entropy

MDPI

Article

Moment Estimation in Paired Comparison Models with a Growing Number of Subjects

Qiuping Wang [1,*], Lu Pan [2] and Ting Yan [3]

[1] School of Mathematics and Statistics, Zhaoqing University, Zhaoqing 526000, China
[2] School of Mathematics and Statistics, Shangqiu Normal University, Shangqiu 476000, China; panlulu@mails.ccnu.edu.cn
[3] Department of Statistics, Central China Normal University, Wuhan 430079, China; tingyanty@mail.ccnu.edu.cn
* Correspondence: qp.wang@mails.ccnu.edu.cn

Abstract

When the number of subjects, n, is large, paired comparisons are often sparse. Here, we study statistical inference in a class of paired comparison models parameterized by a set of merit parameters, under an Erdös–Rényi comparison graph, where the sparsity is measured by a probability p_n tending to zero. We use the moment estimation base on the scores of subjects to infer the merit parameters. We establish a unified theoretical framework in which the uniform consistency and asymptotic normality of the moment estimator hold as the number of subjects goes to infinity. A key idea for the proof of the consistency is that we obtain the convergence rate of the Newton iterative sequence for solving the estimator. We use the Thurstone model to illustrate the unified theoretical results. Further extensions to a fixed sparse comparison graph are also provided. Numerical studies and real data analysis illustrate our theoretical findings.

Keywords: consistency; central limit theorem; method of moments; paired comparison; parameter estimation; sparse

1. Introduction

Subjects are repeatedly compared in pairs in a wide spectrum of situations, including sports games [1], ranking of scientific journals [2,3], the quality of product brands [4] and crowdsourcing [5]. For instance, one team plays with another team in basketball; papers in one journal cite papers in another journal; one consumer chooses one product over another; workers in a crowdsourcing setup are asked to compare pairs of items.

One of the fundamental problems in paired comparison analysis is to derive a fair and reliable ranking of all subjects based on observed comparison data. In round-robin tournaments—where every pair of subjects competes sufficiently many times—a natural ranking can be directly obtained from the number of wins, as the full pairwise comparisons eliminate biases from incomplete matchups. However, in most practical scenarios (e.g., sports leagues, crowdsourcing evaluations, or journal rankings), comparisons are often sparse (not all pairs interact) and stochastic (outcomes contain random noise), leading to unreliable direct rankings based solely on raw win counts. To address this issue, paired comparison models have been developed to statistically infer the underlying merit parameters of subjects and generate objective rankings; see the classic monograph by [6] for a comprehensive overview of such models and their theoretical foundations. Statistical

Check for updates

Academic Editor: Nikolai Leonenko

Received: 19 February 2026
Revised: 5 March 2026
Accepted: 9 March 2026
Published: 11 March 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

models not only provide a method of ranking all subjects but are also tools for making inferences on the merits of subjects (e.g., testing whether two subjects have the same merit).

Here, we are concerned with a class of paired comparison models that assign one merit parameter to each subject and assume that the win–loss probability of any pair only depends on the difference between their merit parameters. Specifically, the probability of subject i winning j is

$$\mathbb{P}(i \text{ wins } j) = F(\beta_i - \beta_j), \quad i, j = 0, 1, \ldots, n; i \neq j, \tag{1}$$

where F is a known cumulative distribution function satisfying $F(x) = 1 - F(-x)$, β_i is the merit parameter of subject i and $n+1$ is the total number of subjects. The well-known Bradley–Terry model [7], which dates back to at least 1929 [8], and the Thurstone model [9], are two special cases of Model (1). The former postulates the logistic distribution of $F(x)$, while the latter postulates the normal distribution.

In the standard setting that n is fixed and the number of comparisons in each pair goes to infinity, the theoretical properties of Model (1) have been widely investigated in Chapter 4 of [6]. In the opposite scenario that n goes to infinity and each pair has a fixed number of comparisons, ref. [10] proved the uniform consistency and asymptotic normality of the maximum likelihood estimator (MLE) in the Bradley–Terry model.

When the number of subjects is large, paired comparisons are often sparse. Taking the NCAA Division I FBS (Football Bowl Subdivision) regular season, for example, a team plays with at most 14 other teams among a total of 120 teams. The observed comparisons can be represented in a comparison graph with $n+1$ nodes denoting subjects and a weighted edge between two nodes denoting the number of comparisons. The Erdös–Rényi comparison graph has been widely considered in the literature, e.g., [1,11–13], where the number of comparisons between any two subjects follows a binomial distribution (T, p_n), and p_n measures the sparsity. Under a very weak condition on the sparsity on p_n, ref. [13] established the uniform consistency and asymptotic normality of the MLE in the Bradley–Terry model by extending the proof strategies in [10].

Moreover, ref. [14] considered a fixed sparse comparison graph by controlling the length from one subject to another subject with 2 or 3, in which the consistency and asymptotic normality of the MLE also hold. Inference in the high-dimensional setting under the Bradley–Terry model and some generalized versions has also attracted great interest in the machine learning literature; the upper bounds of various errors are established under different conditions [e.g., the ℓ_1 error $\|\hat{\beta} - \beta\|_1$ in [15,16], the mean square error in [17], the bias $\mathbb{E}\|\hat{\beta} - \beta\|_\infty$ in [18]]. Under the assumption that the log-likelihood function is strictly convex, ref. [1] establish the uniform consistency of the MLE in general paired comparison models. However, the asymptotic theory of moment estimation under sparse paired comparison models remains largely underdeveloped in the literature. Existing theoretical developments focus almost exclusively on maximum likelihood estimation (MLE), which relies on strict distributional assumptions and is computationally demanding in high-dimensional settings. In contrast, this paper develops the method of moments (MOM), which avoids full distribution assumptions and maintains computational simplicity while achieving comparable asymptotic properties. The primary novelty of this work is to extend the asymptotic theory of high-dimensional sparse paired comparison models from MLE to the method of moments, establishing a parallel and complementary theoretical framework.

We further elaborate on the advantages of the method of moments (MOM) for high-dimensional sparse paired comparison models. Beyond computational efficiency, MOM has two key strengths over maximum likelihood estimation (MLE) for practical inference: (1) Robustness—the moment estimator is much less sensitive to outliers (e.g., upsets in our NFL data) in sparse settings, where MLE can be biased by extreme observations;

(2) Quasi-likelihood compatibility—MOM relies only on moment conditions and avoids strict distributional assumptions, making it robust to model misspecification in high-dimensional sparse scenarios. These merits justify the use of MOM in this study, and our subsequent analysis establishes its asymptotic properties as the core theoretical contribution.

The main contributions of this paper are as follows. First, we develop the moment estimation, instead of the maximum likelihood estimation (MLE) or Bayesian estimation, based on the scores of subjects (i.e., the number of wins) to estimate the merit parameters in Model (1). The reason why we prioritize moment estimation over MLE is that it is natural to rank subjects according to their scores and the computation based on moment equations is simpler, especially in high-dimensional sparse settings where MLE may suffer from numerical instability due to nonlinear optimization. When $F(\cdot)$ belongs to the exponential family distribution, both estimations are identical. Second, under an Erdös–Rényi comparison graph, we establish a unified theoretical framework in which the uniform consistency and asymptotic normality of the moment estimator hold when n goes to infinity and p_n tends to zero. A key idea for the proof of the consistency is that we obtain the convergence rate of the Newton iterative sequence for solving the estimator. The asymptotic normality is proved by applying Taylor expansions to a series of functions constructed from estimating equations and showing that remainder terms in the expansions are asymptotically neglected. Although each pair of subjects is assumed to have a comparison with the same probability p_n, our proof strategy can be easily extended to the case with different comparison probabilities at the order of p_n. Third, we use the Thurstone model to illustrate the unified theoretical results. Further extensions to a fixed sparse comparison graph in [14] are also derived. Numerical studies and real data analysis illustrate our theoretical findings.

The rest of this paper is organized as follows. In Section 2, we present the moment estimation. In Section 3, we present the consistency and asymptotic normality of the moment estimator. We illustrate our unified results with one application in Section 4. We extend the asymptotic results to a fixed comparison graph in Section 5. In Section 6, we carry out simulations and give real data analysis. We give a summary and further discussion in Section 7. The proofs of the main results are relegated to Appendix A. The proofs of supported lemmas are relegated to Appendix B.

2. Moment Estimation

Assume that $n+1$ subjects that are labeled as "$0, \ldots, n$", are compared in pairs repeatedly. Let t_{ij} be the times that subject i compares with subject j and a_{ij} be the times that subject i wins subject j out of t_{ij} comparisons. As a result, $a_{ij} + a_{ji} = t_{ij}$. By convention, define $t_{ii} = 0$ and $a_{ii} = 0$. The comparison matrix $(t_{ij})_{(n+1)\times(n+1)}$ is generated from an Erdös–Rényi comparison graph, where t_{ij} follows a binomial distribution $Bin(T, p_n)$ with p_n measuring the sparsity of comparisons. More generally, $t_{ij} \sim Bin(T_{ij}, p_n)$. We set T_{ij} to be the same for ease of exposition. Recall that $\beta_0, \ldots, \beta_n$ are the merit parameters of subjects $0, \ldots, n$. The probability in Model (1) implies that the winning probability only depends on the difference in merit between two subjects. For the identification of models, we normalize $\beta_i, i = 0, 1, \ldots, n$ by setting $\beta_0 = 0$ as in [10]. We assume that all paired comparisons are independent and a_{ij} follows a binomial distribution $Bin(t_{ij}, p_{ij})$ conditional on t_{ij}.

Let $a_i = \sum_{j=0}^{n} a_{ij}$ be the total wins of subject i and $a = (a_1, \ldots, a_n)^\top$. To motivate the estimating equations, we compare the maximum likelihood equation and the moment equation under the Thurstone model described in Section 4. The maximum likelihood equations are

$$\sum_{j \neq i} \left[\frac{a_{ij}\phi(\beta_i - \beta_j)}{\Phi(\beta_i - \beta_j)} - \frac{(t_{ij} - a_{ij})\phi(\beta_i - \beta_j)}{1 - \Phi(\beta_i - \beta_j)} \right] = 0, \quad i = 1, \ldots, n.$$

where $\phi(\cdot)$ is the density function of the standard normality and $\Phi(\cdot)$ is its distribution function. The corresponding moment equations are

$$a_i = \sum_{j \neq i} t_{ij} \Phi(\beta_i - \beta_j), \ \ i = 1, \ldots, n.$$

We can see that the latter is simpler and easier to compute. On the other hand, it is natural to rank subjects according to their scores. Thus, we use the moment estimation here. When $F(\cdot)$ in Model (1) belongs to the exponential family distributions, both are the same.

Write $\mu(\cdot)$ as the expectation of $F(\cdot)$ and $\mu_{ij}(\beta) = \mu(\beta_i - \beta_j)$. Then, the estimating equations are

$$a_i = \sum_{j \neq i} t_{ij} \mu_{ij}(\beta), \ \ i = 1, \ldots, n. \tag{2}$$

The solution to the above equations is the moment estimator denoted by $\widehat{\beta} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_n)^\top$ and $\hat{\beta}_0 = 0$. Let

$$\varphi(\beta) = (\sum_{j \neq 1} t_{ij} \mu_{1j}(\beta), \ldots, \sum_{j \neq n} t_{nj} \mu_{nj}(\beta))^\top.$$

If $\varphi(\beta) : \mathbb{R}^n \to (0, \infty)$ is a one-to-one mapping, then $\widehat{\beta}$ exists and is unique, i.e., $\widehat{\beta} = \varphi^{-1}(a)$. When φ^{-1} does not exist (i.e., φ is not one-to-one), any solution $\widehat{\beta}$ of Equation (2) is a moment estimator of β. The Newton–Raphson algorithm can be used to solve Equation (2). Moreover, the R language provides the package "BradleyTerry2" to solve the estimator in the Bradley–Terry model.

