Online Supplement to:

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

November 2 2016

Introduction

This supplement contains technical material and figures that accompany the main paper. Appendices A and B provide, respectively, supplementary material for Sections 3 and 4 in the main paper.

Appendix A The Balanced Model

Appendix A.1 Median Bias

Proposition A.1. *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*

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

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

Proof. By Lemma A.2, given at the end of this section, $\mathrm{med}(\mathrm{F}_{r,r(m-1)}) \leq 1$, with equality if and only if $m = 2$. Using (3.7), 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 (3.8). To prove part (iii), note that the function $b_{\text{med}}(\lambda)$ is continuous over Λ, with

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

and

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

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

Note that, since $\zeta_{r,m} > 0$, the point of maximum (A.1) is negative for any r and for any $m > 2$. It is also worth observing that, in terms of the parameter θ, the point of maximum is $\theta = 1/\zeta_{r,m}$.

Figure 1 displays the exact median bias and the large-m median bias of $\hat{\lambda}_{\text{ML}}$ in a Gaussian pure balanced Group Interaction model, obtained from Proposition A.1, 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.

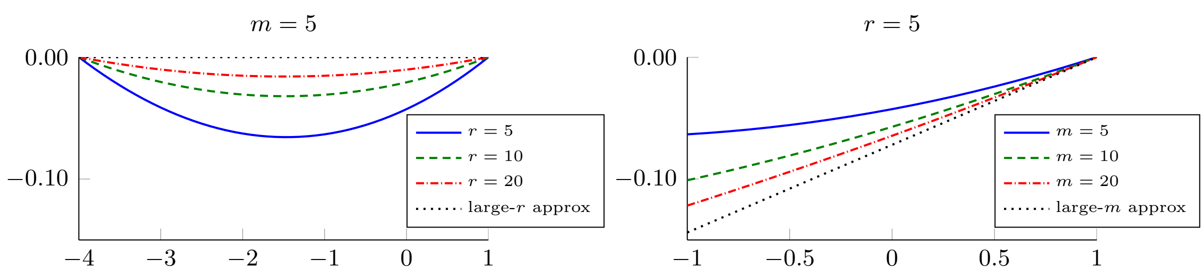

Figure 1: Median bias of $\hat{\lambda}_{\text{ML}}$ for the pure balanced Group Interaction model with $\varepsilon \sim \text{SMN}(0, I_n)$.

Lemma A.2. $\text{med}(\text{F}_{p,q}) = 1$ *if and only if* $p = q$ *and* $\text{med}(\text{F}_{p,q}) < 1$ *if* $p < q$.

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

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

that is, if and only if $\operatorname{med}(\text{Beta}(p/2, q/2)) < \text{E}[\text{Beta}(p/2, q/2)]$. 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 $\text{Beta}(p/2, q/2)$ is positive if and only if $p < q$. □

Appendix A.2 Confidence Intervals

Here we provide two figures that illustrate the properties of the exact confidence intervals introduced in the paper. Figure 2 plots some confidence intervals (3.10), as a function of the observed $\hat{\lambda}_{\text{ML}}$, for $\hat{\lambda}_{\text{ML}} \in \Lambda$, 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 $\hat{\lambda}_{\text{ML}}$ (dotted 45 degree line) as r increases.

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

$$\left(\hat{\lambda}_{\text{ML}} - c_\alpha \sqrt{v_{\hat{\lambda}_{\text{ML}}}}, \hat{\lambda}_{\text{ML}} + c_\alpha \sqrt{v_{\hat{\lambda}_{\text{ML}}}}\right), \tag{A.2}$$

where $v_{\hat{\lambda}_{\text{ML}}}$ is the large-r variance (3.5) evaluated at the MLE, and c_α is the appropriate critical value from the standard normal distribution. Figures 3 compares the confidence intervals (A.2) with the exact confidence intervals (3.10), when $r = 5$, for $m = 5, 50$, and for $\hat{\lambda}_{\text{ML}} \in (-1, 1)$. The general conclusion from this plot, and from similar ones that we do not report, is that, as long as $r > 1$, the two-sided asymptotic confidence intervals provide a good approximation to the equal-tailed exact ones if $\hat{\lambda}_{\text{ML}} \in (-1, 1)$. The large-$r$ approximation may be inaccurate for smaller values of $\hat{\lambda}_{\text{ML}}$, but such values of $\hat{\lambda}_{\text{ML}}$ are rare in applications.

