2020

A Bifactor Approach to Dimensionality Assessment

David Michael Dueber
University of Kentucky, ddueber@gmail.com
Author ORCID Identifier:
https://orcid.org/0000-0001-6734-3268
Digital Object Identifier: https://doi.org/10.13023/etd.2020.078

STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

David Michael Dueber, Student

Dr. Michael D. Toland, Major Professor

Dr. Michael D. Toland, Director of Graduate Studies

A BIFACTOR APPROACH TO DIMENSIONALITY ASSESSMENT

DISSERTATION

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
College of Education
at the University of Kentucky

By
David Michael Dueber
Lexington, Kentucky
Director: Dr. Michael D. Toland, Professor of Educational, School, and
Counseling Psychology
Lexington, Kentucky
2020

https://orcid.org/0000-0001-6734-3268

ABSTRACT OF DISSERTATION

A BIFACTOR APPROACH TO DIMENSIONALITY ASSESSMENT

Bifactor confirmatory factor analysis models and statistical indices computed from them have previously been used to provide evidence for the appropriateness of utilizing a unidimensional interpretation of multidimensional data. However, the ability of bifactor indices to aid in the assessment of subscore strength has not been investigated.

A simulation study was conducted to relate bifactor indices to the strength of subscores corresponding to specific factors. The bifactor indices OmegaHS and ECV_{SS} were found to be strongly predictive of subscore strength conditional upon OmegaS. The number of factors was also found to play a minor role in this relationship. Cutoffs for assessing the appropriateness of interpreting subscores were constructed.

The overarching goal of this work was to extend a framework for using bifactor models and their indices as diagnostic tools for dimensionality assessment. This goal is accomplished in two steps. First, a package for the R statistical computing environment was developed to enable the efficient computation of bifactor indices. Second, the aforementioned simulation study was conducted to discover relationships between bifactor indices and classical test theoretic measures of subscore strength.

KEYWORDS: Bifactor, Dimensionality, Confirmatory Factor Analysis, Simulation, Subscores

David Michael Dueber

04/21/2020

Date

A BIFACTOR APPROACH TO DIMENSIONALITY ASSESSMENT

By

David Michael Dueber

Michael D. Toland
Director of Dissertation

Michael D. Toland
Director of Graduate Studies

04/21/2020

DEDICATION

For Julie, for her long suffering
patience with me during my
seventeen year journey of
on-again, off-again
graduate school

ACKNOWLEDGMENTS

I am indebted first to my advisor, Dr. Michael Toland, for years of mentoring and friendship. I am confident that had I not taken his introduction to psychometrics class, my professional life would be completely different right now.

A second set of thanks goes to my committee, who were always generous and constantly encouraging. Dr. Joe Hammer periodically referred people to me who needed help with bifactor models or indices, which kept the topic near the top of my head during the long dark of dissertation topic selection. Dr. Xin Ma engaged me in entertaining arguments and off-the-wall ideas which kept me from becoming insular in my thinking. Dr. Margaret Mohr-Schroeder was unfailingly encouraging, kind, and patient.

Thirdly, I appreciate all the APS lab members through the years who have been encouraging, offered advice, listened to my ranting, had lunches with me, and generally been fantastic colleagues. A few who have been most influential are Abbey, Caihong, Carol, Caroline, Chen, Hao, John Eric, Rebecca, Rongxiu, and Zijia.

Fourthly, my various stints in graduate school have been made more fun by my wonderful sets of officemates. At the University of Kentucky, this included Ben, Candice, Derek, Emma, Grace, Kerry, Laura, Laura, Lillian, and Maureen. At Illinois State University, this included Craig and Jenny. At the University of Wisconsin this included Ben, Jeremy, John, John, Patrick, Paul, Sharon, and a couple dozen others.

Fifthly and finally, my family has made this whole process possible. Besides providing a loving a nurturing home, my parents also (when I needed it) left me alone to work on problems that captured my curiosity. A little distance was necessary to help me learn to throw myself at a problem and work through the hard parts without needing

constant supervision or help. Julie has been supportive of my quest for a PhD since long before we came to Kentucky, and has helped me navigate the varied roles I have played in our family and professionally for the past many years. I could not have done all of this without you. I love you! Lastly, thanks go to Jonathan and Abigail for making me have fun because, according to Abigail, dissertations are #boring.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1. INTRODUCTION

Psychological research commonly involves the use of a scale, consisting of multiple items, designed to assess a single construct which nevertheless exhibits some elements of multidimensionality. Confirmatory factor analysis (CFA) is a common tool used to assess the dimensionality of a set of data obtained from a scale. CFA models for unidimensional and multidimensional data can be fit and the fit of these models can be compared to determine which model exhibits superior fit. Some quantitative methodologists have observed that such tests are highly sensitive to multidimensionality and may suggest a multidimensional interpretation even when the extent of multidimensionality is not substantively relevant (Reise et al., 2013b; Sellbom & Tellegen, 2019). Instead, auxiliary indices based on hierarchical CFA models have been developed to assist in determining when total scores and unidimensional CFA models can appropriately be used (Rodriguez et al., 2016a). On the other hand, fewer and less exact strategies have been developed for determining when a multidimensional interpretation is appropriate (Chen et al., 2006; Gignac & Kretzschmar, 2017; Reise et al., 2013a). In this dissertation, the aforementioned strategies will be detailed and dissected, and new strategies devised.

1.1 Confirmatory Factor Models for Dimensionality Asssessment

Factor analysis refers to a range of techniques whose purpose is to describe the variation and covariation among a set of observed variables, called indicators, through the use of continuous latent variables called factors. In the contexts and applications considered herein, indicators are almost always individual items to which a research participant responds as part of a scale; the terms “item” and “indicator” are used mostly interchangeably, but “indicator” should be understood as being more general. A “scale” is

defined as a set of items that collectively measure a construct. The "scale score" or "total score" refers to the total of the numeric value of responses to items in the scale.

In CFA, the number of factors (latent variables) underlying a set of indicators and the pattern of associations between indicators and factors is pre-specified, and the corresponding model is estimated using maximum likelihood or other techniques. The specified model for a scale is called a measurement model because it makes a claim about how the scale measures the construct it was designed to assess. When an indicator is declared to be directly linearly related to a factor, it is said to "load" on that factor. Because models are specified before being estimated, CFA can be used to test the hypothesis that the model being estimated is the correct one by examining how well the model describes the sample data. Additionally, CFA is used to examine the strength of relationship between indicators and the factors on which they load as well as the correlational relationship between factors. Thus, CFA is well suited for use in the evaluation and analysis of theoretical models for the composition and dimensionality of constructs being measured.

As implied above, the primary units of factor analysis are a set of latent factors (F_j) and observed indicators (Y_i) which are linearly related. Specifically, the indicators are described using a multiple regression-like equation with the factors as predictors:

$$Y_i = \lambda_{i,1}F_1 + \lambda_{i,2}F_2 + \cdots \lambda_{i,}F_n + \mu_i + e_i, \tag{1.1}$$

where n is the total number of factors, $\lambda_{i,j}$ is a regression-like coefficient called the factor loading of Y_i onto F_j, μ_i is the item's intercept (having the same meaning as in regression), and e_i is a residual. Because this equation holds for each indicator, it is often expressed using matrices and vectors, as

$$\vec{Y} = \Lambda\vec{F} + \vec{M} + \vec{E}, \tag{1.2}$$

where Λ is the matrix of factor loadings referred to as the pattern matrix, $\vec{M}$ is the vector of item intercepts, and $\vec{E}$ is the vector of residual terms, sometimes referred to as unique factors because the variability of e_i is latent and unique to the indicator Y_i. An example of a matrix of factor loadings is provided for a CFA model in which 6 indicators load onto 2 factors such that each factor is measured by only 3 indicators:

$$\begin{array}{c} \\ Y1 \\ Y2 \\ Y3 \\ Y4 \\ Y5 \\ Y6 \end{array} \begin{array}{c} \begin{array}{cc} F1 & F2 \end{array} \\ \begin{bmatrix} .7 & 0 \\ .5 & 0 \\ .6 & 0 \\ 0 & .4 \\ 0 & .6 \\ 0 & .5 \end{bmatrix} \end{array} = \Lambda$$

In addition to a pattern matrix, a fully specified CFA model will include a factor covariance matrix Φ and a covariance matrix of the unique factors, Θ, also called the error covariance matrix of the indicators which represents the indicator (co-)variance which is unexplained by the latent factors. The covariance matrix of unique factors is assumed to be diagonal in all standard models (Kline, 2016); although not discussed herein, it is possible to specify CFA models in which unique factors may correlate. The various CFA models discussed below are largely distinguished by their restriction on the number of factors and the form of factor covariance matrix. Details about the specification and identification of CFA models can be found in any standard psychometrics, latent variable modeling, or structural equation modeling textbook (e.g., Raykov & Marcoulides, 2011; Finch & French, 2015; and Kline, 2016, respectively).

Estimation of CFA models typically involves determining model parameters such that the sample covariance matrix, Σ_{YY}, is reproduced as precisely as possible by the model implied covariance matrix. The model implied covariance matrix, Σ, is related to the

covariance matrix of the common factors, the pattern matrix, and the covariance matrix of the unique factors, according to the following equation:

$$\Sigma = \Lambda\Phi\Lambda^T + \Theta \tag{1.3}$$

Inexact replication of the sample covariance matrix is referred to as misfit and can bias or invalidate interpretations of the CFA model at hand. Because the model is only intended to reproduce the sample covariance matrix, raw data is not used in estimating CFA models; this approach is called limited information estimating since only a small part of the information contained in the data is utilized.

