Exact Properties of the Maximum Likelihood Estimator in Spatial Autoregressive Models

Grant Hillier[a] and Federico Martellosio[b,*]

[a]CeMMAP and Department of Economics, University of Southampton,
Highfield, Southampton, SO17 1BJ, UK
[b]School of Economics, University of Surrey,
Guildford, Surrey, GU2 7XH, UK

17 September 2014

Abstract

The (quasi-) maximum likelihood estimator (MLE) 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 of theoretical and practical interest 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.

*Corresponding author. Tel: +44 (0) 1483 683473
E-mail addresses: ghh@soton.ac.uk (G. Hillier), f.martellosio@surrey.ac.uk (F. Martellosio)

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.[1] In most applications, such models are based on a fixed spatial weights matrix W whose elements reflect the modeler's assumptions about the pairwise interactions between the observational units. A scalar autoregressive parameter λ measures the strength of this cross-sectional interaction. 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 with full column rank, 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. Initially we make no distributional assumptions on ε, but do assume that quasi-maximum likelihood estimation is conducted on the basis of the likelihood that would prevail if the Gaussianity assumption $\varepsilon \sim \mathrm{N}(0, I_n)$ were added to equation (1.1). This setup is identical to that used in Lee (2004), who discusses the *asymptotic* properties of this estimator. Many results we obtain do not require distribution assumptions, but we later add the Gaussianity assumption in order to obtain explicit formulae.

The parameter λ is usually of direct interest in applications. For example, in social interactions analysis measuring the strength of network effects is important to policy makers.[2] Although considerable progress has been made recently in establishing the first-order asymptotic properties of the MLE for λ in such models, there remain some compelling reasons for studying its exact properties - more so, perhaps, than usual. First, exact results reveal explicitly how the properties of the estimator depend on the characteristics of the underlying model. Second, exact results are

[1]For an introduction to spatial autoregressions see, e.g., Cliff and Ord (1973), Cressie (1993), and 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 (2014), to name just a few.

[2]Of course, the parameter β is typically also of interest. The distributional properties of the MLE for β can be deduced from those of the MLE for λ, but will not be considered in this paper.

useful for checking the accuracy of the available asymptotic results. This is important because the distribution of the estimator may (indeed, does) depend crucially on the spatial weights matrix, and on the assumptions made on how it evolves with the sample size. Until now, simulation studies have been virtually the only source of such information. Third, the exact distribution may possess important features that would be impossible to discover by asymptotic methods or Monte Carlo simulation - for example, non-differentiability, non-analyticity, or unboundedness of the density. Finally, exact results are informative when the assumptions needed to obtain asymptotic results are not plausible.

The first-order condition defining the MLE for λ is, in general, a polynomial of high degree from which no closed-form solution can be obtained. Hence, even the calculation of the MLE has been regarded as problematic in this model, let alone study of its exact properties. Ord (1975) presents a simplified procedure for maximum likelihood estimation of model (1.1). 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. Bao (2013) and Yang (2013) extend such approximations to the case when exogenous regressors are included and when ε is not necessarily Gaussian. Several other papers have studied the performance of the MLE by simulation, particularly in relation to competing estimators such as the two-stage least squares (2SLS) estimator or more general GMM estimators.

The key observation that enables us to carry out an exact analysis of the MLE is that, when - as it always is in practice - the likelihood is defined only for an interval of values of λ containing the origin for which the matrix $I_n - \lambda W$ is positive definite, the profile (or concentrated) likelihood after maximizing with respect to (β, σ^2) is, under certain assumptions, *single-peaked*. This fact implies that an exact expression for the cdf of the MLE for λ can easily be written down, notwithstanding the unavailability of the MLE in closed form. This is the main result of the paper.

Starting from this fundamental result, we then present a number of exact results for the MLE that follow from it. In principle, knowledge of the cdf provides a starting point for a full exact analysis of the MLE, for an arbitrary distribution of ε. However, the distribution theory for the MLE is non-standard, and, perhaps not unexpectedly, 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. For this and other reasons, the distribution theory for the MLE that is implied by our main result is, for general (W, X), quite complicated. We give some general results of this nature, including an explicit formula for the cdf in the pure Gaussian case that is valid for any symmetric W. But, we do not attempt a complete general analysis; that is almost certainly best accomplished on a case-by-case basis. We illustrate the usefulness of the main results by examining in detail some popular special cases of model (1.1).

It is intuitive that in model (1.1) the relationship between the matrices W and X must be important, and this will be evident at many points in the paper. The first of these is the observation that there can be (W, X) combinations that lead to non-existence, or non-randomness, of the MLE. These pathological cases, of course, we rule out. The interaction between W and X will also be seen to be fundamental in determining the properties of the MLE. A striking example of this is that the distribution of the MLE may not be supported on the entire parameter space. This result implies that the estimator cannot be uniformly consistent in such circumstances. Our main result, Theorem 1 below, applies for any pair (W, X) for which the MLE exists, and W has real eigenvalues. Some consequences of the Theorem also hold generally, but in order to obtain exact analytic results we usually need to make assumptions about (W, X). For instance, we sometimes assume that W is symmetric, or similar to a symmetric matrix, and sometimes also that $M_X W$ is symmetric ($M_X := I_n - X(X'X)^{-1}X'$). Some of these assumptions may be reasonable in some applications, such as those in which W is the adjacency matrix of a graph, but unreasonable in others. Their virtue lies in revealing important properties of the MLE that can be expected to hold more generally.

The rest of the paper is organized as follows. Section 2 describes the assumptions we make on the spatial weights matrix W and the parameter space for λ, and introduces some examples that will be used to illustrate the theoretical results. Section 3 discusses some key, and novel, properties of the profile log-likelihood for λ. Section 4 gives the main results, along with a number of important consequences. Section 5 gives an explicit expression for the cdf of the MLE under particular conditions. The main results are then applied in Section 6 to the examples introduced earlier. The analysis up to this point is carried out under the assumption that the eigenvalues of W are real; the case of complex eigenvalues is discussed briefly in Section 7. Section 8 concludes by discussing generalizations and further work that our results suggest. The Appendices contain auxiliary results and proofs of the results that are not established directly in the main text.

All matrices considered in this paper are real, unless otherwise stated. For an $n \times p$ matrix A, we denote the space spanned by the columns of A by $\mathrm{col}(A)$, and the null space of A by $\mathrm{null}(A)$. Finally, "a.s." stands for almost surely, with respect to the Lebesgue measure on $\mathbb{R}^n$.

2 Assumptions and Examples

2.1 Assumptions on the Weights Matrix

The following assumptions on W are maintained throughout the paper: (a) W is entrywise nonnegative; (b) W is non-nilpotent; (c) the diagonal entries of W are zero; (d) W is normalized so that its spectral radius is one.[3] Assumptions (a), (b), and (c) are virtually always satisfied in practical applications. Assumption (d) is automatically satisfied if W is row-stochastic; otherwise, the normalization can be accomplished by rescaling, provided only that the spectral radius of W is nonzero, and this is guaranteed under Assumptions (a) and (b). We remark that Assumption (b) captures the "spatial" character of the models we wish to discuss. Given nonnegativity of W, assuming non-nilpotency is equivalent to requiring that there is no permutation of the observational units that would make W triangular, i.e., would make the autoregressive process unilateral (see Martellosio, 2011). Also, if W is nilpotent and nonnegative it can be shown that the ML and OLS estimators for λ coincide, in which case study of the MLE is straightforward.

The four assumptions above are not contentious, and will not be referred to in the statements of the formal results in the paper. Additional assumptions on the structure of W will be made from time-to-time; these will be explicitly stated in the statement of results. In particular, the main results in Section 4 are proved under the assumption that the eigenvalues of W are real. This assumption is very often satisfied in applications of the model, but some consequences of its removal will be discussed in Section 7.

Two assumptions that imply that all eigenvalues of W are real, and will be useful to simplify the results, are that W is similar to a symmetric matrix, or, more restrictively, that W is itself symmetric. The former assumption covers the common case in which W is the row-standardized version of a symmetric matrix,[4] and is equivalent to the assumption that W has real eigenvalues and is diagonalizable. An important context in which all eigenvalues of W are real is when W is the adjacency matrix of a simple graph, possibly row-standardized (a simple graph is an unweighted and undirected graph containing no loops or multiple edges).

2.2 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. Thus, the

[3] Recall that the spectral radius of a matrix is the largest of the absolute values of its eigenvalues.

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

values of λ at which S_λ is singular must be ruled out for the model to be complete, so the reciprocals of the nonzero real eigenvalues of W must be excluded as possible values for λ. This we assume throughout, but in practice the parameter space for λ is usually restricted much further, as explained next.

The normalization of the spectral radius to unity (Assumption (d) above) implies that the largest eigenvalue of W is 1.[5] We also assume that W has at least one real negative eigenvalue, and denote the smallest real eigenvalue of W by $\omega_{\min}$, the value of which must be in $[-1, 0)$. The interval $\Lambda := (\omega_{\min}^{-1}, 1)$ is the largest interval containing the origin in which S_λ is nonsingular.[6] 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 MLE considered in this paper is obtained by maximizing the likelihood over Λ. The consequences of adopting a different parameter space are discussed after equation (3.3) below.

2.3 Examples

To illustrate our results the following examples will be used, chosen for their simplicity and their popularity in the literature.

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

$$B_m := \frac{1}{m-1}\left(\iota_m \iota_m' - I_m\right),$$

where ι_m is the $m-$vector of ones. A model involving r such groups of equal size, with no between-group interactions, involves the $rm \times rm$ spatial weights matrix

$$W = I_r \otimes B_m. \tag{2.1}$$

We call this a *balanced Group Interaction model*; it is popular in applications, and is also often used to illustrate (by simulation) theoretical work (see, e.g., Baltagi, 2006, Kelejian et al., 2006, Lee, 2007). The eigenvalues of W are: 1, with multiplicity r, and $-1/(m-1)$, with multiplicity $r(m-1)$. Here the sample size is $n = rm$, and the parameter space is $\Lambda = (-(m-1), 1)$.

[5]This follows by the Perron-Frobenius Theorem for nonnegative matrices (see, e.g., Horn and Johnson, 1985).

[6]If W does not have any (real) negative eigenvalues one could set $\lambda_{\min} = -\infty$. Note that if all eigenvalues of W are real, then W certainly has at least one negative eigenvalue because of the assumption that $\mathrm{tr}(W) = 0$.

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 (see 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

$$W = \begin{bmatrix} 0_{pp} & \frac{1}{q}\iota_p \iota_q' \\ \frac{1}{p}\iota_q \iota_p' & 0_{qq} \end{bmatrix}. \tag{2.2}$$

This is not symmetric unless $p = q$. Alternatively, A can be rescaled by its spectral radius, yielding the symmetric weights matrix

$$W = (pq)^{-\frac{1}{2}} A. \tag{2.3}$$

We refer to the SAR models 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 the parameter space is $\Lambda = (-1, 1)$.

These two examples will be used to illustrate theoretical results in Sections 3 and 4. Notice that for Group Interaction models W has full rank, while in the Complete Bipartite class it has rank 2 (the minimum possible, since we assume $\operatorname{tr}(W) = 0$). In Section 6 we provide brief details of the properties of the MLE for λ in each case. More extensive treatment of the examples will be given elsewhere.

