# Modified-likelihood estimation of fixed-effect models for dyadic data

Koen Jochmans*

Toulouse School of Economics, Université Toulouse Capitole

This version: May 9, 2023

**Abstract**

We consider point estimation and inference based on modifications of the profile likelihood in models for dyadic interactions between $n$ agents featuring agent-specific parameters. The maximum-likelihood estimator of such models has bias and standard deviation of order $n^{-1}$ and so is asymptotically biased. Estimation based on modified likelihoods leads to estimators that are asymptotically unbiased and likelihood-ratio tests that exhibit correct size.

**JEL Classification:** C23

**Keywords:** asymptotic bias, dyadic data, fixed effects, undirected random graph

## 1 Introduction

A growing literature has uncovered the importance of interactions between agents through networks as drivers for economic and social outcomes. A leading approach to statistical modelling of dyadic interaction is through the inclusion of agent-specific parameters (see, e.g., Snijders 2011 for many references). A specific example that has received substantial attention in the recent literature is the $\beta$-model for network formation. There, agent fixed effects serve to capture degree heterogeneity in link formation and the inclusion of

*Address: Toulouse School of Economics, 1 esplanade de l'Université, 31080 Toulouse, France. E-mail: `koen.jochmans@tse-fr.eu`.

Financial support from the European Research Council, grant ERC-2016-STG-715787, and from the French Government and the ANR under the Investissements d' Avenir program, grant ANR-17-EURE-0010 is gratefully acknowledged. The guest editor and referee gave much appreciated comments on an earlier version of this paper.

dyad-level covariates reflects homophily; see, e.g., Graham (2017), Jochmans (2018), and Dzemski (2019).

Estimation of fixed-effect models for dyadic data is non-standard as the number of parameters grows with the sample size in a similar manner as in the classic incidental-parameter problem for one-way panel data discussed in Neyman and Scott (1948). Under so-called dense-network asymptotics, common parameters in regular models can be consistently estimated but inference is plagued by asymptotic bias; see Fernández-Val and Weidner (2016, 2018) and Graham (2017), for examples, discussion, and approaches to bias-correct the point estimator.

In this paper we look at generic estimation problems for undirected dyadic data and consider inference based on modifying the likelihood function in the spirit of Pace and Salvan (2006) and Arellano and Hahn (2006, 2007). In its most general form, the modified likelihood is a bias-corrected version of the profile likelihood, that is, of the likelihood after having profiled-out the nuisance parameters. The adjustment is both general and simple in form, involving only the score and Hessian of the likelihood with respect to the nuisance parameters. The adjustment term removes the leading bias from the profile likelihood and leads to asymptotically-unbiased inference and likelihood ratio statistics that are $\chi^2$-distributed under the null. The form of the adjustment can be specialized by using the likelihood structure, as in DiCiccio, Martin, Stern and Young (1996).

We work out the modifications to the profile likelihood in a linear version of the $\beta$-model and (in the appendix) in a linear version of the Bradley and Terry (1952) model for paired comparisons. These simple illustrations give insight in how the adjustments work. We next apply them to the $\beta$-model in the simulation designs of Graham (2017). We find that both modifications dramatically improve on maximum likelihood in terms of bias and mean squared error as well as reliability of statistical inference, and that they are considerably more reliable than ex-post bias-correction of the maximum-likelihood estimator.

## 2 Fixed-effect models for dyadic data

We consider data on dyadic interactions between $n$ agents. For each of ${}^{n(n-1)}/_{2}$ distinct agent pairs $(i,j)$ with $i<j$ we observe the random variable $z_{ij}$, which may be a vector. The density of $z_{ij}$ (relative to some dominating measure) takes the form $f(z_{ij};\vartheta,\beta_i,\beta_j)$, where $\vartheta$ and $\beta_1,\ldots,\beta_n$ are unknown Euclidean parameters. We may observe an outcome $y_{ij}$ generated by pair $(i,j)$ together with a vector of dyad characteristics $x_{ij}$, in which case we have $z_{ij}=(y_{ij},x_{ij}')'$, and we could consider the distribution of the outcome conditional on the covariates. In what follows, we take the $z_{ij}$ to be (conditionally) independent. Models of this form are relevant in many areas. Examples include the analysis of network formation as mentioned before, but also the study of strategic behavior among agents (Bajari, Hong and Nekipelov 2010), and the construction of rankings (Bradley and Terry 1952). Our goal is to perform inference on $\vartheta$ treating the $\beta_i$ as fixed effects.

The log-likelihood is

$$\ell(\vartheta,\beta)=\sum_{i=1}^{n}\sum_{i<j}\log f(z_{ij};\vartheta,\beta_i,\beta_j),$$

where we let $\beta=(\beta_1,\ldots,\beta_n)'$. For simplicity of exposition we ignore any normalization that may be needed on $\beta$ to achieve identification. When a normalization of the form $c(\beta)=0$ is needed, everything to follow goes through on replacing $\ell(\vartheta,\beta)$ by the constrained likelihood $\ell(\vartheta,\beta)-\lambda\, c(\beta)$, where $\lambda$ denotes the Lagrange multiplier. We give a detailed example in the appendix.

It is useful to recall that the maximum-likelihood estimator of $\vartheta$ can be expressed as

$$\hat{\vartheta}=\arg\max_{\vartheta}\hat{\ell}(\vartheta),$$

where $\hat{\ell}(\vartheta)=\ell(\vartheta,\hat{\beta}(\vartheta))$, with

$$\hat{\beta}(\vartheta)=\arg\max_{\beta}\ell(\vartheta,\beta),$$

is the profile likelihood.

Inference based on the profile likelihood performs poorly because the estimation noise in $\hat{\beta}(\vartheta)$ introduces non-negligible bias. Moreover, in regular settings,

$$\mathbb{E}(\hat{\vartheta} - \vartheta) = O(n^{-1}), \qquad \mathbb{E}((\hat{\vartheta} - \mathbb{E}(\hat{\vartheta}))^2) = O(n^{-2}),$$

so that bias and standard deviation are of the same order of magnitude. Consequently, the maximum-likelihood estimator is asymptotically biased.

