Promoting Data Science

Volume 4 2024 Number 2

Journal of Behavioral Data Science V4N2 (2024)

https://isdsa.org

JOURNAL OF BEHAVIORAL DATA SCIENCE

ISBN: 2575-8306 (Print) 2574-1284 (Online)

https://jbds.isdsa.org

JOURNAL OF BEHAVIORAL DATA SCIENCE

No Publication Charge and Open Access

jbds@isdsa.org

List of Articles

Journal of Behavioral Data Science, 2024, *4* (2), 1–28.
DOI: https://doi.org/10.35566/jbds/yuan

Modeling Data with Measurement Errors but without Predefined Metrics: Fact versus Fallacy

Ke-Hai Yuan and Zhiyong Zhang

University of Notre Dame
kyuan@nd.edu

Abstract. Data in social and behavioral sciences typically contain measurement errors and also do not have predefined metrics. Structural equation modeling (SEM) is commonly used to analyze such data. This article discuss issues in latent variable modeling as compared to regression analysis with composite-scores. Via logical reasoning and analytical results as well as the analyses of two real datasets, several misconceptions related to bias and accuracy of parameter estimates, standardization of variables, and result interpretation are clarified. The results are expected to facilitate better understanding of the strength and limitations of SEM and regression analysis with weighted composites, and to advance social and behavioral data science.

Keywords: Measurement error · Attenuation · Standardization · Scales of latent variables

1 Introduction

Two key features of data in social and behavioral sciences are measurement errors and no predefined metrics. Associated with the features are latent variables whose scales need to be subjectively chosen. These features pose challenges to data analysis and result interpretation. A conventional method to address the issue of measurement errors is structural equation modeling (SEM), while standardized solution is used to address the issue of lack of metrics. In particular, textbooks contain formulas showing that the least-squares (LS) method yields attenuated or biased regression coefficients when predictors contain measurement errors, and SEM effectively addresses the issue. Textbooks on regression analysis and SEM also contain formulas for computing the regression coefficients with standardized variables (e.g., Bollen, 1989; Cohen, Cohen, West, & Aiken, 2003; Loehlin & Beaujean, 2017), which are available in the output of commonly used software and routinely reported in papers. Because the notions that *measurement errors cause biased estimates* and *standardized solutions facilitate result interpretation* have been imprinted and routinely taught in the discipline, a rigorous examination on their validity not only facilitates better understanding of these concepts but also advances behavioral data science.

The purpose of this article is to bring attention of both quantitative and applied researchers to these potential issues, and for proper and better applications of multivariate methods. In particular, we aim to answer the following questions.

Q1. In social and behavioral sciences, measurements and latent variables typically do not have predefined scales. We need to fix the scales of latent variables subjectively for the parameters of a model to be identified. Under SEM, what is the effect of different scaling options on the accuracy of parameter estimates and the related z-statistics? How can we use the information to serve our purpose?

Q2. Do measurement errors cause attenuated or biased estimates for the LS method of regression analysis with weighted composites? A weighted composite in this article is a linear combination of values of indicators for a latent variable.

Q3. Does SEM yield more accurate parameter estimates than LS regression with weighted composites?

Q4. Between SEM and regression analysis with weighted composites, to what level the estimated regression coefficients can be compared with their SEM counterparts?

Q5. Does standardization advance result interpretation or only facilitate model identification?

We will answer the above questions by combining recent findings from the literature, logical reasoning and analytical results, and fresh numerical results via the analyses of two real datasets. As we are going to show, results do not necessarily support the widely held notions regarding attenuation, bias, accuracy and efficiency of parameter estimates for SEM and regression analysis with composites that contain measurement errors.

Most existing studies comparing SEM and regression analysis with composites are conducted by comparing the values of parameter estimates and their standard errors (SEs). For logical and proper analyses of data that do not have predefined metrics, we propose a new approach under which methods are compared by the sizes of the signal-to-noise ratio (SNR) of their estimates. For a parameter estimate $\hat{\gamma}$ based on a sample of size N, the SNR is defined as

$$\tau = \frac{\gamma}{\mathrm{SD}},$$

where γ and SD are respectively the expected values of $\hat{\gamma}$ and $[\mathrm{Var}(\sqrt{N}\hat{\gamma})]^{1/2}$ or their probability limits as N increases. The new approach is a natural product of our effort in answering the above 5 questions.

The rest of the article is arranged as the following. First, we review the literature for related work and to clarify our contributions. Second, we describe our view on SEM and regression analysis with weighted composites, including key elements to answer the posed questions. Third, we provide a logical analysis on the utility of standardization. Fourth, two empirical examples are provided and patterns over the results are summarized. Fifth, our answers to the above five questions are subsequently presented by combing the results from our analysis and the results of the examples. Sixth, summary, discussion and take-home messages are provided to conclude the article.

2 Review of the Literature and Clarification of Contributions

There are studies in the psychometric literature that might be regarded as related to the development of the current article. We will review them below to clarify the differences between the existing studies and the topics we are going to cover. Some of our results as well as the framework under which our study is conducted will also be previewed in this section.

2.1 Parameters and z-statistics are scale dependent

It is well known that, in SEM, the scale of a latent variable can be set by 1) fixing one of the loadings of its indicators at a given value, typically 1.0; or 2) for an independent latent variable, fixing its variance at a given value. The choices among the indicators as well as between 1) and 2) are equivalent in the sense that the resulting model implied covariance matrix remains the same. But they can yield quite different parameter estimates. Gonzalez and Griffin (2001) noted that different ways to scaling latent variables can also result in different z-statistics. This implies that the results of null hypothesis testing by the Wald test (or z-test for a single parameter) depend on how the scales of the latent variables are fixed. Subsequently, Gonzalez and Griffin recommended using the likelihood ratio statistic (T_{ml}) or its difference for parameter inference, because T_{ml} remains the same across different scalings of latent variables. However, one has to run a separate model to conduct the likelihood ratio test for each single parameter; whereas the z-statistics for all the parameters are in the default output of standard software following a single run of the base model. This might be why the z-test is widely used in practice. In addition, the validity of T_{ml} as a χ^2 statistic depends on the normal distribution assumption[1] even asymptotically. In contrast, SEs and the corresponding z-statistics based on the sandwich-type covariance matrix are asymptotically valid without the need for the normality assumption.

The sensitivity of z-statistics to the scales of latent variables reflects the dependency of statistical power of the Wald test on model parameterization. Instead of treating this sensitivity as an undesired feature, we should make use of it to serve the purpose of data analysis. In particular, if a test with a greater

[1] While there exist conditions for T_{ml} to follow a chi-square distribution when data are not normally distributed, there is not an effective way to verify the so-called asymptotic robustness conditions in practice.

power is desired, we can choose scales that correspond to the greatest z-statistics. Let's call an indicator whose loading is fixed at 1.0 an *anchor*. An analytical result in Yuan and Fang (2023b) implies that the SNR for the path coefficient between two latent variables increases as the anchor of the dependent latent variable becomes more reliable, where they assumed that all the indicators for the independent latent variable are parallel. A better understanding of the relationship between the z-statistics and the properties of the anchors is needed for the general case. For such a purpose, we will further study the following two characteristics of the z-statistics: 1) What properties of the anchors affect the value of z-statistics? and 2) Are the z-statistics for all path coefficients of the structural model equally affected by the changes of scales of the latent variables? These characteristics are not examined in Gonzalez and Griffin (2001).

2.2 Model identification versus theoretical assumption

Steiger (2002) discussed scenarios for fixing the scale of a latent variable using equality constraints, and fixing a factor loading at 1.0 is regarded as a particular constraint. He emphasized that additional constraints beyond the minimal need to fix the scales of latent variables will affect the value of T_{ml}. That is, the model implied covariance matrix will vary when different extra constraints are implemented. The same message has also been given by others (e.g., Bentler, 2006). In this article, we are not interested in the effect of extra constraints beyond the minimal need for scaling latent variables. Instead, for the scaling issue, we examine how the values of the SNRs and z-statistics are affected by the psychometric properties of the indicators used to fix the scales of latent variables. The value of T_{ml} remains the same among these choices. Steiger (2002) also discussed statistical issues due to interactions of extra constraints in standardized solutions. We will also discuss standardization but our interest is on issues related to substantive and statistical interpretations instead of issues caused by interactions of extra constraints.

2.3 Accuracy and precision of parameter estimates

For a mediation model with three latent variables, Ledgerwood and Shrout (2011) compared bias and SEs of parameter estimates between SEM and regression analysis via average scores. They used "accuracy" and "precision" to substitute for the statistical terminology bias and SEs, and showed that SEM yields estimates with greater accuracy but less precision. While Ledgerwood and Shrout (2011) contain several interesting observations, they missed two key points. The first is that values of parameters under SEM depend on the scales of the latent variables and those under regression analysis depend on the scales of the composites (Yuan & Deng, 2021). Thus, accuracy or bias is not a substantively grounded concept for statistical modeling of variables whose metrics are artificially assigned. The second is that SEs are typically proportional to the values of the parameter estimates, precision is also not a meaningful quantity to compare between SEM and regression analysis with composite scores. Consequently, the conclusions of Ledgerwood and Shrout (2011) for the comparison between SEM and regression analysis with composites are problematic. In addition, the use of parallel indicators in their Monte Carlo studies also made their results of regression analysis with the average scores too optimistic, since the average score enjoys the maximum reliability (Bentler, 1968; Yuan & Bentler, 2002). Existing results indicate that, following regression analysis with composites, the estimates of the regression coefficients, their SEs, and the resulting R-square are all related to the reliabilities of the composites (Cochran, 1970; Fuller, 2009).

Instead of comparing different methods by the accuracy or precision of their parameter estimates, we compare methods via their SNRs. We will argue that the SNR is a natural quantity to compare for modeling variables without predefined metrics.

2.4 Factor score vs average score

McNeish and Wolf (2020) discussed rationales in forming composites and suggested treating sum scores as factor scores based on a factor model with parallel measurements. They also recommended factor-scoring items according to the factor model under which the scales are validated instead of using the sum scores by default. A followup discussion by Widaman and Revelle (2022) gave a different perspective on the merit of sum scores. They compared parameter estimates by different scoring methods and noticed little difference. In the current article, we are interested in comparing SEM and regression analysis with weighted composites, and regard both sum scores and factor scores as special cases of weighted composites. In particular, results on SNRs for the estimated path coefficients indicate that larger differences exist between SEM and regression analyses with weighted composites than among regression analyses with differently formulated composites (Yuan & Fang, 2023b).

2.5 Standardized score versus raw score

Variable standardization and treating the standardized coefficients as effect-size measures are common practices in social and behavioral sciences. Their pros and cons have been discussed under different contexts. Aiming to set out guidelines for what to report and how best to report effect sizes, Baguley (2009) listed advantages of simple (unstandardized) effect sizes but the measures need to have metrics that are well understood or substantiated if not predefined. Pek and Flora (2018) provided an informed discussion on why unstandardized effect sizes tend to be more informative than standardized ones in primary research studies. While their focus is on effect sizes with manifest variables, they stated (p. 214) "We agree that standardization of effects associated with latent variables (e.g., factors in a factor analysis) is useful, but assert that observed variables, and consequently effect sizes based on them, should not always be standardized."

Kim and Ferree (1981) distinguished the operation of standardization of scales from the use of standardized coefficients. If the groups under consideration do not have comparable distributions, their discussion discourages standardizing variables on the basis of group-specific means and variances. Olejnik and Algina (2000) showed that measures of effect size are affected by the research design used, and warned that effect sizes may not be comparable across different designs when different random components (e.g., individual difference factors) are included in computing the pooled variances for standardizing the effect sizes. They also reviewed various factors that may contribute to the misinterpretation/understanding of effect size.

McGrath and Meyer (2006) discussed the differences of Cohen's d and the point-biserial correlation coefficient (r_{pb}). Both of which can be used when one variable is dichotomous and the other is quantitative. Termed the proportions of 0 and 1 for the dichotomous variable as the base rates, they showed that r_{pb} is a base-rate-sensitive effect-size measure, whereas d is base-rate-insensitive. Standardization is also widely used in epidemiology when estimating and comparing group means, where it is a different operation than z-scoring the variables. Still the group distributions matter in standardizing the means, as was showed by (Schoenbach & Rosamond, 2000, Chapter 6).

While the pros and cons of standardization have been extensively addressed, none of the articles discuss the issue of standardization in SEM. Part of our interest in this article is to examine the aspects of the usefulness of standardizing latent variables. In particular, our focus is on variables that do not have predefined metrics.

2.6 Properties and results of factor scores

Since factor scores will be repeatedly mentioned in our discussion, we briefly review their properties here. First, because latent variables are not observable, there exists an issue of indeterminacy with their scales and orientation. However, parameters of a factor or SEM model can be uniquely estimated once the scales of all latent variables are fixed and the model is identified. Then both the Bartlett-factor scores (BFSs) and the regression-factor scores (RFSs) based on the parameter estimates can be uniquely computed (Lawley & Maxwell, 1971). Unless explicitly mentioned, factor scores in this article always refer to either the BFSs or the RFSs.

It is well-known that the BFS possesses the maximal reliability among all weighted composites (see e.g., Yuan & Bentler, 2002). Yuan and Deng (2021) showed that the RFSs are proportional to the BFSs. That is, one can get the values of the BFSs from those of the RFSs via a linear transformation, conditional on the estimated factor loadings, factor covariances, and error variances. Thus, the RFSs also possess the maximum reliability. In addition, Yuan and Deng (2021) noted that the two types of factor scores are also equivalent in conducting regression analysis in the sense that they yield the same R-square value. Note that the RFSs can be computed jointly for all the factors or separately for each single factor. Yuan and Deng (2021) also noted that when the RFSs are computed separately, the regression coefficients following RFS-regression are proportional to those following BFS-regression. When the RFSs are computed jointly, the two sets of regression coefficients can still be computed from each others but using a linear transformation.

Skrondal and Laake (2001) noted an important property of factor-score (FS) regression. That is, regression analysis with BFSs as the outcome variables and the jointly computed RFSs as the predictors yields path coefficients that are mathematically identical to those under SEM. This property was also discussed in Croon (2014) and Devlieger, Mayer, and Rosseel (2016), and will be noted in our following discussion.

2.7 Monte Carlo studies comparing parameter estimates

There are Monte Carlo studies comparing the empirical bias, SEs or mean-squared errors (MSE) of parameter estimates across methods (e.g., Forero & Maydeu-Olivares, 2009). For a model whose population values of parameters are held constant, the method that yields the smallest bias or SE or MSE is preferred under the conditions being considered. The finding is still statistically meaningful even when the variables do not have predefined metrics, as in typical simulation studies (e.g., Shi & Tong, 2017; Zhang & Yang, 2020). However, because bias, SE and MSE are scale dependent, additional thoughts are needed to compare the estimates under regression analysis against those under SEM, especially when variables do not have predefined metrics.

Note that bias is defined as the expected value of an estimator minus its population value. The regression coefficients under regression analysis with composites naturally don't have the same population values as their SEM counterparts. But for data that do not have predefined metrics, we can make the two sets of population values identical by properly choosing the scales for the composites or the latent variables. We will formally discuss how to compare estimates across the two classes of methods in the following sections.

3 Structural Equation Modeling versus Regression Analysis with Composites

In this section we present the elements for logically comparing SEM against regression analysis with weighted composites.

3.1 Theoretical constructs by latent variables versus by composites

While latent variables and composites are conceptually different, we regard both of them as representatives of the theoretical constructs. But their degrees of alignment are subject to judgment. It is commonly believed that theoretical constructs are virtually modeled under SEM while weighted composites always contain measurement errors. However, the goodness of fit[2] of an SEM model is unlikely to be perfect, implying that the Greek letters in a path diagram only approximately represent the theoretical constructs. Such a discrepancy systematically changes the values of the parameters from those of an ideal model-population match (Yuan, Marshall, & Bentler, 2003), causing biased parameter estimates and biased interpretation. Similarly, the reliabilities of the composites also vary[3] as the number of indicators vary, implying that the degree of alignment between the theoretical constructs and the weighted composites also varies. However, the bias due to model misspecification under SEM has been often ignored whereas "bias" caused by measurement errors has been repeatedly warned in textbooks (e.g., Allen & Yen, 1979).

In the development of this article, we do not explicitly consider the approximating nature of latent variable models. But we admit that there always exist differences between theoretical constructs and the latent variables in practice. Although the setup implicitly favors latent variable models, as was typically done in the field, we will discuss the rationale and provide the evidence that regression analysis with weighted composites yields different but more efficient parameter estimates than SEM instead of biased estimates.

3.2 Consistency

Measurements in social and behavioral sciences typically do not have pre-defined metrics in the first place. For such data, Yuan and Deng (2021) and Yuan and Fang (2023b) pointed out that the sizes of parameters under SEM do not enjoy a substantive interpretation since they are determined by the scales of latent variables that are subjectively assigned. Two researchers modeling the same dataset via the same model estimated by the same method (e.g., normal-distributed-based ML) can have very different parameter estimates for the same path coefficient. For this path coefficient, it is impossible for a third researcher who conducts regression analysis with composites to obtain an estimate that is consistent with those under SEM before the two SEM modelers have their difference dissolved. Thus, it does not make sense to claim that LS regression analysis with weighted composites yields biased estimates without a unique set of target population values even under SEM.

Let us consider a simple case to understand the details. With a measurement model

$$y = \lambda_y \eta + e_y, \quad x = \lambda_x \xi + e_x,$$

[2] Even if a test statistic for the overall model structure is statistically not significant, we are still unable to confirm that the model is correctly specified, since a non-significant statistic only implies that there is not enough power to reject the current model.

[3] The content of the true score of a composite might also vary as more indicators are added even when a single factor model fits the corresponding variables adequately (Bentler, 2017).

one can estimate the regression relationship between the two latent variables via the structural model

$$\eta = \gamma_0 + \gamma_1 \xi + \zeta,$$

where ξ and η are the latent variables, and e_x, e_y and ζ are the error terms. Alternatively, one can also directly work with the regression model for the observed variables

$$y = a + bx + e.$$

Under the commonly used assumptions about the independence of the error terms for SEM and using σ to denote the population covariance of the variables in its subscript, standard covariance algebra yields

$$b = \frac{\sigma_{xy}}{\sigma_{xx}} = \frac{\lambda_x \lambda_y \sigma_{\xi\eta}}{\lambda_x^2 \sigma_{\xi\xi} + \sigma_{e_x e_x}} = \frac{\rho_x \lambda_x \lambda_y \sigma_{\xi\eta}}{\lambda_x^2 \sigma_{\xi\xi}} = \frac{\rho_x \lambda_y \gamma_1}{\lambda_x}, \tag{1}$$

where ρ_x is the reliability of x. Equation (1) implies that, regardless of the value of ρ_x, $b = \gamma_1$ whenever $\lambda_x = \rho_x \lambda_y$ holds. We can also adjust the values of λ_x and λ_y by rescaling ξ and η to make the value of b greater than that of γ_1. Alternatively, given the population values of ρ_x, σ_{xy} and σ_{xx}, the value of γ_1 can be made equal to any pre-specified value (except 0) by adjusting the value of λ_x and λ_y. We will further illustrate this via a real data example in a following section.

3.3 Parameter comparison

A common theme in Ledgerwood and Shrout (2011), McNeish and Wolf (2020), Widaman and Revelle (2022), and others is the comparison of the raw values of parameter estimates by different methods. In this article we emphasize that the population values of parameters as well as their estimates are not of substantive interest for models involving latent variables measured by indicators that do not have predefined metrics. In particular, population values of the model parameters depend on scales that are artificially assigned.

Although there is no point to directly compare the values of parameters under regression analysis with weighted composites against those under SEM, they have the following relationship:

1) $\boldsymbol{\gamma} = \mathbf{0}$ if and only if $\boldsymbol{\gamma}_w = \mathbf{0}$, where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_p)'$ and $\boldsymbol{\gamma}_w = (\gamma_{w1}, \gamma_{w2}, \ldots, \gamma_{wp})'$ are the vectors of regression coefficients of a given dependent latent variable under SEM and regression analysis with weighted composites, respectively (Buonaccorsi, 2010, Eq. 5.7 on page 109).
2) When the (joint) regression-factor scores are used as the predictors and the Bartlett-factor score is used as the dependent variable in regression analysis with weighted composites, there also exist $\boldsymbol{\gamma}_w = \boldsymbol{\gamma}$ and $\hat{\boldsymbol{\gamma}}_w = \hat{\boldsymbol{\gamma}}$ (Skrondal & Laake, 2001). Then the two types of estimates can be substituted for each other, although they are obtained by different methods.

3) Regression analysis with equally weighted composites can also yield parameter estimates that are mathematically identical to those under SEM according to

$$\begin{aligned} y &= a + b_1 x_1 + \ldots + b_p x_p + e \\ &= a + (h_1 b_1)(x_1/h_1) + \ldots + (h_p b_p)(x_p/h_p) + e, \end{aligned} \tag{2}$$

where the y and x_js are equally weighted composites and the h_js are subject to choice (e.g., $h_j = \gamma_j/b_j$).

However, for arbitrarily chosen scales, b_j and γ_j may not be equal. They even can have different signs.

Instead of judging the goodness of an estimate by its accuracy or precision, we propose to compare the efficiency of parameter estimates by the size of their SNR[4], which plays a key role in statistical inference. Let's term this proposal the new framework in contrast to the old framework that compares methods by precision and accuracy of parameter estimates, the following remarks are in order.

4) The new framework allows us to face and address the issue of dependency for the values of parameters or their estimates on the scales of latent variables, especially for data and variables that have no predefined metrics.
5) The new framework naturally facilitates the comparison of efficiencies of estimates for parameters that do not have the same population value. For example, the population value of a path coefficient under SEM depends on how the scales of the involved latent variables are fixed. However, the SNRs for the estimates of this path coefficient can be compared across different scaling options, and the one with the largest SNR is the most efficient estimate.
6) When the population value of a parameter is uniquely defined across methods, as in a Monte Carlo study where the scales of latent variables are identically scaled across methods, an estimate with a smaller SE is more efficient and corresponds to a greater SNR than that with a greater SE. We will further discuss the issue of empirical bias in Monte Carlo studies in the concluding section.
7) Parallel to Cohen's d, the SNR also serves as a summary statistic.

In addition, strengths of SEM and regression analysis with composites can be fairly compared and utilized under the new framework. We will have more results on this point via examples in a following section.

For a model with one dependent latent variable and one independent latent variable, Yuan and Fang (2023b) rigorously compared the SNRs of the estimated path coefficients under regression analyses with weighted composites against those under SEM. They found that, conditional on the population weights, the SNR under factor score (FS) regression is mathematically greater than that under SEM. They also defined a multivariate version of SNR and conjectured

[4] For a parameter estimate, the SNR is estimated by $z/N^{1/2}$, where z is the z-statistic and N is the sample size.

that its values for the path coefficients under FS regression would be greater than that under SEM. Note that the SNR plays the role of Cohen's d in null hypothesis testing for parameters. Meta analytical results in Deng and Yuan (2023) showed that, across nine different real datasets and eleven models, SEM yields the least powerful test, even weaker than path analysis with equally weighted composites.

3.4 Different utilities

SEM and regression analysis with weighted composites are different not only in their approaches to modeling the theoretical constructs but also in aims and utilities. For SEM, the relationship among the latent variables is modeled. The corresponding parameters and their estimates are to govern the relationship among the latent variables at the *population* level. In practice, an *individual* with greater pretest scores is expected to perform better on the post-tests. This expected relationship is of interest in many disciplines. We may want to plug the path coefficients estimated under SEM in the regression equation to predict the values of individuals corresponding to the latent variable. For such a purpose, we will have to substitute the independent latent variables by composites of the individuals. However, except in rare situations, values of composites are not error-free, including the factor scores that are psychometrically most reliable.

Alternatively, we can start with regression analysis via weighted composites, and use the estimates of the path coefficients to construct an equation for prediction. The new outcome variable is then predicted according to this equation using the newly observed scores of the independent variables via weighted composites. Fuller (2009) noted that, even when the independent variables contain measurement errors, LS estimates of the regression model still yield the best linear unbiased predictor in the sense that the corresponding MSE is the smallest. Yuan and Fang (2023a) contain the details showing that the predicted value based on the SEM estimates becomes less accurate as the reliabilities of the weighted composites decrease.

Thus, one should start by regression analysis with weighted composites if the purpose is for prediction. However, SEM is preferred if the aim is to describe the relationship among the latent variables at the population level. These different characterizations might be more fundamental than which method generates more accurate parameter estimates.

The analyses and discussions in this section contain our answers to questions Q2, Q3 and Q4.

4 Standardization and Bias-correction

The notion that standardized solutions facilitate result interpretation has been rooted in psychometrics, especially with latent variable modeling and regression analysis. In this section, we conduct a logical analysis on the utility of standardization. We will also discuss the utility of bias correction. For such a purpose, we distinguish measurements that have predefined metrics from those that have no predefined metrics.

To facilitate understanding of the issue, let's consider the relationship between height ξ (inches) and weight η (pounds), and they are assumed to follow the relationship

$$\eta = \gamma_0 + \gamma_1 \xi + \zeta, \tag{3}$$

where ζ is the error term in predicting weight by height. In practice, we only observe x (inches) and y (pounds) due to the deficiencies in technology. A reasonable and also logical measurement model[5] in this case is

$$x = \xi + \delta \quad \text{and} \quad y = \eta + \varepsilon, \tag{4}$$

where δ and ε are measurement errors, and they are statistically independent with the latent variables ξ and η. With a sample (x_i, y_i) of size N, if we estimate the model

$$y_i = \gamma_{*0} + \gamma_{*1} x_i + e_i \tag{5}$$

by the LS method, then the LS estimate $\hat{\gamma}_{*1}$ is expected to be smaller than the γ_1 of Equation (3). In the measurement error literature (e.g., Fuller, 2009), emphasis was placed on getting a consistent estimate of γ_1 by correcting $\hat{\gamma}_{*1}$. Let the corrected estimate be denoted by $\tilde{\gamma}_1$. Then the value of $\tilde{\gamma}_1$ provides us the information that, with one inch increase in height, a person is expected to increase by $\tilde{\gamma}_1$ pounds.

Let's standardize the variables in Equations (3) and (4), resulting in

$$\eta_s = \gamma_{s1} \xi_s + \zeta_s, \quad x_s = \lambda_x \xi_s + \delta_s, \quad y_s = \lambda_y \eta_s + \varepsilon_s. \tag{6}$$

By the first equation in Equation (6), we would conclude that an individual with an increase of one SD (inches) in height is expected to increase by γ_{s1} SD (pounds) in weight. Such standardized scales might not prevent us from understanding the relationship between height and weight if we are familiar with the two SD units. However, if the SDs are as inexplicable as the values of ξ or η, then standardization does not facilitate interpretation but only facilitates identifying a set of unique values of the γ and λ.

In the case of height and weight, the use of standardized scales certainly hinders our understanding of their relationship. When there are no established metrics for x and y, the values of their SDs are at least as inexplicable as the values of x and y themselves. Standardization may even block possible attempts to think about the issue because it is hard to get a sense out of the SDs even in the case of height and weight, and each SD depends on the distribution shape as well as the range of the variable.

For height and weight, the bias-corrected estimate $\tilde{\gamma}_1$ does provide a more accurate quantification for the relationship between the two variables. When the

[5] When repeated measurements on height and weight for each individual are available, a typical practice is to substitute the x and y in Equation (4) by their respective averages.

scales of x or y are arbitrary, as for typical variables in social and behavioral sciences where data are obtained via Likert items or the averages/sums of such items, parameters under standardized scales do not advance the understanding of the relationship of the involved variables. Bias-corrected estimates may not help with better substantive interpretation either.

Standardized regression coefficients across groups might be comparable if the ranges and distribution shapes of the groups are similar. Otherwise, equal standardized coefficients may still imply different relationships between the outcome variable and the predictors in separate groups.

5 Real-data Examples

In this section we use two real data examples to illustrate some of the points noted in the previous sections. Because the value of a z-statistic is simply the value of the SNR multiplied by the square root of the sample size[6], comparison of the z-statistics between different methods will be directly followed from our comparison of the SNRs.

5.1 Example 1

Data Mardia, Kent, and Bibby (1979; Table 1.2.1) contain test scores on 5 topics from $N = 88$ students. The five topics are: C_1=Mechanics, C_2=Vectors, O_1=Algebra, O_2=Analysis, and O_3=Statistics. The scores for the first two topics were obtained with closed-book exams and for the last three were with open-book exams. Tanaka, Watadani, and Ho Moon (1991) fitted the dataset by a two-factor model, one factor represents the trait for taking closed-book tests, and the other for taking open-book tests. This dataset has been used to illustrate new developments in SEM and other multivariate methods (e.g., Cadigan, 1995). We will use it to show how different methods perform in estimating the regression parameter between the two constructs. Because this dataset is open to public, we expect readers to easily replicate our results.

The path diagram for the two-factor model is given in Figure 1, where ξ_o and ξ_c represent the latent traits for taking the open- and closed-book tests, respectively. Let $\phi_o = \text{Var}(\xi_o)$ and $\phi_c = \text{Var}(\xi_c)$. Fitting the model implied covariance matrix with $\phi_o = 1.0$ and $\phi_c = 1.0$ to the sample covariance matrix by normal-distribution-based maximum likelihood (NML) yields $T_{ml} = 2.073$, indicating that the model fits the data very well when referred to χ^2_4. The parameter estimates, their SEs, and the corresponding z-statistics for the confirmatory factor model are reported in Table 1. The reliabilities of the individual indicators estimated via the factor model are also included in Table 1 and so are those of the two Bartlett-factor scores (BFS_c, BFS_o).

[6] The values of the SNR in Tables 2, 3 and 5 are obtained by dividing the z-statistics by $(N-1)^{1/2}$, where N is the original sample size.

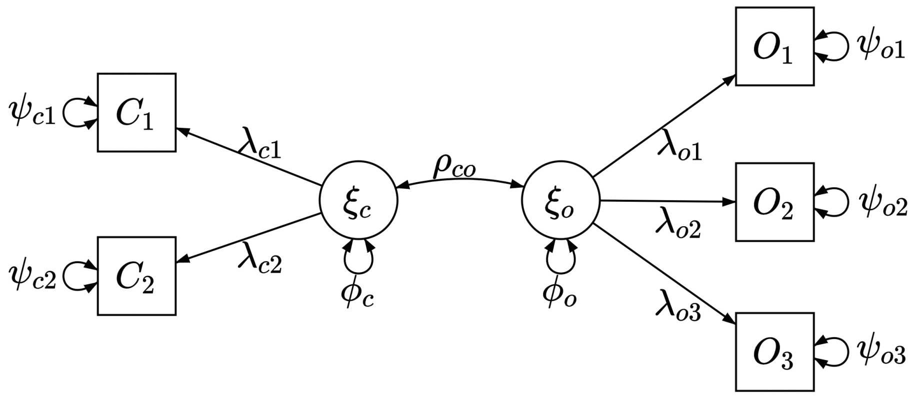

Figure 1. A two-factor model for the open- and closed-book test dataset.

Model For illustration purpose, let's consider the following two structural models under SEM

$$\xi_c = \gamma_{co}\xi_o + \zeta_c, \quad \text{and} \quad \xi_o = \gamma_{oc}\xi_c + \zeta_o. \tag{7}$$

That is, we predict the latent trait for the closed-book test by that for the open-book test, and the latent trait for the open-book test by that for the closed-book test, respectively. Note that our purpose here is to illustrate the properties of different methods rather than to testify the causal directions of the two traits. Actually, the two models in Equation (7) are mathematically equivalent to the confirmatory factor model in Figure 1 with respect to the overall model structure.

Parallel to the two structural models in Equation (7), we also estimate the following regression models

$$\hat{\xi}_c = \gamma_{*co}\hat{\xi}_o + e_c, \quad \text{and} \quad \hat{\xi}_o = \gamma_{*oc}\hat{\xi}_c + e_o \tag{8}$$

by the LS method, where $\hat{\xi}_c$ and $\hat{\xi}_o$ are composite-scores. There are many ways to formulate composite-scores, we will only consider equally-weighted composites (EWC), the BFSs and the RFSs in the study. Note that the EWCs are least selective among all composites since they don't use any of the psychometric properties of the individual indicators, whereas the two types of factor scores are most selective since they optimally use these properties. Also note that both the sum scores and the simple averages are special cases of EWCs.

For the structural models in Equation (7), we need to fix the scales of ξ_o and ξ_c in order for the models to be identified. There are 6 different options to identify each model by fixing two factor loadings at 1.0; 2 options to identify the model $\xi_o \to \xi_c$ via fixing $\phi_o = 1.0$; and 3 options to identify the model $\xi_c \to \xi_o$ via fixing $\phi_c = 1.0$. Thus, there are 8 different sets of scalings to identify the model $\xi_o \to \xi_c$; and 9 different sets of scalings to identify the model $\xi_c \to \xi_o$. The overall model remains equivalent among these different scalings, with T_{ml} being the same as that for the confirmatory factor model in Figure 1.

