Comment on "Generalized Factor Model for Ultra-High Dimensional Correlated Variables with Mixed Types"

Zhen Li*

Abstract

A recent paper by Schaeffer et al. (2025) argues that the rapid proliferation of machine learning research, combined with the fallibility of peer review, has allowed "misleading, incorrect, flawed or perhaps even fraudulent studies" to be accepted and, at times, highlighted at major ML conferences. The authors further contend that conferences "currently lack official processes for rectifying such errors once they are part of the scientific record," and that reforming peer review itself is "impractical due to reviewer incentives and institutional trust in the status quo". As a remedy, they propose establishing a dedicated *"Refutations and Critiques" Track* to provide "a high-profile, reputable platform to support vital research that critically challenges prior research," with the stated goal of fostering "a dynamic self-correcting research ecosystem". They conclude that only through such a mechanism can machine learning conferences "add professionalism and visibility to the process of refuting prior work" and prevent the persistence of flawed results.

This broader issue resonates deeply within statistics, a discipline that relies heavily on journal publications to establish new discoveries and is often perceived as more rigorous than conference-based research. This comment highlights the urgent need for formal mechanisms for refutation and critique in statistics, as the absence of such processes allows even strikingly obvious errors to persist and propagate. A prior note raising concerns about mistakes in Wang et al. (2025) is not an isolated case. To provide more compelling evidence, we now examine the mathematical correctness of another paper, Liu et al. (2023). The focus here is not on assessing methodological novelty, but on rigorously evaluating the validity of the underlying theory, with the broader aim of raising awareness among researchers about the necessity of taking stronger actions to enhance the reliability and credibility of statistical research.

1 Introduction

The work of Liu et al. (2023) develops a generalized factor model (GFM) designed to handle ultra-high-dimensional datasets containing mixed-type variables (continu-

*Independent Researcher. E-mail: statsbeginner@yahoo.com.

ous, binary, categorical, counts, etc.), moving beyond the limitations of traditional factor models that assume continuous, normally distributed data. To address computational challenges, the authors propose a two-step parallelizable algorithm, with an additional one-step correction to improve efficiency. They establish theoretical guarantees, including convergence rates for factor and loading estimators, asymptotic normality, and a penalized criterion for consistently estimating the number of factors when both sample size n and dimension p diverge.

This work is most closely related to the high-dimensional linear factor model literature, beginning with the seminal contribution of Bai and Ng (2002), while also drawing on the latent trait and categorical factor modeling tradition. It effectively merges these two streams by generalizing high-dimensional factor models to exponential family data with mixed types. The proposed methods are undoubtedly useful in certain application settings. We do not aim to, nor are we in the position to, evaluate the methodological novelty of the work; rather, our purpose is solely to examine the correctness of the mathematical proofs.

2 Proof of Lemma 1

The proof of Theorem 1 begins with Lemma 1, stated on page 8 of the Supplementary Material (SM). The proof of this lemma is **entirely incorrect**. For clarity, we first restate the lemma below.

Lemma 1 (Lemma 1 in Liu et al. (2023)). *Under Conditions (C1)–(C6), we have*

$$\sup_i \|\tilde{\mathbf{h}}_i - \mathbf{h}_{i0}\| = O_p(1), \qquad \sup_j \|\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}\| = O_p(1).$$

The existence of this auxiliary result is reasonable, since $(\tilde{\mathbf{h}}_i)_{i=1,\ldots,n}$ and $(\tilde{\boldsymbol{\gamma}}_j)_{j=1,\ldots,p}$ are local maxima of a non-convex objective function. To establish convergence rates, it is natural to first expect that these quantities are stochastically bounded. However, the mathematical arguments presented in the proof of this lemma deviate substantially from a valid line of reasoning.

