Econometric Theory, **35**, 2019, 417–463.
doi:10.1017/S0266466618000142

ESTIMATION OF SPATIAL AUTOREGRESSIONS WITH STOCHASTIC WEIGHT MATRICES

ABHIMANYU GUPTA
University of Essex

We examine a higher-order spatial autoregressive model with stochastic, but exogenous, spatial weight matrices. Allowing a general spatial linear process form for the disturbances that permits many common types of error specifications as well as potential 'long memory', we provide sufficient conditions for consistency and asymptotic normality of instrumental variables, ordinary least squares, and pseudo maximum likelihood estimates. The implications of popular weight matrix normalizations and structures for our theoretical conditions are discussed. A set of Monte Carlo simulations examines the behaviour of the estimates in a variety of situations. Our results are especially pertinent in situations where spatial weights are functions of stochastic economic variables, and this type of setting is also studied in our simulations.

1. INTRODUCTION

Spatial autoregressive (SAR) models, due to Cliff and Ord (1973), have recently become very popular in applied and theoretical research. Their main feature is the modeling of spatial dependence by allowing a form of direct interaction between units through spatial weight matrices W_j, $j = 1, \ldots, p$. The elements of W_j measure distance between units, which may be geographic but in general can be (inverse) economic distances.

While the majority of the literature on estimation and inference for SAR models, e.g., Kelejian and Prucha (1998, 1999, 2001), Lee (2002, 2003, 2004), Robinson (2010), Lee and Liu (2010), Su and Jin (2010), Lee and Yu (2013), and Gupta and Robinson (2015), has assumed the W_j to be deterministic, examples abound that imply stochastically generated W_j. Most commonly a typical element of W_j is determined by economic variables that may themselves be stochastic. Conley and Ligon (2002) study cross-country spillovers in long-run growth rates using several distance measures. While one of them, geographic

I am grateful to co-editor Guido Kuersteiner and three anonymous referees for excellent comments that improved the article substantially. I thank Peter Robinson for several comments and am also grateful to Nicolas Debarsy, Javier Hidalgo, Ingmar Prucha, and Renata Rabovic for useful suggestions and discussions. Address correspondence to Abhimanyu Gupta, Department of Economics, University of Essex; e-mail: a.gupta@essex.ac.uk.

© Cambridge University Press 2018

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

distance, is evidently fixed, the other two measures, United Parcel Service shipping costs and airfare, are more difficult to justify to be fixed in repeated sampling. Both, at the very least, are subject to random shocks in the economic conditions of each pair of countries, among many other factors. Conley and Dupor (2003) take input–output relations as a measure of economic distance, and it is reasonable to imagine that these relations are stochastic and not fixed. In Yuzefovich (2003) spatial weight matrices are constructed using a variety of economic distances, e.g., trade between two countries and competition in borrowing from a common lender. These variables would generally be considered stochastic in econometric analyses that use such data. Another example is Baltagi, Fingleton, and Pirotte (2014), who construct a weight matrix using commuting frequencies between districts in the United Kingdom. Commuting frequencies between two districts depend heavily on macro and microeconomic factors that are stochastic, and therefore may be anticipated to be stochastic too. Souza (2015) considers a SAR model in which networks may form stochastically, captured by nonzero spatial weight matrix elements. However he treats these elements as unobserved heterogeneity, and hence unknown, whereas in our treatment they are known. Robinson (2008) discusses a SAR with stochastic weights in the context of correlation testing. In a recent survey, Boucher and Fortin (2016) discuss the concept of stochastic spatial weight matrices from the perspective of impact caused by a policy shock. They stress that social interaction models based on a stochastic network structure, such as the SAR model considered in this article, should include a description of the network formation process, and that researchers should discuss the impact of proposed policy shocks on the assumed network.

In this article we will justify instrumental variables (IV), ordinary least squares (OLS), and pseudo maximum likelihood estimates (PMLE) with stochastic but exogenous W_j. Asymptotic theory for IV estimates of SAR model parameters was introduced first in Kelejian and Prucha (1998), and subsequently also studied by Lee (2003). IV is employed because of endogeneity problems, but Lee (2002) demonstrated that OLS can deliver consistent and asymptotically normal estimates of SAR model parameters under certain circumstances, thus correcting a tendency to casually discard OLS as a suitable method for SAR estimation and inference. A more general treatment by Gupta and Robinson (2015) examined IV and OLS estimates for an increasing order version of the SAR model, but with iid disturbances. For PMLE, Lee (2004) developed asymptotic theory for SAR models with a single weight matrix, while Gupta and Robinson (2018) consider SAR models in which both the number of weight matrices and regressors can diverge with sample size. Both articles take independent disturbances that are also identically distributed in the case of Lee (2004) but need not be in Gupta and Robinson (2018). On the other hand, Delgado and Robinson (2015) provide asymptotic theory for estimation of spatial models of a rather general type, including SAR, spatial moving average (SMA), and spatial ARMA, allowing a linear process type disturbance structure for their distributional results.

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

Theory has been developed for estimation with endogenous W_j, e.g., Kelejian and Piras (2014) consider such a model and develop IV type estimates. Qu and Lee (2015) were critical of their restrictive assumptions, and instead used the near epoch dependence (NED) theory of Jenish and Prucha (2012) to establish consistency and asymptotic normality of estimates in a more general setting. Building on the spatial martingale limit theory in Kuersteiner and Prucha (2013), endogenous weight matrices are also permitted in a generalized method of moments framework in Kuersteiner and Prucha (2015). However the intermediate case, with stochastic but exogenous W_j has received little theoretical attention. This case can cover situations of economic interest where spatial weights are generated by exogenous regressors and can be examined in a very general framework that does not require NED process theory. For observations recorded at locations r and s the latter essentially requires the locations to be geographic (in the sense that they are in Euclidean space) due to a notion of dependence reducing as the distance between r and s increases. Thus it is not generally applicable to data whose locations do not have a geographic interpretation. On the other hand, the SAR model has been considered to be particularly appealing because of its ability to handle data in general economic spaces, such as income space, where geographical interpretations may not be natural and, indeed, locations may be unknown. If the locations indeed have a geographical interpretation, NED based theory provides powerful results and a greater ability to handle nonlinear models, see, e.g., Xu and Lee (2015). Thus neither approach dominates the other. The results in this article justify the use of methods developed for fixed W_j when the W_j are stochastic and, as a result, should provide reassurance to practitioners faced with such a problem, although the theory requires additional conditions to account for the stochastic nature of the weight matrices.

An additional innovation is that we allow for a general 'spatial linear process' structure in the disturbances, cf. Robinson and Thawornkaiwong (2012) and Delgado and Robinson (2015). The former do not consider models with spatial lags in the dependent variables explicitly, nor do they provide theory for OLS or PMLE, while the latter do not consider models with regressors. The spatial weights are deterministic in both articles. In this sense we make a novel contribution to the literature also in the deterministic W_j case that we formally cover.

The article is organized as follows: Section 2 contains asymptotic theory for IV estimates, Section 3 for OLS and Section 4 for PMLE. We discuss implications of common weight matrix normalizations and structures in Section 5, while Section 6 contains a Monte Carlo simulation study that includes two different regimes for generating the spatial weight matrices, with an additional regime discussed in an online appendix associated with this article, available at Cambridge Journals Online (journals.cambridge.org/ect). Section 7 concludes the article. Two appendices contain theorem proofs and technical lemmas.

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

2. IV ESTIMATION

Assume that, for an $n \times 1$ vector of observations y_n, an $n \times k$ matrix of regressors X_n and $n \times n$ weight matrices W_{jn}, $j = 1, \ldots, p$, there exist unknown vectors $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_p)'$ and $\beta \neq 0$ such that

$$y_n = \sum_{j=1}^{p} \lambda_j W_{jn} y_n + X_n \beta + u_n, \tag{2.1}$$

where u_n is an $n \times 1$ vector of unobserved disturbances. The W_{jn} in (2.1) are sometimes normalized in ways that make their elements dependent on n, e.g., row-normalization, and X_n may contain spatial lags of basic explanatory variables. Both points imply triangular arrays and justify the n subscripting in (2.1), but the linear process type structure we permit for the disturbances also entails n subscripting on these. Subsequently we will drop n subscripts for brevity, but will occasionally remind the reader of the n-dependence of certain quantities.

Let $Z = Z_n$ be a matrix of instruments with dimension $n \times p_1$, $p_1 \geq p$. Denoting $\theta = (\lambda', \beta')'$, define the IV estimate of θ as

$$\hat{\theta} = n^{-1} \bar{\bar{Q}}^{-1} \bar{\bar{K}}' \bar{J}^{-1} [Z, X]' y = \theta + \bar{\bar{Q}}^{-1} \bar{\bar{K}}' \bar{J}^{-1} q, \tag{2.2}$$

where $\bar{\bar{Q}} = \bar{\bar{Q}}_n = \bar{\bar{K}}' \bar{J}^{-1} \bar{\bar{K}}$ (dimension $p+k$) and $\bar{\bar{K}} = \bar{\bar{K}}_n = n^{-1} [Z, X]' [R, X]$ (dimension $(p_1+k) \times (p+k)$), with $R = [W_1 y, \ldots, W_p y]$, $\bar{J} = \bar{J}_n = n^{-1} [Z, X]' [Z, X]$ (dimension p_1+k), $q = n^{-1} [Z, X]' u$. Throughout the article C denotes a generic positive constant, arbitrarily large but bounded and independent of n.

Assumption 1. (2.1) holds with $u = u_n = (u_{1n}, \ldots, u_{nn})'$ and

$$u_{rn} = u_r = \sum_{l=1}^{\infty} c_{rl} \epsilon_l, \quad r = 1, \ldots, n, \quad n \geq 1, \tag{2.3}$$

where ϵ_l are scalar independent random variables with zero mean and variance σ^2, $c_{rl} = c_{rln}$, and satisfy

$$\sum_{l=1}^{\infty} c_{rl}^2 < C, \quad r = 1, \ldots, n, \quad n \geq 1. \tag{2.4}$$

Assumption 2. The elements of W_j, $j = 1, \ldots, p$, are random variables that are uniformly $\mathcal{O}_p(h_{W_n}^{-1})$, as $n \to \infty$, with $h_{W_n} = h_W$ a bounded or divergent sequence that satisfies $h_W = o(n)$ if it is divergent. h_W is bounded away from zero.

Assumption 1 permits a wide variety of disturbance processes including SAR and SMA, and implies that each u_r forms a triangular array. The square summability of linear process coefficients in (2.4) allows spatial 'long-memory', while

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

identity of distribution of the ϵ_l is avoided. Robinson and Thawornkaiwong (2012), who introduced this assumption, discussed it in detail. The time series literature commonly allows for martingale ϵ_l, but this is avoided in spatial settings as there may be no natural ordering available. Assumption 2 is an extension to stochastic weights of a commonly employed assumption that controls spatial weights, cf. Lee (2002, 2004), Gupta and Robinson (2015). Allowing h_W to be bounded can imply a fixed number of neighbours as $n \to \infty$.

Write $S = I_n - \sum_{j=1}^{p} \lambda_j W_j$, I_n denoting the n-dimensional identity matrix, and introduce:

Assumption 3. $P(S \text{ is nonsingular}) = 1$, for all sufficiently large n, at the true parameter value $\lambda = \lambda_0$.

For a generic matrix M, define $\|M\|$ as the square root of the largest eigenvalue of MM' (the spectral norm), $\|M\|_R$ as the largest absolute row-sum of M (the maximum row-sum norm) and $\|M\|_F = \left(tr\left(M'M\right)\right)^{\frac{1}{2}}$ (the Frobenius norm). Assumption 3 ensures that a reduced form exists almost everywhere (a.e.) for y, and is satisfied if $P\left(\left\|\sum_{j=1}^{p} \lambda_j W_j\right\| < 1\right) = 1$. Indeed, we can write (2.1) as

$$Sy = X\beta + u \tag{2.5}$$

or, equivalently, $y = R\lambda + X\beta + u$. Assumption 3 implies that

$$y = S^{-1}X\beta + S^{-1}u, \text{ a.e.,} \tag{2.6}$$

so $R = A + B$, where $A = \left[G_1 X\beta, \ldots, G_p X\beta\right]$, $B = \left[G_1 u, \ldots, G_p u\right]$, and $G_j = W_j S^{-1}$ for $j = 1, \ldots, p$. Also define $\bar{K} = \bar{K}_n = n^{-1}[Z, X]'[A, X]$, $\bar{Q} = \bar{Q}_n = \bar{K}'\bar{J}^{-1}\bar{K}$ and introduce user chosen real numbers ζ_i, $i = 1, \ldots, 12$, (because twelve such numbers will be needed in the subsequent results) such that $1 < \zeta_i < C$ for each i and $\zeta_j^{-1} + \zeta_{j+1}^{-1} = 1$ for odd j. The ζ_i will be used in Hölder inequalities in the proofs.

Assumption 4. X, W_j and z_r are independent of ϵ_l, $r = 1, \ldots, n, l = 1, \ldots, j = 1, \ldots, p$, where z_r is the r-th column of Z'. Let $a_{rjn} = a_{rj}$ denote the (r, j)-th element of $[Z, X]$. Then

$$\max_{1 \le r \le n,\ 1 \le j \le p_1 + k} \mathbb{E}\left|a_{rj}\right|^{2\zeta_1} < C, \tag{2.7}$$

and, as $n \to \infty$,

$$\bar{K} \xrightarrow{p} K,\ \bar{J} \xrightarrow{p} J, \tag{2.8}$$

where K and J are full-rank constant matrices, with J symmetric.

A consequence of Assumption 4 is that $\bar{Q} - Q = o_p(1)$, with $Q = K'J^{-1}K$. Condition (2.7) implies finite $2\zeta_1$-th moments for instruments and regressors. The

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

requirement of the whole regressor matrix X being independent of the ϵ_l stems from the fact the instruments are typically constructed using linearly independent columns of $W_j^s X$, $j = 1, \ldots, p$, $s \geq 1$, cf. Kelejian and Prucha (1998). Evidently a given instrument vector then contains elements from different rows of X, as was noted by Gupta and Robinson (2015).

Denote

$$\chi_n = n^{-\frac{1}{2}} \max_{1 \leq j \leq p} \left(\mathbb{E} \left\| W_j \right\|^{2\zeta_2\zeta_3} \right)^{\frac{1}{2\zeta_2\zeta_3}} \left(\mathbb{E} \left\| S^{-1} \right\|^{2\zeta_2\zeta_4} \right)^{\frac{1}{2\zeta_2\zeta_4}}. \tag{2.9}$$

THEOREM 2.1. *Let Assumptions 1–4 hold and*

$$\chi_n \to 0, \text{ as } n \to \infty. \tag{2.10}$$

Then $\hat{\theta} - \theta \xrightarrow{p} 0$, *as* $n \to \infty$.

The condition (2.10) limits the extent of spatial correlation. Note that (2.10) does not impose that $\left\| W_j \right\|$ or $\left\| S^{-1} \right\|$ have finite $2\zeta_2\zeta_3$-th or $2\zeta_2\zeta_4$-th moments, but allows these to grow with n. In this sense it is not as strong as may be imagined at first glance. We will look at a specific example with potentially unbounded moments in Section 5. There is an implication of being able to 'trade-off' the magnitude of moments of $\left\| W_j \right\|$ and $\left\| S^{-1} \right\|$ in χ_n and a_{rj} by choices of ζ_i, $i = 1, 2, 3, 4$. Some of the existing literature on SAR models with fixed weights imposes restrictions on $\left\| W_j \right\|_R$ or $\left\| S^{-1} \right\|_R$, but these are evidently stronger than those based on the spectral norm. Indeed, taking $\zeta_i = 2$, $i = 1, 2, 3, 4$, for simplicity, the inequality $\|F\|^2 \leq \|F\|_R \left\| F' \right\|_R$ immediately implies $\chi_n = \mathcal{O}\left(\chi_{n,R}\right)$ with

$$\chi_{n,R} = n^{-\frac{1}{2}} \max_{1 \leq j \leq p} \left(\mathbb{E} \left\| W_j \right\|_R^8 \mathbb{E} \left\| W_j' \right\|_R^8 \mathbb{E} \left\| S^{-1} \right\|_R^8 \mathbb{E} \left\| S'^{-1} \right\|_R^8 \right)^{\frac{1}{16}}.$$

Let $\mathbb{1}(\cdot)$ denote the indicator function.

Assumption 5. $\sup_{l \geq 1} \mathbb{E}\left(\epsilon_l^2 \mathbb{1}\left(|\epsilon_l| > \delta\right)\right) \to 0$, as $\delta \to \infty$.

Assumption 6. With a_r' denoting the r-th row of $[Z, X]$ and Φ a positive definite (p.d.) constant matrix,

$$n^{-1} \sum_{r,s=1}^{n} \sum_{l=1}^{\infty} c_{rl} c_{sl} a_r a_s' \xrightarrow{p} \Phi, \text{ as } n \to \infty, \tag{2.11}$$

$$n^{-1} \sup_{l \geq 1} \left\| \sum_{r=1}^{n} a_r c_{rl} \right\|^2 \xrightarrow{p} 0, \text{ as } n \to \infty. \tag{2.12}$$

Assumption 5 avoids identity of distribution for the ϵ_l, (2.11) simply asserts convergence of the covariance matrix of $n^{-\frac{1}{2}}[Z, X]'u$ while (2.12) is the form of the Lindeberg condition required for the central limit theorem.

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

THEOREM 2.2. *Let Assumptions 1–6 and (2.10) hold. Then*

$$n^{\frac{1}{2}}\left(\hat{\theta}-\theta\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2 Q^{-1} K' J^{-1} \Phi J^{-1} K Q^{-1}\right), \text{ as } n \to \infty.$$

3. OLS ESTIMATION

Define the OLS estimate

$$\tilde{\theta} = n^{-1} \bar{\bar{L}}^{-1} [R, X]' y = \theta + \bar{\bar{L}}^{-1} w, \tag{3.1}$$

where $\bar{\bar{L}} = \bar{\bar{L}}_n = n^{-1}[R, X]'[R, X]$ (dimension $p+k$), $\ell = \ell_n = n^{-1}[R, X]' u$. Also define $\bar{L} = \bar{L}_n = n^{-1}[A, X]'[A, X]$. Assumption 2 needs to be strengthened to the following sufficient condition:

Assumption 7. The ζ_i are chosen such that $\zeta_5\zeta_7 = 2\zeta_{11}$ and

$$\max_{1\le j\le p} \mathbb{E}\left(\left\{\max_{1\le r,s\le n} \left|w_{rs,j}\right|\right\}^{2\zeta_{11}}\right) = \mathcal{O}\left(h_W^{-2\zeta_{11}}\right),$$

where $w_{rs,j}$ is the (r,s)-th element of W_j, $j = 1, \ldots, p$.

This assumption implies $\max_{1\le r,s\le n,\ 1\le j\le p} \left|w_{rs,j}\right| = \mathcal{O}_{\mathrm{p}}\left(h_W^{-1}\right)$. Various bounds depending on the distribution of $w_{rs,j}$ exist in the extreme value literature for the expectation, but Assumption 7 ensures also that the familiar case with fixed $w_{rs,j} = \mathcal{O}\left(h_W^{-1}\right)$, uniformly in r, s, is formally covered. The restriction $\zeta_5\zeta_7 = 2\zeta_{11}$ is satisfied in the case where the Cauchy Schwarz inequality is used in place of the Hölder inequality, implying that $\zeta_i = 2$ for all i.

Assumption 8. X and W_j are independent of ϵ_l, $l = 1, \ldots, j = 1, \ldots, p$. Let $t_{rjn} = t_{rj}$ denote the (r, j)-th element of $[A, X]$. Then

$$\max_{1\le r\le n,\ 1\le j\le p+k} \mathbb{E}\left|t_{rj}\right|^{2\zeta_1} < C, \tag{3.2}$$

and, as $n \to \infty$,

$$\bar{L} \xrightarrow{p} L, \tag{3.3}$$

where L is a constant, symmetric and nonsingular matrix.

Define

$$\pi_n = h_W^{-\frac{1}{2}} \left(\max_{1\le j\le p} \mathbb{E}\left\| W_j' \right\|_R^{\zeta_6\zeta_9}\right)^{\frac{1}{2\zeta_6\zeta_9}} \left(\mathbb{E}\left\| S'^{-1} \right\|_R^{\zeta_5\zeta_8}\right)^{\frac{1}{2\zeta_5\zeta_8}} \left(\mathbb{E}\left\| S'^{-1} \right\|_R^{\zeta_6\zeta_{10}}\right)^{\frac{1}{2\zeta_6\zeta_{10}}}. \tag{3.4}$$

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

THEOREM 3.1. *Let Assumptions 1–3, 7, 8 hold and*

$$h_W^{-1} + \chi_n + \pi_n \longrightarrow 0 \text{ as } n \to \infty. \tag{3.5}$$

Then $\tilde{\theta} - \theta \xrightarrow{p} 0$, *as* $n \to \infty$.

For consistency of OLS estimates $h_W \to \infty$ is necessary even with deterministic W_j (cf. Lee, 2002; Gupta and Robinson, 2015), and (3.5) strengthens the restrictions on spatial correlation relative to h_W. Our conditions match those in the aforementioned literature except the additional requirements on moments of $\|W_j\|$ and $\|S^{-1}\|$ which arise because these are stochastic.

Taking $p = 1$ for simplicity (and writing $W_1 = W$), an example of nonstochastic W for which $h_W \to \infty$ can be constructed as in Case (1991), where each of d districts contains m farmers, so $n = md$. Interdistrict independence is assumed, as is equal reaction within districts, implying

$$W = I_d \otimes B_m, \text{ with } B_m = (m-1)^{-1}\left(l_m l_m' - I_m\right), \tag{3.6}$$

where l_m is the m-dimensional vector of ones and $\otimes$ denotes Kronecker product. Here $h_W = m - 1$, which diverges as $m \to \infty$, and thus satisfies (3.5). More general versions of such specifications are also studied recently in Hillier and Martellosio (2018).

The specification given in (3.6) consists of a block-diagonal W with very dense blocks. On the other hand, specifications of W can be of the more sparse variety, in which case h_W will be bounded and (3.5) will fail. As an example consider the case of circulant W, where each unit has only one neighbour on 'either side'. Mathematically, define W^* as the symmetric circulant matrix with first row elements given by

$$w_{1r}^* = \begin{cases} 0 & \text{if } r = 1 \text{ or } r = 3, \ldots, n-1; \\ \omega & \text{if } r = 2 \text{ or } r = n, \end{cases} \tag{3.7}$$

with ω some known real number, see, e.g., Das, Kelejian, and Prucha (2003). A binary neighbourhood criterion would take $\omega = 1$. Now define

$$W = W^* / \|W^*\|, \tag{3.8}$$

where $\|W^*\| = 2\omega$ because W^* is a symmetric, circulant matrix (see, e.g., Davis, 1979, p. 73). Then W is also a symmetric circulant matrix with first row elements given by $w_{1r}^*/2\omega$. In this case $h_W = 2\omega$, which does not diverge with n in general.

