



Discussion Papers in Economics
**EXACT PROPERTIES OF THE MAXIMUM LIKELIHOOD
ESTIMATOR IN
SPATIAL AUTOREGRESSIVE MODELS**

By

Grant Hillier,
(CeMMAP and University of Southampton)

&

Federico Martellosio
(University of Surrey)

DP 07/16

School of Economics
University of Surrey
Guildford
Surrey GU2 7XH, UK
Telephone +44 (0)1483 689380
Facsimile +44 (0)1483 689548
Web www.econ.surrey.ac.uk
ISSN: 1749-5075

Exact Properties of the Maximum Likelihood Estimator in Spatial Autoregressive Models

Grant Hillier

CeMMAP and University of Southampton

ghh@soton.ac.uk

Federico Martellosio*

University of Surrey

f.martellosio@surrey.ac.uk

May 2015

Abstract: The (quasi-) maximum likelihood estimator (QMLE) for the autoregressive parameter in a spatial autoregressive model cannot in general be written explicitly in terms of the data. The only known properties of the estimator have hitherto been its first-order asymptotic properties (Lee, 2004, *Econometrica*), derived under specific assumptions on the evolution of the spatial weights matrix involved. In this paper we 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. A number of immediate consequences of this result are discussed, and some examples are analyzed in detail. The examples are of interest in their own right, but also serve to illustrate some unexpected features of the distribution of the MLE. In particular, we show that the distribution of the MLE may not be supported on the entire parameter space, and may be nonanalytic at some points in its support.

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

JEL Classification: C12, C21.

1 Introduction

Spatial autoregressive processes have enjoyed considerable recent popularity in modelling cross-sectional data in economics and in several other disciplines, among which are geography, regional science, and politics.² In most applications, such models are

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

²For an introduction to spatial autoregressions see, e.g., LeSage and Pace (2009). Empirical applications of spatial autoregressions in economics can be found in Case (1991), Besley and Case (1995), Audretsch and Feldmann (1996), Bell and Bockstael (2000), Bertrand, Luttmer and Mullainathan (2000), Topa (2001), Pinkse, Slade, and Brett (2002), Liu, Patacchini, Zenou, and Lee (2015), to name just a few.

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 *exact* properties of the (quasi-)maximum likelihood estimator (MLE) for this parameter that is implied by assuming a Gaussian likelihood.

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

$$y = \lambda W y + X \beta + \sigma \varepsilon, \tag{1.1}$$

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 will 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. We refer to the model with the regression component ($X\beta$) missing as the pure SAR model.

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) propose higher-order refinements for the distribution of the QMLE. 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.

Despite the above, and other, contributions, there remain some compelling reasons for studying its exact properties - more so, perhaps, than usual. First, the exact distribution of the MLE for λ may possess important features that would be impossible to discover by Monte Carlo simulation or asymptotic methods - for example, non-differentiability, non-analyticity, or unboundedness of the density. Second, exact results can assist and complement simulation methods to check the accuracy of the available asymptotic results. This is particularly important in the present context, as asymptotic results depend on the assumptions made on how the spatial weights matrix evolves with the sample size.

Due to the fact that the QMLE for λ is in general unavailable in closed form, even the calculation of the QMLE has been regarded as problematic in this model, let alone the study of its exact properties. The key observation that enables us to carry out an exact analysis of the MLE is the observation that, under a condition that is

satisfied in particular whenever all eigenvalues of W are real, the profile likelihood for λ is *single-peaked* on the relevant parameter space. This fact is important in itself, because it simplifies the computation of the estimator greatly. And, it implies that the cdf of the MLE for λ can be written down in terms of the cdf of a quadratic form, notwithstanding the unavailability of the estimator in closed form. This representation of the cdf provides a starting point for a full exact analysis of the MLE, for an arbitrary distribution of ε . We show that the distribution theory for the QMLE is non-standard, and turns out to have key 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.

Because the exact distribution theory for the QMLE is complicated for general (W, X) , the present paper focuses on exact consequences of the cdf representation that hold under some restrictions on (W, X) . The virtue of imposing such restrictions lies in revealing important properties of the MLE that can be expected to hold more generally. For the case of arbitrary (W, X) our results are useful for computational purposes, and to derive approximations to the distribution of the MLE, but these matters are not pursued here.

The rest of the paper is organized as follows. Section 2 discusses the parameter space for λ , and introduces some examples that are used throughout the paper to illustrate the theoretical results. Section 3 rules out the cases when the MLE does not exist, or is non-random, and derives some key properties of the profile log-likelihood for λ . Section 4 presents the representation of the cdf of the MLE, 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. For these cases, we provide simple explicit formulae for the cdf of the MLE, which also prove useful to clarify the asymptotic behavior of the estimator in cases not covered by Lee’s (2004) assumptions.

All quantities considered in this paper are real unless otherwise noted. We denote the column space of a matrix A by $\text{col}(A)$, and its null space by $\text{null}(A)$. Finally, “a.s.” stands for almost surely, with respect to the Lebesgue measure on \mathbb{R}^n .

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.

There are two components to Assumption A. The crucial component is that W is not allowed to be nilpotent (a matrix is nilpotent if all its eigenvalues are zero). If W were nilpotent the study of the MLE would be trivial, as in that case maximizing the Gaussian likelihood is equivalent to minimizing the residual sum of squares $(S_\lambda y - X\beta)'(S_\lambda y - X\beta)$, and hence the MLE would coincide with the OLS estimator.⁴ The second component of Assumption A, 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.

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). The estimator we study in this paper is the maximum of the likelihood on 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 popularity 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. We refer to the

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.

⁴If W is nonnegative, as it usually is 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).

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., Kelejian et al., 2006, Lee, 2007). 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 the graph known as a *star*, a particularly important case in 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 two nonzero eigenvalues (1 and -1 , each with multiplicity 1), and $n - 2$ zero eigenvalues, so that $\Lambda = (-1, 1)$.

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

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

⁵Particularly in the unbalanced case, it may be preferable to use different autoregressive parameters for each group. Generalizations of this type will not be considered in the present paper.

where additive constants have been omitted. After maximizing $l(\beta, \sigma^2, \lambda)$ with respect to β and σ^2 we obtain the profile, or concentrated, log-likelihood

$$l_p(\lambda) := -\frac{n}{2} \ln(y'S'_\lambda M_X S_\lambda y) + \ln(|\det(S_\lambda)|), \quad (3.2)$$

where $M_X := I_n - X(X'X)^{-1}X'$. The profile log-likelihood function $l_p(\lambda)$ is a.s. differentiable for any λ such that $\det(S_\lambda) \neq 0$ (see Hillier and Martellosio, 2014b), with first derivative

$$\dot{l}_p(\lambda) = n \frac{y'W'M_X S_\lambda y}{y'S'_\lambda M_X S_\lambda y} - \text{tr}(G_\lambda), \quad (3.3)$$

where $G_\lambda := WS_\lambda^{-1}$. We note for future reference that the profile score can be rewritten as

$$\dot{l}_p(\lambda) = \frac{n}{2} \frac{y'S'_\lambda Q_\lambda S_\lambda y}{y'S'_\lambda M_X S_\lambda y}, \quad (3.4)$$

where

$$Q_\lambda := M_X C_\lambda + C'_\lambda M_X, \quad (3.5)$$

and

$$C_\lambda := G_\lambda - \frac{\text{tr}(G_\lambda)}{n} I_n. \quad (3.6)$$

We can now define the MLE 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 $l_p(\lambda)$ is a.s. defined for λ on the whole real line with the exception of a finite number of isolated points (and the unrestricted maximizer of $l_p(\lambda)$ can, in general, be anywhere on this set). The estimator we consider in this paper is

$$\hat{\lambda}_{\text{ML}} := \arg \max_{\lambda \in \Lambda} l_p(\lambda), \quad (3.7)$$

provided that the maximum exists and is unique. This is the MLE in most common use, but of course it might not be the MLE under a different specification of the parameter space for λ . In particular, several authors suggest that λ should be restricted to $(-1, 1)$ (see, e.g., Kelejian and Prucha, 2010). If W is nonnegative, $(-1, 1) \subseteq \Lambda$ by the Perron-Frobenius Theorem. If $(-1, 1)$ is a proper subset of Λ , the estimator $\tilde{\lambda}_{\text{ML}} := \arg \max_{\lambda \in (-1, 1)} l_p(\lambda)$ 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 $\tilde{\lambda}_{\text{ML}}$.

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.

Proposition 1. *If $M_X(\omega I_n - W) = 0$ for some real eigenvalue ω of W , then the profile score $\dot{l}_p(\lambda)$ does not depend on y . Also, for any real nonzero eigenvalue ω of W , $\lim_{\lambda \rightarrow \omega^{-1}} l_p(\lambda) = +\infty$ a.s. if $M_X(\omega I_n - W) = 0$, and $\lim_{\lambda \rightarrow \omega^{-1}} l_p(\lambda) = -\infty$ a.s. if $M_X(\omega I_n - W) \neq 0$.*

Proposition 1 implies that, if $M_X(\omega I_n - W) = 0$ for some real eigenvalue ω of W , a maximizer of $l_p(\lambda)$ is nonrandom, subject to existence. In fact, the proposition also says that a maximizer of $l_p(\lambda)$ may not even exist, in the sense that $l_p(\lambda)$ may be a.s. unbounded from above.⁶ Fortunately, the condition $M_X(\omega I_n - W) = 0$ is usually not met in applications. but does occur in some apparently reasonable models. Two examples follows.

Example 3. The weights matrix $W = I_r \otimes B_m$ of a balanced Group Interaction model (see Example 1 above) has two eigenspaces: $\text{col}(I_r \otimes \iota_m)$, associated to the eigenvalue 1, and its orthogonal complement, associated to the eigenvalue $\omega_{\min} = -1/(m - 1)$. For this W , $\text{col}(\omega_{\min} I_n - W) = \text{col}(I_r \otimes \iota_m)$ (directly from the definition of B_m). Thus, Proposition 1 implies that, if $\text{col}(I_r \otimes \iota_m) \subseteq \text{col}(X)$, then $l_p(\lambda)$ does not depend on y and $l_p(\lambda) \rightarrow +\infty$ as $\lambda \rightarrow \omega_{\min}^{-1}$. Since the matrix $I_r \otimes \iota_m$ represents group specific fixed effects, it follows that, in the balanced Group Interaction model, $\hat{\lambda}_{\text{ML}}$ fails to exist in the presence of group fixed effects.⁷

Example 4. Consider a symmetric or row-standardized Complete Bipartite model (see Example 2 above) with X containing an intercept for each of the two groups (and any other regressors). In that case $M_X W = 0$, so Proposition 1 applies with $\omega = 0$, and implies that $\hat{\lambda}_{\text{ML}}$ is a constant.

