Journal of Applied Econometrics and Statistics
Vol. 2, No. 1, 2023, pp. 1-21
https://DOI:10.47509/JAES.2023.v02i01.01

ARF INDIA
Academic Open Access Publishing
www.arfjournals.com

An Inferential Approach to Model-based Reliability: Applications with Health Survey Data

Junyu Nie[1] and Jihnhee Yu[1*]

[1]Department of Biostatistics, University at Buffalo, State University of New York, NY 14214, USA
**Correspondence E-mail: jinheeyu@buffalo.edu*

Received: 21 January 2023; Revised: 12 February 2023;
Accepted 15 March 2023; Publication: 26 June 2023

Abstract: We propose methods to obtain the variability of model-based reliability estimates for survey data analysis. Health questionnaires play a role in health economics research by providing a means to collect data on individuals' health-related information, preferences, behaviors, and outcomes. Model-based reliability, or coefficient omega, have become a popular concept to estimate test score reliability for various health care and health utility research instruments. Notably, these instruments are often embedded in data collection for nationwide health surveys. Data analysis of survey data needs to be capable of incorporating the survey design, where the data commonly accompany with unequal probabilities caused by clustering and post-stratification. Methods for estimating the variability of coefficient omega estimates for survey data analysis have not been investigated in the statistical literature, although it is a widely used tool to assess instrument reliability. In this article, we discuss a generally applicable linearization scheme for the relevant inference of such estimates based on the influence function approach when applied to complex survey data. Through the Monte Carlo study, we show adequate coverage rates for the confidence intervals based on scenarios of stratified multistage cluster sampling. Using data from the Medical Expenditure Panel Survey (MEPS), we provide the confidence intervals for the two types of coefficient omega (i.e., omega hierarchical and omega total) for the Short Form-12 version 2 (SF-12v2), a widely used health survey instrument to assess quality of life, and compare reliabilities of the instrument by different demographics.

Keywords: Coefficient omega, composite reliability, complex survey, influence function, linearization.

1. INTRODUCTION

Health questionnaires have a significant role in health economics research as they serve as a valuable tool for gathering data on individuals' health-

for citation :
Junyu Nie & Jihnhee Yu (2023). An Inferential Approach to Model-based Reliability: Applications with Health Survey Data. *Journal of Applied Econometrics and Statistics*, Vol. 2, No. 1, pp. 1-21. https://DOI: 10.47509/JAES.2023.v02i01.01

related information, preferences, behaviors, and outcomes (e.g., Brazier et al., 2002). Model-based reliability evaluation has gained popularity for assessing the test score reliability of various questionnaire instruments for health care and health utility research (Flora, 2020; Gignac et al., 2019; Watkins, 2017; Wiriyakijja et al., 2020). Model-based reliability, also referred to as coefficient omega, has several definitions, but is generally explained by the factor models with one or multiple latent factors. Many health surveys designed for large populations include various physical/mental health instruments, e.g., Center for Epidemiologic Studies depression scale (Carleton et al., 2013) in the National Health and Nutrition Examination Survey and Sheehan disability scale (Sheehan et al., 1996) in the National Comorbidity Survey. However, inference regarding the general class of model-based reliability has not been sufficiently addressed, particularly in the context of complex survey data analysis, in which the data sets commonly accompany with unequal probabilities caused by clustering and post-stratification (Lohr, 1999). In this article, we propose the generally applicable linearization method (Chauvet & Goga, 2018; Demnati & Rao, 2004; Deville, 1999; Yu, Chen, et al., 2019) in complex survey data analyses for model-based reliability estimates. The linearization is applicable to any sampling design to implement unbiased variance estimators of the Horvitz-Thompson estimator (Demnati & Rao, 2004).

In health research on populations, questionnaires are often used as research instruments to quantify information about study participants. These questionnaires consist of multiple items that measure a few aspects of constructs of interest, which often requires high reliability reflecting the characteristics of true constructs (Bentler, 2009). Reliability is defined as the ratio of the variance associated with the true construct over the total variance (Raykov & Marcoulides, 2010). A large reliability value indicates that the item responses in an instrument are not random noise in measuring a construct, and true responses within a same individual are highly correlated when relying on the parallel test form, or tests with same latent structures (McNeish, 2018; Sijtsma, 2009). The formulation of items, such as phrases in questionnaires and who received the tests, may affect instrument reliability (Bentler, 2009; Deng & Chan, 2017). In practice, since the part of the true construct within the scores is not observable, estimating reliability is not straightforward even when we have data with the observed scores.

Some measures of reliability of the total score have been developed based on administrating a single test instead of parallel test forms (Novick & Lewis, 1967). Coefficient alpha (Cronbach, 1951) – a metric based on a single test is known to be the average of the random split-half reliabilities. While coefficient alpha has been a widely used measure of reliability (Heo et al., 2015), its limitation/misunderstanding in the interpretation has been heavily discussed, e.g., requirement of certain assumptions such as tau-equivalence for validity (Novick & Lewis, 1967; Sijtsma, 2009; Trizano-Hermosilla & Alvarado, 2016) as well as a recent supporting discussion of its usage by Sijtsma (Sijtsma & Pfadt, 2021). Some alternative reliability metrics are available such as a series of Guttman's lower bounds of reliabilities (Guttman, 1945), the greatest lower bound (GLB) and model-based reliability (Bentler & Woodward, 1980; Jackson & Agunwamba, 1977). The GLB is shown to be equal to or larger than the alpha, while it is less than the true reliability or the product moment correlation of two parallel tests (Sijtsma, 2009; Zinbarg et al., 2005).

Model-based reliability, or coefficient omega (McDonald, 1981), has been received much attentions from practitioners (Carleton et al., 2013; Flora, 2020; Watkins, 2017; Wiriyakijja et al., 2020) and an extensive list of literature on the topic is available. Estimation of such metrics can be carried out by readily available software, e.g., the psych package (Revelle, 2019) and coefficientalpha package (Zhang & Yuan, 2016) in R. Coefficient omega does not require tau-equivalence allowing varying degree of loadings from a construct. A version of coefficient omega such as omega total (Zinbarg et al., 2005) that accounts for all factors tends to exceed the GLB, which is known for overcoming coefficient alpha's underestimation of reliability (McNeish, 2018; Revelle & Zinbarg, 2009; Sijtsma, 2009). Since its inception, coefficient omega (McDonald, 1970) has commonly been based on single latent factor models or unidimensional factor models (Flora, 2020). We note that the variance and confidence interval estimates for coefficient omega based on the one factor model are found in the extant literature (Garcia-Garzon et al., 2021; Padilla & Divers, 2016; Zhang & Yuan, 2016). For more general modeling schemes, variance estimation for the coefficient omega estimate, specifically accounting for the general factor (i.e., omega hierarchical), is available (Raykov & Zinbarg, 2011) but not in the context of incorporating unequal probabilities caused by survey designs and post-stratification. Overall, although coefficient omega estimation may require to fit a multidimensional factor model (Deville, 1999; Reise et al., 2018,), variance estimation for this general category of model-based reliability has not been sufficiently discussed in the relevant literature.

In this article, we propose an inferential method for a general form of model-based reliability estimates based on the influence function approach. The influence function evaluates the effect of a change in a data point on an estimator and can be used for nonparametric variance estimation. Recently, it has been successfully implemented to estimate variances of many statistics in survey data analysis (Deville, 1999; Yu, Vexler, et al., 2019).