3 Properties of the Profile Log-Likelihood

Quasi-maximum likelihood of the parameters in model (1.1) is based on the log-likelihood obtained under the assumption $\varepsilon \sim \mathrm{N}(0, I_n)$. For any λ such that $\det(S_\lambda) \neq 0$, this log-likelihood is

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

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

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

where $M_X := I_n - X(X'X)^{-1}X'$. For any λ such that $\det(S_\lambda) \neq 0$, $l_p(\lambda)$ is undefined if and only if $y'S'_\lambda M_X S_\lambda y = 0$, a zero probability event according to the Lebesgue measure on $\mathbb{R}^n$ (since, for any λ such that $\det(S_\lambda) \neq 0$, null $(S'_\lambda M_X S_\lambda)$ has dimension $k < n$). The estimator we consider in this paper is

$$\hat{\lambda}_{\mathrm{ML}} := \arg\max_{\lambda \in \Lambda} l_p(\lambda), \tag{3.3}$$

provided that the maximum exists and is unique.[7] 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. Indeed, the unrestricted maximizer of $l_p(\lambda)$ can, in general, be anywhere on the entire real line (with the points where $\det(S_\lambda) = 0$ excluded). Some authors suggest that λ should be restricted to $(-1, 1)$ (see, e.g., Kelejian and Prucha, 2010). When $(-1, 1)$ is a proper subset of Λ, the estimator $\bar{\lambda}_{\mathrm{ML}} := \arg\max_{\lambda \in (-1,1)} l_p(\lambda)$ is a censored version of $\hat{\lambda}_{\mathrm{ML}}$. Since $\Pr(\bar{\lambda}_{\mathrm{ML}} = -1) = \Pr(\hat{\lambda}_{\mathrm{ML}} < -1)$, and $\Pr(\bar{\lambda}_{\mathrm{ML}} < z) = \Pr(\hat{\lambda}_{\mathrm{ML}} < z)$, for any $z \in (-1, 1)$, it is clear that the properties of $\bar{\lambda}_{\mathrm{ML}}$ follow from those of $\hat{\lambda}_{\mathrm{ML}}$.

3.1 Existence of the MLE

Before embarking on a study of the properties of $\hat{\lambda}_{\mathrm{ML}}$ 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. It turns out that there are combinations of the matrices W and X for which neither of these is true.

Since $l_p(\lambda)$ is a.s. continuous on the interior of Λ, to establish boundedness of $l_p(\lambda)$ over Λ we only need to examine its behavior near the endpoints, $\omega_{\min}^{-1}$ and 1. The following lemma, which will also be needed later in the paper, determines the behavior of $l_p(\lambda)$ not only near the endpoints of Λ, but near each of the points where S_λ is singular (the points $\lambda = \omega^{-1}$, for the real nonzero eigenvalue ω of W).

Lemma 3.1. *For any real nonzero eigenvalue ω of W, a.s.*

$$\lim_{\lambda \to \omega^{-1}} l_p(\lambda) = \begin{cases} -\infty, & \text{if } M_X(\omega I_n - W) \neq 0 \\ +\infty, & \text{if } M_X(\omega I_n - W) = 0. \end{cases}$$

Thus, the profile log-likelihood $l_p(\lambda)$ diverges a.s. to either $-\infty$ or $+\infty$ at each of the points where S_λ is singular. The implications for $\hat{\lambda}_{\mathrm{ML}}$ are as follows. If $M_X(\omega I_n - W) \neq 0$ for $\omega = \omega_{\min}^{-1}$ and $\omega = 1$, then $\hat{\lambda}_{\mathrm{ML}}$ exists a.s. If $M_X(\omega I_n - W) = 0$ for $\omega = \omega_{\min}^{-1}$ or $\omega = 1$, then $l_p(\lambda)$ is a.s. unbounded above near one of the endpoints of Λ, in which case we say that $\hat{\lambda}_{\mathrm{ML}}$ does not exist.[8]

[7] Note that when $\lambda \in \Lambda$ the absolute value in (3.1) and (3.2) is not needed as $\det(S_\lambda) > 0$.

[8] When $\lim_{\lambda \to \omega^{-1}} l_p(\lambda) = +\infty$, one could alternatively set $\hat{\lambda}_{\mathrm{ML}} = \omega^{-1}$. This would not change the conclusion in Proposition 3.2 below.

Clearly, the case when $l_p(\lambda)$ is a.s. unbounded from above demands more attention. Under the corresponding condition $M_X(\omega I_n - W) = 0$, we have $M_X S_\lambda = (1 - \lambda\omega)M_X$, and hence equation (3.2) reduces to

$$l_p(\lambda) = \ln(|\det(S_\lambda)|) - n\ln(|1 - \lambda\omega|) - \frac{n}{2}\ln(y' M_X y). \tag{3.4}$$

Note that the only term in equation (3.4) that depends on y does not involve λ. This immediately gives the following result.

Proposition 3.2. *If $M_X(\omega I_n - W) = 0$ for some real eigenvalue ω of W, then $\hat{\lambda}_{\mathrm{ML}}$ is, if it exists, a constant (i.e., does not depend on y).*

Fortunately, the condition $M_X(\omega I_n - W) = 0$ appearing in Lemma 3.1 and Proposition 3.2 is usually not met in applications. It is useful, however, to mention a couple of examples in which it is met. The weights matrix of a Group Interaction model (Example 1 above) has two eigenspaces: $\mathrm{col}(I_r \otimes \iota_m)$, associated to the eigenvalue 1, and its orthogonal complement, associated to the eigenvalue $\omega_{\min} = -1/(m-1)$. Observe that $\mathrm{col}(\omega_{\min} I_n - W) = \mathrm{null}^{\perp}(\omega_{\min} I_n - W) = \mathrm{col}(I_r \otimes \iota_m)$. Lemma 3.1 then implies that, if $\mathrm{col}(I_r \otimes \iota_m) \subseteq \mathrm{col}(X)$, then $l_p(\lambda)$ does not depend on y and $l_p(\lambda) \to +\infty$ as $\lambda \to \omega_{\min}^{-1}$. Since the matrix $I_r \otimes \iota_m$ represents group specific fixed effects, it follows that, *in the balanced Group Interaction model, $\hat{\lambda}_{\mathrm{ML}}$ fails to exist in the presence of group fixed effects.*[9] Another example is a symmetric or row-standardized Complete Bipartite model (Example 2 above) when X includes an intercept for each of the two groups. In this case $M_X W = 0$, so Proposition 3.2 applies (with $\omega = 0$).

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 $\hat{\lambda}_{\mathrm{ML}}$ does not exist or does not depend on the data y.[10]

Remark 3.3. For any real eigenvalue ω of W, $M_X(\omega I_n - W) = 0$ is equivalent to $\mathrm{col}(\omega I_n - W) \subseteq \mathrm{col}(X)$. A necessary condition for $M_X(\omega I_n - W) = 0$ is that $\mathrm{rank}(\omega I_n - W) \leq k$, i.e., the geometric multiplicity of ω as an eigenvalue of W must be at least $n - k$. Also, note that the condition $M_X(\omega I_n - W) = 0$ can be satisfied at most for one real eigenvalue ω of W.

Remark 3.4. The a.s. qualification in Lemma 3.1 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 $\lim_{\lambda \to \omega^{-1}} l_p(\lambda) = +\infty$. 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.

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

[10] For more details on the identifiability failure that occurs when $M_X(\omega I_n - W) = 0$ see Hillier and Martellosio (2014b).

3.2 The Profile Score

The profile log-likelihood $l_p(\lambda)$ is a.s. differentiable on Λ, with first derivative given by

$$\dot{l}_p(\lambda) = n\left[\frac{y'W'M_X S_\lambda y}{y'S'_\lambda M_X S_\lambda y} - \frac{1}{n}\mathrm{tr}(G_\lambda)\right], \tag{3.5}$$

where $G_\lambda := WS_\lambda^{-1}$. This matrix plays an important role in the sequel.

Differentiability of $l_p(\lambda)$ and the fact that Λ is an open set imply that the MLE must be a root of the equation $\dot{l}_p(\lambda) = 0$. The following result establishes an important property of $l_p(\lambda)$.

Lemma 3.5. *The first-order condition defining the* MLE, $\dot{l}_p(\lambda) = 0$, *is a.s. equivalent to a polynomial equation of degree equal to the number of distinct eigenvalues of* W.

Thus, the equation $\dot{l}_p(\lambda) = 0$ has, for any W, a number of complex roots (counting multiplicities) equal to the number of distinct eigenvalues of W. Any real roots lying in Λ are candidates for $\hat{\lambda}_{\mathrm{ML}}$. Since there is no explicit algebraic solution of polynomial equations of degree higher than four, Lemma 3.5 explains why $\hat{\lambda}_{\mathrm{ML}}$ cannot in general be obtained "in closed form". In spite of this, we shall see in the next section that the cdf of $\hat{\lambda}_{\mathrm{ML}}$ can be represented explicitly. The following result is the basis of the main theorem - Theorem 1 below.

Lemma 3.6. *If all eigenvalues of* W *are real, the function* $l_p(\lambda)$ *a.s. has a single critical point in* Λ, *and that point corresponds to a maximum.*

The key to this result is the observation that, when the pathological cases referred to in Lemma 3.1 are excluded, $l_p(\lambda) \to -\infty$ at both endpoints of Λ. Since $l_p(\lambda)$ is a.s. continuous on the interior of Λ, this implies that Λ must contain *at least one real zero* of $\dot{l}_p(\lambda)$. Under the assumption that all eigenvalues of W are real there is *exactly one* such critical point in Λ. The assumption that all eigenvalues of W are real is stronger than needed for the result in Lemma 3.6, but is convenient for expository purposes, and is satisfied in many applications. We defer a discussion of the possibility of extending the result to complex eigenvalues to Section 7.

Geometrically, Lemma 3.6 says that, when all eigenvalues of W are real, *the profile log-likelihood* $l_p(\lambda)$ *is a.s. single-peaked* on Λ, with no stationary inflection points. The result has clear computational advantages, as it makes numerical optimization of the likelihood much easier.

Remark 3.7. 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 3.5 establishes a connection between the algebraic complexity of $\hat{\lambda}_{\mathrm{ML}}$ and the degree of symmetry satisfied by the graph underlying W.

3.3 Invariance Properties

This section derives some general properties of the MLE for λ that can be deduced directly from the invariance properties of the model and of the profile score equation (3.5) (see, e.g., Lehmann and Romano, 2005). To begin with, observe that the profile score equation (3.5), and hence $\hat{\lambda}_{\mathrm{ML}}$, is invariant to scale transformations $y \to \kappa y$, for any $\kappa > 0$, in the sample space. A first important consequence of this type of invariance is stated next.

Proposition 3.8. *The distribution of $\hat{\lambda}_{\mathrm{ML}}$ induced by a particular distribution of y is constant on the family of distributions generated by forming scale mixtures of the initial distribution of y.*

In particular, all results obtained under Gaussian assumptions continue to hold under scale mixtures of the Gaussian distribution for y, i.e., under spherically symmetric distributions for ε. Thus, assuming (as we will later) a Gaussian distribution for the vector ε is far less restrictive on the generality of the results obtained that it would usually be.

A second consequence of the invariance of $\hat{\lambda}_{\mathrm{ML}}$ is a reduction in the number of parameters indexing the distribution of $\hat{\lambda}_{\mathrm{ML}}$. We denote by θ the finite or infinite dimensional parameter upon which the distribution of ε depends. All parameters $(\beta, \lambda, \sigma^2, \theta)$ are assumed to be identifiable, as this is required for the application of the invariance argument in the proof of Proposition 3.9. 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 3.9. *Assume that the distribution of ε does not depend on β or σ^2. Then,*

(i) if $\operatorname{col}(X)$ is not an invariant subspace of W, the distribution of $\hat{\lambda}_{\mathrm{ML}}$ depends on $(\beta, \lambda, \sigma^2, \theta)$ only through $(\beta/\sigma, \lambda, \theta)$;

(ii) if $\operatorname{col}(X)$ is an invariant subspace of W, the distribution of $\hat{\lambda}_{\mathrm{ML}}$ depends only on (λ, θ).

The condition that $\operatorname{col}(X)$ is an invariant subspace of W holds trivially in the case of pure SAR models (with $\operatorname{col}(X)$ being the trivial invariant subspace $\{0\}$). When there are regressors, the condition is certainly restrictive, but it does hold in important cases. For models in which $X = \iota_n$, for example, the condition holds whenever W is row-stochastic. For any W and X, an easy to check necessary and sufficient condition for $\text{col}(X)$ to be an invariant subspace of W is $M_X W X = 0$.