Table 1. Estimates (Est) of factor loadings (λ), error variances (ψ), factor correlation (ρ), and reliability (Rel) of the indicators (Ind) and Bartlett-factor scores (BFS) under the common factor model.

Param.	Est	SE	z	Param.	Est	SE	z	Ind/BF	Rel Est
λ_{c1}	12.253	1.843	6.649	ψ_{c1}	155.632	31.679	4.913	C1	.491
λ_{c2}	10.383	1.379	7.530	ψ_{c2}	65.036	18.099	3.593	C2	.624
λ_{o1}	9.834	0.929	10.588	ψ_{o1}	16.186	7.261	2.229	O1	.857
λ_{o2}	11.490	1.403	8.192	ψ_{o2}	88.352	16.773	5.268	O2	.599
λ_{o3}	12.517	1.667	7.508	ψ_{o3}	141.074	24.881	5.670	O3	.526
ρ_{co}	0.818	0.073	11.258					BFS_c	.724
								BFS_o	.896

Table 2. Estimates of the path coefficient, its standard deviation (SD) and the corresponding signal-to-noise ratio (SNR) by SEM, factor-score (FS) regression and equally-weighted-composite (EWC) regression for the relationship $\xi_o \to \xi_c$.

Method	Identification	Est.	SD	SNR
SEM	$\lambda_{c1} = 1.0$, $\lambda_{o1} = 1.0$	1.019	1.676	0.608
	$\lambda_{c1} = 1.0$, $\lambda_{o2} = 1.0$	0.872	1.528	0.571
	$\lambda_{c1} = 1.0$, $\lambda_{o3} = 1.0$	0.800	1.459	<u>0.548</u>
	$\lambda_{c2} = 1.0$, $\lambda_{o1} = 1.0$	0.863	1.216	**0.710**
	$\lambda_{c2} = 1.0$, $\lambda_{o2} = 1.0$	0.739	1.132	0.653
	$\lambda_{c2} = 1.0$, $\lambda_{o3} = 1.0$	0.678	1.094	0.620
	$\phi_o = 1.0$, $\lambda_{c1} = 1.0$	10.019	16.818	0.596
	$\phi_o = 1.0$, $\lambda_{c2} = 1.0$	8.490	12.286	0.691
FS-reg	$\lambda_{o1} = 1.0$, $\lambda_{c1} = 1.0$			
	BFS(ξ_o) & BFS(ξ_c)	0.912	1.055	0.865
	RFS(ξ_o) & RFS(ξ_c)	0.738	0.853	0.865
	BFS(ξ_o) & RFS(ξ_c)	0.661	0.764	0.865
	RFS(ξ_o) & BFS(ξ_c)	1.019	1.178	0.865
	$\phi_o = 1.0$, $\lambda_{c1} = 1.0$			
	BFS(ξ_o) & BFS(ξ_c)	8.973	10.377	0.865
	RFS(ξ_o) & RFS(ξ_c)	7.253	8.388	0.865
	BFS(ξ_o) & RFS(ξ_c)	6.496	7.512	0.865
	RFS(ξ_o) & BFS(ξ_c)	10.019	11.586	0.865
EWC-reg	sum(ξ_o) & sum(ξ_c)	0.428	0.587	0.730
	ave(ξ_o) & ave(ξ_c)	0.643	0.880	0.730
	sum(ξ_o) & ave(ξ_c)	0.214	0.293	0.730
	ave(ξ_o) & sum(ξ_c)	1.285	1.760	0.730

Note. BFS=Bartlett-factor score, RFS=regression-factor score; sum=sum score, ave=average score; the estimate with the largest SNR under SEM is in bold while the one with the smallest SNR is underlined.

Results The NML estimates of γ for the two models in Equation (7) are given in the upper panel of Tables 2 and 3, respectively. The standard deviation (SD) of each $\hat{\gamma}$ and the corresponding SNR are also included in the tables. Clearly, the value of $\hat{\gamma}$ under SEM changes as the scalings vary. But we do not regard their differences as problematic because their population counterparts are different, and each $\hat{\gamma}$ is consistent and efficient for a different γ (assuming data are normally distributed). As a matter of fact, we can choose the scale of ξ_o or that of ξ_c to make γ_{co} or γ_{oc} to equal any pre-specified (nonzero) value while the test statistic for the overall model structure remains at $T_{ml} = 2.073$.

Table 3. Estimates (Est) of the path coefficient, its standard deviation (SD) and the corresponding signal-to-noise ratio (SNR) by SEM, factor-score (FS) regression and equally-weighted-composite (EWC) regression for the relationship $\xi_c \to \xi_o$.

Method	Identification	Est	SD	SNR
SEM	$\lambda_{c1} = 1.0$, $\lambda_{o1} = 1.0$	0.656	1.101	0.596
	$\lambda_{c1} = 1.0$, $\lambda_{o2} = 1.0$	0.767	1.433	0.535
	$\lambda_{c1} = 1.0$, $\lambda_{o3} = 1.0$	0.835	1.614	<u>0.517</u>
	$\lambda_{c2} = 1.0$, $\lambda_{o1} = 1.0$	0.774	1.266	0.612
	$\lambda_{c2} = 1.0$, $\lambda_{o2} = 1.0$	0.905	1.656	0.546
	$\lambda_{c2} = 1.0$, $\lambda_{o3} = 1.0$	0.986	1.868	0.528
	$\phi_c = 1.0$, $\lambda_{o1} = 1.0$	8.040	10.424	**0.771**
	$\phi_c = 1.0$, $\lambda_{o2} = 1.0$	9.395	14.421	0.651
	$\phi_c = 1.0$, $\lambda_{o3} = 1.0$	10.235	16.500	0.620
FS-reg	$\lambda_{c1} = 1.0$, $\lambda_{o1} = 1.0$			
	BFS(ξ_c) & BFS(ξ_o)	0.475	0.549	0.865
	RFS(ξ_c) & RFS(ξ_o)	0.588	0.680	0.865
	BFS(ξ_c) & RFS(ξ_o)	0.425	0.492	0.865
	RFS(ξ_c) & BFS(ξ_o)	0.656	0.759	0.865
	$\phi_c = 1.0$, $\lambda_{o1} = 1.0$			
	BFS(ξ_c) & BFS(ξ_o)	5.821	6.732	0.865
	RFS(ξ_c) & RFS(ξ_o)	7.201	8.328	0.865
	BFS(ξ_c) & RFS(ξ_o)	5.213	6.029	0.865
	RFS(ξ_c) & BFS(ξ_o)	8.040	9.299	0.865
EWC-reg	sum(ξ_c) & sum(ξ_o)	0.824	1.128	0.730
	ave(ξ_c) & ave(ξ_o)	0.549	0.752	0.730
	sum(ξ_c) & ave(ξ_o)	0.275	0.376	0.730
	ave(ξ_c) & sum(ξ_o)	1.648	2.257	0.730

Note. BFS=Bartlett-factor score, RFS=regression-factor score; sum=sum score, ave=average score; the estimate with the largest SNR under SEM is in bold while the one with the smallest SNR is underlined.

In Tables 2 and 3, the largest SNR under SEM was put in bold and the smallest was underlined. The largest SNR in Table 2 corresponds to the condition when both ξ_o and ξ_c are anchored by the most reliable indicators, whereas the largest SNR in Table 3 corresponds to the condition when ξ_c is scaled by $\phi_c = 1.0$. In both Tables 2 and 3, the smallest SNR under SEM corresponds to the conditions when both ξ_o and ξ_c are anchored by the least reliable indicators.

The middle panel of Table 2 contains the estimates of γ_{*co} for the first regression model in Equation (8), where $\hat{\xi}_o$ and $\hat{\xi}_c$ are factor scores computed following the NML estimates of the parameters under the identification rules $\lambda_{o1} = \lambda_{c1} = 1.0$ and $\phi_o = \lambda_{c1} = 1.0$, respectively. Parallel results for the second regression model in Equation (8) are displayed in the middle panel of Table 3. In particular, the SNRs by the LS method for the two regression models in Equation (8) have the same value. To save space, we did not include the results of FS regression corresponding to all sets of scalings of the two latent variables, while their values of the SNR remain to be 0.865. For each identification condition, there is also a $\hat{\gamma}_*$ that has the same value as its SEM counterpart (i.e., 1.019 and 10.019 in Table 2, and 0.656 and 8.040 in Table 3), verifying the noted result by Skrondal and Laake (2001).

The lower panels of Tables 2 and 3 contain the LS estimates for the regression models in Equation (8) when $\hat{\xi}_o$ and $\hat{\xi}_c$ are the sum and/or average scores. They are denoted by EWC regression (EWC-reg) in the tables. Clearly, the values of the estimates of γ_{*co} and γ_{*oc} for the regression models in Equation (8) depend on the scales of the composites and so do their corresponding SDs. However, unlike SEM, the values of the SNR (as well as the corresponding z-statistic) under FS regression or EWC regression remain the same across different scalings. Note that the SNRs under EWC regression are smaller than those under FS regression, because the sum scores are not as reliable as the factor scores.

In Table 2, all the eight SNRs for $\hat{\gamma}$ by SEM are smaller than those for $\hat{\gamma}^*$ by EWC regression and by FS regression. In Table 3, all the nine SNRs for $\hat{\gamma}$ by SEM are smaller than that for $\hat{\gamma}_*$ by FS regression; and only one SNR for $\hat{\gamma}$ by SEM is larger than that for $\hat{\gamma}^*$ by EWC regression. Thus, using SNR as a measure for the efficiency of parameter estimates, among the 17 options for identifying the two SEM models, only in one option SEM outperforms EWC regression. None of the 17 SEM identification options yields a greater SNR than FS regression.

Results in Tables 2 and 3 also illustrate the fact that an LS estimate under regression analysis with weighted composites does not have to be smaller than its counterpart under SEM. Another interesting fact is that, unlike in regression analysis under which the SNRs for the coefficients of $x \to y$ and $y \to x$ are the same, the SNR under SEM differentiates the path coefficients between $\xi_o \to \xi_c$ and $\xi_c \to \xi_o$ even if the two latent variables are scaled by the same set of anchors (e.g., $\lambda_{c1} = \lambda_{o1} = 1.0$).

The results for FS regression in Tables 2 and 3 were obtained by treating the factor scores as the observed variables, which is widely used in practice (DiStefano, Zhu, & Mindrila, 2009; Widaman & Revelle, 2022). But parameter estimates used to compute the factor scores contain sampling errors, which affect the SEs of the resulting $\hat{\gamma}_*$. Results in Yuan and Fang (2023b) indicate that the SEs of the FS regression coefficients by considering the sampling errors in the estimated weights tend to be smaller than those by treating weights as being given, and FS regression may become even more powerful in detecting the existence of a relationship if the sampling errors in weights are accounted.

5.2 Example 2

Data Table 2 of Weston and Gore (2006) contains a sample covariance matrix for a dataset with $N = 403$ cases and $p = 12$ variables. The dataset was part of a survey of college students who participated in a vocational psychology research project. With three indicators for each construct, the 12 variables are respectively measures of 1) self-efficacy beliefs, 2) outcome expectations, 3) career-related interests, and 4) occupational considerations. Weston and Gore Jr considered two structural models. Deng and Yuan (2023) compared the values of z-statistics of parameter estimates for each model by different methods, where each latent variable was scaled by only one option. We consider only one of their models, and our purpose here is to see how the SNRs react to different options for scaling the latent variables.

Model The path diagram in Figure 2 corresponds to the first model of Weston and Gore (2006), which posits that the effect of self-efficacy beliefs on career-related interests is partially mediated by outcome expectations, while the effect of self-efficacy beliefs on occupational considerations is completely through the two mediator variables (outcome expectations & career-related interests). The structural model has four path coefficients: γ_{11}, γ_{21}, β_{21}, β_{32}.

For each latent variable in Figure 2, we can select one of the three factor loadings and fix it at 1.0 to anchor its scale. So by factor loadings alone there are $3^4 = 81$ different sets of scalings to identify the model. For the independent latent variable self-efficacy beliefs (ξ_1), we can also fix its variance at 1.0, which provides additional $3^3 = 27$ different sets of scalings to identify the model. With a total of $81 + 27 = 108$ ways of model identification, we will only study a subset of them to illustrate our point, and the selected subsets allow us to see how and what parameters are affected by the properties of the anchors.

Results Letting $\phi_{11} = \text{Var}(\xi_1) = 1.0$ and $\lambda_{y_1 1} = \lambda_{y_4 2} = \lambda_{y_7 3} = 1.0$, fitting the model in Figure 2 to the vocational-psychology dataset by NML results in $T_{ml} = 416.061$, which corresponds to a p-value that is essentially 0 when referred to χ^2_{50}. With CFI=.913, and RMSEA=.135, the model might not be regarded as fitting the data adequately although it is substantively derived (see Weston and Gore, 2006 and references therein). Such a discrepancy between theory and goodness of model-fit is not unusual in empirical modeling, reflecting our earlier observation that the theoretical constructs may not be perfectly[7] represented by

[7] When letting all the four latent variables be freely correlated in Figure 2, we have $T_{ml} = 361.848$, which corresponds to a p-value that is essentially 0 when referred to χ^2_{48}, indicating that discrepancy between the theoretical constructs and the latent variables also exists in the measurement model.

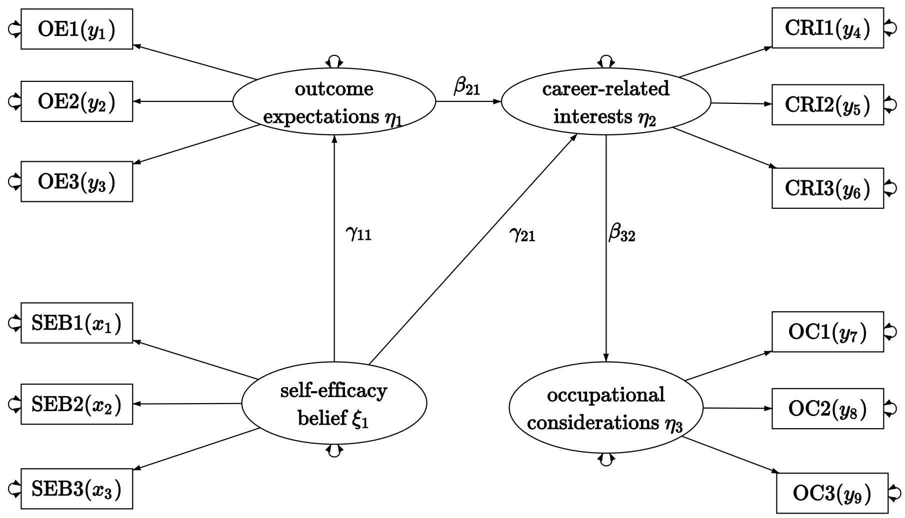

Figure 2. A mediated model for self-efficacy belief on occupational considerations (Weston and Gore, 2006, $N = 403$).

Table 4. Parameter estimates (Est), their SEs (SE) and z-statistics for the model in Figure 2 ($p = 12$, $N = 403$, $T_{ml} = 416.061$, $df = 50$, p-value=.000; RMSEA=.135, and CFI=.913). The right column is the reliability (ρ) of the 12 indicators.

Param.	Est	SE	z	Param.	Est	SE	z	Rel	Est
$\lambda_{x1,1}$	2.381	0.111	21.467	ψ_{x1}	1.840	0.179	10.264	ρ_{x_1}	0.755
$\lambda_{x2,1}$	2.365	0.108	21.903	ψ_{x2}	1.628	0.167	9.768	ρ_{x_2}	0.775
$\lambda_{x3,1}$	2.402	0.104	23.145	ψ_{x3}	1.186	0.148	8.013	ρ_{x_3}	0.830
$\lambda_{y1,1}$	1.000			ψ_{y1}	0.882	0.078	11.318	ρ_{y_1}	0.808
$\lambda_{y2,1}$	0.976	0.030	32.930	ψ_{y2}	0.325	0.048	6.827	ρ_{y_2}	0.916
$\lambda_{y3,1}$	0.993	0.032	30.628	ψ_{y3}	0.580	0.060	9.621	ρ_{y_3}	0.863
$\lambda_{y4,2}$	1.000			ψ_{y4}	0.044	0.003	12.853	ρ_{y_4}	0.394
$\lambda_{y5,2}$	1.144	0.094	12.229	ψ_{y5}	0.027	0.002	11.300	ρ_{y_5}	0.580
$\lambda_{y6,2}$	1.011	0.098	10.326	ψ_{y6}	0.049	0.004	12.974	ρ_{y_6}	0.372
$\lambda_{y7,3}$	1.000			ψ_{y7}	0.795	0.100	7.918	ρ_{y_7}	0.835
$\lambda_{y8,3}$	0.963	0.040	24.212	ψ_{y8}	1.350	0.126	10.705	ρ_{y_8}	0.734
$\lambda_{y9,3}$	0.795	0.033	23.934	ψ_{y9}	0.962	0.089	10.867	ρ_{y_9}	0.725
γ_{11}	1.186	0.096	12.364						
γ_{21}	0.046	0.009	5.161	$\sigma^2_{\zeta_1}$	2.302	0.211	10.913		
β_{21}	0.057	0.006	10.003	$\sigma^2_{\zeta_2}$	0.008	0.001	5.542		
β_{32}	10.368	0.822	12.615	$\sigma^2_{\zeta_3}$	0.964	0.151	6.396		

Table 5. Values of the signal-to-noise ratio (SNR) of $\hat{\gamma}_{11}$, $\hat{\gamma}_{21}$, $\hat{\beta}_{21}$ and $\hat{\beta}_{32}$ for the model in Figure 2 when ξ_1, η_1, η_2, and η_3 are anchored by fixing one of their loadings (λ) at 1.0, or by letting $\phi_{11} = \mathrm{Var}(\xi_1) = 1.0$.

anchors of	SNR				anchors of	SNR			
ξ_1, η_1, η_2, η_3	$\hat{\gamma}_{11}$	$\hat{\gamma}_{21}$	$\hat{\beta}_{21}$	$\hat{\beta}_{32}$	η_1, η_2, η_3	$\hat{\gamma}_{11}$	$\hat{\gamma}_{21}$	$\hat{\beta}_{21}$	$\hat{\beta}_{32}$
x_1, y_1, y_4, y_7	0.631	0.258	0.499	0.629	($\phi_{11} = 1.0$)				
x_2, y_1, y_4, y_7	0.636	0.259	0.499	0.629					
x_3, y_1, y_4, y_7	0.646	0.259	0.499	0.629	y_1, y_4, y_7	0.617	0.257	0.499	0.629
x_1, y_2, y_4, y_7	0.652	0.258	0.507	0.629	y_2, y_4, y_7	0.636	0.257	0.507	0.629
x_1, y_3, y_4, y_7	0.642	0.258	0.504	0.629	y_3, y_4, y_7	0.626	0.257	0.504	0.629
x_1, y_1, y_5, y_7	0.631	0.267	0.569	0.785	y_1, y_5, y_7	0.617	0.266	0.569	0.785
x_1, y_1, y_6, y_7	0.631	0.257	0.489	0.609	y_1, y_6, y_7	0.617	0.256	0.489	0.609
x_1, y_1, y_4, y_8	0.631	0.258	0.499	0.607	y_1, y_4, y_8	0.617	0.257	0.499	0.607
x_1, y_1, y_4, y_9	0.631	0.258	0.499	0.605	y_1, y_4, y_9	0.617	0.257	0.499	0.605
x_2, y_2, y_5, y_8	0.656	0.267	0.582	0.743					
x_1, y_2, y_5, y_8	0.652	0.267	0.582	0.743					
x_3, y_2, y_5, y_8	0.668	0.268	0.582	0.743	y_2, y_5, y_8	0.636	0.266	0.582	0.743
x_2, y_1, y_5, y_8	0.636	0.267	0.569	0.743	y_1, y_5, y_8	0.617	0.266	0.569	0.743
x_2, y_3, y_5, y_8	0.646	0.267	0.577	0.743	y_3, y_5, y_8	0.626	0.266	0.577	0.743
x_2, y_2, y_4, y_8	0.656	0.259	0.507	0.607	y_2, y_4, y_8	0.636	0.257	0.507	0.607
x_2, y_2, y_6, y_8	0.656	0.257	0.496	0.589	y_2, y_6, y_8	0.636	0.256	0.496	0.589
x_2, y_2, y_5, y_7	0.656	0.267	0.582	0.785	y_2, y_5, y_7	0.636	0.266	0.582	0.785
x_2, y_2, y_5, y_9	0.656	0.267	0.582	0.739	y_2, y_5, y_9	0.636	0.266	0.582	0.739
x_3, y_3, y_6, y_9	0.657	0.258	0.493	0.587					
x_1, y_3, y_6, y_9	0.642	0.257	0.493	0.587					
x_2, y_3, y_6, y_9	0.646	0.257	0.493	0.587	y_3, y_6, y_9	0.626	0.256	0.493	0.587
x_3, y_1, y_6, y_9	0.646	0.258	0.489	0.587	y_1, y_6, y_9	0.617	0.256	0.489	0.587
x_3, y_2, y_6, y_9	0.668	0.258	0.496	0.587	y_2, y_6, y_9	0.636	0.256	0.496	0.587
x_3, y_3, y_4, y_9	0.657	0.259	0.504	0.605	y_3, y_4, y_9	0.626	0.257	0.504	0.605
x_3, y_3, y_5, y_9	0.657	0.268	0.577	0.739	y_3, y_5, y_9	0.626	0.266	0.577	0.739
x_3, y_3, y_6, y_7	0.657	0.258	0.493	0.609	y_3, y_6, y_7	0.626	0.256	0.493	0.609
x_3, y_3, y_6, y_8	0.657	0.258	0.493	0.589	y_3, y_6, y_8	0.626	0.256	0.493	0.589

Note. The underlined row in each block is for reference against which the other lines of the block are compared.

the Greek letters in Figure 2. But the discrepancy between the model and data has little to do with our illustration, since the results are essentially the same even if the model fits the data perfectly (i.e., letting the sample covariance matrix equal the model implied covariance matrix). For reference, Table 4 contains the estimates of the factor loadings (λ), error variances (ψ), the path coefficients of the structural model, and the variances of the three prediction errors (σ_ζ^2). The last column of Table 4 indicate that x_3 is the most reliable indicator for ξ_1 while y_2, y_5 and y_7 are the most reliable indicators for η_1, η_2 and η_3, respectively. All the parameter estimates in Table 4 are statistically significant at the level of .05.

Table 5 contains the values of the SNR for the four path coefficients under 48 different identification conditions (out of 108 options). Results on the left side of the table are obtained when one of the loadings of ξ_1 is fixed at 1.0, while those on the right side are obtained by letting $\phi_{11} = \text{Var}(\xi_1) = 1.0$. Note that the values of the four parameter estimates vary across the 48 sets of scalings, while $T_{ml} = 416.061$. Our main interest with this example is the pattern of the SNRs while they vary with the parameter estimates when the latent variables are scaled differently.

There are 6 blocks of results in Table 5, and each block has one set of scalings underlined, which serves the condition for reference. In particular, for each set of scalings within a given block, only one of the four latent variables is rescaled compared to the reference condition. The results in Table 5 exhibit the following patterns.

1) When ξ_1 is anchored by x_1, x_2, x_3 or by $\phi_{11} = 1.0$, the SNRs for $\hat{\beta}_{21}$ and $\hat{\beta}_{32}$ are not affected by the scale change of ξ_1. In addition, results not included in the table also indicate that the values of $\hat{\beta}_{21}$ and $\hat{\beta}_{32}$ as well as their SDs are not affected by the scale change of ξ_1 either. This is because the paths represented by β_{21} and β_{32} are not directly connected with ξ_1 in Figure 2. However, the SNRs or equivalently the z-statistics for both $\hat{\gamma}_{11}$ and $\hat{\gamma}_{21}$ become greater when ξ_1 is anchored by a more reliable indicator.
2) In Figure 2, the paths represented by γ_{21} and β_{32} are not directly connected with η_1. When η_1 is anchored by different indicators, the SNRs for $\hat{\gamma}_{21}$ and $\hat{\beta}_{32}$ are not affected. However, the SNRs for $\hat{\gamma}_{11}$ and $\hat{\beta}_{21}$ become greater as the anchor of η_1 is more reliable.
3) In Figure 2, the path represented by γ_{11} is not directly connected with η_2. When η_2 is anchored by different indicators, the SNR for $\hat{\gamma}_{11}$ is not affected by the scale change of η_2. However, the SNRs for $\hat{\gamma}_{21}$, $\hat{\beta}_{21}$ and $\hat{\beta}_{32}$ become greater as η_2 is anchored by a more reliable indicator.
4) For the same reason, the SNRs for $\hat{\gamma}_{11}$, $\hat{\gamma}_{21}$ and $\hat{\beta}_{21}$ are not affected by the scale change of η_3. The SNR for $\hat{\beta}_{32}$ becomes greater as η_3 is anchored by a more reliable indicator.

The results in Table 5 suggested that, the SNR or z-statistic for a parameter estimate is invariant to the scale changes of the latent variables that are not directly connected with the path that the parameter represents. In contrast, the SNR for a parameter estimate becomes greater when the directly connected latent variables are anchored by more reliable indicators. In particular, the greatest

SNRs for $\hat{\gamma}_{11}$, $\hat{\gamma}_{21}$, $\hat{\beta}_{21}$ and $\hat{\gamma}_{32}$ are respectively $\text{SNR}_{\gamma_{11}} = .668$, $\text{SNR}_{\gamma_{21}} = .268$, $\text{SNR}_{\beta_{21}} = .582$, and $\text{SNR}_{\beta_{32}} = .785$. They are simultaneously obtained when all the latent variables are anchored by indicators with the greatest reliability (i.e., x_3 for ξ_1, y_2 for η_1, y_5 for η_2, and y_7 for η_3).

Note that, while the values of $\hat{\gamma}$, $\hat{\beta}$ and SNR change when different indicators are used as anchors, the value of the SNR under SEM will remain the same once the anchors are chosen regardless of the particular values of the factor loadings. That is, $\lambda_{o1} = 1.0$ or $\lambda_{o1} = 2.3$ leads to the same SNR in Tables 2 and 3. Similarly, the value of the SNR (or z-statistic) remains the same once the scale of ξ_c is determined by fixing the value of ϕ_c regardless of its particular value, e.g., $\phi_c = 1.0$ or $\phi_c = 3.5$ corresponds to the same SNR. More systematic results in this direction are presented in Yuan, Ling, and Zhang (2024).

The results of the two examples provide the fact for answering questions Q1, Q2 and Q3.

6 Answers to Questions Q1 to Q5

Our analysis and results might have already answered the questions posed in the introduction of the article. As a summary, we will answer them directly in this section. For clarity, we will also include the original questions.

Q1. *Under SEM, what is the effect of different scaling options on the accuracy of parameter estimates and the related z-statistics? How can we use the information to serve our purpose?*

When a latent variable is anchored by an indicator with greater reliability, the SNRs and consequently the z-statistics for path coefficients that are directly related to the latent variables are expected to be greater. However, estimates of the path coefficients that are not directly related to the latent variables are not affected nor their SEs. Also, scaling an independent latent variable by fixing its variance at 1.0 may result in even greater SNRs. Thus, under SEM, we can obtain more efficient parameter estimates of path coefficients by selecting more reliable anchors for latent variables.

Q2. *Do measurement errors cause attenuated or biased estimates for the LS method of regression analysis with weighted composites?*

Measurement errors alone do not cause biased or attenuated regression coefficients. It is the artificially chosen scales that make the path coefficients under regression analysis different from those under SEM.

Q3. *Does SEM yield more accurate parameter estimates than LS regression with weighted composites?*

SEM does not yield more accurate parameter estimates than LS regression with weighted composites. When measured by SNR, it is more likely the other way around. That is, even regression analysis with EWCs may yield more efficient estimates of path coefficients than SEM, especially when the indicators for each latent variable are approximately parallel.

Q4. *Between SEM and regression analysis with weighted composites, to what level the estimated regression coefficients can be compared with their SEM counterparts?*

There are three levels that the path coefficients under regression analysis can be compared with their counterparts under SEM: (1) $\boldsymbol{\gamma}_w = \mathbf{0}$ if and only if $\boldsymbol{\gamma} = \mathbf{0}$, where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_p)'$ and $\boldsymbol{\gamma}_w = (\gamma_{w1}, \gamma_{w2}, \ldots, \gamma_{wp})'$ are respectively the path coefficients of a given dependent variable under SEM and regression analysis with weighted composites. Inference for one set of parameters can be done by a statistical test on the other set of parameters, simultaneously. (2) $\boldsymbol{\gamma} = \boldsymbol{\gamma}_w$ when a BFS is used as the dependent variable and the joint RFSs are used as the independent variables in FS regression. Then, the two sets of parameters as well as their respective estimates can be substituted for each other. (3) Regardless of the scales chosen for the variables under each modeling technique, the size of the multivariate SNR determines the statistical power in testing $\boldsymbol{\gamma} = \mathbf{0}$, and FS regression is expected to correspond to a greater multivariate SNR and consequently to be more powerful than SEM.

Q5. *Does standardization advance result interpretation or only facilitate model identification?*

Whether the measurements have predefined metrics or not, standardization only facilitates model identification. Unless standard deviation is a widely used and well-understood unit in a given context, standardization does not advance our understanding nor facilitates interpretation of the parameter estimates in a SEM or regression model.

7 Discussion and Conclusion

We studied several widely circulated notions in modeling data with measurement errors but without predefined metrics. While modeling such data poses many challenges, SEM still offers important information that regression analysis with composite scores is unable to provide. In particular, SEM gives a platform to assess the goodness of the overall model structure, unidimensionality of different subscales, individual indicator reliability, and measurement invariance for group comparison, etc. However, while inference on parameter estimates under SEM can be statistically sound, substantive interpretation on the size of a path coefficient becomes a challenge if the measurements do not have predefined or well-understood scales to start with.

Because path coefficients under both SEM and regression analysis with composites depend on subjectively assigned scales of the involved variables, there is no point to demand regression analysis to yield estimates that are consistent with those under SEM. For parameters whose population values depend on subjectively assigned scales, a natural criterion for comparing their estimates is SNR, which is parallel to Cohen's d and serves as an index for the efficiency of the estimate. When all the path coefficients of a dependent variable are considered simultaneously, FS regression is expected to correspond to a greater (multivariate) SNR than SEM. If the indicators for each latent variable do not greatly deviate from parallel, EWC regression is also expected to outperform SEM with respect to efficiency of the estimated regression coefficients.

In addition to numerical differences between estimates by different methods, parameters under SEM are to govern the relationship of variables representing the population, where individuals are treated equally (i.e., a random representation). In practice, individuals with greater pretest scores are expected to perform better on the outcome variable, and parameters under regression analysis with weighted composites are to govern such a relationship. In particular, conditional on the metrics of the observed and latent variables, the predicted values according to the uncorrected LS estimates of the regression model still have the smallest MSE even if the target is the latent-outcome variable and the pretest scores are not error-free. But the SNR and R^2 of the regression model as well as the MSE of the predicted value are related to the size of measurement errors. More reliable composites correspond to more accurate estimates of path coefficients, greater R^2 values, and smaller prediction errors.

We have shown that standardization of latent or manifest variables is useful for model identification, not necessarily advancing our understanding of the relationship among the involved variables nor better interpretation of the parameters of the model either. That said, we do not exclude a context under which the distribution of a variable can be well understood in a standardized scale with mean 0 and variance 1.0. For example, we might transform the distribution of IQ (ξ) by $z_\xi = (\xi - 100)/15$ and the value of z_ξ allows us to judge the standing of the corresponding ξ in percentile according to the standard normal distribution $N(0, 1)$. However, this might belong to the case where the scale of the measurement was known in advance.

Bias can be easily defined and explained for parameter estimates in modeling variables that have predefined metrics. Empirical bias can also be easily evaluated in Monte Carlo studies even when data do not have predefined metrics. However, it is not clear how to interpret bias substantively if the scales of the variables need to be subjectively assigned. In particular, two researchers who conduct SEM analyses can have very different parameter estimates for the same path coefficient while they both are consistent/unbiased. Similarly, one researcher can choose the sum scores while another can choose the average scores in regression analyses, and they have identical t-statistics and R^2. But their parameter estimates are different. More generally, suppose Researcher A gets an estimate $\hat{\gamma}_a$ while Researcher B get an estimator $\hat{\gamma}_b = c\hat{\gamma}_a$, where $c > 0$ is a constant. We are unable to compare the two estimators with respect to bias since (if needed) Researcher B can always rescale his estimator to $\tilde{\gamma}_b = \hat{\gamma}_b/c$ if the involved variables do not have predefined metrics. Regardless, we recommend the one with a greater SNR, and $\tilde{\gamma}_b$ and $\hat{\gamma}_b$ are equivalent in the sense that they have the same SNR.

