Electronic Research Archive

https://www.aimspress.com/journal/era

ERA, 34(8): 5395–5420.
DOI: 10.3934/era.2026241
Received: 20 April 2026
Revised: 10 June 2026
Accepted: 22 June 2026
Published: 29 June 2026

Research article

Average calibration of high-dimensional jackknife empirical likelihood

Mengdong Shang, Xia Chen, Weimin Yang and Li Yan*

School of Mathematics and Statistics, Shaanxi Normal University, Xi'an 710119, China

* **Correspondence:** Email: lyan@snnu.edu.cn.

Abstract: This paper investigates calibration of the jackknife empirical likelihood for U-statistics and U-shaped estimating equations. By adjusting the mean and variance of the asymptotic distribution, we propose two calibration methods that control the Type I error rate in high-dimensional settings. To overcome their limitations, we develop an average calibration method for the high-dimensional jackknife empirical likelihood. Numerical results show that this average calibration significantly outperforms standard approaches.

Keywords: jackknife empirical likelihood; high-dimensional data; U-statistics; U-shaped estimating equation

1. Introduction

Empirical likelihood (EL) is a data-driven nonparametric inference method introduced by Owen [1], who defined the EL ratio function based on the nonparametric likelihood of the data and demonstrated its use in constructing confidence intervals for sample means, M-estimators, and differentiable statistical functionals, with a comprehensive treatment given in his monograph [2]. This result provides a nonparametric extension of Wilks's theorem on likelihood ratios. Subsequently, Owen [3] extended EL to vector-valued statistical functionals and constructed confidence regions for parameters. Like other nonparametric methods, such as the bootstrap and the jackknife, EL does not require prior knowledge of the data distribution. It can seamlessly incorporate constraints to reflect prior information while automatically determining the shape of confidence regions.

A major advance came with Qin and Lawless [4], who assumed only a parametric form for certain population characteristics and incorporated general estimating equations into the EL framework, providing a unified paradigm for parameter estimation and inference. Building on this foundation, EL has been successfully applied to a wide range of statistical models. Shi and Lau [5] combined EL with partially linear models and derived the asymptotic distribution of the EL ratio statistic. Xue and Zhu [6] studied EL for varying coefficient models with longitudinal data and established the

asymptotic theory of the corresponding EL ratio statistic. Chen and Cui [7] examined Bartlett and Bartlett-type corrections of EL in the presence of nuisance parameters, providing refined inference procedures for semiparametric settings. Beyond these core developments, EL has been extended in several important directions. In the dependent-data setting, Kitamura [8] extended EL to weakly dependent processes, laying a rigorous theoretical foundation for its application in time series analysis. Wang et al. [9] developed Bayesian empirical likelihood inference and order shrinkage for hysteretic autoregressive models, extending these methods to nonlinear time series with regime-dependent dynamics. EL has also achieved notable success across disciplines: Applications range from censored linear regression in survival analysis [10] to econometric moment-condition models, demonstrating its versatility as a nonparametric inference tool. Regarding higher-order properties, Newey and Smith [11] proposed generalized EL estimators and showed that they possess favorable higher-order efficiency compared with the generalized method of moments (GMM). Data simulations demonstrate that, compared to traditional methods, EL yields smaller confidence intervals or regions with better coverage accuracy. Concerning computational improvements, DiCiccio et al. [12] proved that EL admits Bartlett correction, which reduces the coverage error from $O(n^{-1})$ to $O(n^{-2})$. Chen and Cui [13] further demonstrated that EL constructed from estimating equations is also Bartlett correctable. In high-dimensional settings, Tang and Leng [14] combined penalty functions with EL to propose the penalized empirical likelihood (PEL) method.

EL has extensive applications in high-dimensional settings. Hjort et al. [15] applied EL to high-dimensional estimating equations and proved the asymptotic normality of the EL ratio under the condition that the parameter dimension and sample size satisfy $p = o(n^{1/3})$. Chen et al. [16] assessed the impact of data dimension on the asymptotic normality of the EL ratio for high-dimensional means in a general multivariate model, raising the allowable growth rate to $p = o(n^{1/2})$. Their study emphasized that $o(n^{1/2})$ is the maximum rate for which asymptotic normality holds.

EL faces computational challenges with nonlinear statistics such as U-statistics [17]. A key solution is linearization. The jackknife empirical likelihood (JEL) addresses this by representing U-statistics as averages of jackknife pseudo-values, making EL computationally feasible [18]. Subsequent work has expanded JEL's applicability: Li et al. [19] used it for confidence intervals with nuisance parameters; Peng [20] proposed an approximate method for implicit nuisance parameters; and Li et al. [21] applied it to nonsmooth U-shaped estimation functions. Wei et al. [22] further studied its asymptotic properties in high dimensions. The combination of U-statistics, U-type estimating equations, and JEL has significantly broadened the applicability of these methods. Wang et al. [23] proposed a JEL-based test for the equality of two high-dimensional means; Liu et al. [24] developed a JEL test for the equality of two distributions using characteristic functions; Peng and Fei [25] constructed JEL goodness-of-fit tests for U-statistics-based general estimating equations; and Cheng et al. [26] introduced a balanced augmented JEL method for two-sample U-statistics.

When combined with high-dimensional U-statistics and U-type estimating equations, JEL offers clear advantages. However, in high-dimensional settings, the coverage probability of JEL-based confidence regions falls significantly below the nominal level—an issue also observed in standard EL. Liu et al. [27] attributed this to a severe lack of fit when approximating the EL ratio with the asymptotic normal distribution, leading to inflated Type I error rates. The core problem lies in underestimation of the centered and standardized quantities in the likelihood ratio. Correcting these expectations and variances has been shown to improve high-dimensional EL performance. Guo et

al. [28] applied such corrections to high-dimensional regression models and proved asymptotic validity under random and fixed designs. He and Chen [29] extended this calibration to semiparametric varying-coefficient partially linear models with a diverging number of parameters.

This paper studies the JEL method for U-statistics and U-shaped estimating equations in high-dimensional settings. The calibrated JEL approach improves performance by adjusting the mean and variance of the asymptotic normal distribution to control the Type I error rate, thereby addressing the challenges posed by high-dimensional data and U-statistics. Under certain conditions, we establish the asymptotic normality of the high-dimensional JEL with U-shaped estimating equations. However, our simulations show that using normal quantiles for confidence regions yields low coverage rates when p/n is large, due to discrepancies between the JEL statistic and its asymptotic approximation. While two calibration methods based on mean and variance adjustments are proposed, simulations indicate that they perform poorly when used alone. By integrating these two methods, we propose an average calibration approach that effectively overcomes these shortcomings and achieves superior numerical performance.

It is instructive to note that likelihood-based and maximum-likelihood-estimation-based calibration ideas have found success in other high-dimensional kernel modeling contexts. Cavoretto and De Rossi [30] proposed an adaptive residual subsampling algorithm for kernel interpolation in which maximum likelihood estimation is used to tune kernel hyperparameters, thereby calibrating the model to the data—a strategy analogous to our moment-based calibration of the JEL statistic. Cavoretto et al. [31] studied partition-of-unity Kriging estimators whose locality parameters are selected via likelihood criteria. Although their application domains differ from ours, both works illustrate a common principle: Replacing fixed asymptotic benchmarks with data-driven, likelihood-guided corrections improves finite-sample accuracy in high-dimensional settings, which is precisely the motivation underlying the average calibration of high-dimensional jackknife empirical likelihood (ACJEL).

The rest of this paper is organized as follows. Section 2 introduces the high-dimensional JEL. In Sections 3 and 4, we present the calibration and average calibration of the high-dimensional JEL, respectively. Sections 5 and 6 evaluate the performance of the proposed method on simulated and real data. Section 7 concludes the paper with a discussion of future work.

2. High-dimensional jackknife empirical likelihood

Let $X_1, \cdots, X_n \in \mathbb{R}^d$ be independent and identically distributed (i.i.d.) random vectors. Suppose that the parameter $\theta = (\theta_1, \ldots, \theta_p)^{\mathrm{T}} \in \Theta$ can be estimated by solving the U-shaped estimating equations $U_n(\theta) := U(X_1, \ldots, X_n; \theta) = 0$. The l-th component of $U_n(\theta)$ is

$$U_{nl}(\theta) = \binom{n}{m}^{-1} \sum_{1 \le i_1 < \cdots < i_m \le n} h_l(X_{i_1}, \ldots, X_{i_m}; \theta), \quad l = 1, 2, \ldots, p,$$

where m is a fixed constant and h is a symmetric kernel function. To apply EL, we define the jackknife pseudo-values as

$$\hat{V}_i(\theta) = nU_n(\theta) - (n-1)U_{n-1}^{(-i)}(\theta), \quad i = 1, \ldots, n,$$

where $U_{n-1}^{(-i)}(\theta) := U\left(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n; \theta\right)$ is the estimating function obtained by omitting the i-th observation. Based on these jackknife pseudo-values, the JEL function is defined as

$$L\left(\theta\right) = \left\{ \prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \hat{V}_i\left(\theta\right) = 0 \right\},$$

and the corresponding JEL ratio function is formulated as

$$R(\theta) = \left\{ \prod_{i=1}^{n} (np_i) : \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \hat{V}_i(\theta) = 0 \right\}. \tag{2.1}$$

The maximum is achieved via the method of Lagrange multipliers, yielding

$$p_i = \frac{1}{n} \cdot \frac{1}{1 + \lambda^{\mathrm{T}} \hat{V}_i\left(\theta\right)}, \quad i = 1, \ldots, n, \tag{2.2}$$

where $\lambda = \lambda(\theta)$ is the p-dimensional Lagrange multiplier satisfying

$$0 = g\left(\lambda\right) := \frac{1}{n} \sum_{i=1}^{n} \frac{\hat{V}_i\left(\theta\right)}{1 + \lambda^{\mathrm{T}} \hat{V}_i\left(\theta\right)}.$$

Substituting Eq (2.2) into Eq (2.1) yields

$$-2 \log R\left(\theta\right) = 2 \sum_{i=1}^{n} \log\left\{1 + \lambda^{\mathrm{T}} \hat{V}_i\left(\theta\right)\right\}.$$

For convenience, let $\ell(\theta) = -\log R(\theta)$. When the parameter dimension p is fixed, $2\ell(\theta_0) \xrightarrow{d} \chi_p^2$ as $n \to \infty$, where θ_0 is the true value of θ. Therefore, confidence regions for the parameter can be constructed using the quantiles of the chi-square distribution with p degrees of freedom.

In high-dimensional settings, the parameter dimension p grows with the sample size n, so the data form a triangular array rather than having a fixed dimension. Consequently, all key quantities depend on n : $p = p_n, d = d_n, X_i = X_{n,i}, h(x_1, \ldots, x_m; \theta_0) = h_n(x_1, \ldots, x_m; \theta_0)$, and $\theta = \theta_n$. In this scenario, the asymptotic properties of $2\ell(\theta_0)$ must be re-evaluated. As $p \to \infty$, the χ_p^2 distribution approaches a normal distribution with mean p and variance $2p$. This naturally raises the question of whether

$$(2p)^{-1/2}\left\{2\ell\left(\theta_0\right) - p\right\} \xrightarrow{d} N\left(0, 1\right) \tag{2.3}$$

holds as $p \to \infty$.

To formalize this result, we introduce the following notation. Let $a_n = (p/n)^{1/2}$, and let $\gamma\{A\}$ denote the eigenvalues of a matrix A. In the proofs, we decompose the jackknife pseudo-value $\hat{V}_i(\theta_0)$ into a leading term $mg(X_i, \theta_0)$, which depends only on the individual observation X_i, and a remainder term $R_{ni}(\theta_0)$. Let Σ_n be the covariance matrix of $mg(X_i, \theta_0)$. We assume the following regularity conditions:

(A1) The parameter space $\Theta \in \mathbb{R}^p$ is compact, and $\mathrm{E}h(X_1, \ldots, X_m; \theta) = 0$ has a unique solution $\theta_0 \in \Theta$.

