ISSN 1755-5361

University of Essex

Department of Economics

Discussion Paper Series

No. 773 October 2015

Pseudo Maximum Likelihood Estimation of Spatial Autoregressive Models with Increasing Dimension

Abhimanyu Gupta
and Peter M. Robinson

Pseudo Maximum Likelihood Estimation of Spatial Autoregressive Models with Increasing Dimension

Abhimanyu Gupta*
Department of Economics
University of Essex, UK

Peter M. Robinson †‡
Department of Economics
London School of Economics, UK

October 22, 2015

Abstract

Pseudo maximum likelihood estimates are developed for higher-order spatial autoregressive models with increasingly many parameters, including models with spatial lags in the dependent variables and regression models with spatial autoregressive disturbances. We consider models with and without a linear or nonlinear regression component. Sufficient conditions for consistency and asymptotic normality are provided, the results varying according to whether the number of neighbours of a particular unit diverges or is bounded. Monte Carlo experiments examine finite-sample behaviour.

JEL classifications: C21, C31, C36

Keywords: Spatial autoregression; increasingly many parameters; consistency; asymptotic normality; pseudo Gaussian likelihood; Monte Carlo

* *Email*: a.gupta@essex.ac.uk.

† Corresponding author. *Email*: p.m.robinson@lse.ac.uk, *Telephone*: +44-20-7955-7516 *Fax*: +44-20-7955-6592.

‡ Research supported by ESRC Grant ES/J007242/1.

1 Introduction

Spatial autoregressive (SAR) models, introduced by Cliff and Ord (1973), have the ability to describe spatial dependence parsimoniously even when data are irregularly-spaced or when economic (not necessarily geographic) distances between units are known, and information on locations is unavailable. They have been widely used in modelling economic and geographic data. The first-order SAR model, which involves a single weight matrix, consisting of (inverse) distances and a single correlation parameter, has been the focus of much research. Greater flexibility, at the cost of less parsimony, is afforded by higher-order SAR models, which incorporate two or more weight matrices and corresponding parameters. These have been studied in both theoretical and applied research. Brandsma and Ketellapper (1979) introduced a second-order model, and discussed its estimation. Blommestein (1983, 1985), Blommestein and Koper (1992, 1997), Anselin and Smirnov (1996), LeSage and Pace (2011), Elhorst, Lacombe and Piras (2012) and others explored various issues in the specification and estimation of higher order SAR models, the latter two references listing a number of others. A recent purely empirical study is in Kolympiris, Kalaitzandonakes, and Miller (2011). A book length exposition can be found in Anselin (1988).

In the present paper we investigate large sample statistical inference on higher order SAR models, in which the number of parameters is allowed to increase slowly with sample size, denoted n. From this perspective we find it convenient to consider four specifications that have somewhat different theoretical implications. For an $n \times 1$ vector y_n of observations and an integer $p_n \geq 1$, possibly regarded as increasing as n increases, let W_{in}, $i = 1, \ldots, p_n$, be $n \times n$ known weight matrices whose elements are inverse economic distances, let $\lambda_{0n} = (\lambda_{01n}, \ldots, \lambda_{0p_n n})'$, the prime denoting transposition, be a vector of unknown parameters, and let u be an $n \times 1$ vector of independent, zero-mean, homoscedastic unobservable random variables. The basic p_nth-order SAR model, denoted SAR(p_n), is

$$y_n = \sum_{i=1}^{p_n} \lambda_{0in} W_{in} y_n + u. \tag{1.1}$$

Let l_n be a $n \times 1$ vector of ones and let τ_0 be an unknown scalar. The SAR(p_n) with intercept is

$$y_n = \sum_{i=1}^{p_n} \lambda_{0i} W_{in} y_n + \tau_0 l_n + u. \tag{1.2}$$

For given integers $k_n \geq 1$ (possibly regarded as increasing with n) and fixed $q \geq 1$ let β_{0n} be an unknown $k_n \times 1$ vector, let δ_0 be a known or unknown $q \times 1$ vector and let $X_n(\delta_0)$ be an $n \times k_n$ matrix of functions of δ_0 and of explanatory variables, with reference to the latter suppressed. The SAR(p_n) with regressors is

$$y_n = \sum_{i=1}^{p_n} \lambda_{0in} W_{in} y_n + X_n(\delta_0) \beta_{0n} + u. \tag{1.3}$$

Finally, for an $n \times 1$ vector v_n of unobservable random variables, the regression with SAR(p_n) errors is

$$y_n = X_n(\delta_0)\beta_{0n} + v_n, \; v_n = \sum_{i=1}^{p_n} \lambda_{0in} W_{in} v_n + u. \tag{1.4}$$

These models correspond to versions of p_nth-order autoregressive time series models, where competing approaches to introducing both autocorrelation and explanatory variables are mirrored by (1.3) and (1.4).

When τ_0 is known (1.2) nests (1.1) (which is sometimes referred to as 'pure SAR'), while (1.2) is nested in both (1.3) and (1.4) when $X_n(\delta_0)$ contains a subvector l_n. In most spatial autoregression literature, SAR(1) versions of these models have been studied, and previous higher-order SAR literature has almost exclusively assumed that p_n and k_n are fixed. In the the bulk of the literature on (1.3) and (1.4) the regression component is linear, formally covered by regarding δ_0 as known. However, (1.3) and (1.4) allow for nonlinear regression, which features widely in statistics (cf. eg Jennrich (1969)) and econometrics but apparently not in the SAR literature. For example, the elements of $X_n(\delta_0)$ may be parametric Box-Cox, arcsinh or other nonlinear transformations of basic explanatory variables. The separation of β_{0n} from δ_0 follows much of the nonlinear regression literature in expressing the likely presence of an unknown scaling vector. The n-subscripting in $X_n(\delta_0)$ allows it to depend on spatial lags of explanatory variables, which entail weight matrices. The model (1.4) may be included in (1.3) by replacing $X_n(\delta_0)$ by a function of both δ_0 and λ_{0n}, but (1.4) is of sufficient practical importance to warrant separate consideration.

Interest centres on statistical inference on λ_{0n}, β_{0n} and, when it is unknown, δ_0. Consider what is known or anticipated from the literature that regards p_n and k_n as fixed. In (1.1) and (1.2), despite the linearity in parameters, least squares estimates are well known to be inconsistent, for typical W_{in}, which differ from the ones which deliver consistency in the autoregressive time series models formally covered; however, for (1.1) Kelejian and Prucha (1999) established consistency of a generalized method of moments estimate. For the same reason consistency of least squares estimates of all parameters in (1.3) is problematic, though from Lee (2002) (who assumed $p_n = 1$ and linear regression) we may expect consistency to be achieved under certain asymptotic conditions on the W_{in}. Under milder such conditions, again when the regression is linear, use of instrumental variables, when available, can produce closed form consistent estimates in (1.3), see e.g. Kelejian and Prucha (1998); for nonlinear regression one expects to be able to extend, eg, Amemiya (1974). As under many other relaxations of Gauss-Markov conditions, least squares estimates of β_{0n} in the first equation of (1.4) (or nonlinear least squares estimates of β_{0n} and δ_0) are expected to be consistent, though those of λ_{0n} based on residuals inconsistent; see eg Kelejian and Prucha (1997). When estimates are consistent, one expects them to satisfy a central limit theorem under additional conditions. The models (1.1)-(1.4) are somewhat idealised, some of the literature considering ones that are more general. In 'SARAR' versions of (1.1), (1.2) or (1.3), u is replaced by v_n, defined as in (1.4) but with p_n possibly replaced by some other order r_n, say. However after transformation they are still essentially covered by (1.1)-(1.3), albeit offering more parsimony, having SAR order $p_n r_n$ with coefficients depending on only $p_n + r_n$ unknowns. In a SARAR version of (1.3), Lee and Liu (2010) established asymptotic theory for generalized method of moments estimates, as did Badinger and Egger (2011, 2013), allowing respectively for error heteroscedasticity and panel structure. Spatial ARMA models are not covered in (1.1)-(1.4); in this setting Huang (1984), Anselin (2001) respectively discussed maximum likelihood estimation and developed Lagrange multiplier tests to determine model order.

A single type of estimate which can be expected to deliver consistency, and asymptotic normality, in (1.1)-(1.4), and without recourse to instrumental variables, is the Gaussian pseudo-maximum likelihood estimate (PMLE). This maximizes what would be the likelihood were u Gaussian, and as well as enjoying the classical asymptotic properties of maximum likelihood, is consistent and asymptotically normal under more general conditions on u, though in some settings the limiting covariance matrix can be affected. Brandsma and Ketellapper (1979) discussed Gaussian maximum likelihood estimation in the SAR(2) version of (1.1), describing, without rigorous proofs, asymptotic statistical properties, see also Huang (1984). These were established for the PMLE by Lee (2004) in case of SAR(1) versions of (1.1)-(1.3) with linear regression in the latter model. The PMLE is asymptotically efficient when u is Gaussian, though otherwise more efficient estimates have been justified in fixed parameter dimension SAR models, see Lee and Liu (2010), Robinson (2010). Note that our allowance for nonlinear regression does not greatly impact on methods and theory for the PMLE, which is in any case only implicitly defined.

In practice the specification of p_n, and of k_n, may reflect the amount of data n available, as is the case with other multiparameter statistical models. A larger data set affords the possibility of achieving reasonably precise inference on a richer model, which may reflect a degree of model uncertainty. Correspondingly, in a number of other multiparameter models, asymptotic statistical theory has been developed with the number of parameters increasing slowly with sample size, cf. Huber (1973), Berk (1974), Robinson (1979), Portnoy (1984, 1985), Robinson (2003). Gupta and Robinson (2015) have argued that regarding p_n as increasing with n is natural in SAR models with some kinds of weight matrix, and have established asymptotic theory for least squares and instrumental variables estimates of (1.3) in the linear regression case.

The present paper establishes consistency and asymptotic normality for the PMLE in the models (1.1)-(1.4) with p_n and k_n allowed to increase slowly with n. Asymptotic theory for implicitly-defined extremum estimates, requiring an initial consistency proof, is unusual in the literature on increasing parameter dimension with sample size, especially so when combined with nonlinear regression, and our proof of consistency of the PMLE is rather delicate. Our results lead to rules of statistical inference which are also valid when p_n and k_n are regarded as fixed, and to some extent provide a novel contribution in this setting also. In particular we know of no asymptotic theory for the PMLE in the models (1.1)-(1.4) with fixed $p_n > 1$ and k_n. We keep the dimension q of δ_0 fixed as otherwise regression would effectively be nonparametric.

The following section covers models (1.1) and (1.2), with (1.3) and (1.4) covered in Sections 3 and 4, respectively. Section 5 contains a Monte Carlo study of finite sample performance. Proofs are included in two Appendices and an additional online supplementary appendix.

2 SAR with and without intercept

Consider (1.1), stressing the dependence of $p = p_n$ on n. We can rewrite (1.1) as

$$S_n y_n = u \tag{2.1}$$

where $S_n = I_n - \sum_{i=1}^{p_n} \lambda_{0in} W_{in}$. The notation S_n follows a convention we adopt for evaluation of objects at true parameters: $A(\alpha_0) \equiv A$ for any matrix, vector or scalar A and any true parameter α_0. In the sequel we suppress reference to n for individual parameters to simplify notation. We now introduce some basic assumptions.

Assumption 1. *$u = (u_1, \ldots, u_n)'$ has iid elements with zero mean, finite variance σ_0^2 and finite fourth moment.*

Assumption 2. *For $i = 1, \ldots, p_n$, the diagonal elements of each W_{in} are zero and the off-diagonal elements of W_{in} are uniformly $\mathscr{O}\left(h_n^{-1}\right)$, where h_n is a positive sequence which is bounded away from zero and which may be bounded or divergent, with $n/h_n \to \infty$ as $n \to \infty$.*

It is possible to employ different h_{in} for each of the W_{in}, some bounded and some divergent. However we maintain Assumption 2 for notational simplicity. For any rectangular matrix A, we define $\|A\| = \left\{\overline{\zeta}\left(A'A\right)\right\}^{\frac{1}{2}}$, where $\bar{\zeta}(B)$ (respectively $\underline{\zeta}(B)$) is the largest (smallest) eigenvalue of a square, symmetric matrix B.

Definition For $i = 1, \ldots, p_n$, W_{in} are said to have *'single nonzero diagonal block'* structure if, for some set of $m_i \times m_i$ matrices V_{in} such that $\sum_{i=1}^{p_n} m_i = n$, W_{in} has V_{in} as the ith diagonal block and zeros elsewhere.

Let c, C denote throughout generic positive constants, arbitrarily small and large, respectively, that do not depend on n or λ.

Assumption 3. *S_n is non-singular for all sufficiently large n, and $\left\|S_n^{-1}\right\| + \|W_{in}\| \leq C$, $i = 1, \ldots, p_n$, for sufficiently large n. If the W_{in} do not have 'single nonzero diagonal block' structure, then*

$$\sum_{i=1}^{p_n} \|W_{in}\|^2 \leq C. \tag{2.2}$$

The first part of this assumption ensures that (2.1) can be solved for y_n, asymptotically. The restriction on $\left\|S_n^{-1}\right\|$ ensures the limitation of spatial correlation to a manageable degree because the covariance matrix of y_n is $\sigma_0^2 S_n^{-1} S_n^{-1\prime}$, while those for the $\|W_{in}\|$ are satisfied if, for each i, the elements of W_{in} decline fast enough with n. The process is controlled over increasingly many lags by (2.2). A sufficient condition for the non-singularity of S_n is

$$\left\|\sum_{i=1}^{p_n} \lambda_{0i} W_{in}\right\| < 1. \tag{2.3}$$

Depending on the structure of W_{in} more primitive sufficient conditions can be given for (2.3). In the 'single nonzero diagonal block' case we have $\|\sum_{i=1}^{p_n} \lambda_{0i} W_{in}\| \le \max_{i=1,\ldots,p_n} (|\lambda_{0i}| \|V_{in}\|)$, in which case one could take Λ_n such that $\max_{i=1,\ldots,p_n} |\lambda_i| < 1$ and take normalized V_{in} such that $\|V_{in}\| = 1$. For more general W_{in} we have $\|\sum_{i=1}^{p_n} \lambda_{0i} W_{in}\| \le \|\lambda_0\| \left(\sum_{i=1}^{p_n} \|W_{in}\|^2\right)^{\frac{1}{2}}$, and then we may choose Λ_n such that $\|\lambda\| < 1$ and normalize W_{in} such that $\sum_{i=1}^{p_n} \|W_{in}\|^2$=1. In any case, for the identification of the λ_i some normalization of the W_{in} is necessary, so it is essentially costless to do this operation. A similar discussion applies after Assumption 12 below, with row-sum norm used instead. Denote by $\lambda = (\lambda_1, \ldots, \lambda_{p_n})'$ and σ^2 any admissible values of λ_{0n} and σ_0^2. Define the negative Gaussian log-likelihood function as

$$\log\left(2\pi\sigma^2\right) - 2n^{-1} \log|S_n(\lambda)| + \sigma^2 n^{-1} y_n' S_n(\lambda) S_n(\lambda) y_n, \tag{2.4}$$

where $S_n(\lambda) = I_n - \sum_{i=1}^{p_n} \lambda_i W_{in}$. For given λ, (2.4) is minimised with respect to σ^2 by

$$\bar{\sigma}_n^2(\lambda) = n^{-1} y_n' S_n(\lambda) S_n(\lambda) y_n. \tag{2.5}$$

Define the PMLEs of λ_{0n}, σ_0^2 as $\hat{\lambda}_n = \arg\min_{\lambda\in\Lambda_n} \mathcal{Q}_n(\lambda), \hat{\sigma}_n^2 \equiv \bar{\sigma}_n^2\left(\hat{\lambda}_n\right)$ respectively, where

$$\mathcal{Q}_n(\lambda) = \log \bar{\sigma}_n^2(\lambda) + n^{-1} \log\left|S_n^{-1}(\lambda) S_n^{-1\prime}(\lambda)\right|, \tag{2.6}$$

with Λ_n satisfying

Assumption 4. *Λ_n is a subset of $\mathbb{R}^{p_n}$ such that $-\varepsilon \le \lambda_i \le 1-\varepsilon$, for $i = 1, \ldots, p$ when the W_{in} have 'single nonzero diagonal block' structure and $\|\lambda\| \le 1-\varepsilon$ if not, for some fixed $\varepsilon \in (0,1)$.*

The first specification reflects the necessity in our proof that the volume of Λ_n remain bounded as $n \to \infty$, and the likelihood that the λ_{0i} are non-negative, but could be replaced by others.

Assumption 5. *$\lambda_{0n} \in \Lambda_n$, for all sufficiently large n.*

Denote

$$\sigma_n^2(\lambda) = n^{-1} \sigma_0^2 tr\left(S_n^{-1\prime} S_n'(\lambda) S_n(\lambda) S_n^{-1}\right). \tag{2.7}$$

Assumption 6. *For $\lambda \in \Lambda_n$ and all sufficiently large n, $c \le \sigma_n^2(\lambda) \le C$.*

$\sigma_n^2(\lambda)$ is nonnegative by inspection and finite by Assumption 3. For a generic matrix define $\|A\|_F = \{tr(A'A)\}^{\frac{1}{2}}$ and introduce

Assumption 7. *For any $\eta > 0$,*

$$\underline{\lim}_{n\to\infty} \inf_{\lambda\in\overline{\mathcal{N}}_n^{\lambda}(\eta)} n^{-1} \|T_n(\lambda)\|_F^2 / |T_n(\lambda)|^{2/n} > 1, \tag{2.8}$$

where $T_n(\lambda) = S_n(\lambda)S_n^{-1}$, $\overline{\mathcal{N}}_n^{\lambda}(\eta) = \Lambda_n \setminus \mathcal{N}_n^{\lambda}(\eta)$, $\mathcal{N}_n^{\lambda}(\eta) = \{\lambda : \|\lambda - \lambda_{0n}\| < \eta\} \cap \Lambda_n$.

The ratio in (2.8) is guaranteed ≥ 1 due to the inequality between arithmetic and geometric means. Assumption 7 is an identification condition related to the uniqueness of the covariance matrix of y_n, introduced in Delgado and Robinson (2014) who discuss it and compare it to the identification condition employed by Lee (2004) in his asymptotic theory.

Theorem 2.1. *Let Assumptions 1-7 hold, and p_n be allowed to diverge as $n \to \infty$. Then*

$$\left\|\hat{\lambda}_n - \lambda_{0n}\right\| \xrightarrow{p} 0, \text{ as } n \to \infty.$$

Theorem 2.2. *Let Assumptions 1-7 hold, and p_n be allowed to diverge as $n \to \infty$. Then $\hat{\sigma}_n^2 - \sigma_0^2 = o_p(1)$, as $n \to \infty$.*

To establish asymptotic normality, we introduce the derivatives of (2.4). The second derivative matrix of (2.4) at (λ, σ^2) is denoted $H_n(\lambda, \sigma^2)$, and defined in (A.23) in Appendix A. Writing $P_{1n}(\lambda)$, $P_{2n}(\lambda)$ for the $p_n \times p_n$ matrices with (i,j)-th element given by $tr\left(G_{jn}(\lambda)G_{in}(\lambda)\right)$, $tr\left(G'_{jn}(\lambda)G_{in}(\lambda)\right)$, respectively, with $G_{in}(\lambda) = W_{in}S_n^{-1}(\lambda)$ for $i = 1, \ldots, p_n$, we deduce (more details in appendix) that

$$\Xi_n = \mathbb{E}(H_n) = 2n^{-1}(P_{1n} + P_{2n}). \tag{2.9}$$

Write F_n for the $n \times p_n$ matrix with (i,j)-th element $c_{ii,jn}$, where $c_{pq,in}$ is the (p,q)-th element of $G_{in} + G'_{in}$, and define $\Omega_n = (\mu_4 - 3\sigma_0^4)\sigma_0^{-4}n^{-1}F'_nF_n$, where $\mu_l = \mathbb{E}(u_i^l)$. The covariance matrix of the first derivative of (2.4) is $n^{-1}(2\Xi_n + \Omega_n)$. The following assumption is standard:

Assumption 8. *λ_{0n} is in the interior of Λ_n, for all sufficiently large n.*

If h_n diverges with n, we need to account for the normalisation that will yield a non-degenerate and finite asymptotic distribution as follows:

Assumption 9. *$h_n \to \infty$ as $n \to \infty$. $\overline{\lim}_{n\to\infty} \overline{\zeta}(h_n\Xi_n) < \infty$ and $\underline{\lim}_{n\to\infty} \underline{\zeta}(h_n\Xi_n) > 0$.*

Assumption 10. *h_n is bounded as $n \to \infty$. $\overline{\lim}_{n\to\infty} \overline{\zeta}\left(\Xi_n^{-1}\Omega_n\Xi_n^{-1}\right) < \infty$,
$\underline{\lim}_{n\to\infty} \underline{\zeta}\left(2\Xi_n^{-1} + \Xi_n^{-1}\Omega_n\Xi_n^{-1}\right) > 0$ and $\underline{\lim}_{n\to\infty} \underline{\zeta}(\Xi_n) > 0$.*

The rank conditions here strongly restrict the W_{in} in higher-order SAR models, even with fixed p_n. Such problems are transparently avoided with weight matrices having 'single nonzero diagonal block' structure. Blommestein (1985) discusses the possibility of 'circularity' when W_{in} represent orders of contiguity, causing rank condition failure. By way of an illustration, W_{1n} could assign 1 to an element if the relevant units share a common boundary, W_{2n} could assign 1 to an element if the relevant units do not share a boundary with each other but have a common neighbour, and so on. In this case, there is a risk of high-order W_{in} 'circling' back to W_{1n}.

