p = 4$ in (2.4.2). Each reflects a different channel of contagion. The first weight matrix reflects how country i is connected to country j through bilateral trade, measured by exports from country i to country j. It is row-normalised, so denoting exports from country i to country j as $Exports_{ij}$ we have

$$w_{ij}^{Trade} = \frac{Exports_{ij}}{\sum_{h=1}^{n} Exports_{ih}}.$$

The second channel of contagion is financial links, measured by competition for funds from a common lender. The common lenders are defined as the three financial centres given by the set $C = \{Europe,\ Japan,\ US\}$. Define

$$d_{ij}^{C} = 2\frac{\min\{b_{j,C}, b_{i,C}\}}{b_j + b_i}$$

where $b_{i,C}$ is the debt of country i to common lender C and $b_i = \sum_C b_{i,C}$, i.e. the total foreign debt of country i. The financial links matrix is defined as

$$w_{ij}^{Fin} = \frac{1}{3}\sum_{C}\frac{d_{ij}^{C}}{\sum_{h=1}^{N} d_{ih}^{C}}.$$

The third weight matrix is a similarity in risk matrix, defined as

$$w_{ij}^{Sim} = \frac{\exp\left(-|x_i - x_j|\right)}{\sum_{h=1}^{N}\exp\left(-|x_i - x_h|\right)}$$

where x_i is a measure of the credit rating for country i. Finally, the fourth spatial weight matrix is a row-normalized neighbourhood matrix. Countries are divided into five regions: Europe, South and South-East Asia, Latin America, Middle East and North Africa, and Sub-Saharan Africa (see Table 2.3 for a list of countries by region). The neighbourhood matrix is defined as

$$w_{ij}^{Nbd} = \frac{\delta_{ij}}{\sum_{h=1}^{N} \delta_{ih}}$$

where δ_{ij} takes the value 1 when country i and country j belong to the same region, and 0 otherwise. Yuzefovich (2003) demonstrates that all the weight matrices are absolutely bounded in row and column sums. The diagonal elements are also normalised to zero. The common shocks are of two types: those propagated through trade linkages (in F_N) and those through financial linkages (in $X_{t,N}$). Suppose that

$$F_N = (f_1, \ldots, f_N)', \; X_{t,N} = (x_{t,1}, \ldots, x_{t,N})'.$$

Then the common shocks used are

$$\begin{aligned} f_i' &= \left(\frac{Exports_{i,Europe}}{GDP_i}, \frac{Exports_{i,Japan}}{GDP_i}, \frac{Exports_{i,US}}{GDP_i} \right) \\ x_{t,i}' &= \left(\frac{b_{i,Europe}}{GDP_i} y_{Europe,t}, \frac{b_{i,Japan}}{GDP_i} y_{Japan,t}, \frac{b_{i,US}}{GDP_i} y_{US,t} \right) \end{aligned}$$

where $Exports_{i,C}$ and $b_{i,C}$ are the exports from country i to financial centre C while $y_{C,t}$ is the weekly stock market return in financial centre C at time t. Yuzefovich (2003) suggests that the $x_{t,i}$ may be endogenous but a Hausman test conducted by him indicates otherwise and we treat them as exogenous, as he does in his final specification. The model estimated is therefore

$$\begin{aligned} y_n &= (I_T \otimes l_N)\,\alpha + X_n\beta + (I_T \otimes F_N)\,\gamma + \lambda_1 \left(I_T \otimes W^{Trade}\right) y_n \\ &+ \lambda_2 \left(I_T \otimes W^{Fin}\right) y_n + \lambda_3 \left(I_T \otimes W^{Sim}\right) y_n \\ &+ \lambda_4 \left(I_T \otimes W^{Nbd}\right) y_n + U_n \end{aligned} \tag{2.5.1}$$

with $k_1 = k_2 = 3$.

We will restrict ourselves to the Russian crisis in this analysis, and concentrate on the spatial autoregressive parameters. We consider a 12 week period starting in July 1998 as our sampling period. Yuzefovich (2003) found that only the spatial lag corresponding to the financial links matrix is statistically significant (at the 5% level) and we arrive at the same conclusion when replicating his results (see Table 2.1). There is some quantitative difference because data for 2 out of the 52 countries he used was unavailable. It is also possible that he used different stock market indices as compared to us, since these were not specified by him.

λ_1	λ_2	λ_3	λ_4
0.2386	-0.0562	0.2779	**1.2365***
(0.1925)	(0.1174)	(0.1631)	(0.2137)

Table 2.1: Summary of estimates of coefficients corresponding to weighting matrices in specification (2.5.1)

Standard errors are reported in parentheses, starred estimates in bold are significant at 5% level

We now allow for different spatial parameters for different phases of the sampling period, as indicated in Section 2.4.2. In particular, we split the sampling period into three four-week periods, henceforth referred to as months. For the insignificant spatial lags, we find that the lags remain insignificant in each month even after allowing for different spatial parameters. As a result, we estimate the model with different spatial parameters for each month corresponding only to the significant financial links matrix. Specifically, the model estimated is:

$$\begin{aligned} y_n &= (I_T \otimes l_N)\alpha + X_n\beta + (I_T \otimes F_N)\gamma + \lambda_1 \left(I_T \otimes W^{Trade}\right) y_n \\ &+ \sum_{j=1}^{3} \lambda_{j2}\left(I_T \otimes W_j^{Fin}\right) y_n + \lambda_3 \left(I_T \otimes W^{Sim}\right) y_n \\ &+ \lambda_4 \left(I_T \otimes W^{Nbd}\right) y_n + U_n \end{aligned} \tag{2.5.2}$$

where

$$W_1^{Fin} = diag\left[W^{Fin}, W^{Fin}, W^{Fin}, W^{Fin}, 0, \ldots, 0\right]$$

and W_2^{Fin} and W_3^{Fin} are defined analogously using the 5th-8th and 9th-12th diagonal blocks respectively. IV estimation is used, with the linearly independent columns of

$$X_n, I_T \otimes W^{Trade} X_n, I_T \otimes W^{Sim} X_n, I_T \otimes W^{Nbd} X_n, I_T \otimes W_j^{Fin} X_n, j = 1, 2, 3$$

and

$$\begin{aligned} &I_T \otimes F_N, \left(I_T \otimes W^{Trade}\right)(I_T \otimes F_N), \left(I_T \otimes W^{Sim}\right)(I_T \otimes F_N), \\ &\left(I_T \otimes W^{Nbd}\right)(I_T \otimes F_N), \left(I_T \otimes W_j^{Fin}\right)(I_T \otimes F_N), j = 1, 2, 3 \end{aligned}$$

being used as instruments, so that r_n from Section 2.3.1 may also be taken as $T/4$ here. Table 2.2 reports the estimates of the spatial autoregressive parameters. These indicate that λ_{12} and λ_{32} are significant at the 5% level, while λ_{22} is not. This would indicate that contagion through financial links occurs immediately after the onset of a crisis, followed by a lull, and then another period of contagion. This could be due to

λ_1	λ_{12}	λ_{22}	λ_{32}	λ_3	λ_4
0.2327	**1.0541***	0.4714	**1.4938***	-0.0236	0.2767
(0.1827)	(0.4182)	(0.4394)	(0.2508)	(0.1162)	(0.1576)

Table 2.2: Summary of estimates of coefficients corresponding to weighting matrices in specification (2.5.2)

Standard errors are reported in parentheses, starred estimates in bold are significant at 5% level

the fact that immediately after a crisis, country i immediately increases its borrowing from a financial centre but this stabilises after a few weeks. However some domestic businesses hit hard by the crisis may only start to feel financial hardship some time after the initial shock and create a second wave of demand for borrowing. A t-test was also conducted for the null hypothesis

$$H_0 : \lambda_{12} = \lambda_{32}$$

which failed to reject the null returning a test statistic value of 0.9358, indicating that both 'waves' seem to have equal impact on stock market returns through the financial channel.

Remark In another piece of research, not published in this thesis, we consider the problem of testing increasingly many linear restrictions on the parameters $\theta_{(n)} = \left(\lambda'_{(n)}, \beta'_{(n)}\right)'$ of (2.2.1). Such a problem is natural not only because of the increasing-parameter definition of the model, but also because the increasing autoregressive order can arise from a partitioning of the data (see Chapter 2). In the latter case, Chapter 2 prescribes an approach that takes into account heterogeneity between the clusters by recommending that the autoregressive parameters be allowed to vary across clusters. On the other hand, practitioners have a preference for a parsimonious model where possible. As a result there is great interest in testing null hypotheses of the type

$$H_0 : \lambda_1 = \lambda_2 = \ldots = \lambda_{p_n}. \tag{2.5.3}$$

Bearing in mind that p_n increases with n, a more meaningful way of writing the above null hypothesis would be

$$H_0 : \sum_{\substack{i,j=1 \\ i<j}}^{p_n} (\lambda_i - \lambda_j)^2 = 0. \tag{2.5.4}$$

We focus on Lagrange Multiplier (LM) tests. This principle is particularly attractive when testing such hypotheses because it requires only an estimation of the model under the null hypothesis. This may even reduce the model to a finite-dimensional one, as is the case if the null hypothesis is as in (2.5.3) and k_n is fixed. When testing a fixed number of restrictions, say q, the LM test statistic has an asymptotic χ_q^2 distribution under mild regularity conditions. However, it is well known that

$$\frac{\chi_q^2 - q}{\sqrt{2q}} \xrightarrow{d} N(0,1)$$

as $q \to \infty$. We will use this result as our motivation to propose standardized LM statistics and establish their asymptotic normality. Results of this type indicate that critical values from a standard Normal distribution may be employed to conduct inference on the parameters of (2.2.1). From a practitioners point of view, this is an attractive result because critical values from both the chi-squared and Normal distributions can be used to carry out inference. This adds another layer to the diagnostic procedures available to the practitioner. The results can also be viewed as an analogue to familiar statistical results, where either the t distribution or the Normal distribution can be used to obtain critical values for t-tests, but these get arbitrarily close asymptotically.

The idea of using a standardised version of the LM test dates back to at least De Jong and Bierens (1994), who used a similar idea for testing increasingly many conditional moment restrictions. They employed a proof using a central limit theorem of Hall (1984) for degenerate U-statistics, owing to the i.i.d. nature of their data. The regularity implied by i.i.d. data does not obtain in our setting, however, and we use direct martingale central limit theorem arguments to establish the limiting distribution of our test statistics.

It should be mentioned that there are other ways to construct test statistics for testing the equality of the λ_i, but the LM approach allows us to accommodate general linear restrictions and delivers standard asymptotics. Motivated by the extreme-value literature, given some consistent and asymptotically normal estimates λ_i^e of λ_i, a leading candidate for a test statistic to test (2.5.4) is

$$\sup_{i=1,\dots,p_n} |\lambda_i^e| - \inf_{i=1,\dots,p_n} |\lambda_i^e| . \tag{2.5.5}$$

Rejection of the null hypothesis can be based on large values of (2.5.5), but this statistic suffers from major disadvantages as opposed to the LM approach. First, the LM approach does not require the estimation of the unrestricted model so only a model of fixed dimension needs to be estimated. Secondly, the asymptotic distribution of (2.5.5) is extremely hard to derive and will be non-standard, leading to complications in terms of obtaining critical values. As a result, we feel that LM tests are very useful and easy to handle in this context.

Europe	South and South-East Asia	Latin America	Middle East and North Africa	Sub-Saharan Africa
Bulgaria	Australia	Argentina	Egypt	Mauritius
Croatia	China	Brazil	Israel	South Africa
Cyprus	India	Chile	Jordan	Zimbabwe
Czech Republic	Indonesia	Colombia	Kuwait	Kenya
Estonia	Malaysia	Ecuador	Lebanon	Nigeria
Greece	New Zealand	Mexico	Morocco	
Hungary	Pakistan	Peru	Saudi Arabia	
Iceland	Philippines	Venezuela	Tunisia	
Latvia	Singapore			
Lithuania	South Korea			
Malta	Sri Lanka			
Poland	Thailand			
Portugal				
Romania				
Russia				
Slovakia				
Slovenia				
Turkey				
Ukraine				

Table 2.3: List of countries and region classification for Section 2.5

2.A Matrix norm inequalities and notation

There are several inequalities relating the matrix norms used in this thesis. First

$$\|A\| \le \|A\|_F$$

where. This inequality relating the spectral and Frobenius norms is used repeatedly without explicit reference to the Frobenius norm.

Another useful inequality that relates the spectral norm to the maximum row and column sum norms is

$$\|A\| \le \sqrt{\|A\|_R \|A\|_C}. \tag{2.A.1}$$

This allows us to conclude that a matrix that is uniformly bounded in row and column sums is also uniformly bounded in spectral norm. Finally the spectral, maximum row-sum and maximum column-sum norms are all sub-multiplicative.

We also denote $a_n = p_n + k_n$, $b_n = r_n + k_n$, $c_n = p_n k_n^2 + k_n$ and $\tau_n = n^{\frac{1}{2}}/a_n^{\frac{1}{2}}$ to conserve space.

2.B Proofs of results in Section 2.3

Proof of Lemma 2.1.

(i) By definition

$$\underline{\eta}\left(Q_n\right) = \underline{\eta}\left(K_n' J_n^{-1} K_n\right) = \min_{\|x_n\|=1} x_n' K_n' J_n^{-1} K_n x_n$$

while for a $a_n \times 1$ vector x_n satisfying $\|x_n\| = 1$

$$\frac{x_n' K_n' J_n^{-1} K_n x_n}{x_n' K_n' K_n x_n} \ge \underline{\eta}\left(J_n^{-1}\right)$$

so that

$$x_n' K_n' J_n^{-1} K_n x_n \ge \underline{\eta}\left(J_n^{-1}\right) x_n' K_n' K_n x_n \ge \frac{\underline{\eta}\left(K_n' K_n\right)}{\overline{\eta}\left(J_n\right)} \ge c,$$

for large n by Assumption 7, where c denotes a positive but arbitrarily small real number that does not depend on n. Then the result follows because the calculations above indicate that minimization of $x_n' K_n' J_n^{-1} K_n x_n$ over x_n is bounded away from zero for large n and, therefore, so is the limit inferior.

(ii) Similar.

$\square$

Proof of Theorem 2.1. Write

$$\begin{aligned}
\hat{\theta}_{(n)} - \theta_{(n)} &= \left(\hat{Q}_n^{-1} - Q_n^{-1}\right)\hat{K}_n' J_n^{-1} q_n + Q_n^{-1}\hat{K}_n' J_n^{-1} q_n \\
&= Q_n^{-1}\left(Q_n - \hat{Q}_n\right)\hat{Q}_n^{-1}\hat{K}_n' J_n^{-1} q_n + Q_n^{-1}\left(\hat{K}_n - K_n\right)' J_n^{-1} q_n \\
&+ Q_n^{-1} K_n' J_n^{-1} q_n \\
&= Q_n^{-1}\left(Q_n - \hat{Q}_n\right)\left(\hat{\theta}_{(n)} - \theta_{(n)}\right) + Q_n^{-1}\left(\hat{K}_n - K_n\right)' J_n^{-1} q_n \\
&+ Q_n^{-1} K_n' J_n^{-1} q_n. \qquad (2.B.1)
\end{aligned}$$

By elementary norm inequalities

$$\left\|\hat{Q}_n - Q_n\right\| \leq \left\|\hat{K}_n - K_n\right\| \left\| J_n^{-1} \right\| \left(\left\|\hat{K}_n - K_n\right\| + 2\left\|K_n\right\|\right), \qquad (2.B.2)$$

where $E\left\|\hat{K}_n - K_n\right\|^2$ is bounded by

$$\frac{\sigma^2}{n^2}\sum_{i=1}^{b_n}\sum_{j=1}^{p_n}\left|p_{in}' G_{jn} G_{jn}' p_{in}\right| \leq \frac{\sigma^2}{n^2}\sum_{i=1}^{b_n}\left\|p_{in}\right\|^2 \sum_{j=1}^{p_n}\left\|G_{jn}\right\|^2 \leq C\frac{p_n b_n}{n}$$

by Assumptions 5 and 6 and Lemma 2.C1, denoting by p_{in} the $i-th$ column of (Z_n, X_n). We conclude that

$$\left\|\hat{K}_n - K_n\right\| = \mathscr{O}_p\left(\frac{p_n^{\frac{1}{2}} b_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}\right) \qquad (2.B.3)$$

by Markov's inequality. Then

$$\left\|\hat{Q}_n - Q_n\right\| = \mathscr{O}_p\left(\max\left\{\frac{p_n b_n}{n}, \frac{p_n^{\frac{1}{2}} b_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}\right\}\right) = \mathscr{O}_p\left(\frac{p_n^{\frac{1}{2}} b_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}\right), \qquad (2.B.4)$$

by Assumption (8) because $\left\| J_n^{-1} \right\| = \left(\underline{\eta}\left(J_n\right)\right)^{-1}$ and $\left\|K_n\right\| = \overline{\eta}\left(K_n K_n'\right)$. Likewise

$$E\left\|q_n\right\|^2 = E\left\|\frac{1}{n}\sum_{i=1}^{n} a_{in} u_i\right\|^2 = \frac{\sigma^2}{n^2}\sum_{i=1}^{n}\left\|a_{in}\right\|^2 = \mathscr{O}\left(\frac{b_n}{n}\right),$$

where a_{in}' is the $i-th$ row of (Z_n, X_n), since the elements of a_{in}' are uniformly bounded by Assumptions 5 and 6. By Markov's inequality

$$\left\|q_n\right\| = \mathscr{O}_p\left(\frac{b_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}\right). \qquad (2.B.5)$$

Upon taking norms of (2.B.1) and rearranging we get

$$\begin{aligned}\left(1-\left\|Q_n^{-1}\right\|\left\|\hat{Q}_n-Q_n\right\|\right)\left\|\hat{\theta}_{(n)}-\theta_{(n)}\right\| &\leq \left\|Q_n^{-1}\right\|\left\|\hat{K}_n-K_n\right\|\left\|J_n^{-1}\right\|\left\|q_n\right\| \\ &+ \left\|Q_n^{-1}\right\|\left\|K_n\right\|\left\|J_n^{-1}\right\|\left\|q_n\right\|,\end{aligned} \tag{2.B.6}$$

using the submultiplicative property of the spectral norm. By (2.B.4) the first factor on the LHS above converges in probability to one by (2.3.7) and Lemma 2.1 (i), and because

$$\frac{p_n b_n}{n}=\frac{p_n r_n+p_n k_n}{n}.$$

This also ensures that the first factor in the first term on the RHS of (2.B.6) is bounded, as well as the third factor by Assumption 8. The second and fourth factors have orders given in (2.B.3) and (2.B.5) respectively, implying that the first term is $\mathscr{O}_p\left(\frac{p_n^{\frac{1}{2}} b_n}{n}\right)$. The order of the second term on the RHS is determined similarly to be $\mathscr{O}_p\left(\frac{b_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}\right)$ so that

$$\left\|\hat{\theta}_{(n)}-\theta_{(n)}\right\|=\mathscr{O}_p\left(\max\left\{\frac{p_n^{\frac{1}{2}} b_n}{n},\frac{b_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}\right\}\right)=\mathscr{O}_p\left(\frac{b_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}\right). \tag{2.B.7}$$

This is negligible by Assumption 2.3.7. The proof of Corollary 2.2 is similar. □

Proof of Theorem 2.2. Write

$$\begin{aligned}\hat{\sigma}_{(n)}^2 &= \frac{1}{n}\left(U_n-(R_n,X_n)\left(\hat{\theta}_{(n)}-\theta_{(n)}\right)\right)'\left(U_n-(R_n,X_n)\left(\hat{\theta}_{(n)}-\theta_{(n)}\right)\right) \\ &= \frac{1}{n}U_n'U_n-\frac{2}{n}\left(\hat{\theta}_{(n)}-\theta_{(n)}\right)'\begin{bmatrix}R_n'\\X_n'\end{bmatrix}U_n \\ &\quad +\frac{1}{n}\left(\hat{\theta}_{(n)}-\theta_{(n)}\right)'\begin{bmatrix}R_n'\\X_n'\end{bmatrix}[R_n,X_n]\left(\hat{\theta}_{(n)}-\theta_{(n)}\right) \\ &= \frac{1}{n}U_n'U_n-2\left(\hat{\theta}_{(n)}-\theta_{(n)}\right)'w_n+\left(\hat{\theta}_{(n)}-\theta_{(n)}\right)'\hat{L}_n\left(\hat{\theta}_{(n)}-\theta_{(n)}\right).\end{aligned}$$

By the Khinchin Law of Large Numbers, we have $\frac{1}{n}U_n'U_n=\sigma^2+o_p(1)$. Also by (2.B.7) and (2.B.16) the modulus of the second term above is bounded by

$$\left\|\hat{\theta}_{(n)}-\theta_{(n)}\right\|\left\|w_n\right\|=\mathscr{O}_p\left(\max\left\{\frac{b_n^{\frac{1}{2}}c_n^{\frac{1}{2}}}{n},\frac{p_n^{\frac{1}{2}}b_n^{\frac{1}{2}}}{n^{\frac{1}{2}}h_n}\right\}\right)$$

while the third term has modulus bounded by

$$\begin{aligned}
& \left| \left(\hat{\theta}_{(n)} - \theta_{(n)}\right)' L_n \left(\hat{\theta}_{(n)} - \theta_{(n)}\right) \right| \\
& + \left| \left(\hat{\theta}_{(n)} - \theta_{(n)}\right)' \left(\hat{L}_n - L_n\right) \left(\hat{\theta}_{(n)} - \theta_{(n)}\right) \right| \\
\leq \; & \left\| \hat{\theta}_{(n)} - \theta_{(n)} \right\|^2 \left\| L_n \right\| + \left\| \hat{\theta}_{(n)} - \theta_{(n)} \right\|^2 \left\| \hat{L}_n - L_n \right\| \\
= \; & \mathscr{O}_p \left(\max \left\{ \frac{b_n}{n}, \frac{p_n^{\frac{1}{2}} k_n a_n^{\frac{1}{2}} b_n}{n^{\frac{3}{2}}}, \frac{p_n b_n}{n h_n} \right\} \right) \\
= \; & \mathscr{O}_p \left(\max \left\{ \frac{b_n}{n}, \frac{p_n b_n}{n h_n} \right\} \right)
\end{aligned}$$

using (2.B.7), (2.B.20) and Assumption 9. Thus, noting that $\frac{p_n b_n}{n h_n}$ and $\frac{p_n^{\frac{1}{2}} k_n a_n^{\frac{1}{2}} b_n}{n^{\frac{3}{2}}}$ are dominated by $\frac{p_n^{\frac{1}{2}} b_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n}$ under (2.3.10), we have

$$\hat{\sigma}^2_{(n)} - \sigma^2 = \mathscr{O}_p \left(\max \left\{ \frac{b_n^{\frac{1}{2}} c_n^{\frac{1}{2}}}{n}, \frac{p_n^{\frac{1}{2}} b_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n}, \frac{b_n}{n} \right\} \right), \tag{2.B.8}$$

which is negligible by (2.3.10) and because h_n is bounded away from zero, noting that

$$\frac{b_n c_n}{n^2} \leq C \left(\frac{p_n r_n k_n^2 + p_n k_n^3}{n^2} \right).$$

□

Proof of Theorem 2.3. Let α be any $s \times 1$ vector of constants and write

$$\begin{aligned}
\tau_n \alpha' \Psi_n \left(\hat{\theta}_{(n)} - \theta_{(n)}\right) \; &= \; \tau_n \alpha' \Psi_n Q_n^{-1} \left(\hat{Q}_n - Q_n\right) \left(\hat{\theta}_{(n)} - \theta_{(n)}\right) \\
& + \; \tau_n \alpha' \Psi_n Q_n^{-1} \left(\hat{K}_n - K_n\right)' J_n^{-1} q_n \\
& + \; \tau_n \alpha' \Psi_n Q_n^{-1} K_n' J_n^{-1} q_n.
\end{aligned} \tag{2.B.9}$$

We first show that first term on the RHS of (2.B.9) is negligible in probability. It has modulus bounded by

$$\tau_n \left\| \alpha \right\| \left\| \Psi_n \right\| \left\| \hat{\theta}_{(n)} - \theta_{(n)} \right\| \left\| Q_n^{-1} \right\| \left\| \hat{Q}_n - Q_n \right\| = \mathscr{O}_p \left(\frac{p_n^{\frac{1}{2}} b_n}{n^{\frac{1}{2}}} \right),$$

from (2.B.4), (2.B.7) and Assumption 8. This is negligible by (2.3.11) because, by elementary inequalities,

$$\frac{p_n b_n^2}{n} \le C\left(\frac{p_n r_n^2 + p_n k_n^2}{n}\right).$$

Similarly the second term on the right side of (2.B.9) is bounded in absolute value by

$$\tau_n \|\alpha\| \|\Psi_n\| \left\|Q_n^{-1}\right\| \left\|\hat{K}_n - K_n\right\| \|q_n\| = \mathscr{O}_p\left(\frac{p_n^{\frac{1}{2}} b_n}{n^{\frac{1}{2}}}\right)$$

so we have to prove asymptotic normality only for the third term on the RHS of (2.B.9).

Now

$$\tau_n \alpha' \Psi_n Q_n^{-1} K_n' J_n^{-1} q_n = \frac{1}{n^{\frac{1}{2}} a_n^{\frac{1}{2}}} \sum_{i=1}^{n} \alpha' \Psi_n Q_n^{-1} K_n' J_n^{-1} a_{in} u_i$$

has mean zero and variance

$$\frac{\sigma^2}{n a_n} \sum_{i=1}^{n} (\alpha' \Psi_n Q_n^{-1} K_n' J_n^{-1} a_{in})^2.$$

Thus consider

$$\frac{n \alpha' \Psi_n Q_n^{-1} K_n' J_n^{-1} q_n}{\sigma \left\{\sum_{i=1}^{n} (\alpha' \Psi_n Q_n^{-1} K_n' J_n^{-1} a_{in})^2\right\}^{\frac{1}{2}}} = \sum_{i=1}^{n} c_{in} u_i,$$

where

$$c_{in} = \frac{\alpha' \Psi_n Q_n^{-1} K_n' J_n^{-1} a_{in}}{\sigma \left\{\sum_{i=1}^{n} (\alpha' \Psi_n Q_n^{-1} K_n' J_n^{-1} a_{in})^2\right\}^{\frac{1}{2}}}.$$

We now verify the Lindeberg condition for $c_{in} u_i$. We have

$$\sum_{i=1}^{n} \mathbb{E}\left\{(c_{in} u_i)^2 \mathbf{1}(|c_{in} u_i| > \epsilon)\right\} \le \max_{1\le i\le n} \mathbb{E}\left\{u_i^2 \mathbf{1}\left(u_i^2 > \frac{\epsilon^2}{\max\limits_{1\le i\le n} c_{in}^2}\right)\right\} \sum_{i=1}^{n} c_{in}^2$$

Note that assuming 2nd moments for the u_i ensures that u_i^2 are uniformly integrable since they are iid. Therefore it is sufficient to show that $\max\limits_{1\le i\le n} c_{in}^2 \to 0$ as $n \to \infty$, as the last factor on the RHS of the above displayed inequality equals $1/\sigma^2$. Consider

$$\begin{aligned}\max_{1\le i\le n} c_{in}^2 &= \max_{1\le i\le n} \frac{(\alpha'\Psi_n Q_n^{-1}K_n'J_n^{-1}a_{in})^2}{\sigma^2\sum_{i=1}^{n}(\alpha'\Psi_n Q_n^{-1}K_n'J_n^{-1}a_{in})^2}\\ &\le \frac{\left\|Q_n^{-1}K_n'J_n^{-1}\right\|^2\left\|\Psi_n'\alpha\right\|^2\max_{1\le i\le n}\left\|a_{in}\right\|^2}{\sigma^2\sum_{i=1}^{n}(\alpha'\Psi_n Q_n^{-1}K_n'J_n^{-1}a_{in})^2}. \qquad (2.B.10)\end{aligned}$$

The denominator of (2.B.10) equals σ^2 times

$$\begin{aligned}&\alpha'\Psi_n Q_n^{-1}K_n'J_n^{-1}\sum_{i=1}^{n}a_{in}a_{in}'J_n^{-1}K_nQ_n^{-1}\Psi_n'\alpha\\ \ge\ & \left\|\Psi_n'\alpha\right\|^2\underline{\eta}\left(Q_n^{-1}K_n'J_n^{-1}\sum_{i=1}^{n}a_{in}a_{in}'J_n^{-1}K_nQ_n^{-1}\right)\\ =\ & n\left\|\Psi_n'\alpha\right\|^2\ \underline{\eta}\left(Q_n^{-1}K_n'J_n^{-1}K_nQ_n^{-1}\right)\\ =\ & n\left\|\Psi_n'\alpha\right\|^2\ \underline{\eta}\left(Q_n^{-1}\right)\\ \ge\ & nc\left\|\Psi_n'\alpha\right\|^2\end{aligned}$$

for sufficiently large n by Lemma 2.1 (ii), noting that $\sum_{i=1}^{n}a_{in}a_{in}' = nJ_n$, so (2.B.10) is $\mathscr{O}\left(\frac{b_n}{n}\right)$ by Assumptions 5 and 6, which is negligible by (2.3.7). The Lindeberg condition is then satisfied. The proof of the consistency of the covariance matrix estimate is omitted, while the proof of Corollary 2.3 is similar. □

Proof of Theorem 2.4. We can write

$$\begin{aligned}\tilde{\theta}_{(n)}-\theta_{(n)} &= \left(\hat{L}_n^{-1}-L_n^{-1}\right)w_n + L_n^{-1}w_n\\ &= L_n^{-1}\left(\hat{L}_n-L_n\right)\hat{L}_n^{-1}w_n + L_n^{-1}w_n\\ &= L_n^{-1}\left(\hat{L}_n-L_n\right)\left(\tilde{\theta}_{(n)}-\theta_{(n)}\right)+L_n^{-1}w_n. \qquad (2.B.11)\end{aligned}$$

It is clear that

$$\left\|w_n\right\| \le \left\|\frac{1}{n}\begin{bmatrix}A_n'\\ X_n'\end{bmatrix}U_n\right\| + \left\|\frac{1}{n}\begin{bmatrix}B_n'\\ 0\end{bmatrix}U_n\right\|. \qquad (2.B.12)$$

Now

$$\mathbb{E}\left\|\frac{1}{n}\begin{bmatrix}A_n'\\ X_n'\end{bmatrix}U_n\right\|^2 = \mathscr{O}\left(\frac{c_n}{n}\right), \qquad (2.B.13)$$

as in Section 2.1 since the elements of A_n are uniformly $\mathscr{O}\left(k_n\right)$ (Lemma 2.C5). Under

Assumption 11, the square of the second term on the RHS of (2.B.12) has expectation

$$\frac{1}{n^2}\sum_{i=1}^{p_n}\mathbb{E}\left(U_n'G_{in}'U_n\right)^2, \tag{2.B.14}$$

which, using the proof of Lemma 2.C4 and denoting $\mathbb{E}u_i^4=\mu_4$, equals $\sum_{i=1}^{4}\Delta_{in}$ where

$$\begin{aligned}
\Delta_{1n} &= \frac{\mu_4}{n^2}\sum_{i=1}^{p_n}\sum_{j=1}^{n}g_{jj,in}^2=\mathscr{O}\left(\frac{p_n}{nh_n^2}\right)\\
\Delta_{2n} &= \frac{\sigma^4}{n^2}\sum_{i=1}^{p_n}\sum_{j=1}^{n}\sum_{k=1}^{n}g_{jj,in}g_{kk,in}=\mathscr{O}\left(\frac{p_n}{h_n^2}\right)\\
\Delta_{3n} &= \frac{\sigma^4}{n^2}\sum_{i=1}^{p_n}\sum_{j=1}^{n}\sum_{k=1}^{n}g_{jk,in}g_{kj,in}=\mathscr{O}\left(\frac{p_n}{h_n^2}\right)\\
\Delta_{4n} &= \frac{\sigma^4}{n^2}\sum_{i=1}^{p_n}\sum_{j=1}^{n}\sum_{k=1}^{n}g_{jk,in}^2=\mathscr{O}\left(\frac{p_n}{h_n^2}\right),
\end{aligned}$$

by Lemma 2.C2, where $g_{rs,in}$ denotes the (r,s)-th element of G_{in}. Hence

$$\left\|\frac{1}{n}\begin{bmatrix}B_n'\\0\end{bmatrix}U_n\right\|=\mathscr{O}_p\left(\frac{p_n^{\frac{1}{2}}}{h_n}\right) \tag{2.B.15}$$

so that

$$\|w_n\|=\mathscr{O}_p\left\{\max\left(\frac{c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}},\frac{p_n^{\frac{1}{2}}}{h_n}\right)\right\}. \tag{2.B.16}$$

However, under Assumption 1 we have

$$\left\|\frac{1}{n}\begin{bmatrix}B_n'\\0\end{bmatrix}U_n\right\|\leq\frac{1}{n}\left\|[B_n,0]\right\|\left\|U_n\right\|=\mathscr{O}_p\left(\frac{p_n^{\frac{1}{2}}}{h_n^{\frac{1}{2}}}\right) \tag{2.B.17}$$

by calculations used for bounding the first term on the RHS of (2.B.19) and so

$$\|w_n\|=\mathscr{O}_p\left\{\max\left(\frac{c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}},\frac{p_n^{\frac{1}{2}}}{h_n^{\frac{1}{2}}}\right)\right\}. \tag{2.B.18}$$

Also

$$\begin{aligned}
\hat{L}_n - L_n &= \frac{1}{n}\begin{bmatrix} R_n' \\ X_n' \end{bmatrix}[R_n, X_n] - \frac{1}{n}\begin{bmatrix} A_n' \\ X_n' \end{bmatrix}[A_n, X_n] \\
&= \frac{1}{n}\left(\begin{bmatrix} R_n' \\ X_n' \end{bmatrix} - \begin{bmatrix} A_n' \\ X_n' \end{bmatrix}\right)([R_n, X_n] - [A_n, X_n]) \\
&+ \frac{1}{n}\begin{bmatrix} A_n' \\ X_n' \end{bmatrix}([R_n, X_n] - [A_n, X_n]) \\
&+ \frac{1}{n}\left(\begin{bmatrix} R_n' \\ X_n' \end{bmatrix} - \begin{bmatrix} A_n' \\ X_n' \end{bmatrix}\right)[A_n, X_n] \\
&= \frac{1}{n}\begin{bmatrix} B_n' \\ 0 \end{bmatrix}[B_n, 0] + \frac{1}{n}\begin{bmatrix} A_n' \\ X_n' \end{bmatrix}[B_n, 0] + \frac{1}{n}\begin{bmatrix} B_n' \\ 0 \end{bmatrix}[A_n, X_n]
\end{aligned}$$

so we have

$$\left\| \hat{L}_n - L_n \right\| \leq \frac{1}{n}\left\|[B_n, 0]\right\|^2 + \frac{2}{n}\left\| \begin{bmatrix} A_n' \\ X_n' \end{bmatrix}[B_n, 0] \right\|. \tag{2.B.19}$$

The first term in the last displayed expression expectation bounded by

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p_n}\mathbb{E}\left(e_{i,n}' G_{jn} U_n\right)^2 &= \frac{1}{n}\sum_{j=1}^{p_n}\mathbb{E}\left(U_n' G_{jn}\sum_{i=1}^{n} e_{i,n} e_{i,n}' G_{jn}' U_n\right) \\
&\leq \frac{\sigma^2}{n}\sum_{j=1}^{p_n} tr\left(G_{jn} G_{jn}'\right) \leq C\frac{p_n}{h_n},
\end{aligned}$$

using Lemmas 2.C2 and 2.C3. For the second term in (2.B.19) note that

$$\left\| \frac{1}{n}\begin{bmatrix} A_n' \\ X_n' \end{bmatrix}[B_n, 0] \right\|^2 \leq \frac{1}{n^2}\sum_{i=1}^{a_n}\sum_{j=1}^{p_n} h_{in}' G_{jn} U_n U_n' G_{jn}' h_{in},$$

where h_{in} is the $i-th$ column of (A_n, X_n). Then by Lemma 2.C5, Assumption 5 and Lemma 2.C1 we have

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{n^2}\sum_{i=1}^{a_n}\sum_{j=1}^{p_n} h_{in}' G_{jn} U_n U_n' G_{jn}' h_{in}\right) &= \frac{\sigma^2}{n^2}\sum_{i=1}^{a_n}\sum_{j=1}^{p_n} h_{in}' G_{jn}' G_{jn}' h_{in} \\
&\leq \frac{\sigma^2}{n^2}\sum_{i=1}^{a_n}\|h_{in}\|^2\sum_{j=1}^{p_n}\|G_{jn}\|^2 \\
&\leq C\frac{p_n k_n^2 a_n}{n}
\end{aligned}$$

so that

$$\left\| \hat{L}_n - L_n \right\| = \mathscr{O}_p \left(\max \left\{ \frac{p_n}{h_n}, \frac{p_n^{\frac{1}{2}} k_n a_n^{\frac{1}{2}}}{n^{\frac{1}{2}}} \right\} \right). \tag{2.B.20}$$

Note that the bound derived above required only second order moments for the u_i and using fourth order moments (Assumption 11) will not improve the bound because

$$\frac{1}{n^2}\mathbb{E}\left\| \begin{bmatrix} B_n' \\ 0 \end{bmatrix} [B_n, 0] \right\|^2 \leq \frac{1}{n^2} \sum_{i=1}^{p_n} \sum_{j=1}^{p_n} \mathbb{E}\left(U_n' G_{in}' G_{jn} U_n \right)^2$$

which is $\mathscr{O}\left(\frac{p_n^2}{h_n^2}\right)$ in exactly the same way as we bounded (2.B.14) since the elements of $G_{in}' G_{jn}$ are $\mathscr{O}\left(\frac{1}{h_n}\right)$ uniformly in i, j and n by Lemma 2.C3.

