

Exact and higher-order properties of the MLE in spatial autoregressive models, with applications to inference.

Grant Hillier
CeMMAP and
University of Southampton, UK
ghh@soton.ac.uk

Federico Martellosio*
University of Surrey, UK
f.martellosio@surrey.ac.uk

January 11, 2018

Abstract

The quasi-maximum likelihood estimator for the autoregressive parameter in a spatial autoregression usually cannot be written explicitly in terms of the data. A rigorous analysis of the first-order asymptotic properties of the estimator, under some assumptions on the evolution of the spatial design matrix, is available in Lee (2004), but very little is known about its exact or higher-order properties. In this paper we first show that the exact cumulative distribution function of the estimator can, under mild assumptions, be written in terms of that of a particular quadratic form. Simple examples are used to illustrate important exact properties of the estimator that follow from this representation. In general models a complete exact analysis is not possible, but a higher-order (saddlepoint) approximation is made available by the main result. We use this approximation to construct confidence intervals for the autoregressive parameter. Coverage properties of the proposed confidence intervals are studied by Monte Carlo simulation, and are found to be excellent in a variety of circumstances.

Keywords: spatial autoregression, maximum likelihood estimation, group interaction, social networks, complete bipartite graph, saddlepoint.

JEL Classification: C12, C21.

Acknowledgements: We would like to thank the Editor (Oliver Linton), the Associate Editor, the anonymous referees, and Benedikt Pötscher for very helpful comments and suggestions.

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

1 Introduction

Spatial autoregressive processes have enjoyed considerable recent popularity in modelling cross-sectional and panel data in economics, social network analysis, and in several other disciplines.¹ In most applications to date, such models are based on a fixed spatial weights matrix W whose elements reflect the modeler's assumptions about the pairwise interactions between the observational units. A scalar autoregressive parameter λ measures the strength of this cross-sectional interaction, and is often of direct interest. For example, in social interaction analysis measuring the strength of network effects may be important for policy purposes. This paper is concerned with the properties of the quasi- (or pseudo-) maximum likelihood estimator (QMLE) for this parameter that is implied by assuming a Gaussian likelihood.

The particular class of spatial autoregressive models we discuss have the form

$$y = \lambda W y + X\beta + \sigma\varepsilon, \quad (1.1)$$

where y is the $n \times 1$ vector of observed random variables, X is a fixed $n \times k$ matrix of regressors of full column rank, with $n > k + 1$, ε is a mean-zero $n \times 1$ random vector, $\beta \in \mathbb{R}^k$ and $\sigma > 0$ are parameters. We refer to model (1.1) simply as the SAR (spatial autoregressive) model; it is also known as the spatial lag model, or as the mixed regressive, spatial autoregressive model. The matrices X and W are assumed to be non-stochastic; alternatively, if they were random and exogenous, one could interpret inferential result as conditional on the two matrices. 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.² The model is said to be pure if the regression component ($X\beta$) is missing. The notation in (1.1) may also be used for a fixed effects panel data model, in which case W is a block-diagonal matrix and the fixed effects, along one or two of the panel dimensions, are incorporated into $X\beta$ (see, e.g., Lee and Yu, 2010).

There is a vast literature on maximum likelihood estimation of model (1.1), an early reference being Ord (1975). A rigorous first-order asymptotic analysis of the estimator was given only much later, in an influential paper by Lee (2004). Bao and Ullah (2007) provide analytical formulae for the second-order bias and mean squared error of the MLE for λ in the Gaussian pure SAR model, and Bao (2013) and Yang (2015) extend such approximations to the case when exogenous regressors are included and when ε is not necessarily Gaussian. Robinson and Rossi (2015) investigate higher-order (Edgeworth) refinements for the asymptotic distribution of the QMLE in a fixed effects spatial panel data model. Several other papers have studied the performance of the QMLE by simulation, particularly in relation to competing estimators such as the two-stage least squares (2SLS) estimator or more general GMM estimators.

Because the QMLE for λ is generally unavailable in closed form, even the calculation of the QMLE has been regarded as problematic in this model, see LeSage and Pace, (2009, Chapters 3 and 4), let alone the study of its exact properties. The starting point for the present paper is the observation that the profile likelihood for λ is, under mild

¹For an introduction to spatial autoregressions see, e.g., Cliff and Ord (1973) and LeSage and Pace (2009).

²Technically, the case of contextual effects is included in equation (1.1), but we will see later that the presence of such effects can affect the accuracy of the asymptotic approximation to the distribution of the QMLE for λ .

assumptions, *single-peaked* on the usual parameter space for λ . This fact is important in itself, because it simplifies the computation of the estimator. It also implies that the cdf of the QMLE for λ can be written down in terms of the cdf of a family of quadratic forms. Such a representation of the cdf can provide exact distribution properties of the estimator in sufficiently simple models, and some examples of this are given. The examples illustrate important distribution properties of the QMLE for λ , some of which have aspects in common with that for serial correlation coefficients (von Neumann, 1941, Koopmans, 1942). In particular, the cdf can be non-analytic at certain points of its domain, and can have a different functional form in the intervals between those points. The exact results are derived, for simplicity, under Gaussian assumptions on ε , but we note that an invariance property of the estimator implies that these hold under all scale mixtures of the normal distribution for y . The representation result for the cdf of the QMLE also permits an asymptotic analysis for the estimator in models where Lee’s (2004) assumptions are not satisfied.

However, as the examples also illustrate, exact results for general models with arbitrary (W, X) are either intractable, or too complex to be useful for inference. The explicit representation for the cdf is nevertheless useful, because it provides access to a simple higher-order (saddlepoint) approximation to the distribution. We explain this in Section 6 below, and then go on to use it to produce confidence intervals for λ . These intervals are easily implemented, and Monte Carlo simulation suggests that they have very accurate coverage rates, even under conditions which might be expected to produce difficulties (for example, incidental parameters, a dense spatial design matrix W , non-symmetric distribution).

The rest of the paper is organized as follows. Section 2 discusses some assumptions, and introduces some examples that are used to illustrate the theoretical results. Section 3 derives the key properties of the profile log-likelihood for λ . Section 4 presents the representation of the cdf of the QMLE, and discusses a number of consequences. Section 5 applies the main results first to the case of a Gaussian pure SAR model with symmetric W , and then to the examples introduced earlier. Section 6 presents and evaluates the saddlepoint approximation that emanates from the main result, and Section 7 concludes. Proofs are relegated to either Appendix B or the online Supplementary Material. The Supplementary Material also contains various technical material related to the paper.

2 Assumptions and examples

2.1 The parameter space for λ

In order for model (1.1) to uniquely determine the vector y (given $X\beta$ and ε) it is necessary and sufficient that the matrix $S_\lambda := I_n - \lambda W$ is nonsingular, or, equivalently, that $\lambda \neq \omega^{-1}$, for all nonzero real eigenvalues ω of W .³ This we assume throughout, but in practice the parameter space for λ is usually restricted much further, as we discuss next. Throughout the paper, we maintain the following assumption.

Assumption A. W has at least one negative eigenvalue and at least one positive eigenvalue.

³The condition that S_λ is nonsingular is equivalent to the condition that $1 - \lambda\omega \neq 0$ for all eigenvalues ω of W , which in turn is equivalent to $\lambda \neq \omega^{-1}$, for all nonzero real eigenvalues ω of W , because λ is assumed to be real and ω^{-1} is real if and only if ω is.

Assumption A requires W not to be nilpotent (a square matrix is nilpotent if all its eigenvalues are zero). If W were nilpotent the study of the QMLE would be trivial, as in that case maximizing the Gaussian likelihood would be equivalent to minimizing the residual sum of squares $(S_\lambda y - X\beta)'(S_\lambda y - X\beta)$, and hence the QMLE would coincide with the OLS estimator.⁴ The additional requirement that two of the nonzero real eigenvalues of W have opposite sign, is mainly made for simplicity, and is in any case virtually always satisfied in applications when W is non-nilpotent. It only rules out cases when W has no real negative or no real positive eigenvalues, which is used to ensure that the parameter space defined below is bounded.

Given Assumption A, we normalize, without loss of generality, the largest real eigenvalue of W to be equal to 1, and we denote the smallest real eigenvalue of W by ω_{\min} . The interval $\Lambda := (\omega_{\min}^{-1}, 1)$ is the largest interval containing the origin in which S_λ is nonsingular. Either Λ or a subset thereof is, implicitly or explicitly, virtually always regarded as the relevant parameter space for λ (see, e.g., Lee, 2004, and Kelejian and Prucha, 2010), and we do so here. The estimator we study in this paper is thus the maximizer of the (quasi) likelihood over Λ .

2.2 Examples

To illustrate our results the following examples will be used throughout the paper, and in particular in Section 5. The examples are chosen for their simplicity and their importance in the literature. In the first example W has full rank, while in the second example W has rank two (the minimum possible, given Assumption A).

Example 1 (Group Interaction Model). The relationships between a group of m members, all of whom interact uniformly with each other, may be represented by a matrix whose elements are all unity except for a zero diagonal. When normalized so that its row sums are unity, such a matrix has the form $B_m := (m - 1)^{-1}(\iota_m \iota_m' - I_m)$, where ι_m denotes an $m \times 1$ vector of ones. Suppose there are r such groups, of sizes $m_1 \leq m_2 \leq \dots \leq m_r$, and there are no between-group interactions. This is a fundamental configuration in social network analysis. We refer to the SAR model with block-diagonal spatial weights matrix

$$W = \text{diag}(B_{m_i}, i = 1, \dots, r) \quad (2.1)$$

as the *Group Interaction model* (see, e.g., Case, 1992, Kelejian et al., 2006, Lee, 2007, Davezies et al., 2009, Carrell et al., 2013, and Boucher et al., 2014). It is sometimes also known as the districts model (Kyriakou et al., 2017). For this model $\Lambda = (-(m_1 - 1), 1)$. We say that the model is *balanced* if the groups are of the same size, *unbalanced* otherwise.⁵

Example 2 (Complete Bipartite Model). In a *complete bipartite graph* the n observational units are partitioned into two groups of sizes p and q , say, with all individuals within a group interacting with all in the other group, but with none in their own group (e.g., Bramoullé et al., 2009, Lee et al., 2010). For $p = 1$ or $q = 1$ this corresponds to

⁴If W is nonnegative, as it is usually the case in applications, then it is nilpotent if and only if there is a permutation of the observational units that makes W triangular, i.e., makes the autoregressive process unilateral (see Martellosio, 2011).

⁵In the unbalanced case, particularly, it may be more realistic to assume a different autoregressive parameter for each group, but this generalization will not be considered in the present paper.

the graph known as a *star*, a particularly important case in social network theory (e.g., Jackson, 2008). The adjacency matrix of a complete bipartite graph is

$$A := \begin{bmatrix} 0_{pp} & \iota_p \iota'_q \\ \iota_q \iota'_p & 0_{qq} \end{bmatrix}.$$

The corresponding row-standardized weights matrix is

$$W = \begin{bmatrix} 0_{pp} & \frac{1}{q} \iota_p \iota'_q \\ \frac{1}{p} \iota_q \iota'_p & 0_{qq} \end{bmatrix}. \quad (2.2)$$

Alternatively, A can be rescaled by its largest eigenvalue, yielding the symmetric weights matrix

$$W = \frac{1}{\sqrt{pq}} A. \quad (2.3)$$

We refer to the SAR model with weights matrix (2.2) or (2.3), as, respectively, the *row-standardized Complete Bipartite model* and the *symmetric Complete Bipartite model*. In both cases, W has eigenvalues $-1, 0$, and 1 , so $\Lambda = (-1, 1)$.

3 The profile log-likelihood

Quasi-maximum likelihood of the parameters in model (1.1) is based on the log-likelihood obtained under the assumption $\varepsilon \sim N(0, I_n)$. For any λ such that S_λ is nonsingular, this log-likelihood is

$$l(\beta, \sigma^2, \lambda) := -\frac{n}{2} \log(\sigma^2) + \log(|\det(S_\lambda)|) - \frac{1}{2\sigma^2} (S_\lambda y - X\beta)'(S_\lambda y - X\beta), \quad (3.1)$$

where additive constants have been omitted. Maximizing $l(\beta, \sigma^2, \lambda)$ with respect to β and σ^2 for fixed λ yields the values

$$\hat{\beta}_{\text{ML}}(\lambda) := (X'X)^{-1}X'S_\lambda y, \quad \hat{\sigma}_{\text{ML}}^2(\lambda) := \frac{1}{n} y'S'_\lambda M_X S_\lambda y, \quad (3.2)$$

and the profile, or concentrated, log-likelihood

$$l(\lambda) := -\frac{n}{2} \log(y'S'_\lambda M_X S_\lambda y) + \log(|\det(S_\lambda)|), \quad (3.3)$$

where $M_X := I_n - X(X'X)^{-1}X'$. The profile log-likelihood function $l(\lambda)$ is a.s. differentiable for any λ such that $\det(S_\lambda) \neq 0$, and the profile score is

$$s(\lambda) := n \left(\frac{y'W'M_X S_\lambda y}{y'S'_\lambda M_X S_\lambda y} - \bar{g}(\lambda) \right), \quad (3.4)$$

where $G_\lambda := WS_\lambda^{-1}$ and $\bar{g}(\lambda) := \text{tr}(G_\lambda)/n$.⁶

We can now define the QMLE of λ precisely. Recall that the condition that S_λ is nonsingular is equivalent to $\lambda \neq \omega^{-1}$, for all nonzero real eigenvalues ω of W . That is, the function $l(\lambda)$ is a.s. defined for λ on the whole real line with the exception of a finite

⁶Throughout the paper, a.s. stands for almost surely with respect to Lebesgue measure on \mathbb{R}^n .

number of isolated points (and the unrestricted maximizer of $l(\lambda)$ can, in general, be anywhere on this set). The estimator we consider in this paper is

$$\hat{\lambda}_{\text{ML}} := \arg \max_{\lambda \in \Lambda} l(\lambda). \quad (3.5)$$

Issues of existence and uniqueness are addressed below. This is the QMLE in most common use, but of course it might not be the QMLE under a different specification of the parameter space for λ . In particular, several authors have suggested that λ should be restricted to $(-1, 1)$, which is convenient for interpretation of λ and because it does not depend on n (see, e.g., Kelejian and Prucha, 2010). Note that if W is nonnegative, $(-1, 1) \subseteq \Lambda$ by the Perron-Frobenius Theorem. If $(-1, 1)$ is a proper subset of Λ , the estimator $\tilde{\lambda}_{\text{ML}} := \arg \max_{\lambda \in (-1, 1)} l(\lambda)$ is a censored version of $\hat{\lambda}_{\text{ML}}$. That is, $\Pr(\tilde{\lambda}_{\text{ML}} = -1) = \Pr(\hat{\lambda}_{\text{ML}} < -1)$, and $\Pr(\tilde{\lambda}_{\text{ML}} < z) = \Pr(\hat{\lambda}_{\text{ML}} < z)$, for any $z \in (-1, 1)$, and it is clear that the results for $\hat{\lambda}_{\text{ML}}$ given below induce those for $\tilde{\lambda}_{\text{ML}}$.