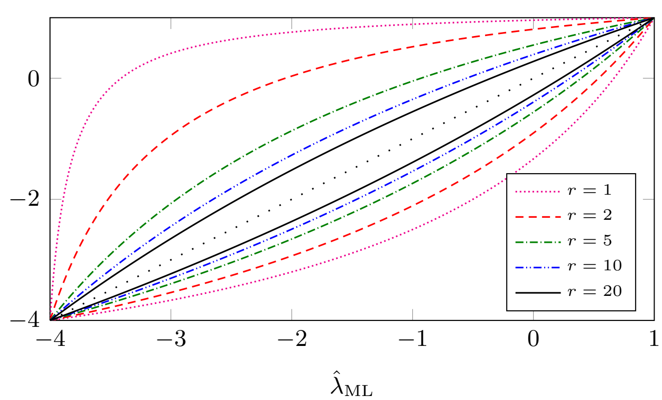

Figure 2: Exact equal-tailed 95% confidence belts for λ in the pure balanced Group Interaction model with $\varepsilon \sim \text{SMN}(0, I_n)$, when $m = 5$.

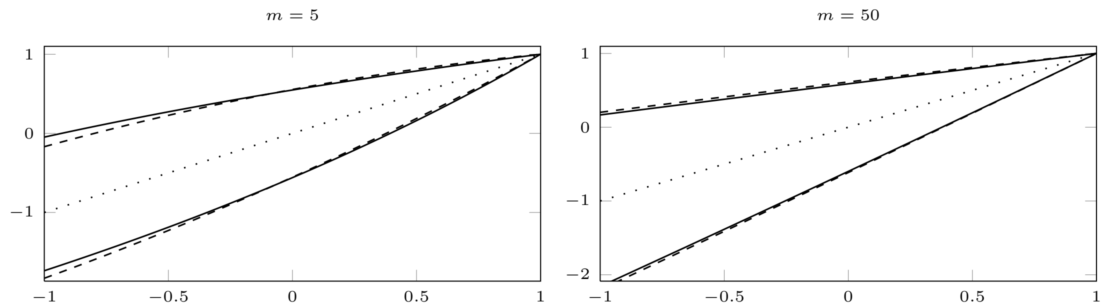

Figure 3: Equal-tailed 95% exact (solid lines) and large-r (dashed lines) confidence belts for λ based on $\hat{\lambda}_{\mathrm{ML}}$, for $\hat{\lambda}_{\mathrm{ML}} \in (-1, 1)$, when $r = 5$.

Appendix A.3 Bias

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[1]

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

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 (3.9) above, except that $\sqrt{\mathrm{med}(\mathrm{F}_{r,r(m-1)})}$ is here replaced by $k_1\sqrt{m-1}$.[2] This suggests that we consider a family of estimators of the form

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

where ϕ is a constant (possibly dependent on (r, m)) to be chosen.[3] 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 asymptotically 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[4]

$$\mathrm{E}(\hat{\lambda}_\phi) \simeq 1 - \alpha(1 + \mu_2 - \mu_3). \tag{A.3}$$

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

$$\mathrm{var}(\hat{\lambda}_\phi) \simeq \alpha^2 \left(\mu_2 - 2\mu_3 + \left(3\mu_4 - \mu_2^2\right)\right). \tag{A.4}$$

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

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

Focussing now on the MLE (the case $\phi = 1$), including only the term μ_2 in (A.3) 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 (A.4) reproduces the exact variance very well. Figure 4 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 (A.4), the large-r approximation (3.5) and the large-m approximation (3.12). The third order approximation seems to be vastly superior to the two asymptotic ones.

Appendix A.3.1 Bias Correction

From equation (A.3), 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

$$\hat{\lambda}_{\mathrm{BC}} := 2\hat{\lambda}_{\mathrm{ML}} - 1 + \hat{\alpha}(2 + \hat{\mu}_2). \tag{A.5}$$

We call this a *direct* bias correction.[5] The variance of the corrected estimator can also be obtained by the same methods, but we omit the details. Instead, in Figure 5 we plot the mean bias, for $\lambda \in (-1, 1)$, of $\hat{\lambda}_{\text{ML}}$, and of the three bias-reducing estimators we have introduced, $\hat{\lambda}_{\text{med}}$, $\hat{\lambda}_{\text{mean}}$, $\hat{\lambda}_{\text{BC}}$. This is obtained by straightforward simulation (cf. Section 3.1.1). Figures 6 and 7 do the same for the RMSE function and the median bias function.