Upon taking norms of (2.B.11) and rearranging we get

$$\left(1 - \left\| L_n^{-1} \right\| \left\| \hat{L}_n - L_n \right\|\right) \left\| \tilde{\theta}_{(n)} - \theta_{(n)} \right\| \leq \left\| L_n^{-1} \right\| \left\| w_n \right\| \tag{2.B.21}$$

using the submultiplicative property of the spectral norm. By (2.B.20) the first factor on the LHS above converges in probability to one by (2.3.14) and Assumption 10, the last being useful since $\left\| L_n^{-1} \right\| = (\underline{\eta}(L_n))^{-1}$. Again, the first factor on the RHS of (2.B.21) is bounded by Assumption 10 for sufficiently large n and so we have

$$\left\| \tilde{\theta}_{(n)} - \theta_{(n)} \right\| = \mathscr{O}_p \left\{ \max\left(\frac{c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}, \frac{p_n^{\frac{1}{2}}}{h_n} \right) \right\} \tag{2.B.22}$$

by (2.B.16) under Assumptions 1 and 11 but

$$\left\| \tilde{\theta}_{(n)} - \theta_{(n)} \right\| = \mathscr{O}_p \left\{ \max\left(\frac{c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}, \frac{p_n^{\frac{1}{2}}}{h_n^{\frac{1}{2}}} \right) \right\} \tag{2.B.23}$$

by (2.B.18) under Assumption 1 only. These are negligible by (2.3.14). □

Proof of Theorem 2.5. As in the IV case, we write

$$\tilde{\sigma}_{(n)}^2 = \frac{1}{n} U_n' U_n - 2\left(\tilde{\theta}_{(n)} - \theta_{(n)}\right)' w_n + \left(\tilde{\theta}_{(n)} - \theta_{(n)}\right)' \hat{L}_n \left(\tilde{\theta}_{(n)} - \theta_{(n)}\right).$$

From (2.B.22) and (2.B.16) the second term has modulus bounded by

$$\begin{aligned}
&\left\| \tilde{\theta}_{(n)} - \theta_{(n)} \right\| \|w_n\| \\
= \;& \mathscr{O}_p\left\{ \max\left(\frac{c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}, \frac{p_n^{\frac{1}{2}}}{h_n} \right) \right\} \mathscr{O}_p\left\{ \max\left(\frac{c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}, \frac{p_n^{\frac{1}{2}}}{h_n} \right) \right\} \\
= \;& \mathscr{O}_p\left(\max\left\{ \frac{c_n}{n}, \frac{p_n}{h_n^2}, \frac{p_n^{\frac{1}{2}} c_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n} \right\} \right)
\end{aligned}$$

while the modulus of the third term is bounded by

$$\begin{aligned}
& \left| \left(\tilde{\theta}_{(n)} - \theta_{(n)} \right)' \left(\hat{L}_n - L_n \right) \left(\tilde{\theta}_{(n)} - \theta_{(n)} \right) \right| \\
& + \left| \left(\tilde{\theta}_{(n)} - \theta_{(n)} \right)' L_n \left(\tilde{\theta}_{(n)} - \theta_{(n)} \right) \right| \\
\leq \;& \left\| \tilde{\theta}_{(n)} - \theta_{(n)} \right\|^2 \left\| \hat{L}_n - L_n \right\| + \left\| \tilde{\theta}_{(n)} - \theta_{(n)} \right\|^2 \|L_n\| \\
= \;& \mathscr{O}_p\left(\max\left\{ \frac{c_n}{n}, \frac{p_n}{h_n^2}, \frac{p_n^{\frac{1}{2}} c_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n} \right\} \right) \mathscr{O}_p\left(\max\left\{ \frac{p_n}{h_n}, \frac{k_n p_n^{\frac{1}{2}} a_n^{\frac{1}{2}}}{n^{\frac{1}{2}}} \right\} \right) \\
& + \mathscr{O}_p\left(\max\left\{ \frac{c_n}{n}, \frac{p_n}{h_n^2}, \frac{p_n^{\frac{1}{2}} c_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n} \right\} \right) \\
= \;& \mathscr{O}_p\left(\max\left\{ \frac{c_n}{n}, \frac{p_n}{h_n^2}, \frac{p_n^{\frac{1}{2}} c_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n} \right\} \right),
\end{aligned}$$

using (2.B.22), (2.B.20) and Assumption 9. We conclude that

$$\tilde{\sigma}_{(n)}^2 - \sigma^2 = \mathscr{O}_p\left(\max\left\{ \frac{c_n}{n}, \frac{p_n}{h_n^2}, \frac{p_n^{\frac{1}{2}} c_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n} \right\} \right). \tag{2.B.24}$$

This is negligible by (2.3.16). □

Proof of Theorem 2.6. First, with α any $s \times 1$ vector, write

$$\begin{aligned}
\tau_n \alpha' \Psi_n \left(\tilde{\theta}_{(n)} - \theta_{(n)} \right) \;=\;& \tau_n \alpha' \Psi_n \left(\hat{L}_n^{-1} - L_n^{-1} \right) w_n \\
+\;& \tau_n \alpha' \Psi_n L_n^{-1} w_n.
\end{aligned} \tag{2.B.25}$$

We first show that first term on the RHS of (2.B.25) is negligible in probability. This term has modulus bounded by τ_n times

$$\begin{aligned} & \|\alpha\| \, \|\Psi_n\| \left\| \hat{L}_n^{-1} w_n \right\| \, \| L_n^{-1} \| \, \left\| \hat{L}_n - L_n \right\| \\ = \; & \|\alpha\| \, \|\Psi_n\| \left\| \tilde{\theta}_{(n)} - \theta_{(n)} \right\| \, \| L_n^{-1} \| \, \left\| \hat{L}_n - L_n \right\| . \end{aligned}$$

The second factor on the RHS is $\mathscr{O}\left(a_n^{\frac{1}{2}}\right)$, the third is $\mathscr{O}_p\left(\max\left\{\frac{c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}, \frac{p_n^{\frac{1}{2}}}{h_n}\right\}\right)$ by (2.B.22), the fourth is bounded for sufficiently large n by Assumption 10 and the fifth is $\mathscr{O}_p\left(\max\left\{\frac{p_n}{h_n}, \frac{p_n^{\frac{1}{2}} k_n a_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}\right\}\right)$ by (2.B.20). The total order of the first term on the RHS of (2.B.25) is the order of the last displayed expression times τ_n, which is $\mathscr{O}_p\left(\max\left\{\frac{p_n^{\frac{1}{2}} k_n a_n^{\frac{1}{2}} c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}, \frac{p_n c_n^{\frac{1}{2}}}{h_n}, \frac{p_n k_n a_n^{\frac{1}{2}}}{h_n}, \frac{n^{\frac{1}{2}} p_n^{\frac{3}{2}}}{h_n^2}\right\}\right)$, all of which are negligible by (2.3.17) because

$$\begin{aligned} \frac{p_n k_n^2 a_n c_n}{n} & \leq C\left(\frac{p_n^3 k_n^4 + p_n^2 k_n^5}{n}\right), \; \frac{p_n^2 c_n}{h_n^2} \leq C \frac{p_n^3 k_n^2}{h_n^2}, \; \frac{n p_n^3}{h_n^4} = n^2 \frac{p_n^2}{h_n^4} \frac{p_n}{n} \\ \frac{p_n^2 k_n^2 a_n}{h_n^2} & = \frac{p_n^3 k_n^2 + p_n^2 k_n^3}{h_n^2} = n \frac{p_n}{h_n^2}\left(\frac{p_n^2 k_n^2}{n} + \frac{p_n k_n^3}{n}\right). \end{aligned}$$

The second term on the RHS of (2.B.25) is

$$\tau_n \alpha' \Psi_n L_n^{-1} w_n = \tau_n \alpha' \Psi_n L_n^{-1} \frac{1}{n} \begin{bmatrix} A_n' \\ X_n' \end{bmatrix} U_n + \tau_n \alpha' \Psi_n L_n^{-1} \frac{1}{n} \begin{bmatrix} B_n' \\ 0 \end{bmatrix} U_n. \tag{2.B.26}$$

The modulus of the second term on the RHS of (2.B.26) is bounded by τ_n times

$$\|\alpha\| \, \|\Psi_n\| \; \| L_n^{-1} \| \; \left\| \frac{1}{n} \begin{bmatrix} B_n' \\ 0 \end{bmatrix} U_n \right\| . \tag{2.B.27}$$

The second factor on the RHS above is $\mathscr{O}\left(a_n^{\frac{1}{2}}\right)$, the third is bounded for sufficiently large n by Assumption 10, and the fourth is $\mathscr{O}_p\left(\frac{p_n^{\frac{1}{2}}}{h_n}\right)$ by (2.B.15). Therefore (2.B.27) is $\mathscr{O}_p\left(\frac{p_n^{\frac{1}{2}} a_n^{\frac{1}{2}}}{h_n}\right)$ and so the modulus of the second term on the RHS of (2.B.26) is $\mathscr{O}_p\left(n^{\frac{1}{2}} \frac{p_n^{\frac{1}{2}}}{h_n}\right)$. Under (2.3.17) this is negligible in probability and so we need to compute only the asymptotic distribution of the first term in (2.B.26). Now

$$\tau_n \alpha' \Psi_n L_n^{-1} t_n = \frac{1}{n^{\frac{1}{2}} a_n^{\frac{1}{2}}} \sum_{i=1}^{n} \alpha' \Psi_n L_n^{-1} t_{in} u_i$$

has mean zero and variance

$$\frac{\sigma^2}{na_n}\sum_{i=1}^{n}(\alpha'\Psi_n L_n^{-1}t_{in})^2,$$

where t'_{in} is the $i-th$ row of (A_n, X_n). Thus consider

$$\frac{n\alpha'\Psi_n L_n^{-1}t_n}{\sigma\left\{\sum_{i=1}^{n}(\alpha'\Psi_n L_n^{-1}t_{in})^2\right\}^{\frac{1}{2}}}=\sum_{i=1}^{n}f_{in}u_i,$$

where

$$f_{in}=\frac{\alpha'\Psi_n L_n^{-1}t_{in}}{\sigma\left\{\sum_{i=1}^{n}(\alpha'\Psi_n L_n^{-1}t_{in})^2\right\}^{\frac{1}{2}}}.$$

We now verify the Lindeberg condition for $f_{in}u_i$. We have

$$\sum_{i=1}^{n}\mathbb{E}\left\{(f_{in}u_i)^2\mathbf{1}(|f_{in}u_i|>\epsilon)\right\}\le\max_{1\le i\le n}\mathbb{E}\left\{u_i^2\mathbf{1}\left(u_i^2>\frac{\epsilon^2}{\max\limits_{1\le i\le n}f_{in}^2}\right)\right\}\sum_{i=1}^{n}f_{in}^2$$

Note that assuming 4th moments for the u_i ensures that u_i^2 are uniformly integrable. Therefore it is sufficient to show that $\max\limits_{1\le i\le n}f_{in}^2\to 0$ as $n\to\infty$, as the last factor on the RHS of the above displayed inequality equals $1/\sigma^2$. Consider

$$\begin{aligned}\max_{1\le i\le n}f_{in}^2 &= \max_{1\le i\le n}\frac{(\alpha'\Psi_n L_n^{-1}t_{in})^2}{\sigma^2\sum_{i=1}^{n}(\alpha'\Psi_n L_n^{-1}t_{in})^2}\\ &\le \frac{\left\|L_n^{-1}\right\|^2\left\|\Psi'_n\alpha\right\|^2\max\limits_{1\le i\le n}\left\|t_{in}\right\|^2}{\sigma^2\sum_{i=1}^{n}(\alpha'\Psi_n L_n^{-1}t_{in})^2}.\end{aligned}\tag{2.B.28}$$

For the denominator of (2.B.28), note that

$$\begin{aligned}\sum_{i=1}^{n}(\alpha'\Psi_n L_n^{-1}t_{in})^2 &= \alpha'\Psi_n L_n^{-1}\sum_{i=1}^{n}t_{in}t'_{in}L_n^{-1}\Psi'_n\alpha\\ &\ge n\left\|\Psi'_n\alpha\right\|^2(\overline{\eta}(L_n))^{-1}\ge nc\left\|\Psi'_n\alpha\right\|^2,\end{aligned}$$

using Assumption 9. Thus (2.B.28) is $\mathscr{O}\left(\frac{c_n}{n}\right)$ by Assumptions 5, 9 and Lemma 2.C5. This is negligible by (2.3.17) and therefore the Lindeberg condition is satisfied. The proof of the consistency of the covariance matrix estimate is omitted. $\square$

2.C Technical lemmas

Lemma 2.C1. *Let Assumptions 3 and 4 hold. Then,*

1. $\|G_{in}\|_R$ *and* $\|G_{in}\|_C$ *are uniformly bounded for all* $i = 1, \ldots, p_n$ *and* $n \geq 1$.

2. $\|G_{in}\|$ *is uniformly bounded for all* $i = 1, \ldots, p_n$ *and* $n \geq 1$.

Proof. 1. For any $i = 1, \ldots, p_n$,

$$\|G_{in}\|_R = \left\| S_n^{-1} W_{in} \right\|_R \leq \left\| S_n^{-1} \right\|_R \|W_{in}\|_R \leq C$$

where the last inequality follows from Assumption 4. The claim for the maximum column-sum norm follows similarly.
2. Follows using (2.A.1). □

Lemma 2.C2. *Let Assumptions 2, 3 and 4 hold. Then, for all* $i = 1, \ldots, p_n$, *the elements of* G_{in} *are uniformly* $\mathscr{O}\left(\frac{1}{h_n}\right)$ *as* $n \to \infty$.

Proof. Denote by $w'_{j,in}$ the j-th row of W_{in}. Then the (j,k)-th element of G_{in} is given by $w'_{j,in} S_n^{-1} e_{k,n}$, where $e_{k,n}$ is the n-dimensional vector with unity in the k-th position and zeros elsewhere. Then

$$\begin{aligned} \left| w'_{j,in} S_n^{-1} e_{k,n} \right| = \left\| w'_{j,in} S_n^{-1} e_{k,n} \right\|_C &\leq \left\| w'_{j,in} \right\|_C \left\| S_n^{-1} \right\|_C \|e_{k,n}\|_C \\ &= \mathscr{O}\left(\frac{1}{h_n}\right). \end{aligned}$$

where the last inequality follows from Assumptions 2 and 4. □

Lemma 2.C3. *Let Assumptions 2, 3 and 4 hold. Then, for all* $i = 1, \ldots, p_n$, *the elements of a product consisting of any finite number of the* G_{in} *or their transposes are uniformly* $\mathscr{O}\left(\frac{1}{h_n}\right)$ *as* $n \to \infty$. *In particular* $G'_{in} G_{jn}$ *and* $G'_{in} G_{jn}$ *have elements that are* $\mathscr{O}\left(\frac{1}{h_n}\right)$ *uniformly in* $i, j = 1, \ldots, p_n$ *as* $n \to \infty$.

Proof. Similar to proof of Lemma 2.C2. □

Lemma 2.C4. *Suppose that* v_n *is a* $n \times 1$ *random vector with i.i.d. elements* v_{in} *with zero mean and finite fourth moment. Let* D_n *be a* $n \times n$ *non-random matrix with elements* $d_{ij,n}$. *Denote* $\nu_4 = \mathbb{E} v_{in}^4$ *and* $\vartheta^2 = \mathbb{E} v_{in}^2$. *Then*

$$var\left(v'_n D_n v_n\right) = \left(\nu_4 - 3\vartheta^4\right) \sum_{i=1}^{n} d_{ii,n}^2 + \vartheta^4 \left[tr(D_n D'_n) + tr\left(D_n^2\right)\right].$$

Proof. See Lee (2004). □

Lemma 2.C5. *Let Assumptions 3-5 hold. Then the elements of A_n are uniformly $\mathscr{O}(k_n)$.*

Proof. Let $g'_{i,jn}$ be the $i-th$ row of G_{jn}. Then a typical $(i,j)-th$ element of A_n is $g'_{i,jn}X_n\beta$. Now $\left|g'_{i,jn}X_n\beta\right| \leq \left\|g'_{i,jn}\right\|_R \|X_n\beta\|_R = \mathscr{O}(k_n)$ since $\|G_{jn}\|_R$ is uniformly bounded by Lemma (2.C1) and by Assumption 5. □

Lemma 2.C6. *Let Assumptions 2, 3 and 4 hold. Then, for all $i = 1, \ldots, p_n$, the elements of C_{in} are uniformly $\mathscr{O}\left(\frac{1}{h_n}\right)$.*

Proof. Follows trivially from Lemma 2.C2. □

2.D Proofs of sundry claims

Proposition 2.7. *A sufficient condition for invertibility of $S_n(\lambda_{(n)})$ when $\|W_{in}\|_R \leq 1$ for each $i = 1, \ldots, p_n$ and have a single non-zero diagonal block structure is that $|\lambda_{in}| < 1$ for each $i = 1, \ldots, p_n$.*

Proof. Let each W_{in} have a single non-zero diagonal block of dimension $q \times q$. Since $S_n(\lambda_{(n)})$ is block-diagonal, invertibility can be proved by showing that each block is invertible. Let B_{in} denote the i^{th} block in W_{in}, i.e. this is the only non-zero block in W_{in}. Then $S_n(\lambda_{(n)}) = I_n - diag(\lambda_{1n}B_{1n}, \ldots.., \lambda_{p_n n}B_{pn})$.

By the normalization of diagonal elements of each W_{in} in Assumption 2, the diagonal elements of $S_n(\lambda_{(n)})$ are 1. Consider the i^{th} block in $S_n(\lambda)$. Then

$$\sum_{s\neq r} |\lambda_{in}||w_{rs,in}| \sum_{m\neq l} |\lambda_{in}||w_{lm,in}| < 1$$

if $\lambda_{in}^2 < 1$, by row-normalization. The claim follows from Horn and Johnson (1985), page 381, Corollary 6.4.11 (b)). □

Proposition 2.8. *An analogous result to Theorem 2.6 is not possible with only Assumption 1 holding true (i.e. without fourth moments).*

Proof. We demonstrate this keeping k_n fixed for simplicity. Note that in this case $a_n/p_n,\ c_n/p_n \to 1$ as $n \to \infty$. If only Assumption 1 holds then (2.B.23) implies that the bound for the first term on the RHS of (2.B.25) worsens from

$$\mathscr{O}_p\left(\max\left\{\frac{p_n^{\frac{3}{2}}}{n^{\frac{1}{2}}}, \frac{p_n^{\frac{3}{2}}}{h_n}, \frac{n^{\frac{1}{2}}p_n^{\frac{3}{2}}}{h_n^2}\right\}\right)$$

to

$$\mathscr{O}_p\left(\max\left\{\frac{p_n^{\frac{3}{2}}}{n^{\frac{1}{2}}}, \frac{p_n^{\frac{3}{2}}}{h_n^{\frac{1}{2}}}, \frac{n^{\frac{1}{2}}p_n^{\frac{3}{2}}}{h_n^{\frac{3}{2}}}\right\}\right).$$

Even if suitable conditions are assumed to ensure that these are negligible, the bound for the second term on the RHS of (2.B.26) also worsens to $\mathscr{O}_p\left(n^{\frac{1}{2}}\frac{p_n^{\frac{1}{2}}}{h_n^{\frac{1}{2}}}\right)$. For this to be negligible it is required that $n\frac{p_n}{h_n} \to 0$ as $n \to \infty$ which is impossible as this equals $p_n\frac{n}{h_n}$, which is the the product of two divergent sequences. Hence the $n^{\frac{1}{2}}/p_n^{\frac{1}{2}}$-norming is not appropriate. Suppose then that we norm by $n^{\varphi}/p_n^{\frac{1}{2}}$, where $0 < \varphi < 1/2$. Now the bound for the second term on the RHS of (2.B.26) becomes $\mathscr{O}_p\left(n^{\varphi}\frac{p_n^{\frac{1}{2}}}{h_n^{\frac{1}{2}}}\right)$, where

$$n^{\varphi}\frac{p_n^{\frac{1}{2}}}{h_n^{\frac{1}{2}}} = \frac{n^{\frac{1}{2}}}{h_n^{\frac{1}{2}}}\frac{p_n^{\frac{1}{2}}}{n^{\frac{1}{2}-\varphi}}.$$

Since the first factor on the RHS above diverges, a necessary condition for the term to be negligible is that the second factor must converge to zero. But if the latter is assumed the first term on the RHS of (2.B.26) converges to zero in probability as it is $\mathscr{O}_p\left(\frac{p_n^{\frac{1}{2}}}{n^{\frac{1}{2}-\varphi}}\right)$ using (2.B.13), implying a degenerate distribution. $\square$

3 Pseudo maximum likelihood estimation of higher-order SAR models

3.1 Introduction

Maximum likelihood estimation has long been considered appropriate for (1.2.1), starting with the work of Cliff and Ord (1973). We concentrate on Gaussian pseudo-maximum likelihood estimation (PMLE), where a Gaussian likelihood is employed but Gaussianity is not actually assumed. In particular, this means that the parameters of interest must be identifiable from the first two moments of y_n.

Define the Gaussian log-likelihood function as

$$\mathcal{Q}_n\left(\theta_{(n)}, \sigma^2\right) = \log\left(2\pi\sigma^2\right) - \frac{2}{n}\log\left|S_n\left(\lambda_{(n)}\right)\right| + \frac{1}{n\sigma^2} y_n' S_n\left(\lambda_{(n)}\right) M_n S_n\left(\lambda_{(n)}\right) y_n, \quad (3.1.1)$$

where $M_n = I_n - X_n\left(X_n' X_n\right)^{-1} X_n'$. Lee (2004) studies the asymptotic properties of the PMLE for the model (1.2.1) in detail, for the case $p = 1$. A well-known drawback of using the PMLE method is the considerable computational cost due to the inversion of an $n \times n$ matrix in the computation of the Jacobean term $\log\left|S_n\left(\lambda_{(n)}\right)\right|$. While this problem can be somewhat alleviated (see e.g. Pace and Barry (1997)) by taking advantage of sparsity in $S_n\left(\lambda_{(n)}\right)$, the computational burden is still high due to the estimate being implicitly defined.

In this chapter we first analyse the properties of the PMLE when $p > 1$ by means of theoretical results describing conditions for the consistency of the PMLE (Section 3.2). This is done for both SAR and pure SAR models. In a Monte Carlo study in Section 3.3 we find that there are problems with using this estimate due to identification problems in even reasonable sample sizes. The theoretical results are compared to the results in Chapter 2 to attempt to explain this behaviour. In Section 3.4 we propose a one Newton-type step approximation to the Gaussian PMLE, starting from initially consistent estimators such as the IV or OLS ones considered in Chapter 2. This has the advantage of providing a closed-form estimate with the same asymptotic properties as the PMLE. Finite sample properties of such estimates are examined in a Monte Carlo study.

We do not consider the Spatial Moving Average (SMA) or the Spatial ARMA models in this chapter. However, the exploration of PML estimation for these models is a natural step from the results of this chapter. For a discussion of the definition and estimation of the SMA model, see Haining (1978). Yao and Brockwell (2006) present theory for the estimation of a spatial ARMA model, but they consider processes defined on a regularly-spaced lattice.

In this chapter, it important to distinguish between true and admissible parameters, so we denote the true parameters with a 0 subscript. A notational convention for evaluation of objects at the true parameters is also introduced. In general this is of the form $A(\delta_0) \equiv A$ for any matrix or vector A and any true parameter δ_0. For instance, $G_{in}\left(\lambda_{(n)}\right)$ now indicates evaluation at an admissible $\lambda_{(n)}$ whereas G_{in} is the result of evaluation at the true parameter $\lambda_{0(n)}$. In addition we suppress reference to n for individual parameters to simplify notation.

3.2 Pseudo ML estimation of Higher-Order SAR Models

In this section sufficient conditions are provided for consistency of of estimates based on the minimization of (3.1.1). We first analyse models with regressors, and then consider Pure SAR models. Of course, for the former IV and OLS estimates are available but we provide conditions for the consistency of the PMLE to compare these to those of Theorems 2.1 and 2.4.

3.2.1 Mixed-Regressive SAR Models

In the case of the model (2.2.1), we will work with the concentrated likelihood obtained by concentrating out $\beta_{(n)}$ and σ^2. This has advantages in terms of not only reduction in computational burden, but also analytical ease. From a technical standpoint, concentrating out these parameters enables us to avoid compactness assumptions on their parameter spaces, these being standard requirements for definitions of implicitly defined estimates. Concentrating out $\beta_{(n)}$ and σ^2 yields

$$\check{\beta}_{(n)}\left(\lambda_{(n)}\right) = \left(X_n' X_n\right)^{-1} X_n' S_n\left(\lambda_{(n)}\right) y_n \tag{3.2.1}$$

$$\check{\sigma}^2_{(n)}\left(\lambda_{(n)}\right) = \frac{1}{n} y_n' S_n'\left(\lambda_{(n)}\right) M_n S_n\left(\lambda_{(n)}\right) y_n, \tag{3.2.2}$$

The concentrated log-likelihood function (of $\lambda_{(n)}$) is

$$\mathcal{Q}_n^c\left(\lambda_{(n)}\right) = \log \check{\sigma}^2_{(n)}\left(\lambda_{(n)}\right) + \frac{1}{n}\log\left|T_n\left(\lambda_{(n)}\right) T_n'\left(\lambda_{(n)}\right)\right|, \tag{3.2.3}$$

where $T_n\left(\lambda_{(n)}\right) = S_n^{-1}\left(\lambda_{(n)}\right)$. The PMLE of $\lambda_{(n)}$ is defined as

$$\check{\lambda}_{(n)} = \underset{\lambda_{(n)}\in\Lambda_n}{\arg\min}\, \mathcal{Q}_n^c\left(\lambda_{(n)}\right). \tag{3.2.4}$$

The PMLEs of $\beta_{(n)}$ and σ^2 are defined as $\check{\beta}_{(n)}\left(\check{\lambda}_{(n)}\right)$ and $\check{\sigma}^2_{(n)}\left(\check{\lambda}_{(n)}\right)$ respectively.

Assumption 12. Λ_n is a compact subset of the p_n-fold Cartesian product of the open interval $(-1, 1)$. In particular it is assumed that there exist real numbers k_1 and k_2 such that $k_1 \leq \lambda_i \leq k_2$, for all $i = 1, \ldots, p_n$, and with $-1 < k_1 < k_2 < 1$.

Assumption 13. $\lambda_{0(n)} \in \Lambda_n$, for all sufficiently large n.

Assumptions 12 and 13 are standard for proving the consistency of implicity defined estimates. Assuming that each λ_i, $i = 1, \ldots, p_n$, lies in a closed interval inside (-1,1) is sufficient to ensure compactness of the parameter space, by the Heine-Borel Theorem and Tychonoff's Theorem (see e.g. Munkres (2000)). Also define

$$\sigma_n^2\left(\lambda_{(n)}\right) = \frac{\sigma_0^2}{n}\operatorname{tr}\left(T_n' S_n'\left(\lambda_{(n)}\right) S_n\left(\lambda_{(n)}\right) T_n\right)$$

and

$$\Theta_n\left(\lambda_{(n)}; \sigma_0^2\right) = \sigma_n^2\left(\lambda_{(n)}\right) T_n\left(\lambda_{(n)}\right) T_n'\left(\lambda_{(n)}\right)$$

so that

$$\Theta_n = \sigma_0^2 T_n T_n'.$$

Define $P_{j'i,n}$ to be the $p_n \times p_n$ matrix with (i, j)-th element $\operatorname{tr}\left(G_{jn}' G_{in}\right)$. Also write f_n for the $p_n \times 1$ vector with i-th element $\operatorname{tr} G_{in}$ and introduce

Assumption 14.

$$\lim_{n\to\infty} \frac{1}{n} f_n' P_{j'i,n}^{-1} f_n < 1. \tag{3.2.5}$$

By the proof of Lemma 3.1 in the appendix,

$$1 - \lim_{n\to\infty} \frac{f_n' P_{j'i,n}^{-1} f_n}{n} \geq 0$$

is always satisfied since it is proportional to a sum of squares. We assume that this limit is strictly bounded below by zero. This can be checked in the case $p_n = 1$. Indeed, in this case (3.2.5) becomes

$$\lim_{n\to\infty} \frac{1}{n} \frac{tr^2 G_n}{tr\left(G_n G_n'\right)} < 1. \tag{3.2.6}$$

The matrix trace Cauchy-Schwarz inequality (see e.g. Liu and Neudecker (1995)) implies that

$$tr^2 G_n \leq tr\left(G_n G_n'\right) tr I_n = n tr\left(G_n G_n'\right) \tag{3.2.7}$$

with equality if and only if

$$G_n = \psi I_n \tag{3.2.8}$$

for some scalar ψ. But if (3.2.8) holds then $W_n S_n^{-1} = \psi I_n$, implying that

$$W_n = \frac{\psi}{1+\psi\lambda} I_n,$$

which is a contradiction unless $\psi = 0$ because the diagonal elements of W_n are normalised to 0 (see Assumption 2). As a result, there cannot exist a natural number n_0 such that (3.2.7) holds with equality, implying that $1 - \frac{1}{n}\frac{tr^2 G_n}{tr(G_n G_n')} \geq c$ whence (3.2.6) follows.

Assumption 15. The limits

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{p_n} \mathrm{tr} G_{in}$$

and

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{p_n} \sum_{j=1}^{p_n} \mathrm{tr}\left(G_{in}' G_{jn}\right)$$

exist and are finite.

Lemma 3.1. *Suppose that Assumptions 14 and 15 hold. Then, for all sufficiently large n,*

$$c \leq \sigma_n^2\left(\lambda_{(n)}\right) \leq C,$$

where c and C are positive constants that do not depend on n or $\lambda_{(n)}$.

We now introduce assumptions needed for the identification of $\lambda_{0(n)}$.

Assumption 16. The limit

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{p_n} \sum_{j=1}^{p_n} \sum_{k=1}^{p_n} \sum_{l=1}^{p_n} b_{in'} M_n G_{jn} G_{kn}' M_n b_{ln}$$

exists and is finite.

Also, note that Assumption 10 implies that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{p_n} \sum_{j=1}^{p_n} b_{in'} M_n b_{jn} > 0 \tag{3.2.9}$$

because, using the partitioned matrix inversion formula, Assumption 10 implies that

$$\lim_{n\to\infty} \frac{1}{n} x_n' A_n' M_n A_n x_n > 0, \text{ for } x_n \neq 0,$$

so that choosing $x_n = l_{p_n}$ (the p_n-dimensional vector of ones) yields (3.2.9). Defining

$$H_n\left(\lambda_{(n)}\right) = \Theta_n^{-\frac{1}{2}}\left(\lambda_{(n)}; \sigma_0^2\right)\Theta_n \Theta_n^{-\frac{1}{2}}\left(\lambda_{(n)}; \sigma_0^2\right)$$

and writing

$$r\left(\lambda_{(n)}\right) = \frac{1}{n}\mathrm{tr} H_n\left(\lambda_{(n)}\right) - \frac{1}{n}\log\left|H_n\left(\lambda_{(n)}\right)\right| - 1,$$

we introduce

Assumption 17. For any $\delta > 0$,

$$\varliminf_{n\to\infty} \inf_{\left\{\left\|\lambda_{(n)}-\lambda_{0(n)}\right\|>\delta\right\}\cap\Lambda_n} r\left(\lambda_{(n)}\right) > 0.$$

We can write

$$r\left(\lambda_{(n)}\right) = \frac{1}{n}\sum_{i=1}^{n}\left(\eta_i - \log\eta_i - 1\right), \tag{3.2.10}$$

where η_i are eigenvalues of $H_n\left(\lambda_{(n)}\right)$. Because $H_n\left(\lambda_{(n)}\right)$ is positive definite, the η_i, $i = 1, \ldots, n$, are positive, and for all i, the i-th summand in (3.2.10) is non-negative, and positive $\eta_i \neq 1$. Since $\eta_i = 1$ for all n only when $\Theta_n\left(\lambda_{(n)}; \sigma_0^2\right) = \Theta_n$, so that Assumption 17 is an identification condition related to the uniqueness of the covariance matrix of y_n. Lee (2004) employs a similar assumption in his asymptotic theory but expressed in a somewhat different way.

Also, we have

$$\mathcal{Q}_n^c = \log\check{\sigma}_{(n)}^2 + \frac{1}{n}\log\left|T_n T_n'\right|, \tag{3.2.11}$$

where

$$\begin{aligned}\check{\sigma}_{(n)}^2 &= \frac{y_n' S_n' M_n S_n y_n}{n} = \frac{U_n' M_n U_n}{n} \\ &= \frac{U_n' U_n}{n} - \frac{U_n' X_n}{n}\left(\frac{X_n' X_n}{n}\right)^{-1}\frac{X_n' U_n}{n} = \sigma_0^2 + o_p(1),\end{aligned}$$

if $k_n/n \to 0$ as $n \to \infty$ because

$$\mathbb{E}\left\|\frac{X_n' U_n}{n}\right\|^2 = \mathcal{O}\left(\frac{k_n}{n}\right),$$

by Assumptions 1 and 5. Thus (3.2.11) becomes

$$\mathcal{Q}_n^c = \log\sigma_0^2 + \frac{1}{n}\log\left|T_n T_n'\right| + o_p^\lambda(1) = \frac{1}{n}\log\left|\Theta_n\right| + o_p^\lambda(1), \tag{3.2.12}$$

where the $o_p^\lambda(1)$ signifies a uniform order in $\lambda_{(n)} \in \Lambda_n$.

Theorem 3.1. *Suppose that Assumptions 1-5, and 10-17 hold together with*

$$\frac{1}{p_n} + \frac{1}{k_n} + \frac{p_n^2 k_n^2}{n^{\frac{1}{2}}} \to 0 \text{ as } n \to \infty. \tag{3.2.13}$$

Then

$$\left\|\check{\theta}_{(n)} - \theta_{0(n)}\right\| \xrightarrow{p} 0.$$

The conditions of the theorem can be compared to those of Theorem 2.1 and 2.4. The requirement of finite fourth order moments is not imposed for consistency of the IV and OLS estimator, where second moments suffice. Fourth moments were assumed to exist to establish asymptotic normality of the OLS estimate, but we had mentioned that these were not required for the consistency result. On the other hand, the only restriction imposed on h_n here is that it be bounded away from zero uniformly in n .

The restrictions on the rate of growth of p_n and k_n are stronger in Theorem 3.1, as compared to Theorems 2.1 and 2.4 where, with k_n fixed for illustrative purposes, it sufficed that $p_n = o\left(n^{\frac{1}{2}}\right)$ compared to $p_n = o\left(n^{\frac{1}{4}}\right)$ in this case. This is not surprising due to the implicitly defined nature of the estimate.

3.2.2 Pure SAR Models

We now consider the SAR model without regressors, given by

$$y_n = \sum_{i=1}^{p_n} \lambda_{in} W_{in} y_n + U_n. \tag{3.2.14}$$

The Gaussian pseudo-likelihood function is now

$$\mathcal{Q}_n^p\left(\lambda_{(n)}, \sigma^2\right) = \log\left(2\pi\sigma^2\right) - \frac{2}{n}\log\left|S_n\left(\lambda_{(n)}\right)\right| + \frac{1}{n\sigma^2} y_n' S_n\left(\lambda_{(n)}\right) S_n\left(\lambda_{(n)}\right) y_n, \tag{3.2.15}$$

while concentrating out σ^2 yields

$$\check{\sigma}_{(n)}^{2,p}\left(\lambda_{(n)}\right) = \frac{1}{n} y_n' S_n\left(\lambda_{(n)}\right) S_n\left(\lambda_{(n)}\right) y_n, \tag{3.2.16}$$

implying that the concentrated likelihood is

$$\mathcal{Q}_n^{p,c}\left(\lambda_{(n)}\right) = \log \check{\sigma}_{(n)}^{2,p} + \frac{1}{n}\log\left|T_n\left(\lambda_{(n)}\right) T_n'\left(\lambda_{(n)}\right)\right|. \tag{3.2.17}$$

Define the PMLE of $\lambda_{(n)}$ as

$$\check{\lambda}_{(n)}^p = \underset{\lambda_{(n)} \in \Lambda_n}{\arg\min}\, \mathcal{Q}_n^{p,c}\left(\lambda_{(n)}\right). \tag{3.2.18}$$

Note that now the PMLE of σ^2 is $\check{\sigma}^{2,p}_{(n)}\left(\check{\lambda}^p_{(n)}\right)$.

Theorem 3.2. *Suppose that Assumptions 1-4, 11-15 and 17 hold together with*

$$\frac{1}{p_n} + \frac{p_n^2}{n^{\frac{1}{2}}} \to 0 \text{ as } n \to \infty. \tag{3.2.19}$$

Then

$$\left\|\check{\lambda}^p_{(n)} - \lambda_{0(n)}\right\| \xrightarrow{p} 0.$$

Theorem 3.2 may be viewed as a particular case of Theorem 3.1 with $k_n = 0$. The conditions of the theorem are amended accordingly.

3.3 Finite-sample performance of PMLE

In this section we study the finite-sample properties of the Pseudo ML estimates defined above. We focus on the spatial scenario of Case (1991, 1992), described in Chapter 1. The model considered is the biparametric mixed-regressive SAR

$$y_n = \lambda_1 W_{1n} y_n + \lambda_2 W_{2n} y_n + X_n \beta + U_n,$$

with $U_n \sim N(0, \sigma^2 I_n)$, so that the estimates are in fact MLE. The weighting matrices are given as in (1.2.4). The regressors are generated from a uniform distribution on $(0, 1)$ and then kept fixed to reflect the non-stochastic nature of Assumption 5. We generate data from $\lambda_1 = 0.70$ and $\lambda_2 = 0.80$, with two regressors included and $\beta_1 = 1$, $\beta_2 = 0.50$. In addition, we set $\sigma^2 = 1$. We experiment with $m = 50,\ 150,\ 300$. Note here that we are simply considering (1.2.6) with $r = 2$, so that $n = 100,\ 300,\ 600$. There are 500 replications for each case. Tables 3.1-3.3 present the results of our experiment.

We first discuss the results in Table 3.1, which reports the empirical mean and bias for each parameter estimate. The estimates of λ_1 and λ_2 are very poor, and exhibit high (negative) bias. The estimates of β_1, β_2 are rather good. However, as we increase m, we see that the estimates of λ_1 and λ_2 become somewhat better for the former but exhibit no improvement for the latter. In particular, it is interesting to note that increasing m does not significantly improve the estimates of the spatial parameters at least up to $m = 150$. Indeed Lee (2004) showed that the MLE is inconsistent under if r is fixed while m diverges, while simulations conducted by Hillier and Martellosio (2013) illustrate that the estimate is centred around the true value with a non-degenerate distribution. Results of this type have counterparts in the spatial statistics literature, where asymptotics when observations become dense in a bounded region is called 'infill-asymptotics'. Asymptotics under such conditions can lead to inconsistent estimation of parameters of interest and non-standard limiting behaviour of the estimates, see

$\frac{m}{r}$		25		75		150	
r		Mean	Bias	Mean	Bias	Mean	Bias
2	λ_1	0.3226	-0.3774	0.3182	-0.3818	0.4730	-0.2270
	λ_2	0.4708	-0.3292	0.4786	-0.3214	0.4565	-0.3435
	β_1	0.9924	-0.0076	1.0004	0.0004	1.0024	0.0024
	β_2	0.4982	-0.0018	0.5002	0.0002	0.4981	-0.0019

Table 3.1: Monte Carlo Mean and Bias of ML Estimates $\check{\theta}_{(n)}$.

$\frac{m}{r}$		25		75		150	
r		$\hat{\theta}_{(n)}$	$\check{\theta}_{(n)}$	$\hat{\theta}_{(n)}$	$\check{\theta}_{(n)}$	$\hat{\theta}_{(n)}$	$\check{\theta}_{(n)}$
2	λ_1	-0.0090	-0.3774	-0.0046	-0.3818	0.0003	-0.2270
	λ_2	-0.0041	-0.3292	-0.0019	-0.3214	0.0005	-0.3435
	β_1	-0.0096	-0.0076	-0.0025	0.0004	-0.0026	0.0024
	β_2	-0.0033	-0.0018	0.0012	0.0002	-0.0044	-0.0019

Table 3.2: Monte Carlo Bias of IV and ML estimates $\hat{\theta}_{(n)}$ and $\check{\theta}_{(n)}$

e.g. Lahiri (1996). On the other hand, the block-diagonality of the model implies that the number of observations available to estimate λ_1 and λ_2 increases one-to-one with m. We carried out further experiments with larger m which revealed better estimates of the autoregressive parameters (see also discussion of Figure 3.3 below), but we do not report these as our interest lies in comparing the properties of the MLE to those of the IV and one-step estimates in smallish samples, besides the fact that under such circumstances estimates are not consistent in view of the discussion above.

In fact, there seem to be some identifiability problems when m is not very large relative to r in (1.2.6), even though both need to increase to avoid a problem with infill-asymptotics. The likelihood-surface has a distinct ridge, rather than a peak, leading to poor estimates for the spatial parameters. This problem and the improvements by increasing m for fixed r are illustrated in Figures 3.1-3.3. Figure 3.1 has $m = 50$, and clearly shows the ridge that causes the identifiability problems. Figure 3.2 has $m = 150$ and shows a rather better defined peak, with the situation improving further in Figure 3.3 where $m = 300$. The figures should only be interpreted in terms of the parameters being centred around the true values under fixed r asymptotics, and therefore do not indicate that the estimates are consistent.

These concerns indicate that Pseudo ML estimates are not reliable for higher-order SAR models even when $p = 2$ in (1.2.1). It is clear that from the above discussion that estimates improve very slowly, and so it can be anticipated that the problems

$\frac{m}{r}$		25	75	150
r		$\frac{MSE(\hat{\theta}_{(n)})}{MSE(\check{\theta}_{(n)})}$	$\frac{MSE(\hat{\theta}_{(n)})}{MSE(\check{\theta}_{(n)})}$	$\frac{MSE(\hat{\theta}_{(n)})}{MSE(\check{\theta}_{(n)})}$
2	λ_1	0.0360	0.0089	0.0088
	λ_2	0.0162	0.0046	0.0021
	β_1	14.1809	12.6237	10.8006
	β_2	15.3910	15.0110	11.4652

Table 3.3: Monte Carlo Relative MSE of IV and ML estimates, $MSE\left(\hat{\theta}_{(n)}\right)/MSE\left(\check{\theta}_{(n)}\right)$

$\frac{m}{r}$		25	75	150
r		$\frac{Var(\hat{\theta}_{(n)})}{Var(\check{\theta}_{(n)})}$	$\frac{Var(\hat{\theta}_{(n)})}{Var(\check{\theta}_{(n)})}$	$\frac{Var(\hat{\theta}_{(n)})}{Var(\check{\theta}_{(n)})}$
2	λ_1	0.0518	0.0126	0.0112
	λ_2	0.0222	0.0063	0.0029
	β_1	14.2650	12.6224	10.8325
	β_2	15.3951	15.0109	11.4770

Table 3.4: Monte Carlo Relative Variance of IV and ML estimates, $Var\left(\hat{\theta}_{(n)}\right)/Var\left(\check{\theta}_{(n)}\right)$

intensify with higher values of r. Indeed for higher-lag orders matters are even worse, with estimates bordering on the disastrous. Further simulations for $r = 4$ and $r = 6$ (with m chosen to deliver ratios of 25, 75 and 150 for m/r in each case) confirmed this, and the results are too poor to report. In addition, the experiments proved to be very expensive computationally even on very high-specification computers. Optimization routines even failed to converge in many replications. It is worth mentioning here that identification problems are more severe the closer the spatial parameters used to generate the data are to zero. Large negative biases are common, and estimates are generally volatile.

As alluded to earlier, it can be argued that that the computational burden can be lessened by taking advantage of sparsity in the weight matrices and, therefore, in the matrix to be inverted $S_n\left(\lambda_{(n)}\right)$. This will help to ease the computational cost, but will not alleviate the identification issues in reasonably sized finite-samples that have been discussed above. If extremely large data sets are available, as may be the case in the analysis of spatial data, sparse matrix routines can be employed and the identification properties will also improve. However, if there is not enough sparsity this solution may not be practical and even if there is enough sparsity explicitly defined estimates will perform better.

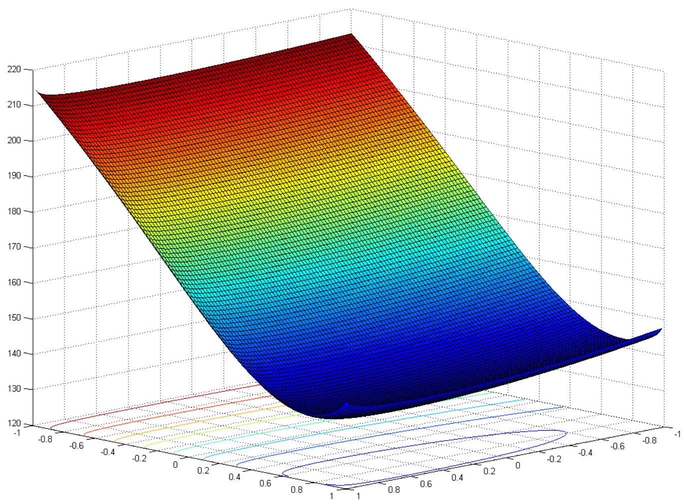

Figure 3.1: Sample Log-Likelihood Surface, $r = 2$, $m = 50$, with $\lambda_1 = 0.70$, $\lambda_2 = 0.80$.