Assumption 9. $\sup_{l \geq 1} \mathbb{E}\left(\epsilon_l^4 \mathbb{1}\left(|\epsilon_l| > \delta\right)\right) \to 0$, as $\delta \to \infty$.

This assumption seems hard to relax for a CLT. Indeed, even for (2.1) with $p = 1$, W_j fixed and no linear process structure, Lee (2002) required $\mathbb{E}|u_r|^{4+\eta} < C$, for some $\eta > 0$. Gupta and Robinson (2015) relaxed this slightly to $\mathbb{E}u_r^4 < C$, with increasing p, k but restricted themselves to iid u_r. Here we avoid identity

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

of distribution of ϵ_l, and u_r, but require the uniform integrability of the ϵ_l^4 that Assumption 9 entails.

Assumption 10. With t_r' denoting the r-th row of $[A, X]$ and Ψ a p.d. constant matrix,

$$n^{-1}\sum_{r,s=1}^{n}\sum_{l=1}^{\infty}c_{rl}c_{sl}t_r t_s' \xrightarrow{p} \Psi, \text{ as } n\to\infty, \tag{3.9}$$

$$n^{-1}\sup_{l\geq 1}\left\|\sum_{r=1}^{n}t_r c_{rl}\right\|^2 \xrightarrow{p} 0, \text{ as } n\to\infty. \tag{3.10}$$

THEOREM 3.2. *Let Assumptions 1–3, 7–9, and* (3.5) *hold and*

$$h_W^{-1}n^{\frac{1}{2}}\left(\mathbb{E}\left\|S'^{-1}\right\|_R^{2\zeta_{12}}\right)^{\frac{1}{2\zeta_{12}}} \longrightarrow 0, \text{ as } n\to\infty. \tag{3.11}$$

Then

$$n^{\frac{1}{2}}\left(\tilde{\theta}-\theta\right)\xrightarrow{d}\mathcal{N}\left(0,\sigma^2L^{-1}\Psi L^{-1}\right), \text{ as } n\to\infty.$$

The proof requires some care to ensure that (2.4) does not need strengthening. Lee (2002) established that asymptotic normality of OLS relies not just on divergence of h_W, but sufficiently fast divergence, viz. $n^{\frac{1}{2}} = o(h_W)$. Condition (3.11) indicates the additional requirement that arises when the W_j are stochastic and reduces to the condition in Lee (2002) for nonstochastic W_j. In the W described in (3.6) we have $h_W = m-1$ and $n = md$, so (3.11) is satisfied if $d^{\frac{1}{2}}/m^{\frac{1}{2}}\to 0$ as $d, m\to\infty$. On the other hand, the W of (3.8) will not satisfy (3.11) in general.

With all $\zeta_i = 2$, we obtain

$$\chi_n = n^{-\frac{1}{2}}\left(\max_{1\leq j\leq p}\mathbb{E}\left\|W_j\right\|^8\mathbb{E}\left\|S^{-1}\right\|^8\right)^{\frac{1}{8}}, \tag{3.12}$$

$$\pi_n = h_W^{-\frac{1}{2}}\left(\max_{1\leq j\leq p}\mathbb{E}\left\|W_j'\right\|_R^4\right)^{\frac{1}{8}}\left(\mathbb{E}\left\|S'^{-1}\right\|_R^4\right)^{\frac{1}{4}}, \tag{3.13}$$

(3.11) becomes $h_W^{-1}n^{\frac{1}{2}}\left(\mathbb{E}\left\|S'^{-1}\right\|_R^4\right)^{\frac{1}{4}}\to 0$, as $n\to\infty$, and (2.7), (3.2) require finite fourth moments for the a_{rj} and t_{rj}, respectively.

In both Theorems 2.2 and 3.2, if the W_j are taken to be deterministic and the u_r are iid, the limit distributions are identical to those obtained by Gupta and Robinson (2015), taking p and k to be fixed in their results.

4. PSEUDO MAXIMUM LIKELIHOOD ESTIMATION

In this section, we parameterize the coefficients in the linear process given in Assumption 1. In particular, suppose there is an unknown $q\times 1$ vector $\tau_0\in\mathcal{T}$ such

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

that $c_{rl} = c_{rl}(\tau_0)$, all $r = 1, \ldots, n$ and $l \geq 1$. Writing $\sigma_0^{-2}\mathbb{E}\left(u_n u_n'\right) = \Xi_n(\tau_0) \equiv \Xi(\tau_0)$, the pseudo likelihood function based on Gaussianity and conditional on W_j, $j = 1, \ldots, p$, X, is

$$\log(2\pi\sigma^2) - \frac{2}{n}\log|S(\lambda)| + \frac{1}{n}\log|\Xi(\tau)| + \frac{1}{\sigma^2 n}(S(\lambda)y - X\beta)'\,\Xi(\tau)^{-1}(S(\lambda)y - X\beta), \tag{4.1}$$

at any admissible point $\left(\varrho', \sigma^2\right)'$ with $\varrho = \left(\theta', \tau'\right)'$, for nonsingular $S(\lambda)$ and $\Xi(\tau)$, although Gaussianity is nowhere assumed. True parameter values are identified by a zero subscript and for any parameter α and any quantity $x(\alpha)$ we adopt the convention $x(\alpha_0) \equiv x$.

For given $\gamma = \left(\lambda', \tau'\right)'$, (4.1) is minimised with respect to β and σ^2 by

$$\bar{\beta}(\gamma) = \left(X'\Xi(\tau)^{-1}X\right)^{-1}X'\Xi(\tau)^{-1}S(\lambda)y, \tag{4.2}$$

$$\bar{\sigma}^2(\gamma) = n^{-1}y'S'(\lambda)\,\Xi(\tau)'^{-\frac{1}{2}}M(\tau)\Xi(\tau)^{-\frac{1}{2}}S(\lambda)y, \tag{4.3}$$

where $M(\tau) = I_n - \Xi(\tau)^{-\frac{1}{2}}X\left(X'\Xi(\tau)'^{-\frac{1}{2}}\Xi(\tau)^{-\frac{1}{2}}X\right)^{-1}X'\Xi(\tau)'^{-\frac{1}{2}}$ and $\Xi(\tau)^{-\frac{1}{2}} \equiv \Xi_n(\tau)^{-\frac{1}{2}}$ is the $n \times n$ matrix such that $\Xi(\tau)'^{-\frac{1}{2}}\Xi(\tau)^{-\frac{1}{2}} = \Xi(\tau)^{-1}$. This is a slight abuse of notation but makes the proofs easier to read. The PMLE of γ_0 is $\check{\gamma} = \arg\min_{\gamma\in\Gamma}\mathcal{L}(\gamma)$, where

$$\mathcal{L}(\gamma) = \log\bar{\sigma}^2(\gamma) + n^{-1}\log\left|S'^{-1}(\lambda)\,\Xi(\tau)S^{-1}(\lambda)\right|, \tag{4.4}$$

and $\Gamma = \Lambda \times \mathcal{T}$ is taken to be a compact subset of $\mathbb{R}^{p+q}$. The PMLEs of β_0 and σ_0^2 are defined as $\bar{\beta}\left(\check{\gamma}\right) \equiv \check{\beta}$ and $\bar{\sigma}^2\left(\check{\gamma}\right) \equiv \check{\sigma}^2$, respectively.

Assumption 11. For all sufficiently large n, the u_r have uniformly bounded fourth moment, $r = 1, \ldots, n$.

Assumption 12. $\max_{j=1,\ldots,p}\left\|W_j\right\| + \left\|S^{-1}\right\| = \mathcal{O}_{\mathrm{p}}(1)$.

Assumption 13. $\overline{\lim}_{n\to\infty}\sup_{\tau\in\mathcal{T}}\left(\|\Xi(\tau)\| + \left\|\Xi(\tau)^{-1}\right\|\right) < \infty$.

Assumption 14. For any $\tau^\dagger \in \mathcal{T}$ and any $\eta > 0$, there exists $\varepsilon > 0$ such that

$$\overline{\lim_{n\to\infty}}\sup_{\tau\in\{\tau:\|\tau-\tau^\dagger\|<\varepsilon\}\cap\mathcal{T}}\left\|\Xi(\tau) - \Xi\left(\tau^\dagger\right)\right\| < \eta. \tag{4.5}$$

Assumption 12 implies that $\max_{j=1,\ldots,p}\left\|G_j\right\| = \mathcal{O}_{\mathrm{p}}(1)$. The approach to conditions on the W_j and S^{-1} differs somewhat in this section as compared to the previous two. There we focused on more primitive conditions based on moments of $\left\|W_j\right\|$ and $\left\|S^{-1}\right\|$, and also allowed a great amount of flexibility both in the way the Hölder inequality is applied (exemplified by the twelve different ζ_i) and the fact that the moments may diverge at a certain rate. The PMLE

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

considered in this section is inherently harder to derive asymptotic properties for because it is only implicitly defined, and matters are complicated further also by the presence of $\Xi(\tau)$ in the objective function (4.1). Thus to maintain tractability we prefer conditions of the type given in Assumption 12. Assumption 13 is a standard type of regularity condition that may be regarded as guaranteeing asymptotic boundedness and nonsingularity. Assumption 14 also holds with $\Xi(\tau) - \Xi\left(\tau^{\dagger}\right)$ replaced by $\Xi(\tau)^{-1} - \Xi\left(\tau^{\dagger}\right)^{-1}$ due to Assumption 13 and is useful in an equicontinuity argument in the consistency proof. Further discussion of it may be found in Delgado and Robinson (2015), who also discuss sufficient conditions for it to hold. Write $T(\lambda) = S(\lambda)S^{-1}$ and define $\sigma^2(\gamma) = n^{-1}\sigma_0^2 tr\left(T'(\lambda)\Xi(\tau)^{-1}T(\lambda)\Xi\right) = n^{-1}\sigma_0^2\left\|\Xi(\tau)^{-\frac{1}{2}}T(\lambda)\Xi^{\frac{1}{2}}\right\|_F^2$, which is nonnegative by definition and bounded in probability by Assumptions 12 and 13.

Assumption 15. There exist positive and finite constants c and C such that $c \leq \sigma^2(\gamma) \leq C$ with probability one, for all $\gamma \in \Gamma$.

Assumption 16. $\gamma_0 \in \Gamma$ and, for any $\eta > 0$,

$$P\left(\lim_{n\to\infty}\inf_{\gamma\in\overline{\mathcal{N}}^{\gamma}(\eta)}\frac{n^{-1}tr\left(T'(\lambda)\Xi(\tau)^{-1}T(\lambda)\Xi\right)}{\left|T'(\lambda)\Xi(\tau)^{-1}T(\lambda)\Xi\right|^{1/n}} > 1\right) = 1, \tag{4.6}$$

where $\overline{\mathcal{N}}^{\gamma}(\eta) = \Gamma \setminus \mathcal{N}^{\gamma}(\eta)$ and $\mathcal{N}^{\gamma}(\eta) = \{\gamma : \|\gamma - \gamma_0\| < \eta\} \cap \Gamma$.

The trace in the numerator inside the probability in (4.6) equals $\left\|\Xi(\tau)^{-\frac{1}{2}}T(\lambda)\Xi^{\frac{1}{2}}\right\|_F^2$, while the determinant in the denominator is similarly $\left|\Xi(\tau)^{-\frac{1}{2}}T(\lambda)\Xi^{\frac{1}{2}}\right|^2$. Thus the inequality between arithmetic and geometric means implies that the ratio inside the probability is at least one. On the other hand, when $\gamma = \gamma_0$ the ratio equals one, so (4.6) essentially requires that this situation not arise outside an arbitrarily small neighbourhood of γ_0. Therefore it is an identification condition which can be compared to those already in the literature. The closest conditions may be found in Gupta and Robinson (2018) and Delgado and Robinson (2015), the latter showing the equivalence of their condition to that of Lee (2004) in the case of SAR models. With deterministic weight matrices, the latter interprets the condition as related to the uniqueness of the covariance matrix of y. Our setup is rather different, with a SAR model coupled with linear process disturbances and stochastic weight matrices, for which our condition accounts. A Lee (2004)-like interpretation is possible in terms of the uniqueness of the covariance matrix of y conditional on the W_j and X, $j = 1, \ldots, p$, viz. $S^{-1}\Xi S'^{-1}$, because the logarithm of the ratio inside the probability equals $n^{-1}\log\left|\sigma^2(\gamma)S(\lambda)^{-1}\Xi(\tau)S'(\lambda)^{-1}\right| - n^{-1}\log\left|\sigma_0^2 S^{-1}\Xi S'^{-1}\right|$, cf. also equation (37) and the discussion thereafter in Delgado and Robinson (2015).

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

Assumption 17. $\beta_0 \neq 0$ and, for any $\eta > 0$,

$$P\left(\lim_{n\to\infty} \inf_{(\lambda',\tau')'\in\Lambda\times\overline{\mathcal{N}}^{\tau}(\eta)} n^{-1}\beta_0' X'T'(\lambda)\Xi(\tau)'^{-\frac{1}{2}} M(\tau)\,\Xi(\tau)^{-\frac{1}{2}} T(\lambda)X\beta_0 / \|\beta_0\|^2 > 0\right) = 1. \tag{4.7}$$

Because the event inside the probability is equivalent (by straightforward minimization) to the event

$$\lim_{n\to\infty} \min_{\beta\in\mathbb{R}^k} \inf_{(\lambda',\tau')'\in\Lambda\times\overline{\mathcal{N}}^{\tau}(\eta)} n^{-1}(X\beta - T(\lambda)X\beta_0)'\,\Xi(\tau)^{-1}\,(X\beta - T(\lambda)X\beta_0)\,/\,\|\beta_0\|^2 > 0,$$

Assumption 17 is in fact an identification condition similar to one used for nonlinear regression in a SAR model by Gupta and Robinson (2018). Let $\overline{\varphi}(A)$ (respectively $\underline{\varphi}(A)$) denote the largest (respectively smallest) eigenvalue of a generic square nonnegative definite matrix A. Then a necessary condition for the event in (4.7) is that

$$\lim_{n\to\infty} \inf_{(\lambda',\tau')'\in\Lambda\times\overline{\mathcal{N}}^{\tau}(\eta)} \underline{\varphi}\left(n^{-1}X'T'(\lambda)\Xi(\tau)'^{-\frac{1}{2}} M(\tau)\,\Xi(\tau)^{-\frac{1}{2}} T(\lambda)X\right) > 0.$$

The final identification condition is a familiar asymptotic full-rank condition:

Assumption 18. $\left\{\underline{\varphi}\left(n^{-1}X'X\right)\right\}^{-1} = \mathcal{O}_{\mathrm{p}}(1)$.

This assumption implies $\left\{\sup_{\tau\in\mathcal{T}} \underline{\varphi}\left(n^{-1}X'\Xi(\tau)^{-1}X\right)\right\}^{-1} = \mathcal{O}_{\mathrm{p}}(1)$ in view of Assumption 13.

THEOREM 4.1. *Let Assumptions 1–3, 8, and 11–17 hold. Then* $\left(\breve{\varrho}', \breve{\sigma}^2\right)' - \left(\varrho_0', \sigma_0{}^2\right)' \xrightarrow{p} 0$ *as* $n \to \infty$.

Assumption 19. γ_0 is an interior point of Γ.

Assumption 20. The ϵ_j in Assumption 1 have finite third and fourth moments μ_3 and μ_4, respectively, finite eighth moment and, denoting by $\xi_{rs}(\tau)$ the (r,s)-th element of $\Xi(\tau)$ and defining

$$c_{rj}^* = c_{rj}/\xi_{rr}^{\frac{1}{2}},\ r = 1,\ldots,n,\ n \geq 1, l \geq 1,$$

we have

$$\overline{\lim_{n\to\infty}} \sup_{r=1,\ldots,n} \sum_{l=1}^{\infty} \left|c_{rl}^*\right| + \sup_{l\geq 1} \overline{\lim_{n\to\infty}} \sum_{r=1}^{n} \left|c_{rl}^*\right| < \infty. \tag{4.8}$$

Assumption 19 is standard for a central limit theorem for implicitly defined estimates. Assumption 20 is introduced by Delgado and Robinson (2015), and is discussed therein. It implies that Assumption 11 holds.

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

Assumption 21. $\max_{j=1,\ldots,p} \|W_j\|_R + \max_{j=1,\ldots,p} \|W_j'\|_R + \|S^{-1}\|_R + \|S'^{-1}\|_R + \|X\|_R = \mathcal{O}_p(1)$, as $n \to \infty$.

Assumption 21 is quite standard even in the literature with nonstochastic W_j and X, where the $\mathcal{O}_p(1)$ requirement is replaced by simply $\mathcal{O}(1)$. Note that $\|X\|_R = \mathcal{O}(1)$ would then be satisfied if X has uniformly bounded, constant elements. Let $\Xi_{i_1}(\tau) = \partial \Xi(\tau)/\partial \tau_{i_1}$, $\Xi_{i_1 i_2}(\tau) = \partial \Xi_{i_1}(\tau)/\partial \tau_{i_2}$, and $\Xi_{i_1 i_2 i_3}(\tau) = \partial \Xi_{i_1 i_2}(\tau)/\partial \tau_{i_3}$, $i_1, i_2, i_3 = 1, \ldots, q$, where the matrices are differentiated element-wise and the existence of the derivatives is guaranteed by Assumption 22 below, with typical elements ξ_{rsi_1}, $\xi_{rsi_1 i_2}$, and $\xi_{rsi_1 i_2 i_3}$, respectively.

Assumption 22. For $i_1, i_2, i_3 = 1, \ldots, q$ and all sufficiently large n, the elements of Ξ are thrice continuously differentiable, for any $\eta > 0$

$$\overline{\lim_{n\to\infty}} \sup_{\tau \in \mathcal{N}^\tau(\eta)} \left(\left\| \Xi(\tau)^{-1} \right\|_R + \left\| \Xi_{i_1}(\tau) \right\|_R + \left\| \Xi_{i_1 i_2}(\tau) \right\|_R + \left\| \Xi_{i_1 i_2 i_3}(\tau) \right\|_R \right) < \infty, \tag{4.9}$$

and for a positive sequence $h_\Xi \equiv h_{\Xi n}$ that is either bounded or divergent and satisfies $h_\Xi / n \to 0$ as $n \to \infty$, we have

$$\overline{\lim_{n\to\infty}} \sup_{\tau \in \mathcal{N}^\tau(\eta)} \sup_{r,s=1,\ldots,n} h_\Xi \left(|\xi_{rs}(\tau)| + \left|\xi_{rsi_1}(\tau)\right| + \left|\xi_{rsi_1 i_2}(\tau)\right| + \left|\xi_{rsi_1 i_2 i_3}(\tau)\right| \right) < \infty. \tag{4.10}$$

Assumptions 21 and 22 serve to control the extent of spatial correlation to a greater degree than Assumptions 12 and 13 in view of the inequalities $\|A\|^2 \leq \|A\|_R \|A'\|_R$ and $\|A\| \leq \|A\|_F$ for a generic matrix A. Assumption 22 also imposes local smoothness conditions on $\Xi(\tau)$.

The second derivative matrix of (4.1) at (ϱ, σ^2) is denoted $H(\varrho, \sigma^2)$. Let $P_{G1}(\lambda)$ and $P_{G2}(\lambda)$ denote the $p \times p$ matrices with (i,j)-th elements $tr\left(G_j(\lambda) G_i(\lambda)\right)$, $tr\left(G_j'(\lambda) G_i(\lambda)\right)$, respectively, $P_\Xi(\tau)$ the $q \times q$ matrix with (i,j)-th element $tr\left(\Xi(\tau)^{-1} \Xi_i(\tau) \Xi(\tau)^{-1} \Xi_j(\tau)\right)$ and $\Upsilon(\tau)$ the $q \times q$ matrix with (i,j)-th element $tr\left(G_i \Xi(\tau)^{-1} \Xi_j(\tau)\right)$. Then

$$\Sigma = \mathbb{E}(H) = 2\sigma_0^{-2} n^{-1} \begin{bmatrix} \sigma_0^2 (P_{G1} + P_{G2}) + A' \Xi^{-1} A & A' \Xi^{-1} X & \sigma_0^2 \Upsilon \\ * & X' \Xi^{-1} X & 0 \\ * & * & 2^{-1} \sigma_0^2 P_\Xi \end{bmatrix}, \tag{4.11}$$

where the expectation is taken conditional on W_j and X, $j = 1, \ldots, p$. We also need the conditional covariance matrix of the first derivative of (4.1) at the true parameter values, which we write as $n^{-1}(2\Sigma + \Omega)$. The next paragraph is devoted to defining Ω, for which purpose let $c_l(\tau)$ be the $n \times 1$ vector with r-th element $c_{rl}(\tau)$.

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

Introduce the following matrices: the $n \times p$ and $p \times p$ (respectively) matrices F_1 and F_2 with j-th column $\sum_{l=1}^{\infty} \Xi^{-1} c_l c_l' G_j \Xi^{-1} c_l$ and (i,j)-th element $\sum_{l=1}^{\infty} c_l' G_i \Xi^{-1} c_l c_l' G_j \Xi^{-1} c_l$, respectively, the $n \times p$ matrix F_3 with j-th column $\sum_{l=1}^{\infty} \Xi^{-1} c_l c_l' \Xi^{-1} G_j' c_l$, the $n \times q$ and $p \times q$ (respectively) matrices F_4 and F_5 with j-th column $\sum_{l=1}^{\infty} \Xi^{-1} c_l c_l' \Xi^{-1} \Xi_j \Xi^{-1} c_l$ and (i,j)-th element $\sum_{l=1}^{\infty} c_l' G_i \Xi^{-1} c_l c_l' \Xi^{-1} \Xi_j \Xi^{-1} c_l$, respectively, the $n \times q$ matrix F_6 with j-th column $\sum_{l=1}^{\infty} \Xi^{-1} c_l c_l' \Xi^{-1} \Xi_i \Xi^{-1} c_l$ and the $q \times q$ matrix F_7 with (i,j)-th element given by $\sum_{l=1}^{\infty} c_l' \Xi^{-1} \Xi_i \Xi^{-1} c_l c_l' \Xi^{-1} \Xi_j \Xi^{-1} c_l$. Then

$$\Omega = \sigma_0^{-4} n^{-1} \begin{bmatrix} 4\mu_3 \left(A' F_1 + F_1' A\right) + 4\left(\mu_4 - 3\sigma_0^4\right) F_2 & 4\mu_3 F_3' X & 2\mu_3 A' F_4 + 2\left(\mu_4 - 3\sigma_0^4\right) F_5 \\ * & 0 & 2\mu_3 X' F_6 \\ * & * & \left(\mu_4 - 3\sigma_0^4\right) F_7 \end{bmatrix}. \tag{4.12}$$

Let $N = diag\left[I_p, I_k, I_q h_\Xi^{\frac{1}{2}}\right]$.

