Adjusted QMLE for the Spatial Autoregressive Parameter

Federico Martellosio*

University of Surrey, UK

`f.martellosio@surrey.ac.uk`

Grant Hillier

CeMMAP and

University of Southampton, UK

`ghh@soton.ac.uk`

February 27, 2019

Abstract

One simple, and often very effective, way to attenuate the impact of nuisance parameters on maximum likelihood estimation of a parameter of interest is to recenter the profile score for that parameter. We apply this general principle to the quasi-maximum likelihood estimator (QMLE) of the autoregressive parameter λ in a spatial autoregression. The resulting estimator for λ has better finite sample properties compared to the QMLE for λ, especially in the presence of a large number of covariates. It can also solve the incidental parameter problem that arises, for example, in social interaction models with network fixed effects, or in spatial panel models with individual or time fixed effects. However, spatial autoregressions present specific challenges for this type of adjustment, because recentering the profile score may cause the adjusted estimate to be outside the usual parameter space for λ. Conditions for this to happen are given, and implications are discussed. For inference, we propose confidence intervals based on a Lugannani–Rice approximation to the distribution of the adjusted QMLE of λ. Based on our simulations, the coverage properties of these intervals are excellent even in models with a large number of covariates.

Keywords: adjusted maximum likelihood estimation, fixed effects, group interaction, networks, spatial autoregression. *JEL Classification*: C12, C21.

1 Introduction

Among the several difficulties posed by nuisance parameters, a fundamental problem in a frequentist framework is that the profile likelihood function for a parameter of interest is typically not a genuine likelihood, in the sense that it does not correspond to the density of

---

*Corresponding author. School of Economics, University of Surrey, Guildford, Surrey, GU2 7XH, UK. Tel: +44 (0) 1483 683473

any observable random variable. To tackle this issue, a number of *modified profile likelihoods* have been proposed (see, e.g., Laskar and King, 2001; Pace and Salvan, 2006). Such modified profile likelihoods are genuine likelihoods only in special cases, but they can often be interpreted as approximations to a genuine likelihood. In practice, they tend to perform better than the profile likelihood, particularly when there is little information in the data about the nuisance parameters, which is likely to happen, for example, if the number of nuisance parameters is large relative to the sample size. The better performance of modified profile likelihoods is not necessarily captured by first-order asymptotic theory. Indeed, if the number of nuisance parameters does not depend on the sample size, modified profile likelihoods usually produce estimators that are first-order asymptotically equivalent to the maximum likelihood estimator (MLE). On the other hand, if the number of nuisance parameters increases with the sample size, a modified profile likelihood may be preferable even in terms of first-order asymptotic properties (see, e.g., Jiang, 1996; Sartori, 2003; Arellano and Hahn, 2007).

One consequence of a profile likelihood not being a genuine likelihood is that the expectation of the profile score is generally nonzero, which means that setting the profile score to zero does not provide an unbiased estimating equation. A simple modified profile likelihood is therefore obtained by centering the profile score. We refer to the modified profile likelihood obtained in this way as the *adjusted profile likelihood*, and to the associated MLE as the *adjusted MLE*. The principle underlying this type of adjustment has been motivated from various perspectives (e.g., Neyman and Scott, 1948; Conniffe, 1987; McCullagh and Tibshirani, 1990), and has been applied to several statistical models (e.g., Macaskill, 1993; Dhaene and Jochmans, 2015).

The present paper is concerned with the adjusted MLE of the autoregressive parameter λ in a spatial autoregression. Reliable estimation of λ is important in many applications in economics, as well as in other fields. For example, in social interaction models λ captures the endogenous effect, the assessment of which may be crucial for policy purposes (see Manski, 1993; Moffitt, 2001; Lee et al., 2010). More precisely, we will consider the quasi-MLE (QMLE), that is, the MLE obtained by maximizing the Gaussian likelihood, but without assuming that the error distribution is truly Gaussian. The literature on estimating (cross-sectional or panel) spatial autoregressions is now very large, the QMLE being perhaps the most popular estimator. An early reference for the QMLE in spatial autoregressions is Ord (1975), whereas a rigorous first-order asymptotic analysis of the QMLE was given only much later, in an influential paper by Lee (2004a), under conditions that have become standard in the literature. Recently, higher-order approximations to the distribution of the QMLE of λ have become available (Robinson and Rossi, 2015), and, motivated by the fact that the QMLE of λ can suffer from substantial bias, a number of studies have suggested bias reduction procedures (e.g., Bao and Ullah, 2007; Bao, 2013; Yang, 2015). In fact, part of the bias in the QMLE of a parameter of interest can be attributed to the bias in the profile

likelihood estimating equation, so centering the profile score can itself be interpreted as a bias reduction technique. Lee and Yu (2010), while discussing bias reduction in spatial panel data models, state explicitly that correction methods based on modifying the score "might be possible and would be of interest in future research" (see their footnote 24). Yu et al. (2015) show that, in the absence of incidental parameters, the adjusted QMLE of λ is first-order asymptotically equivalent to the QMLE of λ. Yu et al. (2015) also derive the second-order bias of the adjusted QMLE, and compare by simulation the adjusted QMLE with other bias reduction techniques. Prior to that article, Durban and Currie (2000) had provided a preliminary investigation of the adjusted QMLE in a class of models that include spatial autoregressions. Notwithstanding the last two papers, the general distributional properties of the adjusted QMLE of λ remain unclear. Indeed, Yu et al. (2015) conclude their analysis by suggesting that the adjusted QMLE of λ "deserves a deep study in future".

Our main contributions are as follows. *First*, on studying the properties of the adjusted profile likelihood, we find that the distributions of the QMLE of λ and its adjusted version may be supported on different intervals. This is due to the fact that, in spatial autoregressions, the parameter λ is usually restricted to a certain interval containing the origin. Such a restriction is incorporated in the QMLE (which, strictly speaking, should therefore be referred to as a restricted QMLE), but may be violated once the profile score is recentered. We discuss the implications of the supports being different, and give conditions for this to happen. *Second*, for inference about λ, we propose confidence intervals based on the Lugannani and Rice (1980) saddlepoint approximation to the cdf of the adjusted QMLE. Contrary to the commonly employed Wald confidence intervals, these confidence intervals have excellent coverage properties even when the dimension of the nuisance parameter is large. *Third*, we consider the social interaction model of Lee et al. (2010), and show that the adjusted QMLE solves the incidental parameter problem due to the network fixed effects without requiring the condition on W (row normalization) that is needed for the estimator in Lee et al. (2010). *Fourth*, we compare the QMLE and the adjusted QMLE by Monte Carlo simulation, and find evidence that the latter is preferable in a variety of circumstances.

The rest of the paper is structured as follows. Section 2 introduces the spatial autoregressive (SAR) model and the QMLE of λ. Section 3 discusses the properties of the adjusted profile likelihood for λ, and introduces the confidence intervals based on the adjusted QMLE. Section 4 briefly considers the spatial error model, which is less popular than the SAR model in economic applications, but provides important motivation for the adjusted ML procedure. Section 5 contains simulation evidence on the performance of the adjusted QMLE and associated confidence intervals. Section 6 discusses extensions. Appendix A contains some auxiliary results needed for the proofs, which can be found in Appendix B. Various technical materials related to the paper can be found in the online Supplement.

Throughout the paper, all vectors and matrices are real valued unless otherwise indicated. The null space of a matrix A is denoted by null(A), the column space by col(A), and the

orthogonal complement of $\text{col}(A)$ by $\text{col}^\perp(A)$. Also, M_A denotes the orthogonal projector onto $\text{col}^\perp(A)$ $(M_A := I_n - A(A'A)^{-1}A'$ if A has full column rank). Finally, $\mu_{\mathbb{R}^n}$ denotes the Lebesgue measure on $\mathbb{R}^n$, and "a.s." stands for almost surely, with respect to $\mu_{\mathbb{R}^n}$.

2 Preliminaries

2.1 The SAR model

We consider the spatial autoregressive (SAR) model

$$y = \lambda W y + X\beta + \sigma\varepsilon, \tag{2.1}$$

where y is the $n \times 1$ vector of observed random variables, λ is a scalar parameter, W is a spatial weights matrix, X is an $n \times k$ matrix of regressors with full column rank and with $k \leq n - 2$, $\beta \in \mathbb{R}^k$, σ is a positive scale parameter, and ε is a zero mean $n \times 1$ random vector. For simplicity, we take X and W to be non-stochastic, and we assume that W is completely known. Alternatively, we could allow X and W to be random, and interpret the analysis as conditional on them, provided that they are independent of ε. Some of the columns of X may be spatial lags of some other columns, to allow for the estimation of, in the terminology of social network analysis, contextual effects. When there is no X, the model is called a pure SAR model. We allow for equation (2.1) to also represent a spatial panel data model, or a model in which individuals are divided in several networks. In both cases, W is a block diagonal matrix, with the number of blocks being given by, respectively, the number of time points and the number of networks. Additive fixed effects along one or both of the panel dimensions, or network fixed effects, can be added. For the purpose of estimation, fixed effects are treated as parameters, and hence can be included in β. Throughout the paper we assume that W has at least one (real) negative eigenvalue and at least one (real) positive eigenvalue. This assumption is virtually always satisfied in applications, especially because the diagonal entries of W are usually set to zero. The smallest real eigenvalue of W is denoted by $\omega_{\min}$, and the largest real eigenvalue is normalized, without loss of generality, to 1.

On rewriting equation (2.1) as $S(\lambda)y = X\beta + \sigma\varepsilon$, where $S(\lambda) := I_n - \lambda W$, it is clear that in order for y to be uniquely determined, given X and ε, it is necessary that $S(\lambda)$ is nonsingular. This requires $\lambda \neq \omega^{-1}$, for any nonzero real eigenvalue ω of W (nonreal complex eigenvalues of W do not need to be considered here, because λ is assumed to be real, and ω^{-1} is real if and only ω is). In both applications and theoretical studies, the parameter space for λ is usually restricted much further, namely to the largest interval containing the origin in which $S(\lambda)$ is nonsingular, that is,

$$\Lambda := (\omega_{\min}^{-1}, 1),$$

or a subset thereof (possibly independent of n) such as $(-1, 1)$. Without such restrictions the models are believed to be too erratic to be useful in practice, and λ is difficult to interpret.[1]

The following assumption is required to rule out some pathological combinations of W and X.

Assumption 1. There is no real eigenvalue ω of W for which $M_X(\omega I_n - W) = 0$.

To provide some intuition for Assumption 1, we note that the condition $M_X(\omega I_n - W) = 0$ is equivalent to $\operatorname{col}(\omega I_n - W) \subseteq \operatorname{col}(X)$, and we distinguish two cases. First, if Assumption 1 is violated for the eigenvalue $\omega = 0$ (i.e., $\operatorname{col}(W) \subseteq \operatorname{col}(X)$), it is evident from equation (2.1) that estimating β and λ separately must be problematic. Second, if Assumption 1 is violated for some eigenvalue $\omega \neq 0$, then for any $y \in \mathbb{R}^n$ it is possible to find a $\beta \in \mathbb{R}^k$ such that $S(\omega^{-1})y = X\beta$, meaning that a SAR model with $\lambda = \omega^{-1}$ can provide perfect fit for any y. It is clear that in this case any sensible inferential procedure should suggest that $\lambda = \omega^{-1}$ and $\sigma = 0$, for any y, and whatever the true values of λ and σ are.[2]

The following example provides a simple illustration of Assumption 1.

Example 2.1 (Group interaction). There are $R \geq 1$ groups of individuals. Individuals interact uniformly within their group, and do not interact across different groups. If each group has the same size, say $m > 1$, this type of interaction can be represented by the block-diagonal weights matrix

$$W = I_R \otimes B_m, \tag{2.2}$$

where $B_m := \frac{1}{m-1}(\iota_m \iota_m' - I_m)$, with ι_m an $m \times 1$ vector of all ones. See Lee (2007b) and Hillier and Martellosio (2018b) for theoretical studies of this model, and Carrell et al. (2013) and Boucher et al. (2014) for recent applications. For matrix (2.2), one can easily verify that $\omega_{\min} = -\frac{1}{m-1}$ and $\operatorname{col}(\omega_{\min}I_n - W) = \operatorname{col}(I_R \otimes \iota_m)$. Noting that $I_R \otimes \iota_m$ is the design matrix of the group fixed effects, it follows that, in a SAR model with weights matrix (2.2), Assumption 1 is violated (for $\omega = \omega_{\min}$) whenever group fixed effects are included in the model (recall that the fixed effects are treated as parameters to be estimated and hence are included in β). It should be noted that the presence of group intercepts does not cause a violation of Assumption 1 when the model is unbalanced, that is, not all groups have the same size.

Indeed, it is well known that the model of Example suffers from an identifiability problem. Specifically, Lee (2007b) and Bramoullé et al. (2009) show that the parameters of a SAR model with weights matrix (2.2), group fixed effects, and contextual effects are not identifiable after removal of the group fixed effects by a within transformation. It is easily verified that, in this model, the identifiability problem occurs even without contextual effects.

---

[1] From a large sample perspective, the restriction to $(-1, 1)$ is often imposed, along with a uniform boundedness condition on W, to guarantee that the variances of the y_i's do not explode as n grows. Also, note that Λ and the entries of W are allowed to depend on n, although this is not emphasized in our notation.

[2] For the specific case of the QMLE, see part (i) of Lemma S.1.2 in Section S.1 of the Supplement. Section S.1 of the Supplement also contains further technical remarks about Assumption 1.

2.2 The QMLE

We now define the QMLE of the parameters in model (2.1). By quasi-likelihood we mean the likelihood that would prevail under the condition $\varepsilon \sim N(0, I_n)$. Omitting additive constants, the quasi-log-likelihood is

$$l(\beta, \sigma^2, \lambda) := -\frac{n}{2} \log(\sigma^2) + \log|\det(S(\lambda))| - \frac{1}{2\sigma^2}(S(\lambda)y - X\beta)'(S(\lambda)y - X\beta), \qquad (2.3)$$

for any λ such that $S(\lambda)$ is nonsingular. To avoid tedious repetitions, in the remainder of the paper we will often omit the "quasi-" in front of "log-likelihood".

The QMLE in most common use is the maximizer of $l(\beta, \sigma^2, \lambda)$ under the condition that λ is in Λ (or in a subset thereof). That is, the QMLE is

$$(\hat{\beta}_{\mathrm{ML}}, \hat{\sigma}^2_{\mathrm{ML}}, \hat{\lambda}_{\mathrm{ML}}) = \operatorname*{argmax}_{\beta \in \mathbb{R}^k,\, \sigma^2 > 0,\, \lambda \in \Lambda} l(\beta, \sigma^2, \lambda).$$

Maximization with respect to β and σ^2 gives $\hat{\beta}_{\mathrm{ML}}(\lambda) := (X'X)^{-1}X'S(\lambda)y$ and $\hat{\sigma}^2_{\mathrm{ML}}(\lambda) := \frac{1}{n}y'S'(\lambda)M_X S(\lambda)y$. The corresponding profile, or concentrated, log-likelihood for λ is, again omitting additive constants,

$$l(\lambda) := l(\hat{\beta}_{\mathrm{ML}}(\lambda), \hat{\sigma}^2_{\mathrm{ML}}(\lambda), \lambda) = -\frac{n}{2} \log(\hat{\sigma}^2_{\mathrm{ML}}(\lambda)) + \log|\det(S(\lambda))|. \qquad (2.4)$$

The QMLE of λ can be equivalently defined as

$$\hat{\lambda}_{\mathrm{ML}} := \operatorname*{argmax}_{\lambda \in \Lambda} l(\lambda). \qquad (2.5)$$

The function $l(\lambda)$ is a.s. well defined (see Section S.2 in the Supplement), and, clearly, is continuously differentiable whenever it is well defined. Also, it is easy to see that, under Assumption 1, $l(\lambda)$ a.s. goes to $-\infty$ at each real zero of $\det(S(\lambda))$ (cf. Hillier and Martellosio, 2018a). Thus, $l(\lambda)$ has a.s. at least one critical point corresponding to a maximum in any interval between two consecutive real zeros of $\det(S(\lambda))$. We also define the unrestricted QMLE $\hat{\lambda}_{\mathrm{uML}} := \operatorname{argmax}_{\lambda \in \Lambda_u} l(\lambda)$, where $\Lambda_u := \{\lambda \in \mathbb{R} : \det(S(\lambda)) \neq 0\}$. The estimators $\hat{\lambda}_{\mathrm{ML}}$ and $\hat{\lambda}_{\mathrm{uML}}$ are different, because the global maximum of $l(\lambda)$ over Λ_u is not necessarily in Λ, even when the true value of λ is in Λ.