As a result, it is natural to study alternatives to the PMLE, where available, and compare their asymptotic and finite-sample properties to those of the PMLE. This is especially crucial in applied work, as the reliability of Pseudo ML estimates has been put in some doubt by the findings of this section. The Monte Carlo study in Lee (2004) was carried out for the case $p = 1$ only, and these concerns were not flagged as a result. We conclude that it is desirable to use a closed-form estimator if one is available, as is the case with the mixed-regressive SAR model. For the pure SAR model, there is no alternative at present to the PMLE or another implicitly-defined estimate such as the GMM estimate of Kelejian and Prucha (1999).

Our theoretical results had also indicated that estimating higher-order models using the PMLE would incur a bias that vanishes at a slower rate than the bias in the IV and OLS estimators. As discussed after Theorem 3.1, this is due the fact that now the restrictions placed on the rate of growth of the parameter space are much more stringent. It is natural that this be reflected in poor finite-sample performance, as the ratio p_n^4/n declines much slower than p_n^2/n. In addition, the consistency of the PMLE requires the additional identification conditions given in Assumptions 15-17, and these will also have an impact on finite sample identification.

In Tables 3.2-3.4, we compare the IV estimate to the MLE. We omit a comparison with the OLS estimate as the results are similar and summarising these would entail unnecessary repetition. Table 3.2 compares the bias in the IV estimate and the MLE. It is clear that the IV estimate has far superior properties with respect to the spatial

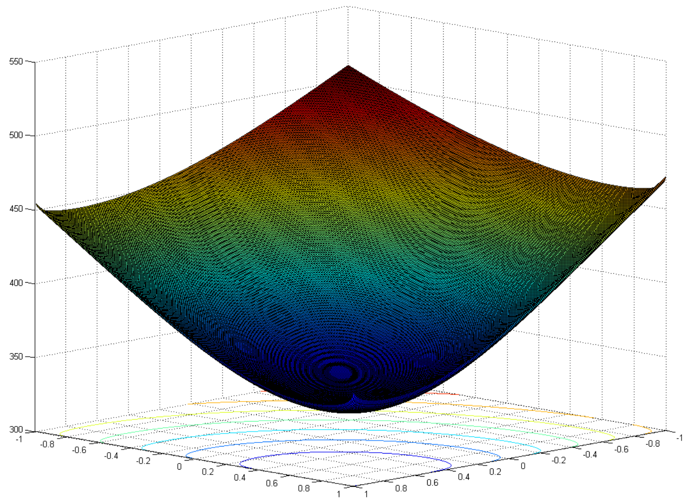

Figure 3.2: Sample Log-Likelihood Surface, $r = 2,\ m = 150$, with $\lambda_1 = 0.70$, $\lambda_2 = 0.80$.

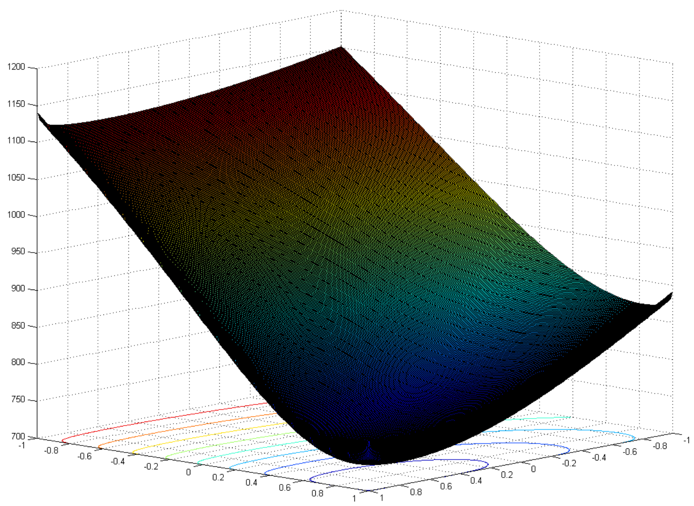

Figure 3.3: Sample Log-Likelihood Surface, $r = 2,\ m = 300$, with $\lambda_1 = 0.70$, $\lambda_2 = 0.80$.

autoregressive parameters λ_1 and λ_2 in this regard. The MLE does a better job with the regression coefficients β_1 and β_2, but the overall performance of the IV is much better as the IV estimates for the regression coefficients are more acceptable than the ML estimates of the autoregression coefficients.

In Tables 3.3 and 3.4 we report relative mean-squared error (MSE) and variance respectively. The conclusions are similar to the bias analysis. Indeed while it looks like the MLE outperforms the IV comfortably for the regression coefficients it should be noted that the much more dramatic advantage of the IV over the MLE for the spatial parameters more than compensates for this. For instance, with $m/r = 25$, we have that MSE for β_1 and β_2 are, respectively, 14 and 15 times those of the MLE for IV. However, MSE for λ_1 and λ_2 are, respectively, nearly 28 and 62 times those of the IV for MLE. Similar conclusions hold for the variance. Moreover, as m/r increases the IV estimate improves for both the autoregression and regression coefficients in both variance and MSE comparisons.

In the next two sections, we propose closed-form estimates with the same asymptotic properties as the PMLE and examine their finite-sample performance in comparison to the estimates considered in this section.

3.4 Approximations to Gaussian PMLE

Pseudo ML estimation involves a highly non-linear optimization problem and is computationally costly. The previous section also indicates substantive concerns about the performance of the PMLE in finite samples. Given $n^{\frac{1}{2}}/\left(p_n+k_n\right)^{\frac{1}{2}}$-consistent preliminary estimates as in Section 2.3, we can consider a one Newton-step approximation to the Gaussian PMLE. This has the advantage of providing a closed-form estimate with the asymptotic properties of the PMLE.

Denote

$$t_n = \frac{1}{n}\begin{bmatrix} A_n' \\ X_n' \end{bmatrix} U_n \tag{3.4.1}$$

and

$$\phi_n = \begin{pmatrix} \frac{1}{n}trC_{1n} - \frac{1}{n\sigma_0^2}U_n'C_{1n}U_n \\ \frac{1}{n}trC_{2n} - \frac{1}{n\sigma_0^2}U_n'C_{2n}U_n \\ \vdots \qquad \vdots \\ \vdots \qquad \vdots \\ \frac{1}{n}trC_{p_nn} - \frac{1}{n\sigma_0^2}U_n'C_{p_nn}U_n \\ 0 \end{pmatrix}, \tag{3.4.2}$$

where $C_{in} = G_{in} + G'_{in}$. Then

$$\xi_n \equiv \frac{\partial \mathcal{Q}_n}{\partial \theta} = \phi_n - \frac{2}{\sigma_0^2} t_n \tag{3.4.3}$$

while the Hessian at any admissible point in the parameter space is

$$\frac{\partial^2 \mathcal{Q}_n(\theta_{(n)}, \sigma^2)}{\partial\theta\partial\theta'} = \begin{pmatrix} \frac{2}{n} P_{ji,n}(\lambda_{(n)}) + \frac{2}{n\sigma^2} R'_n R_n & \frac{2}{n\sigma^2} R'_n X_n \\ \\ \frac{2}{n\sigma^2} X'_n R_n & \frac{2}{n\sigma^2} X'_n X_n \end{pmatrix} \tag{3.4.4}$$

where $P_{ji,n}(\lambda_{(n)})$ is the $p_n \times p_n$ matrix with (i,j)-th element given by $tr\left(G_{jn}(\lambda_{(n)}) G_{in}(\lambda_{(n)})\right)$. So

$$\frac{\partial^2 \mathcal{Q}_n(\theta_{0(n)}, \sigma_0^2)}{\partial\theta\partial\theta'} = \begin{pmatrix} \frac{2}{n} P_{ji,n} + \frac{2}{n\sigma_0^2} R'_n R_n & \frac{2}{n\sigma_0^2} R'_n X_n \\ \\ \frac{2}{n\sigma_0^2} X'_n R_n & \frac{2}{n\sigma_0^2} X'_n X_n \end{pmatrix}. \tag{3.4.5}$$

In Section 2.3, we stated that $\hat{\sigma}^2_{(n)}$ and $\tilde{\sigma}^2_{(n)}$ are consistent estimates of σ_0^2. For subsequent theorems the stochastic orders needed are in terms of p_n, r_n, k_n, h_n and n as opposed to simply $o_p(1)$. However we restrict reference to these orders to appendices.

Define the 'one-step' estimates $\hat{\hat{\theta}}_{(n)}$ and $\tilde{\tilde{\theta}}_{(n)}$ by the following equations

$$\begin{aligned} \hat{\hat{\theta}}_{(n)} &= \hat{\theta}_{(n)} - \hat{H}_n^{-1} \hat{\xi}_n, && (3.4.6) \\ \tilde{\tilde{\theta}}_{(n)} &= \tilde{\theta}_{(n)} - \tilde{H}_n^{-1} \tilde{\xi}_n. && (3.4.7) \end{aligned}$$

where

$$\begin{aligned} \hat{H}_n &= \frac{\partial^2 \mathcal{Q}_n(\hat{\theta}_{(n)}, \hat{\sigma}^2_{(n)})}{\partial\theta\partial\theta'} \quad , \quad \tilde{H}_n = \frac{\partial^2 \mathcal{Q}_n(\tilde{\theta}_{(n)}, \tilde{\sigma}^2_{(n)})}{\partial\theta\partial\theta'}, \\ \hat{\xi}_n &= \frac{\partial \mathcal{Q}_n(\hat{\theta}_{(n)}, \hat{\sigma}^2_{(n)})}{\partial\theta} \quad , \quad \tilde{\xi}_n = \frac{\partial \mathcal{Q}_n(\tilde{\theta}_{(n)}, \tilde{\sigma}^2_{(n)})}{\partial\theta}. \end{aligned}$$

Robinson (2010) considered estimates of the type defined above, in a more general setting where the error distribution is of unknown or perhaps known parametric form. From a practical point of view, more iterations may be desirable and could also have implications for higher-order efficiency. It should be noted that the estimates (3.4.6) and (3.4.7) incur additional bias in finite samples relative to the preliminary estimate.

Indeed by the mean value theorem (3.4.6) implies that

$$\begin{aligned} \hat{\hat{\theta}}_{(n)} - \theta_{0(n)} &= \left[I_{p_n+k_n} - \hat{H}_n^{-1}\overline{H}_n\right]\left(\hat{\theta}_{(n)} - \theta_{0(n)}\right) - \hat{H}_n^{-1}\xi_n \\ &= \hat{\theta}_{(n)} - \theta_{0(n)} - \hat{H}_n^{-1}\overline{H}_n\left(\hat{\theta}_{(n)} - \theta_{0(n)}\right) - \hat{H}_n^{-1}\xi_n \end{aligned} \tag{3.4.8}$$

where

$$\overline{H}_n = \frac{\partial^2 \mathcal{Q}_n(\overline{\theta}_{(n)}, \hat{\sigma}^2_{(n)})}{\partial\theta\partial\theta'}$$

and $\left\|\overline{\theta}_{(n)} - \theta_{0(n)}\right\| \leq \left\|\hat{\theta}_{(n)} - \theta_{0(n)}\right\|$, with each row of the Hessian matrix evaluated at possibly different $\overline{\theta}_{(n)}$. The latter point is a technical comment that we take as given in the remainder of the thesis whenever a mean-value theorem is applied to vector of values. The last two terms on the right of (3.4.8) have norm bounded by

$$\left\|\hat{H}_n^{-1}\right\|\left\|\overline{H}_n^{-1}\right\|\left\|\hat{\theta}_{(n)} - \theta_{0(n)}\right\| \tag{3.4.9}$$

and

$$\left\|\hat{H}_n^{-1}\right\|\left\|\xi_n\right\| \tag{3.4.10}$$

respectively. In the appendix we prove that $\left\|\hat{H}_n^{-1}\right\|$ and $\left\|\overline{H}_n^{-1}\right\|$ are uniformly bounded as $n \to \infty$ under extra conditions, while the third factor in (3.4.9) and the second factor in (3.4.10) are $\mathscr{O}_p\left(r_n^{\frac{1}{2}}/n^{\frac{1}{2}}\right)$ and $\mathscr{O}_p\left(\max\left\{p_n^{\frac{1}{2}}/n^{\frac{1}{2}}h_n^{\frac{1}{2}}, p_n^{\frac{1}{2}}/n^{\frac{1}{2}}\right\}\right)$ respectively. Thus the bias will decline with n under suitable conditions on the rates of p_n and r_n, but represents an additional bias as opposed to $\hat{\theta}_{(n)}$.

The computation of $tr\left(G_{in}\left(\lambda_{(n)}\right)\right)$ can be quite expensive, due to the inversion of the $n \times n$ matrix $S_n\left(\lambda_{(n)}\right)$. However, in the setting of Section 2.4.1 this expression is extremely easy to compute because

$$S_n\left(\lambda_{(n)}\right) = diag\left[I_m - \lambda_{1n}B_m, I_m - \lambda_{2n}B_m, \ldots, I_m - \lambda_{p_n n}B_m\right]$$

so that

$$G_{in}\left(\lambda_{(n)}\right) = diag\left[0, \ldots, B_m\left(I_m - \lambda_{in}B_m\right)^{-1}, \ldots, 0\right] \tag{3.4.11}$$

and

$$tr\left(G_{in}\left(\lambda_{(n)}\right)\right) = tr\left\{B_m\left(I_m - \lambda_{in}B_m\right)^{-1}\right\} = \frac{m\lambda_{in}}{(m-1+\lambda_{in})(1-\lambda_{in})}.$$

From (3.4.11) it is also obvious that $G_{jn}\left(\lambda_{(n)}\right)G_{in}\left(\lambda_{(n)}\right) = 0$ for $j \neq i$. This reduces

$P_{ji,n}(\lambda_{(n)})$ to a diagonal matrix with i-th diagonal element

$$tr\left(G_{in}\left(\lambda_{(n)}\right)^2\right) = \frac{m\left(m-1+\lambda_{in}^2\right)}{\left(m-1+\lambda_{in}\right)^2\left(1-\lambda_{in}\right)^2}.$$

Theorem 3.3. *Consider any $s\times(p_n+k_n)$ matrix of constants Ψ_n with full row-rank.*

(i) Let Assumptions 1-7 and 9-11 hold along with

$$\frac{1}{p_n}+\frac{1}{r_n}+\frac{1}{k_n}+\frac{p_n^3k_n^4}{n}+\frac{p_nr_n}{n}+\frac{p_n^{\frac{3}{2}}k_n}{h_n}\to 0 \text{ as } n\to\infty \tag{3.4.12}$$

and

$$\frac{r_n^2k_n^2}{n}+\frac{r_n^3}{np_n}+\frac{k_n^2}{p_n^{\frac{1}{2}}n^{\frac{1}{2}}}\left(r_n+k_n\right) \text{ bounded as } n\to\infty. \tag{3.4.13}$$

Then

$$\frac{n^{\frac{1}{2}}}{\left(p_n+k_n\right)^{\frac{1}{2}}}\Psi_n\left(\hat{\hat{\theta}}_{(n)}-\theta_{0(n)}\right)\xrightarrow{d} N\left(0,\lim_{n\to\infty}\frac{\sigma_0^2}{p_n+k_n}\Psi_nL_n^{-1}\Psi_n'\right),$$

where the asymptotic covariance matrix exists, and is positive definite, by Assumptions 9 and 10.

(ii) Let Assumptions 1-5 and 9-11 hold. For $\gamma\in[3/2,\infty)$, suppose also that

$$\frac{1}{p_n}+\frac{1}{k_n}+\frac{p_n^{\frac{5}{2}}k_n^4\left(p_n^{\frac{1}{2}}+k_n\right)}{n}+\frac{p_n^{\gamma}k_n^{\frac{2\gamma}{3}}}{h_n}\to 0 \tag{3.4.14}$$

and

$$\frac{k_n^{\frac{1}{3}}n^{\frac{1}{2}}}{h_n^{3-5/2\gamma}} \text{ is bounded as } n\to\infty. \tag{3.4.15}$$

Further, if $\gamma\in[3/2,9/4)$, also assume that

$$p_n^{5-4\gamma}k_n^{6-8\gamma/3} \text{ is bounded as } n\to\infty. \tag{3.4.16}$$

Then

$$\frac{n^{\frac{1}{2}}}{\left(p_n+k_n\right)^{\frac{1}{2}}}\Psi_n\left(\tilde{\tilde{\theta}}_{(n)}-\theta_{0(n)}\right)\xrightarrow{d} N\left(0,\lim_{n\to\infty}\frac{\sigma_0^2}{p_n+k_n}\Psi_nL_n^{-1}\Psi_n'\right),$$

where the asymptotic covariance matrix exists, and is positive definite, by Assumptions 9 and 10.

The asymptotic covariance matrix may be consistently estimated as in Theorem 2.6.

Corollary 3.2. *Suppose $p_n = r_n$. Let the conditions of Theorem 3.3 (i) hold but with Assumption 7 weakened to*

$$\underline{\lim}_{n\to\infty} \underline{\eta}(K_n' K_n) > 0$$

only. Also assume that

$$\frac{1}{p_n} + \frac{1}{k_n} + \frac{p_n^3 k_n^4}{n} + \frac{p_n^{\frac{3}{2}} k_n}{h_n} \to 0 \text{ as } n \to \infty \tag{3.4.17}$$

and

$$\frac{k_n}{p_n^2} \text{ bounded as } n \to \infty.$$

Then

$$\frac{n^{\frac{1}{2}}}{(p_n + k_n)^{\frac{1}{2}}} \Psi_n \left(\hat{\hat{\theta}}_{(n)} - \theta_{0(n)} \right) \xrightarrow{d} N \left(0, \lim_{n\to\infty} \frac{\sigma_0^2}{p_n + k_n} \Psi_n L_n^{-1} \Psi_n' \right),$$

where the asymptotic covariance matrix exists, and is positive definite, by Assumptions 9 and 10.

From Theorem 3.3 (ii), it is clear that the while the same distributional result is obtained as in Theorem 2.6 weaker conditions are imposed on the relative rates of h_n and $n^{\frac{1}{2}}$. For fixed p_n and k_n, the asymptotic normality result relies only on $n^{\frac{1}{2}}/h_n^3 \to 0$ as $n \to \infty$ since $3 - 5/2\gamma \to 3$ as $\gamma \to \infty$. This is a weaker requirement as compared to Lee (2002), who assumed $n^{\frac{1}{2}}/h_n \to 0$ as $n \to \infty$. The reason for this favourable outcome is the cancellation of higher order terms when using the one-step approximation. The key difference is in the rates

$$\left\| \frac{1}{n} \begin{bmatrix} B_n' \\ 0 \end{bmatrix} U_n \right\| = \mathscr{O}_p \left(\frac{p_n^{\frac{1}{2}}}{h_n} \right)$$

and

$$\|\phi_n\| = \mathscr{O}_p \left(\frac{p_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n^{\frac{1}{2}}} \right),$$

the latter being sharper since $n/h_n \to \infty$ as $n \to \infty$.

If k_n is fixed while p_n diverges, the condition (3.4.16) is guaranteed for $\gamma \geq 3/2$, since this implies $5 - 4\gamma < 0$. However, if p_n is fixed and k_n diverges then we must have $\gamma \geq 9/4$ for (3.4.16) to hold.

If h_n is bounded as $n \to \infty$, a more complicated analysis is required, because the information equality does not hold asymptotically. Denote $\mu_l = \mathbb{E}\left(u_i^l\right)$ for natural numbers l, and introduce, with $i, j = 1, \ldots, p_n$, the $p_n \times p_n$ matrix $\Omega_{\lambda\lambda,n}$ with (i,j)-th element

$$\frac{4\mu_3}{n\sigma_0^4}\sum_{r=1}^{n} c_{rr,in}b_{r,jn}X_n\beta_{0n} + \frac{\left(\mu_4 - 3\sigma_0^4\right)}{n\sigma_0^4}\sum_{r=1}^{n} c_{rr,in}c_{rr,jn} \tag{3.4.18}$$

and the $k_n \times p_n$ matrix $\Omega_{\lambda\beta,n}$ with i-th column

$$\frac{2\mu_3}{n\sigma_0^4}\sum_{r=1}^{n} c_{rr,in}x_{r,n} \tag{3.4.19}$$

where $c_{pq,in}$ is the (p,q)-th element of C_{in}, $b_{jn} = G_{jn}X_n\beta_{0(n)}$ with t-th element $b_{t,jn}$ ($j = 1,\ldots,p_n$ and $t = 1,\ldots,n$) and $x_{p,n}$ is the p-th column of X_n'. Define

$$\Omega_n = \begin{pmatrix} \Omega_{\lambda\lambda,n} & \Omega_{\lambda\beta,n}' \\ & \\ \Omega_{\lambda\beta,n} & 0 \end{pmatrix}. \tag{3.4.20}$$

Then

$$\mathbb{E}\left(\xi_n\xi_n'\right) = \frac{1}{n}\left(2\Xi_n + \Omega_n\right) \tag{3.4.21}$$

where

$$\Xi_n = \mathbb{E}\left(H_n\right) = \begin{pmatrix} \frac{2}{n}\left(P_{ji,n} + P_{j'i,n} + \frac{1}{\sigma_0^2}A_n'A_n\right) & \frac{2}{n\sigma_0^2}A_n'X_n \\ & \\ \frac{2}{n\sigma_0^2}X_n'A_n & \frac{2}{n\sigma_0^2}X_n'X_n \end{pmatrix}. \tag{3.4.22}$$

Theorem 3.4. *Consider any $s\times(p_n+k_n)$ matrix of constants Ψ_n with full row-rank. Let the conditions of Theorem 2.1 (i) hold. Suppose that h_n is bounded away from zero and that there is a real number $\delta > 0$ such that*

$$\mathbb{E}\left|u_i\right|^{4+\delta} \le C \tag{3.4.23}$$

for $i = 1,\ldots,n$. In addition, assume that

$$\varliminf_{n\to\infty} \eta\left(2\Xi_n^{-1} + \Xi_n^{-1}\Omega_n\Xi_n^{-1}\right) > 0 \text{ and } \varliminf_{n\to\infty} \eta\left(\Xi_n\right) > 0. \tag{3.4.24}$$

Suppose also that the rate conditions from Theorem 2.1 (i) are strengthened to

$$\frac{1}{p_n}+\frac{1}{r_n}+\frac{1}{k_n}+\frac{p_nk_n^2\left(p_n^3k_n^5 + p_n^3k_n^4r_n + p_n^2k_n^2r_n^2 + r_n^{\frac{3}{2}}\right)}{n}+\frac{\left(p_nk_n\right)^{\frac{8}{\delta}+2}}{n\left(p_n+k_n\right)^{\frac{4}{\delta}+1}} \to 0 \text{ as } n\to\infty. \tag{3.4.25}$$

Then

$$\frac{n^{\frac{1}{2}}}{(p_n+k_n)^{\frac{1}{2}}}\Psi_n\left(\hat{\hat{\theta}}_{(n)}-\theta_{0(n)}\right)\xrightarrow{d}N\left(0,\lim_{n\to\infty}\frac{\sigma_0^2}{p_n+k_n}\Psi_n\left(2\Xi_n^{-1}+\Xi_n^{-1}\Omega_n\Xi_n^{-1}\right)\Psi_n'\right),$$

where the asymptotic covariance matrix exists, and is positive definite, by (3.4.24).

Robinson (2010) studied the estimate defined in (3.4.7) as a particular case and derived, for $p_n = 1$, the same result as in Theorem 3.3 (*ii*). His requirement that $h_n \to \infty$ is weaker than our condition, but we do not impose symmetry of the weighting matrix nor do we assume a symmetric distribution for the errors as he did. In the setting of Section 2.4.1 the weight matrix is symmetric by construction, so his results are more incisive. He also conducted a Monte Carlo experiment in the configuration of Section 2.4.1 and indicated a substantive concern with $\tilde{\tilde{\theta}}_{(n)}$, namely that the trace terms in the score vector over-correct the bias in preliminary OLS estimate.

In view of the poor finite sample properties of the PMLE (see Section 3.3 above) we do not prove the asymptotic distribution of the PMLE, but we assume that under suitable conditions it may be shown that

$$\frac{n^{\frac{1}{2}}}{(p_n+k_n)^{\frac{1}{2}}}\Psi_n\left(\check{\theta}_{(n)}-\theta_{0(n)}\right)\xrightarrow{d}N\left(0,\lim_{n\to\infty}\frac{\sigma_0^2}{p_n+k_n}\Psi_nL_n^{-1}\Psi_n'\right)$$

or

$$\frac{n^{\frac{1}{2}}}{(p_n+k_n)^{\frac{1}{2}}}\Psi_n\left(\check{\theta}_{(n)}-\theta_{0(n)}\right)\xrightarrow{d}N\left(0,\lim_{n\to\infty}\frac{\sigma_0^2}{p_n+k_n}\Psi_n\left(2\Xi_n^{-1}+\Xi_n^{-1}\Omega_n\Xi_n^{-1}\right)\Psi_n'\right)$$

according as h_n is divergent or bounded. This conjecture is reasonable due to the definition of the asymptotic covariance matrix in the standard central limit theorem for implicitly defined estimates.

Versions of Theorems 3.3 and 3.4 for a finite-dimensional subset of parameters can also be stated as in Sections 2.3.1 and 2.3.2 but we omit these to avoid repetition.

The rate conditions can be relaxed if the G_{in} are such that

$$G_{in}G_{jn}=0 \text{ and } G_{in}'G_{jn}=0 \text{ for } i\neq j$$

as is the case when, for example, (1.2.4) and (1.2.5) are employed. This is because the only non-zero contributions in certain double-sums will now come from the diagonal terms. We illustrate the implications with k_n fixed for simplicity. In this case the rate condition (3.4.12) reduces to

$$\frac{1}{p_n}+\frac{1}{r_n}+\frac{p_n^{\frac{1}{2}}r_n^{\frac{3}{2}}}{n}+\frac{p_n}{h_n}\to 0 \text{ as } n\to\infty$$

and (3.4.13) to

$$\frac{r_n^2}{n} \text{ bounded as } n \to \infty,$$

while (3.4.17) becomes

$$\frac{1}{p_n} + \frac{p_n^2}{n} + \frac{p_n}{h_n} \to 0.$$

Similarly (3.4.14) reduces to

$$\frac{1}{p_n} + \frac{p_n^2}{n} + \frac{p_n^\gamma}{h_n} \to 0$$

and (3.4.15) to

$$\frac{n^{\frac{1}{2}}}{h_n^{3-2/\gamma}} \text{ is bounded as } n \to \infty$$

with $\gamma \in [1, \infty)$.

We can also have different h_{in} for each W_{in}, some bounded and some divergent. For those h_{in} which diverge at a sufficiently fast rate, the corresponding elements of ϕ_n are negligible and so the asymptotic covariance matrix will simplify. To illustrate, suppose that h_{1n} diverges while the remaining h_{in} are bounded and bounded away from zero. Then the $n^{\frac{1}{2}}$-normed first element of ϕ_n is negligible and we have that

$$\frac{n^{\frac{1}{2}}}{(p_n + k_n)^{\frac{1}{2}}} \Psi_n \xi_n = \frac{n^{\frac{1}{2}}}{(p_n + k_n)^{\frac{1}{2}}} \Psi_n \left(\phi_n^* - \frac{2}{\sigma_0^2} t_n \right) + o_p(1)$$

with ϕ_n^* differing from ϕ_n only in having zero as its first element. The asymptotic covariance matrix becomes

$$\lim_{n\to\infty} \frac{\sigma_0^2}{p_n + k_n} \Psi_n \left(2\Xi_n^{*-1} + \Xi_n^{*-1} \Omega_n^* \Xi_n^{*-1} \right) \Psi_n'$$

where

$$\Omega_n^* = \begin{pmatrix} \Omega_{\lambda\lambda,n}^* & \Omega_{\lambda\beta,n}^{*'} \\ \\ \Omega_{\lambda\beta,n}^* & 0 \end{pmatrix}$$

with $\Omega_{\lambda\lambda,n}^*$ and $\Omega_{\lambda\beta,n}^*$ differing from $\Omega_{\lambda\lambda,n}$ and $\Omega_{\lambda\beta,n}$ in having zeros in their first row and column and first column respectively and Ξ_n^* also being a simplified version of Ξ_n due to the fact that $tr\left(G_{1n}' G_{in}\right)$ and $tr\left(G_{in}' G_{1n}\right)$ are $\mathscr{O}\left(n/h_{1n}\right)$ for $i = 1, \ldots, p_n$.

3.5 Finite-sample performance of one-step estimates

The behaviour of the one-step estimate $\hat{\hat{\theta}}_{(n)}$ in finite samples was examined in a Monte Carlo study. The spatial weight matrices W_{in} given by (1.2.4) and (1.2.5) were employed. The number of regressors was kept fixed at $k_n = 2$ for simplicity, and we experimented with three values of r: 2, 4 and 6. For each value of r three different

$\frac{m}{r}$		25		75		300	
r		$\hat{\theta}_{(n)}$	$\hat{\hat{\theta}}_{(n)}$	$\hat{\theta}_{(n)}$	$\hat{\hat{\theta}}_{(n)}$	$\hat{\theta}_{(n)}$	$\hat{\hat{\theta}}_{(n)}$
2	λ_1	-0.0090	-0.0701	-0.0046	-0.0269	0.0015	-0.0047
	λ_2	-0.0041	-0.0427	-0.0019	-0.0168	0.0000	-0.0043
	β_1	-0.0096	0.1097	-0.0025	0.0433	-0.0013	0.0122
	β_2	-0.0033	0.1284	0.0012	0.0491	-0.0049	0.0087
4	λ_1	-0.0071	-0.0433	-0.0038	-0.0160	-0.0001	-0.0032
	λ_2	-0.0022	-0.0251	-0.0019	-0.0004	-0.0005	-0.0025
	λ_3	-0.0126	-0.0727	-0.0060	-0.0262	-0.0021	-0.0073
	λ_4	-0.0040	-0.0286	-0.0017	-0.0006	-0.0005	-0.0026
	β_1	-0.0054	0.0652	-0.0006	0.0260	0.0014	0.0080
	β_2	0.0023	0.0840	0.0057	0.0312	0.0003	0.0071
6	λ_1	-0.0035	-0.0181	-0.0009	-0.0059	-0.0004	-0.0017
	λ_2	-0.0018	-0.0114	-0.0005	-0.0039	-0.0003	-0.0012
	λ_3	-0.0082	-0.0328	-0.0016	-0.0100	-0.0007	-0.0028
	λ_4	-0.0029	-0.0126	-0.0011	-0.0045	0.0000	-0.0008
	λ_5	-0.0087	-0.0205	0.0008	-0.0030	0.0003	-0.0006
	λ_6	-0.0069	-0.0412	-0.0012	-0.0134	-0.0007	-0.0037
	β_1	-0.0032	0.0258	0.0019	0.0116	0.0000	0.0024
	β_2	0.0024	0.0289	-0.0016	0.0083	0.0002	0.0026

Table 3.5: Monte Carlo Bias of IV and Newton-step estimates $\hat{\theta}_{(n)}$ and $\hat{\hat{\theta}}_{(n)}$, $X_n \sim U(0,1)$ and $U_n \sim N(0,1)$

values of m were chosen to return three values for the ratio m/r: 25, 75 and 300. The reason behind using the same ratios as opposed to the same sample sizes was to check if finite sample properties improve comparably for all values of r with increasing sample size. The explanatory variables in X_n were generated from two distributions: a uniform distribution on $(0,1)$ and a uniform distribution on $(0,5)$. These were then kept fixed throughout to adhere to the non-stochastic aspect of Assumption 5. We experimented only with $\hat{\hat{\theta}}_{(n)}$ since $\tilde{\tilde{\theta}}_{(n)}$ has already been studied by Robinson (2010). The u_i were generated as iid draws from a standard normal ($\sigma_0^2 = 1$) distribution, and instruments were constructed as in (2.3.2) using only first-order spatial lags of the regressors. y_n was generated using (1.2.6) in each of the 1000 replications. We chose $\beta_{01} = 1$ and $\beta_{02} = 0.5$ and the following values for the spatial autoregressive parameters:

$$r = 2; \qquad \lambda_{01} = 0.7; \lambda_{02} = 0.8$$

$$r = 4; \qquad \lambda_{01} = 0.7; \lambda_{02} = 0.8; \lambda_{03} = 0.5; \lambda_{04} = 0.8$$

$$r = 6; \qquad \lambda_{01} = 0.7; \lambda_{02} = 0.8; \lambda_{03} = 0.5; \lambda_{04} = 0.8; \lambda_{05} = 0.4; \lambda_{06} = 0.3$$

$\frac{m}{r}$		25		75		300	
r		$\hat{\theta}_{(n)}$	$\hat{\hat{\theta}}_{(n)}$	$\hat{\theta}_{(n)}$	$\hat{\hat{\theta}}_{(n)}$	$\hat{\theta}_{(n)}$	$\hat{\hat{\theta}}_{(n)}$
2	λ_1	-0.0014	-0.0044	0.0002	-0.0007	0.0004	0.0002
	λ_2	-0.0001	-0.0020	0.0002	-0.0004	0.0000	0.0000
	β_1	0.0021	0.0087	-0.0003	0.0019	-0.0002	0.0002
	β_2	-0.0024	0.0035	-0.0016	0.0006	-0.0009	-0.0004
4	λ_1	0.0002	-0.0011	-0.0001	-0.0007	0.0001	0.0000
	λ_2	0.0002	-0.0007	0.0000	-0.0002	0.0000	0.0000
	λ_3	0.0002	-0.0022	0.0000	-0.0008	0.0000	-0.0002
	λ_4	0.0000	-0.0010	0.0000	-0.0002	0.0000	0.0000
	β_1	-0.0014	0.0017	0.0000	0.0011	-0.0007	-0.0004
	β_2	-0.0004	0.0028	-0.0004	0.0006	0.0004	0.0006
6	λ_1	0.0000	-0.0005	-0.0001	-0.0003	-0.0001	-0.0001
	λ_2	-0.0001	-0.0006	-0.0002	-0.0003	0.0000	0.0000
	λ_3	-0.0006	-0.0017	0.0000	-0.0002	0.0000	0.0000
	λ_4	0.0001	-0.0002	-0.0002	-0.0003	0.0000	0.0000
	λ_5	-0.0003	-0.0007	-0.0003	-0.0005	0.0000	0.0000
	λ_6	-0.0001	-0.0016	0.0000	-0.0005	-0.0001	-0.0002
	β_1	-0.0002	0.0010	0.0005	0.0008	0.0001	0.0002
	β_2	-0.0002	0.0010	-0.0003	0.0000	0.0000	0.0000

Table 3.6: Monte Carlo Bias of IV and Newton-step estimates $\hat{\theta}_{(n)}$ and $\hat{\hat{\theta}}_{(n)}$, $X_n \sim U(0,5)$ and $U_n \sim N(0,1)$

We report Monte Carlo bias, relative mean squared error (MSE) and relative variance for the estimates $\hat{\hat{\theta}}_{(n)}$ and $\hat{\theta}_{(n)}$. Tables 3.5 and 3.6 tabulate the biases for each element of $\hat{\theta}_{(n)}$ and $\hat{\hat{\theta}}_{(n)}$ for three possible values of the ratio m/r and the three choices of r. In Table 3.5, the bias of $\hat{\hat{\theta}}_{(n)}$ is clearly greater (in absolute value) than that of $\hat{\theta}_{(n)}$ in each of the cases, reflecting the presence of the additional bias term that was observed in (3.4.8). While this bias declines with increasing n so does the bias in $\hat{\theta}_{(n)}$ and the former dominates. In Table 3.6, the extra variation in the regressors implies that the additional bias observed in (3.4.8) declines faster with n, as the bias terms are functions of $(X_n' X_n)^{-1}$. As a result, many of the biases of $\hat{\hat{\theta}}_{(n)}$ in the last column of Table 3.6 are less than or equal (to four decimal places) to the biases of $\hat{\theta}_{(n)}$ reported in the second from last column.

Table 3.7 reports relative MSE when the regressors are generated from $U(0,1)$, a distribution with variance equal to 1/12. We compute, for all combinations of m/r and r, the element-wise ratio $MSE\left(\hat{\theta}_{(n)}\right)/MSE\left(\hat{\hat{\theta}}_{(n)}\right)$. $\hat{\hat{\theta}}_{(n)}$ beats $\hat{\theta}_{(n)}$ in just 3 out of 54 places. On the other hand, Table 3.8 reports relative MSE when the regressors are generated from $U(0,5)$. This distribution has variance equal to 25/12. The table indicates that $\hat{\hat{\theta}}_{(n)}$ beats $\hat{\theta}_{(n)}$ in 15 out of 54 places, including an efficiency improvement for all four parameters in the $r = 2$, $m/r = 300$ case.

Tables 3.9 and 3.10 report relative variance analogously, computing the element-wise ratio $Var\left(\hat{\theta}_{(n)}\right)/Var\left(\hat{\hat{\theta}}_{(n)}\right)$. In Table 3.9 $\hat{\hat{\theta}}_{(n)}$ beats $\hat{\theta}_{(n)}$ in 28 out of the 30 places for $r = 2$, 4, but only for the two regression coefficients for $r = 6$. However the ratios are much closer to 1 than in Table 3.7 even when $\hat{\hat{\theta}}_{(n)}$ does not beat $\hat{\theta}_{(n)}$. These results are as expected due to to the greater bias of $\hat{\hat{\theta}}_{(n)}$. The improvement in relative variance is also not monotone in m/r (equivalently the sample size n), with Table 3.9 indicating higher ratios for smaller sample sizes for several parameter estimates for $r = 2$, 4. In Table 3.10 we observe that $\hat{\hat{\theta}}_{(n)}$ beats $\hat{\theta}_{(n)}$ in all 30 out of the 30 places for $r = 2$, 4, but again only for the two regression coefficients for $r = 6$. However, the ratios are extremely close to unity for $r = 6$ and $m/r = 75, 300$. Although the variances of the two estimators seem to be approaching each other, Table 3.6 indicates that the one-step estimator ultimately outperforms the IV estimator as far as bias is concerned. This explains the improvement in relative MSE for the one-step estimator that was reported in Table 3.8 and discussed above.

Convergence of the iterations was typically fast. The results displayed correspond to a single iteration but further iteration (up to six were carried out) did not lead to any serious change in the results. Single iteration convergence was almost exact for the larger sample sizes.

Tables 3.11 and 3.12 compare bias and MSE for the one-step estimate $\hat{\hat{\theta}}_{(n)}$ with the ML estimate discussed in Section 3.3, for $r = 2$. The conclusions are similar to those in that section. The MLE outperforms the one-step estimate for the regression coefficients but is much worse for the autoregression parameters. The one-step estimate improves faster with increasing m/r.

For practitioners, this chapter prescribes that closed-form estimates be used wherever possible. If IV estimation is used, then a one-step approximation to the PMLE will tend to be more efficient for smaller values of r and larger values of m, while efficiency gains for small samples will be greater if the regressors have high variability.