1.1.1 Unidimensional CFA Model

The simplest CFA models, unidimensional CFA models, are those in which a single factor is specified in order to explain the observed covariance matrix of indicators. A one factor CFA model with 4 indicators is displayed in Figure 1.1.

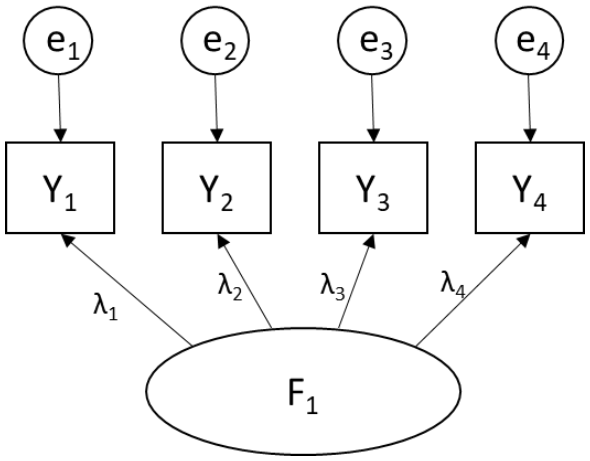

Figure 1.1 Unidimensional CFA Model

For each indicator Y_i, the only sources of indicator variance are from the common factor F_1 and the error, e_i; accordingly, the instantiation of Equation 1.1 for Y_1 of this model is

$Y_1 = \lambda_1 F_1 + e_1$. Likewise, the covariance between items is fully explained by the variance that indicators share with the factor. Using standard path tracing rules, the covariance between items X_1 and X_2 is given by $cov(X_1, X_2) = \lambda_1 \lambda_2$. In more complicated models, the error term is only represented by the incoming arrow to simplify presentation.

In the event that a unidimensional model for a scale is deemed to adequately fit the data, a single total score can reasonably be interpreted (Raykov & Marcoulides, 2011, p. 129). However, the unidimensionality assumption is a strong one which is rarely if ever met perfectly (Bentler, 2009). Even small amounts of construct heterogeneity can manifest as misfit of a unidimensional CFA model (Floyd & Widaman, 1995), but in some cases this heterogeneity does not prohibit interpretation of a total score or unidimensional measurement model (Reise, 2012; Rodriguez et al., 2016a). Thus, while adequate fit of a unidimensional model is sufficient to reasonably interpret a total score, inadequate fit of a unidimensional model does not preclude the justifiable interpretation of a total score.

1.1.2 Correlated Traits CFA Model

Correlated traits CFA models consist of several factors, each of which has several indicators loading on it; typically, indicators are not allowed to load on multiple factors (Kline, 2016, p. 193). A correlated traits model with 3 factors and 3 indicators per factor is shown in Figure 1.2. In this model, covariance between indicators loading on the same factor is explained by the loadings on the factor, whereas covariance between indicators loading on different factors additionally involves the correlations between factors. Unique factors are not shown for simplicity.

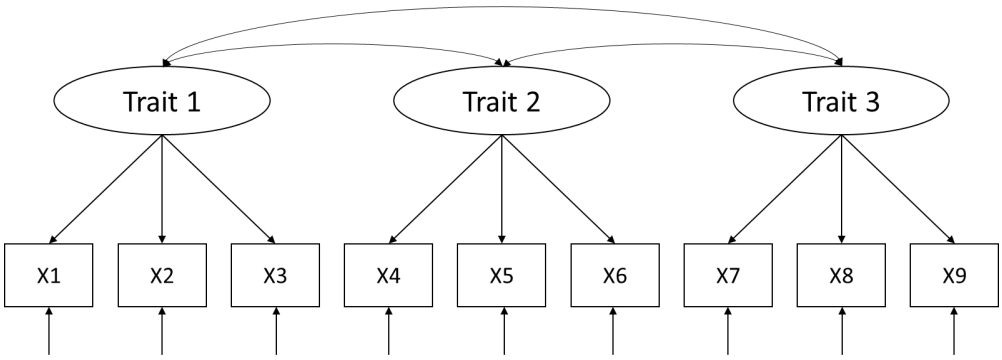

Figure 1.2 Three-Factor Correlated Traits Model

Correlated traits CFA may be used to model multiple scales or to model subdomains of a single scale measuring a multidimensional construct, the latter of which will be the setting for all discussion within this manuscript. While goodness of fit of a correlated traits model is often used as justification for employing a multidimensional interpretation and utilizing subdomain scores (termed subscores), this perspective fails to acknowledge that the degree of multidimensionality may be inconsequential in which case a unidimensional interpretation may be more appropriate (Reise et al., 2013a). Namely, if the factors in the correlated traits model are too strongly correlated, then subscores may be largely redundant and only a total score should be interpreted. For example, Haberman and Sinharay (2010) reported a subscore analysis of an unnamed test measuring skills necessary for paraprofressional. The three subscores had an average inter-correlation of .76 and reliability estimates higher than .80, yet none showed evidence of being interpretable independently of the total score. Within a correlated traits model, it is difficult to determine whether interpreting subscores separately from the total score is likely to be useful. For that reason, models which utilize both a general source of variance (as in a unidimensional model) and subdomain specific sources of variance (as in a correlated traits model) are often used for nuanced discussions of dimensionality.

1.1.3 Second-Order CFA Model

Second-order CFA models (a type of higher-order model) differ from other CFA models in that a separate set of latent variables is included to model the covariance between the first-order factors. To put it another way, the first-order factors comprise a measurement model for the indicators, while the second-order factors comprise a measurement model for the first-order factors. An example of a second-order CFA model with one second order factor and three first order factors can be found in Figure 1.3.

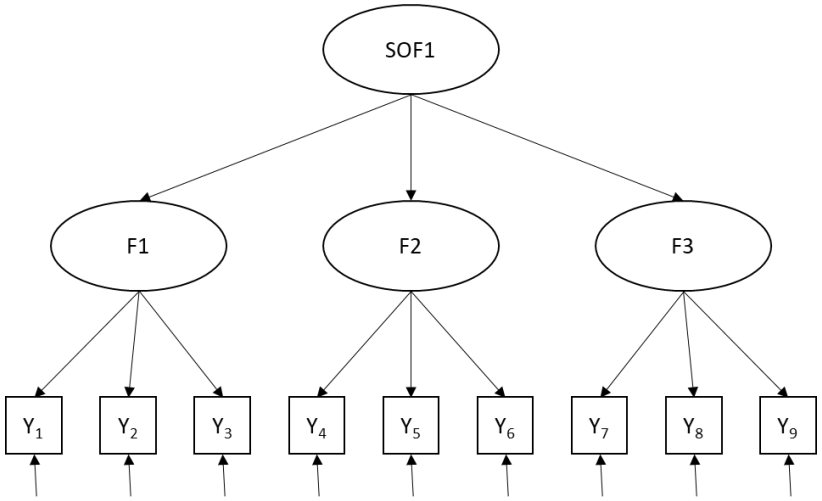

Figure 1.3 Example of a Second-Order CFA Model

In a second-order CFA model, the covariance matrix of first-order factors can be expressed analogously to Equation 1.3 as

$$\Phi_1 = \Gamma\Phi_2\Gamma^T + \Psi \tag{1.4}$$

where Φ_1 is the covariance matrix of first order factors, Φ_2 is the covariance matrix of second order factors, Γ is the factor loading matrix of first-order factors onto second-order factors, Ψ is the residual covariance matrix of first order factors, and ^{T} denotes matrix transpose. As with the covariance matrix of unique factors for indicators, Ψ is assumed to

be diagonal. The model reproduced covariance of indicators can be obtained by substituting Equation 1.4 into Equation 1.3, yielding the expression

$$\Sigma = \Lambda(\Gamma\Phi_2\Gamma^T + \Psi)\Lambda^T + \Theta. \quad (1.5)$$

Second-order factors models are more restrictive than correlated traits models because the covariance amongst first-order factor must itself have the structure of indicators in a CFA model, per Equation 1.4.

Second-order factor models commonly only include one second-order factor which is intended to represent a general factor (Kline, 2016), and only models with a single second-order factor are considered henceforth. Note that, while the second-order factor models the covariance among first-order factors, the first order factors are allowed a residual variance. Accordingly, a second order factor may explain little of the variance of the first order factors if they are weakly correlated, or much of the variance if they are strongly correlated. Thus, indicator explained variance can be split into variance explained by a general, second-order, factor and variance explained by the residual of the appropriate first-order factor. With the exception of specific domains such as intelligence and personality research, the use of second-order models has not been widespread, likely due to the difficulty of their estimation, the likelihood of poor fit given the strictness of constraints placed on the first-order factor covariance matrix, and the difficulty in interpreting relationships between first-order factors and external variables while controlling for the general second-order factor (Chen et al., 2006; Chen et al., 2012).