(A2) There exist constants $\alpha \geq 4$ and $C > 0$ such that $\max_{1\leq k\leq p} \mathrm{E}\,|h_k(X_1, \ldots, X_m; \theta_0)|^\alpha \leq C$.

(A3) There exist constants $b > 0$ and $B < \infty$ such that the eigenvalues of Σ_n satisfy $0 < b \leq \gamma_1\{\Sigma_n\} \leq \cdots \leq \gamma_p\{\Sigma_n\} \leq B < \infty$.

(A4) As $n \to \infty$, $p^t/n \to 0$, where $t = \max\{3\alpha/(\alpha-2), 4\}$.

Condition (A2) controls the tail behavior of the kernel function h. The proof of Theorem 1 requires the moment bound $\mathrm{E}\|(\hat{V}_i - \theta_0)/p^{1/2}\|^\alpha < M$ for some constant M, which is automatically implied by Condition (A2). Condition (A3) assumes that the covariance matrix of the leading term of $\sqrt{n}U_n(\theta_0)$ is positive definite with uniformly bounded eigenvalues. Condition (A4) specifies the permissible growth rate of p relative to n. The data dimension d is omitted here because it is generally a function of p. We emphasize that the kernel degree m is a fixed constant, independent of p and n. The exponent $t = \max\{3\alpha/(\alpha-2), 4\}$ is determined by two separate requirements in the proof of Theorem 1. The first requirement, $t \geq 3\alpha/(\alpha-2)$, arises from controlling the third-order Taylor remainder $r_n(\lambda)$ of $G_n(\lambda)$. The size of this remainder is governed by $\max_i \|\hat{V}_i(\theta_0)\|$, which is of order $O_p(n^{1/\alpha}p^{1/2})$. A smaller α corresponds to a heavier-tailed kernel function h, causing $\max_i \|\hat{V}_i(\theta_0)\|$ to grow faster with n. To keep the remainder negligible relative to the fluctuation scale $p^{1/2}$ of the normal approximation, p must grow more slowly relative to n when α is small. Concretely, this constraint is $p^{3\alpha/(\alpha-2)}/n \to 0$. The second requirement, $t \geq 4$, arises independently from controlling the approximation error between T_{n1} and $K_n \equiv nU_n^T(\theta_0)\Sigma_n^{-1}U_n(\theta_0)$, and does not depend on α. Taking the maximum of these two constraints yields Condition (A4). In particular, since $3\alpha/(\alpha-2)$ is strictly decreasing in α and equals 4 at $\alpha = 8$, when $\alpha \geq 8$ the second constraint dominates and Condition (A4) reduces to $p^4/n \to 0$, i.e., $p = o(n^{1/4})$. When $4 \leq \alpha < 8$, the first constraint is binding, e.g., $\alpha = 4$ gives $p^6/n \to 0$, i.e., $p = o(n^{1/6})$. The asymptotic distribution of $2\ell(\theta_0)$ in this high-dimensional setting is established below.

Theorem 1. *Suppose that conditions* (A1)-(A4) *hold. Then, as* $n \to \infty$,

$$(2p)^{-1/2}\{2\ell(\theta_0) - p\} \xrightarrow{d} N(0, 1).$$

Throughout this paper, d strictly denotes the ambient dimension of the observed data vectors, while p denotes the total number of parameters under inference. Any structural constraint is consistently formulated in terms of the parameter dimension p. Theorem 1 guarantees that confidence regions for the parameter can be constructed using standard normal quantiles. In practice, however, when p is large, the empirical coverage of these regions falls well below the nominal level. To address this, the following section introduces a calibration method for the high-dimensional JEL.

3. Calibration of high-dimensional jackknife empirical likelihood

In high-dimensional EL, using critical values from the normal approximation to construct confidence regions leads to insufficient coverage, particularly when the ratio p/n is large. Calibration can mitigate this issue. As shown in Section 5, the same problem also appears in high-dimensional JEL. Therefore, this section investigates calibration for the JEL.

The insufficient coverage stems from a substantial discrepancy between the true mean and variance (E_n, V_n) of $2\ell(\theta_0)$ and their asymptotic normal counterparts $(p, 2p)$. Section 4 demonstrates that for constructing confidence regions of the covariance matrix, jackknife empirical likelihood suffers from this discrepancy even more severely than standard empirical likelihood. In short, approximating

(E_n, V_n) by $(p, 2p)$ is inadequate (see Table 1). Consequently, the calibration methods developed in this section begin by replacing $(p, 2p)$ in Eq (2.3).

The proof of Theorem 1 proceeds via a two-step approximation. The first step shows that

$$2\ell(\theta_0) - nU_n^{\mathrm{T}}(\theta_0)S_n^{-1}(\theta_0)U_n(\theta_0) = o_p(\sqrt{p}), \tag{3.1}$$

where

$$S_n(\theta_0) = \frac{1}{n}\sum_{i=1}^{n}\hat{V}_i(\theta_0)\hat{V}_i^{\mathrm{T}}(\theta_0).$$

The second step establishes that

$$nU_n^{\mathrm{T}}(\theta_0)\left(\Sigma_n^{-1} - S_n^{-1}(\theta_0)\right)U_n(\theta_0) = o_p\left(\sqrt{p}\right).$$

The theoretical basis for approximating $2\ell(\theta_0)$ with a normal distribution is that $2\ell(\theta_0)$ approaches $K_n \equiv nU_n^{\mathrm{T}}(\theta_0)\Sigma_n^{-1}U_n(\theta_0)$, where $E(K_n) \approx p$ and $\mathrm{Var}(K_n) \approx 2p$. However, in practice, we cannot directly use the true expectation and variance of K_n for calibration. Since the underlying data distribution is unknown, the exact distribution of K_n is inaccessible. Even if the distribution of X were known, computing the mean and variance of K_n would be extremely difficult, especially when the kernel function h is complex. Based on Eq (3.1), we propose approximating $2\ell(\theta_0)$ with

$$T_{n1} \equiv nU_n^{\mathrm{T}}(\theta_0)S_n^{-1}(\theta_0)U_n(\theta_0).$$

The appendix proof shows that, under certain conditions, T_{n1} is the dominant component of $2\ell(\theta_0)$. Thus, we expect its expectation and variance (E_{n1}, V_{n1}) of T_{n1} are closer to those of $2\ell(\theta_0)$ than $(p, 2p)$ are, which should improve the normal approximation. However, simulation results for the covariance matrix show that T_{n1} is an inadequate approximation because it systematically underestimates $2\ell(\theta_0)$.

Figure 1 displays simulation results for the sample covariance matrix using the JEL, where red circles represent the scatter plot of simulated $(2\ell(\theta_0), T_{n1})$ values across $N = 1000$ replications. It is clear that whether $n = 200$, $p = 45$ or $n = 400$, $p = 66$, almost all simulatedT_{n1} values are strictly less than $2\ell(\theta_0)$. Moreover, the minimal dispersion and high concentration of T_{n1} (as shown by the y-axis scale) indicate that both E_{n1} and V_{n1} are excessively small.

An intuitive explanation is as follows. Let $f(\lambda) = 2\sum_{i=1}^{n}\log(1 + \lambda^{\mathrm{T}}\hat{V}_i(\theta_0))$, so that $2\ell(\theta_0) = \sup_\lambda f(\lambda) = f(\lambda_*)$, where λ_* is the maximizer. When $\lambda^{\mathrm{T}}\hat{V}_i(\theta_0)$ is small, a second-order Taylor expansion gives

$$f(\lambda) \approx f_1(\lambda) \equiv 2\sum_{i=1}^{n}\left\{\lambda^{\mathrm{T}}\hat{V}_i(\theta_0) - \frac{1}{2}\left(\lambda^{\mathrm{T}}\hat{V}_i(\theta_0)\right)^2\right\}.$$

Thus, $\sup_\lambda f_1(\lambda) = f_1(S_n^{-1}(\theta_0)U_n(\theta_0)) = T_{n1}$ approximates $2\ell(\theta_0)$. However, for large n and p, this approximation may fail because it requires $\lambda^{\mathrm{T}}\hat{V}_i(\theta_0) \in (-1, 1)$. When p/n is large, $\lambda_*^{\mathrm{T}}\hat{V}_i(\theta_0)$ often exceeds 1. For $x \in (-1, 1), \log(1 + x) \approx x - x^2/2$; but for $x > 1, \log(1 + x) > \log 2 > x - x^2/2$. Consequently, $f(\lambda) \geq f_1(\lambda)$. This reveals a fundamental limitation of Taylor approximations for $2\ell(\theta_0)$: When $\lambda_*^{\mathrm{T}}\hat{V}_i(\theta_0)$ is not uniformly small, higher-order terms only worsen the deviation.

A second calibration method addresses the approximation error of $f_1(\lambda)$. By adding a nonnegative penalty term, we define

$$f_2(\lambda) = f_1(\lambda) + n(\lambda^{\mathrm{T}}U_n(\theta_0))^2.$$

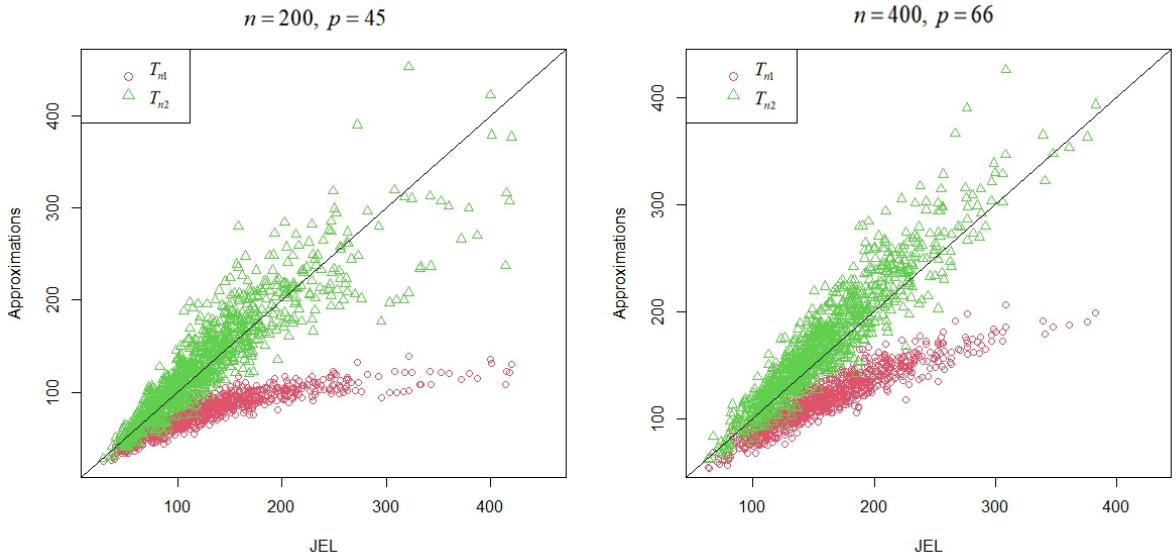

Figure 1. Scatter plots of simulated values $(2\ell(\theta_0), T_{n1})$ and $(2\ell(\theta_0), T_{n2})$. Parameter θ_0 is covariance matrix $\Sigma = (\sigma_{ij})$, where $\sigma_{ij} = 0.5^{|i-j|}$, U-statistic is sample covariance matrix, and data X_i is generated from a multivariate t-distribution.