$\frac{m}{r}$		25	75	300
r		$\frac{MSE(\hat{\theta}_{(n)})}{MSE(\hat{\hat{\theta}}_{(n)})}$	$\frac{MSE(\hat{\theta}_{(n)})}{MSE(\hat{\hat{\theta}}_{(n)})}$	$\frac{MSE(\hat{\theta}_{(n)})}{MSE(\hat{\hat{\theta}}_{(n)})}$
2	λ_1	0.7649	0.8993	1.0014
	λ_2	0.7914	0.9041	0.9818
	β_1	0.9674	1.0030	1.0027
	β_2	0.9505	0.9900	1.0124
4	λ_1	0.7302	0.8579	0.9744
	λ_2	0.7687	0.8702	0.9666
	λ_3	0.7176	0.8559	0.9540
	λ_4	0.7282	0.8764	0.9609
	β_1	0.9333	0.9648	0.9837
	β_2	0.8859	0.9363	0.9891
6	λ_1	0.8536	0.9466	0.9828
	λ_2	0.8569	0.9500	0.9812
	λ_3	0.8442	0.9462	0.9823
	λ_4	0.8427	0.9358	0.9886
	λ_5	0.8683	0.9658	0.9920
	λ_6	0.8605	0.9489	0.9844
	β_1	0.9669	0.9754	0.9965
	β_2	0.9509	0.9922	0.9950

Table 3.7: Monte Carlo Relative MSE of IV and one-step estimates, $MSE\left(\hat{\theta}_{(n)}\right)/MSE\left(\hat{\hat{\theta}}_{(n)}\right)$, $X_n \sim U(0,1)$ and $U_n \sim N(0,1)$

$\frac{m}{r}$		25	75	300
r		$\frac{MSE(\hat{\theta}_{(n)})}{MSE(\hat{\hat{\theta}}_{(n)})}$	$\frac{MSE(\hat{\theta}_{(n)})}{MSE(\hat{\hat{\theta}}_{(n)})}$	$\frac{MSE(\hat{\theta}_{(n)})}{MSE(\hat{\hat{\theta}}_{(n)})}$
2	λ_1	0.9769	0.9998	1.0046
	λ_2	0.9903	1.0019	1.0005
	β_1	0.9953	1.0002	1.0005
	β_2	1.0059	1.0041	1.0027
4	λ_1	0.9921	0.9922	1.0011
	λ_2	0.9940	0.9977	1.0004
	λ_3	0.9910	0.9960	0.9988
	λ_4	0.9863	0.9983	0.9995
	β_1	1.0035	0.9980	1.0036
	β_2	0.9973	1.0008	0.9975
6	λ_1	0.9947	0.9964	0.9981
	λ_2	0.9904	0.9927	1.0005
	λ_3	0.9888	0.9987	0.9994
	λ_4	0.9981	0.9930	0.9997
	λ_5	0.9942	0.9976	0.9999
	λ_6	0.9932	0.9978	0.9988
	β_1	0.9986	0.9972	0.9991
	β_2	0.9988	1.0007	0.9998

Table 3.8: Monte Carlo Relative MSE of IV and one-step estimates, $MSE\left(\hat{\theta}_{(n)}\right)/MSE\left(\hat{\hat{\theta}}_{(n)}\right)$, $X_n \sim U(0,5)$ and $U_n \sim N(0,1)$

$\frac{m}{r}$		25	75	300
r		$\frac{Var\left(\hat{\theta}_{(n)}\right)}{Var\left(\hat{\hat{\theta}}_{(n)}\right)}$	$\frac{Var\left(\hat{\theta}_{(n)}\right)}{Var\left(\hat{\hat{\theta}}_{(n)}\right)}$	$\frac{Var\left(\hat{\theta}_{(n)}\right)}{Var\left(\hat{\hat{\theta}}_{(n)}\right)}$
2	λ_1	0.9830	1.0535	1.0199
	λ_2	1.0159	1.0491	1.0212
	β_1	1.0674	1.0497	1.0162
	β_2	1.0692	1.0496	1.0177
4	λ_1	0.9979	1.0108	1.0038
	λ_2	1.0292	1.0091	1.0035
	λ_3	1.0136	1.0099	1.0041
	λ_4	1.0071	1.0050	1.0051
	β_1	1.0813	1.0357	1.0089
	β_2	1.0972	1.0342	1.0090
6	λ_1	0.9700	0.9938	0.9980
	λ_2	0.9676	0.9914	0.9985
	λ_3	0.9704	0.9927	0.9982
	λ_4	0.9718	0.9925	0.9982
	λ_5	0.8979	0.9684	0.9924
	λ_6	0.9680	0.9926	0.9983
	β_1	1.0136	1.0046	1.0015
	β_2	1.0103	1.0065	1.0012

Table 3.9: Monte Carlo Relative Variance of IV and one-step estimates, $Var\left(\hat{\theta}_{(n)}\right)/Var\left(\hat{\hat{\theta}}_{(n)}\right)$, $X_n \sim U(0,1)$ and $U_n \sim N(0,1)$

$\frac{m}{r}$		25	75	300
r		$\frac{Var(\hat{\theta}_{(n)})}{Var(\hat{\hat{\theta}}_{(n)})}$	$\frac{Var(\hat{\theta}_{(n)})}{Var(\hat{\hat{\theta}}_{(n)})}$	$\frac{Var(\hat{\theta}_{(n)})}{Var(\hat{\hat{\theta}}_{(n)})}$
2	λ_1	1.0080	1.0031	1.0007
	λ_2	1.0091	1.0037	1.0007
	β_1	1.0084	1.0025	1.0005
	β_2	1.0073	1.0028	1.0007
4	λ_1	1.0011	1.0004	1.0001
	λ_2	1.0024	1.0007	1.0002
	λ_3	1.0019	1.0006	1.0001
	λ_4	1.0011	1.0008	1.0001
	β_1	1.0043	1.0014	1.0003
	β_2	1.0039	1.0014	1.0003
6	λ_1	0.9987	0.9997	0.9999
	λ_2	0.9990	0.9997	0.9999
	λ_3	0.9985	0.9997	0.9999
	λ_4	0.9994	0.9997	0.9999
	λ_5	0.9957	0.9988	0.9997
	λ_6	0.9992	0.9997	0.9999
	β_1	1.0006	1.0003	1.0000
	β_2	1.0006	1.0001	1.0000

Table 3.10: Monte Carlo Relative Variance of IV and one-step estimates, $Var\left(\hat{\theta}_{(n)}\right)/Var\left(\hat{\hat{\theta}}_{(n)}\right)$, $X_n \sim U(0,5)$ and $U_n \sim N(0,1)$

$\frac{m}{r}$		25		75		150	
r		$\hat{\hat{\theta}}_{(n)}$	$\check{\theta}_{(n)}$	$\hat{\hat{\theta}}_{(n)}$	$\check{\theta}_{(n)}$	$\hat{\hat{\theta}}_{(n)}$	$\check{\theta}_{(n)}$
2	λ_1	-0.0701	-0.3774	-0.0269	-0.3818	-0.0113	-0.2270
	λ_2	-0.0426	-0.3292	-0.0168	-0.3214	-0.0072	-0.3435
	β_1	0.1097	-0.0076	0.0432	0.0004	0.0210	0.0024
	β_2	0.1284	-0.0018	0.0490	0.0002	0.0216	-0.0019

Table 3.11: Monte Carlo Bias of one-step and ML estimates, $\hat{\hat{\theta}}_{(n)}$ and $\check{\theta}_{(n)}$, $X_n \sim U(0,1)$ and $U_n \sim N(0,1)$

$\frac{m}{r}$		25	75	150
r		$\frac{MSE\left(\hat{\hat{\theta}}_{(n)}\right)}{MSE\left(\check{\theta}_{(n)}\right)}$	$\frac{MSE\left(\hat{\hat{\theta}}_{(n)}\right)}{MSE\left(\check{\theta}_{(n)}\right)}$	$\frac{MSE\left(\hat{\hat{\theta}}_{(n)}\right)}{MSE\left(\check{\theta}_{(n)}\right)}$
2	λ_1	0.0471	0.0099	0.0090
	λ_2	0.0205	0.0051	0.0022
	β_1	14.6594	12.5855	10.7357
	β_2	16.1920	15.1626	11.3499

Table 3.12: Monte Carlo Relative MSE of one-step and ML estimates, $MSE\left(\hat{\hat{\theta}}_{(n)}\right)/MSE\left(\check{\theta}_{(n)}\right)$, $X_n \sim U(0,1)$ and $U_n \sim N(0,1)$

3.A Proofs of theorems

Proof of Theorem 3.1. We can write

$$\frac{1}{n}\log|\Theta_n| = \frac{1}{n}\log\left|H_n\left(\lambda_{(n)}\right)\right| + \frac{1}{n}\log\left|\Theta_n\left(\lambda_{(n)};\sigma_0^2\right)\right|$$

Then

$$\begin{aligned}
\mathcal{Q}_n^c\left(\lambda_{(n)}\right) - \mathcal{Q}_n^c &= \log\check{\sigma}_{(n)}^2\left(\lambda_{(n)}\right) + \frac{1}{n}\log\left|T_n\left(\lambda_{(n)}\right)T_n'\left(\lambda_{(n)}\right)\right| - \frac{1}{n}\log|\Theta_n| + o_p^{\lambda}(1) \\
&= \log\check{\sigma}_{(n)}^2\left(\lambda_{(n)}\right) + \frac{1}{n}\log\left|T_n\left(\lambda_{(n)}\right)T_n'\left(\lambda_{(n)}\right)\right| - \frac{1}{n}\log\left|H_n\left(\lambda_{(n)}\right)\right| \\
&+ \frac{1}{n}\mathrm{tr}H_n\left(\lambda_{(n)}\right) - 1 - \frac{1}{n}\log\left|\Theta_n\left(\lambda_{(n)};\sigma_0^2\right)\right| + o_p^{\lambda}(1). \qquad (3.A.1)
\end{aligned}$$

Now

$$\begin{aligned}
&\frac{1}{n}\log\left|T_n\left(\lambda_{(n)}\right)T_n'\left(\lambda_{(n)}\right)\right| - \frac{1}{n}\log\left|\Theta_n\left(\lambda_{(n)};\sigma_0^2\right)\right| \\
= \; &\frac{1}{n}\log\left|T_n\left(\lambda_{(n)}\right)T_n'\left(\lambda_{(n)}\right)\Theta_n^{-1}\left(\lambda_{(n)};\sigma_0^2\right)\right| \\
= \; &-\log\sigma_n^2\left(\lambda_{(n)}\right),
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{tr}H_n\left(\lambda_{(n)}\right) &= \mathrm{tr}\left(\Theta_n\Theta_n^{-1}\left(\lambda_{(n)};\sigma_0^2\right)\right) \\
&= \frac{\sigma_0^2}{\sigma_n^2\left(\lambda_{(n)}\right)}\mathrm{tr}\left(T_nT_n'S_n'\left(\lambda_{(n)}\right)S_n\left(\lambda_{(n)}\right)\right) \\
&= n, \qquad (3.A.2)
\end{aligned}$$

so that (3.A.1) becomes

$$\mathcal{Q}_n^c\left(\lambda_{(n)}\right) - \mathcal{Q}_n^c = \log\check{\sigma}_{(n)}^2\left(\lambda_{(n)}\right) - \log\sigma_n^2\left(\lambda_{(n)}\right) + r\left(\lambda_{(n)}\right) + o_p^{\lambda}(1). \qquad (3.A.3)$$

Using the approximation $\log a - \log b \approx (a-b)/b$, we can replace

$$\log\check{\sigma}_{(n)}^2\left(\lambda_{(n)}\right) - \log\sigma_n^2\left(\lambda_{(n)}\right)$$

by

$$\left(\check{\sigma}_{(n)}^2\left(\lambda_{(n)}\right) - \sigma_n^2\left(\lambda_{(n)}\right)\right)/\sigma_n^2\left(\lambda_{(n)}\right).$$

As a result,

$$\mathcal{Q}_n^c\left(\lambda_{(n)}\right) - \mathcal{Q}_n^c = \frac{c_n\left(\lambda_{(n)}\right)}{\sigma_n^2\left(\lambda_{(n)}\right)} + \frac{d_n\left(\lambda_{(n)}\right)}{\sigma_n^2\left(\lambda_{(n)}\right)} \qquad (3.A.4)$$

where

$$c_n\left(\lambda_{(n)}\right)=\frac{1}{n}\sum_{i=1}^{p_n}\sum_{j=1}^{p_n}\left(\lambda_{0i}-\lambda_i\right)\left(\lambda_{0j}-\lambda_j\right)b_{in}'M_nb_{jn}+r\left(\lambda_{(n)}\right) \tag{3.A.5}$$

and $d_n\left(\lambda_{(n)}\right)=\sum_{i=1}^{8}d_{in}\left(\lambda_{(n)}\right)+o_p^{\lambda}(1)$ with

$$\begin{aligned}
d_{1n}\left(\lambda_{(n)}\right) &= \frac{2}{n}\sum_{i=1}^{p_n}\left(\lambda_{0i}-\lambda_i\right)b_{in}'M_nU_n && (3.A.6)\\
d_{2n}\left(\lambda_{(n)}\right) &= \frac{2}{n}\sum_{i=1}^{p_n}\sum_{j=1}^{p_n}\left(\lambda_{0i}-\lambda_i\right)\left(\lambda_{0j}-\lambda_j\right)b_{in}'M_nG_{jn}U_n && (3.A.7)\\
d_{3n}\left(\lambda_{(n)}\right) &= \frac{U_n'U_n}{n}-\sigma_0^2 && (3.A.8)\\
d_{4n}\left(\lambda_{(n)}\right) &= \frac{1}{n}U_n'\left(\sum_{i=1}^{p_n}\left(\lambda_{0i}-\lambda_i\right)C_{in}\right)U_n\\
&- \frac{\sigma_0^2}{n}\mathrm{tr}\left(\sum_{i=1}^{p_n}\left(\lambda_{0i}-\lambda_i\right)C_{in}\right) && (3.A.9)\\
d_{5n}\left(\lambda_{(n)}\right) &= \frac{1}{n}U_n'\left(\sum_{i=1}^{p_n}\sum_{j=1}^{p_n}\left(\lambda_{0i}-\lambda_i\right)\left(\lambda_{0j}-\lambda_j\right)G_{in}'G_{jn}\right)U_n\\
&- \frac{\sigma_0^2}{n}\mathrm{tr}\left(\sum_{i=1}^{p_n}\sum_{j=1}^{p_n}\left(\lambda_{0i}-\lambda_i\right)\left(\lambda_{0j}-\lambda_j\right)G_{in}'G_{jn}\right) && (3.A.10)\\
d_{6n}\left(\lambda_{(n)}\right) &= -\frac{U_n'X_n}{n}\left(\frac{X_n'X_n}{n}\right)^{-1}\frac{X_n'U_n}{n} && (3.A.11)\\
d_{7n}\left(\lambda_{(n)}\right) &= \frac{2}{n^2}U_n'\left(\sum_{i=1}^{p_n}\left(\lambda_i-\lambda_{0i}\right)G_{in}'X_n\left(\frac{X_n'X_n}{n}\right)^{-1}X_n'\right)U_n && (3.A.12)\\
d_{8n}\left(\lambda_{(n)}\right) &= \frac{1}{n^2}U_n'\left(\sum_{i=1}^{p_n}\left(\lambda_i-\lambda_{0i}\right)G_{in}'\right)\left(X_n\left(\frac{X_n'X_n}{n}\right)^{-1}X_n'\right)\\
&\times \left(\sum_{i=1}^{p_n}\left(\lambda_i-\lambda_{0i}\right)G_{in}\right)U_n. && (3.A.13)
\end{aligned}$$

By Lemma 3.1 and a standard kind of argument for proving the consistency (in norm) of implicitly defined estimates, it suffices to show that

$$\sup_{\lambda_{(n)}\in\Lambda_n}\left|d_{in}\left(\lambda_{(n)}\right)\right|=o_p(1),\ i=1,\ldots,8, \tag{3.A.14}$$

and, for all $\delta_{\lambda_{(n)}} > 0$,

$$\lim_{n\to\infty} \inf_{\left\{\left\|\lambda_{(n)}-\lambda_{0(n)}\right\|>\delta_{\lambda_{(n)}}\right\}\cap\Lambda_n} c_n\left(\lambda_{(n)}\right) > 0. \tag{3.A.15}$$

To prove (3.A.14), first consider $d_{1n}\left(\lambda_{(n)}\right)$. We first establish pointwise convergence to 0, for any $\lambda_{(n)} \in \Lambda_n$. $d_{1n}\left(\lambda_{(n)}\right)$ has mean zero and variance

$$\begin{aligned} & \frac{4\sigma_0^2}{n^2}\sum_{i=1}^{p_n}\sum_{j=1}^{p_n}\left(\lambda_{0i}-\lambda_i\right)\left(\lambda_{0j}-\lambda_j\right)b_{in}'M_nb_{jn} \\ \leq\ & \frac{C}{n^2}\sum_{i=1}^{p_n}\sum_{j=1}^{p_n}\|b_{in}\|\,\|b_{jn}\|\,\|M_n\| \\ \leq\ & C\frac{p_n^2k_n^3}{n}, \end{aligned} \tag{3.A.16}$$

because

$$\|b_{in}\| \leq \|G_{in}\|\left\|X_n\beta_{0(n)}\right\| \leq Cn^{\frac{1}{2}}k_n$$

by Lemma 2.C2 and Assumptions 5 and 12 and also because by Assumptions 5 and 9 we have

$$\|M_n\| \leq \|I_n\| + \frac{1}{n}\|X_n\|^2\left\|\left(\frac{X_n'X_n}{n}\right)^{-1}\right\| = \mathscr{O}(k_n).$$

As a result,

$$d_{1n}\left(\lambda_{(n)}\right) = \mathscr{O}_p\left(\frac{p_nk_n^{\frac{3}{2}}}{n^{\frac{1}{2}}}\right) \tag{3.A.17}$$

which is negligible by (3.2.13). Uniform convergence follows from an equicontinuity argument. Consider a neighbourhood $\mathscr{N}$ of any $\lambda_{(n)}^*$, such that $\mathscr{N} \subset \Lambda_n$. Then

$$\begin{aligned} \sup_{\lambda_{(n)}\in\mathscr{N}}\left|\frac{d_{1n}\left(\lambda_{(n)}\right)-d_{1n}\left(\lambda_{(n)}^*\right)}{2}\right| &= \sup_{\lambda_{(n)}\in\mathscr{N}}\left|\frac{\sum_{i=1}^{p_n}\left(\lambda_i^*-\lambda_i\right)b_{in}'M_nU_n}{n}\right| \\ \leq \left(\frac{U_n'U_n}{n}\right)^{\frac{1}{2}} & \sup_{\lambda_{(n)}\in\mathscr{N}}\left\{\frac{\sum_{i,j=1}^{p_n}\left(\lambda_i^*-\lambda_i\right)\left(\lambda_j^*-\lambda_j\right)b_{in}'M_nb_{jn}}{n}\right\}^{\frac{1}{2}}. \end{aligned} \tag{3.A.18}$$

Now $\mathbb{E}U_n'U_n/n = \sigma_0^2$, whereas the expression in braces is bounded by a constant times

$$\sum_{i=1}^{p_n}\left(\lambda_i^*-\lambda_i\right)^2$$

by the Cauchy-Schwarz and Hölder inequalities and Assumption 16. This can be made arbitrarily small uniformly on $\mathcal{N}$ by choosing $\mathcal{N}$ small enough. By compactness of Λ_n, any open over has a finite subcover and the proof that $d_{1n}\left(\lambda_{(n)}\right) = o_p(1)$ uniformly in $\lambda_{(n)}$ is completed. Similarly it may be shown that

$$
\begin{aligned}
d_{2n}\left(\lambda_{(n)}\right) &= \mathcal{O}_p\left(\frac{p_n^2 k_n^2}{n^{\frac{1}{2}}}\right) && (3.A.19)\\
d_{3n}\left(\lambda_{(n)}\right) &= o_p^{\lambda}(1) && (3.A.20)\\
d_{4n}\left(\lambda_{(n)}\right) &= \mathcal{O}_p\left(\frac{p_n}{n^{\frac{1}{2}} h_n^{\frac{1}{2}}}\right) && (3.A.21)\\
d_{5n}\left(\lambda_{(n)}\right) &= \mathcal{O}_p\left(\frac{p_n^2}{n^{\frac{1}{2}} h_n^{\frac{1}{2}}}\right) && (3.A.22)\\
d_{6n}\left(\lambda_{(n)}\right) &= \mathcal{O}_p\left(\frac{k_n}{n}\right) && (3.A.23)\\
d_{7n}\left(\lambda_{(n)}\right) &= \mathcal{O}_p\left(\frac{p_n k_n}{n}\right) && (3.A.24)\\
d_{8n}\left(\lambda_{(n)}\right) &= \mathcal{O}_p\left(\frac{p_n^2 k_n}{n}\right), && (3.A.25)\\
& && (3.A.26)
\end{aligned}
$$

which are all negligible by (3.2.13). Uniform equicontinuity arguments will follow as for (3.A.18). The proof of (3.A.15) follows from Assumptions 10 and 17. Indeed, the former, using the partitioned matrix inversion formula, implies that

$$\lim_{n\to\infty} \frac{1}{n} x_n' A_n' M_n A_n x_n > 0, \text{ for } x_n \neq 0,$$

so that choosing $x_n = \lambda_{(n)} - \lambda_{0(n)}$ implies

$$\lim_{n\to\infty} \inf_{\left\{\left\|\lambda_{(n)}-\lambda_{0(n)}\right\|>\delta_{\lambda_{(n)}}\right\}\cap\Lambda_n} \frac{1}{n}\sum_{i=1}^{p_n}\sum_{j=1}^{p_n}\left(\lambda_{0i}-\lambda_i\right)\left(\lambda_{0j}-\lambda_j\right) b_{in}' M_n b_{jn} > 0,$$

for any $\delta_{\lambda_{(n)}} > 0$ since $A_n = (b_{1n}, \ldots, b_{p_n n})$. The consistency of $\check{\lambda}_{(n)}$ is then established. The conclusion that

$$\left\|\check{\beta}_{(n)}\left(\lambda_{(n)}\right) - \beta_{0(n)}\right\| \xrightarrow{p} 0$$

follows from the (3.2.1). □

Proof of Theorem 3.3. (i) For any $s\times 1$ vector α, we can use (3.4.8) to write

$$
\begin{aligned}
\tau_n \alpha' \Psi_n\left(\hat{\hat{\theta}}_{(n)} - \theta_{0(n)}\right) &= \tau_n \alpha' \Psi_n \hat{H}_n^{-1}\left(\hat{H}_n - \overline{H}_n\right)\left(\hat{\theta}_{(n)} - \theta_{0(n)}\right)\\
&- \tau_n \alpha' \Psi_n \hat{H}_n^{-1} \xi_n, && (3.A.27)
\end{aligned}
$$

recalling that $\tau_n = n^{\frac{1}{2}}/a_n^{\frac{1}{2}}$. The first term on RHS above has modulus bounded by

$$\tau_n \|\alpha\| \|\Psi_n\| \left\| \hat{H}_n^{-1} \right\| \left\| \hat{H}_n - \overline{H}_n \right\| \left\| \hat{\theta}_{(n)} - \theta_{0(n)} \right\|,$$

where the second factor in norms is $\mathscr{O}\left(a_n^{\frac{1}{2}}\right)$, the third is bounded for sufficiently large n by Lemma 3.B11, by Lemma 3.B9 the fourth is $\mathscr{O}_p\left(\max\left\{\frac{p_n k_n^2 b_n^{\frac{1}{2}} c_n^{\frac{1}{2}}}{n}, \frac{p_n^{\frac{3}{2}} k_n^2 b_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n}, \frac{p_n k_n^2 b_n}{n}\right\}\right)$ and the fifth is $\mathscr{O}_p\left(\frac{b_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}\right)$ by (2.B.7). We conclude that the first term on the RHS of (3.A.27) is

$$\mathscr{O}_p\left(\max\left\{\frac{p_n k_n^2 b_n c_n^{\frac{1}{2}}}{n}, \frac{p_n^{\frac{3}{2}} k_n^2 b_n}{n^{\frac{1}{2}} h_n}, \frac{p_n k_n^2 b_n^{\frac{3}{2}}}{n}\right\}\right),$$

which is negligible by (3.4.12) and (3.4.13) because

$$\begin{aligned}
\frac{p_n^2 k_n^4 b_n^2 c_n}{n^2} &\leq C\left(\frac{p_n^3 r_n^2 k_n^6 + p_n^3 k_n^8}{n^2}\right), \\
\frac{p_n^3 k_n^4 b_n^2}{n h_n^2} &\leq C\left(\frac{p_n^3 r_n^2 k_n^4 + p_n^3 k_n^6}{n h_n^2}\right), \\
\frac{p_n^2 k_n^4 b_n^3}{n^2} &\leq C\left(\frac{p_n^2 r_n^3 k_n^4 + p_n^2 k_n^7}{n^2}\right)
\end{aligned}$$

where

$$\frac{p_n^3 r_n^2 k_n^6}{n^2} = \frac{p_n^3 k_n^4}{n}\frac{r_n^2 k_n^2}{n}, \quad \frac{p_n^3 r_n^2 k_n^4}{n h_n^2} = \frac{p_n^3 k_n^2}{h_n^2}\frac{r_n^2 k_n^2}{n}, \quad \frac{p_n^2 r_n^3 k_n^4}{n^2} = \frac{p_n^3 k_n^4}{n}\frac{r_n^3}{n p_n}.$$

So we only need to find the asymptotic distribution of $-\tau_n \alpha' \Psi_n \hat{H}_n^{-1} \xi_n$. We can write

$$-\tau_n \alpha' \Psi_n \hat{H}_n^{-1} \xi_n = \frac{2}{\sigma_0^2} \tau_n \alpha' \Psi_n \hat{H}_n^{-1} t_n - \tau_n \alpha' \Psi_n \hat{H}_n^{-1} \phi_n. \tag{3.A.28}$$

Then

$$\begin{aligned}
\mathbb{E}\|\phi_n\|^2 &\leq \sum_{i=1}^{p_n} \mathbb{E}\left(\frac{1}{n} tr C_{in} - \frac{1}{n\sigma_0^2} U_n' C_{in} U_n\right)^2 \\
&= \sum_{i=1}^{p_n} \operatorname{var}\left(\frac{1}{n} U_n' C_{in} U_n\right) = \mathscr{O}\left(\frac{p_n}{n h_n}\right),
\end{aligned}$$

by Lemmas 2.C4 and 2.C2 so that

$$\|\phi_n\| = \mathscr{O}_p\left(\frac{p_n^{\frac{1}{2}}}{n^{\frac{1}{2}}h_n^{\frac{1}{2}}}\right). \tag{3.A.29}$$

Therefore the second term on the right of (3.A.28) has modulus bounded by τ_n times

$$\|\alpha\| \|\Psi_n\| \left\|\hat{H}_n^{-1}\right\| \|\phi_n\|, \tag{3.A.30}$$

where the second factor is $\mathscr{O}\left(a_n^{\frac{1}{2}}\right)$, the third is bounded for sufficiently large n by Lemma 3.B11 and the last is $\mathscr{O}_p\left(\frac{p_n^{\frac{1}{2}}}{n^{\frac{1}{2}}h_n^{\frac{1}{2}}}\right)$. Thus (3.A.30) is $\mathscr{O}_p\left(\frac{p_n^{\frac{1}{2}}a_n^{\frac{1}{2}}}{n^{\frac{1}{2}}h_n^{\frac{1}{2}}}\right)$ and the second term on the right of (3.A.28) is $\mathscr{O}_p\left(\frac{p_n^{\frac{1}{2}}}{h_n^{\frac{1}{2}}}\right)$ which is negligible by (3.4.12). Then the asymptotic distribution required is that of

$$\frac{2}{\sigma_0^2}\tau_n\alpha'\Psi_n\hat{H}_n^{-1}t_n = \sum_{i=1}^{3}\Upsilon_{in} + \tau_n\alpha'\Psi_nL_n^{-1}t_n \tag{3.A.31}$$

where

$$\begin{aligned}
\Upsilon_{1n} &= \frac{2}{\sigma_0^2}\tau_n\alpha'\Psi_n\hat{H}_n^{-1}\left(\hat{H}_n - H_n\right)H_n^{-1}t_n,\\
\Upsilon_{2n} &= \frac{2}{\sigma_0^2}\tau_n\alpha'\Psi_n\Xi_n^{-1}\left(H_n - \Xi_n\right)H_n^{-1}t_n,\\
\Upsilon_{3n} &= \tau_n\alpha'\Psi_nL_n^{-1}\left[\frac{\sigma_0^2}{2}\Xi_n - L_n\right]\left(\frac{\sigma_0^2}{2}\Xi_n\right)^{-1}t_n.
\end{aligned}$$

We will demonstrate that $|\Upsilon_{in}| = o_p(1)$, $i = 1, 2, 3$. First we observe that

$$|\Upsilon_{1n}| \le \frac{2}{\sigma_0^2}\tau_n\|\alpha\|\|\Psi_n\|\left\|\hat{H}_n^{-1}\right\|\left\|\hat{H}_n - H_n\right\|\left\|H_n^{-1}\right\|\|t_n\|,$$

where the second factor in norms is $\mathscr{O}\left(a_n^{\frac{1}{2}}\right)$, the third and fifth are bounded for sufficiently large n by Lemma 3.B11, the fourth is $\mathscr{O}_p\left(\max\left\{\frac{p_nk_n^2b_n^{\frac{1}{2}}c_n^{\frac{1}{2}}}{n}, \frac{p_n^{\frac{3}{2}}k_n^2b_n^{\frac{1}{2}}}{n^{\frac{1}{2}}h_n}, \frac{p_nk_n^2b_n}{n}\right\}\right)$ from the proof of Lemma 3.B9 and the last is $\mathscr{O}_p\left(\frac{c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}\right)$ with the last bound having been derived in (2.B.13). Then

$$|\Upsilon_{1n}| = \mathscr{O}_p\left(\max\left\{\frac{p_nk_n^2b_n^{\frac{1}{2}}c_n}{n}, \frac{p_n^{\frac{3}{2}}k_n^2b_n^{\frac{1}{2}}c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}h_n}, \frac{p_nk_n^2b_nc_n^{\frac{1}{2}}}{n}\right\}\right),$$

which is negligible by (3.4.12) and (3.4.13) because

$$\begin{aligned}
\frac{p_n^2 k_n^4 b_n c_n^2}{n^2} &\leq C\left(\frac{p_n^4 r_n k_n^8 + p_n^4 k_n^9}{n^2}\right), \\
\frac{p_n^3 k_n^4 b_n c_n}{n h_n^2} &\leq C\left(\frac{p_n^4 r_n k_n^6 + p_n^4 k_n^7}{n h_n^2}\right),
\end{aligned}$$

and

$$\begin{aligned}
\frac{p_n^4 r_n k_n^8}{n^2} &= \frac{p_n^3 k_n^4}{n}\frac{p_n^{\frac{3}{2}} k_n^2}{n^{\frac{1}{2}}}\frac{r_n k_n^2}{n^{\frac{1}{2}} p_n^{\frac{1}{2}}}, \quad \frac{p_n^4 k_n^9}{n^2} = \frac{p_n^3 k_n^4}{n}\frac{p_n^{\frac{3}{2}} k_n^2}{n^{\frac{1}{2}}}\frac{k_n^3}{n^{\frac{1}{2}} p_n^{\frac{1}{2}}}, \\
\frac{p_n^4 r_n k_n^6}{n h_n^2} &= \frac{p_n^3 k_n^2}{h_n^2}\frac{p_n^{\frac{3}{2}} k_n^2}{n^{\frac{1}{2}}}\frac{r_n k_n^2}{n^{\frac{1}{2}} p_n^{\frac{1}{2}}}, \quad \frac{p_n^4 k_n^7}{n h_n^2} = \frac{p_n^3 k_n^2}{h_n^2}\frac{p_n^{\frac{3}{2}} k_n^2}{n^{\frac{1}{2}}}\frac{k_n^3}{n^{\frac{1}{2}} p_n^{\frac{1}{2}}},
\end{aligned}$$

while $\frac{p_n^2 k_n^4 b_n^2 c_n}{n^2}$ has been dealt with earlier. Next

$$|\Upsilon_{2n}| \leq \frac{2}{\sigma_0^2}\tau_n \|\alpha\| \|\Psi_n\| \left\|H_n^{-1}\right\| \|H_n - \Xi_n\| \left\|\Xi_n^{-1}\right\| \|t_n\|,$$

where the second factor in norms is $\mathscr{O}\left(a_n^{\frac{1}{2}}\right)$, the third and fifth are bounded for sufficiently large n by Lemma 3.B11, the fourth is $\mathscr{O}_p\left(\frac{p_n k_n}{n^{\frac{1}{2}}}\right)$ by Lemma 3.B10 and the last is $\mathscr{O}_p\left(\frac{c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}\right)$ as above. Then

$$|\Upsilon_{2n}| = \mathscr{O}_p\left(\frac{p_n k_n c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}\right)$$

which is negligible by (3.4.12) because

$$\frac{p_n^2 k_n^2 c_n}{n} \leq C\frac{p_n^3 k_n^4}{n}.$$

Similarly $|\Upsilon_{3n}| = \mathscr{O}_p\left(\frac{p_n c_n^{\frac{1}{2}}}{h_n}\right)$ by Lemma 3.B10, which is negligible by (3.4.12) because

$$\frac{p_n^2 c_n}{h_n^2} \leq C\frac{p_n^3 k_n^2}{h_n^2}.$$

Then we only need to find the asymptotic distribution of the last term term in (3.A.31), but this is precisely what we derived in the proof of Theorem 2.6. Replicating those leads to the theorem.

(ii) In view of Lemmas 3.B10, 3.B12 and 3.B13, the theorem is proved exactly like Theorem 3.3 (i), except for different orders of magnitudes of various expressions. In this case two of the orders will be different from the analogous ones considered in the the proof of Theorem 3.3 (i). Indeed, the analogue of the bound for the first term in (3.A.27) is

$$\begin{aligned}
&\mathscr{O}_p\left(n^{\frac{1}{2}}\max\left\{\frac{p_n^{\frac{3}{2}}k_n^2c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}h_n},\frac{p_n^2k_n^2}{h_n^2},\frac{p_n^{\frac{7}{4}}c_n^{\frac{1}{4}}}{n^{\frac{1}{4}}h_n^{\frac{3}{2}}},\frac{p_nk_n^2c_n}{n}\right\}\max\left\{\frac{c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}},\frac{p_n^{\frac{1}{2}}}{h_n}\right\}\right)\\
= \; &\mathscr{O}_p\left(\max\left\{\pi_{1n},\pi_{2n},\pi_{3n},\pi_{4n},\pi_{5n},\pi_{6n}\right\}\right),
\end{aligned}$$

where

$$\begin{aligned}
\pi_{1n} &= \frac{p_n^{\frac{3}{2}}k_n^2c_n}{n^{\frac{1}{2}}h_n},\ \pi_{2n} = \frac{p_n^2k_n^2c_n^{\frac{1}{2}}}{h_n^2},\ \pi_{3n} = \frac{p_n^{\frac{7}{4}}c_n^{\frac{3}{4}}}{n^{\frac{1}{4}}h_n^{\frac{3}{2}}},\\
\pi_{4n} &= \frac{p_nk_n^2c_n^{\frac{3}{2}}}{n},\ \pi_{5n} = \frac{n^{\frac{1}{2}}p_n^{\frac{5}{2}}k_n^2}{h_n^3},\ \pi_{6n} = \frac{n^{\frac{1}{4}}p_n^{\frac{9}{4}}c_n^{\frac{1}{4}}}{h_n^{\frac{5}{2}}}.
\end{aligned}$$

Now

$$\pi_{1n}^2 = \frac{p_n^3k_n^4c_n^2}{nh_n^2} \leq C\frac{p_n^5k_n^8}{nh_n^2}$$

which is negligible under (3.4.14) and (3.4.16) as we may write

$$\frac{p_n^5k_n^8}{nh_n^2} = \frac{p_n^{\frac{5}{2}}k_n^5}{n}\left(\frac{p_n^{\gamma}k_n^{2\gamma/3}}{h_n}\right)^2 p_n^{5/2-2\gamma}k_n^{3-4\gamma/3}$$

where $5/2-2\gamma<0$ since $\gamma\geq 3/2$ and $\frac{7}{2}-\frac{4\gamma}{3}\leq 0$ if $\gamma\geq\frac{9}{4}$. If the latter condition does not hold then we need to employ the extra condition (3.4.16). Secondly

$$\begin{aligned}
\pi_{2n}^2 &\leq C\frac{p_n^5k_n^6}{h_n^4}\\
&= C\left\{\frac{p_n^{\gamma}k_n^{2\gamma/3}}{h_n}\right\}^4 p_n^{5-4\gamma}k_n^{6-8\gamma/3}
\end{aligned}$$

which is negligible under (3.4.14) and (3.4.16) since $6-8\gamma/3\leq 0$ if $\gamma\geq 9/4$ and $5-4\gamma<0$ always, while if $\gamma\geq 9/4$ then (3.4.16) delivers convergence to zero. Third, we have

$$\pi_{3n}^4 \leq C\frac{p_n^{10}k_n^6}{nh_n^6} = C\frac{p_n^3k_n^4}{n}\frac{p_n^7k_n^6}{h_n^6}$$

which is negligible by (3.4.14). Fourth

$$\pi_{4n}^2 \leq C \frac{p_n^5 k_n^{10}}{n^2}$$

which is negligible by (3.4.14). Fifth,

$$\pi_{5n}^2 \leq C \left\{ \frac{n^{\frac{1}{2}} p_n^{\frac{5}{2}} k_n^2}{h_n^3} \right\}^2 .$$

This is negligible under (3.4.14) and (3.4.15) since

$$\frac{n^{\frac{1}{2}} p_n^{\frac{5}{2}} k_n^2}{h_n^3} = \left(\frac{p_n^{\gamma} k_n^{\frac{2\gamma}{3}}}{h_n} \right)^{\frac{5}{2\gamma}} \frac{k_n^{\frac{1}{3}} n^{\frac{1}{2}}}{h_n^{3-5/2\gamma}}.$$

Finally

$$\pi_{6n}^4 \leq C \left\{ \frac{n^{\frac{1}{2}} p_n^5 k_n}{h_n^5} \right\}^2 ,$$

which is negligible under (3.4.14) and (3.4.15) since

$$\frac{n^{\frac{1}{2}} p_n^5 k_n^2}{h_n^5} = \left(\frac{p_n^{\gamma} k_n^{\frac{2\gamma}{3}}}{h_n} \right)^{\frac{5}{\gamma}} \frac{n^{\frac{1}{2}}}{k_n^{\frac{7}{3}} h_n^{5-5/\gamma}} = \left(\frac{p_n^{\gamma} k_n^{\frac{2\gamma}{3}}}{h_n} \right)^{\frac{5}{\gamma}} \frac{k_n^{\frac{1}{3}} n^{\frac{1}{2}}}{h_n^{3-5/2\gamma}} \frac{1}{k_n^{\frac{8}{3}} h_n^{2-5/2\gamma}}.$$

where $2 - 5/2\gamma \geq 0$ as $\gamma \geq 3/2$. The analogue of the bound for Υ_{1n} is

$$\begin{aligned} & \mathscr{O}\left(n^{\frac{1}{2}}\right) \mathscr{O}_p \left(\max \left\{ \frac{p_n^{\frac{3}{2}} k_n^2 c_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n}, \frac{p_n^2 k_n^2}{h_n^2}, \frac{p_n^{\frac{7}{4}} c_n^{\frac{1}{4}}}{n^{\frac{1}{4}} h_n^{\frac{3}{2}}}, \frac{p_n k_n^2 c_n}{n} \right\} \right) \mathscr{O}_p \left(\frac{c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}} \right) \\ = \; & \mathscr{O}_p \left(\max \left\{ \pi_{1n}, \pi_{2n}, \pi_{3n}, \pi_{4n} \right\} \right), \end{aligned}$$

which was shown to be negligible under the assumed conditions. All other bounds remain unchanged and will be also be negligible under under (3.4.14), (3.4.15) and (3.4.16) as in the proof of Theorem 3.3 (i).

□

Proof of Theorem 3.4. Proceeding as in the proof of Theorem 3.3 (i), we can write

$$\begin{aligned} \tau_n \alpha' \Psi_n \left(\hat{\hat{\theta}}_{(n)} - \theta_{0(n)} \right) &= \tau_n \alpha' \Psi_n \hat{H}_n^{-1} \left(\hat{H}_n - \overline{H}_n \right) \left(\hat{\theta}_{(n)} - \theta_{0(n)} \right) \\ &- \tau_n \alpha' \Psi_n \left(\hat{H}_n^{-1} - \Xi_n^{-1} \right) \xi_n - \tau_n \alpha' \Psi_n \Xi_n^{-1} \xi_n. \quad (3.A.32) \end{aligned}$$

As in the proof of Theorem 3.3 (i), the first term on the RHS above is negligible by

(3.4.25). Lemma 3.B11 (for bounded h_n) indicates that the second term on the RHS of (3.A.32) is bounded in modulus by a constant times

$$\tau_n \left\|\Psi_n\right\| \left(\left\|t_n\right\| + \left\|\phi_n\right\|\right) \left(\left\|\hat{H}_n - H_n\right\| + \left\|H_n - \Xi_n\right\|\right) \tag{3.A.33}$$

which is

$$\mathscr{O}_p\left(n^{\frac{1}{2}} \max\left\{\frac{c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}, \frac{p_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n^{\frac{1}{2}}}\right\} \max\left\{\frac{p_n k_n^2 b_n^{\frac{1}{2}} c_n^{\frac{1}{2}}}{n}, \frac{p_n^{\frac{3}{2}} k_n^2 b_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n}, \frac{p_n k_n^2 b_n}{n}\right\}\right),$$

by (2.B.13), (3.A.29) and Lemmas 3.B9 and 3.B10 (i). This is negligible by (3.4.25). Thus we need to establish the asymptotic distribution of

$$-\tau_n \Psi_n \Xi_n^{-1} \xi_n \tag{3.A.34}$$

which has zero mean and variance

$$\Psi_n \left(2\Xi_n^{-1} + \Xi_n^{-1} \Omega_n \Xi_n^{-1}\right) \Psi_n'.$$

Hence we consider the asymptotic normality of

$$\frac{-n^{\frac{1}{2}} \alpha' \Psi_n \Xi_n^{-1} \xi_n}{\left\{a_n \alpha' \Psi_n \left(2\Xi_n^{-1} + \Xi_n^{-1} \Omega_n \Xi_n^{-1}\right) \Psi_n' \alpha\right\}^{\frac{1}{2}}}, \tag{3.A.35}$$

where α is any $s \times 1$ vector of constants. It is convenient to write

$$\varsigma_n = \left\{a_n \alpha' \Psi_n \left(2\Xi_n^{-1} + \Xi_n^{-1} \Omega_n \Xi_n^{-1}\right) \Psi_n' \alpha\right\}^{\frac{1}{2}}$$

for the denominator of (3.A.35). Then

$$\varsigma_n \geq a_n^{\frac{1}{2}} \left\|\Psi_n' \alpha\right\| \left\{\underline{\eta}\left(2\Xi_n^{-1} + \Xi_n^{-1} \Omega_n \Xi_n^{-1}\right)\right\}^{\frac{1}{2}} \geq c\, a_n^{\frac{1}{2}} \left\|\Psi_n' \alpha\right\| \tag{3.A.36}$$

by (3.4.24). The numerator of (3.A.35) can be written as

$$-\frac{2}{\sigma_0^2 n^{\frac{1}{2}}} m_n' U_n - \frac{1}{\sigma_0^2 n^{\frac{1}{2}}} U_n' D_n U_n + \frac{1}{n^{\frac{1}{2}}} tr D_n \tag{3.A.37}$$

where

$$\begin{aligned} D_n &= \sum_{j=1}^{p_n} \left(\alpha' \Psi_n \zeta_n^j\right) C_{jn}, \\ m_n &= \sum_{j=1}^{p_n} \left(\alpha' \Psi_n \zeta_n^j\right) G_{jn} X_n \beta_{0(n)} + \sum_{j=p_n+1}^{a_n} \left(\alpha' \Psi_n \zeta_n^j\right) \chi_{(j-p_n),n}, \end{aligned}$$

with ζ_n^j and $\chi_{j,n}$ denoting the j-th columns of Ξ_n^{-1} and X_n respectively. We also denote by $d_{ij,n}$ and $m_{i,n}$ the (i,j)-th and i-th elements of D_n and m_n respectively.