The first take-home message from this article is that sizes of parameter estimates and their SEs are not meaningful quantities for models involving (latent or manifest) variables that do not have predefined metrics, and SNR is a logical and also a natural measure of efficiency of parameter estimates. For the same reason, the MSE of parameter estimates is not a logical criterion to compare across methods unless the population values of the parameters are held constant among the methods or when all the involved variables are on the same metrics. When the estimands become equal, the most efficient and accurate estimates have the greatest SNR. The 2nd take-home message is that standardization does not advance interpretation but offers a way to avoid dealing with the issues of lack of metrics. Effort needs to be made to develop substantively rationalized metrics under which parameter estimates are interpreted.

Acknowledgments

This work was supported by a grant from the Department of Education (R305D210023). However, the contents of the study do not necessarily represent the policy of the funding agencies, and you should not assume endorsement by the Federal Government. Correspondence concerning this article should be addressed to Ke-Hai Yuan (kyuan@nd.edu).

References

Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory.* Monterey, CA: Brooks-Cole.

Baguley, T. (2009). Standardized or simple effect size: What should be reported? *British Journal of Psychology*, *100*(3), 603–617. doi: https://doi.org/10.1348/000712608x377117

Bentler, P. M. (1968). Alpha-maximized factor analysis (alphamax): Its relation to alpha and canonical factor analysis. *Psychometrika*, *33*(3), 335–345. doi: https://doi.org/10.1007/bf02289328

Bentler, P. M. (2006). *EQS 6 structural equations program manual.* Encino, CA: Multivariate Software.

Bentler, P. M. (2017). Specificity-enhanced reliability coefficients. *Psychological Methods*, *22*(3), 527–540. doi: https://doi.org/10.1037/met0000092

Bollen, K. A. (1989). *Structural equations with latent variables.* New York, NY: John Wiley & Sons.

Buonaccorsi, J. P. (2010). *Measurement error: models, methods, and applications.* Boca Raton, FL: Chapman & Hall.

Cadigan, N. G. (1995). Local influence in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *2*(1), 13–30. doi: https://doi.org/10.1080/10705519509539992

Cochran, W. G. (1970). Some effects of errors of measurement on multiple correlation. *Journal of the American Statistical Association*, *65*(329), 22–34. doi: https://doi.org/10.1080/01621459.1970.10481059

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Hillsdale NJ: Lawrence Erlbaum Associates.

Croon, M. (2014). Using predicted latent scores in general latent structure models. In G. A. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure models* (pp. 195–223). Psychology Press.

Deng, L., & Yuan, K.-H. (2023). Which method is more powerful in testing the relationship of theoretical constructs? a meta comparison of structural equation modeling and path analysis with weighted composites. *Behavior Research Methods*, *55*(3), 1460–1479. doi: https://doi.org/10.3758/s13428-022-01838-z

Devlieger, I., Mayer, A., & Rosseel, Y. (2016). Hypothesis testing using factor score regression: A comparison of four methods. *Educational and Psychological Measurement*, *76*(5), 741–770. doi: https://doi.org/10.1177/0013164415607618

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, *14*(20), 1–11. doi: https://doi.org/10.7275/DA8T-4G52

Forero, C. G., & Maydeu-Olivares, A. (2009). Estimation of irt graded response models: Limited versus full information methods. *Psychological Methods*, *14*(3), 275–299. doi: https://doi.org/10.1037/a0015825

Fuller, W. A. (2009). *Measurement error models.* John Wiley & Sons.

Gonzalez, R., & Griffin, D. (2001). Testing parameters in structural equation modeling: Every "one" matters. *Psychological Methods*, *6*(3), 258–269. doi: https://doi.org/10.1037/1082-989x.6.3.258

Kim, J.-O., & Ferree, G. D. (1981). Standardization in causal analysis. *Sociological Methods & Research*, *10*(2), 187–210. doi: https://doi.org/10.1177/004912418101000203

Lawley, D. N., & Maxwell, M. A. (1971). *Factor analysis as a statistical method* (2nd ed.). London: Butterworths.

Ledgerwood, A., & Shrout, P. E. (2011). The trade-off between accuracy and precision in latent variable models of mediation processes. *Journal of Personality and Social Psychology*, *101*(6), 1174–1188. doi: https://doi.org/10.1037/a0024776

Loehlin, J. C., & Beaujean, A. A. (2017). *Latent variable models: An introduction to factor, path, and structural equation analysis* (5th ed.). New York: Routledge.

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis.* New York: Academic Press.

McGrath, R. E., & Meyer, G. J. (2006). When effect sizes disagree: The case of r and d. *Psychological Methods*, *11*(4), 386–401. doi: https://doi.org/10.1037/1082-989x.11.4.386

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, *52*(6), 2287–2305. doi: https://doi.org/10.3758/s13428-020-01398-0

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: Applications, interpretations, and limitations. *Contemporary Educational*

Psychology, *25*(3), 241–286. doi: https://doi.org/10.1006/ceps.2000.1040

Pek, J., & Flora, D. B. (2018). Reporting effect sizes in original psychological research: A discussion and tutorial. *Psychological Methods*, *23*(2), 208–225. doi: https://doi.org/10.1037/met0000126

Schoenbach, V. J., & Rosamond, W. D. (2000). *Understanding the fundamentals of epidemiology: an evolving text.* Chapel Hill:: University of North Carolina.

Shi, D., & Tong, X. (2017). The impact of prior information on bayesian latent basis growth model estimation. *SAGE Open*, *7*(3), 215824401772703. doi: https://doi.org/10.1177/2158244017727039

Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, *66*(4), 563–575. doi: https://doi.org/10.1007/bf02296196

Steiger, J. H. (2002). When constraints interact: A caution about reference variables, identification constraints, and scale dependencies in structural equation modeling. *Psychological Methods*, *7*(2), 210–227. doi: https://doi.org/10.1037/1082-989x.7.2.210

Tanaka, Y., Watadani, S., & Ho Moon, S. (1991). Influence in covariance structure analysis: with an application to confirmatory factor analysis. *Communications in Statistics - Theory and Methods*, *20*(12), 3805–3821. doi: https://doi.org/10.1080/03610929108830742

Weston, R., & Gore, P. A. (2006). A brief guide to structural equation modeling. *The Counseling Psychologist*, *34*(5), 719–751. doi: https://doi.org/10.1177/0011000006286345

Widaman, K. F., & Revelle, W. (2022). Thinking thrice about sum scores, and then some more about measurement and analysis. *Behavior Research Methods*, *55*(2), 788–806. doi: https://doi.org/10.3758/s13428-022-01849-w

Yuan, K.-H., & Bentler, P. M. (2002). On robusiness of the normal-theory based asymptotic distributions of three reliability coefficient estimates. *Psychometrika*, *67*(2), 251–259. doi: https://doi.org/10.1007/bf02294845

Yuan, K.-H., & Deng, L. (2021). Equivalence of partial-least-squares sem and the methods of factor-score regression. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(4), 557–571. doi: https://doi.org/10.1080/10705511.2021.1894940

Yuan, K.-H., & Fang, Y. (2023a). Replies to comments on "which method delivers greater signal-to-noise ratio: Structural equation modelling or regression analysis with weighted composites?" by yuan and fang (2023). *British Journal of Mathematical and Statistical Psychology*, *76*(3), 695–704. doi: https://doi.org/10.1111/bmsp.12323

Yuan, K.-H., & Fang, Y. (2023b). Which method delivers greater signal-to-noise ratio: Structural equation modelling or regression analysis with weighted composites? *British Journal of Mathematical and Statistical Psychology*, *76*(3), 646–678. doi: https://doi.org/10.1111/bmsp.12293

Yuan, K.-H., Ling, L., & Zhang, Z. (2024). Scale-invariance, equivariance and dependency of structural equation models. *Struc-*

tural Equation Modeling: A Multidisciplinary Journal, 1–16. doi: https://doi.org/10.1080/10705511.2024.2353168

Yuan, K.-H., Marshall, L. L., & Bentler, P. M. (2003). 8. assessing the effect of model misspecifications on parameter estimates in structural equation models. *Sociological Methodology*, *33*(1), 241–265. doi: https://doi.org/10.1111/j.0081-1750.2003.00132.x

Zhang, Q., & Yang, Y. (2020). Autoregressive mediation models using composite scores and latent variables: Comparisons and recommendations. *Psychological Methods*, *25*(4), 472–495. doi: https://doi.org/10.1037/met0000251

Journal of Behavioral Data Science, 2024, *4* (2), 29–50.
DOI: https://doi.org/10.35566/jbds/ogasawara2

Rephrasing the Lengthy and Involved Proof of Kristof's Theorem: A Tutorial with Some New Findings

Haruhiko Ogasawara[0000−0002−5029−2086]

Otaru University of Commerce, Otaru, Japan
emt-hogasa@emt.otaru-uc.ac.jp

Abstract. Kristof's theorem gives the global maximum and minimum of the trace of some matrix products without using calculus or Lagrange multipliers with various applications in psychometrics and multivariate analysis. However, the underutilization has been seen irrespective of its great use in practice. This may partially be due to the lengthy and involved proof of the theorem. In this tutorial, some known or new lemmas are rephrased or provided to understand the essential points in the proof. Ten Berge's generalized Kristof theorem is also addressed. Then, the modified Kristof and Ten Berge theorems using parent orthonormal matrices are shown, which may be of use to see the properties of the Kristof and Ten Berge theorems.

Keywords: von Neumann's trace inequality · Generalized Kristof theorem · Suborthonormal · Semiorthonormal · Singular value decomposition

1 Introduction

Kristof's (1969; 1970) theorem for the global maximum of the trace of matrix products gives simple derivations of the least square solutions for various problems in psychometrics and multivariate analysis. A special but basic case of the theorem for two sets of matrix products yielding a bilinear form was given by von Neumann (1937, Theorem 1) known as his trace inequality, which was introduced in psychometrics by Green Jr (1969, p. 317) based on the comment of Ingram Olkin.

In spite of its great use, von Neumann's derivation using sophisticated mathematics was not easy for applied researchers to follow. Simplifications or elementary derivations of the theorem has been given by e.g., Kristof (1970) and Mirsky (1975). It is to be noted that these two authors also gave extensions of von Neumann's trace inequality to those with more than two sets of matrix products in different forms.

While the proof by Kristof of his theorem is based on elementary linear algebra using mostly self-contained materials, the proof is long and involved.

This may be one of the reasons for the relatively small frequency of citations as commented by Levin (1979, p. 109) "Kristof's theorem has not received its due attention in the psychometric literature", which was also cited by Waller (2018, Introduction), who also stated that "Underutilization of this method likely stems, in part, to the mathematical complexity of Kristof's (1964; 1970) writings" (Abstract).

One of the purposes of this tutorial is to break down Kristof's long and involved derivation using independent lemmas to provide a transparent structure of the proof. Note that the lemmas may also be of interest as general results in elementary linear algebra. The second purpose is to introduce a short derivation of von Neumann's trace inequality obtained by Mirsky (1975) as mentioned earlier, where Fan (1951)'s lemma with a self-contained didactic proof is introduced. Note that von Neumann's trace inequality and its extensions have wide applications in various fields e.g., applied linear algebra, mathematical physics and the hyperelasticity of isotropic materials as well as psychology as reviewed by Miranda and Thompson (1993), who cited Kristof (1970) among associated references.

The remainder of this article is organized as follows. In Section 2, some lemmas are introduced for Kristof's theorem followed by a didactic derivation of von Neumann's trace inequality. Section 3 gives didactic proofs of Kristof's theorem for the 3-fold or tri-linear case and the general case. In Section 4, Ten Berge's generalized theorem and modifications of the Kristof and Ten Berge theorems are presented. Some applications of these theorems are shown in Section 5. Section 6 gives discussions. In the appendix, technical details are provided.

2 Lemmas for Kristof's theorem and a didactic derivation of von Neumann's trace inequality

In this section, six lemmas and a theorem with a didactic derivation of von Neumann's trace inequality in line with the later derivation of Kristof's theorem will be shown. Lemma 1 gives the maximum of the sum of products of two quantities required for von Neumann's trace inequality, followed by Lemma 1A corresponding to Kristof (1970, Lemma 1), for the similar maximum of the sum of products of more than two quantities for the derivation of Kristof's theorem. Lemma 2 shows the same ranges of the traces irrespective of their absolute values of associated diagonal elements with permutation, which corresponds to Kristof (1970, Lemma 2).

Lemma 3 is a new independent lemma corresponding to the symmetric condition for a maximized trace in Kristof (1970, (iv) of the proof of Theorem (first version)). Lemma 4 is the second independent lemma for the property that symmetric $\mathbf{AD}$ and $\mathbf{DA}$ with $\mathbf{D}$ being diagonal make $\mathbf{A}$ diagonal (Kristof, 1970, (iv) of the proof of Theorem (first version)).

Lemma 5 is the third independent lemma when we have two products of orthonormal and diagonal matrices (a special case of Kristof (1970, (iv) of the proof of Theorem (first version))) for von Neumann's trace inequality, which was provided to understand the inequality as a special case of Kristof's theorem. Theorem 1 is for von Neumann's trace inequality with the derivation similar to the later one for Kristof's general theorem.

Lemma 1 : The maximum of the sum of products of two quantities (Hardy, Littlewood, and Pólya (1934, 1952, Subsection 10.2); von Neumann (1937, Theorem 1); Simon (2005, Lemma 1.8)). *For two sets of m numbers with $a_1 \geq \cdots \geq a_m \geq 0$ and $b_1 \geq \cdots \geq b_m \geq 0$, let $a_1^*, ..., a_m^*$ and $b_1^*, ...b_m^*$ be arbitrary cases in each set of $m!$ permutations including possibly the same ones. Then, the maximum of $\sum_{i=1}^m a_i^* b_i^*$ over the permutations is given by $\sum_{i=1}^m a_i b_i$.*

Proof. Without loss of generality, consider the maximum of $\sum_{i=1}^m a_i b_i^*$. Suppose that $b_1^* \neq b_1$. Then, exchanging b_1^* and $b_k^* = b_1 (k \neq 1)$ in the permutation, $\sum_{i=1}^m a_i b_i^*$ increases if $b_1^* \neq b_k^*$ and is unchanged if $b_1^* = b_k^*$ since $(a_i - a_j)(b_i - b_j) \geq 0$ and consequently $a_i b_i + a_j b_j - (a_i b_j + a_j b_i) \geq 0$ $(1 \leq i \leq j \leq m)$. Using this possibly exchanged permutation, redefine $b_1^*, ..., b_m^*$. Then, when $b_2^* \neq b_2$, exchange b_2^* and $b_k^* = b_2 (k > 2)$. Repeat this process until the possible exchange of b_{m-1}^* and $b_m^* = b_{m-1}$ when $b_{m-1}^* \neq b_{m-1}$. The final permutation gives $\sum_{i=1}^m a_i b_i^* = \sum_{i=1}^m a_i b_i$, which is the maximum since no permutation $b_1^*, ..., b_m^*$ using pairwise exchanges after the final one increases $\sum_{i=1}^m a_i b_i^*$. □

The above proof is a "heuristic" one finding the maximum successively. Waller (2018, Topic III) also used a similar "constructive proof" for the above lemma based on the proof of Simon (2005, Lemma 1.8., p. 4), which is elementary though of interest. However, the above heuristic proof seems to be simpler than Waller's didactic one. Note that "heuristic" is synonymous with "constructive" in this case.

Lemma 1A: The maximum of the sum of products with arbitrary number of factors Kristof (1970, Lemma 1). *For n sets of m numbers with $a_1^{(j)} \geq \cdots \geq a_m^{(j)} \geq 0$ $(j = 1, ..., n;\ n \geq 2)$, let $a_1^{(j)*}, ..., a_m^{(j)*}$ be an arbitrary case in the j-th set of $m!$ permutations including possibly the same ones. Then, the maximum of $\sum_{i=1}^m a_i^{(1)*} \cdots a_i^{(n)*}$ over the permutations is given by $\sum_{i=1}^m a_i^{(1)} \cdots a_i^{(n)}$.*

Proof. Consider the case of $n = 3$. For two sets $a_1^{(j)*}, ..., a_m^{(j)*}$ $(j = 1, 2)$ in the three sets, Lemma 1 gives the maximum of the sum of the products as $\sum_{i=1}^m a_i^{(1)} a_i^{(2)}$. Then, for the two sets of m products $a_1^{(1)} a_1^{(2)}, ..., a_m^{(1)} a_m^{(2)}$ and m numbers $a_1^{(3)*}, ..., a_m^{(3)*}$, the maximum of $\sum_{i=1}^m a_i^{(1)} a_i^{(2)} a_i^{(3)*}$ over the $m!$ permutation in the third set is similarly obtained by $\sum_{i=1}^m a_i^{(1)} a_i^{(2)} a_i^{(3)}$. Since any permutation for the maximized one including the first two sets decreases the product sum or remains unchanged as seen in Lemma 1, $\sum_{i=1}^m a_i^{(1)} a_i^{(2)} a_i^{(3)}$ is the global maximum. The cases with $n \geq 4$ is similarly obtained. □

Lemma 2 : The same ranges of the traces irrespective of their absolute values of the diagonal elements with permutation (Kristof, 1970, Lemma 2). *Let* $\mathbf{\Gamma}_1^*$ *and* $\mathbf{\Gamma}_2^*$ *be* $m \times m$ *diagonal matrices; and* $\mathbf{\Gamma}_1$ *and* $\mathbf{\Gamma}_2$ *be those with the corresponding diagonal elements replaced by their absolute values located in the weakly descending (non-increasing) orders, respectively. Suppose that* $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *independently vary over all the* $m \times m$ *orthonormal matrices. Then,* $\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1^*\mathbf{X}_2\mathbf{\Gamma}_2^*)$ *has the same range as that of* $\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2\mathbf{\Gamma}_2)$.

Proof. Kristof's derivation is didactically repeated. Note that $\mathbf{\Gamma}_i$ is obtained by $\mathbf{\Gamma}_i = \mathbf{P}_i\mathbf{S}_i\mathbf{\Gamma}_i^*\mathbf{P}_i^{\mathrm{T}}$, where $\mathbf{S}_i$ is the signed identity matrix replacing the diagonal elements of $\mathbf{\Gamma}_i^*$ by their corresponding absolute values, and $\mathbf{P}_i$ is the permutation matrix to have the weakly descending order mentioned earlier ($i = 1, 2$). Noting that $\mathbf{\Gamma}_i^* = \mathbf{S}_i\mathbf{P}_i^{\mathrm{T}}\mathbf{\Gamma}_i\mathbf{P}_i$, we have

$$\begin{aligned}\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1^*\mathbf{X}_2\mathbf{\Gamma}_2^*) &= \mathrm{tr}\{\mathbf{X}_1(\mathbf{S}_1\mathbf{P}_1^{\mathrm{T}}\mathbf{\Gamma}_1\mathbf{P}_1)\mathbf{X}_2(\mathbf{S}_2\mathbf{P}_2^{\mathrm{T}}\mathbf{\Gamma}_2\mathbf{P}_2)\} \\ &= \mathrm{tr}\{(\mathbf{P}_2\mathbf{X}_1\mathbf{S}_1\mathbf{P}_1^{\mathrm{T}})\mathbf{\Gamma}_1(\mathbf{P}_1\mathbf{X}_2\mathbf{S}_2\mathbf{P}_2^{\mathrm{T}})\mathbf{\Gamma}_2)\}.\end{aligned}$$

In the last result, $\mathbf{P}_2\mathbf{X}_1\mathbf{S}_1\mathbf{P}_1^{\mathrm{T}}$ and $\mathbf{P}_1\mathbf{X}_2\mathbf{S}_2\mathbf{P}_2^{\mathrm{T}}$ are products of orthonormal matrices and consequently orthonormal with the same variations of $\mathbf{X}_1$ and $\mathbf{X}_2$, which shows the required same ranges of $\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1^*\mathbf{X}_2\mathbf{\Gamma}_2^*)$ and $\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2\mathbf{\Gamma}_2)$. □

The following lemma gives a derivation of Kristof's theorem shown later when $n = 1$, which will also be used in the cases when $n = 2, 3, ...$ in the derivation of the theorem by induction.

Lemma 3 A symmetric condition for a maximized trace (Kristof, 1970, (iv) of the proof of Theorem (first version)). *Let* $\mathbf{G}$ *be a square matrix of full rank whose singular value decomposition (SVD) is* $\mathbf{G} = \mathbf{U\Lambda V}^{\mathrm{T}}$*, where* $\mathbf{U}$ *and* $\mathbf{V}$ *are orthonormal, and* $\mathbf{\Lambda}$ *is a diagonal matrix with positive diagonal elements in a prescribed order. When* $\mathrm{tr}(\mathbf{G})$ *is maximized with a given* $\mathbf{\Lambda}$*,* $\mathbf{G}$ *becomes symmetric.*

Proof. Since $\mathrm{tr}(\mathbf{G}) = \mathrm{tr}(\mathbf{U\Lambda V}^{\mathrm{T}}) = \mathrm{tr}(\mathbf{V}^{\mathrm{T}}\mathbf{U\Lambda})$ with $\mathbf{V}^{\mathrm{T}}\mathbf{U}$ being orthonormal, $\mathrm{tr}(\mathbf{G})$ is maximized when $\mathbf{V}^{\mathrm{T}}\mathbf{U}$ is an identity matrix, which indicates that $\mathbf{U} = \mathbf{V}$ and consequently $\mathbf{G} = \mathbf{U\Lambda U}^{\mathrm{T}}$ is symmetric. □

In the above proof, the identity matrix maximizing the trace is didactically obtained in two ways as follows. (i) Note that $\mathrm{tr}(\mathbf{V}^{\mathrm{T}}\mathbf{U\Lambda}) = \sum_{i=1}^{m}\mathbf{v}_i^{\mathrm{T}}\mathbf{u}_i\lambda_i$, where $\mathbf{v}_i$ and $\mathbf{u}_i$ are the i-th columns of $\mathbf{V}$ and $\mathbf{U}$, respectively with $\mathbf{v}_i^{\mathrm{T}}\mathbf{v}_i = \mathbf{u}_i^{\mathrm{T}}\mathbf{u}_i = 1$ $(i = 1, ..., m)$ due to the orthonormality of $\mathbf{V}$ and $\mathbf{U}$; and $\lambda_i > 0$ $(i = 1, ..., m)$ by assumption. Using $\mathbf{v}_i^{\mathrm{T}}\mathbf{u}_i = \mathbf{v}_i^{\mathrm{T}}\mathbf{u}_i/\{(\mathbf{v}_i^{\mathrm{T}}\mathbf{v}_i)^{1/2}(\mathbf{u}_i^{\mathrm{T}}\mathbf{u}_i)^{1/2}\} \le 1$ $(i = 1, ..., m)$ by the Cauchy-Schwarz (C-S) inequality, we have $\mathrm{tr}(\mathbf{V}^{\mathrm{T}}\mathbf{U\Lambda}) = \sum_{i=1}^{m}\mathbf{v}_i^{\mathrm{T}}\mathbf{u}_i\lambda_i \le \sum_{i=1}^{m}\lambda_i$. Since the maximum in the C-S inequality is obtained when $\mathbf{v}_i = \mathbf{u}_i$ $(i = 1, ..., m)$, we have $\mathbf{V}^{\mathrm{T}}\mathbf{U} = \mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}_m$ (the $m \times m$ identity matrix) for maximizing $\mathrm{tr}(\mathbf{V}^{\mathrm{T}}\mathbf{U\Lambda})$.

(ii) The second derivation is given without using the C-S inequality. Since $\mathbf{W} = \{w_{ij}\} \equiv \mathbf{V}^{\mathrm{T}}\mathbf{U}$ is orthonormal as seen from $(\mathbf{V}^{\mathrm{T}}\mathbf{U})^{\mathrm{T}}(\mathbf{V}^{\mathrm{T}}\mathbf{U}) = \mathbf{U}^{\mathrm{T}}\mathbf{V}\mathbf{V}^{\mathrm{T}}\mathbf{U} = \mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{I}_m$, we have $w_{ij} \le 1$ $(i, j = 1, ..., m)$. Consequently, $\mathrm{tr}(\mathbf{V}^{\mathrm{T}}\mathbf{U}\boldsymbol{\Lambda}) = \sum_{i=1}^{m} w_{ii}\lambda_i \le \sum_{i=1}^{m} \lambda_i = \mathrm{tr}(\boldsymbol{\Lambda})$, where the maximum is attained when $w_{ii} = 1$ $(i = 1, ..., m)$, which is the case of $\mathbf{W} = \mathbf{V}^{\mathrm{T}}\mathbf{U} = \mathbf{I}_m$.

(iii) It is of interest to find that (ii) gives the C-S inequality in (i) via Kristof's theorem, which will be addressed later as an application of Kristof's theorem.

Lemma 4 Symmetric AD and DA with diagonal D make A diagonal (Kristof, 1970, (iv) of the proof of Theorem (first version)). *Let* $\mathbf{A} = \{a_{ij}\}$ *and* $\mathbf{D}$ *be* $m \times m$ *matrices with* $\mathbf{D}$ *being diagonal. Suppose that the diagonal elements* $d_i (i = 1, ..., m)$ *of* $\mathbf{D}$ *are nonzero and* $|d_i|$*'s are mutually different. Suppose further that* $\mathbf{AD}$ *and* $\mathbf{DA}$ *are both symmetric. Then,* $\mathbf{A}$ *is diagonal.*

Proof. By assumption, $\mathbf{AD} = (\mathbf{AD})^{\mathrm{T}} = \mathbf{DA}^{\mathrm{T}}$ and $\mathbf{DA} = (\mathbf{DA})^{\mathrm{T}} = \mathbf{A}^{\mathrm{T}}\mathbf{D}$. From the last equation, we have $\mathbf{DAD}^{-1} = \mathbf{A}^{\mathrm{T}}$. Using $\mathbf{A}^{\mathrm{T}} = \mathbf{DAD}^{-1}$ on the right hand-side of the first equation $\mathbf{AD} = \mathbf{DA}^{\mathrm{T}}$ and post-multiplying $\mathbf{D}^{-1}$ on both sides of the equation, we obtain $\mathbf{A} = \mathbf{D}^2\mathbf{AD}^{-2}$, which indicates that

$$a_{ij} = a_{ij}d_i^2/d_j^2 (i, j = 1, ..., m).$$

Since $d_i^2/d_j^2 \ne 1$ when $i \ne j$ by assumption, $a_{ij} = 0$ $(i \ne j)$ follow. □

When $\mathbf{A}$ is diagonal, we have $\mathbf{AD} = \mathbf{DA}$, where $\mathbf{A}$ and $\mathbf{D}$ are said to commute. Using this formulation, a lemma equivalent to Lemma 4 was stated by Kiers and Ten Berge (1989, Lemma 1).

Lemma 5 : Two products of orthonormal and diagonal matrices (a special case of Kristof (1970, (iv) of the proof of Theorem (first version))). *Let* $\mathbf{X}_i$, $\mathbf{D}_i$ *and* $\boldsymbol{\Delta}_i$ *be* $m \times m$ *matrices, where* $\mathbf{X}_i$ *is orthonormal while* $\mathbf{D}_i$ *and* $\boldsymbol{\Delta}_i$ *are diagonal and of full rank* $(i = 1, 2)$*. Suppose that*

$$\mathbf{X}_1\mathbf{D}_1\mathbf{X}_2 = \boldsymbol{\Delta}_1 \quad \textit{and} \quad \mathbf{X}_2\mathbf{D}_2\mathbf{X}_1 = \boldsymbol{\Delta}_2.$$

Then, $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *are the* $m \times m$ *signed and/or permuted identity matrices with* m *nonzero elements being* ± 1*, where permutation indicates row- or column-wise one. When without permutation,* $\mathbf{X}_1$ *and* $\mathbf{X}_2$ *are diagonal matrices with their diagonal elements being* ± 1*, and* $\boldsymbol{\Delta}_1\mathbf{D}_2 = \boldsymbol{\Delta}_2\mathbf{D}_1$*.*

Proof. Left-multiplying $\mathbf{X}_1\mathbf{D}_1$ on both sides of $\mathbf{X}_2\mathbf{D}_2\mathbf{X}_1 = \boldsymbol{\Delta}_2$ using $\mathbf{X}_1\mathbf{D}_1\mathbf{X}_2 = \boldsymbol{\Delta}_1$, we have $\boldsymbol{\Delta}_1\mathbf{D}_2\mathbf{X}_1 = \mathbf{X}_1\mathbf{D}_1\boldsymbol{\Delta}_2$ giving $\boldsymbol{\Delta}_1\mathbf{D}_2 = \mathbf{X}_1\mathbf{D}_1\boldsymbol{\Delta}_2\mathbf{X}_1^{\mathrm{T}}$. The last result shows the spectral decomposition of the diagonal matrix $\boldsymbol{\Delta}_1\mathbf{D}_2$, which indicates that the orthonormal matrix $\mathbf{X}_1$ becomes a signed and/or permuted identity matrix and when without permutation $\boldsymbol{\Delta}_1\mathbf{D}_2 = \boldsymbol{\Delta}_2\mathbf{D}_1$. For $\mathbf{X}_2$, exchanging the subscripts "1" and "2" due to symmetry, we obtain $\boldsymbol{\Delta}_2\mathbf{D}_1 = \mathbf{X}_2\mathbf{D}_2\boldsymbol{\Delta}_1\mathbf{X}_2^{\mathrm{T}}$ indicating the same results as for $\mathbf{X}_1$. □

Theorem 1 : Two-fold or bilinear case ($n = 2$) (von Neumann (1937, Theorem 1); Kristof (1970, Theorem (first version))). *Let $\mathbf{\Gamma}_1^*$ and $\mathbf{\Gamma}_2^*$ be fixed diagonal matrices of full rank with the absolute values of the diagonal elements being mutually different in each matrix. Consider the maximum and minimum of* $\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1^*\mathbf{X}_2\mathbf{\Gamma}_2^*)$ *which are attained, where $\mathbf{X}_1$ and $\mathbf{X}_2$ independently vary over all $m \times m$ orthonormal matrices. Then,*

$$-\mathrm{tr}(\mathbf{\Gamma}_1\mathbf{\Gamma}_2) \leq \mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1^*\mathbf{X}_2\mathbf{\Gamma}_2^*) \leq \mathrm{tr}(\mathbf{\Gamma}_1\mathbf{\Gamma}_2),$$

where $\mathbf{\Gamma}_i = \mathrm{diag}(\gamma_{i1}, ..., \gamma_{im})$ is given by $\mathbf{\Gamma}_i^*$ with their diagonal elements replaced by the corresponding absolute values with possible permutation to have the descending order i.e., $\gamma_{i1} > \cdots > \gamma_{im} > 0$ $(i = 1, 2)$.

Proof. By Lemma 2, since the range of $\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2\mathbf{\Gamma}_2)(= \mathrm{tr}(\mathbf{\Gamma}_2\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2))$ is the same as that of $\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1^*\mathbf{X}_2\mathbf{\Gamma}_2^*)$, we consider the former range. Due to Lemma 3, when the former trace is maximized, $\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2\mathbf{\Gamma}_2$ and similarly $\mathbf{\Gamma}_2\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2$ become symmetric. Then, using Lemma 4 we find that $\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2$ is diagonal. Due to symmetry with $\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2\mathbf{\Gamma}_2) = \mathrm{tr}(\mathbf{\Gamma}_2\mathbf{X}_2\mathbf{\Gamma}_1\mathbf{X}_1)$, $\mathbf{X}_2\mathbf{\Gamma}_2\mathbf{X}_1$ is also found to be diagonal. By Lemma 5, these two diagonal conditions give the maximum $\mathrm{tr}(\mathbf{\Gamma}_1\mathbf{\Gamma}_2)$ of $\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2\mathbf{\Gamma}_2)$ when $\mathbf{X}_1$ and $\mathbf{X}_2$ are the same signed identity matrices. The global maximum $\mathrm{tr}(\mathbf{\Gamma}_1\mathbf{\Gamma}_2)$ among the permuted diagonal elements of $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ is shown by Lemma 1. The minimum is given by replacing e.g., $\mathbf{X}_1$ by $-\mathbf{X}_1$. □

Remark 1 The second simple proof of Theorem 1. (Mirsky, 1975) using an associated property of the doubly stochastic matrix obtained by Fan (1951, Lemma 1A) will be shown with Fan's lemma in the appendix. Note that a doubly stochastic matrix is a square one, where the sum of each row and that of each column are unities. An example is the matrix consisting of the squared elements of an orthonormal matrix. Mirsky's proof has been known in the mathematical community as a short and simple derivation of von Neumann's trace inequality.