3 The Adjustment

3.1 The adjusted profile likelihood

A profile likelihood for a parameter of interest does not take into account the sampling variability associated to the estimation of the nuisance parameters, and hence, as mentioned in the introduction, is generally not a genuine likelihood. As a consequence, a profile score

estimating equation is generally biased, which is likely to induce bias in the QMLE of the parameter of interest. The adjusted QMLE solves the estimating equation obtained by recentering the profile score. We now apply this general adjustment to the estimation of (σ^2, λ) in the SAR model. The score for (σ^2, λ) is centered assuming only that $\mathrm{E}(\varepsilon) = 0$ and $\mathrm{var}(\varepsilon) = I_n$. Note that we treat only β as the nuisance parameter, not (β, σ^2), because an adjusted estimator of σ^2 is required for estimation of λ.[3]

On concentrating just β out of the Gaussian log-likelihood (2.3), the profile log-likelihood for (σ^2, λ) is

$$l(\sigma^2, \lambda) := l(\hat{\beta}_{\mathrm{ML}}(\lambda), \sigma^2, \lambda) = -\frac{n}{2}\log(\sigma^2) + \log|\det(S(\lambda))| - \frac{1}{2\sigma^2}y'S'(\lambda)M_X S(\lambda)y, \quad (3.1)$$

with profile score

$$s(\sigma^2, \lambda) := \begin{bmatrix} -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}y'S'(\lambda)M_X S(\lambda)y \\ \frac{1}{\sigma^2}y'W'M_X S(\lambda)y - \mathrm{tr}(G(\lambda)) \end{bmatrix}, \quad (3.2)$$

where $G(\lambda) := WS^{-1}(\lambda)$. We now compute the expectation of $s(\sigma^2, \lambda)$ under the SAR model $y = \lambda W y + X\beta + \sigma\varepsilon$. Assuming that $\mathrm{E}(\varepsilon) = 0$ and $\mathrm{var}(\varepsilon) = I_n$, we have[4]

$$\mathrm{E}(s(\sigma^2, \lambda)) = \begin{bmatrix} -\frac{n}{2\sigma^2} + \frac{n-k}{2\sigma^2} \\ \mathrm{tr}(M_X G(\lambda)) - \mathrm{tr}(G(\lambda)) \end{bmatrix}. \quad (3.3)$$

For a pure model, $\mathrm{E}(s(\sigma^2, \lambda)) = 0$ for any λ such that $S(\lambda)$ is nonsingular, meaning that the estimating equation $s(\sigma^2, \lambda) = 0$ is unbiased.[5] When regressors are present, however, the unaccounted variability in the estimation of β causes the estimating equation $s(\sigma^2, \lambda) = 0$ to be biased. Note that the expectation (3.3) does not depend on β, so recentering the score is straightforward. The adjusted profile score for (σ^2, λ), defined as $s_{\mathrm{a}}(\sigma^2, \lambda) := s(\sigma^2, \lambda) - \mathrm{E}(s(\sigma^2, \lambda))$, is

$$s_{\mathrm{a}}(\sigma^2, \lambda) = \begin{bmatrix} s_{\mathrm{a}1}(\sigma^2, \lambda) \\ s_{\mathrm{a}2}(\sigma^2, \lambda) \end{bmatrix} = \begin{bmatrix} -\frac{n-k}{2\sigma^2} + \frac{1}{2\sigma^4}y'S'(\lambda)M_X S(\lambda)y \\ \frac{1}{\sigma^2}y'W'M_X S(\lambda)y - \mathrm{tr}(M_X G(\lambda)) \end{bmatrix}. \quad (3.4)$$

Setting $s_{\mathrm{a}1}(\sigma^2, \lambda) = 0$ gives $\hat{\sigma}^2_{\mathrm{aML}}(\lambda) := \frac{n}{n-k}\hat{\sigma}^2_{\mathrm{ML}}(\lambda)$. That is, recentering the score automatically delivers the usual degrees of freedom correction for $\hat{\sigma}^2_{\mathrm{ML}}(\lambda)$. The adjusted

---

[3]Treating σ^2 as a nuisance parameter too, and consequently recentering the score for λ only (i.e., the score associated to the log-likelihood (2.4)) would not produce an adjusted estimator for σ^2. Also, the (exact) recentering of $s(\lambda)$ would require stronger assumptions than what required for the recentering of $s(\sigma^2, \lambda)$; see Section S.7 of the Supplement.

[4]If the matrices X or W were stochastic, but independent of ε, we would be conditioning on them here.

[5]Depending on the sample size and on the structure of W, $\hat{\lambda}_{\mathrm{ML}}$ can be considerably biased even in the pure case, despite the fact that the estimating equation $s(\sigma^2, \lambda) = 0$ is unbiased in that case. The adjustment studied in this paper is not designed to reduce this type of bias in $\hat{\lambda}_{\mathrm{ML}}$. Instead, it aims at reducing the bias due to the nuisance parameter β.

QMLE for λ must then be a zero of

$$s_{a2}(\lambda) := s_{a2}(\hat{\sigma}^2_{aML}(\lambda), \lambda) = (n - k)\frac{y'W'M_X S(\lambda)y}{y'S'(\lambda)M_X S(\lambda)y} - \mathrm{tr}(M_X G(\lambda)), \qquad (3.5)$$

or, which is a.s. the same, must solve the estimating equation

$$y'S'(\lambda)R(\lambda)S(\lambda)y = 0, \qquad (3.6)$$

where

$$R(\lambda) := M_X\left(G(\lambda) - \frac{\mathrm{tr}(M_X G(\lambda))}{n - k}I_n\right).$$

We emphasize that, by construction, the estimating equation (3.6) is exactly unbiased provided that $\mathrm{E}(\varepsilon) = 0$ and $\mathrm{var}(\varepsilon) = I_n$ (no further distributional assumptions are required).

Given the adjusted profile score $s_a(\sigma^2, \lambda)$, one can define the function with gradient equal to $s_a(\sigma^2, \lambda)$, which we refer to as the adjusted likelihood for (σ^2, λ), denoted by $l_a(\sigma^2, \lambda)$. Letting $\mathrm{Re}[\cdot]$ denote the real part of a complex number, the following result gives a closed form expression for $l_a(\sigma^2, \lambda)$.

Proposition 3.1. *In a SAR model, the adjusted log-likelihood for (σ^2, λ), up to an additive constant, is*

$$l_a(\sigma^2, \lambda) = -\frac{n - k}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}y'S'(\lambda)M_X S(\lambda)y + \mathrm{Re}[\mathrm{tr}(M_X \log S(\lambda))], \qquad (3.7)$$

for any λ such that $S(\lambda)$ is nonsingular, and for a suitable choice (made in the proof) of the branch of the matrix logarithm $\log S(\lambda)$.

From expression (3.7), we immediately obtain the adjusted likelihood for λ only,

$$l_a(\lambda) := l_a(\hat{\sigma}^2_{aML}(\lambda), \lambda) = -\frac{n - k}{2}\log\big(y'S'(\lambda)M_X S(\lambda)y\big) + \mathrm{Re}[\mathrm{tr}(M_X \log S(\lambda))] \qquad (3.8)$$

(note that this is the likelihood with score $s_{a2}(\lambda)$). Equations (3.7) and (3.8) may be useful for graphical or optimization purposes, but are not used for the analytical results that follow in the paper. Indeed, even the results that (for simplicity and to aid intuition) are stated in terms of the adjusted likelihood, are proved using only the expression for the adjusted score, and could equally be formulated in terms of the adjusted score.

Remark 3.1. Like its unadjusted version $l(\sigma^2, \lambda)$, the adjusted profile log-likelihood $l_a(\sigma^2, \lambda)$ is not, in general, a genuine likelihood function. Further adjustments could be implemented to make $l_a(\sigma^2, \lambda)$ closer to a genuine likelihood. In particular, one could normalize $s_a(\sigma^2, \lambda)$ to make it information unbiased (i.e., variance equal to minus the expectation of the derivative of the score), as suggested by McCullagh and Tibshirani (1990). Such adjustments might improve the performance of first-order asymptotic approximations to estimators or

test statistics, but are not considered in this paper because they do not affect the location of the zeros of $s_a(\sigma^2, \lambda)$.

So far, the adjusted QMLE of λ has been introduced as a zero of $s_{a2}(\lambda)$, but of course $s_{a2}(\lambda)$ may have many (real) zeros. The question therefore arises as to how exactly the adjusted QMLE should be defined. Recall from Section 2.2 that $\hat{\lambda}_{ML}$ is defined as the maximizer of $l(\lambda)$ over Λ. One may therefore be tempted to define the adjusted QMLE of λ as the maximizer of $l_a(\lambda)$ over Λ. However, we shall see that $s_{a2}(\lambda)$ may have no zeros in Λ (or, equivalently, the adjusted profile likelihood $l_a(\lambda)$ may have no maximum on Λ), so it makes sense to define the adjusted QMLE on an interval larger than Λ, Λ_a say. By analogy with the unadjusted case, we shall define Λ_a to be the shortest open interval containing the origin with the property that $l_a(\lambda) \to -\infty$ a.s. at both extremes of Λ_a. But first, in order to fully understand the problem, we need to study the behavior of the profile score $s_{a2}(\lambda)$, or, equivalently, of $l_a(\lambda)$, near the zeros of $\det(S(\lambda))$.

3.2 Behavior of $l_a(\lambda)$ near a zero of $\det(S(\lambda))$

Perhaps unexpectedly, the profile score $s(\sigma^2, \lambda)$ and its adjusted version $s_a(\sigma^2, \lambda)$ may have very different behavior as λ approaches a real zero of $\det(S(\lambda))$. Obviously, different behavior of $s(\sigma^2, \lambda)$ and $s_a(\sigma^2, \lambda)$ implies different behavior of the functions $l(\sigma^2, \lambda)$ and $l_a(\sigma^2, \lambda)$, and hence of their profile versions $l(\lambda)$ and $l_a(\lambda)$. For simplicity, we state the results in this section in terms of the univariate functions $l(\lambda)$ and $l_a(\lambda)$. We shall see that the different behavior of $l(\lambda)$ and $l_a(\lambda)$ accounts for important differences in the properties of the associated estimators for λ.

To start with, note that the analysis is straightforward for the pure model. In that case, we trivially have $l_a(\lambda) = l(\lambda)$, and plugging $\hat{\sigma}^2_{ML}(\lambda) = \frac{1}{n}y'S'(\lambda)S(\lambda)y$ in equation (2.4) reveals that $l(\lambda)$ a.s. approaches $-\infty$ near any real zero of $\det(S(\lambda))$. The presence of regressors complicates the analysis. We shall confine attention to semisimple eigenvalues of W. An eigenvalue is said to be *semisimple* if its algebraic and geometric multiplicities are equal (for the matrix theoretic definitions and results used in this section see, for instance, Meyer, 2000). While it simplifies the analysis considerably, the restriction to semisimple eigenvalues does not imply a great loss of generality. For example, all eigenvalues of a diagonalizable matrix are semisimple, and any simple eigenvalue is semisimple (an eigenvalue being *simple* if it has algebraic multiplicity equal to one). The behavior of $l_a(\lambda)$ close to the eigenvalue $\omega = 1$ is often particularly important, and that eigenvalue is in most cases semisimple in applications. For example, it is semisimple when W is *row stochastic* (even if W is not diagonalizable, and the algebraic multiplicity of $\omega = 1$ is larger than one), or when W is *irreducible*.[6]

---

[6]A row stochastic matrix is a square nonnegative matrix whose row sums are all equal to 1. Irreducibility can be defined in terms of the graph of a matrix as follows. Let the graph of an $n \times n$ matrix A be the directed graph on n vertices in which there is an an edge from vertex i to vertex j if and only $A(i,j) \neq 0$. Also, call a graph *strongly connected* if there is a sequence of directed edges from any vertex i to any vertex j. Then, A

The simplification afforded by the restriction to semisimple eigenvalues is that we can express the conditions for $l_a(\lambda)$ to diverge or be bounded near a real zero of $\det(S(\lambda))$ in terms of a projector onto an eigenspace of W. Without the restriction to semisimple eigenvalues, the conditions would need to be stated in terms of projections onto generalized eigenspaces. If the eigenvalue ω of W is semisimple, then $\mathrm{null}(W - \omega I_n)$ (the eigenspace of W associated to the eigenvalue ω) and $\mathrm{col}(W - \omega I_n)$ are complementary subspaces of $\mathbb{C}^n$, and therefore we can define a (unique) projector, denoted by Q_ω, onto $\mathrm{null}(W - \omega I_n)$ along $\mathrm{col}(W - \omega I_n)$.[7] Recalling that the zeros of $\det(S(\lambda))$ are the reciprocals of the nonzero eigenvalues of W, we can prove the following result.

Theorem 1. *Suppose Assumption 1 holds. In a SAR model, for any semisimple nonzero real eigenvalue ω of W, $\lim_{\lambda \to \omega^{-1}} l_a(\lambda)$ is a.s.*

   (i) $-\infty$ if $\mathrm{tr}(M_X Q_\omega) > 0$;

   (ii) bounded if $\mathrm{tr}(M_X Q_\omega) = 0$;

  (iii) $+\infty$ if $\mathrm{tr}(M_X Q_\omega) < 0$.

We can now compare the behavior of the adjusted profile log-likelihood $l_a(\lambda)$ near the real zeros of of $\det(S(\lambda))$, as established by Theorem 1, with the behavior of the unadjusted profile log-likelihood $l(\lambda)$ near those points. Recall from Section 2.2 that $\lim_{\lambda \to \omega^{-1}} l(\lambda) = -\infty$ a.s., for any nonzero real eigenvalue ω of W (under Assumption 1). Thus, $l(\lambda)$ and its adjusted version $l_a(\lambda)$ have the same behavior near a point ω^{-1}, for a semisimple nonzero real eigenvalue ω, *only* in case (i) of Theorem 1. In case (ii), $l_a(\lambda)$ can be extended to a function that is a.s. continuous at $\lambda = \omega^{-1}$. This can be achieved by extending the domain of $l_a(\lambda)$ to include ω^{-1}, and setting $l_a(\omega^{-1}) := \lim_{\lambda \to \omega^{-1}} l_a(\lambda)$. From now on, when we say that $l_a(\lambda)$ is continuous at $\lambda = \omega^{-1}$ we implicitly assume that this extension has been performed. In case (iii) of Theorem 1, $l_a(\lambda)$ is unbounded from above near ω^{-1}. From unreported numerical experiments, it appears that $\mathrm{tr}(M_X Q_\omega) < 0$ is an extremely rare occurrence for pairs (W, X) which are likely to be encountered in applications.[8] We shall also see shortly that $\mathrm{tr}(M_X Q_\omega) < 0$ cannot occur if W is symmetric.

Ruling out the pathological case (iii), it is useful to try and understand which of cases (i) and (ii) in Theorem 1 is likely to occur in applications. At first sight, the condition $\mathrm{tr}(M_X Q_\omega) = 0$ in case (ii) may look very restrictive. Indeed, Lemma A.1 in Appendix A

---

is *irreducible* if and only if the graph of A is strongly connected (e.g., Meyer, 2000, p. 671).

   [7]The matrix Q_ω is a "spectral projector" of W; see, for instance, Meyer (2000), where explicit general representations for such projectors can also be found. Particular cases are given in the proof of Lemma 3.1 and in the proof of Lemma S.4.1 in the Supplement.

