Discussion Papers in Economics

EXACT LIKELIHOOD INFERENCE IN GROUP INTERACTION NETWORK MODELS

By

Grant Hillier,
(CeMMAP and University of Southampton)

&

Federico Martellosio
(University of Surrey)

DP 08/16

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

ISSN: 1749-5075

Exact Likelihood Inference in Group Interaction Network Models

Grant Hillier
CeMMAP and Department of Economics, University of Southampton,
Highfield, Southampton SO17 1BJ, UK
ghh@soton.ac.uk

Federico Martellosio
School of Economics, University of Surrey,
Guildford, Surrey, GU2 7XH, UK
f.martellosio@surrey.ac.uk

April 4 2016

Abstract: The paper studies spatial autoregressive models with group interaction structure, focussing on estimation and inference for the spatial autoregressive parameter λ. The quasi-maximum likelihood estimator for λ usually cannot be written in closed form, but using an exact result obtained earlier by the authors for its distribution function, we are able to provide a complete analysis of the properties of the estimator, and exact inference that can be based on it, in models that are balanced. This is presented first for the so-called pure model, with no regression component, but is also extended to some special cases of the more general model. We then study the much more difficult case of unbalanced models, giving analogues of some, but by no means all, of the results obtained for the balanced case earlier. In both balanced and unbalanced models, results obtained for the pure model generalize immediately to the model with group-specific regression components.

1 Introduction

One important application of spatial autoregressive (SAR) models is to the analysis of social networks, particularly for the case when an outcome variable is observed on a predetermined network; see, for instance Bramoullé, Djebbari and Fortin (2009), Lee, Liu and Lin (2010), and de Paula (2016).[1] Consider a fixed network of n individuals, represented by a $n \times n$ weights matrix W. The matrix W could be a $(0, 1)$ adjacency matrix, a row-standardized adjacency matrix, or could more generally be specified in such a way that the general entry $W_{i,j}$ is a measure of the strength of interaction between individuals i and j. A popular specification of a SAR model for the determination of an $n \times 1$ outcome vector y, given the network and an $n \times k$ matrix X of covariates, is

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

[1] For extensions of SAR models that allow for endogenous network formation, see, e.g., Hsieh and Lee (2016).

where λ is a scalar autoregressive parameter, β and δ are $k \times 1$ parameters, σ is a scale parameter, and ε an $n \times 1$ error term. In the peer effects literature, λ captures the endogenous effect, and δ the exogenous effect; see Manski (1993). In addition to social networks, model (1.1) has been applied to several other cross-sectional contexts. Also, when W is block-diagonal, model (1.1) can be seen as a panel data model with cross-sectional dependence - see for instance the recent paper by Robinson and Rossi (2015), and references therein.[2]

A fundamental, at least conceptually, specification for the matrix W in the social network literature is given by the equal weights matrix $B_n := (n-1)^{-1}(\iota_n \iota_n' - I_n)$, where ι_n denotes the $n \times 1$ vector of ones. In that case, model (1.1) postulates that the outcome variable for individual i is explained by the "leave-own-out" mean $(n-1)^{-1}\sum_{j \neq i} y_j$, the regressors, and the leave-own-out means of the regressors; see, e.g., Moffitt (2001). The weights matrix B_n may be appropriate when all individuals are equally affected by all other individuals, or when no information on how individuals interact is available.

A more general assumption is that individuals interact in groups, with each group member being equally affected by all the other members in that group, and with no links across groups. This results in W having a block diagonal structure, with equal weights matrices as blocks. More precisely, letting m_i be the distinct group sizes, for $i = 1, ..., p$, and r_i the number of groups of size m_i, for $i = 1, ..., p$, the (row-standardized) group interaction weights matrix is

$$W = \operatorname{diag}(I_{r_i} \otimes B_{m_i}, i = 1, .., p). \tag{1.2}$$

Such matrices were used, for example, in Case (1992), Kelejian, Prucha, and Yuzefovich (2006), and Lee (2007), and is the structure we shall consider in this paper.

We focus on inference on λ, which is often the key parameter in applications, and, for simplicity (but without loss of generality), take $\delta = 0$ in (1.1). We call a model

$$y = \lambda W y + X\beta + \sigma\varepsilon, \tag{1.3}$$

with weights matrix (1.2) a *Group Interaction* model. If the group sizes are all equal (i.e., $p = 1$) the Group Interaction model is said to be *balanced*, otherwise, when $p > 1$, it is *unbalanced.* We assume throughout that $m_i \geq 2$ for all i. In the balanced case W consists of $r := \sum_{i=1}^{p} r_i$ copies of B_m, so, letting m be the common group size,

$$W = I_r \otimes B_m. \tag{1.4}$$

The sample size is thus $n = \sum_{i=1}^{r} r_i m_i$, in general, and $n = rm$ in the balanced model. If $\beta = 0$ in equation (1.3) we call this a *pure* model.

The class of Group Interaction models was discussed briefly in Hillier and Martellosio (2013) (hereafter H&M), and some exact results given for the pure balanced case. After

[2] A special case of the model in Robinson and Rossi (2015) is discussed in Section 3.6.1 below.

some preliminaries, given in the next section, in Section 3 we provide a complete analysis of the properties of $\hat{\lambda}_{\text{ML}}$, and of exact inference procedures based upon it, for the pure balanced model. Results for the balanced model are of interest for their own sake, but also because this model is often used to illustrate theoretical results in the literature (see Lee (2004), (2007), and Lee, Liu, and Lin (2010), for instance). However, the balanced model is certainly of limited practical importance, so in Section 4 we go on to discuss the unbalanced model. For reasons to be explained, results for this model are much more complex than those for the balanced model. Thus, although we do give some general results, we often confine ourselves to the case of just two group sizes ($p = 2$) for simplicity. Proofs of the main formal results are in Appendix A, and Appendix B contains some additional figures.

2 Preliminaries

For the present, let W be any matrix assumed to have at least one negative, and one positive, eigenvalue, and normalized to have largest eigenvalue unity. The parameter space for λ is taken to be the largest interval containing the origin within which the matrix $S_\lambda := I_n - \lambda W$ remains non-singular. Letting $\omega_{\min}$ denote the smallest real eigenvalue of W, the parameter space will thus be

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

We assume that the parameters are estimated by (quasi-) maximum likelihood (QML), where the likelihood adopted is that which would apply if, in equation (1.3), $\varepsilon \sim \mathrm{N}(0, I_n)$. We define the QMLE of λ (assuming it exists), $\hat{\lambda}_{\text{ML}}$, by

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

where $l_p(\lambda)$ is the profile (quasi) log-likelihood for λ after maximization with respect to (β, σ^2). This estimator is, in general, a zero of a high degree polynomial in λ, and thus cannot be written in closed form. However, it is shown in H&M that, if W has real eigenvalues - which will be the case in the present paper - the profile likelihood $l_p(\lambda)$ is single-peaked on Λ. This means that, for each $z \in \Lambda$, the event that $\hat{\lambda}_{\text{ML}} \leq z$ is identical to the event that the profile score at z, $\dot{l}_p(z)$, is negative. Thus, notwithstanding its unavailability in closed form, an exact expression for the distribution function (cdf) of $\hat{\lambda}_{\text{ML}}$ can be written down immediately:

$$\Pr(\hat{\lambda}_{\text{ML}} \leq z; \lambda) = \Pr(\dot{l}_p(z) \leq 0), \tag{2.1}$$

where, here and throughout, $\Pr(\hat{\lambda}_{\text{ML}} \leq z; \lambda)$ denotes the cdf of $\hat{\lambda}_{\text{ML}}$ at the point $z \in \Lambda$ when the true parameter value is $\lambda \in \Lambda$. This result is the basis for all of the results in this paper.

In addition to this single-peaked property, it also easy to see that $\dot{l}_p(z) \to -\infty$ as $z \to 1$ (from the left), and $\dot{l}_p(z) \to +\infty$ as $z \to \omega_{\min}^{-1}$ (from the right). Thus, $\Pr(\hat{\lambda}_{\text{ML}} \leq z; \lambda) =$

$\Pr(\dot{l}_p(z) \leq 0) \to 1$ as $z \to 1$, and $\Pr(\dot{l}_p(z) \leq 0) \to 0$ as $z \to \omega_{\min}^{-1}$. In other words, the inequality $\Pr(\dot{l}_p(z) \leq 0)$ does indeed define a distribution function supported on Λ, as one would expect. Note that this argument holds whatever the distribution of y, provided only that the random variable $\dot{l}_p(z)$ is supported on the entire interval Λ.

In the analytical results to follow we take the distribution of ε to be $\mathrm{N}(0, I_n)$ (that is, the likelihood is correctly specified), but, as discussed in H&M, all results obtained under this assumption continue to hold under scale mixtures of the $\mathrm{N}(0, I_n)$ distribution, the family we denote by $\mathrm{SMN}(0, I_n)$. For symmetric pure SAR models, equation (2.1) provides the following representation of the cdf of the MLE:[3]

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z; \lambda) = \Pr\left(\sum_{t=1}^{T} d_{tt}(z,\lambda)\chi^2_{n_t} \leq 0\right), \tag{2.2}$$

where the $\chi^2_{n_t}$ variates are independent. Here, n_t is the algebraic multiplicity of the eigenvalue ω_t of W, T denotes the number of distinct eigenvalues of W, and the coefficient functions $d_{tt}(z, \lambda)$ are given by

$$d_{tt}(z,\lambda) := 2\left(\frac{1 - z\omega_t}{1 - \lambda\omega_t}\right)^2 (g_t(z) - \bar{g}(z)). \tag{2.3}$$

Here,

$$g_t(z) := \frac{\omega_t}{1 - z\omega_t}, \tag{2.4}$$

for $t = 1, ..., T$, are the distinct eigenvalues of $G_z := WS_z^{-1}$, where $S_z := I_n - zW$, while $\bar{g}(z) := (1/n)\sum_{t=1}^{T} n_t g_t(z) = (1/n)\mathrm{tr}(G_z)$ is the average of all eigenvalues of G_z. In what follows we use the notation that, for any matrix A of full column rank, $P_A := A(A'A)^{-1}A'$, and $M_A := I - P_A$. Also, $\mathrm{col}(A)$ denotes the column space of a matrix A. All matrices are assumed to be real.

3 The Balanced Model

In this section we first of all provide a complete analysis of the exact properties of $\hat{\lambda}_{\mathrm{ML}}$, and inference procedures based upon it, for the pure balanced model. Then, we consider some generalizations of these results to balanced models with regressors: we show that, for certain special choices of X, the results obtained for the pure model apply with only minor modifications. We note that in the pure balanced Group Interaction model, because the profile score is a quadratic in λ, $\hat{\lambda}_{\mathrm{ML}}$ is in fact available in closed form. However, its distribution theory is most easily obtained by using equation (2.2), and this also leads naturally to generalizations to the unbalanced model, when the estimator is typically not available in closed form.

[3] If normality is not assumed equation (2.2) involves T quadratic forms in n_t-dimensional vectors; see H&M.

3.1 Distribution Function and Density

For the pure balanced model we have $T = 2$, $n_1 = r(m-1)$, $n_2 = r$, $\Lambda = (-(m-1), 1)$, and the coefficients in equation (2.2) are given by

$$d_{11} = -2\left(\frac{z+m-1}{\lambda+m-1}\right)^2 \frac{1}{(\lambda+m-1)(1-\lambda)} < 0,$$
$$d_{22} = 2\left(\frac{1-z}{1-\lambda}\right)^2 \frac{(m-1)}{(\lambda+m-1)(1-\lambda)} > 0.$$

Eliminating irrelevant scalars in (2.2), we obtain

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \le z; \lambda) = \Pr\left((m-1)\chi_r^2 \le c(z,\lambda)\chi^2_{r(m-1)}\right),$$

where

$$c(z,\lambda) := \left(\frac{(1-\lambda)(z+m-1)}{(1-z)(\lambda+m-1)}\right)^2. \tag{3.1}$$

Thus, as stated in H&M, in the pure balanced Group Interaction model with $\varepsilon \sim \mathrm{SMN}(0, I_n)$, the cdf of $\hat{\lambda}_{\mathrm{ML}}$ is, for any $z, \lambda \in \Lambda$,

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \le z; \lambda) = \Pr(\mathrm{F}_{r,r(m-1)} \le c(z,\lambda)), \tag{3.2}$$

where F_{ν_1,ν_2} denotes a random variable distributed as an F distribution with ν_1 and ν_2 degrees of freedom. The corresponding density function is

$$\mathrm{pdf}_{\hat{\lambda}_{\mathrm{ML}}}(z;\lambda) = \frac{2m\tau^{r(m-1)}}{B\left(\frac{r}{2}, \frac{r(m-1)}{2}\right)} \frac{(1-z)^{r(m-1)-1}(z+m-1)^{r-1}}{\left(\tau^2(1-z)^2 + (z+m-1)^2\right)^{\frac{rm}{2}}}, \tag{3.3}$$

where $\tau := \theta(\lambda)\sqrt{m-1}$, with

$$\theta(\lambda) = \theta := \frac{\lambda+m-1}{1-\lambda} > 0, \tag{3.4}$$

Note that $c(z,\lambda) = (\theta(z)/\theta(\lambda))^2$, and that $c(z,\lambda)$ is monotonic increasing in z. In fact, $c(z,\lambda) \to \infty$ as $z \to 1$, while $c(z,\lambda) \to 0$ as $z \to -(m-1)$. Hence, as noted in the comments following equation (2.1), equations (3.2) and (3.3) define a cdf and pdf supported on Λ. In addition, $\Pr(\hat{\lambda}_{\mathrm{ML}} \le z; \lambda) \to 0$ for all $z \in \Lambda$ as $\lambda \to 1$, because $c(z,\lambda) \to 0$, and $\Pr(\hat{\lambda}_{\mathrm{ML}} \le z; \lambda) \to 1$ for all $z \in \Lambda$ as $\lambda \to -(m-1)$, because $c(z,\lambda) \to \infty$. That is, the distribution of $\hat{\lambda}_{\mathrm{ML}}$ becomes degenerate, i.e., $\mathrm{var}(\hat{\lambda}_{\mathrm{ML}}) \to 0$, as λ approaches either endpoint of Λ.

Finally, observe that, since $c(\lambda,\lambda) = 1$, the probability that $\hat{\lambda}_{\mathrm{ML}}$ underestimates λ, $\Pr(\hat{\lambda}_{\mathrm{ML}} \le \lambda; \lambda)$, is given by $\Pr(\mathrm{F}_{r,r(m-1)} \le 1)$, which does not depend on λ. The fact that $\Pr(\hat{\lambda}_{\mathrm{ML}} \le \lambda; \lambda)$ does not converge to 1 as $\lambda \to 1$, as might have been anticipated, is a consequence of the degeneracy of the distribution of $\hat{\lambda}_{\mathrm{ML}}$ just discussed.

Remark 3.1. *Gaussian pure SAR models - equation (1.3) with the regression component deleted - are members of the 2-parameter exponential family, with parameters* (λ, σ^2), *sufficient statistics the three quadratic forms*

$$q_1 := y'y, \;\; q_2 := y'W'Wy, \;\; q_3 := y'(W + W')y,$$

and canonical parameters

$$\eta_1 := -\frac{1}{2\sigma^2}, \;\; \eta_2 := -\frac{\lambda^2}{2\sigma^2}, \;\; \eta_3 := \frac{\lambda}{2\sigma^2}.$$

Thus, pure SAR models are, in the notation of Barndorff-Nielsen (1980), at worst, $(3, 2)$-*curved exponential models. In the balanced model with* $W = I_r \otimes B_m$ *these three sufficient statistics are not minimal, and can be written in terms of just two statistics,*

$$s_1 := y'(I_r \otimes M_{\iota_m})y, \;\; s_2 := y'(I_r \otimes P_{\iota_m})y.$$

Specifically, $q_1 = s_1 + s_2, q_2 = s_1/(m-1)^2 + s_2$, *and* $q_3 = 2(s_2 - s_1/(m-1))$. *Collecting coefficients, the canonical parameters become*

$$\eta_1^* := -\frac{1}{2\sigma^2}\left(\frac{\lambda + m - 1}{m-1}\right)^2, \;\; \eta_2^* := -\frac{(1-\lambda)^2}{2\sigma^2}.$$

The pure balanced model is thus a regular exponential model, and it is this that makes it amenable to exact inference. We will see later that the unbalanced model cannot be reduced in this way, and so is genuinely curved. It can easily be checked that the two sufficient statistics s_1 *and* s_2 *are independent in the balanced model, and*

$$\frac{s_1(\lambda + m - 1)^2}{\sigma^2(m-1)^2} \sim \chi^2_{r(m-1)}, \;\; \frac{s_2(1-\lambda)^2}{\sigma^2} \sim \chi^2_r.$$

3.1.1 First Consequences

The function $c(z, \lambda)$, defined on $\Lambda \times \Lambda$, is strictly decreasing in λ and strictly increasing in z. The first fact means that the distribution functions for different values of λ *do not cross*, so $\lambda_1 < \lambda_2$ implies that the cdf for $\lambda = \lambda_1$ lies entirely above that for $\lambda = \lambda_2$. That is:

Property 1. *In a pure balanced Group Interaction model with* $\varepsilon \sim \mathrm{SMN}(0, I_n)$, $\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z; \lambda_1) > \Pr(\hat{\lambda}_{\mathrm{ML}} \leq z; \lambda_2)$, *for any* $\lambda_1, \lambda_2 \in \Lambda$ *such that* $\lambda_1 < \lambda_2$, *and for any* $z \in \Lambda$, *that is,* $\hat{\lambda}_{\mathrm{ML}}$ *when* $\lambda = \lambda_2$ *stochastically dominates* $\hat{\lambda}_{\mathrm{ML}}$ *when* $\lambda = \lambda_1$.

Since, in our present setup, the mean of $\hat{\lambda}_{\mathrm{ML}}$ is $-(m-1)$ plus the area above the cdf, Property 1 implies:

Property 2. *The mean of* $\hat{\lambda}_{\mathrm{ML}}$ *is a monotonic increasing function of* λ.

The second property of the function $c(z,\lambda)$ implies that $\Pr(\hat{\lambda}_{\mathrm{ML}} \le z; \lambda) = \Pr(c(\hat{\lambda}_{\mathrm{ML}}, \lambda) \le c(z,\lambda)) = \Pr(\mathrm{F}_{r,r(m-1)} \le c(z,\lambda))$, or that $c(\hat{\lambda}_{\mathrm{ML}}, \lambda) \sim \mathrm{F}_{r,r(m-1)}$. Thus, for the MLE of θ,

$$\hat{\theta}_{\mathrm{ML}} := \frac{\hat{\lambda}_{\mathrm{ML}} + m - 1}{1 - \hat{\lambda}_{\mathrm{ML}}}, \tag{3.5}$$

we have established:

Proposition 3.1. *In the pure balanced Group Interaction model with $\varepsilon \sim \mathrm{SMN}(0, I_n)$,*

$$\hat{\theta}_{\mathrm{ML}} \cong \theta \sqrt{\mathrm{F}_{r,r(m-1)}}.$$

Alternatively, we could write

$$\hat{\theta}_{\mathrm{ML}} \cong \tau \sqrt{f_{r,r(m-1)}},$$

where the random variable $f_{r,r(m-1)} := \chi^2_r / \chi^2_{r(m-1)}$ has density

$$\mathrm{pdf}_{f_{r,r(m-1)}}(f) = \frac{f^{\frac{r}{2}-1}(1+f)^{-\frac{rm}{2}}}{B\left(\frac{r}{2}, \frac{r(m-1)}{2}\right)}.$$

The parameter θ defined in (3.4) is a 1-1 function of λ, and it is clear from equation (3.2) that the properties of $\hat{\lambda}_{\mathrm{ML}}$ depend on λ only through θ. This key parameter can be interpreted as just another way of locating the point λ in the interval Λ, i.e., as a different parameterization of the model.