In the rest of the paper we assume that, unless otherwise specified, $M_X(\omega I_n - W) \neq 0$ for any real eigenvalue ω of W . This amounts to ruling out the pathological cases when $\hat{\lambda}_{\text{ML}}$ does not exist or does not depend on the data y . For a detailed analysis of the identifiability failure that occurs when $M_X(\omega I_n - W) = 0$ see Hillier and Martellosio (2014b).

3.2 The First-Order Condition

Since $l_p(\lambda)$ is almost surely differentiable on Λ , and Λ is an open set, $\hat{\lambda}_{\text{ML}}$, if it exists, must be a root of the first-order condition $\dot{l}_p(\lambda) = 0$.

Lemma 1. *The first-order condition $\dot{l}_p(\lambda) = 0$ is a.s. equivalent to a polynomial equation of degree equal to the number of distinct eigenvalues of W .*

⁶In the case when $\lim_{\lambda \rightarrow \omega^{-1}} l_p(\lambda) = +\infty$ one could define $\hat{\lambda}_{\text{ML}} = \omega^{-1}$, but note that this would be an estimator not depending on y .

⁷See Lee (2007) for a different perspective on the inferential problem in a balanced Group Interaction model with fixed effects.

Thus, the equation $\dot{l}_p(\lambda) = 0$ has a number of complex roots (counting multiplicities) equal to the number of distinct eigenvalues of W , for any W . Of these roots, any real one lying 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 first-order condition $\dot{l}_p(\lambda) = 0$ a.s. has a single solution in Λ , which corresponds to the maximum of $l_p(\lambda)$.*

Geometrically, Lemma 2 says that, under the stated condition, *the profile log-likelihood $l_p(\lambda)$ is a.s. single-peaked* on Λ , with no stationary inflection points. Although this is not our main focus here, the result has clear computational advantages, as it greatly simplifies numerical optimization of the likelihood.

The proof of Lemma 2 relies on an application of a Cauchy-Schwarz inequality, for which only the condition $\text{tr}(C_\lambda^2) > 0$ for all $\lambda \in \Lambda$ is needed. Note that the condition depends only on W , not on X . Importantly, it is satisfied whenever W has *only real eigenvalues*, which is often the case in applications.⁸ 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 5. 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 $l_p(\lambda)$ is a.s. single-peaked by Lemma 2. Other configurations of the a_i can induce multi-peakedness of $l_p(\lambda)$. 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,

⁸If 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 .

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 $l_p(\lambda)$ 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) that $l_p(\lambda)$ is multi-peaked. Figure 3.2 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 . One can check by simulation that, whatever the value of X , $a_4 = 0$ entails a high probability of multi-peakedness.

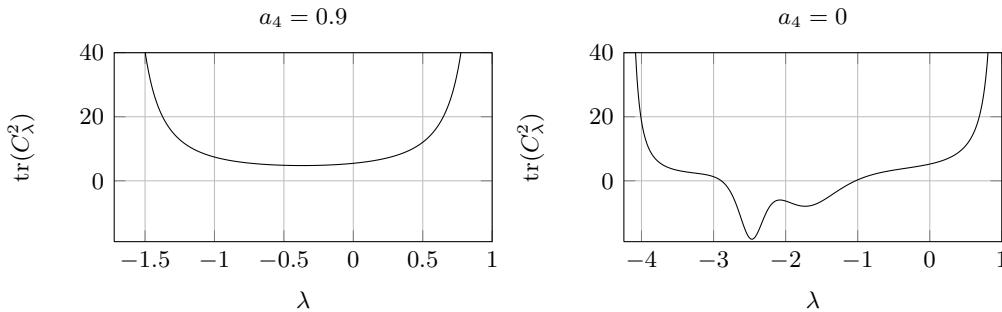


Figure 1: $\text{tr}(C_\lambda^2)$, $\lambda \in \Lambda$, for the weights matrix W in Example 5.

A complete understanding of what causes multi-peakedness 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 first-order condition $\dot{l}_p(\lambda) = 0$ a.s. has at most one solution in $[0, 1)$, which corresponds to a local maximum of $l_p(\lambda)$.*

Thus, provided that W is nonnegative and that $l_p(\lambda)$ has a peak in the open set $(0, 1)$, $l_p(\lambda)$ is unimodal on that interval. In other words, multi-peakedness must always involve peaks at negative values of λ when W is nonnegative. This result is of interest for applications in which it may be natural to restrict attention to positive values of λ .

4 Distributional Properties of the QMLE

4.1 The Main Theorem

The unimodality property established in Lemma 2 has direct consequences for the distribution of $\hat{\lambda}_{\text{ML}}$. Indeed, unimodality of the profile log-likelihood $l_p(\lambda)$ on Λ implies that the single peak of $l_p(\lambda)$ is to the left of a point $z \in \Lambda$ if and only if the slope of $l_p(\lambda)$ at $\lambda = z$ is negative. This observation allows us to derive the main result of the paper.

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 a 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 for any distribution of y (not necessarily that induced by the SAR model). The result reduces the study of $\hat{\lambda}_{\text{ML}}$, an estimator that is generally unavailable in closed form, to the study of a very simple statistic, a quadratic form in y . There are a number of advantages of this representation. In simple cases the result can deliver an explicit formula for the cdf of $\hat{\lambda}_{\text{ML}}$ (examples will be provided below). More generally, the theorem does not provide explicit formulae directly, but it is still useful for computational purposes and to study the distribution of the $\hat{\lambda}_{\text{ML}}$. In the present paper we are interested in *exact* consequences of Theorem 1, but before moving to that, we mention the other possible uses of Theorem 1.

First, 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 directly maximize the likelihood*. This is useful, for example, to study by simulation how the model characteristics (parameters, W and X , and, distribution of ε) affect the distribution of $\hat{\lambda}_{\text{ML}}$.

Second, Theorem 1 can be used to estimate the true cdf of $\hat{\lambda}_{\text{ML}}$, and in particular to construct confidence intervals. This can be done by drawing y from an estimate of the distribution of y (for example by replacing the parameters indexing the distribution of y with some estimates), or by the bootstrap. Note that deriving the bootstrap distribution of $\hat{\lambda}_{\text{ML}}$ directly (that is, without relying on Theorem 1) can be computationally very intensive, given the need to repeatedly maximize the likelihood. Using Theorem 1 it is possible to bootstrap a quadratic form instead, a computationally trivial task.

Third, Theorem 1 provides a direct route to obtaining a higher-order asymptotic approximation to the distribution of $\hat{\lambda}_{\text{ML}}$ - for example, by using a saddlepoint approximation for the distribution of the quadratic form $y' S'_z Q_z S_z y$.⁹

Computational issues and the derivation of approximations are not the focus of this paper. Instead, starting from the next section we study exact consequences of Theorem 1. Not surprisingly, such analysis requires imposing additional structure on the model, which we will do gradually.

⁹Subject to suitable conditions, the first-order asymptotic distribution of $\hat{\lambda}_{\text{ML}}$ can also be obtained 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.

4.2 Some Exact Consequences

4.2.1 General Distributional Properties

Quadratic forms have been much studied in the statistical and econometric literature. Their distribution theory is generally complicated, but some general distributional properties of $\hat{\lambda}_{\text{ML}}$ follow from that literature. The results in Mulholland (1965) and Saldanha and Tomei (1996) show that in general there will be a number of points $z \in \Lambda$ at which the distribution function of $y'S'_z Q_z S_z y$ is non-analytic at the origin. By Theorem 1, it follows that the distribution function of $\hat{\lambda}_{\text{ML}}$ is non-analytic at these values of z , and has a different functional form in the intervals between such points. Importantly, this result does not depend on the distribution assumptions made (see Forchini, 2002). This property of the distribution of $\hat{\lambda}_{\text{ML}}$ is not a mere curiosity: for *any* (W, X) there will usually be a number of points $z \in \Lambda$ at which the cdf of $\hat{\lambda}_{\text{ML}}$ is non-analytic. 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).

Letting $h := y/(y'y)^{1/2}$, a vector distributed on the unit sphere in n dimensions, \mathcal{S}^{n-1} , rewrite (4.1) as $\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr(h'S'_z Q_z S_z h \leq 0)$. This representation allows one to appeal to known results for quadratic forms defined on the sphere (see, e.g., Hillier, 2001, and Forchini, 2005). Such expressions are too complicated for our purposes, but progress can be made by exploiting the special structure of the matrix of the quadratic form.

Let us now, and for the rest of the paper, assume that the distribution of y is that induced by the SAR model, i.e., $S_\lambda y = X\beta + \sigma\varepsilon =: \tilde{y}$. It is convenient to rewrite (4.1) as

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr(\tilde{y}'A(z, \lambda)\tilde{y} \leq 0), \quad (4.2)$$

where

$$A(z, \lambda) := (S_z S_\lambda^{-1})' Q_z (S_z S_\lambda^{-1}). \quad (4.3)$$

The structure of the matrix $A(z, \lambda)$ is evidently crucial in determining the properties of $\hat{\lambda}_{\text{ML}}$. In particular, if $\varepsilon \sim N(0, I_n)$, a spectral decomposition of $A(z, \lambda)$ shows that $\tilde{y}'A(z, \lambda)\tilde{y}$ is distributed as a linear combination of independent (possibly non-central) χ^2 variates, with coefficients the distinct eigenvalues of $A(z, \lambda)$. This would be the “crudest” use of Theorem 1. However, by exploiting the special structure of $A(z, \lambda)$, and imposing some conditions on the relationship between W and X , it is possible to be much more precise.

4.2.2 The Case When W Is Similar to a Symmetric Matrix

We begin by restricting attention to the case when W is similar to a symmetric matrix, a condition that is equivalent to W being diagonalizable and having only real

eigenvalues. It is easy to see that W is similar to a symmetric matrix, for example, whenever it is obtained by row-standardizing a symmetric matrix.¹⁰

Some new notation is needed. Let T denote the number of distinct eigenvalues of W . If the distinct eigenvalues of W are all real we denote them by, in ascending order, $\omega_1, \omega_2, \dots, \omega_T$, the eigenvalue ω_t occurring with algebraic multiplicity n_t (so that $\omega_1 = \omega_{\min}$, $\omega_T = 1$). Also let $g_t(z) := \omega_t / (1 - z\omega_t)$ and $\gamma_t(z) := g_t(z) - \text{tr}(G_z) / n$, for $t = 1, \dots, T$, be the distinct eigenvalues of, respectively, G_z , and C_z . If W is similar to a symmetric matrix we can write $W = HDH^{-1}$, with H a nonsingular matrix (orthogonal if W is symmetric) whose columns are the eigenvectors of W , and $D := \text{diag}(\omega_t I_{n_t}, t = 1, \dots, T)$. We denote by M_{st} the $n_s \times n_t$ submatrix of $M := H' M_X H$ associated to the eigenvalues ω_s and ω_t . Writing $x := H^{-1} \tilde{y}$ and partitioning x conformably with the partition of M (so that x_t is $n_t \times 1$, for $t = 1, \dots, T$), we obtain the following result.