The penalty term $n(\lambda^T U_n(\theta_0))^2$ plays two roles. First, because $f(\lambda) \geq f_1(\lambda)$ holds pointwise, T_{n1} systematically underestimates $2\ell(\theta_0)$. Adding this nonnegative term raises $f_2(\lambda) \geq f_1(\lambda)$, compensating for that underestimation. Second, the specific quadratic form preserves analytical tractability. $f_2(\lambda)$ is still quadratic in λ, so its supremum

$$T_{n2} = \sup_{\lambda} f_2(\lambda) = nU_n^T(\theta_0)\, S_{n2}^{-1}(\theta_0)\, U_n(\theta_0)$$

is attained in closed form, where

$$S_{n2}(\theta_0) = \frac{1}{n}\sum_{i=1}^{n}(\hat{V}_i(\theta_0) - U_n(\theta_0))(\hat{V}_i(\theta_0) - U_n(\theta_0))^{\mathrm{T}}$$

centers each pseudo-value by $U_n(\theta_0)$ rather than zero, thereby accommodating the nonnegligible bias of $U_n(\theta_0)$ when p/n is not small.

This second calibration method can also be used to construct confidence regions, by substituting $(p, 2p)$ with the mean and variance (E_{n2}, V_{n2}) of T_{n2}.

Theorem 2. *Suppose that conditions* (A1)–(A4) *hold, then*

$$2\ell(\theta_0) - T_{nk} = o_p(\sqrt{p}), \quad k = 1, 2.$$

This theorem shows that using T_{n2} to approximate $2\ell(\theta_0)$ is asymptotically equivalent to using T_{n1} or K_n. However, their finite-sample behaviors differ markedly. When p/n is not negligible, a substantial gap emerges between the two calibration methods. Although the T_{n2} calibration improves coverage, it often over-corrects. In Figure 1, the green triangles plot $(2\ell(\theta_0), T_{n2})$ across N replications. Visually,

T_{n2} aligns more closely with $2\ell(\theta_0)$ than T_{n1} does, but it frequently overestimates $2\ell(\theta_0)$ and exhibits much wider dispersion. This indicates that both the mean and variance (E_{n2}, V_{n2}) of T_{n2} are significantly larger than (E_{n1}, V_{n1}). When the variance estimate is excessively large, the normal approximation in Eq (2.3) becomes overly conservative, ultimately producing abnormally high coverage rates.

4. Average calibration of high-dimensional jackknife empirical likelihood

Neither T_{n1} nor T_{n2} calibration is optimal for the JEL. As shown in the previous section, using T_{n1} calibration leads to under-coverage. This is because $f(\lambda) \geq f_1(\lambda)$ implies that $2\ell(\theta_0)$ exceeds T_{n1} with high probability. In contrast, T_{n2} calibration results in over-coverage, since the added term $n(\lambda^{\mathrm{T}} U_n(\theta_0))^2$ is positive, making $f(\lambda) \leq f_2(\lambda)$ with high probability. In simulations with a large p/n ratio, this term becomes large, inflating both the mean and variance of T_{n2}.

To address both issues, we calibrate using the weighted average of T_{n1} and T_{n2}. Specifically, we define

$$T_{n3} = aT_{n1} + (1-a)T_{n2}, \quad a \in [0,1]$$

and replace p and $2p$ with the mean and variance of T_{n3}. The weight a is restricted to $[0,1]$ so that T_{n3} is a convex combination of T_{n1} and T_{n2}. Since T_{n1} systematically underestimates $2\ell(\theta_0)$ whereas T_{n2} systematically overestimates it, this construction effectively compensates for the under-coverage of T_{n1} while correcting the over-coverage of T_{n2}. Based on the proofs of Theorems 1 and 2, the following result follows directly.

Theorem 3. *Suppose that conditions (A1)–(A4) hold; we have*

$$2\ell(\theta_0) - T_{n3} = o_p(\sqrt{p}).$$

Given the above discussion, the mean E_{n3} and variance V_{n3} of T_{n3} provide good approximations to E_n and V_n. Let $(\hat{E}_{ni}, \hat{V}_{ni})$ for $i = 1, 2, 3$ denote the estimates of (E_{ni}, V_{ni}). Parameter confidence regions can be constructed based on

$$(2\ell(\theta_0) - A_n)/\sqrt{B_n} \xrightarrow{d} N(0,1),$$

where (A_n, B_n) can be taken as either $(p, 2p)$ or $(\hat{E}_{ni}, \hat{V}_{ni})$ for $i = 1, 2, 3$. These three methods are compared in simulations, and the results show that the method based on $(\hat{E}_{n3}, \hat{V}_{n3})$ performs best.

A key issue is the rational choice of the weight a. In the simulation experiments of Section 5, we automatically select a using the Kolmogorov-Smirnov (KS) normality test on simulated data. Specifically, we generate N replicates of the data and compute $T_{n1}^{(i)}$ and $T_{n2}^{(i)}$ for each replicate. For a given a, we then obtain $T_{n3}^{(i)} = aT_{n1}^{(i)} + (1-a)T_{n2}^{(i)}, i = 1, \ldots, N$. Using these values, we estimate the moments E_{n3} and V_{n3}, and substitute them for $(p, 2p)$ to construct an approximately normal sequence:

$$(\hat{V}_{n3}^{(i)})^{-1/2} \left\{ 2\ell(\theta_0) - \hat{E}_{n3}^{(i)} \right\}.$$

The KS test is applied to this sequence, and the value of a that yields the largest p-value (i.e., the least evidence against normality) is selected as the weight. Figure 2 illustrates this automatic selection for different n and p in an example covariance matrix. For each (n, p), a is chosen by the KS test, and the error is averaged over 20 repetitions. The figure shows that for fixed n, a decreases as p increases,

indicating that $2\ell(\theta_0)$ becomes increasingly similar to T_{n2}. Conversely, for fixed p, a increases with n, suggesting that $2\ell(\theta_0)$ approaches T_{n1}. Overall, when p/n is small, T_{n1} alone approximates the distribution well; as p/n grows, T_{n1} becomes insufficient, and the larger values of T_{n2} compensate for this deficiency. This demonstrates the advantage of the correction provided by T_{n3}.

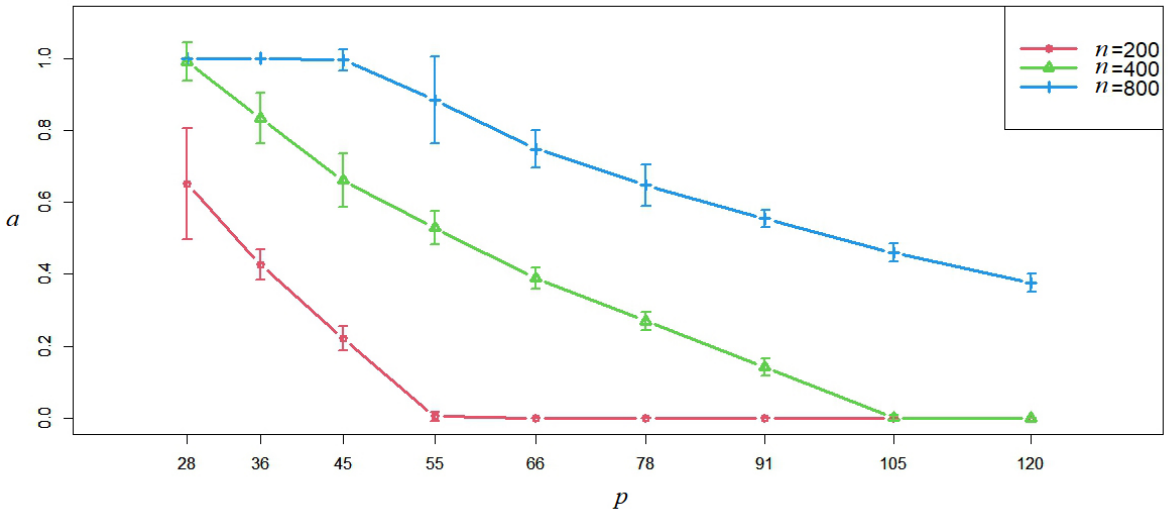

Figure 2. Under different sample sizes n and parameter dimensions p, the changes in weight a over 20 repetitions are obtained via the KS test, where the parameter θ_0 is the covariance matrix $\Sigma = (\sigma_{ij})$, $\sigma_{ij} = 0.5^{|i-j|}$, and the U-statistic is the sample covariance matrix.

In practice, when only a single dataset is available, one may consider using bootstrap or jackknife methods to estimate a, or alternative estimation procedures for E_{n3} and V_{n3}. For a sufficiently large sample size, combining data splitting with the KS test also provides an automatic way to determine $\hat{a}$. Although a formal convergence result for the KS-selected weight $\hat{a}$ remains an open challenge, the following heuristic explains the monotone pattern observed in Figure 2.

When $p/n \to 0$, Theorem 1 guarantees that $2\ell(\theta_0) - T_{n1} = o_p(\sqrt{p})$, so T_{n1} already approximates $2\ell(\theta_0)$ well and the KS test naturally selects $\hat{a}$ close to 1. As p/n grows, the inequality $f(\lambda) \geq f_1(\lambda)$ causes T_{n1} to systematically underestimate $2\ell(\theta_0)$, so the optimally standardized sequence $\{(\hat{V}_{n3}^{(i)})^{-1/2}(2\ell(\theta_0) - \hat{E}_{n3}^{(i)})\}$ becomes increasingly nonnormal under $a = 1$. Adding weight to T_{n2} (decreasing a) corrects this bias, and the KS test selects a smaller $\hat{a}$ accordingly. In the limiting case $p/n \to c > 0$, one would expect $\hat{a} \to 0$ so that ACJEL relies entirely on T_{n2}. This heuristic is consistent with Figure 2 and provides partial motivation for the KS-based selection, even in the absence of a formal theoretical guarantee.

We summarize the complete ACJEL procedure, including the bootstrap-based weight selection strategy for a single observed dataset, in Algorithm 1. In the simulation experiments of Section 5, where multiple fresh datasets are available, the KS-test weight selection in Step 3 is applied to independently generated replicates rather than bootstrap samples; the two approaches yield essentially identical results for $n \geq 200$.

To avoid double-use of a single dataset of size N, a three-step procedure is employed: First, the

sample is partitioned into a tuning set of size $n_1 = \lfloor N/3 \rfloor$ and an inference set of size $n_2 = N - n_1$; second, the weight $\hat{a}$ is selected via a Kolmogorov-Smirnov (KS) test based on bootstrap subsamples drawn exclusively from the tuning set; and third, the ACJEL confidence region is constructed using this fixed $\hat{a}$ by drawing bootstrap subsamples from the disjoint inference set. In small-sample scenarios where splitting might reduce power, a full-sample bootstrap can be utilized; experiments conducted across 10 random seeds confirm that $\hat{a}$ is highly concentrated ($\mathrm{Var}(\hat{a}) < 0.01$), indicating that the resulting mild data overlap does not compromise the reliability of the inference.

Algorithm 1 Average calibration of high-dimensional JEL (ACJEL)

Input: Data $\{X_1, \ldots, X_n\}$; null value θ_0; nominal level $1 - \alpha_0$; weight grid $\mathcal{A} \subset [0, 1]$; bootstrap replicates N (e.g. $N = 5000$).

Output: Reject or retain $H_0 : \theta = \theta_0$.

Step 1: JEL statistic on the observed data

Compute pseudo-values $\hat{V}_i(\theta_0) = nU_n(\theta_0) - (n-1)U_{n-1}^{(-i)}(\theta_0)$, $i = 1, \ldots, n$.

Solve for λ and set $2\ell(\theta_0) = 2\sum_{i=1}^{n} \log\big(1 + \lambda^\top \hat{V}_i(\theta_0)\big)$.

Step 2: Surrogate statistics on the observed data

$S_n \leftarrow \frac{1}{n}\sum_{i=1}^{n} \hat{V}_i \hat{V}_i^\top$; $\quad S_{n2} \leftarrow \frac{1}{n}\sum_{i=1}^{n} (\hat{V}_i - U_n)(\hat{V}_i - U_n)^\top$ (suppressing θ_0).