This article is structured as follows. In Section 2, we propose strategies for estimating the variabilities of model-based reliability estimates in complex surveys. In Section 3, we perform a Monte Carlo study based on scenarios of stratified multistage cluster sampling and evaluate the performance of the proposed methods. In Section 4, the developed methods are used to analyze the Short Form-12 version 2 (SF-12v2) from a national survey data sets, and the estimated reliabilities comparing different demographics are reported. Section 5 presents the concluding remarks.

2. METHODOLOGY

In this section, we first describe a group of coefficient omega and ways to estimate it. Then, we develop a method for estimating variability using the influence function in survey data analyses.

2.1. Reliability Estimates

Several definitions of coefficient omega are available based on the different latent variable model assumptions. Coefficient omega can be defined in the context of the congeneric scale assumption, where the items are explained by the unidimensional factor model (Kelley & Pornprasertmanit, 2016; McDonald, 1978). Other types of coefficient omega can be obtained based on factor models that assume multidimensionality, where coefficient omega can be defined either by accounting for the general factor or by accounting for the general and group factors. Commonly, the former is termed omega hierarchical and the latter is termed omega total or simply omega (McNeish, 2018). The advantages and limitations of these different coefficient omega definitions have been reported in several papers including Rodriquez et al. (Rodriguez et al., 2016), Kelly and Pornprasermanit (Kelley & Pornprasertmanit, 2016) and McNeish (McNeish, 2018). Comparisons between different factor model constructions show that approaches based on higher order models or hierarchical factor models perform similarly, but they are better than principal component based factor analysis in estimating coefficient omega (Zinbarg et al., 2006).

We now describe the methodology development based on the general multidimensional factor model, which is composed of a general factor and group factors. The general factor represents a latent construct that affects all items in the scale, whereas the group factors reflect latent constructs that affect the subdomains of items. Confirmatory factor analysis (CFA) and the Schmid-Leiman procedure (Garcia-Garzon et al., 2021; Schmid & Leiman, 1957) may be used for estimating parameters in the model. The former approach is based on the likelihood function construction using the factor model with pre-defined factors. The latter approach starts with exploratory factor analysis and then the factors are rotated and transformed using the Schmid-Leiman transformation (Schmid & Leiman, 1957), in which researchers must determine the number of factors as well as a threshold in order to determine the meaningful group factor loadings. The Schmid-Leiman procedure also has some limitations regarding the minimum number of group factors or additional constraints on factor loadings (Garcia-Garzon et al., 2021; Revelle & Condon, 2019). It is known that estimates for general factor saturations using the Schmid-Leiman procedure "is positively biased and that a CFA based estimate is more accurate" (Revelle & Condon, 2019) in terms of amount of general factor variance in the test scores. In this paper, we estimate the coefficient omega values based on the likelihood functions in the framework of confirmatory factor analyses, and provide generally applicable solutions for the inference of the group of coefficient omega.

Covariance relationships among different items in the factor model are decomposed to a few underlying unobservable factors. Each item in the response has a loading on a general factor and on some group factors. Let X indicate p-variate random vector, and $Y = X - \mu$. The basic structure of the factor model (Zinbarg et al., 2006) is given as

$$Y = \Lambda_1 F_1 + \Lambda_2 F_2 + DS_p + E,$$

where Y is the $p\times 1$ vector of observed scores on the p scale items, F_1 is a general factor scalar – common to all p scale items, $\Lambda_1 = (\lambda_{11}, \ldots, \lambda_{1p})^T$ is $p\times 1$ general factor loading, F_2 is a $r\times 1$ vector of group factors (i.e., factors that are common to a subgroup of items but not all k items), $\Lambda_2 = (\Lambda_{21}, \ldots, \Lambda_{2r}) = ((\lambda_{21}, \ldots, \lambda_{2p})^T, \ldots, (\lambda_{r+1,1}, \ldots, \lambda_{r+1,p})^T)$ corresponding to group factors' factor loadings, D is a $p\times p$ factor loading for specific factors, S_p is $p\times 1$ vector of specific factors for each scale item with no correlation between items and E is the $p\times 1$ vector of random errors for each item. Typically, S_p and E are confounded and not distinguishable (Revelle & Condon, 2019; Zinbarg et al., 2006). We let $DS_p + E = \varepsilon$ and assume $\varepsilon \sim N_p(0_p, \Psi)$, $\Psi = diag(\psi_1, \ldots, \psi_p)$ where N_p denotes p-variate multinormal distribution, 0_p indicates p-dimension zero vector, and *diag* indicates a diagonal matrix. Some factor analysis literatures call this ε also the specific factors (Johnson & Wichern, 2014). In the model, unstandardized factor loadings Λ_1 and Λ_2 are to describe the relationship of the observed scores with general factor and group factors, respectively. The model assumes that all factors have the variance 1 and are uncorrelated with each other, i.e., $F = (F_1, F_2^T)^T \sim N_m(0, diag(1, \ldots, 1))$, where $m = r+1$. The general and group factors and ε are not correlated, i.e., $Cov(\varepsilon, F) = 0_{p\times m}$, $p \times m$ zero matrix. In confirmatory factor analysis, the latent factors correspond to at least two items (Fan et al., 2016; Kline, 2015) for model identifiability (Zinbarg et al., 2006). We note that some factor models used for coefficient omega may be equivalent to the Schmid-Leiman transformation (Schmid & Leiman, 1957), where the factor loadings for a group factor maintain a proportionality to the factor loadings for the general factor as known as the proportionality constraint (Gignac, 2016; Yung et al., 1999).

The general description of omega hierarchical (ω_h) is based on the major common factor F_1 (Revelle & Zinbarg, 2009). ω_h is defined as

$$\omega_h = \frac{(\sum_{i=1}^{p} \lambda_{1i})^2}{(\sum_{i=1}^{p} \lambda_{1i})^2 + \sum_{k=1}^{r} (\sum_{i=1}^{p} \lambda_{k+1,i})^2 + \sum_{i=1}^{p} \psi_i} = \frac{\mathbf{1}^T \Lambda_1 \Lambda_1{}^T \mathbf{1}}{V_y},$$

where V_y is the variance of the sum of elements in y, and 1 indicates the vector of ones. The numerator describes the true variance of the scales derived from the general factor F_1, whereas the denominator reflects the variances of the scales themselves. A high value of ω_h (i.e., close to 1) suggests that the items are highly saturated by the general factor.

The reliability coefficient can be defined by including group factors (F_2) based on the concept that the inclusion of these extra factors can explain systematic variabilities due to content, method, or certain factors. In this regard, omega total (ω_t) is defined as

$$\omega_t = \frac{(\sum_{i=1}^{p} \lambda_1)^2 + \sum_{k=1}^{r} (\sum_{i=1}^{p} \lambda_{k+1,2})^2}{(\sum_{i=1}^{p} \lambda_1)^2 + \sum_{k=1}^{r} (\sum_{i=1}^{p} \lambda_{k+1,i})^2 + \sum_{i=1}^{p} \psi_i} = \frac{\mathbf{1}^T \Lambda_1 \Lambda_1{}^T \mathbf{1} + \mathbf{1}^T \Lambda_2 \Lambda_2{}^T \mathbf{1}}{V_y}.$$

A high value of ω_t suggests high reliability of an instrument.