Remark 1A In his tutorial, Waller (2018, Equation (21)) explained the result of Theorem 1 using the symmetric condition as given in Lemma 3 with the SVD $\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2\mathbf{\Gamma}_2) = \mathrm{tr}(\mathbf{P\Delta Q}^{\mathrm{T}}) = \mathrm{tr}(\mathbf{PQ}^{\mathrm{T}}\mathbf{\Delta})$, whose optima are attained when $\mathbf{P} = \mathbf{Q}$ or $\mathbf{P} = -\mathbf{Q}$ as $-\mathrm{tr}(\mathbf{\Delta}) \leq \mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2\mathbf{\Gamma}_2) \leq \mathrm{tr}(\mathbf{\Delta})$. This result is correct. Then, Waller (2018, Equation (22)) gave inequalities $-\mathrm{tr}(\mathbf{\Gamma}_1\mathbf{\Gamma}_2) \leq \mathrm{tr}(\mathbf{\Delta}) \leq \mathrm{tr}(\mathbf{\Gamma}_1\mathbf{\Gamma}_2)$ using our notation followed by the statement "the bounds by Kristof's theorem can be achieved". These are also correct. However, the most important result in Theorem 1 is $\mathrm{tr}(\mathbf{\Delta}) = \mathrm{tr}(\mathbf{\Gamma}_1\mathbf{\Gamma}_2)$, whose proof has been shown by using Lemma 5 as well as the second one in the appendix.

3 Didactic proofs of Kristof's theorem

In this section, an independent lemma in linear algebra is provided, which is an extension corresponding to the result in the proof of Kristof (1970). The trilinear case is given as Theorem 3 for didactic purposes, followed by a short proof of Kristof's general theorem using several lemmas.

Lemma 6 : Two products of square, diagonal and orthonormal matrices (an extension of Kristof (1970, (iv) of the proof of Theorem (first version))). *Let* $\mathbf{A}, \mathbf{X}, \mathbf{D}_i$ *and* $\mathbf{\Delta}_i$ *be* $m \times m$ *matrices of full rank, where* $\mathbf{X}$ *is orthonormal while* $\mathbf{D}_i$ *and* $\mathbf{\Delta}_i$ *are diagonal* $(i = 1, 2)$*. Suppose that*

$$\mathbf{A}\mathbf{D}_1\mathbf{X} = \mathbf{\Delta}_1 \quad \textit{and} \quad \mathbf{X}\mathbf{D}_2\mathbf{A} = \mathbf{\Delta}_2.$$

Then, $\mathbf{X}$ *is the* $m \times m$ *signed and/or permuted identity matrices with* m *nonzero elements being* ± 1*, where permutation indicates row- or column-wise one. When without permutation,* $\mathbf{X}$ *is diagonal with its diagonal elements being* ± 1*, and* $\mathbf{\Delta}_1\mathbf{D}_2 = \mathbf{\Delta}_2\mathbf{D}_1$*.*

Proof. Right-multiplying $\mathbf{D}_1\mathbf{X}$ on both sides of $\mathbf{X}\mathbf{D}_2\mathbf{A} = \mathbf{\Delta}_2$ using $\mathbf{A}\mathbf{D}_1\mathbf{X} = \mathbf{\Delta}_1$, we have $\mathbf{X}\mathbf{D}_2\mathbf{\Delta}_1 = \mathbf{\Delta}_2\mathbf{D}_1\mathbf{X}$ giving $\mathbf{\Delta}_2\mathbf{D}_1 = \mathbf{X}\mathbf{\Delta}_1\mathbf{D}_2\mathbf{X}^{\mathrm{T}}$. The last result shows the spectral decomposition of the diagonal matrix $\mathbf{\Delta}_2\mathbf{D}_1$, which indicates that the orthonormal matrix $\mathbf{X}$ becomes a signed and/or permuted identity matrix and when without permutation $\mathbf{\Delta}_1\mathbf{D}_2 = \mathbf{\Delta}_2\mathbf{D}_1$. □

Remark 2 Lemma 5 is seen as a special case of Lemma 6 when $\mathbf{A} = \mathbf{X}_1$ an orthonormal matrix and $\mathbf{X}$ is denoted by $\mathbf{X}_2$. However, in Lemma 5, both $\mathbf{X}_1$ and $\mathbf{X}_2$ were found to be signed and/or permutated identity matrices. Note also that Kristof (1970) dealt with the case when $\mathbf{A} = \mathbf{G}_1\mathbf{\Gamma}_1\mathbf{G}_2\mathbf{\Gamma}_2 \cdots \mathbf{G}_{n-1}\mathbf{\Gamma}_{n-1}\mathbf{G}_n$, $\mathbf{D}_1 = \mathbf{\Gamma}_n$, $\mathbf{D}_2 = \mathbf{\Gamma}_{n+1}$ and $\mathbf{X} = \mathbf{G}_{n+1}$, where $\mathbf{\Gamma}_i$ and $\mathbf{G}_i$ are diagonal and orthonormal matrices, respectively. This specification was necessary for his derivation by induction though the involved expression $\mathbf{G}_1\mathbf{\Gamma}_1\mathbf{G}_2\mathbf{\Gamma}_2 \cdots \mathbf{G}_{n-1}\mathbf{\Gamma}_{n-1}\mathbf{G}_n$ may hide the basic structure in Lemma 6.

Theorem 2 : Three-fold or trilinear case ($n = 3$) (Kristof, 1970, Theorem (first version)). *Let* $\mathbf{\Gamma}_i^*$ $(i = 1, 2, 3)$ *be fixed diagonal matrices of full rank with the absolute values of the diagonal elements being mutually different in each matrix. Consider the maximum and minimum of* $\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1^*\mathbf{X}_2\mathbf{\Gamma}_2^*\mathbf{X}_3\mathbf{\Gamma}_3^*)$ *which are attained, where* $\mathbf{X}_i (i = 1, 2, 3)$ *independently vary over all* $m \times m$ *orthonormal matrices. Then,*

$$-\mathrm{tr}(\mathbf{\Gamma}_1\mathbf{\Gamma}_2\mathbf{\Gamma}_3) \leq \mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1^*\mathbf{X}_2\mathbf{\Gamma}_2^*\mathbf{X}_3\mathbf{\Gamma}_3^*) \leq \mathrm{tr}(\mathbf{\Gamma}_1\mathbf{\Gamma}_2\mathbf{\Gamma}_3),$$

where $\mathbf{\Gamma}_i = \mathrm{diag}(\gamma_{i1}, ..., \gamma_{im})$ *is given by* $\mathbf{\Gamma}_i^*$ *with their diagonal elements replaced by the corresponding absolute values with possible permutation to have the descending order i.e.,* $\gamma_{i1} > \cdots > \gamma_{im} > 0$ $(i = 1, 2, 3)$*.*

Proof. As in the proof for Theorem 1, using Lemma 2 in a similar manner with $\mathbf{\Gamma}_i = \mathbf{P}_i\mathbf{S}_i\mathbf{\Gamma}_i^*\mathbf{P}_i^{\mathrm{T}}$ and unconstrained orthonormal $\mathbf{X}_i (i = 1,\ 2,\ 3)$, the range of $\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2\mathbf{\Gamma}_2\mathbf{X}_3\mathbf{\Gamma}_3)$ $(= \mathrm{tr}(\mathbf{\Gamma}_3\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2\mathbf{\Gamma}_2\mathbf{X}_3))$ is found to be the same as that of $\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1^*\mathbf{X}_2\mathbf{\Gamma}_2^*\mathbf{X}_3\mathbf{\Gamma}_3^*)$. Due to Lemma 3, when the former trace is maximized, $\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2\mathbf{\Gamma}_2\mathbf{X}_3\mathbf{\Gamma}_3 \equiv \mathbf{B}\mathbf{\Gamma}_3$ and similarly $\mathbf{\Gamma}_3\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2\mathbf{\Gamma}_2\mathbf{X}_3 = \mathbf{\Gamma}_3\mathbf{B}$ become symmetric. Then, using Lemma 4 we find that $\mathbf{B} = \mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2\mathbf{\Gamma}_2\mathbf{X}_3 \equiv \mathbf{A}\mathbf{\Gamma}_2\mathbf{X}_3$ is diagonal. Similarly, since $\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2\mathbf{\Gamma}_2\mathbf{X}_3\mathbf{\Gamma}_3) = \mathrm{tr}(\mathbf{A}\mathbf{\Gamma}_2\mathbf{X}_3\mathbf{\Gamma}_3) =$

$\mathrm{tr}(\mathbf{X}_3\mathbf{\Gamma}_3\mathbf{A}\mathbf{\Gamma}_2)$, $\mathbf{X}_3\mathbf{\Gamma}_3\mathbf{A}$ is also diagonal. From Lemma 6, these two diagonal conditions can make $\mathbf{X}_3$ an identity matrix. Then, using Theorem 1, the maximum $\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2\mathbf{\Gamma}_2\mathbf{X}_3\mathbf{\Gamma}_3) = \mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2\mathbf{\Gamma}_2\mathbf{\Gamma}_3)$ is obtained when $\mathbf{X}_1$ and $\mathbf{X}_2$ are identity matrices as $\mathrm{tr}(\mathbf{\Gamma}_1\mathbf{\Gamma}_2\mathbf{\Gamma}_3)$. The minimum $-\mathrm{tr}(\mathbf{\Gamma}_1\mathbf{\Gamma}_2\mathbf{\Gamma}_3)$ is obtained as in Theorem 1. □

Remark 2A In the proof of Theorem 2, the key result is $\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2\mathbf{\Gamma}_2\mathbf{X}_3\mathbf{\Gamma}_3) = \mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2\mathbf{\Gamma}_2\mathbf{\Gamma}_3)$. Since $\mathbf{\Gamma}_2\mathbf{\Gamma}_3 \equiv \mathbf{\Gamma}_{2*3}$ is diagonal, the maximum of $\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2\mathbf{\Gamma}_2\mathbf{\Gamma}_3) = \mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2\mathbf{\Gamma}_{2*3})$ is obtained by Theorem 1 for the bilinear case. This suggests a heuristic proof for the general case, which is employed in the following result.

Theorem 3 : The n-fold case (n = 2, 3,...) (Kristof, 1970, Theorem (first version)). *Let $\mathbf{\Gamma}_i^*$ $(i = 1, ..., n)$ be fixed diagonal matrices. Consider the maximum and minimum of $\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1^* \cdots \mathbf{X}_n\mathbf{\Gamma}_n^*)$ which are attained, where $\mathbf{X}_i(i = 1, ..., n)$ independently vary over all orthonormal matrices. Then,*

$$-\mathrm{tr}(\mathbf{\Gamma}_1 \cdots \mathbf{\Gamma}_n) \le \mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1^* \cdots \mathbf{X}_n\mathbf{\Gamma}_n^*) \le \mathrm{tr}(\mathbf{\Gamma}_1 \cdots \mathbf{\Gamma}_n),$$

where $\mathbf{\Gamma}_i = \mathrm{diag}(\gamma_{i1}, ..., \gamma_{im})$ is given by $\mathbf{\Gamma}_i^*$ with their diagonal elements replaced by the corresponding absolute values with possible permutation to have the weakly descending order i.e., $\gamma_{i1} \ge \cdots \ge \gamma_{im} \ge 0$ $(i = 1, ..., n)$.

Proof. As in Kristof (1970), first suppose that $\gamma_{i1} > \cdots > \gamma_{im} > 0$ $(i = 1, ..., n)$. Using the result of Theorem 2, increase n one by one as n = 4, 5,... When $n = 4$, redefine $\mathbf{B} \equiv \mathbf{X}_1\mathbf{\Gamma}_1 \cdots \mathbf{X}_3\mathbf{\Gamma}_3\mathbf{X}_4 \equiv \mathbf{A}\mathbf{\Gamma}_3\mathbf{X}_4$. Then, from Lemma 4, $\mathbf{B} = \mathbf{A}\mathbf{\Gamma}_3\mathbf{X}_4$ becomes diagonal. Similarly, $\mathbf{X}_4\mathbf{\Gamma}_4\mathbf{A}$ is also diagonal. Then, as before $\mathbf{X}_4$ can become an identity matrix. Using Theorem 2, the maximum of $\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1 \cdots \mathbf{X}_3\mathbf{\Gamma}_3\mathbf{X}_4\mathbf{\Gamma}_4) = \mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1 \cdots \mathbf{X}_3\mathbf{\Gamma}_3\mathbf{\Gamma}_4)$ is given by $\mathrm{tr}(\mathbf{\Gamma}_1 \cdots \mathbf{\Gamma}_4)$. The minimum is similarly obtained as $-\mathrm{tr}(\mathbf{\Gamma}_1 \cdots \mathbf{\Gamma}_4)$. Increasing n successively one by one, we obtain the required results.

Further, consider the weakly ordered case $\gamma_{i1} \ge \cdots \ge \gamma_{im} \ge 0$ $(i = 1, ..., n)$. As in Kristof (1970, (v) of the proof of Theorem (first version) based on the suggestion by Bary G. Wingersky), let $\mathbf{W} = \mathrm{diag}(w_1, ..., w_m)$ with $w_1 > \cdots > w_m > 0$. Redefine $\mathbf{\Gamma}_i$ as $\mathbf{\Gamma}_i + \varepsilon\mathbf{W}(i = 1, ..., n)$ with $\varepsilon > 0$. Then, we have the same above result since $\gamma_{i1} > \cdots > \gamma_{im} > 0$. When ε approaches zero with fixed $\mathbf{X}_i(i = 1, ...n)$, the same required result is given by substituting $\varepsilon = 0$ for $\mathbf{\Gamma}_i + \varepsilon\mathbf{W}$ under the limiting condition of $\gamma_{i1} \ge \cdots \ge \gamma_{im} \ge 0$ $(i = 1, ..., n)$. □

Remark 3 The heuristic derivation of Theorem 3 is essentially equal to that by induction, where the latter was employed by Kristof (1970). The method of successively finding maxima was shown for didactic purposes as well as a direct derivation in Theorem 2 when $n = 2$. Since Theorems 1 and 2 are special cases of Theorem 3, the former results also hold under $\gamma_{i1} \ge \cdots \ge \gamma_{im} \ge 0$. Though when $\gamma_{i1} \ge \cdots \ge \gamma_{im} = 0$, which is given in the limiting case of $\varepsilon = 0$, $\mathbf{\Gamma}_i$ becomes singular while $\mathbf{\Gamma}_i + \varepsilon\mathbf{W}$ with $\varepsilon > 0$ is non-singular, this rank difference does not affect the maximum attained.

4 Generalizations of Kristof's theorem

Ten Berge (1983) gave a generalized version of Kristof's theorem when $\mathbf{X}_i$'s with $\mathrm{rank}(\mathbf{X}_i) \equiv r_i^* \le r_i$ are $m_{i-1} \times m_i$ possibly non-square suborthonormal matrices i.e., submatrices of orthonormal ones ($i = 1, ...n$; $m_0 \equiv m_n$). Let a semiorthonormal matrix be a non-square submatrix of its parent orthonormal one with the same number of the rows or columns (not both) as that of the parent. Note that if $\mathbf{X}_i$ is suborthonormal rather than semi- or fully orthonormal, r_i^* can be 0. For this result, we consider the restricted parent orthonormal matrix taking a block-diagonal form $\mathbf{X}_i^* = \begin{pmatrix} \mathbf{X}_{i11} & \mathbf{X}_{i12} \\ \mathbf{X}_{i21} & \mathbf{X}_{i22} \end{pmatrix} = \begin{pmatrix} \mathbf{X}_{i11} & \mathbf{O} \\ \mathbf{O} & \mathbf{X}_{i22} \end{pmatrix}$, where $\mathbf{X}_i^{*\mathrm{T}}\mathbf{X}_i^* = \mathbf{X}_i^*\mathbf{X}_i^{*\mathrm{T}} = \mathbf{I}_{m_{i-1}+m_i}$, $\mathbf{X}_{i11}^{\mathrm{T}}\mathbf{X}_{i11} = \mathbf{X}_{i11}\mathbf{X}_{i11}^{\mathrm{T}} = \mathbf{I}_{m_{i-1}}$ and $\mathbf{X}_{i22}^{\mathrm{T}}\mathbf{X}_{i22} = \mathbf{X}_{i22}\mathbf{X}_{i22}^{\mathrm{T}} = \mathbf{I}_{m_i}$. Then, when $\mathbf{X}_i$ is one of the two off-block-diagonal zero submatrices $\mathbf{X}_{i12} = \mathbf{X}_{21}^{\mathrm{T}} = \mathbf{O}$, the rank of $\mathbf{X}_i$ is zero i.e., $r_i^* = 0$. It is assumed that $\mathbf{X}_i^*$ varies unrestrictedly under the block-diagonal form. Though Ten Berge did not fully explain the cases with $r_i^* < r_i$, r_i is seen as an upper bound of r_i^*, which is given by $r_i = \min\{m_{i-1}, m_i\}$. Note that $r_i = \min\{m_{i-1}, m_i\}$ is the smallest upper bound when $\mathbf{X}_i$ varies unrestrictedly. In other words, when $r_i < \min\{m_{i-1}, m_i\}$, $\mathbf{X}_i$ does not vary unrestrictedly. If $\mathbf{X}_i$ is semi- or fully orthonormal, we have $r_i^* = r_i = \min\{m_{i-1}, m_i\}$.

Let $\mathbf{\Gamma}_i^*$ and $\mathbf{\Gamma}_i$ be $m_i \times m_i$ fixed diagonal matrices defined as in Theorem 3 with possible different m_i's ($i = 1, ...n$). Let $r = \min(r_1, ..., r_n)$ and $m = \max(m_1, ..., m_n)$. Define $\mathbf{\Delta}_i$ and $\mathbf{E}_r$ as the $m \times m$ diagonal matrices containing $\mathbf{\Gamma}_i$ and $\mathbf{I}_r$ in their upper left corners with zeros elsewhere, respectively. Then, Ten Berge gave the following result.

Theorem 4 : The generalized Kristof theorem (Ten Berge (1983, Theorem 1); see also Kiers and Ten Berge (1989); and Ten Berge (1993, Sections 3.2 and 3.3)). *Under the definitions and assumptions given above, when* $\mathbf{X}_i$ *varies with the condition* $\mathrm{rank}(\mathbf{X}_i) \le r_i$ $(i = 1, ..., n)$*, we have*

$$-\mathrm{tr}(\mathbf{\Delta}_1 \cdots \mathbf{\Delta}_n \mathbf{E}_r) \le \mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1^* \cdots \mathbf{X}_n\mathbf{\Gamma}_n^*) \le \mathrm{tr}(\mathbf{\Delta}_1 \cdots \mathbf{\Delta}_n \mathbf{E}_r).$$

For the proof of Theorem 4, Ten Berge (1983) defined $\mathbf{Y}_i$ as the $m \times m$ matrix containing $\mathbf{X}_i$ in its upper left corner with zeroes elsewhere. Then, he defined its SVD as $\mathbf{Y}_i = \mathbf{P}_i\mathbf{D}_i\mathbf{Q}_i^{\mathrm{T}}$, where "$\mathbf{P}_i$ and $\mathbf{Q}_i$ are orthonormal and $\mathbf{D}_i$ is diagonal" (loc.cit., p. 521). Note that he employed the SVD using the non-negative singular values rather than the positive ones, where $\mathbf{P}_i$, $\mathbf{Q}_i$ and $\mathbf{D}_i$ are $m \times m$ square matrices. This is seen in his inequalities (loc.cit., Equation (10))

$$-\mathrm{tr}(\mathbf{D}_1\mathbf{\Delta}_1 \cdots \mathbf{D}_n\mathbf{\Delta}_n) \le \mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1^* \cdots \mathbf{X}_n\mathbf{\Gamma}_n^*) \le \mathrm{tr}(\mathbf{D}_1\mathbf{\Delta}_1 \cdots \mathbf{D}_n\mathbf{\Delta}_n).$$

For this derivation, he used $\mathbf{Y}_i = \mathbf{P}_i\mathbf{D}_i\mathbf{Q}_i^{\mathrm{T}}$ and $\mathbf{\Delta}_i^*$ defined similarly to $\mathbf{\Delta}_i$ using $\mathbf{\Gamma}_i^*$ $(i = 1, ..., n)$, which gives

$$\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1^* \cdots \mathbf{X}_n\mathbf{\Gamma}_n^*) = \mathrm{tr}(\mathbf{Y}_1\mathbf{\Delta}_1^* \cdots \mathbf{Y}_n\mathbf{\Delta}_n^*) = \mathrm{tr}(\mathbf{P}_1\mathbf{D}_1\mathbf{Q}_1^{\mathrm{T}}\mathbf{\Delta}_1^* \cdots \mathbf{P}_n\mathbf{D}_n\mathbf{Q}_n^{\mathrm{T}}\mathbf{\Delta}_n^*)$$

(loc.cit., Equation (9)). He applied Kristof's theorem to this result supposing that all the $2n$ $m \times m$ matrices $\mathbf{P}_i$ and $\mathbf{Q}_i$ $(i = 1, ..., n)$ vary over all the orthonormal matrices, which is required in Kristof's theorem. Then, he showed his Equation (1) shown earlier.

However, it is found that when $\mathbf{D}_i = \mathrm{diag}(d_1, ..., d_m)$ with $d_1 \geq \cdots \geq d_{r_i^*} > 0$ and $d_{r_i^*+1} = \cdots = d_m = 0$, orthonormal matrices $\mathbf{P}_i$ and $\mathbf{Q}_i$ in $\mathbf{Y}_i = \mathbf{P}_i \mathbf{D}_i \mathbf{Q}_i^{\mathrm{T}}$ should be of the form

$$\mathbf{P}_i = \begin{pmatrix} \mathbf{P}_{i1} & \mathbf{O} \\ \mathbf{O} & \mathbf{P}_{i2} \end{pmatrix} \quad \text{and} \quad \mathbf{Q}_{\mathrm{i}} = \begin{pmatrix} \mathbf{Q}_{i1} & \mathbf{O} \\ \mathbf{O} & \mathbf{Q}_{i2} \end{pmatrix}$$

where $\mathbf{P}_{i1}$ and $\mathbf{Q}_{i1}$ are $r_i^* \times r_i^*$ orthonormal submatrices while $\mathbf{P}_{i2}$ and $\mathbf{Q}_{i2}$ are $(m - r_i^*) \times (m - r_i^*)$ similar ones unless vanishing when $r_i^* = m$. This formulation is due to Ten Berge's special definition of $\mathbf{Y}_i$ whose elements are zero except the upper left submatrix.

Although $\mathbf{P}_i$ and $\mathbf{Q}_i$ are orthonormal rather than semi- or suborthonormal, their block diagonal forms have substantial restrictions in the variations of orthonormal matrices required by Kristof's theorem. One of the severe restrictions is the lack of giving permutations across two sets of variables. Consequently, the upper and lower bounds using Kristof's theorem may not be attained when the diagonal elements of $\mathbf{D}_i \boldsymbol{\Delta}_i$ are not located in the weakly descending order. This restriction was not mentioned by Ten Berge though he did not state that the bounds are attained in his theorem.

The necessity of $\mathbf{E}_r$ in the statement of Theorem 4 is due to the unrestricted rank condition of the diagonal matrix $\boldsymbol{\Gamma}_i$ in the upper left corner of $\boldsymbol{\Delta}_i$ employed by Ten Berge. Since the rank of $\boldsymbol{\Delta}_i$ may be greater than that of $\mathbf{D}_i$, the upper bound in his Equation (10) becomes

$$\begin{aligned} &\mathrm{tr}(\mathbf{X}_1 \boldsymbol{\Gamma}_1^* \cdots \mathbf{X}_n \boldsymbol{\Gamma}_n^*) \leq \mathrm{tr}(\mathbf{D}_1 \boldsymbol{\Delta}_1 \cdots \mathbf{D}_n \boldsymbol{\Delta}_n) \\ &= \mathrm{tr}(\mathbf{D}_1 \cdots \mathbf{D}_n \boldsymbol{\Delta}_1 \cdots \boldsymbol{\Delta}_n) \leq \mathrm{tr}(\boldsymbol{\Delta}_1 \cdots \boldsymbol{\Delta}_n \mathrm{E}_r), \end{aligned}$$

where the last inequality is due to the range [0, 1] of the singular values of suborthonormal matrices (loc.cit., Lemma 2) yielding $\mathbf{D}_1 \cdots \mathbf{D}_n \leq \mathrm{E}_r$ in Löwner's (1934, p. 177) sense. Ten Berge explicitly wrote that 'the statement that "the limits can be attained" has to be omitted' (loc.cit., p. 521). It is to be noted that he added that "the limits ... can be attained if the $\mathbf{X}_i$ are varying independently and (except for the rank) unrestrictedly over the set of suborthonormal matrices" (loc.cit., p. 521).

The meaning of the parenthetical phrase "(except for the rank)" is not clear since when the upper bound r_i of the rank of $\mathbf{X}_i$ (recall the condition $\mathrm{rank}(\mathbf{X}_i) = r_i^* \leq r_i$) is less than $\min\{m_{i-1}, m_i\}$, $\mathbf{X}_i$ does not vary unrestrictedly, but varies over a subset of the suborthonormal matrices satisfying $r_i^* \leq r_i < \min\{m_{i-1}, m_i\}$. That is, in the subset, $\mathbf{X}_i$ cannot be semi- or fully orthonormal. In other words, in this subset the sum of the squared elements in each row or column of $\mathbf{X}_i$ is smaller than 1. Under this restriction, the optima may not be obtained. Note also that Ten Berge mentioned the typical cases with i.e., $r_i^* = r_i$ as "this modification does not affect the validity" (loc.cit., p. 521) of his generalized theorem though the optima may not be attained due to the difficulty of applying Kristof's theorem using constrained parent orthonormal matrices.

In the following modification with attained optima, fully unconstrained suborthonormal matrices are considered. Let $\mathbf{X}_i$ be the $m_{i-1} \times m_i$ $(i = 1, ..., n;\ m_0 \equiv m_n)$ possibly non-square matrix with $\mathrm{rank}(\mathbf{X}_i) = r_i^* \le r_i = \min\{m_{i-1}, m_i\}$, which is supposed to vary unrestrictedly and independently over the set of $m_{i-1} \times m_i$ suborthonormal matrices in the corresponding $m \times m$ parent orthonormal matrix with $m = \max(m_1, ..., m_n)$ as given earlier. The parent orthonormal matrix is denoted by $\mathbf{X}_i^*$, which includes $\mathbf{X}_i$ as a submatrix.

Let $\mathbf{\Gamma}_i^*$ and $\mathbf{\Gamma}_i$ be $m_i \times m_i$ fixed diagonal matrices defined as in Theorems 3 and 4. In the modification, however, $\mathbf{\Gamma}_i^*$ and $\mathbf{\Gamma}_i$ are assumed to be non-singular without loss of generality. This is seen from the form $\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1^*\mathbf{X}_2\mathbf{\Gamma}_2^* \cdots \mathbf{X}_n\mathbf{\Gamma}_n^*)$ to be optimized later, since when $\mathbf{\Gamma}_i^*$ is singular, $\mathbf{\Gamma}_i^*$ can be redefined by deleting the row(s) and column(s) corresponding to the zero diagonal elements of $\mathbf{\Gamma}_i^*$. Then, in the similar manner, the corresponding column(s) of $\mathbf{X}_i$ and row(s) of $\mathbf{X}_{i+1}(\mathbf{X}_{n+1} \equiv \mathbf{X}_1)$ can be deleted without changing the value of $\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1^*\mathbf{X}_2\mathbf{\Gamma}_2^* \cdots \mathbf{X}_n\mathbf{\Gamma}_n^*)$, where $r_i^*(r_{i+1}^*)$ and $r_i(r_{i+1})$ may be adjusted for the reduced $\mathbf{X}_i(\mathbf{X}_{i+1})$ when necessary.

Theorem 5 : A modified generalized Kristof theorem (a modification of Ten Berge (1983, Theorem 1)). *Let $\mathbf{X}_i$, $\mathbf{X}_i^*$, $\mathbf{\Gamma}_i^*$ and $\mathbf{\Gamma}_i$ $(i = 1, ..., n)$ be as defined above. Define $\mathbf{\Delta}_i$ as the diagonal matrix, whose upper left submatrix is $\mathbf{\Gamma}_i$ elsewhere zero, as defined earlier. Then, when the parent orthonormal matrices $\mathbf{X}_i^*$ $(i = 1, ..., n)$ vary independently and unrestrictedly over the set of orthonormal matrices, we have*

$$-\mathrm{tr}(\mathbf{\Delta}_1 \cdots \mathbf{\Delta}_n) \le \mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1^* \cdots \mathbf{X}_n\mathbf{\Gamma}_n^*) \le \mathrm{tr}(\mathbf{\Delta}_1 \cdots \mathbf{\Delta}_n),$$

where the optima are attained.

Proof. Define $\mathbf{\Delta}_i^*$ using $\mathbf{\Gamma}_i^*$ similarly to $\mathbf{\Delta}_i$. Then, we obtain

$$\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1^* \cdots \mathbf{X}_n\mathbf{\Gamma}_n^*) = \mathrm{tr}(\mathbf{X}_1^*\mathbf{\Delta}_1^* \cdots \mathbf{X}_n^*\mathbf{\Delta}_n^*).$$

Noting that the assumption of the independent and unrestricted variations of $\mathbf{X}_i^*$ $(i = 1, ..., n)$ satisfies that of Kristof's theorem, the required results with the attained optima follow. □

Remark 4 In Theorem 5, the assumption for the variations of $\mathbf{X}_i^*$ automatically gives $\mathrm{rank}(\mathbf{X}_i) = r_i^* \le r_i = \min\{m_{i-1}, m_i\}$. The optima are obtained when $r_i^* = r_i$, which indicates that $\mathbf{X}_i$ is semi- or fully orthonormal with non-zero singular value(s) being unity when the optima are attained. This makes the matrix $\mathbf{E}_r$ used in Ten Berge's theorem unnecessary. Recall that the optima may not be obtained in his theorem since the non-zero singular value(s) of $\mathbf{X}_i$ may be less than unity and since the SVD form of $\mathbf{Y}_i$ restricts the permutation of the diagonal elements of $\mathbf{\Gamma}_i^*$ to have $\mathbf{\Gamma}_i$. In other words, when $\mathbf{\Gamma}_i^* = \mathbf{\Gamma}_i$, the latter restriction vanishes. Note that the former restriction corresponds to $r_i^* < r_i$. That is, under this assumption $\mathbf{X}_i$ is suborthonormal (not semi- or fully orthonormal). On the other hand, the case $r_i^* = r_i$, addressed earlier with Ten Berge's statement, indicates that $\mathbf{X}_i$ is semi- or fully orthonormal. Although generally this case does not satisfy the assumption of the unconstrained variation of the parent orthonormal matrix, which is the assumption in Kristof's theorem, the restricted variation also gives the same optima as for the unrestricted case since the non-zero singular value(s) are unities as long as $r_i^* = r_i$. For Ten Berge's generalized theorem, Kiers and Ten Berge (1989, p. 132) stated that "r is the minimum of the ranks of $\mathbf{\Gamma}_1, ..., \mathbf{\Gamma}_k$ and $\mathbf{X}_1, ..., \mathbf{X}_k$", where $k = n$. This is misleading and should be corrected as "r is the minimum of the ranks of $\mathbf{\Gamma}_1, ..., \mathbf{\Gamma}_k$ and $r_1, ..., r_k$" since when $\text{rank}(\mathbf{X}_i) = r_i^* \le r_i$, r_i^* can be smaller than $r_i = \min\{m_{i-1}, m_i\}$ in the subset of the variation of $\mathbf{X}_i$ unless the restricted case of $r_i^* = r_i$ is used.

Remark 4 indicates the following modification of Kristof's theorem.

Theorem 6 : A modification of Kristof's theorem. *In Theorem 3 of Kristof's theorem (first version), redefine the orthonormal matrices* $\mathbf{X}_i (i = 1, ..., n)$ *as*

$$\mathbf{X}_i = \text{Bdiag}(\mathbf{X}_{i1}, ..., \mathbf{X}_{ii_\text{B}}),$$

where $\mathbf{X}_{ij}$ *is the* $i_j \times i_j$ *diagonal block* $(j = 1, ..., i_\text{B})$ *with* $i_1 + ... + i_{i_\text{B}} = m$. *Suppose that* $\mathbf{X}_i (i = 1, ..., n)$ *independently and unrestrictedly vary with* $\text{rank}(\mathbf{X}_{ij}) = i_j$ $(j = 1, ..., i_\text{B})$. *Use* $\mathbf{\Gamma}_i (i = 1, ..., n)$ *as defined in Kristof's theorem. Then, we have*

$$-\text{tr}(\mathbf{\Gamma}_1 \cdots \mathbf{\Gamma}_n) \le \text{tr}(\mathbf{X}_1 \mathbf{\Gamma}_1 \cdots \mathbf{X}_n \mathbf{\Gamma}_n) \le \text{tr}(\mathbf{\Gamma}_1 \cdots \mathbf{\Gamma}_n),$$

where the optima are attained.