Proposition 2. *In a SAR model with W similar to a symmetric matrix,*

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr\left(\sum_{t=1}^T d_{tt}(z, \lambda) x_t' M_{tt} x_t + 2 \sum_{s,t=1, s>t}^T d_{st}(z, \lambda) x_s' M_{st} x_t \leq 0\right), \quad (4.4)$$

for any $z \in \Lambda$, where the coefficients $d_{st}(z, \lambda)$ are given by

$$d_{st}(z, \lambda) := \frac{(1 - z\omega_s)(1 - z\omega_t)}{(1 - \lambda\omega_s)(1 - \lambda\omega_t)} [\gamma_s(z) + \gamma_t(z)] = d_{ts}(z, \lambda). \quad (4.5)$$

Proposition 2 provides a very general representation of the cdf of $\hat{\lambda}_{\text{ML}}$ in terms of a linear combination of simple quadratic and bilinear forms in the vectors x_t . The next result focuses on the cases when representation (4.4) contains only quadratic forms in the x_t , and subvectors of them.

Proposition 3. *(i) In a SAR model with W similar to a symmetric matrix, the bilinear terms in (4.4) all vanish if and only if the matrix $M_X W$ is symmetric. In that case, for any $z \in \Lambda$,*

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr\left(\sum_{t=1}^T d_{tt}(z, \lambda) x_t' M_{tt} x_t \leq 0\right). \quad (4.6)$$

(ii) In a SAR model, if W and $M_X W$ are both symmetric (4.6) simplifies further to

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr\left(\sum_{t=1}^T d_{tt}(z, \lambda) \tilde{x}_t' \tilde{x}_t \leq 0\right), \quad (4.7)$$

¹⁰If R is a diagonal matrix with the row sums of the symmetric matrix A on the diagonal, then the row-standardised matrix $W = R^{-1} A = R^{-1/2} (R^{-1/2} A R^{-1/2}) R^{1/2}$ is similar to the symmetric matrix $R^{-1/2} A R^{-1/2}$.

where \tilde{x}_t is a subvector of x_t of dimension $n_t - n_t(X)$, where $n_t(X)$ is the number of columns of X in the eigenspace associated to ω_t . The vector \tilde{x}_t contains those elements of x_t that correspond to eigenvectors not in $\text{col}(X)$.

The highly structured representations of the cdf in the Propositions 2 and 3 provide the basis for some of the subsequent analysis. The condition that $M_X W$ is symmetric is certainly restrictive in applications, but has the merit to uncover unexpected properties of the QMLE, starting from the support property discussed in the next section. An example when $M_X W$ is symmetric is an unbalanced group interaction model with fixed effects, and all β coefficients varying across groups (that is, X is the direct sum of $m_i \times k_i$ matrices X_i , $i = 1, \dots, r$, with each X_i containing at least a column of ones).¹¹ Other examples will be met later in the paper.

4.2.3 Support of the QMLE

We now discuss another consequence of Theorem 1: *the support of $\hat{\lambda}_{\text{MLE}}$ is not necessarily the entire interval Λ .*¹² This is an unexpected phenomenon that has not been noticed previously, to the best of our knowledge. While it seems difficult to specify general conditions on W and X that lead to restricted support for $\hat{\lambda}_{\text{MLE}}$, it turns out that in the context of Proposition 3 (i) the conditions that do so are straightforward, and we confine ourselves here to that case. The assumptions underlying Proposition 3 (i) are restrictive, but do provide examples when the phenomenon occurs, along with an intuitive interpretation.

To begin with, observe that the first-order condition $\dot{l}_p(\lambda) = 0$ implies that the only possible candidates for the QMLE are the values of λ for which the matrix Q_λ is indefinite (see equation (3.4)). 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 ($\Pr(\hat{\lambda}_{\text{MLE}} \leq z) = 0$), or certain ($\Pr(\hat{\lambda}_{\text{MLE}} \leq z) = 1$). In such cases the support of $\hat{\lambda}_{\text{MLE}}$ 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 $\hat{\lambda}_{\text{MLE}}$ is not supported on the whole 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 $t = 2, \dots, T-1$, z_t denotes the point $z \in \Lambda$ at which $\gamma_t(z) = 0$ (existence and uniqueness of z_t is established by Lemma A.1 in Appendix A).

¹¹See also Hillier and Martellosio (2014a). Note that unbalancedness is essential here, because, as we have seen in Section 3.1, the MLE does not exist in the balanced case when group fixed effects are present.

¹²By support of (the distribution of) $\hat{\lambda}_{\text{MLE}}$ we mean the set on which the density of $\hat{\lambda}_{\text{MLE}}$ is positive, assuming that the density exists.

Proposition 4. *Assume that in a SAR model W is similar to a symmetric matrix and $M_X W$ is symmetric.*

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

We now provide some intuition, and some examples, for Proposition 4. 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).¹³ Now, the eigenvectors of W associated to positive eigenvalues can be interpreted as capturing all positive spatial autocorrelation (as measured by the statistic $u'Wu/u'u$) 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 4 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 fixed effects, and all β coefficients varying across groups (see the end of Section 4.2.2). Also, 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 asymptotic approximations to its distribution that are supported on the entire interval Λ are unlikely to be satisfactory.

4.3 Invariance Properties

This section derives some general properties of the QMLE for λ that can be deduced directly from the invariance properties of the model and of the first-order equation $\dot{l}_p(\lambda) = 0$. To begin with, observe that the profile score equation, and hence $\hat{\lambda}_{\text{ML}}$, is invariant to scale transformations $y \rightarrow \kappa y$, for any $\kappa > 0$, in the sample space. A first important consequence of this type of invariance is stated in the next proposition, where a scale mixture of the pdf $p(y)$ (assuming it exists) is $\int_{\kappa} p(\kappa y)g(\kappa)d\kappa$, where $g(\kappa)$ is the pdf of κ .

Proposition 5. *The distribution of $\hat{\lambda}_{\text{ML}}$ induced by a particular distribution of y is the same for all scale mixtures of that distribution.*

¹³This is because, in that case, z_t in Proposition 4 (i) must be nonpositive, by Lemma A.1 in Appendix A and the fact that $\gamma_t(0) = \omega_t \leq 0$.

A second consequence of the invariance with respect to scale transformations in the sample space is a reduction in the number of parameters indexing the distribution of $\hat{\lambda}_{\text{ML}}$. Suppose the distribution of ε depends on a parameter θ . A subspace \mathcal{U} of \mathbb{R}^n is said to be an *invariant subspace* of a matrix M if $Mu \in \mathcal{U}$ for every $u \in \mathcal{U}$.

Proposition 6. *Assume that the distribution of ε in a SAR model does not depend on β or σ^2 . Then, (i) if $\text{col}(X)$ is not an invariant subspace of W , the distribution of $\hat{\lambda}_{\text{ML}}$ depends on $\beta, \lambda, \sigma^2, \theta$ only through $\beta/\sigma, \lambda, \theta$; (ii) if $\text{col}(X)$ is an invariant subspace of W , the distribution of $\hat{\lambda}_{\text{ML}}$ depends only on λ, θ .*

Note that in case (ii) the distribution of $\hat{\lambda}_{\text{ML}}$ does not depend on σ . Hence, in that case, the distribution of $\hat{\lambda}_{\text{ML}}$ induced by a particular distribution of ε is the same for all scale mixtures of that distribution.

In applications, $\text{col}(X)$ is generally not an invariant subspace of W , and therefore the distribution of $\hat{\lambda}_{\text{ML}}$ depends not only on λ, θ but also on β/σ . Possibly the simplest nontrivial (i.e., $\text{col}(X) \neq \{0\}$) case in which $\text{col}(X)$ is not an invariant subspace of W is a SAR model with $X = \iota_n$ and a W with constant row sums. More generally, an easy to check necessary and sufficient condition for $\text{col}(X)$ to be an invariant subspace of W is $M_X W X = 0$.¹⁴

5 Applications

This section provides simple illustrations of the various aspects of the distribution of $\hat{\lambda}_{\text{ML}}$ we have studied. We assume that ε belongs to the family of scale mixtures of $N(0, I_n)$, denoted by $\varepsilon \sim \text{SMN}(0, I_n)$. Note that these are spherically symmetric distributions for ε , which need not be i.i.d. Also, we focus on the pure model, or the model with a constant mean. Section 5.1 analyzes the case of symmetric W . Then, in Sections 5.2 and 5.3 we consider the balanced Group Interaction model and the Complete Bipartite model, respectively.¹⁵

5.1 Mixed-Gaussian Pure SAR Model with Symmetric W

We now study in some detail the distribution of $\hat{\lambda}_{\text{ML}}$ in a pure SAR model with symmetric W and $\varepsilon \sim \text{SMN}(0, I_n)$. As we shall see, in this case it is possible to derive a relatively simple explicit formula for the cdf of $\hat{\lambda}_{\text{ML}}$.

¹⁴Note that $M_X W X = 0$ if $M_X W$ is symmetric, the condition used in Proposition 3.

¹⁵For the balanced Group Interaction model, and the Complete Bipartite model, $\hat{\lambda}_{\text{ML}}$ is the unique root in Λ of either a quadratic or a cubic (by Lemma 1), and is therefore available in closed form. However, obtaining the exact distribution from such a closed form seems exceedingly difficult. Theorem 1 provides a much more convenient approach.

Propositions 3 (ii) and 6 (ii) imply that for a SAR model with symmetric W , $\varepsilon \sim \text{SMN}(0, I_n)$, and $\text{col}(X)$ spanned by k linearly independent eigenvectors of W ,¹⁶

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr\left(\sum_{t=1}^T d_{tt}(z, \lambda) \chi_{n_t - n_t(X)}^2 \leq 0\right), \quad (5.1)$$

where the $\chi_{n_t}^2$ variates are independent, and, from expression (4.5), $d_{tt}(z, \lambda) = 2\gamma_t(z)(1 - z\omega_t)^2 / (1 - \lambda\omega_t)^2$. Here and elsewhere, χ_ν^2 denotes a (central) χ^2 random variable with ν degrees of freedom, and we use the convention that $\chi_0^2 = 0$. In the rest of this section we focus, for simplicity, on the particular case of a pure model, which we state as a theorem.