For future reference, we rewrite the profile score as

$$s(\lambda) = \frac{n}{2} \frac{y' S'_\lambda Q_\lambda S_\lambda y}{y' S'_\lambda M_X S_\lambda y}, \quad (3.6)$$

where

$$C_\lambda := G_\lambda - \bar{g}(\lambda) I_n, \quad (3.7)$$

and Q_λ is the symmetric matrix

$$Q_\lambda := M_X C_\lambda + C'_\lambda M_X. \quad (3.8)$$

We will also need the following notation. Let N denote the number of distinct eigenvalues of W . If the distinct eigenvalues of W are all real we denote them by, in ascending order, $\omega_1, \omega_2, \dots, \omega_N$, the eigenvalue ω_i occurring with algebraic multiplicity n_i (so that $\omega_1 = \omega_{\min}$, $\omega_N = 1$). The eigenvalues of G_λ and C_λ are then, under the assumption that the eigenvalues of W are all real, respectively $g_i(\lambda) := \omega_i / (1 - \lambda \omega_i)$ and $\gamma_i(\lambda) := g_i(\lambda) - \bar{g}(\lambda)$, for $i = 1, \dots, N$.

3.1 Existence of the QMLE

Before embarking on a study of the properties of $\hat{\lambda}_{\text{ML}}$ it is prudent to check that it exists, i.e., that the profile log-likelihood is bounded above on Λ , and, if it exists, that it is not trivial, i.e., that it depends on the data y . Perhaps unexpectedly, there are combinations of the matrices W and X for which neither of these is true (even for large n), some of which have been noticed by others (see, e.g., Lee, 2007). A complete discussion is omitted for brevity, but details and examples can be found in Section S.1 of the Supplementary Material. In the rest of the paper, we assume that the QMLE exists a.s. and is not trivial (that is, we rule out the pathological cases described in Proposition S.1.1). Note that if it exists a.s., $\hat{\lambda}_{\text{ML}}$ is unique a.s., because, as can be easily verified, $l(\lambda)$ identifies λ .

3.2 The likelihood equation

Since Λ is an open set, and it can be shown that the function $l(\lambda)$ is a.s. differentiable on Λ , $\hat{\lambda}_{\text{ML}}$ must, subject to existence, be a root of the likelihood equation $s(\lambda) = 0$. The fact that $\hat{\lambda}_{\text{ML}}$ is (usually) only implicitly defined by this equation is a consequence of the following result.

Lemma 1. *The likelihood equation $s(\lambda) = 0$ is a.s. equivalent to a polynomial equation of degree equal to the number of distinct eigenvalues of W .*

According to Lemma 1, the equation $s(\lambda) = 0$ has a number of complex roots (counting multiplicities) equal to the number of distinct eigenvalues of W . Of these, any real root in Λ is a candidate for $\hat{\lambda}_{\text{ML}}$. Since there is no explicit algebraic solution of polynomial equations of degree higher than four, Lemma 1 explains why $\hat{\lambda}_{\text{ML}}$ cannot in general be obtained in “closed form”. In spite of this, we shall see in the next section that the cdf of $\hat{\lambda}_{\text{ML}}$ admits a very simple representation in terms of the cdf of a certain quadratic form. The following result is the key to that representation.

Lemma 2. *If $\text{tr}(C_\lambda^2) > 0$ for all $\lambda \in \Lambda$, the likelihood equation $s(\lambda) = 0$ a.s. has a single solution in Λ , which corresponds to the maximum of $l(\lambda)$ on Λ .*

Geometrically, Lemma 2 says that, under the stated condition, *the profile log-likelihood $l(\lambda)$ is a.s. single-peaked* on Λ , with no stationary inflection points. Although not our main focus here, this single-peakedness result has significant implications for the computation of the QMLE, a subject that has itself attracted a considerable literature (see LeSage and Pace, 2009, Chapters 3 and 4). The main advantage is that it guarantees that an optimization algorithm searching for $\hat{\lambda}_{\text{ML}}$ will not settle on a local, rather than global, maximum, thus avoiding the need to start the algorithm on a grid of different initial values.

Note that the condition “ $\text{tr}(C_\lambda^2) > 0$ for all $\lambda \in \Lambda$ ” depends only on W , not on X . Importantly, it is satisfied whenever W has *only real eigenvalues*, which is often the case in applications.⁷ For example, all eigenvalues of W are real when W is the row-standardized version of a symmetric matrix, or, more generally, when it is similar to a symmetric matrix. It seems difficult to provide a simple characterization of the class of matrices W for which $\text{tr}(C_\lambda^2) > 0$ for all $\lambda \in \Lambda$, but, for any given W , one can check the condition graphically. The following example provides some evidence that the condition is considerably more general than the requirement that all eigenvalues of W are real.

Example 3. Consider the weights matrix W obtained by row-standardizing the band matrix

$$A = \begin{bmatrix} 0 & a_3 & a_4 & 0 & \cdots \\ a_1 & 0 & a_3 & a_4 & \\ a_2 & a_1 & 0 & a_3 & \\ 0 & a_2 & a_1 & 0 & \\ \vdots & & & & \ddots \end{bmatrix},$$

for fixed a_1, a_2, a_3, a_4 . If $a_1 = a_3$ and $a_2 = a_4$, all the eigenvalues of W are real and therefore $l(\lambda)$ is a.s. single-peaked by Lemma 2. Other configurations of the a_i can induce multi-peakedness of $l(\lambda)$. To see this, fix $n = 20$, $a_1 = a_2 = a_3 = 1$, and consider values of a_4 in $[0, 1]$. For a_4 larger than about 0.55 and not too close to 1, the eigenvalues of W are not all real. However, for any a_4 larger than about 0.55, $\text{tr}(C_\lambda^2) > 0$ for all $\lambda \in \Lambda$, and hence $l(\lambda)$ is a.s. single-peaked by Lemma 2. For smaller values of a_4 $\text{tr}(C_\lambda^2)$ is not positive for all $\lambda \in \Lambda$, and there is a positive probability (as y ranges over \mathbb{R}^n)

⁷If all eigenvalues of W are real then all eigenvalues of C_λ are real, and hence all eigenvalues of C_λ^2 are nonnegative, which implies $\text{tr}(C_\lambda^2) \geq 0$. But $\text{tr}(C_\lambda^2) = 0$ is impossible given Assumption A, because it requires all eigenvalues of C_λ to be zero, which is the case if and only if G_λ , and hence W , is a scalar multiple of I_n .

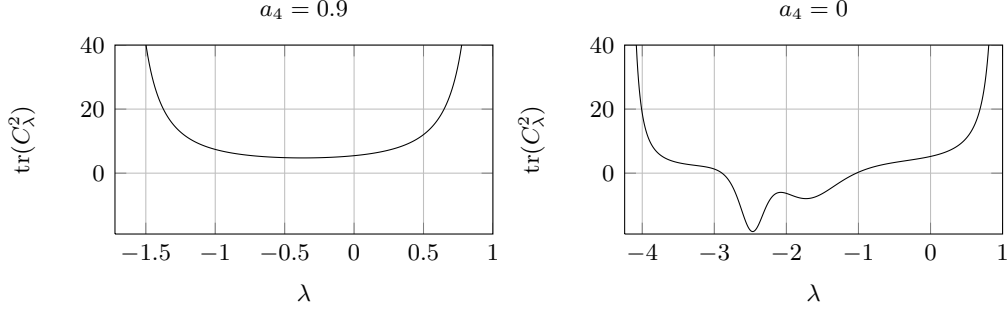


Figure 1: $\text{tr}(C_\lambda^2)$, for $\lambda \in \Lambda$, in the case of the weights matrix W in Example 3.

that $l(\lambda)$ is multi-peaked. Figure 1 displays $\text{tr}(C_\lambda^2)$ when $a_4 = 0.9$ (left panel) and $a_4 = 0$ (right panel). Note that Λ depends on a_4 . Simulation suggests that, whatever the value of X , $a_4 = 0$ entails a high probability of multi-peakedness

We stress that the single-peakedness property implied by Lemma 2 holds for any instance of model (1.1), including panel versions (i.e., W block-diagonal and X possibly containing fixed effects), and models including spatially lagged regressors WX . A complete understanding of what may cause multi-peakedness when the condition in Lemma 2 is not satisfied is beyond the scope of this paper, but the next result is a first step in that direction.

Lemma 3. *If W is nonnegative, the likelihood equation $s(\lambda) = 0$ a.s. has at most one solution in $[0, 1)$, which corresponds to a local maximum of $l(\lambda)$.*

That is, provided that W is nonnegative (which is virtually always the case in applications), if $l(\lambda)$ has a peak on the open interval $(0, 1)$, then it is single-peaked on that interval. In other words, when W is nonnegative, multi-peakedness must always involve peaks at negative values of λ . This result is of interest for applications in which it may be natural to restrict attention to positive values of λ . Some further remarks on Lemma 2 are collected in Section S.2 of the Supplementary Material.

4 Distribution properties of the QMLE

4.1 The main result

The single-peaked property established in Lemma 2 means that $\hat{\lambda}_{\text{ML}}$ is to the left of a point $z \in \Lambda$ if and only if the slope of $l(\lambda)$ at $\lambda = z$ is negative. This observation leads to the following result.

Theorem 1. *If $\text{tr}(C_\lambda^2) > 0$ for all $\lambda \in \Lambda$, the cdf of $\hat{\lambda}_{\text{ML}}$ at any point $z \in \Lambda$ is given by*

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr(y'S'_z Q_z S_z y \leq 0). \quad (4.1)$$

Theorem 1 provides an explicit representation of the cdf of $\hat{\lambda}_{\text{ML}}$, for any W such that $\text{tr}(C_\lambda^2) > 0$, for all $\lambda \in \Lambda$, for any X , and, importantly, for any distribution of y (not necessarily that induced by the SAR model). The result reduces the study of the properties of $\hat{\lambda}_{\text{ML}}$, an estimator that is generally unavailable in closed form, to the study of the properties of the family of quadratic forms $y'S'_z Q_z S_z y$, for $z \in \Lambda$.

There are a number of consequences of this simple result, only some of which will be pursued here. *First*, subject to suitable conditions on (W, X) , the first-order asymptotic distribution of $\hat{\lambda}_{\text{ML}}$ can be obtained directly from Theorem 1 by an application of the results in Kelejian and Prucha (2001) on the asymptotic distribution of quadratic forms. The first-order asymptotic behavior of the QMLE has been comprehensively studied by Lee (2004), using a related methodology, and hence will not be discussed further in this paper.

Second, in cases where the matrix $S_z' Q_z S_z$ of the quadratic form is sufficiently simple, the result can deliver an explicit formula for the cdf of $\hat{\lambda}_{\text{ML}}$. Details for some interesting examples will be given in Section 5. In particular, for the pure model we are able to derive an expression for the exact density of $\hat{\lambda}_{\text{ML}}$ when W is symmetric and $\varepsilon \sim N(0, \sigma^2 I_n)$. But, except under strong restrictions on (W, X) , equation (4.1) does not produce general exact results. This is because, even under Gaussian assumptions on ε in (1.1), distribution theory for quadratic forms is extremely complex, and closed-form formulae for distributions and densities are available only in very special cases. However, because of their ubiquity in inference, this fact has itself motivated a very large literature on the problem of approximating the distributions of quadratic forms, and a *third* consequence of Theorem 1 is that we can appeal directly to that literature for methods that provide simple higher-order asymptotic approximations to the distribution of $\hat{\lambda}_{\text{ML}}$. In Section 6 below we derive a saddlepoint-based approximation to the distribution of $\hat{\lambda}_{\text{ML}}$ and use it to construct confidence intervals for λ . One could, alternatively, use existing results on Edgeworth expansions for the distribution of quadratic forms (cf., Konishi, Niki, and Gupta, 1986, and Robinson and Rossi, 2015), but such expansions are likely to perform less satisfactorily in the tails of the distribution.

It is worth noting that Theorem 1 provides a straightforward way to obtain the cdf of $\hat{\lambda}_{\text{ML}}$ numerically, for *any* completely specified distribution of y (not necessarily that induced by the SAR model). Indeed, using equation (4.1), the whole cdf can be computed very efficiently by simply simulating a quadratic form and counting the proportion of negative realizations, *without the need to maximize the likelihood* at each repetition. This is useful, for example, to study by simulation how the model characteristics (parameters, design matrices W and X , and distribution of ε) affect the distribution of $\hat{\lambda}_{\text{ML}}$. Theorem 1 also simplifies the task of bootstrapping the distribution of $\hat{\lambda}_{\text{ML}}$, again without the need to maximize the likelihood ever. The saddlepoint approach discussed in Section 6 can be viewed as an alternative to bootstrapping the distribution (cf. Butler, 2007, Chapter 14).

Three further (exact) consequences of Theorem 1 are important, and are discussed next. The first is the possibility – exposed by the representation given in Theorem 1 – that the support of the estimator $\hat{\lambda}_{\text{ML}}$ (i.e., the set on which the density of $\hat{\lambda}_{\text{ML}}$ is positive) *may not be the entire interval* Λ . The second is that Theorem 1 implies certain important invariance properties of the estimator. Finally, known properties of quadratic forms imply that the (exact) distribution of $\hat{\lambda}_{\text{ML}}$ will typically have points of non-analyticity, and between such points will have a different functional form. We discuss these briefly in turn.

4.2 Support of the QMLE

The fact that the support of $\hat{\lambda}_{\text{ML}}$ is not necessarily the entire interval Λ is an unexpected phenomenon that has not been noticed previously, to the best of our knowledge. The

phenomenon occurs under certain conditions on the relation between W and X . While it seems difficult to specify general conditions on W and X that lead to restricted support for $\hat{\lambda}_{\text{ML}}$, special cases where this occurs are easily deduced. We give just the simplest of these below.

To begin with, observe that the likelihood equation $s(\lambda) = 0$ implies that the only possible candidates for the QMLE are the values of λ for which the matrix Q_λ is indefinite (see equation (3.6)). More decisively, Theorem 1 shows that if there are values of $z \in \Lambda$ for which Q_z is either positive or negative definite, those will either be impossible ($\Pr(\hat{\lambda}_{\text{ML}} \leq z) = 0$), or certain ($\Pr(\hat{\lambda}_{\text{ML}} \leq z) = 1$). In such cases the support of $\hat{\lambda}_{\text{ML}}$ is a proper subset of Λ . This cannot happen for the pure SAR model, because in that case $Q_z = (G_z + G'_z) - n^{-1} \text{tr}(G_z + G'_z) I_n$, which is necessarily indefinite (since $n^{-1} \text{tr}(G_z + G'_z)$ is the average of the eigenvalues of $G_z + G'_z$). But, when regressors are introduced, there can be choices for (W, X) for which $\hat{\lambda}_{\text{ML}}$ is not supported on the whole interval Λ . The following result illustrates this. For simplicity, the result is based on the assumption that y is supported on the whole of \mathbb{R}^n . For $i = 2, \dots, N - 1$, z_i denotes the point $z \in \Lambda$ at which $\gamma_i(z) = 0$ (existence and uniqueness of each of these points is established by Lemma A.1 in Appendix A), and $\text{col}(A)$ denotes the column space of a matrix A .