   [8]For example, for all simulation designs in Section 5, we find that $l_a(\lambda)$ is always bounded from above on Λ_a. A complete understanding of when $l_a(\lambda)$ may be unbounded from above would be of interest, but is left for future research. Note that the fact that $l_a(\lambda)$ can be unbounded from above is not surprising, since $l(\lambda)$ itself can be unbounded from above (see Lemma S.1.2 in the Supplement).

establishes that $\text{tr}(M_X Q_\omega) \neq 0$ for *generic,* in the measure theoretic sense, full column rank X (and for any fixed W). However, X typically contains an intercept (or group intercepts), and this implies that $\text{tr}(M_X Q_\omega) = 0$ occurs, at $\omega = 1$, for a very large class of matrices W used in practice. To see why this is the case, the next lemma provides a condition for $\text{tr}(M_X Q_\omega) = 0$ in terms of the eigenspace $\text{null}(W - \omega I_n)$ (the eigenspace of W associated to the eigenvalue ω).

Lemma 3.1.

(i) *For any semisimple eigenvalue ω of W, $\text{tr}(M_X Q_\omega) = 0$ if $\text{null}(W - \omega I_n) \subseteq \text{col}(X)$;*

(ii) *For any eigenvalue ω of a symmetric W, $\text{tr}(M_X Q_\omega) = 0$ if $\text{null}(W - \omega I_n) \subseteq \text{col}(X)$, $\text{tr}(M_X Q_\omega) > 0$ otherwise.*

Part (i) of Lemma 3.1 establishes that $\text{null}(W - \omega I_n) \subseteq \text{col}(X)$ is sufficient for case (ii) of Theorem 1 to apply, and hence for $l(\lambda)$ and $l_a(\lambda)$ to have different behaviour near $\lambda = \omega^{-1}$ (for any semisimple nonzero real eigenvalue ω of W, and provided that Assumption 1 holds). It turns out that this sufficient condition is very often satisfied in applications. Two examples are given next, the second one being a generalization of the first.

Example 3.1 (Row stochastic and irreducible weights matrix)**.** In applications of spatial autoregressions, W is often row stochastic and irreducible (cf. footnote 6), and an intercept is included in the regression. By the Perron–Frobenius Theorem (e.g., Horn and Johnson, 1985, Theorem 8.4.4), $\omega = 1$ is a simple (and hence semisimple) eigenvalue of W, and the associated eigenspace $\text{null}(W - I_n)$ is spanned by a vector of identical entries, and therefore is in $\text{col}(X)$. It follows, under Assumption 1, that $l(\lambda)$ a.s. approaches $-\infty$ as $\lambda \to 1$, while $l_a(\lambda)$ is a.s. continuous at $\lambda = 1$.

Example 3.2 (Block diagonal weights matrix)**.** Example 3.1 generalizes immediately to the case when W is a block diagonal matrix whose blocks are row stochastic and irreducible matrices, of, say, size $m_r \times m_r$. That is, using direct sum notation, $W = \bigoplus_{r=1}^R W_r$. This situation arises, for instance, in a social interaction model on R networks (e.g., Lee et al., 2010), or in a spatial panel model where individuals are followed over time (in which case r is the time dimension; see, e.g., Lee and Yu, 2010). When W has this structure, the eigenspace $\text{null}(W - I_n)$ is spanned by the columns of the (network or time) fixed effects matrix $\bigoplus_{r=1}^R \iota_{m_r}$, and therefore is in $\text{col}(X)$ as long as the regressions contains those fixed effects. In that case, and provided that Assumption 1 holds, $l(\lambda)$ a.s. approaches $-\infty$ as $\lambda \to 1$, while $l_a(\lambda)$ is a.s. continuous at $\lambda = 1$.

According to part (ii) of Lemma 3.1, when W is symmetric, the condition $\text{null}(W - \omega I_n) \subseteq \text{col}(X)$ is also necessary for $l_a(\lambda)$ to be a.s. continuous at $\lambda = \omega^{-1}$. Thus, for symmetric W, Theorem 1 reduces to the simple statement that $\lim_{\lambda \to \omega^{-1}} l_a(\lambda)$ is a.s. bounded if $\text{null}(W - \omega I_n) \subseteq \text{col}(X)$, $-\infty$ otherwise, for any semisimple nonzero real eigenvalue ω.

We end this section by providing a graphical comparison of $l(\lambda)$ and $l_\mathrm{a}(\lambda)$. Consider a SAR model with weights matrix W equal to the row normalized adjacency matrix of an Erdős-Rényi $G(n,p)$ graph (Erdős and Rényi, 1959). The $G(n,p)$ graph is a random graph on n vertices, with an edge between any two vertices being present with probability p, independently of every other edge. Suppose that the regression contains an intercept, and that, for simplicity, the graph is connected. Then, since W is row-stochastic and irreducible, $l_\mathrm{a}(\lambda)$ is a.s. continuous at $\lambda = 1$ and a.s. approaches $-\infty$ as λ approaches any other singularity of $S(\lambda)$. Figure 1 displays $l(\lambda)$ and $l_\mathrm{a}(\lambda)$, for one random draw of $G(n,p)$ and for one random draw from the intercept-only model $y = .5Wy + \iota_n + \varepsilon$, with $\varepsilon \sim \mathrm{N}(0, I_n)$. The log-likelihood functions are plotted for $\lambda \in (0, \omega_3^{-1})$, where ω_3 is the third largest eigenvalue of W, and $l_\mathrm{a}(\lambda)$ has been lowered so that the two likelihoods have the same maximum value.[9] Note that, contrary to $l(\lambda)$, $l_\mathrm{a}(\lambda)$ does not go to $-\infty$ as $\lambda \to 1$.
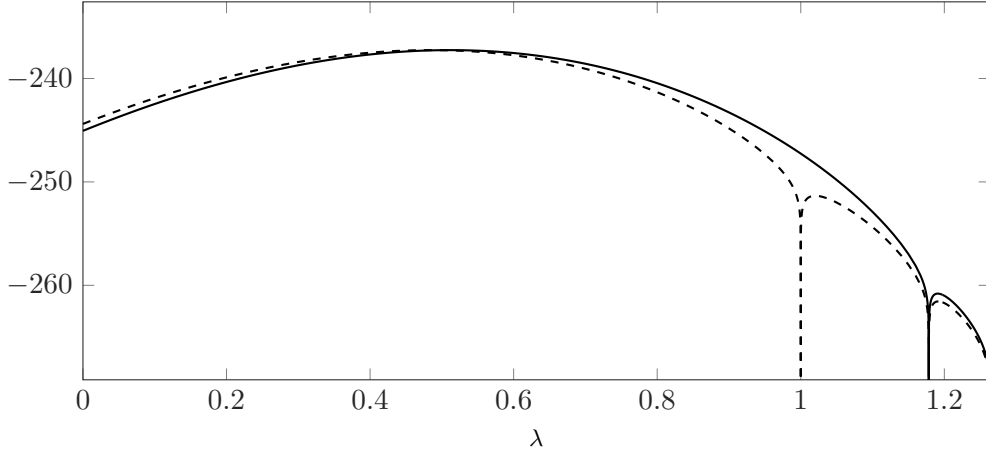

Figure 1: The profile likelihood $l(\lambda)$ (dashed) and its adjusted version $l_\mathrm{a}(\lambda)$ (solid) for a SAR model on a connected $G(100, 0.05)$ graph.

3.3 The adjusted QMLE

Theorem 1 establishes that the adjusted profile log-likelihood $l_\mathrm{a}(\lambda)$ may, in contrast to $l(\lambda)$, be a.s. continuous at the extremes of the parameter space Λ. As a consequence, there is no guarantee that $l_\mathrm{a}(\lambda)$ has a maximum over Λ, which suggests that $l_\mathrm{a}(\lambda)$ should be maximized over a larger set, Λ_a say. As anticipated in Section 3.1, it is natural to define Λ_a as the shortest open interval containing the origin with the property that $l_\mathrm{a}(\lambda) \to -\infty$ a.s. at both extremes of Λ_a. For example, in the case of Figure 1, $\Lambda = (-1.195, 1)$ and $\Lambda_\mathrm{a} = (-1.195, 1.178)$. Note that the extremes of Λ_a must always be zeros of $\det(S(\lambda))$,

---

[9]For the particular random draw of ε underlying Figure 1, $\hat{\lambda}_\mathrm{ML}$ and its adjusted version $\hat{\lambda}_\mathrm{aML}$ (defined as in Section 3.3 below) are about 0.478 and 0.506, respectively. Over 10^6 draws from $\varepsilon \sim \mathrm{N}(0, I_n)$ (and for the same draw of $G(n,p)$ used for Figure 1), empirical bias and RMSE are -0.039 and 0.128 for $\hat{\lambda}_\mathrm{ML}$ and -0.007 and 0.126 for $\hat{\lambda}_\mathrm{aML}$.

because $l_a(\lambda)$ is a.s. continuous between consecutive real zeros of $\det(S(\lambda))$.[10] Hence, our definition of Λ_a requires the following assumption.

Assumption 2. There is at least one negative zero and at least one positive zero of $\det(S(\lambda))$ such that $l_a(\lambda) \to -\infty$ a.s. as λ approaches those zeros.

It is clear from Section 3.2 that Assumption 2 can be violated only in very special cases. In those cases, one could take the left (resp., right) endpoint of Λ_a to be $-\infty$ (resp., $+\infty$), but we refrain from doing this, for simplicity.

It is natural to define the adjusted QMLE of λ as the maximizer of $l_a(\lambda)$ over Λ_a, that is,

$$\hat{\lambda}_{aML} := \operatorname*{argmax}_{\lambda \in \Lambda_a} l_a(\lambda).$$

Maximization of $l_a(\lambda)$ over a subset of Λ_a would yield a censored version of $\hat{\lambda}_{aML}$. In particular, let $\bar{\Lambda}$ be Λ augmented with one of its endpoints if $l_a(\lambda)$ is bounded near that endpoint (augmented with both endpoints if $l_a(\lambda)$ is bounded near both endpoints), and let $\bar{\lambda}_{aML} := \operatorname{argmax}_{\lambda \in \bar{\Lambda}} l_a(\lambda)$ be the estimator that is obtained by maximizing $l_a(\lambda)$ over $\bar{\Lambda}$. Since $\Lambda \subseteq \Lambda_a$, $\bar{\lambda}_{aML}$ is a censored version of $\hat{\lambda}_{aML}$.

Note that the set Λ_a, contrary to Λ, may depend on X (because, by Theorem 1, whether or not $l_a(\lambda) \to -\infty$ a.s. at some zero of $\det(S(\lambda))$ depends on X). For a *fixed* X, $\Lambda_a = \Lambda$ if $l_a(\lambda) \to -\infty$ a.s. as λ approaches the extremes of Λ, and $\Lambda_a \supset \Lambda$ otherwise. But it is also possible to compare Λ and Λ_a for *generic* X (in the measure theoretic sense), or better, since the model typically contains an intercept, for a generic matrix X containing an intercept. The following example is important.

Example 3.3. Suppose W is row-stochastic and irreducible, and that an intercept is included in the model. Let $X = (\iota_n, \widetilde{X})$ and $\widetilde{\mathcal{X}} := \{\widetilde{X} \in \mathbb{R}^{n \times (k-1)} : \operatorname{rank}(X) = k\}$. We know from Example 3.1 that, in this case, $l_a(\lambda)$ is a.s. continuous at $\lambda = 1$, for all $\widetilde{X} \in \widetilde{\mathcal{X}}$. Consider now an arbitrary semisimple nonzero real eigenvalue $\omega \neq 1$ of W. By Lemma A.1 in Appendix A and Theorem 1, $l_a(\lambda)$ is a.s. unbounded from below or from above near ω^{-1} for $\mu_{\mathbb{R}^{n \times (k-1)}}$-almost every $\widetilde{X} \in \widetilde{\mathcal{X}}$. Assuming that $\omega_{\min}$ and the second largest (positive) eigenvalue of W, denoted by ω_2, are semisimple, it follows that $\Lambda_a = (\omega_{\min}^{-1}, \omega_2^{-1})$, for $\mu_{\mathbb{R}^{n \times (k-1)}}$-almost every $\widetilde{X} \in \widetilde{\mathcal{X}} \setminus \{\mathcal{P}_{\omega_{\min}} \cup \mathcal{P}_{\omega_2}\}$, where $\mathcal{P}_\omega$ is the set of pathological $\widetilde{X}$ such that $l_a(\lambda)$ is a.s. unbounded from above near ω^{-1} (i.e., $\mathcal{P}_\omega := \{\widetilde{X} \in \mathbb{R}^{n \times (k-1)} : \operatorname{rank}(X) = k \text{ and } \operatorname{tr}(M_X Q_\omega) < 0\}$). As discussed earlier, $\mathcal{P}_\omega$ is expected to be very small or even $\mu_{\mathbb{R}^{n \times (k-1)}}$-null in cases of interest in applications. By Lemma 3.1, $\mathcal{P}_\omega$ is empty if W is symmetric, so in that case $\Lambda_a = (\omega_{\min}^{-1}, \omega_2^{-1})$ for $\mu_{\mathbb{R}^{n \times (k-1)}}$-almost every $\widetilde{X} \in \widetilde{\mathcal{X}}$.

---

[10]The fact that $l_a(\lambda)$ is a.s. continuous between consecutive real zeros of $\det(S(\lambda))$ also means that, ignoring the pathological cases in which $l_a(\lambda)$ is a.s. unbounded from above, Λ_a is the smallest open set containing the origin on which $l_a(\lambda)$ is guaranteed to have a maximum a.s.

Three remarks about the result in Example 3.3 that $\Lambda_{\mathrm{a}} = (\omega_{\min}^{-1}, \omega_2^{-1})$, for generic non-pathological $\widetilde{X}$ are in order. First, note that $\omega_2^{-1} > 1$ and recall that $\Lambda = (\omega_{\min}^{-1}, 1)$, independently of X. Thus, how much larger Λ_{a} is compared to Λ depends only on the *eigenvalue gap* $1 - \omega_2$. There is considerable evidence in the graph theory literature that the eigenvalue gap $1 - \omega_2$ tends to be large when the graph underlying W has good connectivity and randomness properties (especially when W is the normalized adjacency matrix of a regular, undirected, loopless graph; see, e.g. Brouwer and Haemers, 2011, Chapter 4), and this is indeed something we will come upon in our Monte Carlo experiments later. Second, on replacing the intercept ι_n with group intercepts $\bigoplus_{r=1}^{R} \iota_{m_r}$, as in Example 3.2, the result in Example 3.3 generalizes immediately to the case when W is block diagonal with row stochastic and irreducible blocks. Third, the argument used in Example 3.3 can also be applied to compare Λ_{a} and Λ for weights matrices that are not row-stochastic and irreducible (or are not block diagonal with row-stochastic and irreducible blocks). To do this, note that the result $\Lambda_{\mathrm{a}} = (\omega_{\min}^{-1}, \omega_2^{-1})$ for generic non-pathological $\widetilde{X}$ in Example 3.3 arises because, in that case, $\iota_n \in \mathrm{col}(X)$ and ι_n spans the eigenspace $\mathrm{null}(W - I_n)$, so that the condition in Part (i) of Lemma 3.1 is satisfied. When W is not row-stochastic and irreducible, such a special interaction between $\mathrm{col}(X)$ and W will typically not occur, and as a result $\Lambda_{\mathrm{a}} = \Lambda = (\omega_{\min}^{-1}, 1)$ for generic non-pathological $\widetilde{X}$. This is most easily seen when W is symmetric. In that case, by part (ii) of Lemma 3.1 and Theorem 1, Λ_{a} is different from Λ only if $\mathrm{null}(W - I_n)$ or $\mathrm{null}(W - \omega_{\min} I_n)$ are in $\mathrm{col}(X)$. In general there is no reason why $\mathrm{col}(X)$ should contain those eigenspaces.

3.4   Relationship between the QMLE and the adjusted QMLE

The original motivation for seeking an unbiased estimating equation is, of course, bias reduction, or more generally, improved inference on λ (and possibly σ^2). The simulation evidence reported in Section 5 below is unequivocal that the hoped-for bias reduction is certainly achieved, and without detriment to the mean squared error. However, there is one caveat that must be mentioned here, which we discuss next.