Comparisons between ω_h values or between ω_t values may allow us to investigate characteristics of various scales and subscales as well as the same scales applied to different demographics. For example, Gignac et al. (Gignac et al., 2019) compare two different scales with and without certain subscales using ω_h. Also, comparison of ω_h and ω_t may provide a simple evaluation of one-dimensionality of the scale (Green & Yang, 2015). The high similarity of the two values provides good evidence for one-dimensionality. The relatively low ω_h may be indicative that the simple sum of the scores may not be useful due to weak saturation of the general factor and multi-dimensionality of scales (Zinbarg et al., 2007). The low ω_h in subscales may reveal that an independent use of subscales may be not desirable (Wiriyakijja et al., 2020). The variance of the estimates of ω_h and ω_t can assist decision making on the instrument quality comparison between different scales or application settings.

2.2. Influence function for survey data analysis

Now let us consider a survey sampling setting. A population index set is defined as $U = \{1, ..., N\}$ with population size N. Each unit in the population has a value y_i, $i \in U$. A random sample S of size n is selected from U without replacement by a sampling design $p(s) = Pr\{S = s\}$ for all $s \subset$ U. The sampling scheme is based on certain population characteristics such as strata and clusters. Let us consider the measure M having a mass $1/N$ in each of the points y_i, $i \in U$ indicating that each data point is equally likely to be selected (Deville, 1999; Yu, Vexler, et al., 2019). Then, this measure leads to the classical definitions of the probability measure $\int 1 dM = 1$ and the population expectation $\int y dM = N^{-1}\Sigma_{i\in U}\, y_i := \overline{Y}$ ($:=$ indicates "defined as") (Deville, 1999; Yu, Vexler, et al., 2019). Also, we define the measure $\widehat{M}$ to be the estimator of M allocating a weight w_i/N to any point y_i and zero to any other points.

The full likelihood function construction in survey methodology considers the probability of sample inclusion in its construction (Lawless, 1997). Thus, the full maximum likelihood (ML) approach requires specifying a model of the conditional distribution of the population values given the stratum and cluster identifiers. This approach might require the stratum identifier and the representation of clusters by random effects. The specification of such models can be difficult, and inference cannot be robust to misspecification (Molina & Skinner, 1992). Unlike the full ML approach, the maximum pseudo-likelihood (MPL) approach does not require the specification of a model of the conditional distribution of the population values. Correlations and strata are addressed by incorporating weights in the manner of the design-based approach. Specifically, suppose that $\sum_{i=1}^{N} l(y_i, \boldsymbol{\theta})$ the population log-likelihood function, where $l(y_i, \boldsymbol{\theta})$ is the log of likelihood contribution of y_i and $\boldsymbol{\theta}$ is the parameter of interest. The idea of MPL is that it estimates the population log-likelihood function in the weighted form $\sum_{i\in s} w_i(D) l(y_i, \boldsymbol{\theta})$, where $w_i(D)$ is the i-th observation weight corresponding to the survey design D. This can be understood as a design-consistent estimator of the log-likelihood function based on the Horvitz-Thompson estimator (Krieger & Pfeffermann, 1992; Wang, 2021). The MPL estimator $\hat{\boldsymbol{\theta}}$ is obtained by satisfying $\underset{\boldsymbol{\theta}}{arg\,max} \sum_{i\in s} w_i(D) l(y_i, \boldsymbol{\theta})$.

Under appropriate regularity conditions, $\hat{\boldsymbol{\theta}}$ is consistent for $\boldsymbol{\theta}$ and asymptotically normally distributed (Lawless, 1997). Our development of the coefficients' inference via the influence function is based on the MPL approach that simplifies the computational complexities and can be applied to a wide class of multivariate models and various survey designs. The variance estimation of the MPL estimator needs to be carried out by standard survey sampling procedures such as the linearization method (Molina & Skinner, 1992).

The influence function (IF) is obtained by contaminating the underlying distribution by an increment of a point mass, thus reflecting the effect of a change in a data point on the target parameter. The influence function technique is useful in studying model robustness and calculating variance-covariance matrices of certain types of estimators, especially when more straightforward methods are difficult to implement. In survey methodology, the influence function technique is used as a linearization technique, a standard technique to estimate the variance of a statistic incorporating first and second-order inclusion probabilities (Deville, 1999).

Suppose that a twice differentiable objective function to maximize is $l(y,\theta)$. Maximization is carried out by solving $E[\nabla_\theta l(y,\theta)] = \int \nabla_\theta l(y,\theta) dM = 0$, where ∇_θ is the vector derivative with respect to θ and it is understood that the equation is satisfied at each element of the vector. The distribution perturbed by δ produces the equation $(1-\delta)E[\nabla_\theta l(y,\theta)] + \delta\nabla_\theta l(y,\theta) = 0$ (Kahn, 2015) and θ is influenced by the perturbed distribution. Taking the derivative of the equation with respect to δ produces the relationship

$$-E[\nabla_\theta l(y,\theta)] + (1-\delta)E[\nabla_{\theta\theta} l(y,\theta)]\frac{\partial\theta}{\partial\delta} = -\nabla_\theta l(y,\theta) - \delta\nabla_{\theta\theta} l(y,\theta)\frac{\partial\theta}{\partial\delta}$$

($\nabla_{\theta\theta}$ indicates the second derivative with respect to θ). Letting $\delta = 0$ and solving for $\partial\theta/\partial\delta$ gives the IF of θ in the following form

$$IF(\boldsymbol{\theta}) = E[-\nabla_{\boldsymbol{\theta\theta}} l(y,\boldsymbol{\theta})]^{-1}\nabla_{\boldsymbol{\theta}} l(y,\boldsymbol{\theta}).$$

Now, suppose that $IF_{n,i}(\theta)$ $(i \in s)$ is the empirical influence value based on a sample of observed data. Then, the variance can be estimated in the manner of estimating the quantity $Var(\overline{IF_n(\theta)})$ (Yu, Vexler, et al., 2019), where $\overline{IF_n(\theta)}$ indicates the sample mean of $IF_{n,i}(\theta)$ $(i \in s)$. When the parameter of interest is a joint function of θ (say $f(\theta)$), the IF is obtained using the functional delta method as

$$IF(f(\boldsymbol{\theta})) = \partial f(\boldsymbol{\theta}) / \partial \boldsymbol{\theta} \cdot IF(\boldsymbol{\theta})$$

where f is a differentiable function defined on the space of values for θ (Deville, 1999).