We discuss the existence of $\widehat{\beta}$ from the viewpoint of graph connection. If the comparison graph with the matrix $(t_{ij})_{i,j=0,\ldots,n}$ as its adjacency matrix is not connected, then there are two empty sets such that there are no comparisons between subjects in the first set and those in the second. In this case, there is no basis for ranking subjects in the first set and those in the second set. Further, a necessary condition for the existence of $\widehat{\beta}$ is that the directed graph $\mathcal{G}_n$ with the win–loss matrix $A = (a_{ij})$ as its adjacency matrix is strongly connected. In other words, for every partition of the subjects into two nonempty sets, a subject in the second set beats a subject in the first at least once. To see this, assume that there are two empty sets B_1 and B_2 such that all subjects in B_1 win all comparisons with subjects in B_2. Without loss of generality, we set $B_1 = \{0, \ldots, m\}$ and $B_2 = \{m+1, \ldots, n\}$ with $0 \leq m < n$, where $a_{ij} = t_{ij}$ for $i \in B_1$ and $j \in B_2$. By summing a_i over $i = 0, \ldots, m$, we have

$$\sum_{i=0}^{m} a_i = \sum_{i=0}^{m} \sum_{j=0}^{m} t_{ij} \mu(\beta_i - \beta_j) + \sum_{i=0}^{m} \sum_{j=m+1}^{n} t_{ij} \mu(\beta_i - \beta_j).$$

Because a_i is a sum of a_{ij}, $j = 0, \ldots, n$, and $\mu(\beta_i - \beta_j) + \mu(\beta_j - \beta_i) = 1$, we have

$$\sum_{i=0}^{m} \sum_{j=m+1}^{n} a_{ij} = \sum_{i=0}^{m} \sum_{j=m+1}^{n} t_{ij} \mu(\beta_i - \beta_j).$$

Because $a_{ij} = t_{ij}$ for $i = 0, \ldots, m$ and $j = m+1, \ldots, n$ and at least such one $t_{ij} > 0$, it must be $\mu(\beta_i - \beta_j) = 1$ when $t_{ij} > 0$ in order to guarantee both sides in the above equation to be equal. In this case, at least one such difference $\beta_i - \beta_j$ must go to infinity such that the moment estimate does not exist. The strong connection of $\mathcal{G}_n$ is also sufficient for guaranteeing the existence of the MLE in the Bradley–Terry model [19] in which the moment estimator is equal to the MLE. It is interesting to see whether the strong connection of $\mathcal{G}_n$ is sufficient to guarantee the existence of $\widehat{\beta}$ in a general model. In the next section, we will show that $\widehat{\beta}$ exists with probability approaching one under some mild conditions.

3. Asymptotic Properties

In this section, we present the consistency and asymptotic normality of the moment estimator. We first introduce some notations. For a subset $C \subset \mathbb{R}^n$, let C^0 and $\overline{C}$ denote the interior and closure of C, respectively. For a vector $x = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$, denote $\|x\|$ by a vector norm with the ℓ_∞-norm, $\|x\|_\infty = \max_{1\le i\le n} |x_i|$, and the ℓ_1-norm, $\|x\|_1 = \sum_i |x_i|$. Let $B(x, \epsilon) = \{y : \|x - y\|_\infty \le \epsilon\}$ be an ϵ-neighborhood of x. For an $n \times n$ matrix $J = (J_{ij})$, let $\|J\|_\infty$ denote the matrix norm induced by the ℓ_∞-norm on vectors in $\mathbb{R}^n$, i.e.,

$$\|J\|_\infty = \max_{x\neq 0} \frac{\|Jx\|_\infty}{\|x\|_\infty} = \max_{1\le i\le n} \sum_{j=1}^{n} |J_{ij}|,$$

and let $\|J\|$ be a general matrix norm. Define the matrix maximum norm: $\|J\|_{\max} = \max_{i,j} |J_{ij}|$. We use the superscript "*" to denote the true parameter under which the data are generated. When there is no ambiguity, we omit the superscript "*".

Recall that $\mu(\cdot)$ is the expectation of $F(\cdot)$. We assume that $\mu(\cdot)$ is a continuous function with the third derivative. Write μ' and μ'' as the first and second derivatives of $\mu(\pi)$ on π, respectively. Let ϵ_n be a small positive number. When $\beta \in B(\beta^*, \epsilon_n)$, we assume that there are three positive numbers, b_{n0}, b_{n1}, b_{n2}, such that

$$[\min_{i,j} \mu'(\pi_{ij})] \cdot [\max_{i,j} \mu'(\pi_{ij})] > 0, \tag{3a}$$

$$b_{n0} \le \min_{i,j} |\mu'(\pi_{ij})| \le \max_{i,j} |\mu'(\pi_{ij})| \le b_{n1}, \tag{3b}$$

$$\max_{i,j} |\mu''(\pi_{ij})| \le b_{n2}, \tag{3c}$$

where $\pi_{ij} := \beta_i - \beta_j$.

We use the Bradley–Terry model to illustrate the above inequalities, where $\mu(x) = e^x/(1 + e^x)$. A direct calculation gives that

$$\mu'(x) = \frac{e^x}{(1+e^x)^2}, \quad \mu''(x) = \frac{e^x(1-e^x)}{(1+e^x)^3}.$$

It is easy to show that

$$b_{n0} = \frac{e^{2\|\beta^*\|_\infty + 2\epsilon_n}}{(1 + e^{2\|\beta^*\|_\infty + 2\epsilon_n})^2} \le |\mu'(x)| \le b_{n1} = \frac{1}{4}, \quad |\mu''(x)| \le b_{n2} = \frac{1}{4}. \tag{4}$$

If $\epsilon_n = o(1)$, then $1/b_{n0} = O(e^{2\|\beta^*\|_\infty})$.

3.1. Consistency

To establish the consistency of $\widehat{\beta}$, let us first define a system of functions:

$$H_i(\beta) = \sum_{j=0}^{n} t_{ij}\mu_{ij}(\beta) - a_i, \quad i = 0, \ldots, n, \tag{5}$$

and $H(\beta) = (H_1(\beta), \ldots, H_n(\beta))^\top$. It is clear that $H(\widehat{\beta}) = 0$. Let $H'(\beta)$ be the Jacobian matrix of $H(\beta)$ on the parameter β. The asymptotic behavior of $\widehat{\beta}$ depends crucially on the inverse of $H'(\beta)$. For convenience, denote $H'(\beta)$ as $V = (v_{ij})_{i,j=1,\ldots,n}$, where

$$v_{ij} = -t_{ij}\mu'(\pi_{ij}),\ i \neq j, \quad v_{ii} = \sum_{j=0, j\neq i}^{n} t_{ij}\mu'(\pi_{ij}).$$

Define

$$v_{i0} = v_{0i} := \sum_{j=0, j\neq i}^{n} v_{ij} - v_{ii} = -t_{0i}\mu'(\pi_{0i}),\ \ i = 1, \ldots, n,\ \ v_{00} = \sum_{j=1}^{n} t_{0j}\mu'(\pi_{0j}).$$

When $\beta \in B(\beta^*, \epsilon_n)$ and $\min_{i,j} \mu'(\pi_{ij}) > 0$, in view of inequality (3b), the entries of V satisfy the following inequalities:

$$\begin{array}{ll} \text{if } t_{i0} > 0, & t_{i0}b_{n0} \leq v_{ii} + \sum_{j=1, j\neq i}^{n} v_{ij} \leq t_{i0}b_{n1},\ \ i = 1, \ldots, n, \\ \text{if } t_{ij} > 0, & t_{ij}b_{n0} \leq -v_{ij} \leq t_{ij}b_{n1},\ \ i, j = 1, \ldots, n; i \neq j. \end{array} \tag{6}$$

Without loss of generality, we assume that $\min_{i\neq j} \mu'(\pi_{ij}) > 0$ when $\beta \in B(\beta^*, \epsilon_n)$ hereafter (otherwise, we redefine $H_i(\beta) = a_i - \sum_{j\neq i} \mu_{ij}(\beta)$ and repeat a similar process). Our strategy for the proof of consistency crucially depends on the existence of the inverse of V, which requires that V is a full rank matrix. It is easy to show that V is positively semi-definite. Thus, if V has a full rank, then V must be positively definite. The following lemma assures the existence of the inverse of V.

Lemma 1. *Assume that* $\min_{i,j} \mu'_{ij}(\beta) > 0$. *With probability at least* $1 - (1 - p_n)^{nT}$, $H'(\beta)$ *is positively definite.*

Because $\log(1 - x) \leq -x$ when $x \in (0, 1)$, we have

$$e^{nT\log(1-p_n)} \leq e^{-p_n Tn}.$$

The probability of the nonexistence of V^{-1} is less than $e^{-p_n Tn}$, going exponentially fast to zero. Generally, the inverse of V does not have a closed form. Ref. [20] proposed to approximate the inverse of V, V^{-1}, by the matrix $S = (s_{ij})_{n\times n}$, where

$$s_{ij} = \frac{\delta_{ij}}{v_{ii}} + \frac{1}{v_{00}}. \tag{7}$$

In the above equation, $\delta_{ij} = 1$ if $i = j$; otherwise, $\delta_{ij} = 0$. By extending the proof of [20] to the sparse case, the upper bound of the approximate error $\|V^{-1} - S\|_{\max}$ is given in Lemma A2.

Recall that the main idea of the proof of the consistency in the Bradley–Terry model [10,14] contains two parts. Let $\hat{u}_i = e^{\hat{\beta}_i}$, $u_i = e^{\beta_i}$, $i_0 = \arg\max_i \hat{u}_i/u_i$ and $i_1 = \arg\min_i \hat{u}_i/u_i$. Since $\hat{u}_0/u_0 = 1$, it suffices to show that the ratio of subject i_0, $\hat{u}_{i_0}/u_{i_0}$, and the ratio of i_1, $\hat{u}_{i_1}/u_{i_1}$ are very close. With the nice mathematical properties of the logistic function $\mu(x) = e^x/(1 + e^x)$, the first part is to show that there are a number of subjects satisfying the following inequalities:

$$b \sum_{\{j: t_{i_0,j} > 0\}} (\hat{u}_{i_0}/u_{i_0} - \hat{u}_j/u_j) \leq c,\ \ b \sum_{\{j: t_{i_1,j} > 0\}} (\hat{u}_j/u_j - \hat{u}_{i_0}/u_{i_0}) \leq c, \tag{8}$$

where b and c are certain numbers. The second part is to eliminate common terms $\hat{u}_j/u_j$ based on the condition that the number of the common neighbors between any two subjects, $\min_{i,j} \#\{k : t_{ik} > 0, t_{jk} > 0\}$, is at least τn, where $\tau = 1$ in [10] and $\tau \in (0, 1)$ in [14]. In the Erdös–Rényi comparison graph, [13] further showed that there is at least one subject with its ratio close to both $\hat{u}_{i_0}/u_{i_0}$ and $\hat{u}_{i_1}/u_{i_1}$.

The aforementioned strategies for the proof of consistency are built on the the premise of the existence of the MLE, which is guaranteed by the necessary and sufficient condition that the directed graph with the win–loss matrix as its adjacency matrix is strongly con-

nected [19]. As discussed before, it may be difficult to find the minimal sufficient condition to guarantee the existence of $\widehat{\beta}$ in general paired comparison models. To overcome this difficulty, we aim to obtain the convergence rate of the Newton iterative sequence for solving Equation (2). Under the well-known Newton–Kantorich conditions, the Newton iterative sequence converges, and its limiting point is the solution. We apply an adjusted version of the Newton–Kantorich theorem in [21] to this end, which not only guarantees the existence of the solution but also gives an optimal error bound for the Newton iterative sequence.

Now, we formally state the consistency result.

Theorem 1. *Assume that Conditions (3a), (3b) and (3c) hold. If* $b_{n1}^4 b_{n2}/(b_{n0}^6 p_n^4) = o((n/\log n)^{1/2})$, *then* $\widehat{\beta}$ *exists with probability approaching one and is uniformly consistent in the sense that*

$$\|\widehat{\beta} - \beta^*\|_\infty = O_p\left(\frac{b_{n1}^2}{b_{n0}^3 p_n^2}\sqrt{\frac{\log n}{n}}\right) = o_p(1). \tag{9}$$

To see how small p_n could be, we consider a special case that β^* is a constant vector, in which b_{n0}, b_{n1} and b_{n2} are also constants. According to the above theorem, if $p_n > O((\log n/n)^{1/8})$, then $\|\widehat{\beta} - \beta^*\|_\infty = O_p(p_n^{-2}(\log n/n)^{1/2})$.

3.2. Asymptotic Normality of $\widehat{\beta}$

We establish the asymptotic distribution of $\widehat{\beta}$ by characterizing its asymptotical representation. In detail, we apply a second-order Taylor expansion to $H(\widehat{\beta})$ and find that $\widehat{\beta} - \beta^*$ can be represented as the sum of a main term $V^{-1}(a - \mathbb{E}^* a)$ and an asymptotically neglected remainder term, where $\mathbb{E}^*$ denotes the conditional expectation conditional on $\{t_{ij} : i, j = 0, \ldots, n\}$. Because V^{-1} does not have a closed form, we use the matrix S defined in (7) to approximate it. We formally state the asymptotic normality of $\widehat{\beta}$ as follows.