Remark 3.2. *The result in Proposition 3.1 provides a very efficient method of simulating any properties of $\hat{\lambda}_{\mathrm{ML}}$ (or functions of $\hat{\lambda}_{\mathrm{ML}}$) that are not available exactly, or are too complicated, by simply drawing samples from the $\mathrm{F}_{r,r(m-1)}$ distribution.*

Remark 3.3. *The parameter θ is closely related to the canonical parameters in the exponential family representation of the model, specifically, by $\theta^2 = (m-1)^2 \eta_1^* / \eta_2^*$.*

3.1.2 Asymptotics Under Mixed-Normality

In the case $r \to \infty$ with m fixed (*fixed-domain* asymptotics), the asymptotic distribution of $\hat{\lambda}_{\mathrm{ML}}$ is covered by the results in Lee (2004): $\hat{\lambda}_{\mathrm{ML}}$ is consistent and asymptotically normal as $r \to \infty$ with large-r variance (based on the information matrix, assuming normality) given by

$$v_\lambda := \frac{(1-\lambda)^2(\lambda+m-1)^2}{2rm(m-1)}. \tag{3.6}$$

Note that, as λ goes to either extreme of Λ, this exhibits the same degeneracy as does the exact variance - see Section 3.1. Lee's paper does not fully study the asymptotic properties of $\hat{\lambda}_{\mathrm{ML}}$ when r is fixed and $m \to \infty$ (*infill* asymptotics). Both the large r and the large

m asymptotics are easily deduced, under our present mixed-normal assumptions, from Proposition 3.1, from the following two representations of the $F_{r,r(m-1)}$ random variable involved:

$$F_{r,r(m-1)} = \frac{(m-1)\left(\frac{1}{r}\sum_{i=1}^{r}\chi_1^2\right)}{\frac{1}{r}\sum_{i=1}^{r}\chi_{m-1}^2} = \frac{\chi_r^2}{\frac{1}{m-1}\sum_{i=1}^{m-1}\chi_r^2},$$

where all χ^2 variates are independent. From the first of these expressions, together with the fact that, from Proposition 3.1, $\hat{\theta}_{\mathrm{ML}} \cong \theta\sqrt{F_{r,r(m-1)}}$, we see easily that $F_{r,r(m-1)} \xrightarrow{p} 1$ as $r \to \infty$ with m fixed, which implies that $\hat{\theta}_{\mathrm{ML}} \xrightarrow{p} \theta$, and hence that $\hat{\lambda}_{\mathrm{ML}} \xrightarrow{p} \lambda$, a simple example of Lee's (2004) much more general results. Application of the delta method also produces, from the first of these expressions, the known asymptotic normality result under fixed-domain asymptotics. However, the second expression shows that, as $m \to \infty$ with r fixed, $F_{r,r(m-1)} \xrightarrow{d} \chi_r^2/r$. Thus, in fact $\hat{\lambda}_{\mathrm{ML}}$ converges to a random variable under this regime, so is inconsistent under infill asymptotics. We have:

Property 3. *In a pure balanced Group Interaction model with $\varepsilon \sim \mathrm{SMN}(0, I_n)$, the limiting cdf of $\hat{\lambda}_{\mathrm{ML}}$ as $m \to \infty$ with r fixed is,*

$$\lim_{m\to\infty} \Pr(\hat{\lambda}_{\mathrm{ML}} \le z; \lambda) = \Pr\left(\chi_r^2 \le r\left(\frac{1-\lambda}{1-z}\right)^2\right), \quad -\infty < z < 1,$$

for any $\lambda, z \in \Lambda$, and the associated limiting density is

$$\lim_{m\to\infty} \mathrm{pdf}_{\hat{\lambda}_{\mathrm{ML}}}(z;\lambda) = \frac{r^{\frac{r}{2}}(1-\lambda)^r}{2^{\frac{r}{2}-1}\Gamma(\frac{r}{2})(1-z)^{r+1}} e^{-\frac{r}{2}\left(\frac{1-\lambda}{1-z}\right)^2}. \tag{3.7}$$

The large-m asymptotic moments of $\hat{\lambda}_{\mathrm{ML}}$ can be obtained easily from this asymptotic density, and are given, for $s < r$, by

$$\lim_{m\to\infty} \mathrm{E}(\hat{\lambda}_{\mathrm{ML}}^s) = \sum_{j=0}^{s}\binom{s}{j} h_{rj}(\lambda-1)^j,$$

where

$$h_{rj} := \left(\frac{r}{2}\right)^{\frac{j}{2}} \frac{\Gamma(\frac{r-j}{2})}{\Gamma(\frac{r}{2})},$$

with $h_{r0} := 1$. Thus, the large-m distribution has mean

$$\lim_{m\to\infty} \mathrm{E}(\hat{\lambda}_{\mathrm{ML}}) = 1 + h_{r1}(\lambda - 1),$$

and variance

$$\lim_{m\to\infty} \mathrm{var}(\hat{\lambda}_{\mathrm{ML}}) = (h_{r2} - h_{r1}^2)(1-\lambda)^2. \tag{3.8}$$

The limiting bias is thus $\lim_{m\to\infty}\mathrm{E}(\hat{\lambda}_{\mathrm{ML}} - \lambda) = (1-\lambda)(1-h_{r1})$, which is negative for all r and λ, but diminishes rapidly as r increases. The limiting variances under the two asymptotic regimes can be very different, and we shall see later that neither approximates the exact variance very well.

Figure 1 plots the exact density (3.3) and large-r approximation when $r = m = 10$, for $z \in (-1, 1)$, and $\lambda = -0.5, 0, 0.5$. Here and elsewhere we focus on the interval $(-1, 1)$ because it seems to be most relevant in applications. These plots and similar graphical evidence suggest the tentative conclusion that the density of $\hat{\lambda}_{\mathrm{ML}}$ seems, in general, to be well-centered on the true value of λ. The large-r asymptotic approximation seems unsatisfactory even for this sample size, which is essentially what motivates an exact analysis based on the density (3.3).

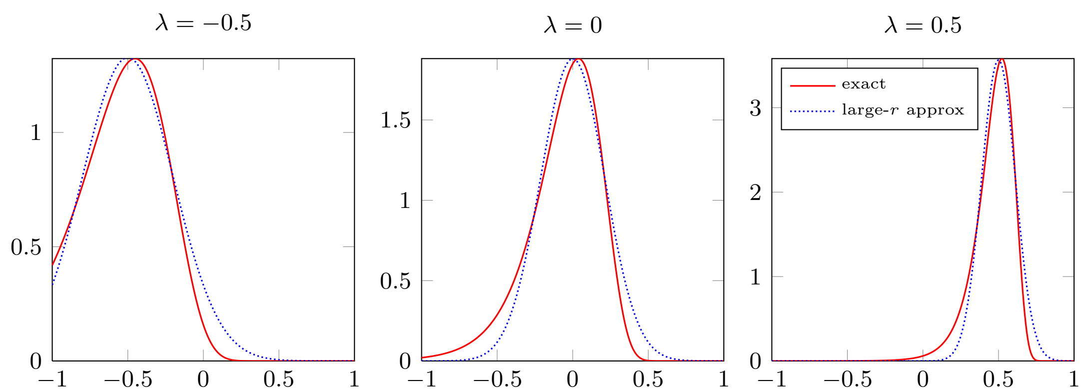

Figure 1: Density of $\hat{\lambda}_{\mathrm{ML}}$ for the pure balanced Group Interaction model with $\varepsilon \sim \mathrm{SMN}(0, I_n)$, when $r = m = 10$.

In Figure 2 we also plot the exact density and the large-m approximation (3.7) when $m = 5, 50$ and $r = 10$ (in this case we plot two different values of m, since the large-m approximation (3.7) is the same for all values of m). Note that when λ is positive the density of $\hat{\lambda}_{\mathrm{ML}}$ is quite insensitive to m, and the large-m density gives an excellent approximation when λ is positive (despite the MLE not converging in probability to a constant as $m \to \infty$). This is due to the fact that in this model information about λ grows very slowly with m. The approximation is less accurate when λ is negative.

3.2 A Median Unbiased Estimator

A second consequence of Proposition 3.1, along with the fact that θ is a monotonic function of λ, is that the median of $\hat{\lambda}_{\mathrm{ML}}$ is defined, in an obvious notation, by the identity $\mathrm{med}(\hat{\theta}_{\mathrm{ML}}) = \theta\sqrt{\mathrm{med}(\mathrm{F}_{r,r(m-1)})}$. Solving this equation yields:

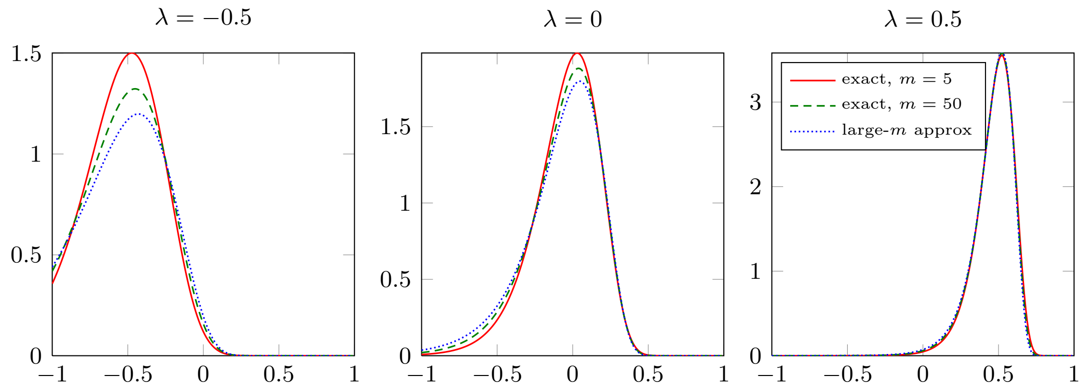

Figure 2: Density of $\hat{\lambda}_{\mathrm{ML}}$ for the pure balanced Group Interaction model with $\varepsilon \sim \mathrm{SMN}(0, I_n)$, when $r = 10$.

Proposition 3.2. *In the pure balanced Group Interaction model with* $\varepsilon \sim \mathrm{SMN}(0, I_n)$, *the median of* $\hat{\lambda}_{\mathrm{ML}}$ *is*

$$\mathrm{med}(\hat{\lambda}_{\mathrm{ML}}) = 1 - \frac{m}{1 + \theta\sqrt{\mathrm{med}(\mathrm{F}_{r,r(m-1)})}}. \tag{3.9}$$

Thus, the median of $\hat{\lambda}_{\mathrm{ML}}$ is a simple function of the median of the F distribution. The median bias of $\hat{\lambda}_{\mathrm{ML}}$ is, for any $\lambda \in \Lambda$,

$$b_{\mathrm{med}}(\lambda) := \mathrm{med}(\hat{\lambda}_{\mathrm{ML}}) - \lambda = \frac{m}{1+\theta} - \frac{m}{1 + \theta\sqrt{\mathrm{med}(\mathrm{F}_{r,r(m-1)})}}. \tag{3.10}$$

The properties of $b_{\mathrm{med}}(\lambda)$ are summarized in the next result, proved in Appendix A.

Proposition 3.3. *In a pure balanced Group Interaction model with* $\varepsilon \sim \mathrm{SMN}(0, I_n)$, $\hat{\lambda}_{\mathrm{ML}}$ *is median-unbiased for all* λ *when* $m = 2$. *For* $m > 2$,

(i) $b_{\mathrm{med}}(\lambda) < 0$ *for all* $\lambda \in \Lambda$;

(ii) $b_{\mathrm{med}}(\lambda) \to 0$ *as* $\lambda \to -(m-1)$ *and as* $\lambda \to 1$, *and also as* $r \to \infty$ *with* m *fixed;*

(iii) $b_{\mathrm{med}}(\lambda)$ *is convex on* Λ, *and* $|b_{\mathrm{med}}(\lambda)|$ *is maximized at*

$$\lambda = \frac{1 - (m-1)\zeta_{r,m}}{1 + \zeta_{r,m}}, \tag{3.11}$$

where $\zeta_{r,m} := (\mathrm{med}(\mathrm{F}_{r,r(m-1)}))^{1/4}$, *with corresponding maximum* $m\,(1 - \zeta_{r,m})\,/(1 + \zeta_{r,m})$.

Note that, since $\zeta_{r,m} > 0$, the point of maximum (3.11) is negative for any r and for any $m > 2$.[4] The asymptotic median bias as $m \to \infty$ can be derived from the fact noted

[4]Also, note that, in terms of θ, the point of maximum is $\theta = 1/\zeta_{r,m}$.

above that $F_{r,r(m-1)} \xrightarrow{p} \chi^2_r/r$ as $m \to \infty$. Thus, (3.9) gives

$$\lim_{m\to\infty} \mathrm{med}(\hat{\lambda}_{\mathrm{ML}}) = 1 - (1-\lambda)\sqrt{\frac{r}{\mathrm{med}\,(\chi^2_r)}}.$$

Figure 3 displays the exact median bias and the large-m median bias of $\hat{\lambda}_{\mathrm{ML}}$ in a Gaussian pure balanced Group Interaction model, obtained from Proposition 3.3, for a range of values of r and m, and plotted against λ. The absolute value of the median bias is large when the number r of groups is small and the group size m is large, and it appears to be decreasing in m and increasing in r.

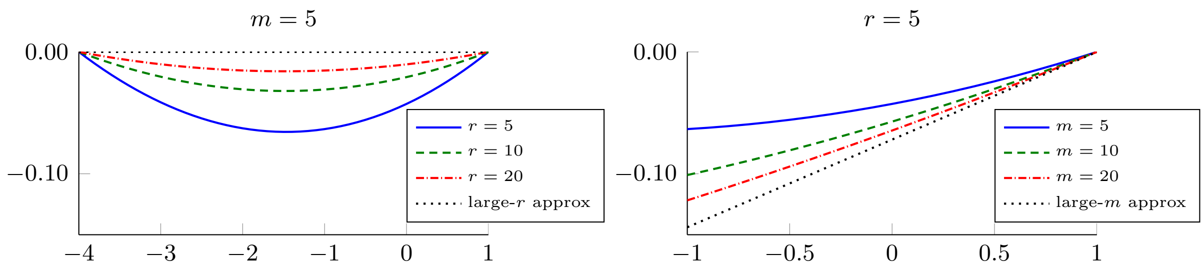

Figure 3: Median bias of $\hat{\lambda}_{\mathrm{ML}}$ for the pure balanced Group Interaction model with $\varepsilon \sim \mathrm{SMN}(0, I_n)$.

Clearly, there can be circumstances in which the median bias of $\hat{\lambda}_{\mathrm{ML}}$ is important, but fortunately this median bias can be *eliminated completely* by exploiting the fact that $\mathrm{med}(\hat{\lambda}_{\mathrm{ML}})$ is known to be a monotonically increasing function of λ. In fact, recalling that $\hat{\theta}_{\mathrm{ML}} \cong \theta\sqrt{F_{r,r(m-1)}}$, we have that $\mathrm{med}(\hat{\theta}_{\mathrm{ML}}/\sqrt{\mathrm{med}(F_{r,r(m-1)})}) = \theta$, i.e., the corrected estimator $\tilde{\theta}_{\mathrm{ML}} := \hat{\theta}_{\mathrm{ML}}/\sqrt{\mathrm{med}(F_{r,r(m-1)})}$ is exactly median-unbiased for θ. Since θ is a monotonically increasing function of λ, we can assert the following:

Proposition 3.4. *In the pure balanced Group Interaction model with* $\varepsilon \sim \mathrm{SMN}(0, I_n)$*, the estimator*

$$\hat{\lambda}_{\mathrm{med}} := \frac{\tilde{\theta}_{\mathrm{ML}} - m + 1}{1 + \tilde{\theta}_{\mathrm{ML}}} = \frac{\hat{\theta}_{\mathrm{ML}} - (m-1)\sqrt{\mathrm{med}(F_{r,r(m-1)})}}{\hat{\theta}_{\mathrm{ML}} + \sqrt{\mathrm{med}(F_{r,r(m-1)})}} \tag{3.12}$$

is exactly median-unbiased for λ.

See Andrews (1993) for a closely related argument in the AR(1) model, but note that here we have the advantage that the median function is known exactly, and is known to be strictly monotonic.

3.3 Exact Confidence Interval for λ

Another consequence of the fact that $\hat{\theta}_{\mathrm{ML}} \cong \theta\sqrt{F_{r,r(m-1)}}$ is that exact confidence sets for λ are immediate. For, from the result in Proposition 3.1, and denoting the α-quantile of the F distribution with (v_1, v_2) degrees of freedom by $\mathrm{F}_{v_1,v_2;\alpha}$, we have[5]

$$\Pr\left(\frac{\hat{\theta}_{\mathrm{ML}}}{\sqrt{\mathrm{F}_{r,r(m-1),1-\alpha/2}}} < \theta < \frac{\hat{\theta}_{\mathrm{ML}}}{\sqrt{\mathrm{F}_{r,r(m-1),\alpha/2}}}\right) = 1 - \alpha.$$

Turning this into a confidence interval for λ, we obtain:

Proposition 3.5. *In the pure balanced Group Interaction model with* $\varepsilon \sim \mathrm{SMN}(0, I_n)$*, a* $100\,(1-\alpha)\,\%$ *exact confidence interval for* λ *is*

$$\left(-(m-1) + \frac{m\hat{\theta}_{\mathrm{ML}}}{\hat{\theta}_{\mathrm{ML}} + \sqrt{\mathrm{F}_{r,r(m-1),1-\alpha/2}}}, 1 - \frac{m\sqrt{\mathrm{F}_{r,r(m-1),\alpha/2}}}{\hat{\theta}_{\mathrm{ML}} + \sqrt{\mathrm{F}_{r,r(m-1),\alpha/2}}}\right). \tag{3.13}$$

Figure 4 plots some confidence bands (3.13), as a function of the observed $\hat{\lambda}_{\mathrm{ML}}$, for $\hat{\lambda}_{\mathrm{ML}} \in \Lambda$, and for $\alpha = 0.05$, $m = 5$, and a series of values of r. When r is small the exact confidence intervals are very wide, but quickly shrink towards $\hat{\lambda}_{\mathrm{ML}}$ (dotted 45 degree line) as r increases.

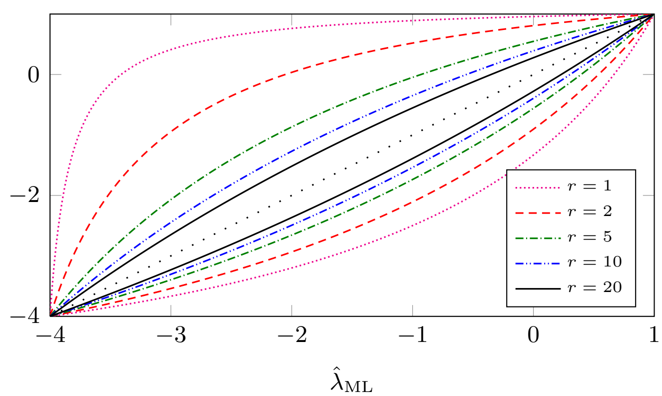

Figure 4: Exact equal-tailed 95% confidence bands for λ in the pure balanced Group Interaction model with $\varepsilon \sim \mathrm{SMN}(0, I_n)$, when $m = 5$.

A commonly used $100\,(1-\alpha)\,\%$ large-r confidence interval for λ, based on the asymptotic normality of $\hat{\lambda}_{\mathrm{ML}}$, is

$$\left(\hat{\lambda}_{\mathrm{ML}} - c_\alpha\sqrt{v_{\hat{\lambda}_{\mathrm{ML}}}}, \hat{\lambda}_{\mathrm{ML}} + c_\alpha\sqrt{v_{\hat{\lambda}_{\mathrm{ML}}}}\right), \tag{3.14}$$

where $v_{\hat{\lambda}_{\mathrm{ML}}}$ is the large-r variance (3.6) evaluated at the MLE, and c_α is the appropriate critical value from the standard normal distribution. Figures 5 compares the confidence intervals (3.14) with the exact confidence intervals (3.13), when $r = 5$, for $m = 5, 50$, and for $\hat{\lambda}_{\mathrm{ML}} \in (-1, 1)$. The general conclusion from this plot, and from similar ones that we do not report, is that, as long as $r > 1$, the asymptotic confidence intervals provide a good approximation to the equal-tailed exact ones if $\hat{\lambda}_{\mathrm{ML}} \in (-1, 1)$. The large-$r$ approximation may be inaccurate for smaller values of $\hat{\lambda}_{\mathrm{ML}}$, but such values of $\hat{\lambda}_{\mathrm{ML}}$ are rare in applications.

[5] As usual, there are many choices for such an interval at a given confidence level. Here we give an interval with equal tail areas, which is not necessarily the shortest, of course.

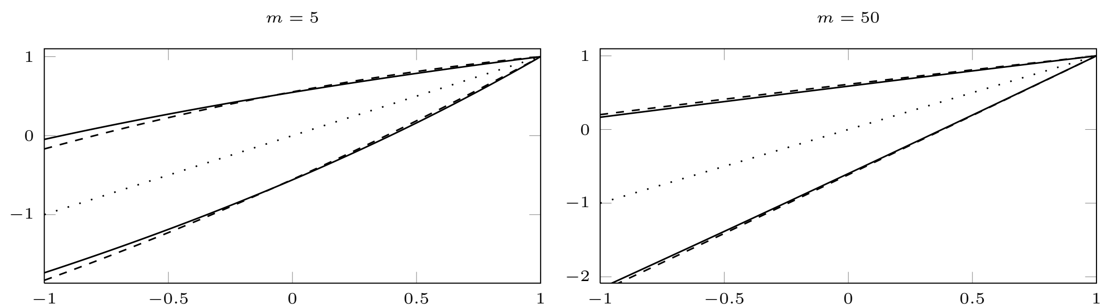

Figure 5: Equal-tailed 95% exact (solid lines) and large-r (dashed lines) confidence bands for λ based on $\hat{\lambda}_{\mathrm{ML}}$, for $\hat{\lambda}_{\mathrm{ML}} \in (-1, 1)$, when $r = 5$.

3.4 Exact Moments

We first discuss the moments of the MLE for θ, $\hat{\theta}_{\mathrm{ML}}$, and, since λ is likely to remain the main parameter of interest, then go on to discuss the moments of $\hat{\lambda}_{\mathrm{ML}}$ itself. From Proposition 3.1 it is easily seen that $\hat{\theta}_{\mathrm{ML}}$ has moments (subject to existence) given by

$$\mathrm{E}(\hat{\theta}_{\mathrm{ML}}^{s}) = \tau^{s} \mathrm{E}\left(f_{r,r(m-1)}^{\frac{s}{2}} \right). \tag{3.15}$$

Evaluating the expectation gives the following result.