Based on the model , we have $S = \Lambda\Lambda^T + \Psi$, where $\Lambda = [\Lambda_1, \Lambda_2] = [\lambda_{ij}], i = 1, \ldots, p,\ j = 1, \ldots, \dim(F)$. Let θ = (Λ, Ψ), the parameters to be estimated. Using the model , the expectation of the objective function (i.e., log-likelihood function) of observed data for the factor analysis can be defined as

$$E\left(l(y, \boldsymbol{\theta})\right) = -\sum_{i \in U} \frac{1}{2N} \{ p \log 2\pi + \log |\Sigma| \} + \frac{1}{2} tr[S\Sigma^{-1}],$$

where $S = \Sigma_{i \in U}\, y_i\, y_i^T / N$ (Yu, Vexler, et al., 2019). Using the pseudo-likelihood approach, the log-likelihood function of observed data **y** for the factor analysis corresponding to $E\,(l(y, \boldsymbol{\theta}))$ is

$$l(\mathbf{y}, \boldsymbol{\theta}) = -\sum_{i \in S} w_i \frac{1}{2N} \left\{ p \log 2\pi + \log |\Sigma| + tr[\hat{S}\Sigma^{-1}] \right\},$$

where w_i is i-th observation weight corresponding to the survey design and $\hat{S}$ is the estimator of population covariance. For the influence function, we need to evaluate the first and second derivatives of the objective function. It should be noted that, if θ_i (i -th element of θ) is a known fixed value, $\partial l(y, \theta) / \partial \theta_i = \partial^2 l(y, \theta) / \partial \theta_i \partial \theta_j = 0$. Then, for $\nabla_\theta l(y, \theta)$, we have

$$vec\left(\frac{\partial l(y, \boldsymbol{\theta})}{\partial \Lambda}\right) = -vec(\Sigma^{-1}\Lambda - \Sigma^{-1} y y^T \Sigma^{-1} \Lambda),$$

$$vec\left(\frac{\partial l(y, \boldsymbol{\theta})}{\partial \Psi}\right) = \frac{1}{2} vec\Big(diag(\Sigma^{-1} - \Sigma^{-1} y y^T \Sigma^{-1})\Big),$$

where *vec* indicates the column-wise vectorization of a matrix. For $E\left(\nabla_{\theta\theta} l(y, \theta)\right)$, we have

$$E\left(\frac{\partial^2 l(y, \boldsymbol{\theta})}{\partial \Lambda \partial \lambda_{jk}}\right) = -\sum_{i \in U} \frac{1}{N} \left[\Sigma^{-1} \left\{ \frac{\partial \Sigma}{\partial \lambda_{jk}} \Sigma^{-1} S - \frac{\partial \Sigma}{\partial \lambda_{jk}} + S\Sigma^{-1} \frac{\partial \Sigma}{\partial \lambda_{jk}} \right\} \Sigma^{-1} \Lambda + (\Sigma^{-1} - \Sigma^{-1} S \Sigma^{-1}) \frac{\partial \Lambda}{\partial \lambda_{jk}} \right],$$

$$E\left(\frac{\partial^2 l(y,\boldsymbol{\theta})}{\partial\Psi\partial\psi_j}\right)=\frac{1}{2}\sum\nolimits_{i\in U}\frac{1}{N}\left[diag\left\{\Sigma^{-1}\left(-\frac{\partial\Sigma}{\partial\psi_j}\Sigma^{-1}S+\frac{\partial\Sigma}{\partial\psi_j}-S\Sigma^{-1}\frac{\partial\Sigma}{\partial\psi_j}\right)\Sigma^{-1}\right\}\right],$$

$$E\left(\frac{\partial^2 l(y,\boldsymbol{\theta})}{\partial\Lambda\partial\psi_j}\right)=-\sum\nolimits_{i\in U}\frac{1}{N}\left[\Sigma^{-1}\left\{\frac{\partial\Sigma}{\partial\psi_j}\Sigma^{-1}S-\frac{\partial\Sigma}{\partial\psi_j}+S\Sigma^{-1}\frac{\partial\Sigma}{\partial\psi_j}\right\}\Sigma^{-1}\Lambda\right],$$

where $j = 1, \ldots, p$, $k = 1, \ldots, dim(F)$. We note that similar derivations of the second derivatives of the log-likelihood function are also found in (Kwan & Fung, 1998). In each observation, the corresponding influence function is $IF_i(\theta)=E[-\nabla_{\theta\theta}l(y,\theta)]^{-1}\nabla_\theta l(y_i,\theta)]_{\theta=\hat{\theta}}:=(I_{\Lambda i},I_{\Psi i})^T$, $i\in s$, where vectors $I_{\Lambda i}$ and $I_{\Psi i}$ correspond to influence functions of Λ and Ψ, respectively. For empirical influence values, E $[-\nabla_{\theta\theta}l(y, \theta)]$ can be replaced by the observed Fisher information matrix. Let ($I^*_{\Lambda i}$ ($i \in s$) indicate the matrix where each element consists of the IF corresponding to the element of $\Lambda = [\Lambda_1, \Lambda_2]$. Also, let 1 indicate $\boldsymbol{p}$-dimensional vector of 1. Then, we have the following IF of ω_t.

Proposition 1. Under the factor model , the IF of ω_t corresponding to each individual is expressed as

$$I_{\omega_t,i}=\frac{1}{\left(\mathbf{1}^T(\Lambda\Lambda^T+\Psi)\mathbf{1}\right)^2}\left\{\mathbf{2}\mathbf{1}^T\Lambda I^{*\,T}_{\Lambda i}\mathbf{1}(\mathbf{1}^T\Psi\mathbf{1})-\mathbf{1}^T\Lambda\Lambda^T\mathbf{1}\mathbf{1}^T I_{\Psi i}\right\},i\in U.$$

The proof of Proposition 1 is found in the Appendix. The IF involves factor loadings and unique factors, thus its estimation requires to replace those parameters by corresponding estimates to obtain the empirical influence values (Davison & Hinkley, 1997). The IF of ω_h can be obtained in a similar manner.

Proposition 2. Under the factor model , the IF of ω_h corresponding to each individual is expressed as

$$I_{\omega_h,i}=\frac{1}{\left(\mathbf{1}^T(\Lambda\Lambda^T+\Psi)\mathbf{1}\right)^2}\left\{\mathbf{2}\mathbf{1}^T\Lambda_1 B I^{*\,T}_{\Lambda i}\mathbf{1}-\mathbf{1}^T\Lambda_1\Lambda_1^T\mathbf{1}\mathbf{1}^T I_{\Psi i}\right\},i\in U,$$

Where $B=[1^T\Lambda_2\Lambda_2^T1+1^T\Psi1,\ -1^T\Lambda_1 11^T\Lambda_2]$ and $I^*_{\Lambda i}$ is the IF vector corresponding to Λ_1.

The proof of Proposition 2 is similar to that of Proposition 1, thus omitted.

Although we discuss the IF of coefficient omega in terms of multi-factor models, the IF for omega coefficient based on the unidimensional factor model (say, ω_m) (Garcia-Garzon et al., 2021; Zhang & Yuan, 2016) can be easily implemented based on the method we discussed. Suppose a one-factor model:

$$Y = \Lambda F + \varepsilon$$

where Λ is the p×1 factor loading vector, $F \sim N(0, 1)$ is the common factor scalar and $\varepsilon \sim N_p(0, \Psi)$, $Cov(\varepsilon, F) = 0$. Based on this setting, we have the IF as follows similar to Propositions 1 and 2.

Proposition 3. Under the factor model , the IF of ω_m corresponding to each individual is expressed as

$$I_{\omega_m, i} = \frac{1}{\left(\mathbf{1}^T(\Lambda\Lambda^T + \Psi)\mathbf{1}\right)^2}\left\{\mathbf{1}^T\Lambda\left(-\mathbf{1}^T I_{\Psi i}\mathbf{1}^T\Lambda - 2\mathbf{1}^T\Lambda\Lambda^T\mathbf{1} + 2\mathbf{1}^T I_{\Lambda i}\mathbf{1}^T(\Lambda\Lambda^T + \Psi)\mathbf{1}\right)\right\}, i \in U.$$

The variance of omega coefficient is estimated as the variance of the sample mean of the influence function based on the variance estimator of the Horvitz-Thomson estimator (e.g., Sen-Yates-Grundy formula (Lohr, 1999)) or the Hansen-Hurwitz estimator (Hansen, 1953) as an approximation of the former. Standard survey software packages (e.g., survey package in R) (Lumley, 2011) provide the variance estimate of the sample mean incorporating complex survey designs. Relevant R codes are available from the authors upon request. Commonly, the Wald type confidence interval is obtained using the variance estimates in survey data analyses (Cochran, 1977; NHANES, 2018). The performance of the confidence interval is investigated in the next section.