Assumption 23. The matrices $\Delta = \text{plim}_{n\to\infty} N(2\Sigma + \Omega)N$ and $\Pi = \text{plim}_{n\to\infty} N\Sigma N$ exist and are positive definite.

THEOREM 4.2. *Let Assumptions 1–3, 8, and 12–23 hold. Then*

$$n^{\frac{1}{2}} N^{-1} \left(\check{\varrho} - \varrho_0\right) \xrightarrow{d} \mathcal{N}\left(0, \Pi^{-1} \Delta \Pi^{-1}\right), \text{ as } n \to \infty.$$

The differential norming entailed by N^{-1} is due to the fact that the elements of P_Ξ are $\mathcal{O}\left(h_\Xi^{-1}\right)$ and will thus all converge to zero when h_Ξ is divergent, causing singularity of the limiting covariance matrix. The implication is that when h_Ξ diverges $\check{\tau}$ is $n^{\frac{1}{2}}/h_\Xi^{\frac{1}{2}}$-consistent, whereas $n^{\frac{1}{2}}$-consistency is achieved when h_Ξ is bounded.

Some simplifications are possible in Π and Δ when h_W is divergent. The reason is as follows: asymptotic normality of $\check{\varrho} - \varrho_0$ is established via the asymptotic normality of the first derivative of (4.1) with respect to ϱ. This derivative, given by (A.36) in Appendix 1, involves the quadratic in u terms $n^{-1} u' G_j \Xi^{-1} u - n^{-1} \sigma_0^2 tr G_j$, $j = 1, \ldots, p$, which are readily shown to be negligible when h_W diverges. We omit the proof of this negligibility to conserve space, and also because it is straightforward given the techniques used in the proof of Theorem 4.2. Thus these quadratic entries can be ignored while establishing asymptotic normality and do not contribute to the asymptotic covariance matrix, eliminating P_{G_2} in Σ and all blocks of Ω involving F_1, F_2, and F_3. Also, because $n^{-1} P_{G_1}$ has elements that are uniformly $\mathcal{O}_{\text{p}}\left(h_W^{-1}\right)$ by Lemma B.6, this part of Σ is negligible if h_W diverges. In sum, we can replace Σ and Ω by

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

$$2\sigma_0^{-2}n^{-1}\begin{bmatrix} A'\Xi^{-1}A & A'\Xi^{-1}X & \sigma_0^2\Upsilon \\ * & X'\Xi^{-1}X & 0 \\ * & * & 2^{-1}\sigma_0^2 P_\Xi \end{bmatrix}$$

and

$$\sigma_0^{-4}n^{-1}\begin{bmatrix} 0\,0\,2\mu_3 A'F_4 + 2\left(\mu_4 - 3\sigma_0^4\right)F_5 \\ *\,0 \qquad 2\mu_3 X'F_6 \\ *\,* \qquad \left(\mu_4 - 3\sigma_0^4\right)F_7 \end{bmatrix},$$

respectively.

Furthermore, suppose that $p = 1$, the u_i in (2.1) are uncorrelated and identically distributed normal random variables. Thus there is no parameterization by τ, $\Omega = 0$ and we can replace Σ by

$$2\sigma_0^{-2}n^{-1}\begin{bmatrix} (GX\beta_0)'(GX\beta_0) & (GX\beta_0)'X \\ * & X'X \end{bmatrix},$$

which matches the distribution obtained in this setting by Lee (2004) in his Theorem 3.2.

5. NORMALIZATIONS OF WEIGHT MATRICES

In this section we discuss the effect of various normalizations of the W_j on (2.10) and (3.5). For simplicity of exposition we will focus on the case when $\zeta_i = 2$, all i, given in (3.12) and (3.13). Due to the $n^{-\frac{1}{2}}$ factor it is not necessary that the elements of W_j and S^{-1} have finite eighth moments for (2.10) to hold, but for given n it is not sufficient either. Similarly due to the $h_W^{-\frac{1}{2}}$ factor and $h_W \to \infty$ finite fourth moments for elements of W_j' or S'^{-1} are neither necessary nor sufficient, for given n, for (3.5) to hold. Both (2.10) and (3.5) can be compared to conditions imposed in Gupta and Robinson (2015) for deterministic W_j elements, where $\max_{1\le j\le p}\|W_j\| + \|S^{-1}\| \le C$ was assumed, for which a necessary condition was boundedness of the elements of W_j, S^{-1}. Thus (2.10) and (3.5) may be viewed as controlling the spatial correlation asymptotically, and in particular controlling the magnitudes of the moments of the W_j and S^{-1}.

Various sufficient conditions may be found for (2.10) and (3.5) to hold. For example, suppose that

$$W_j = W_j^* / \left\| W_j^* \right\|, \tag{5.1}$$

for some matrices W_j^*. Then $\|W_j\|^s = 1$, for any $s > 0$, while

$$\left\| S^{-1} \right\| \le \sum_{l=0}^{\infty} \left(\sum_{j=1}^{p} \|\lambda_j W_j\| \right)^l \le C, \tag{5.2}$$

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

if

$$\sum_{j=1}^{p} |\lambda_j| < 1. \tag{5.3}$$

The W_j can have a special 'single nonzero diagonal block' structure in some applications. Here there are $m_j \times m_j$ matrices V_j, $j = 1, \ldots, p$ and $\sum_{j=1}^{p} m_j = n$, such that each W_j has V_j as its j-th diagonal block and zeroes elsewhere. In this case Gupta and Robinson (2015) prove that

$$\max_{1 \le j \le p} |\lambda_j| < 1 \tag{5.4}$$

can replace the more general condition given in (5.3). Thus under (5.1), condition (2.10) is satisfied always if (5.3) holds for general W_j and (5.4) holds for W_j with 'single nonzero diagonal block' structure. Another popular normalization when the weight matrix has nonnegative elements is row-normalization, where each row of W_j sums to 1, thus implying $\|W_j\|_R = 1$ so negligibility of $\chi_{n,R}$ follows if (5.3) holds in the general case and (5.4) holds in the 'single nonzero diagonal block' case, as illustrated for the spectral norm. More generally, if $\|W_j\|_M \le C$, where $\|\cdot\|_M$ denotes a generic matrix norm, then $\mathbb{E}\|W_j\|_M^s \le C$, any $s > 0$. A sufficient condition for $\mathbb{E}\|S^{-1}\|_M \le C$ is that $\left\|\sum_{j=1}^{p} \lambda_j W_j\right\|_M$ has a moment generating function.

On the other hand, all types of normalizations are not economically justified. For instance Bell and Bockstael (2000) point out that row-normalization is not justified in certain models with real estate data while Lee and Yu (2014) discuss problems in estimation that occur with row-normalized weight matrices and present some simulation evidence using nonnormalized matrices. Thus the moment conditions implied by (2.10) and (3.5) will be different under various normalizations of W_j. The choice of normalization ultimately lies with the practitioner, but we believe (5.1) is the most attractive option. Unlike row-normalization it does not change the economic content of the spatial weight matrices because it preserves relative distances and performs the task of stabilizing moments. Kelejian and Prucha (2010) provide an excellent discussion of normalizations and their implications, particularly for parameter spaces.

For another sufficient condition we focus on (3.6) with generic blocks B_m, where we have $p = 1$ and write $W_1 = W$ and $\lambda_1 = \lambda$. This is sometimes referred to as a Balanced Group Interaction (BGI) setting, see, e.g., Hillier and Martellosio (2013). It implies inter group independence for clustered data. Note that it does *not* have a 'single nonzero diagonal block' structure. We take $d, m \to \infty$, which is a combination of 'increasing domain' and 'infill' asymptotics. Suppose that the elements of B_m are such that $\mathbb{E}\|B_m\|^8 = \mathcal{O}\left(m^{\psi_1}\right)$ and $\mathbb{E}\left\|(I_m - \lambda B_m)^{-1}\right\|^8 = \mathcal{O}\left(m^{\psi_2}\right)$, for some $\psi_1, \psi_2 \ge 0$. Then $\mathbb{E}\|W\|^8 = \mathcal{O}\left(m^{\psi_1}\right)$ and $\mathbb{E}\left\|S^{-1}\right\|^8 = \mathcal{O}(m^{\psi_2})$ due to their block diagonality with equal blocks,

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

implying $\chi_n = \mathcal{O}\left(d^{-\frac{1}{2}} m^{\frac{\psi_1+\psi_2}{8}-\frac{1}{2}}\right) = o(1)$ always if $\psi_1 + \psi_2 \leq 4$ and if $m = o\left(d^{\frac{4}{\psi_1+\psi_2-4}}\right)$ when $\psi_1 + \psi_2 > 4$.

6. MONTE CARLO

Finite sample performance of IV, OLS, and PML estimates with fixed W_j has been examined before, see, e.g., Lee (2004), Gupta and Robinson (2015), and Gupta and Robinson (2018). Our aim in this section is twofold: we wish to compare performance of estimates when W_j are stochastic as opposed to fixed, and we seek information on how estimates behave according to different spatial weight matrix structures.

6.1. Spatial Weights Generated by iid Random Variables

Our first design takes $p = 1$, $k = 2$ with $\lambda = 0.5$, $\beta_1 = 1$, $\beta_2 = 0.7$, X generated as independent draws from $U(0, 1)$ and u_r independent draws from a standard normal distribution in each of the 1,000 replications. In this setup we examine what happens when the spatial matrices are very dense. We define $W_1^* = diag\left[V_{1(m\times m)}, 0_{(m\times m)}\right]$, $W_2^* = diag\left[0_{(m\times m)}, V_{2(m\times m)}\right]$, so $n = 2m$, and generate V_j (both of which have zeros on the diagonals) as iid replications using $|t_\upsilon|$ distributions with $\upsilon = 1, 10, 20, 100$, and $m = 48, 72, 144$. For the fixed design we generate such matrices once and keep them fixed for all 1,000 replications, while for the stochastic design these are generated anew in each trial. The dependent variable is generated as $y = \lambda (W_1 + W_2) y + X\beta + u$ where, for $j = 1, 2$, the W_j are obtained by normalizing W_j^* with the spectral norm as in (5.1). For ease of exposition we refer to the latter procedure as 'spectral-normalization'. We choose to use this block-diagonal specification to address the fact that very dense spatial weight matrices entail $h_W = \mathcal{O}(n)$, a situation that leads to inconsistent estimates, as noted by Lee (2004). Using a block-diagonal weight matrix with very dense blocks mitigates this problem because now we have $h_W = m - 1$. Instruments are selected to be $Z = [X, (W_1 + W_2) X]$, while for the MLE we take $\Lambda = [-0.99, 0.99]$.

We analyze absolute Monte Carlo bias, mean squared error (MSE), and size for the autoregression parameter and absolute average Monte Carlo bias, MSE and size for the regression parameters. By 'average' statistics we mean that in the rows corresponding to $\hat{\theta}$ the columns labelled β contain entries equal to

$$\left(\left|\text{bias}\left(\hat{\beta}_1\right)\right| + \left|\text{bias}\left(\hat{\beta}_2\right)\right|\right)/2, \tag{6.1}$$

for instance, with similar definitions for the rows corresponding to $\tilde{\theta}$ and $\breve{\theta}$. Throughout the section all Monte Carlo statistics for β reported are averages defined like (6.1). Sizes are based on t-ratios, defined again as in (6.1) for β, to be compared against a nominal value of 5%, computed using a critical value of 1.96.

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

Tables 1 and 6.1 summarize the results for stochastic and fixed W_j, respectively. The biases are quite similar for the fixed and stochastic designs for all three estimates. Out of the 72 possible comparisons, bias is greater in the fixed design for 38 cases. A similar conclusion holds for the MSE, which is higher for the fixed design in 34 cases. However, $\hat{\theta}$ seems to perform better in terms of MSE in the fixed design. Out of 24 cases, the stochastic setting yields a lower MSE on only 7 occasions, although the differences are rather small. One aspect that merits attention is the behaviour of estimates of λ when weights are generated by $|t_1|$, a distribution with no defined moments. Also of interest is the behaviour more generally as υ increases. The bias and MSE of all estimates of λ are high when $\upsilon = 1$, indeed for the IV estimate the magnitude of the MSE is alarming when $m = 72$. Increasing m can sometimes help matters, possibly due to smoothing out the effect of extreme draws from a distribution with no finite moments. The theory indicates an important role played by conditions on the moments of W_j, such as Assumptions 7 and 12. Evidently the moment stabilizing transformation (5.1) that we carry out does not sufficiently improve matters when $\upsilon = 1$, but increasing υ does so and broadly speaking helps even in improving the bias of the estimates of β, although not monotonically. The MSE improves dramatically from $\upsilon = 1$ to $\upsilon > 1$ but subsequently stays quite stable across the board. The findings suggest that, in practice, normalization is not enough to stabilize moments of weight matrices when the raw weights are generated by a distribution with no finite moments.

In both Tables 1 and 6.1, it is notable that $\tilde{\theta}$ is less biased than $\breve{\theta}$ in general, even for large m, while the MSE is quite similar. Large m means large h_W, and because the objective function for MLE differs from OLS only in a Jacobian term that is $\mathcal{O}_{\mathrm{p}}\left(h_W^{-1}\right)$, this implies that there is little to choose between the two when h_W is large. The difference in bias possibly stems from the fact that the MLE is only implicitly defined, and such estimates can sometimes have worse finite sample bias properties than closed form estimates. On the other hand the closeness in MSE implies that $\breve{\theta}$ has much smaller variance than $\tilde{\theta}$. Empirical sizes follow the same pattern as the bias and MSE discussed above, in the sense that there is no clear dominance between the fixed and stochastic designs. The MLE tends to undersize for λ, but the problem is worse when $\upsilon = 1$ and more acute in the fixed design.

The next setup considers more sparse spatial weight matrices, whereby the need for a block-diagonal specification to mitigate denseness is dispensed with. Now we take $n = 96, 144, 288$ (to match the sample sizes in the dense case) and generate the $n \times n$ matrix W^* once again with iid $|t_\upsilon|$ elements, $\upsilon = 1, 10, 20, 100$. Next, we generate an $n \times n$ matrix with elements given by $0.75 \times U(0, 1)$ and round all its elements that are less than 0.5 down to zero. Finally all elements of W^* corresponding to zeros in this matrix are set to zero. This produces a sparse W^*, in fact replacing 0.75 with a smaller (larger) factor would produce a more (less) sparse W^*. The fixed and stochastic designs differ as for the dense

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

TABLE 1. Monte Carlo absolute bias, mean squared error and size, nominal size 5%, dense stochastic W_1, W_2, see Section 6.1. $n = 2m$

	m	48		72		144	
Bias	υ	λ	β	λ	β	λ	β
$\hat{\theta}$	1	0.0369	0.0027	0.3397	0.0293	0.0055	0.0048
	10	0.0118	0.0285	0.0037	0.0107	0.0052	0.0103
	20	0.0175	0.0361	0.0029	0.0117	0.0037	0.0060
	100	0.0119	0.0262	0.0083	0.0170	0.0019	0.0030
$\tilde{\theta}$	1	0.0446	0.0035	0.0313	0.0053	0.0258	0.0064
	10	0.0055	0.0044	0.0084	0.0065	0.0009	0.0036
	20	0.0048	0.0080	0.0077	0.0037	0.0014	0.0013
	100	0.0063	0.0033	0.0032	0.0137	0.0045	0.0063
$\breve{\theta}$	1	0.0710	0.0051	0.0642	0.0070	0.0601	0.0076
	10	0.0443	0.0504	0.0362	0.0462	0.0163	0.0209
	20	0.0444	0.0511	0.0357	0.0437	0.0168	0.0236
	100	0.0454	0.0557	0.0316	0.0403	0.0198	0.0286
MSE	υ	λ	β	λ	β	λ	β
$\hat{\theta}$	1	0.9217	0.0808	123.9720	0.7877	0.7418	0.0262
	10	0.0254	0.1212	0.0164	0.0830	0.0087	0.0410
	20	0.0264	0.1272	0.0169	0.0828	0.0081	0.0422
	100	0.0259	0.1267	0.0158	0.0818	0.0085	0.0406
$\tilde{\theta}$	1	0.2254	0.0769	0.2195	0.0532	0.2509	0.0258
	10	0.0240	0.1157	0.0163	0.0820	0.0086	0.0405
	20	0.0250	0.1225	0.0162	0.0807	0.0079	0.0414
	100	0.0248	0.1229	0.0152	0.0801	0.0084	0.0403
$\breve{\theta}$	1	0.1598	0.0765	0.1638	0.0530	0.1731	0.0257
	10	0.0210	0.1071	0.0149	0.0779	0.0080	0.0389
	20	0.0216	0.1128	0.0148	0.0765	0.0074	0.0401
	100	0.0217	0.1144	0.0137	0.0757	0.0080	0.0392
Size	υ	λ	β	λ	β	λ	β
$\hat{\theta}$	1	0.0490	0.0550	0.0490	0.0625	0.0380	0.0625
	10	0.0560	0.0465	0.0520	0.0535	0.0570	0.0475
	20	0.0560	0.0640	0.0610	0.0610	0.0460	0.0565
	100	0.0560	0.0595	0.0440	0.0460	0.0540	0.0485
$\tilde{\theta}$	1	0.0610	0.0530	0.0520	0.0645	0.0610	0.0620
	10	0.0610	0.0485	0.0560	0.0560	0.0590	0.0475
	20	0.0620	0.0640	0.0580	0.0610	0.0470	0.0570
	100	0.0580	0.0585	0.0470	0.0495	0.0600	0.0495
$\breve{\theta}$	1	0.0420	0.0580	0.0310	0.0685	0.0340	0.0640
	10	0.0340	0.0445	0.0430	0.0510	0.0530	0.0460
	20	0.0420	0.0560	0.0480	0.0550	0.0410	0.0580
	100	0.0420	0.0525	0.0360	0.0495	0.0440	0.0470

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

TABLE 2. Monte Carlo absolute bias, mean squared error and size, nominal size 5%, dense fixed W_1, W_2, see Section 6.1. $n = 2m$

	m	48		72		144	
Bias	υ	λ	β	λ	β	λ	β
$\hat{\theta}$	1	0.0832	0.0103	0.0315	0.0061	0.0296	0.0065
	10	0.0058	0.0135	0.0066	0.0126	0.0096	0.0166
	20	0.0126	0.0262	0.0040	0.0170	0.0004	0.0063
	100	0.0067	0.0153	0.0021	0.0090	0.0059	0.0137
$\tilde{\theta}$	1	0.0473	0.0100	0.0286	0.0052	0.0210	0.0067
	10	0.0127	0.0127	0.0051	0.0041	0.0034	0.0077
	20	0.0040	0.0030	0.0063	0.0021	0.0048	0.0016
	100	0.0112	0.0104	0.0131	0.0107	0.0006	0.0044
$\check{\theta}$	1	0.1069	0.0099	0.0750	0.0053	0.0807	0.0067
	10	0.0498	0.0656	0.0327	0.0437	0.0123	0.0150
	20	0.0431	0.0524	0.0344	0.0379	0.0203	0.0237
	100	0.0492	0.0644	0.0408	0.0505	0.0162	0.0181
MSE	υ	λ	β	λ	β	λ	β
$\hat{\theta}$	1	1.9703	0.0770	3.6357	0.0482	1.3637	0.0240
	10	0.0247	0.1198	0.0150	0.0795	0.0081	0.0427
	20	0.0263	0.1264	0.0168	0.0820	0.0091	0.0401
	100	0.0243	0.1213	0.0184	0.0839	0.0083	0.0412
$\tilde{\theta}$	1	0.3590	0.0756	0.3343	0.0447	0.4064	0.0232
	10	0.0243	0.1172	0.0149	0.0787	0.0078	0.0417
	20	0.0255	0.1225	0.0164	0.0808	0.0090	0.0399
	100	0.0231	0.1181	0.0187	0.0838	0.0083	0.0409
$\check{\theta}$	1	0.2100	0.0749	0.2198	0.0446	0.2497	0.0231
	10	0.0219	0.1107	0.0136	0.0750	0.0072	0.0400
	20	0.0220	0.1130	0.0150	0.0763	0.0086	0.0384
	100	0.0207	0.1115	0.0173	0.0797	0.0078	0.0393
Size	υ	λ	β	λ	β	λ	β
$\hat{\theta}$	1	0.0590	0.0515	0.0550	0.0425	0.0500	0.0480
	10	0.0470	0.0540	0.0400	0.0500	0.0480	0.0535
	20	0.0600	0.0560	0.0550	0.0570	0.0540	0.0455
	100	0.0540	0.0490	0.0590	0.0520	0.0480	0.0500
$\tilde{\theta}$	1	0.0750	0.0515	0.0540	0.0450	0.0540	0.0510
	10	0.0610	0.0550	0.0460	0.0515	0.0490	0.0535
	20	0.0640	0.0595	0.0530	0.0555	0.0540	0.0445
	100	0.0540	0.0500	0.0700	0.0525	0.0460	0.0495
$\check{\theta}$	1	0.0290	0.0560	0.0270	0.0460	0.0200	0.0505
	10	0.0510	0.0505	0.0320	0.0485	0.0380	0.0510
	20	0.0480	0.0565	0.0420	0.0490	0.0470	0.0435
	100	0.0440	0.0475	0.0600	0.0510	0.0430	0.0465

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

case. We then take $W = W^*/\|W^*\|$ and generate the dependent variable as $y = \lambda Wy + X\beta + u$, with the other aspects of the experiment identical to the dense case above.

The results are in Tables 3 and 4 for stochastic and fixed designs, respectively. Biases are generally higher for $\tilde{\theta}$ than in the dense case, this being the case because sparser W implies smaller h_W, cf. (3.5). Once again the case with $\upsilon = 1$ produces odd results. For $n = 96$ in the stochastic design, $\tilde{\lambda}$ has less bias in the sparse case than the dense one (cf. Tables 1 and 3) but this counterintuitive feature is reversed as n grows, with a similar smoothing of extreme draws as discussed before. However it persists for all n in the fixed design (cf. Tables 6.1 and 4), undoubtedly because a fresh draw of W is not made in each trial and this inhibits the smoothing effect.

The biases in the stochastic design are higher in 45 cases out of 72, as opposed to the 34 cases of Tables 1 and 6.1. On the other hand, the MSE is higher in the stochastic design in only 30 cases. Thus the estimates are more precise in the stochastic design, but also more biased. Sizes follow the same pattern as in Tables 1 and 6.1, and the undersizing for λ with the MLE noted there is present.

The final iid setup considers extremely sparse W. We choose the circulant W defined in (3.7) and (3.8). Once again ω is generated from iid $|t_\upsilon|$, $\upsilon = 1, 10, 20, 100$, with the fixed and stochastic designs differing because in the latter this generation happens anew in each trial. The rest of the design is as in the sparse and dense cases discussed above.