## 2.1 Modified profile likelihood

In its simplest form, modified likelihoods can be understood as yielding a superior approximation to the target likelihood

$$\ell(\vartheta) = \ell(\vartheta, \beta(\vartheta)), \qquad \beta(\vartheta) = \arg\max_{\beta} \mathbb{E}(\ell(\vartheta, \beta)).$$

Moreover, the profile likelihood is the sample counterpart to this infeasible likelihood. Replacing $\beta(\vartheta)$ with $\hat{\beta}(\vartheta)$ introduces bias that leads to invalid inference. To see this suppose that

$$\hat{\beta}(\vartheta) - \beta(\vartheta) = \Sigma(\vartheta)^{-1} V(\vartheta) + O_p(n^{-1}), \tag{2.1}$$

where we introduce

$$V(\vartheta) = \left. \frac{\partial \ell(\vartheta, \beta)}{\partial \beta} \right|_{\beta = \beta(\vartheta)}, \qquad \Sigma(\vartheta) = -\,\mathbb{E}\left. \left( \frac{\partial^2 \ell(\vartheta, \beta)}{\partial \beta \partial \beta'} \right) \right|_{\beta = \beta(\vartheta)}.$$

An expansion of the profile likelihood around $\beta(\vartheta)$ yields

$$\begin{aligned} \hat{\ell}(\vartheta) - \ell(\vartheta) = \; & (\hat{\beta}(\vartheta) - \beta(\vartheta))' V(\vartheta) \\ & - \frac{1}{2} (\hat{\beta}(\vartheta) - \beta(\vartheta))' \Sigma(\vartheta) (\hat{\beta}(\vartheta) - \beta(\vartheta)) + O_p(n^{-1/2}). \end{aligned}$$

Combining the two expansions and taking expectations then shows that the bias of the profile likelihood is of the form

$$\mathbb{E}(\hat{\ell}(\vartheta) - \ell(\vartheta)) = \frac{1}{2} \mathrm{trace}(\Sigma(\vartheta)^{-1} \Omega(\vartheta)) + O(n^{-1/2}) \tag{2.2}$$

for

$$\Omega(\vartheta) = \mathbb{E}[V(\vartheta)\, V(\vartheta)'],$$

the variance of $V(\vartheta)$.

Equation (2.1) is a conventional asymptotically-linear representation of the estimator of the fixed effects; see, e.g., Rilstone, Srivastava and Ullah (1996). Low-level conditions for it to go through in specific models are provided in Fernández-Val and Weidner (2016, 2018). The difficulty in the current case, as opposed to say the one-way panel data model (as dealt with in Hahn and Newey 2004), is to handle the non-sparse nature of the Hessian matrix.

With (2.2) in hand, a modified likelihood is

$$\dot{\ell}(\vartheta) = \hat{\ell}(\vartheta) - \frac{1}{2}\text{trace}(\hat{\Sigma}(\vartheta)^{-1}\hat{\Omega}(\vartheta)),$$

where we define the plug-in estimators

$$\hat{\Sigma}(\vartheta) = \hat{\Sigma}(\vartheta, \hat{\beta}(\vartheta)), \qquad \hat{\Omega}(\vartheta) = \hat{\Omega}(\vartheta, \hat{\beta}(\vartheta)),$$

for matrices

$$-(\hat{\Sigma}(\vartheta,\beta))_{i,j} = \begin{cases} \sum_{k>i} \frac{\partial^2 \log f(z_{ik};\vartheta,\beta_i,\beta_k)}{\partial \beta_i^2} + \sum_{k<i} \frac{\partial^2 \log f(z_{ki};\vartheta,\beta_k,\beta_i)}{\partial \beta_i^2} & \text{if } i=j \\ \frac{\partial^2 \log f(z_{ij};\vartheta,\beta_i,\beta_j)}{\partial \beta_i \partial \beta_j} & \text{if } i<j \\ \frac{\partial^2 \log f(z_{ji};\vartheta,\beta_j,\beta_i)}{\partial \beta_i \partial \beta_j} & \text{if } i>j \end{cases}$$

and

$$(\hat{\Omega}(\vartheta,\beta))_{i,j} = \begin{cases} \sum_{k>i} \left(\frac{\partial \log f(z_{ik};\vartheta,\beta_i,\beta_k)}{\partial \beta_i}\right)^2 + \sum_{k<i} \left(\frac{\partial \log f(z_{ki};\vartheta,\beta_k,\beta_i)}{\partial \beta_i}\right)^2 & \text{if } i=j \\ \frac{\partial \log f(z_{ij};\vartheta,\beta_i,\beta_j)}{\partial \beta_i} \frac{\partial \log f(z_{ij};\vartheta,\beta_i,\beta_j)}{\partial \beta_j} & \text{if } i<j \\ \frac{\partial \log f(z_{ji};\vartheta,\beta_j,\beta_i)}{\partial \beta_i} \frac{\partial \log f(z_{ji};\vartheta,\beta_j,\beta_i)}{\partial \beta_j} & \text{if } i>j \end{cases}$$

In large samples, this modification removes the leading bias from the profile likelihood. Consequently, in large samples, the likelihood-ratio statistic has correct size and

$$\dot{\vartheta} = \arg\max_{\vartheta} \dot{\ell}(\vartheta),$$

will have bias $o(n^{-1})$. Furthermore, under usual regularity conditions, we have the limit result

$$(\dot{\vartheta} - \vartheta) \overset{a}{\sim} N\left(0, \frac{I(\vartheta)^{-1}}{n(n-1)/2}\right)$$

as $n \to \infty$, where we let

$$I(\vartheta) = \lim_{n\to\infty} \mathbb{E}\left(-\frac{\partial^2 \ell(\vartheta)}{\partial\vartheta\partial\vartheta'}\right) \Big/ \left(\frac{n(n-1)}{2}\right)$$

be the Fisher information for $\vartheta$.

The only point at which the likelihood setting has been used so far is in the statement of the limit distribution of $\dot{\vartheta} - \vartheta$, where the expression for the asymptotic variance exploits the information equality. Bias-corrected estimation—using the same formula for the bias as before—thus carries over to more general extremum-type estimation problems; the only change being that, now, the asymptotic variance is $I(\vartheta)^{-1}\Omega(\vartheta)\, I(\vartheta)^{-1}$.

Alternatively, following the arguments in Arellano and Hahn (2007) we can exploit the likelihood structure to get

$$\frac{1}{2}\text{trace}(\hat{\Sigma}(\vartheta)^{-1}\hat{\Omega}(\vartheta)) = -\frac{1}{2}\log(\det\hat{\Sigma}(\vartheta)) + \frac{1}{2}\log(\det\hat{\Omega}(\vartheta)) + O(n^{-1}),$$

which validates the alternative modified likelihood

$$\ddot{\ell}(\vartheta) = \hat{\ell}(\vartheta) + \frac{1}{2}\log(\det\hat{\Sigma}(\vartheta)) - \frac{1}{2}\log(\det\hat{\Omega}(\vartheta));$$

see DiCiccio, Martin, Stern and Young (1996). Its maximizer, say $\ddot{\vartheta}$, satisfies the same asymptotic properties as $\dot{\vartheta}$.