Proposition 1. *Assume that in a SAR model W is similar to a symmetric matrix and $M_X W$ is symmetric.*

- (i) *If, for some $i = 2, \dots, N - 1$, $\text{col}(X)$ contains all eigenvectors of W associated to the eigenvalues ω_j with $j > i$, then the support of $\hat{\lambda}_{\text{ML}}$ is $(\omega_{\min}^{-1}, z_i)$.*
- (ii) *If, for some $i = 2, \dots, N - 1$, $\text{col}(X)$ contains all eigenvectors of W associated to the eigenvalues ω_j with $j < i$, then the support of $\hat{\lambda}_{\text{ML}}$ is $(z_i, 1)$.*

We now provide some intuition, and some examples, for Proposition 1. One implication of the result is that $\hat{\lambda}_{\text{ML}}$ cannot be positive if $\text{col}(X)$ contains all eigenvectors of W associated to positive eigenvalues (even if the true value of λ is positive).⁸ Now, the eigenvectors of W associated to positive eigenvalues can be interpreted as capturing all positive spatial autocorrelation (as measured by the statistic $u' W u / u' u$) in a zero-mean process u . Also, $\hat{\lambda}_{\text{ML}}$ can be thought of as a measure of the autocorrelation remaining in y after conditioning on the regressors. Hence, Proposition 1 admits the intuitive interpretation that the autocorrelation remaining after conditioning on all eigenvectors of W associated to positive eigenvalues can only be negative. An example of this effect arises in the unbalanced Group Interaction model with group fixed effects, and all β coefficients varying across groups (see Section S.3.3 in the Supplementary Material). In this model, the fixed effects span the eigenspace of W associated to the eigenvalue 1, and 1 is the only positive eigenvalue of W . Hence, in this model $\hat{\lambda}_{\text{ML}}$ can never be positive. Another example of the restricted support phenomenon, in the context of a Complete Bipartite model, will be given in Section 5.3.2.

It is worth remarking that if the support of $\hat{\lambda}_{\text{ML}}$ is restricted, then (i) asymptotic approximations to its distribution that are supported on the entire interval Λ are unlikely to be satisfactory; (ii) $\hat{\lambda}_{\text{ML}}$ clearly cannot be consistent uniformly over λ .

⁸This is because, in that case, z_i in Proposition 1(i) must be nonpositive, by Lemma A.1 in Appendix A and the fact that $\gamma_i(0) = \omega_i \leq 0$.

4.3 Invariance properties

Two invariance properties of the distribution of $\hat{\lambda}_{\text{ML}}$ follow from the score equation (3.4), and are evident in representation (4.1). These derive from the fact that the distribution of $\hat{\lambda}_{\text{ML}}$ is invariant under the group of scale transformations $y \rightarrow \kappa y$. Firstly, assuming that the distribution of ε does not depend on β and σ^2 , this type of scale invariance implies that the distribution of $\hat{\lambda}_{\text{ML}}$ depends on (β, σ) only through β/σ . Thus, in particular, in the pure model the distribution of $\hat{\lambda}_{\text{ML}}$ does not depend on σ^2 . For details, and for a proof that, under certain special conditions on (W, X) , the dependence on parameters may be further reduced, see Section S.3.1 of the Supplementary Material.

A second important consequence of this scale invariance, generalizing the above, is that κ in the transformation $y \rightarrow \kappa y$ may be interpreted as a random variable (independent of y), which leads to the following result.⁹

Proposition 2. *The distribution of $\hat{\lambda}_{\text{ML}}$ induced by a particular distribution of y is the same for all scale mixtures of that distribution.*

This observation means, in particular, that any property of the distribution of $\hat{\lambda}_{\text{ML}}$ that holds under the assumption $y \sim N(S_\lambda^{-1}X\beta, \sigma^2(S'_\lambda S_\lambda)^{-1})$ (a Gaussian SAR model) also holds under every member of the much larger family of scale mixtures of that distribution for y . Assuming normality, as we shall do to derive exact analytical results shortly, is therefore less restrictive than it would usually be. Also note that if it does not depend on σ , the distribution of $\hat{\lambda}_{\text{ML}}$ is invariant to mixing the distribution of ε (see the Supplementary Material).

4.4 Points of non-analyticity

Quadratic forms have been much studied in the statistical and econometric literature. As already remarked, their distribution theory is generally complicated, but some general distributional properties of $\hat{\lambda}_{\text{ML}}$ follow from that literature. One important property in this category is that, in general, the cdf of $\hat{\lambda}_{\text{ML}}$ may fail to be analytic everywhere on Λ . In Section S.3.2 of the Supplementary Material we show that, provided that the distribution of y is sufficiently smooth, the cdf of $\hat{\lambda}_{\text{ML}}$ is analytic everywhere on Λ except at the points of discontinuity of $\text{rank}(Q_z)$. These points depend only on W and X , not on the distribution of y , or on the parameters of the model. In general, Λ contains several nonanalyticity points, their number depending on W and X . In the particular case of a pure SAR with symmetric W , however, the number of nonanalyticity points is always $N - 2$, where, recall, N denotes the number of distinct eigenvalues of W .¹⁰ Examples will be given in Section 5. It is worth remarking that in some cases the non-analyticity persists asymptotically, the Complete Bipartite model being one example (see Section 5.3.1).

5 Some exact analytical results

In this section we study exact properties of $\hat{\lambda}_{\text{ML}}$ under the assumption, common in exact and higher-order asymptotic analysis, of normality. More precisely, we assume that the

⁹The family of scale mixtures of a density $p(y)$ for y (assuming it exists) is defined by $\int_\kappa p(\kappa y)g(\kappa) d\kappa$, where $g(\kappa)$ varies over the set of (for simplicity) proper densities for κ .

¹⁰This is because if W is symmetric and there are no regressors, $Q_z = 2C_z$ so by Lemma A.1 there are $N - 2$ distinct values of z where one of the eigenvalues of Q_z vanishes.

distribution of y is that induced by a SAR model with $\varepsilon \sim N(0, I_n)$. Recall from Section 4.3 that any result obtained for $\hat{\lambda}_{\text{ML}}$ under normality also holds under scale mixtures of the distribution of y (or of ε). We use the notation $\text{SMN}(0, I_n)$ to represent the family of scale mixtures of $N(0, I_n)$.

From Theorem 1 we have

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr(\tilde{y}' A(z, \lambda) \tilde{y} \leq 0), \quad (5.1)$$

where $\tilde{y} := S_\lambda y$, and

$$A(z, \lambda) := (S_z S_\lambda^{-1})' Q_z (S_z S_\lambda^{-1}). \quad (5.2)$$

The structure of the matrix $A(z, \lambda)$ is evidently crucial in determining the properties of $\hat{\lambda}_{\text{ML}}$. In particular, a spectral decomposition of $A(z, \lambda)$ shows that, when $\tilde{y} \sim N(X\beta, \sigma^2 I_n)$, $\tilde{y}' A(z, \lambda) \tilde{y}$ is distributed as a linear combination of independent, possibly non-central, χ^2 variates, with coefficients the distinct eigenvalues of $A(z, \lambda)$. This familiar approach delivers exact results that are of interest in illustrating the various aspects of the distribution of $\hat{\lambda}_{\text{ML}}$ we have mentioned. In Section 5.1 we obtain the exact distribution for the case of a pure model with symmetric W . This provides a clear explanation of the non-analyticity phenomenon. Then, in Sections 5.2 and 5.3, we turn to the two examples of Section 2.2: the balanced Group Interaction model and the Complete Bipartite model.¹¹ These illustrate both the non-analyticity and restricted support phenomena. They also provide examples of cases in which the asymptotic properties of $\hat{\lambda}_{\text{ML}}$ are non-standard.

5.1 Pure SAR model with symmetric W

If the mean component $X\beta$ in equation (1.1) is missing, and W is symmetric, the eigenstructure of the matrix of the quadratic form in (4.1) simplifies dramatically. Let W be a symmetric weights matrix with spectral decomposition $W = HDH'$, where D is diagonal with the eigenvalues of W on the diagonal and H is an orthonormal matrix of eigenvectors. Then, recalling that $\omega_1, \dots, \omega_N$ denote the distinct eigenvalues of W , and n_i their respective multiplicities, the matrix $A(z, \lambda)$ in equation (5.2) has the spectral decomposition

$$A(z, \lambda) = H \text{diag}(d_{ii}(z, \lambda) I_{n_i}, i = 1, \dots, N) H', \quad (5.3)$$

where

$$d_{ii}(z, \lambda) := \gamma_i(z) \left(\frac{1 - z\omega_i}{1 - \lambda\omega_i} \right)^2, \quad i = 1, \dots, N, \quad (5.4)$$

are the eigenvalues of $A(z, \lambda)$ (recall that the $\gamma_i(z)$'s are the eigenvalues of the matrix C_λ defined in equation (3.7)). Thus, the quadratic form $\tilde{y}' A(z, \lambda) \tilde{y}$ reduces to a weighted sum of independent $\chi_{n_i}^2$ variates (here and elsewhere, χ_ν^2 denotes a central χ^2 random variable with ν degrees of freedom). Thus, Theorem 1 has the following consequence:¹²

¹¹By Lemma 1, in the balanced Group Interaction model and in the Complete Bipartite model, $\hat{\lambda}_{\text{ML}}$ is the unique root in Λ of a quadratic and a cubic, respectively, and is therefore available in closed form. However, obtaining the exact distribution from such a closed form seems difficult. Theorem 1 provides a much more convenient approach.

¹²A more general version of Proposition 3 can be obtained under the condition that $\text{col}(X)$ is spanned by k linearly independent eigenvectors of W ; see equation (S.3.5) in the Supplementary Material.

Proposition 3. *In a pure SAR model with symmetric W and $\varepsilon \sim \text{SMN}(0, I_n)$, for any $z \in \Lambda$,*

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr\left(\sum_{i=1}^N d_{ii}(z, \lambda) \chi_{n_i}^2 \leq 0\right), \quad (5.5)$$

where the $\chi_{n_i}^2$ variates are independent.

Proposition 3 shows that, under the stated conditions, the distribution of $\hat{\lambda}_{\text{ML}}$ depends on W only through its eigenvalues and their multiplicities. The properties of the coefficients $d_{ii}(z, \lambda)$ that occur in this expression are described in Lemma A.1 in Appendix A. That lemma establishes, in particular, that, for any $\lambda \in \Lambda$, $d_{11}(z, \lambda) < 0$ and $d_{NN}(z, \lambda) > 0$ for all $z \in \Lambda$, and, for each $i = 2, \dots, N-1$, $d_{ii}(z, \lambda)$ changes sign exactly once, at the point z_i where the eigenvalue $\gamma_i(z)$ of C_z changes sign. Thus, when $N = 2$ neither coefficient changes sign, but when $N > 2$ the number of positive and negative coefficients in (5.5) varies as z varies over Λ . The cdf is therefore defined piecewise, and is non-analytic at the points z_i , $i = 2, \dots, N-1$, where the pieces meet. Note that the points z_i depend on W alone.

The number N of distinct eigenvalues of W must be at least two, by Assumption A. If $N = 2$, equation (5.5) gives

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr\left(F_{n_1, n_2} \leq -\frac{n_2 d_{22}(z, \lambda)}{n_1 d_{11}(z, \lambda)}\right), \quad (5.6)$$

where F_{ν_1, ν_2} denotes a random variable with an F distribution on (ν_1, ν_2) degrees of freedom. Thus, when $N = 2$ the cdf is remarkably simple, and there is no point of non-analyticity in this case (cf. Section 4.4 above). The balanced Group Interaction model has this form.

Using the representation (5.5), it is possible to derive an expression for the cdf in each interval between the points z_i at which the coefficients $d_{ii}(z, \lambda)$ change sign. Such an expression generalizes equation (5.6) to the case $N \geq 2$. To do so, some new notation is needed. For fixed $z \in \Lambda \setminus \{z_2, \dots, z_{N-1}\}$, define $v_1 := \sum_{i=1}^N n_i 1_{d_{ii}(z, \lambda) < 0}$ and $v_2 := n - v_1$ ($1_{\mathcal{K}}$ being the indicator function of the condition \mathcal{K}). The numbers (v_1, v_2) vary with z but not with λ . Next, denote by A_1 the $v_1 \times v_1$ matrix $\text{diag}(-d_{ii}(z, \lambda) I_{n_i}, i \in \{i : d_{ii}(z, \lambda) < 0\})$, and by A_2 the $v_2 \times v_2$ matrix $\text{diag}(d_{ii}(z, \lambda) I_{n_i}, i \in \{i : d_{ii}(z, \lambda) > 0\})$. Finally, $(a)_j := a(a+1)\dots(a+j-1)$ denotes the Pochhammer symbol, and $C_j(A)$ denotes the top-order zonal polynomial of order j in the eigenvalues of a matrix A (see Muirhead, 1982, Chapter 7).

Proposition 4. *In a pure SAR model with symmetric W and $\varepsilon \sim \text{SMN}(0, I_n)$, for z in the interior of any one of the $N-1$ intervals in Λ determined by the points of non-analyticity z_2, \dots, z_{N-1} ,*

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \sum_{j_1, j_2=0}^{\infty} c_{j_1}(\phi_1 A_1) c_{j_2}(\phi_2 A_2) \Pr\left(F_{v_2+2j_1, v_1+2j_2} \leq \frac{(v_1+2j_2)\phi_2}{(v_2+2j_1)\phi_1}\right), \quad (5.7)$$

for arbitrary coefficients $\phi_1, \phi_2 > 0$, and where, for $i = 1, 2$, and for $j = 0, 1, \dots$,

$$c_j(\phi_i A_i) := \frac{\left(\frac{1}{2}\right)_j C_j(I_{v_i} - (\phi_i A_i)^{-1})}{j! (\det(\phi_i A_i))^{\frac{1}{2}}}.$$

When ϕ_i satisfy $\phi_i A_i > I_{v_i}$ (entrywise), for $i = 1, 2$, all weights $c_{j_i}(\phi_i A_i)$ are non-negative and $\sum_{j_1, j_2=0}^{\infty} c_{j_1}(\phi_1 A_1) c_{j_2}(\phi_2 A_2) = 1$; that is, the cdf is a doubly-infinite mixture of F distribution functions.¹³

The weights in the doubly-infinite mixture representation are quite complicated polynomials in the coefficients that appear in (5.5), but efficient computation of the top-order zonal polynomials can be achieved by the methods discussed in Hillier, Kan, and Wang (2009). Even so, the exact expression in Proposition 4 is probably of limited use for inference. What is more useful is that Proposition 3 expresses the cdf of $\hat{\lambda}_{\text{ML}}$ as that of sum of independent random variables, and this is the situation in which the Lugannani-Rice approximation we will use later on can be expected to work well (see, e.g., Wood, Booth, and Butler, 1993).