$T_{n1} \leftarrow nU_n^\top S_n^{-1} U_n$; $\quad T_{n2} \leftarrow nU_n^\top S_{n2}^{-1} U_n$.

Step 3: Bootstrap-based weight selection

for $b = 1, \ldots, N$ **do**

  Draw $\{X_1^*, \ldots, X_n^*\}$ with replacement from $\{X_1, \ldots, X_n\}$.

  Compute $T_{n1}^{(b)}$ and $T_{n2}^{(b)}$ from the bootstrap sample at the fixed null value θ_0 (Steps 1–2).

end for

for each $a \in \mathcal{A}$ **do**

  $T_{n3}^{(b)}(a) \leftarrow a\, T_{n1}^{(b)} + (1-a)\, T_{n2}^{(b)}$, $b = 1, \ldots, N$.

  $\hat{E}_{n3}(a) \leftarrow \frac{1}{N}\sum_b T_{n3}^{(b)}(a)$; $\quad \hat{V}_{n3}(a) \leftarrow \frac{1}{N-1}\sum_b \big(T_{n3}^{(b)}(a) - \hat{E}_{n3}(a)\big)^2$.

  $Z^{(b)}(a) \leftarrow \big(T_{n3}^{(b)}(a) - \hat{E}_{n3}(a)\big) \big/ \sqrt{\hat{V}_{n3}(a)}$, $b = 1, \ldots, N$.

  Apply the KS test to $\{Z^{(b)}(a)\}_{b=1}^{N}$ vs $\mathcal{N}(0, 1)$; record p-value $\mathrm{pv}(a)$.

end for

$\hat{a} \leftarrow \arg\max_{a \in \mathcal{A}} \mathrm{pv}(a)$.

Step 4: Calibrated test

$\hat{E}_{n3} \leftarrow \hat{E}_{n3}(\hat{a})$; $\quad \hat{V}_{n3} \leftarrow \hat{V}_{n3}(\hat{a})$ (retrieved from Step 3).

$Z_{\mathrm{obs}} \leftarrow \big(2\ell(\theta_0) - \hat{E}_{n3}\big) \big/ \sqrt{\hat{V}_{n3}}$.

Reject H_0 if $|Z_{\mathrm{obs}}| > z_{\alpha_0/2}$ (upper $\alpha_0/2$ quantile of $\mathcal{N}(0, 1)$); otherwise **retain** H_0.

5. Simulation studies

In this section, we apply the JEL to several examples and demonstrate the advantages of the average calibration method (ACJEL) through numerical simulations. For EL computation, we use the reference [32], which is implemented in R as the function emplik.

5.1. Covariance matrix

In this numerical study, we consider the covariance matrix of d-dimensional observations. To maintain strict theoretical consistency with the scaling laws established in Section 2, the effective parameter dimension under test is $p = d(d+1)/2$. In the following tables, we report the results indexed by both the data dimension d and its corresponding parameter dimension p to avoid any ambiguity.

Statistical inference for the covariance matrix is a fundamental topic in statistics, with wide applications in principal component analysis, graphical models, discriminant analysis, and factor models. A simple unbiased estimator of the population covariance matrix is the sample covariance matrix, which can be expressed as a U-statistic

$$S_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})^{\mathrm{T}} = \binom{n}{2}^{-1} \sum_{1\le i<j\le n} \frac{1}{2}(X_i - X_j)(X_i - X_j)^{\mathrm{T}}.$$

In this section, we apply the high-dimensional JEL to covariance matrix estimation.

To fully evaluate the ACJEL method, we simulate four representative multivariate distributions for X_i: (1) multivariate normal; (2) multivariate t with 4 degrees of freedom; (3) $X_i = A^{1/2}Z_i$, where Z_i follows a centered chi-squared distribution with one degree of freedom; (4) $X_i = A^{1/2}Z_i$, where Z_i is uniform on $[-1, 1]$. The $d \times d$ covariance matrix $\Sigma = (\sigma_{ij})$ is defined with $\sigma_{ij} = 0.5^{|i-j|}$.

The effective parameter dimension $p = d(d+1)/2$. To study the impact of dimension, we consider sample sizes $n \in \{200, 400, 800\}$ and dimensions p satisfying condition (A4), as detailed in Table 1. Based on $N = 5000$ replications, we compare ACJEL with the two calibrations (CJEL1 and CJEL2) and with standard JEL using normal (JEL-Nor) and chi-square (JEL-Chi) approximations.

Figures 3–5 show the performance of the correction methods under the multivariate normal distribution (Case 1), and Figures B.1–B.3 (Appendix B) present the results for the multivariate t-distribution (Case 2). The scatter and Q-Q plots consistently show that the standard normal approximation deviates markedly from the theoretical quantiles, confirming the need for correction. Among the corrected statistics, T_{n1} systematically underestimates the center of the distribution (especially as p increases), leading to under-coverage, while T_{n2} overestimates it, causing over-coverage. These drawbacks are more severe under the heavy-tailed t-distribution. By balancing the two extremes, the average-corrected statistic T_{n3} (ACJEL) aligns most closely with the normal quantile line $y = x$, effectively bringing the empirical coverage close to the nominal level. Cases 3 and 4 exhibit similar results.

Table 1 reports the coverage probabilities (nominal 95%) across different configurations. Performance generally improves with larger n or smaller p. JEL-Nor and JEL-Chi perform poorly, with coverage dropping to near zero for skewed or heavy-tailed data (Cases 2 and 3), as their asymptotic mean and variance severely underestimate the true moments of $2\ell(\theta_0)$ under heavy-tailed or skewed distributions with large p/n. CJEL1 and CJEL2 improve upon the uncorrected JEL, yet CJEL1 still undercovers while CJEL2 overcovers. This systematic discrepancy arises from the substantial difference in their variance estimates ($\hat{V}_{n1}$ versus $\hat{V}_{n2}$). ACJEL effectively addresses this issue by dynamically selecting an appropriate variance via the averaging method, consistently achieving the desired coverage and outperforming the other methods in nearly all scenarios.

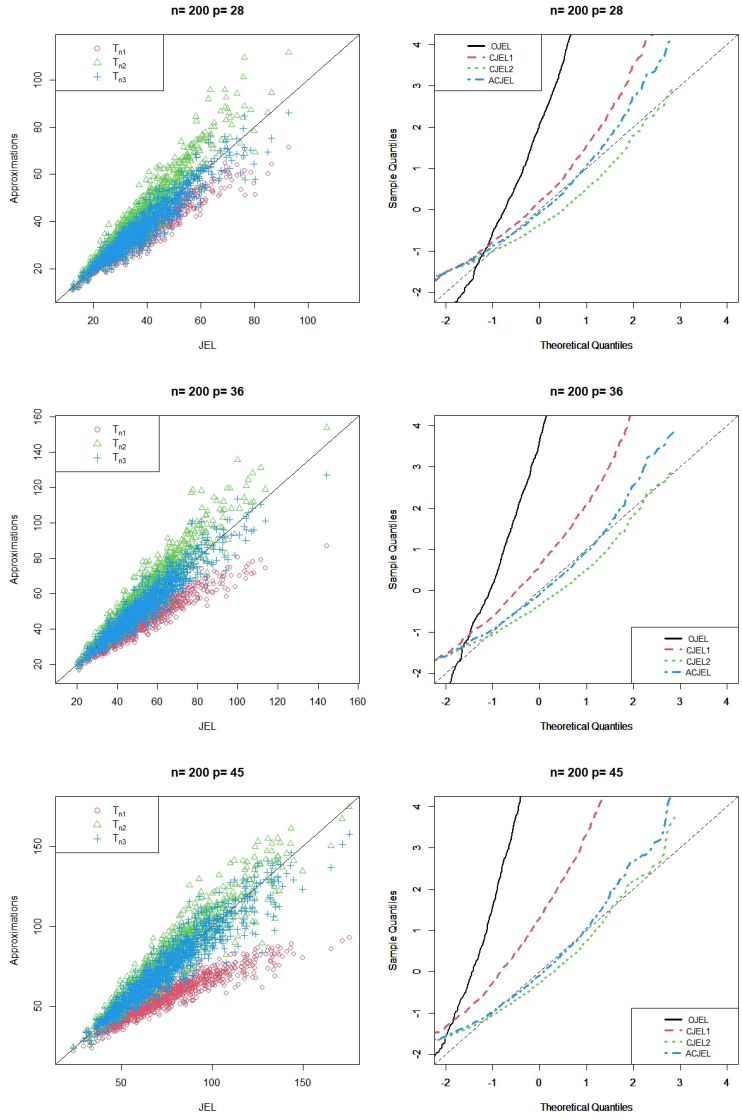

Figure 3. Scatter plots (left) and Q-Q plots (right) generated by the three calibration methods when X_i follows the multivariate normal distribution and $n = 200$.

5.2. *Linear regression*

Having shown the advantages of the ACJEL method for high-dimensional covariance matrices, we now apply it to U-shaped estimating equations, using linear regression as an illustrative example. Consider data generated from $Y_i = \gamma + X_i^{\mathrm{T}}\beta + e_i$ for $i = 1, \ldots, n$, where e_i is a zero-mean random error. An estimate of β can be obtained by solving the U-shaped estimating equation

$$U(\beta) = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} [Y_i - Y_j - (X_i - X_j)^{\mathrm{T}}\beta](X_i - X_j) = 0.$$

In the simulation, X_i is drawn from a multivariate normal distribution with identity covariance matrix, and the true p-dimensional parameter is $\beta_0 = (3, 1, 0.5, 0, \ldots, 0)^{\mathrm{T}}$, where only the first three entries are

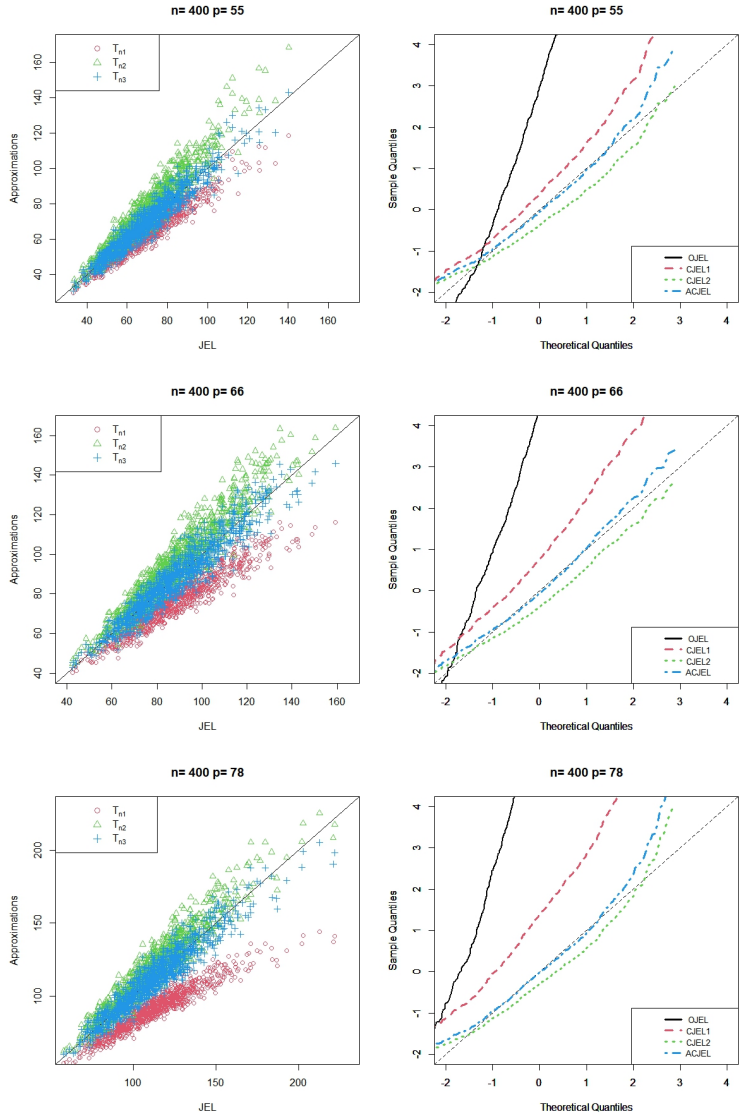

Figure 4. Scatter plots (left) and Q-Q plots (right) generated by the three calibration methods when X_i follows the multivariate normal distribution and $n = 400$.

nonzero. The error e_i follows either a standard normal distribution or a t-distribution with 6 degrees of freedom. The dimension p grows with the sample size n as $p = cn^{0.24}$ with $c \in \{2, 3, 4\}$. Table 2 presents the experimental results.