1.1.4 Bifactor CFA Model

In a bifactor CFA model, all indicators load onto a general factor, and indicators are additionally allowed to load onto uncorrelated specific factors (Holzinger & Swineford,

1937). Thus, like the second-order factor model, explained variance of indicators is partitioned between a general factor and a specific factor. The matrix of factor loadings in a bifactor model is constrained so that all indicators load onto the general factor and at most one specific factor. A diagram for a bifactor model with 6 items and 2 specific factors can be found in Figure 1.4. Arrows for indicator error variances are suppressed so as not to make the diagram needlessly complicated.

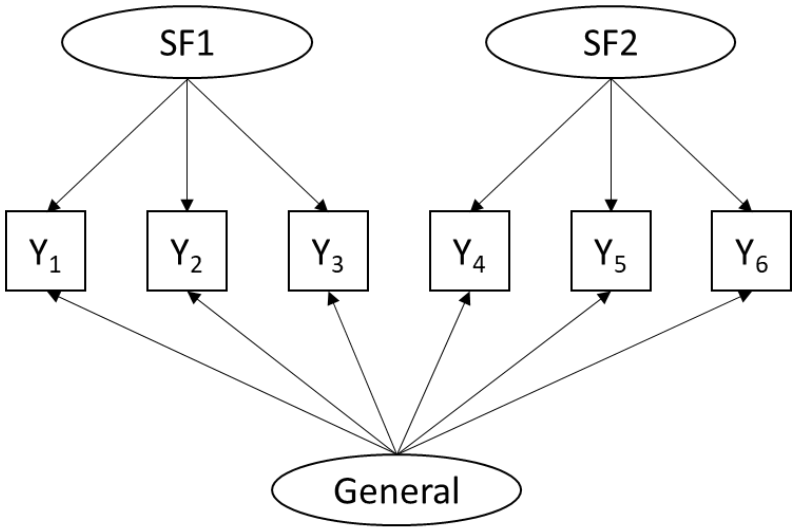

Figure 1.4 Diagram of a Bifactor CFA Model

The general factor of a bifactor model is frequently the only latent variable of interest, while the specific factors are considered as residual, nuisance factors (DeMars, 2013; Reise, 2012; Rodriguez et al., 2016a). However, bifactor models can also be used with a focus on the specific factors, for example to examine the extent to which subscales are distinct from the general factor (Gignac & Kretzschmar, 2017; Gignac & Watkins, 2013; Reise, 2012; Reise et al., 2013a; Rodriguez et al., 2016b), to test whether apparent factors are best interpreted as method factors (e.g., item phrasing factors) or as substantive factors (McKay et al., 2015), or to examine the contribution of specific factors to prediction of external variables (Chen et al., 2006; Gonzalez & MacKinnon, 2018).

As previously noted, second-order and bifactor CFA models are closely related, as both can be interpreted as partitioning indicator explained variance into general and specific sources. In fact, Yung et al. (1999) demonstrate that all second-order CFA models can be reparameterized into a statistically equivalent (i.e., the model implied covariances matrices are the same) bifactor model. Furthermore, correlated traits factor models can be reparameterized as bifactor models exactly when the correlated traits model is statistically equivalent to a second-order model. Informally, the set of second-order factor models is the intersection of the sets of correlated traits and bifactor CFA models, as displayed in Figure 1.5.

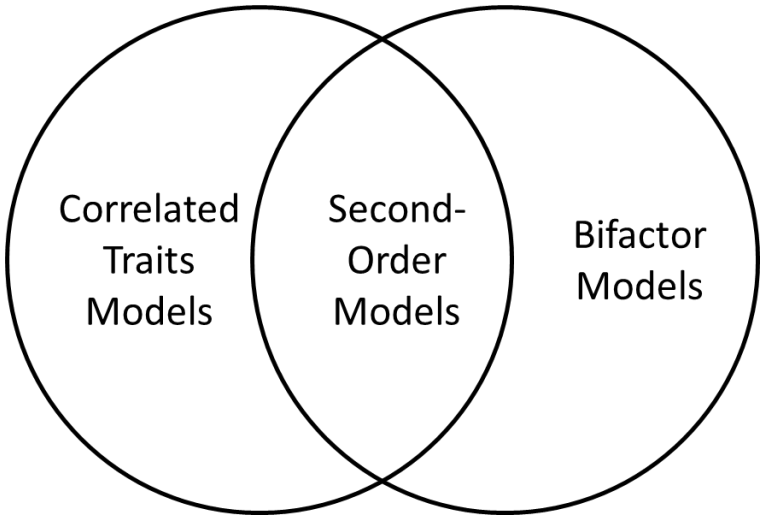

Figure 1.5 Venn Diagram of CFA Model Relationships

As with the other multidimensional models discussed (i.e., correlated traits and second-order CFA models), the existence of a well-fitting bifactor model does not immediately imply that a specific interpretation of the data is appropriate. It is possible that a bifactor model with a weak general factor fits the data well and yet a total score should not be interpreted, but rather the subscores should be interpreted. Contrariwise, a bifactor model with a strong general factor may fit the data such that a total score may be

interpreted, but the subscores are too highly correlated to provide meaningful information above and beyond the total score. To assist in decision making about dimensionality and score interpretations based on bifactor models, a number of statistical indices related to bifactor model parameters have been developed.

1.2 Bifactor Indices for Dimensionality Assessment

In a bifactor CFA model, items are allowed to crossload onto both the general factor and a specific factor; therefore, the variance of each item is split into three components: covariance with the general factor, covariance with the specific factor, and item specific variance, as depicted in Figure 1.6. For the general factor to be interpretable as the primary dimension, it stands to reason that the amount of shared variance explained by the general factor should be substantial relative to the amount of shared variance explained by the specific factor. Various statistical indices estimating the partitioning of variance and covariance across items computed from the parameters of the bifactor model have been devised and can be used for evaluating the appropriateness of making uni- or multi-dimensional interpretations of the data.

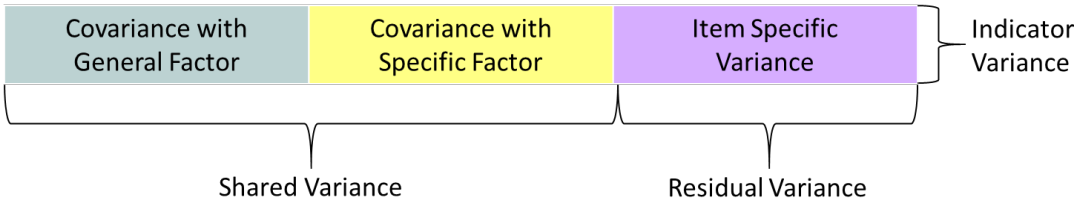

Figure 1.6 Partitioning of Item Variance in a Bifactor Model

1.2.1 Omega Indices

1.2.1.1 Omega

Coefficient omega (Omega, ω; McDonald, 1999) is a model-based estimate of composite reliability of total score. Omega is computed using the estimated parameters

(namely factor loadings and residual variances) of a statistical model; it estimates the proportion of variance in the total score explained by common variance (i.e., all common factors) implied by the model (Bentler, 2009; Raykov, 1997; Revelle & Zinbarg, 2009). While Omega is typically used with unidimensional models, Zinbarg et al. (2005, p. 126, Equation 8) imply the following formula for Omega based on bifactor model parameters:

$$\omega = \frac{\Sigma_i(\lambda_{i,G})^2 + \Sigma_j \Sigma_i(\lambda_{i,S_j})^2}{\Sigma_i(\lambda_{i,G})^2 + \Sigma_j \Sigma_i(\lambda_{i,S_j})^2 + \Sigma_i(1-h_i^2)}, \quad (1.6)$$

where i varies over all items, j varies over all specific factors, $\lambda_{i,G}$ is the loading of item i onto the general factor, λ_{i,S_j} is the loading of item i onto specific factor j, and h_i^2 is the communality of item i. Omega has a slightly different interpretation than many other reliability coefficients; specifically, it includes multidimensional sources of common variance and therefore does not represent the correlation between the total score and a single latent variable. Rather, Omega is simply interpreted as the ratio of variance explained by commonality amongst items to the total variance of the total scale score (McNeish, 2017).

1.2.1.2 OmegaH

Whereas omega estimates the proportion of total score variance that can be explained within the bifactor model, hierarchical omega (OmegaH; ω_H; McDonald, 1999; Zinbarg et al., 2005) estimates the proportion of total score variance that can be explained by the general factor, and is computed similarly as in Equation 1.6 except that only loadings from the general factor are considered in the numerator (Zinbarg et al., 2005, p. 126, Equation 8):

$$\omega_H = \frac{\sum_i \left(\lambda_{i,G}\right)^2}{\sum_i \left(\lambda_{i,G}\right)^2 + \sum_j \sum_i \left(\lambda_{i,S_j}\right)^2 + \sum_i (1 - h_i^2)}. \quad (1.7)$$

From the perspective of OmegaH, variability explained by the group factors is considered as measurement error. While OmegaH does not directly address the issue of unidimensionality (Reise et al., 2007), Gustafsson and Åberg-Bengtsson (2010) and McDonald (1999) argue that high OmegaH indicates that total scores primarily reflect a single dimension, since it may be interpreted as the squared correlation between observed total scale score and the latent general factor (McDonald, 1999). Like many estimates of reliability, OmegaH is strongly influenced by scale length; indeed, for long scales with many specific factors, OmegaH can be high even when the data is plainly multidimensional (Reise et al., 2013b).