Theorem 2. *In a pure SAR model with symmetric W and $\varepsilon \sim \text{SMN}(0, I_n)$, for any $z \in \Lambda$,*

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr\left(\sum_{t=1}^T d_{tt}(z, \lambda) \chi_{n_t}^2 \leq 0\right), \quad (5.2)$$

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

The number T of distinct eigenvalues of W must be at least two, by Assumption A. If $T = 2$ (i.e., W has only two distinct eigenvalues), equation (5.2) gives

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr\left(F_{n_1, n_2} \leq -\frac{n_2 d_{22}(z, \lambda)}{n_1 d_{11}(z, \lambda)}\right), \quad (5.3)$$

where F_{ν_1, ν_2} denotes a random variable with an F-distribution on (ν_1, ν_2) degrees of freedom. Thus, when $T = 2$ the cdf is remarkably simple, and there is no point of non-analyticity in this case (cf. Section 4.2.1 above). We will see in Section 5 that the balanced Group Interaction model has this form. Moving to the case $T > 2$, Lemma A.1 in Appendix A says that each coefficient $d_{tt}(z, \lambda)$, $t = 2, \dots, T - 1$, changes sign exactly once on Λ , at the point points z_t where the eigenvalue $\gamma_t(z)$ of C_z changes sign. By an extension of the argument in Saldanha and Tomei (1996), this implies that the distribution function of $\hat{\lambda}_{\text{ML}}$ in pure SAR models with symmetric W is non-analytic at these $T - 2$ points, and has a different functional form on each interval between those points.¹⁷ We may now use this fact to obtain an explicit form for the cdf (5.2).

¹⁶To see this just note that $\tilde{y} \sim \text{SMN}(0, I_n)$ and $x \sim \text{SMN}(0, (H'H)^{-1})$. But $H'H = I_n$ if W is symmetric. Also, note that if W and $M_X W$ are both symmetric, then $\text{col}(X)$ must be spanned by k linearly independent eigenvectors of W .

¹⁷Saldanha and Tomei (1996) consider a matrix with fixed eigenvalues, and vary the point at which the cdf is to be evaluated. In our case, the point on the cdf is fixed (zero), but the eigenvalues are (continuous) functions of z - they are the $d_{tt}(z, \lambda)$. Reinterpreted, their theorem says that, for any fixed $\lambda \in \Lambda$, whenever an eigenvalue $d_{tt}(z, \lambda)$ changes sign as z varies, the cdf will be non-analytic at the origin, the point of interest for us.

Some new notation is needed. For a fixed $z \in \Lambda$ at which none of the $d_{tt}(z, \lambda)$ vanishes, let T_1 and T_2 denote the numbers of positive and negative terms $d_{tt}(z, \lambda)$, respectively, in (5.2), with the T_1 positive terms first. Let $v_1 := \sum_{t=1}^{T_1} n_t$ and $v_2 := \sum_{t=T_1+1}^T n_t$, with $v_1 + v_2 = n$. The numbers T_1 and T_2 vary with z , as do v_1 and v_2 . Next, let A_1 be the $v_1 \times v_1$ matrix $\text{diag}(d_{tt}(z, \lambda)I_{n_t}; t = 1, \dots, T_1)$, and A_2 the $v_2 \times v_2$ matrix $\text{diag}(-d_{tt}(z, \lambda)I_{n_t}; t = T_1 + 1, \dots, T)$. Finally, we denote by $C_j(A)$ the top-order zonal polynomial of order j in the eigenvalues of a matrix A (Muirhead, 1982, Chapter 7). Given this notation, the following result gives an explicit formula for the cdf (5.2), for any $T \geq 2$.

Corollary 1. *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 $T - 1$ intervals in Λ determined by the points of non-analyticity z_t ,*

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = [\det(\tau_1 A_1) \det(\tau_2 A_2)]^{-\frac{1}{2}} \times \sum_{j,k=0}^{\infty} \frac{\left(\frac{1}{2}\right)_j \left(\frac{1}{2}\right)_k}{j!k!} C_j(\tilde{A}_1) C_k(\tilde{A}_2) \Pr\left(F_{v_1+2j, v_2+2k} \leq \frac{(v_2 + 2k) \tau_1}{(v_1 + 2j) \tau_2}\right), \quad (5.4)$$

where $\tau_i := \text{tr}(A_i^{-1})$ and $\tilde{A}_i := I_{v_i} - (\tau_i A_i)^{-1}$, for $i = 1, 2$.

The top-order zonal polynomials in equation (5.4) can be computed very efficiently by methods described recently in Hillier, Kan, and Wang (2009). Because the matrices A_1 and A_2 vary as z varies over Λ , it is probably impossible to obtain the density function of $\hat{\lambda}_{\text{ML}}$ directly from (5.4), but we this problem can often be avoided by a conditioning argument (see, for example, Section 5.3).

We now discuss some further consequences of Theorem 2. The *spectrum* of an $n \times n$ matrix is defined to be the multiset of its n eigenvalues, each eigenvalue appearing with its algebraic multiplicity. Matrices with the same spectrum are called *cospectral*. According to equation (5.2), the distribution of $\hat{\lambda}_{\text{ML}}$, and hence all of its properties, depends on W only through its spectrum. Two illustrative applications of this observation follow.

Example 6 (Complete Bipartite model). The spectrum of the weights matrix (2.3) of a symmetric Complete Bipartite model depends on p and q only through their sum n . Hence, the distribution of $\hat{\lambda}_{\text{ML}}$ is the same for any pure symmetric mixed-Gaussian Complete Bipartite model on n observational units, regardless of the partition of n into p and q .

Example 7 (Star Graph). In case p or q in a Complete Bipartite model is 1 (i.e., the graph is a star graph), we may also consider the class of all symmetric weights matrices that are “compatible” with a star graph on n vertices (i.e., matrices having

positive (i, j) -th entry if and only if (i, j) is an edge of the star graph). It is a simple exercise to show that all such weights matrices have (after normalization so that the largest positive eigenvalue is 1) eigenvalues 0, with multiplicity $n - 2$, and $-1, 1$, and hence are cospectral with the adjacency matrix of the graph. Hence, the distribution of $\hat{\lambda}_{\text{ML}}$ is the same for any pure mixed-Gaussian SAR model with symmetric weights matrix compatible with a star graph.

Another consequence of representation (5.2) can be deduced for matrices W with *symmetric spectrum*. The spectrum of a matrix is said to be symmetric if, whenever ω is an eigenvalue, $-\omega$ is also an eigenvalue, with the same algebraic multiplicity. The weights matrix of a balanced Group Interaction model with $m = 2$ is an example of this type, as is that of the symmetric Complete Bipartite model.¹⁸

Corollary 2. *In a pure SAR model with $\varepsilon \sim \text{SMN}(0, I_n)$, W symmetric, and the spectrum of W symmetric about the origin, the density of $\hat{\lambda}_{\text{ML}}$ satisfies the symmetry property $\text{pdf}_{\hat{\lambda}_{\text{ML}}}(z; \lambda) = \text{pdf}_{\hat{\lambda}_{\text{ML}}}(-z; -\lambda)$.*

Note that this type of symmetry implies, in particular, that (subject to its existence) the mean of $\hat{\lambda}_{\text{ML}}$ satisfies $E(\hat{\lambda}_{\text{ML}}; \lambda) = -E(\hat{\lambda}_{\text{ML}}; -\lambda)$.

5.2 The Balanced Group Interaction Model

The general theory in this paper gives strikingly simple results for the balanced Group Interaction model. This is mainly a consequence of the fact that the weights matrix $W = I_r \otimes B_m$ has only two distinct eigenvalues. Letting

$$\theta(z) := \left(\frac{z + m - 1}{1 - z} \right)^2,$$

for any $z \in \Lambda$, we obtain the following expressions for the cdf and pdf of $\hat{\lambda}_{\text{ML}}$.

Proposition 7. *In the pure balanced Group Interaction model with $\varepsilon \sim \text{SMN}(0, I_n)$,*

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr\left(F_{r, r(m-1)} \leq \frac{\theta(z)}{\theta(\lambda)}\right), \quad (5.5)$$

and

$$\text{pdf}_{\hat{\lambda}_{\text{ML}}}(z; \lambda) = \frac{2m [(m-1)\theta(\lambda)]^{\frac{r(m-1)}{2}}}{B\left(\frac{r}{2}, \frac{r(m-1)}{2}\right)} \frac{(1-z)^{r(m-1)-1} (z+m-1)^{r-1}}{[(1-z)^2(m-1)\theta(\lambda) + (z+m-1)^2]^{\frac{rm}{2}}}, \quad (5.6)$$

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

¹⁸In fact, for any matrix W that is the adjacency matrix of a graph, it is known that the spectrum is symmetric if and only if the graph is bipartite.

Proposition 7 enables a complete analysis of the exact properties of $\hat{\lambda}_{\text{ML}}$, and the results needed for inference based upon it. For example, exact expressions for the moments and the median of $\hat{\lambda}_{\text{ML}}$, and exact confidence intervals for λ based on $\hat{\lambda}_{\text{ML}}$ can be obtained quite directly; see Hillier and Martellosio (2014a) for details.

Proposition 7 also immediately implies that a necessary condition for consistency of $\hat{\lambda}_{\text{ML}}$ is that $F_{r,r(m-1)} \rightarrow_p 1$. Hence, as already known in the literature, $r \rightarrow \infty$ is sufficient for consistency of $\hat{\lambda}_{\text{ML}}$, but $m \rightarrow \infty$ may not.¹⁹ 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 his results leave open that $\hat{\lambda}_{\text{ML}}$ may be inconsistent in this case. This is an example of so-called infill asymptotics. Equation (5.5), 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 *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 $(-\infty, 1)$. 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 exact result now available, and simple, under mixed-Gaussian assumptions there is no need to invoke either form of asymptotic approximation.

Figure 2 displays the exact density (5.6) 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 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 due to the fact that the left extreme of the support of $\hat{\lambda}_{\text{ML}}$ depends on m).

The results given in Proposition 7 for the pure balanced Group Interaction model are modified only slightly in the case of an unknown constant mean. Indeed, for the

¹⁹ $E(F_{r,r(m-1)}) \rightarrow 1$ as either r or $m \rightarrow \infty$, but $\text{var}(F_{r,r(m-1)}) \rightarrow 0$ when $r \rightarrow \infty$, but not when $m \rightarrow \infty$.