Using (3.A.37), we can write (3.A.35) as $-\sum_{i=1}^n z_{in}$, with

$$z_{in} = \frac{1}{\sigma_0^2 n^{\frac{1}{2}} \varsigma_n} \left(u_i^2 - \sigma_0^2\right) d_{ii,n} + \frac{2}{\sigma_0^2 n^{\frac{1}{2}} \varsigma_n} u_i \sum_{j<i} u_j d_{ij,n} + \frac{2}{\sigma_0^2 n^{\frac{1}{2}} \varsigma_n} m_{i,n} u_i \tag{3.A.38}$$

so that $\{z_{in} : 1 \leq i \leq n, n = 1, 2, \ldots\}$ forms a triangular array of martingale differences with respect to the filtration formed by the σ-field generated by $\{u_j; j < i\}$. Theorem 2 of Scott (1973) is applicable if

$$\sum_{i=1}^n \mathbb{E}\left\{z_{in}^2 1\left(z_{in} \geq \epsilon\right)\right\} \to 0, \ \forall \epsilon > 0 \tag{3.A.39}$$

$$\sum_{i=1}^n \mathbb{E}\left(z_{in}^2 \mid u_j, j < i\right) \xrightarrow{p} 1. \tag{3.A.40}$$

To show (3.A.39) we can check the sufficient Lyapunov condition

$$\sum_{i=1}^n \mathbb{E} \left|z_{in}\right|^{2+\frac{\delta}{2}} \to 0. \tag{3.A.41}$$

The c_r inequality, (3.4.23), (3.A.36) and Markov's inequality indicate that (3.A.41) holds if, as $n \to \infty$, $\mathbb{E}\left(\sum_{i=1}^n \mathbb{E}\left|z_{in}\right|^{2+\frac{\delta}{2}}\right) \to 0$. The latter is bounded by a constant times

$$\frac{\sum_{i=1}^n \left|d_{ii,n}\right|^{2+\frac{\delta}{2}}}{n^{1+\frac{\delta}{4}} a_n^{1+\frac{\delta}{4}} \left\|\Psi_n' \alpha\right\|^{2+\frac{\delta}{2}}} + \frac{\sum_{i=1}^n \mathbb{E}\left|\sum_{j<i} u_j d_{ij,n}\right|^{2+\frac{\delta}{2}}}{n^{1+\frac{\delta}{4}} a_n^{1+\frac{\delta}{4}} \left\|\Psi_n' \alpha\right\|^{2+\frac{\delta}{2}}} + \frac{\sum_{i=1}^n \left|m_{i,n}\right|^{2+\frac{\delta}{2}}}{n^{1+\frac{\delta}{4}} a_n^{1+\frac{\delta}{4}} \left\|\Psi_n' \alpha\right\|^{2+\frac{\delta}{2}}}. \tag{3.A.42}$$

The first term in (3.A.42) is bounded by

$$\frac{\max_i \left|d_{ii,n}\right|^{2+\frac{\delta}{2}}}{n^{\frac{\delta}{4}} a_n^{1+\frac{\delta}{4}} \left\|\Psi_n' \alpha\right\|^{2+\frac{\delta}{2}}}, \tag{3.A.43}$$

while the third term is bounded by

$$\frac{\max_i |m_{i,n}|^{2+\frac{\delta}{2}}}{n^{\frac{\delta}{4}} a_n^{1+\frac{\delta}{4}} \left\|\Psi_n' \alpha\right\|^{2+\frac{\delta}{2}}}. \tag{3.A.44}$$

By the Burkholder, von Bahr/Esseen and elementary ℓ_p-norm inequalities, the second term in (3.A.42) is bounded by a constant times

$$\frac{\max_i \left|\sum_{j<i} d_{ij,n}^2\right|^{1+\frac{\delta}{4}}}{n^{\frac{\delta}{4}} a_n^{1+\frac{\delta}{4}} \left\|\Psi_n' \alpha\right\|^{2+\frac{\delta}{2}}}. \tag{3.A.45}$$

Now, recalling that $e_{i,n}$ is the n-dimensional vector with unity in the i-th position and zeros elsewhere, we can write

$$\begin{aligned}
& \sum_{j=1}^{n} d_{ij,n}^2 = e_{i,n}' D_n^2 e_{i,n} \\
\leq \quad & \left\|D_n\right\|^2 \\
= \quad & \left\|\sum_{j=1}^{p_n} \left(\alpha' \Psi_n \zeta_n^j\right) C_{jn}\right\|^2 \\
\leq \quad & C p_n^2 \left(\max_j \left\|C_{jn}\right\|\right)^2 \left(\max_j \left\|\zeta_n^j\right\|\right)^2 \left\|\Psi_n' \alpha\right\|^2 \\
\leq \quad & C \left\|\Xi_n^{-1}\right\|^2 p_n^2 \left\|\Psi_n' \alpha\right\|^2 \\
= \quad & C \frac{p_n^2 \left\|\Psi_n' \alpha\right\|^2}{\left\{\underline{\eta}\left(\Xi_n\right)\right\}^2} \\
\leq \quad & C p_n^2 \left\|\Psi_n' \alpha\right\|^2,
\end{aligned} \tag{3.A.46}$$

using Lemma 2.C6 and (3.4.24). Also, we can use (3.A.46) to bound

$$|d_{ii,n}| \leq \left(\sum_{j=1}^{n} d_{ij,n}^2\right)^{\frac{1}{2}} \leq C p_n \left\|\Psi_n' \alpha\right\|. \tag{3.A.47}$$

We also note that, for each $i = 1, \ldots, n$,

$$\left(\sum_{j=1}^{n} d_{ij,n}^2\right)^{\frac{1}{2}} \leq \sum_{j=1}^{n} |d_{ij,n}| \leq \left\|D_n\right\|_R \leq C p_n \left\|\Psi_n' \alpha\right\|, \tag{3.A.48}$$

by Lemma 2.C1. Now (3.A.47) and (3.A.46) imply that (3.A.43) and (3.A.45) are both

$\mathscr{O}\left(\frac{p_n^{2+\frac{\delta}{2}}}{n^{\frac{\delta}{4}}a_n^{1+\frac{\delta}{4}}}\right)$. This is negligible by (3.4.25).

Next, writing $b_{i,jn}$ and $x_{ij,n}$ for the i-th elements of $G_{jn}X_n\beta_{0(n)}$ and $\chi_{j,n}$ respectively, we have

$$|m_{i,n}| \leq \sum_{j=1}^{p_n} \left|\alpha'\Psi_n\zeta_n^j\right| |b_{i,jn}| + \sum_{j=p_n+1}^{a_n} \left|\alpha'\Psi_n\zeta_n^j\right| |x_{ij,n}| \leq Ck_n\left(p_n+1\right)\left\|\Psi_n'\alpha\right\|, \quad (3.A.49)$$

using Assumptions 5, (3.4.24) and Lemma 2.C5. Then (3.A.44) is $\mathscr{O}\left(\frac{p_n^{2+\frac{\delta}{2}}k_n^{2+\frac{\delta}{2}}}{n^{\frac{\delta}{4}}a_n^{1+\frac{\delta}{4}}}\right)$, which is negligible by (3.4.25). Hence (3.A.41) is proved.

We now show (3.A.40). First note that we can write

$$\sum_{i=1}^{n}\mathbb{E}\left(z_{in}^2 \mid u_j, j<i\right) - 1 = 4\left(f_{1n}+f_{2n}+f_{3n}\right) \quad (3.A.50)$$

with

$$f_{1n} = \frac{1}{\sigma_0^2 n\varsigma_n^2}\sum_i \sum_{\substack{j,k<i \\ j\neq k}} d_{ij,n}d_{ik,n}u_ju_k, \quad (3.A.51)$$

$$f_{2n} = \frac{1}{\sigma_0^2 n\varsigma_n^2}\sum_i\sum_{j<i} d_{ij,n}^2\left(u_j^2-\sigma_0^2\right), \quad (3.A.52)$$

$$f_{3n} = \frac{1}{\sigma_0^4 n\varsigma_n^2}\sum_i\left(\sigma_0^2 m_{i,n}+\mu_3 d_{ii,n}\right)\sum_{j<i} d_{ij,n}u_j. \quad (3.A.53)$$

f_{1n} has zero mean and variance bounded by $1/n^2\varsigma_n^4$ times

$$\begin{aligned}
& C\sum_{\substack{h,i,j,k \\ j,k<i,h}} |d_{ij,n}d_{ik,n}d_{hj,n}d_{hk,n}| \\
\leq\ & C\sum_{h,i,j,k}|d_{ij,n}d_{ik,n}|\left(d_{hj,n}^2+d_{hk,n}^2\right) \\
\leq\ & C\left(\max_i\sum_k|d_{ik,n}|\right)\left(\max_j\sum_i|d_{ij,n}|\right)\sum_{i,j}d_{ij,n}^2 \\
=\ & C\left\|D_n\right\|_R^2\left\|D_n\right\|_F^2 \\
\leq\ & C\left\|\Psi_n'\alpha\right\|^4 np_n^4, \qquad (3.A.54)
\end{aligned}$$

by (3.A.46) and (3.A.48). (3.A.36) and (3.A.54), together with Markov's inequality, imply that $f_{1n} = \mathscr{O}_p\left(\frac{p_n^2}{n^{\frac{1}{2}}a_n}\right)$, which is negligible by (3.4.25).

Next, f_{2n} has zero mean and variance bounded by $1/n^2\varsigma_n^4$ times

$$\begin{aligned} C\sum_{i,h}\sum_{j<i,h} d_{ij,n}^2 d_{hj,n}^2 &\leq C\sum_{i,h,j} d_{ij,n}^2 d_{hj,n}^2 \\ &= C\sum_{i,j} d_{ij,n}^2 \sum_h d_{hj,n}^2 \\ &\leq C\left(\max_j \sum_h d_{hj,n}^2\right)\|D_n\|_F^2 \\ &\leq C\left\|\Psi_n'\alpha\right\|^4 np_n^4, \end{aligned} \tag{3.A.55}$$

by (3.A.46). (3.A.36) and (3.A.55), together with Markov's inequality, imply that $f_{2n} = \mathscr{O}_p\left(\frac{p_n^2}{n^{\frac{1}{2}}a_n}\right)$ which is negligible by (3.4.25).

Finally f_{3n} has zero mean and variance bounded by $1/n^2\varsigma_n^4$ times

$$\begin{aligned} C\sum_i \left(\sigma_0^2 m_{i,n} + \mu_3 d_{ii,n}\right)^2 \sum_{j<i} d_{ij,n}^2 &\leq C\left(\max_i m_{i,n}^2 + \max_i d_{ii,n}^2\right)\|D_n\|_F^2 \\ &\leq C\left(\max_i m_{i,n}^2 + \max_i \sum_j d_{ij,n}^2\right)\|D_n\|_F^2 \\ &\leq C\left\|\Psi_n'\alpha\right\|^4\left(k_n^2+1\right)np_n^4, \end{aligned} \tag{3.A.56}$$

by (3.A.46) and (3.A.49). (3.A.36) and (3.A.56), together with Markov's inequality, imply that $f_{3n} = \mathscr{O}_p\left(\frac{p_n^2 k_n}{n^{\frac{1}{2}}a_n}\right)$, which is negligible by (3.4.25). □

3.B Proofs of lemmas

Proof of Lemma 3.1. We can write

$$\sigma_n^2\left(\lambda_{(n)}\right) = \frac{\sigma_0^2}{n}\left\{n - 2\left(\lambda_{(n)} - \lambda_{0(n)}\right)' f_n + \left(\lambda_{(n)} - \lambda_{0(n)}\right)' P_{j'i,n}\left(\lambda_{(n)} - \lambda_{0(n)}\right)\right\}. \tag{3.B.1}$$

The minimizers $\mathring{\lambda}_{(n)} - \lambda_{0(n)}$ of (3.B.1) satisfy the first-order condition

$$f_n = P_{j'i,n}\left(\mathring{\lambda}_{(n)} - \lambda_{0(n)}\right),$$

implying that the minimized value of $\sigma_n^2\left(\lambda_{(n)}\right)$ is

$$\sigma_0^2\left(1 - \frac{f_n' P_{j'i,n}^{-1} f_n}{n}\right),$$

which is bounded away from 0 uniformly in n and $\lambda_{(n)}$ by Assumption 14. To show $\sigma_n^2\left(\lambda_{(n)}\right)$ is bounded above uniformly in n and $\lambda_{(n)}$, note that (3.B.1) is bounded above by

$$\sigma_0^2 + C\left(\frac{\sum_{i=1}^{p_n} \mathrm{tr} G_{in}}{n} + \frac{\sum_{i=1}^{p_n}\sum_{j=1}^{p_n} \mathrm{tr}\left(G_{in}' G_{jn}\right)}{n}\right)$$

by Assumption 12. The last displayed expression is uniformly bounded by Assumption 15. □

Lemma 3.B1. *Let Assumptions 3 and 4 hold. Then* $\left\|S_n^{-1}\left(\lambda_{(n)}\right)\right\|_R$ *and* $\left\|S_n^{-1}\left(\lambda_{(n)}\right)\right\|_C$ *are uniformly bounded in a closed neighbourhood of* $\lambda_{0(n)}$.

Proof.

$$\begin{aligned} S_n^{-1}\left(\lambda_{(n)}\right) &= \left(I_n - \sum_{i=1}^{p_n} \lambda_i W_{in}\right)^{-1} = \left(S_n - \sum_{i=1}^{p_n} (\lambda_i - \lambda_{0i}) W_{in}\right)^{-1} \\ &= S_n^{-1}\left(I_n - \sum_{i=1}^{p_n} (\lambda_i - \lambda_{0i}) G_{in}\right)^{-1}. \end{aligned}$$

To admit a Neumann series expansion for the last displayed expression we need $\left\|\sum_{i=1}^{p_n} (\lambda_i - \lambda_{0i}) G_{in}\right\|_R < 1$.
We have

$$\begin{aligned} \left\|\sum_{i=1}^{p_n} (\lambda_i - \lambda_{0i}) G_{in}\right\|_R &\le \sum_{i=1}^{p_n} \left|\lambda_i - \lambda_{0i}\right| \left\|G_{in}\right\|_R \\ &\le \max_{i=1,\ldots,p_n} \left\|G_{in}\right\|_R \sum_{i=1}^{p_n} \left|\lambda_i - \lambda_{0i}\right| \\ &\le C \sum_{i=1}^{p_n} \left|\lambda_i - \lambda_{0i}\right|, \end{aligned}$$

where the last displayed inequality above is obtained through Lemma 2.C1. Let k_1 be a positive real number such that $k_1 < \frac{1}{C}$ and define the set

$$B\left(\lambda_{0(n)}\right) = \left\{\lambda_{(n)} \in \mathbb{R}^{p_n} : \sum_{i=1}^{p_n} \left| \lambda_i - \lambda_{0i}\right| < k_1\right\}.$$

Such a choice is possible due to denseness of the parameter space.
Then $\left\|\sum_{i=1}^{p_n} (\lambda_i - \lambda_{0i}) G_{in}\right\|_R < 1 \;\; \forall\; \lambda_{(n)} \in B\left(\lambda_{0(n)}\right)$. So the series expansion is valid

and for $\lambda_{(n)} \in B\left(\lambda_{0(n)}\right)$ we have

$$\left(I_n - \sum_{i=1}^{p_n}(\lambda_i - \lambda_{0i})G_{in}\right)^{-1} = \sum_{k=0}^{\infty}\left\{\sum_{i=1}^{p_n}(\lambda_i - \lambda_{0i})G_{in}\right\}^k .$$

We use the triangle inequality and the submultiplicative property of the matrix norm $\|\cdot\|_R$ to bound $\left\|\left(I_n - \sum_{i=1}^{p_n}(\lambda_i - \lambda_{0i})G_{in}\right)^{-1}\right\|_R$ by

$$\begin{aligned}
& \sum_{k=0}^{\infty}\left\|\sum_{i=1}^{p_n}(\lambda_i - \lambda_{0i})G_{in}\right\|_R^k \\
\leq\ & \sum_{k=0}^{\infty}\left\{\max_{i=1,\ldots,p_n}\|G_{in}\|_R\right\}^k\left\{\sum_{i=1}^{p_n}|\lambda_i - \lambda_{0i}|\right\}^k \\
\leq\ & \sum_{k=0}^{\infty}C^k\left\{\sum_{i=1}^{p_n}|\lambda_i - \lambda_{0i}|\right\}^k \\
=\ & \sum_{k=0}^{\infty}\left\{C\sum_{i=1}^{p_n}|\lambda_i - \lambda_{0i}|\right\}^k \\
=\ & \frac{1}{1 - C\sum_{i=1}^{p_n}|\lambda_i - \lambda_{0i}|},
\end{aligned}$$

where the last sum is valid as the summands are less than 1 in absolute value by construction of $B\left(\lambda_{0(n)}\right)$. Finally, using the above and Assumption 4, we have

$$\begin{aligned}
\left\| S_n^{-1}\left(\lambda_{(n)}\right)\right\|_R &\leq \left\| S_n^{-1}\right\|_R \left\|\left(I_n - \sum_{i=1}^{p_n}(\lambda_i - \lambda_{0i})G_{in}\right)^{-1}\right\|_R \\
&\leq \frac{C}{1 - C_1\sum_{i=1}^{p_n}|\lambda_i - \lambda_{0i}|} \leq C,
\end{aligned}$$

for any $\lambda_{(n)} \in B\left(\lambda_{0(n)}\right)$, with the last bound following since the denominator is bounded away from zero uniformly in n by choice of $B\left(\lambda_{0(n)}\right)$, whence the result follows if we take a closed subset of $B\left(\lambda_{0(n)}\right)$, denoted $B^c\left(\lambda_{0(n)}\right)$. The claim for column sums follows similarly. □

Corollary 3.B2. *Under the conditions of Lemma 3.B1, we have*

1. *For each $i = 1, \ldots, p_n$, $\left\|G_{in}(\lambda_{(n)})\right\|_R$ and $\left\|G_{in}(\lambda_{(n)})\right\|_C$ are uniformly bounded in $B^c\left(\lambda_{0(n)}\right)$.*

2. For each $i = 1, \ldots, p_n$, the elements of $G_{in}(\lambda_{(n)})$ are uniformly $\mathscr{O}\left(\frac{1}{h_n}\right)$ in $B^c\left(\lambda_{0(n)}\right)$ if also Assumption 2 holds.

Proof. 1. Follows by Lemma 3.B1 together with Assumption 4.
2. Follows by Lemma 3.B1 together with Assumption 2 in exactly the same way as we proved Lemma 2.C2 □

Lemma 3.B3. *Under the conditions of Corollary 3.B2 (2), we have $tr\left(G_{in}(\lambda_{(n)})G_{jn}(\lambda_{(n)})G_{kn}(\lambda_{(n)})\right) = \mathscr{O}\left(\frac{n}{h_n}\right)$ $\forall\ \lambda_{(n)} \in B^c\left(\lambda_{0(n)}\right)$ and for any $i, j, k = 1, \ldots, p_n$.*

Proof. Consider $\lambda_{(n)} \in B^c\left(\lambda_{0(n)}\right)$.
A typical (l, m)-th element of $G_{in}(\lambda_{(n)})G_{jn}(\lambda_{(n)})G_{kn}(\lambda_{(n)})$ is $g'_{l,in}G_{jn}(\lambda_{(n)})G_{kn}(\lambda_{(n)})e_{m,n}$ which is bounded in absolute value by

$$\left\|g'_{l,in}\right\|_C \left\|G_{jn}(\lambda_{(n)})\right\|_C \left\|G_{kn}(\lambda_{(n)})\right\|_C \left\|e_{m,n}\right\|_C .$$

This is uniformly $\mathscr{O}\left(1/h_n\right)$ since the elements of $G_{in}(\lambda_{(n)})$ have that uniform order, and $G_{jn}(\lambda_{(n)})$, $G_{kn}(\lambda_{(n)})$ are uniformly bounded in column sums (Corollary 3.B2). The result now follows by the definition of trace. □

Lemma 3.B4. *Suppose Assumptions 3-5 hold. Then $\|A'_nA_n\| = \mathscr{O}\left(np_nk_n^2\right)$.*

Proof. A'_nA_n has (i, j)-th element $\left(G_{in}X_n\beta_0\right)'\left(G_{jn}X_n\beta_0\right)$.
Then by Cauchy-Schwarz inequality and Lemma 2.C5

$$\begin{aligned}
\left\|A'_nA_n\right\|^2 &\leq \sum_{i=1}^{p_n}\sum_{j=1}^{p_n}\left|\left(G_{in}X_n\beta_0\right)'\left(G_{jn}X_n\beta_0\right)\right|^2 \\
&\leq \sum_{i=1}^{p_n}\sum_{j=1}^{p_n}\left\|G_{in}X_n\beta_0\right\|^2\left\|G_{jn}X_n\beta_0\right\|^2 = \mathscr{O}\left(n^2p_n^2k_n^4\right).
\end{aligned}$$

□

Lemma 3.B5. *Suppose Assumptions 1-5 hold. Then $\|B'_nA_n\| = \|A'_nB_n\| = \mathscr{O}_p\left(n^{\frac{1}{2}}p_nk_n\right)$.*

Proof. B'_nA_n has (i, j)-th element $\left(G_{in}U_n\right)'\left(G_{jn}X_n\beta_0\right)$. Then

$$\begin{aligned}
\mathbb{E}\left\|B'_nA_n\right\|^2 &\leq \sum_{i=1}^{p_n}\sum_{j=1}^{p_n}\mathbb{E}\left|\left(G_{in}U_n\right)'\left(G_{jn}X_n\beta_0\right)\right|^2 \\
&= \sum_{i=1}^{p_n}\sum_{j=1}^{p_n}\mathbb{E}\left\{\left(G_{jn}X_n\beta_0\right)'G_{in}U_nU'_nG'_{in}\left(G_{jn}X_n\beta_0\right)\right\} \\
&\leq \sigma_0^2\sum_{i=1}^{p_n}\left\|G_{in}\right\|^2\sum_{j=1}^{p_n}\left\|G_{jn}X_n\beta_0\right\|^2 \leq Cnp_n^2k_n^2,
\end{aligned}$$

using elementary spectral norm inequalities, Assumption 1 and Lemmas 2.C1 and 2.C5. Using the Markov inequality and noting that the spectral norm is invariant under matrix transposition, the lemma is proved. □

Lemma 3.B6. *Suppose that Assumptions 1-4 hold. Then* $\|B_n' B_n\| = \mathcal{O}_p\left(\frac{np_n}{h_n}\right)$.

Proof. Using elementary spectral norm inequalities, Assumption 1 and Lemma 2.C1, we have

$$\begin{aligned}\mathbb{E}\left\|B_n' B_n\right\| &\leq \mathbb{E}\left\|I_n B_n\right\|^2 \leq \sum_{i=1}^{n}\sum_{j=1}^{p_n}\mathbb{E}\left(e_{in}' G_{jn} U_n\right)^2 \\ &\leq \sigma_0^2 \sum_{j=1}^{p_n} tr\left(G_{jn} G_{jn}'\right) \leq C\frac{np_n}{h_n}\end{aligned}$$

as calculated while bounding the first term on RHS of 2.B.19, whence the lemma follows from Markov's inequality. Note once again that with finite fourth moments this bound will not improve. □

Lemma 3.B7. *Suppose that Assumptions 3-5 hold. Then* $\|X_n' A_n\| = \|A_n' X_n\| = \mathcal{O}\left(np_n^{\frac{1}{2}} k_n^{\frac{3}{2}}\right)$.

Proof. $X_n' A_n$ has (i,j)-th element $x_{i,n}' G_{jn} X_n \beta_0$, where $x_{i,n}'$ is the i-th row of X_n'. Then $\|X_n' A_n\|^2$ is bounded by

$$\sum_{i=1}^{k_n}\sum_{j=1}^{p_n}\left|x_{i,n}' G_{jn} X_n \beta_0\right|^2 \leq \sum_{i=1}^{k_n}\|x_{i,n}\|^2 \sum_{j=1}^{p_n}\|G_{jn} X_n \beta_0\|^2 = \mathcal{O}\left(n^2 p_n k_n^3\right),$$

using Cauchy-Schwarz inequality, Assumption 5 and Lemma 2.C5. The lemma is proved noting that the spectral norm is invariant under matrix transposition. □

Lemma 3.B8. *Suppose Assumptions 1-5 hold. Then* $\|X_n' B_n\| = \|B_n' X_n\| = \mathcal{O}_p\left(n^{\frac{1}{2}} p_n^{\frac{1}{2}} k_n^{\frac{1}{2}}\right)$.

Proof. $X_n' B_n$ has (i,j)-th element $x_{i,n}' G_{jn} U_n$. Then

$$\begin{aligned}\mathbb{E}\left\|X_n' B_n\right\|^2 \leq \sum_{i=1}^{k_n}\sum_{j=1}^{p_n}\mathbb{E}\left|x_{i,n}' G_{jn} U_n\right|^2 &= \sum_{i=1}^{k_n}\sum_{j=1}^{p_n}\mathbb{E}\left\{x_{i,n}' G_{jn} U_n U_n' G_{jn}' x_{i,n}\right\} \\ &\leq \sigma_0^2 \sum_{i=1}^{k_n}\|x_{i,n}\|^2 \sum_{j=1}^{p_n}\|G_{jn}\|^2 \leq Cnp_n k_n\end{aligned}$$

using elementary spectral norm inequalities, Assumptions 1, 5 and Lemma 2.C1. The lemma is proved noting that the spectral norm is invariant under matrix transposition. □

Lemma 3.B9. *Suppose Assumptions 1-5 hold together with (2.3.7). Then*

$$\left\| \hat{H}_n - \overline{H}_n \right\| = \mathscr{O}_p \left(\max \left\{ \frac{p_n k_n^2 b_n^{\frac{1}{2}} c_n^{\frac{1}{2}}}{n}, \frac{p_n^{\frac{3}{2}} k_n^2 b_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n}, \frac{p_n k_n^2 b_n}{n} \right\} \right).$$

Proof. By the triangle inequality

$$\left\| \hat{H}_n - \overline{H}_n \right\| \leq \left\| \hat{H}_n - H_n \right\| + \left\| \overline{H}_n - H_n \right\| .$$

By the triangle inequality again, $\left\| \hat{H}_n - H_n \right\|$ is bounded by

$$\begin{aligned} & \frac{2}{n} \left\| P_{ji,n}(\hat{\lambda}_{(n)}) - P_{ji,n} \right\| \\ + \quad & \frac{2}{n} \left| \frac{1}{\hat{\sigma}^2_{(n)}} - \frac{1}{\sigma_0^2} \right| \left(\left\| R_n' R_n \right\| + 2 \left\| X_n' R_n \right\| + \left\| X_n' X_n \right\| \right) . \end{aligned} \tag{3.B.2}$$

The first term in (3.B.2) is bounded by

$$\left\{ \sum_{i=1}^{p_n} \sum_{j=1}^{p_n} \left(\frac{2}{n} tr(G_{jn}(\hat{\lambda}_{(n)}) G_{in}(\hat{\lambda}_{(n)})) - \frac{2}{n} tr(G_{jn} G_{in}) \right)^2 \right\}^{\frac{1}{2}} \tag{3.B.3}$$

By the mean value theorem,

$$\frac{2}{n} tr(G_{jn}(\hat{\lambda}_{(n)}) G_{in}(\hat{\lambda}_{(n)})) = \frac{2}{n} tr(G_{jn} G_{in}) + \frac{2}{n} \overline{\overline{\zeta}}_n' \left(\hat{\lambda}_{(n)} - \lambda_{0(n)} \right),$$

where

$$\overline{\overline{\zeta}}_{ij,n} = \left(tr\left(\overline{\overline{\zeta}}_{ijn,1} \right), \ldots, tr\left(\overline{\overline{\zeta}}_{ijn,p_n} \right) \right),$$

with

$$\begin{aligned} \overline{\overline{\zeta}}_{ijn,k} = & \, G_{in}\left(\overline{\overline{\lambda}}_{(n)} \right) G_{kn}\left(\overline{\overline{\lambda}}_{(n)} \right) G_{jn}\left(\overline{\overline{\lambda}}_{(n)} \right) \\ & + G_{kn}\left(\overline{\overline{\lambda}}_{(n)} \right) G_{in}\left(\overline{\overline{\lambda}}_{(n)} \right) G_{jn}\left(\overline{\overline{\lambda}}_{(n)} \right), \end{aligned}$$

and $\left\| \overline{\overline{\lambda}}_{(n)} - \lambda_{0(n)} \right\| \leq \left\| \hat{\lambda}_{(n)} - \lambda_{0(n)} \right\|$. Therefore the summands in (3.B.3) are

$$\frac{4}{n^2} \left[\overline{\overline{\zeta}}_{ij,n}' \left(\hat{\lambda}_{(n)} - \lambda_{0(n)} \right) \right]^2 \leq \frac{4}{n^2} \left\| \overline{\overline{\zeta}}_{ij,n} \right\|^2 \left\| \hat{\lambda}_{(n)} - \lambda_{0(n)} \right\|^2,$$

by Cauchy-Schwarz inequality, where the first factor in norms on the RHS is $\mathscr{O}\left(p_n \frac{n^2}{h_n^2} \right)$

by Lemma 3.B3. For the second term,

$$\left\| \hat{\theta}_{(n)} - \theta_{0(n)} \right\|^2 = \mathscr{O}_p\left(\frac{b_n}{n}\right),$$

by (2.B.7). So we conclude that the summands in (3.B.3) are $\mathscr{O}_p\left(\frac{p_n b_n}{n h_n^2}\right)$ and therefore (3.B.3) is $\mathscr{O}_p\left(\frac{p_n^{\frac{3}{2}} b_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n}\right)$ and it follows that so is the first term in (3.B.2).

By (2.B.8),

$$\left| \frac{1}{\hat{\sigma}_{(n)}^2} - \frac{1}{\sigma_0^2} \right| = \mathscr{O}_p\left(\max\left\{\frac{b_n^{\frac{1}{2}} c_n^{\frac{1}{2}}}{n}, \frac{p_n^{\frac{1}{2}} b_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n}, \frac{b_n}{n}\right\}\right), \tag{3.B.4}$$

which handles the second factor in the second term in (3.B.2). We shall now bound the terms inside the parentheses in the second term in (3.B.2). For the first term note that by the definition of R_n and the triangle inequality, $\|R_n' R_n\|$ is bounded by

$$\begin{aligned} \left\|A_n' A_n\right\| + 2\left\|A_n' B_n\right\| + \left\|B_n' B_n\right\| &= \mathscr{O}_p\left(\max\left\{n p_n k_n^2, n^{\frac{1}{2}} p_n k_n, \frac{n p_n}{h_n}\right\}\right) \\ &= \mathscr{O}_p\left(n p_n k_n^2\right) \end{aligned} \tag{3.B.5}$$

by Lemmas 3.B4, 3.B5 and 3.B6.

For the second term inside the parentheses we have

$$\left\|X_n' R_n\right\| \leq \left\|X_n' A_n\right\| + \left\|X_n' B_n\right\| = \mathscr{O}_p\left(\max\left\{n p_n^{\frac{1}{2}} k_n^{\frac{3}{2}}, n^{\frac{1}{2}} p_n^{\frac{1}{2}} k_n^{\frac{1}{2}}\right\}\right) = \mathscr{O}_p\left(n p_n^{\frac{1}{2}} k_n^{\frac{3}{2}}\right), \tag{3.B.6}$$

using Lemmas 3.B7 and 3.B8.

By Assumption 5, the third term inside the parentheses is

$$\left\|X_n' X_n\right\| \leq \|X_n\|^2 \leq \|X_n\|_R \|X_n\|_C = \mathscr{O}(n k_n) \tag{3.B.7}$$

From (3.B.3), (3.B.4), (3.B.5), (3.B.6) and (3.B.7), we conclude that (3.B.2) is

$$\begin{aligned} \mathscr{O}_p\left(\frac{p_n^{\frac{3}{2}} b_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n}\right) &+ \mathscr{O}_p\left(\max\left\{\frac{b_n^{\frac{1}{2}} c_n^{\frac{1}{2}}}{n}, \frac{p_n^{\frac{1}{2}} b_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n}, \frac{b_n}{n}\right\} \max\left\{p_n k_n^2, p_n^{\frac{1}{2}} k_n^{\frac{3}{2}}, k_n\right\}\right) \\ = & \quad \mathscr{O}_p\left(\frac{p_n^{\frac{3}{2}} b_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n}\right) + \mathscr{O}_p\left(\max\left\{\frac{b_n^{\frac{1}{2}} c_n^{\frac{1}{2}}}{n}, \frac{p_n^{\frac{1}{2}} b_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n}, \frac{b_n}{n}\right\} p_n k_n^2\right) \\ = & \quad \mathscr{O}_p\left(\max\left\{\frac{p_n k_n^2 b_n^{\frac{1}{2}} c_n^{\frac{1}{2}}}{n}, \frac{p_n^{\frac{3}{2}} k_n^2 b_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n}, \frac{p_n k_n^2 b_n}{n}\right\}\right). \end{aligned}$$

Then

$$\left\| \hat{H}_n - H_n \right\| = \mathscr{O}_p\left(\max\left\{ \frac{p_n k_n^2 b_n^{\frac{1}{2}} c_n^{\frac{1}{2}}}{n}, \frac{p_n^{\frac{3}{2}} k_n^2 b_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n}, \frac{p_n k_n^2 b_n}{n} \right\} \right).$$

Similarly, it may be shown that $\left\| \overline{H}_n - H_n \right\|$ has the same order, whence the lemma follows.

□

Let

$$\Xi_n = \mathbb{E}(H_n) = \begin{pmatrix} \frac{2}{n}\left(P_{ji,n} + P_{j'i,n} + \frac{1}{\sigma_0^2} A_n' A_n\right) & \frac{2}{n\sigma_0^2} A_n' X_n \\ \\ \frac{2}{n\sigma_0^2} X_n' A_n & \frac{2}{n\sigma_0^2} X_n' X_n \end{pmatrix}$$

with $P_{j'i,n}$ the $p_n \times p_n$ matrix with (i,j)-th element $tr\left(G_{jn}' G_{in}\right)$.

Lemma 3.B10. *Suppose that Assumptions 1-5 hold. Then*

(i) $\|H_n - \Xi_n\| = \mathscr{O}_p\left(\frac{p_n k_n}{n^{\frac{1}{2}}}\right)$ *if also Assumption 11 holds,*

(ii) $\left\| L_n - \frac{\sigma_0^2}{2} \Xi_n \right\| = \mathscr{O}\left(\frac{p_n}{h_n}\right)$.

Proof.

(i) $H_n - \Xi_n$ is

$$\begin{pmatrix} \frac{2}{n}\left(\frac{1}{\sigma_0^2} A_n' B_n + \frac{1}{\sigma_0^2} B_n' A_n + \frac{1}{\sigma_0^2} B_n' B_n - P_{j'i,n}\right) & \frac{2}{n\sigma_0^2} B_n' X_n \\ \\ \frac{2}{n\sigma_0^2} X_n' B_n & 0 \end{pmatrix}$$

which has norm bounded by

$$\frac{2}{\sigma_0^2}\left(\frac{2}{n}\left\|A_n' B_n\right\| + \frac{2}{n}\left\|X_n' B_n\right\| + \frac{1}{n}\left\|B_n' B_n - \sigma_0^2 P_{j'i,n}\right\|\right). \tag{3.B.8}$$

By Lemmas 3.B5 and 3.B8 the first two terms inside parentheses above are at most $\mathscr{O}_p\left(\frac{p_n k_n}{n^{\frac{1}{2}}}\right)$. The last term in parentheses in (3.B.8) has squared expectation bounded by

$$\begin{aligned} & \frac{1}{n^2} \sum_{i=1}^{p_n} \sum_{j=1}^{p_n} \mathbb{E}\left(U_n' G_{jn}' G_{in} U_n - \sigma_0^2 tr\left(G_{jn}' G_{in}\right)\right)^2 \\ = & \frac{1}{n^2} \sum_{i=1}^{p_n} \sum_{j=1}^{p_n} var\left(U_n' G_{jn}' G_{in} U_n\right) = \mathscr{O}\left(\frac{p_n^2}{n h_n}\right), \end{aligned}$$

by Lemmas 2.C3 and 2.C4. This implies that the last term in parentheses in (3.B.8) is $\mathscr{O}_p\left(\frac{p_n}{n^{\frac{1}{2}}h_n^{\frac{1}{2}}}\right)$. Therefore $\|H_n - \Xi_n\| = \mathscr{O}_p\left(\max\left\{\frac{p_n k_n}{n^{\frac{1}{2}}}, \frac{p_n}{n^{\frac{1}{2}}h_n^{\frac{1}{2}}}\right\}\right) = \mathscr{O}_p\left(\frac{p_n k_n}{n^{\frac{1}{2}}}\right)$ since h_n is bounded away from zero.

(ii)

$$L_n - \frac{\sigma_0^2}{2}\Xi_n = \begin{pmatrix} -\frac{\sigma_0^2}{n}\left(P_{ji,n} + P_{j'i,n}\right) & 0 \\ & \\ 0 & 0 \end{pmatrix},$$

which has squared norm bounded by a constant times

$$\frac{1}{n^2}\sum_{i=1}^{p_n}\sum_{j=1}^{p_n} tr^2\left(C_{jn}G_{in}\right). \tag{3.B.9}$$

Now $tr\left(C_{jn}G_{in}\right) = tr\left(G_{jn}G_{in} + G'_{jn}G_{in}\right)$, which is $\mathscr{O}\left(\frac{n}{h_n}\right)$ from Lemma 2.C3. Then (3.B.9) is $\mathscr{O}\left(\frac{p_n^2}{h_n^2}\right)$ which implies the result.

□

Lemma 3.B11. *Under Assumptions 1-5 and 10, 11 along with*

$$\frac{1}{p_n} + \frac{1}{r_n} + \frac{1}{k_n} + \frac{p_n^3 k_n^4}{n} + \frac{p_n r_n}{n} + \frac{p_n^{\frac{3}{2}} k_n}{h_n} \to 0 \text{ as } n \to \infty \tag{3.B.10}$$

and

$$\frac{r_n^2 k_n^2}{n} \text{ bounded as } n \to \infty, \tag{3.B.11}$$

the following inequalities are satisfied:

$$\operatorname{plim}\left\|\hat{H}_n^{-1}\right\| \le C \operatorname{plim}\left\|H_n^{-1}\right\| \le C \lim_{n\to\infty}\left\|\Xi_n^{-1}\right\| \le C\frac{\sigma_0^2}{2}\left(\varliminf_{n\to\infty}\eta(L_n)\right)^{-1} \le C.$$

If h_n does not diverge, the above result becomes

$$\operatorname{plim}\left\|\hat{H}_n^{-1}\right\| \le C \operatorname{plim}\left\|H_n^{-1}\right\| \le C\left(\varliminf_{n\to\infty}\eta(\Xi_n)\right)^{-1} \le C,$$

if also $\varliminf_{n\to\infty}\eta(\Xi_n) > 0$.