Proposition 3.6. *In the pure balanced Group Interaction model with $\varepsilon \sim \mathrm{SMN}(0, I_n)$, the s-th moment of $\hat{\theta}_{\mathrm{ML}}$ exists only for $s < r(m-1)$, and in that case is given by*

$$\mathrm{E}(\hat{\theta}_{\mathrm{ML}}^{s}) = k_s(r, m)\tau^{s}, \;\; s < r(m-1),$$

with

$$k_s(r, m) = k_s := \frac{\Gamma(\frac{r+s}{2})\Gamma(\frac{r(m-1)-s}{2})}{\Gamma(\frac{r}{2})\Gamma(\frac{r(m-1)}{2})}.$$

Thus, although $\hat{\theta}_{\mathrm{ML}}$ itself is biased, the bias, because it is linear in θ, is easily removed, and the variance of the corrected estimator easily computed. Turning to $\hat{\lambda}_{\mathrm{ML}}$ itself, since the sample space for $\hat{\lambda}_{\mathrm{ML}}$ is bounded (and the density is bounded), it is clear that the moments of all orders of $\hat{\lambda}_{\mathrm{ML}}$ exist. However, it is difficult to express the integral defining the moments in terms of the density (3.3) in a useful closed form. It is possible, though, to use the integral expressions to plot the exact mean and variance as functions of λ, and we use these for comparison.

Before considering the moments of $\hat{\lambda}_{\mathrm{ML}}$ itself, we note the following. The mean of $\hat{\theta}_{\mathrm{ML}}$ is a known, monotonically increasing, function of λ, namely $(k_1\sqrt{m-1})\theta$. Inverting that function gives a modified "indirect" estimator of the same form as the median-unbiased estimator $\tilde{\lambda}_{\mathrm{ML}}$ defined above, namely[6]

$$\hat{\lambda}_{\mathrm{mean}} := \frac{\hat{\theta}_{\mathrm{ML}} - (m-1)k_1\sqrt{m-1}}{\hat{\theta}_{\mathrm{ML}} + k_1\sqrt{m-1}}.$$

This correction might be expected to reduce the bias in $\hat{\lambda}_{\mathrm{ML}}$, and is exactly analogous to the median correction given in equation (3.12) above, except that $\sqrt{\mathrm{med}(\mathrm{F}_{r,r(m-1)})}$ is here replaced by $k_1\sqrt{m-1}$.[7] This suggests that we consider a family of estimators of the form

$$\hat{\lambda}_{\phi} := \frac{\hat{\theta}_{\mathrm{ML}} - (m-1)\phi}{\hat{\theta}_{\mathrm{ML}} + \phi}, \tag{3.16}$$

where ϕ is a constant (possibly dependent on (r, m)) to be chosen.[8] The MLE $\hat{\lambda}_{\mathrm{ML}}$ itself corresponds to $\phi = 1$, the median unbiased estimator to $\phi = \sqrt{\mathrm{med}(\mathrm{F}_{r,r(m-1)})}$, and the indirect estimator to $\phi = k_1\sqrt{m-1}$. Note that for both $\hat{\lambda}_{\mathrm{med}}$ and $\hat{\lambda}_{\mathrm{mean}}$, $\phi \to 1$ as $r \to \infty$, so all three estimators are equivalent under fixed-domain asymptotics. We shall consider the moments of $\hat{\lambda}_{\phi}$ generally, thereby covering all three cases.

Taylor expansion of $\hat{\lambda}_{\phi}$ as a function of $\hat{\theta}_{\mathrm{ML}}$ about the mean of $\hat{\theta}_{\mathrm{ML}}$, $k_1\tau$, gives:

$$\hat{\lambda}_{\phi} = 1 - \frac{m\phi}{\phi + k_1\tau} - \frac{m\phi}{\phi + k_1\tau}\sum_{i=1}^{\infty}(-1)^i\left(\frac{\hat{\theta}_{\mathrm{ML}} - k_1\tau}{\phi + k_1\tau}\right)^i.$$

To simplify the notation, put

$$\alpha := \frac{m\phi}{\phi + k_1\tau}, x := \frac{\hat{\theta}_{\mathrm{ML}} - k_1\tau}{\phi + k_1\tau}, \mu_i := E(x^i),$$

so that $\mu_1 = 0$, and

$$\hat{\lambda}_{\phi} = 1 - \alpha - \alpha\sum_{i=1}^{\infty}(-1)^i x^i.$$

Truncating the series at the third order term, and taking expectations using Proposition 3.6, gives[9]

$$\mathrm{E}(\hat{\lambda}_{\phi}) \simeq 1 - \alpha(1 + \mu_2 - \mu_3). \tag{3.17}$$

Similarly, the expansion for $\mathrm{var}(\hat{\lambda}_{\phi})$ up to terms of order 4, is

$$\mathrm{var}(\hat{\lambda}_{\phi}) \simeq \alpha^2\left(\mu_2 - 2\mu_3 + \left(3\mu_4 - \mu_2^2\right)\right). \tag{3.18}$$

[6]Kyriakou, Phillips, and Rossi (2014) consider a different indirect estimator for λ based on the OLS estimator.

[7]The two correction terms seem to be fairly close, except when r is very small, and it seems that $\sqrt{\mathrm{med}(\mathrm{F}_{r,r(m-1)})} < k_1\sqrt{m-1}$.

[8]Note that $\hat{\lambda}_{\phi}$ is supported on Λ for any ϕ .

[9]The approximation cannot be extended to the entire Taylor expansion, because the moments of $\hat{\theta}_{ML}$ exist only up to order $r(m-1) - 1$. However, only the first few terms are needed to obtain an excellent approximation, so this is unimportant.

In these expressions the usual formulae for moments about the mean in terms of raw moments give:

$$\mu_2 = \frac{(k_2 - k_1^2)\tau^2}{(\phi + k_1\tau)^2},\ \mu_3 = \frac{(k_3 - 3k_1k_2 + 2k_1^3)\tau^3}{(\phi + k_1\tau)^3},\ \mu_4 = \frac{(k_4 - 4k_1k_3 + 6k_1^2k_2 - 3k_1^4)\tau^4}{(\phi + k_1\tau)^4}.$$

Fucussing now on the MLE (the case $\phi = 1$), including only the term μ_2 in (3.17) reproduces very accurately the exact mean, over the entire parameter space Λ, and for any r and m. For the variance, using only the first two terms is inadequate, but the three term approximation given in (3.18) reproduce the exact variance very well. Figure 6 plots the exact variance of $\hat{\lambda}_{\mathrm{ML}}$ (obtained by numerical integration) for $\lambda \in (-1, 1)$, along with three different approximations: the third order approximation (3.18), the large-r approximation (3.6) and the large-m approximation (3.8). The third order approximation seems to be vastly superior to the two asymptotic ones.

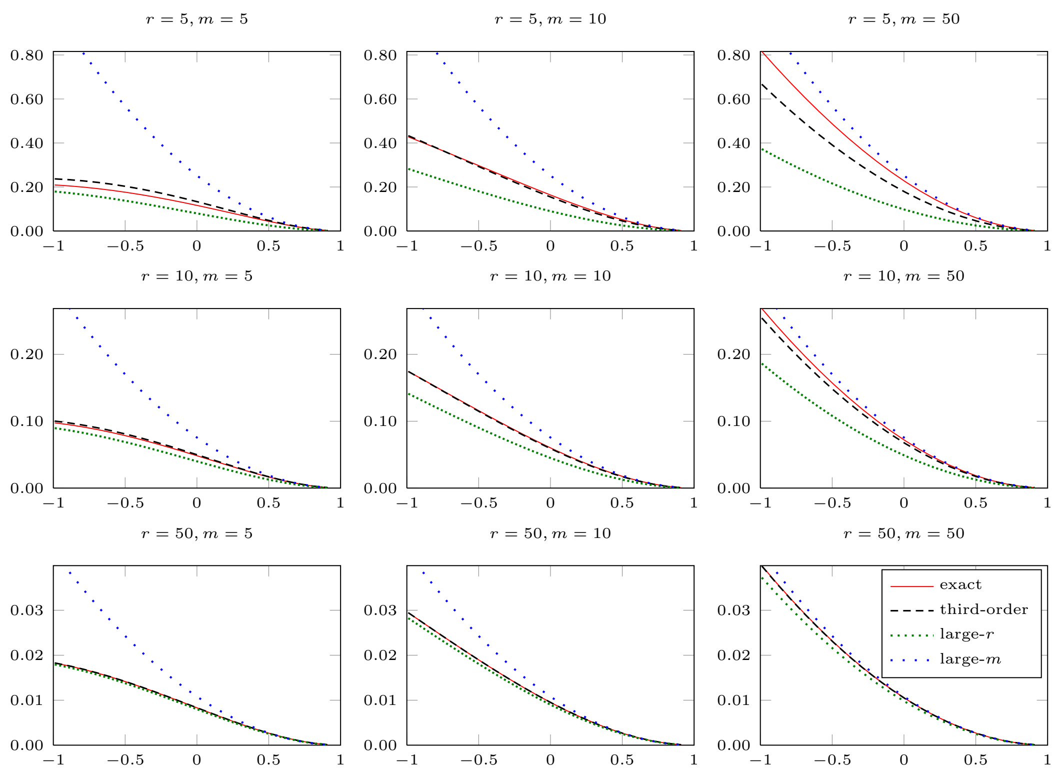

Figure 6: Exact variance of $\hat{\lambda}_{\mathrm{ML}}$, as a function of λ, along with three different approximations.

3.4.1 Bias and Bias Correction

From equation (3.17), omitting the final term μ_3, the approximate bias of $\hat{\lambda}_{\mathrm{ML}}$ is, to this order,

$$b_{\mathrm{mean}}(\lambda) := -\left(\alpha + \lambda - 1 + \alpha(1 + \mu_2)\right),$$

where α and μ_2 are evaluated with $\phi = 1$. Evidently, the bias is negative for all λ if $\alpha + \lambda - 1 > 0$, or $k_1\sqrt{m-1} < 1$, which is so if $m \geq 4$. Thus, based on this approximation, $\hat{\lambda}_{\mathrm{ML}}$ is *negatively biased for all* λ if $m \geq 4$. As might be expected, for moderate r the estimator is almost unbiased for small m, but can be quite biased when m is larger: the matrix W becomes more "dense" as m increases for fixed r.

An alternative approach to bias-correcting $\hat{\lambda}_{\mathrm{ML}}$ is to simply subtract an estimate of the approximate bias $b_{\mathrm{mean}}(\lambda)$ from $\hat{\lambda}_{\mathrm{ML}}$, replacing λ by $\hat{\lambda}_{\mathrm{ML}}$ in $b_{\mathrm{mean}}(\lambda)$. Denoting the estimates of α and μ_2 by $\hat{\alpha}$ and $\hat{\mu}_2$, this means using

$$\hat{\lambda}_{\mathrm{BC}} := 2\hat{\lambda}_{\mathrm{ML}} - 1 + \hat{\alpha}(2 + \hat{\mu}_2). \tag{3.19}$$

We call this a *direct* bias correction.[10] The variance of the corrected estimator can also be obtained by the same methods, but we omit the details. Instead, in Figure 9 in Appendix B we plot the mean bias, for $\lambda \in (-1, 1)$, of $\hat{\lambda}_{\mathrm{ML}}$, and of the three bias-reducing estimators we have introduced, $\hat{\lambda}_{\mathrm{med}}$, $\hat{\lambda}_{\mathrm{mean}}$, $\hat{\lambda}_{\mathrm{BC}}$. This is obtained by straightforward simulation (cf. Remark 3.2). Figures 10 and 11 in Appendix B do the same for the RMSE function and the median bias function.

These figures show that $\hat{\lambda}_{\mathrm{ML}}$ can be significantly biased, but that direct bias correction ($\hat{\lambda}_{\mathrm{BC}}$) essentially removes the entire mean bias. However, $\hat{\lambda}_{\mathrm{BC}}$ performs poorly in terms of the median bias. The estimator $\hat{\lambda}_{\mathrm{med}}$ does not perform as well as $\hat{\lambda}_{\mathrm{BC}}$ in terms of mean bias, but it does reduce a good portion of the mean bias of $\hat{\lambda}_{\mathrm{ML}}$, and is median unbiased by construction. These differing effects reflect the fact that the distribution of $\hat{\lambda}_{\mathrm{ML}}$ can be quite skewed. the estimator $\hat{\lambda}_{\mathrm{mean}}$ appears to be dominated by $\hat{\lambda}_{\mathrm{med}}$ in terms of both mean and median bias. The variances of the four estimators are all virtually identical, and the three bias corrected estimators appear to have lower RMSE than $\hat{\lambda}_{\mathrm{ML}}$, at least when $\lambda \in (-1, 1)$. To conclude, then, bias correction does seem desirable, particularly when r is small and/or m is large, and several methods are available to accomplish this, with varying degrees of success. Which to choose obviously depends on one's preferences.

3.5 Hypothesis Testing: Best Invariant Test

As we have seen, the pure balanced model is a two-parameter regular exponential model. In the canonical parameterization of Remark 3.1 the two sufficient statistics are $s_1 :=$

[10] In greater detail, putting $a := \sqrt{m-1}$, the bias-corrected estimator is

$$\hat{\lambda}_{\mathrm{BC}} = \hat{\lambda}_{\mathrm{ML}} + \frac{m\hat{\theta}_{\mathrm{ML}}}{1 + ak_1\hat{\theta}_{\mathrm{ML}}}\left(\frac{1 - ak_1}{1 + \hat{\theta}_{\mathrm{ML}}} + \frac{a^2(k_2 - k_1^2)\hat{\theta}_{\mathrm{ML}}}{(1 + ak_1\hat{\theta}_{\mathrm{ML}})^2}\right).$$

$y'(I_r \otimes M_{\iota_m})y$, and $s_2 := y'(I_r \otimes P_{\iota_m})y$, with the distribution properties stated in Remark 3.1. The problem of testing $H_0 : \lambda = 0$ is invariant under the group of scale changes $s_1 \to as_1, s_2 \to as_2$, $a > 0$, applied to the sufficient statistics, and under this group the statistic s_2/s_1 is a (single) maximal invariant. The MLE $\hat{\lambda}_{\mathrm{ML}}$ is itself invariant, therefore also maximal, since both are one-dimensional. The class of invariant tests in this model therefore coincides with the class of tests based on $\hat{\lambda}_{\mathrm{ML}}$.[11] Since we know the distribution of $\hat{\lambda}_{\mathrm{ML}}$ (under the $\mathrm{SMN}(0, I_n)$ assumption), we can apply the Neyman Pearson Lemma to the distribution of $\hat{\lambda}_{\mathrm{ML}}$ to obtain the uniformly most powerful invariant (UMPI) test of H_0 against each one-sided alternative. The resulting test can be shown to coincide with the Moran test (see King (1981), who gives an analogous result for the case $r = 1$).[12]

The Neyman-Pearson Lemma applied to the density of $\hat{\lambda}_{\mathrm{ML}}$ given in (3.3) gives a best critical region consisting of large values of the likelihood ratio

$$\frac{\mathrm{pdf}_{\hat{\lambda}_{\mathrm{ML}}}(z;\lambda)}{\mathrm{pdf}_{\hat{\lambda}_{\mathrm{ML}}}(z;0)} \propto \left(\frac{1+\frac{1}{m-1}U(z)}{1+\frac{m-1}{\theta^2}U(z)}\right)^{\frac{rm}{2}},$$

where $U(z) := ((z+m-1)/((m-1)(1-z)))^2$. This ratio is increasing or decreasing in $U(z)$ as $\theta/(m-1) \gtrless 1$, so the best invariant test rejects H_0 against alternatives $\lambda > 0$ when $U(\hat{\lambda}_{\mathrm{ML}}) = (\hat{\theta}_{\mathrm{ML}}/(m-1))^2$ is large, and rejects against alternatives $\lambda < 0$ when $U(z)$ is small. The critical values for a two-sided test can be derived directly from the $F_{r,r(m-1)}$ distribution, since, under H_0, $U(z) \sim F_{r,r(m-1)}$. Noting that, in the canonical representation of the model, $(m-1)s_2/s_1$ is the MLE for the parameter $(\theta/(m-1))^2$, we can therefore state:

Proposition 3.7. *In the pure balanced Group Interaction model with $\varepsilon \sim \mathrm{SMN}(0, I_n)$, the UMPI test of $H_0 : \lambda = 0$ against alternatives $H_1^+ : \lambda > 0$ ($H_1^- : \lambda < 0$) rejects H_0 when $U(\hat{\lambda}_{\mathrm{ML}}) = (m-1)s_2/s_1$ is large (small).*[13] *The test is exact, and critical values can be obtained from the fact that, under H_0, $U(\hat{\lambda}_{\mathrm{ML}}) \sim F_{r,r(m-1)}$.*

When H_0 is false the test statistic $U(\hat{\lambda}_{\mathrm{ML}})$ has the distribution

$$U(\hat{\lambda}_{\mathrm{ML}}) \sim \left(\frac{\lambda+m-1}{(m-1)(1-\lambda)}\right)^2 F_{r,r(m-1)},$$

so that, for any critical value t_α,

$$\Pr\left(U(\hat{\lambda}_{\mathrm{ML}}) > t_\alpha\right) = \Pr\left(\mathrm{F}_{r,r(m-1)} > t_\alpha \left(\frac{(m-1)(1-\lambda)}{\lambda+m-1}\right)^2\right), \tag{3.20}$$

[11] The likelihood ratio test is also invariant, therefore also based on s_1/s_2, or $\hat{\lambda}_{ML}$, as can be shown directly. The same applies to a test based on a Studentized version of $\hat{\lambda}_{ML}$, using, say, the estimated asymptotic variance as $r \to \infty$.

[12] As usual, of course, there is no *uniformly* best test against two-sided alternatives.

[13] The last equality here follows from the fact that, in the canonical representation of the model, $(m-1)s_2/s_1$ is the MLE for the parameter $(\theta/(m-1))^2$.

with a similar expression for the other tail. For the one-sided test against $H_1^+ : \lambda > 0$, therefore, it is clear that the power $\to 1$ as $\lambda \to 1$, and the analogous conclusion holds as $\lambda \to -(m-1)$ for a one-sided test against $H_1^- : \lambda < 0$. Exact power curves for the test(s) are easily obtained from equation (3.20).[14]

3.6 Balanced Models with Regressors

The exact results derived above for the pure model do not generalize easily to the case of an arbitrary regressor matrix X. However, extensions are straightforward under certain specific assumptions on X, and we give some examples of this next. These examples are important in their own right, but also because they might suggest approximations for the case of an arbitrary X. Before continuing, we note that some care is required in dealing with the models with regressors, because there are choices for X that mean that the number of sufficient statistics is less than the number of parameters, in which case inference (on the full parameter) is impossible. See Arnold (1979), Lee (2007), and H&M for further discussion of this issue. This problem arises in the present balanced model when $\text{col}(X)$ contains either of the two eigenspaces of W, which are $\text{col}(I_r \otimes \iota_m)$ and $\text{col}(I_r \otimes L_m)$, where L_m is a matrix whose columns span the orthogonal complement of ι_m.[15] To rule this out we need the following assumption.

Assumption A. Neither $\text{col}(I_r \otimes \iota_m)$ nor $\text{col}(I_r \otimes L_m)$ is in $\text{col}(X)$.

Note that $I_r \otimes \iota_m$ is the group fixed effect matrix. Hence, Assumption A requires, in particular, that the model does not contain group fixed effects. In the general model (1.3) with regressors the random part of the log-likelihood is, under Gaussian assumptions,

$$(S_\lambda y - X\beta)' (S_\lambda y - X\beta) = (y - \lambda W y)'(y - \lambda W y) + \beta' X' X \beta - 2\beta' X'(y - \lambda W y).$$

In general this cannot be written in terms of fewer than $2k + 3$ sufficient statistics, but in certain special cases reduction is possible. In the balanced Group Interaction model the first component can, as we have seen, be written in terms of s_1, s_2. The last term is in general a combination of both $X'y$ and $X'Wy$, but it can be reduced to a single k-vector if $W'X = XA$ for some $k \times k$ matrix A (including $A = 0$), that is, if $\text{col}(X)$ is an invariant subspace of W'. In this case the statistic $X'y$ is sufficient. The case $A = 0$ requires that the column space of X is orthogonal to the column space of W, which, assuming X is of full column rank k, can only be so if $\text{rank}(W) \leq n - k$. This possibility therefore does not arise for the models studied in this paper, in which W has full rank. But, for the balanced model, the column space of X can indeed be an invariant subspace of W'.

[14] An alternative approach would be to apply the NPL to the distribution of the statistic s_2/s_1 directly. It is straightforward but tedious to show that this yields exactly the same test as $\hat{\lambda}_{\text{ML}}$ itself.

[15] If, for instance, $\text{col}(X)$ contains $\text{col}(I_r \otimes \iota_m)$, the only terms in the profile log-likelihood that involve λ are $-n \log(\lambda + m - 1) + \log(\det(S_\lambda))$, so the profile score does not depend on the data.