Theorem 2. *Let* $V = \partial H(\beta^*)/\partial\beta$ *and* $U = (u_{ij}) := \mathrm{Var}(a|t_{ij}, 0 \le i, j \le n)$. *If* $b_{n2} b_{n1}^6 b_{n0}^{-9} p_n^{-6} = o(n^{1/2}/\log n)$, *then for fixed* k, *the vector* $((\widehat{\beta}_1 - \beta^*), \ldots, (\widehat{\beta}_k - \beta_k^*))$ *follows a* k*-dimensional multivariate normal distribution with mean zero and the covariance matrix* $\Sigma = (\sigma_{ij})_{k\times k}$, *where*

$$\sigma_{ij} = \frac{\delta_{ij} u_{ii}}{v_{ii}^2} + \frac{u_{00}}{v_{00}^2}. \tag{10}$$

Remark 1. *If* $U = V$, *then* σ_{ij} *is equal to* $\delta_{ij}/v_{ii} + 1/v_{00}$. *When* $F(\cdot)$ *belongs to the exponential family distribution (e.g., the Bradley–Terry model),* U *is identical to* V. *If* $U \neq V$, *then the asymptotic variance of* $\widehat{\beta}_i$ *is involved with an additional factor* u_{ii}. *The asymptotic variance of* $\widehat{\beta}_i$ *is on the order of* $(np_n)^{-1/2}$ *if* β^* *is bounded above by a constant.*

4. Application to the Thurston Model

In this section, we illustrate the unified theoretical result by the application to the Thurston model.

The original Thurston model has a variance σ^2 in the normal distribution, i.e., the probability that subject i is preferred over j is $F((\beta_i - \beta_j)/\sigma)$. Since the merit parameters are scale invariable, we simply set $\sigma = 1$ hereafter. Recall that $\phi(x) = (2\pi)^{1/2} e^{-x^2/2}$ is the standard normal density function and $\Phi(x) = \int_{-\infty}^{x} \phi(x)dx$ is the distribution function of the standard normality. In the Thurston model, $\mu(x) = \Phi(x)$. Then,

$$\mu'(x) = \phi(x), \mu''(x) = \frac{x}{\sqrt{2\pi}} e^{-x^2/2}.$$

Since $\phi(x) = (2\pi)^{1/2}e^{-x^2/2}$ is an decreasing function on $|x|$, we have when $|x| \le Q_n$,

$$\frac{1}{2\pi}e^{-Q_n^2/2} \le \phi(x) \le \frac{1}{2\pi}.$$

Let $h(x) = xe^{-x^2/2}$. Then, $h'(x) = (1-x^2)e^{-x^2/2}$. Therefore, when $x \in (0,1)$, $h(x)$ is an increasing function on its argument x; when $x \in (1,\infty)$, $h(x)$ is an decreasing function on x. As a result, $h(x)$ attains its maximum value at $x = 1$ when $x > 0$. Since $h(x)$ is a symmetric function, we have $|h(x)| \le e^{-1/2} \approx 0.6$. Therefore,

$$b_{n0} = \frac{1}{2\pi}e^{-(\|\beta^*\|_\infty+\epsilon_n)^2/2}, b_{n1} = \frac{1}{2\pi}, b_{n2} = (2\pi e)^{-1/2}.$$

In view of Theorems 1 and 2, we have the following corollary.

Corollary 1. *If $b_{n1}^4 b_{n2}/(b_{n0}^6 p_n^2) = o((n/\log n)^{1/2})$ and $p_n > 24\log n/n$, then $\widehat{\beta}$ exists with probability approaching one and is uniformly consistent in the sense that*

$$\|\widehat{\beta} - \beta^*\|_\infty = O_p\left(\frac{b_{n1}^2}{b_{n0}^3 p_n}\sqrt{\frac{\log n}{n}}\right) = o_p(1). \tag{11}$$

Let $V = \partial H(\beta^*)/\partial\beta$. If $b_{n2}b_{n1}^6 b_{n0}^{-9} = o(n^{1/2}/\log n)$, then for fixed k, the vector $((\widehat{\beta}_1 - \beta^*), \ldots, (\widehat{\beta}_k - \beta_k^*))$ follows a k-dimensional multivariate normal distribution with mean zero and the covariance matrix $\Sigma = (\sigma_{ij})_{k\times k}$ defined at (10).

5. Extension to a Fixed Sparse Design

In some applications such as sports, the comparison graph may be fixed, not random. For example, in the regular season of the National Football League (NFL), games are scheduled in advance. More specially, there are 32 teams in the 2 conferences of the NFL that are divided into 8 divisions, each consisting of 4 teams. In the regular season, each team plays 16 matches, 6 within the division and 10 between the divisions. Motivated by the design, ref. [14] proposed a sparse condition to control the length from one subject to another subject with 2 or 3:

$$\tau_n := \min_{0\le i<j\le n} \frac{\#\{k : t_{ik} > 0, t_{jk} > 0\}}{t}.$$

That is, τ_n is the minimum ratio of the total number of paths between any i and j with length 2 or 3. Under the Erdös–Rényi comparison graph, there are similar sparsities. Specifically, the set of common neighbors of any two subjects i and j has at least the following size:

$$\#\{k : t_{ik} > 0, t_{jk} > 0\} \ge \tfrac{1}{2}(n-1)p_n^2,$$

with a probability of at least $1 - O(1/n)$ if $np_n^2 \ge 24\log n$; see (A12) in the proof of A2.

We assume that if two subjects have comparisons, they are compared T times, in accordance with the aforementioned setting for easy of exposition. Similar to Lemma A2, the approximate error of using S to approximate V^{-1} is

$$\|V^{-1} - S\|_{\max} \le \frac{2T^2 b_{n1}^2 \rho_{\max}}{b_{n0}^3 \tau_n^3 n^2},$$

where $\rho_{\max} = t_{\max}/n$ and $t_{\max} = \max_i t_i$. With similar lines of argument as in the proofs of Theorems 1 and 2, we have the following theorem, whose proof is omitted.

Theorem 3. *Assume that conditions (3a), (3b) and (3c) hold. If* $b_{n1}^4 b_{n2}/(b_{n0}^6\tau_n^2) = o((n/\log n)^{1/2})$, *then* $\widehat{\beta}$ *exists with probability approaching one and is uniformly consistent in the sense that*

$$\|\widehat{\beta} - \beta^*\|_\infty = O_p\left(\frac{b_{n1}^2}{b_{n0}^3\tau_n}\sqrt{\frac{\log n}{n}}\right) = o_p(1). \tag{12}$$

Let $V = \partial H(\beta^*)/\partial\beta$ *and* $U = (u_{ij}) := \mathrm{Var}(a|t_{ij}, 0 \le i, j \le n)$. *If* $b_{n2}b_{n1}^6 b_{n0}^{-9}\tau_n^{-3} = o(n^{1/2}/\log n)$, *then for fixed* k, *the vector* $((\widehat{\beta}_1 - \beta^*), \ldots, (\widehat{\beta}_k - \beta_k^*))$ *follows a* k*-dimensional multivariate normal distribution with mean zero and the covariance matrix* $\Sigma = (\sigma_{ij})_{k\times k}$, *where* σ_{ij} *is given in (7).*

6. Numerical Studies

In this section, we evaluate the asymptotic results of the moment estimator in the Thurston model through simulation studies and a real data example.

6.1. Simulation Studies

We carry out simulations to evaluate the finite sample performance of the moment estimator in the Thurston model. We set $T = 1$, which means that any pair has one comparison with probability p_n and no comparison with probability $1 - p_n$. Let c be a constant. We set the merit parameters to be a linear form, i.e., $\beta_i^* = ic\log n/n$ for $i = 1, \ldots, n$, where $\beta_0^* = 0$. We considered four different values for c as $c = 0.3, 0.5, 0.8$. By allowing $\|\beta^*\|_\infty$ to grow with n, we intended to assess the asymptotic properties under different asymptotic regimes.

In order to see how small p_n could be, we first evaluate the fail frequency that the "win–loss" graph G_n is strongly connected, which is the necessary condition to guarantee the moment estimator exists. We set c to be fixed with $c = 0.4$. The results are shown in Table 1 with 1000. We can see that the necessary condition did not hold in each simulation when $p_n = \log n/n$, while it holds with almost 100% frequency when $p_n = (\log n/n)^{1/2}$. This shows that it is necessary to control the rate of p_n tending to zero.

Table 1. The fail frequency (×100%).

p_n	$n = 100$	$n = 500$	$n = 1000$
$(\log n/n)^{1/2}$	0.5	0	0
$(\log n/n)^{2/3}$	25.8	0.2	0
$\log n/n$	100	100	100

Based on Theorem 2, $\hat{\xi}_{ij} = [\widehat{\beta}_i - \widehat{\beta}_j - (\beta_i^* - \beta_j^*)]/(\hat{u}_{ii}/\hat{v}_{ii}^2 + \hat{u}_{00}/\hat{v}_{00}^2)^{1/2}$ converges in distribution to the standard normality, where $\hat{u}_{ii}$ and $\hat{v}_{i,i}$ are the estimates of u_{ii} and v_{ii} by replacing β^* with $\widehat{\beta}$. Therefore, we assessed the asymptotic normality of $\hat{\xi}_{ij}$ via the coverage probability of the 95% confidence interval and the length of the confidence interval. The times that $\widehat{\beta}$ failed to exist were also recorded. Two values, $n = 100$ and $n = 200$, were considered for the number of subjects. Each simulation was repeated 10,000 times.

The simulation results are shown in Table 2. When $p_n = (\log n/n)^{1/4}$, all simulated coverage probabilities are very close to the target level 95%. On the other hand, when $p_n = (\log n/n)^{1/2}$, they are a little lower than the normal level in the case $n = 100$ and are very close to 95% in the case $n = 200$. The length of the confidence interval decreases as n increases, which qualitatively agrees with the theory. It is also expected that the length of the confidence interval increases as p_n decreases when n is fixed. Another phenomenon is that the length of the confidence interval under three distinct c has little difference when n and p_n are fixed.

Table 2. The reported values are the coverage frequency (×100%) for $\beta_i - \beta_j$ for a pair (i, j)/length of the confidence interval/fail probabilities (×100%).

$p_n = (\log n/n)^{1/4}$				
n	i	$c = 0.2$	$c = 0.5$	$c = 0.8$
100	$(1, 2)$	94.5/1.04/0	94.62/1.06/0	94.62/1.06/0
	$(49, 50)$	94.97/1.04/0	94.34/1.04/0	94.34/1.04/0
	$(99, 100)$	95.12/1.04/0	94.99/1.06/0	94.99/1.06/0
200	$(1, 2)$	95.18/0.78/0	94.71/0.79/0	94.71/0.79/0
	$(99, 100)$	95.14/0.78/0	94.49/0.78/0	94.49/0.78/0
	$(199, 200)$	94.82/0.78/0	94.64/0.79/0	94.64/0.79/0
$p_n = (\log n/n)^{1/2}$				
n	i	$c = 0.2$	$c = 0.4$	$c = 0.6$
100	$(1, 2)$	93.28/1.58/0.19	93.48/1.60/0.49	93.48/1.60/1.15
	$(49, 50)$	93.85/1.58/0.19	93.94/1.58/0.49	93.94/1.58/1.15
	$(99, 100)$	93.80/1.58/0.19	93.96/1.61/0.49	93.96/1.61/1.15
200	$(1, 2)$	94.04/1.26/0	93.89/1.28/0	93.89/1.28/0
	$(99, 100)$	94.14/1.26/0	94.22/1.26/0	94.22/1.26/0
	$(199, 200)$	94.38/1.26/0	93.98/1.28/0	93.98/1.28/0

6.2. A Real Data Example

We use the 2018 NFL regular season data as an illustrative example, which is available from https://www.espn.com/nfl/schedule/_/year/2018 (accessed on 20 August 2025). The NFL league consists of thirty-two teams that are divided evenly into two conferences, and each conference has four divisions that have four teams each. In the regular season, each team plays with three intra-division teams (each twice) and ten games with ten inter-division teams (each once). As discussed in [14], the design of the NFL regular season satisfies the sparsity condition of the fixed comparison graph, where $\tau_n = 1/16$. We removed two ties before our analysis. The fitted merits that were obtained from fitting the Thurstone model for the remaining data are given in Table 3, where we used "Arizona Cardinals" with the smallest number of wins as the baseline (with $\hat{\beta}_0 = 0$).