These figures show that $\hat{\lambda}_{\text{ML}}$ can be significantly biased, but that direct bias correction ($\hat{\lambda}_{\text{BC}}$) essentially removes the entire mean bias. However, $\hat{\lambda}_{\text{BC}}$ performs poorly in terms of the median bias. The estimator $\hat{\lambda}_{\text{med}}$ does not perform as well as $\hat{\lambda}_{\text{BC}}$ in terms of mean bias, but it does reduce a good portion of the mean bias of $\hat{\lambda}_{\text{ML}}$, and is median unbiased by construction. These differing effects reflect the fact that the distribution of $\hat{\lambda}_{\text{ML}}$ can be quite skewed. The estimator $\hat{\lambda}_{\text{mean}}$ appears to be dominated by $\hat{\lambda}_{\text{med}}$ 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 $\hat{\lambda}_{\text{ML}}$, 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.

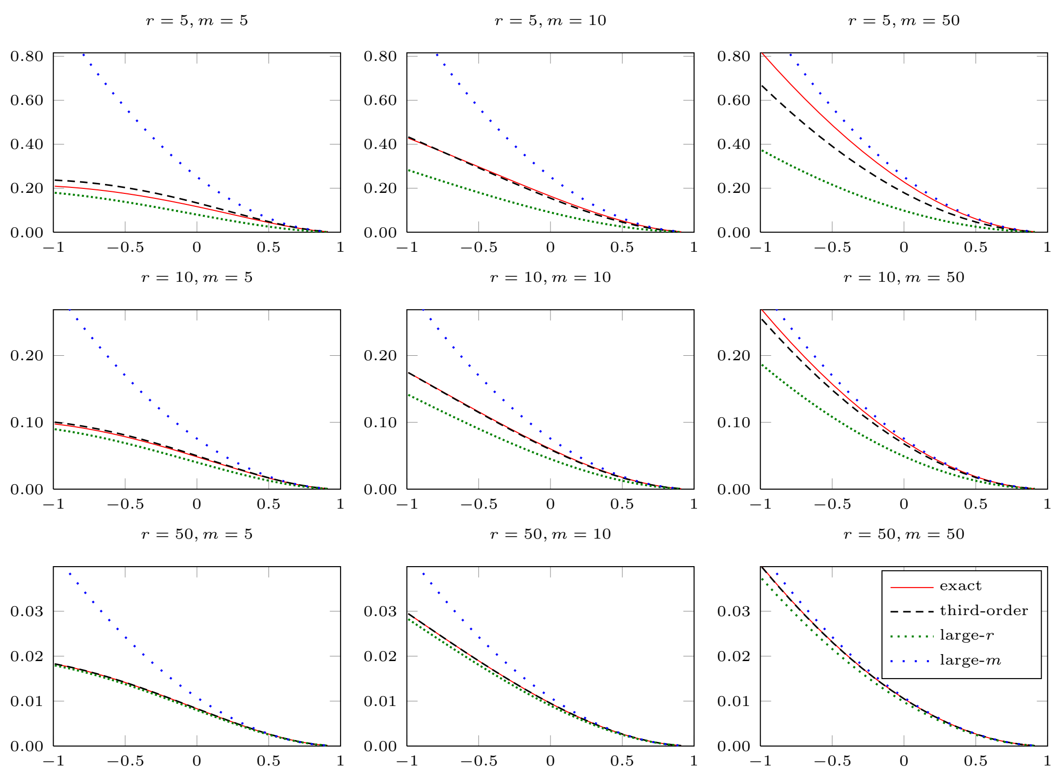

Figure 4: Exact variance of $\hat{\lambda}_{\text{ML}}$, as a function of λ, along with three different approximations.