## 2.2 Illustration: A linear $\beta$-model

Consider the following extension of the classic many normal means problem of Neyman and Scott (1948). Data are generated as

$$z_{ij} \sim N(\beta_i + \beta_j, \vartheta),$$

and are independent across dyads. The likelihood function for all parameters (ignoring additive constants) is

$$\ell(\vartheta, \beta) = -\frac{1}{2}\frac{n(n-1)}{2}\log\vartheta - \frac{1}{2}\sum_{i=1}^{n}\sum_{i<j}\frac{(z_{ij} - \beta_i - \beta_j)^2}{\vartheta}.$$

Its first two derivatives with respect to the $\beta_i$ are

$$\frac{\partial \ell(\vartheta, \beta)}{\partial \beta_i} = \sum_{j>i}\frac{z_{ij} - \beta_i - \beta_j}{\vartheta} + \sum_{j<i}\frac{z_{ji} - \beta_j - \beta_i}{\vartheta}$$

and

$$\frac{\partial^2 \ell(\vartheta, \beta)}{\partial \beta_i \partial \beta_j} = \begin{cases} -\frac{(n-1)}{\vartheta} & \text{if } i = j \\ -\frac{1}{\vartheta} & \text{if } i \neq j \end{cases}.$$

Let $\tilde{z}_i = (n-2)^{-1}\sum_{j>i} z_{ij} + (n-2)^{-1}\sum_{j<i} z_{ji}$ and $\overline{z} = (2(n-1)^{-1}\sum_{i=1}^{n}\tilde{z}_i$. Solving for the maximum-likelihood estimator of $\beta_i$ gives $\hat{\beta}_i = \tilde{z}_i - \overline{z}$ for any $\vartheta$. The profile likelihood is therefore

$$\hat{\ell}(\vartheta) = -\frac{1}{2}\frac{n(n-1)}{2}\log\vartheta - \frac{1}{2}\sum_{i=1}^{n}\sum_{i<j}\frac{(z_{ij} - (\tilde{z}_i - \overline{z}) - (\tilde{z}_j - \overline{z}))^2}{\vartheta},$$

and its maximizer is

$$\hat{\vartheta} = \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{i<j}(z_{ij} - (\tilde{z}_i - \overline{z}) - (\tilde{z}_j - \overline{z}))^2.$$

Some tedious but straightforward calculations yield

$$\mathbb{E}(\hat{\vartheta} - \vartheta) = -\frac{2}{n-1}\,\vartheta, \qquad \operatorname{var}(\hat{\vartheta}) = \frac{n-3}{n-1}\,\frac{2\,\vartheta^2}{n(n-1)/2},$$

which confirms that the maximum-likelihood estimator of $\vartheta$ suffers from asymptotic bias. Moreover,

$$\sqrt{\frac{n(n-1)}{2}}(\hat{\vartheta} - \vartheta) \xrightarrow{d} N\big(-\sqrt{2}\vartheta, (\sqrt{2}\vartheta)^2\big),$$

as $n \to \infty$.

To set up the modified likelihood, first note that

$$(\hat{\Sigma}(\vartheta))_{i,j} = \begin{cases} \frac{n-1}{\vartheta} & \text{if } i = j \\ \frac{1}{\vartheta} & \text{if } i \neq j \end{cases}, \quad (\hat{\Sigma}(\vartheta)^{-1})_{i,j} = \begin{cases} \frac{\vartheta}{2}\frac{2n-3}{(n-1)(n-2)} & \text{if } i = j \\ -\frac{\vartheta}{2}\frac{1}{(n-1)(n-2)} & \text{if } i \neq j \end{cases},$$

and that

$$(\hat{\Omega}(\vartheta))_{i,j} = \begin{cases} \sum_{i<k} \frac{(z_{ik}-(\tilde{z}_i-\overline{z})-(\tilde{z}_k-\overline{z}))^2}{\vartheta^2} + \sum_{i>k} \frac{(z_{ki}-(\tilde{z}_k-\overline{z})-(\tilde{z}_i-\overline{z}))^2}{\vartheta^2} & \text{if } i = j \\ \frac{(z_{ij}-(\tilde{z}_i-\overline{z})-(\tilde{z}_j-\overline{z}))^2}{\vartheta^2} & \text{if } i < j \\ \frac{(z_{ji}-(\tilde{z}_j-\overline{z})-(\tilde{z}_i-\overline{z}))^2}{\vartheta^2} & \text{if } i > j \end{cases} .$$

It is then easily seen that

$$\frac{1}{2}\text{trace}(\hat{\Sigma}(\vartheta)^{-1}\hat{\Omega}(\vartheta)) = \frac{1}{2}\,\frac{2}{n-1}\sum_{i=1}^{n}\sum_{i<j}\frac{(z_{ij}-(\tilde{z}_i-\overline{z})-(\tilde{z}_j-\overline{z}))^2}{\vartheta}.$$

From this we obtain

$$\dot{\ell}(\vartheta) = -\frac{1}{2}\frac{n(n-1)}{2}\log\vartheta - \left(1+\frac{2}{n-1}\right)\frac{1}{2}\sum_{i=1}^{n}\sum_{i<j}\frac{(z_{ij}-(\tilde{z}_i-\overline{z})-(\tilde{z}_j-\overline{z}))^2}{\vartheta},$$

and its maximizer

$$\dot{\vartheta} = \frac{n+1}{n-1}\,\hat{\vartheta} = \hat{\vartheta} + \frac{2}{n-1}\hat{\vartheta}.$$

Clearly, this estimator removes the leading bias from the maximum-likelihood estimator. Moreover,

$$\mathbb{E}(\dot{\vartheta}-\vartheta) = -\left(\frac{2}{n-1}\right)^2\vartheta, \qquad \text{var}(\dot{\vartheta}) = \frac{n(n(n-1)-5)}{(n-1)^3}\,\frac{2\,\vartheta^2}{n(n-1)/2},$$

which shows that the remaining bias in the point estimator is small relative to its standard deviation.

As an alternative correction, we may exploit the likelihood structure to adjust the profile likelihood by the term

$$-\frac{1}{2}\log(\det\hat{\Sigma}(\vartheta)) + \frac{1}{2}\log(\det\hat{\Omega}(\vartheta)) = \frac{n}{2}\,\log\,\vartheta + c,$$

where $c$ is a constant that does not depend on $\vartheta$. This yields the modification

$$\ddot{\ell}(\vartheta) = -\frac{1}{2}\frac{n(n-3)}{2}\log\vartheta - \frac{1}{2}\sum_{i=1}^{n}\sum_{i<j}\frac{(z_{ij}-(\tilde{z}_i-\overline{z})-(\tilde{z}_j-\overline{z}))^2}{\vartheta},$$

whose maximizer satisfies

$$\mathbb{E}(\ddot{\vartheta} - \vartheta) = 0, \qquad \text{var}(\ddot{\vartheta}) = \frac{2\,\vartheta^2}{n(n-3)/2}.$$

This estimator is exactly unbiased.