1.2.1.3 OmegaS

An estimate of composite reliability of subscores (OmegaS; ω_S) can also be computed for each specific factor. The OmegaS index has a formula (Reise et al., 2013a, p. 134, Equation 5) similar to the formula for Omega, except that only items from a particular specific factors are included:

$$\omega_S = \frac{\sum_i \left(\lambda_{i,G}\right)^2 + \sum_i \left(\lambda_{i,S}\right)^2}{\sum_i \left(\lambda_{i,G}\right)^2 + \sum_i \left(\lambda_{i,S}\right)^2 + \sum_i (1 - h_i^2)}, \quad (1.8)$$

where i varies only over the items loading on specific factor S. Like Omega, OmegaS is not generally interpretable as the squared correlation between a total score and a latent factor, but rather as the proportion of variance in the subscale score explained by common variance implied by the model (Rodriguez et al., 2016a). However, in the special case that specific factor loadings are proportional to general factor loadings, the items on the specific

factor can be modeled using a unidimensional model (Yung et al., 1999). In this case, therefore, while OmegaS does not necessarily equal the Omega estimate from a unidimensional model, it will nevertheless have the same interpretation as Omega for a unidimensional model, which is an estimate of the squared correlation between a total score and the single latent factor (McDonald, 1999; Zinbarg et al., 2005).

1.2.1.4 OmegaHS

Hierarchical omega for a subscale (OmegaHS; ω_{HS}) estimates the proportion of subscore variance that can be explained by the corresponding specific factor, and is computed (Reise et al., 2013a, p. 134, Equation 6) as

$$\omega_{HS} = \frac{\sum_i \left(\lambda_{i,S}\right)^2}{\sum_i \left(\lambda_{i,G}\right)^2 + \sum_i \left(\lambda_{i,S}\right)^2 + \sum_i (1 - h_i^2)}, \tag{1.9}$$

where i varies only over the items on the subscale S. According to Reise et al. (2013a), OmegaHS reflects the reliability of a subscore after partialling out variability explained by the general factor, consistent with the interpretation of specific factors as being residuals after the general factor is accounted for (DeMars, 2013). However, Perreira et al. (2018) note that this interpretation of OmegaHS would require variability explained by the general factor to also be removed from the denominator of Equation 1.9. Therefore, in the form of Equation 1.9, OmegaHS is not a reliability coefficient, as it is not the ratio of true score variance to observed score variance for any set of scores. Yet, Reise et al. (2013a) and Rodriguez et al. (2016b) suggest that OmegaHS can be interpreted as a measure of dimensional uniqueness of the specific factor, while Sellbom and Tellegen (2018) instead recommend applying this interpretation to the ratio of OmegaHS to OmegaS. Finally, Gignac and Kretzschmar (2017, p. 138) refer to OmegaHS as "an effect size index of unique latent variable strength." Consequently, these different interpretations can lead to

confusion among applied researchers who are leaning on methodologists for guidance on how to properly interpret such indices.

1.2.2 Explained Common Variance Indices

1.2.2.1 ECV

Explained common variance (ECV; Sijtsma, 2009; ten Berge & Socan, 2004) is the proportion of all common variance explained by the general factor,

$$ECV = \frac{\sum_i \lambda_{i,G}^2}{\sum_i \lambda_{i,G}^2 + \sum_j \sum_i \lambda_{i,S_j}^2}, \tag{1.10}$$

where i varies over all items, j varies over all specific factors, $\lambda_{i,G}$ is the loading of item i onto the general factor, and λ_{i,S_j} is the loading of item i onto specific factor j. While ECV has some similarities to OmegaH, they assess somewhat different things. Whereas OmegaH can be viewed as a measure of unidimensionality of a total score, ECV is a measure of the unidimensionality of the data from a latent variable modeling perspective (Reise et al., 2013b; Rodriguez, et al., 2016b). Notably, ECV does not depend on the residual variances of the items, and therefore can be high even when items have little shared common variance. Finally, since loadings are squared before being summed, ECV also differs from OmegaH in that it is independent of scale length and the number of specific factors.

1.2.2.2 ECV_{GS}, ECV_{SS}, and ECV_{SG}

Explained common variance indices can also be computed for each specific factor. The proportion of common variance of the items in specific factor S explained by the general factor is referred to as "within-domain ECV" by Stucky and Edelen (2015, p. 201) and is computed using the same formula as ECV (i.e., Equation 1.10), except that only

items belonging to the specific factor of interest are used. This specific-factor ECV index is denoted ECV_{GS}, which is meant to be interpreted as the proportion of common variance explained by the general factor amongst items in the specific factor. High values of ECV_{GS} indicate that common variance in the subscale is largely subsumed by the general factor, while low values indicate that the subscale is more independent of the general factor. In the case that the bifactor model is statistically equivalent to a second-order model, ECV_{GS} will be equal to the square of the second-order factor loading for that subscale, per Equation 1.10 and Quinn's (2014, Appendix A) description of the correspondence between second-order and bifactor models.

The complement of ECV_{GS} is the proportion of common variance of the items in specific factor S explained by specific factor S, ECV_{SS}. The formula for ECV_{SS} is

$$ECV_{SS} = \frac{\sum_i \lambda_{i,S}^2}{\sum_i \lambda_{i,G}^2 + \sum_i \lambda_{i,S}^2}, \tag{1.11}$$

where sums are taken only over items loading on the specific factor of interest, and $ECV_{SS} + ECV_{GS} = 1$ for specific factors. ECV_{SS} is the proportion of common variance in a subdomain which is unique to that subdomain's specific factor. An advantage offered by ECV_{SS} is that it can be interpreted as the proportion of common variance of items on a factor explained by that factor even in models more general than bifactor models. For example, in a model with multiple general factors and multiple specific factors (two-tier; Cai, 2010), ECV_{SS} for each general factor can be interpreted as though it were the ECV of a corresponding bifactor model with only that one general factor. In the case that the bifactor model is statistically equivalent to a second-order model, ECV_{SS} will be equal to the residual variance of the first-order factor corresponding to that subscale.

A third specific factor ECV index is ECV_{SG}, which is the proportion of common variance explained by specific factor S with respect to all items from the general factor. The formula for ECV_{SG} is the same as the equation for ECV_{SS} (i.e., Equation 1.11) except that all items in the scale are used. Stucky and Edelen (2015, p. 199) refer to ECV_{SG} as "specific-dimension ECV" and claim it is a measure of the uniqueness of the specific factor. This interpretation is suspect, as the loadings of items from other subdomains are unrelated to how well a subdomain is differentiated from the general construct. Instead, ECV_{SG} measures the portion of total item explained variance captured by the specific factor. As such, high ECV_{SG} in a specific factor will lead to lower ECV of the general factor, and thereby diminish the appropriateness of applying a unidimensional measurement model (Bonifay et al., 2015; Reise et al., 2013b).

As the various indices have similar formulae and are therefore easily confused, the following example is provided for clarification. Consider a bifactor model with two specific factors each being comprised of three indicators, as pictured in Figure 1.7. The standardized general factor loadings are all .60, the standardized loadings on the first specific factor are all .30, and the standardized loadings on the second specific factor are all .20. Computations for ECV indices can be found in Table 1.1. From the results in Table 1.1, it can be seen that for a given specific factor, $ECV_{GS} + ECV_{SS} = 1$. Also, the sum of ECV and all ECV_{SG} indices is always 1, since all of the common variance is partitioned into either a general source (ECV) or a specific source (ECV_{SG}).

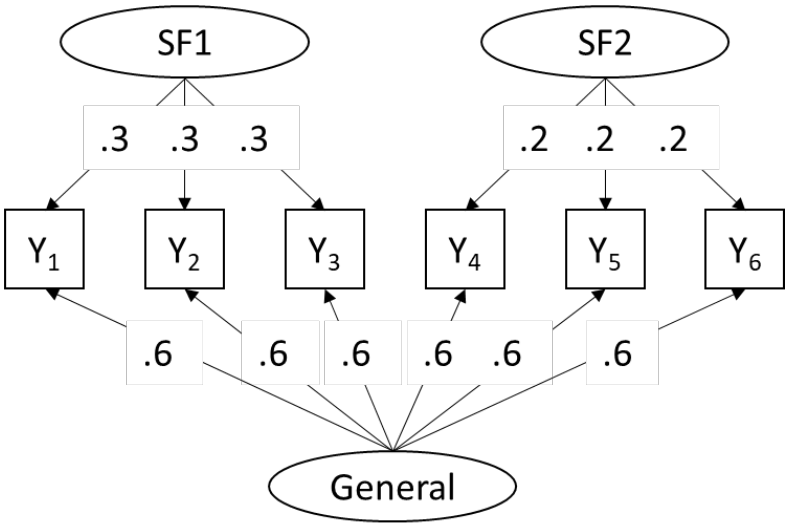

Figure 1.7 Example Bifactor Model with Standardized Loadings

Table 1.1 ECV Indices for Example Bifactor Model

Index	**Factor**	**Formula**	**Value**
ECV	General	$\frac{6 \times .6^2}{6 \times .6^2 + 3 \times .3^2 + 3 \times .2^2}$	.847
ECV_{GS}	SF1	$\frac{3 \times .6^2}{3 \times .6^2 + 3 \times .3^2}$	.800
ECV_{GS}	SF2	$\frac{3 \times .6^2}{3 \times .6^2 + 3 \times .2^2}$	.900
ECV_{SS}	SF1	$\frac{3 \times .3^2}{3 \times .6^2 + 3 \times .3^2}$	.200
ECV_{SS}	SF2	$\frac{6 \times .2^2}{3 \times .6^2 + 3 \times .2^2}$	.100
ECV_{SG}	SF1	$\frac{3 \times .3^2}{6 \times .6^2 + 3 \times .3^2 + 3 \times .2^2}$	.106
ECV_{SG}	SF2	$\frac{3 \times .2^2}{6 \times .6^2 + 3 \times .3^2 + 3 \times .2^2}$	.047

1.2.2.3 I-ECV

An explained common variance index can also be computed for each item. The proportion of common variance for an item which is explained by the general factor is (Stucky & Edelen, 2015, p. 201)

$$I - ECV = \frac{\lambda_{i,G}^2}{\lambda_{i,G}^2 + \lambda_{i,S}^2}, \tag{1.12}$$

where i is the index of a single item. I-ECV is a measure of how well an item's common variance is explained by the general factor (Stucky et al., 2013). Stucky and Edelen (2015) recommend using I-ECV to select a subset of items to include in a shortened scale which is essentially unidimensional, claiming that a set of items with I-ECV "greater than 0.80 or 0.85 will typically yield a fairly unidimensional item set" (p. 202).