Proof. We first observe that:

$$\begin{aligned} \left\|\hat{H}_n^{-1}\right\| &\le \left\|\hat{H}_n^{-1} - H_n^{-1}\right\| + \left\|H_n^{-1}\right\| \\ &\le \left\|\hat{H}_n^{-1}\right\|\left\|\hat{H}_n - H_n\right\|\left\|H_n^{-1}\right\| + \left\|H_n^{-1}\right\|. \end{aligned}$$

Therefore
$$\left\| \hat{H}_n^{-1} \right\| \left(1 - \left\| \hat{H}_n - H_n \right\| \, \| H_n^{-1} \| \right) \leq \| H_n^{-1} \| \, .$$
A similar argument yields
$$\| H_n^{-1} \| \left(1 - \| H_n - \Xi_n \| \, \| \Xi_n^{-1} \| \right) \leq \| \Xi_n^{-1} \| \, ,$$
and
$$\| \Xi_n^{-1} \| \left(1 - \frac{\sigma_0^2}{2} \left\| \frac{\sigma_0^2}{2} \Xi_n - L_n \right\| \, \| L_n^{-1} \| \right) \leq \frac{\sigma_0^2}{2} \, \| L_n^{-1} \| \, ,$$
The result now follows from the last three expressions above, by taking probability limits of the expressions starting from the last displayed expression and using Lemmas 3.B9 and 3.B10 together with (3.B.10), (3.B.11) and Assumption 10. An analogous result clearly holds for $\left\| \overline{H}_n^{-1} \right\|$. □

Lemma 3.B12. *Suppose that Assumptions 1-5, 10, 11 and condition (2.3.14) hold. Then*
$$\left\| \tilde{H}_n - H_n \right\| = \mathscr{O}_p \left(\max \left\{ \frac{p_n k_n^2 c_n}{n}, \frac{p_n^2 k_n^2}{h_n^2}, \frac{p_n^{\frac{3}{2}} k_n^2 c_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n}, \frac{p_n^{\frac{7}{4}} c_n^{\frac{1}{4}}}{n^{\frac{1}{4}} h_n^{\frac{3}{2}}} \right\} \right) .$$

Proof. In this case we need to bound
$$\begin{aligned} & \frac{2}{n} \left\| P_{ji,n}(\tilde{\lambda}_{(n)}) - P_{ji,n} \right\| \\ + \; & \frac{2}{n} \left| \frac{1}{\tilde{\sigma}_{(n)}^2} - \frac{1}{\sigma_0^2} \right| \left(\| R_n' R_n \| + 2 \| X_n' R_n \| + \| X_n' X_n \| \right) . \end{aligned} \tag{3.B.12}$$

Everything follows as in the IV case except now have the different orders
$$\tilde{\sigma}_{(n)}^2 - \sigma_0^2 = \mathscr{O}_p \left(\max \left\{ \frac{c_n}{n}, \frac{p_n}{h_n^2}, \frac{p_n^{\frac{1}{2}} c_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n} \right\} \right)$$
and
$$\left\| \tilde{\theta}_{(n)} - \theta_{0(n)} \right\| = \mathscr{O}_p \left(\max \left\{ \frac{c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}, \frac{p_n^{\frac{1}{2}}}{h_n} \right\} \right)$$
from (2.B.24) and (2.B.22) respectively. The first term in (3.B.12) is then
$$\mathscr{O}_p \left(\max \left\{ \frac{p_n^{\frac{3}{2}} c_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n}, \frac{p_n^2}{h_n^2}, \frac{p_n^{\frac{7}{4}} c_n^{\frac{1}{4}}}{n^{\frac{1}{4}} h_n^{\frac{3}{2}}} \right\} \right)$$

while the second is

$$\mathscr{O}_p\left(\max\left\{\frac{p_n k_n^2 c_n}{n}, \frac{p_n^2 k_n^2}{h_n^2}, \frac{p_n^{\frac{3}{2}} k_n^2 c_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n}\right\}\right).$$

□

We may then argue in a similar way that the Hessian evaluated at the OLS estimate differs from its value at an intermediate point in norm by the same. We skip the details because they replicate those for the proof of Lemma 3.B9 above.

Lemma 3.B13. *Under Assumptions 1-5 and 10, 11 along with*

$$\frac{1}{p_n} + \frac{1}{k_n} + \frac{p_n^2 k_n^4}{n} + \frac{p_n k_n}{h_n} \to 0 \text{ as } n \to \infty, \tag{3.B.13}$$

the following inequalities hold:

$$\operatorname{plim}\left\|\tilde{H}_n^{-1}\right\| \le \operatorname{plim}\left\|H_n^{-1}\right\| \le \lim_{n\to\infty}\left\|\Xi_n^{-1}\right\| \le \frac{\sigma_0^2}{2}\left(\varliminf_{n\to\infty} \underline{\eta}(L_n)\right)^{-1} \le C.$$

Proof. Similar to proof of Lemma 3.B11. □

4 Results for covariances of autoregressive random fields defined on regularly-spaced lattices

4.1 Introduction

In this chapter we present some results on the covariance structure of random fields defined on a d-dimensional regularly-spaced lattice, with $d \geq 1$. Section 4.2 derives bounds for absolute moments of partial sums of spatial processes defined on regularly-spaced lattices. Section 4.3 demonstrates that when the spatial process is stationary and has a half-plane representation the covariance structure satisfies a generalisation of the Toeplitz property familiar from the theory of stationary time series. Section 4.4 provides an upper bound on the number of unique autocovariances that occur in the covariance matrix of stationary and unilateral processes. These results are crucial for proving the claims in Chapter 5 and also provide scope for further extension of time series theory.

Denote by $\mathbb{Z}$ the set of integers. We consider processes indexed by elements of $\mathbb{Z}^d$, which are denoted by a multiple index e.g. $t = (t_1, \ldots, t_d)$ with $t_j \in \mathbb{Z}$, $j = 1, \ldots, d$. Define the rectangular lattice

$$\mathfrak{L} = \left\{ t \in \mathbb{Z}^d : -n_{L_i} \leq t_i \leq n_{U_i}, i = 1, \ \ldots, d \right\},$$

where $n_{U_i}, n_{L_i} \geq 0$ for $i = 1, \ldots, d$.

Lattice data has been the subject of a rich literature in statistics (see references in Chapter 1 and below). In fact there is some scope for the extension of methods for lattice data, such as those outlined in this chapter and the next, to stationary spatial processes observed on a continuum. For a continuous time bandlimited process ($d = 1$) uniform sampling (i.e. sampling at regular intervals) can be employed and the spectral density of the sampled process used to recover the spectrum of the original process. This approach is not possible without bandlimiting, leading to inconsistent estimates of the spectral density. If the original process is not bandlimited, irregular sampling is preferred, an example of irregular sampling being Poisson sampling. A reference for these issues is Srivastava and Sengupta (2010). In a similar way lattice data may be viewed as sampled data from continuous space process.

4.2 Bounds for w-th absolute moments of partial sums, $w \in (1, 2]$

In this section, bounds are derived for the w-th absolute moment of partial sums of the realizations of random fields defined on a lattice, with $w \in (1, 2]$. We first impose certain conditions to reduce the class of processes under consideration to one that arises in many applications, and then obtain the bounds. The result in this section extends Lemma 1 of Robinson (1978) from time series to lattice processes. Consider a zero-mean lattice process $\{\zeta_t : t \in \mathfrak{L}\}$ defined by

$$\zeta_t = \sum_{s^1 \in \mathbb{Z}^d} \cdots \sum_{s^q \in \mathbb{Z}^d} \xi_{\mathbf{s}t}, \quad t \in \mathfrak{L},$$

where $\mathbf{s} = \left(s^1, \ldots, s^q\right)$. This definition covers situations where certain statistics of spatial processes may be expressible in terms of products of sums of random variables. Assume that this process satisfies the following conditions:

Assumption 18. $\xi_{\mathbf{s}t}$ are mean-zero and independent over t.

Assumption 19. For some $w \in (1, 2]$ there exist positive constants $\left\{\eta_{ks} : s \in \mathbb{Z}^d, 1 \leq k \leq q\right\}$, $\{a_t : t \in \mathfrak{L}\}$, such that

$$E\left|\xi_{\mathbf{s}t}\right|^w < \eta_{\mathbf{s}}^w a_t^w, \tag{4.2.1}$$

where $\eta_{\mathbf{s}} = \prod_{k=1}^q \eta_{ks^k}$ and

$$\sum_{s \in \mathbb{Z}^d} \eta_{ks} < \infty, \quad 1 \leq k \leq q. \tag{4.2.2}$$

Assumption 18 can be relaxed to an appropriate lattice martingale type condition and indeed, lattice martingales have been introduced in Cairoli and Walsh (1975), Tjøstheim (1983) and Kallianpur and Mandrekar (1983). However, extensions to lattice martingales require an assumption of the existence of an ordering in the lattice. We prefer to avoid such assumptions for the moment, and in any case this can be rather arbitrary.

Before we can introduce our result, we need to establish some more notation and illustrate it with examples. Write $N = (N_1, \ldots, N_d)$, $0 < N_i \leq n_{L_i} + n_{U_i}$ for $i = 1, \ldots, d$, and define

$$S_N = {\sum_{t(N)}}' \zeta_t,$$

where $\sum_{t(N)}'$ runs over t satisfying $-n_{L_i} < t_i \leq N_i - n_{L_i}$.There are $\prod_{i=1}^d N_i$ summands in this sum. Also write $M = (M_1, \ldots, M_d)$, M_i possibly negative, with $|M_i| < N_i$, and define

$$S_{MN} = \sum_{t(|M|,N)}{}^{\prime\prime} \zeta_t,$$

where $\sum''_{t(|M|,N)}$ runs over t satisfying

$$\begin{aligned} -n_{L_i} < t_i \leq N_i - |M_i| - n_{L_i}; \quad \text{if} \quad M_i < 0, \\ M_i - n_{L_i} < t_i \leq N_i - n_{L_i}; \quad \text{if} \quad M_i \geq 0, \end{aligned}$$

indicating that there are $\prod_{i=1}^{d} (N_i - |M_i|)$ summands in this sum.

If $M_i \geq 0$ for each $i = 1, \ldots, d$ then, unlike in time series, $S_{MN} \neq S_N - S_M$. In the d-dimensional lattice case we may write $S_{MN} = S_N - S^*_{MN}$ with $S^*_{MN} = \sum^*_{t(M,N)} \zeta_t$, $\sum^*_{t(M,N)}$ running over t satisfying $-n_{L_i} < t_i \leq N_i$ with at least one $i = 1, \ldots, d$ for which $t_i \leq M_i - n_{L_i}$. There are $\prod_{i=1}^{d} N_i - \prod_{i=1}^{d} (N_i - M_i)$ summands in this sum.

For $d = 2$, S_N consists of the sum of observations at those points in the intersection of points to the north-east of $(-n_{L_1} + 1, -n_{L_2} + 1)$ and to the south-west of (N_1, N_2). S_M is visualised similarly. S_{MN} consists of the sum of observations at those points in the intersection of points to the north-east of $(-n_{L_1} + M_1 + 1, -n_{L_2} + M_2 + 1)$ and to the south-west of (N_1, N_2). Figure 4.1 illustrates these definitions for $d = 2$; $n_{L_1} = n_{L_1} = 0$; $n_{U_1} = n_{U_2} = 6$; $(N_1, N_2) = (4, 4)$ and $(M_1, M_2) = (2, 2)$. Observations summed in S_N are those recorded at points within the solid-bordered boxed area. For S_M, S^*_{MN} and S_{MN} the points of observation are in the solid-bordered circular area, dashed polygonal area and dotted circular area respectively.

An alternative way of writing $\sum''_{t(|M|,N)}$ is $\sum_{t, t-M \in \mathfrak{L}_N}$ where

$$\mathfrak{L}_N = \left\{ t \in \mathbb{Z}^d : -n_{L_i} \leq t_i \leq N_i - n_{L_i}, i = 1, \ \ldots, d \right\}.$$

Now define $b_{wN} = 0$ if $N = (N_1, \ldots, N_d)$, $N_i \geq 0$ for $i = 1, \ldots, d$ with at least one $N_i = 0$, and $b_{wN} = \sum'_{t(N)} a^w_t$ if $N = (N_1, \ldots, N_d)$, $N_i > 0$ for $i = 1, \ldots, d$. Similarly define $b_{wMN} = 0$ if $N - |M| = (N_1 - |M_1|, \ldots, N_d - |M_d|)$, $N_i - |M_i| \geq 0$ for $i = 1, \ldots, d$ with at least one $N_i - |M_i| = 0$, and $b_{wMN} = \sum''_{t(|M|,N)} a^w_t$ if $N - |M| = (N_1 - |M_1|, \ldots, N_d - |M_d|)$, $N_i - |M_i| > 0$ for $i = 1, \ldots, d$. We are now in a position to prove the main result of this section.

Lemma 4.1. *Let Assumptions 18 and 19 hold. Then*

$$E\,|S_{MN}|^w < C\, b_{wMN}. \tag{4.2.3}$$

Note that we did not impose stationarity of ζ_t, nor did we use any half-plane representation for ζ_t. In view of this Lemma 4.1 is quite general.

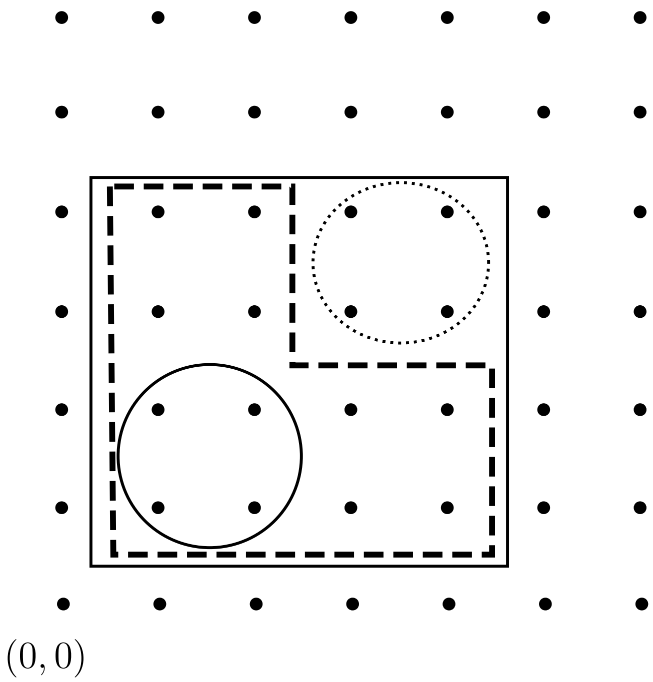

Figure 4.1: Illustration of S_N, S_M, S^*_{MN} and S_{MN} for the case $d = 2$ $n_{L_1} = n_{L_1} = 0$; $n_{U_1} = n_{U_2} = 6$; $(N_1, N_2) = (4, 4)$ and $(M_1, M_2) = (2, 2)$.

4.3 Covariance structure of stationary lattice processes with autoregressive half-plane representation

In this section we generalize the Toeplitz property of covariance matrices for stationary time series with finite autoregressive representations to stationary spatial processes with finite half-plane or quarter-plane representations.

As in Tjøstheim (1983) we define the half-space used in our representation as

$$S^{\infty}_{1+} = \Big\{t \in \mathbb{Z}^d : t_1 > 0;\ t_1 = 0, t_2 > 0;\ \cdots;\ t_1 = \cdots = t_{d-1} = 0, t_d > 0\Big\}. \quad (4.3.1)$$

We will also write $\mathbf{0}$ for the d-dimensional zero vector.

For non-negative integers $p_{L_i},\ p_{U_i}$, $i = 1, \ldots, d$, we now introduce compact notation for an AR $(p_{L_1}, p_{U_1}; \ldots; p_{L_d}, p_{U_d})$ model. First, in view of the half-plane representation we can *a priori* set, say, $p_{L_1} = 0$. Now define

$$S\left[-p_L, p_U\right] = \Big\{t \in \mathbb{Z}^d : -p_{L_i} \leq t_i \leq p_{U_i},\ i = 1, \ldots, d\Big\} \cap S^{\infty}_{1+}, \quad (4.3.2)$$

which is the truncated set of dependence 'lags'. Consider a process $\{x_t : t \in \mathfrak{L}\}$. Then we assume the existence of real numbers d_s, $s \in S\left[-p_L, p_U\right]$, such that

$$x_t = \sum_{s \in S[-p_L, p_U]} d_s x_{t-s} + \epsilon_t, \quad t \in \mathbb{Z}^d, \tag{4.3.3}$$

with ϵ_t a white-noise error term.

Denote $p_i = p_{L_i} + p_{U_i}$, $i = 1, \ldots, d$, with $p_1 \equiv p_{U_1}$ since $p_{L_1} = 0$ in by our definition of half plane, and also write $p = (p_{L_2}, \ldots, p_{L_d}, p_{U_1}, \ldots, p_{U_d})$. Let $\mathfrak{h}(p)$ denote the total number of autoregressive parameters in (4.3.3). Then

$$\mathfrak{h}(p) = p_{U_d} + \sum_{j=1}^{d-1} \prod_{i=j+1}^{d} (p_i + 1)\, p_{U_j}, \tag{4.3.4}$$

which generalizes the formulae given in Tjøstheim (1983).

Assuming that the process x_t is stationary, we can define the autocovariances

$$\gamma(k) = E x_t x_{t+k}, \; k \in \mathbb{Z}^d.$$

It is necessary to introduce an ordering of the elements of $\mathbb{Z}^d$ in order to write the objects of interest in matrical and vectorial form. Such an ordering can be carried out in many ways and as long as a consistent ordering is followed it should not matter which particular ordering is used. From a practical point of view however, certain orderings may be more beneficial in that they allow us to get a clearer picture of the structure of the covariance matrix for a truncated process on a half-plane. A clear picture of the structure will also help us in the proofs of our results in Chapter 5. We consider the cases $d = 2$ and $d = 3$, and then discuss the situation for general d. We also illustrate the relevant quarter-plane situations first and then build on this treatment to explain the differences in the half-plane case, the latter being more complicated due to negative entries in the indices. The definitions are recursive in nature.

4.3.1 d=2

This case is discussed quite extensively in the signal-processing literature for instance in Tjøstheim (1981) and Wester et al. (1990). Examples abound of two-dimensional processes, for instance with spatio-temporal data as also data with no temporal component. Examples of the latter include agricultural and horticultural data of the type used by Whittle (1954). These data were recorded on an equally-spaced grid set on a wheat field and a rectangular lattice of 1000 orange trees respectively.

Quarter-plane representations

Consider a quarter-plane representation. In this case $p_{L_2} = 0$. For each $l = 0, \ldots, p_{U_1}$, define $\check{\psi}_l^{(1)}(p)$ to be the $(p_{U_2}+1) \times 1$ vector with typical i-th element given by $\gamma(l,i)$, $i = 0, \ldots, p_{U_2}$. To illustrate,

$$\check{\psi}_l^{(1)}(p) = \begin{pmatrix} \gamma(l,0) \\ \gamma(l,1) \\ \vdots \\ \gamma(l,p_{U_2}) \end{pmatrix}.$$

Now, define $\check{\psi}^{(2)}(p)$ to be the nested vector of dimension $(p_{U_2}+1) \times (p_{U_1}+1)$ and i-th sub-vector given by $\check{\psi}_i^{(1)}(p)$, $i = 0, \ldots, p_{U_1}$. So we have

$$\check{\psi}^{(2)}(p) = \begin{pmatrix} \check{\psi}_0^{(1)}(p) \\ \check{\psi}_1^{(1)}(p) \\ \vdots \\ \check{\psi}_{p_{U_1}(p)}^{(1)} \end{pmatrix}.$$

Finally denote by $\psi^{(2)}(p)$ the $(p_{U_1}+1)(p_{U_2}+1) - 1 \times 1$ vector got by removing the first element of $\check{\psi}^{(2)}(p)$. This is now an $\mathfrak{h}(p)$-dimensional vector of covariances.

For each $l = 0, \ldots, p_{U_1}$, define $\check{\Psi}_l^{(1)}(p)$ to be the $(p_{U_2}+1) \times (p_{U_2}+1)$ Toeplitz matrix with typical (i,j)-th element given by $\gamma(l, i-j)$, $i, j = 0, \ldots, p_{U_2}$. To illustrate,

$$\check{\Psi}_l^{(1)}(p) = \begin{pmatrix} \gamma(l,0) & \gamma(l,-1) & \cdots\cdots & \gamma(l,-p_{U_2}) \\ \gamma(l,1) & \gamma(l,0) & \cdots\cdots & \gamma(l,-p_{U_2}+1) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma(l,p_{U_2}) & \gamma(l,p_{U_2}-1) & \cdots\cdots & \gamma(l,0) \end{pmatrix}.$$

Now, define $\check{\Psi}^{(2)}(p)$ to be the block-Toeplitz matrix of (block) dimension $(p_{U_1}+1)$ and (i,j)-th block given by $\Psi_{i-j}^{(1)}(p)$, $i, j = 0, \ldots, p_{U_1}$. So we have

$$\check{\Psi}^{(2)}(p) = \begin{pmatrix} \check{\Psi}_0^{(1)}(p) & \check{\Psi}_{-1}^{(1)}(p) & \cdots\cdots & \check{\Psi}_{-p_{U_1}(p)}^{(1)} \\ \check{\Psi}_1^{(1)}(p) & \check{\Psi}_0^{(1)}(p) & \cdots\cdots & \check{\Psi}_{-p_{U_1}+1}^{(1)}(p) \\ \vdots & \vdots & \vdots & \vdots \\ \check{\Psi}_{p_{U_1}}^{(1)}(p) & \check{\Psi}_{p_{U_1}-1}^{(1)}(p) & \cdots\cdots & \check{\Psi}_0^{(1)}(p) \end{pmatrix}.$$

Denote by $\Psi^{(2)}(p)$ the $(p_{U_1}+1)(p_{U_2}+1) - 1 \times (p_{U_1}+1)(p_{U_2}+1) - 1$ matrix formed by deleting the first row and first column of $\check{\Psi}^{(2)}(p)$. Then the dimension of $\Psi^{(2)}(p)$ is

$\mathfrak{h}(p) \times \mathfrak{h}(p)$.

Half-plane representations

Consider now the half-plane situation, where we have $p_{L_2} > 0$. Here, we have similar definitions with the indices running over different ranges. For each $l = 0, \ldots, p_{U_2}$, define $\breve{\psi}_l^{(1)}(p)$ to be the $(p_2+1) \times 1$ vector with typical i-th element given by $\gamma(l,i)$, $i = -p_{L_2}, \ldots, p_{U_2}$. So we have

$$\breve{\psi}_l^{(1)}(p) = \begin{pmatrix} \gamma\left(l, -p_{L_2}\right) \\ \gamma\left(l, -p_{L_2}+1\right) \\ \vdots \\ \gamma\left(l, p_{U_2}\right) \end{pmatrix}.$$

Define $\breve{\psi}^{(2)}(p)$ to be the $(p_2+1) \times (p_{U_1}+1)$-dimensional nested vector with i-th sub-vector given by $\breve{\psi}_i^{(1)}(p)$, $i = 0, \ldots, p_{U_1}$. $\breve{\psi}^{(2)}(p)$ has dimension $(p_{U_1}+1)(p_2+1) \times 1$ with $(p_{U_1}+1)(p_2+1) = \mathfrak{h}(p) + p_{L_2} + 1$. Therefore, unlike in the quarter-plane situation, we will now denote by $\psi^{(2)}(p)$ the $\mathfrak{h}(p) \times 1$ vector formed by deleting the first $p_{L_2}+1$ elements of $\breve{\psi}^{(2)}(p)$.

For each $l = 0, \ldots, p_{U_1}$, define $\breve{\Psi}_l^{(1)}(p)$ to be the $(p_2+1) \times (p_2+1)$ Toeplitz matrix with typical (i,j)-th element given by $\gamma(l, i-j)$, $i,j = 0, \ldots, p_2$. Now, define $\breve{\Psi}^{(2)}(p)$ to be the block-Toeplitz matrix of (block) dimension $(p_{U_1}+1) \times (p_{U_1}+1)$ and (i,j)-th block given by $\Psi_{i-j}^{(1)}(p)$, $i,j = 0, \ldots, p_{U_1}$. So we have

$$\breve{\Psi}^{(2)}(p) = \begin{pmatrix} \breve{\Psi}_0^{(1)}(p) & \breve{\Psi}_{-1}^{(1)}(p) & \ldots\ldots & \breve{\Psi}_{-p_{U_1}}^{(1)}(p) \\ \breve{\Psi}_1^{(1)}(p) & \breve{\Psi}_0^{(1)}(p) & \ldots\ldots & \breve{\Psi}_{-p_{U_1}+1}^{(1)}(p) \\ \vdots & \vdots & \vdots & \vdots \\ \breve{\Psi}_{p_{U_1}}^{(1)}(p) & \breve{\Psi}_{p_{U_1}-1}^{(1)}(p) & \ldots\ldots & \breve{\Psi}_0^{(1)}(p) \end{pmatrix}.$$

$\Psi^{(2)}(p)$ has dimension $(p_{U_1}+1)(p_2+1) \times (p_{U_1}+1)(p_2+1)$ and

$$(p_{U_1}+1)(p_2+1) = \mathfrak{h}(p) + p_{L_2} + 1.$$

Again, unlike in the quadrant situation, we will denote by $\Psi^{(2)}(p)$ the $\mathfrak{h}(p) \times \mathfrak{h}(p)$ matrix formed by deleting the first $p_{L_2}+1$ rows and columns of $\breve{\Psi}^{(2)}(p)$.

4.3.2 d=3

Three-dimensional lattice data are frequently encountered in physical sciences, but are not restricted to this field. For instance, two-dimensional agricultural data observed over a period of time constitutes a three-dimensional lattice dataset.

Quarter-plane representations

In this case we have $p_{L_2} = p_{L_3} = 0$. We build the definitions analogously to the $d = 2$ case. For $l = 0, \ldots, p_{U_1}$ and $m = 0, \ldots, p_{U_2}$, define $\check{\psi}^{(1)}_{l,m}(p)$ to be the $(p_{U_3} + 1) \times 1$ vector with typical i-th element given by $\gamma(l, m, i)$, $i = 0, \ldots, p_{U_3}$, that is

$$\check{\psi}^{(1)}_{l,m}(p) = \begin{pmatrix} \gamma(l, m, 0) \\ \gamma(l, m, 1) \\ \vdots \\ \gamma(l, m, p_{U_3}) \end{pmatrix}.$$

Defining $\check{\psi}^{(2)}_{m}(p)$ to be the $(p_{U_3} + 1) \times (p_{U_1} + 1)$-dimensional nested vector with i-th sub-vector given by $\check{\psi}^{(1)}_{i,m}(p)$, $i = 0, \ldots, p_{U_1}$, we get

$$\check{\psi}^{(2)}_{m}(p) = \begin{pmatrix} \check{\psi}^{(1)}_{0,m}(p) \\ \check{\psi}^{(1)}_{1,m}(p) \\ \vdots \\ \check{\psi}^{(1)}_{p_{U_1},m}(p) \end{pmatrix}.$$

Finally, define $\check{\psi}^{(3)}(p)$ to be the twice nested vector of dimension $\prod_{i=1}^{3}(p_{U_i} + 1)$ and i-th block given by $\check{\psi}^{(2)}_{i}(p)$, $i, j = 0, \ldots, p_{U_2}$, yielding

$$\check{\psi}^{(3)}(p) = \begin{pmatrix} \check{\psi}^{(2)}_{0}(p) \\ \check{\psi}^{(2)}_{1}(p) \\ \vdots \\ \check{\psi}^{(2)}_{p_{U_2}}(p) \end{pmatrix}.$$

Denote by $\psi^{(3)}(p)$ the $\prod_{i=1}^{3}(p_{U_i} + 1) - 1$-dimensional vector formed by deleting the first element of $\check{\psi}^{(3)}(p)$. Then the dimension of $\psi^{(3)}(p)$ is $\mathfrak{h}(p) \times \mathfrak{h}(p)$.

We now define the matrices. For $l = 0, \ldots, p_{U_1}$ and $m = 0, \ldots, p_{U_2}$, define $\check{\Psi}^{(1)}_{l,m}(p)$ to be the $(p_{U_3} + 1) \times (p_{U_3} + 1)$ Toeplitz matrix with typical (i, j)-th element given by $\gamma(l, m, i - j)$, $i, j = 0, \ldots, p_{U_3}$, so that

$$\check{\Psi}^{(1)}_{l,m}(p) = \begin{pmatrix} \gamma(l, m, 0) & \gamma(l, m, -1) & \cdots\cdots & \gamma(l, m, -p_{U_3}) \\ \gamma(l, m, 1) & \gamma(l, m, 0) & \cdots\cdots & \gamma(l, m, -p_{U_3} + 1) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma(l, m, p_{U_3}) & \gamma(l, m, p_{U_3} - 1) & \cdots\cdots & \gamma(l, m, 0) \end{pmatrix}.$$

Now, define $\breve{\Psi}_m^{(2)}(p)$ to be the block-Toeplitz with Topelitz blocks matrix of (block) dimension $(p_{U_1}+1)$ and (i,j)-th block given by $\Psi_{i-j,m}^{(1)}(p)$, $i,j=0,\ldots,p_{U_1}$. So we have

$$\breve{\Psi}_m^{(2)}(p)=\begin{pmatrix} \breve{\Psi}_{0,m}^{(1)}(p) & \breve{\Psi}_{-1,m}^{(1)}(p) & \ldots\ldots & \breve{\Psi}_{-p_{U_1},m}^{(1)}(p) \\ \breve{\Psi}_{1,m}^{(1)}(p) & \breve{\Psi}_{0,m}^{(1)}(p) & \ldots\ldots & \breve{\Psi}_{-p_{U_1}+1,m}^{(1)}(p) \\ \vdots & \vdots & \vdots & \vdots \\ \breve{\Psi}_{p_{U_1},m}^{(1)}(p) & \breve{\Psi}_{p_{U_1}-1,m}^{(1)}(p) & \ldots\ldots & \breve{\Psi}_{0,m}^{(1)}(p) \end{pmatrix}.$$

Finally, define $\breve{\Psi}^{(3)}(p)$ to be the (thrice) block-Toeplitz matrix of (block) dimension $(p_{U_2}+1)\times(p_{U_2}+1)$ and (i,j)-th block given by $\breve{\Psi}_{i-j}^{(2)}(p)$, $i,j=0,\ldots,p_{U_2}$. So we have

$$\breve{\Psi}^{(3)}(p)=\begin{pmatrix} \breve{\Psi}_0^{(2)}(p) & \breve{\Psi}_{-1}^{(2)}(p) & \ldots\ldots & \breve{\Psi}_{-p_{U_2}}^{(2)}(p) \\ \breve{\Psi}_1^{(2)}(p) & \breve{\Psi}_0^{(2)}(p) & \ldots\ldots & \breve{\Psi}_{-p_{U_2}+1}^{(2)}(p) \\ \vdots & \vdots & \vdots & \vdots \\ \breve{\Psi}_{p_{U_2}}^{(2)}(p) & \breve{\Psi}_{p_{U_2}-1}^{(2)}(p) & \ldots\ldots & \breve{\Psi}_0^{(2)}(p) \end{pmatrix}.$$

Now denote by $\Psi^{(3)}(p)$ the $\prod_{i=1}^3(p_{U_i}+1)-1$-dimensional matrix formed by deleting the first row and first column of $\breve{\Psi}^{(3)}(p)$. Then the dimension of $\Psi^{(3)}$ is $\mathfrak{h}(p)\times\mathfrak{h}(p)$.

Half-plane representations

Now $p_{L_2}>0$ or/and $p_{L_3}>0$. For $l=0,\ldots,p_{U_1}$ and $m=-p_{L_2},\ldots,p_{U_2}$, define $\breve{\psi}_{l,m}^{(1)}(p)$ to be the $(p_3+1)\times 1$ vector with typical i-th element given by $\gamma(l,m,i)$, $i=-p_{L_3},\ldots,p_{U_3}$. This gives

$$\breve{\psi}_{l,m}^{(1)}(p)=\begin{pmatrix} \gamma(l,m,-p_{L_3}) \\ \gamma(l,m,-p_{L_3}+1) \\ \vdots \\ \gamma(l,m,p_{U_3}) \end{pmatrix}.$$

Now, define $\breve{\psi}_m^{(2)}(p)$ to be the $(p_3+1)\times(p_{U_1}+1)$-dimensional nested vector with i-th sub-vector given by $\breve{\psi}_{i,m}^{(1)}(p)$, $i=0,\ldots,p_{U_1}$. So we have

$$\breve{\psi}_m^{(2)}(p)=\begin{pmatrix} \breve{\psi}_{0,m}^{(1)}(p) \\ \breve{\psi}_{1,m}^{(1)}(p) \\ \vdots \\ \breve{\psi}_{p_{U_1},m}^{(1)}(p) \end{pmatrix}.$$

Finally, write $\breve{\psi}^{(3)}(p)$ for the $\prod_{i=1}^3(p_i+1)$-dimensional nested vector with i-th sub-vector given by $\breve{\psi}_i^{(2)}(p)$, $i=-p_{L_2},\ldots,p_{U_2}$, , where by definition of half-plane $p_{L_1}=0$, giving

$$\check{\psi}^{(3)}(p) = \begin{pmatrix} \check{\psi}^{(2)}_{-p_{L_2}}(p) \\ \check{\psi}^{(2)}_{-p_{L_2}+1}(p) \\ \vdots \\ \check{\psi}^{(2)}_{p_{U_2}}(p) \end{pmatrix}.$$

$\check{\psi}^{(3)}(p)$ has dimension $\prod_{i=1}^{3}(p_i+1)$ and also $\prod_{i=1}^{3}(p_i+1) = \mathfrak{h}(p) + p_{L_3} + p_{L_2}(p_3+1) + 1$. Therefore, unlike in the quarter-plane situation, we will now denote by $\psi^{(3)}(p)$ the $\mathfrak{h}(p) \times 1$ vector formed by the following procedure:

1. Delete each of the $\check{\psi}^{(1)}_{0,m}(p)$, $m = -p_{L_2}, \ldots, -1$.

2. Delete the first $p_{L_3}+1$ elements from $\check{\psi}^{(2)}_{0}(p)$.

The total elements then deleted are $p_{L_2}(p_3+1) + p_{L_3} + 1$ in number, and the dimension of $\psi^{(3)}(p)$ follows.

For the matrices, we again proceed similarly. For $l = 0, \ldots, p_{U_1}$ and $m = -p_{L_2}, \ldots, p_{U_2}$, define $\check{\Psi}^{(1)}_{l,m}(p)$ to be the $(p_3+1)\times(p_3+1)$ Toeplitz matrix with typical (i,j)-th element given by $\gamma(l,m,i-j)$, $i,j = -p_{L_3}, \ldots, p_{U_3}$. To illustrate, $\check{\Psi}^{(1)}_{l,m}(p)$ is

$$\begin{pmatrix} \gamma(l,m,0) & \gamma(l,m,-1) & \ldots & \gamma(l,m,-p_3) \\ \gamma(l,m,1) & \gamma(l,m,0) & \ldots & \gamma(l,m,-p_3+1) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma(l,m,p_3) & \gamma(l,m,p_3-1) & \ldots & \gamma(l,m,0) \end{pmatrix}.$$

Defining $\check{\Psi}^{(2)}_{m}(p)$ to be the block-Toeplitz with Toeplitz blocks matrix of (block) dimension $(p_{U_1}+1)$ and (i,j)-th block given by $\Psi^{(1)}_{i-j,m}(p)$, $i,j = 0, \ldots, p_{U_1}$, we get

$$\check{\Psi}^{(2)}_{m}(p) = \begin{pmatrix} \check{\Psi}^{(1)}_{0,m}(p) & \check{\Psi}^{(1)}_{-1,m}(p) & \ldots\ldots & \check{\Psi}^{(1)}_{-p_{U_1},m}(p) \\ \check{\Psi}^{(1)}_{1,m}(p) & \check{\Psi}^{(1)}_{0,m}(p) & \ldots\ldots & \check{\Psi}^{(1)}_{-p_{U_1}+1,m}(p) \\ \vdots & \vdots & \vdots & \vdots \\ \check{\Psi}^{(1)}_{p_{U_1},m}(p) & \check{\Psi}^{(1)}_{p_{U_1}-1,m}(p) & \ldots\ldots & \check{\Psi}^{(1)}_{0,m}(p) \end{pmatrix}.$$

Lastly, define $\check{\Psi}^{(3)}(p)$ to be the (thrice) block-Toeplitz matrix of (block) dimension $(p_2+1)\times(p_2+1)$ and (i,j)-th block given by $\check{\Psi}^{(2)}_{i-j}(p)$, $i,j = -p_{L_2}, \ldots, p_{U_2}$. So we have

$$\check{\Psi}^{(3)}(p) = \begin{pmatrix} \check{\Psi}^{(2)}_{0}(p) & \check{\Psi}^{(2)}_{-1}(p) & \ldots\ldots & \check{\Psi}^{(2)}_{-p_2}(p) \\ \check{\Psi}^{(2)}_{1}(p) & \check{\Psi}^{(2)}_{0}(p) & \ldots\ldots & \check{\Psi}^{(2)}_{-p_2+1}(p) \\ \vdots & \vdots & \vdots & \vdots \\ \check{\Psi}^{(2)}_{p_2}(p) & \check{\Psi}^{(2)}_{p_2-1}(p) & \ldots\ldots & \check{\Psi}^{(2)}_{0}(p) \end{pmatrix}.$$

Now denote by $\Psi^{(3)}(p)$ the $\prod_{i=1}^{3}(p_{U_i}+1)-1$-dimensional matrix formed by deleting those rows and columns of $\check{\Psi}^{(3)}(p)$ corresponding to the elements of $\check{\psi}^{(3)}(p)$ deleted earlier. For instance, if the i-th element of $\check{\psi}^{(3)}(p)$ was deleted then we delete the i-th row and i-column of $\check{\Psi}^{(3)}(p)$. We repeat this for each deleted element of $\check{\psi}^{(3)}(p)$. Then the dimension of $\Psi^{(3)}(p)$ is $\mathfrak{h}(p)\times\mathfrak{h}(p)$.