Assumption 11. *$\mathbb{E}\left|u_i\right|^{4+\chi} \leq C$ for $i = 1, \ldots, n$, for some $\chi > 0$.*

For any $s \times q$ matrix $A = [a_{ij}]$ we define $\|A\|_R = \max_{i=1,\ldots,s} \sum_{j=1}^{q} |a_{ij}|$, the maximum absolute row-sum norm.

Assumption 12. *S_n is non-singular and $\left\|S_n^{-1}\right\|_R + \left\|S_n'^{-1}\right\|_R + \|W_{in}\|_R + \|W_{in}'\|_R \leq C$, $i = 1, \ldots, p_n$, for all sufficiently large n. If the W_{in} do not have a 'single nonzero diagonal block' structure, then*

$$\sum_{i=1}^{p_n} \left(\|W_{in}\|_R^2 + \|W_{in}'\|_R^2 \right) \leq C. \tag{2.10}$$

This is a strengthening of Assumption 3 due to the inequality $\|A\|^2 \leq \|A\|_R \|A'\|_R$.

Denote throughout by Ψ_n a matrix of constants with full row rank m and columns equal to the number of parameters for which a central limit theorem is being established.

Theorem 2.3. *Let Assumptions 1, 2, 4-9, 11 and 12 hold, $h_n^{1+\frac{4}{\chi}}/n \to 0$ as $n \to \infty$, p_n be allowed to diverge as $n \to \infty$ such that*

$$(a)\ \frac{p_n^5}{n} + \frac{p_n}{h_n} \to 0 \text{ and either } (b)\ \frac{h_n}{p_n^4} = \mathscr{O}(1) \text{ or } (c)\ \frac{p_n^{\frac{1}{2}} h_n}{n} \to 0, \text{ as } n \to \infty, \tag{2.11}$$

and

$$\left\|S_n^{-1}\right\|_F + \sum_{i=1}^{p_n} \|W_{in}\|_F^2 \leq C, \text{ for all sufficiently large } n. \tag{2.12}$$

Then

$$\frac{n^{\frac{1}{2}}}{h_n^{\frac{1}{2}} p_n^{\frac{1}{2}}} \Psi_n \left(\hat{\lambda}_n - \lambda_{0n} \right) \xrightarrow{d} N\left(0, \Delta_1\right), \text{ as } n \to \infty,$$

where $\Delta_1 = 2 \lim_{n\to\infty} p_n^{-1} \Psi_n \left(h_n \Xi_n\right)^{-1} \Psi_n'$.

The rate condition (c) in (2.11) is a strengthening of the last part on the LHS of (a). Condition (2.12) controls the spatial dependence, and is imposed to avoid practically infeasible conditions relating p_n, h_n and n. If $G_{jn}G_{in} = 0$ and $G_{jn}'G_{in} = 0$ for $i \neq j$, as with 'single nonzero diagonal block' weight matrices, then any finite-dimensional subset of estimates will be asymptotically distributed as independent normal random variables with mean zero and variances $\left\{\lim_{n\to\infty} \left(h_n/n\right) tr\left(G_{in}^2 + G_{in}'G_{in}\right)\right\}^{-1}$. If p_n is fixed then the restrictions on p_n in (2.11)(a) are redundant and (2.11)(b) cannot hold but by Assumption 2 (2.11)(c) holds. In this case the same proof, with $m = 1$, implies $\left(n^{\frac{1}{2}}/h_n^{\frac{1}{2}}\right)\left(\hat{\lambda}_n - \lambda_{0n}\right) \xrightarrow{d} N\left(0, 2\lim_{n\to\infty}\left(h_n\Xi_n\right)^{-1}\right)$, by the Cramer-Wold device. We may derive similar results for fixed parameter spaces from the subsequent central limit theorems in this section.

Theorem 2.4. *Let Assumptions 1, 2, 4-8, 10-12 hold, and p_n be allowed to diverge as $n \to \infty$ such that*

$$\frac{p_n^5}{n} + \frac{p_n^{\frac{8}{\chi}+2}}{n} \to 0, \text{ as } n \to \infty. \tag{2.13}$$

Then

$$\frac{n^{\frac{1}{2}}}{p_n^{\frac{1}{2}}} \Psi_n \left(\hat{\lambda}_n - \lambda_{0n}\right) \xrightarrow{d} N\left(0, \Delta_2\right), \text{ as } n \to \infty,$$

where $\Delta_2 = \lim_{n\to\infty} p_n^{-1} \Psi_n \left(2\Xi_n^{-1} + \Xi_n^{-1}\Omega_n\Xi_n^{-1}\right) \Psi_n'$.

The parameter growth restrictions may be simplified if moments of a certain order exist. For instance when $\chi \geq 8/3$, (2.13) only requires $p_n^5/n \to 0$. Covariance matrix estimation for Theorems 2.3 and 2.4 can be based on $H_n\left(\lambda, \sigma_n^2\right)$ and $\Omega_n\left(\lambda, \sigma_n^2\right)$ evaluated at $\hat{\lambda}_n$, $\hat{\sigma}_n^2$ and empirical moments.

We now consider (1.2). For any admissible values λ, τ and σ^2 the negative Gaussian pseudo log-likelihood function is defined as $\log\left(2\pi\sigma^2\right) - 2n^{-1}\log\left|S_n\left(\lambda\right)\right| + \left(n\sigma^2\right)^{-1}\left\|S_n\left(\lambda\right)y_n - l_n\tau\right\|^2$, which for given λ is minimised with respect to τ and σ^2 by $\bar{\tau}_n\left(\lambda\right) = n^{-1}l_n' S_n\left(\lambda\right)y_n$ and $\bar{\sigma}_n^2\left(\lambda\right) = n^{-1}y_n' S_n'\left(\lambda\right)M_{l_n}S_n\left(\lambda\right)y_n$, where we write $M_A = I_n - A(A'A)^{-1}A'$ for any $n \times s$ matrix A of rank s, with I_n denoting the identity matrix of dimension n. The PMLE of λ is $\hat{\lambda}_n = \arg\min_{\lambda\in\Lambda_n} \mathcal{Q}_n\left(\lambda\right)$, where $\mathcal{Q}_n\left(\lambda\right) = \log\bar{\sigma}_n^2\left(\lambda\right) + n^{-1}\log\left|S_n^{-1}\left(\lambda\right)S_n^{-1\prime}\left(\lambda\right)\right|$, and the PMLEs of τ_0 and σ_0^2 are as $\hat{\tau}_n = \bar{\tau}_n\left(\hat{\lambda}_n\right)$ and $\hat{\sigma}_n^2 = \bar{\sigma}_n^2\left(\hat{\lambda}_n\right)$ respectively. The first and second derivatives evaluated at $\left(\lambda_{0n}, \tau_0, \sigma_0^2\right)$ are written ξ_n^I and H_n^I respectively. Both now include derivatives with respect to τ, and explicit expressions can be obtained by taking $X_n = l_n$ in (1.3) and using the formulae subsequently provided. The covariance matrix of the first derivative of the likelihood function is $n^{-1}\left(2\Xi_n^I + \Omega_n^I\right)$, with $\Xi_n^I = \mathbb{E}\left(H_n^I\right)$.

A feature of this model noted by Lee (2004) is potential multicollinearity. For example, if the W_{in} are row-normalised (with each row containing $n-1$ non-zero elements) then $W_{in}l_n = l_n$, so that $G_{in}l_n\tau_0 = \tau_0 l_n\left(1 - \sum_{i=1}^{p_n}\lambda_{0i}\right)^{-1}$ for each i. Then $M_{l_n}G_{in}l_n\tau_0 = 0$ for every i and multicollinearity ensues. Indeed when h_n diverges and $p_n = o\left(h_n\right)$, $\left\|\Xi_n^I\right\| = o(1)$, as $n \to \infty$, implying that $\underline{\zeta}\left(\Xi_n^I\right) = o(1)$ also (see Lee (2004) for justification when $p_n \equiv 1$, extension to divergent p_n being obvious). While the consistency of the estimates as established in the following section is preserved as long as Assumption 7 continues to hold (τ_0 is identified if λ_{0n} is identified), the central limit theorem entails a different norming.

Theorem 2.5. *Let Assumptions 1-7 hold, and p_n be allowed to diverge as $n \to \infty$. Then*

$$\left\|\left(\hat{\lambda}_n', \hat{\tau}_n\right) - \left(\lambda_{0n}', \tau_0\right)\right\| \xrightarrow{p} 0, \text{ as } n \to \infty.$$

Theorem 2.6. *Let (1.2) hold with $h_n \to \infty$ as $n \to \infty$. Let Assumptions 1, 2, 4-8, 11, 12 and (2.12) hold, $\underline{\zeta}\left(\Xi_n^I\right) + h_n^{1+\frac{4}{\chi}}/n \to 0$ as $n \to \infty$, $\overline{\lim}_{n\to\infty}\overline{\zeta}\left(\left(h_n\Xi_n^I\right)^{-1}h_n\Omega_n^I\left(h_n\Xi_n^I\right)^{-1}\right) < \infty$, $\underline{\lim}_{n\to\infty}\underline{\zeta}\left(h_n\Xi_n^I\right) > 0$, $\underline{\lim}_{n\to\infty}\underline{\zeta}\left(2\left(h_n\Xi_n^I\right)^{-1} + \left(h_n\Xi_n^I\right)^{-1}h_n\Omega_n^I\left(h_n\Xi_n^I\right)^{-1}\right) > 0$, and p_n be* *allowed to diverge as* $n \to \infty$ *such that*

$$(a)\ \frac{p_n^5}{n} \to 0,\ \textit{and either}\ (b)\ \frac{h_n}{p_n^4} = \mathscr{O}(1)\ \textit{or}\ (c)\ \frac{p_n^{\frac{1}{2}} h_n}{n} \to 0,\ \textit{as}\ \ n \to \infty. \tag{2.14}$$

Then

$$\frac{n^{\frac{1}{2}}}{h_n^{\frac{1}{2}} p_n^{\frac{1}{2}}} \Psi_n \left(\left(\hat{\lambda}_n', \hat{\tau}_n \right)' - \left(\lambda_{0n}', \tau_0 \right)' \right) \xrightarrow{d} N\left(0, \Delta_3\right),\ \textit{as}\ n \to \infty,$$

where $\Delta_3 = \lim_{n\to\infty} p_n^{-1} \Psi_n \left(2 \left(h_n \Xi_n^I \right)^{-1} + \left(h_n \Xi_n^I \right)^{-1} h_n \Omega_n^I \left(h_n \Xi_n^I \right)^{-1} \right) \Psi_n'$.

If either multicollinearity does not arise or if h_n is bounded the asymptotic distribution of the PMLE for the parameters of (1.2) is covered under the theorems of the following section.

3 SAR with regressors

We now consider (1.3). Let $X_n(\delta)$ have i-th row $x_{in}'(\delta) = (x_{i1n}(\delta), \ldots, x_{ik_n n}(\delta))$, for some known functions $x_{ijn}(\delta)$, $j = 1, \ldots, k_n$, and unknown vector $\delta = (\delta_1, \ldots, \delta_q)'$. A special case of (1.3) is one where δ_0 is known, in which case the regression is linear. By way of a nonlinear illustration when $k_n = q = 1$, in the Box -Cox case we have, for an explanatory variable z_{i1n}, the i-th element of $X_n(\delta)$ is $x_{i1n}(\delta) = \left(z_{i1n}^{\delta} - 1\right)/\delta$. Generally, the vector β_{0n} is distinguished from δ_0, playing a similar scaling role as in a linear model (and unlike δ_0, β_{0n} need not be assumed an element of a prescribed compact set, cf Robinson (1972)). Recall also that q is assumed fixed as n increases.

With $X_n \equiv X_n(\delta_0)$ we have $S_n y_n = X_n \beta_{0n} + u$ and, denoting by $\theta = (\lambda', \beta', \delta')'$ any admissible values of $\theta_{0n} = (\lambda_{0n}', \beta_{0n}', \delta_0')'$, redefine the negative Gaussian pseudo log-likelihood function as

$$\log\left(2\pi\sigma^2\right) - 2n^{-1} \log\left|S_n(\lambda)\right| + \sigma^{-2} n^{-1} \left\| S_n(\lambda) y_n - X_n(\delta)\beta \right\|^2. \tag{3.1}$$

For given $\gamma = (\lambda', \delta')'$, (3.1) is minimised with respect to β and σ^2 by

$$\begin{aligned} \bar{\beta}_n(\gamma) &= \left(X_n'(\delta) X_n(\delta) \right)^{-1} X_n'(\delta) S_n(\lambda) y_n && (3.2)\\ \bar{\sigma}_n^2(\gamma) &= n^{-1} y_n' S_n'(\lambda) M_n(\delta) S_n(\lambda) y_n, && (3.3) \end{aligned}$$

with $M_n(\delta) = I_n - X_n(\delta) \left(X_n'(\delta) X_n(\delta) \right)^{-1} X_n'(\delta)$. The PMLE of γ is $\hat{\gamma}_n = \arg\min_{\gamma \in \Gamma_n} \mathcal{Q}_n(\gamma)$, where we have redefined

$$\mathcal{Q}_n(\gamma) = \log \bar{\sigma}_n^2(\gamma) + n^{-1} \log\left| S_n^{-1}(\lambda) S_n^{-1\prime}(\lambda) \right|, \tag{3.4}$$

$\Gamma_n = \Lambda_n \times \mathcal{D}$, with $\mathcal{D}$ a compact subset of $\mathbb{R}^q$ and $\hat{\delta}_n \equiv \hat{\delta}$. The PMLEs of β_{0n} and σ_0^2 are defined as $\bar{\beta}_n(\hat{\gamma}_n) \equiv \hat{\beta}_n$ and $\bar{\sigma}_n^2(\hat{\gamma}_n) \equiv \hat{\sigma}_n^2$ respectively.

Assumption 13. $\delta_0 \in \mathcal{D}$.

Assumption 14. *$x_{ijn}(\delta)$ are uniformly bounded constants, $i = 1, \ldots, n$, $j = 1, \ldots, k_n$, $\delta \in \mathcal{D}$, and*

$$\varliminf_{n\to\infty} n^{-1} \sup_{\delta\in\mathcal{D}} \underline{\zeta} \left(X_n'(\delta)X_n(\delta)\right) > 0, \text{ as } n \to \infty. \tag{3.5}$$

(3.5) is an asymptotic non-multicollinearity condition.

Assumption 15. *The $x_{ijn}(\delta)$ are uniformly continuous on $\mathcal{D}$: that is, for any $\varepsilon > 0$ and any $\delta_* \in \mathcal{D}$, there exists $\rho > 0$ such that $\varlimsup_{n\to\infty} \max_{1\le i\le n, 1\le j\le k_n} \sup_{\|\delta-\delta_*\|<\rho;\ \delta\in\mathcal{D}} |x_{ijn}(\delta) - x_{ijn}(\delta_*)| < \varepsilon$.*

Assumption 16. *When δ_0 is unknown,*

$$\|\beta_{0n}\| \sim k_n^{1/2} \text{ as } n \to \infty, \tag{3.6}$$

and for any $\eta > 0$,

$$\varliminf_{n\to\infty} \inf_{(\lambda',\,\delta')'\in\, \Lambda_n\times\overline{\mathcal{N}}_n^{\delta}(\eta)} n^{-1}\beta_{0n}'X_n'T_n'(\lambda)M_n(\delta)T_n(\lambda)X_n\beta_{0n}/\|\beta_{0n}\|^2 > 0. \tag{3.7}$$

We could rewrite (3.7) as

$$\varliminf_{n\to\infty} \inf_{(\lambda',\,\beta',\,\delta')'\in\, \Lambda_n\times\mathbb{R}^{k_n}\times\overline{\mathcal{N}}_n^{\delta}(\eta)} n^{-1}\|X_n(\delta)\beta - T_n(\lambda)X_n\beta_{0n}\|^2/\|\beta_{0n}\|^2 > 0, \tag{3.8}$$

which is analogous to the identification condition for the nonlinear regression model $y_n = X_n(\delta)\beta_{0n} + u$ (take $\lambda = \lambda_{0n}$) with a parametric linear factor in Robinson (1972), and (3.8) may be easier to comprehend than (3.7). A sufficient condition is: for any $\eta > 0$

$$\varliminf_{n\to\infty} \inf_{(\lambda',\,\delta')'\in\, \Lambda_n\times\overline{\mathcal{N}}_n^{\delta}(\eta)} n^{-1}\underline{\zeta}\left(X_n'T_n'(\lambda)M_n(\delta)T_n(\lambda)X_n\right) > 0. \tag{3.9}$$

Theorem 3.1. *Let Assumptions 1-7, 13-16 hold, and p_n, k_n be allowed to diverge as $n \to \infty$ such that*

$$\frac{k_n}{n} \longrightarrow 0, \text{ as } n \to \infty. \tag{3.10}$$

Then

$$\left\|\hat{\theta}_n - \theta_{0n}\right\| \xrightarrow{p} 0, \text{ as } n \to \infty.$$

As discussed after Theorem 2.1 the same proof holds when p_n and k_n remain fixed, and the restriction on k_n in (3.10) becomes redundant. The conditions of the theorem can be compared to those in Gupta and Robinson (2015). The requirement of finite fourth order moments for u_i is not imposed for consistency of the IV and OLS estimates, where second moments suffice. On the other hand, the only restriction imposed on h_n here is that it be bounded away from zero uniformly in n. For $\epsilon > 0$, define $\mathcal{N}^{\delta}(\epsilon) = \{\delta : \|\delta - \delta_0\| < \epsilon\}$.

Assumption 17. *For some $\epsilon > 0$, $\partial x_{ijn}(\delta)/\partial\delta_l$ exist and are uniformly bounded in absolute value for all $\delta \in \mathcal{N}^\delta(\epsilon) \cap \mathcal{D}$, $i = 1\ldots,n$, $j = 1,\ldots,k_n$, $l = 1,\ldots,q$. As $n \to \infty$, $\overline{\lim}_{n\to\infty} n^{-1}\bar{\zeta}(X_n'X_n) < \infty$.*

This assumption implies $\sup_{\delta\in\mathcal{N}^\delta(\epsilon)\cap\mathcal{D}} \|\partial x_{ijn}(\delta)/\partial\delta\| < C$.

Theorem 3.2. *Let Assumptions 1-7, 13-17 hold, and p_n, k_n be allowed to diverge as $n \to \infty$ such that $p_n k_n^5/n \to 0$ as $n \to \infty$. Then $\hat{\sigma}_n^2 - \sigma_0^2 = o_p(1)$, as $n \to \infty$. If δ_0 is known (i.e. the regression is linear), the sufficient rate can be improved to $p_n k_n^3/n \to 0$ as $n \to \infty$.*

Assumption 18. *For some $\epsilon > 0$, $\partial^2 x_{ijn}(\delta)/\partial\delta_{l_1}\partial\delta_{l_2}$ and $\partial^3 x_{ijn}(\delta)/\partial\delta_{l_1}\partial\delta_{l_2}\partial\delta_{l_3}$ exist and are uniformly bounded in absolute value for all $\delta \in \mathcal{N}^\delta(\epsilon) \cap \mathcal{D}$, $i = 1,\ldots,n$, $j = 1,\ldots,k_n$, $l_1, l_2, l_3 = 1,\ldots,q$. As $n \to \infty$,*

$$\overline{\lim_{n\to\infty}}\, n^{-1} \max_{l=1,,\ldots,q} \bar{\zeta}\left\{(\partial X_n'/\partial\delta_l)(\partial X_n/\partial\delta_l)\right\} < \infty, \tag{3.11}$$

$$\overline{\lim_{n\to\infty}}\, n^{-1} \max_{l_1,l_2=1,,\ldots,q} \bar{\zeta}\left\{\left(\partial^2 X_n'/\partial\delta_{l_1}\partial\delta_{l_2}\right)\left(\partial^2 X_n/\partial\delta_{l_1}\partial\delta_{l_2}\right)\right\} < \infty. \tag{3.12}$$

Together (3.11) and (3.12) imply $n^{-\frac{1}{2}}\left(\|\partial X_n/\partial\delta_{l_1}\|, \left\|\partial^2 X_n/\partial\delta_{l_1}\partial\delta_{l_2}\right\|\right) = \mathscr{O}(1)$, uniformly in $l_1, l_2 = 1\ldots,q$.

Let $\Pi_n(\theta)$ be the $n \times q$ matrix with i-th column $(\partial X_n(\delta)/\partial\delta_i)\beta$, where the matrix is differentiated element-by-element. Redefine H_n to be the second derivative matrix of (3.1), so

$$\Xi_n = \mathbb{E}(H_n) = 2\sigma_0^{-2} n^{-1} \begin{bmatrix} \sigma_0^2(P_{1n} + P_{2n}) + A_n'A_n & A_n'X_n & A_n'\Pi_n \\ * & X_n'X_n & X_n'\Pi_n \\ * & * & \Pi_n'\Pi_n \end{bmatrix}, \tag{3.13}$$

where $A_n = [a_{1n},\ldots,a_{p_n n}]$ with $a_{in} = G_{in}X_n\beta_{0n}$. Assumption 14 implies $a_{ijn} = \mathscr{O}(k_n)$, uniformly in $i = 1,\ldots,n$, $j = 1,\ldots,k_n$, where a_{ijn} is the (i,j)-th element of A_n. More details on derivatives are in Appendix A, where their components are used in the proofs of the central limit theorems stated below. Define $L_n = n^{-1}\left([A_n, X_n, \Pi_n]'[A_n, X_n, \Pi_n]\right)$, which equals $\sigma_0^2\Xi_n/2 - \sigma_0^2(P_{1n} + P_{2n})$, with some abuse of notation.

Assumption 19. $\underline{\lim}_{n\to\infty}\, \underline{\zeta}(L_n) > 0$ *and* $\overline{\lim}_{n\to\infty}\, \bar{\zeta}(L_n) < \infty$.