The simplest example of this is, as noted in H&M, the case of a constant mean, i.e., $k = 1$ and $X = \iota_n = \iota_r \otimes \iota_m$. We then have

$$W'X = (I_r \otimes B_m)\,\iota_n = \iota_r \otimes B_m \iota_m = \iota_r \otimes \iota_m = X,$$

because ι_m is an eigenvector of B_m corresponding to the eigenvalue 1. More generally, we may have $X = (I_r \otimes \iota_m)\,R$, for some $r \times k$ matrix R ($k < r$), in which case

$$W'X = (I_r \otimes B_m)\,(I_r \otimes \iota_m)\,R = (I_r \otimes B_m \iota_m)\,R = (I_r \otimes \iota_m)\,R = X,$$

for the same reason. These cases entail that $\mathrm{col}(X)$ is spanned by eigenvectors of W associated to the unit eigenvalue. Alternatively, $\mathrm{col}(X)$ may be spanned by eigenvectors associated to the eigenvalue $-1/(m-1)$, or more generally, some combination of the two. If so we will have $X = (X_1, X_2)$, say, with X_1 of dimension $n \times k_1$ ($k_1 < r$), $\mathrm{col}(X_1) \subseteq \mathrm{col}(I_r \otimes \iota_m)$, and X_2 of dimension $n \times k_2$ ($k_2 < r(m-1)$), $\mathrm{col}(X_2) \subseteq \mathrm{col}(I_r \otimes L_m)$.[16]

In this circumstance the term $(y - \lambda W y)' M_X (y - \lambda W y)$ that appears in the profile likelihood, and yields all of the results discussed earlier for the pure model, can instead be written as a linear combination of the two statistics

$$\tilde{s}_1 := \tilde{y}_1' M_{\tilde{X}_1} \tilde{y}_1, \;\; \tilde{s}_2 := \tilde{y}_2' M_{\tilde{X}_2} \tilde{y}_2,$$

with the same coefficients as earlier. Here, $\tilde{y} := H'y$ and $\tilde{X} := H'X$, where $H := (I_r \otimes L_m, I_r \otimes l_m)$, with $l_m := \iota_m / \sqrt{m}$, is the orthogonal matrix of eigenvectors of W. Thus, $\tilde{X}_1 = (I_r \otimes L_m)' X_1$ is $r(m-1) \times k_1$, and $\tilde{X}_2 = (I_r \otimes l_m)' X_2$ is $r \times k_2$. It is easily checked that

$$\frac{\tilde{s}_1 (\lambda + m - 1)^2}{\sigma^2 (m-1)^2} \sim \chi^2_{r(m-1)-k_2}, \; \frac{\tilde{s}_2 (1-\lambda)^2}{\sigma^2} \sim \chi^2_{r-k_1}.$$

Thus, the only changes needed to all of the above results, for models of this structure, are to the respective degrees of freedom of the F- variate involved in the expressions for the cdf. Thus, we have established the following general result.[17]

Proposition 3.8. *Suppose Assumption A holds. In the balanced Group Interaction model with $\varepsilon \sim \mathrm{SMN}(0, I_n)$, if $\mathrm{col}(X) \subset \mathrm{col}(I_r \otimes \iota_m) \cup \mathrm{col}(I_r \otimes L_m)$, with $k_1 := \dim(\mathrm{col}(X) \cap \mathrm{col}(I_r \otimes \iota_m)) < r$, and $k_2 := k - k_1 < r(m-1)$, then the cdf of $\hat{\lambda}_{\mathrm{ML}}$ is, for any $\lambda, z \in \Lambda$,*

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z; \lambda) = \Pr\left(\mathrm{F}_{v_1, v_2} \leq \frac{v_2}{v_1} \frac{c(z,\lambda)}{m-1} \right),$$

with $v_1 := r - k_1$ and $v_2 := r(m-1) - k_2$, and the corresponding density is

$$\mathrm{pdf}_{\hat{\lambda}_{\mathrm{ML}}}(z;\lambda) = \frac{2m\tau^{v_2}}{B\left(\frac{v_1}{2}, \frac{v_2}{2}\right)} \frac{(1-z)^{v_2 - 1} (z + m - 1)^{v_1 - 1}}{\left(\tau^2 (1-z)^2 + (z+m-1)^2\right)^{\frac{n-k}{2}}}. \tag{3.21}$$

[16] The inequalities $k_1 < r$ and $k_2 < r(m-1)$ must be strict for Assumption A to be satisfied.

[17] Alternatively Proposition 3.8 can be derived directly using results in H&M. Note that if $v_1 = 1$ the limit of the density (3.21) as $z \downarrow -(m-1)^{-1}$ is not zero. This is because the case $v_1 = 1$ is "close" to the degenerate case $v_1 = 0$, in which case $l_p(\lambda)$ is unbounded in a neighborhood of $-(m-1)^{-1}$.

It is certainly true that the conditions needed in Proposition 3.8 are restrictive, but they are met in some simple cases of practical interest, in addition to the constant mean case $X = \iota_n$. We briefly describe two of these next.

Remark 3.4. *Another consequence of the condition that* $\mathrm{col}(X) \subset \mathrm{col}(I_r \otimes \iota_m) \cup \mathrm{col}(I_r \otimes L_m)$, *but Assumption A holds, is that the Cliff-Ord test for* $H_0 : \lambda = 0$ *is UMPI against a one sided alternative in a mixed-Gaussian Group Interaction model. Here invariance is with respect to the group of transformations* $y \to \kappa y + X\delta$ *in the sample space, for any* $\kappa > 0$, *any* $\delta \in \mathbb{R}^k$*; see King (1981).*[18]

3.6.1 Individual Fixed Effects

The model is

$$y_i = \lambda B_m y_i + \mu + \varepsilon_i,\ i = 1, .., r, \tag{3.22}$$

where $y_i \in \mathbb{R}^m$, for each for $i = 1, .., r$, is the subvector of y corresponding to i-th group, μ is $m \times 1$, so the groups (districts) have a common mean $(I_m - \lambda B_m)^{-1}\mu$, and a common autoregressive parameter λ. This is model (1.3) with $W = I_r \otimes B_m$, $X = \iota_r \otimes I_m$, and $\beta = \mu$. Proposition 3.8 applies with $k_1 = 1$ and $k_2 = m - 1$, and gives

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z; \lambda) = \Pr\left(\mathrm{F}_{r-1,(r-1)(m-1)} \leq c(z, \lambda)\right). \tag{3.23}$$

That is, as one might have expected, this case is analogous to a pure model having $r - 1$ rather than r copies of a complete graph on m vertices. The asymptotics are thus the same as in Section 3.1.2: $\hat{\lambda}_{\mathrm{ML}}$ is consistent and asymptotically normal as $r \to \infty$, and converges in distribution to a random variable as $m \to \infty$ with r fixed.

The model (3.22) is a special case of the spatial/panel model studied in the recent paper by Robinson and Rossi (2015), the difference being that in their paper B_m in (3.22) is replaced by a general weights matrix W, common to the blocks, our μ is their c, and our (r, m) are their (T, n). Under Robinson and Rossi's assumptions, $\hat{\lambda}_{\mathrm{ML}}$ is consistent and asymptotically normal as (their) n goes to infinity, and they are able to obtain an Edgeworth expansion for the distribution of $\hat{\lambda}_{\mathrm{ML}}$. These results do not conflict with those just discussed, because, crucially, the matrix B_m does not satisfy the key assumption, Assumption 3 (iv) in Robinson and Rossi (2015).

3.6.2 Group-Specific Regressions

Consider now consider a balanced Group Interaction model with group specific β coefficients:

$$y_i = \lambda B_m y_i + X_i \beta_i + \varepsilon_i,\ \ i = 1, .., r, \tag{3.24}$$

[18] If Assumption A does not hold, the Cliff-Ord statistic is degenerate, in the same sense as the profile score is. As a consequence, the final paragraph of King (1981) needs to be interpreted with great care.

where the matrices X_i are $m \times k_i$, with $k_i \leq m$, for all i. In this case $X = \bigoplus_{i=1}^r X_i$ ($\bigoplus$ denoting matrix direct sum), $k = \sum_{i=1}^r k_i$, and $\beta' = (\beta_1', .., \beta_r')$ in equation (1.3). For each group one can check that the $k_i + 3$ statistics $s_{1i} = y_i' M_{\iota_m} y_i, s_{2i} = y_i' P_{\iota_m} y_i,\ X_i' y_i,$ and $\iota_m' y_i$ are sufficient for the k_i+2 parameters. The sums $s_1 = \sum_{i=1}^r s_{1i}$ and $s_2 = \sum_{i=1}^r s_{2i}$, together with the $X_i' y_i$, $i = 1, .., r$, are therefore sufficient in the full model. If $\mathrm{col}(X_i)$ contains ι_m the statistic $\iota_m' y_i$ is already accounted for in $X_i' y_i$, so under this condition the model is regular for a single group. However, the condition $\iota_m \in \mathrm{col}(X_i)$ cannot be permitted for every i, for this would mean that $\mathrm{col}(I_r \otimes \iota_m)$ were a subspace of $\mathrm{col}(X)$, violating Assumption A. The alternative that also produces a regular model is that for those i for which $\iota_m \notin \mathrm{col}(X_i)$, $\mathrm{col}(X_i) \subset \mathrm{col}(L_m)$. In this case the term involving $\iota_m' y_i$ does not appear, and $X_i' y_i$ is sufficient, again giving a regular model for that group. Note that $\mathrm{col}(X_i) \subset \mathrm{col}(L_m)$ would hold, for instance, if the elements of X_i were deviations of the raw data from their respective within-group sample means. Assuming, therefore, that $\iota_m \in \mathrm{col}(X_i)$ for $r - s$ groups, with $s > 0$, and that, for the remaining s groups, $\mathrm{col}(X_i) \subset \mathrm{col}(L_m)$, the conditions of Proposition 3.8 are satisfied with $k_1 = r - s$, $k_2 = k - r + s$. The cdf is therefore

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z; \lambda) = \Pr\left(\mathrm{F}_{s, rm-k-s} \leq \frac{rm - k - s}{s(m-1)} c(z, \lambda)\right). \tag{3.25}$$

The asymptotics are then easily established. As $m \to \infty$, $\hat{\lambda}_{\mathrm{ML}}$ converges in distribution to a random variable (because $\mathrm{F}_{s,rm-k-s} \xrightarrow{d} \chi_s^2/s$). When $r \to \infty$, consider first the case when both s and $rm - k - s$ diverge. Then $\mathrm{F}_{s,rm-k-s} \xrightarrow{p} 1$. We distinguish two cases, according to whether $(rm - k - s)/s$ is bounded or not as $r \to \infty$. In the former case, let $\rho := \lim_{r\to\infty} (rm - k - s)/(s(m-1))$, and define $\gamma := \rho^{-1/2}\theta$. Then, for any $z, \lambda \in \Lambda$,

$$\lim_{r\to\infty} \Pr(\hat{\lambda}_{\mathrm{ML}} \leq z; \lambda) = I\left(1 \leq \rho c(z, \lambda)\right) = I\left(z \geq \frac{\gamma - m + 1}{1 + \gamma}\right),$$

where $I(\cdot)$ is the indicator function, taking value 1 when its argument is true and 0 otherwise. Thus, if $\lim_{r\to\infty} (rm - k - s)/s$ is bounded, $\hat{\lambda}_{\mathrm{ML}} \xrightarrow{p} (\gamma - m + 1)/(1 + \gamma) \in \Lambda$. Thus, we have convergence in probability, but to an (in general) incorrect point in Λ. If, on the other hand, $(rm - k - s)/s \to \infty$, then the representation (3.25) implies that $\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z; \lambda) \to 1$, for any $\lambda, z \in \Lambda$, that is, $\hat{\lambda}_{\mathrm{ML}} \xrightarrow{p} -(m-1)$. Finally note that the result $\hat{\lambda}_{\mathrm{ML}} \xrightarrow{p} -(m-1)$ is also obtained if s is fixed as $r \to \infty$ (because $\mathrm{F}_{s,rm-k-s} \xrightarrow{d} \chi_s^2/s$ and $(r(m-1) - s)/(s(m-1))c(z, \lambda) \to \infty$).

We shall see below that in the unbalanced case there is no need to rule out the presence of group specific fixed effects. This will enable us to obtain a general exact representation of the cdf of $\hat{\lambda}_{\mathrm{ML}}$ in the case of group specific regressions.

3.7 Conclusion on the Balanced Model

The balanced Group Interaction model, the key property of which is that the spatial design matrix W has just two distinct eigenvalues, is obviously a “toy” model, of the same status, perhaps, as the simple Gaussian regression model, and the AR(1) model in the time-series literature. Indeed, within the class of models in which W is the adjacency matrix of a graph, it is the *only* model with just two distinct eigenvalues. Its practical relevance is obviously limited, but, as with the other examples mentioned, one hopes that study of its properties will be informative more generally. It goes without saying that one can only hope to obtain exact results under very restrictive assumptions, and we make no apology for beginning the study of inference in this class of models with its simplest member. However, in the interests of pragmatism, we now move on to the much more realistic, and therefore more complicated, unbalanced case.

4 The Unbalanced Model

The unbalanced Group Interaction model - with groups of different sizes - presents a much greater challenge, even for the pure model. In this section we present an exact result for the distribution of $\hat{\lambda}_{\mathrm{ML}}$, and some approximations to it. But, so far, we are unable to extend the detailed inference results obtained above for the balanced model to this more difficult case. The key difficulty is that some of the coefficients $d_{tt}(z, \lambda)$ in the expression for the cdf in equation (2.2) change sign as z varies in Λ. This means that there are points in Λ at which the cdf is *non-analytic*, and that the distribution has a different functional form in different sub-intervals of Λ. This makes analytical work with the exact distribution extremely difficult, if not impossible. Nevertheless, it is possible to make some progress by other means.

On the other hand, the presence of groups of different sizes has a favorable consequence: contrary to the balanced case, inference about λ remains possible if (all) group specific fixed effects are included in the regression. We shall see that this immediately implies a simple representation of the cdf of $\hat{\lambda}_{\mathrm{ML}}$ that holds for general regressors, provided only that all β parameters are group specific and that group specific fixed effects are included.

In Sections 4.1-4.4 we restrict ourselves to the pure case, and often, for simplicity, we focus on the case of two group sizes. As is clear in equation (1.3), the interest-parameter λ is still assumed constant across groups. The case of group specific regressions is discussed briefly in Section 4.7.

Remark 4.1. *As noted in Remark 3.1, the (Gaussian) unbalanced model is also a member of the curved exponential family. Indeed the likelihood is the product of p versions of that for the balanced model, with different group sizes, and different multiplicities. Each of these has sufficient statistics and canonical parameters of the same type as those given earlier for the balanced model. That is, the exponent of the exponential part of the likelihood is of the form*

$$\eta_1 \sum_{i=1}^{p}(s_{1i} + s_{2i}) + \eta_2 \sum_{i=1}^{p}\left(s_{2i} + \frac{s_{1i}}{(m_i - 1)^2}\right) + 2\eta_3 \sum_{i=1}^{p}\left(s_{2i} - \frac{s_{1i}}{(m_i - 1)}\right).$$

It is not possible to rewrite this as a linear combination of two statistics with constant coefficients, so the model is a $(3, 2)$ curved model, as mentioned. In this representation of the model the statistics s_{1i}, s_{2i} are all independent of each other, and are proportional to χ^2 variates. Note that the sum can be written as

$$-\frac{1}{2\sigma^2}\left((1-\lambda)^2 s_2 + \sum_{i=1}^{p} s_{1i}\left(\frac{\lambda+m_i-1}{m_i-1}\right)^2\right),$$

with $s_2 = \sum_{i=1}^{p} s_{2i}$, a linear combination of $p+1$ independent multiples of χ^2 variates.

Remark 4.2. *The estimating equation $\dot{l}_p(\lambda) = 0$ is, in this case, a polynomial of degree $p+1$ in λ, and has no explicit solution if $p > 3$. It must be solved numerically. However, the fact that the equation is known to have a single zero in Λ makes this a much simpler task than it would otherwise be.*

4.1 Exact Representation

In the unbalanced Group Interaction model each different group size introduces an extra distinct eigenvalue of W. If there are r_i groups of distinct sizes m_i, $i = 1, ..., p$, with m_1 the smallest group size, the eigenvalues of W are: 1, with multiplicity $r = \sum_{i=1}^{p} r_i$, and, for each $i = 1, .., p$, $-1/(m_i - 1)$ with multiplicity $n_i = r_i(m_i - 1)$.[19] The total sample size is $n = \sum_{i=1}^{p} r_i m_i$, and the number of distinct eigenvalues of W is $T = p + 1$. Since, for any group-interaction model, W is symmetric, the cdf of $\hat{\lambda}_{\mathrm{ML}}$ is, under mixed-Gaussian assumptions on ε, given by (2.2).

We will need the following property of the coefficient functions $g_t(z) - \bar{g}(z)$ in equation (2.3), proved in H&M for any W having only real eigenvalues: for any $z \in \Lambda$, the coefficients $g_t(z) - \bar{g}(z)$, $t = 1, .., T$, are in increasing order (i.e., $s > t$ implies $g_s(z) > g_t(z)$). For any $z \in \Lambda$, $g_1(z) - \bar{g}(z) < 0$, $g_T(z) - \bar{g}(z) > 0$, and, for any $t = 2, ..., T - 1$, $g_t(z) - \bar{g}(z)$ changes sign exactly once on Λ.

We can divide the left-hand term in the inequality in (2.2) by the (positive) coefficient in the final term in the sum, giving the equivalent exact representation of the cdf,

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z; \lambda) = \Pr\left(\sum_{t=1}^{T} c_t(z,\lambda)\left(g_t(z) - \bar{g}(z)\right)\chi^2_{n_t} \leq 0\right), \tag{4.1}$$

where, for the Group Interaction model considered here,

$$c_t(z,\lambda) := \left(\frac{(1-\lambda)(z+m_t-1)}{(1-z)(\lambda+m_t-1)}\right)^2, \; t = 1, .., p, \tag{4.2}$$

and $c_T(z, \lambda) := 1$, all reducing to $c(z, \lambda)$ in equation (3.1) when the model is balanced. Since some of the $g_t(z) - \bar{g}(z)$ are positive and some are negative, for any given $z \in \Lambda$, it

[19] The corresponding eigenspaces are $\mathrm{col}(\bigoplus_{i=1}^{p}(I_{r_i} \otimes \iota_{m_i}))$ associated to the eigenvalue 1 and $\mathrm{col}(I_{r_i} \otimes L_{m_i})$ associated to $-1/(m_i - 1)$, $i = 1, ..., p$. It is easily verified that when $p = 1$ the eigenstructure reduces to the one given in Section 3.

follows that, for any $z \in \Lambda$, $\sum_{t=1}^{T} c_t(z,\lambda)(g_t(z)-\bar{g}(z))\chi^2_{n_t}$ reduces to the difference between two positive linear combinations of independent χ^2 variates.

Remark 4.3. *Notice that, for all $z \in \Lambda$, $c_t(z,\lambda) \to 0$ as $\lambda \to 1$, for each $t = 1,..,p$, while $c_T(z,\lambda) = 1$. Since $g_T(z) - \bar{g}(z) > 0$ for all z, $\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z; \lambda) \to 0$ as $\lambda \to 1$. Likewise, as $\lambda \to -(m_1 - 1)$, all coefficients in (4.1), other than the first, remain finite, while $c_1(z,\lambda) \to \infty$ for all $z \in \Lambda$. Since $g_1(z) - \bar{g}(z) < 0$ for all z, $\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z; \lambda) \to 1$ as $\lambda \to -(m_1-1)$. Thus, as in the balanced case, the distribution of $\hat{\lambda}_{\mathrm{ML}}$ becomes degenerate as λ approaches the endpoints of Λ.*

The eigenvalues of G_z are $g_t(z) = -1/(z+m_t-1)$, $t = 1,..,p$, and $g_{p+1}(z) = 1/(1-z)$, so that

$$\bar{g}(z) = \frac{z}{n(1-z)} \sum_{i=1}^{p} \frac{r_i m_i}{z+m_i-1},$$

and

$$g_t(z) - \bar{g}(z) = -\frac{1}{n(1-z)} \sum_{i=1}^{p} \left(r_i m_i \left(\frac{1-z}{z+m_t-1} + \frac{z}{z+m_i-1} \right) \right), \tag{4.3}$$

for $t = 1,..,p$, while

$$g_{p+1}(z) - \bar{g}(z) = \frac{1}{n(1-z)} \sum_{i=1}^{p} \frac{r_i m_i (m_i - 1)}{z+m_i-1}.$$

Note that $\bar{g}(z)$ has the sign of z, and $\bar{g}(0) = 0$. As noted earlier, $g_1(z) - \bar{g}(z) < 0$ for all $z \in \Lambda$, $g_{p+1}(z) - \bar{g}(z) > 0$ for all $z \in \Lambda$, and the remaining terms all change sign exactly once as z traverses Λ. Thus, the number of positive and negative terms in the representation (4.1) varies with z. If the model is balanced ($p = 1$) the exact representation given here reduces to the result for the balanced case discussed earlier.

For any $p \geq 2$, let z_t denote the unique point in Λ at which $g_t(z) - \bar{g}(z) = 0$, for each $t = 2,...,p$. The distribution of $\hat{\lambda}_{\mathrm{ML}}$ is non-analytic at the points z_t, and has a different functional form in each interval between successive points. The number of positive and negative terms in (4.1) remains the same within an interval, but the numbers of each differ in the different intervals.

Example 1 (Two Group Sizes)**.** *In the case $p = 2$ we have, after simplification,*[20]

$$\begin{aligned}
g_1(z) - \bar{g}(z) &= -\frac{n(m_2-1) + z(n - r_2 m_2 (m_2 - m_1))}{n(1-z)(z+m_1-1)(z+m_2-1)} \\
g_2(z) - \bar{g}(z) &= -\frac{n(m_1-1) + z(n + r_1 m_1 (m_2 - m_1))}{n(1-z)(z+m_1-1)(z+m_2-1)} \\
g_3(z) - \bar{g}(z) &= \frac{r_1 m_1 (m_1-1)(z+m_2-1) + r_2 m_2 (m_2-1)(z+m_1-1)}{n(1-z)(z+m_1-1)(z+m_2-1)}.
\end{aligned}$$

[20] The common denominators of the coefficients here could obviously be dropped, but to economize on notation we do not do so.