The table shows that for large p and n, ACJEL outperforms CJEL1, JEL-Chi, and JEL-Nor. Unlike the earlier examples with means and covariance matrices, ACJEL and CJEL2 yield identical results in most cases here—the weight a is effectively set to 0 in the simulation. This happens primarily because the coverage rate of CJEL2 does not exceed the nominal level, causing ACJEL to rely entirely on CJEL2. In practice, whether CJEL2 will over-cover is often unknown. When CJEL2 does over-cover, ACJEL averages CJEL1 and CJEL2 to achieve a more accurate coverage; when under-coverage is a risk, ACJEL favors CJEL2. Consequently, ACJEL remains consistently more robust and advantageous.

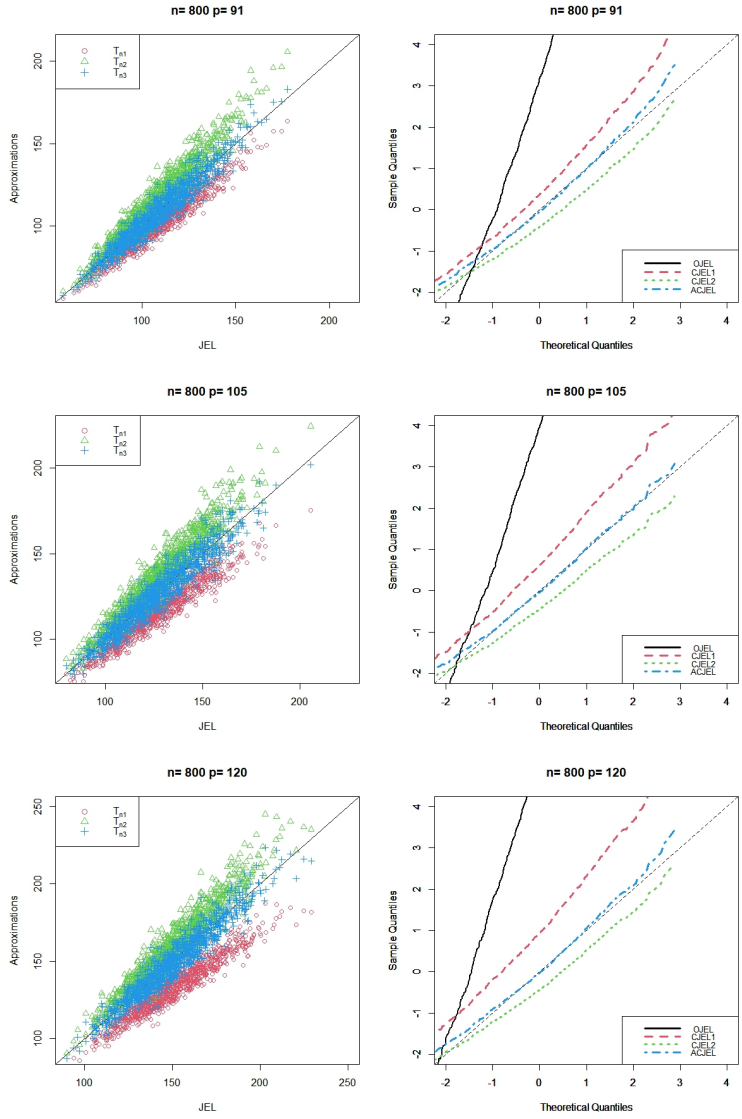

Figure 5. Scatter plots (left) and Q-Q plots (right) generated by the three calibration methods when X_i follows the multivariate normal distribution and $n = 800$.

5.3. Wilcoxon rank regression

To further stress-test the proposed method on a nonsmooth kernel, we consider Wilcoxon rank regression. The corresponding estimating equation is defined as a U-statistic of degree 2

$$U(\beta) = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} h(Z_i, Z_j; \beta) = 0,$$

where the symmetric kernel function $h(Z_i, Z_j; \beta)$ is given by

$$h(Z_i, Z_j; \beta) = (X_i - X_j) \left[I\{e_i(\beta) < e_j(\beta)\} - I\{e_i(\beta) > e_j(\beta)\} \right],$$

with $e_i(\beta) = Y_i - X_i^{\mathrm{T}}\beta$.

Table 1. The coverage probability of the covariance matrix

	n	d	p	ACJEL	CJEL2	CJEL1	JEL-Chi	JEL-Nor
(1)	200	7	28	0.952	0.984	0.898	0.696	0.734
		8	36	0.942	0.976	0.831	0.516	0.555
		9	45	0.937	0.969	0.647	0.316	0.344
	400	10	55	0.952	0.983	0.875	0.633	0.681
		11	66	0.951	0.973	0.812	0.486	0.538
		12	78	0.946	0.964	0.664	0.306	0.359
	800	13	91	0.951	0.969	0.899	0.670	0.720
		14	105	0.953	0.968	0.845	0.580	0.652
		15	120	0.958	0.982	0.788	0.490	0.544
(2)	200	7	28	0.925	0.995	0.751	0.039	0.043
		8	36	0.918	0.975	0.466	0.004	0.006
		9	45	0.891	0.901	0.164	0.000	0.000
	400	10	55	0.936	0.996	0.733	0.001	0.001
		11	66	0.932	0.988	0.479	0.000	0.000
		12	78	0.921	0.973	0.188	0.000	0.000
	800	13	91	0.948	0.969	0.872	0.002	0.002
		14	105	0.949	0.971	0.718	0.000	0.000
		15	120	0.936	0.952	0.498	0.000	0.000
(3)	200	7	28	0.906	0.975	0.754	0.112	0.126
		8	36	0.901	0.965	0.577	0.039	0.047
		9	45	0.879	0.879	0.264	0.008	0.010
	400	10	55	0.953	0.985	0.787	0.067	0.080
		11	66	0.948	0.986	0.628	0.020	0.025
		12	78	0.948	0.960	0.398	0.003	0.004
	800	13	91	0.963	0.991	0.875	0.077	0.094
		14	105	0.965	0.988	0.801	0.032	0.041
		15	120	0.965	0.981	0.678	0.011	0.021
(4)	200	7	28	0.939	0.984	0.908	0.810	0.838
		8	36	0.942	0.976	0.831	0.680	0.711
		9	45	0.934	0.963	0.686	0.475	0.528
	400	10	55	0.938	0.983	0.881	0.751	0.783
		11	66	0.961	0.980	0.830	0.652	0.704
		12	78	0.951	0.972	0.702	0.505	0.565
	800	13	91	0.952	0.972	0.900	0.810	0.832
		14	105	0.949	0.965	0.871	0.733	0.781
		15	120	0.950	0.968	0.821	0.615	0.680

To investigate the sensitivity of the performance to the ratio of the parameter dimension to the sample size (p/n), we let the dimension grow as $p = cn^{0.24}$ with $c \in \{5, 6, 7\}$ under sample sizes $n \in \{100, 200, 400\}$. This configuration yields a wider range of p/n ratios than those in the previous subsections. All other simulation settings remain identical to those described in Section 5.2.

Table 3 reports the empirical coverage probabilities at the nominal 95% level. Both JEL-Nor and JEL-Chi exhibit substantial under-coverage across all configurations, confirming that the standard asymptotic approximation becomes unreliable when the kernel lacks smoothness. In contrast, ACJEL attains an empirical coverage closest to 0.95 in nearly every scenario, outperforming CJEL1, JEL-Chi, and JEL-Nor. Under $t(6)$ errors, where heavier tails amplify the higher-order remainder terms, the advantages of ACJEL over the single-calibration alternatives become even more pronounced.

Furthermore, as the p/n ratio increases, the performance of all methods deteriorates, yet ACJEL degrades at the slowest rate. For the most challenging case considered ($n = 100$, $p = 21$), the coverage probability of CJEL1 drops to approximately 0.546, whereas ACJEL maintains a reliable coverage level above 0.860. This consistent pattern across both error distributions demonstrates that ACJEL

Table 2. The coverage probability of the linear model coefficient

	n	p	ACJEL	CJEL2	CJEL1	JEL-Chi	JEL-Nor
$N(0,1)$	200	7	0.923	0.932	0.910	0.910	0.914
		10	0.908	0.908	0.868	0.892	0.898
		14	0.896	0.896	0.843	0.849	0.863
	400	8	0.949	0.949	0.938	0.941	0.947
		12	0.933	0.933	0.912	0.908	0.915
		16	0.936	0.936	0.891	0.908	0.931
	800	9	0.948	0.954	0.948	0.941	0.948
		14	0.943	0.951	0.935	0.933	0.943
		19	0.936	0.936	0.919	0.928	0.930
$t(6)$	200	7	0.910	0.910	0.888	0.890	0.898
		10	0.892	0.892	0.855	0.873	0.885
		14	0.849	0.849	0.772	0.805	0.826
	400	8	0.933	0.933	0.923	0.933	0.940
		12	0.904	0.904	0.882	0.887	0.894
		16	0.913	0.913	0.875	0.877	0.898
	800	9	0.938	0.941	0.932	0.939	0.948
		14	0.940	0.940	0.929	0.920	0.930
		19	0.923	0.923	0.905	0.891	0.905

exhibits strong robustness against a high dimension-to-sample-size ratio, even when dealing with nonsmooth kernels.

In summary, the Wilcoxon rank-regression experiment confirms that the advantages of ACJEL identified in smooth-kernel settings successfully carry over to the practically important class of non-smooth U-statistics, maintaining robust coverage across a broad range of p/n ratios.

Table 3. The coverage probability of Wilcoxon rank regression coefficients

	n	p	ACJEL	CJEL2	CJEL1	JEL-Chi	JEL-Nor
$N(0,1)$	100	15	0.927	0.927	0.776	0.729	0.744
		18	0.890	0.890	0.668	0.604	0.622
		21	0.863	0.863	0.546	0.469	0.509
	200	17	0.925	0.941	0.880	0.862	0.873
		21	0.943	0.943	0.859	0.814	0.829
		24	0.940	0.940	0.808	0.754	0.783
	400	21	0.952	0.968	0.938	0.917	0.929
		25	0.949	0.965	0.916	0.901	0.923
		29	0.931	0.955	0.894	0.862	0.873
$t(6)$	100	15	0.913	0.913	0.755	0.720	0.738
		18	0.888	0.888	0.674	0.617	0.642
		21	0.871	0.871	0.547	0.464	0.493
	200	17	0.937	0.953	0.892	0.866	0.884
		21	0.940	0.944	0.855	0.805	0.834
		24	0.940	0.941	0.814	0.766	0.791
	400	21	0.941	0.961	0.929	0.898	0.915
		25	0.943	0.959	0.928	0.896	0.910
		29	0.945	0.962	0.904	0.856	0.868

5.4. Computational cost

All three methods—CJEL1, CJEL2, and ACJEL—share the same dominant computational step: evaluating the N jackknife pseudo-value arrays $\widehat{V}_i(\theta_0)$ and the corresponding EL statistic $2\ell(\theta_0)$. The per-replicate cost depends on the setting. For the covariance matrix, computing all n pseudo-values

requires $n + 1$ calls to the sample covariance function, costing $O(n^2 d^2)$ in total. For linear regression, the closed-form identity

$$\widehat{U}_n(\theta) = \frac{2\big[n(X^\top e) - (\mathbf{1}^\top X)(\mathbf{1}^\top e)\big]}{n(n-1)}, \quad e_i = Y_i - X_i^\top \theta,$$

with rank-1 leave-one-out updates reduces the pseudo-value cost to $O(np)$; the EL Newton step $O(np^2)$ then dominates. For Wilcoxon rank regression, the sign-function kernel requires an $O(n^2)$ outer product followed by a single matrix multiplication $XS = X^\top S$ of cost $O(n^2 p)$; each pseudo-value then costs only $O(p)$. In all three settings, the dominant cost is shared identically by CJEL1, CJEL2, and ACJEL.