Tables 5 and 6 contain the results for stochastic and fixed designs, respectively. Theorems 3.1 and 3.2 in this article and the results of Lee (2002) indicate that the performance of OLS estimates is poor in this situation. This is confirmed in the tables, with $\tilde{\theta}$ exhibiting unacceptably poor performance in terms of bias and size. The MSE seems more acceptable, but is much higher than that for $\breve{\theta}$ and $\hat{\theta}$. Given the highly biased nature of the estimates, a low variance provides scant comfort. On the other hand, both $\hat{\theta}$ and $\breve{\theta}$ perform well for both the fixed and stochastic designs. Within each table, $\breve{\theta}$ is much more efficient, as is well-known to be the case for the MLE. Comparing the biases of $\hat{\theta}$ and $\breve{\theta}$ between Tables 5 and 6, we find that bias is bigger in the stochastic design for 31 out of 48 comparisons. A similar MSE comparison returns an equal verdict of 24 each. Thus we may conclude that there is no substantial difference between the fixed and stochastic designs. The MLE still tends to undersize for λ, while sizes for the IV estimate are perfectly acceptable with one exception: the $\upsilon = 10$ and $n = 96$ entry in Table 5. The online appendix contains a simulation study when the spatial weights are possibly dependent, as opposed to the iid weights considered in this section.

6.2. Spatial Weights Generated by Explanatory Variables

We mentioned that an important motivation of stochastic spatial weights is the fact that these are frequently generated using economic variables that are

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

TABLE 3. Monte Carlo absolute bias, mean squared error and size, nominal size 5%, sparse stochastic W, see Section 6.1

	n	96		144		288	
Bias	υ	λ	β	λ	β	λ	β
$\hat{\theta}$	1	0.0710	0.0074	0.0949	0.0077	0.0405	0.0034
	10	0.0026	0.0153	0.0092	0.0063	0.0005	0.0020
	20	0.0072	0.0154	0.0066	0.0051	0.0020	0.0033
	100	0.0078	0.0055	0.0091	0.0079	0.0020	0.0025
$\tilde{\theta}$	1	0.0313	0.0063	0.0554	0.0061	0.0679	0.0037
	10	0.0189	0.0128	0.0236	0.0250	0.0080	0.0101
	20	0.0275	0.0238	0.0197	0.0233	0.0090	0.0113
	100	0.0263	0.0284	0.0234	0.0262	0.0095	0.0108
$\check{\theta}$	1	0.1397	0.0064	0.1423	0.0059	0.1327	0.0044
	10	0.0373	0.0373	0.0367	0.0430	0.0154	0.0207
	20	0.0459	0.0481	0.0329	0.0416	0.0164	0.0219
	100	0.0442	0.0525	0.0365	0.0443	0.0169	0.0213
MSE	υ	λ	β	λ	β	λ	β
$\hat{\theta}$	1	3.6937	0.0816	7.0059	0.0513	3.5902	0.0238
	10	0.0250	0.1116	0.0162	0.0765	0.0081	0.0407
	20	0.0251	0.1153	0.0173	0.0774	0.0087	0.0425
	100	0.0246	0.1170	0.0169	0.0768	0.0082	0.0413
$\tilde{\theta}$	1	2.0082	0.0777	5.0983	0.0495	2.7021	0.0231
	10	0.0244	0.1084	0.0160	0.0757	0.0079	0.0402
	20	0.0232	0.1117	0.0168	0.0757	0.0085	0.0420
	100	0.0227	0.1136	0.0169	0.0760	0.0084	0.0415
$\check{\theta}$	1	0.3377	0.0771	0.3585	0.0492	0.3414	0.0230
	10	0.0227	0.1043	0.0154	0.0739	0.0077	0.0396
	20	0.0218	0.1080	0.0160	0.0739	0.0083	0.0413
	100	0.0216	0.1106	0.0163	0.0744	0.0082	0.0408
Size	υ	λ	β	λ	β	λ	β
$\hat{\theta}$	1	0.0550	0.0525	0.0460	0.0575	0.0390	0.0435
	10	0.0480	0.0485	0.0470	0.0440	0.0490	0.0535
	20	0.0410	0.0495	0.0520	0.0495	0.0570	0.0570
	100	0.0470	0.0505	0.0520	0.0515	0.0580	0.0545
$\tilde{\theta}$	1	0.0460	0.0565	0.0470	0.0600	0.0450	0.0445
	10	0.0550	0.0495	0.0500	0.0455	0.0540	0.0495
	20	0.0470	0.0500	0.0580	0.0525	0.0610	0.0560
	100	0.0470	0.0490	0.0630	0.0495	0.0550	0.0530
$\check{\theta}$	1	0.0200	0.0585	0.0190	0.0610	0.0130	0.0455
	10	0.0540	0.0485	0.0430	0.0490	0.0510	0.0475
	20	0.0420	0.0500	0.0570	0.0530	0.0570	0.0525
	100	0.0460	0.0530	0.0530	0.0485	0.0490	0.0550

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

TABLE 4. Monte Carlo absolute bias, mean squared error and size, nominal size 5%, sparse fixed W, see Section 6.1

	n	96		144		288	
Bias	υ	λ	β	λ	β	λ	β
$\hat{\theta}$	1	0.2813	0.0077	0.0213	0.0019	0.0354	0.0033
	10	0.0000	0.0082	0.0033	0.0031	0.0073	0.0066
	20	0.0011	0.0063	0.0045	0.0042	0.0025	0.0017
	100	0.0066	0.0041	0.0016	0.0083	0.0040	0.0054
$\tilde{\theta}$	1	0.0021	0.0109	0.0106	0.0025	0.0115	0.0031
	10	0.0231	0.0224	0.0186	0.0208	0.0149	0.0173
	20	0.0180	0.0170	0.0183	0.0149	0.0101	0.0124
	100	0.0271	0.0258	0.0174	0.0155	0.0116	0.0095
$\check{\theta}$	1	0.1725	0.0114	0.0391	0.0040	0.0518	0.0031
	10	0.0415	0.0469	0.0316	0.0386	0.0222	0.0276
	20	0.0371	0.0428	0.0320	0.0337	0.0174	0.0227
	100	0.0462	0.0517	0.0311	0.0346	0.0191	0.0202
MSE	υ	λ	β	λ	β	λ	β
$\hat{\theta}$	1	33.4155	0.0878	0.6563	0.0528	1.9369	0.0252
	10	0.0254	0.1131	0.0153	0.0757	0.0084	0.0392
	20	0.0252	0.1135	0.0177	0.0810	0.0081	0.0390
	100	0.0250	0.1179	0.0168	0.0787	0.0093	0.0438
$\tilde{\theta}$	1	2.3193	0.0744	0.1910	0.0510	0.2607	0.0240
	10	0.0245	0.1099	0.0150	0.0741	0.0084	0.0390
	20	0.0238	0.1102	0.0173	0.0796	0.0081	0.0389
	100	0.0244	0.1153	0.0159	0.0761	0.0090	0.0428
$\check{\theta}$	1	0.4626	0.0733	0.1493	0.0508	0.1846	0.0239
	10	0.0229	0.1059	0.0144	0.0724	0.0083	0.0386
	20	0.0219	0.1058	0.0165	0.0771	0.0079	0.0383
	100	0.0228	0.1113	0.0152	0.0739	0.0088	0.0421
Size	υ	λ	β	λ	β	λ	β
$\hat{\theta}$	1	0.0450	0.0480	0.0510	0.0570	0.0340	0.0485
	10	0.0550	0.0460	0.0480	0.0505	0.0420	0.0475
	20	0.0510	0.0485	0.0550	0.0595	0.0450	0.0535
	100	0.0500	0.0490	0.0540	0.0525	0.0580	0.0575
$\tilde{\theta}$	1	0.0470	0.0490	0.0500	0.0585	0.0560	0.0510
	10	0.0650	0.0535	0.0580	0.0490	0.0470	0.0530
	20	0.0490	0.0465	0.0580	0.0600	0.0470	0.0545
	100	0.0570	0.0495	0.0590	0.0520	0.0530	0.0560
$\check{\theta}$	1	0.0140	0.0500	0.0270	0.0620	0.0250	0.0520
	10	0.0550	0.0465	0.0530	0.0500	0.0480	0.0505
	20	0.0410	0.0475	0.0550	0.0575	0.0490	0.0530
	100	0.0550	0.0495	0.0540	0.0490	0.0500	0.0590

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

TABLE 5. Monte Carlo absolute bias, mean squared error and size, nominal size 5%, circulant stochastic W, see Section 6.1

	n	96		144		288	
Bias	υ	λ	β	λ	β	λ	β
$\hat{\theta}$	1	0.0197	0.0344	0.0099	0.0201	0.0066	0.0114
	10	0.0165	0.0340	0.0076	0.0153	0.0053	0.0117
	20	0.0217	0.0383	0.0053	0.0111	0.0029	0.0093
	100	0.0220	0.0417	0.0132	0.0287	0.0028	0.0061
$\tilde{\theta}$	1	0.1065	0.1585	0.1088	0.1624	0.1137	0.1687
	10	0.1396	0.2113	0.1365	0.2022	0.1444	0.2158
	20	0.1328	0.1973	0.1352	0.2004	0.1460	0.2190
	100	0.1400	0.2104	0.1402	0.2124	0.1481	0.2197
$\breve{\theta}$	1	0.0115	0.0157	0.0080	0.0106	0.0057	0.0088
	10	0.0091	0.0088	0.0120	0.0185	0.0034	0.0051
	20	0.0156	0.0239	0.0109	0.0167	0.0053	0.0085
	100	0.0117	0.0151	0.0088	0.0084	0.0048	0.0075
MSE	υ	λ	β	λ	β	λ	β
$\hat{\theta}$	1	0.0265	0.1379	0.0167	0.0859	0.0084	0.0418
	10	0.0239	0.1247	0.0175	0.0907	0.0086	0.0427
	20	0.0288	0.1345	0.0169	0.0859	0.0086	0.0446
	100	0.0277	0.1370	0.0190	0.0888	0.0083	0.0426
$\tilde{\theta}$	1	0.0244	0.1297	0.0227	0.0978	0.0214	0.0700
	10	0.0312	0.1364	0.0282	0.1127	0.0285	0.0867
	20	0.0308	0.1370	0.0289	0.1099	0.0292	0.0904
	100	0.0324	0.1451	0.0306	0.1132	0.0297	0.0885
$\breve{\theta}$	1	0.0054	0.0880	0.0035	0.0568	0.0019	0.0275
	10	0.0052	0.0816	0.0037	0.0605	0.0017	0.0276
	20	0.0055	0.0834	0.0038	0.0573	0.0020	0.0305
	100	0.0055	0.0871	0.0037	0.0540	0.0018	0.0283
Size	υ	λ	β	λ	β	λ	β
$\hat{\theta}$	1	0.0510	0.0550	0.0380	0.0515	0.0560	0.0490
	10	0.0320	0.0465	0.0400	0.0590	0.0500	0.0510
	20	0.0480	0.0535	0.0490	0.0595	0.0470	0.0550
	100	0.0470	0.0565	0.0520	0.0570	0.0490	0.0435
$\tilde{\theta}$	1	0.3190	0.1150	0.4030	0.1470	0.5790	0.2260
	10	0.4090	0.1295	0.5180	0.1730	0.6790	0.3045
	20	0.4030	0.1250	0.5290	0.1645	0.7120	0.3050
	100	0.4320	0.1410	0.5430	0.1840	0.7240	0.3135
$\breve{\theta}$	1	0.0270	0.0575	0.0290	0.0495	0.0310	0.0445
	10	0.0250	0.0390	0.0310	0.0585	0.0220	0.0480
	20	0.0270	0.0515	0.0280	0.0510	0.0380	0.0585
	100	0.0260	0.0555	0.0290	0.0420	0.0280	0.0445

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

TABLE 6. Monte Carlo absolute bias, mean squared error and size, nominal size 5%, circulant fixed W, see Section 6.1

	n	96		144		288	
Bias	υ	λ	β	λ	β	λ	β
$\hat{\theta}$	1	0.0163	0.0341	0.0068	0.0163	0.0043	0.0074
	10	0.0093	0.0171	0.0055	0.0129	0.0041	0.0073
	20	0.0181	0.0338	0.0122	0.0305	0.0061	0.0150
	100	0.0121	0.0268	0.0066	0.0198	0.0075	0.0153
$\tilde{\theta}$	1	0.1319	0.1984	0.1980	0.2958	0.1404	0.2072
	10	0.0905	0.1320	0.1210	0.1798	0.0014	0.0066
	20	0.1185	0.1771	0.2092	0.3162	0.0857	0.1299
	100	0.0240	0.0399	0.1342	0.2028	0.2191	0.3284
$\breve{\theta}$	1	0.0122	0.0144	0.0098	0.0130	0.0057	0.0096
	10	0.0139	0.0228	0.0099	0.0137	0.0054	0.0078
	20	0.0123	0.0164	0.0101	0.0085	0.0037	0.0053
	100	0.0152	0.0174	0.0091	0.0078	0.0018	0.0016
MSE	υ	λ	β	λ	β	λ	β
$\hat{\theta}$	1	0.0282	0.1380	0.0174	0.0851	0.0090	0.0450
	10	0.0249	0.1247	0.0168	0.0841	0.0078	0.0436
	20	0.0267	0.1352	0.0192	0.0904	0.0082	0.0426
	100	0.0234	0.1178	0.0175	0.0840	0.0091	0.0418
$\tilde{\theta}$	1	0.0249	0.1270	0.0454	0.1451	0.0224	0.0727
	10	0.0149	0.0987	0.0193	0.0887	0.0019	0.0307
	20	0.0217	0.1220	0.0496	0.1586	0.0097	0.0464
	100	0.0069	0.0805	0.0233	0.0986	0.0508	0.1348
$\breve{\theta}$	1	0.0055	0.0848	0.0039	0.0546	0.0019	0.0290
	10	0.0054	0.0803	0.0035	0.0550	0.0019	0.0307
	20	0.0057	0.0877	0.0036	0.0569	0.0019	0.0292
	100	0.0060	0.0779	0.0039	0.0554	0.0017	0.0257
Size	υ	λ	β	λ	β	λ	β
$\hat{\theta}$	1	0.0560	0.0510	0.0410	0.0455	0.0580	0.0580
	10	0.0400	0.0420	0.0440	0.0475	0.0450	0.0600
	20	0.0410	0.0535	0.0420	0.0555	0.0440	0.0490
	100	0.0400	0.0400	0.0470	0.0530	0.0460	0.0405
$\tilde{\theta}$	1	0.3960	0.1195	0.7600	0.2465	0.8050	0.2465
	10	0.2300	0.0735	0.4330	0.1230	0.0550	0.0625
	20	0.3430	0.0990	0.7960	0.2810	0.4840	0.1240
	100	0.0790	0.0415	0.5220	0.1410	0.9780	0.5080
$\breve{\theta}$	1	0.0230	0.0475	0.0210	0.0430	0.0360	0.0550
	10	0.0270	0.0435	0.0240	0.0405	0.0540	0.0615
	20	0.0270	0.0530	0.0150	0.0425	0.0410	0.0595
	100	0.0510	0.0440	0.0320	0.0500	0.0170	0.0340

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

themselves stochastic. In this subsection we wish to analyse the performance of estimates when the spatial weights are generated by elements of X. The rank conditions for our asymptotic theorems rule out any identification problems arising from this, but we wish to study the finite sample implications of this type of feature.

We generate the spatial weight matrices as

$$w^*_{rs,1} = \frac{\|x_r - x_s\|}{1 + \|x_r - x_s\|^2} \mathbb{1}(r \neq s), \tag{6.2}$$

$$w^*_{rs,2} = \exp(-\|x_r - x_s\|)\, \mathbb{1}(r \neq s), \tag{6.3}$$

for $r, s = 1, \ldots, n$ with $w^*_{rs,j}$ a typical element of W^*_j, x'_r the r-th row of X and $n = 48, 96, 144$. Robinson (2008) originally suggested the formula given in (6.2), though no finite-sample implications were studied in that article. Now we generate $y = \lambda_1 W_1 y + \lambda_2 W_2 y + X\beta + u$ with $\lambda_1 = 0.2$ and $\lambda_2 = 0.3$, with W_j denoting the spectral-normalized W^*_j, and search over $\Lambda = [-0.99, 0.99] \times [-0.99, 0.99]$ in each of the 1,000 replications. Table 7 reports the results, reporting average statistics as in (6.1) for the two-dimensional λ also. These paint a positive picture, in the sense that estimates exhibit bias and MSE that declines with sample size to quite acceptable levels. The bias reaches zero to two decimal places for IV for instance, and the MSE likewise reaches zero to one decimal place quite comfortably. Sizes are perfectly acceptable except for oversizing of OLS and PMLE when $n = 48$, which is alleviated when $n = 96$.

TABLE 7. Monte Carlo absolute (average) bias, mean squared error and size, nominal size 5%, W_1 and W_2 generated as in Section 6.2 using (6.2) and (6.3), respectively

n	48		96		144	
Bias	λ	β	λ	β	λ	β
$\hat{\theta}$	0.0244	0.0058	0.0025	0.0060	0.0098	0.0042
$\tilde{\theta}$	0.1792	0.0804	0.0754	0.0384	0.0399	0.0255
$\breve{\theta}$	0.1747	0.1131	0.0832	0.0576	0.0463	0.0388
MSE	λ	β	λ	β	λ	β
$\hat{\theta}$	0.1898	0.1487	0.0812	0.0696	0.0556	0.0464
$\tilde{\theta}$	0.2385	0.1706	0.0905	0.0741	0.0587	0.0486
$\breve{\theta}$	0.1852	0.1598	0.0852	0.0727	0.0565	0.0479
Size	λ	β	λ	β	λ	β
$\hat{\theta}$	0.0570	0.0590	0.0525	0.0480	0.0510	0.0500
$\tilde{\theta}$	0.0860	0.0710	0.0530	0.0560	0.0555	0.0505
$\breve{\theta}$	0.0585	0.0670	0.0485	0.0545	0.0515	0.0550

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

6.3. Summary of Monte Carlo Findings

To conclude, in our design we find that there is no clear pattern of dominance between stochastic and fixed designs when spatial weight matrices are dense. On the other hand, in the sparse case the stochastic design can yield more precise estimates but with larger absolute bias than the fixed design. We also find that finite sample performance of estimates is satisfactory even when spatial weight matrices are functions of the explanatory variables that form the regression component of the model, which is anticipated from our theoretical results when multicollinearity is ruled out. However, a caveat applies to the interpretation of these findings, which are dependent on the particular way of comparing fixed and stochastic designs that we have chosen.

Finally, we note that weight matrix normalization may not stabilize moments sufficiently when raw weights come from distributions with no finite moments. Due to normalization, the theoretical conditions, such as (2.10) and (3.5), do hold in this case but finite sample performance is not up to the mark, especially so in the fixed W_j case. A possible reason for this difference is the smoothing out of extreme draws in the stochastic design.

7. CONCLUSION AND EXTENSIONS

We examined IV, OLS, and PML estimates for the parameters of SAR models with stochastic weight matrices and spatial linear process dependence in disturbances. We also discussed the implications of popular weight matrix normalizations on our conditions. In the dependent disturbances setup that we consider, heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimation is an important problem. With deterministic W_j this has been considered in the literature, see, e.g., Kelejian and Prucha (2007, 2010) and Robinson and Thawornkaiwong (2012) for HAC estimation with SAR/SMA disturbances and disturbances satisfying Assumption 1, respectively. These approaches are straightforward to extend to the case with stochastic W_j and using them in practice requires no change in earlier techniques.

REFERENCES

Baltagi, B.H., B. Fingleton, & A. Pirotte (2014) Spatial lag models with nested random effects: An instrumental variable procedure with an application to English house prices. *Journal of Urban Economics* 80, 76–86.

Bell, K.P. & N.E. Bockstael (2000) Applying the generalized-moments estimation approach to spatial problems involving micro-level data. *Review of Economics and Statistics* 82, 72–82.

Boucher, V. & B. Fortin (2016) Some challenges in the empirics of the effects of networks. In Y. Bramoullé, A. Galeotti, & B. Rogers (eds.), *The Oxford Handbook of the Economics of Networks*, chapter 12. Oxford University Press.

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

Cliff, A.D. & J.K. Ord (1973) *Spatial Autocorrelation*. Pion.

Conley, T.G. & B.Dupor (2003) A spatial analysis of sectoral complementarity. *Journal of Political Economy* 111, 311–352.

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

Conley, T.G. & E.Ligon (2002) Economic distance and cross-country spillovers. *Journal of Economic Growth* 7, 157–187.
Das, D., H.H. Kelejian, & I.R. Prucha (2003) Finite sample properties of estimators of spatial autoregressive models with autoregressive disturbances. *Papers in Regional Science* 82, 1–26.
Davis, P.J. (1979) *Circulant Matrices*. Wiley Interscience.
Delgado, M. & P.M. Robinson (2015) Non-nested testing of spatial correlation. *Journal of Econometrics* 187, 385–401.
Gupta, A. & P.M. Robinson (2015) Inference on higher-order spatial autoregressive models with increasingly many parameters. *Journal of Econometrics* 186, 19–31.
Gupta, A. & P.M. Robinson (2018) Pseudo maximum likelihood estimation of spatial autoregressive models with increasing dimension. *Journal of Econometrics* 202, 92–107.
Hillier, G. & F. Martellosio (2013) *Properties of the Maximum Likelihood Estimator in Spatial Autoregressive Models*. Mimeo.
Hillier, G. & F. Martellosio (2018) Exact likelihood inference in group interaction network models. *Econometric Theory* 34, 383–415.
Jenish, N. & I.R. Prucha (2012) On spatial processes and asymptotic inference under near-epoch dependence. *Journal of Econometrics* 170, 178–190.
Kelejian, H.H. & G. Piras (2014) Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes. *Regional Science and Urban Economics* 46, 140–149.
Kelejian, H.H. & I.R. Prucha (1998) A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics* 17, 99–121.
Kelejian, H.H. & I.R. Prucha (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40, 509–533.
Kelejian, H.H. & I.R. Prucha (2001) On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics* 104, 219–257.
Kelejian, H.H. & I.R. Prucha (2007) HAC estimation in a spatial framework. *Journal of Econometrics* 140, 131–154.
Kelejian, H.H. & I.R. Prucha (2010) Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157, 53–67.
Kuersteiner, G.M. & I.R. Prucha (2013) Limit theory for panel data models with cross sectional dependence and sequential exogeneity. *Journal of Econometrics* 174, 107–126.
Kuersteiner, G.M. & I.R. Prucha (2015) Dynamic Spatial Panel Models: Networks, Common Shocks, and Sequential Exogeneity. CESifo Working paper 5445.
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. (2003) Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews* 22, 307–335.
Lee, L.F. (2004) Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72, 1899–1925.
Lee, L.F. & X. Liu (2010) Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. *Econometric Theory* 26, 187–230.
Lee, L.F. & J. Yu (2013) Near unit root in the spatial autoregressive model. *Spatial Economic Analysis* 8, 314–351.
Lee, L.F. & J. Yu (2014) Efficient GMM estimation of spatial dynamic panel data models with fixed effects. *Journal of Econometrics* 180, 174–197.
Qu, X. & L.F. Lee (2015) Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics* 184, 209–232.
Robinson, P.M. (2008) Correlation testing in time series, spatial and cross-sectional data. *Journal of Econometrics* 147, 5–16.
Robinson, P.M. (2010) Efficient estimation of the semiparametric spatial autoregressive model. *Journal of Econometrics* 157, 6–17.