Proof. The case when $\mathbf{\Gamma}_i^* = \mathbf{\Gamma}_i (i = 1, ..., n)$ can be used in Kristof's theorem. Although $\mathbf{X}_i (i = 1, ..., n)$ take the block-diagonal forms, the optima with $\mathbf{X}_i = \mathbf{I}_m$ $(i = 1, ..., n)$ are attained in the subset of varying $\mathbf{X}_i (i = 1, ..., n)$ under the block diagonal restriction. □

The usage of $\mathbf{\Gamma}_i$ in place of $\mathbf{\Gamma}_i^* (i = 1, ..., n)$ is due to the lack of permutation across the different diagonal blocks when using the block diagonal $\mathbf{X}_i$. Note that Kiers and Ten Berge (1989) used the assumptions that $\mathbf{\Gamma}_i$'s are available when $n = 2$ stating that "in most applications the assumptions are satisfied, if they are not satisfied only minor modifications will typically be involved" (pp. 126-127). Note also that when $i_\text{B} = 1$, $\mathbf{X}_i$ becomes a usual orthonormal matrix as in Kistof's theorem, and when $i_\text{B} = m$, $\mathbf{X}_i$ is the signed identity matrix whose diagonal elements are ± 1, where $\mathbf{X}_i$ and its diagonal elements i.e., ± 1 are $m \times m$ and 1×1 orthonormal matrices with unit singular value(s), respectively.

An extension in Theorems 4 and 5, is to relax $\mathbf{\Gamma}_i$ to be an unconstrained square matrix of the same rank as that of $\mathbf{\Gamma}_i$ $(i = 1, ..., n)$, which is mathematically immaterial. Since under the relaxed condition the SVD of $\mathbf{\Gamma}_i \equiv \mathbf{U}_i\mathbf{\Lambda}_i\mathbf{V}_i^{\mathrm{T}}$ with $\mathbf{U}_i^{\mathrm{T}}\mathbf{U}_i$ and $\mathbf{V}_i^{\mathrm{T}}\mathbf{V}_i$ being an identity matrix of the same order as rank$(\mathbf{\Gamma}_i)$ $(i = 1, ..., n)$, redefine $\mathbf{X}_1$ and $\mathbf{X}_i$ as $\mathbf{V}_n\mathbf{X}_1\mathbf{U}_1^{\mathrm{T}}$ and $\mathbf{V}_{i-1}\mathbf{X}_i\mathbf{U}_i^{\mathrm{T}}$ $(i = 2, ..., n)$. Then, the problem becomes the same as that using the diagonal matrix $\mathbf{\Lambda}_i$ $(i = 1, ...n)$. Note that this unconstrained condition was used in von Neumann's (1937) trace inequality and Kristof's (1970) Theorem (second version). Kristof (1970, p. 523) stated that "A distinction of the two versions will not be emphasized", where the second version uses the unconstrained square matrix $\mathbf{\Gamma}_i$.

5 Applications of Kristof's theorem and its generalizations

While as mentioned in the introductory section, "underutilization" of Kristof's theorem and its generalization seem to be still true considering its simplicity, generality and tractability yielding solutions in various applications. In this section, basic or important cases in multivariate analysis showing advantages of Kristof's theorem and its generalizations are provided. Examples or applications are shown below mostly in the chronological order. Although in Ten Berge's generalized Kristof's theorem, the optima may not be attained, all his three examples in multivariate analysis and an illustration of the Cauchy-Schwarz inequality using his theorem are the cases with attained optima.

Maximization of $\mathrm{tr}(\mathbf{M\Lambda L})$ or $\mathrm{tr}(\mathbf{A}^{\mathrm{T}}\mathbf{\Lambda})$: Green Jr (1969, Appendix B) (n = 1). In this problem, $\mathbf{M}, \mathbf{L}$ and $\mathbf{A}$ are fixed $m \times r$, $s \times m$ and $r \times s$ $(r \geq s)$ matrices, respectively while $\mathbf{\Lambda}$ is the $r \times s$ matrix varying over all the semiorthonormal matrices when $r > s$ or orthonormal when $r = s$. This is seen as an extended application of von Neumann's trace inequality when $n = 1$ with unconstrained and possibly rectangular $\mathbf{A} = (\mathbf{LM})^{\mathrm{T}}$. An application of this problem to have an optimal linear combination with a specified correlation matrix was investigated. This problem will also be addressed later.

Orthogonal procrustes transformation: Kristof's (1970, Example 1) (n = 1). This problem is to minimize $\mathrm{tr}\{(\mathbf{A} - \mathbf{BT})(\mathbf{A} - \mathbf{BT})^{\mathrm{T}}\}$, where $\mathbf{A}$ and $\mathbf{B}$ are fixed $r \times s$ $(r \geq s)$ matrices while $\mathbf{T}$ is an orthonormal matrix to be derived, which reduces to maximizing $\mathrm{tr}(\mathbf{A}^{\mathrm{T}}\mathbf{BT})$, a case with $n = 1$. Note that $\mathbf{T}$ gives permutations with sign changes (reflections) of the columns of $\mathbf{B}$ as well as the rotation of $\mathbf{B}$. Although the term "procrustes rotation" is usually used as stated by Kristof (1970, p. 523) especially in factor analysis using factor rotation, it is important that $\mathbf{T}$ can give permutations and reflections of the columns of $\mathbf{B}$.

On the other hand, we have a problem of "procrustes transformation" without rotation. This happens in e.g., simulations when $\mathbf{B}$ is one of sample loading matrices rotated by a fixed method e.g., the geomin, which is to be matched to the population geomin-rotated $\mathbf{A}$ to see the sampling variation of $\mathbf{B}$, where only permutations and reflections of the columns of $\mathbf{B}$ are possible while unconstrained rotation is not allowed since otherwise $\mathbf{BT}$ no longer becomes a geomin-rotated loading matrix. This problem is seen as a subproblem of Kristof's theorem when $n = 1$ using a constrained orthonormal matrix $\mathbf{T}$ considering only permutations and reflections of the columns of $\mathbf{B}$. Problems using constrained orthonormal matrices are included in Ten Berge's (1983) generalized Kristof theorem, as addressed earlier.

At the end of Example 1, Kristof (1970, p. 524) stated that "The generalization of the present problem to allow $\mathbf{A}^{\mathrm{T}}\mathbf{B}$ to be singular is immediate and does not require special discussion." It is found that the singular case with $\mathrm{rank}(\mathbf{A}^{\mathrm{T}}\mathbf{B}) \equiv s^* < s \leq r$ gives the s^* positive singular values, say, $\gamma_1 \geq \cdots \geq \gamma_{s*} > 0$. Then, using the SVD $\mathbf{A}^{\mathrm{T}}\mathbf{B} = \mathbf{U}\,\mathrm{diag}(\gamma_1, ..., \gamma_{s*})\mathbf{V}^{\mathrm{T}}$, we obtain

$$\begin{aligned}&\max\{\mathrm{tr}(\mathbf{A}^{\mathrm{T}}\mathbf{BT})\} = \max[\mathrm{tr}\{\mathbf{U}\,\mathrm{diag}(\gamma_1, ..., \gamma_{s*})\mathbf{V}^{\mathrm{T}}\mathbf{T}\}]\\&= \max[\mathrm{tr}\{\,\mathrm{diag}(\gamma_1, ..., \gamma_{s*})\mathbf{V}^{\mathrm{T}}\mathbf{TU}\}] = \mathrm{tr}\{\,\mathrm{diag}(\gamma_1, ..., \gamma_{s*})\}\\&= \textstyle\sum_{i=1}^{s*} \gamma_i,\end{aligned}$$

where $\mathbf{V}^{\mathrm{T}}\mathbf{TU}$ is a suborthonormal matrix since the product of suborthonormal matrices is suborthonormal (Ten Berge, 1983, Lemma 4). Actually, $\mathbf{U}$ and $\mathbf{V}$ of full column rank by definition are semiorthonormal (Ten Berge, 1983, Definition 2) while $\mathbf{T}$ is orthonormal as well as suborthonormal.

Kristof (1970, p. 524) added three examples with $n = 1$ ("the trivial case of the theorem" in his terminology) for determinations of e.g., orthogonal matrices with specific properties developed in the 1960s in psychometrics.

The two-sided orthogonal procrustes problem: Kristof's (1970, Example 2) ($n = 2$). This problem based on Schönemann (1968) minimizes $\mathrm{tr}\{(\mathbf{B} - \mathbf{T}^{\mathrm{T}}\mathbf{AS})(\mathbf{B} - \mathbf{T}^{\mathrm{T}}\mathbf{AS})^{\mathrm{T}}\}$, where $\mathbf{A}$ and $\mathbf{B}$ are fixed square matrices while $\mathbf{T}$ and $\mathbf{S}$ are orthonormal matrices such that $\mathbf{A}$ is to be matched to $\mathbf{B}$. This problem reduces to maximizing $\mathrm{tr}(\mathbf{T}^{\mathrm{T}}\mathbf{ASB}^{\mathrm{T}})$, a case with unconstrained $\mathbf{T}$ and $\mathbf{S}$ when $n = 2$. The maximum of $\mathrm{tr}(\mathbf{T}^{\mathrm{T}}\mathbf{ASB}^{\mathrm{T}})$ is obtained as the product of the positive singular values of $\mathbf{A}$ and $\mathbf{B}$.

Multivariate multiple regression: Kristof's (1970, Example 4) and Ten Berge's (1983, Application 1) ($n = 2$). Kristof's example included an unnatural restriction of the same numbers of the dependent and independent variables, which was removed by Ten Berge's application. The problem is to minimize $\mathrm{tr}\{(\mathbf{Y} - \mathbf{XB})(\mathbf{Y} - \mathbf{XB})^{\mathrm{T}}\}$, where $\mathbf{Y}$ is an $s \times u$ matrix for u dependent variables and $\mathbf{X}$ of full column rank is an matrix for t independent variables. The well-known solution $\hat{\mathbf{B}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$ was obtained without calculus when $n = 2$ though the method is somewhat tedious.

Principal components analysis: Ten Berge's (1983, Application 2) ($n = 1$). This application deals with the jointly determining p principal components when $\mathbf{R}$ is a $k \times k$ correlation matrix of rank $r \leq k$. The problem is to maximize the sum of squared loadings $\mathrm{tr}(\mathbf{B}^{\mathrm{T}}\mathbf{R}^2\mathbf{B})$, where $\mathbf{B}$ is a $k \times p$ loading matrix with $p \leq r$ subject to the uncorrelated components with unit variances $\mathbf{B}^{\mathrm{T}}\mathbf{R}\mathbf{B} = \mathbf{I}_p$. Note that an atypical assumption of the possibly singular $\mathbf{R}$ is used.

Let $\mathbf{R} = \mathbf{K}\boldsymbol{\Lambda}\mathbf{K}^{\mathrm{T}}$ with $\mathbf{K}^{\mathrm{T}}\mathbf{K} = \mathbf{I}_r$ be the SVD using the positive diagonal elements of the $r \times r$ diagonal matrix $\boldsymbol{\Lambda}$. The solution can be given by maximizing

$$\begin{aligned}\mathrm{tr}(\mathbf{B}^{\mathrm{T}}\mathbf{R}^2\mathbf{B}) &= \mathrm{tr}(\mathbf{B}^{\mathrm{T}}\mathbf{K}\boldsymbol{\Lambda}^2\mathbf{K}^{\mathrm{T}}\mathbf{B}) = \mathrm{tr}\{(\mathbf{B}^{\mathrm{T}}\mathbf{K}\boldsymbol{\Lambda}^{1/2})\boldsymbol{\Lambda}(\boldsymbol{\Lambda}^{1/2}\mathbf{K}^{\mathrm{T}}\mathbf{B})\} \\ &= \mathrm{tr}\{(\boldsymbol{\Lambda}^{1/2}\mathbf{K}^{\mathrm{T}}\mathbf{B})(\mathbf{B}^{\mathrm{T}}\mathbf{K}\boldsymbol{\Lambda}^{1/2})\boldsymbol{\Lambda}\}\end{aligned}$$

with $n = 1$, where $(\boldsymbol{\Lambda}^{1/2}\mathbf{K}^{\mathrm{T}}\mathbf{B})(\mathbf{B}^{\mathrm{T}}\mathbf{K}\boldsymbol{\Lambda}^{1/2})$ is suborthonormal since

$$\mathbf{I}_p = \mathbf{B}^{\mathrm{T}}\mathbf{R}\mathbf{B} = \mathbf{B}^{\mathrm{T}}\mathbf{K}\boldsymbol{\Lambda}\mathbf{K}^{\mathrm{T}}\mathbf{B} = (\mathbf{B}^{\mathrm{T}}\mathbf{K}\boldsymbol{\Lambda}^{1/2})(\boldsymbol{\Lambda}^{1/2}\mathbf{K}^{\mathrm{T}}\mathbf{B}).$$

Noting that $\mathbf{B}^{\mathrm{T}}\mathbf{K}\boldsymbol{\Lambda}^{1/2}$ is a semiorthonormal matrix, Ten Berge's generalized Kristof theorem gives $\max\{\mathrm{tr}(\mathbf{B}^{\mathrm{T}}\mathbf{R}^2\mathbf{B})\} = \sum_{i=1}^{p} \lambda_i$, where $\lambda_1 \geq \cdots \geq \lambda_r > 0$ and the unrotated loading matrix $\mathbf{B}$ is a submatrix of $\mathbf{K}\boldsymbol{\Lambda}^{-1/2}$ taking its first p columns to give $\mathbf{B}^{\mathrm{T}}\mathbf{K}\boldsymbol{\Lambda}^{1/2}$ a submatrix of $\mathbf{I}_r$ consisting its first p rows.

The formulation using suborthonormal matrices $\mathrm{tr}(\mathbf{B}^{\mathrm{T}}\mathbf{R}^2\mathbf{B})$ is of interest since the restriction $\mathbf{B}^{\mathrm{T}}\mathbf{R}\mathbf{B} = \mathbf{I}_p$ is cleverly used in the maximizing function without calculus or Lagrange multipliers. While the above example was employed by Ten Berge to show an application of his generalized Kristof theorem, when all the k components including minor ones are obtained in the usual non-singular case of $\mathbf{R}$, the maximum of $\mathrm{tr}(\mathbf{B}^{\mathrm{T}}\mathbf{R}^2\mathbf{B})$ is obtained by Kristof's theorem as $\mathrm{tr}(\boldsymbol{\Lambda})$, where $\boldsymbol{\Lambda}$ is the $k \times k$ diagonal matrix with positive diagonals. In this case, $\boldsymbol{\Lambda}^{1/2}\mathbf{K}^{\mathrm{T}}\mathbf{B}$ becomes $\mathbf{I}_k$ yielding the well-known unrotated loading matrix $\mathbf{B} = \mathbf{K}\boldsymbol{\Lambda}^{-1/2}$.

Canonical correlations: Kristof (1970, General comment (a)) ($n = 1$), Ten Berge (1983, Application 3) ($n = 1$), Ogasawara (2000, with errata, Theorem 1) ($n = 2$), and Waller (2018, pp. 195-196) ($n = 1$). Kristof (1970) suggested a formulation of canonical correlations applying his theorem when $n = 1$, which was fully described by Ten Berge (1983) using an associated SVD as in principal components. Ogasawara (2000) used Kristof's theorem with $n = 2$ to have a lower bound of the mean squared canonical correlations between factors in factor analysis and the corresponding principal components. Waller also showed an application of Kristof's theorem for canonical correlations using two sets of principal components in the two sets of original variables. However, his results are those when the numbers of original variables in each set are the same. Note that use of the principal components needs justification when the numbers of the original variables are not equal or the number of the canonical correlations is less than the minimum of the numbers of the original variables.

The Cauchy-Schwarz inequality: Ten Berge (1983, p. 522) ($n = 2$); Paragraph (iii) after Lemma 3 ($n = 1$) and Theorem 5 of the current

article ($n = 2$). This example is simple but impressive in that the example well shows the simplicity and generality of Kristof's theorem and its extensions without calculus. Let $\mathbf{x}$ and $\mathbf{y}$ be non-zero vectors of order k. Suppose that $\mathbf{x}$ and $\mathbf{y}$ vary independently and unrestrictedly. Then,

$$\begin{aligned}\mathbf{x}^{\mathrm{T}}(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1/2}\mathbf{y}(\mathbf{y}^{\mathrm{T}}\mathbf{y})^{-1/2} &= \mathbf{x}^{\mathrm{T}}(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1/2}\mathbf{I}_k\mathbf{y}(\mathbf{y}^{\mathrm{T}}\mathbf{y})^{-1/2}1 \\ &= \mathbf{x}^{\mathrm{T}}(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1/2}\boldsymbol{\Gamma}_1^*\mathbf{y}(\mathbf{y}^{\mathrm{T}}\mathbf{y})^{-1/2}\boldsymbol{\Gamma}_2^*,\end{aligned}$$

where $\boldsymbol{\Gamma}_1^* = \boldsymbol{\Gamma}_1 = \mathbf{I}_k$ and $\boldsymbol{\Gamma}_2^* = \boldsymbol{\Gamma}_2 = 1$. Define $k \times k$ diagonal matrices $\boldsymbol{\Delta}_1 = \boldsymbol{\Gamma}_1 = \mathbf{I}_k$ and $\boldsymbol{\Delta}_2 = \mathbf{E}_{11}$, where the first diagonal element of $\mathbf{E}_{11}$ is unity elsewhere zero. Since the vectors $\mathbf{x}^{\mathrm{T}}(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1/2}$ and $\mathbf{y}^{\mathrm{T}}(\mathbf{y}^{\mathrm{T}}\mathbf{y})^{-1/2}$ are semiorthonormal, their non-zero singular values are unity. Consequently, when applying Ten Berge's theorem, the maximum is attained as $\mathrm{tr}(\boldsymbol{\Delta}_1\boldsymbol{\Delta}_2) = \mathrm{tr}(\mathbf{I}_k\mathbf{E}_{11}) = 1$ with minimum obtained similarly. Then, we have the Cauchy-Schwarz inequality $-1 \le \mathbf{x}^{\mathrm{T}}(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1/2}\mathbf{y}(\mathbf{y}^{\mathrm{T}}\mathbf{y})^{-1/2} = \frac{\mathbf{x}^{\mathrm{T}}\mathbf{y}}{(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{1/2}(\mathbf{y}^{\mathrm{T}}\mathbf{y})^{1/2}} \le 1$.

As addressed in Paragraph (iii) after Lemma 3, the inequality is given via Kristof's theorem ($n = 1$). The same result is obtained by using Theorem 5 of this article. Define the $k \times k$ parent orthonormal matrices $\mathbf{X}$ and $\mathbf{Y}$, whose first rows are $\mathbf{x}^{\mathrm{T}}(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1/2}$ and $\mathbf{y}^{\mathrm{T}}(\mathbf{y}^{\mathrm{T}}\mathbf{y})^{-1/2}$, respectively, where $\mathbf{X}$ and $\mathbf{Y}$ independently and unrestrictedly vary over the sets of orthonormal matrices. Then, using Theorem 5, the maximum of $\mathbf{x}^{\mathrm{T}}(\mathbf{x}^{\mathrm{T}}\mathbf{x})^{-1/2}\mathbf{y}(\mathbf{y}^{\mathrm{T}}\mathbf{y})^{-1/2}$ is attained as $\mathrm{tr}(\mathbf{X}\mathbf{I}_k\mathbf{Y}\mathbf{E}_{11}) \le \mathrm{tr}(\mathbf{I}_k\mathbf{E}_{11}) = 1$ with the minimum $\mathrm{tr}(\mathbf{X}\mathbf{I}_k\mathbf{Y}\mathbf{E}_{11}) \ge -\mathrm{tr}(\mathbf{I}_k\mathbf{E}_{11}) = -1$ obtained similarly.

Generalized linear form: Ten Berge (1993, Equation (48)), Yanai and Takane (2007, Property 11) and Adachi (2020, Theorem A.4.2) ($n = 1$). This is a problem maximizing $\mathrm{tr}(\mathbf{X}^{\mathrm{T}}\mathbf{A})$, where $\mathbf{X}$ and fixed $\mathbf{A}$ are $p \times q$ ($p \ge q$) matrices with the constraint $\mathbf{X}^{\mathrm{T}}\mathbf{X} = \mathbf{I}_q$ and $\mathrm{rank}(\mathbf{A}) \le q$. To the author's knowledge, this problem was first solved by Green Jr (1969) as mentioned in the first example. After Kristof (1970), Ten Berge (1993) reformulated the problem as the generalized linear form. He defined the SVD $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^{\mathrm{T}}$ with $\mathbf{U}^{\mathrm{T}}\mathbf{U} = \mathbf{V}^{\mathrm{T}}\mathbf{V} = \mathbf{V}\mathbf{V}^{\mathrm{T}} = \mathbf{I}_q$ and $\boldsymbol{\Lambda}$ being the diagonal matrix with the non-negative diagonals using an application of his generalized Kristof theorem with $n = 1$. Then, the maximum is given by

$$\max\{\mathrm{tr}(\mathbf{X}^{\mathrm{T}}\mathbf{A})\} = \max\{\mathrm{tr}(\mathbf{X}^{\mathrm{T}}\mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^{\mathrm{T}})\} = \max\{\mathrm{tr}(\mathbf{V}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{U}\boldsymbol{\Lambda})\} = \mathrm{tr}(\boldsymbol{\Lambda})$$

since $\mathbf{V}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{U}$ is suborthonormal. The maximum is attained when $\mathbf{V}^{\mathrm{T}}\mathbf{X}^{\mathrm{T}}\mathbf{U} = \mathbf{I}_q$ or $\mathbf{X} = \mathbf{U}\mathbf{V}^{\mathrm{T}}$.

Note that the definition of the SVD using the non-negative rather than positive singular values is important considering the case $\mathrm{rank}(\mathbf{A}) \equiv q^* < q$. That is, in the last case when using only positive singular values, $\mathbf{U}$ and $\mathbf{V}$ become $p \times q^*$ and $q \times q^*$ semiorthonormal matrices, respectively, yielding $\mathbf{X}^{\mathrm{T}}\mathbf{X} \ne \mathbf{I}_q$, which does not satisfy the assumption. Note that Green Jr (1969, p. 317) correctly considered the two cases $q^* < q$ and $q^* = q$ as well as the cases of multiple or equal positive singular values in terms of the uniqueness of $\mathbf{U}$, $\mathbf{V}$ and $\mathbf{U}\mathbf{V}^{\mathrm{T}}$. For this example, Neudecker's (2004, Section 2) derivation as "a Kristof-type theorem" with a correction and added explanation will be shown in the appendix.

6 Discussion

(a) The trivial case ($n = 1$) and the bilinear case i.e., von Neumann's trace inequality ($n = 2$). In the previous section, only the examples of $n = 1$ or 2 are shown. Though Kristof (1970) used the term "trivial case" when $n = 1$, its applications are meaningful ones as shown earlier. Note that only the derivation of Kristof's theorem is trivial or self-evident when $n = 1$. A case of $n = 4$ was provided by Kristof (1970, Example 6) as a generalization of Meredith's (1964) problem for a multivariate selection of subpopulations from a common parent. However, most of the applications of Kristof's theorem and Ten Berge's generalized one seem to be those of $n = 1$ or 2. Kiers and Ten Berge (1989, p. 126) stated that "All practical applications we have encountered so far apply to the cases $k = 1$ or $k = 2$", where k is used for n. That said, it is to be noted that Ten Berge (1983, p. 509) stated that "Theorems should be derived in the greatest possible generality".

(b) Alternative proofs. Proofs in seminal papers tend to be complicated. After the discoveries, alternative simple or short proofs follow. For von Neumann's (1937) trace inequality, the elementary alternative proof by Mirsky (1975) with Fan's (1951) lemma may be the simplest one as shown in the appendix. In this tutorial, rephrasing or breaking down the proof by Kristof for his theorem has been shown. However, the logic is essentially the same as Kristof's one using induction. Marshall, Olkin, and Arnold (2011, Chapter 20, Theorem B.2) also showed a similar proof by induction though they stated that "We give an inductive proof that is elementary, though still somewhat lengthy" (p. 791). Finding alternative simple, self-contained and hopefully short proofs of Kristof's theorem when $n \geq 3$ is an open problem. Mirsky used the doubly stochastic matrix. Applications or generalizations of Mirsky's proof to the inequalities when $n \geq 3$ seem to be difficult as far as the author conjectures.

(c) Equivalent and inequivalent cases of the Kristof and Ten Berge theorems. As mentioned earlier, Ten Berge (1983, Theorem 2) extended Kristof's theorem. For the differences of the theorems, he stated that "The most striking difference is that the $\mathbf{X}_i$ are no longer required to be orthonormal. Second, the $\mathbf{X}_i$ need no longer to be square" (p. 521), which are advantages of Ten Berge's theorem over Kristof's one claimed by Ten Berge. The third difference is the lack of the attained optima, which is not an advantage. The two claimed advantages may be handled by Kristof's theorem by considering the parent orthonormal matrices as used by Ten Berge and Theorem 5 in this article. So, the two theorems may be seen as equivalent when the optima are attained. Note that all the four examples in Ten Berge (1983) are the cases of attained optima. Probably, the cases of unattained optima due to $\mathrm{rank}(\mathbf{X}_i) = r_i^* < r_i = \min\{m_{i-1}, m_i\}$ may be theoretical or special, if any, in practice.

It is conjectured that even in this special case, some adjustment giving $r_i^* \leq r_i$ may be obtained. For instance, consider canonical correlation analysis for two sets of standardized data matrices i.e., $\mathbf{X}_1$ $(n \times r_1)$ of rank r_1^* and $\mathbf{X}_2$ $(n \times r_2)$ of rank r_2^*. Then, when $r^* < \min\{r_1^*, r_2^*\}$ canonical correlations are optimally derived in a least squares sense among $\min\{r_1^*, r_2^*\}$ possible ones, this seems to yield a similar problem. Actually, as Ten Berge (1983, p. 523) formulated the situation using the coefficient matrices $\mathbf{B}_1$ $(r_1 \times r^*)$ and $\mathbf{B}_2$ $(r_2 \times r^*)$ with other ones, the maximum of $\mathrm{tr}(\mathbf{B}_1^{\mathrm{T}}\mathbf{X}_1^{\mathrm{T}}\mathbf{X}_2\mathbf{B}_2)$ was attained.

References

Adachi, K. (2020). *Matrix-based introduction to multivariate data analysis* (2nd ed.). Singapore: Springer Singapore.

Fan, K. (1951). Maximum properties and inequalities for the eigenvalues of completely continuous operators. *Proceedings of the National Academy of Sciences*, *37*(11), 760–766.

Green Jr, B. F. (1969). Best linear composites with a specified structure. *Psychometrika*, *34*(3), 301–318.

Hardy, G. H., Littlewood, J. E., & Pólya, G. (1934). *Inequalities*. New York: Cambridge University Press.

Hardy, G. H., Littlewood, J. E., & Pólya, G. (1952). *Inequalities* (2nd ed). New York: Cambridge University Press. (Reprinted 1994)

Kiers, H. A. L., & Ten Berge, J. M. F. (1989). Optimality conditions for the trace of certain matrix products. *Linear Algebra and its Applications*, *126*, 125–134.

Kristof, W. (1964). Die beste orthogonale transformation zur gegenseitigen Überführung zweier faktorenmatrizen. *Diagnostica*, *10*, 87–90.

Kristof, W. (1969). *A theorem on the trace of certain matrix products and some applications.* ETS Research Bulletin Series, RB-69-21. Educational Testing Service: Princeton, NJ. doi: https://doi.org/https://doi.org/10.1002/j.2333-8504.1969.tb00400

Kristof, W. (1970). A theorem on the trace of certain matrix products and some applications. *Journal of Mathematical Psychology*, *7*(3), 515–530.

Levin, J. (1979). Applications of a theorem on the traces of certain matrix products. *Multivariate Behavioral Research*, *14*(1), 109–113.

Löwner, C. (1934). Über monotone matrixfunctionen. *Mathematische Zeitschrif*, *38*, 177–216.

Marshall, A. W., Olkin, I., & Arnold, B. C. (2011). *Inequalities: theory of majorization and its applications* (2nd ed.). New York: Springer.

Meredith, W. (1964). Rotation to achieve factorial invariance. *Psychometrika*, *29*(2), 187–206.

Miranda, H., & Thompson, R. C. (1993). A trace inequality with a subtracted term. *Linear Algebra and its Applications*, *185*, 165–172.

Mirsky, L. (1975). A trace inequality of john von neumann. *Monatshefte für mathematik*, *79*(4), 303–306.

Mori, K., & Kurata, H. (2013). Optimal correlation preserving linear predictors of factor scores in factor analysis. *Journal of the Japan Statistical Society*, *43*(1), 79–89.

Neudecker, H. (2004). On best affine unbiased covariance-preserving prediction of factor scores. *SORT*, *28*(1), 27–36.

Ogasawara, H. (2000). Some relationships between factors and components. *Psychometrika*, *65*, 167–185. (errata, *65*, 551)

Schönemann, P. H. (1968). On two-sided orthogonal procrustes problems. *Psychometrika*, *33*(1), 19–33.

Simon, B. (2005). *Trace ideals and their applications* (2nd ed.) (No. 120). Providence, RI: American Mathematical Society.

Ten Berge, J. M. (1983). A generalization of kristof's theorem on the trace of certain matrix products. *Psychometrika*, *48*(4), 519–523.

Ten Berge, J. M. (1993). *Least squares optimization in multivariate analysis.* Leiden, The Netherlands: DSWO Press, Leiden University. (pdf version with minor corrections, 2005, https://kiers.webhosting.rug.nl/leastsquaresbook.pdf)

Uchida, T. (2023). *von Neumann's trace inequality.* Unpublished document available at https://note.com/utaka233/n/n511cbf78f89d. (in Japanese).

von Neumann, J. (1937). Some matrix inequalities and metrization of matric-space. *Tomsk University Review*, *1*, 286–300. Reprinted in A. H. Taub (Ed.), *John von Newmann collected works* (1962), Vol. IV, pp. 205-219. New York: Pergamon Press.

Waller, N. (2018). An introduction to Kristof's theorem for solving least-square optimization problems without calculus. *Multivariate Behavioral Research*, *53*(2), 190–198.

Yanai, H., & Takane, Y. (2007). Matrix methods and their applications to factor analysis. In S.-Y. Lee (Ed.), *Handbook of latent variable and related models* (pp. 345–366). Amsterdam: North-Holland.

Appendix A Technical results

Lemma A: Fan's (1951, Lemma 1A) inequality for the doubly substochastic matrix. *Let* $\mathbf{a} = (a_1, ..., a_n)^{\mathrm{T}}$ *and* $\mathbf{b} = (b_1, ..., b_n)^{\mathrm{T}}$ *be fixed vectors with* $a_1 \geq \cdots \geq a_n \geq 0$ *and* $b_1 \geq \cdots \geq b_n \geq 0$*. Define an* $n \times n$ *doubly substochastic matrix* $\mathbf{P} = \{p_{ij}\}$ *with non-negative elements satisfying* $\sum_{j=1}^{n} p_{ij} \leq 1$ $(i = 1, ..., n)$ *and* $\sum_{i=1}^{n} p_{ij} \leq 1$ $(j = 1, ..., n)$ *(see e.g., Marshall et al. (2011, Section 2.C)). Then,* $\mathbf{a}^{\mathrm{T}}\mathbf{P}\mathbf{b} \leq \mathbf{a}^{\mathrm{T}}\mathbf{b}$*.*

Proof (a slight extension of Uchida (2023)). Let $c_1, ..., c_n$ and $d_1, ..., d_n$ be non-negative numbers. Then, we can write $a_i = \sum_{k=i}^{n} c_k$ and $b_i = \sum_{k=i}^{n} d_k$. Using these expressions,

$$\mathbf{a}^{\mathrm{T}}\mathbf{b} - \mathbf{a}^{\mathrm{T}}\mathbf{P}\mathbf{b} = \sum_{i,j=1}^{n} (\delta_{ij} - p_{ij}) a_i b_j = \sum_{i,j=1}^{n} (\delta_{ij} - p_{ij}) \sum_{k=i}^{n} c_k \sum_{l=j}^{n} d_l$$

$$= \sum_{k,l=1}^{n} c_k d_l \sum_{i=1}^{k} \sum_{j=1}^{l} (\delta_{ij} - p_{ij})$$

follows, where δ_{ij} is the Kronecker delta. In the above expression, define the doubly stochastic (rather than substochastic) matrix $\mathbf{P}^* = \{p_{ij}^*\}$ satisfying $p_{ij}^* \geq p_{ij}$ $(i, j = 1, ...n)$. Then, consider the case $k \geq l$ in the above expression. We obtain

$$\begin{aligned}
&\textstyle\sum_{i=1}^{k}\sum_{j=1}^{l}(\delta_{ij} - p_{ij}) = \sum_{i=1}^{k}\sum_{j=1}^{l}\delta_{ij} - \sum_{i=1}^{k}\sum_{j=1}^{l}p_{ij} \\
&= \textstyle\sum_{i=1}^{l}\delta_{ii} - \sum_{i=1}^{k}\sum_{j=1}^{l}p_{ij} = l - \sum_{i=1}^{k}\sum_{j=1}^{l}p_{ij} \\
&\geq l - \textstyle\sum_{i=1}^{n}\sum_{j=1}^{l}p_{ij}^* \geq l - l = 0.
\end{aligned}$$

When $k \leq l$, in a similar manner we obtain $\sum_{i=1}^{k}\sum_{j=1}^{l}(\delta_{ij} - p_{ij}) \geq 0$. These two inequalities give the required result $\mathbf{a}^{\mathrm{T}}\mathbf{b} - \mathbf{a}^{\mathrm{T}}\mathbf{P}\mathbf{b} \geq 0$. □

Remark A. The original proof by Fan (1951) is a short one though it is not self-contained in that "Abel's lemma" is used. The author could not identify the Abel lemma with an associated reference among Abel's formulas. The above proof is a slight extension of the result by Uchida (2023) who dealt with only the doubly stochastic matrix $\mathbf{P}^*$, which is a special case of the doubly substochastic matrix $\mathbf{P}$.