1.2.3 Other Bifactor Indices

1.2.3.1 Percent uncontaminated correlations (PUC)

Percent uncontaminated correlations (PUC) provides the proportion of elements of the covariance matrix which are only modeled by the general factor (Rodriguez et al., 2016a, Equation 8, p. 232),

$$PUC = 1 - \frac{Number\ of\ correlations\ among\ items\ within\ group\ factors}{Total\ number\ of\ correlations} \tag{1.13}$$

Since a unidimensional model attempts to replicate all the elements of the data's covariance matrix and the general factor of a bifactor model influences all elements of the model-implied covariance matrix, it stands to reason that if only few covariances are influenced by specific factors, the unidimensional model and general factor will be similar (Bonifay et al., 2015; Reise et al., 2013b). As an example, consider a test with 9 items allocated

evenly among three subdomains. The covariance matrix (see Figure 1.7) has 9 variance terms (highlighted in lavender) and 9(9 – 1)/2 = 36 covariance terms, of which only 9 (highlighted in blue) are affected by the specific factors. Thus, the remaining 27 (highlighted in tan) are uncontaminated by the subdomains; for this model, PUC = 0.75, which means that 75% (27 out of 36) of the covariance terms in the covariance matrix are modeled only by the general factor of the bifactor model.

v1								
c12	v2							
c13	c23	v3						
c14	c24	c34	v4					
c15	c25	c35	c45	v5				
c16	c26	c36	c46	c56	v6			
c17	c27	c37	c47	c57	c67	v7		
c18	c28	c37	c48	c58	c68	c78	v8	
c19	c29	c39	c49	c59	c69	c79	c89	v9

Figure 1.8 Partitioning of Covariance Matrix into Contaminated and Uncontaminated Covariances

1.2.3.2 Average relative parameter bias (ARPB)

Average relative parameter bias (ARPB) is an overall index of the difference between factor loadings on the bifactor general factor and the factor loadings of a unidimensional CFA model of the same data. Specifically, ARPB is the average of the difference in factor loadings between the models relative to the factor loadings on the unidimensional model,

$$ARPB = \frac{\sum_i \left| \frac{\lambda_{i,U} - \lambda_{i,G}}{\lambda_{i,G}} \right|}{n} \tag{1.16}$$

where the sum is taken overall all items, $\lambda_{i,U}$ is the factor loading of item i onto the unidimensional factor, $\lambda_{i,G}$ is the factor loading of item i onto the bifactor general factor, and n is the total number of items. When ARPB is small, the general factor of the bifactor solution and the single factor of the unidimensional solution will be roughly equivalent models of the data; therefore, the simpler unidimensional model can be used in SEM contexts without substantially biasing structure coefficients (Bonifay et al., 2015; Rodriguez et al., 2016a).

1.3 Use of Bifactor Models for Unidimensionality Assessment

Bifactor indices have been utilized in simulation and theoretical studies to develop cutoffs for when data can reasonably be interpreted in a unidimensional manner. As these simulations provide much of the motivation and inform much of the design of the simulation proposed herein, they will be described in detail.

Reise et al. (2013b) compared the accuracy with which the factor in a unidimensional model predicts a criterion when the true measurement model was bifactor. Specifically, they specified a bifactor population model along with a criterion variable with a fixed latent correlation (.50) to the general factor. Then, an analysis model consisting of a unidimensional model which predicted the criterion in a structural equation model (SEM) was used. The primary outcome of interest in this study was relative bias in the structural coefficient. Relative bias, also known as percent bias, in a coefficient is defined as the ratio of the difference between theoretical and estimated coefficients and the theoretical coefficient (Bonifay et al., 2015, p. 5). Reise et al. (2013b) manipulated the number of factors (3, 6, 12), the number of items per factor (3, 6, 12), standardized loadings on the general factor (.3, .4, .5, .6, .7), and standardized loadings on specific factors (.3, .4, .5, .6).

Notably, rather than simulating data, Reise et al. (2013b) generated a population correlation matrix for each condition and performed analyses using this correlation matrix as the data. Because this study was based on population models, it was not a simulation study, as no data was simulated. Rather, the results are relevant to hypothetical true models; in practice, sampling error will add another source of error to the prediction of external correlates. The design of this study guarantees that all measurement models considered are second-order models (Yung et al., 1999). Additionally, in Reise et al.'s (2013b) design, all specific factor bifactor indices were the same for all specific factors in a given model since all relevant parameters (number of items, magnitude of factor loadings on general factor, magnitude of factor loadings on specific factor) were the same for all specific factors..

Reise et al. (2013b) found that relative bias in the structural coefficient was predicted by ECV, and that this relationship was moderated by PUC. Reise et al. (2013b, p. 22) suggest that when PUC > .80 or when ECV > .60 and OmegaH > .7, relative bias in structural coefficients induced by using a unidimensional measurement model rather than a bifactor measurement model in an SEM framework is likely to be slight. When these cutoffs are met, therefore, use of a unidimensional latent variable model may be justified without too much concern about structural parameter bias.

Bonifay, et al. (2015) compared loadings of a unidimensional model to loadings on the general factor of a bifactor model in situations where the bifactor model was the true population model. Specifically, they simulated data from a bifactor population model and then analyzed that data using a unidimensional measurement model. The primary outcome of interest in this study was average relative bias in loadings between the estimated unidimensional model and the population bifactor model. Here average relative bias is the

average of relative biases in loadings across all the items in the model. Bonifay et al. (2015) manipulated the number of factors and the number of items per factor to create 15 different bifactor structures. They then manipulated standardized loadings on the general factor (.3, .4, .5, .6, .7), and standardized loadings on specific factors (.3, .4, .5, .6). Notably, rather than simulating many samples of data per condition, Bonifay et al. (2015) generated a single large sample (N = 10,000) which they analyzed, making the assumption that loadings and bifactor indices will be precisely estimated with such a large sample. In the same way as Reise et al.'s (2013b) design, the design of Bonifay et al.'s (2015) simulation guarantees that all measurement models considered are second-order models and that all specific factor bifactor indices are the same for all specific factors in a given model. Bonifay et al. (2015) found that ECV predicted average relative bias in factor loadings, and that this relationship was moderated by PUC, mirroring the results of Reise et al. (2013). No cutoffs were provided.

Finally, it should be noted that neither Reise et al. (2013) nor Bonifay et al. (2015) make any claims about the appropriateness of interpreting a total score; instead, their claims are limited to measurement models. The only claim concerning using bifactor indices to justify interpretation of a total score found in the literature is given by Rodriguez et al. (2016a), who claim OmegaH > .80 is sufficient to claim that total scores can be considered as effectively unidimensional, and therefore may be interpreted. In making this recommendation, Rodriguez et al. (2016a) do not refer to any research literature. However, it seems likely that their rationale was inspired by Nunnally and Bernstein's (1994, p. 265) well-known claim that reliabilities above .80 are adequate for research "concerned with the size of correlations and with mean differences."

1.4 Use of Bifactor Models for Multidimensionality Assessment

While no explicit simulations have been conducted using bifactor indices to measure dimensional uniqueness of subscales, numerous researchers have recommended using OmegaHS for this purpose (Gignac & Kretzschmar, 2017; Gignac & Watkins, 2013; Reise et al., 2012; Rodriguez et al., 2016b; Sellbom & Tellegen, 2019). Gignac and Watkins (2013) and Reise et al. (2013a) recommend not interpreting a subscore with OmegaHS < .50, while Reise et al. additionally suggest that higher values such as OmegaHS = .75 would be preferred. Both of these recommendations are based upon interpreting OmegaHS as a reliability coefficient, which has already been demonstrated to be a flawed interpretation (Perreira et al., 2018). Gignac and Kretschmar (2017) utilize the results of a literature review performed by Rodriguez et al. (2016a) to suggest when OmegaHS values may be considered "small," "medium," and "large," but attach no meaning to those labels other than in relation to each other.