The lemma is proved by contradiction in two steps. In the first step, suppose that neither $\sup_i \|\tilde{\mathbf{h}}_i - \mathbf{h}_{i0}\|$ nor $\sup_j \|\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}\|$ is $O_p(1)$. Then, with some positive probability, both the factor and loading estimates can diverge arbitrarily. More precisely, there exists $\epsilon > 0$ such that for any $M_1 > 0$, there exist integers $n_1, p_1 \geq 1$ for which

$$P\left(\sup_{1\leq i\leq n} \|\tilde{\boldsymbol{\kappa}}_i - \boldsymbol{\kappa}_{i0}\| > M_1\right) > \epsilon \quad \text{and} \quad P\left(\sup_{1\leq j\leq p} \|\tilde{\boldsymbol{\gamma}}_j - \boldsymbol{\gamma}_{j0}\| > M_1\right) > \epsilon \tag{2.1}$$

hold for all $n \geq n_1$ and $p \geq p_1$. (The authors' original statement of this condition is inaccurate, though the issue is minor.) From (2.1), the authors claim that there exists some pair (i, j) such that the inner product $|\tilde{\boldsymbol{\gamma}}_j^T \tilde{\boldsymbol{\kappa}}_i|$ can grow arbitrarily large. This reasoning is logically **wrong**. Even if both norms $\|\tilde{\boldsymbol{\kappa}}_i\|$ and $\|\tilde{\boldsymbol{\gamma}}_j\|$ are extremely large, their inner product may remain small if the two vectors are nearly orthogonal. The Cauchy-Schwarz inequality provides only the upper bound $|\tilde{\boldsymbol{\kappa}}_i^T \tilde{\boldsymbol{\gamma}}_j| \leq \|\tilde{\boldsymbol{\kappa}}_i\| \, \|\tilde{\boldsymbol{\gamma}}_j\|$, but offers no nontrivial lower bound. For instance, if the first $q/2$ (assuming q is even) components of $\tilde{\boldsymbol{\kappa}}_i$ are large and the remaining $q/2$ components are zero, while for $\tilde{\boldsymbol{\gamma}}_j$ the first $q/2$ components are zero and the last $q/2$ are large, then their inner product is identically zero regardless of the magnitudes of the individual norms.

Another issue arises when considering the gradient

$$\frac{\partial l(\boldsymbol{\theta}, \boldsymbol{X})}{\partial \boldsymbol{\gamma}_j} = \sum_{i=1}^{n} \frac{x_{ij} - m_j(\boldsymbol{\gamma}_j^T \boldsymbol{\kappa}_i)}{c_j} \boldsymbol{\kappa}_i.$$

Even if $|\tilde{\boldsymbol{\kappa}}_i^T \tilde{\boldsymbol{\gamma}}_j|$ were to become arbitrarily large, it is not clear how this would imply

$$P\left(\sum_{i=1}^{n} \frac{x_{ij} - m_j(\tilde{\boldsymbol{\gamma}}_j^T \tilde{\boldsymbol{\kappa}}_i)}{c_j} \tilde{\boldsymbol{\kappa}}_i = \frac{\partial l(\tilde{\boldsymbol{\theta}}, \boldsymbol{X})}{\partial \tilde{\boldsymbol{\gamma}}_j} \neq 0\right) \geq \epsilon.$$

Even in the linear case with $m(x) = x$, the fact that $|\tilde{\boldsymbol{\gamma}}_j^T \tilde{\boldsymbol{\kappa}}_i|$ is large for some pair (i, j) does not by itself guarantee that

$$\sum_{i=1}^{n} \frac{x_{ij} - \tilde{\boldsymbol{\gamma}}_j^T \tilde{\boldsymbol{\kappa}}_i}{c_j} \tilde{\boldsymbol{\kappa}}_i \neq 0?$$

In short, Step 1 rests on a series of flawed inferences. Step 2 relies on the same arguments as Step 1 and is therefore subject to the same flaws.