The second proof of Theorem 1 (von Neumann's trace inequality; Mirsky (1975, Section 3, p. 305)). $|\,\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2\mathbf{\Gamma}_2)| \leq \mathrm{tr}(\mathbf{\Gamma}_1\mathbf{\Gamma}_2)$ is derived. Let $\mathbf{X}_i = \{x_{(i)jk}\}$ $(j, k = 1, ..., m)$ and $\mathbf{\Gamma}_i = \mathrm{diag}(\gamma_{(i)1}, ..., \gamma_{(i)m})$ $(i = 1,\ 2)$. Then, using Fan's (1951) inequality for doubly (sub)stochastic matrices, we have

$$\begin{aligned}
&|\,\mathrm{tr}(\mathbf{X}_1\mathbf{\Gamma}_1\mathbf{X}_2\mathbf{\Gamma}_2)| \ = \ |\textstyle\sum_{j,k=1}^{m} x_{(1)jk}\gamma_{(1)k}x_{(2)kj}\gamma_{(2)j}| \\
&\leq \textstyle\sum_{j,k=1}^{m}|x_{(1)jk}x_{(2)kj}|\gamma_{(1)k}\gamma_{(2)j} \\
&\leq \tfrac{1}{2}\textstyle\sum_{j,k=1}^{m}x_{(1)jk}^2\gamma_{(1)k}\gamma_{(2)j} + \tfrac{1}{2}\sum_{j,k=1}^{m}x_{(2)kj}^2\gamma_{(1)k}\gamma_{(2)j} \\
&\leq \tfrac{1}{2}\textstyle\sum_{j=1}^{m}\gamma_{(1)j}\gamma_{(2)j} + \tfrac{1}{2}\sum_{j=1}^{m}\gamma_{(1)j}\gamma_{(2)j} = \mathrm{tr}(\mathbf{\Gamma}_1\mathbf{\Gamma}_2).
\end{aligned}$$

□

The Kristof-type theorem for correlation preserving predictors of factor scores: Neudecker (2004, Section 2) (n = 1). Neudecker obtained the same solution of the first example by Green and the last reformulated one by Ten Berge in Section 5 as "A Kristof-type theorem" in the context of the derivations of correlation preserving predictors of factor scores (for these predictors see the references in Neudecker (2004); and Mori and Kurata (2013)). He did not mention or use Ten Berge's theorem, but employed calculus and Lagrange multipliers. Neudecker also used the SVD $\mathbf{A} = \mathbf{U}_0\mathbf{\Lambda}_0\mathbf{V}_0^{\mathrm{T}}$, where he employed only positive singular values i.e., $\mathbf{\Lambda}_0 > \mathbf{O}$ in Löwner's sense.

An advantage of Neudecker's derivation is to give the set of explicit expressions of $\mathbf{X}$ maximizing $\mathrm{tr}(\mathbf{X}^{\mathrm{T}}\mathbf{A})$ as $\mathbf{X} = \mathbf{U}_0\mathbf{V}_0^{\mathrm{T}} = \mathbf{U}\mathbf{V}^{\mathrm{T}}$ when $\mathrm{rank}(\mathbf{A}) = q^* = q$ and $\mathbf{X} = \mathbf{U}_0\mathbf{V}_0^{\mathrm{T}} + \mathbf{Q}(\mathbf{I}_q - \mathbf{V}_0\mathbf{V}_0^{\mathrm{T}})$ with a $p \times q$ matrix $\mathbf{Q}$ when $q^* < q$ as "the general solution" (Neudecker, 2004, Equation (2.5)). In these expressions,

$$\mathbf{U}_0\mathbf{V}_0^{\mathrm{T}} = \mathbf{X}\{(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{1/2}\}^{+} = \mathbf{X}\{(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{+}\}^{1/2} \equiv \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{(+)1/2}$$

and

$$\mathbf{V}_0\mathbf{V}_0^{\mathrm{T}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{(+)1/2}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{1/2} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{1/2}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{(+)1/2},$$

where $(\cdot)^{1/2}$ is the matrix square root of a matrix; and $(\cdot)^{+}$ is the Moore-Penrose generalized (MP g-) inverse of a possibly rectangular matrix, which is obtained by using the SVD. That is, when the SVD of a matrix is $\mathbf{Y} = \mathbf{P\Gamma Q}^{\mathrm{T}}$ employing only the positive singular values, we have $\mathbf{Y}^{+} = \mathbf{Q\Gamma}^{-1}\mathbf{P}^{\mathrm{T}}$, which satisfies the conditions of the MP g-inverse: $\mathbf{YY}^{+}\mathbf{Y} = \mathbf{Y},\ \mathbf{Y}^{+}\mathbf{YY}^{+} = \mathbf{Y}^{+},\ \mathbf{YY}^{+} = (\mathbf{YY}^{+})^{\mathrm{T}}$ and $\mathbf{Y}^{+}\mathbf{Y} = (\mathbf{Y}^{+}\mathbf{Y})^{\mathrm{T}}$.

Neudecker (2004, Equation (2.5)) stated that "$\mathbf{Q}$ arbitrary". This is misleading since when $\mathbf{Q}$ is a zero matrix, the rank of $\mathbf{X} = \mathbf{U}_0\mathbf{V}_0^{\mathrm{T}}$ becomes $q^*(< q)$ and does not satisfy $\mathbf{X}^{\mathrm{T}}\mathbf{X} = \mathbf{I}_q$ though this will give the same maximum. Instead, $\mathbf{Q}$ should be defined as a $p \times q$ arbitrary suborthonormal matrix $\mathbf{Q} = \mathbf{U}_1\mathbf{V}_1^{\mathrm{T}}$ of rank $q - q^*$, where $\mathbf{U}_1$ and $\mathbf{V}_1$ are $p \times (q - q^*)$ and $q \times (q - q^*)$ semiorthonormal matrices, respectively such that $\mathbf{U}_1^{\mathrm{T}}\mathbf{U}_1 = \mathbf{V}_1^{\mathrm{T}}\mathbf{V}_1 = \mathbf{I}_{q-q^*}$, $\mathbf{U}_0^{\mathrm{T}}\mathbf{U}_1 = \mathbf{O}$ and $\mathbf{V}_0^{\mathrm{T}}\mathbf{V}_1 = \mathbf{O}$, which shows the arbitrary property of $\mathbf{Q}$ stated earlier in that when $\mathbf{U}_1$ and $\mathbf{V}_1$ are replaced by $\mathbf{U}_1\mathbf{U}_1^*$ and $\mathbf{V}_1\mathbf{V}_1^*$ with $\mathbf{U}_1^*$ and $\mathbf{V}_1^*$ being arbitrary $(q - q^*) \times (q - q^*)$ orthonormal matrices, these can be used with a different $\mathbf{Q}^* \equiv \mathbf{U}_1\mathbf{U}_1^*(\mathbf{V}_1\mathbf{V}_1^*)^{\mathrm{T}} \neq \mathbf{Q}$. Note that these arbitrary $\mathbf{U}_1$ and $\mathbf{V}_1$ give $\mathbf{V}_0^{\mathrm{T}}\mathbf{V}_0 + \mathbf{V}_1^{\mathrm{T}}\mathbf{V}_1 = (\mathbf{V}_0 : \mathbf{V}_1)^{\mathrm{T}}(\mathbf{V}_0 : \mathbf{V}_1) = (\mathbf{V}_0 : \mathbf{V}_1\mathbf{V}_1^*)^{\mathrm{T}}(\mathbf{V}_0 : \mathbf{V}_1\mathbf{V}_1^*) = \mathbf{I}_q$.

Then, it is found that

$$\begin{aligned}\mathbf{X} &= \mathbf{U}_0\mathbf{V}_0^{\mathrm{T}} + \mathbf{Q}(\mathbf{I}_q - \mathbf{V}_0\mathbf{V}_0^{\mathrm{T}}) = \mathbf{U}_0\mathbf{V}_0^{\mathrm{T}} + \mathbf{U}_1\mathbf{V}_1^{\mathrm{T}}(\mathbf{I}_q - \mathbf{V}_0\mathbf{V}_0^{\mathrm{T}})\\ &= \mathbf{U}_0\mathbf{V}_0^{\mathrm{T}} + \mathbf{U}_1\mathbf{V}_1^{\mathrm{T}} = (\mathbf{U}_0 : \mathbf{U}_1)(\mathbf{V}_0 : \mathbf{V}_1)^{\mathrm{T}}\end{aligned}$$

satisfying

$$\begin{aligned}\mathbf{X}^{\mathrm{T}}\mathbf{X} &= (\mathbf{V}_0 : \mathbf{V}_1)(\mathbf{U}_0 : \mathbf{U}_1)^{\mathrm{T}}(\mathbf{U}_0 : \mathbf{U}_1)(\mathbf{V}_0 : \mathbf{V}_1)^{\mathrm{T}}\\ &= (\mathbf{V}_0 : \mathbf{V}_1)(\mathbf{V}_0 : \mathbf{V}_1)^{\mathrm{T}} = \mathbf{I}_q.\end{aligned}$$

Using the above $\mathbf{X}$, the maximum is given by

$$\begin{aligned}&\mathrm{tr}(\mathbf{A}^{\mathrm{T}}\mathbf{X}) = \mathrm{tr}\{(\mathbf{U}_0\mathbf{\Lambda}_0\mathbf{V}_0^{\mathrm{T}})^{\mathrm{T}}(\mathbf{U}_0 : \mathbf{U}_1)(\mathbf{V}_0 : \mathbf{V}_1)^{\mathrm{T}}\}\\ &= \mathrm{tr}\{\mathbf{V}_0\mathbf{\Lambda}_0\mathbf{U}_0^{\mathrm{T}}(\mathbf{U}_0\mathbf{V}_0^{\mathrm{T}} + \mathbf{U}_1\mathbf{V}_1^{\mathrm{T}})\}\\ &= \mathrm{tr}(\mathbf{V}_0\mathbf{\Lambda}_0\mathbf{V}_0^{\mathrm{T}}) = \mathrm{tr}(\mathbf{V}_0^{\mathrm{T}}\mathbf{V}_0\mathbf{\Lambda}_0)\\ &= \mathrm{tr}(\mathbf{\Lambda}_0),\end{aligned}$$

where $\mathrm{tr}(\mathbf{\Lambda}_0) = \mathrm{tr}(\mathbf{\Lambda})$ and the singular diagonal matrix $\mathbf{\Lambda}$ of rank q^* was used earlier. Define the semiorthonormal matrix $\mathbf{U} \equiv (\mathbf{U}_0 : \mathbf{U}_1)$ and orthonormal $\mathbf{V} \equiv (\mathbf{V}_0 : \mathbf{V}_1)$. Then, $\mathbf{X} = \mathbf{UV}^{\mathrm{T}}$ is equal to that obtained by Ten Berge as shown earlier.

Recall the expression $\mathbf{X} = \mathbf{U}_0\mathbf{V}_0^{\mathrm{T}} + \mathbf{Q}(\mathbf{I}_q - \mathbf{V}_0\mathbf{V}_0^{\mathrm{T}})$. The matrix $\mathbf{Q}$ can be an arbitrary $p \times q$ semiorthonormal one denoted by $\mathbf{U}^*$ with $\mathbf{U}_0^{\mathrm{T}}\mathbf{U}^* = \mathbf{O}$ and $\mathbf{U}^{*\mathrm{T}}\mathbf{U}^* = \mathbf{I}_q$. This is seen by

$$\begin{aligned}\mathbf{X}^{\mathrm{T}}\mathbf{X} &= \{\mathbf{U}_0\mathbf{V}_0^{\mathrm{T}} + \mathbf{U}^*(\mathbf{I}_q - \mathbf{V}_0\mathbf{V}_0^{\mathrm{T}})\}^{\mathrm{T}}\{\mathbf{U}_0\mathbf{V}_0^{\mathrm{T}} + \mathbf{U}^*(\mathbf{I}_q - \mathbf{V}_0\mathbf{V}_0^{\mathrm{T}})\} \\ &= \mathbf{V}_0\mathbf{U}_0^{\mathrm{T}}\mathbf{U}_0\mathbf{V}_0^{\mathrm{T}} + (\mathbf{I}_q - \mathbf{V}_0\mathbf{V}_0^{\mathrm{T}})\mathbf{U}^{*\mathrm{T}}\mathbf{U}^*(\mathbf{I}_q - \mathbf{V}_0\mathbf{V}_0^{\mathrm{T}}) \\ &= \mathbf{V}_0\mathbf{V}_0^{\mathrm{T}} + \mathbf{I}_q - \mathbf{V}_0\mathbf{V}_0^{\mathrm{T}} = \mathbf{I}_q,\end{aligned}$$

satisfying the assumption, where the idempotent property $(\mathbf{I}_q - \mathbf{V}_0\mathbf{V}_0^{\mathrm{T}})^2 = \mathbf{I}_q - \mathbf{V}_0\mathbf{V}_0^{\mathrm{T}}$ is used. The matrix $\mathbf{X}$ gives the same maximum

$$\begin{aligned}\mathrm{tr}(\mathbf{A}^{\mathrm{T}}\mathbf{X}) &= \mathrm{tr}\left[\mathbf{V}_0\mathbf{\Lambda}_0\mathbf{U}_0^{\mathrm{T}}\{\mathbf{U}_0\mathbf{V}_0^{\mathrm{T}} + \mathbf{U}^*(\mathbf{I}_q - \mathbf{V}_0\mathbf{V}_0^{\mathrm{T}})\}\right] \\ &= \mathrm{tr}(\mathbf{V}_0\mathbf{\Lambda}_0\mathbf{U}_0^{\mathrm{T}}\mathbf{U}_0\mathbf{V}_0^{\mathrm{T}}) = \mathrm{tr}(\mathbf{V}_0\mathbf{\Lambda}_0\mathbf{V}_0^{\mathrm{T}}) \\ &= \mathrm{tr}(\mathbf{\Lambda}_0).\end{aligned}$$

In the expression $\mathbf{X} = \mathbf{U}_0\mathbf{V}_0^{\mathrm{T}} + \mathbf{Q}(\mathbf{I}_q - \mathbf{V}_0\mathbf{V}_0^{\mathrm{T}})$, the term $\mathbf{V}_0\mathbf{V}_0^{\mathrm{T}}$ is the projection matrix onto the space spanned by the q^* orthonormal columns of $\mathbf{V}_0$ given by $\mathbf{V}_0(\mathbf{V}_0^{\mathrm{T}}\mathbf{V}_0)^{-1}\mathbf{V}_0^{\mathrm{T}} = \mathbf{V}_0\mathbf{I}_{q*}^{-1}\mathbf{V}_0^{\mathrm{T}} = \mathbf{V}_0\mathbf{V}_0^{\mathrm{T}}$. Recall that $\mathbf{I}_q = \mathbf{V}_0\mathbf{V}_0^{\mathrm{T}} + \mathbf{V}_1\mathbf{V}_1^{\mathrm{T}}$. Then, $\mathbf{I}_q - \mathbf{V}_0\mathbf{V}_0^{\mathrm{T}} = \mathbf{V}_1\mathbf{V}_1^{\mathrm{T}}$ is also the projection matrix onto the space spanned by the $q - q^*$ orthonormal columns of $\mathbf{V}_1$. In the term $\mathbf{Q}(\mathbf{I}_q - \mathbf{V}_0\mathbf{V}_0^{\mathrm{T}}) = \mathbf{Q}\mathbf{V}_1\mathbf{V}_1^{\mathrm{T}}$, each row of $\mathbf{Q}$ is projected onto the space spanned by $\mathbf{V}_1$.

An arbitrary property of $\mathbf{U}_1$ and $\mathbf{V}_1$ in $\mathbf{X} = \mathbf{U}\mathbf{V}^{\mathrm{T}} = \mathbf{U}_0\mathbf{V}_0^{\mathrm{T}} + \mathbf{U}_1\mathbf{V}_1^{\mathrm{T}}$ of Ten Berge's (1983) solution similar to that with $\mathbf{Q}$ is that $\mathbf{U}_1$ and $\mathbf{V}_1$ can be replaced by $\mathbf{U}_1\mathbf{W}_{\mathbf{U}1}$ and $\mathbf{V}_1\mathbf{W}_{\mathbf{V}1}$, respectively, where $\mathbf{W}_{\mathbf{U}1}$ and $\mathbf{W}_{\mathbf{V}1}$ are arbitrary $(q - q^*) \times (q - q^*)$ orthonormal matrices yielding $\mathbf{X}^* \equiv \mathbf{U}_0\mathbf{V}_0^{\mathrm{T}} + \mathbf{U}_1\mathbf{W}_{\mathbf{U}1}(\mathbf{V}_1\mathbf{W}_{\mathbf{V}1})^{\mathrm{T}} \neq \mathbf{X} = \mathbf{U}_0\mathbf{V}_0^{\mathrm{T}} + \mathbf{U}_1\mathbf{V}_1^{\mathrm{T}}$.

For completeness, arbitrary aspects in $\mathbf{A} = \mathbf{U}_0\mathbf{\Lambda}_0\mathbf{V}_0^{\mathrm{T}}$ of rank $q^* \leq q$ are noted. Under the standard definition of $\mathbf{\Lambda}_0 = \mathrm{diag}(\lambda_1, ..., \lambda_{q^*})$ with $\lambda_1 \geq \cdots \geq \lambda_{q^*} \geq 0$, $\mathbf{\Lambda}_0$ is identified while $\mathbf{U}_0$ and $\mathbf{V}_0$ are identified up to the sign changes (orientations or reflections) of the pairs of their corresponding columns. Further, suppose that some positive singular values are multiple e.g., $\lambda_j = \lambda_{j+1} = \cdots = \lambda_{j+k-1} \equiv \lambda_{j(k)}$ with multiplicity $k(> 1)$, we have

$$\begin{aligned}\mathbf{A} &= \mathbf{U}_{0(-k)}\mathbf{\Lambda}_{0(-k)}\mathbf{V}_{0(-k)}^{\mathrm{T}} + \mathbf{U}_{0(k)}\mathbf{\Lambda}_{0(k)}\mathbf{V}_{0(k)}^{\mathrm{T}} \\ &= \mathbf{U}_{0(-k)}\mathbf{\Lambda}_{0(-k)}\mathbf{V}_{0(-k)}^{\mathrm{T}} + \mathbf{U}_{0(k)}\lambda_{j(k)}\mathbf{I}_k\mathbf{V}_{0(k)}^{\mathrm{T}} \\ &= \mathbf{U}_{0(-k)}\mathbf{\Lambda}_{0(-k)}\mathbf{V}_{0(-k)}^{\mathrm{T}} + \lambda_{j(k)}\mathbf{U}_{0(k)}\mathbf{W}_{(k)}(\mathbf{V}_{0(k)}\mathbf{W}_{(k)})^{\mathrm{T}},\end{aligned}$$

where $\mathbf{U}_{0(k)}$ and $\mathbf{V}_{0(k)}$ are semiorthonormal submatrices of $\mathbf{U}_0$ and $\mathbf{V}_0$, respectively corresponding to the multiple $\lambda_{j(k)}$ with $\mathbf{\Lambda}_{0(k)}$ defined similarly; and $\mathbf{U}_{0(-k)}$, $\mathbf{V}_{0(-k)}$ and $\mathbf{\Lambda}_{0(-k)}$ are matrices given by using the singular values except $\lambda_j = \lambda_{j+1} = \cdots = \lambda_{j+k-1}$; and $\mathbf{W}_{(k)}$ is an arbitrary $k \times k$ orthonormal matrix. The last arbitrary property is similar to that in the so-called "rotational indeterminacy" in factor analysis.

Journal of Behavioral Data Science, 2024, *4* (2), 51–63.
DOI: `https://doi.org/10.35566/jbds/lin`

Lord's Paradox Illustrated in Three-Wave Longitudinal Analyses: Cross Lagged Panel Models Versus Linear Latent Growth Models

Hua Lin and Robert E. Larzelere

Department of Human Development and Family Science
Oklahoma State University, USA
`hua.lin@okstate.edu` and `robert.larzelere@okstate.edu`

Abstract. Lord's (1967) paradox showed that two basic ways to analyze change longitudinally can produce contradictory results in 2-occasion nonrandomized studies. This study extends that paradox to difference-score and ANCOVA-type residualized change score analyses across three waves of data for four corrective actions thought to be effective: corrective disciplinary actions by parents (timeout and reasoning) and corrective actions by professionals (psychotherapy and hospitalization). All significant findings indicated that these corrective actions were harmful according to cross-lagged panel models but beneficial according to linear latent growth models. One type of analysis may not generalize to the other type of analysis. These results are consistent with recent recognition that ANCOVA-type analyses are biased by invariant between-person differences, but difference-score analyses can have their own biases. Recognition of these biases is needed to discriminate between stronger and weaker causal evidence in longitudinal analyses.

Keywords: Causal inference · Cross-lagged panel analysis · Latent growth modeling

1 Introduction

Most longitudinal studies have found that corrective actions by parents and by professionals appear to be harmful in analyses that control for initial differences with ANCOVA-type analyses of residualized change scores $Y_2|Y_1$ (i.e., Y_2 conditional on Y_1; Larzelere, Lin, Payton, & Washburn, 2018). Examples include parent-youth discussions about the risks of unprotected sex (Lin & Larzelere, 2020), psychotherapy for children, and methylphenidate (i.e., Ritalin: Larzelere, Ferrer, Kuhn, & Danelia, 2010). Although all of these corrective actions have looked harmful according to analyses of residualized change score analyses (i.e., predicting Wave-2 outcomes Y_2 while controlling for Wave-1 outcome scores Y_1), difference-score analyses have often made them look beneficial from the same data (predicting $Y_2 - Y_1$; Larzelere et al., 2018).

This inconsistency is an example of Lord's (1967) paradox. In Lord's original hypothetical study, females' and males' weight gains were compared with each other using the two types of change-score analyses. Initial average weights differed significantly for females and males, and their average weights stayed the same from pretest to posttest for both genders. Difference-score analyses indicated no gender difference in weight gained, as expected. However, ANCOVA indicated that males gained more weight than females who started at the same weight. Although both results are correct for their corresponding *predictive* research questions, both cannot provide correct causal inferences about the effect of manipulating a causal variable (e.g., for a corrective action of interest). Consistent with Lord's original paradox, causally relevant coefficients from residualized change score analyses are generally biased in the direction of the pretest group means, *relative to* the difference-score coefficients, regardless of which analysis is least biased (Angrist & Pischke, 2009; Larzelere et al., 2018; Lin & Larzelere, 2020). This corresponds to recent documentations that longitudinal analyses of residualized change scores are biased by between-person differences that do not change during the study (Berry & Willoughby, 2017; Hamaker, Kuiper, & Grasman, 2015; Hoffman, 2015).

Despite being discussed for over 50 years, the implications of Lord's paradox have been insufficiently recognized in developmental psychology. Longitudinal analyses have preferred analyzing residualized change scores since Cronbach and Furby's 1970 recommendation. Two-wave residualized change score and difference-score analyses are building blocks for more complex models such as cross-lagged panel analyses and linear growth models. Therefore, this problem of contradictory, potentially biased estimates likely generalizes to advanced statistical models. However, little is known about how Lord's paradox applies to more complex statistical models (e.g., cross-lagged panel models and latent growth models) or how to minimize these biases to approximate valid causal estimates more closely. Like ANCOVA, cross-lagged panel models predict residualized change scores (e.g., predicting y_t controlling for y_{t-1}) between adjacent occasions across three or more occasions. Therefore, cross-lagged panel models could be considered a series of $T - 1$ ANCOVAs. In contrast, the most basic latent growth model typically predicts a simple difference score from Wave 1 to Wave T based on the best-fitting linear slope of the outcome scores across the T waves. In this article, we modify the latent growth model to predict simple difference scores between adjacent waves. This modified latent growth model is more similar to cross-lagged panel models by modeling change in the outcome scores from Wave $t - 1$ to Wave t across T waves.

1.1 Cross-Lagged Panel Model

Cross-lagged panel models estimate the bidirectional effects between the treatment condition and the outcome score over time (Selig & Little, 2012). The cross-lagged panel model provides information about how variations in one variable (typically treatment vs. control) predict changes in another variable (the outcome) over time. The multi-wave cross-lagged panel model can be described as follows:

$$X_{i,t} = \alpha_0 + \alpha_1 X_{i,t-1} + \alpha_2 Y_{i,t-1} + \varepsilon_{i,xt}$$

$$Y_{i,t} = \theta_0 + \theta_1 X_{i,t-1} + \theta_2 Y_{i,t-1} + \varepsilon_{i,yt}$$

where X and Y represent the treatment and outcome variables at a given time t, predicted from these variables at the immediately preceding time $t-1$. These are adjacent-wave ANCOVA functions for both variables - as predictor and outcome at adjacent time points. The primary interest is the treatment effect θ_1 of X_{t-1} on the outcome at the next time point, controlling for the preceding outcome score $Y_t|Y_{t-1}$.

1.2 Latent Growth Model

Whereas cross-lagged panel models predict residualized change scores $Y_t|Y_{t-1}$, the linear growth model uses difference scores as its basic building block for analyzing change. Linear latent growth models analyze how individuals' scores change over time and how treatment conditions influence such changes using the difference-score approach:

$$\text{Level 1}: Y_{ti} = \beta_{0i} + \beta_{1i} T_{ti} + r_{ti}$$

$$\text{Level 2}: \beta_{0i} = \gamma_{00} + \gamma_{01} X_j + \epsilon_{0i}$$

$$\beta_{1i} = \gamma_{10} + \gamma_{11} X_j + \epsilon_{1i}$$

where Level 1 represents how individual scores change linearly over time, and Level 2 predicts initial scores and within-individual changes from between-person differences in the causal variable of interest X_j. At Level 1, Y_{ti} represents individual i's outcome at time t; β_{0i} represents the starting point (when $T_{ti} = 0$) on individual i's best-fitting straight line across time; β_{1i} represents the individual's linear slope across time T_{ti}, and r_{ti} represents the unexplained error in the individual's outcome Y_{ti}. At level 2, γ_{00} represents the mean of the individual starting points on the outcome when $X_j = 0$; γ_{01} is the effect of the predictor X_j on the starting point (or intercept) β_{0i}; ϵ_{0i} represents the deviation of the individual's starting point from what is predicted by the rest of that equation (the fixed-effects part); X_j is the treatment condition, e.g., with $j = 2$ for the treatment group and $j = 1$ for the comparison group; γ_{10} is the mean linear slope across the waves when $X_j = 0$; γ_{11} is the effect of the predictor X_j on the average individual slope β_{1i}; and ϵ_{1i} is the deviation of the individual's slope from the slope predicted from the fixed-effects part of that equation. With person-mean centering, the latent growth model estimates pure within-person changes at Level 1. Level 2 then estimates between-person differences in those changes. In two-wave analyses, the slope is the difference score from Wave 1 to Wave 2 ($Y_{2i} - Y_{1i}$). In three-wave analyses, each individual's slope is the estimated linear change per unit of time in that person's best-fitting straight line across their scores at all three waves. The primary interest of the latent growth model is the effect of the treatment on change in the slope γ_{11}.

1.3 The Current Study

The current study used four examples of corrective actions thought to be effective to illustrate Lord's paradox in three-wave longitudinal analyses. The four examples involve the apparent effect of (1) disciplinary time-out on subsequent child aggression, (2) disciplinary reasoning on subsequent child aggression, 3) psychotherapy on subsequent maternal depression, and 4) hospitalization on subsequent physical health. Each example was analyzed with a cross-lagged panel model and a latent growth model across three waves of data: Although standard latent growth models typically predict one linear slope from the first to the last wave, the two-slope latent growth model in this study was designed to be more similar to a cross-lagged panel by predicting simple difference scores between adjacent waves. The intercept was modeled as usual (all loadings set to 1), but Slope 1 specified the simple change from Wave 1 to Wave 2 (with loadings set at -1 and 0), whereas Slope 2 specified the simple change from Wave 2 to Wave 3 (loadings set at 0 and 1). The model then estimated the effect of each correction action at one wave (Wave 1 or 2) on simple change in the outcome from that wave to the next wave.

It was hypothesized that cross-lagged panel models would make corrective actions appear to be harmful, whether implemented by parents (time-out, reasoning) or by professionals (psychotherapy, hospitalizations). In contrast, latent growth models would indicate that all these corrective actions would lead to improvements in the same outcomes.

2 Method

2.1 Participants

This study used the Fragile Families and Child Wellbeing (FFCW) dataset which started with baseline data for mostly unmarried couples with children born from 1998 to 2000 in 20 large cities of the United States (Reichman, Teitler, Garfinkel, & McLanahan, 2001). It includes a wide range of data on household characteristics, physical and mental health, and parenting, first when the children were born, and later when the children were approximately 1, 3, 5, 9, 15, and 22 years old. The current study uses corrective action data when the children were 3 and 5 years old and outcome data when they were 3, 5, and 9 years old. At baseline (when the child was born), the 4588 mothers in these 3-wave analyses averaged

25.2 years old and had some college on average, and consisted of 21.2% White, 48.0% Black, 27.0% Hispanic, and 3.8% others. Missingness ranged from 8% to 28%. Full information maximum likelihood was used to adjust for missing data in the 3-wave analyses, which assumes that those data were missing at random. The FFCW data set (https://ffcws.princeton.edu/documentation) is available from Princeton University's Office of Population Research (OPR) data archive.

2.2 Measures

Time-out Disciplinary time-out was assessed by mothers' self-report on one item from the Parent-Child Conflict Tactics Scale (Straus, Hamby, Finkelhor, Moore, & Runyan, 1998), which asks how often in the past year mothers put their child in time-out or sent them to their room. The frequency was reported on a 8-point scale, ranging from never (0) to 11-20 times (6) to more than 20 times (7). We created a dummy variable indicating whether the time-out frequency was above the median frequency or not: 11 or more times (1), or less than 11 times (0).

Reasoning Disciplinary reasoning was also assessed by mothers' self-report from one item of the Parent-Child Conflict Tactics Scale (Straus et al., 1998), using the same response options. The item asks how often in the past year mothers explained why something was wrong. We created a dummy variable indicating whether reasoning occurred more frequently than the median or not: 11 or more times (1), or less than 11 times (0).

Child Aggression The FFCW measure of child aggression was a modified version of the aggression subscale of the Child Behavior Checklist (CBCL; Achenbach, 1991; Achenbach & Rescorla, 2000) with 19 items at age 3, 13 items at age 5, and 17 items at age 9. Mothers reported whether various behaviors were not true, somewhat/sometimes true, or often/very true of the child. Sample questions include destroying things, being disobedient, hitting others, getting in many fights, screaming a lot, and threatening people. The scale demonstrated excellent reliability with coefficient alphas of 0.88 (age 3), 0.82 (age 5), and 0.88 (age 9).

Psychotherapy for Depression Psychotherapy for depression was measured by two questions. Mothers reported whether they had received counseling/therapy for personal problems in the past year. If "yes," they were asked whether the counseling/therapy was for depression or for a range of other problems. Reported counseling/therapy for depression was coded 1, and other answers were coded 0. Two dummy codes indicated whether mothers received psychotherapy for depression when the child was 3 and 5 years old.