The first is always negative, the last always positive, for $z \in \Lambda$, while the second changes sign at

$$z_2 = -\frac{n(m_1 - 1)}{n + r_1 m_1 (m_2 - m_1)} < 0, \tag{4.4}$$

being negative for $z > z_2$, positive for $z < z_2$.

After briefly discussing the asymptotic properties of $\hat{\lambda}_{\mathrm{ML}}$ that follow from the exact representation, we next discuss (again briefly) the distribution properties of linear combinations of independent χ^2 variates with positive coefficients, a subject upon which there is a large literature.

4.2 Asymptotics in the Unbalanced Group Interaction Model

The representation of the cdf of $\hat{\lambda}_{\mathrm{ML}}$ in equation (4.1) provides a useful starting point for deriving asymptotic properties of $\hat{\lambda}_{\mathrm{ML}}$ under the mixed Gaussian assumption. Different asymptotic regimes are possible, depending on how the m_i's and the r_i's are assumed to behave as the total sample size grows. To understand the issues we rewrite equation (4.1) in the form

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z; \lambda) = \Pr\left(\chi_r^2 + \sum_{t=1}^{p} \psi_t(z,\lambda)\chi_{n_t}^2 \leq 0\right). \tag{4.5}$$

where

$$\psi_t(z,\lambda) := c_t(z,\lambda)\frac{g_t(z) - \bar{g}(z)}{g_{p+1}(z) - \bar{g}(z)}, \tag{4.6}$$

for $t = 1, ..., p$. Assuming p, the number of different group sizes, is fixed, one can again consider two types of asymptotic regime. The first, infill asymptotics, holds the r_i fixed (hence also r), and assumes one or more of the m_i produce the increased sample size. The second, fixed-domain asymptotics, holds the m_i fixed and assumes an increase in one or more of the r_i. This second case satisfies the assumptions in Lee (2004). Hence, it is already known that, under regularity conditions, $\hat{\lambda}_{\mathrm{ML}}$ is consistent and asymptotically normal. In the first case Lee's (2004) results leave the properties of $\hat{\lambda}_{\mathrm{ML}}$ open.

In fact, the situation is very much as in the balanced case: it is clear from (4.5) that in the first case, convergence will be to a random variable, because the term χ_r^2 in (4.5) will be unaffected. Precise details for this situation depend on exactly what is assumed about the behaviour of the m_i, but $\hat{\lambda}_{\mathrm{ML}}$ is clearly again inconsistent under infill asymptotics. In the second case the known results are easily deduced from the representation (4.5) by a characteristic function argument.

4.3 Exact distribution of a Positive Linear Combinations of χ^2 Variates

As we have just seen, we need to deal with pairs of statistics of the form

$$Q_s := \sum_{i=1}^{s} a_i \chi_{n_i}^2,$$

with all the $a_i > 0$. In our case these coefficients are functions of z.

Define the $n \times n$ diagonal matrix ($n = \sum_{i=1}^{s} n_i$)

$$A = A_{n_1,...,n_s}(a_1, ..., a_s) := \mathrm{diag}\left(a_i I_{n_i}, i = 1, .., s\right).$$

It is well known that the cumulants of Q_s of all orders exist are given by

$$\kappa_l := 2^{l-1}(l-1)!\mathrm{tr}(A^l) = 2^{l-1}(j-1)!\pi_l, \tag{4.7}$$

where, $\pi_l := \sum_{i=1}^{s} n_i a_i^l = \mathrm{tr}(A^l)$. These properties are quite simple, but, despite that, exact distribution theory for Q_s is not straightforward, and there is a very large literature dealing with the subject. We briefly introduce some of this next. Let ϕ be a positive number such that $\phi a_i \geq 1$ for all i. An expression for the exact density is

$$\mathrm{pdf}_{Q_s}(q) = \frac{|\phi A|^{-\frac{1}{2}}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \exp\left(-\frac{1}{2}\phi q\right) q^{\frac{n}{2}-1} {}_1F_1\left(\frac{1}{2}, \frac{n}{2}; \frac{1}{2} q\phi \left(I_n - (\phi A)^{-1}\right)\right), \tag{4.8}$$

(see James (1964), and Ruben (1962)). The hypergeometric function here is a confluent hypergeometric function with matrix argument (Muirhead (1982), Chapter 7), and it is this that make the distribution difficult. For ϕ such that $\phi a_i > 1$ for all i, the distribution of ϕQ_s can be expressed as a *mixture* of central χ^2 distributions with weights

$$p_j(\phi A) := \frac{\left(\frac{1}{2}\right)_j}{j!} |\phi A|^{-\frac{1}{2}} C_j(I_n - (\phi A)^{-1}), \tag{4.9}$$

where $(a)_j := a(a+1)...(a+j-1)$ is the Pochhammer symbol, and $C_j(\cdot)$ denotes the top-order zonal polynomial of order j in the indicated matrix. It is easy to confirm that the $p_j(\phi A)$ are non-negative and sum to unity. The choice of $\phi > 0$ is arbitrary subject to $\phi \min\{a_i, i = 1, ..., s\} > 1$. The weights $p_j(\phi A)$ are relatively complicated polynomials in the a_i, and are difficult to interpret.[21] See Ruben (1962) and Johnson, Kotz, and Balakrishnan (1994) for further details of these and related expansions. There is some incentive, therefore, to seek approximations to the distribution, and we discuss some of these briefly below.

In the case $s = 2$, however, the result is reasonably simple. Without loss of generality we consider the distribution of a statistic of the form $Q = a_1\chi^2_{v_1} + a_2\chi^2_{v_2}$, with $0 < a_1 < a_2$.

Proposition 4.1. *Let $Q := a_1\chi^2_{v_1} + a_2\chi^2_{v_2}$, with $0 < a_1 < a_2$. The density of Q is given by*

$$\mathrm{pdf}_Q(q) = \frac{\phi^{\frac{v}{2}}\psi^{\frac{v_2}{2}} \exp\left(-\frac{\phi q}{2}\right) q^{\frac{v}{2}-1}}{2^{\frac{v}{2}}\Gamma(\frac{v}{2})} {}_1F_1\left(\frac{v_2}{2}, \frac{v}{2}; \frac{1}{2}\phi q\left(1-\psi\right)\right), \tag{4.10}$$

where $\phi = 1/a_1$, $v := v_1 + v_2$, and $\psi := a_1/a_2 < 1$.

[21] Recall that the non-central χ^2 distribution also has this form, but with a Poisson mixing distribution with mean equal to the non-centrality parameter. This is obviously simpler than the present case.

The distribution function follows at once. Note that the hypergeometric function in (4.10) has scalar argument, and is a built-in function in most modern mathematical packages.

Equation (4.10) can be rewritten as

$$\mathrm{pdf}_Q(q) = \phi\psi^{\frac{v_2}{2}} \sum_{k=0}^{\infty} \frac{\left(\frac{v_2}{2}\right)_k (1-\psi)^k}{k!} g_{v+2k}(\phi q), \tag{4.11}$$

where $g_\xi(\cdot)$ denotes the density function of a χ^2_ξ random variable. This representation as a mixture of χ^2 densities is useful for some calculations, and for interpretation, but is perhaps less so for computation purposes.

4.3.1 Approximations for Positive Definite Forms

Because the exact distribution of a positive definite quadratic form is quite complicated, there is a clear incentive to approximate. And, because such forms are ubiquitous throughout statistics, there is a very large literature on the subject. The simplest approximation, usually attributed to Fisher, is to treat Q_s as a multiple of a χ^2 variate, $Q_s = \alpha\chi^2_v$, choosing α and v so that the first two cumulants of the two distributions agree. This entails the choices $\alpha = \pi_2/\pi_1$ and $v = \pi_1^2/\pi_2$, where, as above, $\pi_l = \sum_{i=1}^{s} n_i a_i^l$.

A more sophisticated approximation due to Hall (1983) and Buckley and Eagleson (1988), is to use three parameters, with $Q_s = \alpha\chi^2_v + \beta$, and choosing (α, β, v) so that the first three cumulants agree. This entails the choices $\alpha = \pi_3/\pi_2, \beta = \pi_1 - \pi_2^2/\pi_3$, and $v = \pi_2^3/\pi_3^2$. Buckley and Eagleson (1988) show that this representation can be formally justified by an argument based on Edgeworth expansions of the two distributions involved, and give explicit bounds on the error involved in approximating the distribution function in this way. Hall (1983) calls this a "penultimate" approximation to the distribution of Q_s, which of course, when suitably standardised, converges to a standard normal variate. For our purposes, the simpler two-cumulant approximation is more useful, and seems to work quite well. A number of other, typically more complicated, approximations are extant - for a comprehensive discussion, see Johnson, Kotz, and Balakrishnan (1994).

4.4 Exact Distribution of $\hat{\lambda}_{\mathrm{ML}}$

From the exact results for a pair of independent positive linear combinations like those given above, one can easily obtain an exact formula for the probability $\Pr(Q_{1t} \leq Q_{2t})$, with Q_{it} based on matrix A_{it}, by simple transformation and integration. The result is given in H&M, Section 5, and has the following form: for $z \in (z_t, z_{t+1})$, between successive points z_t,

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z; \lambda) = \mathrm{E}_{J,K}\left(\Pr\left(f_{v_{1t}+2J, v_{2t}+2K} \leq \frac{\phi_1}{\phi_2}\right)\right), \tag{4.12}$$

where $v_{1t} := \sum_{i=1}^{t} n_i$, $v_{2t} := \sum_{i=t+1}^{p+1} n_i$. The symbol $\mathrm{E}_{J,K}$ here denotes the operation of applying two independent weightings of the form (4.9), with suitably defined matrices

A_{1t}, A_{2t}, to the "conditional" probability involved. That is, in each interval we have a different representation of the distribution that, conditionally, is analogous to the result for the balanced model.

Obviously, the "conditional" formulae for each subinterval of Λ are simple enough, but it remains true that the unconditional result, after averaging with respect to the distributions of J, K, is forbiddingly complicated, and, worse, not particularly informative about the properties of the estimator. Moreover, it is probably impossible to obtain the density directly from equation (4.12), simply because of the complexity of the polynomials involved. There is therefore considerable interest in obtaining valid approximations to the exact results that are more easily interpreted, and more informative. Before considering that further, in the next section we give the exact results for the case of just two group sizes (i.e., $p = 2$), which are reasonably tractable.

4.4.1 Two Group Sizes (continued)

When there are $p = 2$ different group sizes the coefficients of the three χ^2 variates in the sum in (4.1) have the following signs:

	$\chi^2_{r_1(m_1-1)}$	$\chi^2_{r_2(m_2-1)}$	χ^2_r
$z < z_2$	$-$	$+$	$+$
$z > z_2$	$-$	$-$	$+$

Using the coefficients in (4.6), we have, for $z < z_2$, where $\psi_2(z, \lambda) > 0$,

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z; \lambda) = \Pr\left(\psi_2(z, \lambda)\chi^2_{n_2} + \chi^2_r \leq (-\psi_1(z, \lambda))\chi^2_{n_1}\right), \tag{4.13}$$

while for $z > z_2$, where $\psi_2(z, \lambda) < 0$,

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z; \lambda) = \Pr\left(\chi^2_r \leq -\psi_1(z, \lambda)\chi^2_{n_1} + \psi_2(z, \lambda)\chi^2_{n_2}\right). \tag{4.14}$$

Each of these involves a linear combination of two χ^2 random variables with positive coefficients, and a third, independent χ^2 variate. Expressions for the distribution functions in the two intervals can be obtained by applying the results in the previous subsection, but it is difficult to use those expressions to obtain information about the properties of $\hat{\lambda}_{\mathrm{ML}}$, in particular, its density.[22] Here we pursue an alternative conditioning argument that is more successful.

Remark 4.4. *Noting that $\psi_2(z_2, \lambda) = 0$, and, as is easily verified, $-\psi_1(z_2, \lambda) = r/n_1$, we have, on setting $z = z_2$ in either of equations (4.13) or (4.14),*

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z_2; \lambda) = \Pr(\mathrm{F}_{r,n_1} < c_1(z_2, \lambda)).$$

[22]The difficulty is that both the conditional distribution, given $J = j$, and the mixing probabilities, are functions of z.

For values of r_1, r_2 that are not too small this function (of λ) is near 1 for $\lambda < z_2$, and near zero for $\lambda > z_2$, falling sharply from 1 to 0 in the neighborhood of z_2. That is, for values $\lambda < z_2$ $\hat{\lambda}_{\mathrm{ML}}$ is almost certainly below z_2, and for values $\lambda > z_2$ it is almost certainly above z_2. If $z_2 < -1$, and $\lambda \in (-1, 1)$, this implies that the distribution of $\hat{\lambda}_{\mathrm{ML}}$ will be almost entirely confined to the interval $z > z_2$. For $\lambda = z_2$, $\Pr(\hat{\lambda}_{\mathrm{ML}} \le z_2; z_2) = \Pr(\mathrm{F}_{r,n_1} < 1)$, which is near .5 as long as r/n_1 is near 1. Other evidence about the median will be discussed shortly.

Let q_v denote a χ^2_v random variable. All such variables in the expressions to follow are independent. For $z < z_2$, we can condition on the variables q_r and q_{n_2} on the left in the expression for $\Pr(\hat{\lambda}_{\mathrm{ML}} \le z; \lambda)$, giving the conditional result

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \le z | q_r, q_{n_2}, \lambda) = 1 - \mathcal{G}_{n_1}\left(\frac{q_r + \psi_2(z,\lambda) q_{n_2}}{-\psi_1(z,\lambda)}\right), \quad -(m_1 - 1) < z < z_2,$$

where $\mathcal{G}_v$ denotes the cdf of the χ^2_v random variable. For $z > z_2$, we can condition instead on (q_{n_1}, q_{n_2}), giving

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \le z | q_{n_1}, q_{n_2}, \lambda) = \mathcal{G}_r\left(-(\psi_1(z,\lambda) q_{n_1} + \psi_2(z,\lambda) q_{n_2})\right), \ z_2 < z < 1,$$

Expressions for the unconditional cdf's can be obtained from these by averaging, but we shall focus instead on the unconditional density in each interval. The reason that this is straightforward is that expressions for the conditional *density* are easily obtained from these conditional cdf's, and these can then be converted into the (components of the) unconditional density.

To state the results for this situation recall the notation introduced above: $A_{n_1,n_2}(a_1, a_2)$ denotes the matrix $\mathrm{diag}\,(a_i I_{n_i}, i = 1, 2)$, and $C_j(A)$ denotes the top-order zonal polynomial of degree j of a matrix A. We need the following lemma.

Lemma 4.2. *We have*

$$\left(\frac{1}{2}\right)_j C_j(A_{n_1,n_2}(a_1, a_2)) = \sum_{k=0}^{j} \binom{j}{k} \left(\frac{n_1}{2}\right)_k \left(\frac{n_2}{2}\right)_{j-k} a_1^k a_2^{j-k}.$$

We can then obtain the following result.

Proposition 4.3. *Let $a(z)$ and $c(z)$ be strictly positive functions of z on some interval Λ_0. Let $q_1 \sim \chi^2_\alpha$, $q_2 \sim \chi^2_\beta$ be independent, and let w be a random variable with conditional cdf, given (q_1, q_2), given by*

$$\Pr(w \le z | q_1, q_2) = \mathcal{G}_\gamma(a(z) q_1 + c(z) q_2)$$

for $z \in \Lambda_0$. The conditional density of w, given (q_1, q_2), is therefore

$$\mathrm{pdf}_w(z|q_1, q_2) = \frac{\exp\left(-\frac{1}{2}\left(a(z) q_1 + c(z) q_2\right)\right)}{2^{\frac{\gamma}{2}} \Gamma(\frac{\gamma}{2})} \left(a(z) q_1 + c(z) q_2\right)^{\frac{\gamma}{2} - 1} \left(\dot{a}(z) q_1 + \dot{c}(z) q_2\right), \tag{4.15}$$

where the dot denotes the derivative with respect to z. Then, denoting the unconditional density of w at $w = z$ when the parameters are (α, β, γ) by $\mathrm{pdf}_w(z; \alpha, \beta, \gamma)$, we have (omitting the argument of $a(\cdot)$ and $c(\cdot)$ for simplicity):

(i) for $\gamma = 2$,

$$\mathrm{pdf}_w(z; \alpha, \beta, 2) = \frac{\alpha \frac{\dot{a}}{1+a} + \beta \frac{\dot{c}}{1+c}}{2(1+a)^{\frac{\alpha}{2}}(1+c)^{\frac{\beta}{2}}}; \tag{4.16}$$

(ii) for $\gamma = 2s + 2$, with $s = 1, 2, ...,$

$$\mathrm{pdf}_w(z; \alpha, \beta, 2s+2) = \frac{\left(\frac{1}{2}\right)_s}{2s!(1+a)^{\frac{\alpha}{2}}(1+c)^{\frac{\beta}{2}}} \times \left(\frac{\alpha \dot{a}}{1+a} C_s \left(A_{\alpha+2,\beta}(\varphi_1, \varphi_2) \right) + \frac{\beta \dot{c}}{1+c} C_s \left(A_{\alpha,\beta+2}(\varphi_1, \varphi_2) \right) \right), \tag{4.17}$$

where $\varphi_1 := a/(1+a)$, and $\varphi_2 := c/(1+c)$.

Note that the two terms in (4.17) are finite polynomials, not infinite series. The statement of Proposition 4.3 is restricted to even degrees of freedom γ for simplicity; the corresponding formulae for odd γ are considerably more complicated, and are given in the proof of Proposition 4.3 in Appendix A. Under a certain restriction on the parameters a general expression for the density, valid for all γ, can be obtained that is analogous to equation (4.17), but in which the final term is a linear combination of two hypergeometric functions. This result is given in Lemma A.3, reported after the proof of Proposition 4.3.

Applying Proposition 4.3 to the unbalanced model, we require two applications of the result, as summarized in the following table:

Interval	α	a	φ_1	β	c	φ_2	γ
$-(m_1 - 1) < z < z_2$	r	$-\frac{1}{\psi_1(z,\lambda)}$	$\frac{1}{1-\psi_1(z,\lambda)}$	n_2	$-\frac{\psi_2(z,\lambda)}{\psi_1(z,\lambda)}$	$\frac{\psi_2(z,\lambda)}{\psi_2(z,\lambda)-\psi_1(z,\lambda)}$	n_1
$z_2 < z < 1$	n_1	$-\psi_1(z,\lambda)$	$-\frac{\psi_1(z,\lambda)}{1-\psi_1(z,\lambda)}$	n_2	$-\psi_2(z,\lambda)$	$-\frac{\psi_2(z,\lambda)}{1-\psi_2(z,\lambda)}$	r

In Figure 7 we display the exact density for the case when $r_1 = r_2 = 1$ (so $r = 2$) and one of the two groups has fixed size 2, varying the size of the other group, and hence varying n. The density is plotted for three different values of λ. When the model is balanced ($r = m = 2$, so that $n = 4$) the density is analytic on $\Lambda = (-1, 1)$. On the other hand, when the model is unbalanced there is a clearly visible point of non-analyticity at z_2. Using expression (4.4), this point is $-.4545$ for $n = 10$, and it approaches $-1/3$ from the left as $n \to \infty$.

The plots show clearly that the density has a single component only when the model is balanced. As the difference between m_1 and m_2 increases, the difference between that two components becomes more apparent, and the density becomes less smooth at the point z_2. Incidentally, in this model, the point of non-analyticity is not an asymptote, but a point of non-differentiablity. In other models the reverse can occur. Note that this phenomenon

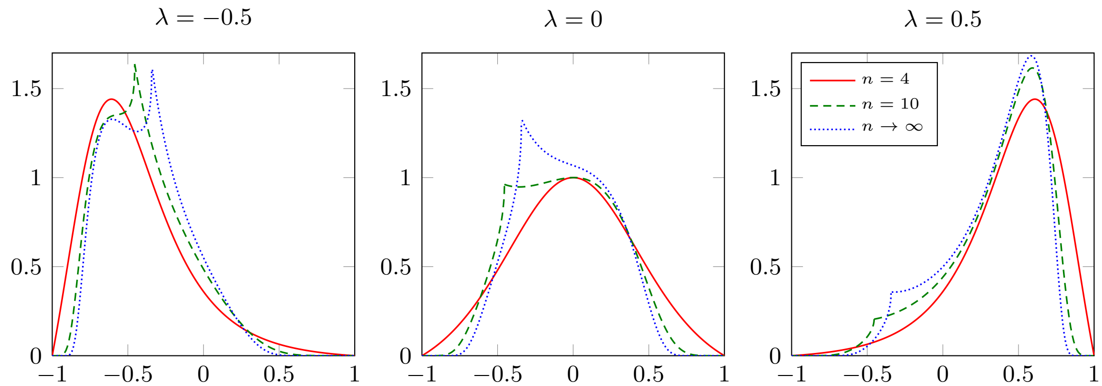

Figure 7: Density of $\hat{\lambda}_{\mathrm{ML}}$ for the Gaussian pure Group Interaction model with two groups, one of which has size $m_1 = 2$.

could be regarded as a consequence of imposing the same parameter λ on the two different groups.