Two further consequences of Proposition 3 are worth pointing out. First, based on equation (5.5), the cdf of $\hat{\lambda}_{\text{ML}}$ can be simulated very efficiently, for a given (possibly estimated) value of λ , without ever maximizing the likelihood; all that is needed is to simulate N independent χ^2 random variables. Second, representation (5.5) reveals the behavior of $\hat{\lambda}_{\text{ML}}$ as λ approaches the extremes of Λ . Indeed, dividing the sum $\sum_{i=1}^N d_{ii}(z, \lambda) \chi_{n_i}^2$ by $(1-z)^2/(1-\lambda)^2$, we obtain

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr\left(\gamma_N(z) \chi_{n_N}^2 + \sum_{i=1}^{N-1} \left(\frac{(1-z\omega_i)(1-\lambda)}{(1-\lambda\omega_i)(1-z)}\right)^2 \gamma_i(z) \chi_{n_i}^2 \leq 0\right). \quad (5.8)$$

As $\lambda \uparrow 1$, all terms in the sum in (5.8) go to zero for any $z \in \Lambda$, and hence, since $\gamma_N(z) > 0$ for any $z \in \Lambda$ by Lemma A.1 in Appendix A, $\Pr(\hat{\lambda}_{\text{ML}} \leq z; \lambda) \rightarrow 0$ for any $z \in \Lambda$. Similarly, $\Pr(\hat{\lambda}_{\text{ML}} \leq z; \lambda) \rightarrow 0$ for any $z \in \Lambda$ as $\lambda \downarrow \omega_1^{-1}$. That is, the distribution of $\hat{\lambda}_{\text{ML}}$ tends, as $\lambda \uparrow 1$ (resp., $\lambda \downarrow \omega_1^{-1}$), to a degenerate distribution with mass at 1 (resp., ω_1^{-1}).

More general versions of Proposition 3 can be derived for the general model (with X included) when W is assumed only to be similar to a symmetric matrix - see the Supplementary Material. Not surprisingly, these representations are complicated, although they do simplify under certain restrictive assumptions of (W, X) .

5.2 Balanced Group Interaction model

Equation (5.5) produces a strikingly simple result for the balanced Group Interaction model, because the weights matrix $W = I_r \otimes B_m$ has only two distinct eigenvalues, 1 with multiplicity r , and $-1/(m-1)$ with multiplicity $r(m-1)$. Using equation (5.5), we obtain the following expressions for the cdf and pdf of $\hat{\lambda}_{\text{ML}}$, where $c(z, \lambda) := \theta(z)/\theta(\lambda)$, with $\theta(z) := ((z+m-1)/(1-z))^2$, for any $z, \lambda \in \Lambda$, and a dot denotes differentiation with respect to z .

Proposition 5. *In the pure balanced Group Interaction model with $\varepsilon \sim \text{SMN}(0, I_n)$,*

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr(F_{r, r(m-1)} \leq c(z, \lambda)), \quad (5.9)$$

and, differentiating,

$$\text{pdf}_{\hat{\lambda}_{\text{ML}}}(z; \lambda) = \frac{\dot{c}(z, \lambda) c(z, \lambda)^{\frac{r}{2}-1}}{(m-1)^{\frac{r}{2}} B\left(\frac{r}{2}, \frac{r(m-1)}{2}\right)} \left(1 + \frac{c(z, \lambda)}{m-1}\right)^{-\frac{rm}{2}}, \quad (5.10)$$

¹³Expression (5.7) holds for arbitrary positive numbers ϕ_1 and ϕ_2 . However, if these coefficients are to satisfy the mixture conditions $\phi_i A_i > I_{v_i}$, $i = 1, 2$, they must depend on z , because one of the $d_{ii}(z, \lambda)$ in each matrix A_1 and A_2 goes to zero as z approaches one of the z_i .

for any $\lambda, z \in \Lambda$.

Proposition 5 enables a complete analysis of the exact properties of $\hat{\lambda}_{\text{ML}}$, and the results needed for inference based upon it. These are explored in detail in Hillier and Martellosio (2016), so we restrict ourselves here to a brief summary.

First, one has from equation (5.9) that $\Pr(\hat{\lambda}_{\text{ML}} \leq \lambda) = \Pr(F_{r,r(m-1)} \leq 1)$, which does not depend on λ . For $m = 2$ this is exactly .5, so $\hat{\lambda}_{\text{ML}}$ is median-unbiased when $m = 2$, but not otherwise. For $m > 2$ it can be shown that $\Pr(F_{r,r(m-1)} \leq 1) > .5$, so that $\Pr(\hat{\lambda}_{\text{ML}} \leq \lambda) > .5$, which means that the median of $\hat{\lambda}_{\text{ML}}$ is below the true value: there is a *negative median bias* when $m > 2$. The exact median of $\hat{\lambda}_{\text{ML}}$, and a median-unbiased estimator based on it, exact confidence intervals for λ based on $\hat{\lambda}_{\text{ML}}$, and expressions for the moments of $\hat{\lambda}_{\text{ML}}$, are all given in Hillier and Martellosio (2016), which also provides a detailed analysis of the unbalanced case and other generalizations. One particular case of interest is a balanced model with a constant mean, so that $X = \iota_n$. In this model, existence of the MLE requires $r > 1$ (cf. Section 3.1). Provided that $r > 1$, Proposition S.3.3 in the Supplementary Material yields, when applied to this model, the result

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr\left(F_{r-1,r(m-1)} \leq \frac{r}{r-1}c(z, \lambda)\right).$$

Next, equation (5.9) shows that, as is already known in the literature, $r \rightarrow \infty$ is sufficient for consistency of $\hat{\lambda}_{\text{ML}}$ (because then $F_{r,r(m-1)} \rightarrow_p 1$), but $m \rightarrow \infty$ may not be.¹⁴ Indeed, if $r \rightarrow \infty$ is assumed, Lee's (2004) Assumptions 3 and 8' are satisfied, as is his condition (4.3), so $\hat{\lambda}_{\text{ML}}$ is consistent and asymptotically normal by Lee's Theorems 4.1 and 4.2. On the other hand, if $n \rightarrow \infty$ because $m \rightarrow \infty$, Lee's Assumption 3 is *not* satisfied, and $\hat{\lambda}_{\text{ML}}$ may be inconsistent in this case. This case, that is not fully analyzed by Lee (2004), is an example of so-called infill asymptotics.

Equation (5.9), along with the known result $v_1 F_{v_1, v_2} \rightarrow_d \chi_{v_1}^2$ as $v_2 \rightarrow \infty$, shows that, for fixed r ,

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) \xrightarrow{m \rightarrow \infty} \Pr\left(\chi_r^2 \leq r \left(\frac{1-\lambda}{1-z}\right)^2\right), \quad -\infty < z < 1.$$

Thus, $\hat{\lambda}_{\text{ML}}$ is indeed inconsistent under infill asymptotics. The associated limiting density as $m \rightarrow \infty$ with r fixed is

$$\text{pdf}_{\hat{\lambda}_{\text{ML}}}(z; \lambda) \xrightarrow{m \rightarrow \infty} \frac{r^{\frac{r}{2}}(1-\lambda)^r}{2^{\frac{r}{2}-1}\Gamma(\frac{r}{2})(1-z)^{r+1}} e^{-\frac{r}{2}\left(\frac{1-\lambda}{1-z}\right)^2},$$

so $\hat{\lambda}_{\text{ML}}$ converges to a random variable supported on $(-\infty, 1)$.

Figure 2 displays the exact density (5.10) for $\lambda = 0.5$, and for $m = 10$ and various values of r (left panel), and for $r = 10$ and various values of m (right panel). For convenience the densities are plotted for $z \in (-1, 1) \subseteq \Lambda$. It is apparent that the density

¹⁴This is not unexpected: the partial information on λ in this model is

$$i_{\lambda\lambda.\sigma^2} = \frac{2rm(m-1)}{(1-\lambda)^2(\lambda+m-1)^2},$$

which increases without bound as $r \rightarrow \infty$, m fixed, but approaches the finite limit $2r/(1-\lambda)^2$ as $m \rightarrow \infty$ with r fixed.

is much more sensitive to r (the number of groups) than to m (the group size). Analogs of these plots for other positive values of λ exhibit similar characteristics (when λ is negative the density can be quite sensitive to m , mainly because the left extreme of the support of $\hat{\lambda}_{\text{ML}}$ depends on m). It is clear from Figure 2 that increasing m but not r provides very little extra information on λ , at least as embodied in the MLE, and that the effective sample size under this asymptotic regime is r , and *not* $n = rm$. However, with the simple exact result available, under (mixed-) Gaussian assumptions there is no need to invoke either form of asymptotic approximation in this model.

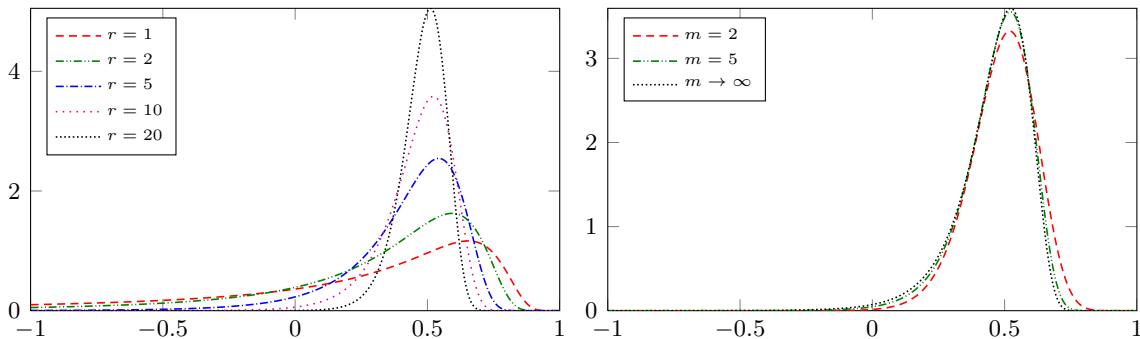


Figure 2: Density of $\hat{\lambda}_{\text{ML}}$ for the Gaussian pure balanced Group Interaction model with $\lambda = 0.5$, and with $m = 10$ (left panel), $r = 10$ (right panel).

5.3 The Complete Bipartite model

We now consider the Complete Bipartite model introduced in Section 2.2. We first discuss the pure model with symmetric W , for which a simple version of equation (5.7) applies. Then, in Section 5.3.2, we consider the case of row-standardized (therefore non-symmetric) W in a model with a constant mean (i.e., $X = \iota_n$). This latter case provides an example of the restricted support phenomenon described in Section 4.2.

5.3.1 Symmetric W , zero mean

In the symmetric Complete Bipartite model, W has eigenvalues 1 and -1 , each with multiplicity 1, and 0, with multiplicity $n - 2$. Thus, $N = 3$, and the parameter space is $\Lambda = (-1, 1)$. Several general properties of the distribution of $\hat{\lambda}_{\text{ML}}$ follow easily in this case. *First*, since the eigenvalues and their multiplicities depend on n , but not on the partition into (p, q) , Proposition 3 implies that the distribution of $\hat{\lambda}_{\text{ML}}$ does not depend on that partition. *Second*, in this model the linear combination of χ^2 variates in equation (5.5) becomes, after dropping irrelevant terms,

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z; \lambda) = \Pr(-\phi_1 \chi_1^2 - \phi_2 \chi_{n-2}^2 + \phi_3 \chi_1^2 \leq 0), \quad (5.11)$$

where $\phi_1 = \phi_1(z, \lambda) := (1 + \lambda)^{-2}(1 + z)^2(n - (n - 2)z)$, $\phi_2 = \phi_2(z) := 2z$, $\phi_3 = \phi_3(z, \lambda) := (1 - \lambda)^{-2}(1 - z)^2(n + (n - 2)z)$, and the three χ^2 random variables involved are independent. The coefficients ϕ_1 , ϕ_3 are positive for any $z \in (-1, 1)$, while ϕ_2 obviously changes sign at the origin. Thus, the distribution has a different functional form on each side of the origin, and is non-analytic at the origin. *Third*, for $z = 0$, $\Pr(\hat{\lambda}_{\text{ML}} \leq 0; \lambda) = \Pr(F_{1,1} \leq (1 - \lambda)^2 / (1 + \lambda)^2)$. Since this does not depend on n it

is also the limiting cdf at $z = 0$ as $n \rightarrow \infty$ (so the estimator cannot have standard asymptotic properties). Evidently, $\Pr(\hat{\lambda}_{\text{ML}} \leq 0; \lambda)$ is 0.5 if $\lambda = 0$, < 0.5 if $\lambda > 0$, and > 0.5 if $\lambda < 0$, so the estimator is median unbiased when $\lambda = 0$, and has positive (resp., negative) median when λ is positive (resp., negative). *Fourth*, for $z \in (0, 1)$ we have the expression $\Pr(\hat{\lambda}_{\text{ML}} \leq z; \lambda) = \Pr(\phi_3 \chi_1^2 \leq \phi_1 \chi_1^2 + \phi_2 \chi_{n-2}^2) =: F^+(z; \lambda)$, say. Noting that $\phi_3(z, \lambda) = \phi_1(-z, -\lambda)$, the corresponding expression for $z \in (-1, 0)$ is $\Pr(\hat{\lambda}_{\text{ML}} \leq z; \lambda) = 1 - F^+(-z; -\lambda)$, so we need only consider $F^+(z; \lambda)$.¹⁵ *Finally*, from the standard formula for the mean of a random variable X in terms of its cdf, $F(x)$ say,

$$E(X) = \int_0^\infty (1 - F(x)) dx - \int_{-\infty}^0 F(x) dx, \quad (5.12)$$

we have that

$$E(\hat{\lambda}_{\text{ML}}; \lambda) = \int_0^\infty (F^+(s; -\lambda) - F^+(s; \lambda)) ds. \quad (5.13)$$

It follows that

$$E(\hat{\lambda}_{\text{ML}}; -\lambda) = -E(\hat{\lambda}_{\text{ML}}; \lambda), \quad (5.14)$$

and hence that $\hat{\lambda}_{\text{ML}}$ is mean-unbiased when $\lambda = 0$, but may be biased otherwise.

Expression (5.11) can also be used to derive closed form expressions for the exact cdf and density of $\hat{\lambda}_{\text{ML}}$. For $z \in (0, 1)$,

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr(\chi_1^2 \leq \psi_1 \chi_1^2 + \psi_2 \chi_{n-2}^2), \quad (5.15)$$

with $\psi_1 := \phi_1/\phi_3$ and $\psi_2 := \phi_2/\phi_3$ both positive, from which an exact closed form solution for the cdf follows easily. The exact density is difficult to obtain by differentiation of the cdf, but an alternative strategy is available (that can also be applied in more general models): first condition on the random variables involved on the right hand side of the inequality $\chi_1^2 \leq \psi_1 \chi_1^2 + \psi_2 \chi_{n-2}^2$, giving the simple expression for the conditional cdf, $\Pr(\hat{\lambda}_{\text{ML}} \leq z | \chi_1^2, \chi_{n-2}^2) = G_1(\psi_1 \chi_1^2 + \psi_2 \chi_{n-2}^2)$, where G_v denotes the cdf of a χ_v^2 variate. This can be differentiated with respect to z to produce the conditional density given (χ_1^2, χ_{n-2}^2) , a relatively simple expression. The unconditional result then emerges by averaging with respect to the conditioning variates. The derivations for the exact cdf and density are given in Section S.4 of the Supplementary Material. Both are, even in this relatively simple model, quite complicated (as expected), and are difficult to use.