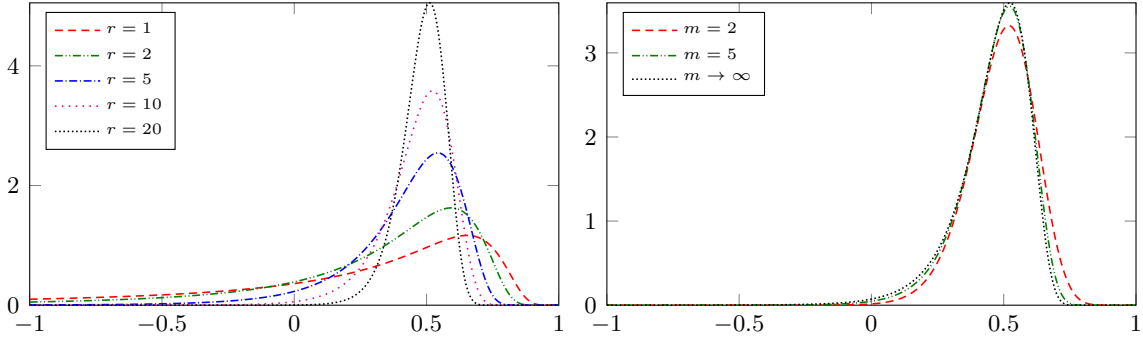


Figure 2: Density of $\hat{\lambda}_{\text{ML}}$ for the Gaussian pure balanced Group Interaction model with $\lambda = 0.5$, and with $m = 10$ (left panel), $r = 10$ (right panel).

balanced Group Interaction model with $X = \iota_n$ and $\varepsilon \sim \text{SMN}(0, I_n)$, equation (5.1) gives

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

Extensions to more general matrices X and to the unbalanced case are considered in Hillier and Martellosio (2014a). One important difference between balanced and unbalanced cases is that distribution of $\hat{\lambda}_{\text{ML}}$ is smooth over Λ in the former case, but contains points of non-analyticity in the latter.

5.3 The Complete Bipartite Model

We now apply the general results to the Complete Bipartite model introduced in Section 2.2. Section 5.3.1 discusses the pure case with symmetric W , which provides a particularly simple illustration of the general theory. Then, in Section 5.3.2, we consider the case of row-standardized W and constant mean (i.e., $X = \iota_n$). This latter case provides an example of the restricted support phenomenon described in Section 4.2.3.

5.3.1 Symmetric W , Zero Mean

In the symmetric Complete Bipartite model, W again has three distinct eigenvalues: $-1, 0, 1$. According to Corollary 1, the pdf of $\hat{\lambda}_{\text{ML}}$ in the pure Gaussian case is analytic everywhere on $\Lambda = (-1, 1)$ except at the point z_2 , and it is readily verified that $z_2 = 0$. Moreover, since the spectrum of W is symmetric, the symmetry established in Corollary 2 may be used to obtain the density for $z \in (-1, 0)$ from that for $z \in (0, 1)$.

Proposition 8. *In the pure symmetric Complete Bipartite model with $\varepsilon \sim \text{SMN}(0, I_n)$,*

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr(\phi_1 \chi_1^2 \leq \phi_2 \chi_1^2 + 2z \chi_{n-2}^2), \quad (5.7)$$

for $-1 < z < 1$, where

$$\phi_1 := \frac{(1-z)^2 [n + (n-2)z]}{(1-\lambda)^2}, \phi_2 := \frac{(1+z)^2 [n - (n-2)z]}{(1+\lambda)^2},$$

and the three χ^2 random variables involved are independent.

For $z \in (0, 1)$ the corresponding density is

$$\begin{aligned} \text{pdf}_{\hat{\lambda}_{\text{ML}}}(z; \lambda) = \frac{B\left(\frac{1}{2}, \frac{n}{2}\right) c}{2\pi a^{\frac{1}{2}} (1+c)^{\frac{n}{2}}} & \left[\frac{\alpha \dot{a}}{a} {}_2F_1\left(\frac{n}{2}, \frac{3}{2}, \frac{n+1}{2}; \eta\right) \right. \\ & \left. + \frac{\beta \dot{c}}{c} {}_2F_1\left(\frac{n}{2}, \frac{1}{2}, \frac{n+1}{2}; \eta\right) \right], \quad (5.8) \end{aligned}$$

where $a := \phi_2/\phi_1$, $c := 2z/\phi_1$, $\eta := \phi_1(\phi_2 - 2z)/[\phi_2(\phi_1 + 2z)]$, and ${}_2F_1(\cdot)$ denotes the Gauss hypergeometric function. For $z \in (-1, 0)$ the density is defined by $\text{pdf}_{\hat{\lambda}_{\text{ML}}}(z; \lambda) = \text{pdf}_{\hat{\lambda}_{\text{ML}}}(-z; -\lambda)$.

Proposition 8 confirms that the properties of $\hat{\lambda}_{\text{ML}}$, depends on p and q only through their sum n - cf. Example 6. We also note that taking $z = 0$ in (5.7) gives $\Pr(\hat{\lambda}_{\text{ML}} \leq 0) = \Pr(|\xi| \leq (1-\lambda)(1+\lambda))$, where ξ has a Cauchy distribution. This very simple formula for the probability that $\hat{\lambda}_{\text{ML}}$ is negative does not depend on the sample size.

The asymptotic distribution as $n \rightarrow \infty$ can be obtained easily, as follows. For every fixed $z \in \Lambda$, the characteristic function of the random variable $V_n := (\phi_1 \chi_1^2 - \phi_2 \chi_1^2 - 2z \chi_{n-2}^2)/(n-2)$ is easily seen to converge to that of

$$\bar{V}_n := \bar{\phi}_1 \chi_1^2 - \bar{\phi}_2 \chi_1^2 - 2z,$$

where $\bar{\phi}_1 := \lim_{n \rightarrow \infty} (\phi_1/(n-2)) = (1-z)^2(1+z)/(1-\lambda)^2$ and $\bar{\phi}_2 := \lim_{n \rightarrow \infty} (\phi_2/(n-2)) = (1+z)^2(1-z)/(1+\lambda)^2$. Therefore, $V_n \rightarrow_d \bar{V}_n$, and so, from Proposition 8,

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) \rightarrow \Pr(\chi_1^2 \leq \psi_1 \chi_1^2 + \psi_2),$$

with

$$\psi_1 := \left(\frac{1+z}{1-z}\right) \left(\frac{1-\lambda}{1+\lambda}\right)^2, \psi_2 := \frac{2z(1-\lambda)^2}{(1+z)(1-z)^2},$$

for $z \in (0, 1)$, and the two χ_1^2 variates are independent. For $z \in (0, 1)$, therefore, by conditioning on $q_1 \equiv \chi_1^2$,

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) \rightarrow \mathbb{E}_{q_1} [\mathcal{G}_1(\psi_1 q_1 + \psi_2)]. \quad (5.9)$$

Thus, as in the case when $m \rightarrow \infty$ in a balanced Group Interaction model, $\hat{\lambda}_{\text{ML}}$ is not consistent, but converges in distribution to a random variable as $n \rightarrow \infty$. The limiting density can be obtained from (5.9), but is omitted for brevity.

The density (5.8) is plotted in Figure 3 for $\lambda = -0.5, 0, 0.5$, for $n = 5, 10$, and for $n \rightarrow \infty$. It is clear from the plots that the density is again 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 $z = 0$.

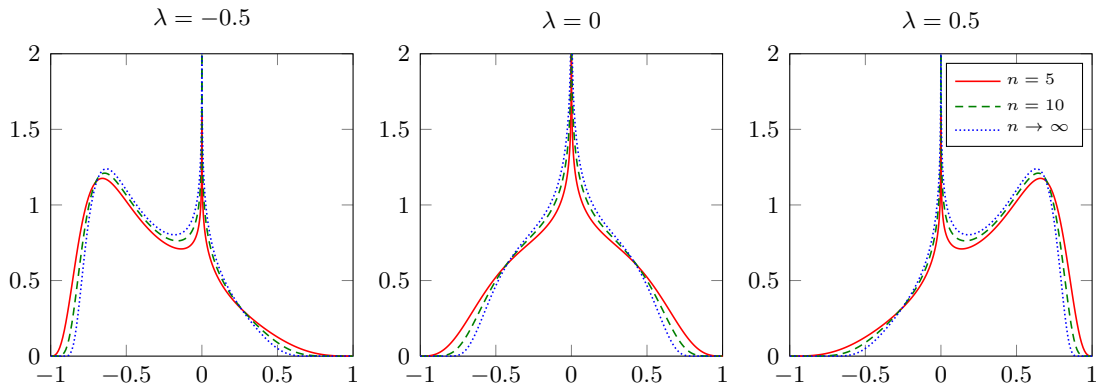


Figure 3: Density of $\hat{\lambda}_{\text{ML}}$ for the Gaussian pure symmetric Complete Bipartite model.

Given the cdf and pdf, other exact properties of $\hat{\lambda}_{\text{ML}}$ can be derived following techniques similar to those used in Hillier and Martellosio (2014a) for the Group Interaction model, but this is not pursued here.

5.3.2 Row-Standardized W , Constant Mean

In a Complete Bipartite model with 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 Proposition 4, because ι_n spans the eigenspace of W corresponding to the only positive eigenvalue of W . Hence, in this model $\hat{\lambda}_{\text{ML}}$ can never be positive, regardless of the true value of λ .

Proposition 9. *For the row-standardized Complete Bipartite model with $X = \iota_n$ and $\varepsilon \sim \text{SMN}(0, I_n)$,*

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

where

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

The corresponding density is

$$\text{pdf}_{\hat{\lambda}_{\text{ML}}}(z; \lambda) = \begin{cases} \frac{1}{B(\frac{1}{2}, \frac{n-2}{2})} \frac{\dot{g}(z, \lambda)}{g(z, \lambda)^{\frac{1}{2}} [1-g(z, \lambda)]^{\frac{n-1}{2}}}, & \text{if } z \in (-1, 0) \\ 0, & \text{if } z \in [0, 1). \end{cases} \quad (5.10)$$

The limiting cdf and pdf as $n \rightarrow \infty$ can be obtained immediately from Proposition 9. Letting

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

we obtain 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 (*maybe explain what goes wrong in terms of Lee's theory*): the Complete Bipartite model with symmetric W and constant mean does satisfy the assumptions for consistency and asymptotic normality in Lee (2004).

The density (5.10) 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 Conclusion

We have proposed a novel approach to the study of the distributional properties of the QMLE for the spatial autoregressive parameter λ , based on representing the cdf of the estimator as the probability that a certain quadratic form is negative. This paper has focused on exact properties of the QMLE. We have shown that under a very general condition the profile likelihood of the model is single peaked. We have also studied some general properties of the QMLE and we have been able to obtain exact formulae for some relatively simple cases. The cdf representation appears to be also useful to derive accurate approximations to the distribution of the QMLE, for any W and any X , and for computational purposes, but these topics are left for future research.

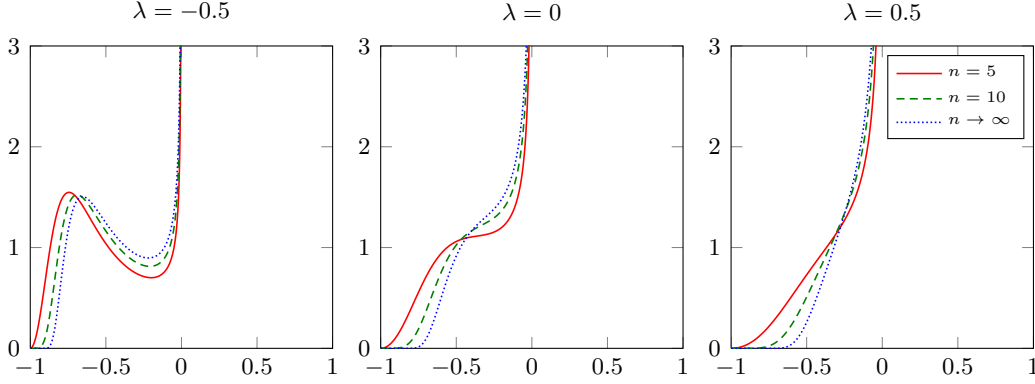


Figure 4: Density of $\hat{\lambda}_{\text{ML}}$ for the Gaussian row-standardized Complete Bipartite model with constant mean.