It is interesting to compare the ordering of six playoff seeds of the two conferences with the NFL rules with the ordering by their fitted merits in Table 3. The NFL rules are based on the regular season won–lost percentage record (PCT) and can be briefly summarized as follows: the teams in each division with the best PCT are seeded one through four; another two teams from each conference are seeded five and six based on their PCT. The six playoff seeds in the American Football Conference from No. 1 to No. 6 based on the PCT are Kansas City Chiefs, New England Patriots, Pittsburgh Steelers, Houston Texans, Los Angeles Chargers, and Indianapolis Colts, while the selected teams based on the fitted merits are Kansas City Chiefs, New England Patriots, Houston Texans, Baltimore Ravens, Los Angeles Chargers, and Pittsburgh Steelers. The corresponding six playoff seeds in the National Football Conference based on the PCT are Los Angeles Rams, New Orleans Saints, Chicago Bears, Washington Redskins, Seattle Seahawks, and Carolina Panthers, while the selected teams base on the fitted merits are Los Angeles Rams, New Orleans Saints, Chicago Bears, Dallas Cow boys, Seattle Seahawks, and Philadelphia Eagles. As we can see, the selected top three teams based on the PCT and the fitted merits in each conference are the same, and the selected teams from No. 4, No. 5 and No. 6 are not all the same.

Table 3. The fitted merit $\hat{\beta}_i$, the number of wins a_i, and the standard error $\hat{\sigma}_i$.

	American Football Conference					National Football Conference			
Division	**Team**	$\hat{\beta}_i$	a_i	$\hat{\sigma}_i$	**Division**	**Team**	$\hat{\beta}_i$	a_i	$\hat{\sigma}_i$
East	New England Patriots	1.452	11	0.519	East	Dallas Cow boys	1.284	10	0.512
	New York Jets	0.338	4	0.530		Philadelphia Eagles	1.193	9	0.511
	Miami Dophins	0.718	7	0.509		New York Giants	0.423	5	0.514
	Buffalo Bills	0.673	6	0.514		Washington Redskins	0.756	7	0.511
North	Baltimore Ravens	1.382	10	0.514	North	Chicago Bears	1.494	12	0.532
	Cincinnati Bengals	0.769	6	0.514		Green Bay Packers	0.595	6	0.522
	Pittsburgh Steelers	1.337	9	0.52		Minnesota Vikings	1.059	8	0.523
	Cleveland Browns	0.968	7	0.519		Detroit Lions	0.579	6	0.516
South	Indianapolis Colts	1.244	10	0.512	South	New Orleans Saints	1.908	13	0.544
	Houston Texans	1.430	11	0.516		Atlanta Falcons	0.767	7	0.512
	Tennessee Titans	1.205	9	0.51		Carolina Panthers	0.840	7	0.511
	Jacksonville Jaguars	0.591	5	0.517		Tampa Bay Buccaneers	0.506	5	0.52
West	Kansas City Chiefs	1.762	12	0.537	West	Los Angeles Rams	1.963	13	0.559
	Denver Broncos	0.713	6	0.523		San Francisco 49ers	0.166	4	0.54
	Oakland Raiders	0.365	4	0.537		Seattle Seahawks	1.305	10	0.525
	Los Angeles Chargers	1.748	12	0.536		Arizona Cardinals	0	3	0.555

7. Summary and Discussion

We have presented the moment estimation based on the scores of subjects in the paired comparison model under sparse comparison graphs. We have established the uniform consistency and asymptotic normality of the moment estimator. The consistency is shown by obtaining the convergence rate of the Newton iterative sequence. This leads to a condition on the sparsity parameter p_n requiring that $p_n \geq O((\log n/n)^{1/8})$ if β^* is a constant vector. We note that this condition looks much stronger than that in the Bradley–Terry model in [13]. Since we consider a general model, it would seem to be suitable that a more severe condition is imposed. On the other hand, the condition imposed on b_{n0} may not be the best possible condition. In particular, the conditions for guaranteeing the asymptotic normality seem stronger than those needed for the consistency. Note that the asymptotic behavior of the moment estimator depends not only on b_{n0} but also on the configuration of all parameters. It would be of interest to investigate whether these conditions could be relaxed.

In this paper, we assume that given the comparison graph, all paired comparisons are independent. Note that the moment equation holds regardless of whether comparisons are independent.

When comparisons are not independent, the moment estimation still works. The consistency result in Theorem 1 still holds as long as there is the same order of the upper bound of $a_i - \mathbb{E}^* a_i$ in Lemma A4. In fact, the independence assumption is not directly used when checking our proofs. It is only used in Lemma A4 to derive the upper bound of $a_i - \mathbb{E}^* a_i$ using the Hoeffding inequality. Analogously, the independence assumption is used to derive the central limit theorem of $a_i - \mathbb{E}^* a_i$. In the dependence case, there are also many Hoeffding-type exponential tail inequalities (e.g., [22–24]) and cental limit theorems for sums of a sequence of random variables (e.g., [25,26]) to apply.

Building on the theoretical and empirical findings of this study, we identify two promising avenues for further exploration, which not only address the current limitations but also extend the moment estimation framework to more complex and practical scenarios: 1. Moment estimation for paired comparison models with dependent outcomes. This paper assumes that all paired comparison outcomes are independent, which is a standard but

restrictive assumption in many real-world settings (e.g., crowdsourcing evaluations where raters may have consistent biases, or sports leagues where team performance is serially correlated). A natural extension is to develop moment estimation methods for models with dependent outcomes, such as incorporating Markovian dependence or exchangeable correlation structures. Key challenges include deriving unbiased moment equations under dependence and establishing asymptotic properties (consistency, asymptotic normality) using tools from dependent random variable theory (e.g., Hoeffding-type inequalities for associated variables [22–24]). 2. High-dimensional sparse paired comparison models with structured merit parameters. This study focuses on unstructured merit parameters (i.e., β_i are independent). However, in many applications, merit parameters often exhibit inherent structures, such as group-level homogeneity (e.g., teams in the same sports division share similar strengths) or sparsity (e.g., only a few subjects have distinct merits in large-scale crowdsourcing). Extending the moment estimation framework to incorporate such structures (e.g., group lasso-penalized moment estimation, sparse merit parameter inference) would improve estimation efficiency. Key research questions include designing computationally feasible penalized moment equations and establishing oracle properties for the structured estimators. These directions align with the core theme of sparse paired comparison inference and address practical limitations of the current work. We believe pursuing these avenues will not only extend the theoretical scope of moment estimation but also broaden its applicability to more complex real-world problems.

Author Contributions: Conceptualization, Q.W.; Data curation, Q.W.; Formal analysis, Q.W.; Funding acquisition, Q.W.; Investigation, Q.W. and L.P.; Methodology, Q.W. and T.Y.; Software, Q.W. and L.P.; Supervision, T.Y.; Validation, Q.W. and T.Y.; Visualization, Q.W. and L.P.; Writing—original draft, Q.W.; Writing—review and editing, Q.W., L.P. and T.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 12301386.

Informed Consent Statement: Not Applicable.

Data Availability Statement: We use the 2018 NFL regular season data as an example, which is available from https://www.espn.com/nfl/schedule/_/year/2018 (accessed on 20 August 2025).

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

In this section, we present the proofs of the theorems.

Appendix A.1. Preliminaries

We state two preliminary results firstly, which will be used in the proofs. The first result is the optimal error bound in the Newton method in [21] under the Kantorovich conditions [27].

Lemma A1 ([21]). *Let X and Y be Banach spaces, D be an open convex subset of X and $F: D \subseteq X \to Y$ be Fréchet differentiable. Assume that, at some $x_0 \in D$, $F'(x_0)$ is invertible and that*

$$\|F'(x_0)^{-1}(F'(x) - F'(y))\| \leq K\|x - y\|, \quad x, y \in D, \tag{A1}$$

$$\|F'(x_0)^{-1}F(x_0)\| \leq \eta, h = K\eta \leq 1/2, \tag{A2}$$

$$\bar{S}(x_0, t^*) \subseteq D, t^* = 2\eta/(1 + \sqrt{1 - 2h}).$$

Then: (1) The Newton iterates $x_{n+1} = x_n - F'(x_n)^{-1}F(x_n)$, $n \geq 0$ are well-defined, lie in $\bar{S}(x_0, t^*)$ and converge to a solution x^* of $F(x) = 0$.

(2) The solution x^* is unique in $S(x_0, t^{**}) \cap D$, $t^{**} = (1+\sqrt{1-2h})/K$ if $2h < 1$ and in $\bar{S}(x_0, t^{**})$ if $2h = 1$.
(3) $\|x^* - x_n\| \le t^*$ if $n = 0$ and $\|x^* - x_n\| \le 2^{1-n}(2h)^{2^n-1}\eta$ if $n \ge 1$.

The second result is the approximate error of using S to approximate V^{-1}, whose proof is in the Appendix B.

Lemma A2. *If $p_n \ge 24\log n/n$, then for sufficiently large n with a probability of at least $1 - O(n^{-1})$, we have*

$$\|V^{-1} - S\|_{\max} \le \frac{12Tb_{n1}^2}{b_{n0}^3 n(n-1)p_n^3}.$$

Appendix A.2. Proof of Theorem 1

We aim to show Theorem 1 by obtaining the convergence rate of the Newton iterative sequence in view of Lemma A1, which requires us to verify the Kantovorich conditions (A1) and (A2). Condition (A1) depends on the Lipschitz continuous form of $H_i'(\beta)$. Recall that $t_{\max} = \max_{i=0,\ldots,n} t_i$ and $t_{\min} = \min_{i=0,\ldots,n} t_i$.

Lemma A3. *Let $D = B(\beta^*, \epsilon_n)(\subset \mathbb{R}^n)$ be an open convex set containing the true point β^*. For any given set $\{t_{ij}, 0 \le i, j \le n\}$, if Inequality (3c) holds, then*

$$\max_{i=0,\ldots,n} \|H_i'(x) - H_i'(y)\|_1 \le 4b_{n2}t_{\max}\|x - y\|_\infty.$$

Moreover, Condition (A1) also depends on the magnitudes of $|a_i - \mathbb{E}(a_i|t_{ij}, j = 0, \ldots n)|$, $i = 0, \ldots, n$, which are stated below.

Lemma A4. *With a probability of at least $1 - O(1/n)$, we have*

$$\max_{i=0,\ldots,n} |a_i - \mathbb{E}(a_i|t_{ij}, j = 0, \ldots, n)| \le \sqrt{2\log n t_{\max}}.$$

The following results are the lower bound of $t_{\min}$ and the upper bounds of $t_{\max}$ and of $\sum_i t_i$.

Lemma A5. *(1)With probability at least $1 - (n+1)\exp(-\frac{1}{8}nTp_n)$,*

$$t_{\min} = \min_{i=0,\ldots,n} t_i \ge \tfrac{1}{2}nTp_n.$$

(2) With probability at least $1 - (n+1)\exp(-\frac{1}{10}nTp_n)$,

$$t_{\max} = \max_{i=0,\ldots,n} t_i \le \tfrac{3}{2}nTp_n.$$

(3) With probability at least $1 - \exp(-\frac{1}{10}n(n+1)Tp_n$,

$$\sum_{i=0}^{n} t_i \le 3n(n+1)Tp_n.$$

We are now ready to prove Theorem 1.