Motivated by the complexity of exact results, even in this simple model, one naturally looks for approximations that are more enlightening. The (large- n) asymptotic result, which again is not covered by Lee (2004), is easily deduced, as follows. For every fixed $z \in \Lambda$, the characteristic function of the random variable $V_n := (\phi_3 \chi_1^2 - 2z \chi_{n-2}^2 - \phi_1 \chi_1^2)/(n-2)$ is readily seen to converge to that of $\bar{V}_n := \bar{\phi}_3 \chi_1^2 - 2z - \bar{\phi}_1 \chi_1^2$, where $\bar{\phi}_1 := \lim_{n \rightarrow \infty} (\phi_1/(n-2)) = (1+z)^2(1-z)/(1+\lambda)^2$ and $\bar{\phi}_3 := \lim_{n \rightarrow \infty} (\phi_3/(n-2)) = (1-z)^2(1+z)/(1-\lambda)^2$. Therefore, $V_n \rightarrow_d \bar{V}_n$, and so, by Proposition 3,

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) \rightarrow \Pr(\bar{\phi}_3 \chi_1^2 - 2z - \bar{\phi}_1 \chi_1^2 \leq 0).$$

For $z \in (0, 1)$, therefore,

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) \rightarrow \Pr(\chi_1^2 \leq \bar{\psi}_1 \chi_1^2 + \bar{\psi}_2), \quad (5.16)$$

¹⁵The weights matrix W in this model is a simple example of a matrix with a *symmetric spectrum*, and the result just given holds for all such models - see Corollary S.3.1 in the Supplementary Material.

with

$$\bar{\psi}_1 := \left(\frac{1+z}{1-z} \right) \left(\frac{1-\lambda}{1+\lambda} \right)^2, \quad \bar{\psi}_2 := \frac{2z(1-\lambda)^2}{(1+z)(1-z)^2}.$$

The expression for $z \in (-1, 0)$ is obtained as for the exact case above. The limiting distribution is certainly non-standard, and has much in common with (though is simpler than) the exact distribution. In particular, the cdf has a different functional form for positive and negative z , and is non-analytic at the origin. And, as in the case where $m \rightarrow \infty$ in a balanced Group Interaction model, $\hat{\lambda}_{\text{ML}}$ is not consistent, but converges in distribution to a random variable as $n \rightarrow \infty$.¹⁶

An alternative, very simple, approach (attributed to Fisher) is to treat the term $\psi_1 \chi_1^2 + \psi_2 \chi_{n-2}^2$ in equation (5.15) as a multiple $\alpha \chi_v^2$ of a χ_v^2 variate, choosing α and v so that the means and variances of the true and approximate random variables coincide. This has been found to work very well when less than five terms are involved.¹⁷ It also works well in the present model - see the Supplementary Material for details. A third approximation based on saddlepoint methods that is much more generally applicable is introduced and discussed in the next section.

Figure 3 displays the exact density for $n = 5, 10$, and $n \rightarrow \infty$, and for $\lambda = -0.5, 0, 0.5$, using the exact expressions given in the Supplementary Material. It is clear from the plots that, similarly to the case when $m \rightarrow \infty$ in a balanced Group Interaction model, the density is very insensitive to the sample size, so in this model increasing the sample size yields little extra information about λ . As a consequence, the non-standard asymptotic density is an excellent approximation to the actual distribution under mixed-normal assumptions. The expected non-analyticity at $z = 0$ is evident, and in fact for this model the density of $\hat{\lambda}_{\text{ML}}$ is unbounded at $z = 0$.

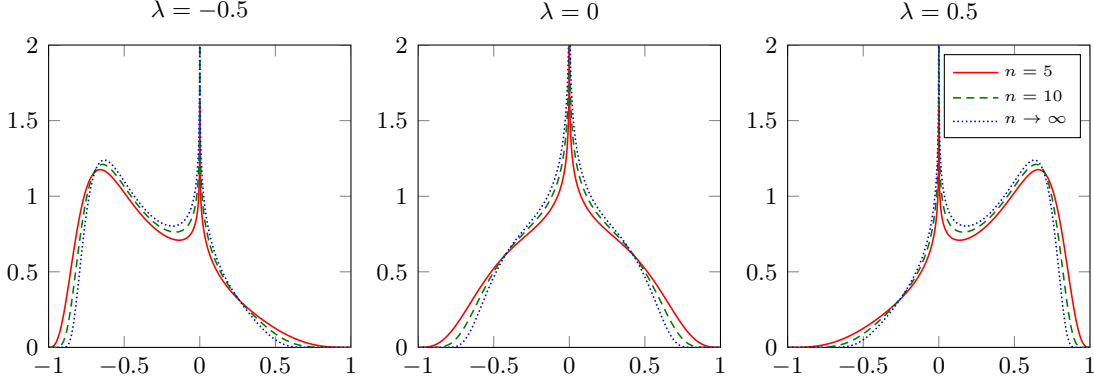


Figure 3: Density of $\hat{\lambda}_{\text{ML}}$ for the Gaussian pure symmetric Complete Bipartite model.

5.3.2 Row-standardized W , constant mean

In a Complete Bipartite model with row-standardized W and constant mean the support of $\hat{\lambda}_{\text{ML}}$ is not the entire interval $\Lambda = (-1, 1)$, but the subset $(-1, 0)$. This follows from

¹⁶The failure of the standard limit behavior, and the fact that the distribution is insensitive to n , are suggested by the fact that the partial information on λ in this model is $i_{\lambda\lambda.\sigma^2} = 4(1+\lambda^2-1/n)/(1-\lambda)^2$. This evidently increases with n , but to a *finite* limit.

¹⁷This and other approximations to the distribution of a sum of positive multiples of χ^2 random variables are discussed in Johnson, Kotz, and Balakrishnan (1995).

Proposition 1, because ι_n spans the eigenspace of W corresponding to the only positive eigenvalue of W . In other words, in this model $\hat{\lambda}_{\text{ML}}$ can never be positive, regardless of the true value of λ .

Proposition 6. *In the row-standardized Complete Bipartite model with $X = \iota_n$ and $\varepsilon \sim \text{SMN}(0, I_n)$,*

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \begin{cases} \Pr(F_{1,n-2} > -(n-2)g(z, \lambda)), & \text{if } z \in (-1, 0) \\ 1, & \text{if } z \in [0, 1), \end{cases} \quad (5.17)$$

where

$$g(z, \lambda) := \frac{2z(1+\lambda)^2}{(1+z)^2(n-(n-2)z)}.$$

The corresponding density is

$$\text{pdf}_{\hat{\lambda}_{\text{ML}}}(z; \lambda) = \begin{cases} \frac{1}{B(\frac{1}{2}, \frac{n-2}{2})} \frac{\dot{g}(z, \lambda)}{g(z, \lambda)^{\frac{1}{2}} (1-g(z, \lambda))^{\frac{n-1}{2}}}, & \text{if } z \in (-1, 0) \\ 0, & \text{if } z \in [0, 1). \end{cases} \quad (5.18)$$

The limiting cdf and pdf as $n \rightarrow \infty$ can be obtained from Proposition 6. Letting

$$h(z, \lambda) := \lim_{n \rightarrow \infty} (-(n-2)g(z, \lambda)) = -\frac{2z(1+\lambda)^2}{(1+z)^2(1-z)},$$

we have that under mixed Gaussianity, and for $z \in (-1, 0)$,

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) \rightarrow \Pr(\chi_1^2 > h(z, \lambda)),$$

and

$$\text{pdf}_{\hat{\lambda}_{\text{ML}}}(z; \lambda) \rightarrow -\frac{\dot{h}(z, \lambda)}{\sqrt{2\pi h(z, \lambda)}} e^{-\frac{h(z, \lambda)}{2}}.$$

Again, $\hat{\lambda}_{\text{ML}}$ is not consistent, but converges in distribution to a random variable supported on the non-positive real line as $n \rightarrow \infty$. Note that row-standardization of W is critical here: the Complete Bipartite model with symmetric W and constant mean does satisfy the assumptions for consistency and asymptotic normality in Lee (2004).

The density (5.18) is plotted in Figure 4 for $\lambda = -0.5, 0, 0.5$, for $n = 5, 10$, and for $n \rightarrow \infty$. Note that the shape of the density for $z < 0$ is similar to the case of the pure symmetric Complete Bipartite model (Figure 3).

6 Saddlepoint approximation

The complexity of a distribution theory for $\hat{\lambda}_{\text{ML}}$ derives from the fact that we need to deal with the quadratic forms $R = R(y, z) := y' S_z' Q_z S_z y$. However, fortunately, and precisely because the distribution theory for quadratic forms is so complex, there is a very large literature on the problem of approximating the distribution of such a statistic. This literature contains both first-order approximations (Kelejian and Prucha, 2001, Lee, 2004), and higher-order (Edgeworth and Saddlepoint) approximations. Examples include the fundamental work of Daniels (1954), (1956), Phillips (1978), Durbin (1980), Lieberman (1994a,b), Marsh (2002), Butler and Paoletta (2008), and Robinson and Rossi (2015).

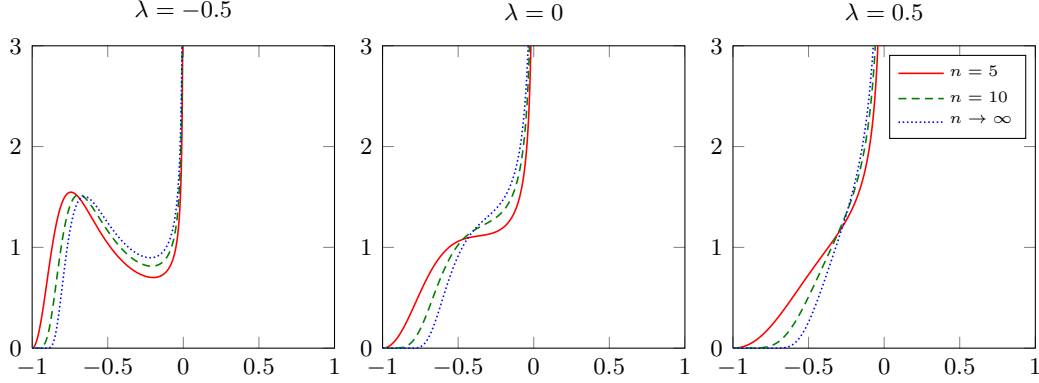


Figure 4: Density of $\hat{\lambda}_{\text{ML}}$ for the Gaussian row-standardized Complete Bipartite model with constant mean.

The first-order asymptotic properties of $\hat{\lambda}_{\text{ML}}$ are, at least under some regularity conditions, known (Lee, 2004). And, Robinson and Rossi (2015) have recently given an Edgeworth expansion for the distribution in a fixed-effects panel data model without regressors. The result in Theorem 1 provides us with a very direct route to obtaining saddlepoint expansions for the distribution $\hat{\lambda}_{\text{ML}}$.¹⁸ Thus, we focus here on the saddlepoint approach, appealing mainly to Daniels (1956, 1987), and Lugannani and Rice (1980). The approach was used by Lieberman (1994b) to approximate the distribution of a serial correlation coefficient, and that application has much in common with the present case. In particular, the matrix of the key quadratic form involved in that paper, $D(\hat{\alpha})$, is a function of both the point at which the cdf is evaluated, and the parameter of interest, as is our matrix $A(z; \lambda)$. We begin by explaining the method very briefly.

6.1 The Lugannani-Rice approximation

The density of R , for a fixed $z \in \Lambda$, may be expressed as the inverse Fourier transform

$$\text{pdf}_R(r) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \exp(K_R(\theta) - r\theta) d\theta, \quad (6.1)$$

where $K_R(\theta) = \log E(\exp(\theta R(y, z)))$ is the cumulant generating function (cgf) of R . Note that we have suppressed the dependence of K_R on the model parameters $(\lambda, \beta, \sigma^2)$, but of course they are present, as will shortly become important. Integrating the density, the tail probability $\Pr(R \geq r)$ is given by

$$\Pr(R \geq r) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \exp(K_R(\theta) - r\theta) \frac{d\theta}{\theta}, \quad (6.2)$$

(see Daniels, 1956, 1987). Lugannani and Rice (1980) derived an approximation to integrals of this type by saddlepoint methods, but accounting for the fact that the integrand has a pole at the origin (see Daniels, 1987, for an excellent detailed discussion).

¹⁸There are other higher-order approximations of the same type that could be used, see Reid (1996, 2003) for surveys of such methods. Those, however, do not exploit the explicit representation of the exact density, and we would expect them therefore to be inferior.

The first-order Lugannani-Rice approximation has the form (expressing it in terms of the cdf)

$$\Pr(R \leq r) \simeq \Phi(\hat{w}) + \phi(\hat{w}) \left(\frac{1}{\hat{w}} - \frac{1}{\hat{u}} \right), \quad (6.3)$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cdf and pdf of the standard normal distribution,¹⁹ respectively, and

$$\hat{w} := \text{sign}(\hat{r}) \sqrt{-2K_R'(\hat{r})}, \quad \hat{u} := \hat{r} \sqrt{K_R''(\hat{r})}, \quad (6.4)$$

where the saddlepoint value \hat{r} is determined by $K_R'(\hat{r}) = r$, with a prime denoting a derivative. In our case we are only concerned with the value $r = 0$, so the saddlepoint \hat{r} is the solution to $K_R'(\hat{r}) = 0$. An analogous second-order expansion is also available (see Daniels, 1987), and is given in Section S.5 of the Supplementary Material, where the modification needed when $K_R'(0) = 0$ is also given.

Remark 1. For general statistics based on non-i.i.d. data, such as ours, it can be difficult to formally justify the validity of the saddlepoint approximation. One approach that has seen some success - see Butler, Huzurbazar, and Booth (1992a,b) - is to embed the problem into the exponential family, and then exploit the known results on saddlepoint approximations in that context. However, this is beyond the scope of the present paper. See also Section S.5 of the Supplementary Material.

In the case of interest here, the cgf of R assuming that $\tilde{y} \sim N(X\beta, \sigma^2 I_n)$ is given by

$$K_R(r) = -\frac{1}{2} \log(\det(I_n - 2rA(z, \lambda))) - \frac{1}{2\sigma^2} \beta' X' \left(I_n - (I_n - 2rA(z, \lambda))^{-1} \right) X\beta, \quad (6.5)$$

for $r \in (1/(2a_{\min}), 1/(2a_{\max}))$, where a_{\min} and a_{\max} denote, respectively, the smallest and the largest eigenvalues of $A(z, \lambda)$ (see equation (5.2) for the definition of $A(z, \lambda)$).²⁰ In the pure model with $\beta = 0$, only the first term remains, so $K_R(r)$ is free of nuisance parameters and depends only on λ .

Remark 2. Like the QMLE $\hat{\lambda}_{\text{ML}}$ itself, the cgf $K_R(r)$ that we propose to use as the basis for the approximation is based on Gaussian assumptions. However, we intend the approximation derived under these assumptions to be used more generally, and will confirm later by simulation that it is quite robust, at least in the models we have studied. Even when the errors have a gamma distribution - i.e., are very far from Gaussian - the approximation works extremely well, and in many cases is much closer to the exact cdf than the first-order asymptotic cdf.

6.2 Examples

We first illustrate the utility of the Lugannani-Rice approximation by applying it to the two examples discussed in the previous section. Recall that these are both cases where, under certain conditions, standard first-order asymptotics does not apply. In the

¹⁹The base distribution in this expression need not be Gaussian; see Wood, Booth, and Butler (1993). In our limited experience, using a chi-squared base seems to make very little difference to the performance of the approximation, possibly because we are only concerned with the cdf of R at the origin.