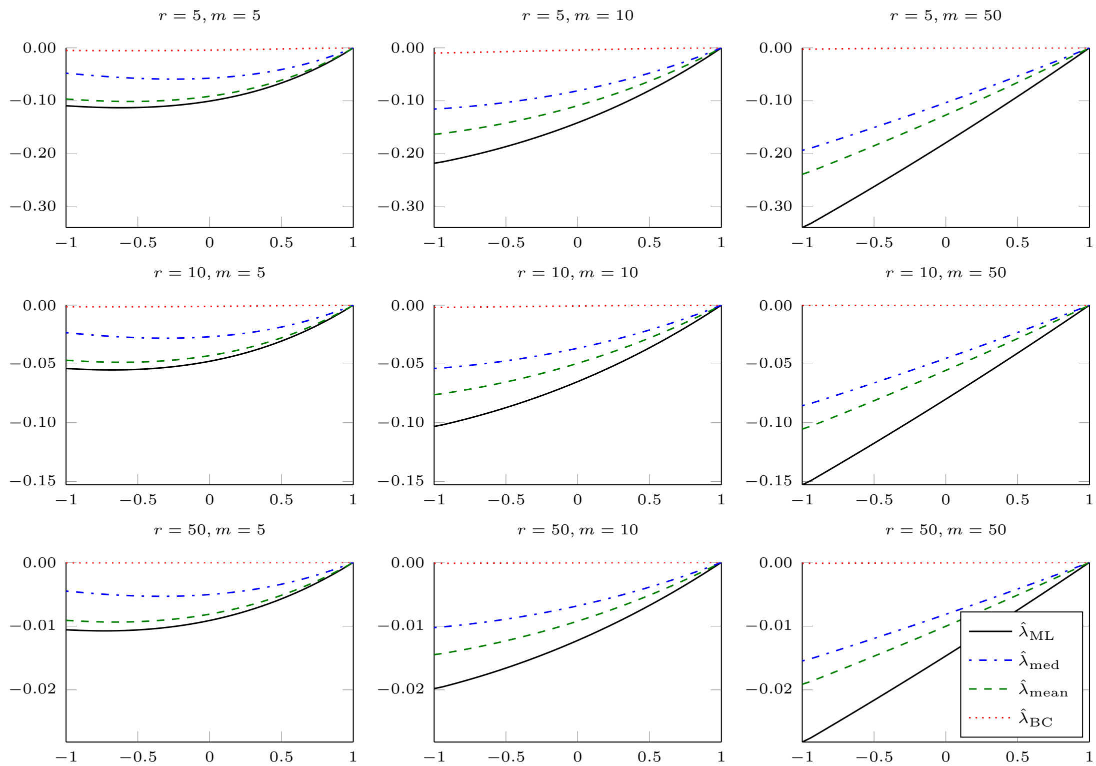

Figure 5: Bias function of the MLE ($\hat{\lambda}_{\mathrm{ML}}$), the median unbiased estimator ($\hat{\lambda}_{\mathrm{med}}$), the indirect estimator obtained by inverting the mean function ($\hat{\lambda}_{\mathrm{mean}}$), and the direct bias corrected MLE ($\hat{\lambda}_{\mathrm{BC}}$).

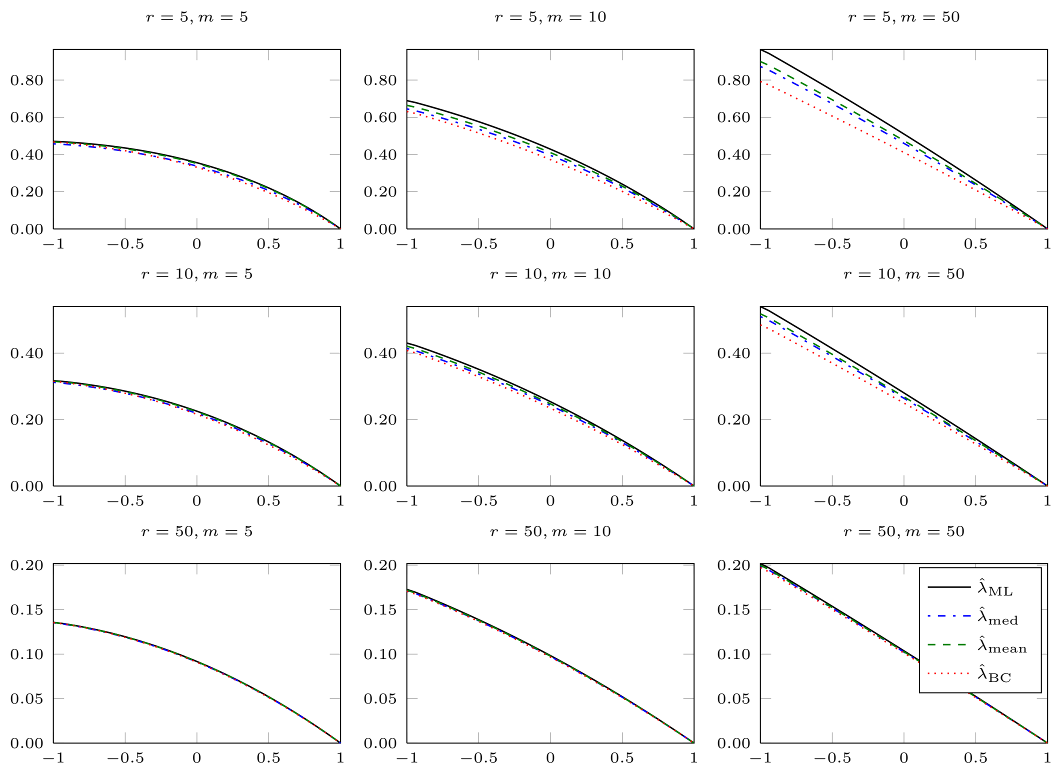

Figure 6: RMSE function of the MLE ($\hat{\lambda}_{\mathrm{ML}}$), the median unbiased estimator ($\hat{\lambda}_{\mathrm{med}}$), the indirect estimator obtained by inverting the mean function ($\hat{\lambda}_{\mathrm{mean}}$), and the direct bias corrected MLE ($\hat{\lambda}_{\mathrm{BC}}$).