Appendix A Auxiliary Results

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

- (i) *For any $z \in \Lambda$, the distinct eigenvalues $\gamma_1(z), \gamma_2(z), \dots, \gamma_T(z)$ of C_z are in increasing order (i.e., $s > t$ implies $\gamma_s(z) > \gamma_t(z)$ for any $z \in \Lambda$). For any $z \in \Lambda$, $\gamma_1(z) < 0, \gamma_T(z) > 0$, and, for any $t = 2, \dots, T-1$, $\gamma_t(z)$ changes sign exactly once on Λ .*
- (ii) *For $T \geq 2$, $d_{11}(z, \lambda) < 0$ and $d_{TT}(z, \lambda) > 0$ for all $z, \lambda \in \Lambda$. If $T > 2$, $d_{tt}(z, \lambda) > 0$ if $z < z_t$ and $d_{tt}(z, \lambda) < 0$ if $z > z_t$, for any $\lambda \in \Lambda$, where z_t denotes the unique value of $z \in \Lambda$ at which $\gamma_t(z) = 0$.*

Proof of Lemma A.1. (i) Obviously, $\omega_s > \omega_t$ implies $g_s(z) > g_t(z)$ for all $z \in \Lambda$, which in turn implies $\gamma_s(z) > \gamma_t(z)$. If $\omega_t = 0$, $g_t(z) = 0$ for all $z \in \Lambda$. For the nonzero eigenvalues, since $dg_t(z)/dz = g_t^2(z) > 0$, each of these functions is strictly increasing on Λ . The function $g_1(z) = \omega_{\min}/(1 - z\omega_{\min})$ is bounded at $z = 1$, and approaches $-\infty$ as $z \downarrow \omega_{\min}^{-1}$. Likewise, the function $g_T(z) = 1/(1 - z)$ is bounded at $z = \omega_{\min}^{-1}$, and approaches $+\infty$ as $z \uparrow 1$. The remaining functions $g_t(z)$ are all bounded at both endpoints of the interval Λ . Now,

$$\frac{1}{n} \text{tr}(G_z) = \frac{1}{n} \sum_{t=1}^T n_t g_t(z).$$

Since this is a convex combination of the $g_t(z)$, $t = 1, \dots, T$, $\text{tr}(G_z)/n$ must be between the smallest and largest of the $g_t(z)$, for all $z \in \Lambda$, i.e., $g_1(z) < \text{tr}(G_z)/n < g_T(z)$, or

$\gamma_1(z) < 0 < \gamma_T(z)$ for all $z \in \Lambda$. Next, the properties of the $g_t(z)$ imply that $\text{tr}(G_z)/n$ is monotonic increasing on Λ , going to $-\infty$ as $z \downarrow \omega_{\min}$, and to $+\infty$ as $z \uparrow 1$. It follows that $\text{tr}(G_z)/n$ crosses all $T - 2$ of the functions $g_t(z)$, for $t \neq 1, T$, at least once, somewhere in Λ . To show that the two functions can only cross once, simply observe that, at a point z where $\gamma_t(z) = 0$ (so that $g_t(z) = \text{tr}(G_z)/n$),

$$\frac{dg_t(z)}{dz} = g_t^2(z) = \left(\frac{1}{n} \sum_{t=1}^T n_t g_t(z) \right)^2 < \frac{1}{n} \sum_{t=1}^T n_t g_t^2(z) = \frac{d}{dz} \left(\frac{1}{n} \text{tr}(G_z) \right)$$

(the inequality is strict because the $g_t(z)$ cannot all be equal, by Assumption A). That is, at every point of intersection, $\text{tr}(G_z)/n$ intersects $g_t(z)$ from below, which implies that there can be only one such point. (ii) This follows from part (i) and the fact that the signs of the $d_{tt}(z, \lambda)$ are those of the $\gamma_t(z)$, for all $\lambda \in \Lambda$. \square

Lemma A.2. *If, for any given y, X, W , the equation $M_X S_\lambda y = 0$ is satisfied by two distinct values of $\lambda \in \mathbb{R}$, then it is satisfied by all $\lambda \in \mathbb{R}$.*

Proof of Lemma A.2. If $M_X(I - \lambda_1 W)y = M_X(I - \lambda_2 W)y = 0$ for two real numbers λ_1 and λ_2 , then $(\lambda_1 - \lambda_2) M_X W y = 0$. If $\lambda_1 \neq \lambda_2$, then $M_X W y = M_X y = 0$, which in turn implies that $M_X S_\lambda y = 0$ for all $\lambda \in \mathbb{R}$. \square

Appendix B Proofs and Remarks

Proof of Proposition 1. If $M_X(\omega I_n - W) = 0$, then $M_X S_\lambda = (1 - \lambda\omega)M_X$, and hence equation (3.2) reduces to

$$l_p(\lambda) = \ln(|\det(S_\lambda)|) - n \ln(|1 - \lambda\omega|) - \frac{n}{2} \ln(y' M_X y). \quad (\text{B.1})$$

The first part of the proposition follows on noticing that the only term in equation (B.1) that depends on y does not involve λ . Moving to the second part, assume that $M_X(\omega I_n - W) = 0$ for some real nonzero eigenvalue ω of W . The profile log-likelihood is a.s. defined by equation (B.1). Letting n_\varkappa denote the algebraic multiplicity of an eigenvalue \varkappa , and $\text{Sp}(W)$ the spectrum of W (defined as the set of distinct eigenvalues), we obtain

$$\begin{aligned} l_p(\lambda) &= \ln \left(\frac{|\prod_{\varkappa \in \text{Sp}(W)} (1 - \lambda\varkappa)^{n_\varkappa}|}{(y' M_X y)^{\frac{n}{2}}} \right) - n \ln(|1 - \lambda\omega|), \\ &= \ln \left(\frac{|\prod_{\varkappa \in \text{Sp}(W) \setminus \{\omega\}} (1 - \lambda\varkappa)^{n_\varkappa}|}{(y' M_X y)^{\frac{n}{2}}} \right) - (n - n_\omega) \ln(|1 - \lambda\omega|), \end{aligned} \quad (\text{B.2})$$

The first term in equation (B.2) is a.s. bounded as $\lambda \rightarrow \omega^{-1}$. The second term goes to $+\infty$ as $\lambda \rightarrow \omega^{-1}$, because $n_\omega < n$ by Assumption A. Thus, $\lim_{\lambda \rightarrow \omega^{-1}} l_p(\lambda) = +\infty$ a.s. Suppose now that $M_X(\omega I_n - W) \neq 0$ for some nonzero eigenvalue ω of W . Then $M_X(\omega I_n - W)y$ is a.s. nonzero. For any y such that $M_X(\omega I_n - W)y \neq 0$, the term $-(n/2) \ln(y' S'_\lambda M_X S_\lambda y)$ in equation (3.2) is defined in a neighborhood of $\lambda = \omega^{-1}$, because it is defined at $\lambda = \omega^{-1}$ and, by Lemma A.2, cannot be undefined at more than one value of $\lambda \neq \omega^{-1}$. Hence, $-(n/2) \ln(y' S'_\lambda M_X S_\lambda y)$ is a.s. continuous at $\lambda = \omega^{-1}$. The other term in equation (3.2), $\ln(|\det(S_\lambda)|)$, goes to $-\infty$ as $\lambda \rightarrow \omega^{-1}$. It follows that $\lim_{\lambda \rightarrow \omega^{-1}} l_p(\lambda) = -\infty$ a.s. \square

Remark 1. The a.s. qualification in the second part of Proposition 1 is required whether $M_X(\omega I_n - W)$ is zero or not. Details are omitted for brevity, but it is easy to show that, if $M_X(\omega I_n - W) \neq 0$, then there is a zero probability (according to the Lebesgue measure on \mathbb{R}^n) set of values of y such that $\lim_{\lambda \rightarrow \omega^{-1}} l_p(\lambda) = +\infty$. If $M_X(\omega I_n - W) = 0$, then there is a zero probability set of values of y such that $l_p(\lambda)$ is undefined for all values of λ .

Proof of Lemma 1. Let ω_t , $t = 1, \dots, T$, denote the distinct (possibly complex) eigenvalues of W , ordered arbitrarily, let $e_t = e_t(W)$ denote the t -th elementary symmetric function in the T distinct eigenvalues of W , and let $e_{t,j}$ be that with the j -th eigenvalue omitted. The polynomial

$$\prod_{t=1}^T (1 - \lambda \omega_t) = \sum_{t=0}^T (-\lambda)^t e_t$$

is a generating function for the e_t , and we have accordingly $e_0 = 1$, and $e_r = 0$ for $r > T$. Correspondingly, the polynomial

$$\prod_{\substack{t=1 \\ t \neq j}}^T (1 - \lambda \omega_t) = \sum_{t=0}^{T-1} (-\lambda)^t e_{t,j}$$

is a generating function for the $e_{t,j}$, and it can easily be checked (by equating coefficients of suitable powers of λ) that

$$\omega_j e_{t-1,j} = e_t - e_{t,j}, \tag{B.3}$$

for $t = 1, \dots, T-1$, and

$$\omega_j e_{T-1,j} = e_T. \tag{B.4}$$

We can therefore write the first-order condition (see equation (3.3) as

$$n(b - a\lambda) \sum_{t=0}^T (-\lambda)^t e_t - (a\lambda^2 - 2b\lambda + c) \sum_{j=1}^T \left(n_j \omega_j \sum_{\substack{t=0 \\ t \neq j}}^{T-1} (-\lambda)^t e_{t,j} \right) = 0, \tag{B.5}$$

where $a := y'W'M_XWy$, $b := y'W'M_Xy$, and $c := y'M_Xy$. We now show that the polynomial equation (B.5) has degree T . Using (B.4) and $\sum_{j=1}^T n_j = n$, the coefficient of λ^{T+1} is

$$na(-1)^{T+1}e_T + (-1)^T a \sum_{j=1}^T n_j \omega_j e_{T-1,j} = 0.$$

On the other hand, the coefficient of λ^T is

$$a(-1)^T \left(ne_{T-1} - \sum_{j=1}^T n_j \omega_j e_{T-2,j} \right) + nb(-1)^{T-1}e_T,$$

which, on using (B.3), reduces to

$$a(-1)^T \left(\sum_{j=1}^T n_j e_{T-1,j} \right) + nb(-1)^{T-1}e_T.$$

This will a.s. not vanish: the term e_T can vanish if one eigenvalue is zero, but at least one term in the sum in the first term will not vanish, since only one eigenvalue can be zero. \square

Remark 2. In many applications, W is the adjacency matrix of a (unweighted and undirected) graph. It is well known in graph theory that the number of distinct eigenvalues of an adjacency matrix is related to the degree of symmetry of the graph (see Biggs, 1993). On the other hand, in algebraic statistics the degree of the score equation is regarded as an index of algebraic complexity of ML estimation (see Drton et al., 2009). Thus Lemma 1 establishes a connection between the algebraic complexity of $\hat{\lambda}_{\text{ML}}$ and the degree of symmetry satisfied by the graph underlying W .