²⁰It can be shown that the saddlepoint equation $K_R'(r) = 0$ has a unique root, \hat{r} , in the given interval, but that root cannot be written down explicitly. Phillips (1978) and Lieberman (1994b) (and others) have suggested using the first few terms of a series reversion to approximate the root, but, since in our case it is straightforward to locate the root numerically, we do not pursue that here.

Group Interaction model this occurs when the sample size increases because group sizes increase, with the number of groups remaining small. In the complete bipartite model Lee’s (2004) standard asymptotic results do not apply. The density of the MLE has a point of non-analyticity (in fact a pole) at the origin in the pure model, and can have restricted support in the row-standardized model with intercept. Thus, both examples have peculiarities that might lead one to expect that any approximation must perform poorly.

On the contrary, the Lugannani-Rice approximation (6.3) works very well in both models, and reproduces the peculiarities of the respective distributions quite faithfully. Figure 5 displays, on the left, the exact and approximated cdf of $\hat{\lambda}_{ML}$ in a pure balanced group interaction model with $r = 5, m = 40$, and, on the right, the cdf of $\hat{\lambda}_{ML}$ in an intercept-only row-standardized bipartite model with $n = 10$. In both cases, $\lambda = 0$, $\varepsilon \sim N(0, I_n)$, and the cdf’s are plotted over $(-1, 1)$.²¹ We display the exact cdf (obtained using equations (5.9) and (5.17)), the Lugannani-Rice approximation, and, for comparison, the cdf that would be used by (mistakenly) using the usual asymptotic normal approximation. The saddlepoint approximation is graphically almost indistinguishable from the exact cdf in both cases, and embodies the restricted support property accurately in the second case. More evidence of the performance of the saddlepoint approximation for the balanced Group Interaction model can be found in Section S.5.1 of the Supplementary Material.

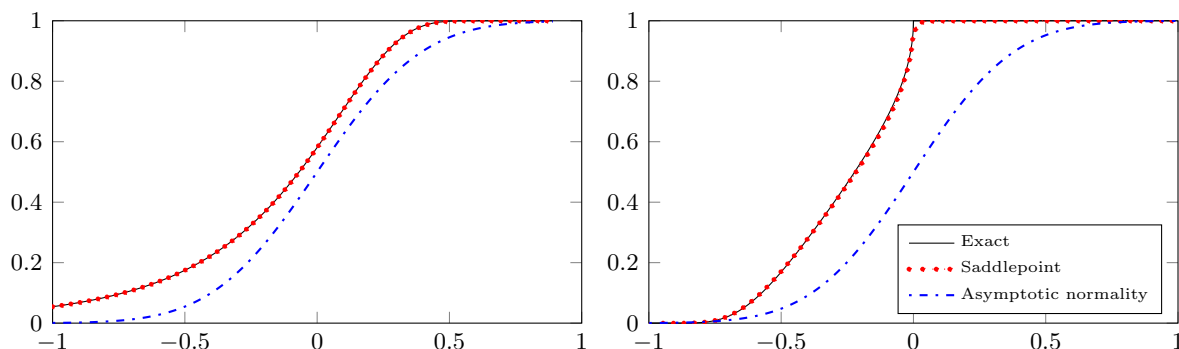


Figure 5: Saddlepoint approximation, and approximation based on asymptotic normality, to the cdf of $\hat{\lambda}_{ML}$ for a pure balanced group interaction model with $(r, m) = (5, 40)$ (left panel), and for an intercept-only row-standardized bipartite model with $n = 10$ (right panel). In both cases $\lambda = 0$.

Turning to a case where standard asymptotics does apply, Figure 6 displays the Lugannani-Rice approximation (6.3) and the first-order asymptotic cdf (Lee, 2004) for a pure SAR model with a row-standardized h -ahead h -behind circular matrix (before row-standardization, this matrix has (i, j) -th entry equal to 1 if $0 < |i - j| \bmod ((n - h - 1) \leq h)$, and to 0 otherwise). We choose $n = 200$, $h = 5, 10$, and $\lambda = 0.5$, take $\varepsilon \sim N(0, I_n)$, and plot the cdf’s over the interval $(0, 1)$. The “exact” cdf here is obtained by simulation. The Lugannani-Rice approximation is virtually exact. Unreported simulations suggest that, as regressors are introduced (and hence nuisance parameters need to be estimated), the saddlepoint approximation remains extremely accurate even

²¹Note that the distribution of $\hat{\lambda}_{ML}$ is invariant to σ in the former model, to both σ and β in the latter model; cf. Section 4.3.

in the tails, provided that n is not too small relative to k .

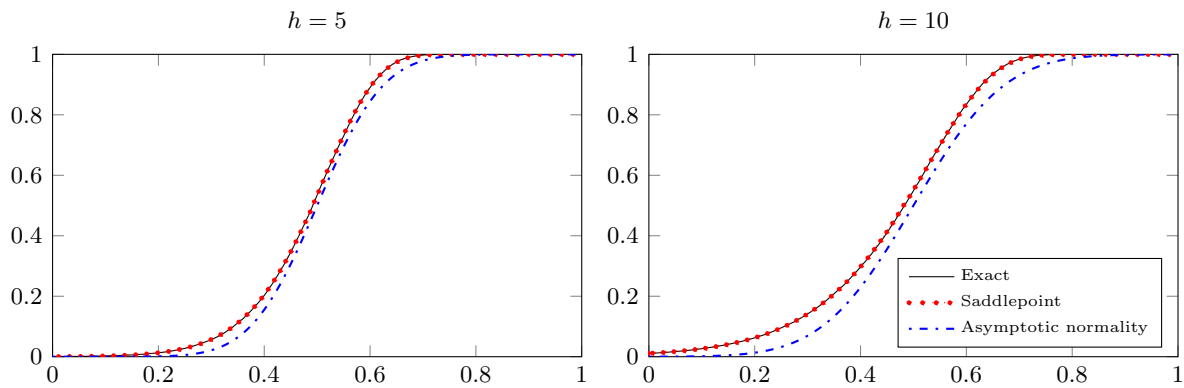


Figure 6: Saddlepoint approximation, and approximation based on asymptotic normality, to the cdf of $\hat{\lambda}_{\text{ML}}$, for a pure SAR model, when W is a h -ahead h -behind circular matrix, $n = 200$, and $\lambda = 0.5$.

6.3 Approximate confidence intervals for λ

The saddlepoint approximation to the cdf of $\hat{\lambda}_{\text{ML}}$ can be used to construct confidence intervals for λ . In this section we discuss and evaluate the use of such intervals. It is important to note from the outset that K_R , and hence all the terms in (6.3), depend not just on λ , but also on the nuisance parameters β and σ^2 . To overcome this obstacle, we suggest replacing (β, σ^2) in the terms that appear in (6.3) not by their maximum likelihood estimates, but by the profile MLEs $\hat{\beta}_{\text{ML}}(\lambda)$ and $\hat{\sigma}_{\text{ML}}^2(\lambda)$ given in equation (3.2). This is motivated by, but is not formally equivalent to, the use of quantities obtained from the profile likelihood to approximate the cdf of the MLE for a scalar parameter in the presence of nuisance parameters - see Reid (2003, Section 3) for discussion and references.²²

With these substitutions, the terms \hat{w} and \hat{u} in equation (6.3) will be functions of just z , λ , and the data. To obtain a confidence interval for λ from the approximation we simply replace z in the approximation (6.3) by the observed value of $\hat{\lambda}_{\text{ML}}$, say \hat{z} . Given the data, the (approximate) cdf is then a function of λ alone, and we may, for a $(1 - \alpha)\%$ equi-tailed two-sided confidence interval, read off upper and lower values of λ at which $\Pr(\hat{\lambda}_{\text{ML}} \leq \hat{z}; \lambda) = \alpha/2$ and $1 - \alpha/2$ (say). We will also consider right-sided confidence intervals $(-\infty, \lambda_U)$ where λ_U is the value of λ such that $\Pr(\hat{\lambda}_{\text{ML}} \leq \hat{z}; \lambda) = 1 - \alpha$.

6.3.1 Simulation evidence

We conduct Monte Carlo experiments to investigate the performance of the saddlepoint confidence intervals for λ . We take W in equation (1.1) to be a row-standardized h -ahead h -behind circular matrix, and we generate the errors ε_i from either (a) a standard normal distribution, or (b) a gamma distribution with shape parameter 1 and scale parameter 1, demeaned by the population mean. Mean, variance, skewness, and kurtosis

²²The difference is that we are taking explicit account of the structure of the profile score function, which the general approach does not. When k is large compared to n , some adjustments may be prudent, the simplest of which would be to use the bias-corrected profile estimator $n\hat{\sigma}_{\text{ML}}^2(\lambda)/(n - k)$ for σ^2 .

are 0, 1, 0, 3 in case (a), and 0, 1, 2, 9 in case (b). The number of replications is 50,000 for all experiments.

We start from a small sample setting, with no regressors. We set $n = 50$, $h = 5$, and we consider a pure model ($k = 0$) and a model with intercept ($k = 1$). Note that in both these models the distribution of the MLE, and hence the confidence intervals, do not depend on β or σ (cf. Section 4.3). Table 1 reports empirical coverages of 95% saddlepoint confidence intervals, for a variety of values of λ , and compares them to the coverages of 95% Wald confidence intervals based on first-order asymptotic normality.²³ From the results in the table, it is clear that the coverages of the saddlepoint intervals are uniformly excellent in the Gaussian case, and still very satisfactory in the gamma case, whereas the first-order asymptotic approximation is not sufficiently accurate at this sample size.

Table 1: Empirical coverages of saddlepoint and first-order Wald 95% confidence intervals for λ , for a pure ($k = 0$) or intercept only ($k = 1$) model, when $n = 50$, W is a 5-ahead 5-behind circular matrix.

		$k = 0$				$k = 1$			
		two-sided		right-sided		two-sided		right-sided	
	λ	Saddl.	Wald	Saddl.	Wald	Saddl.	Wald	Saddl.	Wald
Normal									
	-0.9	0.950	0.936	0.950	0.935	0.950	0.916	0.949	0.890
	-0.5	0.951	0.937	0.951	0.934	0.949	0.912	0.949	0.882
	0	0.951	0.938	0.950	0.938	0.950	0.911	0.950	0.875
	0.5	0.950	0.944	0.950	0.945	0.950	0.912	0.950	0.865
	0.9	0.949	0.954	0.949	0.955	0.951	0.907	0.950	0.821
Gamma									
	-0.9	0.952	0.938	0.946	0.942	0.958	0.926	0.955	0.900
	-0.5	0.950	0.937	0.945	0.943	0.958	0.924	0.956	0.893
	0	0.950	0.939	0.946	0.947	0.959	0.922	0.955	0.886
	0.5	0.949	0.942	0.944	0.952	0.957	0.925	0.954	0.879
	0.9	0.944	0.946	0.938	0.961	0.958	0.917	0.955	0.834

Next, we move to a more realistic setting, by increasing the sample size and adding regressors, and we also allow for larger values of h . Table 2 reports empirical coverages for $n = 200, 400$, $h = 5, 20, 50$, and $\lambda = 0, 0.5$, and for a pure model ($k = 0$), a model with intercept ($k = 1$), and a model with intercept and two regressors ($k = 3$). In the case $k = 3$, the results depend on β/σ , and we set the true value of β/σ to ι_k . Also note that, when $k = 3$, the two regressors are randomly drawn, one from a standard normal distribution and the other from a uniform on $[0, 1]$, and then kept constant across repetitions.²⁴ The results in the table show that as h increases the coverage of the Wald tests moves away from the nominal coverage, particularly for the one-sided intervals. This is consistent with the well known fact that the first-order asymptotic approximation to the distribution of $\hat{\lambda}_{\text{ML}}$ deteriorates as W becomes more dense. On the other hand,

²³For nominal coverage 95%, the latter intervals are $\hat{\lambda}_{\text{ML}} \pm 1.960\sqrt{\hat{v}}$ (two-sided) and $(-\infty, \hat{\lambda}_{\text{ML}} + 1.645\sqrt{\hat{v}})$ (right-sided), where \hat{v} denotes the asymptotic variance of $\hat{\lambda}_{\text{ML}}$ (see Theorem 3.2 of Lee, 2004) evaluated at the MLEs for λ, β, σ^2 .

²⁴Generating the regressors in each repetition, which we prefer not to do since the saddlepoint confidence intervals are constructed assuming that X is fixed, does not change the simulation results substantially. The results are also not very sensitive to the true value of β/σ .

the coverage properties of the saddlepoint confidence intervals are quite insensitive to h , k , and λ , and again remain very satisfactory under the gamma assumption.

Larger number of regressors We now consider models with a larger number of regressors. Table 3 is the analog of Table 2 for a model with intercept and $k-1$ regressors, when $k = 5, 11, 21$. Half of the regressors are drawn from a standard normal distribution, half from a uniform distribution on $[0, 1]$. Due to the larger number of regressors, we use the adjusted profile estimator $n\hat{\sigma}_{\text{ML}}^2(\lambda)/(n-k)$ for σ^2 given λ for the saddlepoint confidence intervals, and, correspondingly, the estimator $n\hat{\sigma}_{\text{ML}}^2/(n-k)$ for the Wald tests. Table 4 repeats this exercise for the important case in which contextual effects are included in the model (see the Introduction). That is, we take $X = (\iota_n, \tilde{X}, W\tilde{X})$, where half of the columns of the $n \times (\tilde{k} - 1)$ matrix \tilde{X} are drawn from a standard normal distribution, half from a uniform distribution on $[0, 1]$. Note that the number k of columns of X is then $2\tilde{k} - 1$.

The empirical coverages of the saddlepoint confidence intervals are uniformly excellent in both Tables 3 and 4, even when the number of nuisance parameters is large. On the other hand, the empirical coverages of the first-order Wald confidence intervals can be very far from the nominal coverage, especially when contextual effects are included. One important difference between the case of Table 3 (no contextual effects) and that of Table 4 (contextual effects) is that the coverage properties of the first-order Wald confidence intervals improve slightly as k increases (for $k > 0$) in the former case, but get worse as k increases in the latter case. This is connected to the fact that, without contextual effects, additional regressors help to accurately estimate λ , whereas they make estimation of λ more difficult in the case of contextual effects.

7 Conclusion

We have shown that, although the QMLE for λ in the SAR model (1.1) cannot be written in closed form, its cumulative distribution function admits a relatively simple representation. This representation exposes some unexpected properties of the estimator, including that its support may be constrained to a subset of the parameter space, and its cdf may have a different functional form on different parts of the parameter space, and be non-analytic where they join. Such properties are illustrated in some simple, but relevant, examples, which also provide examples of cases where standard asymptotics do not apply. An exact expression for the cdf in a symmetric pure model is given, but this itself illustrates that exact results based on the main representation result, Theorem 1, will always be extremely complex, except in very special cases.