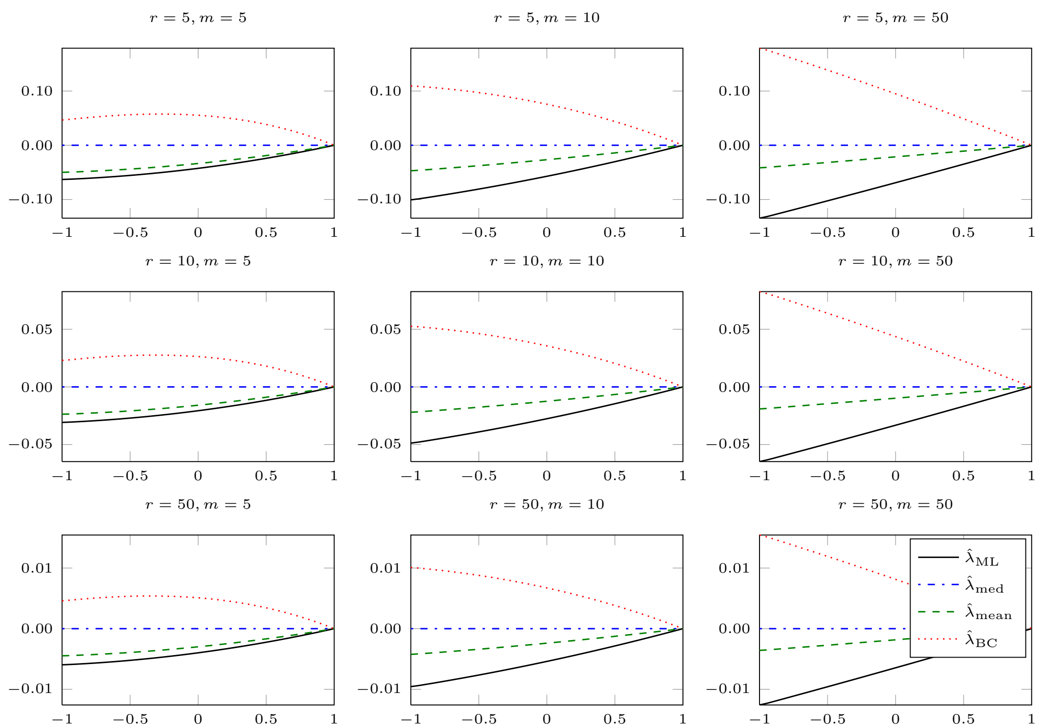

Figure 7: Median function of the MLE ($\hat{\lambda}_{\text{ML}}$), the median unbiased estimator ($\hat{\lambda}_{\text{med}}$), the indirect estimator obtained by inverting the mean function ($\hat{\lambda}_{\text{mean}}$), and the direct bias corrected MLE ($\hat{\lambda}_{\text{BC}}$).

Appendix B Unbalanced Model

Remark B.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 B.2. The estimating equation $\dot{l}_p(\lambda)=0$ is, for the unbalanced model, a polynomial of degree $p+1$ in λ, and has no explicit solution if $p>3$. The fact that the equation is known to have a single zero in Λ makes the numerical computation of the solution a much simpler task than it would otherwise be.

Appendix B.1 Exact Distribution

Figures 8 and 9 complement Figure 3 in the paper. They were produced using the result given in Proposition 4.3 in the text. Each of the three rows of Figure 8 displays $\mathrm{pdf}_{\hat{\lambda}_{\mathrm{ML}}}(z;\lambda)$ for a fixed value of m_1 and varying n, while Figure 9 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 $\hat{\lambda}_{\mathrm{ML}}$ is 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.

Appendix B.2 Probability of Underestimation

The values of z_2 relevant for Figure 10 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 $r_1=r_2$).