Additional figures for values of $m_1 > 2$ are given in Appendix B. All of these figures show that the properties of $\hat{\lambda}_{\mathrm{ML}}$ are, in this model with just two groups, almost invariant to the sample size, a property related to, but not implied by the asymptotic properties for a fixed number of groups mentioned earlier. However, even though the estimator is not consistent under some asymptotic regimes, there is certainly no evidence here that suggests not using maximum likelihood in this model.

4.5 Probability of Underestimation: the median

We next consider the special case of equation (4.1) with $z = \lambda$, so that the object of interest becomes $\Pr(\hat{\lambda}_{\mathrm{ML}} \leq \lambda; \lambda)$, the probability of underestimating λ. This seems to be the only available method for examining the median bias of $\hat{\lambda}_{\mathrm{ML}}$ in this unbalanced model. When $z = \lambda$, we have $c_t(\lambda, \lambda) = 1$ for all t and all λ, so that

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq \lambda; \lambda) = \Pr\left(\sum_{t=1}^{T} \left(g_t(\lambda) - \bar{g}(\lambda)\right) \chi^2_{n_t} \leq 0\right). \tag{4.18}$$

If $\lambda \geq z_p$, which includes all values $\lambda \geq 0$, all of the coefficients in this expression are negative, except the last. Thus, for $\lambda \geq z_p$ we have

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq \lambda; \lambda) = \Pr\left(\chi^2_r \leq \sum_{t=1}^{p} \psi_t(\lambda) \chi^2_{n_t}\right),$$

where

$$\psi_t(\lambda) := \psi_t(\lambda, \lambda) = -\frac{g_t(\lambda) - \bar{g}(\lambda)}{g_{p+1}(\lambda) - \bar{g}(\lambda)}, \ t = 1, .., p.$$

Using the exact expression for the density of the variate on the right given in equation (4.8), it is straightforward to deduce a formula for the required probability. The expression is

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq \lambda; \lambda) = |\phi A|^{-\frac{1}{2}} \sum_{j=0}^{\infty} \frac{(\frac{1}{2})_j}{j!} C_j \left(I_{n_1+n_2} - (\phi A)^{-1}\right) \times \Pr\left(\mathrm{Beta}\left(j + \frac{n_1+n_2}{2}, \frac{r}{2}\right) \leq \frac{\phi}{1+\phi}\right). \quad (4.19)$$

Here, $\mathrm{Beta}(a, c)$ denotes a beta variate with parameters a, c. Regrettably, this formula is just as complicated as the exact density itself in (4.8), and does not easily yield conclusions about the median of the estimator. A simpler, more helpful approach, is to use the Fisher approximation for the linear combination on the right, i.e., to assume

$$\sum_{t=1}^{p} \psi_t(\lambda) \chi^2_{n_t} \cong \alpha \chi^2_v,$$

where $\alpha = \pi_2/\pi_1$, and $v(\lambda) := \pi_1^2/\pi_2$, and $\cong$ denotes equality in distribution. In this case things simplify greatly, because $\pi_1 = r$, so $\alpha = \pi_2/r, v(\lambda) = r^2/\pi_2$, which produces the approximation, for $\lambda \geq z_p$,

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq \lambda; \lambda) \simeq \Pr\left(\chi^2_r \leq \alpha \chi^2_{v(\lambda)}\right) = \Pr\left(\mathrm{F}_{r,v(\lambda)} \leq \frac{\alpha v(\lambda)}{r}\right) = \Pr\left(\mathrm{F}_{r,v(\lambda)} \leq 1\right), \quad (4.20)$$

an analogue of the result given earlier for the balanced model. But, as we have noted earlier, $\Pr(F_{r,v} \leq 1) > .5$ if $v > r$, and vice versa. That is, up to the accuracy of this approximation, $\mathrm{med}(\hat{\lambda}_{\mathrm{ML}}) < \lambda$ if $v(\lambda) > r$, and $\mathrm{med}(\hat{\lambda}_{\mathrm{ML}}) > \lambda$ if $v(\lambda) < r$. There is therefore a negative median-bias when λ is in the set $\{\lambda : \lambda > z_p, v(\lambda) > r\}$, and a positive median-bias when $\lambda \in \{\lambda : \lambda > z_p, v(\lambda) < r\}$.

For the interval at the lower end of Λ, i.e., $-(m_1 - 1) < \lambda < z_2$, the opposite situation occurs: all coefficients in the linear combination are positive, except the first. In this case we have an expression for $\Pr(\hat{\lambda}_{\mathrm{ML}} \leq \lambda; \lambda)$ of the form

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq \lambda; \lambda) = 1 - \Pr\left(\chi^2_{n_1} \leq \sum_{t=2}^{p+1} \tilde{\psi}_t(\lambda) \chi^2_{n_t}\right),$$

with $n_{p+1} = r$, and

$$\tilde{\psi}_t(\lambda) := -\frac{g_t(\lambda) - \bar{g}(\lambda)}{g_1(\lambda) - \bar{g}(\lambda)}, t = 2, .., p+1.$$

In this interval the appropriate parameters for the approximation are $\tilde{\alpha} = \tilde{\pi}_2/n_1$ and $\tilde{v}(\lambda) = n_1^2/\tilde{\pi}_2$, with

$$\tilde{\pi}_2 := \sum_{t=2}^{p+1} n_t \tilde{\psi}_t^2(\lambda) = \frac{\sum_{t=2}^{p+1} (g_t(\lambda) - \bar{g}(\lambda))^2}{(g_1(\lambda) - \bar{g}(\lambda))^2}, \quad (4.21)$$

and we have the approximation, for $-(m_1-1)<\lambda<z_2$,

$$\Pr(\hat{\lambda}_{\text{ML}} \le \lambda; \lambda) \simeq 1 - \Pr\left(F_{n_1,\tilde{v}(\lambda)} \le 1\right),$$

For values of λ between z_2 and z_p the expression for $\Pr(\hat{\lambda}_{\text{ML}} \le \lambda; \lambda)$ will involve the difference between two positive linear combinations of χ^2 variates. Each can separately be approximated as above, and an approximation for the probability easily obtained. For each interval the approximation takes the form, in obvious notation,

$$\Pr(\hat{\lambda}_{\text{ML}} \le \lambda; \lambda) \simeq \Pr\left(\mathrm{F}_{v_L,v_R} \le 1\right),$$

so the only things needed are the pairs (v_L, v_R) appropriate to each interval. The reason for this is as follows: when the approximation is used for both sides of an inequality we have, symbolically,

$$\begin{aligned}\Pr(\alpha_L\chi^2_{v_L} \le \alpha_R\chi^2_{v_R}) &= \Pr\left(\mathrm{F}_{v_L,v_R} \le \frac{v_R\alpha_R}{v_L\alpha_L}\right)\\ &= \Pr\left(\mathrm{F}_{v_L,v_R} \le -\frac{\pi_{1R}}{\pi_{1L}}\right)\\ &= \Pr\left(\mathrm{F}_{v_L,v_R} \le 1\right),\end{aligned}$$

since it is always the case that $\pi_{1R}+\pi_{1L}=0$. For example, in the case $p=4$ we have four intervals to accommodate, and the following results for the approximation to $\Pr(\hat{\lambda}_{\text{ML}} \le \lambda; \lambda)$ are typical of the general case:

$$\begin{aligned}
-(m_1-1)<\lambda<z_2 :&\quad \Pr\left(\mathrm{F}_{v_A,n_1} \le 1\right), && v_A := \tfrac{(n_1\psi_1)^2}{\left(r+\sum_{i=2}^4 n_i\psi_i^2\right)}\\
z_2<\lambda<z_3 :&\quad \Pr\left(\mathrm{F}_{v_{BR},v_{BL}} \le 1\right), && v_{BR} := \tfrac{(n_1\psi_1+n_2\psi_2)^2}{n_1\psi_1^2+n_2\psi_2^2},\ v_{BL} := \tfrac{(n_1\psi_1+n_2\psi_2)^2}{(n_3\psi_3^2+n_4\psi_4^2+r)}\\
z_3<\lambda<z_4 :&\quad \Pr\left(\mathrm{F}_{v_{CR},v_{CL}} \le 1\right), && v_{CR} := \tfrac{(n_4\psi_4+r)^2}{\sum_{i=1}^3 n_i\psi_i^2},\ v_{CL} := \tfrac{(n_4\psi_4+r)^2}{n_4\psi_4^2+r}\\
z_4<\lambda<1 :&\quad \Pr\left(\mathrm{F}_{r,v_D} \le 1\right), && v_D := \tfrac{r^2}{\sum_{i=1}^4 n_i\psi_i^2}.
\end{aligned}$$

Evidence on the accuracy of the approximation is given in the following table, where we compare exact results (obtained by simulating (4.18)) with those obtained by the approximation, for the case $p=4$, and three different combinations of the group sizes (design 1: $m_1=5, m_2=10, m_3=15, m_4=20$; design 2: $m_1=10, m_2=20, m_3=30, m_4=40$; design 3: $m_1=5, m_2=50, m_3=100, m_4=150$).

	$\lambda=-.9$		$\lambda=0$		$\lambda=.9$	
Design	Exact	Approx.	Exact	Approx.	Exact	Approx.
1	.561	.561	.580	.579	.582	.583
2	.581	.580	.587	.587	.588	.589
3	.553	.553	.585	.585	.592	.592

Note that for all cases considered in the table $\Pr(\hat{\lambda}_{\text{ML}} \le \lambda; \lambda) > .5$, i.e., the median bias is negative. Based on our calculations using the approximation developed in this section, this seems a general result for whenever $\lambda \in (z_p, 1)$ (similarly, the median bias seems to be always positive for $\lambda \in (-(m_1-1), z_2)$).

4.5.1 Probability of Underestimation: Two Group Sizes

Returning now to exact results, in the case of two distinct group sizes ($p = 2$) the two intervals $-(m_1 - 1) < \lambda < z_2$, and $z_2 < \lambda < 1$ make up all of Λ, and each of the above expressions involves a positive linear combination of just two χ^2 variates. We can therefore use the result in equation (4.19), together with Lemma 4.2, twice, to obtain expressions for the required probability in each of these intervals. For the first (upper) interval, $\phi = 1/\psi_1(\lambda)$, and $\phi A = \mathrm{diag}(I_{n_1}, (\psi_2(\lambda)/\psi_1(\lambda))\, I_{n_2})$, and the expression reduces to

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq \lambda; \lambda) = \left(\frac{\psi_2(\lambda)}{\psi_1(\lambda)}\right)^{-\frac{n_2}{2}} \sum_{j=0}^{\infty} \frac{(\frac{n_2}{2})_j}{j!} \left(1 - \frac{\psi_1(\lambda)}{\psi_2(\lambda)}\right)^j \times \Pr\left(\mathrm{Beta}\left(j + \frac{n_1 + n_2}{2}, \frac{r}{2}\right) \leq \frac{1}{1 + \psi_1(\lambda)}\right). \quad (4.22)$$

For the lower interval, $\phi = 1/\tilde{\psi}_2(\lambda)$ and $\phi A = \mathrm{diag}(I_{n_2}, (\tilde{\psi}_3(\lambda)/\tilde{\psi}_2(\lambda))I_r)$, so that

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq \lambda; \lambda) = 1 - \left(\frac{\tilde{\psi}_3(\lambda)}{\tilde{\psi}_2(\lambda)}\right)^{-\frac{r}{2}} \sum_{j=0}^{\infty} \frac{(\frac{r}{2})_j}{j!} \left(1 - \frac{\tilde{\psi}_2(\lambda)}{\tilde{\psi}_3(\lambda)}\right)^j \times \Pr\left(\mathrm{Beta}\left(j + \frac{n_2 + r}{2}, \frac{n_1}{2}\right) \leq \frac{1}{1 + \tilde{\psi}_2(\lambda)}\right). \quad (4.23)$$

These formulae can be used to plot the probability $\Pr(\hat{\lambda}_{\mathrm{ML}} \leq \lambda; \lambda)$ as a function of λ. Figure 8 plots (a truncated version of) the formulae (4.22) and (4.23) in the case of two group sizes, for $\lambda \in (-1, 1)$, and for a variety of values of r_1, r_2, m_1, m_2. The results in Figure 8 were compared to simulation results, and also to the approximation based on Fishers method discussed above. All three methods give virtually identical results. In the left panel the two group sizes are $m_1 = 10$ and $m_2 = 20$, and the three lines are for different values of the numbers r_1 and r_2 of groups of sizes m_1 and m_2. In the right panel, there are two groups, and the four lines are for different combinations of m_1 and m_2 such that $m_1 + m_2 = 30$. Note that the solid line in the right panel corresponds to a balanced case, in which case $\Pr(\hat{\lambda}_{\mathrm{ML}} \leq \lambda; \lambda)$ does not depend on λ (see Section 3.1).[23] The left panel shows that as r_1 and r_2 increase the probability of underestimation converges to .5. The right panel shows that the probability of underestimation can be very sensitive to λ, even for values of λ in $(-1, 1)$.

4.6 Approximating the distribution

The approach used above to approximate $\Pr(\hat{\lambda}_{\mathrm{ML}} \leq \lambda; \lambda)$ can be applied to the expressions for the cdf itself, in each interval of its domain. Considering just the case $p = 2$, we simply

[23] The values of z_2 relevant for Figure 8 are -2.0769 when $m_1 = 10$ and $m_2 = 20$, -0.9231 when $m_1 = 5$ and $m_2 = 25$, -0.3659 when $m_1 = 2$ and $m_2 = 28$ (note that z_0 does not depend on r_1 if $r_1 = r_2$).

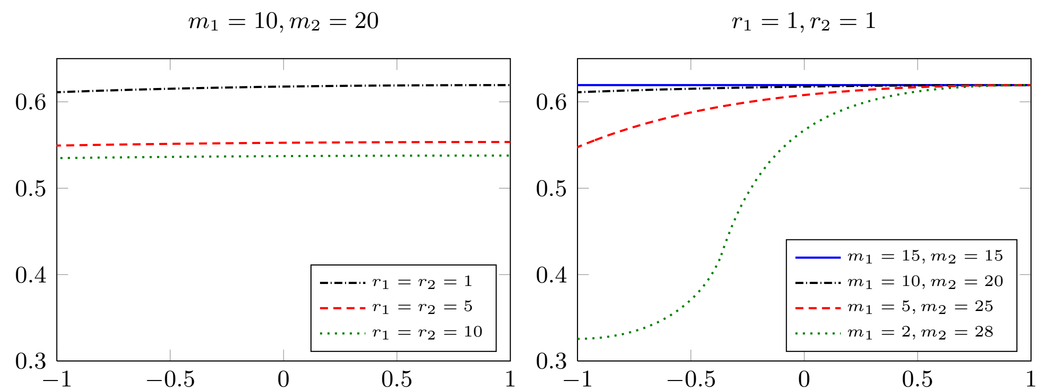

Figure 8: The probability that $\hat{\lambda}_{\mathrm{ML}}$ underestimates λ as a function of λ, in the two-groups case.

need to replace ψ_1 and ψ_2 by $\psi_1(z,\lambda)$ and $\psi_2(z,\lambda)$ in the definitions of π_1, π_2, although, in the case of the distribution function the results are not quite so simple as those given above for $\Pr(\hat{\lambda}_{\mathrm{ML}} \leq \lambda; \lambda)$. The relevant expressions for the cdf are, in the case $p = 2$,

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z; \lambda) \simeq \Pr\left(\mathrm{F}_{v_1(z,\lambda),n_1} \leq u_1(z,\lambda)\right),$$

for $\lambda < z_2$, and

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z; \lambda) \simeq \Pr\left(\mathrm{F}_{r,v_2(z,\lambda)} \leq u_2(z,\lambda)\right),$$

for $\lambda > z_2$, where

$$u_1(z,\lambda) := -\frac{n_1\psi_1(z,\lambda)}{n_2\psi_2(z,\lambda) + r},$$
$$v_1(z,\lambda) := \frac{(n_2\psi_2(z,\lambda) + r)^2}{n_2\psi_2^2(z,\lambda) + r},$$

and

$$u_2(z,\lambda) := \frac{n_1\psi_1(z,\lambda) + n_2\psi_2(z,\lambda)}{n_1\psi_1(z,\lambda) + n_2\psi_2(z,\lambda)},$$
$$v_2(z,\lambda) := \frac{(n_1\psi_1(z,\lambda) + n_2\psi_2(z,\lambda))^2}{n_1\psi_1^2(z,\lambda) + n_2\psi_2^2(z,\lambda)}.$$

Analytic differentiation to obtain the density is messy, but easily accomplished by a symbolic mathematical package, and again can be extended to cases with $p > 2$ without difficulty.

4.7 Group-Specific Regressions

We now consider generalizations to the pure unbalanced Group Interaction model with regressors. Compared to the balanced case, unbalanceness has the favorable consequence that group fixed effects do not render inference on the full parameter impossible.[24]

Similarly to Section 3.6.2, we focus on the case in which all β coefficients are group specific. We show that in this case the cdf of $\hat{\lambda}_{\mathrm{ML}}$ admits a very simple representation when fixed effects are present, regardless of the values of the regressors. Within each group the model is a Balanced group Interaction model, or, stacking groups of same size,

$$y_i = \lambda(I_{r_i} \otimes B_{m_i})y_i + \bigoplus_{j=1}^{r_i} X_{ij}\beta_{ij} + \varepsilon_i,\ i = 1, .., p, \tag{4.24}$$

where y_i is $r_i m_i \times 1$, X_{ij} is an $m_i \times k_{ij}$ matrix containing a column of ones (with $k_{ij} \leq m_i$), and β_i is $\sum_{j=1}^{r_i} k_{ij} \times 1$ (that is, for each of the p distinct group sizes, the model is a balanced model with group specific regressors). This correspond to an unbalanced Group Interaction model with $X = \bigoplus_{i=1}^{p} \bigoplus_{j=1}^{r_i} X_{ij}$, $k = \sum_{i=1}^{p} \sum_{j=1}^{r_i} k_{ij}$, and $\beta' = (\beta'_{11}, .., \beta'_{1r_1}, ..., \beta'_{p1}, .., \beta'_{pr_p})$. By Lemma A.2 in Appendix A if the model contains group fixed effects, then $\mathrm{col}(X)$ is spanned by k eigenvectors of $W = \mathrm{diag}(I_{r_i} \otimes B_{m_i}, i = 1, .., p)$. Then, provided only that $\mathrm{col}(X)$ does not contain all eigenvectors of W associated with eigenvalues other than ω (to avoid degeneracy of the score), by the same argument as in Section 3.6.2 we obtain $\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z; \lambda) = \Pr(\sum_{t=1}^{p} d_{tt}(z, \lambda)\chi^2_{n_t - n_t(X)} \leq 0)$, where the $\chi^2_{n_t - n_t(X)}$ variates are independent, $n_t(X) := \dim(\mathrm{col}(X) \cap \mathrm{col}(I_r \otimes L_{m_t}))$, and we use the convention that $\chi^2_0 = 0$. Using the definition (2.3) of the coefficients $d_{tt}(z, \lambda)$, we have

$$\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z; \lambda) = \Pr\left(\sum_{t=1}^{p}(g_t(z) - \bar{g}(z))\left(\frac{z + m_t - 1}{\lambda + m_t - 1}\right)^2 \chi^2_{n_t - n_t(X)} \leq 0\right), \tag{4.25}$$

where the coefficients $g_t(z) - \bar{g}(z)$ are given in equation (4.3). Representation (4.25) reveals an unexpected property of $\hat{\lambda}_{\mathrm{ML}}$. Specifically, recalling from Section 4.1 that $g_t(z) - \bar{g}(z) < 0$ for any $z \in (z_p, 1)$ and for any $t = 1, ..., p$, representation (4.25) implies that $\Pr(\hat{\lambda}_{\mathrm{ML}} \leq z; \lambda) = 1$ for any $z > z_p$ (recall also that z_p denotes the point at which the coefficient $g_p(z) - \bar{g}(z)$ changes sign). That is, for this model the support of the distribution of $\hat{\lambda}_{\mathrm{ML}}$ is not the entire Λ, but its subset $(-(m_1 - 1), z_p)$.

Similarly to what was done in Section 3.6.2, one can study the distribution of $\hat{\lambda}_{\mathrm{ML}}$ under different asymptotic regimes, but we omit these calculations for the sake of brevity.

5 Concluding Remarks

In Hillier and Martellosio (2013) we presented a general result, equation (2.1) above, giving an expression for the exact distribution function of the quasi-maximum likelihood estimator for λ in equation (1.3), valid for any distribution of ε. Some examples of the application

[24] In the unbalanced case, the columns of the fixed effects matrix span an eigenspace of W (as in the balanced case). However, when $p > 1$, the presence of fixed effects, i.e., $\mathrm{col}(\bigoplus_{i=1}^{p} (I_{r_i} \otimes \iota_{m_i})) \subseteq \mathrm{col}(X)$, does not imply the same degeneracy that occurs when $p = 1$. This is a consequence of the fact that W has more than two eigenspaces when $p > 1$.

of the result to particular cases were given in H&M, but the earlier paper concentrated mainly on its more general consequences. In the present paper we have explored the application of the result to a particular, important, class of models - those based on spatial weights matrices that embody group-interaction. These models are important in various areas of application to the study of networks, and to panels with a spatial autoregressive component. Starting from equation (2.1) we have been able to present a very complete set of results for likelihood-based inference in the pure balanced Group Interaction model under mixed-Gaussian assumptions. We have also been able to generalize these simple results to some special cases of models involving regressors, for example, models with a common mean across all observations, and models with group-specific regressors satisfying certain assumptions.