To give an idea of the magnitude of the bias in this problem, Table 1 contains the bias and standard deviation of the estimators $\hat{\vartheta}$, $\dot{\vartheta}$, and $\ddot{\vartheta}$ for various sample sizes $n$ and variance parameter fixed to $\vartheta = 1$. These results are invariant to the value of the $\beta_i$ and can be interpreted as relative bias for general values of $\vartheta$.

Table 1: Many normal means

| $n$ | $\hat{\vartheta}$ | $\dot{\vartheta}$ | $\ddot{\vartheta}$ | $\hat{\vartheta}$ | $\dot{\vartheta}$ | $\ddot{\vartheta}$ |
|---|---|---|---|---|---|---|
| | | bias | | | standard deviation | |
| 5 | -0.5000 | -0.2500 | 0.0000 | 0.3162 | 0.4743 | 0.6325 |
| 10 | -0.2222 | -0.0494 | 0.0000 | 0.1859 | 0.2272 | 0.2390 |
| 15 | -0.1429 | -0.0204 | 0.0000 | 0.1278 | 0.1460 | 0.1491 |
| 20 | -0.1053 | -0.0111 | 0.0000 | 0.0970 | 0.1073 | 0.1085 |
| 25 | -0.0833 | -0.0069 | 0.0000 | 0.0782 | 0.0847 | 0.0853 |
| 50 | -0.0408 | -0.0017 | 0.0000 | 0.0396 | 0.0412 | 0.0413 |
| 75 | -0.0270 | -0.0007 | 0.0000 | 0.0265 | 0.0272 | 0.0272 |
| 100 | -0.0202 | -0.0004 | 0.0000 | 0.0199 | 0.0203 | 0.0203 |

Simulation results for the maximum-likelihood estimator ($\hat{\vartheta}$), the modified profile-likelihood estimator ($\dot{\vartheta}$), and the modified profile-likelihood estimator exploiting the likelihood structure ($\ddot{\vartheta}$)

# 3 Application to the $\beta$-model

The $\beta$-model of network formation models Bernoulli outcome variables as having success probability

$$\mathbb{P}(y_{ij} = 1 | x_{ij}; \vartheta, \beta_i, \beta_j) = F(\beta_i + \beta_j + x_{ij}'\vartheta),$$

where $F(a) = (1 + e^{-a})^{-1}$ is the logit link function. We now present the results from a Monte Carlo experiment. The designs are borrowed from Graham (2017). All designs are of the following form. Let $u_i \in \{-1, 1\}$ so that $\mathbb{P}(u_i = 1) = \frac{1}{2}$. We generate the dyad covariate as

$$x_{ij} = u_i \, u_j,$$

and the fixed effects as

$$\beta_i = \mu + \gamma_1 \, \frac{1 + u_i}{2} + \gamma_2 \, \frac{1 - u_i}{2} + v_i,$$

where $v_i \sim \text{Beta}(\lambda_1, \lambda_2)$. We set $\mu = -\lambda_1(\lambda_1 + \lambda_2)^{-1}$, so that $\mu + v_i$ has mean zero, and will consider several choices for the parameters $(\gamma_1, \gamma_2)$ and $(\lambda_1, \lambda_2)$. The parameter choices are summarized in Table 2. In the first four designs (A1–A4), the $\beta_i$ are drawn independently of $x_{ij}$ from symmetric Beta distributions. In the next four designs (B1–B4) the $\beta_i$ are generated from skewed distributions that depend on $u_i$ (and thus correlate with the regressor $x_{ij}$). For both the designs labelled A and B, the average number of observed links per agent goes down as we move from the first design (A1 and B1) to the fourth design (A4 and B4). The average number of links decreases from about 50% to 12%. This is clear from the second block of Table 2, which contains the average, minimum, and maximum number of links per agent (in percentages).

We simulate 10,000 data sets for each design for $n \in \{25, 50, 75, 100\}$ and $\vartheta = 1$. Because the results across designs are qualitatively very similar, we present the full set of results only for Design A1 (Table 3). Tables 4 and 5 provide the results for $n \in \{50, 100\}$ for all designs. Each table contains the mean and median bias of $\vartheta$, $\dot{\vartheta}$, and $\ddot{\vartheta}$, along with their standard deviation and their interquartile range (both across the Monte Carlo replications). The tables also provide the empirical size of the likelihood ratio test for the null that $\vartheta = 1$ for theoretical size $\alpha \in \{.05, .10\}$. Inference results based on the Wald statistic, using a plug-in estimator of $I(\vartheta)$, are very similar and not reported for brevity.

Because the results for $n = 100$ can be compared (up to Monte Carlo error) to the numerical results collected in Graham (2017, Table 2), Table 5 contains two additional columns in which we reproduce the results for his analytically bias-corrected maximum-

Table 2: Simulation designs for the $\beta$-model

| Design | $\gamma_1$ | $\gamma_2$ | $\lambda_1$ | $\lambda_2$ | degree (%) | | |
|---|---|---|---|---|---|---|---|
| | | | | | mean | min | max |
| A1 | 0 | 0 | 1 | 1 | 50 | 32 | 67 |
| A2 | -0.25 | -0.25 | 1 | 1 | 40 | 24 | 57 |
| A3 | -0.75 | -0.75 | 1 | 1 | 23 | 10 | 38 |
| A4 | -1.25 | -1.25 | 1 | 1 | 12 | 3 | 22 |
| B1 | 0 | 0.50 | 0.25 | 0.75 | 60 | 40 | 78 |
| B2 | -0.50 | 0 | 0.25 | 0.75 | 40 | 21 | 62 |
| B3 | -1.00 | -0.50 | 0.25 | 0.75 | 24 | 8 | 44 |
| B4 | -1.50 | -1.00 | 0.25 | 0.75 | 12 | 2 | 28 |

likelihood estimator ($\hat{\vartheta}_{\mathrm{BC}}$) and his 'tetrad logit' estimator ($\hat{\vartheta}_{\mathrm{TETRAD}}$). The latter is based on moment conditions that are free of $\beta_i$ using a sufficiency argument. Bias correcting $\hat{\vartheta}$ does not salvage the likelihood-ratio statistic, and the conditional likelihood function of the 'tetrad logit' estimator is a quasi likelihood and, therefore, does not satisfy the information equality. Hence, the results on size for these two estimators are based on the Wald statistic.