We have seen that the intervals Λ and Λ_{a} on which the distributions of $\hat{\lambda}_{\mathrm{ML}}$ and $\hat{\lambda}_{\mathrm{aML}}$ are supported are different in many cases of interest. That is, there are circumstances in which $\hat{\lambda}_{\mathrm{aML}}$ lies outside Λ. This raises the question of whether, from the point of view of interpreting the parameter λ, that outcome is acceptable.[11] Ultimately, this will depend on the context, but censoring the estimator to ensure that it lies in Λ will clearly entail sacrifice in terms of bias. The extent of the censoring would usually be greater the larger is λ in absolute value, and would also depend on characteristics of W such as its sparseness (see the

---

[11]The situation is somewhat similar to, but more subtle than, that in which the MLE for a variance (or covariance matrix) satisfies the expected non-negativity (or positive definiteness) requirement, but a bias-corrected version of it does not. For example, subtraction of the estimated first (bias) term in an asymptotic expansion for the expectation of the MLE can have this undesirable outcome. We also note that several other estimators of λ, for example IV, GMM, or indirect inference estimators, may have support larger than Λ (see e.g., Kelejian and Prucha, 1998; Lee, 2007a; Kyriacou et al., 2017).

Monte Carlo simulations in Section 5). But, presumably, the greater the censoring, the more sacrifice there will be in terms of bias-reduction. In fact, simulations reported in Section S.5.1 of the Supplement suggest that, when $\hat{\lambda}_{\mathrm{aML}}$ is outside Λ, often the unrestricted maximizer of $l(\lambda)$ is also outside Λ (i.e., $\hat{\lambda}_{\mathrm{ML}} \neq \hat{\lambda}_{\mathrm{uML}}$). This suggests that, whenever $\Lambda_{\mathrm{a}} \neq \Lambda$, $\hat{\lambda}_{\mathrm{aML}}$ should be regarded as a modification to the QMLE that maximizes $l(\lambda)$ over Λ_{a}, not over Λ.

3.5 Confidence intervals

Confidence intervals for a parameter of interest based on the QMLE may perform poorly if the data contains little information about the nuisance parameters. This may be the case, for instance, when the number of nuisance parameters is large relative to the sample size. The theory presented in the paper so far allows the construction of confidence intervals for λ with accurate coverage even in the case of little information about β.

We say that a differentiable function is *single-peaked* on an open interval if, on that interval, it has a maximum and no other stationary points corresponding to minima or maxima (so the function is non-decreasing to the left of the peak, and non-increasing to the right of the peak). We shall see later in this subsection that $l_{\mathrm{a}}(\lambda)$ is single-peaked on Λ_{a} quite generally (Proposition 3.2), but before doing that we show how single-peakedness can be used to construct confidence intervals for λ. If $l_{\mathrm{a}}(\lambda)$ is single-peaked on Λ_{a}, then the cdf of $\hat{\lambda}_{\mathrm{aML}}$ admits the representation

$$\Pr(\hat{\lambda}_{\mathrm{aML}} \leq z; \beta, \sigma^2, \lambda) = \Pr(y'S'(z)R(z)S(z)y \leq 0), \tag{3.9}$$

for any $z \in \Lambda_{\mathrm{a}}$.[12] That is, the cdf of $\hat{\lambda}_{\mathrm{aML}}$ at z equals the cdf of the quadratic form $y'S'(z)R(z)S(z)y$ at zero. Thus, one can use some approximation to the cdf (at zero) of $y'S'(z)R(z)S(z)y$ to obtain an approximation to the cdf (at z) of $\hat{\lambda}_{\mathrm{aML}}$. Since the cumulant generating function of a quadratic form is available in closed form, at least under normality, one natural candidate is the Lugannani–Rice saddlepoint approximation (Lugannani and Rice, 1980). In fact, using a saddlepoint approximation to the cdf of a score to recover the cdf of the corresponding estimator had already been investigated in Daniels (1983) (see also Butler, 2007, Chapter 12). The approximate cdf of $\hat{\lambda}_{\mathrm{aML}}$ can then be inverted to obtain confidence intervals for λ. This approach to constructing confidence intervals has been applied by Hillier and Martellosio (2018a) to the (unadjusted) QMLE of λ. Whilst revising the present paper we discovered that essentially the same approach had previously been suggested by Paige et al. (2009) for general quadratic estimating equations, and then by Jeganathan et al. (2015) specifically for some spatial models.

Remark 3.2. Let $\hat{\theta}$ an estimator obtained from the estimating equation $q(\theta) = 0$. Paige

---

[12]The notation $\Pr(\hat{\lambda}_{\mathrm{aML}} \leq z; \beta, \sigma^2, \lambda)$ emphasizes that, of course, we are assuming that y is generated by the SAR model (2.1). But it is worth remarking that, for a general random vector y, we would still have $\Pr(\hat{\lambda}_{\mathrm{aML}} \leq z) = \Pr(y'S'(z)R(z)S(z)y \leq 0)$; all that is required is that the likelihood (2.3) is used for estimation.

et al. (2009) use the device $\Pr(\hat{\theta} \leq z) = \Pr(q(z) \leq 0)$ under the assumption that $q(\theta)$ is monotonically decreasing in θ, whereas we have used it under the more general assumption that the function with derivative $q(\theta)$ is single-peaked. It is worth pointing out that, for the case of $\hat{\lambda}_{\mathrm{ML}}$ in a SAR model, the difference between the two assumptions is immaterial when all the eigenvalues of W are real. Indeed, the score associated to the profile log-likelihood (2.4) is $s(\lambda) := n(y'W'M_X S_\lambda y)/(y'S_\lambda' M_X S_\lambda y) - \mathrm{tr}(G_\lambda)$, and hence the equation $s(\lambda) = 0$ is a.s. equivalent to $f(\lambda) = 0$, where the function $f(\lambda) := ny'W'M_X S_\lambda y - \mathrm{tr}(G_\lambda)y'S_\lambda' M_X S_\lambda y$ is monotonically decreasing in λ if all eigenvalues of W are real (see Li et al., 2013).

Let $\widetilde{\Pr}(\hat{\lambda}_{\mathrm{aML}} \leq z; \beta, \sigma^2, \lambda)$ denote the approximation to the cdf of $\hat{\lambda}_{\mathrm{aML}}$, obtained by the Lugannani–Rice formula (see Section S.6 of the Supplement for details). The parameters β and σ^2 in the approximation can be replaced with their QMLEs given λ, $\hat{\beta}_{\mathrm{ML}}(\lambda)$ and $\hat{\sigma}^2_{\mathrm{aML}}(\lambda)$, to give the approximation $\widetilde{\Pr}(\hat{\lambda}_{\mathrm{aML}} \leq z; \lambda) := \widetilde{\Pr}(\hat{\lambda}_{\mathrm{aML}} \leq z; \hat{\beta}_{\mathrm{ML}}(\lambda), \hat{\sigma}^2_{\mathrm{aML}}(\lambda), \lambda)$. Confidence intervals for λ based on $\hat{\lambda}_{\mathrm{aML}}$ can then be constructed by inverting $\widetilde{\Pr}(\hat{\lambda}_{\mathrm{aML}} \leq z; \lambda)$. More specifically, replace z with the observed value of $\hat{\lambda}_{\mathrm{aML}}$, say $\hat{\lambda}_{\mathrm{aML}}^{\mathrm{obs}}$, and let $\lambda_1 := \inf\{\lambda : \widetilde{\Pr}(\hat{\lambda}_{\mathrm{aML}} \leq \hat{\lambda}_{\mathrm{aML}}^{\mathrm{obs}}; \lambda) = 1 - \alpha_1\}$ and $\lambda_2 = \sup\{\lambda : \widetilde{\Pr}(\hat{\lambda}_{\mathrm{aML}} \leq \hat{\lambda}_{\mathrm{aML}}^{\mathrm{obs}}; \lambda) = \alpha_2\}$. Then (λ_1, λ_2) is an approximate $(1 - \alpha_1 - \alpha_2)\%$ two-sided confidence interval for λ.[13]

The Lugannani–Rice approximation we employ is, as is frequently the case, a normal based one. That is, it is constructed on the basis of the cumulant generating function of $y'S'(z)R(z)S(z)y$ that obtains if $\varepsilon \sim \mathrm{N}(0, I_n)$. It is important to emphasize, however, that we intend the approximation to be used generally. Indeed, there is considerable evidence in the literature that a normal-based Lugannani–Rice approximation is typically very accurate for the distribution of a statistic that has a limiting normal distribution. There is also evidence that a normal-based Lugannani–Rice approximation works generally very well for the distribution of a quadratic form in nonnormal variables as long as we are far from the lower tail of the distribution (Wood et al., 1993). This seems to be relevant in our case as we are only interested in the cdf of $y'S'(z)R(z)S(z)y$ at 0, and $R(z)$ is indefinite.

The key condition for representation (3.9) to hold is that $l_a(\lambda)$ is single-peaked on Λ_a. We now discuss this condition. The following very mild assumption allows us to make use of the results derived earlier for semisimple eigenvalues.[14]

Assumption 3. Every eigenvalue of W corresponding to a zero of $\det(S(\lambda))$ in Λ_a is semisimple.

Proposition 3.2. *Suppose Assumptions 1, 2, and 3 hold. If*

$$(n-k)\mathrm{tr}(M_X G^2(\lambda)) > [\mathrm{tr}(M_X G(\lambda))]^2 \ \text{for all } \lambda \in \Lambda_a \text{ such that } \det(S(\lambda)) \neq 0, \quad (3.10)$$

---

[13]Note that the set $\{\lambda : \alpha_1 \leq \widetilde{\Pr}(\hat{\lambda}_{\mathrm{aML}} \leq z; \lambda) \leq \alpha_2\}$ is in general a union of intervals. It is an interval if $\widetilde{\Pr}(\hat{\lambda}_{\mathrm{aML}} \leq z; \lambda)$ is monotonic in λ. It seems reasonable to expect this monotonicity to hold in many cases, at least approximately.

[14]When $\Lambda_a = \Lambda$, there is nothing to assume, because in that case Λ does not contain any zero of $\det(S(\lambda))$. When $\Lambda_a \supset \Lambda$, in most cases of practical interest the only zero of $\det(S(\lambda))$ in Λ_a is the one corresponding to the eigenvalue 1, which, as pointed out earlier, is virtually always semisimple in applications.

the adjusted profile log-likelihood function $l_{\mathrm{a}}(\lambda)$ is a.s. single-peaked on Λ_{a}.

It turns out that if W is symmetric, condition (3.10) is satisfied for any X (see Lemma A.2 in Appendix A), and hence Λ_{a} is a.s. single-peaked in that case. If W is not symmetric, the condition depends on both W and X.[15] Extensive numerical experimentation with nonsymmetric W suggests that for the vast majority of pairs (W, X) likely to be met in applications, $l_{\mathrm{a}}(\lambda)$ is a.s. single-peaked on Λ_{a}. And, for pairs (W, X) such that condition (3.10) is not satisfied, the probability (for some distribution of y that is absolutely continuous w.r.t. $\mu_{\mathbb{R}^n}$) that $l_{\mathrm{a}}(\lambda)$ is multi-peaked is generally very small. Note that the right hand side of representation (3.9) is likely to provide a good approximation to the cdf of $\hat{\lambda}_{\mathrm{aML}}$ whenever the probability that $l_{\mathrm{a}}(\lambda)$ is multi-peaked is nonzero but small. The performance of the saddlepoint confidence intervals is assessed by numerical simulation in Section 5. Naturally, the confidence intervals can be inverted to obtain hypothesis tests on λ, but we do not investigate the power properties of such tests here.

3.6 Network fixed effects

In this section we apply the score adjustment to a social interaction model with network fixed effects and contextual effects (e.g., Lee, 2007b; Bramoullé et al., 2009; Lee et al., 2010; Lin, 2015; Fortin and Yazbeck, 2015). There are R networks, with network r having m_r individuals. The model is

$$y_r = \lambda W_r y_r + \widetilde{X}_r \gamma + W_r \widetilde{X}_r \delta + \alpha_r \iota_{m_r} + \sigma \varepsilon_r, \ r = 1, ..., R, \tag{3.11}$$

where W_r is the weights matrix of network r, α_r is an unobserved network fixed effect, $\widetilde{X}_r$ is an $m_r \times \tilde{k}$ matrix of regressors, γ and δ are $\tilde{k} \times 1$ parameters. In terms of equation (2.1), $y = (y_1', ..., y_R')'$, $W = \bigoplus_{r=1}^{R} W_r$, $\beta = (\gamma', \delta', \alpha_1, ..., \alpha_R)'$, $\varepsilon = (\varepsilon_1', ..., \varepsilon_R')'$, and $X = (\widetilde{X}, W\widetilde{X}, \bigoplus_{r=1}^{R} \iota_{m_r})$, with $\widetilde{X} := (\widetilde{X}_1', ..., \widetilde{X}_R')'$. We assume that X has full column rank.

To avoid the incidental parameter problem that arises in the case of many networks of fixed dimensions, inference in this model usually proceeds by first eliminating the network fixed effects. Define the orthogonal projector $M_r := I_{m_r} - \frac{1}{m_r} \iota_{m_r} \iota_{m_r}'$, and let F_r be the $m_r \times (m_r - 1)$ matrix of orthonormal eigenvectors of M_r corresponding to the eigenvalue 1, so that $F_r' F_r = I_{m_r - 1}$, $F_r F_r' = M_r$, and $F_r' \iota_{m_r} = 0$. In a likelihood framework, the standard procedure is that proposed by Lee et al. (2010), which eliminates the fixed effects by premultiplying each equation in (3.11) by F_r'. Under the condition that each weights matrix W_r has all row sums equal to one (i.e., $W_r \iota_{m_r} = \iota_{m_r}$ for all $r = 1, ..., R$), we have $F_r' W_r = F_r' W_r F_r F_r'$, and therefore the transformed model is

$$y_r^* = \lambda W_r^* y_r^* + \widetilde{X}_r^* \gamma + W_r^* \widetilde{X}_r^* \delta + \sigma \varepsilon_r^*, \ r = 1, ..., R, \tag{3.12}$$

---

[15]It is interesting to note that, for the unadjusted profile likelihood $l(\lambda)$, single-peakedness over Λ holds if $ntr(G^2(\lambda)) > [\mathrm{tr}(G(\lambda))]^2$ for all $\lambda \in \Lambda$, a condition that does *not* depend on X (see Hillier and Martellosio, 2018a).

where $y_r^* := F_r' y_r$, $W_r^* := F_r' W_r F_r$, $\widetilde{X}_r^* := F_r' \widetilde{X}_r$, and $\varepsilon_r^* := F_r' \varepsilon_r$. Note that $\mathrm{var}(\varepsilon_r^*) = I_{m-1}$ if $\mathrm{var}(\varepsilon_r) = I_m$. We denote the profile (quasi) log-likelihood for (σ^2, λ) based on the transformed model (3.12) by $l_{\mathrm{LLL}}(\sigma^2, \lambda)$. If the condition $W_r \iota_{m_r} = \iota_{m_r}$, for each $r = 1, ..., R$, is not satisfied, the transformation by F_r' does not yield a reduced form and hence a likelihood. In that case, only non-likelihood procedures, such as the GMM of Liu and Lee (2010), are available.

The score adjustment discussed in the present paper provides an alternative likelihood solution to the large-R incidental parameter problem. It provides a large-R consistent estimator of all model parameters (see Remark 3.3), like the Lee et al. (2010) procedure, but, contrary to Lee et al. (2010), it does not require the constant row sums condition. The score adjustment is performed exactly as for the general SAR model, treating the α_r's as parameters, profiling out the parameter $\beta = (\gamma', \delta', \alpha_1, ..., \alpha_R)'$ as in equation (3.1), and recalling that the expectation (3.3) does not depend on β.

The next proposition shows that the score adjustment method is equivalent to the Lee et al. (2010) method, if the latter applies (i.e., $W_r \iota_{m_r} = \iota_{m_r}$ for all r) and there are no covariates (i.e., the model is $y_r = \lambda W_r y_r + \alpha_r \iota_{m_r} + \sigma \varepsilon_r$, $r = 1, ..., R$).