Theorem 3.3. *Let $h_n \to \infty$ as $n \to \infty$, Assumptions 1, 2, 4-8, 12, 14-19 hold, δ_0 be in the interior of $\mathcal{D}$, and p_n, k_n be allowed to diverge as $n \to \infty$ such that*

$$\frac{p_n^2 k_n^4}{n h_n}\left(\frac{p_n^3}{h_n} + k_n^3\right) + \frac{p_n^4 k_n^6}{n} + \frac{p_n^3 k_n^2}{h_n^2} \longrightarrow 0, \text{ as } n \to \infty. \tag{3.14}$$

Then

$$\frac{n^{\frac{1}{2}}}{(p_n + k_n)^{\frac{1}{2}}}\Psi_n\left(\hat{\theta}_n - \theta_{0n}\right) \xrightarrow{d} N(0, \Delta_4), \text{ as } n \to \infty,$$

where $\Delta_4 = \sigma_0^2 \lim_{n\to\infty} (p_n + k_n)^{-1} \Psi_n L_n^{-1} \Psi_n'$.

$n^{-1} \left[W_{1n}y_n, \ldots, W_{p_n n}y_n, X_n\left(\hat{\delta}\right), \Pi_n\left(\hat{\theta}_n\right)\right]' \left[W_{1n}y_n, \ldots, W_{p_n n}y_n, X_n\left(\hat{\delta}\right), \Pi_n\left(\hat{\theta}_n\right)\right]$ and $\hat{\sigma}_n^2$ can replace L_n and σ_0^2 respectively to obtain a consistent estimate of Δ_4. When p_n and k_n are fixed we obtain $n^{\frac{1}{2}}\left(\hat{\theta}_n - \theta_{0n}\right) \xrightarrow{d} N\left(0, \sigma_0^2 \lim_{n\to\infty} L_n^{-1}\right)$ via the Cramer-Wold device, as discussed after Theorem 2.3. Similar comments apply after the other central limit theorems presented subsequently both in this section and the next one. If h_n is bounded as $n \to \infty$ a more complicated analysis is required because the information equality does not hold asymptotically. Define

$$\Omega_n = \sigma_0^{-4} n^{-1} \begin{bmatrix} F_n'\left(4\mu_3 A_n + \left(\mu_4 - 3\sigma_0^4\right) F_n\right) & 2\mu_3 F_n' X_n & 2\mu_3 F_n' \Pi_n \\ * & 0 & 0 \\ * & * & 0 \end{bmatrix}. \tag{3.15}$$

Again $n^{-1}(2\Xi_n + \Omega_n)$ is the covariance matrix of the first derivative of (3.1). The asymptotic distribution relies on the following non-multicollinearity and boundedness condition:

Assumption 20. $\overline{\lim}_{n\to\infty} \overline{\zeta}\left(\Xi_n^{-1}\Omega_n\Xi_n^{-1}\right) < \infty$, $\underline{\lim}_{n\to\infty} \underline{\zeta}\left(2\Xi_n^{-1} + \Xi_n^{-1}\Omega_n\Xi_n^{-1}\right) > 0$ *and* $\underline{\lim}_{n\to\infty} \underline{\zeta}\left(\Xi_n\right) > 0$.

Theorem 3.4. *Let h_n be bounded as $n \to \infty$, Assumptions 1, 2, 4-8, 11, 12, 14-18, 20 hold, δ_0 be in interior of $\mathcal{D}$, and p_n, k_n be allowed to diverge as $n \to \infty$ such that*

$$\frac{p_n^2 k_n^4}{n}\left(p_n^3 + k_n^3\right) + \frac{(p_n k_n)^{\frac{8}{\chi}+2}}{n} \longrightarrow 0, \text{ as } n \to \infty. \tag{3.16}$$

Then

$$\frac{n^{\frac{1}{2}}}{(p_n + k_n)^{\frac{1}{2}}} \Psi_n\left(\hat{\theta}_n - \theta_{0n}\right) \xrightarrow{d} N\left(0, \Delta_5\right), \text{ as } n \to \infty,$$

where $\Delta_5 = \lim_{n\to\infty} (p_n + k_n)^{-1} \Psi_n \left(2\Xi_n^{-1} + \Xi_n^{-1}\Omega_n\Xi_n^{-1}\right) \Psi_n'$.

The parameter space growth restriction (3.16) can be simplified according to the value of χ. For example $\chi \geq 8/3$ implies that $p_n^5 k_n^7/n = o(1)$ suffices for (3.16) to hold while if $\chi \geq 8/5$ and p_n is fixed then $k_n^7/n \to 0$ is sufficient.

4 Regression with SAR errors

From (1.4), we get the model

$$S_n(\lambda_0) y_n = X_n(\gamma_0)\beta_0 + u, \tag{4.1}$$

where with some abuse of notation $X_n(\gamma) = S_n(\lambda) X_n(\delta)$. Thus consider $\mathcal{Q}_n(\gamma)$ defined as before but with

$$\begin{aligned} \overline{\sigma}_n^2(\gamma) &= n^{-1} y_n' S_n'(\lambda) M_n(\gamma) S_n(\lambda) y_n, \\ M_n(\gamma) &= I_n - X_n(\gamma)\left(X_n'(\gamma) X_n(\gamma)\right)^{-1} X_n'(\gamma). \end{aligned}$$

Write $X_n = X_n(\gamma_0)$ and introduce

Assumption 21. *When δ_0 is unknown, (3.6) holds and for any $\eta > 0$*

$$\underline{\lim}_{n\to\infty} \inf_{(\lambda',\delta')\in\Lambda_n\times\overline{\mathcal{N}}_n^{\delta}(\eta)} n^{-1}\beta_{0n}'X_n'T_n'(\lambda)M_n(\gamma)T_n(\lambda)X_n\beta_{0n}/\left\|\beta_{0n}\right\|^2 > 0.$$

Theorem 4.1. *Let Assumptions 1-7, 13-15 and 21 hold, and p_n, k_n be allowed to diverge as $n \to \infty$ such that*

$$\frac{k_n}{n} \longrightarrow 0, \text{ as } n \to \infty. \tag{4.2}$$

Then

$$\left\|\hat{\theta}_n - \theta_{0n}\right\| \xrightarrow{p} 0, \text{ as } n \to \infty.$$

Under similar regularity conditions as in the previous section we may obtain asymptotic distributions of $\hat{\theta}_n = \left(\hat{\lambda}_n', \hat{\beta}_n'\right)'$, with formulae for asymptotic covariance matrices adjusted accordingly, but there is a key finding for the case where h_n diverges. We provide the derivatives in Appendix A, from where

$$\Xi_n = 2\sigma_0^{-2}n^{-1}\begin{bmatrix} \sigma_0^2(P_{1n}+P_{2n}) & 0 & 0 \\ * & X_n'S_n'S_nX_n & X_n'S_n'\Pi_n \\ * & * & \Pi_n'\Pi_n \end{bmatrix}, \tag{4.3}$$

which is block diagonal between λ and $(\beta', \delta')'$ and, notably, the top left block can have spectral norm going to zero when $h_n \to \infty$ because it is identical to (2.9), which entailed a different norming in Theorems 2.3 and 2.6.

Assumption 22. *For some $\epsilon > 0$, $\partial x_{ijn}(\gamma)/\partial\gamma_l$ exist and are uniformly bounded in absolute value for all $\gamma \in \mathcal{N}^{\gamma}(\epsilon)\cap\Gamma$, $i = 1\ldots, n$, $j = 1,\ldots,k_n$, $l = 1,\ldots,p_n+q$. As $n \to \infty$, $\overline{\lim}_{n\to\infty} n^{-1}\bar{\zeta}(X_n'X_n) < \infty$.*

Assumption 23. *For some $\epsilon > 0$, $\partial^2 x_{ijn}(\gamma)/\partial\gamma_{l_1}\partial\gamma_{l_2}$ and $\partial^3 x_{ijn}(\gamma)/\partial\gamma_{l_1}\partial\gamma_{l_2}\partial\gamma_{l_3}$ exist and are uniformly bounded in absolute value for all $\gamma \in \mathcal{N}^{\gamma}(\epsilon)\cap\mathcal{D}$, $i = 1,\ldots,n$, $j = 1,\ldots,k_n$, $l_1, l_2, l_3 = 1,\ldots,p_n+q$. As $n \to \infty$,*

$$\overline{\lim}_{n\to\infty} n^{-1} \max_{l=1,,\ldots,p_n+q} \bar{\zeta}\left\{(\partial X_n'/\partial\gamma_l)(\partial X_n/\partial\gamma_l)\right\} < \infty, \tag{4.4}$$

$$\overline{\lim}_{n\to\infty} n^{-1} \max_{l_1,l_2=1,,\ldots,p_n+q} \bar{\zeta}\left\{\left(\partial^2 X_n'/\partial\gamma_{l_1}\partial\gamma_{l_2}\right)\left(\partial^2 X_n/\partial\gamma_{l_1}\partial\gamma_{l_2}\right)\right\} < \infty. \tag{4.5}$$

In the two central limit theorems stated below identification conditions are taken to hold for the changed definitions of Ξ_n and Ω_n in this section. The latter is described in Appendix A, but the key feature of the next theorem is the differential norming that implies a slower rate of convergence for $\hat{\lambda}_n$ as compared to $\left(\hat{\beta}_n', \hat{\delta}_n'\right)'$. Define $\Phi_n = diag\left[h_nI_{p_n}, I_{k_n}, I_q\right]$ and write $B_n^{\Phi} = \Phi_n^{\frac{1}{2}}B_n\Phi_n^{\frac{1}{2}}$ for a generic matrix B_n.

Theorem 4.2. *Let $h_n \to \infty$ as $n \to \infty$, $h_n^{1+\frac{4}{\chi}}/n \to 0$ as $n \to \infty$, Assumptions 1, 2, 4-8, 11, 12, 14, 15 and 21-23 hold, δ_0 be in the interior of $\mathcal{D}$, $\overline{\lim}_{n\to\infty} \overline{\zeta}\left(\Xi_n^{\Phi-1}\Omega_n^{\Phi}\Xi_n^{\Phi-1}\right) < \infty$, $\underline{\lim}_{n\to\infty} \underline{\zeta}\left(\Xi_n^{\Phi}\right) > 0$, $\underline{\lim}_{n\to\infty} \underline{\zeta}\left(2\Xi_n^{\Phi-1} + \Xi_n^{\Phi-1}\Omega_n^{\Phi}\Xi_n^{\Phi-1}\right) > 0$, and (2.11), (2.12), (3.14) hold if p_n, k_n are allowed to diverge as $n \to \infty$. Then*

$$\frac{n^{\frac{1}{2}}}{(p_n + k_n)^{\frac{1}{2}}} \Psi_n \Phi_n^{-\frac{1}{2}} \left(\hat{\theta}_n - \theta_{0n}\right) \xrightarrow{d} N\left(0, \Delta_6\right), \text{ as } n \to \infty,$$

where $\Delta_6 = \lim_{n\to\infty} (p_n + k_n)^{-1} \Psi_n \left(2\Xi_n^{\Phi-1} + \Xi_n^{\Phi-1}\Omega_n^{\Phi}\Xi_n^{\Phi-1}\right) \Psi_n'$.

Theorem 4.3. *Let h_n be bounded as $n \to \infty$, Assumptions 1, 2, 4-8, 11, 12, 14, 15 and 20-23 hold, δ_0 be in interior of $\mathcal{D}$, and (3.16) hold if p_n, k_n are allowed to diverge as $n \to \infty$. Then*

$$\frac{n^{\frac{1}{2}}}{(p_n + k_n)^{\frac{1}{2}}} \Psi_n \left(\hat{\theta}_n - \theta_{0n}\right) \xrightarrow{d} N\left(0, \Delta_7\right), \text{ as } n \to \infty,$$

where $\Delta_7 = \lim_{n\to\infty} (p_n + k_n)^{-1} \Psi_n \left(2\Xi_n^{-1} + \Xi_n^{-1}\Omega_n\Xi_n^{-1}\right) \Psi_n'$.

Covariance matrix estimation follows in much the same manner as Section 3. σ_0^2 is estimated by $\overline{\sigma}_n^2\left(\hat{\gamma}_n\right)$, while higher moments in Ω_n are estimated by empirical counterparts.

5 Finite-sample performance

In this section we study the finite-sample properties of the estimates considered above in a Monte Carlo study, in two distinct settings considered earlier e.g. in Gupta and Robinson (2015). In the first setting, from Case (1991, 1992) take the 'single nonzero diagonal block' specification

$$W_{kn}^f = diag\left[0, \ldots, \underbrace{V_m}_{k-th \text{ diagonal block}}, \ldots, 0\right], \; k = 1, \ldots, p, \tag{5.1}$$

with $V_m = (m-1)^{-1}\left(l_m l_m' - I_m\right)$. In the second setting these were taken to be

$$W_{kn}^c = \left(\|W_{kn}^*\|\right)^{-1} W_{kn}^*, \tag{5.2}$$

with W_{kn}^* the symmetric circulant matrix with first row elements given by

$$w_{1j,kn}^* = \begin{cases} 0 & \text{if } j = 1 \text{ or } j = k+2, \ldots, n-k; \\ 1 & \text{if } j = 2, \ldots, k+1 \text{ or } j = n-k+1, \ldots, n. \end{cases} \tag{5.3}$$

Thus W_{kn}^c is also a symmetric circulant matrix with first row elements given by $w_{1j,kn}^*/2k$. In both experiments we took $p = 2, 4, 6$. We first analyse the pure and intercept SAR cases. y_n was generated using (1.1) or (1.2) in each of the 1000 replications. We chose $\lambda_{01} = 0.7, \lambda_{02} =$

$u \sim N(0,1)$							
n		108		216		432	
	p	Bias	MSE	Bias	MSE	Bias	MSE
W^c_{kn}	2	0.0169	0.0267	0.0036	0.0138	0.0017	0.0069
	4	0.0464	0.1300	0.0592	0.0861	0.0181	0.0404
	6	0.0449	0.2325	0.1068	0.2298	0.0284	0.1058
s		12		24		36	
W^f_{kn}	2	0.0396	0.0132	0.0177	0.0040	0.0114	0.0023
	4	0.1047	0.0710	0.0453	0.0198	0.0288	0.0105
	6	0.2017	0.1982	0.0962	0.0703	0.0593	0.0352
$u \sim t_6$							
n		108		216		432	
	p	Bias	MSE	Bias	MSE	Bias	MSE
W^c_{kn}	2	0.0141	0.0274	0.0026	0.0135	0.0012	0.0069
	4	0.0501	0.1277	0.0499	0.0880	0.0121	0.0364
	6	0.0350	0.2296	0.0917	0.2189	0.0356	0.1099
s		12		24		36	
W^f_{kn}	2	0.0343	0.0114	0.0178	0.0040	0.0108	0.0023
	4	0.0991	0.0685	0.0441	0.0180	0.0262	0.0093
	6	0.2001	0.2014	0.0923	0.0661	0.0574	0.0336

Table 5.1: Monte Carlo (average) bias and (average) MSE for PMLE, model (1.1)

$0.8, \lambda_{03} = 0.5, \lambda_{04} = 0.8, \lambda_{05} = 0.4$ and $\lambda_{06} = 0.3$, when using W^f_{kn} while the values chosen when using W^c_{kn} were $\lambda_{01} = 0.1, \lambda_{02} = 0.2, \lambda_{03} = 0.2, \lambda_{04} = 0.1, \lambda_{05} = 0.1$ and $\lambda_{06} = 0.2$ (because a sufficient condition for S_n^{-1} to exist in this case is $\sum_{i=1}^{p_n} |\lambda_i| < 1$). One set of u_i was generated from $N(0,1)$ (here PMLE is MLE), and another set from t_6 ($\sigma_0^2 = 3/2$), having thicker tails.

Tables 5.1 and 5.2 display Monte Carlo (absolute) bias and MSE for (1.1) and (1.2) respectively, with $\tau_0 = 1$. Table 5.2 considers only W^c_{kn}, the inclusion of an intercept not being possible with W^f_{kn} (cf Kelejian, Prucha, and Yuzefovich (2006)). Averages (averaging over bias and MSE for λ_{0i}, $i = 1, \ldots, p$) are reported for the spatial parameter estimates to conserve space, and in fact all subsequent tables will report such average statistics for the λ_{0i}. We report results for $m = 16$ ($m = 8, 24$ were also simulated) only when using W^f_{kn}, and also take the number of districts s_n (implying $n = 16s_n$) to grow faster than p_n. Indeed Theorems 2.3 and 2.4 indicate that when m is either bounded or divergent the PMLE is $s_n^{\frac{1}{2}}/p_n^{\frac{1}{2}}$-consistent for the farmer-district setting. We take $s = 12, 24, 36$, and this implies the need to combine spatial weight matrices by imposing the same spatial parameter for some blocks. Combinations are made according to equal numbers of blocks. When using W^c_{kn} we took $n = 108, 216, 432$.

In Table 5.1, bias and MSE decline with sample size for both MLE and PMLE of (1.1), using

$u \sim N(0,1)$	n	108		216		432	
p		Bias	MSE	Bias	MSE	Bias	MSE
2	λ	0.0275	0.0277	0.0088	0.0141	0.0043	0.0069
	τ	0.0386	0.0420	0.0181	0.0192	0.0079	0.0092
4	λ	0.0445	0.1327	0.0607	0.0844	0.0177	0.0404
	τ	0.4455	1.7600	0.4286	1.5636	0.1772	0.6106
6	λ	0.0356	0.2187	0.0856	0.2115	0.0272	0.1030
	τ	1.3373	7.0599	1.4352	5.7450	0.5206	2.0209
$u \sim t_6$							
p		Bias	MSE	Bias	MSE	Bias	MSE
2	λ	0.0253	0.0286	0.0080	0.0137	0.0027	0.0069
	τ	0.0384	0.0468	0.0207	0.0225	0.0093	0.0107
4	λ	0.0433	0.1270	0.0511	0.0865	0.0099	0.0350
	τ	0.3932	1.5885	0.3694	1.3702	0.0952	0.3675
6	λ	0.0370	0.2233	0.0715	0.1998	0.0271	0.1044
	τ	1.3176	7.3067	1.5214	6.1180	0.5600	2.1321

Table 5.2: Monte Carlo bias and MSE for PMLE, model (1.2), with W_{kn}^c only. Average bias and MSE reported for λ_i.

either W_{kn}^c or W_{kn}^f, although with the former the decline in bias is not necessarily monotonic. Generally biases for W_{kn}^f exceed those for W_{kn}^c but MSEs tend to be smaller, indicating that variances are smaller. Table 5.2 indicates a similar, non-monotonic, pattern of reduction for (1.2). However the bias and MSE for $\hat{\tau}_n$ can be very high for large p, e.g. for $p = 6$ the bias and MSE are not acceptable even when $n = 432$.

Tables 5.3 and 5.4 similarly display Monte Carlo size and power for (1.1) and (1.2) respectively. The sizes should be compared with the nominal 5%. Power was computed using the false null hypothesis $\lambda_i, \tau = 0.5$, for each i. With W_{kn}^c in (1.1), size approaches the nominal value non-monotonically with n, but with (1.2) the behaviour is rather more erratic. For $p = 2$ the oversizing is moderate, but dramatically worsens for $p = 4, 6$. However in each case it gets closer to the nominal size as the sample size increases, although not necessarily monotonically. On the contrary, with W_{kn}^f there is considerable undersizing. Greater values of s do not give much indication of an approach to the nominal 5%. The sizes are closer to the nominal value for larger values of p, the best results arising when $p = 6$. The behaviour is not too different according to whether $N(0, 1)$ or t_6 disturbances are employed. On the other hand power increases monotonically in each of the various settings, and would be much higher for the $p = 4, 6$ cases if not for the dilution due to $\lambda_{03} = 0.5$, which effectively caps power at around 83%. Nevertheless substantial improvements can be noted as n and s increase, according to the weight matrices being employed.

When generating y_n using (1.3), we set $k_n = 2$ and $\beta_{01} = 1$, $\beta_{02} = 0.7$. In X_n we took

$u \sim N(0,1)$

n		108		216		432	
	p	Size	Power	Size	Power	Size	Power
W^c_{kn}	2	0.0475	0.5800	0.0460	0.7935	0.0400	0.9470
	4	0.0660	0.2715	0.0998	0.4047	0.0760	0.5830
	6	0.0335	0.1860	0.1197	0.3043	0.1027	0.4040
s		12		24		36	
W^f_{kn}	2	0.0085	0.5835	0.0060	0.7805	0.0060	0.8855
	4	0.0103	0.3520	0.0083	0.4925	0.0073	0.5867
	6	0.0300	0.2035	0.0242	0.2807	0.0215	0.3422

$u \sim t_6$

n		108		216		432	
	p	Size	Power	Size	Power	Size	Power
W^c_{kn}	2	0.0560	0.5695	0.0425	0.7970	0.0490	0.9480
	4	0.0583	0.2745	0.0985	0.4233	0.0660	0.5775
	6	0.0313	0.1818	0.1148	0.3062	0.1078	0.4085
s		12		24		36	
W^f_{kn}	2	0.0090	0.5910	0.0060	0.7755	0.0070	0.8770
	4	0.0123	0.3602	0.0067	0.4080	0.0070	0.5900
	6	0.0303	0.2087	0.0230	0.2855	0.0237	0.3480

Table 5.3: Monte Carlo average size and average power for PMLE, model (1.1)

$x_{i1n}(\delta) = (z_{i1}^{\delta} - 1)/\delta$ and $x_{i2n}(\delta) = z_{i2}$, with $(z_{i1}, z_{i2})' \sim U(0,5)$, $i = 1, \ldots, n$, generating these in each of the 1000 replications, and $\delta_0 = 0.7$. In previous versions of the paper we permitted stochastic $X_n(\delta)$ that were independent of u, but this only led to more complicated, not more illuminating, conditions. When using W^f_{kn} equal blocks of size m were used, while three different values of m were chosen for each value of p: 48, 96 and 144. We also simulated a model with $x_{i1n}(\delta) = \mathrm{e}^{\delta z_{i1}}$ and $x_{i2n}(\delta) = z_{i2}$ and found similar results to those reported below.