Table 3 clearly shows that both the bias and standard deviation of $\hat{\vartheta}$ are of order $n^{-1}$. Consequently, the likelihood ratio test is size distorted even in large samples. Point estimation through the modified likelihoods gives estimators with small bias relative to their standard error. Even for $n = 25$, the bias is only about 20% of the bias in maximum likelihood estimator. In larger samples, the estimators are essentially unbiased. Both $\dot{\vartheta}$ and $\ddot{\vartheta}$ are also less volatile than is $\hat{\vartheta}$. This phenomenon has been observed elsewhere; we refer to Schumann, Severini and Tripathi (2022). Thus, here, bias correction does not come at the cost of an increase in dispersion. Together with the substantial decrease in mean squared error, inference, too, improves dramatically. The likelihood ratio statistics for $\dot{\ell}(\vartheta)$ and $\ddot{\ell}(\vartheta)$ have near-theoretical size for all $n$.

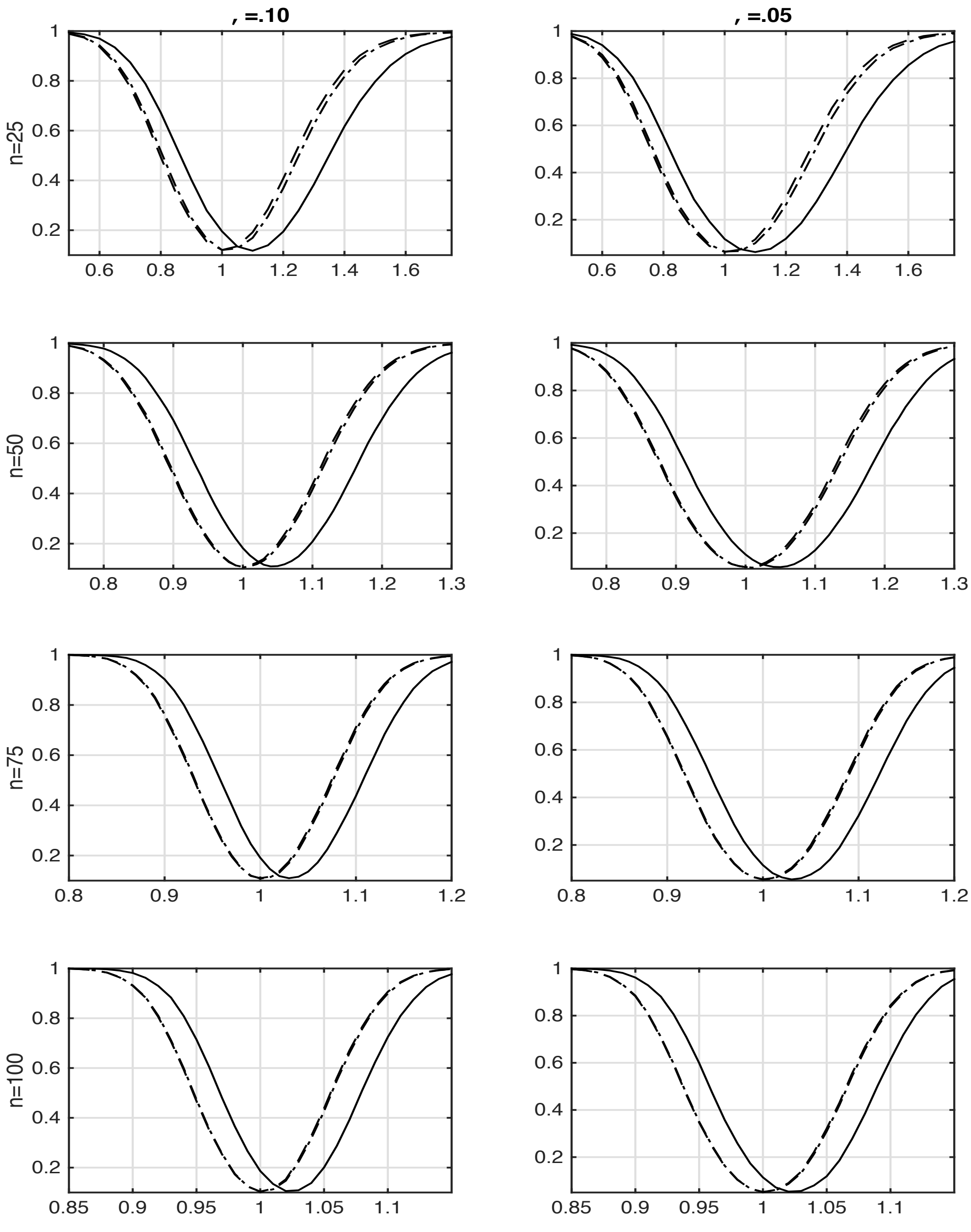
To give a more complete picture on inference via modifying the profile likelihood Figure

Table 3: $\beta$-model. Design A1 for all $n$

| $n$ | $\hat{\vartheta}$ | $\dot{\vartheta}$ | $\ddot{\vartheta}$ | $\hat{\vartheta}$ | $\dot{\vartheta}$ | $\ddot{\vartheta}$ |
|---|---|---|---|---|---|---|
| | mean bias | | | standard deviation | | |
| 25 | 0.1098 | 0.0204 | 0.0304 | 0.1897 | 0.1560 | 0.1572 |
| 50 | 0.0492 | 0.0045 | 0.0071 | 0.0717 | 0.0679 | 0.0681 |
| 75 | 0.0320 | 0.0020 | 0.0032 | 0.0467 | 0.0450 | 0.0451 |
| 100 | 0.0237 | 0.0011 | 0.0017 | 0.0341 | 0.0332 | 0.0332 |
| | median bias | | | interquartile range | | |
| 25 | 0.1029 | 0.0154 | 0.0253 | 0.2069 | 0.1873 | 0.1889 |
| 50 | 0.0487 | 0.0042 | 0.0067 | 0.0961 | 0.0913 | 0.0914 |
| 75 | 0.0316 | 0.0017 | 0.0028 | 0.0630 | 0.0607 | 0.0608 |
| 100 | 0.0236 | 0.0010 | 0.0017 | 0.0464 | 0.0450 | 0.0451 |
| | empirical size ($\alpha = .10$) | | | empirical size ($\alpha = .05$) | | |
| 25 | 0.1937 | 0.1134 | 0.1147 | 0.1142 | 0.0627 | 0.0637 |
| 50 | 0.1896 | 0.1128 | 0.1125 | 0.1178 | 0.0558 | 0.0555 |
| 75 | 0.1866 | 0.1092 | 0.1081 | 0.1142 | 0.0575 | 0.0569 |
| 100 | 0.1890 | 0.1042 | 0.1025 | 0.1103 | 0.0520 | 0.0513 |

Simulation results for the maximum-likelihood estimator ($\hat{\vartheta}$), the modified profile-likelihood estimator ($\dot{\vartheta}$), and the modified profile-likelihood estimator exploiting the likelihood structure ($\ddot{\vartheta}$)

Figure 1: Power curves. Design A1 for all $n$




Power curves for the likelihood-ratio statistic based on the profile likelihood $\hat{\ell}(\vartheta)$ (solid lines), the modified profile likelihood $\dot{\ell}(\vartheta)$ (dashed lines), and the modified profile likelihood that exploits the likelihood structure $\ddot{\ell}(\vartheta)$ (dashed-dotted lines) for different sample sizes (vertically; $n \in \{25, 50, 75, 100\}$) and size (horizontally; left: $\alpha = .10$, right: $\alpha = .05$).