The case when $\text{col}(X)$ is an invariant subspace of W and the distribution of ε is completely specified (e.g., $\varepsilon \sim \text{N}(0, I_n)$) provides an important theoretical benchmark. In that case, according to Proposition 3.9*(ii)*, the distribution of $\hat{\lambda}_{\text{ML}}$ is completely free of nuisance parameters, making the statistic an ideal basis for inference on λ. Of course, in practice this case is too restrictive, and the distribution of $\hat{\lambda}_{\text{ML}}$ generally depends on any parameter θ affecting in the distribution of ε.

4 Main Results

4.1 The Main Theorem

The key to the main result is the simple observation that the single-peaked property of $l_p(\lambda)$ established in Lemma 3.6 implies that, for any $z \in \Lambda$,

$$\Pr(\hat{\lambda}_{\text{ML}} \le z) = \Pr(\dot{l}_p(z) \le 0),$$

because the single peak of $l_p(\lambda)$ is to the left of a point $z \in \Lambda$ if and only if the slope of $l_p(z)$ is negative. The log-likelihood derivative $\dot{l}_p(\lambda)$ in equation (3.5) can be rewritten as

$$\dot{l}_p(\lambda) = \frac{n}{2} \frac{y' S_\lambda' Q_\lambda S_\lambda y}{y' S_\lambda' M_X S_\lambda y}, \tag{4.1}$$

where

$$Q_\lambda := M_X C_\lambda + C_\lambda' M_X. \tag{4.2}$$

with

$$C_\lambda := G_\lambda - (\text{tr}(G_\lambda)/n) I_n. \tag{4.3}$$

Since only the sign of $\dot{l}_p(z)$ matters, we have the following representation for the cdf of $\hat{\lambda}_{\text{ML}}$.

Theorem 1. *If all eigenvalues of W are real, the* cdf *of $\hat{\lambda}_{\text{ML}}$ at each point $z \in \Lambda$ is given by*

$$\Pr(\hat{\lambda}_{\text{ML}} \le z) = \Pr(y' S_z' Q_z S_z y \le 0). \tag{4.4}$$

Theorem 1 reduces the study of the properties of $\hat{\lambda}_{\text{ML}}$ to the study of the properties of a quadratic form in y. Since quadratic forms have been much-studied in the statistical literature, such a reduction has several computational and analytical advantages, some of which we mention briefly next.

First, equation (4.4) provides a simple way of obtaining the cdf of $\hat{\lambda}_{\mathrm{ML}}$ numerically, *without the need to directly maximize the likelihood*. Indeed, using equation (4.4), it is possible to compute the whole cdf of $\hat{\lambda}_{\mathrm{ML}}$ very efficiently by simply simulating a quadratic form and counting the proportion of negative realizations. This can be done for any parameter configuration, any choices of W and X, and, importantly, any (completely specified) distribution of ε.

Second, equation (4.4) facilitates the construction of bootstrap confidence intervals. Deriving the bootstrap distribution of $\hat{\lambda}_{\mathrm{ML}}$ directly can be very intensive computationally, given the need to repeatedly maximize the likelihood. Theorem 1 says that it is possible to bootstrap a quadratic form instead, a computationally trivial task.

Third, subject to suitable conditions, the first-order asymptotic distribution of $\hat{\lambda}_{\mathrm{ML}}$ follows from Theorem 1 by an application of the results in Kelejiian and Prucha (2001) on the asymptotic distribution of quadratic forms. These properties have been comprehensively studied by Lee (2004), using a related methodology, so need not be repeated here. But Theorem 1 also provides a direct route to obtaining a more accurate approximation to the distribution of $\hat{\lambda}_{\mathrm{ML}}$ - for example, by using a saddlepoint approximation for the distribution of the quadratic form $y'S_z'Q_zS_zy$ - but these matters are not our focus here.

In the present paper we are instead concerned with the *exact* consequences of Theorem 1. Not surprisingly, such analysis requires imposing additional structure on the model, which we will do gradually.[11] We begin by pointing out some simple but important general results that can be seen immediately from (4.4).

4.2 Some Exact Consequences

It is convenient to rewrite (4.4) as

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z) = \Pr\left(\tilde{y}'A(z,\lambda)\tilde{y} \leq 0\right), \tag{4.5}$$

where $\tilde{y} := S_\lambda y = X\beta + \sigma\varepsilon$, and

$$A(z,\lambda) := (S_z S_\lambda^{-1})' Q_z (S_z S_\lambda^{-1}). \tag{4.6}$$

The structure of the matrix $A(z,\lambda)$ is evidently crucial in determining the properties of the MLE. In particular, if $\varepsilon \sim \mathrm{N}(0, I_n)$, 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 noncentral) χ^2 variates, with coefficients the distinct eigenvalues of $A(z,\lambda)$. This would

[11] It is worth noting here that fairly strong assumptions - particularly about the evolution of W, but also about the relationship of W to X - are also needed for the asymptotic analysis of $\hat{\lambda}_{\mathrm{ML}}$ - see Lee (2004).

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. This will become clearer as we proceed.[12]

Next, observe that, because only the sign of the quadratic form in (4.5) matters, we can divide the statistic $\tilde{y}'A(z, \lambda)\tilde{y}$ by any positive quantity, without altering the probability. Dividing by $\tilde{y}'\tilde{y}$, we obtain the form $h'A(z, \lambda)h$, where $h := \tilde{y}/(\tilde{y}'\tilde{y})^{1/2}$ is distributed on the unit sphere in n dimensions, $\mathcal{S}^{n-1}$. This representation allows one to appeal to known results for quadratic forms defined on the sphere. In particular, with the added assumption that the distribution of ε is spherically symmetric, h is uniformly distributed on $\mathcal{S}^{n-1}$ in the pure SAR model, but in general non-uniformly distributed on $\mathcal{S}^{n-1}$ in the presence of regressors. An expression for the cdf suitable for the latter case is given in Forchini (2005), while the uniformly distributed case was dealt with in Hillier (2001).

In both of these cases the results in Mulholland (1965) and Saldanha and Tomei (1996) suggest that there may be a number of points $z \in \Lambda$ at which the distribution function of $\tilde{y}'A(z, \lambda)\tilde{y}$ (and hence of $\hat{\lambda}_{\mathrm{ML}}$) will be *non-analytic*, and the cdf will have a different functional form in the intervals between such points. This is indeed the case, and this property of the distribution of $\hat{\lambda}_{\mathrm{ML}}$ is not a mere curiosity: for any (W, X) there will usually be a number of points at which the cdf is non-analytic. Importantly, this result does not depend on the distribution assumptions made (see Forchini, 2002), and in some cases these properties of the distribution persist asymptotically, the Complete Bipartite model being one example. We will come back to the analyticity issue in Section 5 for the case of a pure model. An example will be given in Section 6.2.1.

Before continuing, we remark that the argument used to obtain Theorem 1 has implications for the relationship between $\hat{\lambda}_{\mathrm{ML}}$ and the ordinary least squares estimator, $\hat{\lambda}_{\mathrm{OLS}}$.

Proposition 4.1. *When all eigenvalues of W are real the distribution function of $\hat{\lambda}_{\mathrm{OLS}}$ is above that of $\hat{\lambda}_{\mathrm{ML}}$ for $\hat{\lambda}_{\mathrm{OLS}} < 0$, crosses it at $\hat{\lambda}_{\mathrm{OLS}} = \hat{\lambda}_{\mathrm{ML}} = 0$, and is below it for $\hat{\lambda}_{\mathrm{OLS}} > 0$.*

The proof is immediate from the fact that, when defined, $\hat{\lambda}_{\mathrm{OLS}}$ is the solution to $y'W'M_X S_\lambda y = 0$, so that $\dot{l}_p(\hat{\lambda}_{\mathrm{OLS}}) = -\mathrm{tr}(G_{\hat{\lambda}_{\mathrm{OLS}}})$, and the easily established fact that, if all the eigenvalues of W are real, $\mathrm{tr}(G_\lambda)$ has the same sign as λ.[13] The

[12]The particular case $z = \lambda$, corresponding to $\Pr(\hat{\lambda}_{\mathrm{ML}} \leq \lambda)$, is especially important. In that case $A(\lambda, \lambda) = Q_\lambda$, so $\Pr(\hat{\lambda}_{\mathrm{ML}} \leq \lambda) = \Pr(\tilde{y}'Q_\lambda\tilde{y} \leq 0)$. Apart from providing a simple device for computing the probability of underestimating λ, it is also clear that the asymptotic behavior of $\hat{\lambda}_{\mathrm{ML}}$ is governed by that of the quadratic form $\tilde{y}'Q_\lambda\tilde{y}$.

[13]When all eigenvalues of W are real $d\mathrm{tr}(G_\lambda)/d\lambda = \mathrm{tr}(G_\lambda^2) > 0$, so that $\mathrm{tr}(G_\lambda)$ is monotonic increasing in λ, and $\mathrm{tr}(G_0) = 0$.

single-peaked property of $l_p(\lambda)$ means that $\hat{\lambda}_{\text{OLS}} < 0$ implies $\dot{l}_p(\hat{\lambda}_{\text{OLS}}) > 0$ so that $\hat{\lambda}_{\text{OLS}} < \hat{\lambda}_{\text{ML}}$, $\hat{\lambda}_{\text{OLS}} = 0$ implies $\hat{\lambda}_{\text{OLS}} = \hat{\lambda}_{\text{ML}}$, and $\hat{\lambda}_{\text{OLS}} > 0$ implies $\dot{l}_p(\hat{\lambda}_{\text{OLS}}) < 0$ so that $\hat{\lambda}_{\text{OLS}} > \hat{\lambda}_{\text{ML}}$. It is worth emphasizing that Proposition 4.1 holds for any X, and any distribution of ε.[14]

Thus, for instance, $\Pr(\hat{\lambda}_{\text{OLS}} < \lambda)$ is greater than (less than) $\Pr(\hat{\lambda}_{\text{ML}} < \lambda)$ for any negative (positive) value of λ, and the two coincide when $\lambda = 0$. Also, the density of $\hat{\lambda}_{\text{ML}}$ is necessarily above that of $\hat{\lambda}_{\text{ML}}$ at the origin. We do not investigate the properties of the OLS estimator further in the present paper.

4.3 A Canonical Form

It is clear that while Theorem 1 permits, in principle at least, an exact analysis of the properties of $\hat{\lambda}_{\text{ML}}$ for any given W and X, the distribution theory is complicated, and probably intractable. However, by imposing some additional structure on the problem we can use the result to gain more insight into the exact distributional properties of $\hat{\lambda}_{\text{ML}}$. In particular, we assume now that W is similar to a symmetric matrix, i.e., that it is diagonalizable and has real eigenvalues. Recall that the condition that W is similar to a symmetric matrix is satisfied whenever W is a row-standardized version of a symmetric matrix.

In the remainder of the paper we first discuss some further general results that, under this additional assumption, are reasonably straightforward consequences of Theorem 1, and then, in Section 6, explore the detailed consequences of Theorem 1 for the examples described earlier. First we show that, under this new assumption, the quadratic form in equation (4.4) can be expressed in a canonical form which helps to simplify analysis of its consequences.

To begin with, let us fix some notation. Let T denote the number of distinct eigenvalues of W. If the distinct eigenvalues of W are real we denote them by, in ascending order, $\omega_1, \omega_2, ..., \omega_T$, the eigenvalue ω_t occurring with algebraic multiplicity n_t (so that $\sum_{t=1}^{T} n_t = n$). Thus, $\omega_1 = \omega_{\min}$ and $\omega_T = 1$. Also, let

$$\gamma_t(z) := \frac{\omega_t}{1 - z\omega_t} - \frac{1}{n} \sum_{t=1}^{T} \frac{n_s \omega_s}{1 - z\omega_s},$$

$t = 1, ..., T$, be the distinct eigenvalues of the matrix C_z in (4.3). 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 $D := \operatorname{diag}(\omega_t I_{n_t}, t = 1, ..., T)$. Under this assumption the matrix $A(z; \lambda)$ in (4.6) can be

[14]The support of $\hat{\lambda}_{\text{OLS}}$ can be larger than Λ, but this single-crossing property also applies for $\hat{\lambda}_{\text{OLS}}$ outside Λ, where the cdf of $\hat{\lambda}_{\text{ML}}$ must necessarily be either 0 or 1.

reduced to the form $A(z, \lambda) = (H')^{-1}B(z; \lambda)H^{-1}$, with

$$B(z; \lambda) = \{d_{st}M_{st}; s, t = 1, ..., T\}, \tag{4.7}$$

where M_{st} is the $n_s \times n_t$ submatrix of $M := H'M_X H$ associated to the eigenvalues (ω_s, ω_t), and the coefficients d_{st} are given by

$$d_{st} := \frac{(1 - z\omega_s)(1 - z\omega_t)}{(1 - \lambda\omega_s)(1 - \lambda\omega_t)} [\gamma_s(z) + \gamma_t(z)] = d_{ts}. \tag{4.8}$$

(see Appendix A for details). Note that the coefficients d_{st} are functions of z, λ, and W, but do not depend on X, and $d_{tt} = -2\text{tr}(G_z)/n$ for all $z \in \Lambda$ if $\omega_t = 0$. Some useful properties of the coefficient functions d_{tt} are given in Proposition A.1.