Depression Severity Depression severity was based on maternal self-reports about symptoms of a Major Depressive Episode, derived from the Composite International Diagnostic Interview—Short Version (Kessler, Andrews, Mroczek, Ustun, & Wittchen, 1998). The CIDI is a standardized survey for assessing mental disorders such as depression. The depression items included two stem questions and seven additional questions for those exceeding the threshold on the stem questions. We constructed a 13-point scale from none (0) through subthreshold symptoms (1 to 4) to the number of symptoms above the threshold, including the stem questions (5 to 12).

Hospitalization Hospitalization was measured by a single dummy-coded item indicating whether mothers visited an emergency room or had an overnight hospital stay during the past year.

Physical Health Mothers' physical health was based on mothers' self-reports on their health condition on a five-point scale (0 = poor to 4 = great).

3 Results

The results showed contradictory results from cross-lagged panel models compared to linear latent growth models. The four examples estimate the apparent effects of corrective actions by parents (time-out and reasoning) and by professionals (psychotherapy and inpatient hospitalized treatments).

3.1 Time-out and Subsequent Child Aggression

Cross-lagged panel models made time-out at Wave 2 look significantly harmful by increasing child aggression at Wave 3, after controlling for the preceding aggression scores: $b = 0.03, p = 0.002$, Figure 1, Plot A. (Time-out at Wave 1 also predicted higher aggression at Wave 2 controlling for Time-1 aggression, but only marginally, $b = 0.02, p < 0.01$.) In contrast, two-slope latent growth models made time-out look helpful in reducing child aggression from each wave to the next wave: $b = -0.06$ (Wave 1 time-out predicting change in aggression from Wave 1 to Wave 2), and $b = -0.03$ (Wave 2 time-out predicting change in aggression from Wave 2 to Wave 3), ps < 0.01, Figure 1, Plot B.

3.2 Reasoning and Subsequent Child Aggression

Cross-lagged panel models also made disciplinary reasoning at Wave 2 look harmful by predicting more child aggression at Wave 3, after controlling for Wave 1 and Wave 2 aggression scores: $b = 0.04, p = 0.001$, Figure 2, Plot A. In contrast, the 2-slope latent growth model made reasoning at Wave 1 look helpful in reducing child aggression from Wave 1 to Wave 2: $b = -0.07, p < .001$, Figure 2, Plot B.

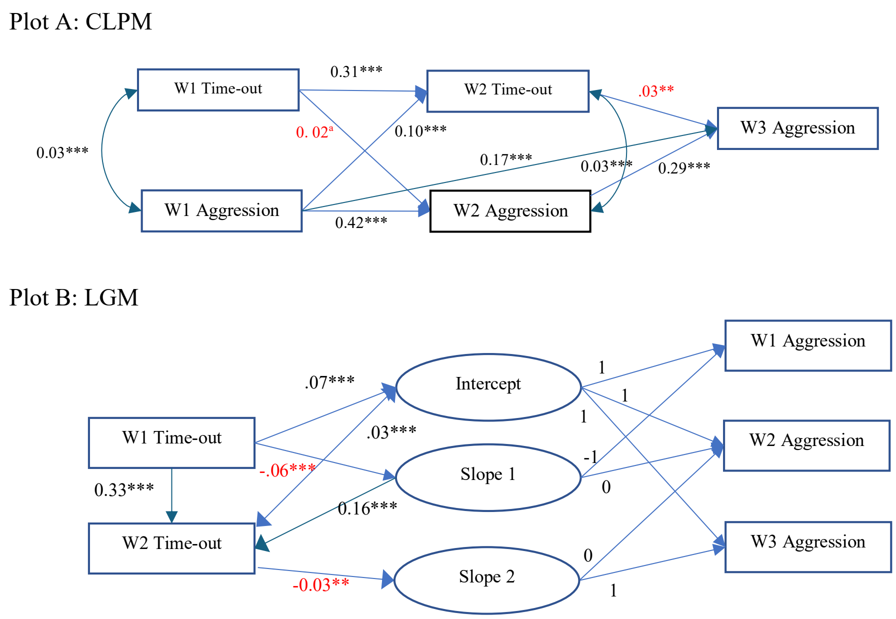

Figure 1. Cross-Lagged Panel Model (CLPM) and Latent Growth Model (LGM) of Time-out and Child Aggression across three waves of data. $^{a}p < .10; ^{*}p < .05; ^{**}p < .01; ^{***}p < .001. N = 4153$

3.3 Psychotherapy and Subsequent Depression

The results followed a similar pattern for professional treatments. A cross-lagged panel model made therapy for depression look significantly harmful by predicting higher depression severity at the next wave, even after controlling for the preceding depression severity score: $b = 0.70$ (Wave 1 psychotherapy predicting Wave 2 depression severity) and $b = 1.60$ (Wave 2 psychotherapy predicting Wave 3 depression severity), all $ps < 0.05$, Figure 3, Plot A. In contrast, 2-slope latent growth models made therapy look helpful in reducing depression severity from each wave to the next wave: $b = -3.05$ (Wave 1 psychotherapy predicting a decrease in depression from Wave 1 to Wave 2) and $b = -0.73$ (Wave 2 psychotherapy predicting a decrease in depression from Wave 2 to Wave 3), $ps < 0.05$, Figure 3, Plot B.

3.4 Hospitalization and Subsequent Physical Health

A cross-lagged panel model made hospitalization look harmful by predicting worse physical health in mothers at the next wave, after controlling for mothers' preceding physical health score: $b = -0.15$ (Wave 1 hospitalization predicting worse health at Wave 2) and $b = -0.09$ (Wave 2 hospitalization predicting worse health at Wave 3), all $ps < 0.05$, Figure 4, Plot A. In contrast, a 2-slope latent

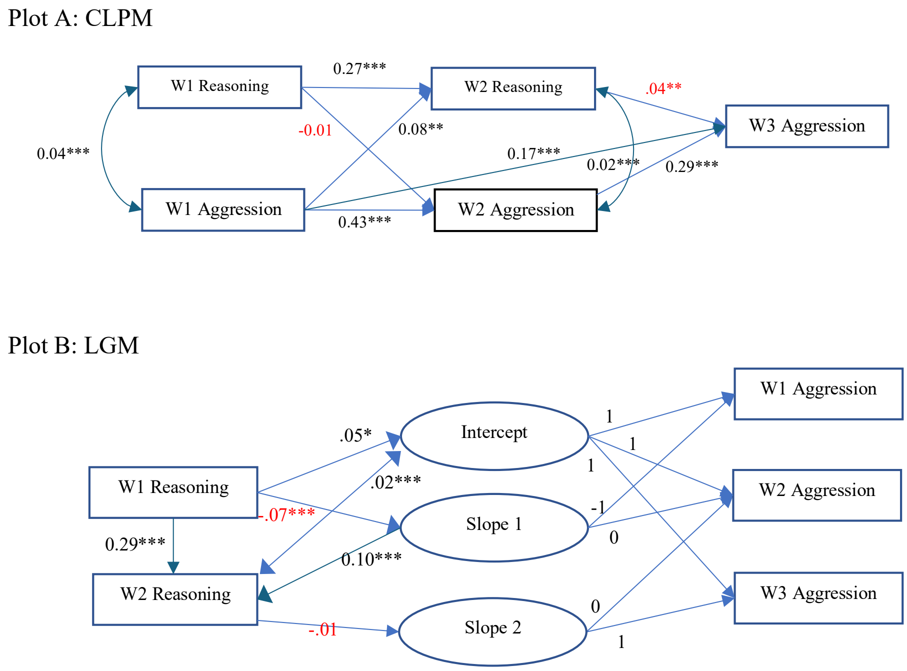

Figure 2. Cross-Lagged Panel Model (CLPM) and Latent Growth Model (LGM) of Reasoning and Child Aggression across three waves of data. $^{*}p < .05;^{**}p < .01;^{***}p < .001. N = 4153$

growth model made hospitalization look helpful in improving mothers' health from each wave to the next wave: $b = 0.20$ (Wave 1 hospitalization predicting improving health from Wave 1 to Wave 2) and $b = 0.12$ (Wave 2 hospitalization predicting improving health from Wave 2 to Wave 3), $ps < 0.01$, Figure 4, Plot B.

4 Discussion

Despite being well-known for over 50 years, the implications of Lord's (1967) paradox for multi-wave longitudinal analyses have not been well understood. The current study used four examples of corrective actions to illustrate Lord's paradox in three-wave longitudinal analyses. As expected, results from the difference-score approach (e.g., latent growth models) contradicted results from the residualized change score approach (cross-lagged panel models), just as in two-wave analyses. All four corrective actions looked effective according to latent growth models but harmful according to cross-lagged panel models. This may help explain why longitudinal analyses of residualized change scores have been unable to find effective parental responses to perceived child problems, such as persistent defiance, smoking, and precocious sex (Larzelere et al., 2018). The bias

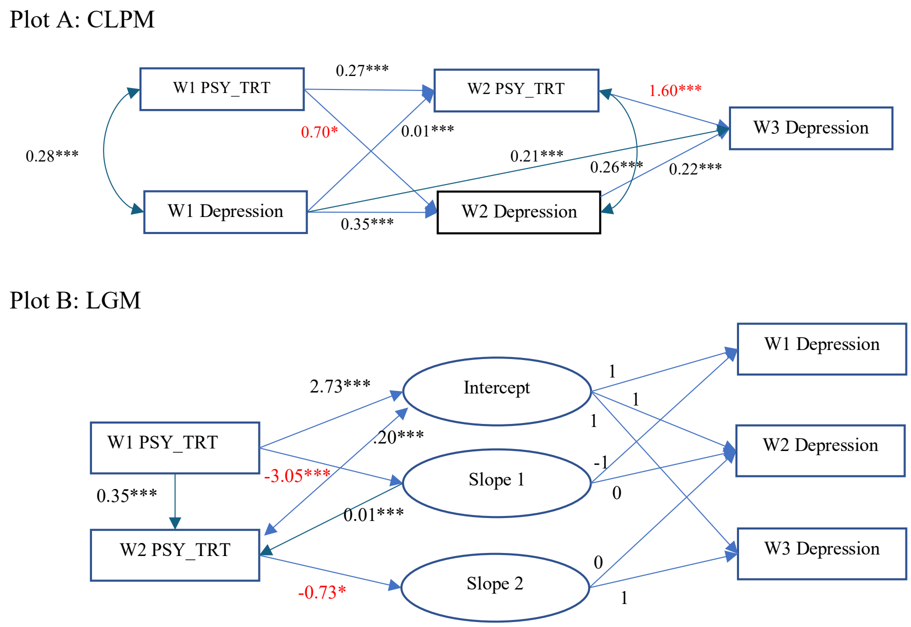

Figure 3. Cross-Lagged Panel Model (CLPM) and Latent Growth Model (LGM) of Mothers' Depression across three waves of data, predicted by Psychotherapy (PSY_TRT). $^{*}p < .05;^{**}p < .01;^{***}p < .001. N = 4588$

in residualized change score analyses helps parenting researchers confirm what they oppose (e.g., spanking), but hinders their efforts to document more effective corrective actions to replace it. The failure to find more effective corrective disciplinary responses in basic parental discipline research may help explain why clinical treatments for conduct problems in children (mostly implemented by parents) have not improved in effectiveness over the past 50 years (Weisz et al., 2019). In any case, it is worrisome that the kinds of analyses considered to be sufficient causal evidence to oppose harsh discipline practices such as spanking make most corrective actions by professionals look harmful also (Larzelere et al., 2018). These results can be explained by systematic biases recently elucidated in ANCOVA-type longitudinal analyses, because they confound within-person changes with invariant between-person differences, which are already reflected in the initial outcome scores (Berry & Willoughby, 2017; Hamaker et al., 2015; Hoffman, 2015).

Note that the four corrective actions in this study are all considered to be effective on average. Their effectiveness has been demonstrated in meta-analyses of randomized trials for psychotherapy for depression (Cuijpers et al., 2023) and time-out for oppositional defiance (Larzelere, Gunnoe, Roberts, Lin, & Ferguson, 2020) , whereas disciplinary reasoning and hospital-based treatments are widely considered to be effective in most cases. These results add to a wide range of

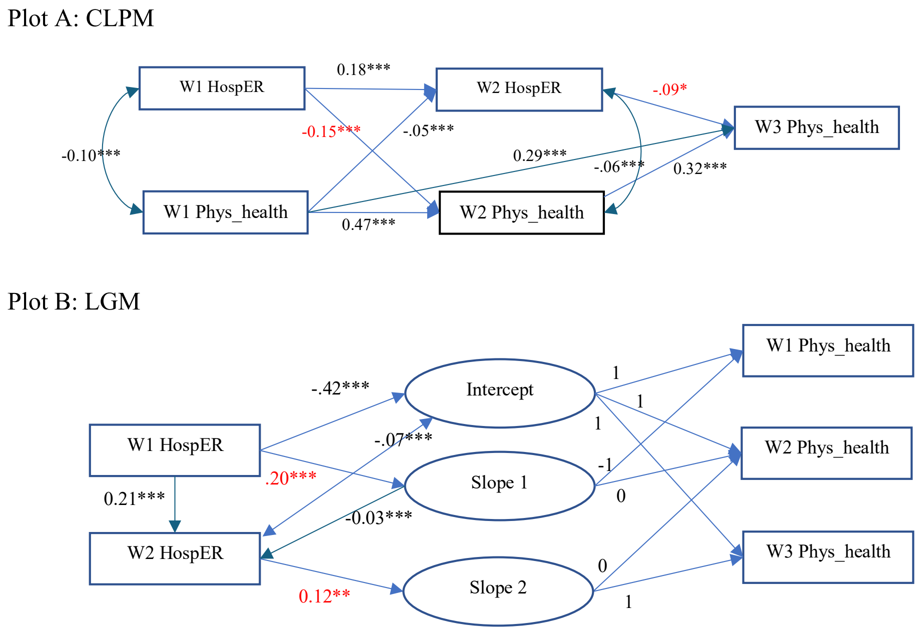

Figure 4. Cross-Lagged Panel Model (CLPM) and Latent Growth Model (LGM) of Mothers' Health across three waves of data. HospER = Overnight hospitalization or Emergency room visit. $^{*}p < .05;^{**}p < .01;^{***}p < .001. N = 4588$

corrective actions shown (incorrectly) to be significantly harmful in longitudinal analyses of residualized change scores whether implemented by parents or professionals (Larzelere et al., 2018).

Does this mean that difference-score analyses are always less biased than residualized change score analyses? Not necessarily. If the covariates account perfectly for selection into treatment conditions, ANCOVA is unbiased (Van Breukelen, 2013). The problem is that covariates fall short of this ideal in comprehensiveness, validity, and reliability in most longitudinal analyses. Steiner, Cook, Shadish, and Clark (2010) showed that ANCOVA can approximate unbiased causal effects when the covariates include baseline scores on the outcome and variables that account for self-selection into treatment conditions. Their study compared self-selection by college students into exercises to improve either math or vocabulary. It is unclear how well their results generalize to other situations in which self-selection is less well understood and poorly represented in the covariates. Note that the current study adjusted only for baseline scores on the outcome, with no additional covariates to account for why people with the same problem severity selected the corrective action of interest or not.

One factor is that the bias in ANCOVA-type analyses of residualized change scores is larger when the treatment conditions differ greatly in baseline scores on the outcome. When pretest group mean scores differ, the assumption of independence between covariates and treatment of ANCOVA is violated. Violations of this assumption usually imply invariant between-group differences, which have been shown recently to bias analyses of residualized change scores (Berry & Willoughby, 2017; Hamaker et al., 2015). The bias occurs because within-person change following the corrective action is confounded or "smushed" with between-person differences that are unchanging (Hoffman, 2015).

In contrast, independence of treatment condition and baseline scores is not an assumption of the difference-score approach, making it free from that particular bias. Nor are difference-score analyses biased by measurement error in its baseline scores, whereas residualized change score analyses are known to be biased by that measurement error. In contrast to residualized change score approaches, the difference-score approach ignores between-person differences except for differences due to within-person changes in the time period studied. In randomized studies, between-person differences that precede the treatment are removed, so that any between-person differences at post-test are due *only* to within-person changes due to the treatment conditions. In non-randomized studies, difference-score analyses can have their own unique biases, such as regression toward the mean, but the results of this and other studies of corrective actions suggest that difference-score models are often less biased than are residualized change score models, such as cross-lagged panel models.

What can be done to improve the causal validity of longitudinal analyses? The first step is to recognize the problem. One improvement would be to follow the example of econometricians in checking the robustness of results across multiple types of analyses (Duncan, Engel, Claessens, & Dowsett, 2014). Angrist and Pischke (2009) showed that these two types of change-score analyses will bracket the true causal effect under some assumptions, but it can be difficult to tell whether those assumptions are satisfied. For example, the true effect of job training programs was outside this bracket for both men and women in Lalonde's (1986) classic study. Robustness across both types of change-score analyses is therefore consistent with an unbiased causal effect, but does not guarantee it (Lin & Larzelere, 2020).

Statisticians continue to expand the options for improving the capability of longitudinal analyses to approximate less biased causal inferences (e.g., Zyphur et al., 2020). Whereas simulations of statistical innovations are generally based on conditions that may not apply to real data (e.g., the possibility of avoiding all specification errors), illustrations with actual data have rarely shown which types of analyses can correctly recover the same direction of effectiveness for corrective actions that have been documented in randomized trials. Confidence in causal inferences from longitudinal analyses can be strengthened by showing that statistical innovations can make longitudinal analyses agree with unbiased causal evidence from randomized trials.

Acknowledgments

This research was supported by a National Institutes of Health (Grant number R03 HD107307) NICHD R03 grant.

References

Achenbach, T. M. (1991). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. University of Vermont Department of Psychiatry.

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA Preschool Forms & Profiles*. University of Vermont Department of Psychiatry.

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly harmless econometrics: An empiricist's approach*. Princeton University Press. doi: https://doi.org/10.1515/9781400829828

Berry, D., & Willoughby, M. T. (2017). On the practical interpretability of cross-lagged panel models: Rethinking a developmental workhorse. *Child Development*, *88*(4), 1186–1206. doi: https://doi.org/10.1111/cdev.12660

Cronbach, L. J., & Furby, L. (1970). How we should measure "change": Or should we? *Psychological Bulletin*, *74*(1), 68–80. doi: https://doi.org/10.1037/h0029382

Cuijpers, P., Miguel, C., Harrer, M., Plessen, C. Y., Ciharova, M., Ebert, D., & Karyotaki, E. (2023). Cognitive behavior therapy vs. control conditions, other psychotherapies, pharmacotherapies and combined treatment for depression: A comprehensive meta-analysis including 409 trials with 52,702 patients. *World Psychiatry*, *22*(1), 105–115. doi: https://doi.org/10.1002/wps.21069

Duncan, G. J., Engel, M., Claessens, A., & Dowsett, C. J. (2014). Replication and robustness in developmental research. *Developmental Psychology*, *50*(11), 2417–2425. doi: https://doi.org/10.1037/a0037996

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102–116. doi: https://doi.org/10.1037/a0038889

Hoffman, L. (2015). *Longitudinal analysis: Modeling within-person fluctuation and change*. Routledge.

Kessler, R. C., Andrews, G., Mroczek, D., Ustun, B., & Wittchen, H. U. (1998). The World Health Organization Composite International Diagnostic Interview – Short Form (CIDI-SF). *International Journal of Methods in Psychiatric Research*, *7*(4), 171–185. doi: https://doi.org/10.1002/mpr.47

LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review*, *76*(4), 604–620.

Larzelere, R. E., Ferrer, E., Kuhn, B. R., & Danelia, K. (2010). Differences in causal estimates from longitudinal analyses of residualized versus simple gain scores: Contrasting controls for selection and regression artifacts. *International Journal of Behavioral Development*, *34*(2), 180–189. doi: https://doi.org/10.1177/0165025409351386

Larzelere, R. E., Gunnoe, M. L., Roberts, M. W., Lin, H., & Ferguson, C. J. (2020). Causal evidence for exclusively positive parenting and for timeout: Rejoinder to Holden, Grogan-Kaylor, Durrant, and Gershoff (2017). *Marriage & Family Review*, *56*(4), 287–319. doi: https://doi.org/10.1080/01494929.2020.1712304

Larzelere, R. E., Lin, H., Payton, M. E., & Washburn, I. J. (2018). Longitudinal biases against corrective actions. *Archives of Scientific Psychology*, *6*(1), 243–250. doi: https://doi.org/10.1037/arc0000052

Lin, H., & Larzelere, R. E. (2020). Dual-centered ANCOVA: Resolving contradictory results from Lord's paradox with implications for reducing bias in longitudinal analyses. *Journal of Adolescence*, *85*, 135–147. doi: https://doi.org/10.1016/j.adolescence.2020.11.001

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin*, *68*, 304–305. doi: https://doi.org/10.1037/h0025105

Reichman, N. E., Teitler, J. O., Garfinkel, I., & McLanahan, S. S. (2001). Fragile families: Sample and design. *Children and Youth Services Review*, *23*(4-5), 303–326. doi: https://doi.org/10.1016/S0190-7409(01)00141-4

Selig, J. P., & Little, T. D. (2012). Autoregressive and cross-lagged panel analysis for longitudinal data. In B. Laursen, T. D. Little, & N. A. Card (Eds.), *Handbook of developmental research methods* (pp. 265–278). New York, NY, US: Guilford Press.

Steiner, P. M., Cook, T. D., Shadish, W. R., & Clark, M. H. (2010). The importance of covariate selection in controlling for selection bias in observational studies. *Psychological Methods*, *15*(3), 250–267. doi: https://doi.org/10.1037/a0018719

Straus, M. A., Hamby, S. L., Finkelhor, D., Moore, D. W., & Runyan, D. (1998). Identification of child maltreatment with the Parent-Child Conflict Tactics Scales: Development and psychometric data for a national sample of American parents. *Child Abuse and Neglect*, *22*(4), 249–270. doi: https://doi.org/10.1016/s0145-2134(97)00174-9

Van Breukelen, G. J. P. (2013). ANCOVA versus CHANGE from baseline in nonrandomized studies: The difference. *Multivariate Behavioral Research*, *48*(6), 895–922. doi: https://doi.org/10.1080/00273171.2013.831743

Weisz, J. R., Kuppens, S., Ng, M. Y., Vaughn-Coaxum, R. A., Ugueto, A. M., Eckshtain, D., & Corteselli, K. A. (2019). Are psychotherapies for young people growing stronger? Tracking trends over time for youth anxiety, depression, attention-deficit/hyperactivity disorder, and conduct problems. *Perspectives on Psychological Science*, *14*(2), 216–237. doi: https://doi.org/10.1177/1745691618805436

Zyphur, M. J., Allison, P. D., Tay, L., Voelkle, M. C., Preacher, K. J., Zhang, Z., . . . Diener, E. (2020). From data to causes I: Building a general cross-lagged panel model (GCLM). *Organizational Research Methods*, *23*(4), 651–687. doi: https://doi.org/10.1177/1094428119847278

Journal of Behavioral Data Science, 2024, *4* (2), 64–79.
DOI: https://doi.org/10.35566/jbds/shan

Exploring the Impact of Social Media Usage and Sports Participation on High School Students' Mental Health and Academic Confidence

Yilin Elaine Shan[1] and Xin Tong[2]

[1] Piedmont Hills High, USA
elaineshan70@gmail.com
[2] Department of Psychology, University of Virginia, USA
xt8b@virginia.edu

Abstract. This study investigates the effects of sports participation and social media use on high school students' mental health and self-perception, with a focus on understanding their unique contributions to happiness and academic confidence. Structural equation modeling was applied to analyze the relationships between sports participation, time spent on social media, and self-reported levels of happiness and confidence, while accounting for potential gender differences. The results indicate that sports participation is positively associated with happiness, but does not significantly affect academic confidence. In contrast, the use of social media is negatively associated with academic confidence, although it does not significantly impact happiness. Gender differences were observed, with female students reporting a lower level of happiness but a higher level of academic confidence. These findings suggest that while extracurricular activities, such as sports varsity involvement, can support students' well-being, the excessive use of social media apps may undermine their academic confidence.

Keywords: High school students · Social media usage · Sport participation · Mental health · Happiness · Academic confidence

1 Introduction

High school students face numerous challenges, including academic pressures and social dynamics, that significantly impact their mental health (Pascoe et al., 2020). In recent years, awareness of mental health issues within this demographic has increased substantially. According to the *Youth Risk Behavior Survey Data Summary* & *Trends Report: 2013–2023* (Centers for Disease Control and Prevention (CDC), 2023), 40% of high school students reported experiencing persistent feelings of sadness or hopelessness in 2023. Contributing factors include pressure to achieve high grades, participation in extracurricular activities, and the complexities of maintaining social relationships. In addition, the transition from adolescence to adulthood is a critical period that requires targeted mental health support to promote resilience. To alleviate stress, the use of social media and the participation in sports have been identified as potential resources (e.g., Eather, Wade, Pankowiak, & Eime, 2023; Orsolini et al., 2022).

Social media has become an integral part of daily life, enabled by the widespread accessibility of modern devices. It offers valuable opportunities for social connection and emotional support, fostering relationships and community in the digital age. A recent survey indicates that nearly 93% of teens in America use social media platforms (Sentiment.io, 2024), engaging in activities such as entertainment, social interaction, and maintaining interpersonal relationships (Ifinedo, 2016). However, its widespread use has sparked ongoing debate about its impact on mental well-being (Abiddine, Aljaberi, Gadelrab, Lin, & Muhammed, 2022; Bekalu, McCloud, & Viswanath, 2019).

Social media plays a dual role in the lives of high school students, offering both significant benefits and notable challenges. On the one hand, it facilitates social connectivity, enabling students to maintain relationships with friends and family, even across long distances, and to expand their networks by connecting with like-minded individuals. It also provides emotional support, as students can share their feelings and experiences and often receive encouragement from their peers (Shensa et al., 2016). In addition, social media grants students access to an abundance of information and resources, helping them stay updated on current events, educational opportunities, and tools that foster learning and personal growth (Westerman, Spence, & Van Der Heide, 2014). These positive aspects highlight social media's potential to enhance social, emotional, and intellectual development.

On the other hand, the extensive use of social media also presents significant challenges. Issues such as cyberbullying, pressure to maintain an idealized online image, and reduced face-to-face interactions can have detrimental effects on users. A recent meta-analysis by Marciano, Lin, Sato, Saboor, and Viswanath (2024) explored the relationship between social media use and positive well-being. The study found that hedonic well-being characterized by positive emotions and life satisfaction is positively associated with social media communication and positive online experiences, but negatively correlated with problematic social media use, highlighting the dual nature of social media's impact on mental well-being. Moreover, excessive engagement with social media for video content, often distracts students from academic responsibilities and learning activities (Akter, 2014).

In contrast to the challenges posed by social media, sports participation provides a unique and effective outlet for high school students, combining physical activity and teamwork as a means of stress relief. Sports offer substantial benefits for mental and physical health (e.g., Fossati et al., 2021; Pascoe et al., 2020). Engaging in sports allows students to channel stress through physical activity, which releases endorphins and improves mood (Alam & Rufo, 2019; Fossati et al., 2020). Moreover, being part of a team fosters a sense of community and belonging, boosting self-esteem and confidence (Haim-Litevsky, Komemi, & Lipskaya-Velikovsky, 2023). The discipline and structure inherent in sports also promote the development of time management skills and resilience (Martín-Rodríguez et al., 2024). Regular physical activity has also been linked to better concentration, improved sleep quality, and enhanced overall physical fitness, all of which contribute to a healthier lifestyle. By incorporating sports into their routines, high school students can adopt a balanced approach to navigating academic and social pressures, ultimately improving their overall well-being.

Despite a growing body of research, several gaps remain in the literature regarding the impact of social media and sports on high school students. The mechanisms through which these activities influence mental health-whether positively or negatively-are still not fully understood. For example, while some studies suggest that social media fosters social connectivity and emotional support, others highlight its potential to contribute to feelings of inadequacy, anxiety, and depression. A more nuanced understanding of the contexts, patterns, and types of social media use (e.g., active versus passive use, positive versus negative interactions) is needed to clarify these conflicting findings.

Furthermore, limited knowledge exists about which social media platforms are most popular among high school students, how much time they spend on these platforms, and how platform-specific features influence their mental well-being. Factors such as algorithm-driven content exposure, platform design, and peer interactions may play a significant role but remain under explored. Similarly, the role of individual differences-such as gender, socioeconomic status, or personality traits-in moderating the effects of social media is not well-documented.

To address these gaps, the current study investigates the complex relationships between social media use, sports participation, mental health, and academic confidence among high school students. Specifically, it explores how different patterns of social media engagement and various types of sports activities contribute to students' well-being, including their mental health, happiness, and academic confidence. By examining these factors, the study seeks to clarify the mechanisms underlying these influences and to identify potential areas where targeted interventions and support strategies can enhance the well-being of high school students.

Building on this objective, we collected data on high school students' happiness, academic confidence, social media usage, and sports participation. The study examines how these activities influence students' happiness and academic confidence. Additionally, it explores the mediating role of happiness in these relationships, providing deeper insights into the mechanisms through which these activities affect students' overall well-being. Since happiness and academic confidence are latent constructs, each measured by multiple indicators, a structural equation modeling (SEM) approach is used to model the relationships among these variables while accounting for measurement error (e.g., Bollen, 1989; Lee & Song, 2012; Merkle & Rosseel, 2015; Muthén & Asparouhov, 2012).

The remainder of the article is structured as follows. First, we introduce the high school student mental health dataset collected for this study. Next, we describe the data analysis procedures and present the results. Finally, we conclude with a discussion of the findings and their implications.

2 High School Student Mental Health: An Empirical Example

In this section, we introduce the participants, describe the measurements used for data collection, and summarize the key characteristics of the dataset.

2.1 Participants

This study involved 51 students from a public high school in San Jose, California. Among the participants, there were 26 female students, 24 male students, and one non-binary student. The grade levels ranged as follows: 2 ninth graders, 11 tenth graders, 17 eleventh graders, and 21 twelfth graders. Additionally, 25 participants were members of high school varsity sports teams, while the remaining 26 were non-athletes.

2.2 Procedures

To understand the impact of sports participation and social media usage on happiness and confidence among high school students, data collection was conducted using a Google Form containing items on demographics, sports participation, social media usage, and self-reported happiness and confidence levels. The survey was distributed among students, specifically targeting members of sports teams and their classmates, to gather diverse perspectives.

2.3 Measurement

We collected self-reported data on students' happiness and academic confidence, both measured as latent constructs using multiple indicators. Additionally, we gathered data on students' sports participation and social media usage, including their preferred platforms and the among of time spent on these platforms. Students also provided their perceptions of how sports activities affected their happiness and academic confidence.

Happiness and Academic Confidence To measure happiness and academic confidence, we developed a 10-item scale [3].

Each of the two latent constructs-happiness and academic confidence-is measured by 5 items. Example items include, "I generally feel happy in my daily

[3] Refer to the Appendix

life" for happiness and "I believe in my ability to perform well in exams and assessments" for academic confidence. All 10 items were rated on a 5-point Likert scale (1 = strongly disagree, 2 = somewhat disagree, 3 = neither agree nor disagree, 4 = somewhat agree, and 5 = strongly agree).

The reliability (α, Cronbach, 1951) for the happiness construct is 0.79, with the 95% confidence interval [0.68, 0.87]. For academic confidence, α is 0.62, with the 95% confidence interval [0.42, 0.76].

Social Media Usage To understand social media usage among high school students, we included an open-ended question asking them to report their favorite social media apps. Students were allowed to list more than one app if applicable. Figure 1 illustrates the frequencies of the social media apps indicated by the students as their favorite apps.

In this data set, students reported a total of nine social media applications as their favorite applications. Among the nine applications, YouTube was the most popular, with 34 students (around two-thirds) listing it as their favorite app. Instagram was also highly favored, with 29 students listing it as their favorite app.