1 presents power curves for the likelihood ratio statistic that go along with Table 3. The curves for $\hat{\ell}(\vartheta)$ (solid lines) are symmetric but not correctly centered, reflecting the fact that they are size distorted. This is so for all sample sizes and significance levels considered. Modifying the likelihood shifts the power curve so that the likelihood ratio test is (approximately) size correct. This is done without significantly altering the shape of the power curves. For the smallest sample size considered ($n = 25$; upper two plots) there is a small difference in power between the likelihood ratio test for $\dot{\ell}(\vartheta)$ (dashed lines) and $\ddot{\ell}(\vartheta)$ (dashed-dotted lines); the former has slightly higher power than the latter for alternatives $\vartheta > 1$, and slightly less power for $\vartheta < 1$. This difference vanished rapidly as $n$ increases, however, which is in line with the similar performance of both corrections observed in Table 3.

Tables 4 and 5 show that all conclusions from Design A1 carry over to the other designs. Moreover, the introduction of correlation between regressors and heterogenous coefficients or skewing the distribution from which the latter are drawn does not prevent the modified likelihood to improve on maximum likelihood both in terms of point estimation and inference. A comparison of the two tables clearly shows that both the bias and standard deviation of $\hat{\vartheta}$ shrink by a factor of one half as $n$ doubles, again illustrating that both are of order $n^{-1}$. The subsequent reduction in bias by considering $\dot{\vartheta}$ and $\ddot{\vartheta}$ and improvement in size are manifest for all designs.

Table 5 further shows that the modified-likelihood approach outperforms bias correction of the maximum-likelihood estimator in Designs A3 and B3 and, in particular, in Designs A4 and B4. There, bias correction of maximum likelihood introduces rather substantial additionial bias relative to $\hat{\vartheta}$. The additional bias also leads to a large deterioration of the empirical size of the Wald statistic associated with $\hat{\vartheta}_{\text{BC}}$, with actual sizes ranging up to seven times the nominal size. This type of sensitivity of analytical bias correction has equally been observed in panel data applications; see Dhaene and Jochmans (2015) and Higgins and Jochmans (2023). The performance of the modified likelihood is comparable to Graham's 'tetrad logit' estimator $\hat{\vartheta}_{\text{TETRAD}}$ in terms of bias, and it tends to be somewhat more accurate in terms of the empirical size of the associated hypothesis tests. Moreover,

Table 4: $\beta$-model. All designs for $n = 50$

| Design | $\hat{\vartheta}$ | $\dot{\vartheta}$ | $\ddot{\vartheta}$ | $\hat{\vartheta}$ | $\dot{\vartheta}$ | $\ddot{\vartheta}$ |
|---|---|---|---|---|---|---|
| | | mean bias | | | standard deviation | |
| A1 | 0.0492 | 0.0045 | 0.0071 | 0.0717 | 0.0679 | 0.0681 |
| A2 | 0.0499 | 0.0054 | 0.0079 | 0.0742 | 0.0704 | 0.0705 |
| A3 | 0.0467 | 0.0033 | 0.0047 | 0.0933 | 0.0890 | 0.0891 |
| A4 | 0.0497 | 0.0049 | 0.0024 | 0.1391 | 0.1335 | 0.1335 |
| B1 | 0.0526 | 0.0073 | 0.0096 | 0.0768 | 0.0728 | 0.0729 |
| B2 | 0.0490 | 0.0035 | 0.0059 | 0.0747 | 0.0707 | 0.0708 |
| B3 | 0.0493 | 0.0046 | 0.0060 | 0.0936 | 0.0891 | 0.0891 |
| B4 | 0.0500 | 0.0043 | 0.0005 | 0.1380 | 0.1320 | 0.1316 |
| | | median bias | | | interquartile range | |
| A1 | 0.0487 | 0.0042 | 0.0067 | 0.0961 | 0.0913 | 0.0914 |
| A2 | 0.0482 | 0.0040 | 0.0064 | 0.0995 | 0.0943 | 0.0945 |
| A3 | 0.0441 | 0.0008 | 0.0022 | 0.1247 | 0.1191 | 0.1191 |
| A4 | 0.0412 | -0.0032 | -0.0059 | 0.1827 | 0.1748 | 0.1748 |
| B1 | 0.0513 | 0.0061 | 0.0084 | 0.1034 | 0.0981 | 0.0982 |
| B2 | 0.0479 | 0.0024 | 0.0049 | 0.0999 | 0.0948 | 0.0949 |
| B3 | 0.0470 | 0.0024 | 0.0039 | 0.1252 | 0.1195 | 0.1196 |
| B4 | 0.0438 | -0.0018 | -0.0052 | 0.1827 | 0.1740 | 0.1743 |
| | | empirical size ($\alpha = .10$) | | | empirical size ($\alpha = .05$) | |
| A1 | 0.1896 | 0.1128 | 0.1125 | 0.1178 | 0.0558 | 0.0555 |
| A2 | 0.1857 | 0.1135 | 0.1118 | 0.1139 | 0.0602 | 0.0603 |
| A3 | 0.1565 | 0.1098 | 0.1082 | 0.0878 | 0.0581 | 0.0563 |
| A4 | 0.1287 | 0.1095 | 0.1083 | 0.0664 | 0.0594 | 0.0592 |
| B1 | 0.1902 | 0.1141 | 0.1112 | 0.1146 | 0.0582 | 0.0579 |
| B2 | 0.1801 | 0.1081 | 0.1049 | 0.1040 | 0.0574 | 0.0564 |
| B3 | 0.1498 | 0.1052 | 0.1030 | 0.0830 | 0.0554 | 0.0538 |
| B4 | 0.1236 | 0.1064 | 0.1067 | 0.0634 | 0.0543 | 0.0551 |

Simulation results for the maximum-likelihood estimator ($\hat{\vartheta}$), the modified profile-likelihood estimator ($\dot{\vartheta}$), and the modified profile-likelihood estimator exploiting the likelihood structure ($\ddot{\vartheta}$)