The use of OmegaHS for assessment of individual specific factors has been observed as a common practice in applied literature. Specifically, a search was made of PsychINFO for peer reviewed articles published in 2018 using the search term "bifactor." This search revealed 149 articles reporting at least one exploratory or confirmatory bifactor model, of which 58 interpreted OmegaHS coefficients for the purpose of dimensionality assessment. The most common interpretation given to OmegaHS among these articles was consistent with Reise et al.'s (2013a) interpretation as reliability of the subscore after partialling out variability explained by the general factor. Less common was to consider OmegaHS as an estimate of the reliability of the subscore. Generally, small OmegaHS values were considered as evidence that only a total score should be interpreted; for

example, Shihata et al. (2018) argue that small OmegaHS values (.07 - .24) indicate that only a total score of intolerance of uncertainty should be interpreted, but later interpret the inhibitory subscore anyways. Decisions based on moderate levels of OmegaHS were more variable. On the one hand, Bruner and Benson (2018) use OmegaHS values between .27 and .41 as evidence to support interpreting only a general factor of social identity and total score despite low ECV (.50) and marginal OmegaH (.78). Several other researchers interpret OmegaHS between .40 and .50 as being inadequate to interpret a subscore (e.g., Dagnall et al., 2018; Isoard-Gautheur et al., 2018; Naser et al., 2018). On the other hand, Folberg et al. (2019) argue that similarly sized OmegaHS values (.40 and .49) indicate that dominance and self-direction goals specific factors need to be included in a measurement model for agentic goals (of which they are subdomains) in an SEM. It is noteworthy that in three of these studies, both a general factor or total score and at least one subscore are interpreted. Thus, in practice, scales are sometimes treated as both effectively unidimensional and effectively multidimensional, even within a single study.

1.5 Purpose

Applied researchers employ bifactor CFA models and indices based on model parameter estimates for dimensionality assessment purposes, including determining whether to interpret subscales. However, to date no rigorous guidelines for accomplishing this determination have been developed. The present research endeavors to partially close this gap between methodology literature and research practice by aligning bifactor indices with a classical test theoretic approach to subscore analysis common in educational testing (Sinharay, 2019). It is hoped that the present research will provide results that aid psychology, education, and, more broadly, social science researchers in making rigorous

decisions about whether to interpret subscores for use in research concerning, for example, group mean differences or covariance structures such as regression. Therefore, the purpose of this dissertation is to accomplish two closely related goals:

(1) Develop and disseminate a package for the R statistical computing environment (R Core Team, 2019) to efficiently compute bifactor indices from bifactor CFA, EFA, or exploratory SEM model parameters, with special convenience functions for inputting fitted model results from Mplus (Muthén & Muthén, 2019), the lavaan R package (Rosseel & Jorgensen, 2018), and the psych R package (Revelle, 2020). Dissemination will involve publication of the package on the Comprehensive R Archive Network (CRAN) and preparation of a vignette manuscript suitable for publication in the Journal of Statistical Software (impact factor = 11.655, 5 year impact factor = 20.539; Journal Citation Reports (JCR), March 19, 2020) or the R Journal (impact factor = 2.682, 5 year impact factor = 3.377; JCR, March 19, 2020).

(2) Use of simulation techniques to devise a strategy for which bifactor indices can be used to determine whether interpretation of subscores is appropriate. Specifically, cutoffs will be devised for a specific factor's bifactor indices, possibly conditioned upon general bifactor indices, such that exceeding these cutoffs provides empirical evidence for the appropriateness of interpreting that factor's subscore separately from the total score.

CHAPTER 2. AN R PACKAGE FOR COMPUTING BIFACTOR INDICES

2.1 Introduction

Many psychological constructs are measured using multi-item scales. In this case, it is common for researchers to model data arising from those using a latent variable model. Frequently, unidimensional confirmatory and exploratory factor analysis models are found to exhibit poor fit to questionnaire data due to multidimensionality resulting from clusters of items belonging to subdomains of the general construct (Chen et al., 2006). However, the extent of this multidimensionality may be ignorable, so that a unidimensional interpretation is warranted despite the poor fit (Reise et al., 2013b; Sellbom & Tellegen, 2018). In order to investigate the extent of multidimensionality in data, the use of ancillary indices computed from parameter estimates of a bifactor measurement model has become common (Rodriguez et al., 2016a).

Bifactor models are a specific type of latent variable model in which each indicator loads on a general factor and at most one orthogonal specific factor. Thus, bifactor models partition the variance of the indicators into general, specific, and unique sources (Reise, 2012). This partitioning of variance enables two primary purposes for bifactor models. First, the biasing effects of multidimensionality can be accounted for, allowing accurate estimation of coefficients related to the general factor (Reise et al., 2010). Second, the partitioning of variance can be studied to determine the extent of multidimensionality; if a general factor explains the vast majority of the variance of items, then multidimensional data can be treated as unidimensional without causing too much bias (Bonifay et al., 2015; Reise et al., 2013a; Rodriguez et al., 2016a). To accomplish this latter goal, a variety of auxiliary statistical indices for bifactor models have been developed. These bifactor indices

include various forms of reliability indices, aggregate variance explained indices, and others; bifactor indices have been successfully used to determine when data is unidimensional enough to be interpreted as unidimensional or multidimensional enough that subdomain scores may be interpreted instead of (or even in addition to) a total score (Bonifay et al., 2015; Dueber, 2019; Reise et al., 2013a; Reise et al., 2013b; Rodriguez et al., 2016a; Stucky & Edelen, 2015).

A search of PsychINFO for peer-reviewed articles published in 2018 using the search term "bifactor" revealed 65 papers which utilized ancillary bifactor indices to aid in decision-making about dimensionality. These studies spanned a wide variety of psychological constructs, including gender roles (Hammer et al., 2018), memory (McGill & Dombrowski, 2018), burnout (Isoard-Gautheur et al., 2018), intelligence (Fenollar-Cortés et al., 2019), emotional distress (Marshall et al., 2018), belief in the paranormal (Drinkwater et al., 2018), personality (Dagnall et al., 2018), racial attitudes (Keum et al., 2018), and many others. As such, bifactor indices are used across a wide range of psychological sciences.

2.2 Computing Bifactor Indices

Model based reliability and explained common variance indices can be computed for bifactor models. Coefficient omega (ω; McDonald, 1999) is a model-based estimate of composite reliability of total score, typically computed for unidimensional models. However, Zinbarg et al. (2005, p. 126) provide an extension of the logic for omega such that it can be computed for bifactor models as

$$\omega = \frac{\sum_i \left(\lambda_{i,G}\right)^2 + \sum_j \sum_i \left(\lambda_{i,S_j}\right)^2}{\sum_i \left(\lambda_{i,G}\right)^2 + \sum_j \sum_i \left(\lambda_{i,S_j}\right)^2 + \sum_i (1 - h_i^2)}, \tag{2.1}$$

where *i* varies over all items, *j* varies over all specific factors, $\lambda_{i,G}$ is the loading of item *i* onto the general factor, λ_{i,S_j} is the loading of item *i* onto specific factor *j*, and h_i^2 is the communality of item *i*. Whereas omega estimates the proportion of total score variance that can be explained within the bifactor model, hierarchical omega (OmegaH; ω_H; McDonald, 1999; Zinbarg et al., 2005) estimates the proportion of total score variance that can be explained by the general factor, and is computed similarly as in Equation 2.1 except that only loadings from the general factor are considered in the numerator (Zinbarg et al., 2005, p. 126, Equation 8):

$$\omega_H = \frac{\sum_i \left(\lambda_{i,G}\right)^2}{\sum_i \left(\lambda_{i,G}\right)^2 + \sum_j \sum_i \left(\lambda_{i,S_j}\right)^2 + \sum_i (1 - h_i^2)}. \tag{2.2}$$

Explained common variance (ECV; ten Berge & Socan, 2004) is the proportion of all common variance explained by the general factor,

$$ECV = \frac{\sum_i \lambda_{i,G}^2}{\sum_i \lambda_{i,G}^2 + \sum_j \sum_i \lambda_{i,S_j}^2}, \tag{2.3}$$

where i varies over all items, j varies over all specific factors, $\lambda_{i,G}$ is the loading of item i onto the general factor, and λ_{i,S_j} is the loading of item i onto specific factor j.

Average absolute relative parameter bias (ARPB; Bonifay et al., 2015; Rodriguez et al., 2016a) is a measure of the extent of deviation between loadings in a unidimensional measurement model and loadings of the general factor in a bifactor model:

$$ARPB = \frac{\sum_i \left| \frac{\lambda_{i,U} - \lambda_{i,G}}{\lambda_{i,G}} \right|}{n} \tag{2.4}$$

where the sum is taken overall all items, $\lambda_{i,U}$ is the factor loading of item *i* onto the unidimensional factor, $\lambda_{i,G}$ is the factor loading of item *i* onto the bifactor general factor,

and *n* is the total number of items. Percentage of uncontaminated correlations (PUC) is the proportion of inter-indicator correlations which are modeled only by the general factor. Omega, OmegaH, ECV, PUC, and ARPB have been found to be useful for assessing the strength of general factors (Bonifay et al., 2015; Reise et al., 2013a; Rodriguez et al., 2016a). These overall indices only make sense when there is a single global general factor, and BifactorIndicesCalculator only provides them when it can determine that the model being evaluated has a general factor.