4.3.3 General d

Quarter-plane representations

In this case we have $p_{L_2}=p_{L_3}=\ldots=p_{L_d}=0$. For $l_i=0,\ldots,p_{U_i}$, $i=1,\ldots,d-1$, define $\check{\psi}^{(1)}_{l_1,\ldots,l_{d-1}}(p)$ to be the $(p_{U_d}+1)\times 1$ vector with typical i-th element given by $\gamma(l_1,\ldots,l_{d-1},i)$, $i=0,\ldots,p_{U_d}$. Then

$$\check{\psi}^{(1)}_{l_1,\ldots,l_{d-1}}(p)=\begin{pmatrix}\gamma\left(l_1,\ldots,l_{d-1},0\right)\\ \gamma\left(l_1,\ldots,l_{d-1},1\right)\\ \vdots\\ \gamma\left(l_1,\ldots,l_{d-1},p_{U_d}\right)\end{pmatrix}.$$

Next, for $l_i=0,\ldots,p_{U_i}$, $i=2,\ldots,d-1$ define $\check{\psi}^{(2)}_{l_2,\ldots,l_{d-1}}(p)$ to be the nested vector of (nested) dimension $(p_{U_1}+1)$ and i-th sub-vector given by $\psi^{(1)}_{i,l_2,\ldots,l_{d-1}}(p)$, $i=0,\ldots,p_{U_1}$. So we have

$$\check{\psi}^{(2)}_{l_2,\ldots,l_{d-1}}(p)=\begin{pmatrix}\check{\psi}^{(1)}_{0,l_2,\ldots,l_{d-1}}(p)\\ \check{\psi}^{(1)}_{1,l_2,\ldots,l_{d-1}}(p)\\ \vdots\\ \check{\psi}^{(1)}_{p_{U_1},l_2,\ldots,l_{d-1}}(p)\end{pmatrix}.$$

Proceeding in this manner, for $l_{d-1}=0,\ldots,p_{U_{d-1}}$ we define $\check{\psi}^{(d-1)}_{l_{d-1}}(p)$ to be the nested vector of (nested) dimension $\left(p_{U_{d-2}}+1\right)\times 1$ and i-th sub-vector given by $\check{\psi}^{(d-2)}_{i,l_{d-1}}(p)$, $i=0,\ldots,p_{U_{d-2}}$, yielding

$$\check{\psi}^{(d-1)}_{l_{d-1}}(p)=\begin{pmatrix}\check{\psi}^{(d-2)}_{0,l_{d-1}}(p)\\ \check{\psi}^{(d-2)}_{1,l_{d-1}}(p)\\ \vdots\\ \check{\psi}^{(d-2)}_{p_{U_{d-2}},l_{d-1}}(p)\end{pmatrix}.$$

Finally, define $\check{\psi}^{(d)}(p)$ to be the nested vector of (nested) dimension $(p_{U_d}+1)$ and i-th sub-vector given by $\check{\psi}_i^{(d-1)}(p)$, $i,j=0,\dots,p_{U_{d-1}}$. This implies that

$$\check{\psi}^{(d)}(p)=\begin{pmatrix} \check{\psi}_0^{(d-1)}(p) \\ \check{\psi}_1^{(d-1)}(p) \\ \vdots \\ \check{\psi}_{p_{U_{d-1}}}^{(d-1)}(p) \end{pmatrix}.$$

Now denote by $\psi^{(d)}(p)$ the $\prod_{i=1}^{d}(p_{U_i}+1)-1$-dimensional vector formed by deleting the first element of $\check{\psi}^{(d)}(p)$. Then the dimension of $\psi^{(d)}(p)$ is $\mathfrak{h}(p)\times 1$.
Coming to the matrices, for $l_i=0,\dots,p_{U_i}$, $i=1,\dots,d-1$, we define $\check{\Psi}^{(1)}_{l_1,\dots,l_{d-1}}(p)$ to be the $(p_{U_d}+1)$-dimensional Toeplitz matrix with typical (i,j)-th element given by $\gamma(l_1,\dots,l_{d-1},i-j)$, $i,j=0,\dots,p_{U_d}$. This means that $\check{\Psi}^{(1)}_{l_1,\dots,l_{d-1}}(p)$ equals

$$\begin{pmatrix} \gamma\left(l_1,\dots,l_{d-1},0\right) & \gamma\left(l_1,\dots,l_{d-1},-1\right) & \dots & \gamma\left(l_1,\dots,l_{d-1},-p_{U_d}\right) \\ \gamma\left(l_1,\dots,l_{d-1},1\right) & \gamma\left(l_1,\dots,l_{d-1},0\right) & \dots & \gamma\left(l_1,\dots,l_{d-1},-p_{U_d}+1\right) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma\left(l_1,\dots,l_{d-1},p_{U_d}\right) & \gamma\left(l_1,\dots,l_{d-1},p_{U_d}-1\right) & \dots & \gamma\left(l_1,\dots,l_{d-1},0\right) \end{pmatrix}.$$

Next, for $l_i=0,\dots,p_{U_i}$, $i=2,\dots,d-1$ define $\check{\Psi}^{(2)}_{l_2,\dots,l_{d-1}}(p)$ to be the block Toeplitz with Toeplitz blocks matrix of (nested) dimension $(p_{U_1}+1)$ and (i,j)-th block given by $\check{\Psi}^{(1)}_{i-j,l_2,\dots,l_{d-1}}(p)$, $i,j=0,\dots,p_{U_1}$. So we have

$$\check{\Psi}^{(2)}_{l_2,\dots,l_{d-1}}(p)=\begin{pmatrix} \check{\Psi}^{(1)}_{0,l_2,\dots,l_{d-1}}(p) & \check{\Psi}^{(1)}_{-1,l_2,\dots,l_{d-1}}(p) & \dots & \Psi^{(d)}(p)^{(1)}_{-p_{U_1},l_2,\dots,l_{d-1}}(p) \\ \check{\Psi}^{(1)}_{1,l_2,\dots,l_{d-1}}(p) & \check{\Psi}^{(1)}_{0,l_2,\dots,l_{d-1}}(p) & \dots & \Psi^{(d)}(p)^{(1)}_{-p_{U_1}+1,l_2,\dots,l_{d-1}}(p) \\ \vdots & \vdots & \vdots & \vdots \\ \check{\Psi}^{(1)}_{p_{U_1},l_2,\dots,l_{d-1}}(p) & \check{\Psi}^{(1)}_{p_{U_1}-1,l_2,\dots,l_{d-1}}(p) & \dots & \check{\Psi}^{(1)}_{0,l_2,\dots,l_{d-1}}(p) \end{pmatrix}.$$

Proceeding as above, for $l_{d-1}=0,\dots,p_{U_{d-1}}$ we define $\check{\Psi}^{(d-1)}_{l_{d-1}}(p)$ to be the nested block-Toeplitz matrix of (block) dimension $\left(p_{U_{d-2}}+1\right)\times\left(p_{U_{d-2}}+1\right)$ and (i,j)-th block given by $\check{\Psi}^{(d-2)}_{i-j,l_{d-1}}(p)$, $i,j=0,\dots,p_{U_{d-2}}$, so that

$$\check{\Psi}^{(d-1)}_{l_{d-1}}(p)=\begin{pmatrix} \check{\Psi}^{(d-2)}_{0,l_{d-1}}(p) & \check{\Psi}^{(d-2)}_{-1,l_{d-1}}(p) & \dots & \check{\Psi}^{(d-2)}_{-p_{U_{d-2}},l_{d-1}}(p) \\ \check{\Psi}^{(d-2)}_{1,l_{d-1}}(p) & \check{\Psi}^{(d-2)}_{0,l_{d-1}}(p) & \dots & \check{\Psi}^{(d-2)}_{-p_{U_{d-2}}+1,l_{d-1}}(p) \\ \vdots & \vdots & \vdots & \vdots \\ \check{\Psi}^{(d-2)}_{p_{U_{d-2}},l_{d-1}}(p) & \check{\Psi}^{(d-2)}_{p_{U_{d-2}}-1,l_{d-1}}(p) & \dots & \check{\Psi}^{(d-2)}_{0,l_{d-1}}(p) \end{pmatrix}.$$

The last step consists of defining $\check{\Psi}^{(d)}(p)$ to be the block-Toeplitz matrix of (block) dimension $\left(p_{U_{d-1}}+1\right) \times \left(p_{U_{d-1}}+1\right)$ and (i,j)-th block given by $\check{\Psi}^{(d-1)}_{i-j}(p)$, $i,j = 0,\ldots,p_{U_{d-1}}$. This yields the general form

$$\check{\Psi}^{(d)}(p) = \begin{pmatrix} \check{\Psi}^{(d-1)}_{0}(p) & \check{\Psi}^{(d-1)}_{-1}(p) & \cdots\cdots & \check{\Psi}^{(d-1)}_{-p_{U_{d-1}}}(p) \\ \check{\Psi}^{(d-1)}_{1}(p) & \check{\Psi}^{(d-1)}_{0}(p) & \cdots\cdots & \check{\Psi}^{(d-1)}_{-p_{U_{d-1}}+1}(p) \\ \vdots & \vdots & \vdots & \vdots \\ \check{\Psi}^{(d-1)}_{p_{U_{d-1}}}(p) & \check{\Psi}^{(d-1)}_{p_{U_{d-1}}-1}(p) & \cdots\cdots & \check{\Psi}^{(d-1)}_{0}(p) \end{pmatrix},$$

for the covariance matrix. Now denote by $\Psi^{(d)}(p)$ the $\prod_{i=1}^{d}(p_{U_i}+1)-1$-dimensional matrix formed by deleting the first row and first column of $\check{\Psi}^{(d)}(p)$. Clearly the dimension of $\Psi^{(d)}(p)$ is $\mathfrak{h}(p) \times \mathfrak{h}(p)$.

Half-plane representations

Now $p_{L_i} > 0$ for some $i = 1,\ldots,d$. For $l_i = -p_{L_i},\ldots,p_{U_i}$, $i = 1,\ldots,d-1$; $p_{L_1} = 0$, define $\check{\psi}^{(1)}_{l_1,\ldots,l_{d-1}}(p)$ to be the $(p_d+1)\times 1$ vector with typical i-th element given by $\gamma(l_1,\ldots,l_{d-1},i)$, $i = -p_{L_d},\ldots,p_{U_d}$. Next, for $l_i = -p_{L_i},\ldots,p_{U_i}$, $i = 2,\ldots,d-1$ define $\check{\psi}^{(2)}_{l_2,\ldots,l_{d-1}}(p)$ to be the nested vector of (nested) dimension $(p_{U_1}+1)$ and i-th sub-vector given by $\psi^{(1)}_{i,l_2,\ldots,l_{d-1}}(p)$, $i = 0,\ldots,p_{U_1}$. Proceeding in this manner, for $l_{d-1} = -p_{L_{d-1}},\ldots,p_{U_{d-1}}$ we define $\check{\psi}^{(d-1)}_{l_{d-1}}(p)$ to be the nested vector of (nested) dimension $(p_{d-2}+1)\times 1$ and i-th sub-vector given by $\check{\psi}^{(d-2)}_{i,l_{d-1}}(p)$, $i = -p_{L_{d-2}},\ldots,p_{U_{d-2}}$. Finally, define $\check{\psi}^{(d)}(p)$ to be the nested vector of (nested) dimension (p_d+1) and i-th sub-vector given by $\check{\psi}^{(d-1)}_{i}(p)$, $i = -p_{L_{d-1}},\ldots,p_{U_{d-1}}$. So we have

$$\check{\psi}^{(d)}(p) = \begin{pmatrix} \check{\psi}^{(d-1)}_{-p_{L_{d-1}}}(p) \\ \check{\psi}^{(d-1)}_{-p_{L_{d-1}}+1}(p) \\ \vdots \\ \check{\psi}^{(d-1)}_{p_{U_{d-1}}}(p) \end{pmatrix}.$$

Now $\check{\psi}^{(d)}(p)$ has dimension $\prod_{i=1}^{d}(p_i+1)\times 1$ where we note that $p_{L_1} = 0$, so that $\prod_{i=1}^{d}(p_i+1) = \mathfrak{h}(p) + p_{L_d} + p_{L_{d-1}}(p_d+1) + \ldots + p_{L_2}(p_3+1)\ldots(p_d+1)+1$.

Define $\psi^{(d)}(p)$ as the $\mathfrak{h}(p)\times 1$ vector formed using the following procedure:

(1) Delete each of the $\check{\psi}^{(1)}_{0,l_2,\ldots,l_{d-1}}(p)$, $l_2 = -p_{L_2},\ldots,-1$ and $l_i = -p_{L_i},\ldots,p_{U_i}$, $i = 3,\ldots,d-1$.

(2) Delete each of the $\check{\psi}^{(2)}_{0,l_3,\ldots,l_{d-1}}(p)$, $l_3 = -p_{L_3},\ldots,-1$ and $l_i = -p_{L_i},\ldots,p_{U_i}$, $i = 4,\ldots,d-1$.

$\vdots$

$\vdots$

$(d-2)$ Delete each of the $\check{\psi}_{0,l_{d-1}}^{(d-2)}(p)$, $l_{d-1} = -p_{L_{d-1}}, \ldots, -1$.

$(d-1)$ Delete the first $p_{L_d} + 1$ elements of $\check{\psi}_0^{(d-1)}(p)$.

The total elements thus deleted are

$$p_{L_2}\,(p_3+1)\ldots(p_d+1) + \ldots + p_{L_{d-1}}\,(p_d+1) + p_{L_d} + 1$$

in number, and the dimension of $\psi^{(d)}(p)$ follows. By construction $\psi^{(d)}(p)$ has elements $\gamma(s)$, $s \in S\,[-p_L, p_U]$.

We now define the matrices. For $l_1 = 0, \ldots, p_{U_1}$ and $l_i = -p_{L_i}, \ldots, p_{U_i}$, $i = 2, \ldots, d-1$, define $\check{\Psi}_{l_1,\ldots,l_{d-1}}^{(1)}(p)$ to be the (p_d+1)-dimensional Toeplitz matrix with typical (i,j)-th element given by $\gamma(l_1, \ldots, l_{d-1}, i-j)$, $i, j = -p_{L_d}, \ldots, p_{U_d}$. Next, for $l_i = -p_{L_i}, \ldots, p_{U_i}$, $i = 2, \ldots, d-1$ define $\check{\Psi}_{l_2,\ldots,l_{d-1}}^{(2)}(p)$ to be the block Toeplitz with Toeplitz blocks matrix of (nested) dimension $(p_{U_1}+1)$ and (i,j)-th block given by $\check{\Psi}_{i-j,l_2,\ldots,l_{d-1}}^{(1)}(p)$, $i, j = 0, \ldots, p_{U_1}$. Proceeding in this manner, for $l_{d-1} = -p_{L_{d-1}}, \ldots, p_{U_{d-1}}$ we define $\check{\Psi}_{l_{d-1}}^{(d-1)}(p)$ to be the nested block-Toeplitz matrix of (block) dimension $(p_{d-2}+1) \times (p_{d-2}+1)$ and (i,j)-th block given by $\check{\Psi}_{i-j,l_{d-1}}^{(d-2)}(p)$, $i, j = -p_{L_{d-2}}, \ldots, p_{U_{d-2}}$. So we have

$$\check{\Psi}_{l_{d-1}}^{(d-1)}(p) = \begin{pmatrix} \check{\Psi}_{0,l_{d-1}}^{(d-2)}(p) & \check{\Psi}_{-1,l_{d-1}}^{(d-2)}(p) & \ldots & \check{\Psi}_{-p_{d-2},l_{d-1}}^{(d-2)}(p) \\ \check{\Psi}_{1,l_{d-1}}^{(d-2)}(p) & \check{\Psi}_{0,l_{d-1}}^{(d-2)}(p) & \ldots & \check{\Psi}_{-p_{d-2}+1,l_{d-1}}^{(d-2)}(p) \\ \vdots & \vdots & \vdots & \vdots \\ \check{\Psi}_{p_{d-2},l_{d-1}}^{(d-2)}(p) & \check{\Psi}_{p_{d-2}-1,l_{d-1}}^{(d-2)}(p) & \ldots & \check{\Psi}_{0,l_{d-1}}^{(d-2)}(p) \end{pmatrix}.$$

Finally, define $\check{\Psi}^{(d)}(p)$ to be the block-Toeplitz matrix of (block) dimension $(p_{d-1}+1) \times (p_{d-1}+1)$ and (i,j)-th block given by $\check{\Psi}_{i-j}^{(d-1)}(p)$, $i, j = -p_{L_{d-1}}, \ldots, p_{U_{d-1}}$. So in this case we obtain the general form of the covariance matrix as

$$\check{\Psi}^{(d)}(p) = \begin{pmatrix} \check{\Psi}_0^{(d-1)}(p) & \check{\Psi}_{-1}^{(d-1)}(p) & \ldots\ldots & \check{\Psi}_{-p_{d-1}}^{(d-1)}(p) \\ \check{\Psi}_1^{(d-1)}(p) & \check{\Psi}_0^{(d-1)}(p) & \ldots\ldots & \check{\Psi}_{-p_{d-1}+1}^{(d-1)}(p) \\ \vdots & \vdots & \vdots & \vdots \\ \check{\Psi}_{p_{d-1}}^{(d-1)}(p) & \check{\Psi}_{p_{d-1}-1}^{(d-1)}(p) & \ldots\ldots & \check{\Psi}_0^{(d-1)}(p) \end{pmatrix}.$$

Now denote by $\Psi^{(d)}(p)$ the matrix formed by deleting those rows and columns of

$\check{\Psi}^{(d)}(p)$ corresponding to the elements deleted from $\check{\psi}^{(d)}(p)$ above. Then the dimension of $\Psi^{(d)}(p)$ is $\mathfrak{h}(p) \times \mathfrak{h}(p)$.

4.4 Counting covariances in stationary and unilateral lattice autoregressive models

Autoregressive models on d-dimensional lattices can generate covariance matrices of the form $\Psi^{(d)}(p)$ which differ from those in the time series case in the number of unique covariances amongst their elements. Consider a stationary time series x_t with an AR(k) representation

$$x_t = \sum_{j=1}^{k} a_j x_{t-j} + \epsilon_t, \tag{4.4.1}$$

Then we have

$$\Psi^{(1)}(k) = \begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \gamma(k-1) \\ \gamma(1) & \gamma(0) & \dots & \gamma(k-2) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma(k-1) & \gamma(k-2) & \dots & \gamma(0) \end{pmatrix},$$

which is a Toeplitz matrix with k unique autocovariances, which is also the dimension of the matrix. On the other hand, consider a 2-dimensional lattice process x_t with an AR$(0,1;1,1)$ representation. In this case

$$\Psi^{(2)}(0,1;1,1) = \begin{pmatrix} \gamma(0,0) & \gamma(-1,0) & \gamma(-1,2) & \gamma(-1,1) \\ & \gamma(0,0) & \gamma(0,2) & \gamma(0,1) \\ & & \gamma(0,0) & \gamma(0,-1) \\ & & & \gamma(0,0) \end{pmatrix},$$

which is a 4×4 matrix with 6 unique covariances. While the above may suggest that the number of unique covariances in such matrices is $\prod_{i=1}^{d}(p_i+1)$, this is in fact incorrect as the following example shows. A 2-dimensional lattice process x_t with an AR$(0,2;1,1)$ representation has $\Psi^{(2)}(0,1;2,1)$ given by

$$\begin{pmatrix} \gamma(0,0) & \gamma(-1,0) & \gamma(-2,0) & \gamma(-1,2) & \gamma(-2,2) & \gamma(-1,1) & \gamma(-2,1) \\ & \gamma(0,0) & \gamma(-1,0) & \gamma(0,2) & \gamma(-1,2) & \gamma(0,1) & \gamma(-1,1) \\ & & \gamma(0,0) & \gamma(1,2) & \gamma(0,2) & \gamma(1,1) & \gamma(0,1) \\ & & & \gamma(0,0) & \gamma(-1,0) & \gamma(0,-1) & \gamma(-1,-1) \\ & & & & \gamma(0,0) & \gamma(1,-1) & \gamma(-1,0) \\ & & & & & \gamma(0,0) & \gamma(-1,-1) \\ & & & & & & \gamma(0,0) \end{pmatrix},$$

which is a 7×7 matrix with 11 unique covariances, and the latter obviously does not equal $(p_1+1) \times (p_2+1) = 9$. This indicates the need for a formula which enables us to calculate the number of unique covariances for a lattice autoregressive model. We will provide an upper bound for the number of unique covariances in $\hat{\Psi}^{(d)}(p)$.

Proposition 4.1. *Suppose that $\{x_t; t \in \mathfrak{L}\}$ is a stationary random field with the unilateral representation (4.3.3). Then the number of unique covariances in $\hat{\Psi}^{(d)}(p)$ does not exceed*

$$\mathfrak{C}(p) = 1 + \sum_{l=1}^{d-1} 2^{d-l-1} \sum_{\#(l=0)} \prod_{\substack{k=1 \\ \nsim 0_d^l}}^{d} p_k + 2^{d-1} \prod_{k=1}^{d} p_k, \tag{4.4.2}$$

where $\sum\limits_{\#(l=0)}$ sums over all the possible ways in which the vector $(p_1, p_2, \ldots, p_d)'$ can have l entries equal to 0 and the product $\prod\limits_{\substack{k=1 \\ \nsim 0_d^l}}^{d}$ multiplies over k such that the l zero entries of $(p_1, p_2, \ldots, p_d)'$ are excluded.

The proof follows by a counting argument. Also, it is clear from the formulae (4.3.4) and (4.4.2) that

$$\mathfrak{h}(p) \leq \mathfrak{C}(p), \tag{4.4.3}$$

for all d.

We now illustrate the formula with examples. For $d=1$ with $p_1 = k$ (an AR(k) specification) we have

$$\hat{\Psi}^{(1)}(k) = \begin{pmatrix} \gamma(0) & \gamma(1) & \ldots & \gamma(k) \\ \gamma(1) & \gamma(0) & \ldots & \gamma(k-2) \\ \vdots & \vdots & \vdots & \vdots \\ \gamma(k) & \gamma(k-2) & \ldots & \gamma(0) \end{pmatrix},$$

and the formula (4.4.2) delivers a bound that holds with equality.

For $d=2$ the formula indicates a maximum of

$$1 + 2^0(p_1+p_2) + 2^1 p_1 p_2 = 1 + p_1 + p_2 + 2p_1p_2 \tag{4.4.4}$$

unique covariances, delivering bounds of 8 and 13 for the AR$(0,1;1,1)$ and AR$(0,2;1,1)$ models respectively, while for $d=3$ there are at most

$$1 + 2^0(p_1+p_2+p_3) + 2^1(p_1p_2 + p_1p_3 + p_2p_3) + 2^2 p_1p_2p_3 \tag{4.4.5}$$

unique covariances. If equal truncation lengths are chosen in each dimension, so that $p_{U_i} = p_{L_i} = p$ for each $i = 1, \ldots, d$, we have $p_1 = p$ and $p_i = 2p$ for $i = 2, \ldots, d$. Then the formulae (4.4.4) and (4.4.5) become

$$1 + 2^0 \times 3p + 2^2 p^2 = 1 + 3p + 4p^2 \tag{4.4.6}$$

and

$$1 + 2^0 \times 5p + 2^1 \times 10p^2 + 2^2 \times 4p^3 = 1 + 5p + 20p^2 + 16p^3 \tag{4.4.7}$$

respectively.

4.A Proof of Lemma 4.1

Proof. First note that

$$S_{MN} = \sum_{t(|M|,N)}{}'' \sum_{s^1 \in \mathbb{Z}^d} \cdots \sum_{s^q \in \mathbb{Z}^d} \xi_{\mathbf{st}},$$

which may be rewritten as

$$S_{MN} = \sum_{s^1 \in \mathbb{Z}^d} \cdots \sum_{s^q \in \mathbb{Z}^d} \eta_{1s^1}^{1-1/w} \eta_{1s^1}^{1/w} \sum_{t(|M|,N)}{}'' (\xi_{\mathbf{st}}/\eta_{1s^1})$$

whence from Hölder's inequality

$$|S_{MN}|^w \leq \left(\sum_{s \in \mathbb{Z}^d} \eta_{1s} \right)^{w-1} \sum_{s^1 \in \mathbb{Z}^d} \eta_{1s^1}^{1-w} \left| \sum_{s^2 \in \mathbb{Z}^d} \cdots \sum_{s^q \in \mathbb{Z}^d} \sum_{t(|M|,N)}{}'' \xi_{\mathbf{st}} \right|^w .$$

Similarly $\left| \sum_{s^2 \in \mathbb{Z}^d} \cdots \sum_{s^q \in \mathbb{Z}^d} \sum_{t(|M|,N)}{}'' \xi_{\mathbf{st}} \right|^w$ is bounded by

$$\left(\sum_{s \in \mathbb{Z}^d} \eta_{2s} \right)^{w-1} \sum_{s^2 \in \mathbb{Z}^d} \eta_{2s^2}^{1-w} \left| \sum_{s^3 \in \mathbb{Z}^d} \cdots \sum_{s^q \in \mathbb{Z}^d} \sum_{t(|M|,N)}{}'' \xi_{\mathbf{st}} \right|^w .$$

After q applications of Hölder's inequality and using (4.2.2) we obtain

$$|S_{MN}|^w \leq C \sum_{s^1 \in \mathbb{Z}^d} \cdots \sum_{s^q \in \mathbb{Z}^d} \eta_{\mathbf{s}}^{1-w} \left| \sum_{t(|M|,N)}{}'' \xi_{\mathbf{st}} \right|^w . \tag{4.A.1}$$

Also, from von Bahr and Esseen (1965) and (4.2.1)

$$\begin{aligned} E \left| \sum_{t(|M|,N)}{}'' \xi_{\mathbf{st}} \right|^w &\leq C \sum_{t(|M|,N)}{}'' |\xi_{\mathbf{st}}|^w \\ &\leq C \, \eta_{\mathbf{s}}^w \sum_{t(|M|,N)}{}'' a_t^w . \end{aligned}$$

Taking expectations of (4.A.1) and applying the above and (4.2.2) we conclude

$$\begin{aligned} E\,|S_{MN}|^w &\leq C \sum_{s^1 \in \mathbb{Z}^d} \cdots \sum_{s^q \in \mathbb{Z}^d} \eta_{\mathbf{s}} \sum_{t(|M|,N)}{}'' a_t^w \\ &\leq C \sum_{t(|M|,N)}{}'' a_t^w \\ &= C \, b_{wMN}, \end{aligned}$$

establishing the lemma. $\square$

5 Consistent autoregressive spectral density estimation for stationary lattice processes

5.1 Introduction

This chapter extends the consistency result of Berk (1974) to spatial processes on a lattice. The approach involves fitting an autoregression with the autoregressive order permitted to diverge with sample size. A key ingredient of the proof involves the property that, for a stationary time series x_t with an AR(k) representation

$$x_t = \sum_{j=1}^{k} a_j x_{t-j} + \epsilon_t, \tag{5.1.1}$$

the covariance matrix is Toeplitz. This property allows autoregressive order k to grow with sample size N while satisfying $k^3 = o(N)$. Failure to utilise the Toeplitz property will entail a requirement of $k^4 = o(N)$.

Because the Toeplitz property is so vital to obtain the sharpest possible bounds, the analysis in Chapter 4 plays a crucial role in this chapter. The nested-block Toeplitz structure of the covariance matrix derived therein allows us to derive bounds analogous to the time series case, albeit with one key difference.

In Chapter 1, we discussed the literature on autoregressive spectral density estimation for time series and lattice processes. Some advantages of the autoregressive approach were highlighted in Parzen (1969), pertaining to cross-spectrum estimation multiple time series. These advantages are also relevant here and we enumerate them below.

1. We avoid a debate about the choice of window for smoothed periodogram estimates. How
2. The truncation point can be chosen on the basis that the time series passes a goodness of fit test.
3. If the time series obeys a finite autoregressive scheme (truncation point chosen as above) the the autoregressive estimate has a much smaller bias than the smoothed periodogram estimate.
4. Autoregressive estimates are easily updated for additional observations.

A criticism of the first point is that the burden of choosing the autoregressive order lies on the the practitioner, and this may not be too different from choosing a window.

It has been noted that that autoregressive spectral estimation is also better at estimating peaked spectra, as compared to weighted/windowed periodogram estimation, see e.g. Ensor and Newton (1988). For lattice processes another advantage presents itself, connected with the edge-effect. This effect indicates a bias in covariance estimates that matters when $d = 2$ and worsens with increasing d. While this effect is negligible when $d = 1$, this is not the case when $d > 1$. Although we discuss this in detail in Section 5.3, we state some key points here. Guyon (1982) suggested an incorrectly centred version of the covariance estimates which eliminates the bias (asymptotically), but his device was criticised by Dahlhaus and Künsch (1987) as it could give rise to possible negative spectral density estimates when using kernel based spectral density estimation. The latter suggested tapering the covariance estimates, but introduced ambiguity arising from the choice of an appropriate taper. Robinson and Vidal Sanz (2006) suggest another approach, but again there is an element of ambiguity due to the practitioner having to make a choice of a function.

On the other hand, autoregressive spectral estimation delivers a guaranteed non-negative estimate even when using the device of Guyon (1982) to correct for the edge-effect. This eliminates the need to choose a taper, but the practitioner still has to choose the autoregressive lag-order. This can be achieved by generalisations of various time series information criteria, see e.g. Tjøstheim (1981) for a generalisation of the Final Prediction Error (FPE) criterion of Akaike (1970) and the Bayesian Information Criterion (BIC) of Schwarz (1978).

As was observed first by Whittle (1954), estimation of the parameters of multilateral autoregressive processes by least squares leads to inconsistency. This is due to the presence in the likelihood function of a Jacobean term which depends on the parameters to be estimated and therefore may not be ignored. A representation on a 'half-plane' will allow us to use least squares estimation. From Helson and Lowdenslager (1958, 1961) we know that a stationary process x_t has a moving average representation on a half-plane as long as the log of the spectral density is integrable (i.e. the process is purely non-deterministic in the linear prediction sense). Under conditions that allow us to invert this moving average representation to obtain an autoregressive representation on a half-space, we may truncate the order of the autoregression in each dimension and investigate how fast the parameter space may increase relative to sample size while still yielding consistent estimates for the autoregressive parameters. These consistent estimates can then be used to construct a consistent estimate of the spectral density.

Section 5.2 provides the basic setup and assumptions of the problem. Section 5.3 provides a sequence of lemmas related to the covariances and covariance estimates used in this chapter, and proposes an estimate for the autoregression coefficients. Since we introduce covariance estimates in this section it is also natural to include a small discussion of the edge effect. Section 5.4 contains theorems recording conditions for consistency of autoregression coefficient estimates, and introduces the proposed spectral density estimate and provides conditions under which it is uniformly consistent.

5.2 Truncated approximation of unilateral autoregressive processes

Let t be a multiple index $(t_1, \ldots, t_d)$ with $t_j \in \mathbb{Z}$ where $\mathbb{Z}$ is the set of integers. Consider a stationary zero-mean random field $\left\{x_t : t \in \mathbb{Z}^d\right\}$ with spectral density $f(\lambda)$, $\lambda \in \Pi^d$, $\Pi = (-\pi, \pi]$. Suppose that the x_t are observed on the rectangular lattice $\mathfrak{L} = \{t : -n_{L_i} \le t_i \le n_{U_i}, i = 1, \ \ldots, d\}$, $n_{U_i}, n_{L_i} \ge 0$, $i = 1, \ldots, d$. Defining $n_i = n_{L_i} + n_{U_i} + 1$, the total number of observations are $\prod_{i=1}^d n_i$, which we denote N. Also denote $n = (n_1, n_2, \ldots, n_d)$.

As in Chapter 4 we define the half-space used in our representation as S_{1+}^{∞} consisting of those $t \in \mathbb{Z}^d$ satisfying $t_1 > 0$; $t_1 = 0, t_2 > 0$; $t_1 = t_2 = 0, t_3 > 0$; $\cdots$; $t_1 = \cdots = t_{d-1} = 0, t_d > 0$. We will also write $\mathbf{0}$ for the d-dimensional zero vector.

Suppose that the process $\{x_t : t \in \mathfrak{L}\}$ satisfies

$$x_t = \sum_{s \in S_{1+}^{\infty} \cup \mathbf{0}} b_s \epsilon_{t-s}, \quad \sum_{s \in S_{1+}^{\infty} \cup \mathbf{0}} |b_s| < \infty, \ b_{\mathbf{0}} \neq 0 \tag{5.2.1}$$

with ϵ_t a spatial white noise. This is a linear process-like condition, and as mentioned the existence of such a white noise is guaranteed by the log integrability of the spectral density. The spectral density can be written as

$$f(\lambda) = \frac{\sigma^2}{(2\pi)^d} \left| \sum_{s \in S_{1+}^{\infty} \cup \mathbf{0}} b_s e^{i\lambda' s} \right|^2, \quad \lambda \in \Pi^d. \tag{5.2.2}$$

Assuming that $\left| \sum_{s \in S_{1+}^{\infty} \cup \mathbf{0}} b_s e^{i\lambda' s} \right|$ is bounded and bounded away from zero, this process is invertible and admits the AR representation

$$x_t = \sum_{s \in S_{1+}^{\infty}} d_s x_{t-s} + \epsilon_t, \quad \sum_{s \in S_{1+}^{\infty}} |d_s| < \infty. \tag{5.2.3}$$

We truncate in each dimension and, for $p_{L_i} \ge 0, p_{U_i} \ge 0$, $i = 1, \ldots, d$, concentrate on an AR $(p_{L_1}, p_{U_1}; \ldots; p_{L_d}, p_{U_d})$. Each p_{L_i}, p_{U_i}, $i = 1, \ldots, d$ is treated as a function of N and our asymptotic theory consists of finding functions $p_{L_i} = p_{L_i}(N), p_{U_i} = p_{U_i}(N)$, $i = 1, \ldots, d$ such that we can consistently approximate the infinite representation with the truncated one. For notational convenience, explicit reference to the dependence of the orders on N is suppressed as is the dependence of the total parameter space on the sample size. Consider the following truncated approximation to (5.2.3)

$$x_t = \sum_{s \in S[-p_L, p_U]} d_s x_{t-s} + \epsilon_t, \quad t \in \mathbb{Z}^d, \tag{5.2.4}$$

where $S[-p_L, p_U]$ was defined in (4.3.2). As noted in Chapter 4, in view of the half-space representation we can *a priori* set, say, $p_{L_1} = 0$. We approximate the true model (5.2.3) by the truncated model (5.2.4).

Denote $p_i = p_{L_i} + p_{U_i}$, $i = 1, \ldots, d$, with $p_1 \equiv p_{U_1}$ since $p_{L_1} = 0$ by our definition of half-plane, and also write $p = (p_{L_2}, \ldots, p_{L_d}, p_{U_1}, \ldots, p_{U_d})$. Again, let $\mathfrak{h}(p)$ denote the total number of AR parameters to be estimated. We should mention that the practitioner may prefer to choose only one truncation length for each dimension. In this case $p_{L_i} = p_{U_i} = p$, $i = 1, \ldots, d$, and the formula (4.3.4) indicates that $\mathfrak{h}(p) = p\,(1 + (d-1)(2p+1))$.

For $z = (z_1, \ldots, z_d)$ with complex-valued elements and $s = (s_1, \ldots, s_d)$ with integer-valued elements, introduce the notation $z^s = \prod_{j=1}^d z_j^{s_j}$. Define the multidimensional polynomials

$$\begin{aligned} B(z) &= \sum_{s \in S_{1+}^{\infty} \cup \mathbf{0}} b_s z^s \\ D(z) &= 1 - \sum_{s \in S_{1+}^{\infty}} d_s z^s. \end{aligned}$$

These are called polynomials even though they may involve negative powers. Strictly speaking, they are rational functions unless we have a quarter-plane representation for x_t (see Rosenblatt (1985), p. 228), which ensures that all the entries of s are non-negative. However, we follow the precedent of Robinson and Vidal Sanz (2006) in using this terminology. We now introduce the following assumptions

Assumption 20. $\{x_t : t \in \mathfrak{L}\}$ is a weakly stationary random field with spectral density $f(\lambda)$ satisfying

$$\int_{\Pi} \log f(\lambda) d\lambda > -\infty.$$

Assumption 20 guarantees that the representation (5.2.1) holds, see e.g. Helson and Lowdenslager (1958), Korezlioglu and Loubaton (1986). This is simply a generalisation of the result that every stationary time series that is purely non-deterministic in the linear prediction sense (captured by the log of the spectral density being integrable) has an infinite moving-average representation. We now proceed on the basis of this representation.

Assumption 21. The ϵ_t are i.i.d. with mean zero and variance σ^2 and, for some $v \in (1, 2]$, $E\left|\epsilon_t\right|^{2v} \leq C$ for all $t \in \mathfrak{L}$.

Again martingale assumptions can replace the i.i.d. imposition, but we choose to avoid these. Expressing the moment condition in terms of the number v delivers conditions restricting the rate of growth of the parameter space relative to sample size that become more stringent with $v \to 1$.

Assumption 22. $\sum_{s \in S_{1+}^{\infty}} |d_s| < \infty$.

Assumption 23. $D(z) \neq 0$ for $|z_i| = 1$, $i = 1, \ldots, d$.

Assumption 24. $B(z)$ is bounded away from zero for $|z_i| = 1$, $i = 1, \ldots, d$.

By Wiener's Lemma (see e.g. Rudin (1973) p. 266), Assumptions 22 and 23 imply that $\sum_{s \in S_{1+}^{\infty} \cup \mathbf{0}} |b_s| < \infty$. Together with Assumption 24 this implies that the spectral density $f(\lambda)$ is bounded and bounded away from 0 i.e. there exist real numbers m, M satisfying $0 < m \leq M < \infty$, such that

$$m \leq f(\lambda) \leq M. \tag{5.2.5}$$

This indicates that these assumptions in fact imply a regularity condition on the spectral density. Wiener's Lemma is a generalisation to d dimensions of the original Lemma IIe given on Wiener (1932), p. 14 and used for the proof of the celebrated Tauberian Theorem of Wiener. A discussion of the Tauberian theorem in a general setting may be found in Rudin (1962), Ch. 7.

We can also regard the d_s as coefficients in the Laurent series of the holomorphic function $D(z)$ about $\mathbf{0}$. Minakshisundaram and Szász (1947) show that Assumption 22 may be replaced by a Hölder condition on $D(z)$ of order τ with $\tau > d/2$. Sufficient conditions on the modulus of continuity are available in Konovalov (1979) and Golubov (1985) which imply the condition given in Minakshisundaram and Szász (1947).

5.3 Preliminary results on covariances and covariance estimates

We state in this section some lemmas that will be needed for the proofs in the next section. Many of these are generalisations to spatial processes of the lemmas in the section titled 'Six Lemmas' in Robinson (1979), except that they are proved here in an autoregression context as opposed to a regression context as in the original paper.

Lemma 5.1. *Suppose* $\sum_{s\in S_{1+}^{\infty}\cup\mathbf{0}}|b_s|<\infty$. *Then*

$$\sum_{k\in\mathbb{Z}^d}|\gamma(k)|<\infty.$$

The following lemma is simply a particular case of Lemma 4.1.

Lemma 5.2. *For such* n_i *and* k_i *that satisfy* $n_i>|k_i|$ *for* $i=1,\dots,d$, *let*

$$S_{kn}=\frac{1}{\prod_{i=1}^{d}(n_i-|k_i|)}\sum_{t(|k|,n)}{}^{\prime\prime}u_t,\ u_t=\sum_{r\in\mathbb{Z}^d}\sum_{s\in\mathbb{Z}^d}\xi_{rs,t},\ t\in\mathfrak{L}, \tag{5.3.1}$$

with the $\xi_{rs,t}$ *satisfying Assumption 18. For some* $w'\in(1,2]$, *suppose there exist* $\eta_{1,r},\eta_{2,r}$, $r\in\mathbb{Z}^d$, *such that*

$$E\,|\xi_{rs,t}|^{w'}\le|\eta_{1,r}\eta_{2,s}|^{w'},\sum_{r\in\mathbb{Z}^d}|\eta_{j,r}|<\infty,\ j=1,2, \tag{5.3.2}$$

for all $r,s\in\mathbb{Z}^d$ *and* $t\in\mathfrak{L}$. *Then*

$$E\,|S_{kn}|^{w'}\le K\left(\prod_{i=1}^{d}(n_i-|k_i|)\right)^{1-w'}. \tag{5.3.3}$$

Lemma 5.3. *Suppose that* $A=[a_{ij}]_{i,j=1,\dots,n}$ *is an* $n\times n$ *matrix and* $\tilde{A}=[\tilde{a}_{ij}]_{i,j=1,\dots,n+k}$ *is an* $(n+k)\times(n+k)$ *matrix formed by adding* k *additional rows and columns to* A. *Then*

$$\|A\|_R\le\left\|\tilde{A}\right\|_R. \tag{5.3.4}$$

In view of the stationarity of x_t, define the autocovariances as

$$\gamma(k)=Ex_tx_{t+k},\ k\in\mathbb{Z}^d,$$

and introduce the covariance estimates

$$\hat{\gamma}(k)=\frac{1}{\prod_{i=1}^{d}(n_i-|k_i|)}\sum_{t(|k|,n)}{}^{\prime\prime}x_tx_{t+k},$$

where it is assumed that $n_i>|k_i|\ge 0$ for $i=1,\dots,d$ and the sum $\sum_{t(|k|,n)}''$ is defined analogously to Section 4.2 with respect to n and k.

The estimates $\hat{\gamma}(k)$ incorporate the device for edge-effect correction suggested by Guyon (1982). Consider instead the estimates

$$\tilde{\gamma}(k)=\frac{1}{N}\sum_{t(|k|,n)}{}^{\prime\prime}x_tx_{t+k}.$$

Then for fixed k, as the $n_i \to \infty$, the bias of $\tilde{\gamma}(k)$ for $\gamma(k)$ is of order $\sum_{i=1}^{d} \frac{1}{n_i}$. The inequality between arithmetic and geometric means indicates that

$$\sum_{i=1}^{d} \frac{1}{n_i} \geq dn^{-\frac{1}{d}}$$

with equality implying that the n_i all increase at the same, $n^{\frac{1}{d}}$, rate. This inequality implies that the bias of $\tilde{\gamma}(k)$ is of order no less than $n^{-\frac{1}{d}}$. It is clear that this worsens with increasing d, but for $d = 1$ gives the usual 'parametric' rate of bias. Guyon (1982) suggested the use of

$$\left\{ \frac{N}{\prod_{i=1}^{d} (n_i - |k_i|)} \right\} \tilde{\gamma}(k)$$

to rectify this problem, and the last displayed expression is exactly what we define as $\hat{\gamma}(k)$.

Denote by $\hat{\psi}^{(d)}(p)$ ($\hat{\Psi}^{(d)}(p)$) the $\mathfrak{h}(p) \times 1$ vector ($\mathfrak{h}(p) \times \mathfrak{h}(p)$ matrix) constructed in exactly the same way as $\psi^{(d)}(p)$ ($\Psi^{(d)}(p)$) but using $\hat{\gamma}(k)$ in place of $\gamma(k)$. Also denote by $d(p)$ the $\mathfrak{h}(p) \times 1$ vector formed in exactly the same way as $\psi^{(d)}(p)$ but using d_k in instead of $\gamma(k)$. By construction the elements of $d(p)$ are d_s, $s \in S[-p_L, p_U]$. We then identify

$$d(p) = \Psi^{(d)}(p)^{-1} \psi^{(d)}(p), \tag{5.3.5}$$

assuming that (5.2.3) is the true model.

For n_i and p_i satisfying $n_i > p_i$, define the least squares estimate of $d(p)$ by the $\mathfrak{h}(p) \times 1$ vector

$$\hat{d}(p) = \hat{\Psi}^{(d)}(p)^{-1} \hat{\psi}^{(d)}(p). \tag{5.3.6}$$

Also write

$$\Delta(p) = \hat{\Psi}^{(d)}(p) - \Psi^{(d)}(p),$$

and

$$\delta(p) = \hat{\psi}^{(d)}(p) - \psi^{(d)}(p).$$

The lemmas that follow provide orders of magnitude related to moments of the difference between covariance estimates and true covariances.

Lemma 5.4. *Under Assumptions 20-23,*

$$E \left| \hat{\gamma}(k) - \gamma(k) \right|^{v} \leq C \left(\prod_{i=1}^{d} (n_i - |k_i|) \right)^{1-v}. \tag{5.3.7}$$

Lemma 5.5. *Under Assumptions 20-23,*

$$E\,\|\delta(p)\|^{v} \le C\;\mathfrak{h}(p)^{v}\left(\prod_{i=1}^{d}(n_i - p_i)\right)^{1-v},$$

where $\mathfrak{h}(p)$ is defined as in (4.3.4).

Lemma 5.6. *Under Assumptions 20-23,*

$$E\,\|\Delta(p)\|^{v} \le C\;\mathfrak{C}(p)^{v}\prod_{i=1}^{d}(n_i - p_i)^{1-v},$$

where $\mathfrak{C}(p)$ is defined as in Proposition 4.1.

We are now in a position to state a lemma and a corollary that will, in view of (5.3.5), allow us to identify the true autoregressive parameter $d(p)$.

Lemma 5.7. *Let ρ be any eigenvalue of $\Psi^{(d)}(p)$. Then, under Assumptions 20 and 22-24,*

$$(2\pi)^{d}m \le \rho \le (2\pi)^{d}M.$$

We note that this lemma is a generalization of the statement on Grenander and Szegö (1984), p. 64.

Corollary 5.8. *Under Assumptions 20 and 22-24,*

$$\left\|\Psi^{(d)}(p)^{-1}\right\| \le C.$$

5.4 Uniform consistency of $\hat{f}_p(\lambda)$

The first theorem in this section establishes conditions under which $\hat{d}(p)$ is a consistent estimate of the true autoregressive parameters $d(p)$.