Table 5: $\beta$-model. All designs for $n = 100$

| Design | $\hat{\vartheta}$ | $\dot{\vartheta}$ | $\ddot{\vartheta}$ | $\hat{\vartheta}_{\mathrm{BC}}$ | $\hat{\vartheta}_{\mathrm{TETRAD}}$ | $\hat{\vartheta}$ | $\dot{\vartheta}$ | $\ddot{\vartheta}$ | $\hat{\vartheta}_{\mathrm{BC}}$ | $\hat{\vartheta}_{\mathrm{TETRAD}}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | mean bias | | | | | standard deviation | | |
| A1 | 0.0237 | 0.0011 | 0.0017 | — | — | 0.0341 | 0.0332 | 0.0332 | 0.0340 | 0.0340 |
| A2 | 0.0240 | 0.0015 | 0.0022 | — | — | 0.0352 | 0.0343 | 0.0343 | 0.0350 | 0.0350 |
| A3 | 0.0215 | 0.0003 | 0.0006 | — | — | 0.0444 | 0.0434 | 0.0434 | 0.0460 | 0.0440 |
| A4 | 0.0206 | 0.0001 | -0.0004 | — | — | 0.0645 | 0.0633 | 0.0633 | 0.0790 | 0.0650 |
| B1 | 0.0241 | 0.0013 | 0.0019 | — | — | 0.0356 | 0.0346 | 0.0346 | 0.0360 | 0.0370 |
| B2 | 0.0246 | 0.0015 | 0.0021 | — | — | 0.0359 | 0.0349 | 0.0349 | 0.0340 | 0.0340 |
| B3 | 0.0222 | 0.0004 | 0.0007 | — | — | 0.0450 | 0.0439 | 0.0439 | 0.0470 | 0.0460 |
| B4 | 0.0231 | 0.0020 | 0.0015 | — | — | 0.0651 | 0.0639 | 0.0639 | 0.0810 | 0.0670 |
| | | | median bias | | | | | interquartile range | | |
| A1 | 0.0236 | 0.0010 | 0.0017 | 0.0010 | -0.0010 | 0.0464 | 0.0450 | 0.0451 | — | — |
| A2 | 0.0234 | 0.0010 | 0.0017 | 0.0020 | 0.0000 | 0.0471 | 0.0458 | 0.0458 | — | — |
| A3 | 0.0199 | -0.0013 | -0.0010 | 0.0190 | 0.0010 | 0.0598 | 0.0584 | 0.0584 | — | — |
| A4 | 0.0181 | -0.0022 | -0.0028 | 0.1060 | 0.0050 | 0.0870 | 0.0854 | 0.0855 | — | — |
| B1 | 0.0240 | 0.0012 | 0.0019 | 0.0040 | 0.0010 | 0.0482 | 0.0469 | 0.0469 | — | — |
| B2 | 0.0247 | 0.0017 | 0.0022 | 0.0030 | 0.0010 | 0.0484 | 0.0472 | 0.0472 | — | — |
| B3 | 0.0213 | -0.0007 | -0.0003 | 0.0170 | 0.0000 | 0.0604 | 0.0591 | 0.0591 | — | — |
| B4 | 0.0212 | 0.0002 | -0.0003 | 0.0860 | 0.0020 | 0.0867 | 0.0852 | 0.0851 | — | — |
| | | | empirical size ($\alpha = .10$) | | | | | empirical size ($\alpha = .05$) | | |
| A1 | 0.1890 | 0.1042 | 0.1025 | 0.1020 | 0.0760 | 0.1103 | 0.0520 | 0.0513 | 0.0550 | 0.0360 |
| A2 | 0.1846 | 0.1044 | 0.1026 | 0.0970 | 0.0810 | 0.1081 | 0.0533 | 0.0526 | 0.0570 | 0.0420 |
| A3 | 0.1447 | 0.1051 | 0.1037 | 0.1260 | 0.0800 | 0.0805 | 0.0522 | 0.0515 | 0.0680 | 0.0380 |
| A4 | 0.1149 | 0.1003 | 0.1002 | 0.4670 | 0.0810 | 0.0593 | 0.0505 | 0.0500 | 0.3520 | 0.0360 |
| B1 | 0.1748 | 0.1002 | 0.0978 | 0.1050 | 0.0810 | 0.0999 | 0.0494 | 0.0488 | 0.0610 | 0.0340 |
| B2 | 0.1846 | 0.1064 | 0.1050 | 0.0870 | 0.0610 | 0.1089 | 0.0533 | 0.0521 | 0.0350 | 0.0250 |
| B3 | 0.1437 | 0.1050 | 0.1036 | 0.1400 | 0.0850 | 0.0812 | 0.0522 | 0.0516 | 0.0760 | 0.0380 |
| B4 | 0.1174 | 0.1007 | 0.1003 | 0.3910 | 0.0900 | 0.0621 | 0.0509 | 0.0507 | 0.2960 | 0.0430 |

Simulation results for the maximum-likelihood estimator ($\hat{\vartheta}$), the modified profile-likelihood estimator ($\dot{\vartheta}$), the modified profile-likelihood estimator exploiting the likelihood structure ($\ddot{\vartheta}$, the bias-corrected maximum-likelihood estimator ($\hat{\vartheta}_{\mathrm{BC}}$) and the 'tetrad logit' estimator ($\hat{\vartheta}_{\mathrm{TETRAD}}$).

inference based on the 'tetrad logit' estimator is conservative in all designs even though, with $n = 100$ and therefore $4,950$ dyadic observations, the sample size is large. In addition, the 'tetrad logit' estimator is computationally prohibitive in large networks.

## Appendix: A linear Bradley-Terry model

As an alternative to the specification of the Neyman and Scott (1948) model with complementarities, now suppose that

$$z_{ij} \sim N(\beta_i - \beta_j, \vartheta)$$

independently across dyads. This model is overparametrized as, clearly, the mean of the $\beta_i$ is not identified. A common normalization in this type of model is $\sum_{i=1}^n \beta_i = 0$ (Simons and Yao 1999), and we will maintain it here. The constrained likelihood is

$$-\frac{1}{2}\frac{n(n-1)}{2}\log\vartheta - \frac{1}{2}\sum_{i=1}^n\sum_{i<j}\frac{(z_{ij}-\beta_i+\beta_j)^2}{\vartheta} + \lambda\sum_{i=1}^n\beta_i,$$

where $\lambda$ is the Lagrange multiplier for our normalization constraint. The first-order condition for the constrained problem for $\beta_i$ for a given $\vartheta$ equals

$$\frac{\sum_{j>i} z_{ij} - \sum_{j<i} z_{ji}}{\vartheta} - \frac{n}{\vartheta}\beta_i = 0.$$

This gives

$$\hat{\beta}_i = \frac{\sum_{j>i} z_{ij} - \sum_{j<i} z_{ji}}{n} = \tilde{z}_i \quad \text{(say)}$$

for all $i$ and any $\vartheta$. Observe that the sign of $\hat{\beta}_i$ is driven by the comparison of the magnitudes of $\sum_{i<j} z_{ij}$ and $\sum_{i>j} z_{ji}$. Also note that $\sum_{i=1}^n \hat{\beta}_i = 0$ holds. We therefore have

$$\hat{\ell}(\vartheta) = -\frac{1}{2}\frac{n(n-1)}{2}\log\vartheta - \frac{1}{2}\sum_{i=1}^n\sum_{i<j}\frac{(z_{ij}-\tilde{z}_i+\tilde{z}_j)^2}{\vartheta},$$

and with it, the maximum-likelihood estimator

$$\hat{\vartheta} = \frac{2}{n(n-1)}\sum_{i=1}^n\sum_{i<j}(z_{ij}-\tilde{z}_i+\tilde{z}_j)^2.$$

A calculation shows that $\mathbb{E}(\hat{\vartheta} - \vartheta) = -2n^{-1}\vartheta$.