3. SIMULATION STUDIES

We perform a Monte Carlo study (1000 simulations per scenario) based on several scenarios of data structures and sample sizes to assess the performance of the influence function methods in estimating the variance of omega coefficients in complex survey settings in R. In the simulation, we first generate a finite population with strata and clusters, and then we conduct stratified two-stage cluster sampling. In each stratum (a total of three strata), we construct 500 primary sampling units (PSUs) with common covariance (0.05 throughout the scenarios) within PSUs and 100 secondary sampling units (SSUs) per PSU following the factor model , totaling 150,000 SSUs. For each sample, we conduct simple random sampling in each stratum in the first-stage and in the second-stage.

Both multivariate normal data and correlated ordinal data are considered in the population generation. In the multivariate normal distribution, the data $X = Y + \mu_h$ are generated on the model structure described in using a covariance matrix implied by the model parameters, where different values of $\mu_h (h = 1, 2, 3)$ are applied for different strata similar to the structure of the one-way ANOVA model. We consider three bifactor models (Yung et al., 1999): (i) 8 items ($p = 8$) with one general factor and two group factors (each group factor corresponds to four non-overlapping items), (ii) 12 items ($p = 12$) with one general factor and three group factors (each group factor corresponds to four non-overlapping items) and (iii) the same structure as the model in (ii) with the proportionality constraint between general and group factors as described in Section 2. In the simulation, true omega coefficients are known based on the definitions and . The structure of these models are shown in Figure 1. To generate the ordinal data with correlations between items, we first generate the multinormal data with the same structures described above and then discretize them into 0, 1, 2 and 3 so that the distribution of the categorized data is approximately 22%, 8%, 20%, and 50% for 0, 1, 2 and 3, respectively. This discretization changes the covariance structure and true factor loadings, thus altering true coefficient omega values. Since we assume that the population for the survey is finite, the population coefficient omega values can be obtained from the generated finite population. Using the generated population, we first obtain the factor loadings and specific variances based on the bifactor model, and then we obtain the population coefficient omega values. The weight for each individual (SSU) in each stratum h is $(N_h M_h)/(n_h m_h)$ where N_h is the number of PSUs per stratum, M_h is the number of SSUs per PSU, n_h is the first-stage sample size per stratum and m_h is the sample size in the second stage. Nonzero factor loadings are resulting in varying levels of coefficient omega values where these values mimic our real data analysis, which is described in Section 4.

Tables 1 and 2 show the coverage rates and average widths of the confidence intervals using the proposed IF approach for ω_t and ω_h, respectively. The coverage rate is computed as the percentage of simulation runs where the interval contained the true coefficient omega value. As shown in both tables, the results of the proposed influence function method show that the coverage rates were close to the target confidence level in both multi-normal distributions and correlated ordinal data. When the sample size is increased, the width of the confidence interval became narrower, indicating lower variability in the coefficient omega estimates. We conclude that the influence function method satisfactorily provides confidence intervals of coefficient omega at specified confidence levels.

4. APPLICATION

In this section, we provide a detailed description of our data set and report coefficient omega estimates based on the whole data and some subgroups.

It is possible that different demographic groups (the people taking the test (Revelle & Condon, 2019)) may show different levels of general factor saturation of instruments. A relatively low reliability in a certain group would indicate more variability within questionnaire items than variability of the constructs of interest. On the other hand, a low general factor saturation may indicate that there may be a large distinction in responses between subscales. These estimates may reflect some characteristics of the composition of the participants in the group (Bentler, 2009).

4.1. The Data

We use data on the Short Form-12 version 2 (SF-12v2) from the Household Component of the Medical Expenditure Panel Survey (MEPS-HC) in 2015 and 2016. The MEPS, which is administered by the Agency for Healthcare Research and Quality, is a national survey that represents the health of the non-institutionalized adult population in the US. The MEPS consists of panels that encompass two years in five rounds of mail surveys and/or interviews. To compose the panels, households are chosen annually from the households that participated in the National Health Interview Survey in the previous year. The Household Component (HC) is a prominent constituent of the MEPS data. The HC gathers information from individual household members regarding general demographics, disease states, overall health status, insurance coverage, charges and payments, employment, income, use of and access to healthcare, as well as satisfaction with healthcare.

The SF-12, which is composed of 12 questions, is a widely-used health survey to assess self-reported health-related quality of life. The SF-12v2 (Montazeri et al., 2011) is an improved version of the original SF-12. The instrument consists of two areas: physical health and mental health (Ware Jr et al., 1996). The area on physical health focuses on participants' general overall health, limitations in mobility, work, and other physical activities as well as limitations due to pain. The corresponding scales include general health (GH, one item), physical functioning (PF, two items), role physical (RP, two items), and bodily pain (BP, one item). The area on mental health encompasses limitations in social activity, emotional state, and level of distraction. The corresponding scales include a physical health summary and vitality (VT, one item), social functioning (SF, one item), role emotional (RE, two items), and mental health (MH, two items). In the data analysis, following the common practice (Hays et al., 1993; Kathe et al., 2018), each of coded scores in the SF-12v2 is converted to values ranging from 0 to 100, and inverse-coded scores are corrected so that 100 indicates the best health condition.

The SF-12v2 data are collected in the second and/or fourth rounds in the two-year time span and are included in the Longitudinal File in the 2015–2016 MEPS-HC. The majority of the participants in the second round have SF-12v2 data, which are the primary data used in our analysis. When the data are missing in the second round, we use the data collected in the fourth round. The Longitudinal file provides a weight variable (LONGWT) and design variables (sampling strata: VARSTR, primary sampling units: VARPSU) to produce national estimates based on the data (Cohen et al., 2009). The data set has a total of 165 strata (VARSTR value 1001-1165), which are broken down by state and employment size. Each stratum includes three counties as the primary sampling units (VARPSU value 1-3), totaling 495 PSUs. The proper usage of these variables produces estimates of the civilian noninstitutionalized population for the entire two-year period from 2015–2016. To obtain estimates of coefficient omega and their variances using IFs, we take into account the complex sample design of the MEPS by incorporating these variables. The final data set contains some missing data (31.66%). Weights are adjusted for missing data within each stratum based on nonresponse weight adjustment (Lohr, 1999).

We use the bifactor model as described in Figure 2, which includes one general factor and two group factors, where the two group factors correspond to the physical health scales and mental health scales, respectively as the instrument is originally designed to estimate (Ware Jr et al., 1996). Initial values used to maximize the likelihood function are selected based on loadings and specific variance estimates from the exploratory factor analysis. A domain analysis (Lumley, 2011) is applied to the subgroups (i.e., sex and age groups), in which the weights are adjusted to be 0 for individuals not in the domain while the original weights are kept for individuals in the domain.