The pure balanced model is the simplest example of equation (1.3) one can imagine, and the ability to carry out the above program is due to the fact that this model is a regular exponential family. We have then discussed the much more realistic unbalanced model, a model that is considerably more difficult. Again, that is no doubt because unbalanced model is not a regular exponential family, but a curved exponential family in which the dimension of the sufficient statistic is larger than that of the parameter space. Exact results in this model are available in closed form, but are very complex. Thus, in addition to reporting the exact results, we have given some approximations that appear to work well, and which generalize nicely the simpler result for the balanced model. There is more work to be done on the unbalanced model however.

References

Andrews, D.W.K. (1993) Exactly median-unbiased estimation of first-order autoregressive/unit root models, *Econometrica* 61,1, 139-165.

Arnold, S.F., (1979) Linear models with exchangeably distributed errors, *Journal of the American Statistical Association* 74, 194-199.

Barndorff-Nielsen, O., (1980) Conditionality resolutions, *Biometrika* 67, 293-310.

Buckley, M.J., and Eagleson, G.K. (1988) An approximation to the distribution of quadratic forms in normal random variables, *Australian Journal of Statistics* 30A, 150-159.

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

Case, A. (1992) Neighborhood influence and technological change, *Regional Science and Urban Economics* 22, 491-508.

de Paula (2016) Econometrics of network models, Cemmap working paper CWP06/16.

Groeneveld, R.A. and Meeden, G. (1977) The mode, median, mean inequality. *The American Statistician* 31, 120–121

Hall, P., (1983) Chi squared approximations to the distribution of a sum of independent random variables, *The Annals of Probability* 11, 4, 1028-1036.

Hillier, G.H. and Martellosio, F. (2013) Properties of the maximum likelihood estimator in spatial autoregressive models. Cemmap Working Papers, CWP44/13.

Hsieh, C.-S., Lee, L.F. (2016) A social interactions model with endogenous friendship formation and selectivity, *Journal of Applied Econometrics* 31, 301–319.

James, A.T. (1964) Distributions of matrix variates and latent roots derived from normal samples, *Annals of Mathematical Statistics* 35, 475–501.

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994) *Distributions in Statistics—Continuous Univariate Distributions, Vol. 1 (Second Edition).* New York: John Wiley & Sons.

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

King, M.L. (1981) A small sample property of the Cliff-Ord test for spatial autocorrelation. *Journal of the Royal Statistical Society B* 43, 263-4.

Kyriakou, M., Phillips, P.C.B., and Rossi, F. (2014) Indirect inference in spatial autoregression, Manuscript, University of Southampton.

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.

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

Moffitt, R. (2001) Policy interventions, low-level equilibria and social interactions, in Social Dynamics, ed. by S. Durlauf, and P. Young. MIT Press.

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

Ruben, H. (1962) Probability content of regions under spherical normal distributions, IV: the distribution of homogeneous and non-homogeneous quadratic functions of normal variables, *The Annals of Mathematical Statistics* 33, 542-570.

Appendix A Proofs and Auxiliary Results

Lemma A.1. $\mathrm{med}(\mathrm{F}_{p,q}) = 1$ *if and only if* $p = q$ *and* $\mathrm{med}(\mathrm{F}_{p,q}) < 1$ *if* $p < q$.

Proof. The first part of the lemma is straightforward, because $\mathrm{F}_{p,q} = 1/\mathrm{F}_{q,p}$ implies that $\mathrm{med}(\mathrm{F}_{p,q})\mathrm{med}(\mathrm{F}_{q,p}) = 1$, and hence that $\mathrm{med}(\mathrm{F}_{p,q}) = 1$ if $p = q$. Moving to the second part, $\mathrm{med}(\mathrm{F}_{p,q}) < 1$ if and only if $\Pr(\mathrm{F}_{p,q} < 1) > 1/2$. Using the well-known relationship between the cdf's of the F and beta distributions, $\Pr(\mathrm{F}_{p,q} < 1) = \Pr(\mathrm{Beta}(p/2, q/2) < p/(p+q))$, where $\mathrm{Beta}(p/2, q/2)$ is a beta random variable. But note that $p/(p+q)$ is the mean of $\mathrm{Beta}(p/2, q/2)$. Thus, $\mathrm{med}(\mathrm{F}_{p,q}) < 1$ if and only if

$$\Pr\left(\mathrm{Beta}\left(\frac{p}{2}, \frac{q}{2}\right) < \mathrm{E}\left[\mathrm{Beta}\left(\frac{p}{2}, \frac{q}{2}\right)\right]\right) > 1/2,$$

that is, if and only if $\mathrm{med}(\mathrm{Beta}(p/2, q/2)) < \mathrm{E}[\mathrm{Beta}(p/2, q/2)]$. For the beta distribution the median is smaller than the mean if and only the skewness is positive (e.g., Groeneveld and Meeden, 1977). The desired result follows, because the skewness of $\mathrm{Beta}(p/2, q/2)$ is positive if and only if $p < q$. □

Lemma A.2. *Let A_i, $i = 1, ..., t$, be $m_i \times n_i$ matrices. If $\iota_{m_i} \in \mathrm{col}(A_i)$ for each $i = 1, ..., t$, then $\mathrm{col}(\bigoplus_{i=1}^t A_i)$ is spanned by $\sum_{i=1}^t n_i$ eigenvectors of $\mathrm{diag}(\iota_{m_i}\iota_{m_i}' - I_{m_i}, i = 1, .., t)$.*

Proof. If $\iota_{m_i} \in \mathrm{col}(A_i)$ for each $i = 1, ..., t$, then the t columns of $\bigoplus_{i=1}^t \iota_{m_i}$ and the $\sum_{i=1}^t (n_i) - t$ columns of $\bigoplus_{i=1}^t O_i$, where O_i is an $m_i \times (n_i - 1)$ matrix with $\mathrm{col}(O_i) \subset \mathrm{col}^{\perp}(\iota_{m_i})$, form an orthogonal basis for $\mathrm{col}(\bigoplus_{i=1}^t A_i)$. But these $\sum_{i=1}^t n_i$ columns are orthogonal eigenvectors of $\mathrm{diag}(\iota_{m_i}\iota_{m_i}' - I_{m_i}, i = 1, .., t)$ (see footnote 19). □

Proof of Proposition 3.3. By Lemma A.1, $\mathrm{med}(\mathrm{F}_{r,r(m-1)}) \leq 1$, with equality if and only if $m = 2$. Using (3.9), it follows that $\mathrm{med}(\hat{\lambda}_{\mathrm{ML}}) \leq \lambda$, with equality if and only if $m = 2$, thus establishing part (i). Part (ii) follows immediately from (3.10). To prove part (iii), note that the function $b_{\mathrm{med}}(\lambda)$ is continuous over Λ, with

$$\frac{db_{\mathrm{med}}(\lambda)}{d\lambda} = \frac{m\left(1 - \lambda + \zeta_{r,m}^2(\lambda + m - 1)\right) + m(1-\lambda)\left(\zeta_{r,m}^4 - 1\right)}{\left(1 - \lambda + \zeta_{r,m}^2(\lambda + m - 1)\right)^2} - 1,$$

and

$$\frac{d^2 b_{\mathrm{med}}(\lambda)}{d\lambda^2} = -\frac{2m^2(\zeta_{r,m}^2 - 1)\zeta_{r,m}^2}{(\zeta_{r,m}^2(\lambda + m - 1) + 1 - \lambda)^3}.$$

Clearly, $d^2 b_{\mathrm{med}}(\lambda)/d\lambda^2 > 0$ for any $\lambda \in \Lambda$, because $\zeta_{r,m} < 1$ if $m > 2$ by Lemma A.1. Solving $db_{\mathrm{med}}(\lambda)/d\lambda = 0$ gives two critical points, one inside Λ and one outside. The one inside Λ is $\lambda = (1 - (m-1)\zeta_{r,m})/(1 + \zeta_{r,m})$. □

Proof of Proposition 3.6. From (3.15),

$$\mathrm{E}(\hat{\theta}_{\mathrm{ML}}^s) = \frac{\tau^s}{B\left(\frac{r}{2}, \frac{r(m-1)}{2}\right)} \int_0^\infty f^{\frac{r+s}{2} - 1}(1+f)^{-\frac{rm}{2}}\, df = \frac{\tau^s \Gamma(\frac{r+s}{2})\Gamma(\frac{r(m-1)-s}{2})}{\Gamma(\frac{r}{2})\Gamma(\frac{r(m-1)}{2})},$$

provided $s < r(m-1)$. □

Proof of Proposition 4.1. Let $q_i \sim \chi^2_{v_i}$, $i = 1, 2$, assumed independent, and let $q = a_1 q_1 + a_2 q_2$, with $0 < a_1 < a_2$. In the joint density of (q_1, q_2), transform to $x_1 := a_1 q_1, x_2 := a_2 q_2$. The Jacobian is $(a_1 a_2)^{-1}$, so

$$\mathrm{pdf}(x_1, x_2) = \frac{\exp\left\{-\frac{1}{2}\left(\frac{x_1}{a_1} + \frac{x_2}{a_2}\right)\right\} x_1^{\frac{v_1}{2}-1} x_2^{\frac{v_2}{2}-1}}{a_1^{\frac{v_1}{2}} a_2^{\frac{v_2}{2}} 2^{\frac{v_1+v_2}{2}} \Gamma(\frac{v_1}{2})\Gamma(\frac{v_2}{2})}.$$

Now transform to $q = x_1 + x_2, b = x_1/(x_1 + x_2)$, $0 < b < 1$, so that $x_1 = bq, x_2 = (1-b)q$, and the Jacobian is q. Then,

$$\mathrm{pdf}(q, b) = \frac{\exp\left\{-\frac{1}{2}\left(\frac{q}{a_1} - \frac{(1-b)q}{a_1} + \frac{(1-b)q}{a_2}\right)\right\} q^{\frac{v_1+v_2}{2}-1} b^{\frac{v_1}{2}-1}(1-b)^{\frac{v_2}{2}-1}}{a_1^{\frac{v_1}{2}} a_2^{\frac{v_2}{2}} 2^{\frac{v_1+v_2}{2}} \Gamma(\frac{v_1}{2})\Gamma(\frac{v_2}{2})}.$$

Integrating out b is straightforward, giving the sought-after density:

$$\begin{aligned}
\mathrm{pdf}(q) &= \frac{\exp\left(-\frac{q}{2a_1}\right) q^{\frac{v}{2}-1}}{a_1^{\frac{v_1}{2}} a_2^{\frac{v_2}{2}} 2^{\frac{v}{2}} \Gamma(\frac{v}{2})} \sum_{j=0}^{\infty} \frac{\left(\frac{q}{2a_1}\left(1 - \frac{a_1}{a_2}\right)\right)^j}{j!} \frac{\left(\frac{v_2}{2}\right)_j}{\left(\frac{v}{2}\right)_j} \\
&= \frac{\exp\left(-\frac{q}{2a_1}\right) q^{\frac{v}{2}-1}}{a_1^{\frac{v_1}{2}} a_2^{\frac{v_2}{2}} 2^{\frac{v}{2}} \Gamma(\frac{v}{2})} {}_1F_1\left(\frac{v_2}{2}, \frac{v}{2}; \frac{1}{2a_1} q \left(1 - \frac{a_1}{a_2}\right)\right),
\end{aligned}$$

where $v := v_1 + v_2$. Putting $\phi = 1/a_1$, $\psi := a_1/a_2$, we have

$$\mathrm{pdf}(q) = \frac{\phi^{\frac{v}{2}} \psi^{\frac{v_2}{2}} \exp\left(-\frac{\phi q}{2}\right) q^{\frac{v}{2}-1}}{2^{\frac{v}{2}} \Gamma(\frac{v}{2})} {}_1F_1\left(\frac{v_2}{2}, \frac{v}{2}; \frac{1}{2}\phi q (1-\psi)\right).$$

□

Proof of Lemma 4.2. A generating function for $C_j(A)$ is

$$|I - tA|^{-\frac{1}{2}} = \sum_{j=0}^{\infty} \frac{t^j \left(\frac{1}{2}\right)_j}{j!} C_j(A).$$

But, when A has the form assumed, the left-hand side is

$$\begin{aligned}
(1 - ta_1)^{-\frac{n_1}{2}} (1 - ta_2)^{-\frac{n_2}{2}} &= \sum_{j,k=0}^{\infty} \frac{t^{j+k} \left(\frac{n_1}{2}\right)_j \left(\frac{n_2}{2}\right)_k}{j!k!} a_1^j a_2^k \\
&= \sum_{j=0}^{\infty} \frac{t^j}{j!} \left(\sum_{k=0}^{j} \binom{j}{k} \left(\frac{n_1}{2}\right)_k \left(\frac{n_2}{2}\right)_{j-k} a_1^k a_2^{j-k}\right).
\end{aligned}$$

Equating coefficients of $t^j/j!$ gives the result. □

Proof of Proposition 4.3. In addition to proving Proposition 4.3, here we also derive the corresponding formulae for odd γ (cases (iii) and (iv) below). The conditional density of w given (q_1, q_2) is

$$\mathrm{pdf}_w(z|q_1,q_2) = \frac{\exp\left\{-\frac{1}{2}(aq_1+cq_2)\right\}}{2^{\frac{\gamma}{2}}\Gamma(\frac{\gamma}{2})}(\dot{a}q_1+\dot{c}q_2)(aq_1+cq_2)^{\frac{\gamma}{2}-1}. \tag{A.1}$$

Multiplying by the joint density of (q_1, q_2), and transforming to $x_1 := (1+a)q_1, x_2 := (1+c)q_2$ gives

$$\mathrm{pdf}_w(z,x_1,x_2) = \frac{\exp\left\{-\frac{1}{2}(x_1+x_2)\right\}}{2^{\frac{\gamma+\alpha+\beta}{2}}\Gamma\left(\frac{\gamma}{2}\right)\Gamma\left(\frac{\alpha}{2}\right)\Gamma\left(\frac{\beta}{2}\right)(1+a)^{\frac{\alpha}{2}}(1+c)^{\frac{\beta}{2}}} x_1^{\frac{\alpha}{2}-1}x_2^{\frac{\beta}{2}-1}$$
$$\times\left(\frac{\dot{a}}{1+a}x_1+\frac{\dot{c}}{1+c}x_2\right)\left(\frac{a}{1+a}x_1+\frac{c}{1+c}x_2\right)^{\frac{\gamma}{2}-1}.$$

(i) $\gamma = 2$. In this case the last term is not present and, on integrating out x_1, x_2, we obtain simply

$$\mathrm{pdf}_w(z;\alpha,\beta,2) = \frac{\alpha\frac{\dot{a}}{1+a}+\beta\frac{\dot{b}}{1+b}}{2(1+a)^{\frac{\alpha}{2}}(1+b)^{\frac{\beta}{2}}}.$$

(ii) $\gamma = 2s+2$. When $\gamma = 2s+2$ the final term has the binomial expansion

$$\sum_{j=0}^{s}\binom{s}{j}\left(\frac{a}{1+a}\right)^j\left(\frac{c}{1+c}\right)^{s-j}x_1^j x_2^{s-j}.$$

The term with coefficient $\frac{\dot{a}}{1+a}$ is then

$$\frac{\alpha}{2s!(1+a)^{\frac{\alpha}{2}}(1+c)^{\frac{\beta}{2}}}\sum_{j=0}^{s}\binom{s}{j}\left(\frac{a}{1+a}\right)^j\left(\frac{c}{1+c}\right)^{s-j}\left(\frac{\alpha+2}{2}\right)_j\left(\frac{\beta}{2}\right)_{s-j}$$
$$= \frac{\left(\frac{1}{2}\right)_s}{s!(1+a)^{\frac{\alpha}{2}}(1+c)^{\frac{\beta}{2}}}\frac{\alpha C_s\left(A_{\alpha+2,\beta}\left(\frac{a}{1+a},\frac{c}{1+c}\right)\right)}{2}$$

on using Lemma 4.2. The other term is exactly analogous, and we find, for the case $\gamma = 2s+2$,

$$\mathrm{pdf}_w(z) = \frac{\left(\frac{1}{2}\right)_s}{2s!(1+a)^{\frac{\alpha}{2}}(1+c)^{\frac{\beta}{2}}}\left(\frac{\dot{a}\alpha C_s\left(A_{\alpha+2,\beta}(\frac{a}{1+a},\frac{c}{1+c})\right)}{1+a}+\frac{\dot{c}\beta C_s\left(A_{\alpha,\beta+2}(\frac{a}{1+a},\frac{c}{1+c})\right)}{1+c}\right).$$

(iii) $\gamma = 1$. Starting from equation (A.1) with $\gamma = 1$, and expressing the final term in the form

$$(aq_1+cq_2)^{-\frac{1}{2}} = \frac{1}{\sqrt{2\pi}}\int_{x>0}\exp\{-\frac{1}{2}x(aq_1+cq_2)\}x^{-\frac{1}{2}}\,dx,$$

we have, on integrating out (q_1, q_2), for the first term

$$\frac{\alpha\dot{a}}{2\Gamma(\frac{1}{2})\Gamma(\frac{1}{2})(1+a)^{-\frac{\alpha+2}{2}}(1+c)^{-\frac{\beta}{2}}}\int_{x>0} x^{-\frac{1}{2}}\left(1+\frac{ax}{1+a}\right)^{-\frac{\alpha+2}{2}}\left(1+\frac{cx}{1+c}\right)^{-\frac{\beta}{2}} dx. \tag{A.2}$$

Transforming to $b := x/(1+x)$, the integral in (A.2) becomes

$$\begin{aligned}
&\int_{0<b<1} b^{-\frac{1}{2}}(1-b)^{\frac{\alpha+\beta+1}{2}-1}\left(1-\frac{b}{1+a}\right)^{-\frac{\alpha+2}{2}}\left(1-\frac{b}{1+c}\right)^{-\frac{\beta}{2}} db \\
&= \frac{\Gamma\left(\frac{\alpha+\beta+1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{\alpha+\beta+2}{2}\right)}\sum_{i,j=0}^{\infty}\frac{\left(\frac{1}{2}\right)_{i+j}\left(\frac{\alpha+2}{2}\right)_i\left(\frac{\beta}{2}\right)_j}{i!j!(1+a)^i(1+c)^j\left(\frac{\alpha+\beta+2}{2}\right)_{i+j}} \\
&= \frac{\Gamma\left(\frac{\alpha+\beta+1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{\alpha+\beta+2}{2}\right)}\sum_{j=0}^{\infty}\frac{\left(\frac{1}{2}\right)_j}{j!\left(\frac{\alpha+\beta+2}{2}\right)_j}\sum_{i=0}^{j}\binom{j}{i}\frac{\left(\frac{\alpha+2}{2}\right)_i\left(\frac{\beta}{2}\right)_{j-i}}{(1+a)^i(1+c)^{j-i}} \\
&= \frac{\Gamma\left(\frac{\alpha+\beta+1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{\alpha+\beta+2}{2}\right)}\sum_{j=0}^{\infty}\frac{\left(\frac{1}{2}\right)_j\left(\frac{1}{2}\right)_j}{j!\left(\frac{\alpha+\beta+2}{2}\right)_j}C_j\left(A_{\alpha+2,\beta}\left(\frac{1}{1+a},\frac{1}{1+c}\right)\right),
\end{aligned}$$

and hence (A.2) is

$$\frac{\Gamma\left(\frac{\alpha+\beta+1}{2}\right)}{\Gamma\left(\frac{\alpha+\beta+2}{2}\right)\Gamma\left(\frac{1}{2}\right)(1+a)^{\frac{\alpha}{2}}(1+c)^{\frac{\beta}{2}}}\sum_{j=0}^{\infty}\frac{\left(\frac{1}{2}\right)_j\left(\frac{1}{2}\right)_j}{j!\left(\frac{\alpha+\beta+2}{2}\right)_j}\frac{\alpha\dot{a}C_j\left(A_{\alpha+2,\beta}\left(\frac{1}{1+a},\frac{1}{1+c}\right)\right)}{2(1+a)}.$$

The validity of the series expansions used for the Bessel functions $(1-b/(1+a))^{-\alpha/2}$ and $(1-b/(1+c))^{-\beta/2}$, as well as of the term-by-term integration involved, are readily confirmed (because $1/(1+a)$ and $1/(1+c)$ are both between 0 and 1). The second term is exactly analogous, and we find

$$\begin{aligned}
\mathrm{pdf}_w(z;\alpha,\beta,1) &= \frac{\Gamma\left(\frac{\alpha+\beta+1}{2}\right)}{\Gamma\left(\frac{\alpha+\beta+2}{2}\right)\Gamma\left(\frac{1}{2}\right)(1+a)^{\frac{\alpha}{2}}(1+c)^{\frac{\beta}{2}}}\sum_{j=0}^{\infty}\frac{\left(\frac{1}{2}\right)_j\left(\frac{1}{2}\right)_j}{j!\left(\frac{\alpha+\beta+2}{2}\right)_j} \\
&\quad\times\left(\frac{\alpha\dot{a}C_j\left(A_{\alpha+2,\beta}\left(\frac{1}{1+a},\frac{1}{1+c}\right)\right)}{2(1+a)}+\frac{\beta\dot{c}C_j\left(A_{\alpha,\beta+2}\left(\frac{1}{1+a},\frac{1}{1+c}\right)\right)}{2(1+c)}\right).
\end{aligned} \tag{A.3}$$

(iv) $\gamma = 2s+1$. In this case we have

$$\mathrm{pdf}_w(z|q_1,q_2) = \frac{\exp\left\{-\frac{1}{2}(aq_1+cq_2)\right\}}{2^{\frac{\gamma}{2}}\Gamma\left(\frac{\gamma}{2}\right)}\frac{(\dot{a}q_1+\dot{c}q_2)(aq_1+cq_2)^s}{(aq_1+bq_2)^{\frac{1}{2}}}.$$

After expanding the term $(aq_1 + cq_2)^s$ binomially we can proceed as for the case $\gamma = 1$ above, replacing α by $\alpha + 2i$, and β by $\beta + 2(s - i)$. The result is:

$$\begin{aligned}\mathrm{pdf}_w(z; \alpha, \beta, 2s+1) = \frac{1}{2^s \left(\frac{1}{2}\right)_s} \sum_{i=0}^{s} \binom{s}{i} a^i c^{s-i} \left(\frac{\alpha}{2}\right)_i \left(\frac{\beta}{2}\right)_{s-i} \\ \mathrm{pdf}_w(z; \alpha + 2i, \beta + 2(s-i); 1). \qquad \text{(A.4)}\end{aligned}$$

□

Lemma A.3. *If, in the same context as Proposition 4.3, $a(1+c) \leq 2c(1+a)$ for all $z \in \Lambda$, then the results in Proposition 4.3 can be written more simply as*

$$\begin{aligned}\mathrm{pdf}_w(z; \alpha, \beta, \gamma) = \frac{1}{(\alpha+\beta) B\left(\frac{\gamma}{2}, \frac{\alpha+\beta}{2}\right)} \frac{a^{\frac{\gamma+\beta}{2}}}{c^{\frac{\beta}{2}} (1+a)^{\frac{\alpha+\beta+\gamma}{2}}} \\ \times \left\{ \frac{\dot{a}\alpha}{a} {}_2F_1\left(\frac{\alpha+\beta+\gamma}{2}, \frac{\beta}{2}, \frac{\alpha+\beta+2}{2}, \eta\right) \right. \\ \left. + \frac{\dot{c}\beta}{c} {}_2F_1\left(\frac{\alpha+\beta+\gamma}{2}, \frac{\beta+2}{2}, \frac{\alpha+\beta+2}{2}, \eta\right) \right\},\end{aligned}$$

where $\eta := 1 - a(1+c)/\left(c(1+a)\right)$.