More importantly for inference, therefore, is that the representation of the cdf in Theorem 1 provides access to a simple but powerful approximation based on saddlepoint methods, due to Lugannani and Rice (1980). We have shown that it accurately reproduces the true cdf even in examples where standard asymptotics fails, and when the distribution has points of non-analyticity or restricted support. Finally, we have applied the approximation to the problem of constructing confidence intervals for λ , and have found that the proposed confidence intervals have very satisfactory coverage properties, even under apparently adverse model structure and assumptions. The approximation improves significantly on standard first-order asymptotics, most especially in the important class of models containing contextual effects.

Table 2: Empirical coverages of saddlepoint and first-order Wald 95% confidence intervals for λ , for a model when W is a h -ahead h -behind circular matrix, and $k = 0$ (pure model), $k = 1$ (intercept only), $k = 3$ (intercept and two regressors).

				$k = 0$				$k = 1$				$k = 3$			
				two-sided		right-sided		two-sided		right-sided		two-sided		right-sided	
	n	λ	h	Saddl.	Wald	Saddl.	Wald	Saddl.	Wald	Saddl.	Wald	Saddl.	Wald	Saddl.	Wald
Normal															
	200	0	5	0.950	0.947	0.950	0.946	0.950	0.941	0.950	0.925	0.949	0.942	0.949	0.927
			20	0.950	0.947	0.950	0.944	0.950	0.925	0.950	0.886	0.950	0.929	0.949	0.896
			50	0.950	0.942	0.950	0.938	0.952	0.874	0.951	0.805	0.950	0.893	0.951	0.837
		0.5	5	0.950	0.949	0.950	0.949	0.949	0.942	0.949	0.922	0.949	0.943	0.950	0.925
			20	0.950	0.951	0.950	0.947	0.950	0.923	0.950	0.873	0.949	0.924	0.949	0.883
			50	0.950	0.950	0.948	0.946	0.951	0.858	0.951	0.766	0.951	0.873	0.951	0.798
	400	0	5	0.950	0.949	0.950	0.950	0.949	0.945	0.950	0.935	0.950	0.947	0.951	0.937
			20	0.951	0.949	0.950	0.946	0.949	0.939	0.950	0.912	0.950	0.940	0.950	0.918
			50	0.950	0.949	0.951	0.944	0.949	0.920	0.950	0.874	0.950	0.926	0.951	0.887
		0.5	5	0.950	0.949	0.949	0.950	0.951	0.949	0.951	0.934	0.951	0.947	0.949	0.935
			20	0.950	0.951	0.950	0.947	0.949	0.938	0.950	0.904	0.951	0.939	0.950	0.912
			50	0.951	0.953	0.950	0.947	0.951	0.917	0.951	0.858	0.950	0.919	0.950	0.873
Gamma															
	200	0	5	0.953	0.950	0.950	0.952	0.954	0.946	0.951	0.930	0.951	0.944	0.950	0.930
			20	0.950	0.947	0.949	0.945	0.953	0.928	0.952	0.889	0.951	0.931	0.951	0.898
			50	0.949	0.942	0.947	0.940	0.953	0.876	0.951	0.806	0.952	0.894	0.951	0.838
		0.5	5	0.952	0.951	0.950	0.954	0.954	0.948	0.951	0.929	0.951	0.943	0.950	0.927
			20	0.950	0.951	0.948	0.949	0.953	0.927	0.951	0.877	0.951	0.926	0.951	0.885
			50	0.948	0.948	0.945	0.947	0.953	0.861	0.952	0.768	0.951	0.873	0.951	0.798
	400	0	5	0.952	0.951	0.951	0.953	0.952	0.948	0.949	0.937	0.951	0.948	0.950	0.936
			20	0.950	0.949	0.949	0.947	0.952	0.941	0.953	0.914	0.950	0.942	0.949	0.920
			50	0.949	0.947	0.948	0.945	0.952	0.922	0.950	0.878	0.950	0.926	0.951	0.888
		0.5	5	0.952	0.953	0.949	0.955	0.952	0.949	0.950	0.937	0.950	0.948	0.950	0.936
			20	0.952	0.953	0.951	0.950	0.953	0.941	0.951	0.909	0.949	0.940	0.949	0.914
			50	0.950	0.953	0.948	0.949	0.951	0.919	0.950	0.859	0.951	0.919	0.951	0.872

Table 3: Empirical coverages of saddlepoint and first-order Wald 95% confidence intervals for λ , when W is a h -ahead h -behind circular matrix, there are k regressors (including the intercept), and no contextual effects.

				$k = 5$				$k = 11$				$k = 21$			
				two-sided		right-sided		two-sided		right-sided		two-sided		right-sided	
	n	λ	h	Saddl.	Wald	Saddl.	Wald	Saddl.	Wald	Saddl.	Wald	Saddl.	Wald	Saddl.	Wald
Normal															
	200	0	5	0.950	0.944	0.950	0.929	0.950	0.944	0.949	0.933	0.949	0.944	0.949	0.936
			20	0.951	0.932	0.949	0.901	0.950	0.935	0.950	0.909	0.950	0.939	0.950	0.921
			50	0.951	0.894	0.951	0.840	0.951	0.912	0.950	0.866	0.950	0.923	0.949	0.889
		0.5	5	0.949	0.942	0.950	0.925	0.950	0.944	0.950	0.931	0.950	0.945	0.949	0.935
			20	0.950	0.926	0.949	0.885	0.949	0.930	0.950	0.897	0.951	0.937	0.950	0.912
			50	0.952	0.875	0.952	0.800	0.951	0.893	0.950	0.831	0.951	0.907	0.951	0.860
	400	0	5	0.951	0.948	0.951	0.936	0.950	0.948	0.950	0.939	0.950	0.948	0.950	0.942
			20	0.949	0.939	0.950	0.918	0.949	0.942	0.950	0.923	0.951	0.945	0.951	0.931
			50	0.950	0.928	0.951	0.891	0.950	0.934	0.951	0.904	0.949	0.938	0.948	0.914
		0.5	5	0.949	0.946	0.948	0.934	0.948	0.945	0.949	0.933	0.949	0.948	0.950	0.941
			20	0.949	0.939	0.950	0.911	0.950	0.942	0.950	0.918	0.950	0.945	0.949	0.928
			50	0.949	0.919	0.949	0.873	0.949	0.927	0.948	0.890	0.948	0.931	0.948	0.904
Gamma															
	200	0	5	0.951	0.945	0.951	0.931	0.950	0.944	0.950	0.933	0.949	0.944	0.950	0.936
			20	0.952	0.930	0.951	0.899	0.949	0.936	0.949	0.909	0.950	0.940	0.950	0.920
			50	0.952	0.896	0.952	0.840	0.950	0.912	0.951	0.867	0.950	0.924	0.950	0.889
		0.5	5	0.950	0.943	0.950	0.927	0.949	0.943	0.951	0.929	0.948	0.943	0.949	0.934
			20	0.951	0.927	0.951	0.887	0.951	0.931	0.951	0.898	0.951	0.935	0.950	0.910
			50	0.953	0.875	0.952	0.800	0.951	0.893	0.952	0.832	0.951	0.907	0.950	0.858
	400	0	5	0.950	0.946	0.949	0.935	0.949	0.947	0.949	0.939	0.949	0.947	0.948	0.943
			20	0.952	0.942	0.951	0.919	0.950	0.943	0.950	0.924	0.949	0.944	0.949	0.931
			50	0.951	0.928	0.952	0.892	0.952	0.936	0.951	0.907	0.952	0.938	0.952	0.911
		0.5	5	0.950	0.947	0.951	0.935	0.952	0.948	0.949	0.938	0.951	0.949	0.951	0.938
			20	0.951	0.940	0.948	0.914	0.951	0.942	0.952	0.918	0.949	0.944	0.949	0.925
			50	0.952	0.921	0.951	0.874	0.949	0.927	0.949	0.887	0.953	0.935	0.951	0.903

Table 4: Empirical coverages of saddlepoint and first-order Wald 95% confidence intervals for λ , when W is a h -ahead h -behind circular matrix, there are \tilde{k} regressors (including the intercept), and the model contains contextual effects.

				$\tilde{k} = 3$				$\tilde{k} = 5$				$\tilde{k} = 11$			
				two-sided		right-sided		two-sided		right-sided		two-sided		right-sided	
n	λ	h		Saddl.	Wald	Saddl.	Wald	Saddl.	Wald	Saddl.	Wald	Saddl.	Wald	Saddl.	Wald
Normal															
200	0	5		0.950	0.926	0.949	0.891	0.949	0.903	0.949	0.851	0.947	0.794	0.950	0.699
		20		0.947	0.842	0.951	0.759	0.947	0.718	0.952	0.603	0.950	0.387	0.967	0.262
		50		0.950	0.679	0.960	0.546	0.955	0.546	0.970	0.386	0.959	0.398	0.974	0.225
	0.5	5		0.950	0.927	0.950	0.888	0.950	0.907	0.950	0.854	0.949	0.826	0.950	0.739
		20		0.951	0.832	0.950	0.737	0.951	0.711	0.953	0.587	0.958	0.423	0.964	0.294
		50		0.956	0.616	0.961	0.464	0.961	0.445	0.972	0.283	0.961	0.250	0.975	0.112
	400	5		0.950	0.940	0.948	0.915	0.951	0.928	0.952	0.891	0.950	0.882	0.950	0.813
		20		0.950	0.900	0.951	0.843	0.946	0.840	0.949	0.754	0.944	0.566	0.956	0.439
		50		0.949	0.819	0.951	0.721	0.948	0.666	0.955	0.537	0.951	0.354	0.970	0.223
		5		0.951	0.940	0.951	0.914	0.949	0.932	0.950	0.894	0.950	0.894	0.951	0.832
		20		0.950	0.899	0.949	0.836	0.949	0.841	0.949	0.753	0.950	0.635	0.953	0.512
		50		0.949	0.797	0.949	0.688	0.952	0.653	0.955	0.514	0.961	0.355	0.968	0.228
Gamma															
200	0	5		0.953	0.931	0.950	0.898	0.951	0.906	0.950	0.854	0.948	0.796	0.950	0.704
		20		0.951	0.847	0.952	0.763	0.948	0.721	0.953	0.604	0.952	0.389	0.967	0.264
		50		0.954	0.682	0.961	0.547	0.956	0.550	0.970	0.389	0.962	0.400	0.974	0.224
	0.5	5		0.952	0.930	0.951	0.892	0.952	0.908	0.952	0.856	0.950	0.832	0.951	0.746
		20		0.952	0.835	0.951	0.741	0.951	0.717	0.953	0.594	0.958	0.425	0.965	0.296
		50		0.957	0.619	0.962	0.469	0.962	0.447	0.973	0.282	0.963	0.256	0.976	0.114
	400	5		0.951	0.941	0.949	0.917	0.951	0.930	0.951	0.893	0.950	0.881	0.950	0.814
		20		0.952	0.903	0.952	0.844	0.949	0.843	0.953	0.758	0.947	0.569	0.957	0.441
		50		0.948	0.820	0.951	0.724	0.947	0.670	0.955	0.539	0.953	0.354	0.969	0.220
		5		0.951	0.941	0.949	0.914	0.948	0.929	0.947	0.893	0.948	0.897	0.948	0.834
		20		0.950	0.899	0.949	0.833	0.951	0.842	0.950	0.754	0.951	0.637	0.955	0.508
		50		0.950	0.798	0.952	0.690	0.954	0.657	0.955	0.521	0.960	0.355	0.968	0.227

Appendix A An auxiliary result

The following result establishes some properties of the eigenvalues $\gamma_i(z)$ and $d_{ii}(z, \lambda)$, defined respectively in Section 3 and in Section 5.1.

Lemma A.1. *Assume that all eigenvalues of W are real.*

- (i) *For any $z \in \Lambda$, the distinct eigenvalues $\gamma_1(z), \gamma_2(z), \dots, \gamma_N(z)$ of C_z are in increasing order (i.e., $j > i$ implies $\gamma_j(z) > \gamma_i(z)$ for any $z \in \Lambda$). For any $z \in \Lambda$, $\gamma_1(z) < 0$, $\gamma_N(z) > 0$, and each $\gamma_i(z)$, $i = 2, \dots, N-1$, changes sign exactly once on Λ .*
- (ii) *For $N \geq 2$, $d_{11}(z, \lambda) < 0$ and $d_{NN}(z, \lambda) > 0$ for all $z, \lambda \in \Lambda$. If $N > 2$, $d_{ii}(z, \lambda) > 0$ if $z < z_i$ and $d_{ii}(z, \lambda) < 0$ if $z > z_i$, for any $\lambda \in \Lambda$, and for $i = 2, \dots, N-1$, where z_i denotes the unique zero of $\gamma_i(z)$ in Λ .*

Proof. See Supplementary Material, Section S.6. □

Appendix B Proofs

Proof of Lemma 1. See Supplementary Material, Section S.6. □

Proof of Lemma 2. Recall that we are assuming that $\hat{\lambda}_{\text{ML}}$ exists and is not trivial. By Proposition S.1.1 in the Supplementary Material, $l(\lambda) \rightarrow -\infty$ a.s. at the extremes of Λ , and hence must a.s. have at least one maximum on Λ . Since the function $l(\lambda)$ is a.s. differentiable on Λ , all its maxima must be critical points. We now show that $l(\lambda)$ has a.s. exactly one maximum, and no other critical points, on Λ . The second derivative of $l(\lambda)$ can be written as

$$\frac{d^2 l(\lambda)}{d\lambda^2} = \frac{-n(ac - b^2)}{(a\lambda^2 - 2b\lambda + c)^2} + \frac{n(b - a\lambda)^2}{(a\lambda^2 - 2b\lambda + c)^2} - \text{tr}(G_\lambda^2),$$

where $a := y'W'M_X W y$, $b := y'W'M_X y$, and $c := y'M_X y$. But at any point where $s(\lambda) = 0$,

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

so that, at any critical point,

$$\begin{aligned} \frac{d^2 l(\lambda)}{d\lambda^2} &= \frac{-n(ac - b^2)}{(a\lambda^2 - 2b\lambda + c)^2} + \frac{1}{n} [\text{tr}(G_\lambda)]^2 - \text{tr}(G_\lambda^2) \\ &= \frac{-n(ac - b^2)}{(a\lambda^2 - 2b\lambda + c)^2} - \text{tr}(C_\lambda^2). \end{aligned} \tag{B.1}$$

By the Cauchy-Schwarz inequality the first term in (B.1) is nonpositive. Hence, if $\text{tr}(C_\lambda^2) > 0$ for all $\lambda \in \Lambda$, we have that $d^2 l(\lambda)/d\lambda^2 < 0$ at every point where $s(\lambda) = 0$, that is, $l(\lambda)$ has a.s. exactly one point of maximum in Λ , and no other critical points. □

Proof of Lemma 3. See Supplementary Material, Section S.6. □

Proof of Theorem 1. If $l(\lambda)$ is single-peaked on Λ , its single peak is to the left of a point $z \in \Lambda$ if and only $s(z)$ is negative. That is, single-peakedness of $l(\lambda)$ implies that

the event $\hat{\lambda}_{\text{ML}} \leq z$ is identical to the event $s(z) \leq 0$, and hence that $\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr(s(z) \leq 0)$, for any $z \in \Lambda$. By equation (3.6), $\Pr(s(z) \leq 0) = \Pr(y'S'_z Q_z S_z y \leq 0)$, since $y'S'_z M_X S_z y$ is a.s. positive for any $z \in \Lambda$. The desired result follows by Lemma 2. \square