Theorem 5.1. *Let Assumptions 20-24 hold and also assume that $\mathfrak{C}(p)$ is chosen as a function of N such that*

(i) $$\frac{1}{\mathfrak{C}(p)} + \frac{\mathfrak{C}(p)}{N^{\frac{v-1}{v}}} \to 0 \text{ as } N \to \infty,$$

and

(ii) $$\sum_{t \in S_{1+}^{\infty} \setminus S[-p_L, p_U]} |d_t| \to 0 \text{ as } N \to \infty.$$

Then

$$\left\|\hat{d}(p) - d(p)\right\| \xrightarrow{p} 0.$$

Condition (ii) says that the dependence from 'distant' lags must decline sufficiently fast. It is not restrictive in view of Assumption 22. In fact, from the Cauchy Convergence Criterion of real analysis a series $\sum_{n=1}^{\infty} a_n$ of positive terms (with scalar subscript n) converges if and only if

$$\lim_{r,n\to\infty} \sum_{j=r}^{n} a_j \to 0.$$

This is just the version of condition (ii) for $d = 1$. For multiple-series (i.e. series indexed by vectors), such results do not seem to be available but it seems that an extension may be rather natural.

It is important to note that this extension to $d > 1$ differs from the case $d = 1$ in one important sense. In the case $d = 1$ condition (i) applies to the dimension of the parameter space, because this dimension equals the number of unique covariances in $\psi^{(d)}(p)$. Now this is clearly not the case due to (4.4.3).

We now prove the consistency of the estimate of the error variance based on the least squares estimate considered above, under the same conditions as in Theorem 5.1. Define the error variance estimate as

$$\hat{\sigma}^2(p) = \frac{1}{\prod_{i=1}^{d}(n_i - |k_i|)} \sum_{t(|k|,n)}{}'' \left(x_t - \sum_{s\in S[-p_L,p_U]} \hat{d}_s(p) x_{t-s} \right)^2.$$

Theorem 5.2. *Let Assumptions 20-24 hold and also assume that* $\mathfrak{C}(p)$ *is chosen as a function of* N *such that*

(i) $$\frac{1}{\mathfrak{C}(p)} + \frac{\mathfrak{C}(p)}{N^{\frac{v-1}{v}}} \to 0 \text{ as } N \to \infty,$$

and

(ii) $$\sum_{t\in S_{1+}^{\infty}\backslash S[-p_L,p_U]} |d_t| \to 0 \text{ as } N \to \infty.$$

Then

$$\hat{\sigma}^2(p) \xrightarrow{p} \sigma^2$$

We now introduce spectral density estimates. First, for $\lambda \in \Pi$, the spectral density of x_t under (5.2.4) is given by

$$f(\lambda) = \frac{\sigma^2}{(2\pi)^d \left| 1 - \sum_{s\in S_{1+}^{\infty}} d_s e^{is'\lambda} \right|^2},$$

and we estimate this using

$$\hat{f}_p(\lambda) = \frac{\hat{\sigma}^2(p)}{(2\pi)^d \left|1 - \sum_{s \in S[-p_L, p_U]} \hat{d}_s(p) e^{is'\lambda}\right|^2}.$$

Theorem 5.3. *Let Assumptions 20-24 hold and also assume that $\mathfrak{C}(p)$ and $\mathfrak{h}(p)$ are chosen as functions of N such that*

(i) $$\frac{1}{\mathfrak{C}(p)} + \frac{1}{\mathfrak{h}(p)} + \frac{\mathfrak{C}(p)\mathfrak{h}(p)^{\frac{1}{2}}}{N^{\frac{v-1}{v}}} \to 0 \text{ as } N \to \infty,$$

and

(ii) $$\mathfrak{h}(p)^{\frac{1}{2}} \sum_{t \in S_{1+}^{\infty} \setminus S[-p_L, p_U]} |d_t| \to 0 \text{ as } N \to \infty.$$

Then

$$\sup_{\lambda \in \Pi} \left|\hat{f}_p(\lambda) - f(\lambda)\right| \xrightarrow{p} 0.$$

The conditions we imposed for this theorem were stronger than those for earlier results in two ways. First, the condition restricting the rate of growth of the parameter space relative to sample size is stronger than the one imposed for Theorems 5.1 and 5.2. For example, if $v = 2$ then condition (i) in those theorems required $\mathfrak{C}(p)/N^{1/2} \to 0$ whereas condition (i) in Theorem 5.3 requires $\mathfrak{C}(p)\mathfrak{h}(p)^{1/2}/N^{1/2} \to 0$. Note that for $d = 1$ the latter reduces to the condition established by Berk (1974), which is, in fact, a particular case of the condition in Robinson (1979). The second aspect of difference is the requirement in condition (ii) that the dependence on 'distant' lags decline sufficiently fast to overcome norming by $\mathfrak{h}(p)^{\frac{1}{2}}$.

An aspect in which Theorem 5.3 differs from the consistency result in Berk (1974) is that it is a uniform consistency result. Uniform consistency is possible in the time series case under the same conditions as Theorem 1 of Berk (1974), although this is not stated in that paper. Bhansali (1980) records the uniform consistency under identical conditions in his Theorem 3.1.

5.A Proofs of theorems

Proof of Theorem 5.1: We have

$$\begin{aligned}\hat{d}(p) - d(p) &= \hat{\Psi}^{(d)}(p)^{-1}\left(\hat{\psi}^{(d)}(p) - \hat{\Psi}^{(d)}(p)d(p)\right)\\ &= \hat{\Psi}^{(d)}(p)^{-1}\left(\delta(p) - \Delta(p)d(p) + \psi^{(d)}(p) - \Psi^{(d)}(p)d(p)\right)\end{aligned}$$

so that the norm of the LHS above is bounded by

$$\left\|\hat{\Psi}^{(d)}(p)^{-1}\right\|\left(\|\delta(p)\| + \|\Delta(p)\|\,\|d(p)\| + \left\|\Psi^{(d)}(p)d(p) - \psi^{(d)}(p)\right\|\right). \tag{5.A.1}$$

Now

$$\begin{aligned}\left\|\hat{\Psi}^{(d)}(p)^{-1}\right\| &\leq \left\|\hat{\Psi}^{(d)}(p)^{-1} - \Psi^{-1}\right\| + \left\|\Psi^{(d)}(p)^{-1}\right\|\\ &\leq \left(\left\|\hat{\Psi}^{(d)}(p)^{-1}\right\|\,\|\Delta(p)\| + 1\right)\left\|\Psi^{(d)}(p)^{-1}\right\|,\end{aligned}$$

so

$$\left\|\hat{\Psi}^{(d)}(p)^{-1}\right\|\left(1 - \left\|\Psi^{(d)}(p)^{-1}\right\|\,\|\Delta(p)\|\right) \leq \left\|\Psi^{(d)}(p)^{-1}\right\|.$$

Using Markov's inequality and Lemma 5.6 it follows that $\|\Delta(p)\| \xrightarrow{p} 0$ if

$$\mathfrak{C}(p)^v\left(\prod_{i=1}^{d}(n_i - p_i)\right)^{1-v} \to 0, \text{ i.e., } \mathfrak{C}(p)^v N^{1-v}\left(\prod_{i=1}^{d}\left(1 - \frac{p_i}{n_i}\right)\right)^{1-v} \to 0,$$

which is true by (i). Thus from Corollary 5.8

$$\operatorname{plim}_{N,p_{L_i},p_{U_i}\to\infty,i=1,\ldots,d}\left\|\hat{\Psi}^{(d)}(p)^{-1}\right\| \leq \lim_{N,p_{L_i},p_{U_i}\to\infty,i=1,\ldots,d}\left\|\Psi^{(d)}(p)^{-1}\right\| < \infty.$$

Now we deal with the factor in parentheses in (5.A.1). By Lemma 5.5, Markov's inequality and (i), $\|\delta(p)\| \xrightarrow{p} 0$.

For the second term, we have $\|\Delta(p)\| \xrightarrow{p} 0$ and also

$$\|d(p)\| = \left(\sum_{s\in S[-p_L,p_U]} d_s^2\right)^{\frac{1}{2}} \leq \sum_{s\in S[-p_L,p_U]} |d_s| \leq \sum_{s\in S_{1+}^{\infty}} |d_s| < \infty,$$

by Lemma 5.1. Thus the second term converges to zero in probability. Finally, for the third term note that by (5.2.1) implies that

$$E\epsilon_t x_{t-k} = \sum_{s\in S_{1+}^{\infty}\cup\mathbf{0}} b_s E\epsilon_t\epsilon_{t-k-s} = 0,\ k \in S_{1+}^{\infty},\ t \in \mathfrak{L},$$

because $k+s=0$ is not possible due to our definition of half-plane (4.3.1). This indicates that

$$\gamma(k) = Ex_t x_{t-k} = \sum_{t \in S_{1+}^{\infty}} d_t \gamma(t-k),\ k \in S_{1+}^{\infty},$$

so $\left\|\Psi^{(d)}(p)d(p) - \psi^{(d)}(p)\right\|^2$ is

$$\begin{aligned}
& \sum_{s \in S[-p_L,p_U]} \left(\sum_{t \in S[-p_L,p_U]} d_s \gamma(t-s) - \gamma(s) \right)^2 \\
= & \sum_{s \in S[-p_L,p_U]} \left(\sum_{t \in S[-p_L,p_U]} d_s \gamma(t-s) - \sum_{t \in S_{1+}^{\infty}} d_t \gamma(t-s) \right)^2 \\
= & \sum_{s \in S[-p_L,p_U]} \left(\sum_{t \in S_{1+}^{\infty} \setminus S[-p_L,p_U]} d_t \gamma(t-s) \right)^2 \\
= & \sum_{s \in S[-p_L,p_U]} \left(\sum_{t \in S_{1+}^{\infty} \setminus S[-p_L,p_U]} d_t^2 \right) \left(\sum_{t \in S_{1+}^{\infty} \setminus S[-p_L,p_U]} \gamma(t-s)^2 \right) \\
= & \left(\sum_{s \in S[-p_L,p_U]} \sum_{t \in S_{1+}^{\infty} \setminus S[-p_L,p_U]} \gamma(t-s)^2 \right) \left(\sum_{t \in S_{1+}^{\infty} \setminus S[-p_L,p_U]} d_t^2 \right) \\
\leq & \ C \sum_{s \in \mathbb{Z}^d} \gamma(s)^2 \sum_{t \in S_{1+}^{\infty} \setminus S[-p_L,p_U]} d_t^2 \\
= & \ C \sum_{t \in S_{1+}^{\infty} \setminus S[-p_L,p_U]} d_t^2,
\end{aligned}$$

using Lemma 5.1.

Thus

$$\left\|\Psi^{(d)}(p)d(p) - \psi^{(d)}(p)\right\| \leq C \left(\sum_{t \in S_{1+}^{\infty} \setminus S[-p_L,p_U]} d_t^2 \right)^{\frac{1}{2}} \leq C \sum_{t \in S_{1+}^{\infty} \setminus S[-p_L,p_U]} |d_t|,$$

which converges to zero as $N \to \infty$ due to (ii), completing the proof. Note that we have also shown that

$$\left\|\hat{d}(p) - d(p)\right\| = \mathscr{O}_p\left(\frac{\mathfrak{C}(p)}{N^{\frac{v-1}{v}}}\right), \tag{5.A.2}$$

by Markov's inequality. □

Proof of Theorem 5.2: Write

$$\hat{\gamma}(\mathbf{0}) = \frac{1}{\prod_{i=1}^{d}(n_i - |k_i|)} \sum_{t(|k|,n)}^{\prime\prime} x_t^2.$$

Using standard algebraic manipulation and the definition of least squares we may write $\hat{\sigma}^2(p) - \sigma^2$ as

$$\begin{aligned}
&\frac{1}{\prod_{i=1}^{d}(n_i - |k_i|)} \sum_{t(|k|,n)}^{\prime\prime} \left(x_t - \sum_{s \in S[-p_L, p_U]} \hat{d}_s(p) x_{t-s} \right)^2 - \sigma^2 \\
= \;& \hat{\gamma}(\mathbf{0}) - \hat{d}(p)'\hat{\psi}^{(d)}(p) - \sigma^2 \\
= \;& \hat{\gamma}(\mathbf{0}) - \left(\hat{d}(p) - d(p)\right)' \hat{\psi}^{(d)}(p) - d(p)'\hat{\psi}^{(d)}(p) - \hat{\gamma}(\mathbf{0}) + \sum_{t \in S_{1+}^{\infty}} d_t \gamma(t) \\
= \;& \hat{\gamma}(\mathbf{0}) - \gamma(\mathbf{0}) - \left(\hat{d}(p) - d(p)\right)' \psi^{(d)}(p) - d(p)'\Delta(p) \\
& - \left(\hat{d}(p) - d(p)\right)' \Delta(p) - d(p)'\psi^{(d)}(p) + \sum_{t \in S_{1+}^{\infty}} d_t \gamma(t).
\end{aligned}$$

Since $d(p)'\psi^{(d)}(p) = \sum_{s \in S[-p_L, p_U]} d_s \gamma(s)$, we can write

$$\begin{aligned}
\hat{\sigma}^2(p) - \sigma^2 \;=\; & (\hat{\gamma}(\mathbf{0}) - \gamma(\mathbf{0})) - \left(\hat{d}(p) - d(p)\right)' \psi^{(d)}(p) - d(p)'\Delta(p) \\
& - \left(\hat{d}(p) - d(p)\right)' \Delta(p) + \sum_{t \in S_{1+}^{\infty} \backslash S[-p_L, p_U]} d_t \gamma(t).
\end{aligned}$$

The first term on the RHS converges to 0 in probability by Lemma 5.4 and Markov's inequality. The second $\xrightarrow{p} 0$ by Theorem 5.1 and Lemma 5.1. The third term $\xrightarrow{p} 0$ by Lemma 5.5, (i) and Assumption 22. The fourth term $\xrightarrow{p} 0$ by Theorem 5.1, Lemma 5.5 and (i). For the fifth term, convergence to zero follows by (ii) and Lemma 5.1. $\square$

Proof of Theorem 5.3: Write

$$D_p(\lambda) = 1 - \sum_{s \in S[-p_L, p_U]} d_s(p) e^{is'\lambda},$$

and

$$\hat{D}_p(\lambda) = 1 - \sum_{s \in S[-p_L, p_U]} \hat{d}_s(p) e^{is'\lambda}.$$

Then we have

$$\hat{f}_p(\lambda) - f(\lambda) = \frac{\sigma^2 \left(\left| \hat{D}_p(\lambda) \right|^2 - |D_p(\lambda)|^2 \right) - |D_p(\lambda)|^2 \left(\hat{\sigma}^2(p) - \sigma^2 \right)}{(2\pi)^d \, |D_p(\lambda)|^2 \left| \hat{D}_p(\lambda) \right|^2}. \tag{5.A.3}$$

Because $D_p(\lambda) = \frac{\sigma^2}{(2\pi)^d f(\lambda)}$, by (5.2.5) we have

$$c \leq D_p(\lambda) \leq C, \tag{5.A.4}$$

uniformly in $\lambda \in \Pi$.

On the other hand $\hat{D}_p(\lambda) = \frac{\hat{\sigma}^2}{(2\pi)^d \hat{f}_p(\lambda)}$, so that

$$\sup_{\lambda \in \Pi} \left| \hat{D}_p(\lambda) \right| \leq \sup_{\lambda \in \Pi} \left| \hat{D}_p(\lambda) - D_p(\lambda) \right| + \sup_{\lambda \in \Pi} |D_p(\lambda)| \tag{5.A.5}$$

and

$$\begin{aligned} \inf_{\lambda \in \Pi} \left| \hat{D}_p(\lambda) \right| &\geq \inf_{\lambda \in \Pi} |D_p(\lambda)| + \inf_{\lambda \in \Pi} \left\{ - \left| \hat{D}_p(\lambda) - D_p(\lambda) \right| \right\} \\ &= \inf_{\lambda \in \Pi} |D_p(\lambda)| - \sup_{\lambda \in \Pi} \left| \hat{D}_p(\lambda) - D_p(\lambda) \right|. \end{aligned} \tag{5.A.6}$$

We also have, using the Cauchy-Schwarz inequality,

$$\begin{aligned} \left| \hat{D}_p(\lambda) - D_p(\lambda) \right| &\leq \sum_{s \in S[-p_L, p_U]} \left| \hat{d}_s(p) - d_s \right| \left| e^{is'\lambda} \right| + \sum_{s \in S_{1+}^{\infty} \backslash S[-p_L, p_U]} |d_s| \left| e^{is'\lambda} \right| \\ &\leq \left(\sum_{s \in S[-p_L, p_U]} \left(\hat{d}_s(p) - d_s \right)^2 \right)^{\frac{1}{2}} \left(\sum_{s \in S[-p_L, p_U]} \left| e^{is'\lambda} \right|^2 \right)^{\frac{1}{2}} \\ &+ \sum_{s \in S_{1+}^{\infty} \backslash S[-p_L, p_U]} |d_s| \\ &\leq \left\| \hat{d}(p) - d(p) \right\| \left(\sum_{s \in S[-p_L, p_U]} 1 \right)^{\frac{1}{2}} + \sum_{s \in S_{1+}^{\infty} \backslash S[-p_L, p_U]} |d_s| \\ &= \mathfrak{h}(p)^{\frac{1}{2}} \left\| \hat{d}(p) - d(p) \right\| + \sum_{s \in S_{1+}^{\infty} \backslash S[-p_L, p_U]} |d_s|. \end{aligned} \tag{5.A.7}$$

Utilizing the stronger conditions (i) and (ii), we conclude from (5.A.2) that

$$\mathfrak{h}(p)^{\frac{1}{2}} \left\| \hat{d}(p) - d(p) \right\| = \mathscr{O}_p \left(\frac{\mathfrak{C}(p) \mathfrak{h}(p)^{\frac{1}{2}}}{N^{\frac{v-1}{v}}} \right),$$

implying that (5.A.7) is negligible. We have then shown that

$$\sup_{\lambda\in\Pi}\left|\hat{D}_p(\lambda)-D_p(\lambda)\right| \xrightarrow{p} 0. \tag{5.A.8}$$

Using (5.A.4), (5.A.5) and (5.A.6) together with (5.A.8) implies that

$$c \leq \hat{D}_p(\lambda) \leq C, \tag{5.A.9}$$

uniformly in $\lambda \in \Pi$, with probability approaching 1 as $n \to \infty$.
By the identity $a^2 - b^2 = (a-b)^2 + 2b(a-b)$, we obtain

$$\left|\left|\hat{D}_p(\lambda)\right|^2 - |D_p(\lambda)|^2\right| \leq \left(\hat{D}_p(\lambda)-D_p(\lambda)\right)^2 + 2\,|D_p(\lambda)|\left|\hat{D}_p(\lambda)-D_p(\lambda)\right|, \tag{5.A.10}$$

where the RHS converges to 0 in probability uniformly in λ by (5.A.8) and (5.A.9) so that

$$\sup_{\lambda\in\Pi}\left|\left|\hat{D}_p(\lambda)\right|^2 - |D_p(\lambda)|^2\right| \xrightarrow{p} 0. \tag{5.A.11}$$

Because (5.A.3) implies that

$$\left|\hat{f}_p(\lambda)-f(\lambda)\right| \leq \frac{\sigma^2\left|\left|\hat{D}_p(\lambda)\right|^2 - |D_p(\lambda)|^2\right| + |D_p(\lambda)|^2\left|\hat{\sigma}^2(p)-\sigma^2\right|}{(2\pi)^d\,|D_p(\lambda)|^2\left|\hat{D}_p(\lambda)\right|^2},$$

the theorem now follows by (5.A.4), (5.A.9), (5.A.11) and Theorem 5.2. □

5.B Proofs of lemmas

Proof of Lemma 5.1:

$$\begin{aligned}
\sum_{k\in\mathbb{Z}^d}|\gamma(k)| &\leq 2\sum_{k\in S_{1+}^{\infty}\cup\mathbf{0}}|\gamma(k)| \\
&= 2\sum_{k\in S_{1+}^{\infty}\cup\mathbf{0}}|Ex_t x_{t+k}| \\
&\leq 2\sigma^2\sum_{s\in S_{1+}^{\infty}\cup\mathbf{0}}|b_s|\sum_{k\in S_{1+}^{\infty}\cup\mathbf{0}}|b_{s+k}| < \infty
\end{aligned}$$

□

Proof of Lemma 5.2: The result follows from Lemma 4.1 taking $N = n$, $M = k$, $q = 2$ and $a_t = 1$ for all $t \in \mathfrak{L}$. □

Proof of Lemma 5.3: We may assume without loss of generality that the rows and columns have been added at the bottom and end of A respectively. Because $\|A\|_R =$

$\max_{i=1,\ldots,n}\sum_{j=1}^{n}|a_{ij}|$ and $\left\|\tilde{A}\right\|_R = \max_{i=1,\ldots,n+k}\sum_{j=1}^{n+k}|\tilde{a}_{ij}|$, we have

$$\begin{aligned}\|A\|_R &= \sum_{j=1}^{n}|a_{lj}|, \\ \left\|\tilde{A}\right\|_R &= \sum_{j=1}^{n+k}|\tilde{a}_{mj}|,\end{aligned}$$

for some $l = 1,\ldots,n$ and $m = 1,\ldots,n+k$. We proceed by contradiction. Suppose that

$$\sum_{j=1}^{n}|a_{lj}| > \sum_{j=1}^{n+k}|\tilde{a}_{mj}|,$$

adding $\sum_{j=n+1}^{n+k}|\tilde{a}_{lj}|$ to both sides of which yields

$$\sum_{j=1}^{n}|a_{lj}| + \sum_{j=n+1}^{n+k}|\tilde{a}_{lj}| > \sum_{j=1}^{n}|a_{mj}| + \sum_{j=n+1}^{n+k}|\tilde{a}_{lj}|. \tag{5.B.1}$$

The RHS of the above is greater than or equal to $\sum_{j=1}^{n+k}|\tilde{a}_{mj}|$. This indicates that

$$\sum_{j=1}^{n}|a_{lj}| + \sum_{j=n+1}^{n+k}|\tilde{a}_{lj}| > \left\|\tilde{A}\right\|_R,$$

which is a contradiction since $\left\|\tilde{A}\right\|_R$ is by definition the maximum absolute row sum of $\tilde{A}$. □

Proof of 5.4: For $\hat{\gamma}(k) - \gamma(k)$ to be of the form of S_{kn} in Lemma 5.2, we define $\xi_{rs,t} = b_r b_{r-k}\left(\epsilon_{t-r}^2 - \sigma^2\right),\ s = r-k;\ = b_r b_s \epsilon_{t-r}\epsilon_{t-k-s}, s \neq r-k$. Then the $\xi_{rs,t}$ are clearly zero-mean. They are independent because the ϵ_t are. Therefore, they satisfy Assumption 18.

By the c_r-inequality, Cauchy-Schwarz inequality and Assumption 21,

$$\begin{aligned}E\left|\xi_{rs,t}\right|^v &\leq 2\left|b_r b_{r-k}\right|^v\left(E\left|\epsilon_{t-r}\right|^{2v} + \sigma^{2v}\right) \leq C\ \left|b_r b_{r-k}\right|^v, s = r-k, \\ E\left|\xi_{rs,t}\right|^v &\leq \left|b_r b_s\right|^v\left(E\left|\epsilon_{t-r}\right|^{2v} E\left|\epsilon_{t-s}\right|^{2v}\right)^{\frac{1}{2}} \leq C\ \left|b_r b_s\right|^v, s \neq r-k,\end{aligned}$$

verifying that (5.3.2) holds since the b_r are absolutely summable. The result follows immediately from Lemma 5.2. □

Proof of Lemma 5.5:

$$
\begin{aligned}
E\left\|\delta(p)\right\|^{v} &\leq E\left\{\sum_{s\in S[-p_L,p_U]}\left|\hat{\gamma}(s)-\gamma(s)\right|\right\}^{v}\\
&\leq \mathfrak{h}(p)^{v-1}\sum_{s\in S[-p_L,p_U]}E\left|\hat{\gamma}(s)-\gamma(s)\right|^{v}\\
&\leq C\,\mathfrak{h}(p)^{v-1}\sum_{s\in S[-p_L,p_U]}\left(\prod_{i=1}^{d}\left(n_i-|s_i|\right)\right)^{1-v}\\
&\leq C\,\mathfrak{h}(p)^{v}\left(\prod_{i=1}^{d}\left(n_i-p_i\right)\right)^{1-v},
\end{aligned}
$$

using Hölder's inequality and Lemma 5.4. □

Proof of Lemma 5.6: Write

$$\check{\Delta}(p)=\check{\hat{\Psi}}^{(d)}-\check{\Psi}^{(d)},$$

where $\check{\hat{\Psi}}^{(d)}$ is constructed in the obvious way using estimated covariances. First, since $\Delta(p)$ is symmetric, $\|\Delta(p)\|$ is its greatest eigenvalue. Using Perron's theorem (Gradshteyn and Ryzhik (1994), p. 1155, Eq. 15.816), we have

$$
\begin{aligned}
\|\Delta(p)\|=\overline{\eta}(\Delta(p)) &\leq \|\Delta(p)\|_R\\
&\leq \left\|\check{\Delta}(p)\right\|_R, \qquad (5.B.2)
\end{aligned}
$$

by Lemma 5.3. We will now bound the absolute row-sums of $\check{\Delta}(p)$ uniformly over all rows. Consider a typical row of $\check{\Delta}(p)$. This consists of

$$\hat{\gamma}\left(l_1-\bar{l}_1,l_2-\bar{l}_2,\ldots,l_d-j_d\right)-\gamma\left(l_1-\bar{l}_1,l_2-\bar{l}_2,\ldots,l_d-j_d\right);\ j_d=0,\ldots,p_d,$$

for some $l_1,\ldots,l_d$, $l_i=0,\ldots,p_i$ and all $\bar{l}_1,\ldots,\bar{l}_{d-1}$, $\bar{l}_i=0,\ldots,p_i$. It follows that a typical absolute row sum is

$$\bar{\sum_{d-1}}\sum_{j_d=0}^{p_d}\left|\hat{\gamma}\left(l_1-\bar{l}_1,l_2-\bar{l}_2,\ldots,l_d-j_d\right)-\gamma\left(l_1-\bar{l}_1,l_2-\bar{l}_2,\ldots,l_d-j_d\right)\right| \qquad (5.B.3)$$

with $\bar{\sum}_{d-1}$ running over $\bar{l}_1,\ldots,\bar{l}_{d-1}$, $\bar{l}_i=0,\ldots,p_i$. Since the summands are absolute values of the elements of a row of a Toeplitz matrix (by construction), (5.B.3) is bounded by

$$2\bar{\sum_{d-1}}\sum_{k_d=-p_d}^{p_d}\left|\hat{\gamma}\left(l_1-\bar{l}_1,l_2-\bar{l}_2,\ldots,k_d\right)-\gamma\left(l_1-\bar{l}_1,l_2-\bar{l}_2,\ldots,k_d\right)\right|$$

which in turn is bounded by

$$2 \sum_{\text{unique covariances}} |\hat{\gamma}(k) - \gamma(k)|,$$

there being $\mathfrak{C}(p)$ terms in the sum by Proposition 4.1. This bound is clearly uniform over all possible rows. So using Hölder's inequality and Lemma 5.4

$$\begin{aligned} E\left\|\check{\Delta}(p)\right\|_R^v &\leq 4^v E\left\{\sum_{\text{unique covariances}} |\hat{\gamma}(k) - \gamma(k)|\right\}^v \\ &\leq 8\, \mathfrak{C}(p)^{1-v} \sum_{\text{unique covariances}} E\left|\hat{\gamma}(k) - \gamma(k)\right|^v \\ &\leq C\, \mathfrak{C}(p)^{1-v} \sum_{\text{unique covariances}} \prod_{i=1}^{d} (n_i - |k_i|)^{1-v} \\ &\leq C\, \mathfrak{C}(p)^{v} \prod_{i=1}^{d} (n_i - p_i)^{1-v}. \end{aligned}$$

Then the result follows from the above and (5.B.2). □

Proof of Lemma 5.7: Consider real numbers ξ_s, $s \in S\left[-p_L, p_U\right]$, $\sum_{s \in S[-p_L,p_U]} \xi_s^2 = 1$. The eigenvalues of $\Psi^{(d)}(p)$ are determined through the generalized Toeplitz form

$$T_N\left[\Psi^{(d)}(p)\right] = \sum_{j,k \in S[-p_L,p_U]} \xi_j \gamma(j-k) \xi_k,$$

the sum running over $j, k \in S\left[-p_L, p_U\right]$ by construction of $\Psi^{(d)}(p)$. Since $\gamma(j-k) = \int_\Pi e^{i(j-k)'\lambda} f(\lambda) d\lambda$, we have

$$\begin{aligned} T_N\left[\Psi^{(d)}(p)\right] &= \sum_{j,k \in S[-p_L,p_U]} \int_\Pi e^{i(j-k)'\lambda} f(\lambda) d\lambda\ \ \xi_j \xi_k \\ &= \sum_{j,k \in S[-p_L,p_U]} \int_\Pi e^{ij'\lambda} e^{-ik'\lambda} f(\lambda) d\lambda\ \ \xi_j \xi_k \\ &= \int_\Pi \sum_{j \in S[-p_L,p_U]} e^{ij'\lambda} \xi_j \sum_{k \in S[-p_L,p_U]} e^{-ik'\lambda} \xi_k\ \ f(\lambda) d\lambda \\ &= \int_\Pi \sum_{j \in S[-p_L,p_U]} e^{ij'\lambda} \xi_j\ \overline{\sum_{k \in S[-p_L,p_U]} e^{ik'\lambda} \xi_k}\ \ f(\lambda) d\lambda \\ &= \int_\Pi \left|\sum_{j \in S[-p_L,p_U]} e^{ij'\lambda} \xi_j\right|^2 f(\lambda) d\lambda \end{aligned}$$

$$
\begin{aligned}
&\in \left[m \int_{\Pi} \left| \sum_{j \in S[-p_L, p_U]} \xi_j \right|^2 \, d\lambda \, , \, M \int_{\Pi} \left| \sum_{j \in S[-p_L, p_U]} \xi_j \right|^2 \, d\lambda \right] \\
&= \left[m \int_{\Pi} \sum_{j \in S[-p_L, p_U]} \xi_j^2 \, d\lambda \, , \, M \int_{\Pi} \sum_{j \in S[-p_L, p_U]} \xi_j^2 \, d\lambda \right] \\
&= \left[(2\pi)^d m \, , \, (2\pi)^d M \right] .
\end{aligned}
$$

□

Proof of Corollary 5.8: If $\left\| \Psi^{(d)}(p)^{-1} \right\|$ exists, it is the reciprocal of the smallest eigenvalue, say μ, of $\Psi^{(d)}(p)$. Using Lemma 5.7 we get

$$
\left\| \Psi^{(d)}(p)^{-1} \right\| = \mu^{-1} \leq (2\pi)^{-d} m^{-1} \leq C.
$$

□

References

Akaike, H. (1969). Power spectrum estimation through autoregressive model fitting. *Annals of the Institute of Statistical Mathematics 21*(1), 407–419.

Akaike, H. (1970). Statistical predictor identification. *Annals of the Institute of Statistical Mathematics 22*(1), 203–217.

Anselin, L. (2010). Thirty years of spatial econometrics. *Papers in Regional Science 89*(1), 3–25.

Arbia, G. (2006). *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence*. Springer-Verlag, Berlin.

Berk, K. N. (1974). Consistent autoregressive spectral estimates. *The Annals of Statistics 2*(3), 489–502.

Bhansali, R. J. (1980). Autoregressive and window estimates of the inverse correlation function. *Biometrika 67*(3), 551–566.

Bhattacharjee, A. and C. Jensen-Butler (2013). Estimation of the spatial weights matrix under structural constraints. *Regional Science and Urban Economics 43*(4), 617–634.

Blommestein, H. J. (1985). Elimination of circular routes in spatial dynamic regression equations. *Regional Science and Urban Economics 15*(1), 121–130.

Cairoli, R. and J. B. Walsh (1975). Stochastic integrals in the plane. *Acta Mathematica 134*(1), 111–183.

Case, A. C. (1991). Spatial patterns in household demand. *Econometrica 59*(4), 953–965.

Case, A. C. (1992). Neighborhood influence and technological change. *Regional Science and Urban Economics 22*(3), 491–508.

Cliff, A. D. and J. K. Ord (1973). *Spatial Autocorrelation.* London: Pion.

Cliff, A. D. and J. K. Ord (1981). *Spatial Processes: Models & Applications*, Volume 44. Pion London.

Conley, T. G. and B. Dupor (2003). A spatial analysis of sectoral complementarity. *Journal of Political Economy 111*(2), 311–352.

Conley, T. G. and E. Ligon (2002). Economic distance and cross-country spillovers. *Journal of Economic Growth 7*(2), 157–187.

Cressie, N. A. (1993). *Statistics for Spatial Data*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons.

Dahlhaus, R. and H. Künsch (1987). Edge effects and efficient parameter estimation for stationary random fields. *Biometrika 74*(4), 877–882.

De Jong, R. M. and H. J. Bierens (1994). On the limit behavior of a chi-square type test if the number of conditional moments tested approaches infinity. *Econometric Theory 10*(01), 70–90.

Eaton, M. L. (1989). Group Invariance Applications in Statistics. In *Regional conference series in Probability and Statistics*. IMS, Hayward, CA.

Ensor, K. B. and H. J. Newton (1988). The effect of order estimation on estimating the peak frequency of an autoregressive spectral density. *Biometrika 75*(3), 587–589.

Golubov, B. I. (1985). Absolute convergence of multiple Fourier series. *Mathematical Notes 37*(1), 8–15.

Gradshteyn, I. S. and I. M. Ryzhik (1994). *Table of Integrals, Series and Products* (5th ed.). Academic Press, London.

Grenander, U. and G. Szegö (1984). *Toeplitz Forms and their Applications*. Chelsea, New York.

Guyon, X. (1982). Parameter estimation for a stationary process on a d-dimensional lattice. *Biometrika 69*(1), 95–105.

Haining, R. P. (1978). The moving average model for spatial interaction. *Transactions of the Institute of British Geographers 3*(2), 202–225.

Hall, P. (1984). Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of Multivariate Analysis 14*(1), 1–16.

Hannan, E. J. (1973). The asymptotic theory of linear time-series models. *Journal of Applied Probability 10*(1), 130–145.

Helson, H. and D. Lowdenslager (1958). Prediction theory and Fourier series in several variables. *Acta Mathematica 99*(1), 165–202.

Helson, H. and D. Lowdenslager (1961). Prediction theory and Fourier series in several variables. II. *Acta Mathematica 106*(3), 175–213.

Hernández, L. and R. Valdés (2001). What drives contagion: Trade, neighborhood, or financial links? *International Review of Financial Analysis 10*(3), 203–218.

Hillier, G. and F. Martellosio (2013). Properties of the maximum likelihood estimator in spatial autoregressive models. Mimeo.

Horn, R. A. and C. R. Johnson (1985). *Matrix Analysis.* Cambridge University Press.

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and monte carlo. *The Annals of Statistics 1*(5), 799–821.

Kallianpur, G. and M. Mandrekar (1983). Non-deterministic random fields and Wold and Halmos decomposition for commuting isometries. In V. Mandrekar and H. Salehi (Eds.), *Prediction Theory and Harmonic Analysis: The Pesi Masani Volume*, pp. 165–190. North-Holland, Amsterdam.

Kelejian, H. H. and I. R. Prucha (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics 17*(1), 99–121.

Kelejian, H. H. and I. R. Prucha (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review 40*(2), 509–533.

Kelejian, H. H. and I. R. Prucha (2002). 2SLS and OLS in a spatial autoregressive model with equal spatial weights. *Regional Science and Urban Economics 32*(6), 691–707.

Kelejian, H. H. and I. R. Prucha (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics 157*(1), 53–67.

Kelejian, H. H., I. R. Prucha, and Y. A. Yuzefovich (2006). Estimation problems in models with spatial weighting matrices which have blocks of equal elements. *Journal of Regional Science 46*(3), 507–515.

Kolympiris, C., N. Kalaitzandonakes, and D. Miller (2011). Spatial collocation and venture capital in the US biotechnology industry. *Research Policy 40*(9), 1188–1199.

Konovalov, S. P. (1979). Absolute convergence of multiple Fourier series. *Mathematical Notes 25*(2), 109–112.

Korezlioglu, H. and P. Loubaton (1986). Spectral factorization of wide sense stationary processes on $\mathbb{Z}^2$. *Journal of Multivariate Analysis 19*(1), 24–47.

Kromer, R. E. (1970). *Asymptotic Properties of the Autoregressive Spectral Estimator*. Ph. D. thesis, Stanford University.

Lahiri, S. N. (1996). On inconsistency of estimators based on spatial data under infill asymptotics. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002) 58*(3), 403–417.

Lee, L. F. (2002). Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. *Econometric Theory 18*(2), 252–277.

Lee, L. F. (2003). Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews 22*(4), 307–335.

Lee, L. F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica 72*(6), 1899–1925.

Lee, L. F. and X. Liu (2010). Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. *Econometric Theory 26*, 187–230.

Liu, S. and H. Neudecker (1995). Matrix-trace Cauchy-Schwarz inequalities and applications in canonical correlation analysis. *Statistical Papers 36*(1), 287–298.

Mann, H. B. and A. Wald (1943). On the statistical treatment of linear stochastic difference equations. *Econometrica 11*(3/4), 173–220.

McClellan, J. H. (1982). Multidimensional spectral estimation. *Proceedings of the IEEE 70*(9), 1029–1039.

Minakshisundaram, S. and O. Szász (1947). On absolute convergence of multiple Fourier series. *Transactions of the American Mathematical Society 61*, 36–53.

Moran, P. A. P. (1950). A test for the serial independence of residuals. *Biometrika 37*(1/2), 178–181.

Moreira, M. J. (2009). A maximum likelihood method for the incidental parameter problem. *The Annals of Statistics 37*(6A), 3660–3696.

Munkres, J. R. (2000). *Topology*. Prentice Hall, Incorporated.

Neyman, J. and E. L. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica 16*(1), 1–32.

Pace, R. K. and R. Barry (1997). Quick computation of spatial autoregressive estimators. *Geographical Analysis 29*(3), 232–247.

Parzen, E. (1969). Multiple time series modeling. In P. R. Krishnaiah (Ed.), *Multivariate Analysis II*, pp. 389–410. Academic Press, New York.

Portnoy, S. (1984). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I. Consistency. *The Annals of Statistics 12*(4), 1298–1309.

Portnoy, S. (1985). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large; II. Normal approximation. *The Annals of Statistics 13*(4), 1403–1417.

Ringland, J. T. (1983). Robust multiple comparisons. *Journal of the American Statistical Association 78*(381), 145–151.

Robinson, P. M. (1977). Estimation of a time series model from unequally spaced data. *Stochastic Processes and their Applications 6*(1), 9–24.

Robinson, P. M. (1978). On consistency in time series analysis. *The Annals of Statistics 6*(1), 215–223.

Robinson, P. M. (1979). Distributed lag approximation to linear time-invariant systems. *The Annals of Statistics 7*(3), 507–515.

Robinson, P. M. (2003). Denis Sargan: Some perspectives. *Econometric Theory 19*(3), 481–494.

Robinson, P. M. (2008). Developments in the analysis of spatial data. *Journal of the Japan Statistical Society (special issue in honour of H. Akaike) 38*(1), 87–96.

Robinson, P. M. (2010). Efficient estimation of the semiparametric spatial autoregressive model. *Journal of Econometrics 157*(1), 6–17.

Robinson, P. M. and J. Vidal Sanz (2006). Modified Whittle estimation of multilateral models on a lattice. *Journal of Multivariate Analysis 97*(5), 1090–1120.

Rosenblatt, M. (1985). *Stationary Sequences and Random Fields*. Birkhäuser.

Rudin, W. (1962). *Fourier Analysis on Groups*. Interscience Tracts in Pure and Applied Mathematics.

Rudin, W. (1973). *Functional Analysis*. McGraw-Hill, New York.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics 6*(2), 461–464.

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*(1), 119–137.

Srivastava, R. and D. Sengupta (2010). Consistent estimation of non-bandlimited spectral density from uniformly spaced samples. *IEEE Transactions on Information Theory 56*(8), 3642–3659.

Tjøstheim, D. (1981). Autoregressive modeling and spectral analysis of array data in the plane. *IEEE Transactions on Geoscience and Remote Sensing GE-19*(1), 15–24.

Tjøstheim, D. (1983). Statistical spatial series modelling II: Some further results on unilateral lattice processes. *Advances in Applied Probability 15*(3), 562–584.

von Bahr, B. and C. G. Esseen (1965). Inequalities for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$. *The Annals of Mathematical Statistics 36*(1), 299–303.

Wester, R., M. Tummala, and C. W. Therrien (1990). Multidimensional autoregressive spectral estimation using iterative methods. In *1990 Conference Record Twenty-Fourth Asilomar Conference on Signals, Systems and Computers*, Volume 1, pp. 442–446.

Whittle, P. (1954). On stationary processes in the plane. *Biometrika 41*, 434–449.

Wiener, N. (1932). Tauberian theorems. *Annals of Mathematics 2*(33), 1–100.

Wiener, N. (1949). *Extrapolation, Interpolation and Smoothing of Stationary Time-Series*. Wiley, New York.

Yao, Q. and P. J. Brockwell (2006). Gaussian maximum likelihood estimation for ARMA models II: Spatial processes. *Bernoulli 12*(3), 403–429.

Yohai, V. J. and R. A. Maronna (1979). Asymptotic behavior of M-estimators for the linear model. *The Annals of Statistics 7*(2), 258–268.

Yuzefovich, Y. A. (2003). *Two Essays on Spatial Econometrics*. Ph. D. thesis, University of Maryland.