Under our current assumption, it is through the matrix M that the relationship between W and X is manifest. Writing $x := H^{-1}\tilde{y}$ (where, recall, $\tilde{y} = S_\lambda y$) and partitioning x conformably with the partition of M (so that x_t is $n_t \times 1$, for $t = 1, ..., T$), we obtain the following results.

Proposition 4.2. *(i) If W is similar to a symmetric matrix,*

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr\left(\sum_{t=1}^{T} d_{tt}(x_t' M_{tt} x_t) + 2 \sum_{s,t=1,s>t}^{T} d_{st}(x_s' M_{st} x_t) \leq 0\right). \tag{4.9}$$

(ii) If W is similar to a symmetric matrix, the bilinear terms in (4.9) all vanish if and only if the matrix $M_X W$ is symmetric. In that case,

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr\left(\sum_{t=1}^{T} d_{tt}(x_t' M_{tt} x_t) \leq 0\right). \tag{4.10}$$

(iii) If W and $M_X W$ are both symmetric (4.10) simplifies further to

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z) = \Pr\left(\sum_{t=1}^{T} d_{tt}(\tilde{x}_t' \tilde{x}_t) \leq 0\right), \tag{4.11}$$

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 $\text{col}(X)$.

Equation (4.9) provides a general canonical representation of the cdf of $\hat{\lambda}_{\text{ML}}$ in terms of a linear combination of quadratic and bilinear forms in the vectors x_t. Under the additional conditions in Proposition 4.2 (ii) and (iii), the representation contains only quadratic forms in the x_t, and subvectors of them. Note that, under the assumption that the error ε has a spherical Gaussian distribution, the vectors x_t are independent in part (iii), because H is orthogonal in that case, but not in parts (i) or (ii).

Remark 4.3. If W and $M_X W$ are both symmetric, then $\mathrm{col}(X)$ is spanned by k linearly independent eigenvectors of W. It follows from Proposition 3.9*(ii)* that, assuming that the distribution of ε does not depend on β or σ^2, the distribution defined in (4.11) does not depend on β and σ^2 either.

Two examples where $M_X W$ is symmetric will be met in Section 6: the balanced Group Interaction model with constant mean, and the Complete Bipartite model with row-standardized W and constant mean. Another example is an *unbalanced* Group Interaction model, with r groups of different sizes, m_i, $i = 1, .., r$, and $X = \bigoplus_{i=1}^{r} \iota_{m_i}$ (i.e., X contains an intercept for each of the r groups, and no other regressors).[15]

4.4 Support of the MLE

We are now in a position to discuss another important consequence of Theorem 1: *the support of* $\hat{\lambda}_{\mathrm{ML}}$ *is not necessarily the entire interval* Λ.[16] 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 $\hat{\lambda}_{\mathrm{ML}}$, it turns out that in the context of Proposition 4.2 (ii) the conditions that do so are straightforward, and we confine ourselves here to that case. The assumptions underlying Proposition 4.2 (ii) are certainly 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 MLE are the values of λ for which the matrix Q_λ is *indefinite* (see equation (4.1)). 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 ($\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z) = 0$), or certain ($\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z) = 1$). In such cases the support of $\hat{\lambda}_{\mathrm{ML}}$ is a proper subset of Λ. This cannot happen for the pure SAR model, because in that case $Q_z = (G_z + G_z') - n^{-1}\mathrm{tr}(G_z + G_z')I_n$, which is necessarily indefinite (since $n^{-1}\mathrm{tr}(G_z + G_z')$ 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 $\hat{\lambda}_{\mathrm{ML}}$ 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 $t = 2, ..., T-1$, z_t denotes the unique point $z \in \Lambda$ at which $\gamma_t(z) = 0$ (see Proposition A.1 in Appendix A).

[15] Note that here it is essential that the model is unbalanced: as we have seen in Section 3.1, the MLE does not exist in the balanced case if X includes group fixed effects.

[16] By support of (the distribution of) $\hat{\lambda}_{\mathrm{ML}}$ we mean the set on which the density of $\hat{\lambda}_{\mathrm{ML}}$ is positive, if the density exists. If the density does not exists then we can define the support as the largest subset of $*$ for which every open neighbourhood of every point of the set has positive measure.

Proposition 4.4. *Assume that W is similar to a symmetric matrix and $M_X W$ is symmetric.*

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

It is useful to provide some interpretation, and some examples, for the result in Proposition 4.4. In the context of Proposition 4.4, $\hat{\lambda}_{\text{ML}}$ cannot, in particular, be positive if $\text{col}(X)$ contains all eigenvectors of W associated to positive eigenvalues, even if the true value of λ is positive.[17] 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, our support result 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 with the row-standardized Complete Bipartite model when $X = \iota_n$, because in that case ι_n spans the eigenspace of W corresponding to the eigenvalue 1, and 1 is the only positive eigenvalue of W. Thus in this model $\hat{\lambda}_{\text{ML}}$ cannot be positive, even if the true value of λ is positive - see also Section 6.2.2. Another simple example for which $\hat{\lambda}_{\text{ML}}$ is not supported on the whole Λ is the unbalanced Group Interaction model, when there are group fixed effects and no other regressors (see Hillier and Martellosio, 2014a).

The restricted support phenomenon certainly seems to demand further investigation, but this is beyond the scope of the present paper. We conclude this section with two remarks. Firstly, it is clear 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. Secondly, the restricted support phenomenon is not confined to the MLE, but also applies to other estimators in the SAR model.

5 Gaussian Pure SAR Model with Symmetric W

We now show that the exact results above simplify considerably when (i) there are no regressors, (ii) W is symmetric, and (iii) ε is a scale mixture of the $\text{N}(0, I_n)$ distribution. The resulting model provides a fairly simple context in which to discuss

[17] This is because, in that case, z_t in Proposition 4.4 (i) must be nonpositive, by Proposition A.1 in Appendix A, and the fact that $\gamma_t(0) = \omega_t$.

general properties of the distribution of the MLE. Bao and Ullah (2007) have given finite sample approximations to the moments of the MLE in a Gaussian pure SAR model. Our focus here is on the exact distribution of the MLE.

According to Proposition 3.8 any property of the distribution of $\hat{\lambda}_{\mathrm{ML}}$ that holds under the assumption $\varepsilon \sim \mathrm{N}(0, I_n)$ continues to hold under the assumption that ε belongs to the family of scale mixtures of $\mathrm{N}(0, I_n)$, which we denote by $\varepsilon \sim \mathrm{SMN}(0, I_n)$. Note that these are spherically symmetric distributions for ε, which need not be i.i.d. Letting, here and elsewhere, χ^2_ν denote a (central) χ^2 random variable with ν degrees of freedom, Proposition 4.2 (iii) yields the following result:[18]

Theorem 2. *In a pure SAR model, if W is symmetric and $\varepsilon \sim \mathrm{SMN}(0, I_n)$,*

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z) = \Pr\left(\sum_{t=1}^{T} d_{tt}\chi^2_{n_t} \leq 0\right), \tag{5.1}$$

where the $\chi^2_{n_t}$ variates are independent, for any $z \in \Lambda$.

The highly structured representation of the cdf in Theorem 2 has several consequences. We first discuss two straightforward, but important, corollaries of Theorem 2, and then we will move to derive an explicit formula for the cdf in 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 (5.1), the distribution of $\hat{\lambda}_{\mathrm{ML}}$, and hence all of its properties, depends on W only through its spectrum.

Corollary 5.1. *In a pure SAR model with $\varepsilon \sim \mathrm{SMN}(0, I_n)$, the distribution of $\hat{\lambda}_{\mathrm{ML}}$ is constant on the set of cospectral symmetric weights matrices.*

One simple application of Corollary 5.1 is as follows: since the spectrum of the weights matrix (2.3) depends on p and q only through their sum n, *the distribution of $\hat{\lambda}_{\mathrm{ML}}$ is the same for any pure Gaussian symmetric Complete Bipartite model on n observational units*, regardless of the partition of n into p and q. In case p or q 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).[19] It is a simple exercise to show that all such weights matrices have (after normalization by the spectral radius) eigenvalues 0, with multiplicity $n - 2$, and -1, 1, and hence are

[18] This result can also be obtained directly from equation (4.5), since, under our current assumptions, the d_{tt} are eigenvalues of $A(z; \lambda)$.

[19] That is, W is not restricted to be the $(0, 1)$ adjacency matrix associated to the star graph, but is allowed to be any symmetric matrix compatible with that graph.

cospectral with the adjacency matrix of the graph. We conclude that *the distribution of* $\hat{\lambda}_{\mathrm{ML}}$ *is the same for any Gaussian pure SAR model with symmetric weights matrix compatible with a star graph.*

Another application of Corollary 5.1 is to (non-isomorphic, to avoid trivial cases) cospectral graphs, which are well studied in graph theory; see, e.g., Biggs (1993). Corollary 5.1 implies that the distribution of $\hat{\lambda}_{\mathrm{ML}}$ is constant on the family of pure Gaussian SAR models with weights matrices that are the adjacency matrices of cospectral graphs.

A second corollary to Theorem 2 can be deduced for matrices W with *symmetric spectrum*. The spectrum of a matrix is said to be symmetric if, whenever ω is eigenvalue, $-\omega$ is also an eigenvalue, with the same algebraic multiplicity.[20] The weights matrix of a balanced Group Interaction model with $m = 2$ is an example of this type, as is that of the Complete Bipartite model, when symmetrically normalized.[21]

Corollary 5.2. *In a pure SAR model with* $\varepsilon \sim \mathrm{SMN}(0, I_n)$*,* W *symmetric, and the spectrum of* W *symmetric about the origin, the density of* $\hat{\lambda}_{\mathrm{ML}}$ *satisfies the symmetry property* $\mathrm{pdf}_{\hat{\lambda}_{\mathrm{ML}}}(z; \lambda) = \mathrm{pdf}_{\hat{\lambda}_{\mathrm{ML}}}(-z; -\lambda)$.

That is, under the stated assumptions, the density of $\hat{\lambda}_{\mathrm{ML}}$ when $\lambda = \lambda_0$ is the reflection about the vertical axis of the density when $\lambda = -\lambda_0$. This implies, in particular, that (subject to its existence) the mean of $\hat{\lambda}_{\mathrm{ML}}$ satisfies $\mathrm{E}(\hat{\lambda}_{\mathrm{ML}}; \lambda) = -\mathrm{E}(\hat{\lambda}_{\mathrm{ML}}; -\lambda)$.

5.1 Exact Distribution

Theorem 2 shows that in pure SAR models with symmetric W the cdf of $\hat{\lambda}_{\mathrm{ML}}$ is induced by that of a linear combination of independent χ^2 random variables with coefficients d_{tt}. Proposition A.1 in Appendix A says that, in this representation, except for d_{11} and d_{TT}, each coefficient changes sign exactly once on Λ, so that the number of positive and negative coefficients changes exactly $T-2$ times as z varies in Λ. By an extension of the argument in Saldanha and Tomei (1996),[22] this implies that the distribution function of $\hat{\lambda}_{\mathrm{ML}}$ is non-analytic at these $T-2$ points, but analytic everywhere between them. This is an example of the non-analyticity property of the

[20]Note that if W is non-negative and normalised to have largest eigenvalue 1, then $\Lambda = (-1, 1)$ when W has a symmetric spectrum.