Proof of Theorem 1. Note that $\widehat{\beta}$ is the solution to the equation $H(\beta)$=0. We prove the consistency by obtaining the convergence rate of the Newton iterative sequence: $\beta^{(k+1)} = \beta^{(k)} - [H'(\beta^{(k)})]^{-1}H(\beta^{(k)})$, where we set $\beta^{(0)} := \beta^*$. To apply Lemma A1, we choose the convex set $D = B(\beta^*, \epsilon_n)$. The following calculations are based on the event E_n:

$$\{t_{ij}, 0 \le i,j \le n : \max_i |a_i - \mathbb{E}(a_i|t_{ij}, j = 0,\ldots,n)| \le \sqrt{2t_{\max}\log n},$$
$$H'(\beta) > 0,\ \ t_{\min} \ge \tfrac{T}{2}np_n,\ \ t_{\max} \le \tfrac{3}{2}nTp_n\}.$$

Note that $b_{n0} \le |\mu'_{ij}(\beta)| \le b_{n1}$ when $\beta \in B(\beta^*, \epsilon_n)$. Let $V = (v_{ij})_{n\times n} = H'(\beta^*)$. We use S defined in (7) to approximate V^{-1} and let $W = V^{-1} - S$. We verify the Kantovorich conditions in Lemma A1 as follows. Since $\sum_{i=0}^n H_i(\beta) = 0$, we have

$$\sum_{i=1}^n H_i(\beta) = -H_0(\beta).$$

Based on Lemma A3, we have

$$\begin{aligned}
& \|[H'(\beta^*)]^{-1}[H'(x) - H'(y)]\|_\infty \\
\le\ & \|S[H'(x) - H'(y)]\|_\infty + \|W[H'(x) - H'(y)]\|_\infty \\
\le\ & \max_{i=1,\ldots,n}\frac{1}{v_{ii}}\|H'_i(x) - H'_i(y)\|_1 + \frac{1}{v_{00}}\|H'_0(x) - H'_0(y)\|_1 + \|W\|_\infty\|H'(x) - H'(y)\|_\infty \\
\le\ & [\frac{2}{b_{n0}t_{\min}} + n\cdot\frac{12b_{n1}^2T}{n(n-1)b_{n0}^3p_n^3}] \times 4b_{n2}t_{\max} \times \|x - y\|_\infty \\
=\ & O(\frac{b_{n1}^2b_{n2}}{b_{n0}^3p_n^2}) \times \|x - y\|_\infty.
\end{aligned}$$

Thus, we can set $K = O(b_{n1}^2b_{n2}b_{n0}^{-3}\eta_n^{-1})$ in (A1). Again, based on the event E_n, we have

$$\begin{aligned}
\eta &= \|[H'(\beta^*)]^{-1}H(\beta^*)\|_\infty \\
&\le n\|V^{-1} - S\|_{\max}\|H(\beta^*)\|_\infty + \max_{i=1,\ldots,n}\frac{|H_i(\beta^*)|}{v_{ii}} + \frac{|H_0(\beta^*)|}{v_{00}} \\
&\le \left[O(\frac{b_{n1}^2}{nb_{n0}^3p_n^3}) + O(\frac{1}{b_{n0}t_{\min}})\right] \times O((t_{\max}\log n)^{1/2}) \\
&= O\left(\frac{b_{n1}^2}{b_{n0}^3p_n^2}\sqrt{\frac{\log n}{n}}\right).
\end{aligned}$$

If

$$K\eta = O\left(\frac{b_{n1}^4b_{n2}}{b_{n0}^6p_n^4}\sqrt{\frac{\log n}{n}}\right) = o(1), \tag{A3}$$

then this verifies Condition (A2). Based on Lemma A1, $\lim_k \beta^{(k)}$ exists, denoted by $\widehat{\beta}$, and it satisfies

$$\|\widehat{\beta} - \beta^*\|_\infty = O\left(\frac{b_{n1}^2}{b_{n0}^3p_n^2}\sqrt{\frac{\log n}{n}}\right).$$

Based on Lemmas 1, A2, A4 and A5, the event E_n holds with a probability of at least $1 - O(n^{-1})$ if $p_n \ge 24\log n/n$. Note that (A3) implies $p_n \ge 24\log n/n$. This completes the proof. □

Appendix A.3. Proofs for Theorem 2

Write Var^* and $\mathbb{E}^*$ as the conditional variance and conditional expectation given t_{ij} for $0 \le i,j \le n$. Let $U = (u_{ij}) := \mathrm{Var}^*(a)$. In the Bradley–Terry model, $U = H'(\beta^*)$. Note that a_i is a sum of t_i independent Bernoulli random variables. Based on Lemma A5, we know that $\min_i t_i = O_p(np_n)$. Let $\sigma_{\min} = \min_{i\ne j} p_{ij}(1 - p_{ij})$. If $np_n\sigma_{\min} \to \infty$, then $\min_i u_{ii} \to \infty$. Based on the central limit theorem in the bound case, as in [28] (p. 289), if $np_n\sigma_{\min} \to \infty$, then $u_{ii}^{-1/2}\{a_i - \mathbb{E}^*(a_i)\}$ converges in distribution to the standard normal distribution. When considering the asymptotic behaviors of the vector $(a_1,\ldots,a_r)$ with

a fixed r, one could replace the degrees $a_1, \ldots, a_r$ by the independent random variables $\tilde{a}_i = a_{i,r+1} + \ldots + a_{in}$, $i = 1, \ldots, r$. Therefore, we have the following proposition.

Proposition A1. *If $np_n\sigma_{\min} \to \infty$, then as $n \to \infty$, for any fixed $r \geq 1$, the components of $(a_1 - \mathbb{E}^*(a_1), \ldots, a_r - \mathbb{E}^*(a_r))$ are asymptotically independent and normally distributed with variances $u_{11}, \ldots, u_{rr}$, respectively. Moreover, the first r rows of $S(a - \mathbb{E}^*(a))$ are asymptotically normal with covariance matrix $\Sigma = (\sigma_{ij})$, where*

$$\sigma_{ij} = \frac{\delta_{ij}u_{ii}}{v_{ii}^2} + \frac{u_{00}}{v_{00}^2}.$$

Lemma A6. *Let $V = H'(\beta^*)$ and $W = V^{-1} - S$ and $\mathrm{Cov}^*(\cdot) = \mathrm{Cov}(\cdot|t_{ij}, 0 \leq i, j \leq n)$. Then,*

$$\mathrm{Cov}^*(WH'(\beta^*)) = O_p(\frac{b_{n1}^5}{n^2 b_{n0}^6 p_n^5}).$$

Further, if $U = H'(\beta^*)$, then

$$\mathrm{Cov}^*(WH'(\beta^*)) = O_p(\frac{b_{n1}^2}{n^2 b_{n0}^3 p_n^5}).$$

Now, we are ready to prove Theorem 2.

Proof of Theorem 2. Let $\widehat{\pi}_{ij} = \widehat{\beta}_i - \widehat{\beta}_j$ and $\pi_{ij}^* = \beta_i^* - \beta_j^*$. Based on Theorem 1, $\widehat{\beta} \in B(\beta^*, \epsilon_n)$. To simplify notations, write $\mu_{ij}' = \mu'(\pi_{ij}^*)$. Based on a second-order Taylor expansion, we have

$$t_{ij}\mu(\widehat{\pi}_{ij}) - t_{ij}\mu(\pi_{ij}^*) = t_{ij}\mu_{ij}'(\widehat{\beta}_i - \beta_i^*) - t_{ij}\mu_{ij}'(\widehat{\beta}_j - \beta_j^*) + g_{ij}, \ \ i \neq j, \tag{A4}$$

where g_{ij} is the second-order remainder term:

$$g_{ij} = t_{ij}\mu''(\tilde{\pi}_{ij})[(\widehat{\beta}_i - \beta_i)^2 + (\widehat{\beta}_j - \beta_j)^2 - 2(\widehat{\beta}_i - \beta_i)(\widehat{\beta}_j - \beta_j)].$$

In the above equation, $\tilde{\pi}_{ij}$ lies between π_{ij}^* and $\widehat{\pi}_{ij}$. If $b_{n1}^4 b_{n2} b_{n0}^{-6} p_n^{-4} = o((n/\log n)^{1/2})$, based on Theorem 1, we have

$$\|\widehat{\beta} - \beta^*\|_\infty = O_p\left(\frac{b_{n1}^2}{b_{n0}^3 p_n^2}\sqrt{\frac{\log n}{n}}\right).$$

Therefore, in view of (3c), $|\mu_{ij}''(\tilde{\pi}_{ij})| \leq t_{ij} b_{n2}$ such that

$$|g_{ij}| \leq 4b_{n2}t_{ij}\|\widehat{\beta} - \beta^*\|_\infty^2. \tag{A5}$$

Let $g_i = \sum_{j \neq i} g_{ij}$, $i = 0, \ldots, n$, and $g = (g_1, \ldots, g_n)^\top$. Then, based on Lemma A5 (2), we have

$$\max_{i=0,\ldots,n} |g_i| = 4b_{n2}t_{\max} \cdot O_p\left(\frac{b_{n1}^4 \log n}{n b_{n0}^6 p_n^4}\right) = O_p\left(\frac{b_{n1}^4 b_{n2} \log n}{b_{n0}^6 p_n^3}\right). \tag{A6}$$

By writing the equation in (A4) into a matrix form, we have

$$\mathbb{E}^* a - a = V(\widehat{\beta} - \beta^*) + g. \tag{A7}$$

Equivalently,

$$\widehat{\beta} - \beta^* = V^{-1}(\mathbb{E}^* a - a) + V^{-1}g. \tag{A8}$$

Similarly, we have

$$\begin{aligned}\mathbb{E}^* a_0 - a_0 &= \frac{\partial H_0(\beta^*)}{\partial \beta}(\widehat{\beta} - \beta^*) + \frac{1}{2}(\widehat{\beta} - \beta^*)^\top \frac{\partial^2 H_0(\tilde{\beta})}{\partial \beta \partial \beta^\top}(\widehat{\beta} - \beta^*) \\ &= -\sum_{i=1}^{n} v_{i0}(\widehat{\beta}_i - \beta_i^*) + \frac{1}{2} v_{i0}(\widehat{\beta}_i - \beta_i^*)^2.\end{aligned}$$

Therefore, based on Lemma A5 (2), we have

$$|\mathbb{E}^* a_0 - a_0 + \sum_{i=1}^{n} v_{i0}(\widehat{\beta}_i - \beta_i^*)| = O_p\left(\frac{b_{n1}^4 t_{\max} \log n}{n b_{n0}^6 p_n^4}\right) = O_p\left(\frac{b_{n1}^4 \log n}{b_{n0}^6 p_n^3}\right).$$

Note that $\sum_{i=0}^{n} a_i - \mathbb{E}^*(a_i) = 0$. Multiplying both sides of (A7) by a row vector with all element 1 yields

$$\sum_{i=1}^{n} g_i = a_0 - \mathbb{E}^* a_0 + \sum_{i=1}^{n} v_{i0}(\widehat{\beta}_i - \beta_i^*).$$

Therefore, we have

$$|\sum_{i=1}^{n} g_i| = O_p\left(\frac{b_{n1}^4 \log n}{b_{n0}^6 p_n^3}\right). \tag{A9}$$

Based on (A6) and Lemma A2, we have

$$\begin{aligned}\|V^{-1} g\|_\infty &\leq \|Sg\|_\infty + \|(V^{-1} - S)g\|_\infty \\ &\leq \max_{i=1,\ldots,n} \frac{1}{v_{ii}}|g_i| + \frac{1}{v_{00}}|\sum_{i=1}^{n} g_i| + n\|V^{-1} - S\|_{\max}\|g\|_\infty \\ &= O_p\left(\frac{b_{n2}}{t_{\min} b_{n0}} + \frac{b_{n1}^2}{n b_{n0}^3 p_n^3}\right) \times O_p\left(\frac{b_{n1}^4 b_{n2} \log n}{b_{n0}^6 p_n^3}\right) \\ &= O_p\left(\frac{b_{n2} b_{n1}^6 \log n}{n b_{n0}^9 p_n^6}\right).\end{aligned}$$

If $b_{n2} b_{n1}^6 b_{n0}^{-9} p_n^{-6} = o(n^{1/2}/\log n)$, then we have

$$\widehat{\beta}_i - \beta_i^* = V^{-1}(\mathbb{E}^* a - a) + o_p(n^{-1/2}). \tag{A10}$$

Consequently, in view of Lemma A6, we have

$$\widehat{\beta}_i - \beta_i^* = [S(\mathbb{E}^* a - a)]_i + o_p(n^{-1/2}).$$

Therefore, Theorem 2 immediately comes from Proposition A1. □

Appendix B

In this section, we present the proofs of supported lemmas.