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

Robinson, P.M. & F.J. Hidalgo (1997) Time series regression with long-range dependence. *The Annals of Statistics* 25, 77–104.
Robinson, P.M. & S. Thawornkaiwong (2012) Statistical inference on regression with spatial dependence. *Journal of Econometrics* 167, 521–542.
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.
Souza, P.C.L. (2015) *Estimating Network Effects without Network Data*. Mimeo.
Su, L. & S. Jin (2010) Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models. *Journal of Econometrics* 157, 18–33.
Xu, X. & L.F. Lee (2015) A spatial autoregressive model with a nonlinear transformation of the dependent variable. *Journal of Econometrics* 186, 1–18.
Yuzefovich, Y.A. (2003) Two Essays on Spatial Econometrics. Ph.D. thesis, University of Maryland.

APPENDIX A: Proofs of Theorems

For any matrices $\bar{F}$ and $\bar{\bar{F}}$ of equal dimension, we will write $\bar{\bar{\Delta}}^F = \bar{F} - \bar{\bar{F}}$.

Proof of Theorem 2.1. By (2.2) $\hat{\theta} - \theta = \bar{Q}^{-1}\bar{\bar{\Delta}}^Q\left(\hat{\theta} - \theta\right) - \bar{Q}^{-1}\bar{\bar{\Delta}}^{K'}\bar{J}^{-1}q + \bar{Q}^{-1}\bar{K}'\bar{J}^{-1}q$, so

$$\left(I_{p+k} - \bar{Q}^{-1}\bar{\bar{\Delta}}^Q\right)\left(\hat{\theta} - \theta\right) = -\bar{Q}^{-1}\bar{\bar{\Delta}}^{K'}\bar{J}^{-1}q + \bar{Q}^{-1}\bar{K}'\bar{J}^{-1}q. \tag{A.1}$$

We first show $\bar{\bar{\Delta}}^K = o_p(1)$. Write e_r for the $n \times 1$ vector with unity in the r-th position and zeroes elsewhere and b_i for the i-th column of $[Z, X]$. By the law of iterated expectations, the expectation of the square of a typical (i, j)-th element, $i, j = 1, \ldots, p_1 + k$, of $\bar{\bar{\Delta}}^K = n^{-1}[Z, X]'[B, 0]$ is

$$n^{-2}\mathbb{E}\left(b_i'G_juu'G_j'b_i\right) = n^{-2}\mathbb{E}\left(\sum_{r,s=1}^{n} b_i'G_je_r\mathbb{E}\left(u_{rn}u_{sn}\right)e_s'G_j'b_i\right). \tag{A.2}$$

Now $\mathbb{E}\left(u_{rn}u_{sn}\right) = \sum_{k,l=1}^{\infty} c_{rk}c_{sl}\mathbb{E}\left(\epsilon_k\epsilon_l\right) = \sigma^2\sum_{k=1}^{\infty} c_{rk}c_{sk} \leq \sigma^2\left(\sum_{k=1}^{\infty} c_{rk}^2\right)^{\frac{1}{2}}\left(\sum_{k=1}^{\infty} c_{sk}^2\right)^{\frac{1}{2}} \leq C$, by Assumption 5 and Cauchy–Schwarz inequality, so (A.2) is bounded by Cn^{-2} times

$$\mathbb{E}\left(b_i'G_j\sum_{r,s=1}^{n} e_re_s'G_j'b_i\right) = \mathbb{E}\left(b_i'G_j\sum_{r=1}^{n} e_re_r'G_j'b_i\right) = \mathbb{E}\left(b_i'G_jG_j'b_i\right). \tag{A.3}$$

The term inside the expectation on the far right is bounded by $\|b_i\|^2\left\|G_j\right\|^2$ so, by the Hölder inequality (for expectations), (A.3) is bounded by

$$\left(\mathbb{E}\left\|b_i\right\|^{2\zeta_1}\right)^{\frac{1}{\zeta_1}}\left(\mathbb{E}\left\|G_j\right\|^{2\zeta_2}\right)^{\frac{1}{\zeta_2}}. \tag{A.4}$$

The expectation inside parentheses in the first factor in (A.4) equals $\mathbb{E}\left(\sum_{r=1}^{n} a_{ri}^2\right)^{\zeta_1}$, which, by the Hölder inequality (for sums of real numbers) is bounded by

$$n^{\zeta_1\left(1-\frac{1}{\zeta_1}\right)}\sum_{r=1}^{n}\mathbb{E}\left|a_{ri}\right|^{2\zeta_1} = \mathcal{O}\left(n^{\zeta_1}\right), \tag{A.5}$$

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

by (2.7). The second factor in (A.4) is bounded by

$$\left\{\mathbb{E}\left(\|W_j\|^{2\zeta_2}\left\|S^{-1}\right\|^{2\zeta_2}\right)\right\}^{\frac{1}{\zeta_2}} \leq \left\{\left(\mathbb{E}\|W_j\|^{2\zeta_2\zeta_3}\right)^{\frac{1}{\zeta_3}}\left(\mathbb{E}\left\|S^{-1}\right\|^{2\zeta_2\zeta_4}\right)^{\frac{1}{\zeta_4}}\right\}^{\frac{1}{\zeta_2}} = \mathscr{O}\left(n\chi_n^2\right), \tag{A.6}$$

once again using Hölder's inequality. Combining (A.5) and (A.6) yields

$$\left(\mathbb{E}\|b_i\|^{2\zeta_1}\right)^{\frac{1}{\zeta_1}}\left(\mathbb{E}\|G_j\|^{2\zeta_2}\right)^{\frac{1}{\zeta_2}} = \mathscr{O}\left(n^2\chi_n^2\right), \tag{A.7}$$

whence Markov's inequality implies that

$$\left\|\bar{\tilde{\Delta}}^K\right\| = \mathscr{O}_{\mathrm{p}}(\chi_n) = o_p(1), \tag{A.8}$$

by (2.10). By (2.8) and (A.8),

$$\left\|\bar{\tilde{\Delta}}^Q\right\| \leq \left\|\bar{\tilde{\Delta}}^K\right\|\left\|\bar{J}^{-1}\right\|\left(\left\|\bar{\tilde{\Delta}}^K\right\| + 2\|\bar{K}\|\right) = \mathscr{O}_{\mathrm{p}}\left(\left\|\bar{\tilde{\Delta}}^K\right\|\right) = o_p(1). \tag{A.9}$$

Finally, the expectation of the square of a typical element of q is

$$n^{-2}\mathbb{E}\left(b_i'uu'b_i\right) = n^{-2}\mathbb{E}\left(\sum_{r,s=1}^{n} b_i'e_r\mathbb{E}\left(u_{rn}u_{sn}\right)e_s'b_i\right) \leq Cn^{-2}\mathbb{E}\|b_i\|^2 \leq Cn^{-1},$$

so

$$q = \mathscr{O}_{\mathrm{p}}\left(n^{-\frac{1}{2}}\right). \tag{A.10}$$

Using (A.8), (A.9), (A.10), and (2.8) in (A.1) we obtain the desired result. ■

Proof of Theorem 2.2. In view of (A.1), (A.8), (A.9) and (A.10) it suffices to show $n^{\frac{1}{2}}q \xrightarrow{d} \mathcal{N}(0, \sigma^2\Phi)$. The proof now follows Robinson and Thawornkaiwong (2012), who modified a proof of Robinson and Hidalgo (1997). Write $d = d_n = n^{-\frac{1}{2}}\sum_{r=1}^{n} a_r u_r = n^{-\frac{1}{2}}\sum_{l=1}^{\infty} f_l\epsilon_l$, where $f_l = f_{ln} = \sum_{r=1}^{n} a_r c_{rl}$. By Lemma A1 of Robinson and Thawornkaiwong (2012), there exists a sequence $N = N_n$, increasing in n without bound, such that $d - d_N = o_p(1)$, where $d_N = n^{-\frac{1}{2}}\sum_{l=1}^{N} f_l\epsilon_l$. Writing $E = E_n = n^{-1}\sum_{l=1}^{N} f_l f_l'$, again Lemma A1 of Robinson and Thawornkaiwong (2012) implies that $E \xrightarrow{p} \sigma^2\Phi$, by Assumption 6. Let $\alpha \in \mathbb{R}^{p+k}$ such that $\|\alpha\|^2 = 1$ and $c_N = \alpha' E^{-\frac{1}{2}} d_N$, $v_l = v_{ln} = n^{-\frac{1}{2}}\alpha' E^{-\frac{1}{2}} f_l$. Then $c_N = \sum_{l=1}^{N} v_l\epsilon_l$, and Assumption 6 implies that $\{f_l\epsilon_l, 1 \leq l \leq N\}$ is a martingale difference sequence for each $N \geq 1$. We show $c_N \xrightarrow{d} \mathcal{N}(0,1)$, conditional on X, z_r and W_j, $j = 1, \ldots, p$, which follows by Theorem 2 of Scott (1973) if, conditional on X, z_r and W_j, $j = 1, \ldots, p$, as $n \to \infty$,

$$\mathbb{E}\left(\sum_{l=1}^{N} v_l^2\epsilon_l^2 | \epsilon_j, j < l\right) \xrightarrow{p} 1, \tag{A.11}$$

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

and for all $\xi > 0$,

$$\mathbb{E}\left(\sum_{l=1}^{N} v_l^2 \mathbb{E}\left(\epsilon_l^2 \mathbb{1}\left(|v_l \epsilon_l| > \xi\right) | z_r, X, W_1, \ldots, W_p\right)\right) \to 0. \tag{A.12}$$

The LHS of (A.11) equals 1, while the LHS of (A.12) is bounded by

$$\max_{1 \le l \le N} \mathbb{E}\left\{\epsilon_l^2 \mathbb{1}\left(\epsilon_l^2 > \frac{\xi^2}{\max\limits_{1 \le l \le N} v_l^2}\right)\right\} \mathbb{E}\left(\sum_{l=1}^{N} v_l^2\right).$$

By Assumption 5, it suffices to show that $\max_{1 \le l \le N} v_l^2 = o_p(1)$ as $n \to \infty$, as the rightmost factor equals 1. Now, $\max_{1 \le l \le N} v_l^2 \le n^{-1} \left\| E^{-\frac{1}{2}} \right\|^2 \sup_{l \ge 1} \left\| \sum_{r=1}^{n} a_r c_{rl} \right\|^2 = o_p(1)$ by Assumptions 4, 5, and 6. ■

Proof of Theorem 3.1. By (3.1) $\tilde{\theta} - \theta = \bar{L}^{-1} \bar{\tilde{\Delta}}^L \left(\tilde{\theta} - \theta\right) + \bar{L}^{-1} w$, so

$$\left(I_{p+k} - \bar{L}^{-1} \bar{\tilde{\Delta}}^L\right)\left(\tilde{\theta} - \theta\right) = \bar{L}^{-1} w. \tag{A.13}$$

Note that $\|\ell\| \le \left\| n^{-1}[A, X]' u \right\| + \left\| n^{-1}[B, 0]' u \right\|$, where the first term on the RHS is readily shown to be negligible, as we deduced (A.10), but using (3.2) in Assumption 8 instead of (2.7) in Assumption 4 because here Z is replaced by A. Next $n^{-1}[B, 0]' u = o_p(1)$ by (3.5) and Lemma B.3, so that $\ell = o_p(1)$. It remains to prove that $\bar{\tilde{\Delta}}^L = o_p(1)$, for which first note that

$$\left\| \bar{\tilde{\Delta}}^L \right\| \le n^{-1} \|B\|^2 + 2n^{-1} \left\| [A, X]' [B, 0] \right\|. \tag{A.14}$$

The first term on the RHS is $\mathscr{O}_{\mathrm{p}}(\pi_n)$, by the proof of Lemma B.3, and is negligible by (3.5). The second term on the RHS is bounded exactly like $\left\| \bar{\tilde{\Delta}}^K \right\| = n^{-1}[Z, X]'[B, 0]$ in the proof of Theorem 2.1, but again using (3.2) in Assumption 8 instead of (2.7) in Assumption 4 because here Z is replaced by A, see also the proof of Theorem 4.1 in Gupta and Robinson (2015). Therefore

$$\left\| \bar{\tilde{\Delta}}^L \right\| = \mathscr{O}_{\mathrm{p}}\left(\max\left\{\pi_n, \chi_n\right\}\right) = o_p(1), \tag{A.15}$$

and the theorem is proved. ■

Proof of Theorem 3.2. We claim that it is sufficient to prove

$$n^{-\frac{1}{2}}[A, X]' u \xrightarrow{d} \mathcal{N}(0, \Psi), \tag{A.16}$$

for which, in view of (A.13), (A.15) and $\ell = n^{-1}[A, X]' u + n^{-1}[B, 0]' u$ it is enough to show that

$$n^{-\frac{1}{2}}[B, 0]' u = o_p(1). \tag{A.17}$$

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

To show (A.17), we can exploit Assumption 9 to obtain a sharper bound for $[B,0]'u$ than the one used in the proof of Theorem 3.1, where Assumption 5 sufficed. Indeed, by Lemma B.4, $n^{-\frac{1}{2}}[B,0]'u = \mathscr{O}_{\mathrm{p}}\left(n^{\frac{1}{2}}\left(\mathbb{E}\left\|S'^{-1}\right\|_R^{2\zeta_{12}}\right)^{\frac{1}{2\zeta_{12}}} h_W^{-1}\right) = o_p(1)$, by (3.11). The proof of (A.16) follows exactly as the proof of asymptotic normality of $n^{\frac{1}{2}}q$ in Theorem 2.2, and we omit the details, noting only that here A replaces Z, and Assumption 10 replaces Assumption 6. ■

Proof of Theorem 4.1. We show $\check{\gamma} \xrightarrow{p} \gamma_0$, whence $\check{\beta} \xrightarrow{p} \beta_0$ and $\check{\sigma}^2 \xrightarrow{p} \sigma_0^2$ follow from (4.2) and (4.3), respectively. First note that

$$\mathcal{L}(\gamma) - \mathcal{L} = \log\overline{\sigma}^2(\gamma)/\overline{\sigma}^2 - n^{-1}\log\left|T'(\lambda)\Xi(\tau)^{-1}T(\lambda)\Xi\right| = \log\overline{\sigma}^2(\gamma)/\sigma^2(\gamma) - \log\overline{\sigma}^2/\sigma_0^2 + \log r(\gamma), \tag{A.18}$$

where recall that $\sigma^2(\gamma) = n^{-1}\sigma_0^2 tr\left(T'(\lambda)\Xi(\tau)^{-1}T(\lambda)\Xi\right)$, $\overline{\sigma}^2 = \overline{\sigma}^2(\gamma_0) = n^{-1}u'\Xi'^{-\frac{1}{2}}M\Xi^{-\frac{1}{2}}u$, using (4.3) and also $r(\gamma) = n^{-1}tr\left(T'(\lambda)\Xi(\tau)^{-1}T(\lambda)\Xi\right)/\left|T'(\lambda)\Xi(\tau)^{-1}T(\lambda)\Xi\right|^{1/n}$. From (2.6), $\overline{\sigma}^2(\gamma) = n^{-1}\left\{S^{-1'}(X\beta_0+u)\right\}'S'(\lambda)\Xi(\tau)'^{-\frac{1}{2}}M(\tau)\Xi(\tau)^{-\frac{1}{2}}S(\lambda)S^{-1}(X\beta_0+u) = c_1(\gamma) + c_2(\gamma) + c_3(\gamma)$, where

$$\begin{aligned}
c_1(\gamma) &= n^{-1}\beta_0'X'T'(\lambda)\Xi(\tau)'^{-\frac{1}{2}}M(\tau)\Xi(\tau)^{-\frac{1}{2}}T(\lambda)X\beta_0,\\
c_2(\gamma) &= n^{-1}\sigma_0^2 tr\left(T'(\lambda)\Xi(\tau)'^{-\frac{1}{2}}M(\tau)\Xi(\tau)^{-\frac{1}{2}}T(\lambda)\Xi\right),\\
c_3(\gamma) &= n^{-1}tr\left(T'(\lambda)\Xi(\tau)'^{-\frac{1}{2}}M(\tau)\Xi(\tau)^{-\frac{1}{2}}T(\lambda)\left(uu' - \sigma_0^2\Xi\right)\right)\\
&\quad + 2n^{-1}\beta_0'X'T'(\lambda)\Xi(\tau)'^{-\frac{1}{2}}M(\tau)\Xi(\tau)^{-\frac{1}{2}}T(\lambda)u.
\end{aligned}$$

Then

$$\begin{aligned}
\log\frac{\overline{\sigma}^2(\gamma)}{\sigma^2(\gamma)} &= \log\frac{\overline{\sigma}^2(\gamma)}{(c_1(\gamma)+c_2(\gamma))} + \log\frac{c_1(\gamma)+c_2(\gamma)}{\sigma^2(\gamma)}\\
&= \log\left(1+\frac{c_3(\gamma)}{c_1(\gamma)+c_2(\gamma)}\right) + \log\left(1+\frac{c_1(\gamma)-f(\gamma)}{\sigma^2(\gamma)}\right),
\end{aligned}$$

where $f(\gamma) = n^{-1}\sigma_0^2 tr\left(\Xi'^{\frac{1}{2}}T'(\lambda)\Xi(\tau)'^{-\frac{1}{2}}(I_n - M(\tau))\Xi(\tau)^{-\frac{1}{2}}T(\lambda)\Xi^{\frac{1}{2}}\right)$. Then (A.18) implies

$$\begin{aligned}
P\left(\left\|\hat{\gamma}-\gamma_0\right\| \in \overline{\mathcal{N}}^{\gamma}(\eta)\right) &= P\left(\inf_{\gamma\in\overline{\mathcal{N}}^{\gamma}(\eta)} \mathcal{L}(\gamma) - \mathcal{L} \le 0\right)\\
&\le P\left(\log\left(1+\sup_{\gamma\in\overline{\mathcal{N}}^{\gamma}(\eta)}\left|\frac{c_3(\gamma)}{c_1(\gamma)+c_2(\gamma)}\right|\right) + \left|\log\left(\overline{\sigma}^2/\sigma_0^2\right)\right|\right.\\
&\ge \inf_{\gamma\in\overline{\mathcal{N}}^{\gamma}(\eta)}\left(\log\left(1+\frac{c_1(\gamma)-f(\gamma)}{\sigma^2(\gamma)}\right) + \log r(\gamma)\right)\Bigg),
\end{aligned}$$

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

where recall that $\overline{\mathcal{N}}^{\gamma}(\eta) = \Gamma \backslash \mathcal{N}^{\gamma}(\eta)$, $\mathcal{N}^{\gamma}(\eta) = \{\gamma : \|\gamma - \gamma_0\| < \eta\} \cap \Gamma$. Because $\overline{\sigma}^2/\sigma_0^2 \xrightarrow{p} 1$, the property $\log(1+x) = x + o(x)$ as $x \to 0$ implies that it is sufficient to show that

$$\sup_{\gamma \in \overline{\mathcal{N}}^{\gamma}(\eta)} \left| \frac{c_3(\gamma)}{c_1(\gamma) + c_2(\gamma)} \right| \xrightarrow{p} 0, \tag{A.19}$$

$$\sup_{\gamma \in \overline{\mathcal{N}}^{\gamma}(\eta)} \left| \frac{f(\gamma)}{\sigma^2(\gamma)} \right| \xrightarrow{p} 0, \tag{A.20}$$

$$P\left(\inf_{\gamma \in \overline{\mathcal{N}}^{\gamma}(\eta)} \left\{ \frac{c_1(\gamma)}{\sigma^2(\gamma)} + \log r(\gamma) \right\} > 0 \right) \longrightarrow 1. \tag{A.21}$$

Because $\overline{\mathcal{N}}^{\gamma}(\eta) \subseteq \left\{ \Lambda \times \overline{\mathcal{N}}^{\tau}(\eta/2) \right\} \cup \left\{ \overline{\mathcal{N}}^{\lambda}(\eta/2) \times \mathcal{T} \right\}$, we have

$$\begin{aligned} P\left(\inf_{\gamma \in \overline{\mathcal{N}}^{\gamma}(\eta)} \left\{ \frac{c_1(\gamma)}{\sigma^2(\gamma)} + \log r(\gamma) \right\} > 0 \right) &\geq P\left(\min \left\{ \inf_{\Lambda \times \overline{\mathcal{N}}^{\tau}(\eta/2)} \frac{c_1(\gamma)}{\sigma^2(\gamma)}, \inf_{\overline{\mathcal{N}}^{\lambda}(\eta/2)} \log r(\gamma) \right\} > 0 \right) \\ &\geq P\left(\min \left\{ \inf_{\Lambda \times \overline{\mathcal{N}}^{\tau}(\eta/2)} \frac{c_1(\gamma)}{C}, \inf_{\overline{\mathcal{N}}^{\lambda}(\eta/2)} \log r(\gamma) \right\} > 0 \right), \end{aligned}$$

from Assumption 15, whence Assumptions 16 and 17 imply (A.21). Again using Assumption 15, uniformly in γ, $\left| f(\gamma)/\sigma^2(\gamma) \right| = \mathcal{O}_{\mathrm{p}}(|f(\gamma)|)$ and

$$\begin{aligned} |f(\gamma)| &= \mathcal{O}_{\mathrm{p}}\left(tr\left(\Xi'^{\frac{1}{2}} T'(\lambda) \Xi(\tau)^{-1} X \left(X' \Xi(\tau)^{-1} X \right)^{-1} X' \Xi(\tau)^{-1} T(\lambda) \Xi^{\frac{1}{2}} \right) / n \right) \\ &= \mathcal{O}_{\mathrm{p}}\left(tr\left(\Xi'^{\frac{1}{2}} T'(\lambda) \Xi(\tau)^{-1} X X' \Xi(\tau)^{-1} T(\lambda) \Xi^{\frac{1}{2}} \right) / n^2 \right) = \mathcal{O}_{\mathrm{p}}\left(\left\| X' \Xi(\tau)^{-1} T(\lambda) \Xi^{\frac{1}{2}} / n \right\|_F^2 \right) \\ &= \mathcal{O}_{\mathrm{p}}\left(\|X/n\|_F^2 \, \overline{\varphi}^2\left(\Xi(\tau)^{-1} \right) \|T(\lambda)\|^2 \left\| \Xi^{\frac{1}{2}} \right\|^2 \right) = \mathcal{O}_{\mathrm{p}}\left(\|X/n\|_F^2 \, \|T(\lambda)\|^2 \overline{\varphi}(\Xi) / \underline{\varphi}^2(\Xi(\tau)) \right) \\ &= \mathcal{O}_{\mathrm{p}}\left(\|T(\lambda)\|^2 / n \right), \end{aligned} \tag{A.22}$$

where we have twice made use of the inequality

$$\|AB\|_F \leq \|A\|_F \, \|B\| \tag{A.23}$$

for generic multiplication compatible matrices A and B. (A.20) now follows by Assumption 12 and compactness of Λ because $T(\lambda) = I_n + \sum_{j=1}^{p} (\lambda_{0j} - \lambda_j) G_j$. Finally consider (A.19). We first prove pointwise convergence. For any fixed $\gamma \in \overline{\mathcal{N}}^{\gamma}(\eta)$ and large enough n, Assumptions 15 and 17 imply

$$\{c_1(\gamma)\}^{-1} = \mathcal{O}_{\mathrm{p}}\left(\|\beta_0\|^{-2} \right) = \mathcal{O}_{\mathrm{p}}(1) \tag{A.24}$$

$$\{c_2(\gamma)\}^{-1} = \mathcal{O}_{\mathrm{p}}(1), \tag{A.25}$$

because $\left\{ n^{-1} \sigma_0^2 tr\left(\Xi'^{\frac{1}{2}} T'(\lambda) \Xi(\tau)^{-1} T(\lambda) \Xi^{\frac{1}{2}} \right) \right\}^{-1} = \mathcal{O}_{\mathrm{p}}(1)$ and, proceeding like in the bound for $|f(\gamma)|$, $tr\left(\Xi'^{\frac{1}{2}} T'(\lambda) \Xi(\tau)'^{-\frac{1}{2}} (I - M(\tau)) \Xi(\tau)^{-\frac{1}{2}} T(\lambda) \Xi^{\frac{1}{2}} \right) =$