We now discuss the results for $\hat{\theta}_n$ in Tables 5.5 and 5.6, which report Monte Carlo bias and MSE for $u \sim N(0,1)$ and $u \sim t_6$ respectively. It is interesting to note that for W^f_{kn} increasing m improves the estimates of the spatial parameters in most cases, for fixed p. However, Lee (2004) showed that the PMLE is inconsistent if $p = 1$ while m alone increases, while simulations conducted by Hillier and Martellosio (2013) also correspond with the convergence to a nondegenerate distribution often arising from 'infill asymptotics', see e.g. Lahiri (1996). Similar results will undoubtedly apply if $p > 1$, but fixed, and m alone increases. On the other hand, the block-diagonality of W^f_{kn} implies that the number of observations available to estimate the λ_{0i} increase one-to-one with m. Generally bias and MSE improve with sample size, as expected. For W^c_{kn} the results are as expected. Bias and MSE reduce with larger n and smaller p, and also

$u \sim N(0,1)$	n	108		216		432	
p		Size	Power	Size	Power	Size	Power
2	λ	0.0275	0.5950	0.0088	0.8055	0.0043	0.9505
	τ	0.0750	0.8460	0.0560	0.9920	0.0570	1.0000
4	λ	0.0445	0.2742	0.0607	0.4072	0.0177	0.5837
	τ	0.2630	0.7070	0.2000	0.9320	0.1260	1.0000
6	λ	0.0356	0.1817	0.0856	0.2982	0.0272	0.4038
	τ	0.4050	0.4960	0.5330	0.7830	0.2960	0.9110
$u \sim t_6$							
p		Size	Power	Size	Power	Size	Power
2	λ	0.0253	0.5885	0.0080	0.8090	0.0027	0.9540
	τ	0.0540	0.7950	0.0630	0.9720	0.0610	1.0000
4	λ	0.0433	0.2785	0.0511	0.4265	0.0099	0.5825
	τ	0.2360	0.6210	0.1840	0.9100	0.0890	0.9920
6	λ	0.0370	0.1778	0.0715	0.2968	0.0271	0.4060
	τ	0.4090	0.4820	0.5540	0.7680	0.2970	0.9920

Table 5.4: Monte Carlo size and power for PMLE, model (1.2), with W^c_{kn} only. Average size and power reported for λ_i.

with larger n for fixed p. The values are small and seem acceptable.

Tables 5.7 and 5.8 report Monte Carlo size and power for $u \sim N(0,1)$ and $u \sim t_6$ respectively. Now power is calculated using the incorrect null hypothesis $\theta_i = 0.6$, for each i. For the MLE, sizes when using W^f_{kn} are always between 3.7% and 6.7% but those for W^c_{kn} range from 0.78% to 4.8%, with the best results for $n = 432$ where they range from 2.32% to 4.8%. Matters are much worse for the PMLE, where oversizing persists for both W^c_{kn} and W^f_{kn} no matter the values of p, n. On the other hand, for both MLE and PMLE, the power tends to increase (but not always monotonically) with large n and small p for W^c_{kn} but large m, p, for W^f_{kn}, due to the increase in sample size afforded by increasing p in this setting. Power for δ_0 tends to be low across the board, due in part to the proximity of its true value to the postulated value. This factor doubtless also plays a role in the lower power for β_{02} generally as compared to that for β_{01}.

Finally, Table 5.9 compares $\hat{\theta}_n$ with the IV estimate of Gupta and Robinson (2015) (denoted $\check{\theta}_n$) when W^c_{kn} are employed, $x_{i1n}(\delta) = z_{i1}$ also (i.e. linear regressive SAR) and $(z_{i1}, z_{i2}) \sim U(0,1)$ to match their design. Both $u \sim N(0,1)$ and $u \sim t_6$ are considered. We report relative average MSE (RAMSE) separately for the autoregression and regression components, defining these as average MSE($\hat{\lambda}_n$)/average MSE($\check{\lambda}_n$) and average MSE($\hat{\beta}_n$)/average MSE($\check{\beta}_n$), using the instruments $\{W^c_{jn} z_{i1}, W^c_{jn} z_{i2}\}$, $j = 1, \ldots, p$. The PMLE does very well in general. The IV estimates outperform the PMLE for the regression coefficients β_{01} and β_{02} in 4 out of 6 cases when $p = 6$, but fare much worse for the spatial parameters in all cases. Experiments in which the u were generated from a $\chi^2_6 - 6$ (this having $\sigma^2_0 = 12$, and also being non-symmetric) distribution

W^c_{kn}	n	108		216		432	
p		Bias	MSE	Bias	MSE	Bias	MSE
2	λ	0.0036	0.0110	0.0009	0.0053	0.0003	0.0028
	δ	0.0184	0.0462	0.0212	0.0220	0.0016	0.0089
	β_1	0.0135	0.0238	0.0150	0.0116	0.0025	0.0049
	β_2	0.0018	0.0038	0.0002	0.0017	0.0002	0.0010
4	λ	0.0085	0.0546	0.0029	0.0268	0.0028	0.0132
	δ	0.0176	0.0470	0.0215	0.0222	0.0018	0.0090
	β_1	0.0172	0.0242	0.0171	0.0116	0.0037	0.0049
	β_2	0.0048	0.0043	0.0010	0.0020	0.0006	0.0011
6	λ	0.0073	0.1186	0.0070	0.0648	0.0069	0.0312
	δ	0.0194	0.0492	0.0221	0.0227	0.0023	0.0092
	β_1	0.0220	0.0251	0.0196	0.0118	0.0050	0.0050
	β_2	0.0068	0.0046	0.0024	0.0021	0.0009	0.0012
W^f_{kn}	m	48		96		144	
p		Bias	MSE	Bias	MSE	Bias	MSE
2	λ	0.0009	0.0002	0.0018	0.0003	0.0010	0.0002
	δ	0.0038	0.0214	0.0090	0.0232	0.0090	0.0156
	β_1	0.0083	0.0264	0.0026	0.0129	0.0055	0.0078
	β_2	0.0042	0.0052	0.0066	0.0025	0.0032	0.0019
4	λ	0.0044	0.0006	0.0020	0.0003	0.0012	0.0002
	δ	0.0129	0.0230	0.0070	0.0117	0.0023	0.0074
	β_1	0.0022	0.0129	0.0036	0.0061	0.0001	0.0040
	β_2	0.0109	0.0026	0.0059	0.0013	0.0022	0.0008
6	λ	0.0059	0.0010	0.0027	0.0005	0.0020	0.0003
	δ	0.0142	0.0158	0.0037	0.0074	0.0046	0.0050
	β_1	0.0046	0.0079	0.0000	0.0040	0.0014	0.0028
	β_2	0.0087	0.0020	0.0039	0.0008	0.0028	0.0006

Table 5.5: Monte Carlo bias and MSE for MLE ($u \sim N(0,1)$), model (1.3) with $x_{i1n}(\delta) = (z_{i1}^{\delta} - 1)/\delta$. Average bias and MSE reported for λ_i.

were also carried out and the results follow the same pattern.

W^c_{kn}	n	108		216		432	
p		Bias	MSE	Bias	MSE	Bias	MSE
2	λ	0.0024	0.0134	0.0010	0.0068	0.0033	0.0031
	δ	0.0263	0.0708	0.0148	0.0304	0.0076	0.0143
	β_1	0.0183	0.0358	0.0112	0.0170	0.0053	0.0074
	β_2	0.0028	0.0057	0.0010	0.0026	0.0018	0.0012
4	λ	0.0110	0.0622	0.0112	0.0319	0.0069	0.0165
	δ	0.0250	0.0728	0.0165	0.0308	0.0080	0.0145
	β_1	0.0229	0.0366	0.0143	0.0171	0.0067	0.0074
	β_2	0.0059	0.0066	0.0018	0.0030	0.0013	0.0014
6	λ	0.0140	0.1371	0.0112	0.0743	0.0046	0.0390
	δ	0.0261	0.0748	0.0159	0.0308	0.0083	0.0147
	β_1	0.0283	0.0371	0.0169	0.0173	0.0081	0.0074
	β_2	0.0092	0.0071	0.0035	0.0033	0.0007	0.0015
W^f_{kn}	m	48		96		144	
p		Bias	MSE	Bias	MSE	Bias	MSE
2	λ	0.0036	0.0009	0.0030	0.0004	0.0009	0.0003
	δ	0.0384	0.0881	0.0239	0.0378	0.0072	0.0238
	β_1	0.0132	0.0411	0.0133	0.0195	0.0041	0.0125
	β_2	0.0079	0.0084	0.0098	0.0041	0.0048	0.0026
4	λ	0.0070	0.0009	0.0036	0.0005	0.0018	0.0003
	δ	0.0291	0.0380	0.0165	0.0180	0.0086	0.0102
	β_1	0.0117	0.0196	0.0077	0.0096	0.0026	0.0055
	β_2	0.0159	0.0042	0.0070	0.0019	0.0023	0.0013
6	λ	0.0083	0.0015	0.0041	0.0007	0.0039	0.0005
	δ	0.0142	0.0237	0.0097	0.0102	0.0119	0.0082
	β_1	0.0024	0.0127	0.0018	0.0055	0.0037	0.0040
	β_2	0.0130	0.0027	0.0045	0.0013	0.0061	0.0009

Table 5.6: Monte Carlo bias and MSE for PMLE ($u \sim t_6$), model (1.3) with $x_{i1n}(\delta) = (z_{i1}^{\delta} - 1)/\delta$. Average bias and MSE reported for λ_i.

W^c_{kn}	n	108		216		432	
p		Size	Power	Size	Power	Size	Power
2	λ	0.0210	0.9425	0.0210	0.9985	0.0245	1.0000
	δ	0.0380	0.0390	0.0440	0.0830	0.0410	0.1270
	β_1	0.0480	0.6720	0.0420	0.9100	0.0400	0.9990
	β_2	0.0300	0.2970	0.0270	0.5970	0.0480	0.8920
4	λ	0.0173	0.4608	0.0240	0.6870	0.0233	0.8748
	δ	0.0400	0.0350	0.0410	0.0820	0.0400	0.1200
	β_1	0.0380	0.6540	0.0370	0.9020	0.0390	0.9980
	β_2	0.0320	0.2500	0.0330	0.5280	0.0440	0.8220
6	λ	0.0078	0.2683	0.0203	0.4325	0.0232	0.6443
	δ	0.0450	0.0340	0.0390	0.0780	0.0380	0.1180
	β_1	0.0420	0.6310	0.0310	0.8950	0.0400	0.9980
	β_2	0.0300	0.2200	0.0330	0.4850	0.0440	0.8090
W^f_{kn}	m	48		96		144	
p		Size	Power	Size	Power	Size	Power
2	λ	0.0480	1.0000	0.0515	0.9995	0.0585	1.0000
	δ	0.0600	0.1000	0.0440	0.0660	0.0430	0.1010
	β_1	0.0590	0.7620	0.0600	0.9160	0.0370	0.9770
	β_2	0.0470	0.2860	0.0480	0.5520	0.0670	0.7000
4	λ	0.0480	0.9523	0.0417	0.9963	0.0535	1.0000
	δ	0.0410	0.0630	0.0620	0.1160	0.0510	0.1680
	β_1	0.0610	0.9120	0.0530	0.9950	0.0540	1.0000
	β_2	0.0450	0.5720	0.0480	0.8430	0.0480	0.9450
6	λ	0.0530	0.9843	0.0492	0.9993	0.0512	1.0000
	δ	0.0380	0.1050	0.0490	0.1750	0.0500	0.2680
	β_1	0.0340	0.9760	0.0540	1.0000	0.0560	1.0000
	β_2	0.0600	0.7370	0.0470	0.9520	0.0610	0.9860

Table 5.7: Monte Carlo size and power for MLE ($u \sim N(0,1)$), model (1.3) with $x_{i1n}(\delta) = (z_{i1}^{\delta} - 1)/\delta$. Average size and power reported for λ_i.

W^c_{kn}	n	108		216		432	
p		Size	Power	Size	Power	Size	Power
2	λ	0.1130	0.9665	0.1275	0.9985	0.1095	1.0000
	δ	0.1750	0.1980	0.1660	0.2230	0.1740	0.3020
	β_1	0.1820	0.7380	0.1890	0.9150	0.1740	0.9970
	β_2	0.1690	0.4530	0.1580	0.7160	0.1490	0.9250
4	λ	0.1010	0.6020	0.1150	0.7918	0.1220	0.9288
	δ	0.1760	0.1850	0.1650	0.2240	0.1730	0.3030
	β_1	0.1720	0.7170	0.1910	0.9090	0.1700	0.9970
	β_2	0.1710	0.4130	0.1650	0.6570	0.1530	0.8940
6	λ	0.0785	0.4048	0.1040	0.5797	0.1157	0.7768
	δ	0.1660	0.1790	0.1550	0.2040	0.1690	0.3030
	β_1	0.1620	0.6900	0.1940	0.9030	0.1660	0.9960
	β_2	0.1640	0.3610	0.1680	0.6310	0.1540	0.8740
W^f_{kn}	m	48		96		144	
p		Size	Power	Size	Power	Size	Power
2	λ	0.1775	0.9610	0.2000	0.9985	0.1810	1.0000
	δ	0.1850	0.1810	0.1830	0.2320	0.1950	0.2620
	β_1	0.1900	0.7060	0.1720	0.9000	0.1860	0.9750
	β_2	0.1980	0.4470	0.1990	0.6510	0.1940	0.7630
4	λ	0.1960	0.9625	0.1975	0.9958	0.2055	0.9998
	δ	0.1730	0.2300	0.1880	0.3330	0.1670	0.3700
	β_1	0.1830	0.9020	0.1870	0.9940	0.1560	1.0000
	β_2	0.2000	0.6800	0.1980	0.8750	0.2020	0.9370
6	λ	0.1882	0.9878	0.1938	0.9993	0.1922	1.0000
	δ	0.1860	0.2640	0.1560	0.3720	0.2020	0.4730
	β_1	0.1820	0.9740	0.1490	1.0000	0.2070	1.0000
	β_2	0.2020	0.8080	0.2060	0.9450	0.2050	0.9940

Table 5.8: Monte Carlo size and power for PMLE ($u \sim t_6$), model (1.3) with $x_{i1n}(\delta) = (z_{i1}^{\delta}-1)/\delta$. Average size and power reported for λ_i.

	n	108	216	432	108	216	432
p			$u \sim N(0,1)$			$u \sim t_6$	
2	λ	0.0472	0.0488	0.0507	0.0362	0.0287	0.0284
	β	0.5212	0.5554	0.6202	0.4931	0.5028	0.5649
4	λ	0.0339	0.0413	0.0399	0.0239	0.0231	0.0233
	β	0.4152	0.4706	0.5404	0.4630	0.4022	0.4357
6	λ	0.0353	0.0683	0.0601	0.0300	0.0536	0.0382
	β	0.8069	3.5825	1.5249	0.9315	3.4552	1.3950

Table 5.9: Monte Carlo RAMSE between PMLE and IV with W^c_{kn}, $(z_{i1}, z_{i2}) \sim U(0,1)$.

Appendices

A Proofs of theorems

Proof of Theorem 2.1. This is omitted as it can be deduced from the proof of Theorem 3.1 below, ignoring components of formulae and steps that are not relevant. $\square$

Proof of Theorem 2.2. In supplementary material. $\square$

We drop n subscripts in the appendices. The following inequalities will be useful: $\|A\| \leq \|A\|_F$, $\|A\|^2 \leq \|A\|_R \|A'\|_R$, $\|AB\|_F \leq \|A\| \|B\|_F$. In the sequel write $\nu = n^{\frac{1}{2}}/a^{\frac{1}{2}}$, where a is the number of columns in Ψ. Thus in Section 2, $a = p$ or $p+1$, in Section 3, $a = p+k+q$ and in Section 4, $a = p+k$. Further, for any matrix, vector $E\left(\theta, \sigma^2\right)$, $\tilde{E}$ denotes evaluation at a generic estimate $\left(\tilde{\theta}', \tilde{\sigma}^2\right)'$ and $\tilde{\Delta}^E = \tilde{E} - E$. We can express (1.3) as $y = R\lambda_0 + X\beta_0 + u$ with $R = [W_1 y, \ldots, W_p y]$. Because Assumption 3 implies

$$y = S^{-1} X \beta_0 + S^{-1} u, \tag{A.1}$$

we have $R = A + B$, with $B = [G_{1n} u, \ldots, G_{p_n n} u]$, and for (1.1) the reduced form (A.1) holds with $X = 0$. The proofs of Theorems 3.3 and 3.4 should be read before the next two proofs, which we present at this point to follow the order of the paper, for descriptions of notation and more details.

Proof of Theorem 2.3. For any non-null $m \times 1$ vector of constants α, $\hat{\xi} = 0$ and the MVT imply $\nu h^{-\frac{1}{2}} \alpha' \Psi \left(\hat{\lambda} - \lambda_0\right) = l_1 + l_2 - \nu h^{\frac{1}{2}} \alpha' \Psi \left(h\Xi\right)^{-1} \phi$, where

$$\begin{aligned} l_1 &= \nu h^{\frac{1}{2}} \alpha' \Psi \left(hH\right)^{-1} h\bar{\Delta}^H \left(h\bar{H}\right)^{-1} \phi = \mathscr{O}_p\left(n^{-\frac{1}{2}} p^{\frac{5}{2}}\right), && \text{(A.2)} \\ l_2 &= \nu h^{\frac{1}{2}} \alpha' \Psi \left(h\Xi\right)^{-1} \left(hH - h\Xi\right) \left(hH\right)^{-1} \phi = \mathscr{O}_p\left(\max\left\{n^{-\frac{1}{2}} p^{\frac{3}{2}}, p^{\frac{1}{2}} h/n\right\}\right), && \text{(A.3)} \end{aligned}$$

by Lemmas B.1 (ii), B.2 (i) and B.3 (ii), both being negligible by (2.11). Indeed, the negligibility of l_1 and the first term in braces on the far right of (A.3) follow easily from part (a) of (2.11). The second term in braces on the right of (A.3) is negligible by the condition (c) in (2.11). If instead (b) holds we can write this term as $\left(p^{\frac{5}{2}}/n^{\frac{1}{2}}\right)\left(h^{\frac{1}{2}}/n^{\frac{1}{2}}\right)\left(h^{\frac{1}{2}}/p^2\right) \to 0$. Thus consider

$$-n^{\frac{1}{2}} \varrho^{-1} h^{\frac{1}{2}} \alpha' \Psi \left(h\Xi\right)^{-1} \phi,$$

with $\varrho = \left\{\alpha' \Psi \left(2\left(h\Xi\right)^{-1} + \left(h\Xi\right)^{-1} h\Omega \left(h\Xi\right)^{-1}\right) \Psi' \alpha\right\}^{\frac{1}{2}}$. This can be written as a sum of martingale differences, as in the proof of Theorem 3.4. The arguments thereafter are identical except for changes in stochastic orders due to the different norming and the additional condition (2.12).