[21]In 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.

[22]Saldanha 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}. Reinterpreted, their Theorem says that whenever an eigenvalue vanishes, the cdf will be non-analytic at the origin, the point of interest for us.

distribution mentioned above: *in a pure SAR model with W symmetric and $T > 2$, the cdf of $\hat{\lambda}_{\rm ML}$ is non-analytic at the $T-2$ points z_t where the $\gamma_t(z)$ change sign, and has a different functional form on each interval between those points.* We may now use this fact to obtain an explicit form for the cdf of $\hat{\lambda}_{\rm ML}$ in such models.[23]

Now, for a fixed $z \in \Lambda$ at which none of the d_{tt} vanishes, let $T_1 = T_1(z)$ and $T_2 = T_2(z)$ denote the numbers of positive and negative terms d_{tt}, respectively, in (5.1), 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, partition x into (x_1', x_2'), with x_i of dimension $v_i \times 1$, for $i = 1, 2$, and let A_1 be the $v_1 \times v_1$ matrix $\mathrm{diag}(d_{tt} I_{n_t}; t = 1, .., T_1)$, and A_2 the $v_2 \times v_2$ matrix $\mathrm{diag}(-d_{tt} I_{n_t}; t = T_1 + 1, .., T)$. Both matrices are diagonal with positive diagonal elements, and as z varies the dimensions of the two square matrices A_1 and A_2 necessarily vary (subject to $v_1 + v_2 = n$).

Let $Q_i := x_i' A_i x_i$, for $i = 1, 2$. The statistics Q_1 and Q_2 are independent linear combinations of central χ^2 random variables with positive coefficients. From (5.1),

$$\Pr(\hat{\lambda}_{\rm ML} \leq z) = \Pr(Q_1 \leq Q_2) = \Pr(R \leq 1), \tag{5.2}$$

where $R := Q_1/Q_2$. That is, the distribution of $\hat{\lambda}_{\rm ML}$ in symmetric Gaussian pure SAR models is determined by that of a ratio of positive linear combinations of independent χ^2 random variables *at the fixed point* $r = 1$.

Before giving the general result, notice that if $T = 2$ (i.e., W has only two distinct eigenvalues), then $T_1 = T_2 = 1$, $v_1 = n_1, v_2 = n_2, Q_1 = d_{11}\chi^2_{n_1}$, $Q_2 = d_{22}\chi^2_{n_2}$, and so from (5.2) we obtain

$$\Pr(\hat{\lambda}_{\rm ML} \leq z) = \Pr\left(\mathrm{F}_{n_1,n_2} \leq -\frac{n_2 d_{22}}{n_1 d_{11}}\right). \tag{5.3}$$

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. We will shortly see that the balanced Group Interaction model has this form. For $T > 2$ the distribution will have a different form on each of the $T - 1$ segments of Λ that result from the d_{tt} changing sign for each $t \neq 1, T$.

To state the general result, let $C_j(A)$ denote the top-order zonal polynomial of order j in the eigenvalues of the matrix A (Muirhead, 1982, Chapter 7), i.e., the coefficient of ξ^j in the expansion of $(\det(I_n - \xi A))^{-1/2}$. Then, the result for general T is the following consequence of Theorem 2.

[23]The cdf of the OLS estimator has exactly the same form as equation (5.1), under the same assumptions, but with the d_{tt} replaced by $\omega_t(1 - z\omega_t)/(1 - \lambda\omega_t)^2$. Again, some of these must be positive, some negative, for $z \in \Lambda$. The results to follow therefore also hold for the OLS estimator with this modification.

Corollary 5.3. *If W is symmetric and $\varepsilon \sim \mathrm{SMN}(0, I_n)$, then for any pure SAR model, for z in the interior of any one of the $T-1$ intervals in Λ determined by the points of non-analyticity, z_t,*

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

where $\tau_i := \mathrm{tr}(A_i^{-1})$ and $\tilde{A}_i := I_{v_i} - (\tau_i A_i)^{-1}$, for $i = 1, 2$.[24]

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 $\hat{\lambda}_{\mathrm{ML}}$ directly from (5.4), but we shall see in Section 6 that this problem can often be avoided by a conditioning argument.

The introduction of regressors, or the removal of the assumption that W is symmetric, does not change the general nature of these results. A generalized version of equation (5.4) for the SAR model with arbitrary X can certainly be obtained, but would require lengthy explanation. Instead, to conclude this section we provide a simple generalization of Theorem 2 to the model with W symmetric and regressors present, but subject to a restriction on the relationship between W and X. Indeed, when the assumption $\varepsilon \sim \mathrm{SMN}(0, I_n)$ is added, Proposition 4.2 (iii) assumes the following form.

Theorem 3. *Assume that W is symmetric, $\varepsilon \sim \mathrm{SMN}(0, I_n)$, and $\mathrm{col}(X)$ is spanned by k linearly independent eigenvectors of W. Then the cdf of $\hat{\lambda}_{\mathrm{ML}}$ is given by*

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z) = \Pr\left(\sum_{t=1}^{T} d_{tt} \chi^2_{n_t - n_t(X)} \leq 0\right), \quad (5.5)$$

where the χ^2 variates involved are central, and independent, and $\chi^2_0 = 0$.

It is clear here that the cdf of $\hat{\lambda}_{\mathrm{ML}}$ in equation (5.5) depends only on λ (i.e., is free of (β, σ^2)), as also follows from part (ii) of Proposition 3.9. An explicit expression for the cdf analogous to that in Corollary 5.3 obviously holds, as do the other corollaries of Theorem 2 discussed above, with only minor modifications.

[24] It is easily confirmed that the cdf (5.4) 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 $\mathrm{F}_{v_1+2j, v_2+2k}$. 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 $\Pr(K = k)$.

Remark 5.4. The convention $\chi^2_0 = 0$ means that any term for which $n_t(X) = n_t$ does not appear in the sum on the right in (5.5). For example, in the Complete Bipartite model the eigenspaces associated with the eigenvalues ± 1 are both one-dimensional, so if either of these is in $\mathrm{col}(X)$ that term does not appear. Subject to the other conditions of Theorem 3 holding, the cdf is then particularly simple, involving only two independent χ^2 variates.

Remark 5.5. In some models a special case of the condition used in Theorem 3 holds, in that $\mathrm{col}(X)$ is contained in a single eigenspace of W. In that case the columns of X itself are eigenvectors of W, and the condition needed automatically holds. In that case we have the following simpler form of equation (5.5): if $\mathrm{col}(X)$ is a subspace of the eigenspace associated to the eigenvalue ω_t, then

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z) = \Pr\left(d_{tt}\chi^2_{n_t - k} + \sum_{s=1; s\neq t}^{T} d_{ss}\chi^2_{n_s} \leq 0 \right). \tag{5.6}$$

For example, in the *unbalanced* Group Interaction model with $X = \bigoplus_{i=1}^{r} \iota_{m_i}$ the columns of X are eigenvectors associated with the unit eigenvalue. Hence, equation (5.6) holds with $k = r$.

6 Applications

In this section we apply the general results to the examples introduced earlier. Our main purpose here is to illustrate the various aspects of the distribution of $\hat{\lambda}_{\mathrm{ML}}$ we have studied, but we also provide some completely new exact results for these examples, and some new asymptotic results for cases not covered by Lee's (2004) assumptions. We consider the balanced Group Interaction Model in Section 6.1, and the Complete Bipartite model in Section 6.2.[25] To keep the analysis as simple as possible, we confine ourselves to the pure case and the constant mean case, and we assume $\varepsilon \sim \mathrm{SMN}(0, I_n)$. Extensions to more general cases are certainly possible, but are not pursued here.

[25]For the balanced Group interaction model, and the Complete Bipartite model, $\hat{\lambda}_{\mathrm{ML}}$ is the unique root in Λ of either a quadratic or a cubic (by Lemma 3.5), 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.

6.1 The Balanced Group Interaction Model

6.1.1 Zero Mean

Because the matrix (2.1) has only two distinct eigenvalues, equation (5.3) applies, giving the following strikingly simple result.

Proposition 6.1. *In the pure balanced Group Interaction model with* $\varepsilon \sim \mathrm{SMN}(0, I_n)$*, the* cdf *of* $\hat{\lambda}_{\mathrm{ML}}$ *is, for* $z \in \Lambda$*,*

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z) = \Pr(\mathrm{F}_{r,r(m-1)} \leq c(z,\lambda)), \tag{6.1}$$

where

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

Taking $z = \lambda$, equation (6.1) gives $\Pr(\hat{\lambda}_{\mathrm{ML}} \leq \lambda) = \Pr(\mathrm{F}_{r,r(m-1)} \leq 1)$. Thus, in this model the probability of underestimating λ is independent of the true value of λ. A necessary condition for the consistency of $\hat{\lambda}_{\mathrm{ML}}$ is clearly that $\mathrm{F}_{r,r(m-1)} \to_p 1$, which suggests that $r \to \infty$ will be sufficient, but $m \to \infty$ may not.[26] More on the asymptotics for this model below.

Given the cdf we can immediately obtain the density.

Proposition 6.2. *In the pure balanced Group Interaction model with* $\varepsilon \sim \mathrm{SMN}(0, I_n)$*, the density of* $\hat{\lambda}_{\mathrm{ML}}$ *is, for* $z \in \Lambda$*,*

$$\mathrm{pdf}_{\hat{\lambda}_{\mathrm{ML}}}(z;\lambda) = \frac{2m\delta^{\frac{r}{2}}}{B(\frac{r}{2},\frac{r(m-1)}{2})} \frac{(1-z)^{r(m-1)-1}\,(z+m-1)^{r-1}}{[(1-z)^2+\delta(z+m-1)^2]^{\frac{rm}{2}}}, \tag{6.2}$$

where $\delta := (1-\lambda)^2/[(m-1)(\lambda+m-1)^2]$.

Figure 1 displays the density (6.2) 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 $\hat{\lambda}_{\mathrm{ML}}$ depends on m).

In this model, if $r \to \infty$ is assumed, Lee's (2004) Assumptions 3 and 8' are satisfied, as is his condition (4.3), so $\hat{\lambda}_{\mathrm{ML}}$ is consistent and asymptotically normal by Lee's Theorems 4.1 and 4.2. On the other hand, if $n \to \infty$ because $m \to \infty$

[26] $\mathrm{E}(\mathrm{F}_{r,r(m-1)}) \to 1$ as either r or $m \to \infty$, but $\mathrm{var}(\mathrm{F}_{r,r(m-1)}) \to 0$ when $r \to \infty$, but not when $m \to \infty$.

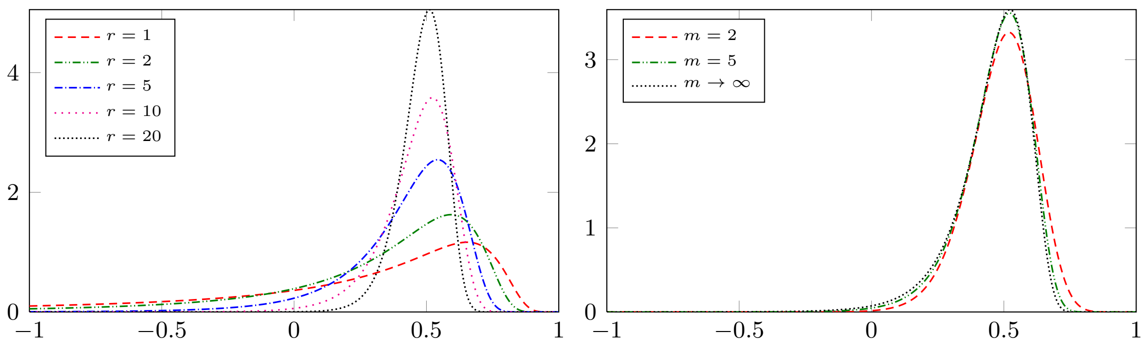

Figure 1: Density of $\hat{\lambda}_{\mathrm{ML}}$ for the Gaussian pure balanced Group Interaction model with $\lambda = 0.5$, and with $m = 10$ (left panel), $r = 10$ (right panel).