Proposition 3.3. *Assume that, in model (3.11), there are no regressors and $W_r \iota_{m_r} = \iota_{m_r}$, for each $r = 1, ..., R$. Then, $l_a(\sigma^2, \lambda) = l_{\mathrm{LLL}}(\sigma^2, \lambda)$, for any $y \in \mathbb{R}^n$.*

When covariates are present, the estimators of σ^2 and λ (and hence of β) obtained by the score adjustment method are different from the ones obtained by the Lee et al. (2010) method. Both methods solve the large-R incidental parameter problem, but, in addition to not requiring the constant row sum condition, the score adjustment approach also implements a correction that deals with the nuisance parameters γ and δ. The two methods are compared by simulation in Section 5.2.[16]

Remark 3.3. A formal consistency proof for the adjusted QMLE in model (3.11) would require stating several regularity conditions, which is not our aim here. Heuristically, however, consistency follows from a standard argument that we can sketch here (see, e.g., Neyman and Scott, 1948; Dhaene and Jochmans, 2015). To start with, note that in SAR models without incidental parameters, the expectation $\mathrm{E}(s(\sigma^2, \lambda))$ in equation (3.3) is $O(1)$ as $n \to \infty$ under standard conditions (see, e.g., Lee, 2004b, Lemma A.9), and therefore the necessary condition $\mathrm{plim}_{n \to \infty} \frac{1}{n} s(\sigma^2, \lambda) = 0$ for $(\hat{\sigma}_{\mathrm{ML}}^2, \hat{\lambda}_{\mathrm{ML}})$ to be consistent is satisfied. On the other hand, when there is an incidental parameter of dimension R, as in model (3.11), and n increases only because R increases, $\mathrm{plim}_{n \to \infty} \frac{1}{n} s(\sigma^2, \lambda) \neq 0$ because the bias of the profile score $s(\sigma^2, \lambda)$ is typically $O(R)$; however, the fact that the expectation (3.3) is independent of

---

[16]Our focus is on finite sample results. From the point of view of first-order asymptotics, we expect the two estimators to be equivalent under the conditions in Lee et al. (2010). Such conditions include, in particular, the fact that $\tilde{k}$ does not vary with n; if $\tilde{k}$ increased with n, the adjusted QMLE might have an advantage even from the point of view of first-order asymptotics.

the nuisance parameter β immediately implies that $\text{plim}_{n\to\infty} \frac{1}{n}\big\{s(\sigma^2, \lambda) - \text{E}(s(\sigma^2, \lambda))\big\} = 0$, which is essentially what delivers consistency of the adjusted QMLE.

4 Spatial error model

The spatial error model

$$y = X\beta + u, \ u = \lambda W u + \sigma\varepsilon \qquad (4.1)$$

is considerably less popular than the SAR model (2.1) in economic applications, but, as we shall see, provides important motivation for the score adjustment considered in this paper. We now briefly consider this model, leaving all details to Section S.8 of the Supplement. It is convenient to generalize the error structure in model (4.1) to $A(\theta)u = \sigma\varepsilon$, where $A(\theta)$ is a square matrix that is invertible for any value of the parameter θ in a subset Θ of $\mathbb{R}^p$. In the case of the spatial error model, we may take $A(\theta) = S(\lambda)$.

The profile (quasi) log-likelihood for (σ^2, θ) based on the assumption $\varepsilon \sim \text{N}(0, I_n)$ is (up to an additive constant)

$$l(\sigma^2, \theta) := l(\hat{\beta}_{\text{ML}}(\theta), \sigma^2, \theta) = -\frac{n}{2}\log(\sigma^2) + \log|\det(A(\theta))| - \frac{1}{2\sigma^2}y'U(\theta)y, \qquad (4.2)$$

where $U(\theta) := A'(\theta)M_{A(\theta)X}A(\theta)$. As for the case of the SAR model, the score associated to (4.2) can be (exactly) recentered assuming only that $\text{E}(\varepsilon) = 0$ and $\text{var}(\varepsilon) = I_n$. The corresponding adjusted likelihood is

$$l_{\text{a}}(\sigma^2, \theta) = -\frac{n-k}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}y'U(\theta)y + \log|\det(A(\theta))| - \frac{1}{2}\log\big(\det\big(X'A'(\theta)A(\theta)X\big)\big). \ (4.3)$$

Remarkably, and contrary to the case of a SAR model, $l_{\text{a}}(\sigma^2, \theta)$ is a genuine likelihood. Hence, the corresponding score provides an unbiased and information unbiased estimating equation (under the assumptions that $\text{E}(\varepsilon) = 0$ and $\text{var}(\varepsilon) = I_n$). In fact, $l_{\text{a}}(\sigma^2, \theta)$ is equivalent to the likelihood used for restricted ML (REML) estimation, which has been shown to be quite generally preferable to ML estimation in a regression model with correlated errors (see, e.g., Thompson, 1962; Patterson and Thompson, 1971; Rahman and King, 1997).

Maximization of (4.3) for fixed θ gives $\hat{\sigma}^2_{\text{aML}}(\theta) = \frac{1}{n-k}y'U(\theta)y$, and thus the profile adjusted likelihood for θ only is

$$l_{\text{a}}(\theta) := l_{\text{a}}(\hat{\sigma}^2_{\text{aML}}(\theta), \theta) = -\frac{n-k}{2}\log(y'U(\theta)y) + \log|\det(A(\theta))| - \frac{1}{2}\log\big(\det\big(X'A'(\theta)A(\theta)X\big)\big).$$
$$(4.4)$$

To fully understand the effect of the score adjustment, it is useful to relate the likelihood $l_{\text{a}}(\theta)$ to invariance properties of the model. Assuming that the distribution of ε does not depend on the parameters β and σ (and that the parameters are identifiable), it is straightforward to check that the model is invariant under the group $\mathcal{G}_X$ of transformations $y \to \kappa y + X\delta$

in the sample space, for any $\kappa > 0$, any $\delta \in \mathbb{R}^k$, and for a fixed X (for the relevant definition of invariance, see Lehmann and Romano, 2005, Chapter 6). The likelihood $l_a(\theta)$ corresponds to the density of a maximal invariant under the group $\mathcal{G}_X$ (and $l_a(\sigma^2, \theta)$ corresponds to the density of a maximal invariant under the group of transformations $y \to y + X\delta$, $\delta \in \mathbb{R}^k$). That is, using the adjusted likelihood $l_a(\theta)$ corresponds to imposing that inference should be invariant with respect to the group under which the model itself is invariant, as advocated by the "principle of invariance".

We do not give details here, but it can be shown that, as λ approaches any real zero of $\det(S(\lambda))$, $l(\lambda) \to -\infty$ a.s., whereas $l_a(\lambda)$ may either a.s. approach $-\infty$ or be a.s. bounded in a spatial error model. Thus, as might have been expected, in the spatial error model the profile log-likelihood and its adjusted version behave very much as in the SAR model.[17] So, in particular, it makes sense to define the adjusted QMLE of λ on a support different from Λ, as in the SAR model. More generally, this implies that when the covariance parameter θ is restricted to a certain set, the REML estimator of θ does not necessarily respect that restriction, something that, to the best of our knowledge, has not been noted before in the literature.

5 Simulation evidence

We conduct Monte Carlo experiments to investigate the performance of the adjusted MLE of λ in the SAR model. First, we consider the case of a single network, and then the case of the multiple networks model of Section 3.6. The number of replications is 10^6 in all experiments.

5.1 Single network

A Watts-Strogatz random graph is formed by rewiring with probability p each link in a h-ahead h-behind circular matrix (Watts and Strogatz, 1998).[18] The extreme cases $p = 0$ and $p = 1$ correspond, respectively, to a h-ahead h-behind circular matrix and to a Erdős-Rényi random graph. Watts-Strogatz graphs are popular in social network analysis because, for relatively small values of p, they can reproduce important characteristics of many real world networks, including high clustering and small distances between most nodes. In our first numerical experiment, we take W in model (2.1) to be the normalized adjacency matrix of (one realization of) a Watts-Strogatz graph. Normalization of W is either a row normalization or a spectral normalization (by spectral normalization we mean that the matrix is rescaled by

---

[17]There is one important difference though. Contrary to the case of the SAR model, in the spatial error model $l_a(\lambda)$ can never approach $+\infty$ a.s. as λ approaches a real zero of $\det(S(\lambda))$. This is because $l_a(\lambda)$ is a genuine likelihood, so a.s. unboundedness from above of $l_a(\lambda)$ as λ approaches a zero of $\det(S(\lambda))$ would imply the existence of a density that diverges to $+\infty$ almost everywhere on $\mathbb{R}^n$ as λ approaches that zero.

[18]A h-ahead h-behind circular matrix A is the adjacency matrix of a graph made of n vertices on a circle such that each vertex is linked to its h nearest neighbors on each side. That is, $A(i,j) = 1$ if $0 < |i - j| \mod(n - 1 - h) \le h)$, and $A(i,j) = 0$ otherwise.

its spectral radius). Note that the two normalizations are equivalent when $p = 0$, due to the fact that a h-ahead h-behind circular matrix has constant row sums. The matrix X contains an intercept, $\tilde{k}$ regressors, and, to allow for contextual effects, the spatially lagged version of those regressors. That is, $X = (\iota_n, \widetilde{X}, W\widetilde{X})$, where $\widetilde{X}$ is $n \times \tilde{k}$. The matrix $\widetilde{X}$ is drawn in each repetition; half of its columns are drawn from independent $N(0, I_n)$ distributions, half from independent uniform distributions on $[0, 1]$.[19] For $p > 0$, W is drawn once and then kept constant across repetitions. We set $\beta = \iota_{2\tilde{k}+1}$, $\sigma = 1$, and $n = 200$. The errors ε_i are generated from independent standard normal distributions.

In the present context, $\hat{\lambda}_{\mathrm{ML}}$ is, subject to regularity conditions, consistent and asymptotically normal as $n \to \infty$ (Lee, 2004a), and $\hat{\lambda}_{\mathrm{ML}}$ and $\hat{\lambda}_{\mathrm{aML}}$ are first-order asymptotically equivalent (Yu et al., 2015). Table 1 compares the finite sample performance of $\hat{\lambda}_{\mathrm{ML}}$ and $\hat{\lambda}_{\mathrm{aML}}$ for a range of values of p, h, and λ, when $\tilde{k} = 2$. The columns headed by $\hat{\lambda}_{\mathrm{ML}}$ and $\hat{\lambda}_{\mathrm{aML}}$ give the bias(s.d.) of the estimators. $\Delta\%$ stands for percentage change (of the absolute bias or RMSE) from $\hat{\lambda}_{\mathrm{ML}}$ to $\hat{\lambda}_{\mathrm{aML}}$. For the case of row normalization, we also report ω_2^{-1}, the percentage of times when $\hat{\lambda}_{\mathrm{aML}} > 1$, denoted by $\%(\hat{\lambda}_{\mathrm{aML}} > 1)$, and bias and s.d. of $\bar{\lambda}_{\mathrm{aML}}$ (which, recall, is the estimator that is set to 1 when $\hat{\lambda}_{\mathrm{aML}} > 1$). According to the results in Section 3.2, in the present setting Λ_{a} must be the same in each repetition, with its right endpoint being equal to ω_2^{-1} when W is row normalized, equal to 1 when W is spectrally normalized (and $p > 0$). Thus, $\hat{\lambda}_{\mathrm{aML}}$ may be greater than 1 when W is row normalized, while $\hat{\lambda}_{\mathrm{aML}} < 1$ in all repetitions when W is spectrally normalized. The results in Table 1 show that, when $p = 0$, $\hat{\lambda}_{\mathrm{aML}}$ provides a significant improvement compared to $\hat{\lambda}_{\mathrm{ML}}$ both in terms of bias and RMSE. As p increases, the reduction in bias afforded by $\hat{\lambda}_{\mathrm{aML}}$ increases but at the cost of greater variability, for both normalizations. The bias of $\hat{\lambda}_{\mathrm{aML}}$ is small, unless h is large and p is small. Note that ω_2^{-1} increases with p and h (cf. the paragraph after Example 3.3), and $\%(\hat{\lambda}_{\mathrm{aML}} > 1)$ is nondecreasing in p, h, and λ, and can be very large when p, h, and λ are large. As $\%(\hat{\lambda}_{\mathrm{aML}} > 1)$ increases, $\bar{\lambda}_{\mathrm{aML}}$ becomes less biased, and more variable, compared to $\hat{\lambda}_{\mathrm{aML}}$.

Table 2 analyzes the impact of the number of regressors. We take $\lambda = .5$ and $h = 5$. For these values of λ and h, $\Pr(\hat{\lambda}_{\mathrm{aML}} > 1)$ is either zero or negligible in the row normalized case, so, contrary to the previous table, we do not report ω_2^{-1}, $\%(\hat{\lambda}_{\mathrm{aML}} > 1)$, and $\bar{\lambda}_{\mathrm{aML}}$. As expected, the bias reduction afforded by $\hat{\lambda}_{\mathrm{aML}}$ increases as $\tilde{k}$ increases. As $\tilde{k}$ increases, the relative performance of $\hat{\lambda}_{\mathrm{aML}}$, compared to $\hat{\lambda}_{\mathrm{ML}}$, improves also in terms of RMSE if p is not large.

---

[19]Given our theoretical framework, it could be argued that a design with $\widetilde{X}$ fixed across repetitions would be more appropriate. However, we prefer to vary $\widetilde{X}$ across repetitions, in an attempt to investigate the performance of the adjusted QMLE in an environment where $\widetilde{X}$ is random.

Table 1: Comparison of $\hat{\lambda}_{\mathrm{ML}}$ and $\hat{\lambda}_{\mathrm{aML}}$ on a Watts-Strogatz network of size $n = 200$, with $\tilde{k} = 2$ regressors. $\Delta\%$ refers to a percentage change from $\hat{\lambda}_{\mathrm{ML}}$ to $\hat{\lambda}_{\mathrm{aML}}$.