The latter implies that

$$\sum_{j=1}^{p}\|C_j\|_R^2 \le C\sum_{j=1}^{p}\left(\|W_j\|_R^2+\left\|W_j'\right\|_R^2\right)\le C, \quad \sum_{j=1}^{p}\|C_j\|_F^2 \le C\left\|S^{-1}\right\|_F^2\sum_{j=1}^{p}\|W_j\|_F^2\le C, \tag{A.4}$$

whence

$$\|D\|^2 \le \|D\|_R^2 \le C\left\|\Psi'\alpha\right\|^2\sum_{j=1}^{p}\|C_j\|_R^2 \le C\left\|\Psi'\alpha\right\|^2 \tag{A.5}$$

and

$$\|D\|_F^2 \le C\left\|\Psi'\alpha\right\|^2\sum_{j=1}^{p}\|C_j\|_F^2 \le C\left\|\Psi'\alpha\right\|^2. \tag{A.6}$$

Using (A.4)-(A.6) eliminates the factors involving p in (A.38), (A.39), (A.41)-(A.43). Next, $p^{-1}\left\|\Psi\left(h\Xi\right)^{-1}h\Omega\left(h\Xi\right)^{-1}\Psi'\right\| \le Cp/h = o(1)$ because $\|\Omega\| \le C\|F\|^2/n = \mathscr{O}\left(p/h^2\right)$ and by Assumption 9, which also guarantees that the asymptotic covariance matrix exists and is positive definite. The proof of Theorem 2.6 is similar and omitted. □

Proof of Theorem 2.4. Again $\nu\alpha'\Psi\left(\hat{\lambda}-\lambda_0\right) = l_1 + l_2 - \nu\alpha'\Psi\Xi^{-1}\phi$ for any non-null $m\times 1$ vector of constants α, where now

$$\begin{aligned} l_1 &= \nu\alpha'\Psi H^{-1}\bar{\Delta}^H\bar{H}^{-1}\phi = \mathscr{O}_p\left(n^{-\frac{1}{2}}h^{-1}p^{\frac{5}{2}}\right), \\ l_2 &= \nu\alpha'\Psi\Xi^{-1}\left(H-\Xi\right)H^{-1}\phi = \mathscr{O}_p\left(n^{-\frac{1}{2}}h^{-1}p^{\frac{3}{2}}\right), \end{aligned}$$

by Lemmas B.1 (ii), B.2 (i) and B.3 (iii), both being negligible by (2.13). The asymptotic distribution of $\nu\alpha'\Psi\Xi^{-1}\phi$ is established as in the proof of Theorem 3.4. The asymptotic covariance matrix exists and is positive definite by Assumption 10. □

Proof of Theorem 2.5. This is omitted for the same reason as Theorem 2.1's proof. □

Proof of Theorem 2.6. This similar to the proof of Theorem 2.3 and therefore omitted. □

Proof of Theorem 3.1. The property $\left\|\hat{\beta}-\beta_0\right\| \xrightarrow{p} 0$ follows using arguments below, the closed form expression (see (3.2)) for $\hat{\beta}$ as a function of $\hat{\gamma}$, and the property $\|\hat{\gamma}-\gamma_0\| \xrightarrow{p} 0$, so we focus on proving the latter From (3.4), (A.1)

$$\begin{aligned} Q\left(\gamma\right)-Q &= \log\overline{\sigma}^2\left(\gamma\right)/\overline{\sigma}^2 - n^{-1}\log\left|T'(\lambda)T(\lambda)\right| \\ &= \log\overline{\sigma}^2\left(\gamma\right)/\sigma^2\left(\lambda\right) - \log\overline{\sigma}^2/\sigma_0^2 + \log r(\lambda), \end{aligned} \tag{A.7}$$

where

$$\sigma^2\left(\lambda\right) = n^{-1}\left\|T(\lambda)\right\|_F^2,\ \overline{\sigma}^2 = \overline{\sigma}^2\left(\gamma_0\right) = n^{-1}u'Mu,$$

using (3.3) and writing $r(\lambda) = n^{-1} \|T(\lambda)\|_F^2 / |T(\lambda)|^{2/n}$. From (A.1)

$$\begin{aligned}
\overline{\sigma}^2(\gamma) &= n^{-1}\left\{S^{-1'}(X\beta_0+u)\right\}' S'(\lambda)M(\delta)S(\lambda)S^{-1}(X\beta_0+u) \\
&= c(\gamma)+d(\gamma)+e(\gamma),
\end{aligned}$$

where

$$\begin{aligned}
c(\gamma) &= n^{-1}\beta_0' X' T'(\lambda) M(\delta) T(\lambda) X\beta_0, \\
d(\gamma) &= n^{-1}\sigma_0^2 tr\left(T'(\lambda)M(\delta)T(\lambda)\right), \\
e(\gamma) &= n^{-1} tr\left(T'(\lambda)M(\delta)T(\lambda)(uu'-\sigma_0^2 I)\right) + 2n^{-1}\beta_0' X' T'(\lambda) M(\delta) T(\lambda) u.
\end{aligned}$$

Then

$$\begin{aligned}
\log\frac{\overline{\sigma}^2(\gamma)}{\sigma^2(\lambda)} &= \log\frac{\overline{\sigma}^2(\gamma)}{(c(\gamma)+d(\gamma))} + \log\frac{c(\gamma)+d(\gamma)}{\sigma^2(\lambda)} \\
&= \log\left(1+\frac{e(\gamma)}{c(\gamma)+d(\gamma)}\right) + \log\left(1+\frac{c(\gamma)-f(\gamma)}{\sigma^2(\lambda)}\right),
\end{aligned}$$

where

$$f(\gamma) = n^{-1}\sigma_0^2 tr\left(T'(\lambda)\left(I-M(\delta)\right)T(\lambda)\right).$$

Then from (A.7) and a standard kind of argument for proving consistency of implicitly defined extremum estimates

$$\begin{aligned}
P\left(\|\hat{\gamma}-\gamma_0\| \in \overline{\mathcal{N}}^{\gamma}(\eta)\right) &= P\left(\inf_{\gamma\in\overline{\mathcal{N}}^{\gamma}(\eta)} Q(\gamma) - Q \le 0\right) \\
&\le P\left(\log\left(1+\sup_{\gamma\in\overline{\mathcal{N}}^{\gamma}(\eta)}\left|\frac{e(\gamma)}{c(\gamma)+d(\gamma)}\right|\right) + \left|\log\left(\overline{\sigma}^2/\sigma_0^2\right)\right|\right. \\
&\qquad \left.\ge \inf_{\gamma\in\overline{\mathcal{N}}^{\gamma}(\eta)}\left(\log\left(1+\frac{c(\gamma)-f(\gamma)}{\sigma^2(\lambda)}\right) + \log r(\lambda)\right)\right),
\end{aligned}$$

where $\overline{\mathcal{N}}^{\gamma}(\eta) = \Gamma\backslash\mathcal{N}^{\gamma}(\eta)$, $\mathcal{N}^{\gamma}(\eta) = \{\gamma : \|\gamma-\gamma_0\| < \eta; \gamma\in\Gamma\}$. From Assumptions 1 and 15 it follows that $\overline{\sigma}^2/\sigma_0^2 \xrightarrow{p} 1$, so using $\log(1+x) = x + o(x)$ as $x\to 0$ it suffices to show that as $n\to\infty$

$$\sup_{\gamma\in\overline{\mathcal{N}}_n^{\gamma}(\eta)}\left|\frac{e(\gamma)}{c(\gamma)+d(\gamma)}\right| \xrightarrow{p} 0, \tag{A.8}$$

$$\sup_{\gamma\in\overline{\mathcal{N}}_n^{\gamma}(\eta)}\left|\frac{f(\gamma)}{\sigma^2(\lambda)}\right| \longrightarrow 0, \tag{A.9}$$

$$\varliminf_{n\to\infty}\inf_{\gamma\in\overline{\mathcal{N}}_n^{\gamma}(\eta)}\left\{\frac{c(\gamma)}{\sigma^2(\lambda)} + \log r(\lambda)\right\} > 0. \tag{A.10}$$

Now $\overline{\mathscr{N}}^{\gamma}(\eta) \subseteq \left\{\Lambda \times \overline{\mathscr{N}}^{\delta}(\eta/2)\right\} \cup \left\{\overline{\mathscr{N}}^{\lambda}(\eta/2) \times \mathcal{D}\right\}$, so

$$\begin{aligned}
\inf_{\gamma \in \overline{\mathscr{N}}^{\gamma}(\eta)} \left\{\frac{c(\gamma)}{\sigma^2(\lambda)} + \log r(\lambda)\right\} &\geq \min\left\{\inf_{\Lambda \times \overline{\mathscr{N}}^{\delta}(\eta/2)} \frac{c(\gamma)}{\sigma^2(\lambda)}, \inf_{\overline{\mathscr{N}}^{\lambda}(\eta/2)} \log r(\lambda)\right\} \\
&\geq \min\left\{\inf_{\Lambda \times \overline{\mathscr{N}}^{\delta}(\eta/2)} \frac{c(\gamma)}{C}, \inf_{\overline{\mathscr{N}}^{\lambda}(\eta/2)} \log r(\lambda)\right\},
\end{aligned}$$

from Assumption 6, whence Assumptions 7 and 16 imply (A.10). Again using Assumption 6, uniformly in γ, $\left|f(\gamma)/\sigma^2(\lambda)\right| \leq |f(\gamma)|/c$ and

$$\begin{aligned}
|f(\gamma)| &\leq Ctr\left(T'(\lambda)X(\delta)\left(X'(\delta)X(\delta)\right)^{-1}X'(\delta)T(\lambda)\right)/n \\
&= \mathscr{O}\left(tr\left(X'(\delta)X(\delta)\right)/n^2\right) = \mathscr{O}(k/n)
\end{aligned}$$

uniformly, by Assumption 14, to check (A.9).

Finally consider (A.8). We first prove pointwise convergence. For any fixed $\gamma \in \overline{\mathscr{N}}^{\gamma}(\eta)$ and large enough n, $c(\gamma) \geq c\|\beta_0\|^2$ from Assumption 16, $d(\gamma) \geq c$ because $n^{-1}\sigma_0^2 tr\left(T'(\lambda)T(\lambda)\right) \geq c$ and $tr\left(T'(\lambda)\left(I - M(\delta)\right)T(\lambda)\right) = \mathscr{O}(k/n)$. Thus $e(\gamma)/\left(c(\gamma) + d(\gamma)\right) = \mathscr{O}_p\left(|e(\gamma)|\right)$, where $e(\gamma)$ has mean 0 and variance

$$\mathscr{O}\left(\|T'(\lambda)M(\delta)T(\lambda)/n\|_F^2 + \sum_{i=1}^{n}\left(t_i'(\lambda)M(\delta)t_i(\lambda)/n\right)^2 + \|\beta_0'X'T'(\lambda)M(\delta)T(\lambda)/n\|^2\right),$$

where $t_i(\lambda)$ is the ith column of $T(\lambda)$. Since $\|M(\delta)\| = 1$ and Assumptions 4 and 12 (we give a bound for the general case, that the same bound holds for the 'single nonzero diagonal block' case is simple to check) imply

$$\|T(\lambda)\| \leq C\|S(\lambda)\| \leq C\sum_{i=1}^{p}|\lambda_i|\,\|W_i\| \leq C\|\lambda\|\left(\sum_{i=1}^{p}\|W_i\|^2\right)^{\frac{1}{2}} = \mathscr{O}(1), \tag{A.11}$$

the first component is $\mathscr{O}\left(\|T(\lambda)/n\|_F^2\right) = \mathscr{O}\left(n^{-1}\right)$. The second one is $\mathscr{O}\left(\sum_{i=1}^{n}\|t_i(\lambda)\|^2/n^2\right) = \mathscr{O}\left(\|T(\lambda)/n\|_F^2\right) = \mathscr{O}\left(n^{-1}\right)$ likewise. The final component is $\mathscr{O}\left(\|X\beta_0/n\|^2\right) = \mathscr{O}\left(\|\beta_0\|^2/n\right) = \mathscr{O}(k/n)$, from (3.6). Thus pointwise convergence is established.

To complete the proof of (A.8) we employ an equicontinuity argument. For arbitrary $\varepsilon > 0$ and finitely many $\gamma_* = \left(\lambda_*', \delta_*'\right)'$, the neighbourhoods $\|\gamma - \gamma_*\| < \varepsilon$ form a sub-cover of the compact Γ. It remains to prove that

$$\sup_{\|\gamma - \gamma_*\| < \varepsilon}\left|\frac{e(\gamma)}{c(\gamma) + d(\gamma)} - \frac{e(\gamma_*)}{c(\gamma_*) + d(\gamma_*)}\right| \xrightarrow{p} 0.$$

Write

$$\frac{e(\gamma)}{c(\gamma)+d(\gamma)}-\frac{e(\gamma_*)}{c(\gamma_*)+d(\gamma_*)}=\frac{e(\gamma)-e(\gamma_*)}{c(\gamma)+d(\gamma)}+e(\gamma_*)\left(\frac{c(\gamma_*)-c(\gamma)+d(\gamma_*)-d(\gamma)}{(c(\gamma)+d(\gamma))(c(\gamma_*)+d(\gamma_*))}\right)$$

whence, denoting the two components of $e(\gamma)$ by $e_1(\gamma)$, $e_1(\gamma)$, the left side is bounded in absolute value by

$$\frac{|e_1(\gamma)-e_1(\gamma_*)|}{d(\gamma)}+\frac{|e_2(\gamma)-e_2(\gamma_*)|}{c(\gamma)}+\frac{|e(\gamma_*)|}{c(\gamma)\,c(\gamma_*)}|c(\gamma_*)-c(\gamma)|+\frac{|e(\gamma_*)|}{d(\gamma)\,d(\gamma_*)}|d(\gamma_*)-d(\gamma)|. \tag{A.12}$$

We prove that

$$\sup_{\|\gamma-\gamma_*\|<\varepsilon}\frac{|e_2(\gamma)-e_2(\gamma_*)|}{c(\gamma)}\xrightarrow{p}0. \tag{A.13}$$

This part of the proof is relatively delicate due to both numerator and denominator increasing with k. The proof for the second term in (A.12) does not involve this feature and uses other arguments in the proof of (A.13). For the third term in (A.12),

$$\frac{|e(\gamma_*)|}{c(\gamma)\,c(\gamma_*)}|c(\gamma_*)-c(\gamma)|\le\frac{|e(\gamma_*)|}{c(\gamma_*)}\left(1+\frac{c(\gamma_*)}{c(\gamma)}\right)\xrightarrow{p}0$$

uniformly on $\|\gamma-\gamma_*\|<\varepsilon$, from the pointwise convergence of $e(\gamma)/(c(\gamma)+d(\gamma))$ and the fact that numerator and denominator of $c(\gamma_*)/c(\gamma)$ are uniformly of the same order of magnitude, namely k, the result for the numerator being straightforward and that for the denominator a consequence of Assumption. The fourth term in (A.12) is uniformly $o_p(1)$ by similar arguments.

To prove (A.13), note that

$$e_2(\gamma)-e_2(\gamma_*)=2n^{-1}\beta_0'\left(X'(\delta)\,T'(\lambda)M(\delta)\,T(\lambda)\right)-X'(\delta_*)\,T'(\lambda_*)M(\delta_*)\,T(\lambda_*))\,u,$$

which can be written

$$\begin{aligned}&2n^{-1}\beta_0'\left\{(X(\delta)-X(\delta_*))'\,T'(\lambda)M(\delta)\,T(\lambda)\right.\\&\left.+X'(\delta_*)\,(T'(\lambda)M(\delta)\,T(\lambda)-T'(\lambda_*)M(\delta_*)\,T(\lambda_*))\right\}u.\end{aligned} \tag{A.14}$$

The first of the two terms in braces has spectral norm bounded by $\|X(\delta)-X(\delta_*)\|\,\|T(\lambda)\|^2$, and by Assumption 15,

$$\|X(\delta)-X(\delta_*)\|^2\le\sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij}(\delta)-x_{ij}(\delta_*))^2=O\left(kn\varepsilon^2\right), \tag{A.15}$$

Thus due to $\|u\|=\mathscr{O}_p\left(n^{1/2}\right)$, it follows that $2n^{-1}\beta_0'\,(X(\delta)-X(\delta_*))'\,T'(\lambda)M(\delta)\,T(\lambda)\,u$ is uniformly $\mathscr{O}_p\left(\|\beta_0\|\,k^{1/2}\varepsilon\right)$. Looking at the second term in braces in (A.14), write $T'(\lambda)M(\delta)\,T(\lambda)-$

$T'(\lambda_*)M(\delta_*)T(\lambda_*)$ as

$$(T(\lambda)-T(\lambda_*))' M(\delta)T(\lambda)+T'(\lambda_*)(M(\delta_*)-M(\delta))T(\lambda)+T'(\lambda_*)M(\delta_*)(T(\lambda)-T(\lambda_*)),$$

whose spectral norm is bounded by

$$\begin{aligned} & \|T(\lambda)-T(\lambda_*)\|(\|T(\lambda)\|+\|T(\lambda_*)\|)+\|T(\lambda_*)\|\|M(\delta_*)-M(\delta)\|\|T(\lambda)\| \\ = \; & \mathscr{O}(\|T(\lambda)-T(\lambda_*)\|+\|M(\delta_*)-M(\delta)\|). \end{aligned} \tag{A.16}$$

Now

$$\begin{aligned} \|T(\lambda)-T(\lambda_*)\| &\le \sum_{i=1}^{p}|\lambda_i-\lambda_{*i}|\|W_i\|\left\|S^{-1}\right\| \\ &\le C\|\lambda-\lambda_*\|\left(\sum_{i=1}^{p}\|W_i\|^2\right)^{1/2}\le C\varepsilon \end{aligned} \tag{A.17}$$

uniformly on $\|\gamma-\gamma_*\|<\varepsilon$. Representing $M(\delta_*)-M(\delta)$ as a sum of terms each with factor $X(\delta)-X(\delta_*)$, or its transpose, with bounds for these typified by

$$n^{-1}\|X(\delta)-X(\delta_*)\|\left\|(X'(\delta)X(\delta)/n)^{-1}\right\|\|X(\delta)\|,$$

where $\|X(\delta)\|\le Cn^{1/2}$, we deduce $\|M(\delta_*)-M(\delta)\|=\mathscr{O}\left(n^{-1/2}\|X(\delta)-X(\delta_*)\|\right)=\mathscr{O}\left(k^{1/2}\varepsilon\right)$, from (A.15). Thus from (A.17), (A.16) has the same bound, so arguing much as before the contribution from the second term in braces in (A.14) is $\mathscr{O}\left(\|\beta_0\|k^{1/2}\varepsilon\right)$. Thus (A.14)=$\mathscr{O}_p\left(\|\beta_0\|k^{1/2}\varepsilon\right)$, and since Assumption 16 implies that as $n\to\infty$, $c(\gamma)\ge c\|\beta_0\|^2$ uniformly and $\|\beta_0\|^{-1}=\mathscr{O}\left(k^{-1/2}\right)$, the left side of (A.13) is $\mathscr{O}_p\left(\|\beta_0\|^{-1}k^{1/2}\varepsilon\right)=\mathscr{O}_p(\varepsilon)$, whence (A.13) follows from arbitrariness of ε, and the proof is completed. □

Proof of Theorem 3.2. In supplementary material. □

Proof of Theorem 3.3. Let $\xi\left(\lambda,\sigma^2\right)$ denote the first derivative vector of (3.1), evaluated at $\left(\lambda,\sigma^2\right)$. Defining $\mathcal{R}^y(\theta)=R\lambda+X(\delta)\beta-y$, the derivative of (3.1) at any admissible $\left(\theta,\sigma^2\right)$ is

$$\xi\left(\theta,\sigma^2\right)=\left(\varphi'(\lambda,\sigma^2),\ 2\sigma^{-2}n^{-1}\mathcal{R}^{y\prime}(\theta)X(\delta),\ 2\sigma^{-2}n^{-1}\mathcal{R}^{y\prime}(\theta)\Pi(\theta)\right)', \tag{A.18}$$

where

$$\varphi\left(\lambda,\sigma^2\right)=2\sigma^{-2}n^{-1}\left(\sigma^2 trG_1+y'W_1'\mathcal{R}^y(\lambda),\ldots,\sigma^2 trG_p+y'W_p'\mathcal{R}^y(\lambda)\right). \tag{A.19}$$

Noting that $\mathcal{R}^y=-u$, denoting $C_i=G_i+G_i'$ and

$$\phi=\sigma_0^{-2}n^{-1}\left(\sigma_0^2 trC_1-u'C_1u,\ldots,\sigma_0^2 trC_p-u'C_pu\right)', \tag{A.20}$$

so

$$\xi = (\phi', 0, 0)' - 2\sigma_0^{-2} t - 2\sigma_0^{-2} \ell, \tag{A.21}$$

with

$$t = n^{-1} [A, X, 0]' u, \qquad \ell = n^{-1} [0, 0, \Pi]' u. \tag{A.22}$$

Denote by $K_1(\theta)$ and $K_2(\theta)$ the $k \times q$ and $q \times q$ matrices with i-th column $(\partial X'(\delta)/\partial \delta_i) \mathcal{R}^y(\theta)$ and (i,j)-th element $\mathcal{R}^{y\prime}(\theta) \left(\partial^2 X(\delta)/\partial \delta_i \partial \delta_j\right) \beta$, respectively. The matrix of second derivatives of (3.1) at any admissible point in the parameter space, denoted $H\left(\theta, \sigma^2\right)$, is

$$2\sigma^{-2} n^{-1} \begin{bmatrix} \sigma^2 P_1(\lambda) + R'R & R'X(\delta) & R'\Pi(\theta) \\ * & X'(\delta) X(\delta) & X'(\delta)\Pi(\theta) + K_1(\theta) \\ * & * & \Pi'(\theta)\Pi(\theta) + K_2(\theta) \end{bmatrix}, \tag{A.23}$$

whence (2.9) and (3.13) follow.