Lee's Assumption 3 is *not* satisfied, and his results leave open that $\hat{\lambda}_{\mathrm{ML}}$ may be inconsistent in this case. This is an example of so-called infill asymptotics. In fact, it may easily be shown (using equation (6.1) and the known result $v_1 \mathrm{F}_{v_1,v_2} \to_d \chi^2_{v_1}$ as $v_2 \to \infty$) that, for fixed r,

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z) \overset{m\to\infty}{\longrightarrow} \Pr\left(\chi^2_r \leq r\left(\frac{1-\lambda}{1-z}\right)^2\right), \quad -\infty < z < 1.$$

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

$$\mathrm{pdf}_{\hat{\lambda}_{\mathrm{ML}}}(z;\lambda) \overset{m\to\infty}{\longrightarrow} \frac{r^{\frac{r}{2}}(1-\lambda)^r}{2^{\frac{r}{2}-1}\Gamma(\frac{r}{2})(1-z)^{r+1}} e^{-\frac{r}{2}\left(\frac{1-\lambda}{1-z}\right)^2},$$

so $\hat{\lambda}_{\mathrm{ML}}$ converges to a random variable supported on $(-\infty, 1)$. It is clear from Figure 1 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.

6.1.2 Constant Mean

The results given above for the pure balanced Group Interaction model can be extended immediately to the case of an unknown constant mean (i.e., $X = \iota_n$) by using Theorem 3 (in fact the stronger version in equation (5.6)), because ι_n is in the eigenspace associated to the unit eigenvalue.

Proposition 6.3. *For the balanced Group Interaction model with $X = \iota_n$ and $\varepsilon \sim \mathrm{SMN}(0, I_n)$, the* cdf *of $\hat{\lambda}_{\mathrm{ML}}$ is, for $z \in \Lambda$,*

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

Because this is only a trivial modification of the result in Proposition 6.1, we omit further details for this case.

The exact results given in Propositions 6.1, 6.2 and 6.3 enable a complete analysis of the exact properties of $\hat{\lambda}_{\mathrm{ML}}$, and the results needed for inference based upon it. For example, exact expressions for the moments and the median of $\hat{\lambda}_{\mathrm{ML}}$, and exact confidence intervals for λ based on $\hat{\lambda}_{\mathrm{ML}}$ can be obtained quite directly; see Hillier and Martellosio (2014a). Hillier and Martellosio (2014a) also provides a detailed analysis of the unbalanced case (groups are not all of the same size). An important consequence of unbalancedness is the introduction in the distribution of $\hat{\lambda}_{\mathrm{ML}}$ of points of non-analyticity.

6.2 The Complete Bipartite Model

We now apply the general results to the Complete Bipartite model introduced in Section 2.3. In Section 6.2.1 we discuss the simple case of a pure symmetric Complete Bipartite model. Then, in Section 6.2.2, we discuss the case of the row-standardized Complete Bipartite model with unknown constant mean (i.e., $X = \iota_n$). This provides an important illustration of the restricted support phenomenon described in Section 4.4.

6.2.1 Symmetric W, Zero Mean

In the symmetric Complete Bipartite model, W again has $T = 3$ distinct eigenvalues: -1, 0, 1. According to Corollary 5.3, the pdf of $\hat{\lambda}_{\mathrm{ML}}$ 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 5.2 may be used to obtain the density for $z \in (-1, 0)$ from that for $z \in (0, 1)$.

Proposition 6.4. *In the pure symmetric Complete Bipartite model with $\varepsilon \sim \mathrm{SMN}(0, I_n)$,*

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z) = \Pr(\phi_1 \chi_1^2 \leq \phi_2 \chi_1^2 + 2z \chi_{n-2}^2), \tag{6.3}$$

for $-1 < z < 1$, where

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

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

Proposition 6.4 confirms the fact remarked upon in the discussion of Corollary 5.1, that the distribution, and hence all the properties of $\hat{\lambda}_{\mathrm{ML}}$, depends on p and q only through their sum n.[27] The coefficients ϕ_1, ϕ_2 in (6.3) are both positive for all $z \in \Lambda = (-1, 1)$, but z changes sign of course. Applying a conditioning argument discussed in Hillier and Martellosio (2014a), we obtain the following proposition, where ${}_2F_1(\cdot)$ denotes the Gaussian Hypergeometric function (e.g., Muirhead, 1982, Chapter 1).

Proposition 6.5. *In the pure symmetric Complete Bipartite model with $\varepsilon \sim \mathrm{SMN}(0, I_n)$ the density of $\hat{\lambda}_{\mathrm{ML}}$ for $z \in (0, 1)$ is*

$$\mathrm{pdf}_{\hat{\lambda}_{\mathrm{ML}}}(z;\lambda) = \frac{B\left(\frac{1}{2},\frac{n}{2}\right)c}{2\pi a^{\frac{1}{2}}(1+c)^{\frac{n}{2}}}\left[\frac{\alpha\dot{a}}{a}{}_2F_1\left(\frac{n}{2},\frac{3}{2},\frac{n+1}{2};\eta\right)\right.$$
$$\left.+\frac{\beta\dot{c}}{c}{}_2F_1\left(\frac{n}{2},\frac{1}{2},\frac{n+1}{2};\eta\right)\right], \quad (6.4)$$

where $a := \phi_2/\phi_1$, $c := 2z/\phi_1$, and $\eta := \phi_1(\phi_2 - 2z)/\phi_2(\phi_1 + 2z)$. For $z \in (-1, 0)$ the density is defined by $\mathrm{pdf}_{\hat{\lambda}_{\mathrm{ML}}}(z;\lambda) = \mathrm{pdf}_{\hat{\lambda}_{\mathrm{ML}}}(-z;-\lambda)$.

The asymptotic distribution as $n \to \infty$ 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 $\bar{\phi}_1 := \lim_{n\to\infty}(\phi_1/(n-2)) = (1-z)^2(1+z)/(1-\lambda)^2$ and $\bar{\phi}_2 := \lim_{n\to\infty}(\phi_2/(n-2)) = (1+z)^2(1-z)/(1+\lambda)^2$. Therefore, $V_n \to_d \bar{V}_n$, and so (from Proposition 6.4), $\Pr(\hat{\lambda}_{\mathrm{ML}} \le z) \to \Pr\left(\chi_1^2 \le \bar{\psi}_1\chi_1^2 + \bar{\psi}_2\right)$, with

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

for $z \in (0, 1)$, and the two χ_1^2 variates are independent. For $z \in (0, 1)$, therefore, the usual conditioning argument yields

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \le z) \to \mathrm{E}_{q_1}\left[\mathcal{G}_1(\bar{\psi}_1 q_1 + \bar{\psi}_2)\right], \quad (6.5)$$

[27]Taking $z = 0$ in (6.3) gives $\Pr(\hat{\lambda}_{\mathrm{ML}} \le 0) = \Pr\left(|\xi| \le (1-\lambda)(1+\lambda)\right)$, where ξ has a Cauchy distribution. Note that this very simple formula for the probability that $\hat{\lambda}_{\mathrm{ML}}$ is negative does not depend on the sample size.

where $q_1 \equiv \chi^2_1$. Thus, as in the case when $m \to \infty$ in a balanced Group Interaction model, $\hat{\lambda}_{\mathrm{ML}}$ is not consistent, but converges in distribution to a random variable as $n \to \infty$. The limiting pdf can be obtained from (6.5), but is omitted for brevity.

The density (6.4) is plotted in Figure 2 for $\lambda = -0.5, 0, 0.5$, for $n = 5, 10$, and for $n \to \infty$. 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 $\hat{\lambda}_{\mathrm{ML}}$ is unbounded at $z = 0$.

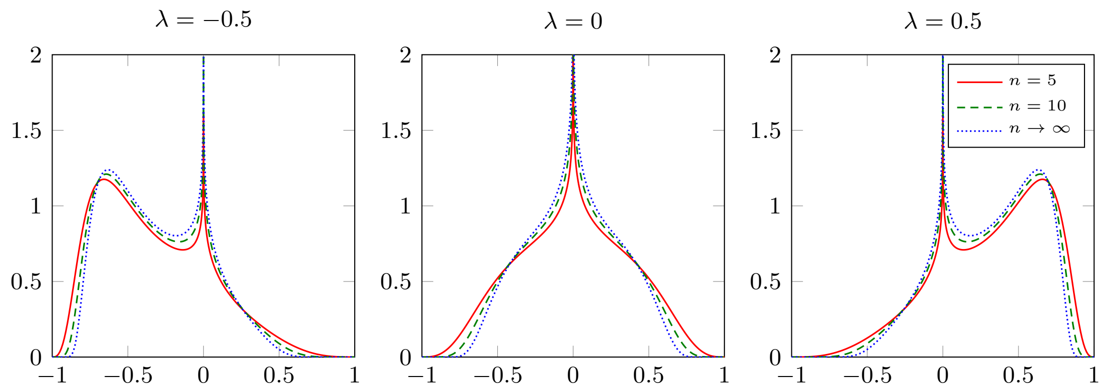

Figure 2: Density of $\hat{\lambda}_{\mathrm{ML}}$ for the Gaussian pure symmetric Complete Bipartite model.

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

6.2.2 Row-Standardized W, Constant Mean

As already anticipated in the discussion of Proposition 4.4, the support of $\hat{\lambda}_{\mathrm{ML}}$ in the row-standardized Complete Bipartite model with constant mean is not the entire interval $\Lambda = (-1, 1)$, but the subset $(-1, 0)$ (regardless of whether the true value of λ is positive or negative).

Proposition 6.6. *For the row-standardized Complete Bipartite model with $X = \iota_n$ and $\varepsilon \sim \mathrm{SMN}(0, I_n)$,*

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

where

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

Differentiating the cdf we obtain the following expression for the density.

Proposition 6.7. *For the row-standardised Complete Bipartite model with* $\varepsilon \sim \mathrm{SMN}(0, I_n)$*, and with* $X = \iota_n$*,*

$$\mathrm{pdf}_{\hat{\lambda}_{\mathrm{ML}}}(z;\lambda) = \frac{1}{B\left(\frac{1}{2}, \frac{n-2}{2}\right)} \frac{\dot{g}(z;\lambda)}{g(z;\lambda)^{\frac{1}{2}}\,[1-g(z;\lambda)]^{\frac{n-1}{2}}}, \tag{6.6}$$

for $z \in (-1, 0)$*. For* $z \in (0,1)$*,* $\mathrm{pdf}_{\hat{\lambda}_{\mathrm{ML}}}(z;\lambda) = 0$*.*

The limiting cdf and pdf as $n \to \infty$ can be obtained immediately from the results above. Letting

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

we obtain that, as $n \to \infty$, and for $z \in (-1, 0)$,

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \le z) \to \Pr\left(\chi_1^2 > h(z;\lambda)\right),$$

and

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

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

The density (6.6) is plotted in Figure 3 for $\lambda = -0.5, 0, 0.5$, for $n = 5, 10$, and for $n \to \infty$. Note that the shape of the density for $z < 0$ is similar to the case of the pure symmetric Complete Bipartite model (Figure 2).