The proof of Lemma 1 does not withstand mathematical scrutiny at any level, yet its conclusion serves as a crucial step in establishing the remainder of the theory, particularly the rate of convergence. That said, deriving such a result is inherently difficult and represents a central challenge when dealing with estimators arising from non-convex optimization problems.

2.1 Proofs of Propositions 1 and 2

Proposition 1 asserts the identifiability of model parameters under Conditions (A1)–(A3). Let $(\mathbf{H}_1, \mathbf{B}_1, \boldsymbol{\mu}_1)$ and $(\mathbf{H}_2, \mathbf{B}_2, \boldsymbol{\mu}_2)$ be two sets of parameters satisfying model (1). To prove that $\mathbf{H}_1 = \mathbf{H}_2$, the authors define

$$\mathbf{H}_3 = \mathbf{H}_2^T \mathbf{H}_1 / n,$$

and claim that $\mathbf{H}_3$ is orthogonal according to Condition (A1). However, Condition (A1) only states that

$$\frac{1}{n} \mathbf{H}_1^T \mathbf{H}_1 = \frac{1}{n} \mathbf{H}_2^T \mathbf{H}_2 = \mathbf{I}_q.$$

By definition,

$$\mathbf{H}_3^T \mathbf{H}_3 = \frac{1}{n^2} \mathbf{H}_1^T \mathbf{H}_2 \mathbf{H}_2^T \mathbf{H}_1,$$

where $\mathbf{H}_2 \mathbf{H}_2^T$ is an $n \times n$ matrix. It is therefore clear that Condition (A1) alone is insufficient to conclude that $\mathbf{H}_3^T \mathbf{H}_3 = \mathbf{I}_q$. Why this matters: $\mathbf{H}_3^T \mathbf{H}_3 = \mathbf{I}_q$ would hold only if col$(\mathbf{H}_1)$ =col$(\mathbf{H}_2)$, which is precisely the fact the proof is attempting to establish. Thus, this conclusion cannot be reached at this stage of the argument.

The step concluding that $\mathbf{H}_3$ is a sign matrix relies on the uniqueness of the diagonalization (or eigendecomposition). However, if the diagonal elements of $\mathbf{B}_\ell^T\mathbf{B}_\ell$ ($\ell = 1, 2$) are merely ordered in decreasing fashion, thus allowing ties, rotations within tied blocks remain possible. In such cases, $\mathbf{H}_3$ need not be a signed permutation matrix. The proof, as written, therefore implicitly requires a strict decrease (or an additional tie-breaking condition) to guarantee uniqueness up to signs.

To prove Proposition 2, the closedness condition (S4) is assumed in general but verified only in the Gaussian special case; the more general mixed-type case is neither proved nor discussed.

3 Additional Remarks on the Remaining Proofs

Lemma 1 sets the foundation for deriving convergence rates, as it establishes that the preliminary estimates–obtained as local maxima of a non-convex objective function–are stochastically bounded. While one might expect the subsequent proofs to become more straightforward once this step is secured, they also contain errors.

Lemma 2 is intended to establish the **uniform** convergence of certain random quantities that arise in the proofs. Specifically, it either provides stochastic bounds uniformly over $i = 1, \ldots, n$, or it bounds the supremum of such quantities across all i. In the proof of Lemma 2, however, many arguments establish convergence only for a fixed i, rather than uniformly over the full index set.

For example, in the proof of (19), showing that $P(\zeta_{i4} > M_0) \le \epsilon$ for any given i does not automatically imply that $\zeta_{i4} = O_p(1)$ uniformly over $i = 1, \ldots, n$. What is required instead is a bound on $P(\sup_i \zeta_{i4} > M_0)$. For the same reason, analogous issues also arise in the proofs of (20) and (25). Nonetheless, these gaps could likely be remedied by adapting the argument used in the proof of (38), perhaps with the addition of stronger or more explicit conditions.