Proof of Lemma 2. Recall that we are assuming that $M_X(\omega I_n - W) \neq 0$ for any real nonzero eigenvalue ω of W . Hence, by Proposition 1, $l_p(\lambda) \rightarrow -\infty$ a.s. at the extremes of Λ . Then, because it is a.s. continuous on Λ , $l_p(\lambda)$ must a.s. have at least one maximum on Λ . Since it is also a.s. differentiable on Λ , all maxima must be critical points. We now show that $l_p(\lambda)$ has a.s. exactly one maximum, and no other stationary points, on Λ . The second derivative of $l_p(\lambda)$ can be written as

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

where $a := y'W'M_XWy$, $b := y'W'M_Xy$, and $c := y'M_Xy$. But at any point where $\dot{l}_p(\lambda) = 0$,

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

so that, at any critical point,

$$\begin{aligned}\ddot{l}_p(\lambda) &= \frac{-n(ac - b^2)}{(a\lambda^2 - 2b\lambda + c)^2} + \frac{1}{n}[\text{tr}(G_\lambda)]^2 - \text{tr}(G_\lambda^2) \\ &= \frac{-n(ac - b^2)}{(a\lambda^2 - 2b\lambda + c)^2} - \text{tr}(C_\lambda^2)\end{aligned}\tag{B.6}$$

By the Cauchy-Schwarz inequality the first term in (B.6) is nonpositive. Hence, if $\text{tr}(C_\lambda^2) > 0$ for all $\lambda \in \Lambda$, we have that $\ddot{l}_p(\lambda) < 0$ at every point where $\dot{l}_p(\lambda) = 0$, that is, $l_p(\lambda)$ has a.s. exactly one point of maximum in Λ , and no other stationary points. \square

Proof of Lemma 3. Assume W is nonnegative. For any $\lambda \in [0, 1)$, G_λ can be expanded as $\sum_{r=0}^{\infty} \lambda^r W^{r+1}$, which shows that G_λ is nonnegative too. Then, for any $\lambda \in [0, 1)$, all off-diagonal entries of C_λ are nonnegative, and hence $\text{tr}(C_\lambda^2) \geq 0$. Since $\text{tr}(C_\lambda^2)$ cannot be zero (cf. footnote 8), it follows by (B.6) that any stationary point of $l_p(\lambda)$ in $[0, 1)$ must be a maximum. \square

Proof of Theorem 1. By Lemma 2, $\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr(\dot{l}_p(z) \leq 0)$, for any $z \in \Lambda$. The desired result follows using expression (3.4), since $y'S'_z M_X S_z y$ is a.s. positive for any $z \in \Lambda$. \square

Proof of Proposition 2. Using the decomposition $W = HDH^{-1}$ we find that $C_z = HD_1 H^{-1}$, and $S_z S_\lambda^{-1} = HD_2 H^{-1}$, with

$$D_1 := \text{diag}(\gamma_t(z) I_{n_t}, t = 1, \dots, T),$$

and

$$D_2 := \text{diag}\left(\frac{1 - z\omega_t}{1 - \lambda\omega_t} I_{n_t}, t = 1, \dots, T\right).$$

As pointed out in the text, if W is similar to a symmetric matrix, Theorem 1 applies, and the matrix $A(z, \lambda)$ in (4.2) can be decomposed as $(H')^{-1} D_2 (D_1 M + M D_1) D_2 H^{-1}$. The proof is completed by expressing $D_2 (D_1 M + M D_1) D_2$ as the block matrix $(d_{st}(z, \lambda) M_{st}; s, t = 1, \dots, T)$. \square

Proof of Proposition 3. (i) Under the assumption that W is similar to a symmetric matrix, the off-diagonal blocks in M vanish if and only if $MD = DM$, where D contains the eigenvalues of W and $M = H' M_X H$ is as in the text, because the eigenvalues in the decomposition of D are distinct. One can then easily check that this is so if and only if $M_X W = W' M_X$. (ii) If W is symmetric, H is orthogonal, and hence $M_{ij} = h'_i M_X h_j$, where h_i denotes the i -th column of H . The diagonal entries M_{ii} are 0 if $h_i \in \text{col}(X)$, 1 if $h_i \notin \text{col}(X)$. Under the assumption that $M_X W$ is also symmetric $\text{col}(X)$ is spanned by k linearly independent eigenvectors of W . Hence, for

each $j = 1, \dots, n$, $M_X h_j$ equals either 0 (if $h_j \in \text{col}(X)$) or h_j (if $h_j \notin \text{col}(X)$). Since $h_i' h_j = 0$ for any $i \neq j$, it follows that all off-diagonal entries of M are zero, which completes the proof. \square

Remark 3. Proposition 3 (ii) can also be obtained directly from equation (4.2), since, under the relevant assumptions, the $d_{tt}(z, \lambda)$ are eigenvalues of $A(z, \lambda)$.

Proof of Proposition 4. Starting from part (i) and recalling that $M := H' M_X H$, under the stated condition all diagonal blocks M_{ss} for $s > t$ vanish. Since, by Proposition 3 (ii), $d_{ss}(z, \lambda) < 0$ for $s \leq t$, $z > z_t$, and for any $\lambda \in \Lambda$, it follows by Proposition 3 (i) that $\Pr(\hat{\lambda}_{\text{ML}} \leq z) = 1$ for $z \geq z_t$. By the same argument, part (ii) is proved by showing that in that case $\Pr(\hat{\lambda}_{\text{ML}} \leq z) = 0$ for $z \leq z_t$. \square

Proof of Proposition 5. For simplicity, assume that all densities exist. We need to show that the distribution of the maximal invariant $v := y(y'y)^{-1/2} \in \mathcal{S}^{n-1}$ under scale transformations in the sample space is invariant under scale mixtures of the distribution of y . Let $f(y)$ denote the density of $y \in \mathbb{R}^n$, and let $q := (y'y)^{1/2} > 0$. We may transform $y \rightarrow (q, v)$, setting $y = qv$. The volume element (Lebesgue measure) (dy) on \mathbb{R}^n decomposes as $(dy) = q^{n-1} dq(v'dv)$, where $(v'dv)$ denotes (unnormalized) invariant measure on \mathcal{S}^{n-1} (see Muirhead, 1982, Theorem 2.1.14 for a more general version of this result). The measure on \mathcal{S}^{n-1} induced by the density $f(y)$ for y is therefore defined, for any subset \mathcal{A} of \mathcal{S}^{n-1} , by

$$\Pr(v \in \mathcal{A}) = \int_{\mathcal{A}} \left\{ \int_{q>0} q^{n-1} f(qv) dq \right\} (v'dv).$$

Now let κ be a random scalar independent of y with density $p(\kappa)$ on \mathbb{R}^+ . The density of $y^* := \kappa y$ is then given by the mixture

$$g(y^*) := \int_{\kappa>0} \kappa^{-n} f(y^*|\kappa) p(\kappa) d\kappa$$

The measure induced by $g(\cdot)$ for $v(y^*) = v(y)$ is therefore

$$\begin{aligned} \int_{q>0} q^{n-1} g(qv) dq &= \int_{q>0} \int_{\kappa>0} q^{n-1} \kappa^{-n} f(qv|\kappa) p(\kappa) d\kappa dq \\ &= \int_{q>0} q^{n-1} f(qv) dq \end{aligned}$$

on transforming to $(q/\kappa, \kappa)$ and integrating out κ . That is, for any (proper) density $p(\cdot)$, $g(\cdot)$ induces the same measure on \mathcal{S}^{n-1} as does $f(\cdot)$, as claimed. \square

Remark 4. For a formal treatment of the argument used to establish Proposition 5 - averaging over a group - see also Eaton (1989), particularly Chapters 4 and 5.

Proof of Proposition 6. Because of the presence of the scale parameter σ , the SAR model (1.1) is invariant with respect to the scale transformations $y \rightarrow \kappa y$, $\kappa > 0$, for any W and X . If the distribution of ε does not depend on β or σ^2 , the transformation $y \rightarrow \kappa y$ induces the transformations $(\beta, \lambda, \sigma^2, \theta) \rightarrow (\kappa\beta, \lambda, \kappa^2\sigma^2, \theta)$ in the parameter space, with maximal invariant $(\beta/\sigma, \lambda, \theta)$. Since, as pointed out earlier in the text, $\hat{\lambda}_{\text{ML}}$ itself is invariant to scale transformations of y , its distribution can depend on $(\beta, \lambda, \sigma^2, \theta)$ only through a maximal invariant in the parameter space (see, e.g., Lehmann and Romano, 2005, Theorem 6.3.2). In fact, if $\text{col}(X)$ is an invariant subspace of W , and only in that case, the SAR model (1.1) is invariant under the larger group \mathcal{G}_X of transformations $y \rightarrow \kappa y + X\delta$, for any $\kappa > 0$, any $\delta \in \mathbb{R}^k$; see Hillier and Martellosio (2014b). The condition that $\text{col}(X)$ is an invariant subspace of W is equivalent to the existence of a $k \times k$ matrix A such that $WX = XA$. Since the eigenvalues of A must be a subset of the eigenvalues of W , $WX = XA$ if and only if $S_\lambda^{-1}X = X(I_k - \lambda A)^{-1}$, for any λ such that S_λ is invertible. The group, say $\bar{\mathcal{G}}_X$, induced by \mathcal{G}_X on the parameter space is that of the transformations $(\beta, \lambda, \sigma^2, \theta) \rightarrow (\kappa\beta + (I_k - \lambda A)\delta, \lambda, \kappa^2\sigma^2, \theta)$. Now, it is easy to see from the profile score equation that (under the conditions stated above) $\hat{\lambda}_{\text{ML}}$ is invariant under \mathcal{G}_X . Since $\bar{\mathcal{G}}_X$ acts transitively on the parameter space for (β, σ^2) , and leaves (λ, θ) invariant, it follows that the distribution of $\hat{\lambda}_{\text{ML}}$ cannot depend on (β, σ^2) , and depends only on (λ, θ) . \square

Remark 5. Proposition 6 implicitly assumes that all parameters $\beta, \lambda, \sigma^2, \theta$ are identifiable. Identifiability is required for the application of the invariance argument in the proof of the Proposition.