7 The Single-Peaked Property Generally

The exact expression for the cdf of $\hat{\lambda}_{\mathrm{ML}}$ given in Theorem 1 depends only upon the fact that the profile log-likelihood $l_p(\lambda)$ is a.s. single-peaked on Λ, which was established in Lemma 3.6 under the condition that all eigenvalues of W are real. That condition makes the single-peaked property easy to prove, but it is certainly

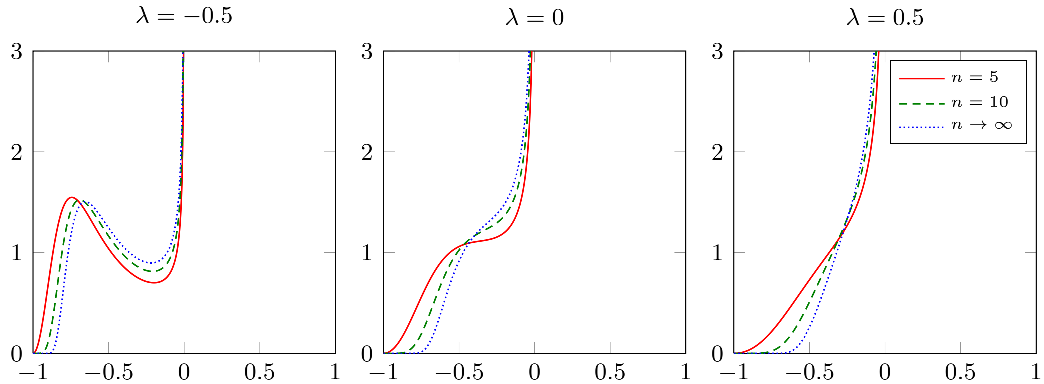

Figure 3: Density of $\hat{\lambda}_{\mathrm{ML}}$ for the Gaussian row-standardized Complete Bipartite model with constant mean.

not necessary. It is desirable to investigate the issue of single/multi-peakedeness of the log-likelihood further. Let

$$\delta(\lambda) := [\mathrm{tr}(G_\lambda)]^2 - n\mathrm{tr}(G_\lambda^2).$$

The proof of Lemma 3.6 shows that whenever W has the property that $\delta(\lambda) < 0$ for all $\lambda \in \Lambda$, every critical point of $l_p(\lambda)$ is a point of local maximum, implying that $l_p(\lambda)$ is again a.s. single-peaked on Λ. Thus, we have the following more general version of Theorem 1.

Theorem 4. *For any W such that $\delta(\lambda) < 0$ for all $\lambda \in \Lambda$, the* cdf *of $\hat{\lambda}_{\mathrm{ML}}$ is as given in Theorem 1.*

Theorem 4 generalizes Theorem 1 to cases in which some eigenvalues of W may be complex. It seems difficult to characterize the class of matrices W for which $\delta(\lambda) < 0$ for all $\lambda \in \Lambda$, but, for any given W, it is straightforward to check graphically whether the condition $\delta(\lambda) < 0$ holds for all $\lambda \in \Lambda$. Note that the condition depends only on W, not on X. The following example provides some evidence that the condition $\delta(\lambda) < 0$ for all $\lambda \in \Lambda$ is considerably more general than requiring real eigenvalues.

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

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

for fixed a_1, a_2, a_3, a_4. If $a_1 = a_3$ and $a_2 = a_4$, all the eigenvalues of W are real and therefore $l_p(\lambda)$ is a.s. single-peaked by Lemma 3.6. Other configurations of the a_i can induce multi-peakedeness 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 any value of a_4 larger than about 0.55, $\delta(\lambda) < 0$ for all $\lambda \in \Lambda$, so, even though not all eigenvalues of W are real, $l_p(\lambda)$ is a.s. single-peaked by Theorem 4. For smaller values of a_4 $\delta(\lambda)$ is not negative for all $\lambda \in \Lambda$, and there is a positive probability that $l_p(\lambda)$ is multi-peaked. Figure 7 displays $\delta(\lambda)$ 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-peakedeness as y ranges over $\mathbb{R}^n$.

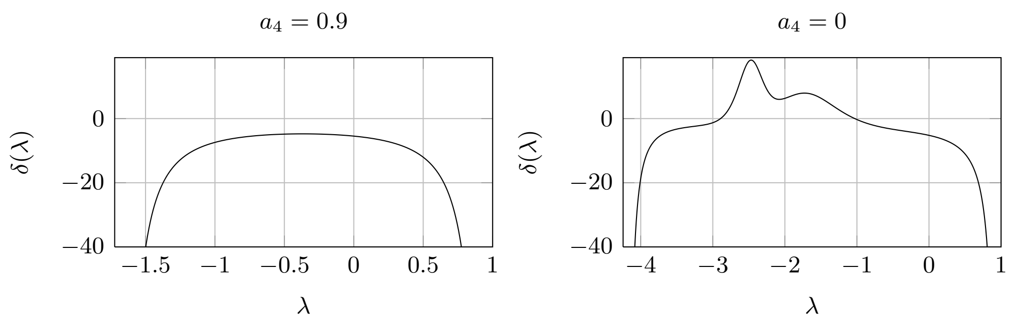

Figure 4: $\delta(\lambda)$, $\lambda \in \Lambda$, for the weights matrix W in Example 3.

A complete understanding of the cases in which the single-peaked property fails to hold is beyond the scope of this paper, but the next result is a first step in that direction. It says multi-peakedeness must always involve peaks at negative values of λ, for any W and X.

Proposition 7.1. *$l_p(\lambda)$ has at most one maximum in the interval* $[0, 1)$.

8 Discussion

The main result in this paper - Theorem 1 - provides a starting point for an examination of the properties of the maximum likelihood estimator for the spatial autoregressive parameter λ. Whatever the matrices W and X involved in a SAR model, and whatever the distribution assumptions entertained for ε, Theorem 1 provides a simple basis for simulation study of the properties of $\hat{\lambda}_{\mathrm{ML}}$. The result is also a useful starting point for the study of the higher-order asymptotic properties of $\hat{\lambda}_{\mathrm{ML}}$, a subject not embarked upon here. Finally, we have seen that in reasonably simple models with a high degree of structure (when W has only a few distinct eigenvalues), it can provide both exact results directly useful for inference, and new asymptotic results for cases not covered by the known results in Lee (2004). The present paper is just a beginning.

The study of quadratic forms of the type involved in Theorem 1 was begun by John von Neumann and Tjalling Koopmans in the 1940's when studying the distribution of serial correlation coefficients. The papers by von Neumann (1941) and Koopmans (1942) both discuss the unusual aspects of the distribution of serial correlation coefficients. Interestingly, the results in this paper show that the distributional properties of the MLE in spatial autoregressive models have closely related characteristics, at least in the Gaussian pure SAR case, a result that perhaps might have been anticipated but was, a priori, certainly not obvious. Two aspects of our results for this model did not occur in that earlier work: the possibility that the MLE might, with probability one, not exist, and the possibility that the support of the estimator might not be the entire parameter space. These are subjects that clearly demand further work.

Appendix A Auxiliary Results

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

(i) For any $z \in \Lambda$, the distinct eigenvalues $\gamma_1(z), \gamma_2(z), ..., \gamma_T(z)$ 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 $t = 2, ..., T-1$, $\gamma_t(z)$ changes sign exactly once on Λ.

(ii) For $T \geq 2$, $d_{11} < 0$ and $d_{TT} > 0$ for all $z \in \Lambda$. If $T > 2$, the coefficients d_{tt}, $t = 2, ..., T-1$, each change sign exactly once on Λ, with $d_{tt} > 0$ if $z < z_t$, $d_{tt} < 0$ if $z > z_t$, where z_t denotes the unique value of $z \in \Lambda$ at which $\gamma_t(z) = 0$.

Proof of Proposition A.1. (i) Let $\gamma_{1t}(z) := \omega_t/(1 - z\omega_t)$, for any $t = 1, ..., T$. Obviously, $\omega_s > \omega_t$ implies $\gamma_{1s}(z) > \gamma_{1t}(z)$ for all $z \in \Lambda$, which in turn implies $\gamma_s(z) > \gamma_t(z)$. If $\omega_t = 0$, $\gamma_{1t}(z) = 0$ for all $z \in \Lambda$. For the non-zero eigenvalues, since $d\gamma_{1t}(z)/dz = \gamma_{1t}^2(z) > 0$, each of these functions is strictly increasing on Λ. The function $\gamma_{11}(z) = \omega_{\min}/(1 - z\omega_{\min}) \to -\infty$ as $z \downarrow \omega_{\min}^{-1}$, and is bounded ($= \omega_{\min}/(1 - \omega_{\min})$) at $z = 1$. Likewise, the function $\gamma_{1T}(z) = 1/(1-z)$ is bounded at $z = \omega_{\min}^{-1}$ ($= \omega_{\min}/(\omega_{\min} - 1)$) and $\gamma_{1T}(z) \to +\infty$ as $z \uparrow 1$. The remaining functions $\gamma_{1t}(z)$ are all bounded at both endpoints of the interval Λ. The average of the γ_{1t} is

$$\frac{1}{n}\mathrm{tr}(G_z) = \frac{1}{n}\sum_{t=1}^{T} \frac{n_t\omega_t}{1 - z\omega_t} = \sum_{t=1}^{T} \alpha_t \gamma_{1t}(z)$$

(with $\alpha_t := n_t/n$). Since this is a convex combination of the $\gamma_{1t}(z)$, it is between the smallest and largest of them, for all $z \in \Lambda$, i.e.,

$$\gamma_{11}(z) < \frac{1}{n}\mathrm{tr}(G_z) < \gamma_{1T}(z),$$

or $\gamma_1(z) < 0 < \gamma_T(z)$ for all $z \in \Lambda$, so these two functions do not change sign on Λ. Next, the properties of the γ_{1t} imply that $\mathrm{tr}(G_z)/n$ is monotonic increasing on Λ, going to $-\infty$ as $z \downarrow \omega_{\min}$, and to $+\infty$ as $z \uparrow 1$. It follows that $\mathrm{tr}(G_z)/n$ crosses all $T-2$ of the functions $\gamma_{1t}(z)$, $t \neq 1, T$, at least once, somewhere in Λ. To show that the two functions can only cross once, simply observe that, at a point z where $\gamma_t(z) = 0$,

$$\dot{\gamma}_{1t}(z) = \gamma_{1t}^2(z) = \left(\sum_{t=1}^{T} \alpha_t \gamma_{1t}(z)\right)^2 < \sum_{t=1}^{T} \alpha_t \gamma_{1t}^2(z) = \frac{d}{dz}\left(\frac{1}{n}\mathrm{tr}(G_z)\right).$$

(the inequality is strict because the $\gamma_{1t}(z)$ cannot all be equal). That is, at every point of intersection, $\mathrm{tr}(G_z)/n$ intersects $\gamma_{1t}(z)$ 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} are those of the γ_t. □

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 $\lambda_1 M_X y = \lambda_2 M_X y$. If $\lambda_1 \neq \lambda_2$, then $M_X y = 0$, and hence $M_X W y = 0$, which in turn implies that $M_X S_\lambda y = 0$ for all $\lambda \in \mathbb{R}$. □

Details for Section 4.3. Using the assumption $W = HDH^{-1}$ we find that $C_z = HD_1H^{-1}$, and $S_z S_\lambda^{-1} = HD_2H^{-1}$, with

$$D_1 := \mathrm{diag}\left(\gamma_t(z) I_{n_t}, t = 1, ..., T\right),$$

and

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

We can now write the matrix of the quadratic form in (4.5) as

$$A(z, \lambda) = (H')^{-1} D_2 \left(D_1 M + M D_1\right) D_2 H^{-1}. \tag{A.1}$$

Next, let $M = (M_{st}; s, t = 1, ..., T)$ be the partition of M conformable with D_1 and D_2, so that the blocks $M_{st} = (M_{ts})'$ are of dimension $n_s \times n_t$. We have

$$D_2 \left(D_1 M + M D_1\right) D_2 = (d_{st} M_{st}; s, t = 1, ..., T),$$

where the coefficients d_{st} are as defined in the text.