| | | | Row normalization | | | | | | | Spectral normalization | | | |
| | | | $\hat{\lambda}_{\mathrm{ML}}$ | $\hat{\lambda}_{\mathrm{aML}}$ | | | | | $\bar{\lambda}_{\mathrm{aML}}$ | $\hat{\lambda}_{\mathrm{ML}}$ | $\hat{\lambda}_{\mathrm{aML}}$ | | |
p	h	λ	bias(s.d.)	bias(s.d.)	$\Delta\%$|bias|	$\Delta\%$RMSE	ω_2^{-1}	$\%(\hat{\lambda}_{\mathrm{aML}} > 1)$	bias(s.d.)	bias(s.d.)	bias(s.d.)	$\Delta\%$|bias|	$\Delta\%$RMSE
0	5	0	−0.072(0.161)	−0.020(0.160)	−72.27	−8.80	1.01	0.00	−0.020(0.160)	−0.072(0.161)	−0.020(0.160)	−72.27	−8.80
		0.5	−0.046(0.095)	−0.015(0.094)	−68.09	−10.58	1.01	0.00	−0.015(0.094)	−0.046(0.095)	−0.015(0.094)	−68.09	−10.58
		0.9	−0.013(0.025)	−0.005(0.024)	−61.17	−12.09	1.01	0.00	−0.005(0.024)	−0.013(0.025)	−0.005(0.024)	−61.17	−12.09
	10	0	−0.168(0.253)	−0.048(0.246)	−71.24	−17.50	1.02	0.00	−0.048(0.246)	−0.168(0.253)	−0.048(0.246)	−71.24	−17.50
		0.5	−0.105(0.153)	0.034(0.145)	−68.07	−19.51	1.02	0.00	−0.034(0.145)	−0.105(0.153)	−0.034(0.145)	−68.07	−19.51
		0.9	−0.032(0.045)	−0.012(0.042)	−62.49	−21.12	1.02	0.00	−0.012(0.042)	−0.032(0.045)	−0.012(0.042)	−62.49	−21.12
	50	0	−1.150(0.734)	−0.215(0.821)	−81.33	−37.78	1.60	3.58	−0.220(0.813)	−1.150(0.734)	−0.215(0.821)	−81.33	−37.78
		0.5	−1.068(0.718)	−0.254(0.696)	−76.24	−42.43	1.60	10.90	−0.270(0.676)	−1.068(0.718)	−0.254(0.696)	−76.24	−42.43
		0.9	−0.849(0.609)	−0.222(0.540)	−73.81	−44.11	1.60	31.36	−0.277(0.490)	−0.849(0.609)	−0.222(0.540)	−73.81	−44.11
0.2	5	0	−0.045(0.151)	−0.010(0.153)	−78.84	−2.91	1.15	0.00	−0.010(0.153)	−0.044(0.149)	−0.009(0.151)	−79.46	−2.68
		0.5	−0.044(0.107)	−0.013(0.108)	−71.27	−6.30	1.15	0.00	−0.013(0.108)	−0.039(0.101)	−0.011(0.102)	−72.62	−5.42
		0.9	−0.035(0.050)	−0.009(0.053)	−74.54	−12.14	1.15	0.59	−0.009(0.053)	−0.017(0.034)	−0.004(0.036)	−78.08	−5.68
	10	0	−0.099(0.221)	−0.022(0.226)	−78.03	−6.29	1.22	0.00	−0.022(0.226)	−0.092(0.214)	−0.019(0.219)	−79.31	−5.73
		0.5	−0.096(0.164)	−0.027(0.166)	−71.55	−11.47	1.22	0.00	−0.027(0.166)	−0.079(0.148)	−0.019(0.151)	−75.36	−9.28
		0.9	−0.082(0.087)	−0.021(0.093)	−74.06	−20.30	1.22	6.49	−0.023(0.090)	−0.033(0.048)	−0.005(0.053)	−84.01	−8.67
	50	0	−0.563(0.582)	−0.071(0.702)	−87.40	−12.77	2.34	5.22	−0.081(0.684)	−0.491(0.542)	−0.061(0.640)	−87.60	−12.16
		0.5	−0.661(0.514)	−0.108(0.637)	−83.65	−22.85	2.34	17.49	−0.150(0.581)	−0.467(0.406)	−0.083(0.474)	−82.31	−22.23
		0.9	−0.737(0.445)	−0.137(0.566)	−81.47	−32.44	2.34	38.25	−0.251(0.451)	−0.164(0.126)	−0.036(0.141)	−78.17	−29.41
0.5	5	0	−0.028(0.151)	−0.002(0.154)	−92.77	0.59	1.49	0.00	−0.002(0.154)	−0.026(0.146)	−0.002(0.149)	−92.81	0.61
		0.5	−0.044(0.127)	−0.008(0.132)	−82.44	−1.43	1.49	0.00	−0.008(0.132)	−0.031(0.107)	−0.005(0.111)	−82.94	−0.90
		0.9	−0.077(0.078)	−0.010(0.098)	−86.53	−9.95	1.49	12.46	−0.015(0.090)	−0.013(0.032)	−0.002(0.034)	−85.61	−0.57
	10	0	−0.064(0.218)	−0.005(0.229)	−91.37	0.66	1.80	0.00	−0.005(0.229)	−0.056(0.207)	−0.005(0.216)	−91.90	0.73
		0.5	−0.098(0.186)	−0.014(0.205)	−85.97	−2.21	1.80	0.27	−0.014(0.205)	−0.066(0.155)	−0.008(0.168)	−87.27	−0.66
		0.9	−0.161(0.125)	−0.020(0.171)	−87.87	−15.89	1.80	25.19	−0.041(0.143)	−0.028(0.047)	−0.003(0.053)	−89.61	−2.84
	50	0	−0.435(0.540)	−0.010(0.708)	−97.64	2.15	4.37	7.69	−0.033(0.665)	−0.352(0.492)	−0.019(0.600)	−94.71	−0.80
		0.5	−0.585(0.469)	−0.021(0.698)	−96.38	−6.75	4.37	23.23	−0.111(0.573)	−0.349(0.351)	−0.045(0.430)	−87.09	−12.84
		0.9	−0.742(0.399)	−0.031(0.684)	−95.86	−18.71	4.37	43.44	−0.243(0.454)	−0.112(0.099)	−0.021(0.114)	−81.04	−22.26
1	5	0	−0.024(0.151)	−0.000(0.155)	−99.25	1.32	1.76	0.00	−0.000(0.155)	−0.021(0.146)	−0.000(0.149)	−99.45	1.26
		0.5	−0.044(0.133)	−0.005(0.140)	−87.73	0.48	1.76	0.00	−0.005(0.140)	−0.029(0.110)	−0.004(0.114)	−85.40	0.39
		0.9	−0.092(0.085)	−0.008(0.115)	−91.07	−8.02	1.76	17.42	−0.018(0.101)	−0.012(0.031)	−0.002(0.033)	−84.70	−0.18
	10	0	−0.052(0.219)	0.000(0.231)	−99.87	2.74	2.45	0.00	0.000(0.231)	−0.043(0.204)	0.001(0.214)	−97.97	2.59
		0.5	−0.097(0.191)	−0.006(0.220)	−93.57	2.60	2.45	0.84	−0.007(0.219)	−0.057(0.153)	−0.003(0.168)	−94.75	2.73
		0.9	−0.186(0.133)	−0.010(0.201)	−94.38	−11.71	2.45	29.94	−0.046(0.156)	−0.023(0.044)	−0.002(0.049)	−91.20	−0.78
	50	0	−0.415(0.531)	0.002(0.710)	−99.59	5.34	7.27	8.04	−0.024(0.661)	−0.303(0.465)	−0.009(0.565)	−97.19	1.92
		0.5	−0.572(0.459)	−0.005(0.709)	−99.08	−3.38	7.27	23.99	−0.104(0.570)	−0.276(0.314)	−0.028(0.383)	−89.81	−8.22
		0.9	−0.739(0.390)	−0.011(0.704)	−98.53	−15.67	7.27	44.14	−0.240(0.452)	−0.079(0.081)	−0.012(0.094)	−84.39	−16.43

5.2 Multiple networks

We now generate data according to model (3.11). For simplicity, we consider a balanced case: each of the R networks has the same number, m, of individuals. The matrices W_r and $\widetilde{X}$ are generated as in Section 5.1. That is, each W_r is the normalized adjacency matrix of (a realization of) a Watts-Strogatz graph, and the regressors are drawn in each repetition, $\tilde{k}/2$ of them from a standard normal distribution, and the other $\tilde{k}/2$ from a Unif$(0,1)$ distribution. Errors and the fixed effects α_r are drawn (in each repetition) from N$(0,1)$ distributions, and we set $\gamma = \delta = \iota_{\tilde{k}}$ and $\sigma = 1$.[20]

Table 3 compares $\hat{\lambda}_{\text{aML}}$ to the estimator $\hat{\lambda}_{\text{LLL}}$ obtained by maximizing the Lee et al. (2010) likelihood $l_{\text{LLL}}(\sigma^2, \lambda)$, for a range of values of R, m, and $\tilde{k}$.[21] We choose $\lambda = 0.5$, and set the parameters of the Watts-Strogatz random graph to $h = 5$ and $p = 0.2$. When W is row normalized, $\hat{\lambda}_{\text{aML}}$ performs similarly to $\hat{\lambda}_{\text{LLL}}$.[22] In fact, it has lower bias (but note that the bias of $\hat{\lambda}_{\text{LLL}}$ is already very small in most of the cases considered in the table) and slightly slower RMSE. As expected, the relative performance of $\hat{\lambda}_{\text{LLL}}$ improves as $\tilde{k}$ increases.[23] When W is spectrally normalized, $\hat{\lambda}_{\text{LLL}}$ cannot be obtained, so only results for $\hat{\lambda}_{\text{aML}}$ are reported. The simulation results show that $\hat{\lambda}_{\text{aML}}$ performs very satisfactorily even in that case.

Table 2: Comparison of $\hat{\lambda}_{\text{ML}}$ and $\hat{\lambda}_{\text{aML}}$ on a Watts-Strogatz network of size $n = 200$, with $\lambda = .5$ and $h = 5$. $\Delta\%$ refers to a percentage change from $\hat{\lambda}_{\text{ML}}$ to $\hat{\lambda}_{\text{aML}}$.

		Row normalization				Spectral normalization			
		$\hat{\lambda}_{\text{ML}}$	$\hat{\lambda}_{\text{aML}}$			$\hat{\lambda}_{\text{ML}}$	$\hat{\lambda}_{\text{aML}}$		
p	$\tilde{k}$	bias(s.d.)	bias(s.d.)	$\Delta\%$|bias|	$\Delta\%$RMSE	bias(s.d.)	bias(s.d.)	$\Delta\%$|bias|	$\Delta\%$RMSE
0	2	−0.053(0.104)	−0.017(0.101)	−67.29	−12.32	−0.053(0.104)	−0.017(0.101)	−67.29	−12.32
	6	−0.072(0.095)	−0.013(0.092)	−81.54	−22.13	−0.072(0.095)	−0.013(0.092)	−81.54	−22.13
	10	−0.086(0.091)	−0.011(0.088)	−87.22	−29.41	−0.086(0.091)	−0.011(0.088)	−87.22	−29.41
0.2	2	−0.050(0.116)	−0.015(0.117)	−71.09	−7.09	−0.038(0.102)	−0.010(0.103)	−72.47	−4.96
	6	−0.055(0.104)	−0.011(0.104)	−80.26	−10.97	−0.039(0.087)	−0.007(0.088)	−82.03	−7.71
	10	−0.059(0.096)	−0.009(0.097)	−84.55	−13.82	−0.039(0.078)	−0.005(0.078)	−86.65	−9.80
0.5	2	−0.049(0.132)	−0.009(0.138)	−82.39	−1.72	−0.028(0.103)	−0.005(0.106)	−81.88	−0.48
	6	−0.048(0.123)	−0.008(0.128)	−84.19	−2.83	−0.022(0.086)	−0.003(0.088)	−84.18	−0.85
	10	−0.048(0.116)	−0.006(0.121)	−86.38	−3.43	−0.018(0.074)	−0.002(0.075)	−86.07	−1.19
1	2	−0.047(0.137)	−0.006(0.146)	−88.13	0.56	−0.025(0.105)	−0.003(0.109)	−87.54	0.61
	6	−0.043(0.129)	−0.005(0.137)	−88.03	0.64	−0.018(0.088)	−0.002(0.090)	−88.64	0.56
	10	−0.041(0.124)	−0.005(0.131)	−88.62	0.47	−0.014(0.075)	−0.002(0.077)	−88.69	0.47

Table 4 reports empirical coverages of Wald confidence intervals based on first-order asymptotic normality of $\hat{\lambda}_{\text{LLL}}$ (in the columns headed by W_{LLL}), empirical coverages of Wald

---

[20]The simulated model is one in which fixed effects and regressors are random and independent of each other, and W is non-stochastic. Note that we could allow for some correlation between fixed effects and regressors, but this is not needed for our purposes.

[21]Some results for larger R are given in Table S.2 in the Supplement.

[22]Note that, as for the design in Section 5.1, in the present setting the set Λ_{a} is the same in all repetitions.

[23]Similarly to the cross-sectional case, the reduction in RMSE is higher for lower values of p (for example, in the extreme case $p = 0$, $\Delta\%$RMSE is -4.417, -14.707, -24.278 when $\tilde{k}$ is $2, 6, 10$, respectively, and $R = 10$, $m = 20$).

Table 3: Model (3.11) with $\lambda = 0.5$, $h = 5$ and $p = 0.2$. $\Delta\%$ refers to a percentage change from $\hat{\lambda}_{\text{LLL}}$ to $\hat{\lambda}_{\text{aML}}$.

			Row normalization				Spectral normalization
			$\hat{\lambda}_{\text{LLL}}$	$\hat{\lambda}_{\text{aML}}$			$\hat{\lambda}_{\text{aML}}$
$\tilde{k}$	R	m	bias(s.d.)	bias(s.d.)	$\Delta\%|\text{bias}|$	$\Delta\%\text{RMSE}$	bias(s.d.)
2	10	20	$-0.032(0.209)$	$-0.016(0.210)$	-49.93	-0.58	$-0.007(0.170)$
		30	$-0.021(0.130)$	$-0.011(0.130)$	-49.68	-0.94	$-0.007(0.111)$
	20	20	$-0.015(0.149)$	$-0.008(0.149)$	-48.86	-0.08	$-0.004(0.119)$
		30	$-0.011(0.091)$	$-0.006(0.090)$	-46.48	-0.60	$-0.003(0.078)$
	30	20	$-0.010(0.120)$	$-0.005(0.120)$	-48.23	-0.25	$-0.002(0.098)$
		30	$-0.007(0.073)$	$-0.004(0.073)$	-44.79	-0.40	$-0.003(0.065)$
6	10	20	$-0.053(0.200)$	$-0.014(0.202)$	-73.94	-2.31	$-0.005(0.141)$
		30	$-0.033(0.117)$	$-0.010(0.117)$	-70.68	-2.98	$-0.004(0.093)$
	20	20	$-0.025(0.140)$	$-0.006(0.141)$	-74.59	-0.87	$-0.002(0.098)$
		30	$-0.016(0.081)$	$-0.005(0.081)$	-70.50	-1.29	$-0.002(0.065)$
	30	20	$-0.017(0.113)$	$-0.004(0.114)$	-76.75	-0.55	$-0.002(0.080)$
		30	$-0.011(0.066)$	$-0.003(0.066)$	-69.45	-0.80	$-0.002(0.053)$
10	10	20	$-0.070(0.196)$	$-0.013(0.199)$	-81.28	-4.12	$-0.004(0.128)$
		30	$-0.041(0.110)$	$-0.008(0.111)$	-80.24	-5.92	$-0.005(0.084)$
	20	20	$-0.033(0.133)$	$-0.005(0.134)$	-83.58	-2.30	$-0.002(0.086)$
		30	$-0.020(0.074)$	$-0.005(0.074)$	-77.00	-3.12	$-0.002(0.057)$
	30	20	$-0.023(0.106)$	$-0.004(0.106)$	-81.81	-2.00	$-0.001(0.069)$
		30	$-0.013(0.059)$	$-0.003(0.059)$	-79.55	-2.36	$-0.001(0.046)$

confidence intervals based on first-order asymptotic normality of $\hat{\lambda}_{\text{aML}}$ (in the columns headed by W_{aML}), and empirical coverages of saddlepoint confidence intervals based on $\hat{\lambda}_{\text{aML}}$ (in the columns headed by s_{aML}). Since $\hat{\lambda}_{\text{LLL}}$ is not available when W is spectrally normalized, we only report the case of row normalized W.[24] The nominal size is 95%. We consider equi-tailed two-sided confidence intervals, and right-sided confidence intervals of the form $(-\infty, \lambda_U)$, where λ_U is a suitably selected upper end-point.[25] The errors ε_{ri} are generated independently from either (a) a standard normal distribution, (b) a gamma distribution with shape parameter 1 and scale parameter 1, demeaned by the population mean. Mean, variance, skewness, and kurtosis are $0, 1, 0, 3$ in case (a) and $0, 1, 2, 9$ in case (b). The Wald confidence intervals based on $\hat{\lambda}_{\text{aML}}$ offer an improvement over Wald confidence intervals based on $\hat{\lambda}_{\text{LLL}}$ when $\tilde{k}$ is not too small, and particularly in the case of right sided confidence intervals. The Lugannani–Rice approximation delivers a further improvement, and indeed the coverages of the saddlepoint confidence intervals based on $\hat{\lambda}_{\text{aML}}$ are excellent in all cases considered in the table, even under the gamma distribution. Conversely, the empirical coverage of the Wald confidence intervals based on $\hat{\lambda}_{\text{LLL}}$ is acceptable in the two-sided case when $R = m = 30$, but quickly deteriorates as $\tilde{k}$ increases or n decreases, and is considerably worse in the right-sided

---

[24]When W is spectrally normalized, coverages of the confidence intervals based on $\hat{\lambda}_{\text{aML}}$ are similar to the case of row normalization.