For any non-null $m \times 1$ vector of constants α, we can use $\hat{\xi} = 0$ and the MVT to write

$$\nu \alpha' \Psi \left(\hat{\theta} - \theta_0\right) = -\nu \alpha' \Psi \bar{H}^{-1} \xi,$$

for some $\bar{\theta}$ such that $\left\|\bar{\theta} - \theta_0\right\| \leq \left\|\hat{\theta} - \theta_0\right\|$, where $\bar{\theta}$ may be different across rows of $\bar{H}^{-1}$. The RHS equals $\sum_{i=1}^4 \Upsilon_i - \nu \alpha' \Psi L^{-1} (t + \ell)$ with

$$\Upsilon_1 = 2\sigma_0^{-2} \nu \alpha' \Psi \bar{H}^{-1} \bar{\Delta}^H H^{-1} (t + \ell), \ \Upsilon_2 = 2\sigma_0^{-2} \nu \alpha' \Psi \Xi^{-1} (H - \Xi) H^{-1} (t + \ell),$$
$$\Upsilon_3 = \nu \alpha' \Psi L^{-1} \left(\sigma_0^2 \Xi/2 - L\right) \left(\sigma_0^2 \Xi/2\right)^{-1} (t + \ell), \ \Upsilon_4 = -\nu \alpha' \Psi \bar{H}^{-1} \phi.$$

We will demonstrate that $\Upsilon_i = o_p(1)$, $i = 1, 2, 3, 4$. First, $\mathbb{E}\|\ell\|^2 = \sigma_0^2 n^{-2} \sum_{r=1}^n \|\pi_r\|^2$, where π_r is the r-th column of Π'. Now

$$\|\pi_r\|^2 = \sum_{i=1}^q \left\{\beta_0' \left(\partial x_r(\delta_0)/\partial \delta_i\right)\right\}^2 \leq \|\beta_0\|^2 \sum_{i=1}^q \sum_{l=1}^k \left(\partial x_{rl}(\delta_0)/\partial \delta_i\right)^2 \leq Ck^2,$$

by Assumption 17. Thus

$$\|\ell\| = \mathcal{O}_p\left(n^{-\frac{1}{2}} k\right), \tag{A.24}$$

by Markov's inequality. By Lemma B.1 we have

$$|\Upsilon_1| \leq 2\sigma_0^{-2} \nu \|\alpha\| \|\Psi\| \left\|\bar{H}^{-1}\right\| \left\|\bar{\Delta}^H\right\| \left\|H^{-1}\right\| (\|t\| + \|\ell\|),$$

where the second factor in norms is $\mathcal{O}\left((p+k)^{\frac{1}{2}}\right)$, the third and fifth are bounded for sufficiently large n by Lemma B.3 (i), the fourth is $\mathcal{O}_p\left(\left\|\hat{\Delta}^H\right\|\right) = \mathcal{O}_p\left(\max\left\{p^2 k/n^{\frac{1}{2}} h, p^{\frac{1}{2}} k^{\frac{5}{2}}/n^{\frac{1}{2}} h^{\frac{1}{2}}\right\}\right)$ by Lemma B.1 (i) and the last is $\mathcal{O}_p\left(p^{\frac{1}{2}} k/n^{\frac{1}{2}}\right)$ (because $\|t\| = \mathcal{O}\left(p^{\frac{1}{2}} k/n^{\frac{1}{2}}\right)$ by (A.13) of

Gupta and Robinson (2015)), so $\Upsilon_1 = \mathscr{O}_p\left(\max\left\{p^{\frac{5}{2}}k^2/n^{\frac{1}{2}}h, pk^{\frac{7}{2}}/n^{\frac{1}{2}}h^{\frac{1}{2}}\right\}\right)$, which is negligible by (3.14). Similarly $\Upsilon_2 = \mathscr{O}_p\left(p^{\frac{3}{2}}k^2/n^{\frac{1}{2}}\right)$ which is negligible by (3.14) and Lemma B.2 (i), and $\Upsilon_3 = \mathscr{O}_p\left(p^{\frac{3}{2}}k/h\right)$ by Lemma B.2 (ii), which is negligible by (3.14). Finally, $\mathbb{E}\left\|\phi\right\|^2 = \sum_{i=1}^{p} \operatorname{var}\left(n^{-1}u'C_iu\right) = \mathscr{O}\left(p/nh\right)$, (shown like (S.17) in the supplementary appendix) so that

$$\|\phi\| = \mathscr{O}_p\left(n^{-\frac{1}{2}}h^{-\frac{1}{2}}p^{\frac{1}{2}}\right), \tag{A.25}$$

by Chebyshev's inequality. So Υ_4 has modulus bounded by $\nu\left\|\Psi\right\|\left\|\bar{H}^{-1}\right\|\left\|\phi\right\|$ times a constant, where the second factor is $\mathscr{O}\left((p+k)^{\frac{1}{2}}\right)$, the third is bounded for sufficiently large n by Lemma B.3 (i) and the last is $\mathscr{O}_p\left(p^{\frac{1}{2}}/n^{\frac{1}{2}}h^{\frac{1}{2}}\right)$. Thus $\Upsilon_4 = \mathscr{O}_p\left(p^{\frac{1}{2}}/h^{\frac{1}{2}}\right)$ which is negligible by (3.14). Then we only need to find the asymptotic distribution of $\nu\alpha'\Psi L^{-1}(t+\ell)$. The theorem now follows by a standard Lindeberg central limit theorem argument. The asymptotic covariance matrix exists, and is positive definite, by Assumption 19. The proof of the consistency of its estimate is omitted. $\square$

Proof of Theorem 3.4. Here we redefine $\mathcal{R}^y(\lambda) = R\lambda - y$ and obtain $\xi = \phi$. Also $H\left(\lambda, \sigma^2\right) = 2n^{-1}P_1(\lambda) + 2\sigma^{-2}n^{-1}R'R$, whence the formulae for H and Ξ follow. Then proceeding as in the proof of Theorem 3.3, we can write

$$\nu\alpha'\Psi\left(\hat{\theta} - \theta_0\right) = \nu\alpha'\Psi\left(\bar{H}^{-1} - \Xi^{-1}\right)\xi - \nu\alpha'\Psi\Xi^{-1}\xi. \tag{A.26}$$

Lemma B.3 (i) indicates that the first term on the RHS of (A.26) is bounded in modulus by a constant times

$$\begin{aligned}&\nu\left\|\Psi\right\|\left(\|t\| + \|\ell\| + \|\phi\|\right)\left(\left\|\bar{\Delta}^H\right\| + \|H - \Xi\|\right) = \\ &\mathscr{O}_p\left(n^{\frac{1}{2}}\max\left\{p^{\frac{1}{2}}k/n^{\frac{1}{2}}, p^{\frac{1}{2}}/n^{\frac{1}{2}}h^{\frac{1}{2}}\right\}\max\left\{p^2k/n^{\frac{1}{2}}h, p^{\frac{1}{2}}k^{\frac{5}{2}}/n^{\frac{1}{2}}h^{\frac{1}{2}}, pk/n^{\frac{1}{2}}\right\}\right),\end{aligned}$$

by (A.24), (A.25) and Lemma B.1 (i). This is negligible by (3.16). Thus we establish the asymptotic distribution of the second term on the RHS of (A.26), which has zero mean and variance $a^{-1}\Psi\left(2\Xi^{-1} + \Xi^{-1}\Omega\Xi^{-1}\right)\Psi'$. Hence we consider the asymptotic normality of

$$\frac{-n^{\frac{1}{2}}\alpha'\Psi\Xi^{-1}\xi}{\left\{\alpha'\Psi\left(2\Xi^{-1} + \Xi^{-1}\Omega\Xi^{-1}\right)\Psi'\alpha\right\}^{\frac{1}{2}}}, \tag{A.27}$$

where α is any $m \times 1$ vector of constants. Write $\varsigma = \left\{\alpha'\Psi\left(2\Xi^{-1} + \Xi^{-1}\Omega\Xi^{-1}\right)\Psi'\alpha\right\}^{\frac{1}{2}}$ for the denominator of (A.27). Then

$$\varsigma \geq \|\Psi'\alpha\|\left\{\underline{\zeta}\left(2\Xi^{-1} + \Xi^{-1}\Omega\Xi^{-1}\right)\right\}^{\frac{1}{2}} \geq c\,\|\Psi'\alpha\| \tag{A.28}$$

by Assumption 20. The numerator of (A.27) can be written as

$$-2\sigma_0^{-2}n^{-\frac{1}{2}}m'u - \sigma_0^{-2}n^{-\frac{1}{2}}u'Du + n^{-\frac{1}{2}}trD \tag{A.29}$$

where $D = \sum_{j=1}^{p}\left(\alpha'\Psi\zeta^j\right)C_j$, $m = \sum_{j=1}^{p}\left(\alpha'\Psi\zeta^j\right)a_j + \sum_{j=p+1}^{p+k}\left(\alpha'\Psi\zeta^j\right)\chi_{j-p}$, with ζ^j and χ_j denoting the j-th columns of Ξ^{-1} and X respectively. We also denote by d_{ij} and m_i the (i,j)-th and i-th elements of D and m respectively. Using (A.29), we can write (A.27) as $-\sum_{i=1}^{n} w_i$, with

$$w_i = \sigma_0^{-2}n^{-\frac{1}{2}}\varsigma^{-1}\left(u_i^2 - \sigma_0^2\right)d_{ii} + 2\sigma_0^{-2}n^{-\frac{1}{2}}\varsigma^{-1}u_i\sum_{j<i}u_jd_{ij} + 2\sigma_0^{-2}n^{-\frac{1}{2}}\varsigma^{-1}m_iu_i. \tag{A.30}$$

$\{w_i, i = 1,\ldots,n, n \geq 1\}$ forms a martingale difference sequence by Assumption 14, so Theorem 2 of Scott (1973) implies $\sum_{i=1}^{n} w_i \xrightarrow{d} N(0,1)$ if

$$\sum_{i=1}^{n}\mathbb{E}\left\{w_i^2 1\left(w_i \geq \epsilon\right)\right\} \xrightarrow{p} 0,\ \forall\epsilon > 0 \tag{A.31}$$

$$\sum_{i=1}^{n}\mathbb{E}\left(w_i^2 \mid u_j, j < i\right) \xrightarrow{p} 1. \tag{A.32}$$

To show (A.31) we can check the sufficient Lyapunov condition

$$\sum_{i=1}^{n}\mathbb{E}\left|w_i\right|^{2+\frac{\chi}{2}} \xrightarrow{p} 0. \tag{A.33}$$

The c_r inequality, (11) and (A.28) indicate that the left side is bounded by a constant times

$$\frac{\sum_{i=1}^{n}\left|d_{ii}\right|^{2+\frac{\chi}{2}}}{n^{1+\frac{\chi}{4}}\left\|\Psi'\alpha\right\|^{2+\frac{\chi}{2}}} + \frac{\sum_{i=1}^{n}\mathbb{E}\left|\sum_{j<i}u_jd_{ij}\right|^{2+\frac{\chi}{2}}}{n^{1+\frac{\chi}{4}}\left\|\Psi'\alpha\right\|^{2+\frac{\chi}{2}}} + \frac{\sum_{i=1}^{n}\left|m_i\right|^{2+\frac{\chi}{2}}}{n^{1+\frac{\chi}{4}}\left\|\Psi'\alpha\right\|^{2+\frac{\chi}{2}}}. \tag{A.34}$$

The first term in (A.34) is bounded by

$$\max_i\left|d_{ii}\right|^{2+\frac{\chi}{2}}/n^{\frac{\chi}{4}}\left\|\Psi'\alpha\right\|^{2+\frac{\chi}{2}}, \tag{A.35}$$

while the third term is bounded by

$$\max_i\left|m_i\right|^{2+\frac{\chi}{2}}/n^{\frac{\chi}{4}}\left\|\Psi'\alpha\right\|^{2+\frac{\chi}{2}}. \tag{A.36}$$

By the Burkholder, von Bahr/Esseen and elementary ℓ_p-norm inequalities, the second term in

(A.34) is bounded by a constant times

$$\max_i \left| \sum_{j<i} d_{ij}^2 \right|^{1+\frac{\chi}{4}} / n^{\frac{\chi}{4}} \left\| \Psi' \alpha \right\|^{2+\frac{\chi}{2}}. \tag{A.37}$$

Now, writing e_i for the n-dimensional vector with unity in the i-th position and zeros elsewhere, we can write $\sum_{j=1}^n d_{ij}^2 = e_i' D^2 e_i \leq \|D\|^2$ which is bounded by

$$\begin{aligned} \left\| \sum_{j=1}^{p} \left(\alpha' \Psi \zeta^j\right) C_j \right\|^2 &\leq Cp^2 \left(\max_j \|C_j\| \right)^2 \left(\max_j \left\|\zeta^j\right\| \right)^2 \|\Psi'\alpha\|^2 \leq C \left\|\Xi^{-1}\right\|^2 p^2 \|\Psi'\alpha\|^2 \\ &= Cp^2 \|\Psi'\alpha\|^2 \left\{ \underline{\zeta}(\Xi) \right\}^{-2} \leq Cp^2 \|\Psi'\alpha\|^2, \end{aligned} \tag{A.38}$$

using Assumption 20. Also, we can use (A.38) to bound

$$|d_{ii}| \leq \left(\sum_{j=1}^{n} d_{ij}^2 \right)^{\frac{1}{2}} \leq Cp \|\Psi'\alpha\|. \tag{A.39}$$

(A.38) and (A.39) imply that (A.35) and (A.37) are both $\mathscr{O}\left(p^{2+\frac{\chi}{2}}/n^{\frac{\chi}{4}}\right)$. This is negligible by (3.16). Next

$$|m_i| \leq \sum_{j=1}^{p} \left|\alpha' \Psi \zeta^j\right| |a_{ij}| + \sum_{j=p+1}^{p+k} \left|\alpha' \Psi \zeta^j\right| |x_{ij}| = \mathscr{O}\left(k(p+1)\|\Psi'\alpha\|\right), \tag{A.40}$$

using Assumptions 14, 20. Then (A.36) is $\mathscr{O}_p\left(p^{2+\frac{\chi}{2}} k^{2+\frac{\chi}{2}} / n^{\frac{\chi}{4}}\right)$, which is negligible by (3.16). Hence (A.33) is proved.

We now show (A.32). Write $\sum_{i=1}^n \mathbb{E}\left(w_i^2 \mid u_j, j<i\right) - 1 = 4\left(f_1+f_2+f_3\right)$ with $f_1 = \sigma_0^{-2} n^{-1} \varsigma^{-2} \sum_i \sum_j \sum_{k\ (j,k<i, j\neq k)} d_{ij} d_{ik} u_j u_k$, $f_2 = \sigma_0^{-2} n^{-1} \varsigma^{-2} \sum_i \sum_{j<i} d_{ij}^2 \left(u_j^2 - \sigma_0^2\right)$ and $f_3 = \sigma_0^{-4} n^{-1} \varsigma^{-2} \sum_i \left(\sigma_0^2 m_i + \mu_3 d_{ii}\right) \sum_{j<i} d_{ij} u_j$. All sums and maxima are taken over 1 to n unless otherwise stated. f_1 has zero mean and variance bounded by $n^{-2}\varsigma^{-4}$ times

$$\begin{aligned} C \sum_{h,i,j,k\ (j,k<i,h)} |d_{ij} d_{ik} d_{hj} d_{hk}| &\leq C \sum_{h,i,j,k} |d_{ij} d_{ik}| \left(d_{hj}^2 + d_{hk}^2\right) \\ &\leq C \left(\max_i \sum_k |d_{ik}| \right) \left(\max_j \sum_i |d_{ij}| \right) \sum_{i,j} d_{ij}^2 \\ &= C \|D\|_R^2 \|D\|_F^2 \leq C \|\Psi'\alpha\|^4 np^4, \end{aligned} \tag{A.41}$$

by (A.38) and because, for each $i = 1, \ldots, n$, $\left(\sum_{j=1}^n d_{ij}^2\right)^{\frac{1}{2}} \leq \sum_{j=1}^n |d_{ij}| \leq \|D\|_R \leq Cp \|\Psi'\alpha\|$

by Assumption 3. (A.28) and (A.41), together with Markov's inequality, imply that $f_1 = \mathscr{O}_p\left(p^2/n^{\frac{1}{2}}\right)$, which is negligible by (3.16). Next, f_2 has zero mean and variance bounded by $n^{-2}\varsigma^{-4}$ times

$$C\sum_{i,h}\sum_{j<i,h} d_{ij}^2 d_{hj}^2 \leq C\sum_{i,h,j} d_{ij}^2 d_{hj}^2 \leq C\left(\max_j \sum_h d_{hj}^2\right)\|D\|_F^2 \leq C\left\|\Psi'\alpha\right\|^4 np^4, \tag{A.42}$$

by (A.38). (A.28) and (A.42), together with Markov's inequality, imply that $f_2 = \mathscr{O}_p\left(p^2/n^{\frac{1}{2}}\right)$ which is negligible by (3.16). Finally f_3 has zero mean and variance bounded by $n^{-2}\varsigma^{-4}$ times

$$\begin{aligned} C & \sum_i \left(\sigma_0^2 m_i + \mu_3 d_{ii}\right)^2 \sum_{j<i} d_{ij}^2 \leq C\left(\max_i m_i^2 + \max_i d_{ii}^2\right)\|D\|_F^2 \\ \leq & \; C\left(\max_i m_i^2 + \max_i \sum_j d_{ij}^2\right)\|D\|_F^2 = \mathscr{O}\left(\left\|\Psi'\alpha\right\|^4\left(k^2+1\right)np^4\right), \end{aligned} \tag{A.43}$$

by (A.38) and (A.40). (A.28) and (A.43), together with Markov's inequality, imply that $f_3 = \mathscr{O}_p\left(p^2 k/n^{\frac{1}{2}}\right)$, which is negligible by (3.16). The asymptotic covariance matrix exists, and is positive definite, by Assumption 20. □

Proofs of Theorems 4.1, 4.2 and 4.3 . These follow like the proofs of Theorems 2.3, 3.1, 3.3 and 3.4, with the replacement of $M(\delta)$ by $M(\gamma)$ requiring only bounds established in the proofs of those theorems and elementary inequalities, but we give some details for the proof of Theorem 4.2 in view of the differential norming applied therein. Also note that now

$$\xi = \left(\phi', 0, 0\right)' - 2\sigma_0^{-2}t - 2\sigma_0^{-2}\ell, \tag{A.44}$$

with SX and $S\partial X/\partial\delta_i$ replacing X and $\partial X/\partial\delta_i$ respectively in the definitions of t and Π, $A = 0$ in t. H is redefined as

$$H = 2\sigma_0^{-2}n^{-1}\begin{bmatrix} \sigma_0^2 P_1 + B'B & B'SX + Q_1' & B'\Pi + Q_2' \\ * & X'S'SX & X'S'\Pi + K_1 \\ * & * & \Pi'\Pi + K_2 \end{bmatrix}, \tag{A.45}$$

where Q_1 has j-th column $X'W_j'u$ and Q_2 has (i,j)-th element $\beta_0'\partial X'/\partial\delta_i W_j'u$, $i = 1,\ldots,q$, $j = 1,\ldots,p$, and $S\partial X/\partial\delta_i$ and $S\partial^2 X(\delta)/\partial\delta_i\partial\delta_j$ replace $\partial X/\partial\delta_i$ and $\partial^2 X(\delta)/\partial\delta_i\partial\delta_j$ respectively in the definitions of K_1 and K_2. Thus Ξ is redefined simply by taking the expectation of (A.45), whence (4.3) follows, and Ω is redefined using the new definitions of X and Π, and also $A = 0$. The inflation of by the h factor in Theorem 4.2 is necessary for a nondegenerate limit distribution, as in Theorems 2.3 and 2.6. Indeed, because the first p elements in both t and ℓ equal zero, the negligibility of ϕ immediately causes singularity of the limiting covariance matrix.

Proceeding like in the proof of earlier theorems, for any non-null $m\times 1$ vector of constants α,

$\hat{\xi} = 0$ and the MVT imply $\nu\alpha'\Psi\Phi^{-\frac{1}{2}}\left(\hat{\theta} - \theta_0\right) = l_1 + l_2 - \nu\alpha'\Psi\Xi^{\Phi-1}\Phi^{\frac{1}{2}}\xi$, where

$$l_1 = \nu\alpha'\Psi H^{\Phi-1}\left(\Phi^{\frac{1}{2}}\bar{\Delta}^H\Phi^{\frac{1}{2}}\right)\bar{H}^{\Phi-1}\Phi^{\frac{1}{2}}\xi, \tag{A.46}$$

$$l_2 = \nu\alpha'\Psi\Xi^{\Phi-1}\left(H^\Phi - \Xi^\Phi\right)H^{\Phi-1}\Phi^{\frac{1}{2}}\xi, \tag{A.47}$$

The top left block of $\Phi^{\frac{1}{2}}\bar{\Delta}^H\Phi^{\frac{1}{2}}$ is identical to that for whose spectral norm Lemma B.1(ii) derives a bound. The spectral norms of remaining blocks are bounded like in the proofs of Section 3, but again with the replacements described in the previous paragraph. Similarly a bound for the spectral norm of the top left block of $H^\Phi - \Xi^\Phi$ is derived in Lemma B.2(i) under (2.12), and indeed the same lemma also accounts for the remaining blocks. Evidently all bounds thus obtained for (A.46) and (A.47) are subsets of those assumed negligible in (2.11) and (3.14), and therefore both l_1 and l_2 are negligible. The asymptotic distribution of $-\nu\alpha'\Psi\Xi^{\Phi-1}\Phi^{\frac{1}{2}}\xi$ is then established by applying a martingale central limit theorem as in earlier proofs. □

B Technical Lemmas

All proofs are contained in the supplementary appendix.

Lemma B.1. *(i) Under the conditions of Theorem 3.3 or 3.4,*

$$\left\|\hat{\Delta}^H\right\| = \mathscr{O}_p\left(n^{-\frac{1}{2}}h^{-\frac{1}{2}}p^{\frac{1}{2}}k\left(h^{-\frac{1}{2}}p^{\frac{3}{2}} + k^{\frac{3}{2}}\right)\right).$$

(ii) Under the conditions of Theorem 2.3, 2.4 or 2.6,

$$h\left\|\hat{\Delta}^H\right\| = \mathscr{O}_p\left(n^{-\frac{1}{2}}p^2\right),$$

or, equivalently, $\left\|\hat{\Delta}^H\right\| = \mathscr{O}_p\left(n^{-\frac{1}{2}}h^{-1}p^2\right)$.

The same bounds hold if we replace $\left\|\hat{\Delta}^H\right\|$ *by* $\left\|\bar{\Delta}^H\right\|$, *where* $\left\|\bar{\theta} - \theta_0\right\| \leq \left\|\hat{\theta} - \theta_0\right\|$.

Lemma B.2. *Suppose that Assumptions 1-14 hold. Then*

(i) $\|H - \Xi\| = \mathscr{O}_p\left(p/n^{\frac{1}{2}}h^{\frac{1}{2}}\right)$ *for the SAR without regressors and bounded h,* $\|H - \Xi\| = \mathscr{O}_p\left(\max\left\{p/n^{\frac{1}{2}}h, 1/n\right\}\right)$ *for the SAR without regressors and divergent h if (2.12) also holds and* $\|H - \Xi\| = \mathscr{O}_p\left(pk/n^{\frac{1}{2}}\right)$ *for the SAR with regressors.*

(ii) $\left\|L - \sigma_0^2\Xi/2\right\| = \mathscr{O}\left(p/h\right)$.