Proof. Multiplying the conditional density (A.1) by the joint density of (q_1, q_2) and transforming to $(x_1, x_2) := ((1+a)q_1, (1+c)q_2)$ gives

$$\begin{aligned}\mathrm{pdf}_Z(z, x_1, x_2) = \frac{\exp\left\{-\frac{1}{2}(x_1 + x_2)\right\}}{2^{\frac{\gamma+\alpha+\beta}{2}} \Gamma(\frac{\gamma}{2}) \Gamma(\frac{\alpha}{2}) \Gamma(\frac{\beta}{2}) (1+a)^{\frac{\alpha}{2}} (1+c)^{\frac{\beta}{2}}} x_1^{\frac{\alpha}{2}-1} x_2^{\frac{\beta}{2}-1} \\ \times \left(\frac{\dot{a}}{1+a} x_1 + \frac{\dot{c}}{1+c} x_2\right) \left(\frac{a}{1+a} x_1 + \frac{c}{1+c} x_2\right)^{\frac{\gamma}{2}-1} \qquad \text{(A.5)}\end{aligned}$$

Note that if $\gamma = 2$ the last term is not present and, on integrating out x_1, x_2, we obtain the result given in Proposition 4.3. For the general case, transforming to $q := x_1 + x_2$ and $b := x_1/q$, and integrating out q gives

$$\begin{aligned}\mathrm{pdf}_Z(z, b) = \frac{\Gamma(\frac{\alpha+\beta+\gamma}{2}) \left(\frac{c}{1+c}\right)^{\frac{\gamma}{2}-1}}{\Gamma(\frac{\gamma}{2}) \Gamma(\frac{\alpha}{2}) \Gamma(\frac{\beta}{2}) (1+a)^{\frac{\alpha}{2}} (1+c)^{\frac{\beta}{2}}} \\ \times \left(\frac{\dot{a}b}{1+a} + \frac{\dot{c}(1-b)}{1+c}\right) b^{\frac{\alpha}{2}-1} (1-b)^{\frac{\beta}{2}-1} (1-\eta b)^{\frac{\gamma}{2}-1}. \qquad \text{(A.6)}\end{aligned}$$

Provided $|\eta| < 1$, integrating out b in the second line of the last display gives two terms (ignoring the first line for the moment):

$$\frac{\dot{a}}{1+a} \frac{\Gamma(\frac{\beta}{2}) \Gamma(\frac{\alpha+2}{2})}{\Gamma(\frac{\alpha+\beta+2}{2})} \left(\frac{a(1+c)}{c(1+a)}\right)^{\frac{\gamma+\beta}{2}-1} {}_2F_1\left(\frac{\alpha+\beta+\gamma}{2}, \frac{\beta}{2}; \frac{\alpha+\beta+2}{2}; \eta\right)$$

and

$$\frac{\dot{c}}{1+c}\frac{\Gamma(\frac{\beta+2}{2})\Gamma(\frac{\alpha}{2})}{\Gamma(\frac{\alpha+\beta+2}{2})}\left(\frac{a(1+c)}{c(1+a)}\right)^{\frac{\gamma+\beta}{2}} {}_2F_1\left(\frac{\alpha+\beta+\gamma}{2},\frac{\beta+2}{2};\frac{\alpha+\beta+2}{2};\eta\right).$$

Simplifying the entire expression gives the result stated. □

Appendix B Additional Figures

Bias, RMSE, and Median of Bias Corrected Estimators

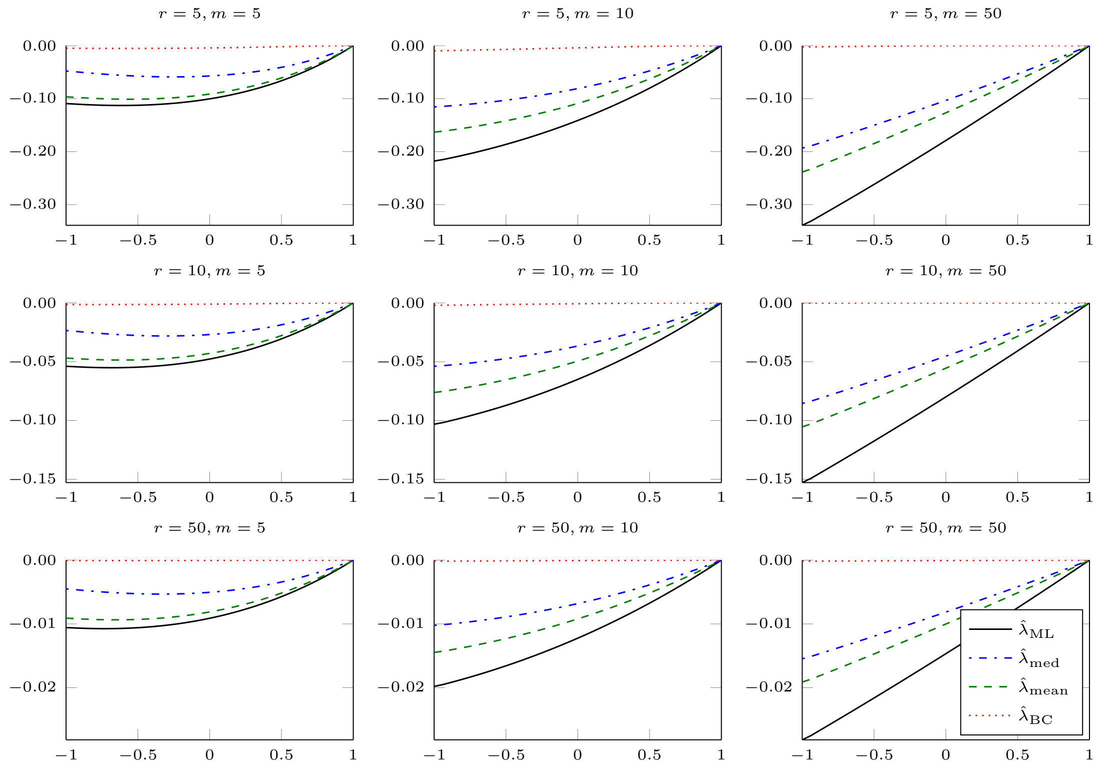

Figure 9: Bias function of the MLE ($\hat{\lambda}_{\mathrm{ML}}$), the median unbiased estimator ($\hat{\lambda}_{\mathrm{med}}$), the indirect estimator obtained by inverting the mean function ($\hat{\lambda}_{\mathrm{mean}}$), and the direct bias corrected MLE ($\hat{\lambda}_{\mathrm{BC}}$).

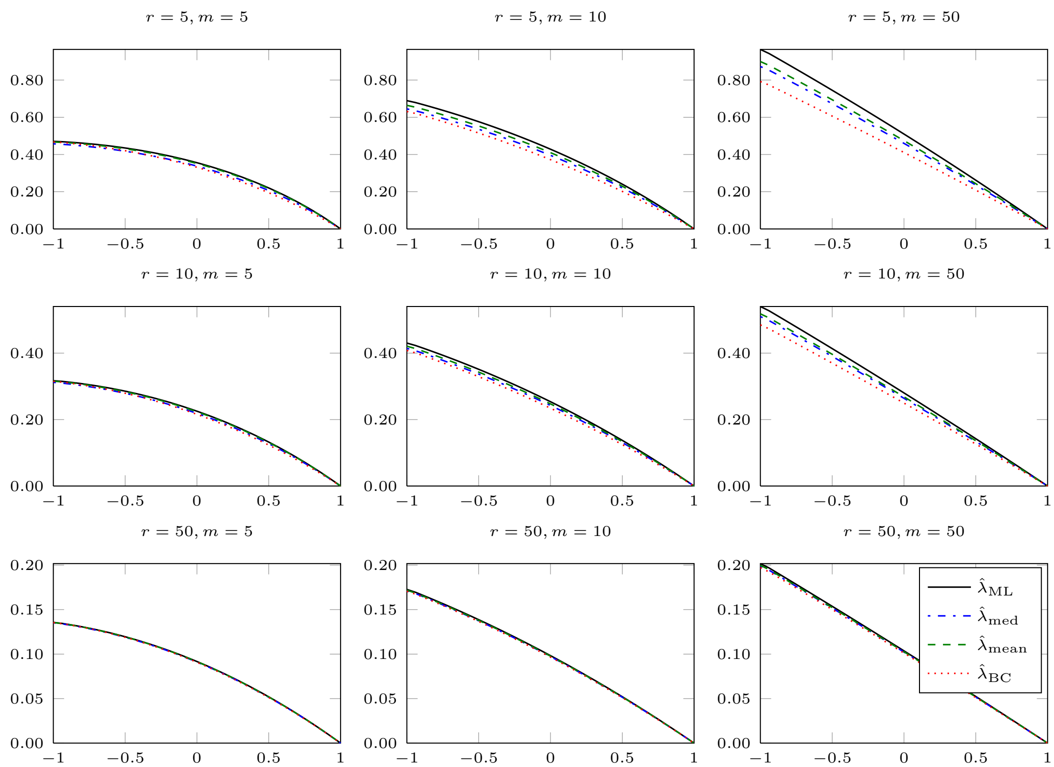

Figure 10: RMSE function of the MLE ($\hat{\lambda}_{\mathrm{ML}}$), the median unbiased estimator ($\hat{\lambda}_{\mathrm{med}}$), the indirect estimator obtained by inverting the mean function ($\hat{\lambda}_{\mathrm{mean}}$), and the direct bias corrected MLE ($\hat{\lambda}_{\mathrm{BC}}$).

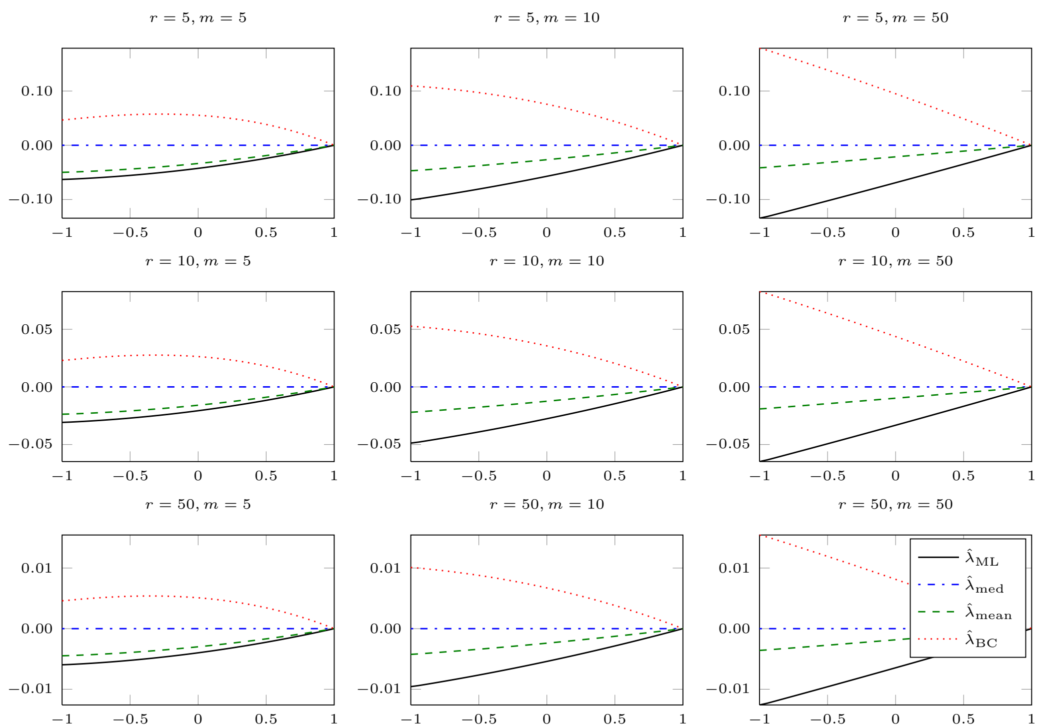

Figure 11: Median function of the MLE ($\hat{\lambda}_{\mathrm{ML}}$), the median unbiased estimator ($\hat{\lambda}_{\mathrm{med}}$), the indirect estimator obtained by inverting the mean function ($\hat{\lambda}_{\mathrm{mean}}$), and the direct bias corrected MLE ($\hat{\lambda}_{\mathrm{BC}}$).

Densities in the Unbalanced Model with Two Groups Figures 12 and 13 complement Figure 7 in the paper. They were obtained using the results given in both Proposition 4.3 in the text, and Lemma A.3 in Appendix A. Each of the three rows of Figure 12 displays $\mathrm{pdf}_{\hat{\lambda}_{\mathrm{ML}}}(z;\lambda)$ for a fixed value of m_1 and varying n, while Figure 13 displays $\mathrm{pdf}_{\hat{\lambda}_{\mathrm{ML}}}(z;\lambda)$ for fixed n and varying m_1. For convenience, all densities are plotted on $(-2,1) \subset \Lambda = (-(m_1-1),1)$. Recall that as long as the model is unbalanced, there is a point $z_2 \in \Lambda$ where the density of $\hat{\lambda}_{\mathrm{ML}}$ in nonanalytic, whatever the sample size n. Graphically, nonanalyticity is clearly visible only for small m_1; at $m_1 = 6$ it is already difficult to detect.

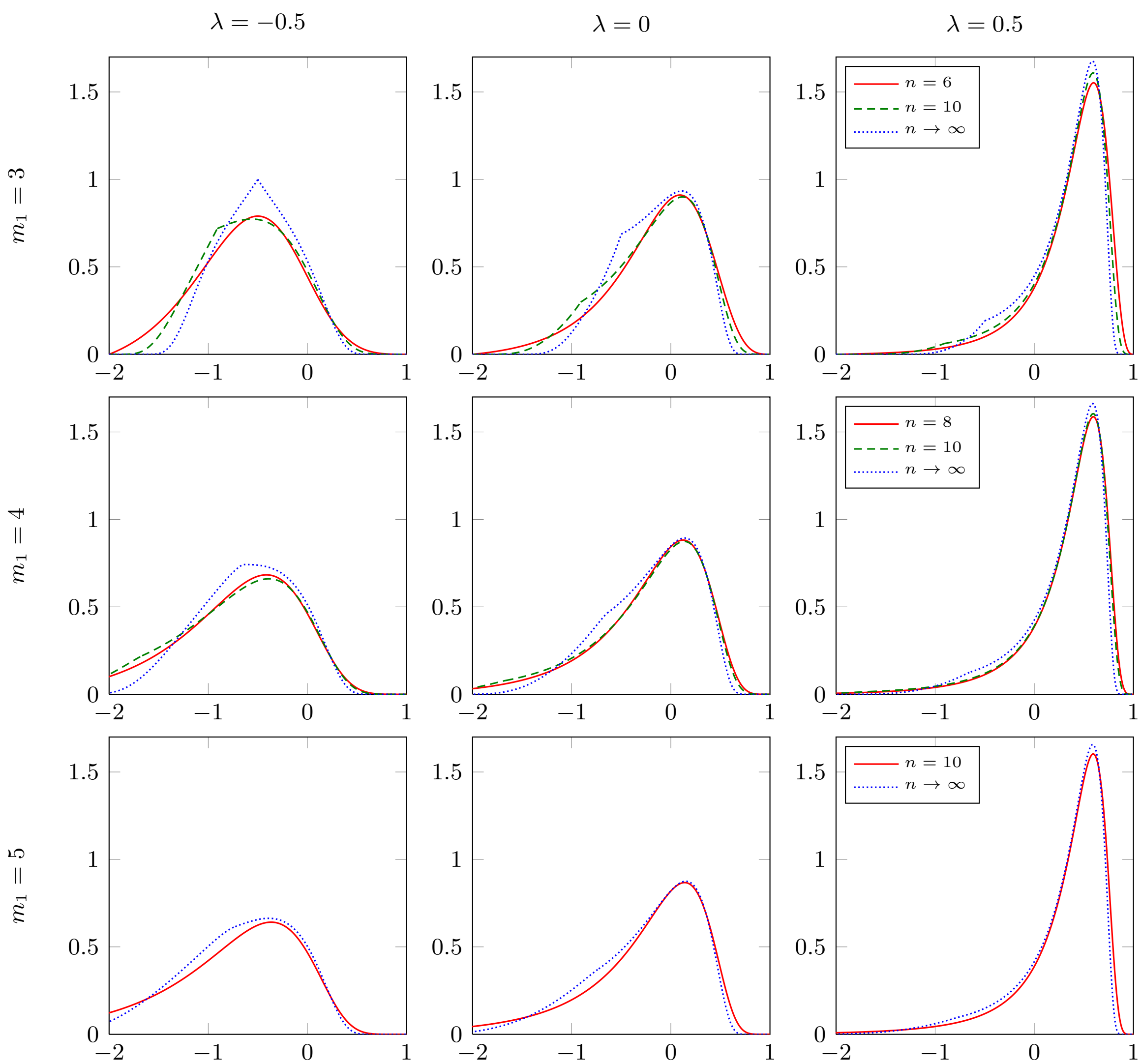

Figure 12: Density of $\hat{\lambda}_{\mathrm{ML}}$ for pure Group Interaction model with two groups, when $\varepsilon \sim \mathrm{SMN}(0, I_n)$.

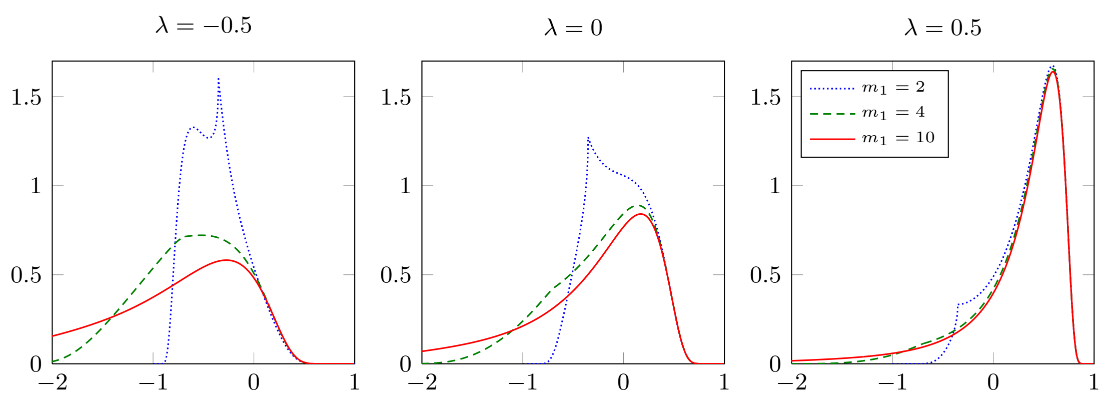

Figure 13: Density of $\hat{\lambda}_{\mathrm{ML}}$ for pure Group Interaction model with two groups and $n = 25$, when $\varepsilon \sim \mathrm{SMN}(0, I_n)$.