[25]The 95% Wald confidence intervals are $\hat{\lambda} \pm 1.96\sqrt{\hat{v}}$ (two-sided) and $(-\infty, \hat{\lambda} + 1.645\sqrt{\hat{v}})$ (right-sided), where $\hat{\lambda}$ is either $\hat{\lambda}_{\text{LLL}}$ or $\hat{\lambda}_{\text{aMLE}}$, and $\hat{v}$ denotes the asymptotic variance given in Proposition 6.1 of Lee et al. (2010) and evaluated at the LLL or aMLE estimates of λ, β, σ^2.

case. Of course, one minus a coverage in the table gives the size of the test for $\lambda = 0$ obtained by inverting the confidence interval. Table S.4 in the Supplement reports coverages under other severely non-normal distributions; again, the saddlepoint confidence intervals based on the adjusted QMLE are very accurate in all cases considered.

Table 4: Empirical coverages of 95% confidence intervals in model (3.11) with $\lambda = 0$, $h = 5$, and $p = 0.2$. The error distribution is either a standard normal or a centered gamma$(1, 1)$, and W is row normalized.

$\tilde{k}$	R	m	Normal						Gamma					
			Two-sided			Right-sided			Two-sided			Right-sided		
			W_{LLL}	W_{aML}	s_{aML}	W_{LLL}	W_{aML}	s_{aML}	W_{LLL}	W_{aML}	s_{aML}	W_{LLL}	W_{aML}	s_{aML}
2	10	20	0.942	0.941	0.949	0.935	0.941	0.949	0.945	0.944	0.951	0.939	0.943	0.951
		30	0.946	0.945	0.950	0.938	0.945	0.949	0.946	0.945	0.950	0.940	0.946	0.949
	20	20	0.946	0.946	0.950	0.940	0.944	0.950	0.948	0.948	0.952	0.942	0.946	0.951
		30	0.947	0.947	0.950	0.942	0.946	0.950	0.949	0.949	0.951	0.943	0.948	0.950
	30	20	0.947	0.947	0.950	0.943	0.946	0.950	0.949	0.949	0.951	0.945	0.948	0.950
		30	0.948	0.947	0.949	0.944	0.948	0.949	0.949	0.949	0.951	0.945	0.949	0.949
6	10	20	0.933	0.933	0.948	0.916	0.934	0.948	0.935	0.935	0.950	0.916	0.936	0.951
		30	0.937	0.938	0.948	0.919	0.941	0.949	0.939	0.940	0.949	0.924	0.943	0.950
	20	20	0.943	0.942	0.949	0.932	0.942	0.949	0.943	0.943	0.950	0.930	0.942	0.949
		30	0.944	0.944	0.949	0.931	0.945	0.949	0.944	0.945	0.950	0.930	0.945	0.950
	30	20	0.944	0.944	0.949	0.934	0.944	0.949	0.946	0.945	0.950	0.934	0.944	0.951
		30	0.945	0.945	0.948	0.934	0.945	0.949	0.946	0.947	0.950	0.934	0.947	0.950
10	10	20	0.921	0.924	0.946	0.898	0.928	0.947	0.922	0.925	0.948	0.897	0.930	0.948
		30	0.925	0.932	0.947	0.895	0.936	0.948	0.928	0.934	0.948	0.899	0.938	0.947
	20	20	0.935	0.937	0.948	0.916	0.939	0.948	0.936	0.938	0.949	0.918	0.940	0.948
		30	0.938	0.941	0.948	0.919	0.943	0.948	0.939	0.942	0.949	0.920	0.943	0.949
	30	20	0.940	0.943	0.949	0.922	0.942	0.950	0.942	0.943	0.949	0.925	0.943	0.949
		30	0.943	0.945	0.949	0.925	0.945	0.949	0.942	0.944	0.949	0.926	0.945	0.949

6   Conclusions

Recentering the profile score for a parameter of interest is one possible way to deal with nuisance parameters. In this paper, we have applied this general principle to the estimation of the autoregressive parameter λ in a spatial autoregression. The resulting adjusted QMLE for λ successfully reduces the bias in the QMLE, provides confidence intervals with excellent coverage properties (and hence tests for λ with excellent size properties) even when the dimension of the nuisance parameter is large, and is as straightforward to compute as the original QMLE. The adjusted QMLE can also solve the incidental parameter problem that occur, for example, in social network models with network fixed effects. However, due to the fact that the parameter space for λ is usually restricted to a certain interval, the spatial autoregressive setting presents challenges for the score adjustment procedure that do not arise in other models. Namely, the distributions of the QMLE and of its adjusted version can be supported on different intervals, which means that a comparison between the two estimators is not straightforward. Our simulations suggest that the adjusted QMLE generally performs

better than the QMLE in terms of RMSE, particularly in models with a large number of covariates.

This paper has focused on a simple version of a spatial autoregressive model. In empirical applications, it is typically desirable to extend the model in various directions. For example, one may want to allow for endogeneity coming from X or W, or for model errors that are subject to a spatial autoregressive structure themselves. Such extensions would not preclude the use of the score adjustment procedure, but would typically imply that the expectation of the profile score for λ depends on nuisance parameters. In that case, as discussed in McCullagh and Tibshirani (1990), the expectation would need to be obtained numerically, rather than analytically, and the resulting estimating equation would be unbiased up to some order, rather than being exactly unbiased.

It is also worth mentioning that it should be possible to use the score adjustment procedure in conjunction with modifications to the QMLE that allow for unknown heteroskedasticity (see, e.g., Liu and Yang, 2015). Finally, the adjusted QMLE should be effective also in models where the number of regressors k is increasing with the sample size (see, e.g. Gupta and Robinson, 2018). If k increases sufficiently quickly with n, then the adjusted QMLE should have advantages, with respect to the QMLE, even in terms of first-order asymptotics.

Appendix A   Auxiliary results

Lemma A.1. *Let W be a weights matrix, and ω a semisimple real eigenvalue ω of W. The set of full column rank matrices X such that $\operatorname{tr}(M_X Q_\omega) = 0$ is a $\mu_{\mathbb{R}^{n \times k}}$-null set.*

Proof. We need to show that $\mathcal{A} := \{X \in \mathbb{R}^{n \times k} : \operatorname{rank}(X) = k \text{ and } \operatorname{tr}(M_X Q_\omega) = 0\}$ is a $\mu_{\mathbb{R}^{n \times k}}$-null set. This is equivalent to showing that $\mathcal{B} := \{X \in \mathbb{R}^{n \times k} : \det(X'X)\operatorname{tr}(M_X Q_\omega) = 0\}$ is a $\mu_{\mathbb{R}^{n \times k}}$-null set, because $\mathcal{A} = \{X \in \mathbb{R}^{n \times k} : \operatorname{rank}(X) = k\} \cap \mathcal{B}$. But $\det(X'X)\operatorname{tr}(M_X Q_\omega)$ is a polynomial in the entries of X, as it is clear from writing $\det(X'X)\operatorname{tr}(M_X Q_\omega) = \det(X'X)\operatorname{tr}(Q_\omega) - \operatorname{tr}(X\operatorname{adj}(X'X)X'Q_\omega)$. Hence $\mathcal{B}$ is an algebraic variety, and as such it is either a $\mu_{\mathbb{R}^{n \times k}}$-null set or the whole $\mathbb{R}^{n \times k}$. The latter case is easily ruled out. $\qquad\square$

Lemma A.2. *If W is symmetric, condition (3.10) is satisfied for any X.*

Proof. If W is symmetric, $v'(G(\lambda) - tI_n)^2 v \geq 0$ for any $t \in \mathbb{R}$, any $v \in \mathbb{R}^n$, and any λ such that $\det(S(\lambda)) \neq 0$. It follows that, for any $t \in \mathbb{R}$, any $u \in \mathbb{R}^m$, any $m \times n$ matrix C, and any λ such that $\det(S(\lambda)) \neq 0$, $u'C(G(\lambda) - tI_n)^2 C'u \geq 0$ and hence $\operatorname{tr}\!\big(C(G(\lambda) - tI_n)^2 C'\big) \geq 0$. Let us now choose C to be an $(n-k) \times n$ matrix such that $CC' = I_{n-k}$ and $C'C = M_X$, and t to be $\operatorname{tr}(M_X G(\lambda))/(n-k)$. Then, $\operatorname{tr}\!\big(C(G(\lambda) - tI_n)^2 C'\big) = \operatorname{tr}(M_X G^2(\lambda)) - [\operatorname{tr}(M_X G(\lambda))]^2/(n-k) = -(n-k)\delta_{\mathrm{a}}(\lambda)$. The proof is completed on noting that $\operatorname{tr}\!\big(C(G(\lambda) - tI_n)^2 C'\big) = 0$ if and only if $G(\lambda)$, and hence W, is a scalar multiple of I_n, which is ruled out by the maintained assumption that W has at least one negative and at least one positive eigenvalue. $\qquad\square$

Lemma A.3. *For any semisimple nonzero real eigenvalue ω of W,*

    (i) if $\mathrm{tr}(M_X Q_\omega) > 0$ then $\lim_{\lambda \uparrow \omega^{-1}} \mathrm{tr}(M_X G(\lambda)) = +\infty$ and $\lim_{\lambda \downarrow \omega^{-1}} \mathrm{tr}(M_X G(\lambda)) = -\infty$;

    (ii) if $\mathrm{tr}(M_X Q_\omega) = 0$ then $\lim_{\lambda \to \omega^{-1}} \mathrm{tr}(M_X G(\lambda))$ is bounded;

    (iii) if $\mathrm{tr}(M_X Q_\omega) < 0$ then $\lim_{\lambda \uparrow \omega^{-1}} \mathrm{tr}(M_X G(\lambda)) = -\infty$ and $\lim_{\lambda \downarrow \omega^{-1}} \mathrm{tr}(M_X G(\lambda)) = +\infty$.

Proof. Let $\mathrm{Sp}(W)$ denote the spectrum (defined as the set of distinct eigenvalues) of W. For any $\lambda \in \mathbb{R} \setminus \mathrm{Sp}(W)$, consider the function $f(z) := z(1 - \lambda z)^{-1}$ from $\mathbb{C} \setminus \{\lambda^{-1}\}$ to $\mathbb{C}$, and let $f^{(i)}(z)$ denote its i-th order derivative. Since $f(z)$ is defined at each eigenvalue of W, the matrix $G(\lambda)$, for any $\lambda \in \mathbb{R} \setminus \mathrm{Sp}(W)$, admits the spectral resolution (e.g., Meyer, 2000, p. 603)

$$G(\lambda) = \sum_{\chi \in \mathrm{Sp}(W)} \sum_{i=0}^{k_\chi - 1} \frac{f^{(i)}(\chi)}{j!} (W - \chi I_n)^i T_\chi, \tag{A.1}$$

where k_χ denotes the index of an eigenvalue χ, and T_χ denotes the projector onto the generalized eigenspace $\mathrm{null}\big((W - \chi I_n)^{k_\chi}\big)$ along $\mathrm{col}\big((W - \chi I_n)^{k_\chi}\big)$. If an eigenvalue ω of W is semisimple (and only in that case), then $k_\omega = 1$, and hence

$$\mathrm{tr}(M_X G(\lambda)) = f(\omega)\mathrm{tr}(M_X Q_\omega) + \sum_{\chi \in \mathrm{Sp}(W) \setminus \{\omega\}} \sum_{i=0}^{k_\chi - 1} \frac{f^{(i)}(\chi)}{j!} \mathrm{tr}\big(M_X (W - \chi I_n)^i T_\chi\big). \tag{A.2}$$

The stated result obtains, because all derivatives $f^{(i)}(\chi)$ are bounded for $\lambda \neq \chi^{-1}$. $\qquad\square$

Lemma A.4. *For any semisimple nonzero real eigenvalue ω of W, if $y \notin \mathrm{null}(M_X S(\omega^{-1}))$, then $\lim_{\lambda \to \omega^{-1}} l_a(\lambda)$ is*

    (i) $-\infty$ if $\mathrm{tr}(M_X Q_\omega) > 0$;

    (ii) bounded if $\mathrm{tr}(M_X Q_\omega) = 0$;

    (iii) $+\infty$ if $\mathrm{tr}(M_X Q_\omega) < 0$.

Proof. The result can be proved by looking at the two terms that make up $s_{a2}(\lambda)$ in expression (3.5): $(n - k)(y'W'M_X S(\lambda)y)/(y'S(\lambda)'M_X S(\lambda)y)$ and $-\mathrm{tr}(M_X G(\lambda))$. Consider an arbitrary nonzero real eigenvalue ω of W. If $y \notin \mathrm{null}(M_X S(\omega^{-1}))$, the function $\lambda \mapsto (n-k)(y'W'M_X S(\lambda)y)/(y'S(\lambda)'M_X S(\lambda)y)$ is continuous at $\lambda = \omega^{-1}$ because it is well defined at $\lambda = \omega^{-1}$, and is well defined in a neighborhood of $\lambda = \omega^{-1}$ (the last claim follows by Lemma S.1.1 in the online supplement to Hillier and Martellosio, 2018a). The proof is completed on using Lemma A.3 to establish the limiting behavior of the term $-\mathrm{tr}(M_X G(\lambda))$ as $\lambda \to \omega^{-1}$.
$\qquad\square$

Lemma A.5. *For any semisimple nonzero real eigenvalue ω of W, $[\operatorname{tr}(M_X G(\lambda))]^2 - (n - k)\operatorname{tr}(M_X G^2(\lambda)) \to +\infty$ as $\lambda \to \omega^{-1}$ if $\operatorname{tr}(M_X Q_\omega) < 0$.*

Proof. This proof is based on the proof of Lemma A.3. For any $\lambda \in \mathbb{R} \setminus \operatorname{Sp}(W)$, consider the function $g(z) = f^2(z) = z^2(1 - \lambda z)^{-2}$ from $\mathbb{C} \setminus \{\lambda^{-1}\}$ to $\mathbb{C}$, and let $g^{(i)}(z)$ denote its i-th order derivative. Then, for any $\lambda \in \mathbb{R} \setminus \operatorname{Sp}(W)$, a spectral resolution of $G^2(\lambda)$ is given by the right hand side of equation (A.1) with all $f^{(i)}(\cdot)$ replaced by $g^{(i)}(\cdot)$. Hence, for an arbitrary semisimple nonzero real eigenvalue ω of W, $\operatorname{tr}(M_X G^2(\lambda))$ can be expressed as in the right hand side of (A.2), with f replaced by g. Since all derivatives $g^{(i)}(\chi)$ are bounded for $\lambda \neq \chi^{-1}$, $\lim_{\lambda \to \omega^{-1}} \operatorname{tr}(M_X G^2(\lambda)) = -\infty$ if $\operatorname{tr}(M_X Q_\omega) < 0$. The proof is completed on observing that, by Lemma A.3, $\lim_{\lambda \to \omega^{-1}} [\operatorname{tr}(M_X G(\lambda))]^2 = +\infty$ if $\operatorname{tr}(M_X Q_\omega) < 0$. $\qquad\square$

Lemma A.6. *Suppose that, in a SAR model, $\operatorname{col}(X)$ is an invariant subspace of W. Then the adjusted log-likelihood function $l_a(\sigma^2, \lambda)$ is the same as the quasi log-likelihood function for (σ^2, λ) based on Dy, for any full rank $(n - k) \times n$ matrix D such that $DX = 0$, and for any $y \in \mathbb{R}^n$.*

Proof. We only provide a sketch of the proof here; full details are given in Section S.9 of the Supplement. If $\operatorname{col}(X)$ is an invariant subspace W, there exists a unique $k \times k$ matrix A such that $WX = XA$, and hence $S^{-1}(\lambda)X = X(I_k - \lambda A)^{-1}$, for any λ such that $S(\lambda)$ is invertible. Thus, when $\operatorname{col}(X)$ is an invariant subspace W, the SAR model $y = S^{-1}(\lambda)X\beta + \sigma S^{-1}(\lambda)\varepsilon$ can be written as $y = X(I_k - \lambda A)^{-1}\beta + \sigma S^{-1}(\lambda)\varepsilon$, which corresponds to the spatial error model (4.1) with β replaced by $(I_k - \lambda A)^{-1}\beta$. Intuitively, the lemma then follows from the fact that, for a spatial error model, the adjusted (quasi) log-likelihood $l_a(\sigma^2, \lambda)$ is the same as the restricted, or residual, quasi log-likelihood for (σ^2, λ) (see Section S.8.2.2 in the Supplement). $\qquad\square$