Appendix B Proofs

Proof of Lemma 3.1. Suppose first that, for some non-zero eigenvalue ω of W, $M_X(\omega I_n - W) \neq 0$. Then $M_X(\omega I_n - W)y$ is a.s. nonzero. It follows that the term $-(n/2)\ln(y'S_\lambda' M_X S_\lambda y)$ in equation (3.2) is a.s. continuous at $\lambda = \omega^{-1}$, because it is a.s. defined at $\lambda = \omega^{-1}$, and, by Lemma A.2, cannot, again a.s., be undefined at more than one value of $\lambda \neq \omega^{-1}$. The other term in equation (3.2), $\ln(|\det(S_\lambda)|)$, goes to $-\infty$ as $\lambda \to \omega^{-1}$. Hence $\lim_{\lambda\to\omega^{-1}} l_p(\lambda) = -\infty$ a.s. Let us now move to the case when, for some real non-zero eigenvalue ω of W, $M_X(\omega I_n - W) = 0$. The profile log-likelihood is a.s. defined by equation (3.4). Letting $n_\varkappa$ denote the algebraic multiplicity of an eigenvalue $\varkappa$, and $\mathrm{Sp}(W)$ the spectrum of W (defined as the set of distinct eigenvalues), we obtain

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

The first term in equation (B.1) is a.s. bounded as $\lambda \to \omega^{-1}$. The second term goes to $+\infty$ as $\lambda \to \omega^{-1}$, because $n_\omega < n$ (since $W \neq I_n$ by the assumption that $\mathrm{tr}(W) = 0$). Thus, $\lim_{\lambda\to\omega^{-1}} l_p(\lambda) = +\infty$ a.s. □

Proof of Lemma 3.5. Let ω_t, $t = 1, ..., T$, 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

$$\omega_j e_{t-1,j} = e_t - e_{t,j}, \qquad \text{(B.2)}$$

for $t = 1, ..., T-1$, and

$$\omega_j e_{T-1,j} = e_T. \tag{B.3}$$

We can therefore write the first-order condition (see equation (3.5) as

$$n\,(b - a\lambda)\sum_{t=0}^{T}(-\lambda)^t e_t - \left(a\lambda^2 - 2b\lambda + c\right)\sum_{j=1}^{T}\left(n_j\omega_j \sum_{\substack{t=0\\ t\neq j}}^{T-1}(-\lambda)^t e_{t,j}\right) = 0, \tag{B.4}$$

where $a := y'W'M_XWy$, $b := y'W'M_Xy$, and $c := y'M_Xy$. We now show that the polynomial equation (B.4) has degree T. Using (B.3) 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.2), 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. □

Proof of Lemma 3.6. Recall that we are assuming that $M_X(\omega I_n - W) \neq 0$ for any real nonzero eigenvalue ω of W. Hence, by Lemma 3.1, $l_p(\lambda) \to -\infty$ 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. Because 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

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

where

$$a := y'W'M_XWy,\ b := y'W'M_Xy,\ c := y'M_Xy.$$

But at any point where $\dot{l}_p(\lambda) = 0$,

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

so that, at any critical point,

$$\ddot{l}_p(\lambda) = \left\{\frac{-n(ac - b^2)}{(a\lambda^2 - 2b\lambda + c)^2}\right\} + \frac{1}{n}\left\{[\mathrm{tr}(G_\lambda)]^2 - n\mathrm{tr}(G_\lambda^2)\right\}. \tag{B.5}$$

By the Cauchy-Schwarz inequality the first term on the right hand side of (B.5) is nonpositive. When the eigenvalues of W are real, the second term in curly brackets is also nonpositive, again by the Cauchy-Schwarz inequality, and cannot be zero because G_λ cannot be a scalar multiple of I_n. That is, at every point where $\dot{l}_p(\lambda)$ vanishes, $\ddot{l}_p(\lambda) < 0$. Thus, $l_p(\lambda)$ has a.s. exactly one point of maximum in Λ, and no other stationary points. □

Proof of Proposition 3.8. For simplicity, assume that all densities exist. We need to show that the distribution of the maximal invariant $v := y(y'y)^{-1/2} \in \mathcal{S}^{n-1}$ 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 $y \to (q, v)$, 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.[28] □

Proof of Proposition 3.9. (i) Because of the presence of the scale parameter σ, the SAR model (1.1) is invariant with respect to the scale transformations $y \to \kappa y$, $\kappa > 0$. If the distribution of ε does not depend on β or σ^2, the transformation $y \to \kappa y$ induces the transformations $(\beta, \lambda, \sigma^2, \theta) \to (\kappa\beta, \lambda, \kappa^2\sigma^2, \theta)$ 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 depends 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).

(ii) Suppose that the distribution of ε does not depend on β or σ^2, and that $\text{col}(X)$ is an invariant subspace of W. Then the SAR model (1.1) is invariant under the group $\mathcal{G}_X$ of transformations $y \to \kappa y + X\delta$, 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 $WX = XA$, which in turn is equivalent to $S_\lambda^{-1}X = (I_k - \lambda A)^{-1}X$, for any λ such that S_λ is invertible. The group, say $\overline{\mathcal{G}}_X$, induced by $\mathcal{G}_X$ on the parameter space is that of the transformations $(\beta, \lambda, \sigma^2, \theta) \to (\kappa\beta + (I_k - \lambda A)\delta, \lambda, \kappa^2\sigma^2, \theta)$. Now, it is easy to see from the profile score equation (3.5) that (under the conditions stated above) $\hat{\lambda}_{\text{ML}}$ is invariant under $\mathcal{G}_X$. Since $\overline{\mathcal{G}}_X$ 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). □

Proof of Proposition 4.2. (i) Proved in the text. (ii) 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 $M = H'M_XH$ 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 $M_XW = W'M_X$. (iii) If W is symmetric, H is orthogonal, and hence $M_{ij} = h_i'M_Xh_j$, 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_XW is also symmetric $\text{col}(X)$ is spanned by k linearly independent eigenvectors of W. Hence, for each $j = 1, ..., n$, M_Xh_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. □

Proof of Theorem 2. The assumption $\varepsilon \sim \text{SMN}(0, I_n)$ gives $\tilde{y} \sim \text{SMN}(0, I_n)$ and $x \sim \text{SMN}(0, (H'H)^{-1})$. But $H'H = I_n$ if W is symmetric, and hence the stated

[28] For a formal treatment of the argument used to establish Proposition 3.8 - averaging over the group - see also Eaton (1989), particularly Chapters 4 and 5.

result follows from (4.11). □

Proof of Proposition 4.4. Under the first stated condition all diagonal blocks M_{ss} for $s > t$ vanish, and for $s \leq t$ the $\gamma_s(z)$ are all negative for $z > z_t$, so that, by Proposition 4.2 (iii), $\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z) = 1$ for $z \geq z_t$. In the second case the $\gamma_s(z)$ are all positive for $z < z_t$ by Proposition A.1, so that $\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z) = 0$ for $z \leq z_t$. □

Proof of Corollary 5.2. We prove this result under Gaussian assumptions. For notational convenience, let us rewrite expression (5.1) as

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z; \lambda) = \Pr\left(\sum_{s=1}^{S} d_{\omega_s}(\lambda, z)\, \chi^2_{n_s} \leq 0\right).$$

Since $d_{\omega_s}(\lambda, z) = -d_{-\omega_s}(-\lambda, -z)$, we have

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z; \lambda) = \Pr\left(\sum_{s=1}^{S} -d_{-\omega_s}(-\lambda, -z)\, \chi^2_{n_s} \leq 0\right),$$

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

Proof of Corollary 5.3. We first note 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 $Q := \sum_{i=1}^{S} a_i \chi^2_{n_i}$ is a linear combination of independent $\chi^2_{n_i}$ random variables with positive coefficients a_i, the density of Q is given by

$$\mathrm{pdf}_Q(q; A) = \frac{\exp\left(-\frac{1}{2}\tau q\right) q^{\frac{n}{2}-1}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})\,(\det(A))^{\frac{1}{2}}}\, {}_1F_1\left(\frac{1}{2}; \frac{n}{2}; \frac{q}{2}A^*\right) \tag{B.6}$$

where $n = \sum_{i=1}^{S} n_i$, $A := \mathrm{diag}(a_i I_{n_i}, i = 1, ..., S)$, $\tau := \mathrm{tr}(A^{-1})$, 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 (R, Q_2), and integrating out the redundant variable 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. □

Proof of Proposition 6.1. 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, $\mathrm{tr}(G_z)/n = (rm)^{-1}\left[r/(1-z) - r(m-1)/(z+m-1)\right] =$

$z/[(1-z)(z+m-1)]$, and hence $d_{11} = 2(m-1)(1-z)/[(1-\lambda)^2(z+m-1)]$ and $d_{22} = -2(z+m-1)/[(\lambda+m-1)^2(1-z)]$. The stated result follows from equation (5.3). □

Proof of Proposition 6.2. By equation (6.1),

$$\mathrm{pdf}_{\hat{\lambda}_{\mathrm{ML}}}(z;\lambda) = \frac{\partial c(z,\lambda)}{\partial z}\mathrm{pdf}_{\mathrm{F}_{r,r(m-1)}}(c(z,\lambda)),$$

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

Proof of Proposition 6.4. For a symmetric Complete Bipartite model $\mathrm{tr}(WS_z^{-1}) = -1/(1+z)+1/(1-z) = 2z/\left(1-z^2\right)$, and hence $\gamma_1(z) = -[n-(n-2)z]/[n\left(1-z^2\right)]$, $\gamma_2(z) = -2z/[n\left(1-z^2\right)]$, and $\gamma_3(z) = [n+(n-2)z]/[n\left(1-z^2\right)]$. The stated result follows by using expression (5.1). □

Proof of Proposition 6.5. For $z \in (0,1)$ the density is obtained as an application of the result 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 on using Corollary 5.2. □

Proof of Proposition 6.6. 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 $\mathrm{E}(x) = \beta\sqrt{n}(n,0,0)'$. Therefore, from equation (4.9), we have

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z) = \Pr\left(d_{22}\chi_1^2 + d_{33}\chi_{n-2}^2 \leq 0\right),$$

i.e.,

$$\Pr\left(-2\left(\phi_2\chi_1^2 + 2z\chi_{n-2}^2\right) \leq 0\right).$$

But, if $z \geq 0$, both coefficients here are non-negative, so for $z \geq 0$, $\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z) = 1$. This yields the result stated. □

Proof of Proposition 7.1. When $\lambda \in [0,1)$, G_λ can be expanded as $\sum_{r=0}^{\infty} \lambda^r W^{r+1}$, showing that nonnegativity of W implies nonnegativity of G_λ. Letting g_{ij} denote (i,j)-th entry of G_λ, we have

$$n\text{tr}\left(G_\lambda^2\right) = n \sum_{i,j=1}^{n} g_{ij} g_{ji} = n \sum_{i=1}^{n} g_{ii}^2 + n \sum_{i,j=1, i \neq j}^{n} g_{ij} g_{ji}.$$

By the sum of squares inequality $n \sum_{i=1}^{n} g_{ii}^2 \geq \left(\sum_{i=1}^{n} g_{ii}\right)^2 = [\text{tr}(G_\lambda)]^2$. Also, note that $n \sum_{i,j=1, i\neq j}^{n} g_{ij} g_{ji} > 0$ when $\lambda \in [0,1)$, by the nonnegativity of G_λ. It follows that, for any $\lambda \in [0,1)$, $n\text{tr}\left(G_\lambda^2\right) > [\text{tr}(G_\lambda)]^2$ or, equivalently, $\delta(\lambda) < 0$. As stated in the proof of Lemma 3.6, the first term on the right hand side of expression (B.5) is nonpositive by the Cauchy-Schwarz inequality. The second term is equal to $\delta(\lambda)$. Hence all stationary points of $l_p(\lambda)$ in $[0,1)$ must be maxima, which completes the proof. □

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.

Baltagi, B.H. (2006) Random effects and spatial autocorrelation with equal weights, *Econometric Theory* 22, 973-984.

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.

Cliff, A.D. and Ord, J.K. (1973) *Spatial autocorrelation.* London, Pion

Cressie, N. (1993) *Statistics for spatial data.* Wiley,

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 Satistic with Applications, *Journal of Econometrics* 104, 219–257.

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

Kelejian, H.H., Prucha, I.R. (2010) Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157, 53-67.

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. (2014) Criminal networks: who is the key player?, 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.

Ross-Parker, H. (1975) Inter-plant competition models and their sampling distributions, *Advances in Applied Probability* 7, 453-4.

Rothenberg, T.J. (1971) Identification in parametric models, *Econometrica* 39, 577-591.

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. (2014) A general method for third-order bias and variance corrections on a nonlinear estimator, manuscript, *Journal of Econometrics*, forthcoming.

von Neumann, J. (1941) Distribution of the ratio of the mean square successive difference to the variance, *Annals of Mathematical Statistics* 12, 367–395.