Proof of Theorem 2. In the pure case $k = 0$ and $n_t(X) = 0$ for all $t = 1, \dots, T$, so the result follows from equation (5.1). \square

Proof of Corollary 1. Partition the vector x in Proposition 3 (ii) into (x'_1, x'_2) , with x_i of dimension $v_i \times 1$, for $i = 1, 2$, and let $Q_i := x'_i A_i x_i$, for $i = 1, 2$. The statistics Q_1 and Q_2 are linear combinations of central χ^2 random variables with positive coefficients. By Theorem 2, $\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr(Q_1 \leq Q_2) = \Pr(Q_1/Q_2 \leq 1)$, with Q_1 and Q_2 independent of each other. We shall now use the following slight modification of a result due to James (1964) for the density of a positive definite quadratic form in standard normal variables: If $Q := \sum_{i=1}^S a_i \chi_{n_i}^2$ is a linear combination of independent $\chi_{n_i}^2$ random variables with positive coefficients a_i , the density of Q is given by

$$\text{pdf}_Q(q; A) = \frac{\exp\left(-\frac{1}{2}\tau q\right) q^{\frac{n}{2}-1} {}_1F_1\left(\frac{1}{2}; \frac{n}{2}; \frac{q}{2}A^*\right)}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right) (\det(A))^{\frac{1}{2}}} \quad (\text{B.7})$$

where $n = \sum_{i=1}^S n_i$, $A := \text{diag}(a_i I_{n_i}, i = 1, \dots, S)$, $\tau := \text{tr}(A^{-1})$, and $A^* := \tau I_n - A^{-1}$. The confluent hypergeometric function here is of matrix argument (see Muirhead,

1982), but, importantly, only top-order zonal polynomials are involved. Using this result for both Q_1 and Q_2 , transforming to $(Q_1/Q_2, Q_2)$, and integrating out Q_2 termwise gives an expression involving only r (it is straightforward to check that the term-by-term integration involved is justified). Integrating this over $0 < r < 1$ gives the result. \square

Remark 6. It is easily confirmed that the cdf in Corollary 1 is a bivariate mixture of the distributions of random variables that are conditionally, given the values of two independent non-negative integer-valued random variables J and K , say, distributed as F_{v_1+2j, v_2+2k} . The probability $\Pr(J = j)$ is the coefficient of t^j in the expansion of $(\det[tI_{v_1} + (1-t)\tau_1 A_1])^{-1/2}$, with a similar expression for $\Pr(K = k)$.

Proof of Corollary 2. For notational convenience, let us rename the coefficients $d_{tt}(z, \lambda)$ as $d_{\omega_t}(z, \lambda)$, for all $t = 1, \dots, T$. Since $d_{\omega_s}(z, \lambda) = -d_{-\omega_s}(-z, -\lambda)$, from expression (5.2) we have

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z; \lambda) = \Pr\left(\sum_{t=1}^T -d_{-\omega_t}(-z, -\lambda) \chi_{n_t}^2 \leq 0\right),$$

which is equal to $\Pr(\hat{\lambda}_{\text{ML}} \geq -z; -\lambda) = 1 - \Pr(\hat{\lambda}_{\text{ML}} \leq -z; -\lambda)$ if the spectrum of W is symmetric. The stated result follows on differentiating $\Pr(\hat{\lambda}_{\text{ML}} \leq z; \lambda) = 1 - \Pr(\hat{\lambda}_{\text{ML}} \leq -z; -\lambda)$ with respect to z . \square

Proof of Proposition 7. For the pure balanced Group Interaction model, W is symmetric, $T = 2$, $n_1 = r$, $n_2 = r(m-1)$, $\omega_1 = 1$, $\omega_2 = -1/(m-1)$. 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 (5.3) now gives $\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr(F_{r, r(m-1)} \leq \theta(z)/\theta(\lambda))$. On differentiating with respect to z , we obtain

$$\text{pdf}_{\hat{\lambda}_{\text{ML}}}(z; \lambda) = \frac{1}{\theta(\lambda)} \frac{\partial \theta(z)}{\partial z} \text{pdf}_{F_{r, r(m-1)}}(\theta(z)/\theta(\lambda)),$$

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

Proof of Proposition 8. For a symmetric Complete Bipartite model $\text{tr}(G_z^{-1}) = -1/(1+z) + 1/(1-z) = 2z/(1-z^2)$, and hence $\gamma_1(z) = -[n - (n-2)z]/[n(1-z^2)]$, $\gamma_2(z) = -2z/[n(1-z^2)]$, and $\gamma_3(z) = [n + (n-2)z]/[n(1-z^2)]$. The stated expression for the cdf then follows from equation (5.2). For $z \in (0, 1)$ the density is obtained as an application of the conditioning argument in Hillier and Martellosio (2014a) for the case $T = 3$, with $\gamma = 1$, $\alpha = n-2$, $\beta = 1$, $a(z) = 2z/\phi_1$, and $c(z) = \phi_2/\phi_1$. The proof is completed using Corollary 2. \square

Proof of Proposition 9. For the row-standardised Complete Bipartite model the matrix H is

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

where $L_{p,p-1}$ ($p \times (p-1)$) satisfies $L'_{p,p-1}\iota_p = 0$ and $L'_{p,p-1}L_{p,p-1} = I_{p-1}$. Thus,

$$M = H' M_{\iota_n} H = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{4pq}{n^2} & 0 \\ 0 & 0 & I_{n-2} \end{bmatrix}.$$

This is certainly block-diagonal, as expected, and in addition the (1,1) block also vanishes. The mean of $x = H^{-1}\tilde{y}$ is $E(x) = \beta\sqrt{n}(n, 0, 0)'$. Therefore, from equation (4.4), we have

$$\begin{aligned} \Pr(\hat{\lambda}_{\text{ML}} \leq z) &= \Pr(d_{22}(z, \lambda)\chi_1^2 + d_{33}(z, \lambda)\chi_{n-2}^2 \leq 0) \\ &= \Pr(-2(\phi_2\chi_1^2 + 2z\chi_{n-2}^2) \leq 0). \end{aligned}$$

But, if $z \geq 0$, both coefficients here are non-negative, so for $z \geq 0$, $\Pr(\hat{\lambda}_{\text{ML}} \leq z) = 1$. This yields the stated result for the cdf. The density is obtained by differentiation, as in the proof of Proposition 7. \square

References

- Arnold, S.F. (1979) Linear models with exchangeably distributed errors, *Journal of the American Statistical Association* 74, 194-199.
- Audretsch, D.B. and Feldmann, M.P. (1996) R&D spillovers and the geography of innovation and production, *American Economic Review* 86, 630-640.
- 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.
- Bell, K.P. and Bockstael, N.E. (2000) Applying the Generalized Method of Moments Approach to Spatial Problems Involving Micro-Level Data, *Review of Economic and Statistics* 82, 72-82.
- Bertrand, M., Luttmer, E.F.P., and Mullainathan S. (2000) Network effects and welfare cultures, *Quarterly Journal of Economics* 115, 1019-1055.
- Besley, T. and Case, A. (1995) Incumbent behavior: vote-seeking, tax-setting, and yardstick competition, *American Economic Review* 85, 25-45.

- Biggs, N.L. (1993) *Algebraic Graph Theory*, second edition. Cambridge University Press.
- Bramoullé, Y., Djebbari, H. and Fortin, B. (2009) Identification of peer effects through social networks, *Journal of Econometrics* 150, 41-55.
- Case, A. (1991) Spatial patterns in household demand, *Econometrica* 59, 953-966.
- Drton, M., Sturmfels, B. and Sullivant, S. (2009). *Lectures on Algebraic Statistics*, Birkhauser.
- Eaton, M.L. (1989) *Group Invariance Applications in Statistics*, Institute of Mathematical Statistics and American Statistical Association.
- Forchini, G. (2002) The exact cumulative distribution function of a ratio of quadratic forms in normal variables, with application to the AR(1) model, *Econometric Theory* 18, 823-852.
- Forchini, G. (2005) The distribution of a ratio of quadratic forms in noncentral normal variables, *Communications in Statistics - Theory and Methods* 34, 999-1008.
- Hillier, G.H. (2001) The density of a quadratic form in a vector uniformly distributed on the n-sphere, *Econometric Theory* 17, 1-28.
- 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. (2014a). Exact likelihood inference in the spatial group-interaction model. Manuscript.
- Hillier, G.H. and Martellosio, F. (2014b). Adjusted MLE for the spatial autoregressive parameter. Manuscript.
- 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.
- 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, & Y. Yuzevovich (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.
- Koopmans, T.C. (1942) Serial correlation and quadratic forms in normal variables. *Annals of Mathematical Statistics*, 12, 14-33.
- 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., Liu, X., Lin, X. (2010) Specification and estimation of social interaction models with network structures, *The Econometrics Journal* 13, 145-176.
- Lehmann, E.L. and Romano, J. (2005) *Testing Statistical Hypotheses*. Springer.
- LeSage, J.P. and Kelley Pace, R. (2009) *Introduction to Spatial Econometrics*, Boca Raton, CRC Press/Taylor & Francis.
- Liu, X., Patacchini, E., Zenou, Y. and Lee, L.F. (2015) Who is the key player? A network analysis of juvenile delinquency, working paper.
- Martellosio, F. (2011) Efficiency of the OLS estimator in the vicinity of a spatial unit root. *Statistics & Probability Letters* 81, 1285–1291.
- Muirhead, R. J. (1982) *Aspects of Multivariate Statistical Theory*, Wiley, New York.
- Mulholland, H.P. (1965) On the degree of smoothness and on singularities in distributions of statistical functions, *Proceedings of the Cambridge Philosophical Society* 61, 721–739.
- Ord, J.K. (1975) Estimation methods for models of spatial interaction, *Journal of the American Statistical Association* 70, 120–126.
- Pinkse, J., Slade, M.E. and Brett, C. (2002) Spatial price competition: a semiparametric approach, *Econometrica* 70, 1111-1153.
- Robinson, P.M. and Rossi, F. (2015) Refinements in maximum likelihood inference on spatial autocorrelation in panel data. *Journal of Econometrics*, doi:10.1016/j.jeconom.2015.03.036.
- Ross-Parker, H. (1975) Inter-plant competition models and their sampling distributions, *Advances in Applied Probability* 7, 453-4.
- Saldanha, N.C. & C. Tomei (1996) The accumulated distribution of quadratic forms on the sphere, *Linear Algebra and Its Applications*, 245, 335–351.
- Topa, G. (2001) Social interactions, local spillovers and unemployment, *The Review of Economic Studies* 68, 261-295.
- 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.