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

$\mathscr{O}_{\mathrm{p}}\left(\|T(\lambda)\|^2/n\right) = \mathscr{O}_{\mathrm{p}}(1/n)$. In fact it is worth noting for the equicontinuity argument presented later that Assumptions 15 and 17 actually imply that (A.24) and (A.25) hold uniformly over $\overline{\mathcal{N}}^{\gamma}(\eta)$, a property not needed for the present pointwise arguments. Thus $c_3(\gamma)/(c_1(\gamma)+c_2(\gamma)) = \mathscr{O}_{\mathrm{p}}(|c_3(\gamma)|)$ where, writing $\mathfrak{B}(\gamma) = T'(\lambda)\Xi(\tau)'^{-\frac{1}{2}}M(\tau)\,\Xi(\tau)^{-\frac{1}{2}}T(\lambda)$ with typical element $\mathfrak{b}_{rs}(\gamma)$, $r,s = 1,\ldots,n$, $c_3(\gamma)$ has mean 0 and variance

$$\mathscr{O}_{\mathrm{p}}\left(\frac{\|\mathfrak{B}(\gamma)\Xi\|_F^2}{n^2} + \frac{\sum_{r,s,t,v=1}^{n}\mathfrak{b}_{rs}(\gamma)\mathfrak{b}_{tv}(\gamma)\kappa_{rstv}}{n^2} + \frac{\left\|\beta_0' X'\mathfrak{B}(\gamma)\Xi^{\frac{1}{2}}\right\|^2}{n^2}\right), \tag{A.26}$$

with κ_{rstv} denoting the fourth cumulant of u_r, u_s, u_t, u_v, $r,s,t,v = 1,\ldots,n$. Under the linear process assumed in Assumption 1 it is known that

$$\sum_{r,s,t,v=1}^{n}\kappa_{rstv}^2 = \mathscr{O}(n). \tag{A.27}$$

Using (A.23) and Assumptions 12 and 13, the first term in parentheses in (A.26) is

$$\begin{aligned}\mathscr{O}_{\mathrm{p}}\left(\|\mathfrak{B}(\gamma)\|_F^2\,\overline{\varphi}^2(\Xi)/n^2\right) &= \mathscr{O}_{\mathrm{p}}\left(\|T(\lambda)\|_F^2\left\|\Xi(\tau)^{-\frac{1}{2}}\right\|^4\|M(\tau)\|^2\,\|T(\lambda)\|^2/n^2\right)\\ &= \mathscr{O}_{\mathrm{p}}\left(\|T(\lambda)\|^4/n\underline{\varphi}^2(\Xi(\tau))\right) = \mathscr{O}_{\mathrm{p}}\left(\|T(\lambda)\|^4/n\right),\end{aligned} \tag{A.28}$$

while the second is similarly

$$\mathscr{O}_{\mathrm{p}}\left\{\left(\|\mathfrak{B}(\gamma)\|_F^2/n\right)\left(\sum_{r,s,t,v=1}^{n}\kappa_{rstv}^2/n^2\right)^{\frac{1}{2}}\right\} = o_p\left(\|T(\lambda)\|^4\right), \tag{A.29}$$

using (A.27). Finally, using (3.2), the third term in parentheses in (A.26) is

$$\mathscr{O}_{\mathrm{p}}\left(\|\mathfrak{B}(\gamma)\|^2/n\right) = \mathscr{O}_{\mathrm{p}}\left(\|T(\lambda)\|^4/n\right). \tag{A.30}$$

By compactness of Λ and Assumption 12, (A.28), (A.29), and (A.30) are negligible, thus pointwise convergence is established.

Uniform convergence will follow from an equicontinuity argument. First, for arbitrary $\varepsilon > 0$ we can find points $\gamma_* = \left(\lambda_*', \tau_*'\right)'$, possibly infinitely many, such that the neighbourhoods $\|\gamma - \gamma^*\| < \varepsilon$ form an open cover of $\overline{\mathcal{N}}^{\gamma}(\eta)$. Since Γ is compact any open cover has a finite subcover and thus we may in fact choose finitely many $\gamma_* = \left(\lambda_*', \tau_*'\right)'$, whence it suffices to prove

$$\sup_{\|\gamma-\gamma_*\|<\varepsilon}\left|\frac{c_3(\gamma)}{c_1(\gamma)+c_2(\gamma)} - \frac{c_3(\gamma_*)}{c_1(\gamma_*)+c_2(\gamma_*)}\right| \xrightarrow{p} 0.$$

Proceeding as in Gupta and Robinson (2018), we denote the two components of $c_3(\gamma)$ by $c_{31}(\gamma)$, $c_{32}(\gamma)$, and are left with establishing the negligibility of

$$\frac{|c_{31}(\gamma)-c_{31}(\gamma_*)|}{c_2(\gamma)} + \frac{|c_{32}(\gamma)-c_{32}(\gamma_*)|}{c_1(\gamma)} + \frac{|c_3(\gamma_*)|}{c_1(\gamma)c_1(\gamma_*)}|c_1(\gamma_*)-c_1(\gamma)| + \frac{|c_3(\gamma_*)|}{c_2(\gamma)c_2(\gamma_*)}|c_2(\gamma_*)-c_2(\gamma)|, \tag{A.31}$$

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

uniformly on $\|\gamma - \gamma_*\| < \varepsilon$. By the fact that (A.24) and (A.25) hold uniformly over Γ, we first consider only the numerators in the first two terms in (A.31). As in the proof of Theorem 1 of Delgado and Robinson (2015), (A.23) implies that $\mathbb{E}\left(\sup_{\|\gamma-\gamma_*\|<\varepsilon}|c_{31}(\gamma)-c_{31}(\gamma_*)|\right)$ is bounded by $n^{-1}\left(\mathbb{E}\|u\|^2+\sigma_0^2 tr\,\Xi\right)\sup_{\|\gamma-\gamma_*\|<\varepsilon}\|\mathfrak{B}(\gamma)-\mathfrak{B}(\gamma_*)\| = \mathcal{O}_{\mathrm{p}}\left(\sup_{\|\gamma-\gamma_*\|<\varepsilon}\|\mathfrak{B}(\gamma)-\mathfrak{B}(\gamma_*)\|\right)$, because $\mathbb{E}\|u\|^2 = \mathcal{O}(n)$ and $tr\,\Xi = \mathcal{O}(n)$. $\mathfrak{B}(\gamma)-\mathfrak{B}(\gamma_*)$ can be written as

$$(T(\lambda)-T(\lambda_*))'\,\Xi(\tau)'^{-\frac{1}{2}}M(\tau)\Xi(\tau)^{-\frac{1}{2}}T(\lambda)+T(\lambda_*)'\Xi'(\tau_*)M(\tau_*)\Xi(\tau_*)^{-\frac{1}{2}}(T(\lambda)-T(\lambda_*))$$
$$+\,T'(\lambda_*)\left(\Xi(\tau)'^{-\frac{1}{2}}M(\tau)\Xi(\tau)^{-\frac{1}{2}}-\Xi(\tau_*)'^{-\frac{1}{2}}M(\tau_*)\Xi(\tau_*)^{-\frac{1}{2}}\right)T(\lambda), \tag{A.32}$$

which, by the triangle inequality, has spectral norm bounded by

$$\|T(\lambda)-T(\lambda_*)\|\left(\left\|\Xi(\tau)^{-\frac{1}{2}}\right\|^2\|T(\lambda)\|+\left\|\Xi(\tau_*)^{-\frac{1}{2}}\right\|^2\|T(\lambda_*)\|\right)$$
$$+\,\|T(\lambda_*)\|\left\|\Xi(\tau)'^{-\frac{1}{2}}M(\tau)\Xi(\tau)^{-\frac{1}{2}}-\Xi(\tau_*)'^{-\frac{1}{2}}M(\tau_*)\Xi(\tau_*)^{-\frac{1}{2}}\right\|\|T(\lambda)\|$$
$$=\mathcal{O}_{\mathrm{p}}\left(\|T(\lambda)-T(\lambda_*)\|+\left\|\Xi(\tau)'^{-\frac{1}{2}}M(\tau)\Xi(\tau)^{-\frac{1}{2}}-\Xi(\tau_*)'^{-\frac{1}{2}}M(\tau_*)\Xi(\tau_*)^{-\frac{1}{2}}\right\|\right) \tag{A.33}$$

By Assumption 12 the first term in parentheses on the right side of (A.33) is bounded uniformly on $\|\gamma-\gamma_*\|<\varepsilon$ by

$$\sum_{j=1}^{p}|\lambda_j-\lambda_{*j}|\,\|G_j\| \le \max_{j=1,\ldots,p}\|G_j\|\,\|\lambda-\lambda_*\| = \mathcal{O}_{\mathrm{p}}(\varepsilon), \tag{A.34}$$

while because $\Xi(\tau)'^{-\frac{1}{2}}M(\tau)\Xi(\tau)^{-\frac{1}{2}} = n^{-1}\Xi(\tau)^{-1}X\left(n^{-1}X'\Xi(\tau)^{-1}X\right)^{-1}X'\Xi(\tau)^{-1}$ for any $\tau\in\mathcal{T}$, the second one can be decomposed into terms with bounds typified by

$$n^{-1}\left\|\Xi(\tau)^{-1}-\Xi(\tau_*)^{-1}\right\|\|X\|^2\left\|\left(n^{-1}X'\Xi(\tau)^{-1}X\right)^{-1}\right\|\left\|\Xi(\tau)^{-1}\right\|^2$$
$$\le n^{-1}\|\Xi(\tau)-\Xi(\tau_*)\|\,\|X\|^2\left\|\left(n^{-1}X'\Xi(\tau)^{-1}X\right)^{-1}\right\|\left\|\Xi(\tau)^{-1}\right\|^3\left\|\Xi(\tau_*)^{-1}\right\|$$
$$=\mathcal{O}_{\mathrm{p}}(\|\Xi(\tau)-\Xi(\tau_*)\|)=\mathcal{O}_{\mathrm{p}}(\varepsilon),$$

uniformly on $\|\gamma-\gamma_*\|<\varepsilon$, by Assumptions 13, 14, 18, (3.2) and the inequality $\|A\|\le\|A\|_F$ for a generic matrix A, so that

$$\sup_{\|\gamma-\gamma_*\|<\varepsilon}\|\mathfrak{B}(\gamma)-\mathfrak{B}(\gamma_*)\| = \mathcal{O}_{\mathrm{p}}(\varepsilon). \tag{A.35}$$

Thus equicontinuity of the first term in (A.31) follows because ε is arbitrary. The equicontinuity of the second term in (A.31) follows in much the same way. Indeed $\sup_{\|\gamma-\gamma_*\|<\varepsilon}c_{32}(\gamma)-c_{32}(\gamma_*) = 2n^{-1}\beta_0'X'\sup_{\|\gamma-\gamma_*\|<\varepsilon}(\mathfrak{B}(\gamma)-\mathfrak{B}(\gamma_*))u = \mathcal{O}_{\mathrm{p}}\left(\sup_{\|\gamma-\gamma_*\|<\varepsilon}\|\mathfrak{B}(\gamma)-\mathfrak{B}(\gamma_*)\|\right) = \mathcal{O}_{\mathrm{p}}(\varepsilon)$, using earlier arguments and (A.35). Because $c_1(\gamma)$ is bounded and bounded away from zero in probability (see A.24) for

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

sufficiently large n and all $\gamma \in \overline{\mathcal{N}}^{\gamma}(\eta)$, the third term in (A.31) may be bounded by $|c_3(\gamma_*)|/c_1(\gamma_*)\,(1+c_1(\gamma_*)/c_1(\gamma))\xrightarrow{p} 0$, convergence being uniform on $\|\gamma-\gamma_*\|<\varepsilon$ by pointwise convergence of $c_3(\gamma)/(c_1(\gamma)+c_2(\gamma))$, cf. Gupta and Robinson (2018). The uniform convergence to zero of the fourth term in (A.31) follows in identical fashion because $c_2(\gamma)$ is bounded and bounded away from zero (see (A.25)) in probability for sufficiently large n and all $\gamma \in \overline{\mathcal{N}}^{\gamma}(\eta)$. This concludes the proof. ■

Proof of Theorem 4.2. We first introduce the derivative of (4.1) evaluated at $\left(\varrho_0', \sigma_0^2\right)'$:

$$d = -n^{-1}\sigma_0^{-2}\Big[2\left(u'G_1'\Xi^{-1}u-\sigma_0^2 trG_1,\ldots,u'G_p'\Xi^{-1}u-\sigma_0^2 trG_p,0',0'\right)'$$
$$+\left(0',0',u'\Xi^{-1}\Xi_1\Xi^{-1}u-\sigma_0^2 tr\left(\Xi^{-1}\Xi_1\right),\ldots,u'\Xi^{-1}\Xi_q\Xi^{-1}u-\sigma_0^2 tr\left(\Xi^{-1}\Xi_q\right)\right)'$$
$$+2\left(u'\Xi^{-1}A,u'\Xi^{-1}X,0'\right)'\Big]. \tag{A.36}$$

The first part of the proof consists of proving that

$$n^{\frac{1}{2}}N\ d \xrightarrow{d} \mathcal{N}(0,\Delta). \tag{A.37}$$

Proving (A.37) requires the establishment of asymptotic normality for normalized linear combinations of the LHS whence the required multivariate normality follows via the Cramer–Wold device, as usual. From (A.36) we observe that such a linear combination will have terms that are linear in u and quadratic in u, and our aim of establishing asymptotic normality will be achieved by writing these terms as sums of martingale differences in the ϵ_l. First, notice that $d=\sum_{l=1}^{\infty} w_l$ with $w_l = w_{1l}+w_{2l}+w_{3l}$ and $w_{1l} = -2n^{-1}\sigma_0^{-2}\left(v_{ll}'^{\lambda}\left(\epsilon_l^2-\sigma_0^2\right)+2\mathbb{1}(l\geq 2)\epsilon_l\sum_{k=1}^{l-1}v_{lk}'^{\lambda}\epsilon_k,0',0'\right)', w_{2l} = -2n^{-1}\sigma_0^{-2}\left(c_l'\Xi^{-1}A\epsilon_l,c_l'\Xi^{-1}X\epsilon_l,0'\right)', w_{3l} = -n^{-1}\sigma_0^{-2}\left(0',0',v_{ll}'^{\tau}\left(\epsilon_l^2-\sigma_0^2\right)+2\mathbb{1}(l\geq 2)\epsilon_l\sum_{k=1}^{l-1}v_{lk}'^{\tau}\epsilon_k\right)'$, where v_{lk}^{λ} is a $p\times 1$ vector with j-th element $c_l'G_j\Xi^{-1}c_k$ and v_{lk}^{τ} is a $q\times 1$ vector with j-th element $c_l'\Xi^{-1}\Xi_j\Xi^{-1}c_k$ $l,k\geq 1$. Notice that here k is used as an index, without risk of confusion with the dimension of β. We first show that

$$n^{\frac{1}{2}}Nd_* \xrightarrow{p} 0, \tag{A.38}$$

where $d_*=d-d_L$, $d_L=\sum_{l=1}^{L}w_l$ and $L=L_n$ is a positive integer sequence that is increasing in n. All expectations in the sequel are taken conditional on W_j and X, $j=1,\ldots,p$.

By Chebyshev's inequality proving

$$\mathbb{E}\left\|n^{\frac{1}{2}}Nd_*\right\|^2 \xrightarrow{p} 0 \tag{A.39}$$

is sufficient to establish (A.38). The LHS of (A.39) is bounded by a constant times

$$n\sum_{l=L+1}^{\infty}\mathbb{E}\|Nw_l\|^2 \leq Cn\sum_{l=L+1}^{\infty}\left(\mathbb{E}\|Nw_{1l}\|^2+\mathbb{E}\|Nw_{2l}\|^2+\mathbb{E}\|Nw_{3l}\|^2\right), \tag{A.40}$$

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

by uncorrelatedness of w_l due to Assumption 1, and the triangle and c_r inequalities. The negligibility of $n\sum_{l=L+1}^{\infty}\mathbb{E}\|Nw_{3l}\|^2$ follows by the proof of (57) in Delgado and Robinson (2015), while the linearity in ϵ_l of w_{2l} implies that $n\sum_{l=L+1}^{\infty}\mathbb{E}\|Nw_{2l}\|^2$ is easily shown to be negligible as in Lemma A1 of Robinson and Thawornkaiwong (2012). It remains to prove that $n\sum_{l=L+1}^{\infty}\mathbb{E}\|Nw_{1l}\|^2$ is negligible. Notice that $\mathbb{E}\|Nw_{1l}\|^2 \le Cn^{-2}\left\|\nu_{ll}^{\lambda}\right\|^2 + Cn^{-2}\mathbb{1}(l\ge 2)\sum_{k=1}^{l-1}\left\|\nu_{lk}^{\lambda}\right\|^2 \le Cn^{-2}\sum_{k=1}^{l}\left\|\nu_{lk}^{\lambda}\right\|^2$, so that

$$n\sum_{l=L+1}^{\infty}\mathbb{E}\|Nw_{1l}\|^2 \le Cn^{-1}\sum_{l=L+1}^{\infty}\sum_{k=1}^{l}\left\|\nu_{lk}^{\lambda}\right\|^2 \le Cn^{-1}\sum_{j=1}^{p}\sum_{l=L+1}^{\infty}c_l'G_j\Xi^{-1}\sum_{k=1}^{l}c_kc_k'\Xi^{-1}G_j'c_l$$

$$\le C\|\Xi\|n^{-1}\sum_{j=1}^{p}\sum_{l=L+1}^{\infty}c_l'G_j\Xi^{-1}\Xi^{-1}G_j'c_l \le Cn^{-1}\sum_{j=1}^{p}\sum_{l=L+1}^{\infty}\sum_{r,s,t=1}^{n}c_{rl}c_{tl}p_{jrs}p_{jts}$$

$$\le Cn^{-1}\sum_{l=L+1}^{\infty}\sum_{r,t=1}^{n}\left(\sup_{\substack{j=1,\dots,p\\ s=1,\dots,n}}|p_{jrs}|\right)|c_{rl}^*||c_{tl}^*|\left(\max_{j=1,\dots,p}\sum_{s=1}^{n}|p_{jts}|\right), \tag{A.41}$$

where p_{jrs} is the (r,s)-th element of $G_j\Xi^{-1}$. Note that

$$|p_{jrs}| \le \sum_{s=1}^{n}|p_{jrs}| \le \left\|G_j\Xi^{-1}\right\|_R = \mathcal{O}_{\mathrm{p}}(1), \tag{A.42}$$

uniformly in r,s,j, by Assumptions 21 and 22. Thus (A.41) is

$$\mathcal{O}_{\mathrm{p}}\left(n^{-1}\sum_{r=1}^{n}\sum_{l=L+1}^{\infty}|c_{rl}^*|\sum_{t=1}^{n}|c_{tl}^*|\right) = \mathcal{O}_{\mathrm{p}}\left(\sup_{r=1,\dots,n}\sum_{l=L+1}^{\infty}|c_{rl}^*|\right), \tag{A.43}$$

by Assumption 20. By the same assumption, there exists L_{rn} such that $\sum_{l=L_{rn}+1}^{\infty}|c_{rl}^*| \le \varepsilon_n$ for any decreasing sequence $\varepsilon_n \to 0$ as $n\to\infty$. Choosing $L = \max_{r=1,\dots,n}L_{rn}$ in d_L, we deduce that (A.43) is $\mathcal{O}_{\mathrm{p}}(\varepsilon_n) = o_p(1)$, proving (A.39). Thus we need only focus on d_L.

Define $D = n\mathbb{E}\left(Nd_Ld_L'N\right) = n\sum_{l=1}^{L}N\mathbb{E}\left(w_lw_l'\right)N$. In view of our subsequent calculations it is straightforward to show that $D \xrightarrow{p} \Delta$, implying that D is positive definite for sufficiently large n and assuring the existence of a square matrix, denoted $D^{\frac{1}{2}}$, such that $D'^{\frac{1}{2}}D^{\frac{1}{2}} = D$. For a $(p+k+q)\times 1$ constant vector α such that $\|\alpha\| = 1$ define $s = n^{\frac{1}{2}}\alpha'D^{-\frac{1}{2}}Nd_L = n^{\frac{1}{2}}\sum_{l=1}^{L}\alpha'D^{-\frac{1}{2}}Nw_l$, which has zero mean and unit variance, and we seek to establish that

$$s \xrightarrow{d} \mathcal{N}(0,1), \text{ as } n\to\infty. \tag{A.44}$$

From Scott (1973), (A.44) follows if

$$\sum_{l=1}^{L}\mathbb{E}\left\|n^{\frac{1}{2}}D^{-\frac{1}{2}}Nw_l\right\|^4 \xrightarrow{p} 0, \text{ as } n\to\infty \tag{A.45}$$

and

$$n\sum_{l=1}^{L}\left[\mathbb{E}\left(Nw_lw_l'N\middle|\epsilon_k, k<l\right) - \mathbb{E}\left(Nw_lw_l'N\right)\right] \xrightarrow{p} 0, \text{ as } n\to\infty. \tag{A.46}$$

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

To prove (A.45) it is sufficient to show

$$\sum_{l=1}^{L}\mathbb{E}\left\|n^{\frac{1}{2}}D^{-\frac{1}{2}}Nw_{il}\right\|^4 \xrightarrow{p} 0, \text{ as } n\to\infty, i=1,2,3, \tag{A.47}$$

by the triangle and c_r inequalities.