It is immediate that

$$\hat{\Sigma}(\vartheta) = \text{diag}\left(\frac{n}{\vartheta}\right), \qquad \hat{\Sigma}(\vartheta)^{-1} = \text{diag}\left(\frac{\vartheta}{n}\right),$$

and that

$$(\hat{\Omega}(\vartheta))_{i,j} = \begin{cases} \sum_{k>i} \frac{(z_{ik}-\tilde{z}_i+\tilde{z}_k)^2}{\vartheta^2} + \sum_{k<i} \frac{(z_{ki}-\tilde{z}_k+\tilde{z}_i)^2}{\vartheta^2} & \text{if } i = j \\ \frac{(z_{ij}-\tilde{z}_i+\tilde{z}_j)^2}{\vartheta^2} & \text{if } i < j \\ \frac{(z_{ji}-\tilde{z}_j+\tilde{z}_i)^2}{\vartheta^2} & \text{if } i > j \end{cases}.$$

Therefore,

$$\dot{\ell}(\vartheta) = -\frac{1}{2}\frac{n(n-1)}{2}\log\vartheta - \frac{1}{2}\left(1+\frac{2}{n}\right)\sum_{i=1}^{n}\sum_{i<j}\frac{(z_{ij}-\tilde{z}_i+\tilde{z}_j)^2}{\vartheta},$$

$$\ddot{\ell}(\vartheta) = -\frac{1}{2}\frac{n(n-3)}{2}\log\vartheta - \frac{1}{2}\sum_{i=1}^{n}\sum_{i<j}\frac{(z_{ij}-\tilde{z}_i+\tilde{z}_j)^2}{\vartheta}.$$

The corresponding estimators are

$$\dot{\vartheta} = \left(1+\frac{2}{n}\right)\hat{\vartheta}, \qquad \ddot{\vartheta} = \frac{n-1}{n-3}\hat{\vartheta} = \left(1+\frac{2}{n-3}\right)\hat{\vartheta}.$$

Both remove the leading bias from the maximum-likelihood estimator, as

$$\mathbb{E}(\dot{\vartheta}-\vartheta) = -\frac{4}{n^2}\vartheta = O(n^{-2}), \qquad \mathbb{E}(\ddot{\vartheta}-\vartheta) = \frac{2}{n(n-3)}\vartheta = O(n^{-2}),$$

but, in this case, neither is exactly unbiased. The first estimator has bias that is strictly negative (for any finite $n$). The second estimator overcorrects and has strictly positive bias. The second-order bias is monotone in $n$. We have

$$\frac{4}{n^2}\vartheta > \frac{2}{n(n-3)}\vartheta$$

for all $n > 7$. As $n \to \infty$,

$$\sqrt{\frac{n(n-1)}{2}}(\dot{\vartheta}-\vartheta) \xrightarrow{d} N(0, 2\vartheta^2),$$

and $\|\ddot{\vartheta}-\dot{\vartheta}\| = o_p(n^{-1})$; that is, the two modifications to the likelihood yield asymptotically-equivalent estimators.

## References

Arellano, M. and J. Hahn (2006). A likelihood-based approximate solution to the incidental parameter problem in dynamic nonlinear models with multiple effects. Unpublished manuscript.

Arellano, M. and J. Hahn (2007). Understanding bias in nonlinear panel models: Some recent developments. In R. W. Blundell, W. K. Newey, and T. Persson (Eds.), *Advances In Economics and Econometrics*, Volume III. Econometric Society: Cambridge University Press.

Bajari, P., H. Hong, and D. Nekipelov (2010). Estimating static models of strategic interactions. *Journal of Business & Economic Statistics 28*, 469–482.

Bradley, R. A. and M. E. Terry (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika 39*, 324–325.

Dhaene, G. and K. Jochmans (2015). Split-panel jackknife estimation of fixed-effect models. *Review of Economic Studies 82*, 991–1030.

DiCiccio, T. J., M. A. Martin, S. E. Stern, and A. Young (1996). Information bias and adjusted profile likelihoods. *Journal of the Royal Statistical Society, Series B 58*, 189–203.

Dzemski, A. (2019). An empirical model of dyadic link formation in a network with unobserved heterogeneity. *Review of Economics and Statistics 101*, 763–776.

Fernández-Val, I. and M. Weidner (2016). Individual and time effects in nonlinear panel models with large $N, T$. *Journal of Econometrics 192*, 291–312.

Fernández-Val, I. and M. Weidner (2018). Fixed effect estimation of large-$t$ panel data models. *Annual Review of Economics 10*, 109–138.

Graham, B. S. (2017). An econometric model of link formation with degree heterogeneity. *Econometrica 85*, 1033–1063.

Hahn, J. and W. K. Newey (2004). Jackknife and analytical bias reduction for nonlinear panel models. *Econometrica 72*, 1295–1319.

Higgins, A. and K. Jochmans (2023). Bootstrap inference in fixed-effect models. TSE

Working Paper 22-1328.

Jochmans, K. (2018). Semiparametric analysis of network formation. *Journal of Business & Economic Statistics 36*, 705–713.

Neyman, J. and E. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica 16*, 1–32.

Pace, L. and A. Salvan (2006). Adjustments of profile likelihood from a new perspective. *Journal of Statistical Planning and Inference 136*, 3554–3564.

Rilstone, P., V. K. Srivastava, and A. Ullah (1996). The second-order bias and mean squared error of nonlinear estimators. *Journal of Econometrics 75*, 369–395.

Schumann, M., T. A. Severini, and G. Tripathi (2022). The role of information bias in panel data likelihoods. Forthcoming in *Journal of Econometrics*.

Simons, G. and Y.-C. Yao (1999). Asymptotics when the number of parameters tends to infinity in the Bradley-Terry model for paired comparisons. *Annals of Statistics 27*, 1041–1060.

Snijders, T. A. B. (2011). Statistical models for social networks. *Annual Review of Sociology 37*, 129–151.