The only overhead specific to ACJEL is Step 3 of Algorithm 1: For a grid $\mathcal{A} \subset [0, 1]$ of size $|\mathcal{A}|$, a KS test is applied to an already-computed length-N vector at each grid point, costing $O(|\mathcal{A}| \cdot N \log N)$ in total—negligible relative to the simulation cost.

Table 4 reports measured times ($N = 5000$ replicates, default grid $\mathcal{A} = \{0, 0.01, \ldots, 1\}$). Each row covers ACJEL, CJEL1, and CJEL2 simultaneously, since they share the simulation cost. "Simulation time" denotes Steps 1–2 of Algorithm 1; "Weight selection time" denotes Step 3 (KS-based weight selection, ACJEL-specific).

Table 4. Time (seconds, $N = 5000$ replicates) for the three simulation settings

Setting	n	p	Simulation time (s)	Weight selection time (s)
Covariance	200	28–45	51–86	<0.08
	400	55–78	221–358	<0.08
	800	91–120	980–1443	<0.10
Linear regression	200	7–14	7–11	<0.08
	400	8–16	12–19	<0.08
	800	9–19	22–38	<0.08
Wilcoxon rank regr.	100	15–21	11–15	<0.09
	200	18–25	20–28	<0.09
	400	21–30	54–74	<0.10

The weight-selection step accounts for less than 0.2% of total runtime in every configuration; the ACJEL procedure therefore adds essentially no computational overhead relative to the individual calibrations CJEL1 and CJEL2. For all three settings, the absolute times grow roughly as $O(n^2 N)$ in the dominant term, consistent with the figures in Table 4.

In Sections 5.1–5.3 we used $N = 5000$ Monte Carlo replicates to obtain stable coverage estimates. For a single applied dataset, a sensitivity analysis shows that $N = 500$ bootstrap replicates yield $\hat{a}$ values within ± 0.05 of those from $N = 5000$, with coverage probabilities differing by at most 0.01. We therefore recommend $N = 500$ as the default for practical applications; this reduces total runtime by a factor of 10 at no meaningful loss in accuracy. A coarser grid $\mathcal{A} = \{0, 0.05, \ldots, 1\}$ ($|\mathcal{A}| = 21$) further accelerates Step 3 without appreciably changing $\hat{a}$.

6. Application to real data example

Duchenne muscular dystrophy (DMD) is a common genetic degenerative muscle disorder, primarily transmitted from mother to child. Affected males typically die at a young age, while female carriers remain asymptomatic. Early diagnosis of female carriers is critical, as no cure currently exists. Studies indicate that carriers often show elevated levels of serum enzymes, such as creatine kinase (CK),

hemopexin (H), lactate dehydrogenase (LD), and pyruvate kinase (PK). Dr. M. Thompson at the Hospital for Sick Children in Toronto conducted a study measuring these four enzymes in 75 carriers and 134 non-carriers, aiming to develop a method for assessing carriers probability using serum markers and family pedigrees. The data, drawn from Table 38.1 of Andrews and Herzberg [33], provide key evidence for an effective DMD carriers screening procedure.

We consider the sample covariance matrix of CK and H levels for non-carriers, where $d = 2$ and the effective dimension is $p = d(d+1)/2 = 3$. Because only one real dataset is available, we use the bootstrap method: In each replication, a subsample of size $n = 33$ is randomly drawn from the 134 non-carrier observations. This satisfies condition (A4) required by the theorem, with $p = cn^{1/4-0.01}$ and $c = 1.3$, and allows us to compute the weight a via the KS test. Repeating the experiment 1000 times yields 95% coverage rates of 0.940 (ACJEL), 0.935 (CJEL1), 0.989 (CJEL2), 0.808 (JEL-Chi), and 0.807 (JEL-Nor). The results show that the calibrated JEL methods substantially outperform JEL-Chi and JEL-Nor, demonstrating their advantage in high-dimensional settings. Furthermore, ACJEL performs better than CJEL2, as the automatic weight selection in the average correction effectively addresses the CJEL2's shortcomings.

Regarding computational cost, the bootstrap procedure with $N = 1{,}000$ replications completes in approximately 3.16 seconds, making the method practically feasible. For settings with larger n or p, the data-splitting alternative described in Section 4 provides a computationally lighter means of estimating a and the moments (E_{n3}, V_{n3}) without any increase in asymptotic bias.

7. Conclusions

This paper improves the JEL method for U-statistics and U-shaped estimating equations. By correcting the mean and variance of the JEL ratio function, Type I error is controlled in high dimensions. Asymptotic normality of the high-dimensional JEL combined with U-shaped estimating equations is established. Simulations show that when p/n is large, using normal quantiles to construct confidence regions yields low coverage. This occurs because the true mean and variance of $2\ell(\theta_0)$ differ substantially from the asymptotic values $(p, 2p)$. Following EL corrections, one can replace $(p, 2p)$ with estimators $(\hat{E}_{ni}, \hat{V}_{ni})$ for $i = 1, 2$ that better approximate the true moments. However, these two corrections perform poorly for JEL. We therefore propose an average correction that combines them, replacing $(p, 2p)$ with $(\hat{E}_{n3}, \hat{V}_{n3})$, which achieves significantly better results.

Despite the progress made in this paper, several issues remain for future research. We propose a data-driven method for selecting the average weight a via hypothesis testing; developing a theoretically justified selection procedure remains an open challenge. Following Liu et al. [27], approximate expressions for the mean and variance of the quadratic forms T_{n1} and T_{n2} could be derived to enable theoretical weight selection. Additionally, Condition (A4) requires $p^t/n \to 0$ with $t = \max\{3\alpha/(\alpha - 2), 4\}$, where $\alpha \geq 4$ is the moment exponent appearing in Condition (A2). Since $3\alpha/(\alpha - 2)$ is strictly decreasing in α and equals 4 at $\alpha = 8$, the condition simplifies to $p = o(n^{1/4})$ when $\alpha \geq 8$. For smaller values of α, the restriction on p is more stringent: For instance, $\alpha = 4$ gives $t = 6$, i.e., $p = o(n^{1/6})$. Since the correction theory for standard high-dimensional EL allows $p = o(n^{1/2})$, relaxing Condition (A4) in the JEL framework is an important direction for future work.

The JEL successfully combines U-statistics with U-shaped estimating equations, broadening the scope of EL. Further work should explore whether other nonlinear statistics can be similarly integrated.

Extending these methods also faces the major challenge of managing the computational complexity inherent in jackknifing high-dimensional data.

Data and code availability

The R code implementing the ACJEL procedure, together with scripts reproducing all simulation studies in Sections 5, is publicly available at https://github.com/shangmd797/ACJEL.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

The authors sincerely thank the referees and the editor for their valuable comments, which substantially improved an earlier version of this paper. This work was supported by the National Social Science Foundation of China (22BTJ017).

Conflict of interest

The authors declare there are no conflicts of interest.

References

1. A. B. Owen, Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, **75** (1988), 237–249. https://doi.org/10.1093/biomet/75.2.237
2. A. B. Owen, *Empirical Likelihood*, Chapman and Hall/CRC, New York, 2001. https://doi.org/10.1201/9781420036152
3. A. B. Owen, Empirical Likelihood Ratio Confidence Regions, *Ann. Stat.*, **18** (1990), 90–120. https://doi.org/10.1214/aos/1176347494
4. J. Qin, J. F. Lawless, Empirical likelihood and general estimating equations, *Ann. Statist.*, **22** (1994), 300–325. https://doi.org/10.1214/aos/1176325370
5. J. Shi, T. S. Lau, Empirical Likelihood for Partially Linear Models, *J. Multivariate Anal.*, **72** (2000), 132–148. https://doi.org/10.1006/jmva.1999.1866
6. L. Xue, L. Zhu, Empirical likelihood for a varying coefficient model with longitudinal data, *J. Amer. Stat. Assoc.*, **102** (2007), 642–654. https://doi.org/10.1198/016214507000000293
7. S. X. Chen, H. J. Cui, On Bartlett and Bartlett-type corrections of empirical likelihood in the presence of nuisance parameters, *Biometrika*, **93** (2006), 215–220. https://doi.org/10.1093/biomet/93.1.215
8. Y. Kitamura, Empirical likelihood methods with weakly dependent processes, *Ann. Stat.*, **25** (1997), 2084–2102. https://doi.org/10.1214/aos/1069362388

9. W. Wang, X. Song, G. Han, K. Yang, Bayesian empirical likelihood inference and order shrinkage for a hysteretic autoregressive model, *Stat. Papers*, **66** (2025), 36. https://doi.org/10.1007/s00362-025-01659-0
10. G. S. Qin, B. Y. Jing, Empirical likelihood for censored linear regression, *Scand. J. Stat.*, **28** (2001), 661–673. https://doi.org/10.1111/1467-9469.00261
11. W. K. Newey, R. J. Smith, Higher order properties of GMM and generalized empirical likelihood estimators, *Econometrica*, **72** (2004), 219–255. https://doi.org/10.1111/j.1468-0262.2004.00482.x
12. T. J. DiCiccio, P. Hall, J. P. Romano, Empirical likelihood is Bartlett-correctable, *Ann. Stat.*, **19** (1991), 1053–1061. https://doi.org/10.1214/aos/1176348137
13. S. X. Chen, H. J. Cui, On the second-order properties of empirical likelihood with moment restrictions, *J. Econometrics*, **141** (2007), 492–516. https://doi.org/10.1016/j.jeconom.2006.10.006
14. C. Y. Tang, C. Leng, Penalized high-dimensional empirical likelihood, *Biometrika*, **97** (2010), 905–919. https://doi.org/10.1093/biomet/asq057
15. N. L. Hjort, I. McKeague, I. Van Keilegom, Extending the scope of empirical likelihood, *Ann. Stat.*, **37** (2009), 1079–1111. https://doi.org/10.1214/07-AOS555
16. S. X. Chen, L. Peng, Y. L. Qin, Effects of data dimension on empirical likelihood, *Biometrika*, **96** (2009), 711–722. https://doi.org/10.1093/biomet/asp037
17. A. T. A. Wood, K. A. Do, B. M. Broom, Sequential linearization of empirical likelihood constraints with application to *U*-statistics, *J. Comput. Graph. Stat.*, **5** (1996), 365–385. https://doi.org/10.1080/10618600.1996.10474718
18. B. Y. Jing, J. Yuan, W. Zhou, Jackknife empirical likelihood, *J. Am. Stat. Assoc.*, **104** (2009), 1224–1232. https://doi.org/10.1198/jasa.2009.tm08260
19. M. Li, L. Peng, Y. Qi, Reduce computation in profile empirical likelihood method, *Can. J. Stat.*, **39** (2011), 370–384. https://doi.org/10.1002/cjs.10101
20. L. Peng, Approximate jackknife empirical likelihood method for estimating equations, *Can. J. Stat.*, **40** (2012), 110–123. https://doi.org/10.1002/cjs.10138
21. Z. Li, J. Xu, W. Zhou, On nonsmooth estimating functions via jackknife empirical likelihood, *Scand. J. Stat.*, **43** (2016), 49–69. https://doi.org/10.1111/sjos.12164
22. Y. Wei, Z. P. Li, F. Yang, Jackknife empirical likelihood for growing dimensional nonsmooth estimating equations, *Sci. Sin. Math.*, **49** (2019), 1103–1118. https://doi.org/10.1360/scm-2018-0239
23. R. Wang, L. Peng, Y. Qi, Jackknife empirical likelihood test for equality of two high dimensional means, *Stat. Sin.*, **23** (2013), 667–690. https://doi.org/10.5705/ss.2011.261
24. Z. Liu, X. Xia, W. Zhou, A test for equality of two distributions via jackknife empirical likelihood and characteristic functions, *Comput. Stat. Data Anal.*, **92** (2015), 97–114. https://doi.org/10.1016/j.csda.2015.06.004
25. H. Peng, T. Fei, Jackknife empirical likelihood goodness-of-fit tests for *U*-statistics based general estimating equations, *Bernoulli*, **24** (2018), 449–464. https://doi.org/10.3150/16-BEJ884