For $i=3$ the proof of (A.47) is identical to the proof of claim (60) in Delgado and Robinson (2015). Writing $z_l^{\lambda}=A'\Xi^{-1}c_l$ and $z_l^{\beta}=X'\Xi^{-1}c_l$, for $i=2$ the left side of (A.47) is equal to a constant times $n^{-2}\sum_{l=1}^{L}\left(\left(z_l'^{\lambda}\epsilon_l, z_l'^{\beta}\epsilon_l, 0'\right)D\left(z_l'^{\lambda}\epsilon_l, z_l'^{\beta}\epsilon_l, 0'\right)'\right)^2$, which is bounded by a constant times $n^{-2}\sum_{l=1}^{L}\left(\left\|z_l^{\lambda}\right\|^4+\left\|z_l^{\beta}\right\|^4\right)\|D\|^2$. The last factor in the latter is $\mathcal{O}_{\mathrm{p}}(1)$ because $\|D\|^2$ converges in probability to $\|\Delta\|^2$, while

$$\begin{aligned}\sum_{l=1}^{L}\left\|z_l^{\lambda}\right\|^4 &= \sum_{l=1}^{L}\left(z_l'^{\lambda}z_l^{\lambda}\right)^2 \le \sum_{l=1}^{\infty}\left(z_l'^{\lambda}z_l^{\lambda}\right)^2 = \sum_{j=1}^{p}e_j'\Xi^{-1}A\sum_{l=1}^{\infty}c_lc_l'A'\Xi^{-1}e_j\\ &= \sum_{j=1}^{p}e_j'\Xi^{-1}A\Xi'A'\Xi^{-1}e_j = \mathcal{O}_{\mathrm{p}}\left(\|A\|^2\left\|\Xi^{-1}\right\|^2\|\Xi\|\right)\\ &= \mathcal{O}_{\mathrm{p}}\left(\|A\|^2\right) = \mathcal{O}_{\mathrm{p}}\left(\max_{j=1,\ldots,p}\left\|G_j\right\|^2\|X\|^2\right) = \mathcal{O}_{\mathrm{p}}(n),\end{aligned}$$

by Assumption 21 and (3.2). Similarly $\sum_{l=1}^{L}\left\|z_l^{\beta}\right\|^4 = \mathcal{O}_{\mathrm{p}}(n)$, so $\sum_{l=1}^{L}\mathbb{E}\left\|n^{\frac{1}{2}}D^{-\frac{1}{2}}Nw_{2l}\right\|^4 = \mathcal{O}_{\mathrm{p}}(n^{-1})$ and (A.47) is proved for $i=2$. Finally consider (A.47) for $i=1$. Evaluating the expectation, using an ℓ_p norm inequality and employing techniques used earlier, it follows that

$$\begin{aligned}&\mathbb{E}\left\|D^{-\frac{1}{2}}Nw_{1l}\right\|^4 \le Cn^{-4}\left\|v_{ll}^{\lambda}\right\|^4 + Cn^{-4}\mathbb{E}\left\|\sum_{k=1}^{l-1}v_{lk}^{\lambda}\epsilon_k\right\|^4 \le Cn^{-4}\left\|v_{ll}^{\lambda}\right\|^4 + Cn^{-4}\left(\sum_{k=1}^{l-1}\left\|v_{lk}^{\lambda}\right\|^2\right)^2\\ &\le Cn^{-4}\left(\sum_{k=1}^{l}\left\|v_{lk}^{\lambda}\right\|^2\right)^2 \le Cn^{-4}\left(\sum_{j=1}^{p}c_l'G_j\Xi^{-1}\sum_{k=1}^{l}c_kc_k'\Xi^{-1}G_j'c_l\right)^2 = \mathcal{O}_{\mathrm{p}}\left(n^{-4}\|c_l\|^4\right),\end{aligned}$$

whence

$$\begin{aligned}\sum_{l=1}^{L}\mathbb{E}\left\|n^{\frac{1}{2}}D^{-\frac{1}{2}}Nw_{2l}\right\|^4 &= \mathcal{O}_{\mathrm{p}}\left(n^{-2}\sum_{l=1}^{L}\|c_l\|^4\right) = \mathcal{O}_{\mathrm{p}}\left(n^{-2}\sum_{l=1}^{L}\left(\sum_{r=1}^{n}c_{rl}^{*2}\right)^2\right)\\ &= \mathcal{O}_{\mathrm{p}}\left(n^{-2}\sum_{l=1}^{L}\left(\sum_{r=1}^{n}\left|c_{rl}^{*}\right|\right)^4\right) = \mathcal{O}_{\mathrm{p}}\left(n^{-2}\sum_{r=1}^{n}\left(\sum_{l=1}^{L}\left|c_{rl}^{*}\right|\right)\right) = \mathcal{O}_{\mathrm{p}}\left(n^{-1}\right),\end{aligned}$$

by Assumption 20 and the ℓ_p norm inequality. Thus (A.47), and hence (A.45), is established.

To prove (A.46) it is sufficient to show

$$n\sum_{l=1}^{L}\left[\mathbb{E}\left(Nw_{il}w_{jl}'N\middle|\epsilon_k, k<l\right)-\mathbb{E}\left(Nw_{il}w_{jl}'N\right)\right]\xrightarrow{p}0, \text{ as } n\to\infty, i,j=1,2,3, j\ge i. \tag{A.48}$$

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

Once again, for $i = j = 3$ the proof is identical to that of (61) in Delgado and Robinson (2015) and for $i = j = 2$ it follows easily as in the proof of a similar assertion in Theorem 2.2. Of the remaining cases the most complex to handle is $i = j = 1$ in view of both its quadraticity in u and involvement of the G_i, so we prove this first.

Notice that the only nonzero block of $\mathbb{E}\left(w_{1l}w'_{1l}\middle|\epsilon_k, k < l\right)$ equals

$$4n^{-2}\sigma_0^{-4}\left\{\left(\mu_4 - \sigma_0^4\right)\nu_{ll}^{\lambda}\nu_{ll}^{\prime\lambda} + \mu_3\mathbb{1}(l \geq 2)\sum_{k=1}^{l-1}\left(\nu_{lk}^{\lambda}\nu_{ll}^{\prime\lambda} + \nu_{ll}^{\lambda}\nu_{lk}^{\prime\lambda}\right)\epsilon_k\right\}$$
$$+ 4n^{-2}\sigma_0^{-2}\mathbb{1}(l \geq 2)\left(\sum_{k=1}^{l-1}\nu_{lk}^{\lambda}\epsilon_k\right)\left(\sum_{k=1}^{l-1}\nu_{lk}^{\lambda}\epsilon_k\right)',$$

with expectation $\mathbb{E}\left(w_{1l}w'_{1l}\right)$ (we only consider the nonzero block) equal to $4n^{-2}\sigma_0^{-4}\left(\mu_4 - \sigma_0^4\right)\nu_{ll}^{\lambda}\nu_{ll}^{\prime\lambda} + 4n^{-2}\mathbb{1}(l \geq 2)\sum_{k=1}^{l-1}\nu_{lk}^{\lambda}\nu_{lk}^{\prime\lambda}$, so that the Frobenius norm of (A.48) for $i = j = 1$ is bounded by a constant times

$$n^{-1}\left\|\sum_{l=2}^{L}\sum_{k=1}^{l-1}\left(\nu_{lk}^{\lambda}\nu_{ll}^{\prime\lambda} + \nu_{ll}^{\lambda}\nu_{lk}^{\prime\lambda}\right)\epsilon_k\right\|_F + n^{-1}\left\|\sum_{l=2}^{L}\left(\left(\sum_{k=1}^{l-1}\nu_{lk}^{\lambda}\epsilon_k\right)\left(\sum_{k=1}^{l-1}\nu_{lk}^{\lambda}\epsilon_k\right)' - \sigma_0^2\sum_{k=1}^{l-1}\nu_{lk}^{\lambda}\nu_{lk}^{\prime\lambda}\right)\right\|_F. \tag{A.49}$$

By transforming the range of summation, the square of the first term in (A.49) has expectation bounded by

$$Cn^{-2}\mathbb{E}\left\|\sum_{k=1}^{L-1}\left(\sum_{l=k+1}^{L}\left(\nu_{lk}^{\lambda}\nu_{ll}^{\prime\lambda} + \nu_{ll}^{\lambda}\nu_{lk}^{\prime\lambda}\right)\right)\epsilon_k\right\|_F^2 \leq Cn^{-2}\sum_{k=1}^{L-1}\left\|\sum_{l=k+1}^{L}\left(\nu_{lk}^{\lambda}\nu_{ll}^{\prime\lambda} + \nu_{ll}^{\lambda}\nu_{lk}^{\prime\lambda}\right)\right\|_F^2, \tag{A.50}$$

where the factor in the norm on the RHS of (A.50) is

$$\sum_{i,j=1}^{p}\sum_{l,m=k+1}^{L} c_l'G_i\Xi^{-1}c_lc_l'G_j\Xi^{-1}c_kc_m'G_i\Xi^{-1}c_mc_m'G_j\Xi^{-1}c_k$$
$$\leq \sum_{l,m=k+1}^{L}\left|\sum_{i=1}^{p}c_l'G_i\Xi^{-1}c_lc_m'G_i\Xi^{-1}c_m\right|\sum_{j=1}^{p}\left|c_l'G_j\Xi^{-1}c_k\right|\left|c_m'G_j\Xi^{-1}c_k\right|$$
$$\leq C\sum_{l,m=k+1}^{L}\max_{i=1,\ldots,p}\sum_{r,s,u,t=1}^{n}|c_{rl}||p_{irs}||c_{sm}||c_{ul}||p_{iut}||c_{tm}|\max_{j=1,\ldots,p}\left|c_l'G_j\Xi^{-1}c_k\right|\left|c_m'G_j\Xi^{-1}c_k\right|$$
$$\leq C\left(\sup_{i,r,s}|p_{irs}|\right)^2\left(\sup_{l\geq 1}\sum_{r=1}^{n}|c_{rl}^*|\right)^4\sum_{l,m=k+1}^{L}\max_{j=1,\ldots,p}\left|c_l'G_j\Xi^{-1}c_k\right|\left|c_m'G_j\Xi^{-1}c_k\right|$$
$$= \mathcal{O}_{\mathrm{p}}\left(\max_{j=1,\ldots,p}\left(\sum_{l=k+1}^{L}\left|c_k'G_j\Xi^{-1}c_l\right|\right)^2\right) = \mathcal{O}_{\mathrm{p}}\left(\max_{j=1,\ldots,p}\left(\sum_{l=k+1}^{L}\sum_{r,s=1}^{n}|c_{rk}^*||p_{jrs}||c_{sl}^*|\right)^2\right),$$

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

where we used Assumptions 20 and (A.42). Now Assumptions 20, 21, 22 and (A.42) imply that, uniformly in j, $\sum_{l=k+1}^{L}\sum_{r,s=1}^{n}|c_{rk}^{*}||p_{jrs}||c_{sl}^{*}| = \sum_{r=1}^{n}|c_{rk}^{*}|\sum_{s=1}^{n}|p_{jrs}|\sum_{l=k+1}^{L}|c_{sl}^{*}| = \mathcal{O}_{\mathrm{p}}\left(\sum_{r=1}^{n}|c_{rk}^{*}|\right)$, so (A.50) is $\mathcal{O}_{\mathrm{p}}\left(n^{-2}\sup_{k\geq 1}\left(\sum_{r=1}^{n}|c_{rk}^{*}|\right)\left(\sum_{r=1}^{n}\left(\sum_{k=1}^{L-1}|c_{rk}^{*}|\right)\right)\right)$. By Assumption 20 the latter is $\mathcal{O}_{\mathrm{p}}\left(n^{-1}\right)$ and therefore the first term in (A.49) is $\mathcal{O}_{\mathrm{p}}\left(n^{-\frac{1}{2}}\right)$.

Once again transforming the range of summation, we can rewrite the square of the second term in (A.49) as

$$n^{-2}\left\|\sum_{k=1}^{L-1}\sum_{l=k+1}^{L}\nu_{lk}^{\lambda}\nu_{lk}^{\prime\lambda}\left(\epsilon_{k}^{2}-\sigma_{0}^{2}\right)+\sum_{k=1}^{L-1}\sum_{m=1}^{k-1}\left(\sum_{l=k+1}^{L}\left(\nu_{lk}^{\lambda}\nu_{lm}^{\prime\lambda}+\nu_{lm}^{\lambda}\nu_{lk}^{\prime\lambda}\right)\right)\epsilon_{k}\epsilon_{m}\right\|_{F}^{2}. \tag{A.51}$$

A calculation along the lines of the proof of assertion (64) in Delgado and Robinson (2015) establishes that (A.51) is $\mathcal{O}_{\mathrm{p}}\left(\varkappa_{1}+\varkappa_{2}\right)$, where

$$\varkappa_{1}=n^{-2}\sum_{k=1}^{L-1}\left(\sum_{j=1}^{p}\sum_{l=k+1}^{L}\sum_{r,s=1}^{n}|c_{rl}^{*}||p_{jrs}||c_{sk}^{*}|\right)^{2}, \tag{A.52}$$

$$\begin{aligned}\varkappa_{2}=&\,n^{-2}\sum_{j,k,l,m=1}^{L}\sum_{i_{1},i_{2}=1}^{p}\sum_{r,s=1}^{n}|c_{rk}^{*}||p_{i_{1}rs}|\left|c_{sj}^{*}\right|\sum_{r,s=1}^{n}|c_{rk}^{*}||p_{i_{2}rs}||c_{sm}^{*}|\\&\times\sum_{r,s=1}^{n}|c_{rl}^{*}||p_{i_{1}rs}|\left|c_{sj}^{*}\right|\sum_{r,s=1}^{n}|c_{rl}^{*}||p_{i_{1}rs}||c_{sm}^{*}|.\end{aligned} \tag{A.53}$$

Now by (A.42), Assumptions 20, 21, 22 and elementary inequalities $\varkappa_{2}$ is

$$\begin{aligned}&\mathcal{O}_{\mathrm{p}}\left(n^{-2}\max_{i_{1},i_{2}=1,\ldots,p}\sum_{j,k,l=1}^{L}\left(\sum_{r,s=1}^{n}|c_{rk}^{*}||p_{i_{1}rs}|\left|c_{sj}^{*}\right|\right)\left(\sum_{r,s=1}^{n}|c_{rk}^{*}||p_{i_{2}rs}|\sum_{m=1}^{L}|c_{sm}^{*}|\right)\right.\\&\left.\times\sum_{r,s=1}^{n}|c_{rl}^{*}||p_{i_{1}rs}|\left|c_{sj}^{*}\right|\sum_{r=1}^{n}|c_{rl}^{*}|\sum_{s=1}^{n}|c_{sm}^{*}|\right)\\&=\mathcal{O}_{\mathrm{p}}\left(n^{-2}\max_{i_{1}=1,\ldots,p}\sum_{j,k=1}^{L}\left(\sum_{r,s=1}^{n}|c_{rk}^{*}||p_{i_{1}rs}|\left|c_{sj}^{*}\right|\right)\sum_{r=1}^{n}|c_{rk}^{*}|\left(\sum_{r,s=1}^{n}\sum_{l=1}^{L}|c_{rl}^{*}||p_{i_{1}rs}|\left|c_{sj}^{*}\right|\right)\right)\\&=\mathcal{O}_{\mathrm{p}}\left(n^{-2}\max_{i_{1}=1,\ldots,p}\sum_{r,s=1}^{n}\left(\sum_{k=1}^{L}|c_{rk}^{*}|\right)|p_{i_{1}rs}|\left(\sum_{j=1}^{L}\left|c_{sj}^{*}\right|\right)\sum_{s=1}^{n}\left(\sum_{r=1}^{n}|p_{i_{1}rs}|\right)\left|c_{sj}^{*}\right|\right)\\&=\mathcal{O}_{\mathrm{p}}\left(n^{-1}\max_{i_{1}=1,\ldots,p}\left\|\Xi^{-1}G_{i_{1}}^{\prime}\right\|_{R}\right)=\mathcal{O}_{\mathrm{p}}\left(n^{-1}\right),\end{aligned}$$

while a similar use of the conditions of the theorem implies that $\varkappa_{1}$ is

$$\mathcal{O}_{\mathrm{p}}\left(n^{-2}\sum_{k=1}^{L-1}\left(\max_{j=1,\ldots,p}\sum_{r,s=1}^{n}|p_{jrs}||c_{sk}^{*}|\right)^{4}\right)=\mathcal{O}_{\mathrm{p}}\left(n^{-2}\sum_{k=1}^{L-1}\left(\sum_{s=1}^{n}|c_{sk}^{*}|\right)^{4}\right)$$

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

$$= \mathcal{O}_{\mathrm{p}}\left(n^{-2}\sum_{s=1}^{n}\sum_{k=1}^{L-1}|c_{sk}^{*}|\left(\sum_{s=1}^{n}|c_{sk}^{*}|\right)^{3}\right) = \mathcal{O}_{\mathrm{p}}\left(n^{-2}\sum_{s=1}^{n}\sum_{k=1}^{L-1}|c_{sk}^{*}|\right) = \mathcal{O}_{\mathrm{p}}\left(n^{-1}\right),$$

proving (A.48) for $i = j = 1$, as desired. For the remaining combinations of i and j we discuss the proof briefly to avoid repetition.

When $i = 1, j = 2$, the only nonzero blocks of $\mathbb{E}\left(w_{1l}w_{2l}'\middle|\epsilon_k, k < l\right)$ equal $4n^{-2}\sigma_0^{-4}\mu_3\nu_{ll}^{\lambda}c_l'\Xi^{-1}A + 8n^{-2}\sigma_0^{-2}\mathbb{1}(l \geq 2)\sum_{k=1}^{l-1}\nu_{lk}^{\lambda}c_l'\Xi^{-1}A\epsilon_k$ and $4n^{-2}\sigma_0^{-4}\mu_3\nu_{ll}^{\lambda}c_l'\Xi^{-1}X + 8n^{-2}\sigma_0^{-2}\mathbb{1}(l \geq 2)\sum_{k=1}^{l-1}\nu_{lk}^{\lambda}c_l'\Xi^{-1}X\epsilon_k$ with expectations, respectively, equal to $4n^{-2}\sigma_0^{-4}\mu_3\nu_{ll}^{\lambda}c_l'\Xi^{-1}A$ and $4n^{-2}\sigma_0^{-4}\mu_3\nu_{ll}^{\lambda}c_l'\Xi^{-1}X$. Thus the Frobenius norm of (A.48) for $i = 1, j = 2$ is bounded by a constant times

$$n^{-1}\left\|\sum_{l=2}^{L}\sum_{k=1}^{l-1}\nu_{lk}^{\lambda}c_l'\Xi^{-1}A\epsilon_k\right\|_F + n^{-1}\left\|\sum_{l=2}^{L}\sum_{k=1}^{l-1}\nu_{lk}^{\lambda}c_l'\Xi^{-1}X\epsilon_k\right\|_F. \tag{A.54}$$

Writing $A_j = G_jX\beta_0$, $j = 1, \ldots, p$, and proceeding as in the earlier proof for $i = j = 1$ we deduce that the expectation of the square of the first term in (A.54) is bounded by

$$\begin{aligned}
Cn^{-2}\sum_{k=1}^{L-1}\left\|\sum_{l=k+1}^{L}\nu_{lk}^{\lambda}c_l'\Xi^{-1}A\right\|_F^2 &= Cn^{-2}\sum_{k=1}^{L-1}\sum_{i,j=1}^{p}\left\{c_k'\Xi^{-1}G_i'\left(\sum_{l=k+1}^{L}c_lc_l'\right)\Xi^{-1}A_i\right.\\
&\quad\left.\times A_j'\Xi^{-1}\left(\sum_{m=k+1}^{L}c_mc_m'\right)G_j\Xi^{-1}c_k\right\}\\
&\leq Cn^{-2}\sum_{k=1}^{L-1}\sum_{i,j=1}^{p}\left\{\left(c_k'\Xi^{-1}G_i'G_i\Xi^{-1}c_kc_k'\Xi^{-1}G_j'G_j\Xi^{-1}c_k\right)^{\frac{1}{2}}\right.\\
&\quad\left.\times\left(A_i'\Xi^{-1}\Xi^{-1}A_iA_j'\Xi^{-1}\Xi^{-1}A_j\right)^{\frac{1}{2}}\right\}\\
&\leq Cn^{-2}\max_{j=1,\ldots,p}\left(A_j'\Xi^{-2}A_j\right)\sum_{k=1}^{L-1}\left(c_k'\Xi^{-1}G_j'G_j\Xi^{-1}c_k\right),
\end{aligned} \tag{A.55}$$

where we use the inequality $x'Zy \leq \|x\|\,\|Z\|\,\|y\|$ and the fact that $\left\|\sum_{l=k+1}^{L}c_lc_l'\right\| \leq \|\Xi\| = \mathcal{O}(1)$. The first factor in parentheses in (A.55) is $\mathcal{O}_{\mathrm{p}}\left(\|A_j\|_R^2\left\|\Xi^{-1}\right\|_R^2\right) = \mathcal{O}_{\mathrm{p}}\left(\|G_j\|_R^2\|X\|_R^2\right) = \mathcal{O}_{\mathrm{p}}(1)$, uniformly in j by Assumptions 21, 22 and (3.2). The sum in (A.55) is bounded by

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

$$\sum_{k=1}^{\infty}\sum_{r,s=1}^{n} c_{rk}c_{sk}\left(\sum_{t=1}^{n} p_{jtr}p_{jts}\right) = tr\left(\Xi\Xi^{-1}G_j'G_j\Xi^{-1}\right) = tr\left(\Xi^{-\frac{1}{2}}G_j'G_j\Xi'^{-\frac{1}{2}}\right) = \left\|\Xi^{-\frac{1}{2}}G_j'\right\|_F^2$$
$$= \mathcal{O}_{\mathrm{p}}\left(\left\|\Xi^{-\frac{1}{2}}\right\|_F^2\right) = tr(\Xi^{-1}) = \mathcal{O}_{\mathrm{p}}(n),$$

uniformly in j by (A.23) and Assumption 12. The first term in (A.54) is thus established to be of order $\mathcal{O}_{\mathrm{p}}\left(n^{-\frac{1}{2}}\right)$, which is negligible. The second term in (A.54) is handled in an identical fashion.