Appendix B.1. Proof of Lemma 1

Proof of Lemma 1. For an arbitrarily given nonzero vector $x = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$, direct calculations give

$$\begin{aligned}x^\top V x &= = \sum_{i=1}^{n} x_i^2 v_{ii} + \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} x_i v_{ij} x_j \\ &= -\sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} x_i^2 v_{ij} - \sum_{i=1}^{n} x_i^2 v_{i0} + \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} x_i v_{ij} x_j \\ &= -\tfrac{1}{2}\sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} (x_i - x_j)^2 v_{ij} - \sum_{i=1}^{n} x_i^2 v_{i0},\end{aligned}$$

where the second equality is due to that $v_{ii} = -\sum_{j\neq i} v_{ij}$. Therefore, $x^\top Vx = 0$ if and only if

$$x_i v_{i0} = 0, \;\; i = 1, \ldots, n, v_{ij}(x_i - x_j) = 0, 1 \leq i \neq j \leq n.$$

Because $\mu_{ij}(\beta) \neq 0$ and $v_{ij} = t_{ij}\mu_{ij}(\beta)$ for $i \neq j$, the above equations are identical to

$$x_i t_{i0} = 0, \;\; i = 1, \ldots, n, t_{ij}(x_i - x_j) = 0, 1 \leq i \neq j \leq n.$$

Let E be the event

$$\{\{t_{ij}\}_{i,j=0,i\neq j}^{n} : x_i t_{i0} = 0, \;\; i = 1, \ldots, n, t_{ij}(x_i - x_j) = 0, 1 \leq i \neq j \leq n\}.$$

To show Lemma 1, it is sufficient to obtain the lower bound of the probability of the event E. We will evaluate the probability of the event E under two cases: there exists some zero element in x and there are no zero elements in x.

Case I: We consider there are zero elements in x. Let $\{0, x_{i_1}, \ldots, x_{i_k}\}$ be $k+1$ different distinct values in $\{x_1, \ldots, x_n\}$, $\Omega_0 = \{i : x_i = 0\}$ and $\Omega_j = \{q : x_q = x_{i_j}\}$, $j = 1, \ldots, k$. Since $x \neq 0$, $k \geq 1$. It is clear that

$$|\Omega_0| > 0, \;\; |\Omega_j| > 0, j = 1, \ldots, k, \;\; \sum_{i=0}^{k} |\Omega_i| = n + 1, \tag{A11}$$

where $|\Omega_i|$ denotes the cardinality of Ω_i. Therefore, we have

$$\begin{aligned} \mathbb{P}(E) &= (1 - p_n)^{\sum_{j=1}^{k} |\Omega_j|} \times \prod_{0\leq i<j\leq k} (1 - p_n)^{T|\Omega_i||\Omega_j|} \\ &= (1 - p_n)^{\sum_{j=1}^{k} |\Omega_j| + \sum_{0\leq i<j\leq k} |\Omega_i||\Omega_j|}. \end{aligned}$$

To obtain the lower bound of $\mathbb{P}(E)$, it is sufficient to solve the minimizer of $\sum_{j=1}^{k} |\Omega_j| + \sum_{0\leq i<j\leq k} |\Omega_i||\Omega_j|$ under the restricted condition (A11). Let $y_i = |\Omega_i|$. Then,

$$\begin{aligned} &\sum_{j=1}^{k} y_i + \sum_{0\leq i<j\leq k} y_i y_j \\ =&\sum_{j=1}^{k} y_i + \tfrac{1}{2} \sum_{i=0}^{n} \sum_{j=0, j\neq i}^{k} y_i y_j \\ =&\tfrac{1}{2} \sum_{i=0}^{n} \sum_{j=0}^{n} y_i y_j - \tfrac{1}{2} \sum_{i=0}^{k} y_i^2 + \sum_{j=1}^{k} y_i \\ =&\tfrac{1}{2} (\sum_{i=0}^{n} y_i)^2 - \tfrac{1}{2} \sum_{i=1}^{k} (y_i - 1)^2 - \tfrac{1}{2} y_0^2 + \tfrac{1}{2} k \\ =&\tfrac{1}{2} ((n+1)^2 + k) - \tfrac{1}{2} (\sum_{i=0}^{k} z_i^2), \end{aligned}$$

where $z_0 = y_0$ and $z_i = y_i - 1$, $i = 1, \ldots, k$. Under the restriction $\sum_i z_i = n + 1 - k > 0$ and $z_i \geq 0$, the function $\sum_{i=0}^{k} z_i^2$ obtains its maximizer at such points $z = (0, \ldots, n + 1 - k, 0, \ldots, 0)$. Therefore, we have

$$\begin{aligned} &\sum_{j=1}^{k} y_i + \sum_{0\leq i<j\leq k} y_i y_j \\ \geq&\tfrac{1}{2} ((n+1)^2 + k - (n+1-k)^2) = \tfrac{1}{2} (2(n+1)k + k - k^2) \\ =&\tfrac{1}{2} [-(k - \frac{2(n+1)+1}{2})^2 + (n + 1 + \frac{1}{2})^2]. \end{aligned}$$

Because $1 \le k \le n$, the above function obtains its minimizer at $k = 1$. That is,

$$\begin{aligned} & -(k - \frac{2(n+1)+1}{2})^2 + ((n+1) + \frac{1}{2})^2 \\ \ge & -(1 - (n+1) - 1/2)^2 + (n+1)^2 + (n+1) + 1/4 \\ = & 2(n+1). \end{aligned}$$

This shows

$$\mathbb{P}(E) \le (1 - p_n)^{2T(n+1)}.$$

Case II: there are no zero elements in x. With the same notation Ω_j as in Case I, we have that $|\omega_0| = 0$ and

$$\begin{aligned} \mathbb{P}(E) &= (1 - p_n)^{\sum_{j=1}^k |\Omega_j|} \times \prod_{1 \le i < j \le k} (1 - p_n)^{T|\Omega_i||\Omega_j|} \\ &= (1 - p_n)^{T(\sum_{j=1}^k |\Omega_j| + \sum_{1 \le i < j \le k} |\Omega_i||\Omega_j|)}. \end{aligned}$$

It is sufficient to obtain the minimizer of $\sum_{j=1}^k |\Omega_j| + \sum_{1 \le i < j \le k} |\Omega_i||\Omega_j|$. under the restriction $\sum_i |\Omega_i| = n$ and $k \ge 1$. Let $y_i = |\Omega_i|$. Then,

$$\begin{aligned} \sum_{j=1}^k y_i + \sum_{1 \le i < j \le k} y_i y_j &= \sum_{j=1}^k y_i + \tfrac{1}{2} \sum_{i=1}^n \sum_{j=1, j \ne i}^k y_i y_j \\ &= \tfrac{1}{2} \sum_{i=1}^k \sum_{j=1}^k y_i y_j - \tfrac{1}{2} \sum_{i=1}^k y_i^2 + (n+1) \\ &= \tfrac{1}{2} (\sum_{i=1}^k y_i)^2 - \tfrac{1}{2} \sum_{i=1}^k y_i^2 + (n+1) \\ &= \tfrac{1}{2} (n+1)^2 + (n+1) - \tfrac{1}{2} (\sum_{i=1}^k y_i^2). \end{aligned}$$

Under the restriction $\sum_i y_i = n + 1 > 0$ and $z_i \ge 0$, the functions $\sum_{i=1}^k z_i^2$ obtain their maximizer at points $z = (0, \ldots, n+1, 0, \ldots, 0)$. Therefore, we have

$$\sum_{j=1}^k y_i + \sum_{1 \le i < j \le k} y_i y_j \ge \tfrac{1}{2} ((n+1)^2 + (n+1) - (n+1)^2) \ge n + 1.$$

This shows

$$\mathbb{P}(E) \le (1 - p_n)^{(n+1)T}.$$

By combining the lower bounds of $\mathbb{P}(E)$ under Cases I and II, we have

$$\mathbb{P}(E^c) \ge 1 - (1 - p_n)^{(n+1)T},$$

where E^c denotes that V is positively definite. □

Appendix B.2. Proof of Lemma A2

Proof of Lemma A2. Based on Lemma 1, V is a positively definite matrix with a probability of at least $1 - e^{-p_n T(n+1)}$. In what follows, we assume that V is positively definite such that its inverse exists. The proof proceeds two parts. The first part is to evaluate the cardinality of the set of the common neighbors of any two subjects i and j. That is, we establish the lower bound:

$$\min_{i,j} \#\{k : t_{ik} > 0, t_{kj} > 0\}.$$

The second part is to show such inequality [c.f. (A16)]

$$a \geq b[\sum_{\{k:t_{\alpha k}>0, t_{k\beta}>0\}} (z_{i\alpha} - z_{i\beta})].$$

We use the method of the proof in [20] with minor modifications that simplify their proofs to show the second part.

Part I. Let $1_{\{\cdot\}}$ be an indicator variable. It is equal to one when the expression in $\{\cdot\}$ is true; otherwise, it is equal to zero. For any given $i \neq j$, define

$$\xi_{ij} = \sum_{k=0, k\neq i,j}^{n} 1_{\{t_{ik}>0, t_{jk}>0\}}.$$

Note that ξ_{ij} is the sum of $n-1$ independent Bernoulli random variables, and for three distinct indices i, j, k,

$$\mathbb{E}\xi_{ij} = \mathbb{P}(t_{ik} > 0)\mathbb{P}(t_{jk} > 0) = (1 - (1 - p_n)^T)^2 := \eta_n.$$

Based on the Chernoff bound in [29], we have

$$\mathbb{P}\Big(\xi_{ij} \leq \tfrac{1}{2}(n-1)\eta_n\Big) \leq \exp\Big(-\tfrac{1}{8}(n-1)\eta_n\Big).$$

It follows that

$$\mathbb{P}\Big(\min_{i,j} \xi_{ij} \leq \tfrac{1}{2}(n-1)\eta_n\Big) \leq \sum_{i,j} \mathbb{P}\Big(\xi_{ij} \leq \tfrac{1}{2}(n-1)\eta_n\Big) \leq \tfrac{(n+1)n}{2} \exp\Big(-\tfrac{1}{8}(n-1)\eta_n\Big).$$

Since $T \geq 1$,

$$\eta_n = (1 - (1 - p_n)^T)^2 \geq (1 - (1 - p_n))^2 = p_n^2.$$

That is, with a probability of at least $1 - \frac{(n+1)n}{2} \exp\Big(-\frac{1}{8}(n-1)\eta_n\Big)$, we have

$$\min_{i,j} \xi_{ij} \geq \tfrac{1}{2}(n-1)\eta_n \geq \tfrac{1}{2}(n-1)p_n^2. \tag{A12}$$

Part II. For convenience, we introduce a non-negative array $\{q_{ij}\}_{i,j=1}^n$, where

$$q_{ij} := -v_{ij},\ i \neq j;\ \ q_{ii} := \sum_{k=1}^{n} v_{ik} = -v_{i0},\ \ i, j = 1, \ldots, n.$$

Let

$$m := \min_{(i,j)\in\{(i,j):q_{ij}>0\}} q_{ij}, \qquad M := \max_{i,j} q_{ij},$$
$$t_{\max} := \max_i t_i, \qquad t_{\min} := \min_i t_i.$$

It is clear that $M \geq m > 0$ and

$$q_{ij} \geq 0,\ \ q_{ij} = q_{ji},\ \ v_{ij} = -q_{ij},\ i \neq j,\ \ Mt_{\max} \geq v_{ii} = \sum_{k=1}^{n} q_{ik} \geq mt_{\min}.$$

Notice that

$$V^{-1} - S = (V^{-1} - S)(I - VS) + S(I - VS),$$

where I is a $n \times n$ identity matrix. Let $X = I - VS$, $Y = SX$ and $Z = V^{-1} - S$; we have the recursion formula

$$Z = ZX + Y.$$

The goal is to give an upper bound of all $|z_{ij}|$.