In the middle of page 11 of the SM, immediately below inequality (24), the goal is to bound the probability

$$\sup_i \left\{ \frac{1}{p} \sum_{j=1}^{p} \|\mathbf{G}_{ji*} - \mathbf{G}_{ji0}\|^2 \cdot \left\| \frac{1}{n^{1/2}} \sum_{\ell \ne i} \mathbf{w}_{j\ell 0} \right\|^2 \right\} > \delta.$$

Under the condition $\sup_{i,j} \|\mathbf{G}_{ji*} - \mathbf{G}_{ji0}\|^2 \le M_1$, the left-hand side is bounded by

$$M_1 \sup_i \left\{ \frac{1}{p} \sum_{j=1}^{p} \left\| \frac{1}{n^{1/2}} \sum_{\ell \ne i} \mathbf{w}_{j\ell 0} \right\|^2 \right\}.$$

In other words, the supremum over i ($\sup_i$) remains because the summation term $\sum_{\ell \ne i}$ also depends on i. Consequently, by Markov's inequality, the correct argument should be

$$P\left[\sup_i \left\{ \frac{1}{p} \sum_{j=1}^{p} \left\| \frac{1}{n^{1/2}} \sum_{\ell \ne i} \mathbf{w}_{j\ell 0} \right\|^2 \right\} > \frac{\delta}{M_1} \right] \le \frac{M_1}{\delta} \mathrm{E} \sup_i \left\{ \frac{1}{p} \sum_{j=1}^{p} \left\| \frac{1}{n^{1/2}} \sum_{\ell \ne i} \mathbf{w}_{j\ell 0} \right\|^2 \right\}.$$

The proofs of Lemma 3, Theorem 1, and Theorem 2 all rely on the conclusions of Lemmas 1 and 2. At this point, we stopped our examination and offer no further comments.

4 Summary

The two papers discussed in this comment and in Li (2025) are by no means the only instances of theoretically flawed work to appear in prestigious statistics journals. It is, of course, beyond the scope of this comment to document every such case. Nonetheless, judging from discussions on social media, there seems to be a widespread sense of fatigue, or even indifference, toward whether the mathematical details of a statistics paper truly matter. The present author is not in a position to determine whether such an attitude will ultimately prove beneficial or detrimental to the field.

If one presumes that theoretical rigor and correctness still matter, the remarks on these two papers are meant to draw attention to the issue. In this case, it may be too strong to claim that the authors deliberately engaged in opportunistic proof-making. However, if much of the proof-writing and proofreading is left to students, one must ask: what, then, is the responsibility of the advisor or corresponding author? Should their role not extend beyond methodological guidance, conceptual framing, and editorial oversight? Regardless of the justification, whether the complexity of the models or the sophistication of the algorithms, introducing an incorrectly proved lemma as the first theoretical contribution in the Supplementary Material is unacceptable.

References

Jushan Bai and Serena Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.

Zhen Li. Comment on "Deep Regression Learning with Optimal Loss Function", 2025. URL https://arxiv.org/abs/2509.03702.

Wei Liu, Huazhen Lin, Shurong Zheng, and Jin Liu. Generalized factor model for ultra-high dimensional correlated variables with mixed types. *Journal of the American Statistical Association*, 118(542):1385–1401, 2023.

Rylan Schaeffer, Joshua Kazdan, Yegor Denisov-Blanch, Brando Miranda, Matthias Gerstgrasser, Susan Zhang, Andreas Haupt, Isha Gupta, Elyas Obbad, Jesse Dodge, Jessica Zosa Forde, Francesco Orabona, Sanmi Koyejo, and David Donoho. Position: Machine learning conferences should establish a "Refutations and Critiques" Track, 2025. URL https://arxiv.org/abs/2506.19882.

Xuancheng Wang, Ling Zhou, and Huazhen Lin. Deep regression learning with optimal loss function. *Journal of the American Statistical Association*, 120(550): 1305–1317, 2025.