4.2. Results

Based on the whole data set, the estimate of ω_t is 0.938 (confidence interval: 0.931~0.945) and that of ω_h is 0.888 (confidence interval: 0.870 ~ 0.906) indicating that overall variability is well explained by the general factor and common factors. As expected, ω_t shows a larger value than ω_h, i.e., throughout the estimates, ω_t is about 5% larger than ω_h. The difference between ω_t and ω_h implies an existence of latent variables representative of certain domains within test items (Revelle & Zinbarg, 2009). In the SF-12v2, a high saturation of the general factor represented by the high ω_h is indicative of closer associations between mental and physical health statuses. Some differences between ω_h and ω_t in the SF-12v2 are reasonable, considering that the questionnaire has two subdomains.

In the subgroup analysis, the estimates of ω_t have a range of 0.89 ~ 0.94 and the estimates of ω_h have a range of 0.84 ~ 0.90. We use 95% confidence intervals to check that there is a noticeable difference between two subgroups. We observe little difference in the coefficient omega estimates between males and females. However, relatively small coefficient omega estimates in the young (≥ 30) age group are shown. In both ω_t and ω_h, the coefficient omega estimates in the young (≥ 30) age group are significantly lower than those in the older (>30) age group (significance level of 5%). The lower ω_t and ω_h indicate that variability explained by the common factors is slightly compromised among the young age group compared to the group over 30 years old suggesting proportionally more random variance in response by the young group compared to the older age groups. In terms of differences between ω_t and ω_h, no particular subgroups show substantially large differences between ω_t and ω_h.

5. CONCLUDING REMARKS

In this paper, we explained how to obtain the IFs to estimate the variability in the general group of model-based reliability metrics which was based on the factor models. We also show how to implement the IF to estimate the variability in such estimates in complex survey settings, where the proper weights of individual observations need to be considered. Through the Monte Carlo study, the proposed approach showed the workable property to carry out an inference of coefficient omega incorporating the survey design. By applying these methods to the SF-12v2 data set from the MEPS, we demonstrated that the proposed methods were useful to compare coefficient omega values as characteristics of response patterns in the groups of interest.

We conclude that the IF approach has a feasible inferential property. The viable inferential property of the proposed approach made it possible to compare model-based reliability values between different groups. We consider that such comparisons can be useful in constructing or modifying survey instruments that target specific groups of interest.

Conflict of interest: The authors declare that they have no conflict of interest.

References

Bentler, P., & Woodward, J. A. (1980). Inequalities among lower bounds to reliability: With applications to test construction and factor analysis. *Psychometrika*, *45*(2), 249-267.

Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika*, *74*(1), 137-143.

Brazier, J, Roberts, J, & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics*, *21*(2), 271-292.

Carleton, R. N., Thibodeau, M. A., Teale, M. J., Welch, P. G., Abrams, M. P., Robinson, T., & Asmundson, G. J. (2013). The center for epidemiologic studies depression scale: a review with a theoretical and empirical examination of item content and factor structure. *PloS one*, *8*(3), e58067.

Chauvet, G., & Goga, C. (2018). Linearization versus bootstrap for variance estimation of the change between Gini indexes. *Survey Methodology*, *44*(1), 17-43.

Cochran, W. G. (1977). *Sampling techniques*. John Wiley & Sons.

Cohen, J. W., Cohen, S. B., & Banthin, J. S. (2009). The medical expenditure panel survey: a national information resource to support healthcare cost research and inform policy and practice. *Medical care*, S44-S50.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge university press.

Demnati, A., & Rao, J. (2004). Linearization variance estimators for survey data. *Survey Methodology*, *30*(1), 17-26.

Deng, L., & Chan, W. (2017). Testing the difference between reliability coefficients alpha and omega. *Educational and Psychological Measurement*, *77*(2), 185-203.

Deville, J.-C. (1999). Variance estimation for complex statistics and estimators: linearization and residual techniques. *Survey Methodology*, *25*(2), 193-204.

Fan, Y., Chen, J., Shirkey, G., John, R., Wu, S. R., Park, H., & Shao, C. (2016). Applications of structural equation modeling (SEM) in ecological studies: an updated review. *Ecological Processes*, *5*(1), 1-12.

Flora, D. B. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Advances in Methods and Practices in Psychological Science*, *3*(4), 484-501.

Garcia-Garzon, E., Abad, F. J., & Garrido, L. E. (2021). On Omega Hierarchical Estimation: A Comparison of Exploratory Bi-Factor Analysis Algorithms. *Multivariate Behavioral Research*, *56*(1), 101-119.

Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence*, *55*, 57-68.

Gignac, G. E., Reynolds, M. R., & Kovacs, K. (2019). Digit Span subscale scores may be insufficiently reliable for clinical interpretation: distinguishing between stratified coefficient alpha and omega hierarchical. *Assessment*, *26*(8), 1554-1563.

Green, S. B., & Yang, Y. (2015). Evaluation of dimensionality in the assessment of internal consistency reliability: Coefficient alpha and omega coefficients. *Educational Measurement: Issues and Practice*, *34*(4), 14-20.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, *10*(4), 255-282.

Hansen, M. H. (1953). *Sample survey methods and theory vol. 1*. John Wiley & Sons.

Hays, R. D., Sherbourne, C. D., & Mazel, R. M. (1993). The rand 36 item health survey 1.0. *Health economics*, *2*(3), 217-227.

Heo, M., Kim, N., & Faith, M. S. (2015). Statistical power as a function of Cronbach alpha of instrument questionnaire items. *BMC medical research methodology*, *15*(1), 1-9.

Jackson, P. H., & Agunwamba, C. C. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: I: Algebraic lower bounds. *Psychometrika*, *42*(4), 567-578.

Johnson, R. A., & Wichern, D. W. (2014). *Applied multivariate statistical analysis* (Vol. 6). Pearson London, UK:.

Kahn, J. (2015). Influence Functions for Fun and Profit. *Ross School of Business, University of Michigan. Available from http://j-kahn. com/files/influencefunctions. pdf (version: July 10, 2015).*

Kathe, N., Hayes, C. J., Bhandari, N. R., & Payakachat, N. (2018). Assessment of reliability and validity of SF-12v2 among a diabetic population. *Value in Health*, *21*(4), 432-440.

Kelley, K., & Pornprasertmanit, S. (2016). Confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. *Psychological methods*, *21*(1), 69.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Krieger, A. M., & Pfeffermann, D. (1992). Maximum likelihood estimation from complex sample surveys. *Survey Methodology*, *18*(2), 225-239.

Kwan, C. W., & Fung, W. K. (1998). Assessing local influence for specific restricted likelihood: Application to factor analysis. *Psychometrika*, *63*(1), 35-46.

Lawless, J. (1997). Likelihood and pseudo likelihood estimation based on response-biased observation. *Lecture Notes-Monograph Series*, 43-55.

Lohr, S. L. (1999). *Sampling : design and analysis / Sharon L. Lohr*. Duxbury Press.

Lumley, T. (2011). *Complex surveys: a guide to analysis using R* (Vol. 565). John Wiley & Sons.

McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology*, *23*(1), 1-21.

McDonald, R. P. (1978). Generalizability in Factorable Domains:" Domain Validity and Generalizability" 1. *Educational and Psychological Measurement*, *38*(1), 75-79.

McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, *34*(1), 100-117.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological methods*, *23*(3), 412.

Molina, E, & Skinner, C. (1992). Pseudo-likelihood and quasi-likelihood estimation for complex sampling schemes. *Computational statistics & data analysis*, *13*(4), 395-405.