Appendix B Proofs

Proof of Proposition 3.1. The adjusted likelihood $l_a(\sigma^2, \lambda)$ is defined by the property that its gradient is the score $s_a(\sigma^2, \lambda)$ given in equation (3.4), for any $\sigma > 0$ and any $\lambda \in \Lambda_u$, where, recall, $\Lambda_u := \{\lambda \in \mathbb{R} : \det(S(\lambda)) \neq 0\}$. Solving the two equations

$$\frac{\partial l_a(\sigma^2, \lambda)}{\partial \sigma^2} = -\frac{n - k}{2\sigma^2} + \frac{1}{2\sigma^4} y' S'(\lambda) M_X S(\lambda) y,$$

$$\frac{\partial l_a(\sigma^2, \lambda)}{\partial \lambda} = \frac{1}{\sigma^2} y' W' M_X S(\lambda) y - \operatorname{tr}(M_X G(\lambda)),$$

gives, up to an additive constant,

$$l_a(\sigma^2, \lambda) = -\frac{n - k}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} y' S'(\lambda) M_X S(\lambda) y + \int \operatorname{tr}(M_X G(\lambda)) \, d\lambda.$$

Using elementary complex analysis, the integral $\int \mathrm{tr}(M_X G(\lambda))\,\mathrm{d}\lambda$ can be expressed in terms of the matrix logarithm. For $\lambda \in (-1, 1)$, $\log S(\lambda)$ admits the convergent power series representation $-\sum_{k=1}^{\infty} \frac{1}{k}\lambda^k W^k$, from which it is immediately clear that $\frac{\mathrm{d}}{\mathrm{d}\lambda}\log S(\lambda) = -G(\lambda)$. This expression for $\frac{\mathrm{d}}{\mathrm{d}\lambda}\log S(\lambda)$ can be extended over the set Λ_u, by selecting a suitable branch of the matrix logarithm. Let $\Xi := \{1 - \lambda\omega : \lambda \in \mathbb{R},\ \omega \in \mathrm{Sp}(W)\}$, where $\mathrm{Sp}(W)$ denotes the set of eigenvalues of W, be the subset of $\mathbb{C}$ formed by the eigenvalues of $S(\lambda)$ as λ ranges over $\mathbb{R}$. Since Ξ is formed by a finite number of lines in $\mathbb{C}$ going through 1, there must exist a half line l in $\mathbb{C}$ starting from the origin that does not intersect Ξ. In the rest of the proof, log denotes the matrix logarithm associated to the branch cut l. Then, $\log S(\lambda)$, and hence its first derivative, is holomorphic over $\mathbb{C}\backslash l$. Since $\mathbb{C}\backslash l$ is a connected set and $\Lambda_u \subset \mathbb{C}\backslash l$, the identity theorem for holomorphic functions implies that $\frac{\mathrm{d}}{\mathrm{d}\lambda}\log S(\lambda) = -G(\lambda)$ for any $\lambda \in \Lambda_u$. It follows that $\mathrm{tr}(M_X \log S(\lambda))$ is an antiderivative of $-\mathrm{tr}(M_X G(\lambda))$, and hence that $\mathrm{Re}[\mathrm{tr}(M_X \log S(\lambda))]$ is a real antiderivative of $-\mathrm{tr}(M_X G(\lambda))$, for any $\lambda \in \Lambda_u$, which completes the proof. $\qquad\square$

Proof of Theorem 1. According to Assumption 1, $M_X S(\omega^{-1}) \neq 0$, for any nonzero real eigenvalue ω of W. Hence $\mathrm{null}(M_X S(\omega^{-1}))$ is a $\mu_{\mathbb{R}^n}$-null set, for any nonzero real eigenvalue ω of W. The result follows by Lemma A.4. $\qquad\square$

Proof of Lemma 3.1. (i) For any semisimple eigenvalue ω of W, Q_ω is a projector onto $\mathrm{col}(Q_\omega) = \mathrm{null}(W - \omega I_n)$, so $\mathrm{null}(W - \omega I_n) \subseteq \mathrm{col}(X)$ if and only if $M_X Q_\omega = 0$, which is obviously sufficient for $\mathrm{tr}(M_X Q_\omega) = 0$. (ii) Let H_ω be a matrix whose columns form an orthonormal basis for $\mathrm{null}(W - \omega I_n)$. If W is symmetric, $Q_\omega = H_\omega H'_\omega$ (the orthogonal projector onto $\mathrm{null}(W - \omega I_n)$), and hence $\mathrm{tr}(M_X Q_\omega) = \mathrm{tr}(H'_\omega M_X H_\omega)$. Observe now that, since M_X is positive semidefinite, all diagonal entries of $H'_\omega M_X H_\omega$ are nonnegative. Thus, $\mathrm{tr}(M_X Q_\omega) \geq 0$, and the desired conclusion follows from (i). $\qquad\square$

Proof of Proposition 3.2. According to Lemma A.5, condition (3.10) implies that $\mathrm{tr}(M_X Q_\omega) \geq 0$ for all semisimple eigenvalues ω such that $\omega^{-1} \in \Lambda_a$. Hence, by Theorem 1, as λ approaches the reciprocals of such eigenvalues, $l_a(\lambda)$ either a.s. diverges to $-\infty$ or is a.s. bounded, provided that Assumption 1 holds. The former case is impossible by the definition of Λ_a, so $l_a(\lambda)$ must be a.s. continuous on the whole Λ_a (after extension of the domain of $l_a(\lambda)$ to include any zeros of $\det(S(\lambda))$ in Λ_a). By the definition of Λ_a we also have that $l_a(\lambda) \to -\infty$ a.s. at the extremes of Λ_a. We now show that the second derivative $\ddot{l}_a(\lambda)$ of $l_a(\lambda)$ is negative at any critical point of $l_a(\lambda)$ in Λ_a. This implies that $l_a(\lambda)$ a.s. has a single critical point in Λ_a, corresponding to a maximum (that is, it is single-peaked with no stationary inflection points). Write

$$\ddot{l}_a(\lambda) = (n - k)\left(-\frac{(ac - b^2)}{(a\lambda^2 - 2b\lambda + c)^2} + \frac{(b - a\lambda)^2}{(a\lambda^2 - 2b\lambda + c)^2}\right) - \mathrm{tr}(M_X G^2(\lambda)),$$

where $a := y'W'M_XWy$, $b := y'W'M_Xy$, $c := y'M_Xy$. The first order condition $s_{a2}(\lambda) = 0$ implies

$$\frac{(b - a\lambda)^2}{(a\lambda^2 - 2b\lambda + c)^2} = \frac{1}{(n-k)^2}[\text{tr}(M_XG(\lambda))]^2,$$

so that, at any critical point of $l_a(\lambda)$,

$$\ddot{l}_a(\lambda) = -(n-k)\frac{(ac - b^2)}{(a\lambda^2 - 2b\lambda + c)^2} + \frac{1}{n-k}[\text{tr}(M_XG(\lambda))]^2 - \text{tr}(M_XG^2(\lambda)). \qquad \text{(B.1)}$$

By the Cauchy-Schwarz inequality the first of the three terms on the right hand side of (B.1) is nonpositive. Hence condition (3.10) is sufficient for $\ddot{l}_a(\lambda)$ to be negative at any zero of $s_{a2}(\lambda)$ in Λ_a, which completes the proof. $\qquad \square$

Proof of Proposition 3.3. If $W_r\iota_{m_r} = \iota_{m_r}$, for all $r = 1, ..., R$, then $\text{col}(\bigoplus_{r=1}^R \iota_{m_r})$ is an invariant subspace of $W = \bigoplus_{r=1}^R W_r$. Since $F_r\iota_{m_r} = 0$, for all $r = 1, ..., R$, the desired claim follows by Lemma A.6. $\qquad \square$

References

Arellano, M., Hahn, J., 2007. Understanding bias in nonlinear panel models: some recent developments. In: Blundell, R., Newey, W., Persson, T. (Eds.), Advances in Economics and Econometrics, Ninth World Congress. Cambridge University Press, Cambridge, Ch. 12, pp. 381–409.

Bao, Y., 2013. Finite-sample bias of the QMLE in spatial autoregressive models. Econometric Theory 29 (1), 68–88.

Bao, Y., Ullah, A., 2007. Finite sample properties of maximum likelihood estimator in spatial models. Journal of Econometrics 137 (2), 396–413.

Boucher, V., Bramoullé, Y., Djebbari, H., Fortin, B., 2014. Do peers affect student achievement? Evidence from Canada using group size variation. Journal of Applied Econometrics 29 (1), 91–109.

Bramoullé, Y., Djebbari, H., Fortin, B., 2009. Identification of peer effects through social networks. Journal of Econometrics 150 (1), 41–55.

Brouwer, A. E., Haemers, W. H., 2011. Spectra of graphs. Springer, New York.

Butler, R. W., 2007. Saddlepoint Approximations with Applications. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Carrell, S. E., Sacerdote, B. I., West, J. E., 2013. From natural variation to optimal policy? The importance of endogenous peer group formation. Econometrica 81 (3), 855–882.

Conniffe, D., 1987. Expected maximum log likelihood estimation. Journal of the Royal Statistical Society. Series D (The Statistician) 36 (4), 317–329.

Daniels, H. E., 1983. Saddlepoint approximations for estimating equations. Biometrika 70 (1), 89–96.

Dhaene, G., Jochmans, K., 2015. Profile-score adjustements for incidental-parameter problems, manuscript.

Durban, M., Currie, I. D., 2000. Adjustment of the profile likelihood for a class of normal regression models. Scandinavian Journal of Statistics 27 (3), 535–542.

Erdős, P., Rényi, A., 1959. On random graphs i. Publicationes Mathematicae (Debrecen) 6, 290–297.

Fortin, B., Yazbeck, M., 2015. Peer effects, fast food consumption and adolescent weight gain. Journal of Health Economics 42, 125–138.

Gupta, A., Robinson, P. M., 2018. Pseudo maximum likelihood estimation of spatial autoregressive models with increasing dimension. Journal of Econometrics 202 (1), 92–107.

Hillier, G., Martellosio, F., 2018a. Exact and higher-order properties of the MLE in spatial autoregressive models, with applications to inference. Journal of Econometrics 205 (2), 402–422.

Hillier, G., Martellosio, F., 2018b. Exact likelihood inference in group interaction network models. Econometric Theory 34 (2), 383–415.

Horn, R. A., Johnson, C. R., 1985. Matrix Analysis. Cambridge University Press, Cambridge.

Jeganathan, P., Paige, R. L., Trindade, A. A., 2015. Saddlepoint-based bootstrap inference for the spatial dependence parameter in the lattice process. Spatial Statistics 12, 1–14.

Jiang, J., 1996. REML estimation: asymptotic behavior and related topics. The Annals of Statistics 24 (1), 255–286.

Kelejian, H. H., Prucha, I. R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. Journal of Real Estate Finance and Economics 17 (1), 99–121.

Kyriacou, M., Phillips, P. C. B., Rossi, F., 2017. Indirect inference in spatial autoregression. The Econometrics Journal 20 (2), 168–189.

Laskar, M., King, M., 2001. Modified likelihood and related methods for handling nuisance parameters in the linear regression model. In: Saleh, A. K. M. E. (Ed.), Data Analysis from Statistical Foundations. Nova Science Publisher Inc., New York, pp. 119–142.

Lee, L.-F., 2004a. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72 (6), 1899–1925.

Lee, L.-F., 2004b. Supplement to: Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Department of Economics, The Ohio State University, Columbus, OH.

Lee, L.-F., 2007a. GMM and 2sls estimation of mixed regressive, spatial autoregressive models. Journal of Econometrics 137 (2), 489–514.

Lee, L.-F., 2007b. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. Journal of Econometrics 140 (2), 333–374.

Lee, L.-F., Liu, X., Lin, X., 2010. Specification and estimation of social interaction models with network structures. Econometrics Journal 13 (2), 145–176.

Lee, L.-F., Yu, J., 2010. Estimation of spatial autoregressive panel data models with fixed effects. Journal of Econometrics 154 (2), 165–185.

Lehmann, E. L., Romano, J. P., 2005. Testing statistical hypotheses, 3rd Edition. Springer Texts in Statistics. Springer, New York.

Li, M., Yu, D., Bai, P., 2013. A note on the existence and uniqueness of quasi-maximum likelihood estimators for mixed regressive, spatial autoregression models. Statistics & Probability Letters 83 (2), 568 – 572.

Lin, X., 2015. Utilizing spatial autoregressive models to identify peer effects among adolescents. Empirical Economics 49 (3), 929–960.

Liu, S. F., Yang, Z., 2015. Modified QML estimation of spatial autoregressive models with unknown heteroskedasticity and nonnormality. Regional Science and Urban Economics 52, 50–70.

Liu, X., Lee, L.-F., 2010. GMM estimation of social interaction models with centrality. Journal of Econometrics 159 (1), 99–115.

Lugannani, R., Rice, S., 1980. Saddle point approximation for the distribution of the sum of independent random variables. Advances in Applied Probability 12 (2), 475–490.

Macaskill, G. T., 1993. A note on adjusted profile likelihoods in non-linear regression. Journal of the Royal Statistical Society. Series B (Methodological) 55 (1), 125–131.

Manski, C. F., 1993. Identification of endogenous social effects: The reflection problem. The Review of Economic Studies 60 (3), 531–542.

McCullagh, P., Tibshirani, R., 1990. A simple method for the adjustment of profile likelihoods. Journal of the Royal Statistical Society B 52, 325–344.

Meyer, C. D. (Ed.), 2000. Matrix Analysis and Applied Linear Algebra. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

Moffitt, R. A., 2001. Policy Interventions, Low-Level Equilibria, and Social Interactions. In: Durlauf, S. N., Young, H. P. (Eds.), Social Dynamics. MIT Press, pp. 45–82.

Neyman, J., Scott, E. L., 1948. Consistent estimates based on partially consistent observations. Econometrica 16 (1), 1–32.

Ord, K., 1975. Estimation methods for models of spatial interaction. Journal of the American Statistical Association 70 (349), 120–126.

Pace, L., Salvan, A., 2006. Adjustments of the profile likelihood from a new perspective. Journal of Statistical Planning and Inference 136 (10), 3554–3564.

Paige, R. L., Trindade, A. A., Fernando, H. P., 2009. Saddlepointbased bootstrap inference for quadratic estimating equations. Scandinavian Journal of Statistics 36 (1), 98–111.

Patterson, H. D., Thompson, R., 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58 (3), 545–554.

Rahman, S., King, M. L., 1997. Marginal-likelihood score-based tests of regression disturbances in the presence of nuisance parameters. Journal of Econometrics 82 (1), 81–106.

Robinson, P. M., Rossi, F., 2015. Refinements in maximum likelihood inference on spatial autocorrelation in panel data. Journal of Econometrics 189 (2), 447–456.

Sartori, N., 2003. Modified profile likelihoods in models with stratum nuisance parameters. Biometrika 90 (3), 533–549.

Thompson, W. A., J., 1962. The problem of negative estimates of variance components. The Annals of Mathematical Statistics 33 (1), 273–289.

Watts, D. J., Strogatz, S., 1998. Collective dynamics of 'small-world' networks. Nature 393, 440–442.

Wood, A., Booth, J., Butler, R., 1993. Saddlepoint approximations to the cdf of some statistics with nonnormal limit distributions. Journal of the American Statistical Association 88, 680–686.

Yang, Z., 2015. A general method for third-order bias and variance corrections on a nonlinear estimator. Journal of Econometrics 186 (1), 178–200.

Yu, D., Bai, P., Ding, C., 2015. Adjusted quasi-maximum likelihood estimator for mixed regressive, spatial autoregressive model and its small sample bias. Computational Statistics and Data Analysis 87, 116–135.