26. C. Cheng, Y. Liu, Z. Liu, W. Zhou, Balanced augmented jackknife empirical likelihood for two sample U-statistics, *Sci. China Math.*, **61** (2018), 1129–1138. https://doi.org/10.1007/s11425-016-9071-y

27. Y. Liu, C. Zou, Z. Wang, Calibration of the empirical likelihood for high dimensional data, *Ann. Inst. Stat. Math.*, **65** (2013), 529–550. https://doi.org/10.1007/s10463-012-0384-7

28. H. Guo, C. L. Zou, Z. J. Wang, B. Chen, Empirical likelihood for high-dimensional linear regression models, *Metrika*, **77** (2014), 921–945. https://doi.org/10.1007/s00184-013-0479-z

29. J. W. He, X. Chen, Calibration of the empirical likelihood for semiparametric varying-coefficient partially linear models with diverging number of parameters, *Hacet. J. Math. Stat.*, **48** (2019), 230–241. https://doi.org/10.15672/HJMS.2018.604

30. R. Cavoretto, A. De Rossi, An adaptive residual sub-sampling algorithm for kernel interpolation based on maximum likelihood estimations, *J. Comput. Appl. Math.*, **418** (2023), 114658. https://doi.org/10.1016/j.cam.2022.114658

31. R. Cavoretto, A. De Rossi, E. Perracchione, Learning with partition of unity-based Kriging estimators, *Appl. Math. Comput.*, **448** (2023), 127938. https://doi.org/10.1016/j.amc.2023.127938

32. A. B. Owen, Self-concordance for empirical likelihood, *Can. J. Stat.*, **41** (2013), 387–397. https://doi.org/10.1002/cjs.11183

33. D. F. Andrews, A. M. Herzberg, *Data: A Collection of Problems from Many Fields for the Student and Research Worker*, Springer-Verlag, New York, 1985. https://doi.org/10.1007/978-1-4612-5098-2

34. A. J. Lee, *U-Statistics: Theory and Practice*, Marcel Dekker, New York, 1990. https://doi.org/10.1201/9780203734520

Appendix

A. Proofs

All proofs in Appendix A are given explicitly for kernel degree $m = 2$ (without loss of generality, as stated at the beginning of that appendix). The extension to any fixed $m \geq 2$ is immediate via the Hoeffding decomposition: For general m, each pseudo-value still decomposes as

$$\hat{V}_{il}(\theta_0) = m g_l(X_i, \theta_0) + R_{ni,l}(\theta_0),$$

where $g(x, \theta_0) = \binom{m}{1}\mathrm{E}[h(x, X_2, \ldots, X_m; \theta_0)]$, and $R_{ni,l}$ is controlled by the same moment bound in Condition (A2). All error bounds in Lemmas 1–5 acquire multiplicative constants depending only on m, and the overall proof structure is unchanged. We use C to denote a generic positive constant whose value may change from line to line. For a matrix or vector A, $\|A\|$ denotes the Frobenius norm or Euclidean norm, respectively.

According to the Hoeffding decomposition, the l-th component of $U_n(\theta_0)$ can be written as

$$U_{nl}(\theta_0) = \frac{2}{n}\sum_{i=1}^{n} g_l(X_i, \theta_0) + \binom{n}{2}^{-1} \sum_{s<t}^{n} \varphi_l(X_s, X_t, \theta_0),$$

where $g(x,\theta_0) = \mathrm{E}[h(x,X_1,\theta_0)]$ and $\varphi(x,y,\theta_0) = h(x,y,\theta_0) - g(x,\theta_0) - g(y,\theta_0)$. Consequently, the jackknife pseudo-value decomposes into

$$\hat{V}_{il}(\theta_0) = 2g_l(X_i,\theta_0) + R_{ni,l}(\theta_0), \tag{A.1}$$

where $R_{ni,l}(\theta_0) = \frac{2}{n-2}\sum_{k\neq i}^{n}\varphi_l(X_i,X_k,\theta_0) - \binom{n-1}{2}^{-1}\sum_{s<t,s\neq i,t\neq i}^{n}\varphi_l(X_s,X_t,\theta_0)$.

Lemma 1. *Let $S_n(\theta_0) = \frac{1}{n}\sum_{i=1}^{n}\hat{V}_i(\theta_0)\hat{V}_i^{\mathrm{T}}(\theta_0)$, and Σ_n be the asymptotic covariance matrix of $\sqrt{n}U_n$, whose (k,l)-th element is $4\,\mathrm{Cov}(g_k(X_1,\theta_0),g_l(X_1,\theta_0))$. Under conditions (A1)–(A3), we have $\|S_n(\theta_0) - \Sigma_n\| = O_p(p/\sqrt{n})$.*

Proof. By definition, $S_n(\theta_0)$ can be expanded as

$$S_n(\theta_0) = \frac{1}{n}\sum_{i=1}^{n}(\hat{V}_i(\theta_0) - U_n(\theta_0))(\hat{V}_i(\theta_0) - U_n(\theta_0))^{\mathrm{T}} + U_n(\theta_0)U_n^{\mathrm{T}}(\theta_0).$$

From Lemma 1 in Jing et al. [18], $\|U_n(\theta_0)\|^2 = O_p(p/n)$, rendering the second term negligible. The first term is equivalent to $(n-1)\widehat{\mathrm{Cov}}(\mathrm{Jack})$. Following Lee [34], the jackknife covariance estimator satisfies $\mathrm{Var}(\widehat{\mathrm{Cov}}_{jk}) = O(n^{-3})$ and $\mathrm{E}[\widehat{\mathrm{Cov}}_{jk}] - \mathrm{Cov}(U_{nj},U_{nk}) = O(n^{-2})$. Applying Markov's inequality to the Frobenius norm of the difference yields

$$\mathrm{P}\left(\|(n-1)\widehat{\mathrm{Cov}}(\mathrm{Jack}) - \Sigma_n\| \geq \frac{p}{\sqrt{n}}M_\epsilon\right) \leq \frac{n}{p^2M_\epsilon^2}\sum_{j,k=1}^{p}O(n^{-3}) = O(M_\epsilon^{-2}).$$

Thus, $\|S_n(\theta_0) - \Sigma_n\| = O_p(p/\sqrt{n})$. □

Lemma 2. *If conditions (A1)–(A4) hold, 0 is an interior point of the convex hull of $\{\hat{V}_1(\theta_0),\ldots,\hat{V}_n(\theta_0)\}$ with probability approaching 1.*

Proof. Under condition (A2), higher-order moments of the U-statistic remainders are uniformly bounded by $O(p/n)$. Thus, $R_{ni,l}(\theta_0)$ in (A.1) converges uniformly to 0 in probability. The problem then reduces to whether 0 is in the convex hull of $\{2g_l(X_i,\theta_0)\}$, which holds directly by Lemma 8 of Liu et al. [27]. □

Lemma 3. *Under condition (A2), $\max_{1\leqslant i\leqslant n}\|\hat{V}_i(\theta_0)\| = O_p(n^{1/\alpha}p^{1/2})$.*

Proof. By Minkowski's inequality and condition (A2), $\mathrm{E}[|\hat{V}_{ik}(\theta_0)|^\alpha] \leq 3^\alpha \max_k \mathrm{E}[|h_k(X_1,X_2;\theta_0)|^\alpha] \leq C$. Using Hölder's inequality, we obtain $\mathrm{E}[(\|\hat{V}_i(\theta_0)\|p^{-1/2})^\alpha] \leq p^{-1}\sum_{k=1}^{p}\mathrm{E}[|\hat{V}_{ik}(\theta_0)|^\alpha] \leq C$. Finally, an application of Chebyshev's inequality over $1\leq i\leq n$ directly yields the uniform bound $\max_i\|\hat{V}_i(\theta_0)\| = O_p(n^{1/\alpha}p^{1/2})$. □

Lemma 4. *Under conditions (A2) and (A4), we have $\frac{p}{\sqrt{n}}\max_{1\leq i\leq n}\|\hat{V}_i(\theta_0)\| \xrightarrow{p} 0$.*

Proof. This follows immediately from Lemma 3 and Chebyshev's inequality. □

Lemma 5. *Under conditions (A1)–(A4), we have $\|\lambda(\theta_0)\| = O_p(a_n)$.*

Proof. Let $\lambda(\theta_0) = \rho u$ with $\rho = \|\lambda(\theta_0)\|$ and $\|u\| = 1$. From the score equation, we obtain $\rho u^{\mathrm{T}}\widetilde{S}_n(\theta_0)u = u^{\mathrm{T}}U_n(\theta_0)$, where $\widetilde{S}_n(\theta_0)$ is the empirically reweighted covariance matrix. Bounding $\widetilde{S}_n$ by S_n yields $\rho(u^{\mathrm{T}}S_n(\theta_0)u) \leq u^{\mathrm{T}}U_n(\theta_0)(1 + \rho \max_i \|\hat{V}_i(\theta_0)\|)$. By Lemma 1 and condition (A3), $u^{\mathrm{T}}S_n(\theta_0)u \asymp O_p(1)$. Combined with Lemma 3 and $u^{\mathrm{T}}U_n(\theta_0) = O_p(a_n)$, we deduce $\|\lambda(\theta_0)\| = O_p(a_n)$. □

Proof of Theorem 1. Let $T_n = 2\ell(\theta_0)$ and $T_{n1} = nU_n^{\mathrm{T}}(\theta_0)S_n^{-1}(\theta_0)U_n(\theta_0)$. Using a standard Taylor expansion of $G_n(\lambda) = 2\sum_{i=1}^n \log\{1 + \lambda^{\mathrm{T}}\hat{V}_i(\theta_0)\}$ around 0, we can express the empirical likelihood ratio as

$$G_n(\lambda) = 2n\lambda^{\mathrm{T}}U_n(\theta_0) - n\lambda^{\mathrm{T}}S_n(\theta_0)\lambda + r_n(\lambda).$$

For λ in the neighborhood $\Omega_n(C) = \{\lambda : \|\lambda\| \leq Ca_n\}$, the remainder term is bounded by

$$|r_n(\lambda)| \leq \frac{4}{3}n\|\lambda\|^3 \max_i \|\hat{V}_i(\theta_0)\|\lambda_{\max}(S_n(\theta_0)),$$

where $\lambda_{\max}$ is the largest eigenvalue. We now verify that $\sup_{\lambda\in\Omega_n}|r_n(\lambda)| = o_p(p^{1/2})$. By Lemma 5, $\|\lambda\| = O_p(a_n)$, and by Lemma 3, $\max_i \|\hat{V}_i(\theta_0)\| = O_p(n^{1/\alpha}p^{1/2})$. Substituting into the bound for $|r_n(\lambda)|$ gives

$$\sup_{\lambda\in\Omega_n}|r_n(\lambda)| = O_p\left(p^2 n^{1/\alpha - 1/2}\right).$$

Requiring this to be $o_p(p^{1/2})$ yields $p^{3\alpha/(\alpha-2)}/n \to 0$, corresponding to the constraint $t \geq 3\alpha/(\alpha-2)$ in Condition (A4). Intuitively, a smaller α implies a heavier-tailed kernel h, causing $\max_i \|\hat{V}_i(\theta_0)\|$ to grow faster, so p must be restricted more stringently. With this bound established, $\sup_{\lambda\in\Omega_n}|r_n(\lambda)| = o_p(p^{1/2})$, and hence $(T_n - T_{n1})/p^{1/2} \xrightarrow{p} 0$.