Proof of Proposition 1. Let us start from part (i). Under the stated condition, the diagonal blocks M_{jj} of the matrix $M := H'M_X H$ vanish for any $j > i$. Since, by Lemma A.1 (ii), $d_{jj}(z, \lambda) < 0$ for $j \leq i$, $z > z_i$, and any $\lambda \in \Lambda$, it follows by Proposition S.3.3 (i) that $\Pr(\hat{\lambda}_{\text{ML}} \leq z) = 1$ for $z \geq z_i$. By the same argument, part (ii) is proved by showing that in that case $\Pr(\hat{\lambda}_{\text{ML}} \leq z) = 0$ for $z \leq z_i$. \square

Proof of Proposition 2. See Supplementary Material, Section S.6. \square

Proof of Proposition 3. Follows immediately from the discussion in the text, and the fact that $H'\tilde{y} \sim N(0, I_n)$. \square

Proof of Proposition 4. According to Proposition 3, $\Pr(\hat{\lambda}_{\text{ML}} < z) = \Pr(Q_2/Q_1 < 1)$, for any $z \in \Lambda$, where $Q_i := u'_i A_i u_i$, with $u_i \sim N(0, I_{v_i})$, for $i = 1, 2$. By a trivial modification of the derivation in James (1964), the density of Q_i , for $z \in \Lambda \setminus \{z_2, \dots, z_{N-1}\}$, may be expressed as

$$\text{pdf}(q_i) = \frac{q_i^{\frac{v_i}{2}-1} \exp\left(-\frac{1}{2}q_i \phi_i\right)}{2^{\frac{v_i}{2}} \Gamma(\frac{v_i}{2}) (\det A_i)^{\frac{1}{2}}} {}_1F_1\left(\frac{1}{2}, \frac{v_i}{2}; \frac{1}{2}q_i \phi_i (I_{v_i} - (\phi_i A_i)^{-1})\right), \quad (\text{B.2})$$

where ϕ_i is a positive constant, and ${}_1F_1$ denotes a matrix-argument confluent hypergeometric function (see, e.g., Muirhead, 1982). If ϕ_i is chosen so that $\phi_i A_i > I_{v_i}$, this expresses the density as a mixture of central χ^2 densities. The joint density of Q_1 and Q_2 is

$$\begin{aligned} \text{pdf}(q_1, q_2) &= \frac{q_1^{\frac{v_1}{2}-1} q_2^{\frac{v_2}{2}-1} \exp\left(-\frac{1}{2}(q_1 \phi_1 + q_2 \phi_2)\right)}{2^{\frac{n}{2}} \Gamma(\frac{v_1}{2}) \Gamma(\frac{v_2}{2}) (\det A_1 \det A_2)^{\frac{1}{2}}} \\ &\times {}_1F_1\left(\frac{1}{2}, \frac{v_1}{2}; \frac{1}{2}q_1 \phi_1 (I_{v_1} - (\phi_1 A_1)^{-1})\right) {}_1F_1\left(\frac{1}{2}, \frac{v_2}{2}; \frac{1}{2}q_2 \phi_2 (I_{v_2} - (\phi_2 A_2)^{-1})\right). \end{aligned}$$

Transforming to $f := q_2/q_1$ and q_1 , and integrating out q_1 gives

$$\begin{aligned} \text{pdf}(f) &= \frac{\Gamma(\frac{n}{2}) f^{\frac{v_1}{2}-1}}{\Gamma(\frac{v_1}{2}) \Gamma(\frac{v_2}{2}) (\det D_1 \det D_2)^{\frac{1}{2}}} \sum_{j_1, j_2=0}^{\infty} \frac{(\frac{1}{2})_{j_1} (\frac{1}{2})_{j_2} (\frac{n}{2})_{j_1+j_2}}{j_1! j_2! (\frac{v_1}{2})_{j_1} (\frac{v_2}{2})_{j_2}} f^{j_1+\frac{v_1}{2}-1} \\ &\times \frac{\phi_1^{j_1} \phi_2^{j_2} C_{j_1}(I_{v_1} - (\phi_1 A_1)^{-1}) C_{j_2}(I_{v_2} - (\phi_2 A_2)^{-1})}{\phi_1^{j_1+j_2+\frac{n}{2}} (1 + \frac{\phi_2}{\phi_1} f)^{j_1+j_2+\frac{n}{2}}}. \end{aligned}$$

The integral over $0 < f < 1$ is thus, transforming to $\frac{\phi_2}{\phi_1} f$,

$$\begin{aligned} \Pr(\hat{\lambda}_{\text{ML}} < z) &= \frac{1}{(\det(\phi_1 A_1) \det(\phi_2 A_2))^{\frac{1}{2}}} \\ &\times \sum_{j_1, j_2=0}^{\infty} \frac{(\frac{1}{2})_{j_1} (\frac{1}{2})_{j_2}}{j_1! j_2!} C_{j_1}(I_n - (\phi_1 A_1)^{-1}) C_{j_2}(I_n - (\phi_2 A_2)^{-1}) \Pr\left(f_{v_2+2j_1, v_1+2j_2} < \frac{\phi_2}{\phi_1}\right), \end{aligned}$$

where $f_{p,q}$ denotes the ratio of two independent χ^2 variates, $f_{p,q} := \chi_p^2/\chi_q^2$. This yields the result stated. \square

Proof of Proposition 5. For the pure balanced Group Interaction model, W is symmetric, $N = 2$, $n_1 = r$, $n_2 = r(m-1)$, $\omega_1 = 1$, $\omega_2 = -1/(m-1)$. Also, by direct computation, $\text{tr}(G_z)/n = (rm)^{-1} [r/(1-z) - r(m-1)/(z+m-1)] = z/[(1-z)(z+m-1)]$, and hence $d_{11}(z) = 2(m-1)(1-z)/[(1-\lambda)^2(z+m-1)]$ and $d_{22}(z, \lambda) = -2(z+m-1)/[(\lambda+m-1)^2(1-z)]$. Equation (5.6) now gives $\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr(F_{r,r(m-1)} \leq c(z, \lambda))$. On differentiating with respect to z , we obtain

$$\text{pdf}_{\hat{\lambda}_{\text{ML}}}(z; \lambda) = \dot{c}(z, \lambda) \text{pdf}_{F_{r,r(m-1)}}(c(z, \lambda)),$$

which yields the stated expression for $\text{pdf}_{\hat{\lambda}_{\text{ML}}}(z; \lambda)$. \square

Proof of Proposition 6. For the row-standardized Complete Bipartite model an orthonormal eigenvector matrix H of W is

$$H = \begin{bmatrix} \iota_p/\sqrt{n} & L_{p,p-1} & 0 & \iota_p/\sqrt{n} \\ -\iota_q/\sqrt{n} & 0 & L_{q,q-1} & \iota_q/\sqrt{n} \end{bmatrix},$$

where $L_{p,p-1}$ is a $p \times (p-1)$ matrix such that $L'_{p,p-1}\iota_p = 0$ and $L'_{p,p-1}L_{p,p-1} = I_{p-1}$. Thus,

$$M = H' M_{\iota_n} H = \text{diag} \left(\frac{4pq}{n^2}, I_{n-2}, 0 \right).$$

As expected from Proposition S.3.3 (i) in the Supplementary Material, this is block-diagonal since $M_X W$ is symmetric in this model, and in addition the (n, n) block also vanishes. The mean of $x := H^{-1}\tilde{y}$, partitioned conformably to M , is $\beta(0, 0, \sqrt{n})'$. Therefore, by equation (S.3.5) in the Supplementary Material, when $\varepsilon \sim \text{SMN}(0, I_n)$,

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr(d_{11}(z, \lambda)\chi_1^2 + d_{22}(z, \lambda)\chi_{n-2}^2 \leq 0). \quad (\text{B.3})$$

If $z \in [0, 1)$, then both $d_{11}(z, \lambda)$ and $d_{22}(z, \lambda)$ are non-positive by Lemma A.1 (ii), for any $\lambda \in \Lambda$, and hence $\Pr(\hat{\lambda}_{\text{ML}} \leq z) = 1$. Next, from equation (B.3), $\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr(-2(\phi_2\chi_1^2 + 2z\chi_{n-2}^2) \leq 0)$, where $\phi_2 := 2z$. This yields the stated result for the cdf when $z \in (-1, 0)$. The density is obtained by differentiation, as in the proof of Proposition 5. \square

References

- Bao, Y. (2013) Finite sample bias of the QMLE in spatial autoregressive models, *Econometric Theory*, 29, 68–88.
- Bao, Y. and Ullah, A. (2007) Finite sample moments of maximum likelihood estimator in spatial models, *Journal of Econometrics* 137, 396–413.
- Boucher, V., Bramoullé, Y., Djebbari, H., and Fortin, B. (2014). Do peers affect student achievement? Evidence from Canada using group size variation, *Journal of Applied Econometrics* 29, 91109.
- Bramoullé, Y., Djebbari, H. and Fortin, B. (2009) Identification of peer effects through social networks, *Journal of Econometrics* 150, 41–55.

- Butler, R.W. (2007), *Saddlepoint Approximations with Applications*, Cambridge University Press.
- Butler, R.W., Huzurbazar, S. and Booth, J. (1992a) Saddlepoint approximations for the Bartlett-Nanda-Pillai trace statistic in multivariate analysis, *Biometrika* 79, 705-715.
- Butler, R.W., Huzurbazar, S. and Booth, J. (1992b) Saddlepoint approximations for the generalized variance and Wilks' statistic, *Biometrika* 79, 157-169.
- Butler, R.W. and Paoletta, M.S. (2008), Uniform saddlepoint approximations for ratios of quadratic forms, *Bernoulli* 14, 140-154.
- Carrell, S.E., Sacerdote, B.I., and West, J.E. (2013). From natural variation to optimal policy? The importance of endogenous peer group formation, *Econometrica* 81, 855-882.
- Case, A. (1992). Neighborhood influence and technological change, *Regional Science and Urban Economics* 22, 491-508.
- Cliff, A. D. and Ord, J. K. (1973) *Spatial Autocorrelation*. Pion, London.
- Davezies, L., DHaultfoeuille, X., and Fougère, D. (2009). Identification of peer effects using group size variation, *Econometrics Journal* 12, 397-413.
- Daniels, H.E. (1954) Saddlepoint approximations in statistics, *Annals of Mathematical Statistics* 25, 631-650.
- Daniels, H.E. (1956) The approximate distribution of the serial correlation coefficient, *Biometrika* 43, 169-185.
- Daniels, H.E. (1983) Saddlepoint approximations for estimating equations, *Biometrika* 70, 89-96.
- Daniels, H.E. (1987) Tail probability approximations. *International Statistical Review* 55, 37-48.
- Durbin, J. (1980) Approximations for densities of sufficient estimators, *Biometrika* 67, 311-333.
- Hillier, G.H., Kan, R., and Wang, X. (2009) Computationally efficient recursions for top-order invariant polynomials, with applications, *Econometric Theory* 25, 211-242.
- Hillier, G.H. and Martellosio, F. (2016). Exact likelihood inference in group interaction network models. *Econometric Theory*, doi:10.1017/S0266466616000505
- Horn, R. and Johnson, C.R. (1985) *Matrix Analysis*. Cambridge University Press, Cambridge.
- Jackson, M.O. (2008) *Social and Economic Networks*. Princeton University Press, Princeton.
- James, A.T. (1964) Distributions of matrix variates and latent roots derived from normal samples. *Annals of Mathematical Statistics* 35, 475-501.
- Johnson, N.L., Kotz, S., and Balakrishnan, N. (1995). Continuous univariate distributions. Vol. 2., 2nd Ed. John Wiley, N.Y.
- Kelejian, H.H., Prucha, I.R. (2001) On the asymptotic distribution of the Moran I test statistic with applications, *Journal of Econometrics* 104, 219-257.
- Kelejian, H.H., I.R. Prucha, and Y. Yuzefovich (2006) Estimation problems in models with spatial weighting matrices which have blocks of equal elements, *Journal of Regional Science* 46, 507-515.
- Kelejian, H.H., Prucha, I.R. (2010) Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances, *Journal of Econometrics* 157, 53-67.
- Konishi, S., Niki, N., and Gupta, A.K. (1988) Asymptotic expansions for the distribution of quadratic forms in normal variables, *Annals of the Institute of Statistical Mathematics* 40, 279-296.
- Koopmans, T.C. (1942) Serial correlation and quadratic forms in normal variables, *Annals of Mathematical Statistics* 12, 14-33.

- Kyriakou, M., Phillips, P.C.B., and Rossi, F. (2017) Indirect inference in spatial autoregression, *The Econometrics Journal* 20, 168–189.
- Lee, L.F. (2004) Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models, *Econometrica* 72, 1899–1925.
- Lee, L.F. (2007) Identification and estimation of econometric models with group interactions, contextual factors and fixed effects, *Journal of Econometrics* 140, 333–374.
- Lee, L.-F. and Yu, J. (2010) Estimation of spatial autoregressive panel data models with fixed effects, *Journal of Econometrics* 154, 165–185.
- Lee, L.-F., Liu, X., Lin, X. (2010) Specification and estimation of social interaction models with network structures, *The Econometrics Journal* 13, 145–176.
- LeSage, J.P. and Pace, K.R. (2009) *Introduction to Spatial Econometrics*, Boca Raton, CRC Press/Taylor & Francis.
- Lieberman, O. (1994a), Saddlepoint approximation for the distribution of a ratio of quadratic forms in normal variables. *Journal of the American Statistical Association* 89, 924–928.
- Lieberman, O. (1994b), Saddlepoint approximation for the least squares estimator in first-order autoregression, *Biometrika* 81, 807–11.
- Lugannani, R. and Rice, S.O. (1980). Saddlepoint approximations for the distribution of the sum of independent random variables. *Advances in Applied Probability* 12, 475–490.
- Martellosio, F. (2011) Efficiency of the OLS estimator in the vicinity of a spatial unit root. *Statistics & Probability Letters* 81, 1285–1291.
- Marsh, P.W.R. (1998), Saddlepoint approximations and non-central quadratic forms. *Econometric Theory* 14, 539–559.
- Muirhead, R. J. (1982) *Aspects of Multivariate Statistical Theory*, Wiley, New York.
- Ord, J.K. (1975) Estimation methods for models of spatial interaction, *Journal of the American Statistical Association* 70, 120–126.
- Phillips, P.C.B. (1978) Edgeworth and saddlepoint approximations in first order noncircular autoregression. *Biometrika* 65, 91–98.
- Reid, N. (1996) Likelihood and higher-order approximations to tail areas: A review and annotated bibliography, *The Canadian Journal of Statistics* 24, 141–166.
- Reid, N. (2003), The 2000 Wald Memorial Lecture: Asymptotics and the theory of inference, *The Annals of Statistics* 31, 1695–1731.
- Robinson, P.M. and Rossi, F. (2015) Refinements in maximum likelihood inference on spatial autocorrelation in panel data. *Journal of Econometrics* 189, 447–456.
- Wood, A.T.A., Booth, J.G. and Butler, R.W. (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* 188, 178–200.
- von Neumann, J. (1941) Distribution of the ratio of the mean square successive difference to the variance, *Annals of Mathematical Statistics* 12, 367–395.