As an item-level version of the general factor ECV index, Stucky and Edelen (2015) compute ECV for each item (I-ECV) given by

$$I - ECV(i) = \frac{\lambda_{i,G}^2}{\lambda_{i,G}^2 + \sum_j \lambda_{i,S_j}^2}, \tag{2.5}$$

where $\lambda_{i,G}$ is the factor loading of item *i* onto the bifactor general factor, j varies over all specific factors, and λ_{i,S_j}^2 is the loading of item i onto specific factor j. Values of I-ECV near one indicate an item that only reflects the general dimension (Stucky et al., 2013). As with model level indices, I-ECV is only computed by BifactorIndicesCalculator when a general factor is present.

In addition to overall model level and item level indices, several indices are also computed at the factor level. These include Omega and ECV indices which are very similar to the model level indices. OmegaS (ω_S; Reise et al., 2013b) is identical to Omega except that only the items loading on specific factor S are utilized:

$$\omega_S = \frac{\sum_i \left(\lambda_{i,G}\right)^2 + \sum_i \left(\lambda_{i,S}\right)^2}{\sum_i \left(\lambda_{i,G}\right)^2 + \sum_i \left(\lambda_{i,S}\right)^2 + \sum_i (1 - h_i^2)}, \tag{2.6}$$

where *i* varies only over the items loading on specific factor *S*. OmegaHS (ω_{HS}; Reise et al., 2013b) is similarly related to OmegaH, this time with the numerator utilizing specific factor loadings:

$$\omega_{HS} = \frac{\sum_i (\lambda_{i,S})^2}{\sum_i (\lambda_{i,G})^2 + \sum_i (\lambda_{i,S})^2 + \sum_i (1 - h_i^2)}, \tag{2.7}$$

where *i* varies only over the items loading on specific factor *S*. OmegaHS is not a reliability index, but rather an indicator of dimensional uniqueness for the specific factor (Reise et al., 2013b; Rodriguez et al., 2016a). Three different ECV indices can be computed for specific factors: ECV_{SG}, ECV_{GS}, and ECV_{SS}. By using specific factor loadings in place of general factor loadings in Equation 2.3, ECV_{SG} is obtained (Stucky & Edelen, 2015):

$$ECV_{SG} = \frac{\sum_i \lambda_{i,S}^2}{\sum_i \lambda_{i,G}^2 + \sum_j \sum_i \lambda_{i,S_j}^2}, \tag{2.8}$$

where i varies over all items, j varies over all specific factors, and S of the numerator is an individual specific factor. While Stucky and Edelen (2015) interpret this index as an indicator of specific factor uniqueness, a more natural approach would be to only consider items loading on the specific factor of interest. This approach yields ECV_{SS}, which has the same formula as ECV_{SG} (i.e., Equation 2.8) except that only the items loading on the specific factor are included in any of the sums.

The complement of ECV_{SS} is the within-domain ECV (ECV_{GS}; Stucky & Edelen, 2015):

$$ECV_{GS} = \frac{\sum_i \lambda_{i,G}^2}{\sum_i \lambda_{i,G}^2 + \sum_j \sum_i \lambda_{i,S_j}^2}, \tag{2.9}$$

where again i only varies over items loading on the specific factor of interest. Notationally, the subscripts of specific factor ECVs are two letters: the first letter indicates whether loadings from the general factor of specific factor appear in the numerator, and the second

letter indicates whether sums are to be taken over all items or just items loading on the specific factor.

Specific factor indices OmegaHS and ECV_{SS} can be interpreted as indicators of dimensional uniqueness for bifactor models. However, when models with multiple general factors, termed "two-tier" models (Cai, 2010), OmegaHS and ECV_{SS} can be interpreted as general factor ECV and OmegaH for those general factors.

2.3 Installation and Examples of Using BifactorIndicesCalculator

The R package BifactorIndicesCalculator (>1.0.0) contains functions for computing bifactor indices for a variety of model types as well as convenience functions for directly utilizing output from various R packages and Mplus (Muthén & Muthén, 2019). The package is publicly available from the Comprehensive R Archive Network (CRAN). Installation and loading of the package can be accomplished with

```
install.packages("BifactorIndicesCalculator")
library(BifactorIndicesCalculator)
```

For confirmatory models, bifactor indices can be computed using the following two functions:

```
bifactorIndices(Lambda, Theta = NULL, UniLambda = NULL,
                                                standardized =  TRUE)
bifactorIndicesMplus(Lambda = file.choose(), UniLambda = NULL,
                                                standardized = TRUE)
```

with the following arguments:

- Lambda – A matrix of factor loadings or an object that BifactorIndicesCalculator can convert to a matrix of factor loadings. Currently, models fit by the R package lavaan (Rosseel & Jorgensen, 2019) are supported in the "bifactorIndices" function

and Mplus .out files are supported in the "bifactorIndicesMplus" function. The default behavior of "bifactorIndicesMplus" is to spawn a file selection window which the user can use to select the desired .out file.

- Theta – an optional vector of indicator residual variances. When Lambda is a model fit by lavaan of Mplus, Theta is extracted from the fit model. Additionally, when standardized coefficients are used, Theta is computed using Lambda. Thus Theta input is only required when inputting a matrix of unstandardized factor loadings for Lambda.
- UniLambda – a matrix of factor loadings from a unidimensional model of the same data or an object that BifactorIndicesCalculator can convert to a matrix of factor loadings (i.e., a model fitted by lavaan or Mplus). UniLambda is used for computing parameter bias and is only needed when that index is desired.
- standardized – a Boolean indicator of whether bifactor indices are to be computed based on standardized coefficients as is standard practice (Rodriguez et al., 2016a). Bifactor indices can also be computed based on unstandardized coefficients (standardized = FALSE).

Additionally, since bifactor indices from exploratory models show promise for being interpretable in a similar way as those from confirmatory models (Murray et al., 2019), the following two functions are provided for computing bifactor indices based on exploratory models:

```
bifactorIndices_expl(Lambda, ItemsBySF = NULL, LoadMin = 0.2)
bifactorIndicesMplus_expl(Lambda = file.choose(), ItemsBySF = NULL,
                                                        LoadMin = 0.2)
```

with the following arguments:

- Lambda – A standardized factor loading matrix from an exploratory model or an object that BifactorIndicesCalculator can convert to a factor loading matrix. Currently, exploratory factor analysis models fit by the *psych* package (Revelle, 2017) are supported in "bifactorIndices_expl", and .out files for exploratory structural equation models fit by Mplus are supported by "bifactorIndices Mplus_expl." The default behavior of "bifactorIndicesMplus_expl" is to spawn a file selection window which the user can use to select the desired .out file.
- ItemsBySF – A list, indexed by specific factors, of items which are intended to load on that specific factor. This list is used for controlling which indicators are included in the sums for specific factor indices. The default input is for this list to be NULL and assign indicators to specific factors based on having large enough factor loadings.
- LoadMin – The factor loading threshold for which an indicator is be considered as loading substantially on a factor. Used for assigning items to specific factors and generate warnings concerning unexpected items loading on factors when ItemsBySF is provided.

2.3.1 Example 1. Confirmatory Bifactor Model

Data from the twenty SRS-22r (Asher et al., 2006) items concerning patient quality of life with scoliosis is provided in the BifactorIndicesCalculator as the built-in data set SRS_data. The SRS-22r has four subdomains, but a total score is often interpreted. A