Next, let $K_n = nU_n^{\mathrm{T}}(\theta_0)\Sigma_n^{-1}U_n(\theta_0)$ and write $S_n(\theta_0) = \Sigma_n + \epsilon_n$. A standard matrix inverse approximation gives $S_n^{-1}(\theta_0) \approx \Sigma_n^{-1} - \Sigma_n^{-1}\epsilon_n\Sigma_n^{-1}$. By Lemma 1,

$$|T_{n1} - K_n| \leq \left|nU_n^{\mathrm{T}}(\theta_0)\Sigma_n^{-1}\epsilon_n\Sigma_n^{-1}U_n(\theta_0)\right| = O_p\left(p^2/\sqrt{n}\right).$$

Requiring $|T_{n1} - K_n| = o_p(p^{1/2})$ yields $p^4/n \to 0$, corresponding to the constraint $t \geq 4$ in Condition (A4). This constraint arises independently from controlling the approximation error between T_{n1} and K_n and does not depend on α. Taking the maximum of the two constraints gives $t = \max\{3\alpha/(\alpha-2), 4\}$. Since $3\alpha/(\alpha-2)$ is strictly decreasing in α and equals 4 at $\alpha = 8$: when $\alpha \geq 8$ the second constraint dominates and Condition (A4) reduces to $p = o(n^{1/4})$; when $4 \leq \alpha < 8$, the first constraint is binding, e.g., $\alpha = 4$ gives $t = 6$, i.e., $p = o(n^{1/6})$.

Finally, with $|T_n - K_n| = o_p(p^{1/2})$ established, asymptotic normality follows from Lemma A.2 in Wei et al. [22]. □

Proof of Theorem 2. From Theorem 1, $2\ell(\theta_0) - T_{n1} = o_p(\sqrt{p})$. Let $T_{n2} \equiv nU_n^{\mathrm{T}}(\theta_0)S_{n2}^{-1}(\theta_0)U_n(\theta_0)$. From Lemma 1, we know that $\|S_{n2}(\theta_0) - \Sigma_n\| = O_p(p/\sqrt{n})$. Denoting $S_{n2}(\theta_0) = \Sigma_n + \epsilon_{n2}$, a similar algebraic expansion gives

$$|T_{n2} - K_n| \leq n\lambda_{\min}^{-2}(\Sigma_n)\|U_n(\theta_0)\|^2\|\epsilon_{n2}\| = nO_p(1)O_p(p/n)O_p(p/\sqrt{n}) = o_p(1).$$

This completes the proof. □

B. Additional numerical results

In the sample covariance matrix example, Figures B.1–B.3 show the correction results when the data follow a multivariate t-distribution.

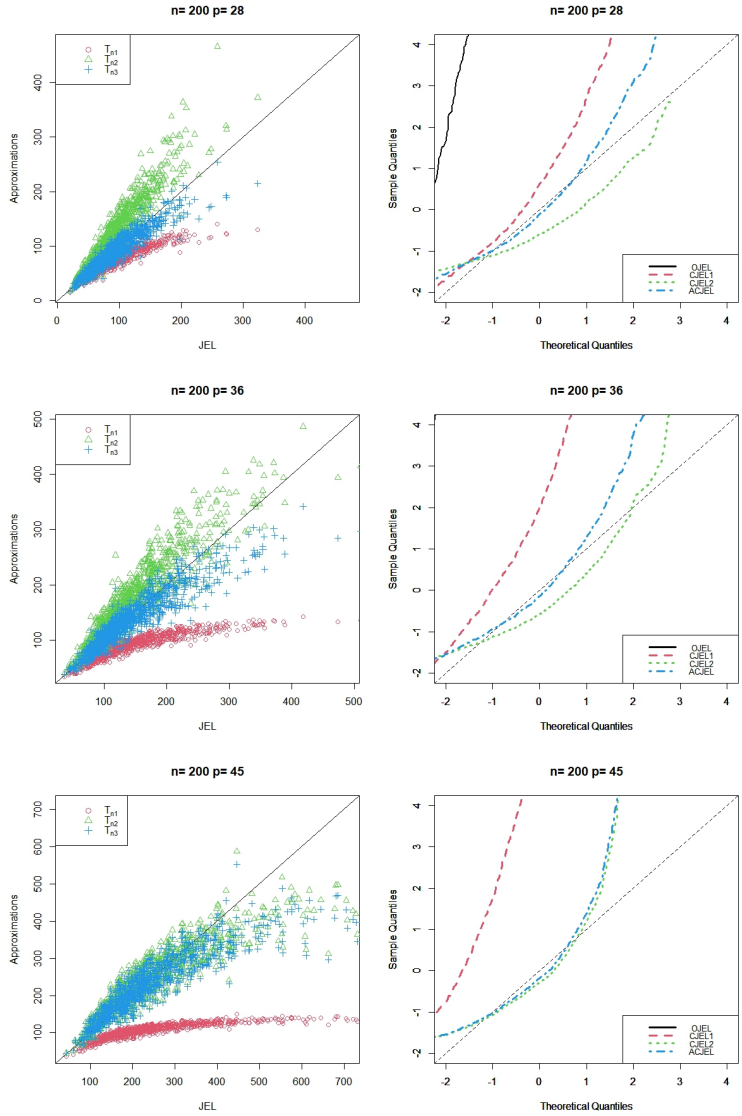

Figure B.1. Scatter plots (left) and Q-Q plots (right) generated by the three calibration methods when X_i follows the multivariate t-distribution and $n = 200$.

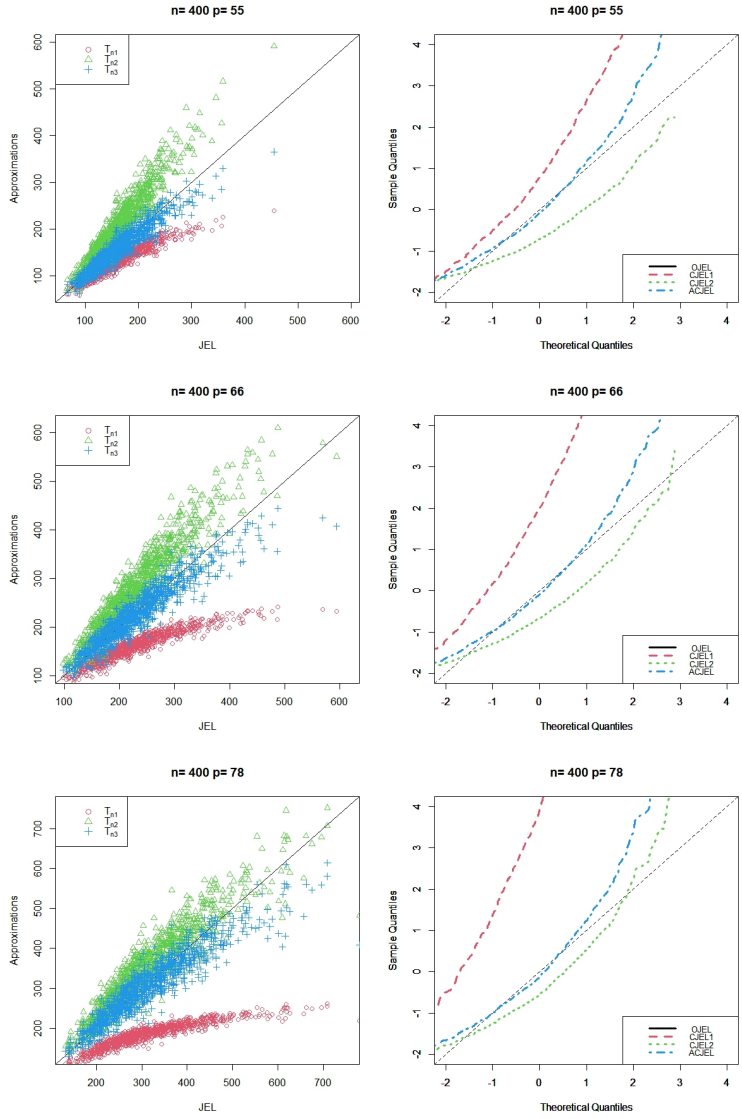

Figure B.2. Scatter plots (left) and Q-Q plots (right) generated by the three calibration methods when X_i follows the multivariate t-distribution and $n = 400$.

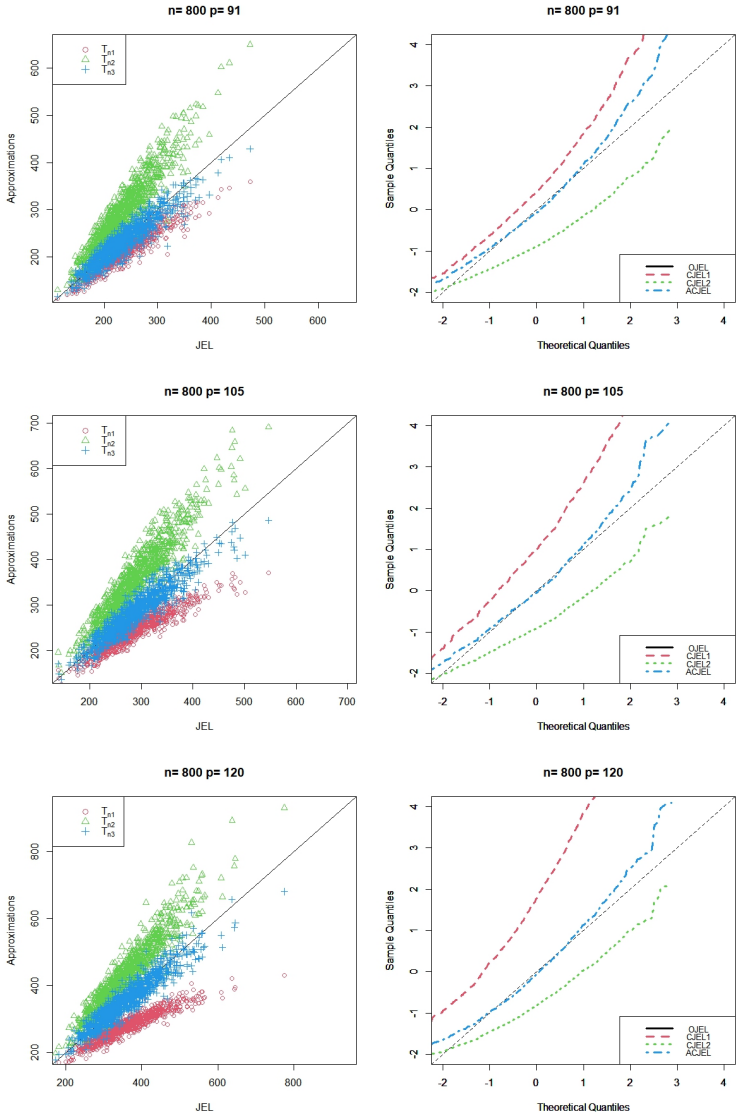

Figure B.3. Scatter plots (left) and Q-Q plots (right) generated by the three calibration methods when X_i follows the multivariate t-distribution and $n = 800$.