According the definitions of S, V and X, we have

$$\begin{aligned} x_{ij} &= \delta_{ij} - \sum_{k=1}^{n} v_{ik} s_{kj} \\ &= \delta_{ij} - \sum_{k=1}^{n} v_{ik} (\frac{\delta kj}{v_{jj}} + \frac{1}{v_{00}}) \\ &= \delta_{ij} - \frac{v_{ij}}{v_{ii}} - \frac{q_{ii}}{v_{00}} \\ &= (1-\delta_{ij}) \frac{q_{ij}}{v_{jj}} - \frac{q_{ii}}{v_{00}}. \end{aligned}$$

and

$$\begin{aligned} y_{ij} &= \sum_{k=1}^{n} s_{ik} x_{kj} = \sum_{k=1}^{n} (\frac{\delta_{ik}}{v_{ii}} + \frac{1}{v_{00}})((1-\delta_{kj}) \frac{q_{kj}}{v_{jj}} - \frac{q_{kk}}{v_{00}}) \\ &= \sum_{k=1}^{n} \frac{\delta_{ik}}{v_{ii}} \left\{ (1-\delta_{kj}) \frac{q_{kj}}{v_{jj}} - \frac{q_{kk}}{v_{00}} \right\} + \sum_{k=1}^{n} \frac{1}{v_{00}} \left\{ (1-\delta_{kj}) \frac{q_{kj}}{v_{jj}} - \frac{q_{kk}}{v_{00}} \right\} \\ &= \frac{(1-\delta_{ij}) q_{ij}}{v_{ii} v_{jj}} - \frac{q_{ii}}{v_{ii} v_{00}} - \frac{q_{jj}}{v_{jj} v_{00}}. \end{aligned}$$

Since

$$0 \le \frac{q_{ij}}{v_{ii} v_{jj}} \le \frac{M}{m^2 t_{\min}^2}, \ \ 0 \le \frac{q_{ij}}{v_{ii} v_{00}} \le \frac{M}{m^2 t_{\min}^2},$$

for any different i, j, k, we have

$$|y_{ij}| \le a := \frac{2M}{m^2 t_{\min}^2}, \ \ |y_{ij} - y_{ik}| \le a. \tag{A13}$$

In view of the expressions of x_{ij} and y_{ij}, we have

$$z_{ij} = \sum_{k=1}^{n} z_{ik} (1-\delta_{kj}) \frac{q_{kj}}{v_{jj}} - \sum_{k=1}^{n} z_{ik} \frac{q_{kk}}{v_{00}} + y_{ij}, \ \ i, j = 1, \ldots, n. \tag{A14}$$

Now, we fix an arbitrary i value and consider the upper bound of $\max_j |z_{ij}|$.

Let α and β be such that

$$z_{i\alpha} = \max_{k=1,\ldots,n} z_{ik}, \ \ z_{i\beta} = \min_{k=1,\ldots,n} z_{ik}.$$

Without loss of generality, we assume $z_{i\alpha} \ge |z_{i\beta}|$ (otherwise, we can invert the sign of z_{ik} and repeat the same process). Below, we will show $z_{i\beta} \le 0$. Note that this conclusion is not investigated in [20]. By multiplying v_{jj} by both sides of (A14), we have

$$v_{jj} z_{ij} = \sum_{k=1}^{n} z_{ik} (1-\delta_{kj}) q_{kj} - \sum_{k=1}^{n} z_{ik} \frac{q_{kk} v_{jj}}{v_{00}} + v_{jj} y_{ij}.$$

Summarizing the above equations with $j = 1, \ldots, n$, we have

$$\sum_{j=1}^{n} v_{jj} z_{ij} = \sum_{k=1}^{n} \sum_{j=1}^{n} z_{ik} (1-\delta_{kj}) q_{kj} - \sum_{k=1}^{n} \sum_{j=1}^{n} z_{ik} \frac{q_{kk} v_{jj}}{v_{00}} + \sum_{j=1}^{n} v_{jj} (\frac{(1-\delta_{ij}) q_{ij}}{v_{ii} v_{jj}} - \frac{q_{ii}}{v_{ii} v_{00}} - \frac{q_{jj}}{v_{jj} v_{00}}).$$

Thus,

$$\begin{aligned}
&\sum_{k=1}^{n}\sum_{j=1}^{n} z_{ik}\frac{q_{kk}v_{jj}}{v_{00}} + \sum_{k=1}^{n} z_{ik}q_{kk} \\
= &\sum_{j=1}^{n} v_{jj}\Big(\frac{(1-\delta_{ij})q_{ij}}{v_{ii}v_{jj}} - \frac{q_{ii}}{v_{ii}v_{00}} - \frac{q_{jj}}{v_{jj}v_{00}}\Big) \\
= &\sum_{j=1}^{n} \frac{(1-\delta_{ij})q_{ij}}{v_{ii}} - \frac{q_{ii}\sum_j v_{jj}}{v_{ii}v_{00}} - \sum_{j=1}^{n}\frac{q_{jj}}{v_{00}} \\
= &-\frac{q_{ii}}{v_{ii}} - \frac{q_{ii}\sum_j v_{jj}}{v_{ii}v_{00}} = -\frac{q_{ii}}{v_{ii}v_{00}}(v_{00} + \sum_{j=1}^{n} v_{jj})
\end{aligned}$$

Thus,

$$\begin{aligned}
-\frac{q_{ii}}{v_{ii}v_{00}}(v_{00} + \sum_{j=1}^{n} v_{jj}) &= \sum_{k=1}^{n} z_{ik}\frac{q_{kk}\sum_{j=1}^{n} v_{jj}}{v_{00}} + \sum_{k=1}^{n} z_{ik}q_{kk} \\
&\geq z_{i\beta}(v_{00} + \sum_{j=1}^{n} v_{jj}),
\end{aligned}$$

This shows

$$z_{i\beta} \leq -\frac{q_{ii}}{v_{ii}v_{00}} \leq 0. \tag{A15}$$

Since $\sum_{k=1}^{n} q_{k\alpha}/v_{\alpha\alpha} = 1$, we have

$$\sum_{k=1}^{n} z_{i\alpha}\frac{q_{k\alpha}}{v_{\alpha\alpha}} = \sum_{k=1}^{n} z_{ik}(1-\delta_{k\alpha})\frac{q_{k\alpha}}{v_{\alpha\alpha}} - \sum_{k=1}^{n} z_{ik}\frac{q_{kk}}{v_{00}} + y_{i\alpha}.$$

In other words,

$$\sum_{k=1}^{n}[z_{i\alpha} - z_{ik}(1-\delta_{k\alpha})]\frac{q_{k\alpha}}{v_{\alpha\alpha}} = -\sum_{k=1}^{n} z_{ik}\frac{q_{kk}}{v_{00}} + y_{i\alpha}.$$

Analogously, we have

$$\sum_{k=1}^{n}[z_{ik}(1-\delta_{k\beta}) - z_{i\beta}]\frac{q_{k\beta}}{v_{\beta\beta}} = \sum_{k=1}^{n} z_{ik}\frac{q_{kk}}{v_{00}} - y_{i\beta}.$$

Therefore,

$$\begin{aligned}
y_{i\alpha} - y_{i\beta} &= \sum_{k=1}^{n}\Big\{[z_{i\alpha} - z_{ik}(1-\delta_{k\alpha})]\frac{q_{k\alpha}}{v_{\alpha\alpha}} + [z_{ik}(1-\delta_{k\beta}) - z_{i\beta}]\frac{q_{k\beta}}{v_{\beta\beta}}\Big\} \\
&\geq \Big[\sum_{\{k:t_{\alpha k}>0, t_{k\beta}>0\}} (z_{i\alpha} - z_{i\beta})\Big] \times \frac{m}{Mt_{\max}}.
\end{aligned} \tag{A16}$$

The following calculations are based on the event E_n:

$$\{\min_{i\neq j}\xi_{ij} \geq \tfrac{1}{2}(n-1)p_n^2,\ \ t_{\min} \geq \tfrac{1}{2}nTp_n,\ \ t_{\max} \leq \tfrac{3}{2}nTp_n\}$$

In view of (A15) and (A16), we have

$$\begin{aligned} z_{i\alpha} &\leq z_{i\alpha} - z_{i\beta} \leq \frac{Mt_{\max}}{m} \times [\min_{i\neq j} \xi_{ij}]^{-1} \times \frac{2M}{m^2 t_{\min}^2} \\ &\leq \frac{M \cdot \frac{3}{2}nTp_n}{m} \times \frac{1}{\frac{1}{2}(n-1)p_n^2} \times \frac{2M}{m^2(\frac{1}{2}nTp_n)^2} \\ &= \frac{24M^2}{n(n-1)m^3Tp_n^3} \end{aligned} \tag{A17}$$

Note that $M = Tb_{n1}$ and $m = b_{n0}$. Based on Lemma 6 in the main text, we have

$$\mathbb{P}(t_{\min} \geq \tfrac{1}{2}nTp_n, t_{\max} \leq \tfrac{3}{2}nTp_n) \geq 1 - 2(n+1)\exp(-\tfrac{1}{8}nTp_n).$$

In view of inequality (A12), we have

$$P(E_n) \geq 1 - \tfrac{(n+1)n}{2}e^{-\frac{1}{8}(n-1)p_n} - 2(n+1)e^{-\frac{1}{8}nTp_n}.$$

If $p_n \geq 24\log n/n$, then

$$\tfrac{(n+1)n}{2}e^{-\frac{1}{8}(n-1)p_n} \leq O(\frac{1}{n}), \;\; (n+1)e^{-\frac{1}{10}np_n} = o(\frac{1}{n^{1.4}}),$$

such as

$$P(E_n) \geq 1 - O(\frac{1}{n}).$$

Let F_n be the event that V^{-1} exists. Based on Lemma 1, if $p_n \geq 24\log n/n$, then

$$\mathbb{P}(F_n) \geq 1 - \exp(p_nT(n+1)) \geq 1 - \frac{1}{n^{24}}.$$

Therefore,

$$\mathbb{P}(E_n \bigcap F_n) \geq 1 - O(\frac{1}{n}).$$

Substituting $M = Tb_{n1}$ and $m = b_{n0}$ into (A17) shows Lemma 2. □

Appendix B.3. Proof of Lemma A3

Proof of Lemma A3. Recall that $\pi_{ij} = \beta_i - \beta_j$ and

$$H_i(\beta) = \sum_{j\neq i} t_{ij}\mu(\pi_{ij}) - a_i, \;\; i = 1, \ldots, n.$$

The Jacobian matrix $H'(\beta)$ of $H(\beta)$ can be calculated as follows. By finding the partial derivative of H_i with respect to β for $i \neq j$, we have

$$\frac{\partial H_i(\beta)}{\partial \beta_j} = -t_{ij}\mu'(\pi_{ij}), \;\; \frac{\partial H_i(\beta)}{\partial \beta_i} = \sum_{j\neq i} t_{ij}\mu'(\pi_{ij}),$$

$$\frac{\partial^2 H_i(\beta)}{\partial \beta_i \partial \beta_j} = -t_{ij}\mu''(\pi_{ij}), \;\; \frac{\partial^2 H_i(\beta)}{\partial \beta_i^2} = \sum_{j\neq i} t_{ij}\mu''(\pi_{ij}).$$

When $\beta \in B(\beta^*, \epsilon_n)$, based on Condition (3c), we have

$$|\frac{\partial^2 H_i(\beta)}{\partial \beta_i \partial \beta_j}| \leq b_{n2}t_{ij}, \;\; i \neq j.$$

Let

$$\mathbf{g}_{ij}(\beta) = (\frac{\partial^2 H_i(\beta)}{\partial\beta_1\partial\beta_j}, \ldots, \frac{\partial^2 H_i(\beta)}{\partial\beta_n\partial\beta_j})^\top.$$

Therefore,

$$|\frac{\partial^2 H_i(\beta)}{\partial\beta_i^2}| \le t_i b_{n2}, \quad |\frac{\partial^2 H_i(\beta)}{\partial\beta_j\partial\beta_i}| \le b_{n2} t_{ij}. \tag{A18}$$

This demonstrates that $\|\mathbf{g}_{ii}(\beta)\|_1 \le 2t_i b_{n2}$. Note that when $i \ne j$ and $k \ne i, j$,

$$\frac{\partial^2 H_i(\beta)}{\partial\beta_k\partial\beta_j} = 0.$$

Therefore, we have $\|\mathbf{g}_{ij}(\beta)\|_1 \le 2t_{ij} b_{n2}$ for $j \ne i$. Consequently, for vectors $x, y \subset D$, we have

$$\begin{aligned}
& \max_{i=0,\ldots,n} \|H_i'(x) - H_i'(y)\|_1 \\
\le\ & \max_{i=0,\ldots,n} \sum_{j=1}^{n} |\frac{\partial H_i(x)}{\partial x_j} - \frac{\partial H_i(y)}{\partial y_j}| \\
=\ & \max_{i=0,\ldots,n} \sum_{j=1}^{n} |\int_0^1 [\mathbf{g}_{ij}(tx + (1-t)y)]^\top (x-y) dt| \\
\le\ & \max_{i=0,\ldots,n} 4 b_{n2} t_i \|x - y\|_\infty \\
=\ & 4 b_{n2} t_{\max} \|x - y\|_\infty.
\end{aligned}$$