When $i = 1, j = 3$, we consider the product of two zero mean blocks that are quadratic in ϵ_l, as in the $i = j = 1$ case, and the desired conclusion follows on proceeding as for that case. Finally the case $i = 2, j = 3$ involves a block that is the product of a linear and a quadratic term in ϵ, so it will also yield a term like (A.54) except with v_{lk}^{λ} replaced with v_{lk}^{τ}, whence the details follow as above and are omitted. Thus (A.37) is established.

For the next step of the proof we will introduce the second derivative matrix of (4.1) at any admissible point $\left(\varrho', \sigma^2\right)'$ in the parameter space. Write $\mathcal{R}^y(\theta) = y - R\lambda - X\beta = S(\lambda)y - X\beta$, and introduce the following matrices: $F_8(\varrho)$ of dimension $p \times q$ and j-th column $R'\Xi(\tau)^{-1}\Xi_j(\tau)\Xi(\tau)^{-1}\mathcal{R}^y(\theta)$, $F_9(\varrho)$ of dimension $k \times q$ with j-th column $X'\Xi(\tau)^{-1}\Xi_j(\tau)\Xi(\tau)^{-1}\mathcal{R}^y(\theta)$, and finally $F_{10}(\varrho)$ and $P_{\Xi 2}(\tau)$ both of dimension $q \times q$ with (i, j)-th elements $\mathcal{R}'^y(\theta)\Xi(\tau)^{-1}\left\{2\Xi_i(\tau)\Xi(\tau)^{-1}\Xi_j(\tau) - \Xi_{ij}(\tau)\right\}\Xi(\tau)^{-1}\mathcal{R}^y(\theta)$ and $tr\left(\Xi(\tau)^{-1}\left\{\Xi_{ij}(\tau) - \Xi_i(\tau)\Xi(\tau)^{-1}\Xi_j(\tau)\right\}\right)$, respectively. Then

$$H(\varrho, \sigma^2) = 2\sigma^{-2}n^{-1}\begin{bmatrix} \sigma^2 P_{G1}(\lambda) + R'\Xi(\tau)^{-1}R & R'\Xi(\tau)^{-1}X & F_8(\varrho) \\ * & X'\Xi(\tau)^{-1}X & F_9(\varrho) \\ * & * & 2^{-1}\left(F_{10}(\varrho) + \sigma^2 P_{\Xi 2}(\tau)\right) \end{bmatrix}, \tag{A.56}$$

so that

$$H = 2\sigma_0^{-2}n^{-1}\begin{bmatrix} \sigma_0^2 P_{G1} + R'\Xi^{-1}R & R'\Xi^{-1}X & F_8 \\ * & X'\Xi^{-1}X & F_9 \\ * & * & 2^{-1}\left(F_{10} + \sigma^2 P_{\Xi 2}\right) \end{bmatrix}, \tag{A.57}$$

whence, noting that $\mathcal{R}^y = u$, the expression in (4.11) follows easily. By the mean value theorem (MVT), we have $0 = d + H\left(\bar{\varrho}, \check{\sigma}^2\right)\left(\check{\varrho} - \varrho_0\right)$, where $H\left(\bar{\varrho}, \check{\sigma}^2\right)$ is obtained from $H\left(\varrho, \sigma^2\right)$ by evaluating each row at possibly different $\bar{\varrho}$ satisfying $\|\bar{\varrho} - \varrho_0\| \leq \left\|\check{\varrho} - \varrho_0\right\|$. Writing $0 = Nd + NH\left(\bar{\varrho}, \check{\sigma}^2\right)NN^{-1}\left(\check{\varrho} - \varrho_0\right)$, we obtain $n^{\frac{1}{2}}N^{-1}\left(\check{\varrho} - \varrho_0\right) = -\left(NH\left(\bar{\varrho}, \check{\sigma}^2\right)N\right)^{-1}n^{\frac{1}{2}}Nd$, indicating that the theorem is proved if we also show that

$$N\left(H\left(\bar{\varrho}, \check{\sigma}^2\right) - H\right)N \xrightarrow{p} 0, N\left(H - \Sigma\right)N \xrightarrow{p} 0, \text{ as } n \to \infty \tag{A.58}$$

because we have shown (A.37). The first part above uses $\bar{\varrho} - \varrho_0 \xrightarrow{p} 0$ and $\check{\sigma}^2 - \sigma_0^2 \xrightarrow{p} 0$ together with the regularity conditions imposed on $\Xi(\tau)$ and $\Xi(\tau)^{-1}$, and the second

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

follows using law of large numbers arguments. The calculations are handled in the same fashion as in the existing SAR literature, cf. Lee (2004), Delgado and Robinson (2015) and Gupta and Robinson (2018), so we omit the details except for the proof of

$$n^{-1}\left(P_{G1}\left(\bar{\lambda}\right)-P_{G1}\right)\xrightarrow{p}0, \tag{A.59}$$

which is the most relevant and novel in this article given our focus on stochastic W_j. The proof of (A.59) commences as in Lee (2004) and Gupta and Robinson (2018), by using the MVT to write the (i,j)-th element of the left side as $n^{-1}tr\left(G_i\left(\bar{\lambda}\right)G_j\left(\bar{\lambda}\right)\right)-n^{-1}tr\left(G_iG_j\right)=n^{-1}\bar{\bar{g}}'\left(\bar{\lambda}-\lambda_0\right)$, where $\bar{\bar{g}}$ is a $p\times 1$ vector with k-th element $tr\left(G_i\left(\bar{\bar{\lambda}}\right)G_k\left(\bar{\bar{\lambda}}\right)G_j\left(\bar{\bar{\lambda}}\right)+G_k\left(\bar{\bar{\lambda}}\right)G_i\left(\bar{\bar{\lambda}}\right)G_j\left(\bar{\bar{\lambda}}\right)\right)$ and $\bar{\bar{\lambda}}$ satisfies $\left\|\bar{\bar{\lambda}}-\lambda_0\right\|\leq\left\|\bar{\lambda}-\lambda_0\right\|\leq\left\|\check{\lambda}-\lambda_0\right\|$. Because $\left\|\bar{\lambda}-\lambda_0\right\|\xrightarrow{p}0$, (A.59) is proved if $n^{-1}\bar{\bar{g}}=\mathscr{O}_{\mathrm{p}}(1)$. Since Lemma B.6 implies that each element of $\bar{\bar{g}}$ is $\mathscr{O}_{\mathrm{p}}\left(n/h_W\right)$, the latter claim is true. The proof of the theorem is now completed. ■

APPENDIX B: Lemmas

LEMMA B.1. *Under the conditions of Theorem 3.1, the expectation of an absolute typical element of $G_j'G_j$ is $\mathscr{O}\left(\pi_n^2\right)$, uniformly in j.*

Proof. For $r,s=1,\ldots,n$, a typical absolute element of $G_j'G_j$ is $\left|g_{r,j}'G_je_s\right|=\left|e_s'G_j'g_{r,j}\right|$, where $g_{r,j}'$ is the r-th row of G_j'. Using Hölder's inequality as before, this has expectation bounded by

$$\left(\mathbb{E}\left\|g_{r,j}\right\|_R^{\zeta_5}\right)^{\frac{1}{\zeta_5}}\left(\mathbb{E}\left\|G_j'\right\|_R^{\zeta_6}\right)^{\frac{1}{\zeta_6}}\leq\left(\mathbb{E}\left\|g_{r,j}\right\|_R^{\zeta_5}\right)^{\frac{1}{\zeta_5}}\left(\mathbb{E}\left(\left\|W_j'\right\|_R^{\zeta_6}\left\|S'^{-1}\right\|_R^{\zeta_6}\right)\right)^{\frac{1}{\zeta_6}}. \tag{B.1}$$

Consider the first factor on the RHS of (B.1). $g_{r,j}$ has elements $w_{s,j}'S^{-1}e_r=e_r'S'^{-1}w_{s,j}$, where $w_{s,j}'$ is the s-th row of W_j, $s=1,\ldots,n$, so this factor is

$$\begin{aligned}\left(\mathbb{E}\left(\max_{1\leq s\leq n}\left|w_{s,j}'S^{-1}e_r\right|^{\zeta_5}\right)\right)^{\frac{1}{\zeta_5}}&\leq\left(\mathbb{E}\left(\max_{1\leq s\leq n}\left\|w_{s,j}\right\|_R^{\zeta_5}\left\|S'^{-1}\right\|_R^{\zeta_5}\right)\right)^{\frac{1}{\zeta_5}}\\&\leq\left(\mathbb{E}\left(\max_{1\leq s\leq n}\left\|w_{s,j}\right\|_R^{\zeta_5\zeta_7}\right)\right)^{\frac{1}{\zeta_5\zeta_7}}\left(\mathbb{E}\left(\left\|S'^{-1}\right\|_R^{\zeta_5\zeta_8}\right)\right)^{\frac{1}{\zeta_5\zeta_8}}\\&=\left(\mathbb{E}\left(\max_{1\leq r,s\leq n}\left|w_{rs,j}\right|^{\zeta_5\zeta_7}\right)\right)^{\frac{1}{\zeta_5\zeta_7}}\left(\mathbb{E}\left(\left\|S'^{-1}\right\|_R^{\zeta_5\zeta_8}\right)\right)^{\frac{1}{\zeta_5\zeta_8}}\\&=\mathscr{O}\left(h_W^{-1}\left(\mathbb{E}\left(\left\|S'^{-1}\right\|_R^{\zeta_5\zeta_8}\right)\right)^{\frac{1}{\zeta_5\zeta_8}}\right),\end{aligned} \tag{B.2}$$

by the Hölder inequality and Assumption 7. The second factor on the RHS of (B.1) is bounded by

$$\left(\max_{1\leq j\leq p}\mathbb{E}\left\|W_j'\right\|_R^{\zeta_6\zeta_9}\right)^{\frac{1}{\zeta_6\zeta_9}}\left(\mathbb{E}\left\|S'^{-1}\right\|_R^{\zeta_6\zeta_{10}}\right)^{\frac{1}{\zeta_6\zeta_{10}}}, \tag{B.3}$$

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

by another application of Hölder's inequality, whence the claim follows from (B.1), (B.2), (B.3) and the definition of π_n. ■

LEMMA B.2. *Under the conditions of Theorem 3.2, the expectation of the absolute product of two typical elements of G_j is $\mathcal{O}\left(\left(\mathbb{E}\left\|S'^{-1}\right\|_R^{2\zeta_{12}}\right)^{\frac{1}{\zeta_{12}}} h_W^{-2}\right)$, uniformly in j.*

Proof. For $p,q,r,s = 1,\ldots,n$, the expectation of the absolute product of typical elements of G_j is

$$\mathbb{E}\left|w'_{r,j}S^{-1}e_s w'_{p,j}S^{-1}e_q\right| \leq \mathbb{E}\left(\max_{1\leq r\leq n}\|w_{r,j}\|_R^2\left\|S'^{-1}\right\|_R^2\right), \tag{B.4}$$

which is bounded by $\left(\mathbb{E}\left(\max_{1\leq r\leq n}\|w_{r,j}\|_R^{2\zeta_{11}}\right)\right)^{\frac{1}{\zeta_{11}}}\left(\mathbb{E}\left\|S'^{-1}\right\|_R^{2\zeta_{12}}\right)^{\frac{1}{\zeta_{12}}}$, whence the result follows by Assumption 7 because $\|w_{r,j}\|_R = \max_{1\leq s\leq n}|w_{rs,j}|$. ■

LEMMA B.3. *Under the conditions of Theorem 3.1, $n^{-1}[B,0]'u = \mathcal{O}_p(\pi_n)$.*

Proof. First note that $\mathbb{E}\|u\|^2 = \sigma^2\sum_{r=1}^n\sum_{l=1}^\infty c_{rl}^2 = \mathcal{O}(n)$, by (2.4), so $\left\|n^{-1}[B,0]'u\right\| = \mathcal{O}_{\mathrm{p}}\left(n^{-\frac{1}{2}}\|B\|\right)$ by Markov's inequality. Next $\mathbb{E}\|B\|^2 \leq \mathbb{E}\left(tr\,B'B\right) = \mathcal{O}\left(\max_{1\leq j\leq p}\mathbb{E}\left(u'G'_jG_ju\right)\right)$, the RHS being

$$\mathcal{O}_{\mathrm{p}}\left(\pi_n^2\sum_{r,s=1}^n\sum_{j,l=1}^\infty c_{rj}c_{sl}\mathbb{E}\left(\epsilon_j\epsilon_l\right)\right) = \mathcal{O}_{\mathrm{p}}\left(\pi_n^2\sum_{r,s=1}^n\sum_{l=1}^\infty\left(c_{rl}^2+c_{sl}^2\right)\right) = \mathcal{O}_{\mathrm{p}}\left(\pi_n^2\sum_{r=1}^n\sum_{l=1}^\infty c_{rl}^2\right) = \mathcal{O}_{\mathrm{p}}\left(n\pi_n^2\right),$$

by Lemma B.1, the inequality $|ab| \leq \left(a^2+b^2\right)/2$ for real numbers a,b and Assumption 1. The claim follows by Markov's inequality, ■

LEMMA B.4. *Under the conditions of Theorem 3.2, $n^{-1}[B,0]'u = \mathcal{O}_p\left(\left(\mathbb{E}\left\|S'^{-1}\right\|_R^{2\zeta_{12}}\right)^{\frac{1}{2\zeta_{12}}} h_W^{-1}\right)$.*

Proof. Write $g_{rs,j}$ for a typical element of G_j, $r,s = 1,\ldots,n$. It is sufficient to evaluate

$$\mathbb{E}\left(n^{-1}u'G_ju\right)^2 = n^{-2}\sum_{r,s,t,v=1}^n\mathbb{E}\left(u_ru_su_tu_v\right)\mathbb{E}\left(g_{rs,j}g_{tv,j}\right), \tag{B.5}$$

with $j = 1,\ldots,p$, and then use Markov's inequality. By Assumption 1 and Lemma B.2 the RHS of (B.5) is

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

$$n^{-2}\sum_{r,s,t,v=1}^{n}\sum_{j,k,l,m=1}^{\infty}c_{rj}c_{sk}c_{tl}c_{vm}\mathbb{E}\left(\epsilon_j\epsilon_k\epsilon_l\epsilon_m\right)\mathbb{E}\left(g_{rs,j}g_{tv,j}\right)$$

$$=\mathcal{O}\left(n^{-2}\left(\mathbb{E}\left\|S'^{-1}\right\|_R^{2\zeta_{12}}\right)^{\frac{1}{\zeta_{12}}}h_W^{-2}\left[\sum_{r,s,t,v=1}^{n}\sum_{j=1}^{\infty}\mathbb{E}\left(\epsilon_j^4\right)c_{rj}c_{sj}c_{tj}c_{vj}+\right.\right.$$

$$\left.\left.+\sum_{r,s,t,v=1}^{n}\sum_{j,k=1}^{\infty}\left(c_{rj}c_{sj}c_{tk}c_{vk}+c_{rj}c_{sk}c_{tj}c_{vk}+c_{rj}c_{sk}c_{tk}c_{vj}\right)\right]\right). \tag{B.6}$$

By Assumption 9 and the ℓ_p norm inequality, the first sum inside square brackets in (B.6) is bounded in absolute value by a constant times

$$\sum_{r,s,t,v=1}^{n}\sum_{j=1}^{\infty}\left|c_{rj}c_{sj}c_{tj}c_{vj}\right|\leq C\sum_{r,s,t,v=1}^{n}\sum_{j=1}^{\infty}\left(c_{rj}^2c_{sj}^2+c_{tj}^2c_{vj}^2\right)$$

$$\leq C\sum_{r,s,t,v=1}^{n}\sum_{j=1}^{\infty}\left(c_{rj}^4+c_{sj}^4+c_{tj}^4+c_{vj}^4\right)$$

$$\leq C\sum_{r=1}^{n}\sum_{j=1}^{\infty}c_{rj}^4\leq C\sum_{r=1}^{n}\left(\sum_{j=1}^{\infty}c_{rj}^2\right)^2\leq Cn.$$

Now consider the first product inside parentheses in the second sum inside square brackets in (B.6). By similar techniques this is bounded in absolute value by

$$\sum_{r,s,t,v=1}^{n}\sum_{j,k=1}^{\infty}\left|c_{rj}c_{sj}\right|\left|c_{tk}c_{vk}\right|=\left(\sum_{r,s=1}^{n}\sum_{j=1}^{\infty}\left|c_{rj}c_{sj}\right|\right)^2\leq C\left(\sum_{r,s=1}^{n}\sum_{j=1}^{\infty}\left(c_{rj}^2+c_{sj}^2\right)\right)^2$$

$$\leq C\left(\sum_{r=1}^{n}\left(\sum_{j=1}^{\infty}c_{rj}^2\right)\right)^2\leq Cn^2.$$

The remaining two products inside parentheses in the second sum inside square brackets in (B.6) are similarly shown to be $\mathcal{O}\left(n^2\right)$. We have established that the term inside square brackets in (B.6) is $\mathcal{O}\left(n^2\right)$, whence the claim follows. ■

LEMMA B.5. *Under the conditions of Theorem 4.2,* $\left\|S(\lambda)^{-1}\right\|_R=\mathcal{O}_p(1)$ *and* $\left\|S(\lambda)'^{-1}\right\|_R=\mathcal{O}_p(1)$ *in a closed neighbourhood of* λ_0*, denoted* $B(\lambda_0)$.

Proof. We prove the claim for $\left\|S(\lambda)^{-1}\right\|_R$ only, the proof for the transpose being similar. We first show that, for given $0<\varepsilon<1$,

$$\lim_{n\to\infty}P\left(\left\|\sum_{j=1}^{p}\left(\lambda_j-\lambda_{0j}\right)G_j\right\|_R<1-\varepsilon\right)=1. \tag{B.7}$$

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

Because $\|G_i\|_R = \mathcal{O}_{\mathrm{p}}(1)$ (Assumption 21) uniformly in i, $\left(\sum_{j=1}^{p} \|G_j\|_R^2\right)^{\frac{1}{2}} = \mathcal{O}_{\mathrm{p}}(1)$ also, implying that there exists $0 < C_1 < \infty$ such that

$$\lim_{n\to\infty} P\left(\left(\sum_{j=1}^{p} \|G_j\|_R^2\right)^{\frac{1}{2}} > C_1\right) = 0. \tag{B.8}$$

The probability on the left side of (B.7) is bounded below by

$$P\left(\left(\sum_{j=1}^{p} (\lambda_j - \lambda_{0j})^2\right)^{\frac{1}{2}} \left(\sum_{j=1}^{p} \|G_j\|_R^2\right)^{\frac{1}{2}} < 1-\varepsilon\right) = P\left(\|\lambda - \lambda_0\| \left(\sum_{j=1}^{p} \|G_j\|_R^2\right)^{\frac{1}{2}} < 1-\varepsilon\right),$$

which in turn is bounded below by

$$1 - P\left(\|\lambda - \lambda_0\| \geq C_1^{-1}(1-\varepsilon)\right) - P\left(\left(\sum_{j=1}^{p} \|G_j\|_R^2\right)^{\frac{1}{2}} \geq C_1\right)$$

$$= P\left(\|\lambda - \lambda_0\| < C_1^{-1}(1-\varepsilon)\right) - P\left(\left(\sum_{j=1}^{p} \|G_j\|_R^2\right)^{\frac{1}{2}} \geq C_1\right), \tag{B.9}$$

where we used the fact that $|cd| > ef$ implies $|c| > e$ or $|d| > f$. Choosing $B(\lambda_0)$ to be a closed subset of a neighbourhood of λ_0 such that $\|\lambda - \lambda_0\| < C_1^{-1}(1-\varepsilon)$ implies that the first probability in (B.9) equals 1, while the second probability in (B.9) converges to zero by (B.8). Thus (B.7) is established.

Because $S(\lambda)^{-1} = S^{-1}\left(I_n - \sum_{j=1}^{p} (\lambda_j - \lambda_{0j}) G_j\right)^{-1}$, we have

$$\left\|S(\lambda)^{-1}\right\|_R \leq \left\|S^{-1}\right\|_R \left\|\left(I_n - \sum_{j=1}^{p} (\lambda_j - \lambda_{0j}) G_j\right)^{-1}\right\|_R, \tag{B.10}$$

with the first factor on the RHS of (B.10) $\mathcal{O}_{\mathrm{p}}(1)$ by Assumption 21. By (B.7), the second factor equals $\sum_{k=0}^{\infty} \left\|\sum_{j=1}^{p} (\lambda_j - \lambda_{0j}) G_j\right\|_R^k \leq \sum_{k=0}^{\infty} (1-\varepsilon)^k = \varepsilon^{-1}$ with probability one on $B(\lambda_0)$. Thus the lemma is proved. ■

LEMMA B.6. *Under the conditions of Theorem 4.2, $G_i(\lambda)G_j(\lambda)$ and $G_i(\lambda)G_j(\lambda)G_k(\lambda)$ both have elements that are $\mathcal{O}_p(h_W^{-1})$ uniformly in $i, j, k = 1, \ldots, p$ and $\lambda \in B(\lambda_0)$.*

Proof. We prove the lemma for the triple product, the double product being similarly handled. Take $\lambda \in B(\lambda_0)$. By Assumption 21 and Lemma B.5, $\|G_i(\lambda)\|_R$ and $\|G_i'(\lambda)\|_R$ are uniformly $\mathcal{O}_{\mathrm{p}}(1)$ in $B(\lambda_0)$. Let $g_{r,i}'(\lambda)$ and $w_{r,i}'$ be the r-th rows of $G_i(\lambda)$ and W_i, respectively, and the (r,s)-th element of G_i be written as $g_{rs,i}$. Then

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press

$g_{rs,i} = w'_{r,i} S(\lambda)^{-1} e_s = \mathcal{O}_{\mathrm{p}}\left(\|w_{r,i}\|_R\right) = \mathcal{O}_{\mathrm{p}}\left(h_W^{-1}\right)$, uniformly in $r, s = 1, \ldots, n$, by Lemma B.5 and Assumption 2. Thus the (r, s)-th element of $G_i(\lambda)G_j(\lambda)G_k(\lambda)$, viz. $g'_{r,i}(\lambda)G_j(\lambda)G_k(\lambda)e_s$, is $\mathcal{O}_{\mathrm{p}}\left(\|g_{r,i}\|_R\right) = \mathcal{O}_{\mathrm{p}}\left(\max_{s=1,\ldots,n} |g_{rs,i}|\right) = \mathcal{O}_{\mathrm{p}}(h_W^{-1})$, uniformly in $r = 1, \ldots, n$ and $i, j, k = 1, \ldots, p$, as claimed. ■

https://doi.org/10.1017/S0266466618000142 Published online by Cambridge University Press