Lemma B.3. *Let Assumptions 1-19 hold.*

(i) If (3.14) holds, then

$$
\begin{aligned}
\left\| \hat{H}^{-1} \right\| &= \mathscr{O}_p\left(\| H^{-1} \|\right) = \mathscr{O}_p\left(\| \Xi^{-1} \|\right) = \mathscr{O}_p\left(\left\{\underline{\zeta}(L)\right\}^{-1}\right) = \mathscr{O}_p(1), \\
\left\| \hat{H} \right\| &= \mathscr{O}_p\left(\|H\|\right) = \mathscr{O}_p\left(\|\Xi\|\right) = \mathscr{O}_p\left(\overline{\zeta}(L)\right) = \mathscr{O}_p(1).
\end{aligned}
$$

If h is bounded and Assumption 20 holds together with (3.16), then

$$
\begin{aligned}
\left\| \hat{H}^{-1} \right\| &= \mathscr{O}_p\left(\| H^{-1} \|\right) = \mathscr{O}_p\left(\left\{\underline{\zeta}(\Xi)\right\}^{-1}\right) = \mathscr{O}_p(1), \\
\left\| \hat{H} \right\| &= \mathscr{O}_p\left(\|H\|\right) = \mathscr{O}_p\left(\overline{\zeta}(\Xi)\right) = \mathscr{O}_p(1)
\end{aligned}
$$

(ii) If $\underline{\lim}_{n\to\infty} \underline{\zeta}(h\Xi) > 0$ and (2.11) holds, then

$$
\left\| \left(h\hat{H}\right)^{-1} \right\| = \mathscr{O}_p\left(\left\| (hH)^{-1} \right\|\right) = \mathscr{O}_p\left(\left\{\underline{\zeta}\left(h\Xi\right)\right\}^{-1}\right) = \mathscr{O}_p(1).
$$

(iii) If h is bounded, $\underline{\lim}_{n\to\infty} \underline{\zeta}(\Xi) > 0$ and (2.13) holds , then

$$
\left\| \hat{H}^{-1} \right\| = \mathscr{O}_p\left(\| H^{-1} \|\right) = \mathscr{O}_p\left(\left\{\underline{\zeta}\left(\Xi\right)\right\}^{-1}\right) = \mathscr{O}_p(1).
$$

The same bounds hold if we replace $\left\| \hat{H} \right\|$ by $\| \bar{H} \|$, where $\| \bar{\theta} - \theta_0 \| \le \left\| \hat{\theta} - \theta_0 \right\|$.

References

Amemiya, T., 1974. The non-linear two-stage least-squares estimator. Journal of Econometrics 82, 105–110.

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Dordrecht.

Anselin, L., 2001. Rao's score test in spatial econometrics. Journal of Statistical Planning and Inference 97, 113–139.

Anselin, L., Smirnov, O., 1996. Efficient algorithms for constructing proper higher order spatial lag operators. Journal of Regional Science 36, 67–89.

Badinger, H., Egger, P., 2011. Estimation of higher-order spatial autoregressive cross-section models with heteroscedastic disturbances. Papers in Regional Science 90, 213–236.

Badinger, H., Egger, P., 2013 Estimation and testing of higher-order spatial autoregressive panel data error component models. Journal of Geographical Systems 15, 453–489.

Berk, K. N., 1974. Consistent autoregressive spectral estimates. The Annals of Statistics 2, 489–502.

Blommestein, H. J., 1983. Specification and estimation of spatial econometric models: A discussion of alternative strategies for spatial economic modelling. Regional Science and Urban Economics 13, 251–270.

Blommestein, H. J., 1985. Elimination of circular routes in spatial dynamic regression equations. Regional Science and Urban Economics 15, 121–130.

Blommestein, H. J., Koper, N. A. M., 1992. Recursive algorithms for the elimination of redundant paths in spatial lag operators. Journal of Regional Science 32, 91–111.

Blommestein, H. J., Koper, N. A. M., 1997. The influence of sample size on the degree of redundancy in spatial lag operators. Journal of Econometrics 82, 317–333.

Brandsma, A. S., Ketellapper R. H. 1979. A biparametric approach to spatial autocorrelation. Environment and Planning 11, 51–58.

Case, A. C., 1991. Spatial patterns in household demand. Econometrica 59, 953–965.

Case, A. C., 1992. Neighborhood influence and technological change. Regional Science and Urban Economics 22, 491–508.

Cliff, A. D., Ord, J. K., 1973. Spatial Autocorrelation. Pion, London.

Delgado, M. A., Robinson, P. M., 2014. Non-nested testing of spatial correlation. Journal of Econometrics, 187, 385-401.

Elhorst, J. P., Lacombe, D. J., Piras, G., 2012. On model specification and parameter space definitions in higher order spatial econometric models. Regional Science and Urban Economics 42, 211–220.

Gupta, A., Robinson, P. M., 2015. Inference on higher-order spatial autoregressive models with increasingly many parameters. Journal of Econometrics, 186, 19-31.

Hillier, G., Martellosio, F., 2013. Properties of the maximum likelihood estimator in spatial autoregressive models. CeMMap Working Paper.

Huang, J. S., 1984. The autoregressive moving average model for spatial analysis. Australian Journal of Statistics 26, 169–178.

Huber, P. J., 1973. Robust regression: asymptotics, conjectures and Monte Carlo. The Annals of Statistics 1, 799–821.

Jennrich, R. I., 1969. Asymptotic properties of non-linear least squares estimators. The Annals of Mathematical Statistics 40, 633–643.

Kelejian, H. H., Prucha, I. R., 1997. Estimation of spatial regression models with autoregressive errors by two-stage least squares procedures: a serious problem. International Regional Science Review 20, 103–111.

Kelejian, H. H., Prucha, I. R., 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, 99–121.

Kelejian, H. H., Prucha, I. R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review 40, 509–533.

Kelejian, H. H., Prucha, I. R., Yuzefovich, Y. A., 2006. Estimation problems in models with spatial weighting matrices which have blocks of equal elements. Journal of Regional Science 46, 507–515.

Kolympiris, C., Kalaitzandonakes, N., Miller, D., 2011. Spatial collocation and venture capital in the US biotechnology industry. Research Policy 40, 1188–1199.

Lahiri, S. N., 1996. On inconsistency of estimators based on spatial data under infill asymptotics. Sankhyā: Series A 58, 403–417.

Lee, L. F., 2002. Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. Econometric Theory 18, 252–277.

Lee, L. F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72, 1899–1925.

Lee, L. F., Liu, X., 2010. Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. Econometric Theory 26, 187–230.

LeSage, J. P., Pace, R. K., 2011. Pitfalls in higher order model extensions of basic spatial regression methodology. The Review of Regional Studies 41, 13–26.

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, 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, 1403–1417.

Robinson, P. M., 1972. Non-linear regression for multiple time-series. Journal of Applied Probability 9, 758–768.

Robinson, P. M., 1979. Distributed lag approximation to linear time-invariant systems. The Annals of Statistics 7, 507–515.

Robinson, P. M., 2003. Denis Sargan: some perspectives. Econometric Theory 19, 481–494.

Robinson, P. M., 2010. The efficient estimation of the semiparametric spatial autoregressive model. Journal of Econometrics 157, 6–17.

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

Supplementary appendix to 'Pseudo Maximum Likelihood Estimation of Spatial Autoregressive Models with Increasing Dimension'

A. Gupta and P.M. Robinson

October 22, 2015

This appendix contains the proof of Theorems 2.2 and 3.2, all lemmas in Appendix B and some supplementary lemmas. Denote throughout $\kappa_i = \kappa_{in} = \lambda_i - \lambda_{0i}$, $\hat{\kappa}_i = \hat{\kappa}_{in} = \hat{\lambda}_i - \lambda_{0i}$, for $i = 1, \ldots, p$, $\kappa = \lambda - \lambda_0$, $\hat{\kappa} = \hat{\lambda} - \lambda_0$, $\hat{\Delta}^{\bar{\beta}} = \hat{\beta} - \beta_0$ and $\hat{\Delta}^{\varphi(\bar{\sigma})} = \varphi(\hat{\sigma}) - \varphi(\sigma_0)$ for a function φ. We will suppress n subscripts. Further, for any matrix, vector $E(\theta, \sigma^2)$, $\tilde{E}$ denotes evaluation at a generic estimate $\left(\tilde{\theta}', \tilde{\sigma}^2\right)'$ and $\tilde{\Delta}^E = \tilde{E} - E$.

Proof of Theorems 2.2 and 3.2. Consider Theorem 3.2 first. Because $\hat{\sigma}^2 = \bar{\sigma}^2(\hat{\gamma}) = n^{-1}y'\hat{S}'\hat{M}\hat{S}y$ and $\hat{S} = \left(I - \sum_{i=1}^p \hat{\kappa}_i G_i\right) S$, so we can use $y = S^{-1}X\beta_0 + S^{-1}u$ to write $\hat{\sigma}^2 - \sigma_0^2 = \sum_{i=1}^{11} \hat{D}_i$, with $\hat{D}_1 = n^{-1}u'u - \sigma_0^2, \hat{D}_2 = -2n^{-1}u'\sum_{i=1}^p \hat{\kappa}_i G_i'\hat{M}u, \hat{D}_3 = 2n^{-1}\sum_{i=1}^p \hat{\kappa}_i a_i'\hat{M}u, \hat{D}_4 = n^{-1}u'\sum_{i,j=1}^p \hat{\kappa}_i\hat{\kappa}_j G_i'\hat{M}G_j u, \hat{D}_5 = n^{-1}\sum_{i,j=1}^p \hat{\kappa}_i\hat{\kappa}_j a_i'\hat{M}a_j, \hat{D}_6 = 2n^{-1}\sum_{i,j=1}^p \hat{\kappa}_i\hat{\kappa}_j a_i'\hat{M}G_j u, \hat{D}_7 = \left(n^{-1}u'\hat{X}\right)\left(n^{-1}\hat{X}'\hat{X}\right)^{-1}\left(n^{-1}\hat{X}'u\right)$,
$\hat{D}_8 = n^{-1}\beta_0'X'\hat{M}X\beta_0, \hat{D}_9 = -2n^{-1}\sum_{i=1}^p \hat{\kappa}_i\beta_0'X'\hat{M}a_i, \hat{D}_{10} = -2n^{-1}\beta_0'X'\hat{M}u, \hat{D}_{11} = -2n^{-1}\sum_{i=1}^p \hat{\kappa}_i\beta_0'X'\hat{M}G_i u$. We claim that

$$\hat{D}_1 = \mathscr{O}_p\left(n^{-\frac{1}{2}}\right), \hat{D}_2 = \mathscr{O}_p\left(\|\xi\|\right), \hat{D}_3 = \mathscr{O}_p\left(\|\xi\| k^{\frac{1}{2}}\right), \hat{D}_4 = \mathscr{O}_p\left(\|\xi\|^2\right),$$
$$\hat{D}_5 = \mathscr{O}_p\left(\|\xi\|^2 k\right), \hat{D}_6 = \mathscr{O}_p\left(\|\xi\|^2 k^{\frac{1}{2}}\right), \hat{D}_7 = \mathscr{O}_p\left(\|\xi\|^2 k\right), \hat{D}_8 = \mathscr{O}_p\left(\|\xi\| k^{\frac{3}{2}}\right),$$
$$\hat{D}_9 = \mathscr{O}_p\left(\|\xi\| k\right), \hat{D}_{10} = \mathscr{O}_p\left(\|\xi\| k\right), \hat{D}_{11} = \mathscr{O}_p\left(\|\xi\| k^{\frac{1}{2}}\right).$$

First note the following properties:

1. *Bound for* $\|\hat{\kappa}\|$: $\|\hat{\kappa}\| \leq \left\|\hat{\theta} - \theta_0\right\|$ where $\left\|\hat{\theta} - \theta_0\right\| = \mathscr{O}_p\left(\|\xi\|\right)$ by Theorem 1 of Robinson (1988).

2. *Bounds for* $\|a_i\|$, $\left\|\hat{\Delta}^X\right\|$ *and* $\left\|\hat{X}\right\|$: By Assumption 17, $\|X\| = \mathscr{O}\left(n^{\frac{1}{2}}\right)$, implying $\|a_i\| = \mathscr{O}_p\left(n^{\frac{1}{2}}k^{\frac{1}{2}}\right)$. Now $\left\|\hat{X}\right\| \leq \left\|\hat{\Delta}^X\right\| + \|X\|$, while by the mean value theorem (MVT) there exists $\bar{\delta}$ (possibly different for each matrix element) satisfying $\left\|\bar{\delta} - \delta_0\right\| \leq \left\|\hat{\delta} - \delta_0\right\|$ such that

$$\left\|\hat{\Delta}^X\right\|^2 \leq \sum_{i=1}^n \sum_{j=1}^k \left\|\partial x_{ij}\left(\bar{\delta}\right)/\partial\delta\right\|^2 \left\|\hat{\delta} - \delta_0\right\|^2 = \mathscr{O}_p\left(\|\xi\|^2 nk\right), \tag{S.1}$$

by Assumption 17 and Cauchy-Schwarz inequality, so

$$n^{-\frac{1}{2}} \left\| \hat{X} \right\| = \mathscr{O}_p (1) \tag{S.2}$$

if $\|\xi\| k^{\frac{1}{2}} = o_p(1)$, which is true by (3.10).

3. *Bound for* $\left\| \left(n^{-1}\hat{X}'\hat{X}\right)^{-1} \right\|$: Because $\left(n^{-1}\hat{X}'\hat{X}\right)^{-1}$ equals

$$\begin{aligned} & \left(n^{-1}X'X\right)^{-1} + \hat{\Delta}^{\left(n^{-1}X'X\right)^{-1}} \\ = & \left(n^{-1}X'X\right)^{-1} - \left(n^{-1}X'X\right)^{-1} \hat{\Delta}^{n^{-1}X'X} \left(n^{-1}\hat{X}'\hat{X}\right)^{-1} \\ = & \left(n^{-1}X'X\right)^{-1} - \left(n^{-1}X'X\right)^{-1} n^{-1} \left(\hat{X}\hat{\Delta}^X + \hat{\Delta}^X X\right) \left(n^{-1}\hat{X}'\hat{X}\right)^{-1}, \end{aligned}$$

we get

$$\begin{aligned} & \left\| \left(n^{-1}\hat{X}'\hat{X}\right)^{-1} \right\| \left(1 + \left\| \left(n^{-1}X'X\right)^{-1} \right\| \left\| n^{-\frac{1}{2}} \hat{\Delta}^X \right\| n^{-\frac{1}{2}} \left(\left\| \hat{X} \right\| + \|X\| \right) \right) \\ \leq & \left\| \left(n^{-1}X'X\right)^{-1} \right\|, \end{aligned} \tag{S.3}$$

By (S.1), Assumptions 14 and 17, $\left\| n^{-\frac{1}{2}} \hat{\Delta}^X \right\| = \mathscr{O}_p \left(\|\xi\| k^{\frac{1}{2}} \right) = o_p(1)$. It follows that $\left\| \left(n^{-1}\hat{X}'\hat{X}\right)^{-1} \right\| = \mathscr{O}_p(1)$, again by Assumption 14.

4. *Bound for* $\left\| n^{-1}\hat{X}'u \right\|$: $\left\| n^{-1}\hat{X}'u \right\| \leq \left\| n^{-1}\hat{\Delta}^{X'}u \right\| + \left\| n^{-1}X'u \right\|$, with the first term on the RHS $\mathscr{O}_p \left(\|\xi\| k^{\frac{1}{2}} \right)$ by (S.1), and the second term readily shown to $\mathscr{O}_p \left(k^{\frac{1}{2}}/n^{\frac{1}{2}} \right)$ by Assumption 14, on evaluating $\mathbb{E} \|X'u\|^2$ and using Markov's inequality. The first order dominates the second by (S.4).

The bound for $\hat{D}_1$ is standard. Next, because $\left\| \hat{M} \right\| = 1$,

$$\left| \hat{D}_2 \right| \leq n^{-1} \|u\|^2 \|\hat{\kappa}\| \left(\sum_{i=1}^{p} \|G_i\|^2 \right)^{\frac{1}{2}} \leq n^{-1} \|u\|^2 \|\hat{\kappa}\| \left(\sum_{i=1}^{p} \left\| S^{-1} \right\|^2 \|W_i\|^2 \right)^{\frac{1}{2}} = \mathscr{O}_p (\|\xi\|),$$

by Cauchy Schwarz inequality, Assumption 3 and point 1. For $\hat{D}_3$ the bound follows similarly using point 2. above. Similarly

$$\begin{aligned}
\left|\hat{D}_4\right| &\leq n^{-1}\|u\|^2\left(\sum_{i,j=1}^{p}\hat{\kappa}_i^2\hat{\kappa}_j^2\right)^{\frac{1}{2}}\left(\sum_{i,j=1}^{p}\|G_i\|^2\|G_j\|^2\right)^{\frac{1}{2}} \\
&\leq n^{-1}\|u\|^2\left(\sum_{i=1}^{p}\hat{\kappa}_i^2\right)^{\frac{1}{2}}\left(\sum_{j=1}^{p}\hat{\kappa}_j^2\right)^{\frac{1}{2}}\left(\sum_{i=1}^{p}\|G_i\|^2\right)^{\frac{1}{2}}\left(\sum_{j=1}^{p}\|G_j\|^2\right)^{\frac{1}{2}} \\
&= \mathscr{O}_p\left(\sum_{i=1}^{p}\hat{\kappa}_i^2\right) = \mathscr{O}_p\left(\|\hat{\kappa}\|^2\right) = \mathscr{O}_p\left(\|\xi\|^2\right).
\end{aligned}$$

A similar argument holds for the bounds on $\hat{D}_5$ and $\hat{D}_6$, again using point 2. Next, $\left|\hat{D}_7\right| = \mathscr{O}_p\left(\left\|n^{-1}\hat{X}'u\right\|^2\left\|\left(n^{-1}\hat{X}'\hat{X}\right)^{-1}\right\|\right)$, whence the stated bound follows from points 3. and 4.

To obtain the next bound decompose $\hat{D}_8 = \hat{D}_{1,8} - \hat{D}_{2,8}$ with

$$\begin{aligned}
\hat{D}_{1,8} &= n^{-2}\beta_0'X'\hat{\Delta}^X\left(n^{-1}X'X\right)^{-1}\left(X+\hat{X}\right)'X\beta_0 = \mathscr{O}_p\left(\|\xi\|\,k^{\frac{3}{2}}\right) \\
\hat{D}_{2,8} &= n^{-3}\beta_0'X'\hat{X}\left(n^{-1}X'X\right)^{-1}\left(\hat{\Delta}^{X\prime}\hat{X}+X'\hat{\Delta}^X\right)\left(n^{-1}\hat{X}'\hat{X}\right)^{-1}\hat{X}'X\beta_0 = \mathscr{O}_p\left(\|\xi\|\,k^{\frac{3}{2}}\right).
\end{aligned}$$

The bound for $\hat{D}_{10}$ is obtained in much the same way, while those for $\hat{D}_9$ and $\hat{D}_{11}$ are derived using the Cauchy Schwarz inequality as for earlier quantities. Thus

$$\max_{i=1,\ldots,11}\hat{D}_i = \mathscr{O}_p\left(\|\xi\|\,k^{\frac{3}{2}}\right),$$

with

$$\|\xi\| = \mathscr{O}_p\left(\max\left\{\|\phi\|,\|t\|,\|\ell\|\right\}\right) = \mathscr{O}_p\left(n^{-\frac{1}{2}}\max\left\{p^{\frac{1}{2}}k, h^{-\frac{1}{2}}p^{\frac{1}{2}}, k\right\}\right) = \mathscr{O}_p\left(p^{\frac{1}{2}}k/n^{\frac{1}{2}}\right), \tag{S.4}$$

using (A.24), (A.25) and $\|t\| = \mathscr{O}_p\left(p^{\frac{1}{2}}k/n^{\frac{1}{2}}\right)$ (see (A.13) in Gupta and Robinson (2015)). Thus

$$\hat{\sigma}^2 - \sigma_0^2 = \mathscr{O}_p\left(\frac{p^{\frac{1}{2}}k^{\frac{5}{2}}}{n^{\frac{1}{2}}}\right) \tag{S.5}$$

If δ_0 is known, $M(\delta)X(\delta) = 0$ so $\hat{D}_i = 0$ for $i \geq 8$ and the order $\|\xi\|\,k^{\frac{1}{2}}$ suffices. The proof of Theorem 2.2 follows in exactly the same manner, except here $\hat{\sigma}^2 - \sigma_0^2 = \hat{D}_1 -$

$2n^{-1}u'\sum_{i=1}^{p}\hat{\kappa}_i G_i' u + n^{-1}u'\sum_{i,j=1}^{p}\hat{\kappa}_i\hat{\kappa}_j G_i' G_j u$ only, whence

$$\hat{\sigma}^2 - \sigma_0^2 = \mathscr{O}_p\left(\frac{p^{\frac{1}{2}}}{n^{\frac{1}{2}}h^{\frac{1}{2}}}\right) \tag{S.6}$$

follows. □

Lemma LS.1. *Let Assumption 12 hold. Then* $\left\|S^{-1}(\lambda)\right\|_R$ *and* $\left\|S'^{-1}(\lambda)\right\|_R$ *are uniformly bounded in a closed neighbourhood of* λ_0*.*

Proof. We can write $S^{-1}(\lambda) = S^{-1}\left(I - \sum_{i=1}^{p}\kappa_i G_i\right)^{-1}$. We will justify $\|\sum_{i=1}^{p}\kappa_i G_i\|_R \leq 1-\varepsilon$, any $\varepsilon > 0$. In the 'single non-zero diagonal block' of Section 1,

$$\left\|\sum_{i=1}^{p}\kappa_i G_i\right\|_R \leq C\max_{i=1,\ldots,p}\left(|\kappa_i|\,\|V_i\|_R\right) \leq C\left(\sum_{i=1}^{p}\kappa_i^2\right)^{1/2}\max_{i=1,\ldots,p}\|V_i\|_R \tag{S.7}$$

whence the result follows by Assumption 12, taking a small enough neighbourhood $B(\lambda_0)$. In the more general, non-block-diagonal, case,

$$\left\|\sum_{i=1}^{p}\kappa_i G_i\right\|_R \leq C\left(\sum_{i=1}^{p}\kappa_i^2\right)^{1/2}\left(\sum_{i=1}^{p}\|W_i\|_R^2\right)^{1/2}, \tag{S.8}$$

the claim following now by (2.10) and a choice of sufficiently small neighbourhood. Thus

$$\left\|S^{-1}(\lambda)\right\|_R \leq \left\|S^{-1}\right\|_R\sum_{j=0}^{\infty}\left\|\sum_{i=1}^{p}\kappa_i G_i\right\|_R^j \leq C/\varepsilon \leq C, \quad \lambda \in B(\lambda_0).$$

The result follows if we take a closed subset of $B(\lambda_0)$, denoted $B^c(\lambda_0)$. The claim for the transpose follows similarly. □

Corollary LS.2. *Under the conditions of Lemma LS.1, we have*

1. *For each* $i = 1,\ldots,p$, $\|G_i(\lambda)\|_R$ *and* $\|G_i'(\lambda)\|_R$ *are uniformly bounded in* $B^c(\lambda_0)$*.*
2. *For each* $i = 1,\ldots,p$, *the elements of* $G_i(\lambda)$ *are uniformly* $\mathscr{O}\left(h^{-1}\right)$ *in* $B^c(\lambda_0)$ *if also Assumption 2 holds.*

Proof. 1. Follows by Lemma LS.1 together with Assumption 12 while *2.* follows by Lemma LS.1 together with Assumption 2. □

Lemma LS.3. *Under the conditions of Corollary LS.2 (2), we have*

$$tr\left(G_i(\lambda)G_j(\lambda)G_k(\lambda)\right) = \mathscr{O}\left(n/h\right) \forall\, \lambda \in B^c\left(\lambda_0\right) \text{ and for any } i,j,k = 1,\ldots,p.$$

Proof. Consider $\lambda \in B^c\left(\lambda_0\right)$. The (l,m)-th element of $G_i(\lambda)G_j(\lambda)G_k(\lambda)$ is $g_{l,i}'G_j(\lambda)G_k(\lambda)e_m$ which is bounded in absolute value by $C\left\|g_{l,i}\right\|_R$, by Corollary LS.2 *1.*, while Corollary LS.2 *2.* indicates that the bound is uniformly $\mathscr{O}\left(1/h\right)$. The result now follows by the definition of trace. □

Proof of Lemma B.1.