This completes the proof. □

Appendix B.4. Proof of Lemma A4

Proof of Lemma A4. Recall that $t_i = \sum_{j\ne i} t_{ij}$, and a_i is the number of wins of subject i out of t_i comparisons. Since all comparisons are mutually independent, a_i is the sum of m_i independent Bernoulli random variables given $t_{ij} = m_{ij}$ for $j = 0, \ldots, n$, where $m_i = \sum_{j\ne i} m_{ij}$. Based on [30]'s (1963) inequality, we have

$$\begin{aligned}
& \mathbb{P}\Big(|a_i - \mathbb{E}(a_{ij}|t_{ij}, j = 0, \ldots, n)| \ge \sqrt{2m_i \log n} | t_{ij} = m_{ij}, j = 0, \ldots, n\Big) \\
\le\ & 2\exp\{-\frac{2m_i \log n}{m_i}\} = \frac{2}{n^2}.
\end{aligned}$$

where $\mathbb{E}(a_{ij}|t_{ij}, j = 0, \ldots, n)$ is the conditional expectation given t_{ij} for $0 \le j \le n$. Note that the upper bound of the above probability does not depend on m_{ij}. With the law of total probability, for fixed i,

$$\begin{aligned}
& \mathbb{P}\left(|a_i - \sum_j \mathbb{E}(a_{ij}|t_{ij}, j = 0, \ldots, n)| \ge \sqrt{2t_i \log n}|\right) \\
=\ & \sum_{m_{i0}=0}^{T} \cdots \sum_{m_{in}=0}^{T} \mathbb{P}(t_{ij} = m_{ij}, j = 0, \ldots, n) \\
& \times \mathbb{P}\left(|a_i - \sum_j \mathbb{E}(a_{ij}|t_{ij}, j = 0, \ldots, n)| \ge \sqrt{2m_i \log n} | t_{ij} = m_{ij}, j = 0, \ldots, n\right) \\
\le\ & 2n^{-2} \sum_{m_{i0}=0}^{T} \cdots \sum_{m_{in}=0}^{T} \mathbb{P}(t_{ij} = m_{ij}, j = 0, \ldots, n) \\
\le\ & 2\exp\{-\frac{2m_i \log n}{m_i}\} = \frac{2}{n^2}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \mathbb{P}\left(\max_{i=1,\ldots,n}|a_i-\sum_j \mathbb{E}(a_{ij}|t_{ij},j=0,\ldots,n)|\geq\sqrt{2t_{\max}\log n}\right)\\
\leq\ & \mathbb{P}\left(\bigcup_i\left\{|a_i-\mathbb{E}a_i|\geq\sqrt{2t_i\log n}\right\}\right)\\
\leq\ & \sum_{i=1}^n\mathbb{P}\left(|a_i-\mathbb{E}a_i|\geq\sqrt{2t_i\log n}\right)\\
\leq\ & n\times\frac{1}{n^2}=\frac{1}{n}.
\end{aligned}$$

This completes the proof. □

Appendix B.5. Proof of Lemma A5

Proof of Lemma A5. We first evaluate the uniform lower bound of t_i, $i=0,\ldots,n$. Note that t_i is the sum of nT independent and identically distributed (i.i.d.) binomial random variables, $Bin(T,p_n)$. It can be also regarded as the sum of Tn, i.i.d., Bernoulli random variables. With the use of the Chernoff bound ([29]), we have

$$\mathbb{P}(\min_{i=0,\ldots,n}t_i<\tfrac{T}{2}np_n)\leq\sum_{i=0}^n\mathbb{P}(t_i<\tfrac{T}{2}np_n)\leq(n+1)\exp(-\tfrac{T}{8}np_n).$$

Thus, with a probability of at least $1-(n+1)\exp(-\frac{T}{8}np_n)$,

$$\min_{i=0,\ldots,n}t_i\geq\tfrac{T}{2}np_n.$$

Analogously, with the use of the Chernoff bound ([29]), we have

$$\mathbb{P}(\max_{i=0,\ldots,n}t_i>\tfrac{3}{2}nTp_n)\leq\sum_{i=0}^n\mathbb{P}(t_i>\tfrac{3}{2}nTp_n)\leq(n+1)\exp(-\tfrac{1}{10}nTp_n),$$

and

$$\mathbb{P}(\tfrac{1}{2}\sum_i t_i>\tfrac{3}{2}(n+1)nTp_n)\leq\exp(-\tfrac{1}{10}(n+1)nTp_n).$$

□

Appendix B.6. Proof of Lemma A6

Proof of Lemma A6. Write

$$\begin{gathered}
H=H(\beta^*),\ \ V^{-1}=H'(\beta^*),\\
\mathbb{E}^*(\cdot)=\mathbb{E}(\cdot|t_{ij},0\leq i,j\leq n),\ \ \mathrm{Cov}^*(\cdot)=\mathrm{Cov}(\cdot|t_{ij},0\leq i,j\leq n).
\end{gathered}$$

Then,

$$H=\mathbb{E}^*(a)-a.$$

Let $W=V^{-1}-S$. Note that $U=\mathrm{Cov}^*(H)$. Via direct calculations, we have

$$\mathrm{Cov}^*(WH)=(V^{-1}-S)U(V^{-1}-S):=I_1+I_2$$

where $I_1=V^{-1}-S+SVS-S$ and $I_2=(V^{-1}-S)(U-V)(V^{-1}-S)$. It is easy to verify

$$(SVS-S)_{ij}=\frac{v_{i0}}{v_{ii}v_{00}}+\frac{v_{0j}}{v_{jj}v_{00}}-\frac{(1-\delta_{ij})v_{ij}}{v_{ii}v_{jj}}.$$

Therefore,

$$\max_{i,j} |(SVS - S)_{ij}| \leq \frac{3b_{n1}}{t_{\min}^2 b_{n0}^2}.$$

Based on Lemmas 3 and 5 in the main text, we have

$$I_1 = O_p(\frac{b_{n1}^2}{n^2 b_{n0}^3 p_n^5}).$$

Now, we evaluate I_2. Direct calculations give

$$\begin{aligned}
& [(V^{-1} - S)(U - V)(V^{-1} - S)]_{ij} \\
= & \sum_{k,s} (V^{-1} - S)_{ik} (U - V)_{ks} (V^{-1} - S)_{sj} \\
= & O((\frac{b_{n1}^2}{n^2 b_{n0}^3 p_n \eta_n})^2) \sum_{k,s} |(U - V)_{ks}| \\
= & O((\frac{b_{n1}^2}{n^2 b_{n0}^3 p_n \eta_n})^2) \times 2(b_{n1} + 1/4) \sum_i t_i \\
= & O(\frac{b_{n1}^5}{n^2 b_{n0}^6 p_n^5}).
\end{aligned}$$

where the second inequality is due to Lemma 3 and the third inequality is due to $p_{ij}(1 - p_{ij}) \leq 1/4$ and $\mu'_{ij}(\beta) \leq b_{n1}$. Therefore, we have

$$I_1 + I_2 = O_p(\frac{b_{n1}^5}{n^2 b_{n0}^6 p_n^5}).$$

□

References

1. Han, R.; Xu, Y.; Chen, K. A general pairwise comparison model for extremely sparse networks. *J. Am. Stat. Assoc.* **2023**, *118*, 2422–2432. [CrossRef]
2. Stigler S.M. Citation patterns in the journals of statistics and probability. *Stat. Sci.* **1994**, *9*, 94–108. [CrossRef]
3. Varin, C.; Cattelan, M.; Firth, D. Statistical modelling of citation exchange between statistics journals. *J. R. Stat. Soc. Ser. A-Stat. Soc.* **2016**, *179*, 1–63. [CrossRef]
4. Radlinski F., Joachims T. Active exploration for learning rankings from clickthrough data. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM: San Jose, CA, USA, 2007; pp. 570–579.
5. Chen B., Escalera S., Guyon I., Ponce-López V., Shah N.B., Simon M.O. Overcoming calibration problems in pattern labeling with pairwise ratings: Application to personality traits. In *European Conference on Computer Vision (ECCV 2016) Workshops*; Springer: Cham, Switzerland, 2016; Volume 9915, pp. 419–432.
6. David H.A. *The Method of Paired Comparisons*, 2nd ed.; Oxford University Press: Oxford, UK, 1988.
7. Bradley, R.A.; Terry, M.E. Rank analysis of incomplete block designs the method of paired comparisons. *Biometrika* **1952**, *39*, 324–345. [CrossRef]
8. Zermelo, E. Die berechnung der turnier-ergebnisse als ein maximumproblem der wahrscheinlichkeitsrechnung. *Math. Z.* **1929**, *29*, 436–460. [CrossRef]
9. Thurstone, L.L. A law of comparative judgment. *Psychol. Rev.* **1927**, *34*, 273–286. [CrossRef]
10. Simons, G.; Yao, Y.C. Asymptotics when the number of parameters tends to infinity in the bradley-terry model for paired comparisons. *Ann. Stat.* **1999**, *27*, 1041–1060. [CrossRef]
11. Chen, Y.; Suh, C. Spectral mle: Top-k rank aggregation from pairwise comparisons. In *Proceedings of the 32nd International Conference on Machine Learning (ICML 2015)*; Bach F., Blei D., Eds.; International Machine Learning Society (IMLS): Stroudsburg, PA, USA, 2015; pp. 371–380.
12. Shah, N.B.; Wainwright M.J. Simple, robust and optimal ranking from pairwise comparisons. *J. Mach. Learn. Res.* **2018**, *18*, 1–38.
13. Han, R.; Ye, R.; Tan, C.; Chen, K. Asymptotic theory of sparse bradley-terry model. *Ann. Appl. Probab.* **2020**, *30*, 2491–2515. [CrossRef]

14. Yan, T.; Yang, Y.; Xu, J. Sparse paired comparisons in the bradley-terry model. *Stat. Sin.* **2012**, *22*, 1035–1318. [CrossRef]
15. Agarwal, A.; Patil, P., Agarwal S. Accelerated spectral ranking. In *Proceedings of the Machine Learning Research, Stockholmsmässan, Stockholm, Sweden, 10–15 July 2018*; PMLR: Cambridge, MA, USA, 2018; Volume *80*, pp. 70–79.
16. Hendrickx, J.; Olshevsky, A.; Saligrama, V. Graph resistance and learning from pairwise comparisons. In *Proceedings of the Conference on Neural Information Processing Systems*; NIPS: San Diego, CA, USA, 1999; pp. 2702–2711.
17. Vojnovic, M.; Yun, S. Parameter estimation for thurstone choice models. *arXiv* **2017**, arXiv:1705.00136. [CrossRef]
18. Wang, J.; Shah, N.; Ravi, R. Stretching the effectiveness of mle from accuracy to bias for pairwise comparisons. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics (AISTATS 2020), Online, 26–28 August 2020*; Chiappa, S., Calandra, R., Eds.; Proceedings of Machine Learning Research; PMLR: Cambridge, MA, USA,2020; Volume 108, pp. 66–76.
19. Ford, L.R. Solution of a ranking problem from binary comparisons. *Am. Math. Mon.* **1957**, *64*, 28–33. [CrossRef]
20. Simons, G.; Yao, Y.C. Approximating the inverse of a symmetric positive definite matrix. *Linear Algebra Appl.* **1998**, *281*, 97–103. [CrossRef]
21. Yamamoto, T. Error bounds for newton s iterates derived from the kantorovich theorem. *Numer. Math.* **1986**, *48*, 91–98. [CrossRef]
22. Delyon, B. Exponential inequalities for sums of weakly dependent variables. *Electron. J. Probab.* **2009**, 752–779. [CrossRef]
23. Roussas, G.G. Exponential probability inequalities with some applications. *Lect. Notes-Monogr. Ser.* **1996**, *30*, 303–319.
24. Ioannides D.A., Roussas G.G. Exponential inequality for associated random variables. *Stat. Probab. Lett.* **1999**, *42*, 423–431. [CrossRef]
25. Cocke, W.J. Central limit theorems for sums of dependent vector variables. *Ann. Math. Statist.* **1972**, *43*, 968–976. [CrossRef]
26. Cox, J.T.; Grimmett, G. Central limit theorems for associated random variables and the percolation model. *Ann. Probab.* **1984**, *12*, 514–528. [CrossRef]
27. Kantorovich, L.V. Functional analysis and applied mathematics. *Uspekhi Mat Nauk* **1948**, *3*, 89–185.
28. Loève, M. *Probability Theory I*, 4th ed.; Springer: New York, NY, USA, 1977.
29. Chernoff, H. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **1952**, *23*, 493–507. [CrossRef]
30. Hoeffding, W. Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.* **1963**, *58*, 13–30. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.