Montazeri, A., Vahdaninia, M., Mousavi, S. J., Asadi-Lari, M., Omidvari, S., & Tavousi, M. (2011). The 12-item medical outcomes study short form health survey version 2.0 (SF-12v2): a population-based validation study from Tehran, Iran. *Health and quality of life outcomes*, *9*(1), 1-8.

NHANES. (2018). *Estimating Variance, Analyzing Subgroups and Calculating Degrees of Freedom for Performing Statistical Tests and Calculating Confidence Intervals*. https://www.cdc.gov/nchs/tutorials/nhanes-cms/variance/variance.htm

Novick, M. R., & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, *32*(1), 1-13.

Padilla, M. A., & Divers, J. (2016). A comparison of composite reliability estimators: coefficient omega confidence intervals in the current literature. *Educational and Psychological Measurement*, *76*(3), 436-453.

Raykov, T., & Marcoulides, G. A. (2010). Group comparisons in the presence of missing data using latent variable modeling techniques. *Structural equation modeling*, *17*(1), 134-149.

Raykov, T., & Zinbarg, R. E. (2011). Proportion of general factor variance in a hierarchical multiple component measuring instrument: A note on a confidence interval estimation procedure. *British Journal of Mathematical and Statistical Psychology*, *64*(2), 193-207.

Reise, S. P., Bonifay, W., & Haviland, M. G. (2018). Bifactor modelling and the evaluation of scale scores. *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development*, 675-707.

Revelle, W. (2019). psych: procedures for psychological, psychometric, and personality research. Northwestern University, Evanston, Illinois. R package version 1.9. 12. *Comprehensive R Archive Network (CRAN), Northwestern University: Evanston, IL, USA*.

Revelle, W., & Condon, D. M. (2019). Reliability from á to ù: A tutorial. *Psychological assessment*, *31*(12), 1395.

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, *74*(1), 145-154.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological methods*, *21*(2), 137.

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, *22*(1), 53-61.

Sheehan, D. V., Harnett-Sheehan, K., & Raj, B. (1996). The measurement of disability. *International clinical psychopharmacology*.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107-120.

Sijtsma, K., & Pfadt, J. M. (2021). Part II: On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha: Discussing Lower Bounds and Correlated Errors. *Psychometrika*, *86*(4), 843-860. https://doi.org/10.1007/s11336-021-09789-8

Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: congeneric and asymmetrical measurements. *Frontiers in psychology*, *7*, 769.

Wang, J. (2021). The Pseudo Maximum Likelihood Estimator for Quantiles of Survey Variables. *Journal of Survey Statistics and Methodology*, *9*(1), 185-201.

Ware Jr, J. E., Kosinski, M., & Keller, S. D. (1996). A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Medical care*, 220-233.

Watkins, M. W. (2017). The reliability of multidimensional neuropsychological measures: From alpha to omega. *The clinical neuropsychologist*, *31*(6-7), 1113-1126.

Wiriyakijja, P., Porter, S., Fedele, S., Hodgson, T., McMillan, R., Shephard, M., & Ni Riordain, R. (2020). Validation of the HADS and PSS 10 and psychological status in patients with oral lichen planus. *Oral diseases*, *26*(1), 96-110.

Yu, J., Chen, Z., Wang, K., & Tezal, M. (2019). Suggestion of confidence interval methods for the Cronbach alpha in application to complex survey data. *Survey Methodology*, *45*(3), 465-485.

Yu, J., Vexler, A., & Jalal, K. (2019). A critical issue of using the variance of the total in the linearization method–In the context of unequal probability sampling. *Statistics in medicine*, *38*(8), 1475-1483.

Yung, Y.-F., Thissen, D., & McLeod, L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika*, *64*(2), 113-128.

Zhang, Z., & Yuan, K.-H. (2016). Robust coefficients alpha and omega and confidence intervals with outlying observations and missing data: Methods and software. *Educational and Psychological Measurement*, *76*(3), 387-411.

Zinbarg, R. E., Revelle, W., & Yovel, I. (2007). Estimating ù h for structures containing two group factors: Perils and prospects. *Applied Psychological Measurement*, *31*(2), 135-157.

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ω H: Their relations with each other and two alternative conceptualizations of reliability. *Psychometrika*, *70*(1), 123-133.

Zinbarg, R. E., Yovel, I., Revelle, W., & McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for ωh. *Applied Psychological Measurement*, *30*(2), 121-144.

Appendix

The proof of Proposition 1. We express ω_t in the equation (3) as

$$\omega_t = \frac{(\sum_{i=1}^{p}\lambda_1)^2 + \sum_{k=1}^{r}(\sum_{i=1}^{p}\lambda_{k+1,2})^2}{(\sum_{i=1}^{p}\lambda_1)^2 + \sum_{k=1}^{r}(\sum_{i=1}^{p}\lambda_{k+1,i})^2 + \sum_{i=1}^{p}\psi_i} = \frac{\mathbf{1}^T\Lambda_1\Lambda_1{}^T\mathbf{1} + \mathbf{1}^T\Lambda_2\Lambda_2{}^T\mathbf{1}}{\mathbf{1}^T\Lambda_1\Lambda_1{}^T\mathbf{1} + \mathbf{1}^T\Lambda_2\Lambda_2{}^T\mathbf{1} + \mathbf{1}^T\Psi}.$$

We now can obtain first derivate of ω_t corresponding to general factor Λ_1, group factor Λ_2 and unique variance Ψ in forms of

$$\frac{\partial\omega_t}{\partial\Lambda_1} = \frac{2\mathbf{1}^T\Lambda_1\mathbf{1}(\mathbf{1}^T\Psi\mathbf{1})}{(\mathbf{1}^T\Lambda_1\Lambda_1{}^T\mathbf{1} + \mathbf{1}^T\Lambda_2\Lambda_2{}^T\mathbf{1} + \mathbf{1}^T\Psi)^2},$$

$$\frac{\partial\omega_t}{\partial\Lambda_{2k}} = \frac{2\mathbf{1}^T\Lambda_{2k}\mathbf{1}(\mathbf{1}^T\Psi\mathbf{1})}{(\mathbf{1}^T\Lambda_1\Lambda_1{}^T\mathbf{1} + \mathbf{1}^T\Lambda_2\Lambda_2{}^T\mathbf{1} + \mathbf{1}^T\Psi)^2},\ k = 1,\dots,r \text{ and}$$

$$\frac{\partial\omega_t}{\partial\Psi} = \frac{-(\mathbf{1}^T\Lambda_1\Lambda_1{}^T\mathbf{1} + \mathbf{1}^T\Lambda_2\Lambda_2{}^T\mathbf{1})\mathbf{1}}{(\mathbf{1}^T\Lambda_1\Lambda_1{}^T\mathbf{1} + \mathbf{1}^T\Lambda_2\Lambda_2{}^T\mathbf{1} + \mathbf{1}^T\Psi)^2}.$$