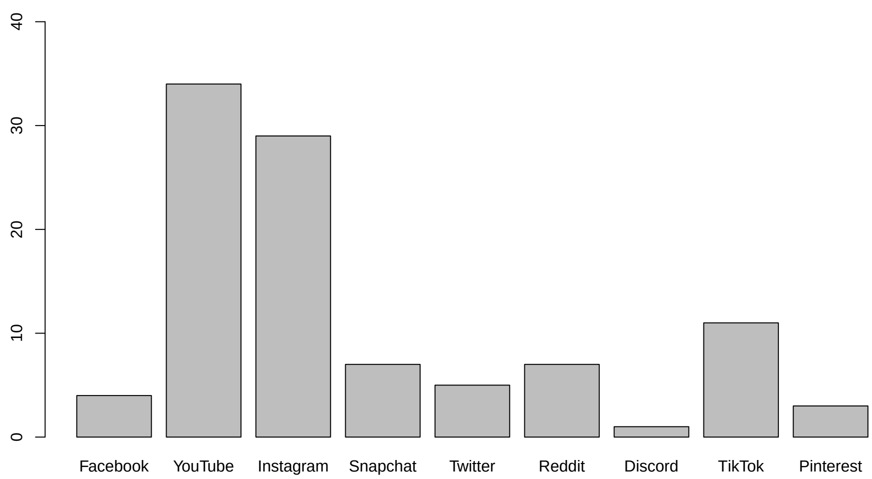

Figure 1. Comparison of student preferences for social media applications

Moreover, students also reported how many hours they spend on social media per day. The distributions are included in Figure 2. The plot illustrates the distribution of time spent on social media among students, categorized into four

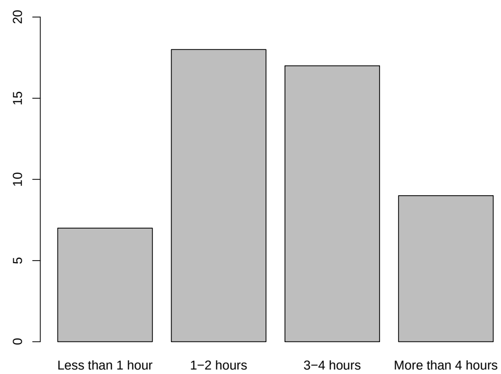

Figure 2. The average time that students spend on social media per day

groups: "Less than 1 hour," "1-2 hours," "3-4 hours," and "More than 4 hours." The majority of students fall into the "1-2 hours" category, indicating that moderate daily social media usage is the most common. As the time spent on social media increases, the number of students decreases, with fewer reporting "3-4 hours" of use and an even smaller proportion in the "More than 4 hours" category. Additionally, a smaller yet notable group of students reported "Less than 1 hour" of daily social media use, suggesting that a subset engages minimally with social media. These findings highlight the varying levels of social media engagement among students.

Sports Participation Among the 51 students, 25 were members of a sports team at the time of the survey, while the remaining 26 were not. Of the students on sports teams, 11 had been members for 1 or 2 years, 12 had been members for 3 or 4 years, and 3 had been members for less than a year.

Perceived Impact of Sports on Happiness and Confidence Students also described how sports activities impacted their happiness and confidence, using a scale ranging from "Sports significantly decrease my happiness /confidence" to "Sports significantly increase my happiness/confidence." The responses are summarized in Figure 3.

The majority of students reported either an increase or a significant increase in happiness as a result of sports participation, with a smaller proportion indicating "No Impact" and only a few reporting negative effects. A similar pattern was observed for academic confidence, where most students experienced positive changes. However, slightly fewer students reported an increase in academic confidence compared to happiness, with a notable portion indicating "No Impact."

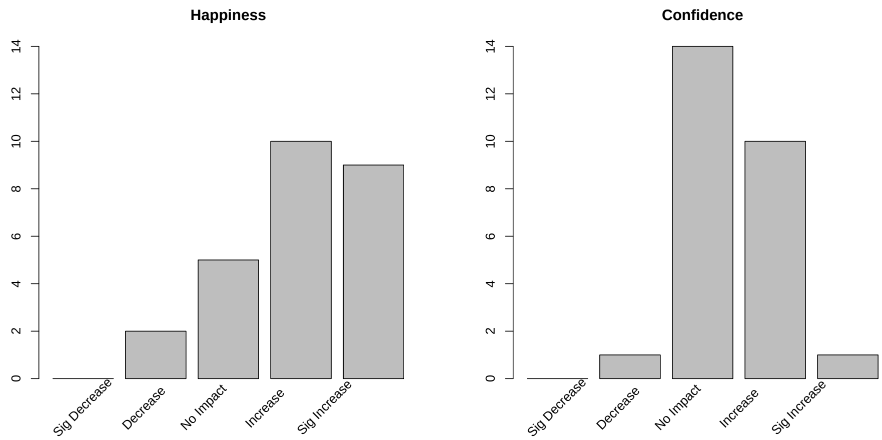

Figure 3. Self-reported impact of sports participation on students' happiness and confidence

2.4 Overview of Data Analysis

This analysis aims to examine the impact of sports participation and social media on high school students' mental health and academic confidence. Specifically, we aim to address the following questions: (1) Does sport participation positively influence students' happiness and/or academic confidence? (2) Does social media enhance students' happiness but hinder academic confidence? (3) How are happiness and academic confidence related? (4) Are there gender-based differences in happiness and the academic confidence among students?

3 Data Analysis and Results

Since both happiness and academic confidence are latent variables measured by five indicators each, we will use structural equation models (Garnier-Villarreal & Jorgensen, 2020) that incorporate a measurement model for the latent variables and a structural model to assess the hypothesized relationships between them.

3.1 Model

In LISREL notation (Byrne, 1989), an SEM model typically includes the measurement and structural components. The measurement model describes the measurement structures of the latent variables:

$$\begin{cases} \mathbf{x} = \mathbf{\Lambda}_x \xi + \boldsymbol{\delta} \\ \mathbf{y} = \mathbf{\Lambda}_y \eta + \boldsymbol{\varepsilon} \end{cases} \tag{1}$$

where $\mathbf{x}$ represents observed indicators for the exogenous latent variables (ξ), $\mathbf{y}$ represents observed indicators for the endogenous latent variables (η), $\mathbf{\Lambda}_x$ and $\mathbf{\Lambda}_y$ are the factor loading matrices for exogenous and endogenous variables, respectively. The symbols $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$ are the measurement errors of indicators.

The structural model assess the relationship among latent variables and with predictors

$$\eta = \mathbf{B}\eta + \mathbf{\Gamma}\xi + \mathbf{\Pi}\mathbf{Z} + \boldsymbol{\zeta} \tag{2}$$

where $\mathbf{B}$ is the matrix of regression coefficients among endogenous latent variables $\mathbf{\Gamma}$ is the matrix of regression coefficients from exogenous latent variables (ξ) to endogenous latent variables (η), $\mathbf{\Pi}$ is the matrix of regression coefficients from observed predictors $\mathbf{Z}$ to endogenous latent variables, and $\boldsymbol{\zeta}$ represents the disturbance terms.

In the current analysis, the endogenous latent variables (η) include happiness and the academic confidence. The manifest predictors $\mathbf{Z}$ includes the sport varsity participation (yes=1, no=0), gender (girl=1, boy=0), and app use (in hours).

We fit the model using the *lavaan* package (Rosseel, 2012) within the *R* platform (R Core Team, 2020). The initial model was specified without cross-loadings or correlated residuals, and all models were estimated using the maximum likelihood approach (Jöreskog, 1967). After fitting the initial model, we conducted modifications based on the largest Modification Index (MI) to improve model fit by incorporating the suggested paths. The final model, refined through this process, is illustrated in Figure 4.

3.2 Results

The model parameter estimates and fit indices are presented in Table 1. When compared to the saturated model (i.e., a model with a perfect fit), the hypothesized model shown in Figure 4 demonstrated a good fit, with a Chi-square statistic of 45.538(df=40) and a p-value of 0.253, indicating that our model is not significantly worse than the saturated model. The fit indices further support this conclusion: the Root Mean Square Error of Approximation (RMSEA) is 0.052, the Comparative Fit Index (CFI) is 0.965, and the Tucker-Lewis Index (TLI) is 0.945, all of which indicate a good model fit.

Examining the parameter estimates, we find that **sports varsity participation** (1=yes, 0=no) positively predicts happiness (Est = 0.371, p-value = 0.035), suggesting that students involved in sport varsity report higher happiness. However, **hours spent on social media applications** (App_hour) did not significantly predict happiness (Est = 0.077, p-value = 0.430). The coefficient estimate of gender (female=1, male=0) is −0.433 with p-value 0.019. The result indicates that female students are less happier than male students.

For academic confidence, sports participation was not a significant predictor (Est = −0.143, p-value = 0.499). In contrast, hours spent on social media negatively predicted confidence (Est = −0.373, p-value = 0.003), suggesting that more time on social media may be associated with lower academic confidence. In

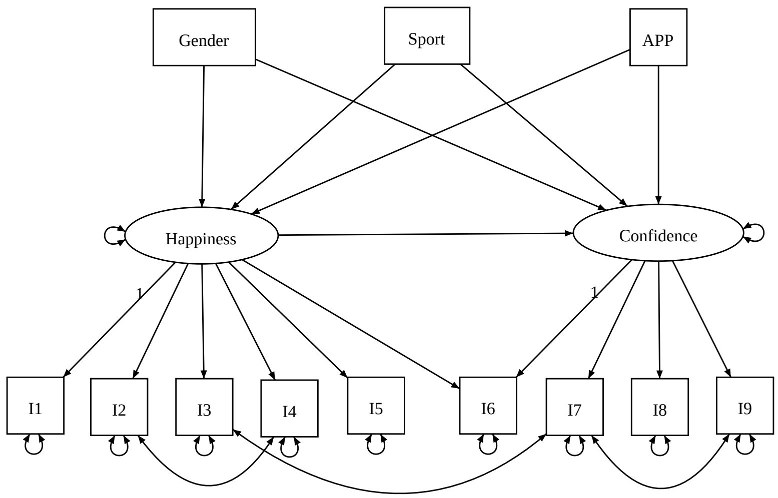

Figure 4. Path diagram of the CFA model of Imagination and Extraversion

addition, female students are more confident in academic performance than male students with the coefficient estimate of gender on confidence being 0.602 and the p-value being 0.013. Moreover, happiness, though not significant, showed a positive trend in predicting confidence (Est = 2.981, p-value = 0.129).

Par		Est	Std.Err	z-value	$P(>\|z\|)$
Happy∼	Sport	0.371	0.176	2.104	0.035
	App_hour	0.077	0.097	0.790	0.430
	Gender	-0.433	0.185	-2.337	0.019
Confidence∼	Sport	-0.143	0.212	-0.676	0.499
	App_hour	-0.373	0.128	-2.925	0.003
	Gender	0.602	0.242	2.484	0.013
	Happy	2.981	1.966	1.517	0.129
$\mathcal{X}^2(40)$	45.538	p-value=0.253			
CFI	0.965		TLI	0.945	
RMSEA	0.052				

Table 1. Parameter estimates of the regression coefficients and the model fit indices

4 Concluding Remarks

This study examined the effects of sports participation and the use of social media on high school students' happiness and academic confidence. By employing a structural equation modeling approach, the research assessed the relationships among these variables while accounting for gender difference, the time spent on social media, and the participation in sports varsity.

4.1 Highlights of the Results

The results reveal that sports participation and social media have distinct influences on high school students' happiness and academic confidence. Specifically, sports participation was positively associated with happiness, indicating that engaging in sports may enhance overall well-being. However, it did not have a significant impact on academic confidence. Conversely, increased time spent on social media was negatively associated with academic confidence but showed no significant effect on happiness levels. While happier students tended to report higher academic confidence overall, this relationship was not statistically significant in the current dataset. Gender differences were observed, with female students reporting a lower levels of happiness and a higher level of academic confidence than male students.

4.2 Discussion

The findings of this study reveal notable patterns in the popularity of social media applications among high school students. YouTube emerges as the most widely favored platform, with a significant majority of students identifying it as their preferred app. Instagram follows as the second most popular choice, emphasizing its appeal for photo sharing and social interaction. While other platforms are less commonly selected, they still maintain a presence among students, reflecting the diverse preferences within this demographic.

In addition to insights on social media preferences, the study highlights the complex relationships between sport activities, social media use, and high school students' mental health and academic confidence. The positive association between sports participation and happiness aligns with existing literature, suggesting that physical activity and team engagement can enhance well-being and provide students with a sense of community and purpose. However, the lack of a significant link between sports participation and academic confidence implies that while sports may enhance general happiness, they do not necessarily impact students' perceptions of their academic abilities.

In contrast, the negative impact of social media use on academic confidence highlights growing concerns about its influence, arising from social comparison or distractions from academic responsibilities. Importantly, the observed gender differences in happiness and confidence, with female students reporting a lower level of happiness and a higher level of academic confidence, emphasize the importance of implementing tailored support strategies to address these disparities.

These findings suggest that promoting balanced social media use and encouraging participation in sports can play a vital role in supporting students' mental well-being. However, effective educational interventions must also consider the unique challenges and needs of different gender groups to foster both mental health and academic confidence.

4.3 Future Considerations

This study offers valuable insights into the relationships between sports participation, social media use, happiness, and academic confidence among high school students, employing a structural equation modeling approach. However, a few of methodological limitations should be addressed in future research.

First, the cross-sectional design limits the ability to draw causal inferences. Future studies should consider longitudinal or experimental designs to better capture the dynamic effects of sports participation and social media over time, helping to clarify causal pathways and temporal relationships.

Second, the reliability and validity of the scales used to measure happiness and academic confidence require further calibration to ensure robust results. Additionally, the reliance on self-reported data introduces potential biases, such as social desirability bias, which may compromise the accuracy of the findings. Future research should consider incorporating objective measures, such as digital tracking of social media use and external assessments of academic confidence, to strength the reliability and validity of the data.

Additionally, this study was conducted in a single school in the Bay Area, which may restrict the ability of generalization of the findings to broader populations, such as students from the Midwest or other regions with distinct cultural and socioeconomic contexts. Expanding the sample to include schools from diverse regions and demographic backgrounds would enhance the external validity and applicability of the results.

Lastly, while this study examined gender differences, future research should explore other moderating factors, such as socioeconomic status, academic achievement, and family support, to gain a more subtle understanding of how these variables interact with social media use and sports participation. Addressing these methodological considerations would provide a more comprehensive understanding of the factors influencing high school students' mental health and academic confidence, offering a stronger foundation for effective interventions.

References

Abiddine, F. Z. E., Aljaberi, M. A., Gadelrab, H. F., Lin, C.-Y., & Muhammed, A. (2022). Mediated effects of insomnia in the association between problematic social media use and subjective well-being among university students during COVID-19 pandemic. *Sleep Epidemiology*, *2*, 100030. doi: https://doi.org/10.1016/j.sleepe.2022.100030

Akter, T. (2014). Social media addiction, resistance, and influence of awareness: Measurement of psychology students' resistance to facebook addiction. *Mediterranean Journal of Social Sciences*, *5*(8), 456–464. doi: https://doi.org/10.18295/squmj.2015.15.03.009

Alam, N., & Rufo, N. (2019, 02). Fitness as a component of mental health intervention. *Health & Social Work*, *44*(2), 129-132. doi: https://doi.org/10.1093/hsw/hlz001

Bekalu, M. A., McCloud, R. F., & Viswanath, K. (2019). Association of social media use with social well-being, positive mental health, and self-rated health: Disentangling routine use from emotional connection to use. *Health Education & Behavior*, *46*(2_suppl), 69S-80S. doi: https://doi.org/10.1177/1090198119863768

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: NY: John Wiley & Sons, Inc.

Byrne, B. (1989). *A primer of LISREL: Basic applications and programming for confirmatory factor analytic models.* Springer-Verlag.

Centers for Disease Control and Prevention (CDC). (2023). *The youth risk behavior survey data summary & trends report: 2013–2023.* Retrieved from `https://www.cdc.gov/healthyyouth/data/yrbs/pdf/YRBSDataSummary-TrendsReport2013-2023.pdf`

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *5*(8), 297–334. doi: https://doi.org/10.1007/BF02310555

Eather, N., Wade, L., Pankowiak, A., & Eime, R. (2023). The impact of sports participation on mental health and social outcomes in adults: A systematic review and the 'mental health through sport'conceptual model. *Systematic Reviews*, *12*(1), 102. doi: https://doi.org/10.1186/s13643-023-02264-8

Fossati, C., Torre, G., Borrione, P., Giombini, A., Fagnani, F., Turchetta, M., . . . Pigozzi, F. (2020). Biohumoral indicators influenced by physical activity in the elderly. *Journal of Clinical Medicine*, *9*(4), 1115. doi: https://doi.org/10.3390/jcm9041115

Fossati, C., Torre, G., Vasta, S., Giombini, A., Quaranta, F., Papalia, R., & Pigozzi, F. (2021). Physical exercise and mental health: The routes of a reciprocal relation. *International journal of environmental research and public health*, *18*(23), 12364. doi: https://doi.org/10.3390/ijerph182312364

Garnier-Villarreal, M., & Jorgensen, T. D. (2020). Adapting fit indices for bayesian structural equation modeling: Comparison to maximum likelihoods. *Psychological Methods*, *25*(1), 46–70. doi: https://doi.org/10.1037/met0000224

Haim-Litevsky, D., Komemi, R., & Lipskaya-Velikovsky, L. (2023). Sense of belonging, meaningful daily life participation, and well-being: Integrated investigation. *International journal of environmental research and public health*, *20*(5), 4121. doi: https://doi.org/10.3390/ijerph20054121

Ifinedo, P. (2016). Applying uses and gratifications theory and social influence processes to understand students' pervasive adoption of social networking sites: Perspectives from the americas. *International Journal of Information Management*, *36*(2), 192-206. doi: https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2015.11.007

Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, *32*(4), 443–482. doi: https://doi.org/10.1007/BF02289658

Lee, S.-Y., & Song, X.-Y. (2012). *Basic and advanced bayesian structural equation modeling: With applications in the medical and behavioral sciences.* John Wiley & Sons.

Marciano, L., Lin, J., Sato, T., Saboor, S., & Viswanath, K. (2024). Does social media use make us happy? a meta-analysis on social media and positive well-being outcomes. *SSM - Mental Health*, *6*, 100331. doi: https://doi.org/https://doi.org/10.1016/j.ssmmh.2024.100331

Martín-Rodríguez, A., Gostian-Ropotin, L. A., Beltrán-Velasco, A. I., Belando-Pedreño, N., Simón, J. A., López-Mora, C., ... Clemente-Suárez, V. J. (2024). Sporting mind: The interplay of physical activity and psychological health. *Sports (Basel, Switzerland)*, *12*(1), 37. doi: https://doi.org/10.3390/sports12010037

Merkle, E. C., & Rosseel, Y. (2015). blavaan: Bayesian structural equation models via parameter expansion. *arXiv preprint arXiv:1511.05604*.

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological methods*, *17*(3), 313-335.

Orsolini, L., Volpe, U., Albert, U., Carmassi, C., Carrà, G., Cirulli, F., ... Fiorillo, A. (2022). Use of social network as a coping strategy for depression among young people during the COVID-19 lockdown: Findings from the COMET collaborative study. *Annals of general psychiatry*, *21*(1), 44. doi: https://doi.org/10.1186/s12991-022-00419-w

Pascoe, M., Bailey, A. P., Craike, M., Carter, P., T., R., N., Stepto, & Parker, A. (2020). Physical activity and exercise in youth mental health promotion: a scoping review. *BMJ open sport & exercise medicine*, *6*(1), e000677. doi: https://doi.org/10.1136/bmjsem-2019-000677

R Core Team. (2020). R: A language and environment for statistical computing

[Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. doi: https://doi.org/10.18637/jss.v048.i02

Sentiment.io. (2024). *How many teens use social media?* Retrieved from https://www.sentiment.io/how-many-teens-use-social-media/#:~:text=Surveys%20suggest%20that%20over%2093,widely%20used%20in%20our%20survey (Accessed: 2024-11-27)

Shensa, A., Sidani, J., Lin, L., Bowman, N., , & Primack, B. (2016). Social media use and perceived emotional support among us young adults. *Journal of community health*, *41*(3), 541-549. doi: https://doi.org/10.1007/s10900-015-0128-8

Westerman, D., Spence, P. R., & Van Der Heide, B. (2014, 01). Social media as information source: Recency of updates and credibility of information. *Journal of Computer-Mediated Communication*, *19*(2), 171-183. doi: https://doi.org/10.1111/jcc4.12041

Appendix

High School Student Happiness and Academic Confidence Survey: Sports Participation

Thank you for taking part in our survey, which aims to explore the relationship between high school students' happiness, academic confidence, and their participation in high school sports teams. Your responses will provide valuable insights into how sports involvement may impact your overall well-being and academic self-assurance. Your responses will remain anonymous and confidential.

Section I

1. Name (OPTIONAL)____________________________
2. Email (OPTIONAL)____________________________
3. Gender: ○ Male ○ Female ○ Non-Binary ○ Other ____________________________
4. Grade: ○ 9th ○ 10th ○ 10th ○ 11th
5. Cumulative GPA (unweighted) ____________________________
6. Which high school do you currently attend? ____________________________
7. Are you currently a member of a high school sports team? ○ Yes ○ No

Section II

Happiness Assessment Please rate the following statements on a scale of 1 to 5, where 1 indicates strongly disagree and 5 indicates strongly agree.

1. I generally feel happy in my daily life. ○ 1 ○ 2 ○ 3 ○ 4 ○ 5
2. I have supportive friendships. ○ 1 ○ 2 ○ 3 ○ 4 ○ 5
3. I am satisfied with my overall well-being. ○ 1 ○ 2 ○ 3 ○ 4 ○ 5
4. I feel a sense of belonging at my school. ○ 1 ○ 2 ○ 3 ○ 4 ○ 5
5. I am optimistic about my future. ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

Section III

Academic Confidence Assessment Please rate the following statements on a scale of 1 to 5, where 1 indicates strongly disagree and 5 indicates strongly agree.

1. I believe I am capable of understanding challenging subjects. ○ 1 ○ 2 ○ 3 ○ 4 ○ 5
2. I feel confident participating in classroom discussions. ○ 1 ○ 2 ○ 3 ○ 4 ○ 5
3. I am comfortable seeking help from teachers when needed. ○ 1 ○ 2 ○ 3 ○ 4 ○ 5
4. I manage my time effectively to balance schoolwork and other activities. ○ 1 ○ 2 ○ 3 ○ 4 ○ 5
5. I believe in my ability to perform well in exams and assessments. ○ 1 ○ 2 ○ 3 ○ 4 ○ 5

Section IV

Social Media Usage

1. On average, how many hours per day do you spend on social media? ○ Less than 1 hour ○ 1-2 hours ○ 3-4 hours ○ More than 4 hours
2. Please select your favorite social media app from the following list ○ Instagram ○ TikTok ○ Snapchat ○ Twitter ○ Facebook ○ Youtube ○ Pinterest ○ Reddit ○ Other____________________________

Section V

Sports Participation and Impact

1. How long have you been a member of a high school sports team? ○ Less than a year ○ 1-2 years ○ 3-4 years ○ Not applicable
2. How do you feel your involvement in a sports team has influenced your overall happiness? Significantly decreased ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 Significantly increased
3. How do you perceive the impact of sports participation on your academic confidence? Significantly decreased ○ 1 ○ 2 ○ 3 ○ 4 ○ 5 Significantly increased

Journal of Behavioral Data Science, 2024, *4* (2), 80–85.
DOI: `https://doi.org/10.35566/jbds/rodrigues`

greekLetters: Routines for Writing Greek Letters and Mathematical Symbols on the RStudio and RGui

Kévin Allan Sales Rodrigues[0000−0003−4925−5883]

University of São Paulo, São Paulo, Brazil
`kevin@alumni.usp.br`

Abstract. This is a brief description of the *greekLetters* R package. In short, *greekLetters* is a package for displaying Greek letters and various mathematical symbols in RStudio and RGui environments.

Keywords: R · Statistical notation · Mathematical symbols · Greek letters

1 Introduction

The R (R Core Team, 2024) ecosystem lacks a dedicated, streamlined package for incorporating Greek letters and mathematical symbols into text outputs seamlessly. The *greekLetters* package (Rodrigues, 2023) addresses this need, offering a comprehensive toolkit for displaying Greek letters and various mathematical symbols in RStudio (Posit team, 2024) and RGui environments.

Designed for ease of use, the package facilitates the inclusion of Greek letters and math equations in RGui and RStudio, enhancing the clarity and presentation of statistical notation. The package ensures compatibility across operating systems by encoding characters in UTF-8. Additionally, it supports the creation of summary functions that incorporate the functional form of fitted models with Greek letters, bridging the gap between statistical theory and practice. The package's simplicity and accessibility make it an essential tool for enhancing the presentation and understanding of statistical concepts in R.

The article is organized as follows. Section 2 provides the rationale for the existence of the package. Section 3 explains the syntax for using Greek letter symbols and mathematical symbols, Section 4 presents some examples of how to use the *greekLetters* package. Section 5 details where the package is available and how to install it. The final section discusses the package documentation.

2 Statement of need

Incorporating Greek letters and mathematical symbols in R outputs is essential for clear and accurate statistical notation, particularly in educational and professional settings. However, R lacks a package that fills this gap. Existing solutions are limited in scope and functionality, especially in environments like RGui where Unicode support is partial.

The *greekLetters* package (Rodrigues, 2023) fills this gap by providing functions to display Greek letters and mathematical symbols consistently across RStudio and RGui. This capability is crucial for creating clear and professional statistical outputs, enhancing the communication and understanding of statistical models and results.

Thus, the *greekLetters* package will be useful as a support package for the development of other packages, serving to create summaries with equations and mathematical symbols, as well as assisting in the communication of other R outputs. An example of a package that depends on *greekLetters* is *diagL1* (Rodrigues & Elian, 2024). It is expected that in the coming years, *greekLetters* will be used in more packages.

The next section deals with the syntax for using Greek letters and mathematical symbols.

3 Syntax

The syntax for using the symbols is simple, just use the *greeks()* function as shown in the following command.

```
# Basic syntax
greeks("math_symbol_or_Greek_letter_name")
```

To write equations, simply concatenate the symbols, which can be done using the *paste()* and *paste0()* functions. The *paste()* function uses a space, " ", as the default separator, but it allows other symbols to be defined as separators. On the other hand, the *paste0()* function does not insert any separators between the concatenated strings.

Figure 1 contains the output of the *print_greeks()* function, which lists the symbols and their respective names, allowing the *greeks()* function to generate these symbols.

The next section will present examples illustrating the utility of the package.

```
               alpha                  beta           gamma
                 "α"                   "β"             "γ"
               delta               epsilon            zeta
                 "δ"                   "ε"             "ζ"
                 eta                 theta            iota
                 "η"                   "θ"             "ι"
               kappa                lambda              mu
                 "κ"                   "λ"             "μ"
                  nu                    xi         omicron
                 "ν"                   "ξ"             "ο"
                  pi                   rho           sigma
                 "π"                   "ρ"             "σ"
                 tau               upsilon             phi
                 "τ"                   "υ"             "φ"
                 chi                   psi           omega
                 "χ"                   "ψ"             "ω"
               Alpha                  Beta           Gamma
                 "Α"                   "Β"             "Γ"
               Delta               Epsilon            Zeta
                 "Δ"                   "Ε"             "Ζ"
                 Eta                 Theta            Iota
                 "Η"                   "Θ"             "Ι"
               Kappa                Lambda              Mu
                 "Κ"                   "Λ"             "Μ"
                  Nu                    Xi         Omicron
                 "Ν"                   "Ξ"             "Ο"
                  Pi                   Rho           Sigma
                 "Π"                   "Ρ"             "Σ"
                 Tau               Upsilon             Phi
                 "Τ"                   "Υ"             "Φ"
                 Chi                   Psi           Omega
                 "Χ"                   "Ψ"             "Ω"
            infinity        leftrightarrow          forall
                 "∞"                   "⇔"             "∀"
               exist              notexist        emptyset
                 "∃"                   "∄"             "∅"
           elementof          notelementof    proportional
                 "∈"                   "∉"             "∝"
asymptoticallyEqual notasymptoticallyEqual      approxEqual
                 "≃"                   "≄"             "≅"
         almostEqual                   leq             geq
                 "≈"                   "≤"             "≥"
            muchless           muchgreater       leftarrow
                 "≪"                   "≫"             "⇐"
          rightarrow                 equal        notEqual
                 "⇒"                   "="             "≠"
            integral        doubleintegral  tripleintegral
                 "∫"                   "∬"             "∭"
          logicalAnd             logicalOr    intersection
                 "∧"                   "∨"             "∩"
               union
                 "∪"
```

Figure 1. Output of *print_greeks()* containing symbols names for *greeks()* function.

4 Examples

Here are some straightforward examples showcasing *greekLetters* utility. To denote the approximation of π, you can use:

```
# pi constant
paste(greeks("pi"), greeks("almostEqual"), "3.14")
```

The linear regression equation, in matrix form, can be elegantly displayed using Greek letters for the coefficients and error term:

```
# Linear regression
paste("y", " = X", greeks("beta"), " + ",
greeks("epsilon"), sep ="")
```

The expected value of a random variable X can be represented as:

```
# Expected value
paste("E[X] = ", greeks("integral"), "xf(x)dx",
 sep = "")
```

The notation for testing a statistical hypothesis can be shown as:

```
# Testing statistical hypothesis
paste(greeks("H_0"), ":", greeks("mu"), "= 0")
paste("versus", greeks("H_1"), ":",
greeks("mu"), greeks("notEqual"), "0" )
```

Figure 2 contains the outputs of the presented commands.

```
> # pi constant
> paste(greeks("pi"), greeks("almostEqual"), "3.14")
[1] "π ≈ 3.14"
>
> # Linear regression
> paste("y", " = X", greeks("beta"), " + ", greeks("epsilon"), sep ="")
[1] "y = Xβ + ε"
>
> # Expected value
> paste("E[X] = ", greeks("integral"), "xf(x)dx", sep = "")
[1] "E[X] = ∫xf(x)dx"
>
> # Testing statistical hypothesis
> paste(greeks("H_0"), ":", greeks("mu"), "= 0")
[1] "H₀ : μ = 0"
> paste("versus", greeks("H_1"), ":", greeks("mu"), greeks("notEqual"), "0" )
[1] "versus H₁ : μ ≠ 0"
```

Figure 2. R console outputs with mathematical equations and Greek letters.

By using the *greekLetters* package, these examples demonstrate how to effectively incorporate Greek letters and mathematical symbols into R outputs, enhancing the clarity of R statistical outputs.

The next section will explain how to install the package and describe its license, detailing what can be done with it.

5 Package availability and license

The *greekLetters* package is hosted on the official CRAN (The Comprehensive R Archive Network) repository, ensuring its reliability and easy access for all R users. The package can be found and downloaded via the following link: `https://cran.r-project.org/package=greekLetters`. Being available on CRAN ensures that the package has undergone rigorous quality and compliance checks, guaranteeing its compatibility with different versions of R and various operating systems. Additionally, its presence on CRAN facilitates the installation and updating of the package directly through R using simple commands like:

```
install.packages("greekLetters")
```

Additionally, the *greekLetters* package is licensed under the GNU General Public License version 3 (GPLv3), which provides several freedoms and responsibilities. This license allows anyone to use the software for any purpose, whether personal, educational, or commercial, and requires the source code to be available, enabling users to study, modify, and adapt the software to their needs.

Users have the right to modify the *greekLetters* code and distribute both the original code and modified versions, as long as they do so under the same terms of the GPLv3. This includes providing or making the source code available with any distribution of the software, ensuring continuous transparency and openness. When distributing greekLetters, it is necessary to keep the copyright notices and the GPLv3 license intact, ensuring that all recipients are aware of their rights and responsibilities.

The GPLv3 also requires license compatibility when combining *greekLetters* with other software and includes provisions that address patents, protecting users against claims that could restrict their freedoms. Adopting this license promotes collaboration in the free software community and ensures that the package remains free and open for future developers and users, which is crucial for a support package like *greekLetters* that can be widely integrated and used by other packages, such as *diagL1* (Rodrigues & Elian, 2024), and many others in the future.

This accessibility and ongoing support make *greekLetters* a valuable tool for anyone needing to incorporate Greek letters and mathematical symbols into their R outputs. The next and final section will provide some comments on the software documentation.

6 Manual and documentation

Extensive documentation accompanies the package, featuring detailed descriptions and examples for each function. This thorough documentation aids users in effectively utilizing the package's capabilities, ensuring they can integrate Greek letters and mathematical symbols into their R outputs with ease. The package's comprehensive testing and documentation guarantee a reliable and user-friendly experience.

The manual is available on the package's page and can be accessed without any registration. Additionally, the documentation for each function can be accessed within R using the *help()* command.

Acknowledgments

The author gratefully acknowledges CNPq (Brazilian National Council for Scientific and Technological Development) for the financial support.

References

Posit team. (2024). Rstudio: Integrated development environment for r [Computer software manual]. Boston, MA. Retrieved from http://www.posit.co/

R Core Team. (2024). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from http://www.R-project.org

Rodrigues, K. A. S. (2023). greekLetters: routines for writing greek letters and mathematical symbols on the RStudio and RGui [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=greekLetters (R package version 1.0.2) doi: https://doi.org/10.32614/CRAN.package.greekLetters

Rodrigues, K. A. S., & Elian, S. N. (2024). diagL1: Routines for fit, inference and diagnostics in linear L1 and LAD models [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=diagL1 (R package version 1.0.0) doi: https://doi.org/10.32614/CRAN.package.diagL1