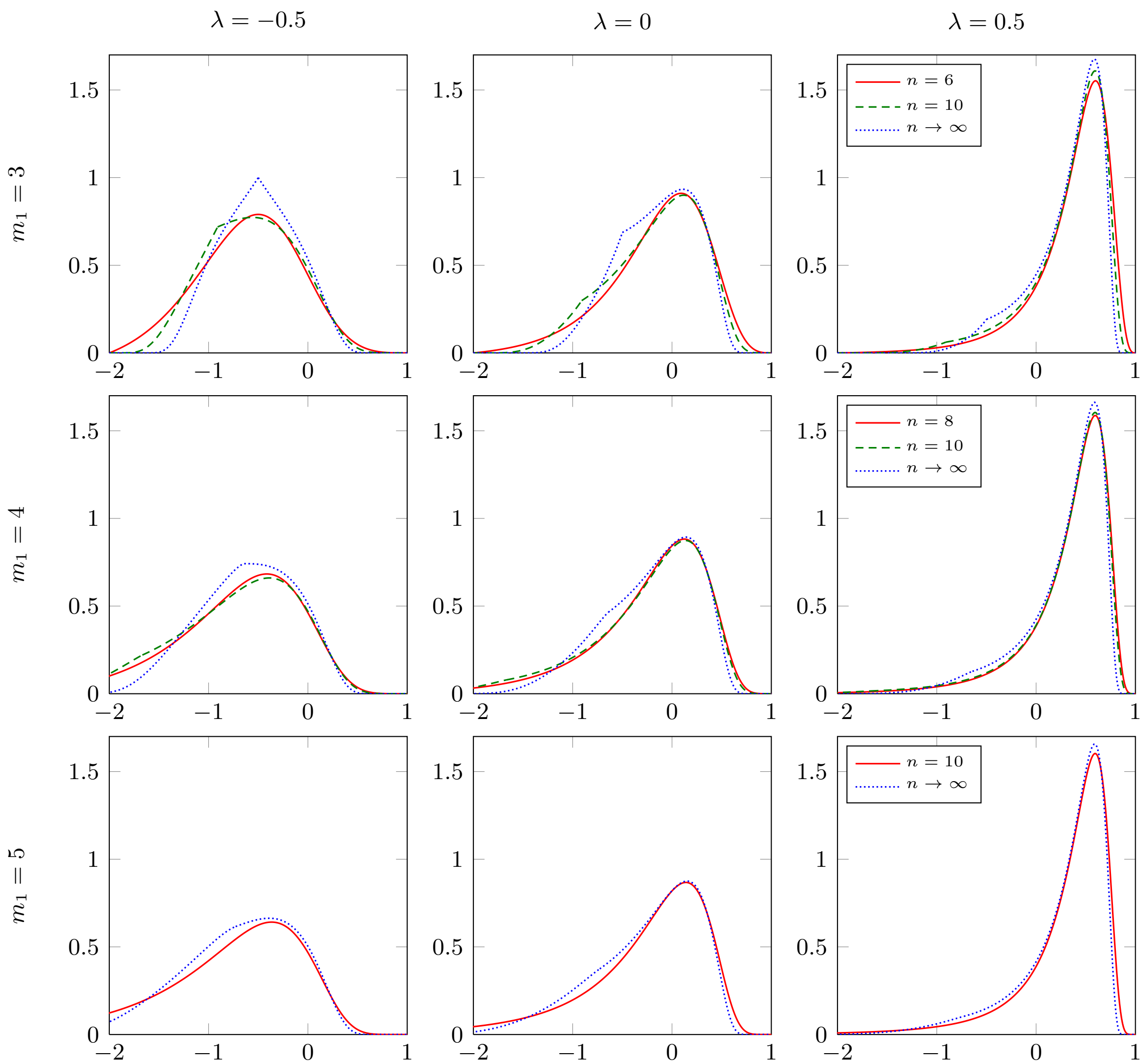

Figure 8: Density of $\hat{\lambda}_{\text{ML}}$ for pure Group Interaction model with two groups, when $\varepsilon \sim \text{SMN}(0, I_n)$.

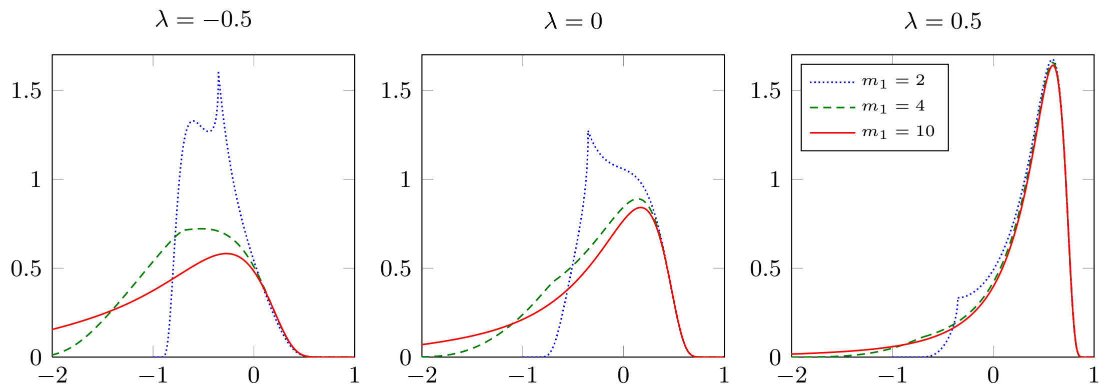

Figure 9: Density of $\hat{\lambda}_{\text{ML}}$ for pure Group Interaction model with two groups and $n = 25$, when $\varepsilon \sim \text{SMN}(0, I_n)$.