(i) First notice that

$$\hat{\Delta}^H = 2n^{-1}\begin{bmatrix} \hat{P}_1 + \hat{\sigma}^{-2}R'R & \hat{\sigma}^{-2}R'\hat{X} & \hat{\sigma}^{-2}R'\hat{\Pi} \\ * & \hat{\sigma}^{-2}\hat{X}'\hat{X} & \hat{\sigma}^{-2}\hat{X}'\hat{\Pi} + \hat{\sigma}^{-2}\hat{K}_1 \\ * & * & \hat{\sigma}^{-2}\hat{\Pi}'\hat{\Pi} + \hat{\sigma}^{-2}\hat{K}_2 \end{bmatrix} \tag{S.9}$$

$$-2n^{-1}\begin{bmatrix} P_1 + \sigma_0^{-2}R'R & \sigma_0^{-2}R'X & \sigma_0^{-2}R'\Pi \\ * & \sigma_0^{-2}X'X & \sigma_0^{-2}X'\Pi + \sigma_0^{-2}K_1 \\ * & * & \sigma_0^{-2}\Pi'\Pi + \sigma_0^{-2}K_2 \end{bmatrix}. \tag{S.10}$$

Consider the block $\hat{\sigma}^{-2}R'\hat{X} - \sigma_0^{-2}R'X$. Adding and subtracting $\sigma_0^{-2}R'\hat{X}$ implies that it equals $\hat{\Delta}^{\bar{\sigma}^{-2}}R'\hat{X} + \sigma_0^{-2}R'\hat{\Delta}^X$. Manipulating blocks similarly, the components of $\hat{\Delta}^H$ are $V_1 = 2n^{-1}\hat{\Delta}^{P_1}$, $V_2 = 2n^{-1}\hat{\Delta}^{\bar{\sigma}^{-2}}R'R$, $V_3 = 2n^{-1}\hat{\Delta}^{\bar{\sigma}^{-2}}R'\hat{X}$,
$V_4 = 2n^{-1}\sigma_0^{-2}R'\hat{\Delta}^X$, $V_5 = 2n^{-1}\hat{\Delta}^{\bar{\sigma}^{-2}}R'\hat{\Pi}$, $V_6 = 2n^{-1}\sigma_0^{-2}R'\hat{\Delta}^{\Pi}$,
$V_7 = 2n^{-1}\hat{X}'\left(\hat{X}\hat{\Delta}^{\bar{\sigma}^{-2}} + \sigma_0^{-2}\hat{\Delta}^X\right)$, $V_8 = 2n^{-1}\left(\hat{X}'\hat{\Delta}^{\bar{\sigma}^{-2}} + \sigma_0^{-2}\hat{\Delta}^{X'}\right)X$,
$V_9 = 2n^{-1}\left(\hat{X}'\hat{\Delta}^{\bar{\sigma}^{-2}} + \sigma_0^{-2}\hat{\Delta}^{X'}\right)\Pi$, $V_{10} = 2n^{-1}\hat{X}'\left(\hat{\Pi}\hat{\Delta}^{\bar{\sigma}^{-2}} + \sigma_0^{-2}\hat{\Delta}^{\Pi}\right)$,
$V_{11}(\text{typical column}) = 2n^{-1}\hat{\sigma}^{-2}\left(\partial\hat{X}/\partial\delta_i\right)'\left(R\hat{\kappa} + \hat{\Delta}^X\beta + X\hat{\Delta}^{\bar{\beta}}\right)$,
$V_{12}(\text{typical column}) = -2n^{-1}\hat{\Delta}^{\bar{\sigma}^{-2}}\left(\partial\hat{X}/\partial\delta_i\right)'u$,
$V_{13}(\text{typical column}) = -2n^{-1}\sigma_0^{-2}\hat{\Delta}^{(\partial X/\partial\delta_i)'}u$, $V_{14} = 2n^{-1}\left(\hat{\Pi}'\hat{\Delta}^{\bar{\sigma}^{-2}} + \sigma_0^{-2}\hat{\Delta}^{\Pi'}\right)\Pi$,
$V_{15}(\text{typical element}) = 2n^{-1}\hat{\sigma}^{-2}\hat{\beta}'\left(\partial^2\hat{X}/\partial\delta_i\partial\delta_j\right)'\left(R\hat{\kappa} + \hat{\Delta}^X\beta + X\hat{\Delta}^{\bar{\beta}}\right)$,
$V_{16}(\text{typical element}) = -2n^{-1}\hat{\Delta}^{\bar{\sigma}^{-2}}\hat{\beta}'\left(\partial^2\hat{X}/\partial\delta_i\partial\delta_j\right)'u$,
$V_{17}(\text{typical element}) = -2n^{-1}\sigma_0^{-2}\hat{\Delta}^{\bar{\beta}'}\left(\partial^2 X/\partial\delta_i\partial\delta_j\right)'u$ and
$V_{18}(\text{typical element}) = -2n^{-1}\sigma_0^{-2}\hat{\beta}'\hat{\Delta}^{\left(\partial^2 X/\partial\delta_i\partial\delta_j\right)'}u$.

By the triangle inequality $\left\| \hat{\Delta}^H \right\| \leq 2\sum_{i=1}^{18} \|V_i\|$. $\|V_1\|$ is bounded by

$$\left\{ \sum_{i,j=1}^{p} \left(2n^{-1}tr\left(\hat{G}_j\hat{G}_i\right) - 2n^{-1}tr(G_jG_i) \right)^2 \right\}^{\frac{1}{2}} \tag{S.11}$$

By the mean value theorem,

$$tr\left(\hat{G}_j\hat{G}_i\right) = tr(G_jG_i) + \overline{\overline{\mu}}'_{ij}\hat{\kappa},$$

where $\overline{\overline{\mu}}_{ij} = \left(tr\left(\overline{\overline{\mu}}_{ij,1}\right), \ldots, tr\left(\overline{\overline{\mu}}_{ij,p}\right)\right)'$, with

$$\overline{\overline{\mu}}_{ij,k} = G_i\left(\overline{\overline{\lambda}}\right)G_k\left(\overline{\overline{\lambda}}\right)G_j\left(\overline{\overline{\lambda}}\right) + G_{kn}\left(\overline{\overline{\lambda}}\right)G_i\left(\overline{\overline{\lambda}}\right)G_j\left(\overline{\overline{\lambda}}\right)$$

and $\left\|\overline{\overline{\lambda}} - \lambda_0\right\| \leq \|\hat{\kappa}\|$. Therefore the summands in (S.11) are

$$4n^{-2}\left(\overline{\overline{\mu}}'_{ij,n}\hat{\kappa}\right)^2 \leq 4n^{-2}\left\|\overline{\overline{\mu}}_{ij,n}\right\|^2\|\hat{\kappa}\|^2,$$

by Cauchy-Schwarz inequality, where the first factor in norms on the RHS is $\mathscr{O}\left(pn^2/h^2\right)$ by Lemma LS.3. The second factor is bounded by $\left\|\hat{\theta} - \theta_0\right\|^2 = \mathscr{O}_p\left(\|\xi\|^2\right)$, so $\|V_1\| = \mathscr{O}_p\left(p^2k/n^{\frac{1}{2}}h\right)$. Assumptions 19/20 also imply $n^{-\frac{1}{2}}\|R\| = \mathscr{O}_p(1)$. Using the last bound and by (S.5), Assumption 14,

$$\|V_2\| = \mathscr{O}_p\left(\left|\hat{\Delta}^{\sigma^2}\right|\right) = \mathscr{O}_p\left(n^{-\frac{1}{2}}h^{-\frac{1}{2}}p^{\frac{1}{2}}k^{\frac{5}{2}}\right).$$

We now derive appropriate bounds for terms involving Π. Indeed by Assumptions 19 or 20 we have $n^{-\frac{1}{2}}\|\Pi\| = \mathscr{O}(1)$. To show

$$n^{-\frac{1}{2}}\left\|\hat{\Pi}\right\| = \mathscr{O}_p(1) \tag{S.12}$$

note that, like in (S.1), we have $\left\|\hat{\Pi}\right\| \leq \left\|\hat{\Delta}^{\Pi}\right\| + \|\Pi\|$, and

$$\left\|\hat{\Delta}^{\Pi}\right\| \leq \left\|\hat{\Delta}^{(\partial X/\partial\delta_i)}\right\|\left\|\hat{\beta}\right\| + \left\|\frac{\partial X}{\partial\delta_i}\right\|\left\|\hat{\Delta}^{\overline{\beta}}\right\| = \mathscr{O}_p\left(n^{\frac{1}{2}}\|\xi\|\max\{k,1\}\right) = \mathscr{O}_p\left(n^{\frac{1}{2}}k\|\xi\|\right), \tag{S.13}$$

so (S.12) follows if $k\|\xi\| = \mathscr{O}_p\left(p^{\frac{1}{2}}k^2/n^{\frac{1}{2}}\right)$ is negligible, which is true by (3.14) or

(3.16). Thus we have $(\|V_3\|, \|V_5\|) = \mathscr{O}_p\left(\left|\hat{\Delta}^{\sigma^2}\right|\right) = \mathscr{O}_p(\|V_2\|)$. Next,

$$\|V_4\| = \mathscr{O}_p\left(n^{-\frac{1}{2}}k^{\frac{1}{2}}\|\xi\|\right) = \mathscr{O}_p\left(n^{-1}h^{-\frac{1}{2}}p^{\frac{1}{2}}k^{\frac{3}{2}}\right),$$

by (S.1), with similar arguments implying $\|V_6\| = \mathscr{O}_p(\|V_4\|)$. Similarly we derive

$$(\|V_7\|, \|V_8\|, \|V_9\|, \|V_{10}\|, \|V_{14}\|) = \mathscr{O}_p\left(\max\left\{\|V_2\|, n^{\frac{1}{2}}\|V_4\|\right\}\right).$$

Assumption 18 implies that

$$\left(\left\|\hat{\Delta}^{(\partial X/\partial\delta_i)}\right\|, \left\|\hat{\Delta}^{\left(\partial^2 X/\partial\delta_i\partial\delta_j\right)}\right\|\right) = \mathscr{O}_p\left(\|\xi\| n^{\frac{1}{2}}k^{\frac{1}{2}}\right), \tag{S.14}$$

proceeding exactly like in (S.1). Assumption 18 also implies that

$$n^{-\frac{1}{2}}\left(\|\partial X/\partial\delta_i\|, \left\|\partial^2 X/\partial\delta_i\partial\delta_j\right\|\right) = \mathscr{O}_p(1), \tag{S.15}$$

so combining (S.14) and (S.15) we obtain

$$n^{-\frac{1}{2}}\left(\left\|\partial^2\hat{X}/\partial\delta_i\right\|, \left\|\partial^2\hat{X}/\partial\delta_i\partial\delta_j\right\|\right) = \mathscr{O}_p(1), \tag{S.16}$$

because $\|\xi\| k^{\frac{1}{2}} = o_p(1)$, just like we obtained (S.2). Assumptions 19 or 20 together with Lemma B.3, (S.14), (S.15) and (S.16) yield

$$\begin{aligned}
(\|V_{11}\|, \|V_{15}\|) &= \mathscr{O}_p\left(\max\left\{\|\xi\|, n^{\frac{1}{2}}\|V_4\|\right\}\right), \\
(\|V_{12}\|, \|V_{13}\|, \|V_{16}\|, \|V_{17}\|, \|V_{18}\|) &= \mathscr{O}_p\left(\max\left\{\|V_2\|, n^{\frac{1}{2}}\|V_4\|\right\}\right).
\end{aligned}$$

Thus

$$\begin{aligned}
\left\|\hat{\Delta}^H\right\| &= \mathscr{O}_p\left(\max\left\{\|V_1\|, \|V_2\|, n^{\frac{1}{2}}\|V_4\|\right\}\right) \\
&= \mathscr{O}_p\left(n^{-\frac{1}{2}}h^{-\frac{1}{2}}p^{\frac{1}{2}}k\left(h^{-\frac{1}{2}}p^{\frac{3}{2}} + k^{\frac{3}{2}}\right)\right).
\end{aligned}$$

The result for $\bar{\Delta}^H$ follows identically because $\left\|\bar{\theta} - \theta_0\right\| \leq \left\|\hat{\theta} - \theta_0\right\|$.

(ii) We omit this because it follows exactly as the proof of (i) noting that $\hat{\kappa} = \mathscr{O}_p(\|\phi\|)$ for the pure SAR model and also utilising (S.6) in place of (S.5).

$\square$

Lemma LS.4. *Suppose Assumptions 1-14 hold. Then*

$$\left\|B'A\right\| = \left\|A'B\right\| = \mathscr{O}_p\left(n^{\frac{1}{2}}p^{\frac{1}{2}}k\right), \qquad \left\|X'B\right\| = \left\|B'X\right\| = \mathscr{O}_p\left(n^{\frac{1}{2}}k^{\frac{1}{2}}\right).$$

Proof. $B'A$ and $X'B$ have (i,j)-th element $(G_i u)' b_j$ and $\chi_i' G_j u$ respectively. Then

$$\begin{aligned}
\mathbb{E}\left\|B'A\right\|^2 &\leq \sum_{i=1}^{p}\sum_{j=1}^{p}\mathbb{E}\left(a_j' G_i u u' G_i' a_j\right) \leq \sigma_0^2 \sum_{i=1}^{p}\|G_i\|^2 \sum_{j=1}^{p}\|a_j\|^2 \leq Cnpk^2, \\
\mathbb{E}\left\|X'B\right\|^2 &\leq \sum_{i=1}^{k}\sum_{j=1}^{p}\mathbb{E}\left(\chi_i' G_j u u' G_j' \chi_i\right) \leq \sigma_0^2 \sum_{j=1}^{p}\|G_j\|^2 \sum_{i=1}^{k}\|\chi_i\|^2 \leq Cnk,
\end{aligned}$$

whence the claim follows by Markov's inequality.

□

Proof of Lemma B.2.

(i) $\|H - \Xi\|$ is bounded by

$$2\sigma_0^{-2}n^{-1}\left(2\left\|A'B\right\| + 2\left\|X'B\right\| + \left\|B'B - \sigma_0^2 P_2\right\|\right).$$

By Lemma LS.4 the first two terms inside parentheses are $\mathscr{O}_p\left(p^{\frac{1}{2}}kn^{\frac{1}{2}}\right)$ while the last is readily shown to be $\mathscr{O}_p\left(pn^{\frac{1}{2}}/h^{\frac{1}{2}}\right)$. Indeed $\mathbb{E}\left\|B'B - \sigma_0^2 P_2\right\|^2$ is bounded by

$$\sum_{i,j=1}^{p}\mathbb{E}\left(u' G_i' G_j u - \sigma_0^2 tr\left(G_i' G_j\right)\right)^2 = \sum_{i,j=1}^{p}\operatorname{var}\left(u' G_i' G_j u\right),$$

the summands on the RHS being

$$\left(\mu_4 - 3\sigma_0^4\right)\sum_{k=1}^{n}\left(G_i G_j'\right)_{kk}^2 + \sigma_0^4\left[tr\left\{\left(G_i G_j'\right)^2\right\} + tr\left(G_i G_j' G_j G_i'\right)\right] = \mathscr{O}\left(n/h\right), \quad \text{(S.17)}$$

by Lemma B.3. of Gupta and Robinson (2015), where $\left(G_i G_j'\right)_{lk}$ denotes the (l,k)-th element of $G_i G_j'$. Hence $\|H - \Xi\| = \mathscr{O}_p\left(\max\left\{p^{\frac{1}{2}}k/n^{\frac{1}{2}}, p/n^{\frac{1}{2}}h^{\frac{1}{2}}\right\}\right) = \mathscr{O}_p\left(pk/n^{\frac{1}{2}}\right)$ since h is bounded away from zero. The claim for $\|H - \Xi\|$ in the case without regressors follows easily when h is bounded, but we need to utilize (2.12) when h is divergent. For the latter case consider the two trace terms in (S.17). We show that the first one is bounded by the second one. Indeed, by Cauchy Schwarz inequality the first trace term is

$$tr\left\{\left(I_n' G_i G_j'\right)^2\right\} \leq tr\left(I_n I_n' G_i G_j' G_j G_i'\right) = tr\left(G_i G_j' G_j G_i'\right).$$

The second one equals $\left\|G_i G_j'\right\|_F^2$, so that

$$\sum_{i,j=1}^{p} \operatorname{var}\left(u' G_j' G_i u\right) = \mathscr{O}\left(\max\left\{p^2 n/h^2, \sum_{i,j=1}^{p} \left\|G_i G_j'\right\|_F^2\right\}\right),$$

the first term in braces arising because $\left(G_i G_j'\right)_{kk} = \mathscr{O}(1/h)$ as earlier. The second term in braces is bounded by

$$\left(\sum_{i=1}^{p} \|G_i\|_F^2\right)^2 \leq \left(\sum_{i=1}^{p} \left\|S^{-1}\right\|_F^2 \|W_i\|_F^2\right)^2 \leq C,$$

by (2.12). Thus $\|H - \Xi\| = \mathscr{O}_p\left(\max\left\{p/n^{\frac{1}{2}}h, 1/n\right\}\right)$, as desired.

(ii) We have

$$L - \sigma_0^2 \Xi/2 = -\left[I_p, 0\right]' \left[\sigma_0^2 n^{-1}\left(P_1 + P_2\right), 0\right],$$

which has squared norm bounded by a constant times $n^{-2}\sum_{i,j=1}^{p} tr^2\left(C_j G_i\right) = \mathscr{O}\left(p^2/h^2\right)$ (using Corollary LS.2).

□

Proof of Lemma B.3.

(i) We have

$$\left\|\hat{H}^{-1}\right\| \leq \left\|\hat{H}^{-1} - H^{-1}\right\| + \left\|H^{-1}\right\| \leq \left\|\hat{H}^{-1}\right\| \left\|\hat{H} - H\right\| \left\|H^{-1}\right\| + \left\|H^{-1}\right\|.$$

Therefore $\left\|\hat{H}^{-1}\right\|\left(1 - \left\|\hat{H} - H\right\| \left\|H^{-1}\right\|\right) \leq \left\|H^{-1}\right\|$. Similarly, we can argue that

$$\left\|H^{-1}\right\|\left(1 - \|H - \Xi\| \left\|\Xi^{-1}\right\|\right) \leq \left\|\Xi^{-1}\right\|$$

and

$$\left\|\Xi^{-1}\right\|\left(1 - \left\|\sigma_0^2 \Xi/2 - L\right\| \left\|L^{-1}\right\|\right) \leq \sigma_0^2 \left\|L^{-1}\right\|/2.$$

The result follows from Lemmas B.1 (i), B.2 together with (3.14) or (3.16) and Assumption 19.

(ii) Similar to (i), except utilising (2.11).

(iii) Again similar to (i), except utilising (2.13).

The claims for $\bar{H}$ follow similarly because $\left\| \bar{\theta} - \theta_0 \right\| \leq \left\| \hat{\theta} - \theta_0 \right\|$. □

References

Gupta, A., Robinson, P. M., 2015. Inference on higher-order spatial autoregressive models with increasingly many parameters. Journal of Econometrics, 186, 19-31.

Robinson, P. M., 1988. The stochastic difference between econometric statistics. Econometrica 56, 531–548.