The equation gives rise to the expression of the IF in a form $I_{\omega_t} = \frac{\partial\omega_t^T}{\partial\Lambda_1} I_{\Lambda_1} + \frac{\partial\omega_t^T}{\partial\Lambda_2} I_{\Lambda_2} + \frac{\partial\omega_t^T}{\partial\Psi} I_{\Psi}$, where $\frac{\partial\omega_t}{\partial\Lambda_2} = \left(\frac{\partial\omega_t^T}{\partial\Lambda_{21}}, \dots, \frac{\partial\omega_t^T}{\partial\Lambda_{2r}}\right)$. Applying this expression, we have

$$I_{\omega_t,i} = \frac{1}{\left(\mathbf{1}^T(\Lambda\Lambda^T + \Psi)\mathbf{1}\right)^2}\left\{2\mathbf{1}^T\Lambda I_{\Lambda i}^{*\,T}\mathbf{1}(\mathbf{1}^T\Psi\mathbf{1}) - \mathbf{1}^T\Lambda\Lambda^T\mathbf{1}\mathbf{1}^T I_{\Psi i}\right\}, i \in s.$$

Table 1: The coverage rates (CR) and confidence interval widths (95% confidence level) of ω_t based on the IF method (Proposition 1). The values of npsu and nssu are the sample sizes for PSUs and SSUs within a PSU, respectively. Three ω_t values in the column ω_t are the population ω_t values for instruments of 8 items (p = 8), 12 items (p = 12) and 12 items with the proportionality constraint (p = 12 with prop.), respectively

Distribution	*(npsu,nssu)*	ω_t	*p=8*		*p=12*		*p=12 with prop.*	
			CR	*Width*	*CR*	*Width*	*CR*	*Width*
Multi-normal	10,20	0.909, 0.894, 0.915	0.940	0.048	0.935	0.025	0.937	0.019
	20,20	0.909, 0.894, 0.915	0.951	0.028	0.939	0.017	0.947	0.014
Ordinal	10,20	0.883, 0.883, 0.895	0.962	0.088	0.953	0.046	0.938	0.024
	20,20	0.864, 0.881, 0.895	0.969	0.047	0.955	0.027	0.947	0.017

Table 2: The coverage rates (CR) and confidence interval widths (95% confidence level) of ω_h based on the IF method (Proposition 2). The values of npsu and nssu are the sample sizes for PSUs and SSUs within a PSU, respectively. Three ω_h values in the column ω_h are the population ω_h values for instruments of 8 items (p = 8), 12 items (p = 12) and 12 items with the proportionality constraint (p = 12 with prop.), respectively

Distribution	*(npsu,nssu)*		*p=8*		*p=12*		*p=12 with prop.*	
			CR	*Width*	*CR*	*Width*	*CR*	*Width*
Multi-normal	10,20	0.891, 0.902, 0.890	0.941	0.038	0.945	0.029	0.949	0.030
	20,20	0.891, 0.902, 0.890	0.942	0.035	0.942	0.020	0.951	0.021
Ordinal	10,20	0.881, 0.883, 0.886	0.948	0.048	0.935	0.038	0.947	0.036
	20,20	0.880, 0.884, 0.885	0.945	0.036	0.949	0.026	0.948	0.025

Table 3: Estimates of ω_t of SF-12v2 from MEPS, 95% confidence intervals (CI), and the number of sample (n) for the whole data and subgroups

Group		*CI*	*n*
Total	0.938	0.930, 0.945	11629
Sex			
Female	0.939	0.932, 0.947	6249
Male	0.936	0.927, 0.945	5380
Age group			
Age<=30	0.890	0.879, 0.901	2894
Age>30	0.940	0.937, 0.944	8696

Table 4: Estimates of ω_h of SF-12v2 from MEPS, 95% confidence intervals (CI), and the number of sample (n) for the whole data and subgroups

Group		*CI*	*n*
Total	0.888	0.870, 0.906	11629
Sex			
Female	0.887	0.871, 0.904	6249
Male	0.888	0.871, 0.905	5380
Age			
Age<=30	0.843	0.821, 0.864	2894
Age>30	0.897	0.881, 0.912	8696

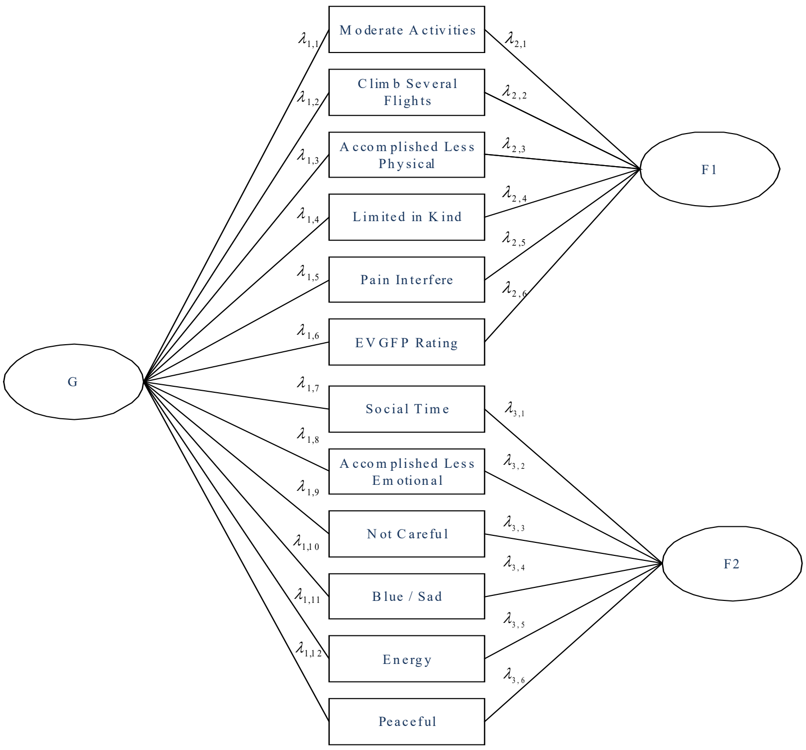

Figure 1: Schematic of data simulations based on multinormal distribution

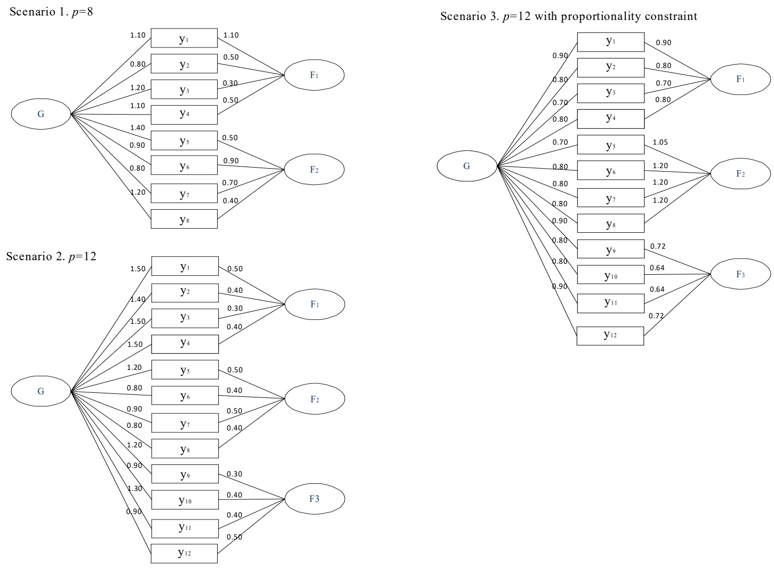

Figure 2: The structure of the bi-factor model of SF-12v2. "G" indicates the general factor and "F1" and "F2" indicate the physical component score (PCS) and mental component score (MCS), respectively