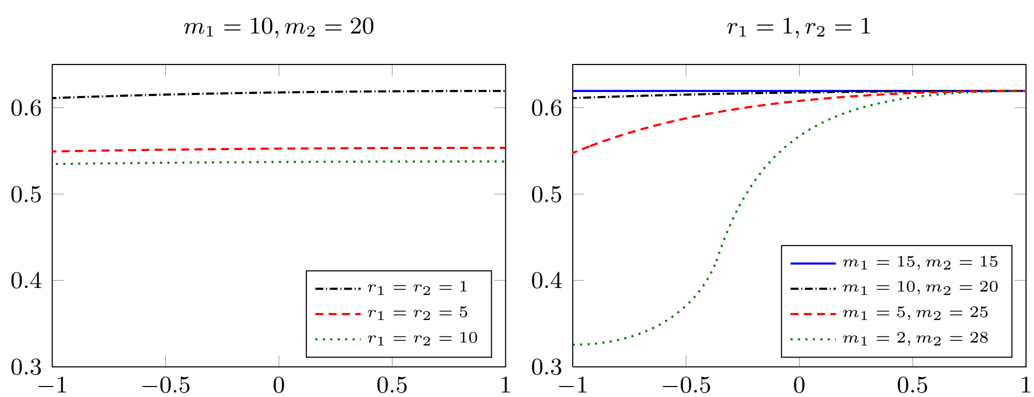

Figure 10: The probability that $\hat{\lambda}_{\text{ML}}$ underestimates λ as a function of λ, in the two-groups case.

Appendix B.3 Proofs for Section 4

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, the integral is a standard form of the confluent hypergeometric function, giving the density:

$$\mathrm{pdf}(q) = \frac{\exp\left(-\frac{q}{2a_1}\right) q^{\frac{v}{2}-1}}{a_1^{\frac{v_1}{2}} a_2^{\frac{v_2}{2}} 2^{\frac{v}{2}} \Gamma(\frac{v}{2})} {}_1F_1\left(\frac{v_2}{2}, \frac{v}{2}; \frac{1}{2a_1} q \left(1 - \frac{a_1}{a_2}\right)\right),$$

where $v := v_1 + v_2$. Obviously, for $a_1 = a_2$ we have the standard result that q is a multiple of a χ^2_v variate. Putting $\phi = 1/a_1$, $\psi := a_1/a_2$, we have

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

as given in the text. □

Proof of Proposition 4.3. We first prove the special cases (4.13) and (4.14) and then the general case. Consider the cdf defined by

$$\Pr(w \le z) = \Pr(\chi^2_\gamma \le a_1 \chi^2_\alpha + a_2 \chi^2_\beta),$$

where a_1, a_2 are positive functions of z, and we assume, without loss of generality, that $a_1 > a_2$ for all z (in case $a_1 < a_2$, simply interchange (α, β) and (a_1, a_2) in everything that follows). Conditioning first on $(q_1 = \chi^2_\alpha, q_2 = \chi^2_\beta)$, we have, on differentiating the conditional cdf, the conditional density is

$$\mathrm{pdf}_w(z|q_1, q_2) = \frac{\exp\left\{-\frac{1}{2}(a_1 q_1 + a_2 q_2)\right\}}{2^{\frac{\gamma}{2}} \Gamma(\frac{\gamma}{2})} (\dot{a}_1 q_1 + \dot{a}_2 q_2)(a_1 q_1 + a_2 q_2)^{\frac{\gamma}{2}-1}. \quad \text{(B.1)}$$

Multiplying by the joint density of (q_1, q_2),

$$\mathrm{pdf}_w(z, q_1, q_2) = \frac{\exp\left\{-\frac{1}{2}((1 + a_1)q_1 + (1 + a_2)q_2)\right\}}{2^{\frac{\alpha+\beta+\gamma}{2}} \Gamma(\frac{\gamma}{2})\Gamma(\frac{\alpha}{2})\Gamma(\frac{\beta}{2})} (\dot{a}_1 q_1 + \dot{a}_2 q_2)(a_1 q_1 + a_2 q_2)^{\frac{\gamma}{2}-1} q_1^{\frac{\alpha}{2}-1} q_2^{\frac{\beta}{2}-1}.$$

The problem is to integrate out (q_1, q_2). In case $\gamma = 2$ the term $(a_1q_1 + a_2q_2)^{\frac{\gamma}{2}-1}$ is missing, and it is immediate that

$$\mathrm{pdf}_w(z) = \frac{\frac{\alpha\dot{a}_1}{1+a_1} + \frac{\beta\dot{a}_2}{1+a_2}}{2(1+a_1)^{\frac{\alpha}{2}}(1+a_2)^{\frac{\beta}{2}}}. \tag{B.2}$$

Likewise, if γ is even, say $\gamma = 2s + 2$, then

$$(a_1q_1 + a_2q_2)^{\frac{\gamma}{2}-1} = (a_1q_1 + a_2q_2)^s.$$

Simple binomial expansion, followed by integration gives, in generalization of (B.2), the following special case of the result given in Proposition 4.3 in the text:

$$\begin{aligned}\mathrm{pdf}_w(z) &= \frac{(\frac{1}{2})_s}{2s!(1+a_1)^{\frac{\alpha}{2}}(1+a_2)^{\frac{\beta}{2}}} \\ &\times \left[\frac{\alpha\dot{a}_1}{1+a_1}C_s\left(A_{\alpha+2,\beta}\left(\frac{a_1}{1+a_1}, \frac{a_2}{1+a_2}\right)\right) + \frac{\beta\dot{a}_2}{1+a_2}C_s\left(A_{\alpha,\beta+2}\left(\frac{a_1}{1+a_1}, \frac{a_2}{1+a_2}\right)\right)\right].\end{aligned} \tag{B.3}$$

Here, we have used the formula given in Lemma 4.4.
Moving now to the general case where γ is arbitrary, start from the conditional density and make the same transformations as in the proof of Proposition 4.1 above. That is, set $x_i := a_iq_i, i = 1, 2$, then $s := x_1 + x_2$ and $b := x_1/(x_1 + x_2)$, we obtain

$$\begin{aligned}\mathrm{pdf}_w(z, s, b) &= \frac{\exp\left\{-\frac{1}{2}s\left(1 + a_2^{-1}\right)\right\} s^{\frac{\alpha+\beta+\gamma}{2}-1}}{2^{\frac{\alpha+\beta+\gamma}{2}}\Gamma(\frac{\alpha}{2})\Gamma(\frac{\beta}{2})\Gamma(\frac{\gamma}{2})a_1^{\frac{\alpha}{2}}a_2^{\frac{\beta}{2}}} \\ &\qquad \times \exp\left\{\frac{1}{2}sb\left(a_2^{-1} - a_1^{-1}\right)\right\} b^{\frac{\alpha}{2}-1}(1-b)^{\frac{\beta}{2}-1}\left[\frac{\dot{a}_1}{a_1}b + \frac{\dot{a}_2}{a_2}(1-b)\right].\end{aligned}$$

Integrating out b in the last line gives a linear combination of two confluent hypergeometric functions:

$$\begin{aligned}\mathrm{pdf}_w(z, s) &= \frac{\exp\left\{-\frac{1}{2}s\left(1 + a_2^{-1}\right)\right\} s^{\frac{\alpha+\beta+\gamma}{2}-1}}{2^{\frac{\alpha+\beta+\gamma}{2}}\Gamma(\frac{\gamma}{2})\Gamma(\frac{\alpha+\beta+2}{2})a_1^{\frac{\alpha}{2}}a_2^{\frac{\beta}{2}}} \\ &\qquad \times \left[\frac{\alpha\dot{a}_1}{2a_1}{}_1F_1\left(\frac{\alpha+2}{2}, \frac{\alpha+\beta+2}{2}; \frac{1}{2}s(a_2^{-1} - a_1^{-1})\right)\right. \\ &\qquad\qquad \left. + \frac{\beta\dot{a}_2}{2a_2}{}_1F_1\left(\frac{\alpha}{2}, \frac{\alpha+\beta+2}{2}; \frac{1}{2}s(a_2^{-1} - a_1^{-1})\right)\right].\end{aligned}$$

Integrating out s then produces the result given in Proposition 4.3 in the text. □

It can be shown (with some algebra) that this general result reduces to the results given above for the special cases $\gamma = 2$ and $\gamma = 2s + 2$.

Proof of Lemma 4.4. 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. □

Appendix B.4 An auxiliary Lemma

The following lemma is used in Section 4.7:

Lemma B.3. *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 14 in the main text). □

Notes

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

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

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

[4]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.

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

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

Additional References

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

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