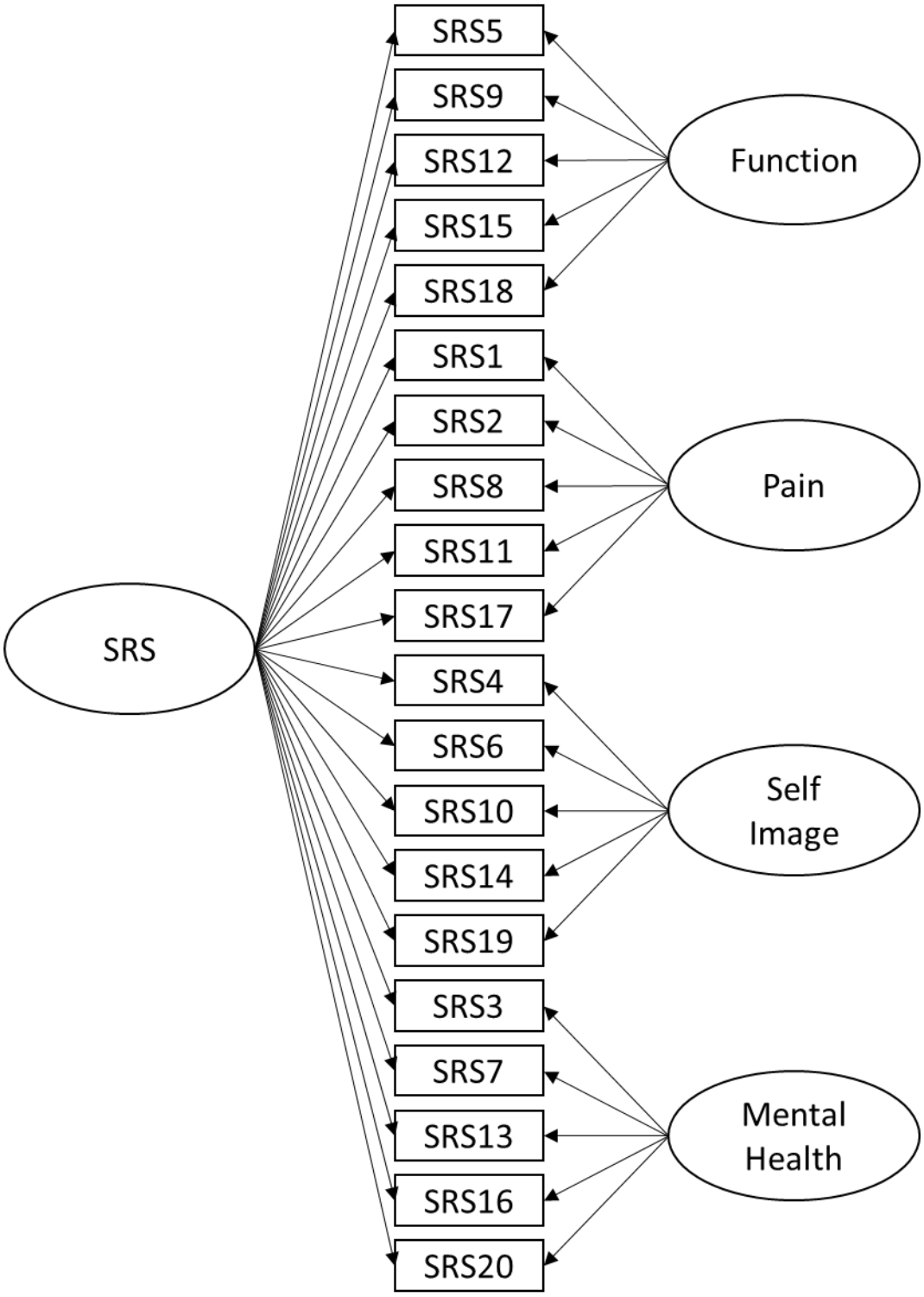

Figure 2.1 Bifactor Model of SRS-22r

diagram of a bifactor model for the SRS-22r in which the specific factors are aligned with the subdomains can be found in Figure 2.1. To assess the appropriateness of utilizing a total score or unidimensional measurement model, the following code fits an appropriate bifactor model in lavaan and computes the bifactor indices:

```
SRS_UnidimensionalModel <-
   "SRS =~ SRS_1  + SRS_2  + SRS_3  + SRS_4  + SRS_5  +
           SRS_6  + SRS_7  + SRS_8  + SRS_9  + SRS_10 +
           SRS_11 + SRS_12 + SRS_13 + SRS_14 + SRS_15 +
           SRS_16 + SRS_17 + SRS_18 + SRS_19 + SRS_20"

SRS_BifactorModel <-
   "SRS =~ SRS_1  + SRS_2  + SRS_3  + SRS_4  + SRS_5  +
           SRS_6  + SRS_7  + SRS_8  + SRS_9  + SRS_10 +
           SRS_11 + SRS_12 + SRS_13 + SRS_14 + SRS_15 +
           SRS_16 + SRS_17 + SRS_18 + SRS_19 + SRS_20
    Function     =~ SRS_5  + SRS_9  + SRS_12 + SRS_15 + SRS_18
    Pain         =~ SRS_1  + SRS_2  + SRS_8  + SRS_11 + SRS_17
    SelfImage    =~ SRS_4  + SRS_6  + SRS_10 + SRS_14 + SRS_19
    MentalHealth =~ SRS_3  + SRS_7  + SRS_13 + SRS_16 + SRS_20"

SRS_Unidimensional <- lavaan::cfa(SRS_UnidimensionalModel,
                                  SRS_data,
                                  ordered = paste0("SRS_", 1:20),
                                  orthogonal = TRUE)

SRS_bifactor <- lavaan::cfa(SRS_BifactorModel,
                            SRS_data,
                            ordered = paste0("SRS_", 1:20),
                            orthogonal = TRUE)

bifactorIndices(SRS_bifactor, UniLambda = SRS_Unidimensional)
```

The output of "bifactorIndices" is a list with three elements: factor level indices, item level indices, and model level indices. According to the guidelines established by Rodriguez et al. (2016a), a total score is interpretable, but use of a unidimensional model is questionable. Abridged output from "bifactorIndices" for this example is:

```
$FactorLevelIndices
                ECV_SS     ECV_SG    ECV_GS     Omega   Omega_H
SRS          0.6728130 0.67281303 0.6728130 0.9614271 0.8702229
Function     0.1972990 0.04153902 0.8027010 0.8342751 0.1011000
Pain         0.4123779 0.11147096 0.5876221 0.9116273 0.3616746
SelfImage    0.3280132 0.08183383 0.6719868 0.8846751 0.2445099
MentalHealth 0.3424358 0.09234316 0.6575642 0.9127146 0.3054461

$ItemLevelIndices
           IECV RelParBias
SRS_1  0.5104022 0.35337859
SRS_2  0.4976737 0.36753848
SRS_3  0.7980893 0.03658076
```

```
            . . .
SRS_19 0.4582466 0.16640639
SRS_20 0.6912559 0.10480146

$ModelLevelIndices
      ECV       PUC      ARPB
0.6728130 0.7894737 0.1209678
```

2.3.2 Example 2. Exploratory Bifactor Model

Using the same dataset, the use of "bifactorIndices_expl" is now illustrated. Of note that when using the psych package for exploratory factor analysis, the factors are automatically named. To ensure that the proper items sets were associated with the appropriate specific factors, the loadings from the exploratory factor analysis solution were reviewed (not shown).

```
Library(psych)
SRS_BEFA <- fa(SRS_data, nfactors = 5, rotate = "bifactor")

ItemsBySF = list(MR4 = paste0("SRS_", c(5, 9, 12, 15, 18)), #Function
                 MR2 = paste0("SRS_", c(1, 2,  8, 11, 17)), #Pain
                 MR3 = paste0("SRS_", c(4, 6, 10, 14, 19)), #SelfImage
                 MR5 = paste0("SRS_", c(3, 7, 13, 16, 20))) #Mental H…

bifactorIndices_expl(SRS_BEFA, ItemsBySF = ItemsBySF)
```

The output of "bifactorIndices_expl" is a list with two elements: factor level indices and model level indices. According to the guidelines established by Rodriguez et al. (2016a), a total score is interpretable, but use of a unidimensional model is questionable. Note that, had "ItemsBySF" not been specified, general factor results would have been unchanged but specific factor results would have been based on different items and not been the same. Abridged output from "bifactorIndices" for this example is:

```
$FactorLevelIndices
        ECV_SS     ECV_SG    ECV_GS     Omega    Omega_H
MR1  0.6528916 0.65289164 0.6528916 0.9402626   0.890523
MR2  0.3052481 0.12171192 0.5911061 0.8757252  0.2698778
MR3  0.4024035 0.10707890 0.5103314 0.8398127  0.3263923
MR4  0.2760142 0.06322365 0.6600196 0.7412911  0.1689451
MR5 0.02662722 0.05509389 0.8479560  0.873287 0.02185901

$ModelLevelIndices
       ECV_SS     Omega  Omega_H
MR1 0.6528916 0.9402626 0.890523
```

Additionally, “bifactorIndices_expl” issued 34 warnings, which suggests that the exploratory factor analysis did not recover the hypothesized structure very well. The first four of these warnings are:

```
Warning messages:
1: In bifactorIndices_expl(SRS_BEFA, ItemsBySF = ItemsBySF) :
  Item SRS_1 loads on factor MR1 above 0.2
2: In bifactorIndices_expl(SRS_BEFA, ItemsBySF = ItemsBySF) :
  Item SRS_2 loads on factor MR1 above 0.2
3: In bifactorIndices_expl(SRS_BEFA, ItemsBySF = ItemsBySF) :
  Item SRS_3 loads on factor MR1 above 0.2
4: In bifactorIndices_expl(SRS_BEFA, ItemsBySF = ItemsBySF) :
  Item SRS_3 loads on factor MR5 below 0.2
. . .
```

2.3.3 Example 3. Two-Tier model

Simulated multitrait-multimethod data was simulated for use as an example and is available in BifactorIndicesCalculator as “MTMM_data”. In this dataset, three traits are each represented by three items for each of three methods. In the model fit below, trait factors are allowed to covary but method factors are orthogonal to all other factors. The following code can be used to compute the bifactor indices for this data set:

```
MTMM_model <- "
Trait1 =~
       T1M1_1+T1M1_2+T1M1_3+T1M2_1+T1M2_2+T1M2_3+T1M3_1+T1M3_2+T1M1_3
Trait2 =~
       T2M1_1+T2M1_2+T2M1_3+T2M2_1+T2M2_2+T2M2_3+T2M3_1+T2M3_2+T2M1_3
Trait3 =~
       T3M1_1+T3M1_2+T3M1_3+T3M2_1+T3M2_2+T3M2_3+T3M3_1+T3M3_2+T3M1_3
Method1 =~ T1M1_1+T1M1_2+T1M1_3+T2M1_1+T2M1_2+T2M1_3+
           T3M1_1+T3M1_2+T3M1_3
Method2 =~ T1M2_1+T1M2_2+T1M2_3+T2M2_1+T2M2_2+T2M2_3+
           T3M2_1+T3M2_2+T3M2_3
Method3 =~ T1M3_1+T1M3_2+T1M3_3+T2M3_1+T2M3_2+T2M3_3+
           T3M3_1+T3M3_2+T3M3_3

Trait1 ~~ 0*Method1
Trait1 ~~ 0*Method2
Trait1 ~~ 0*Method3
Trait2 ~~ 0*Method1
Trait2 ~~ 0*Method2
Trait2 ~~ 0*Method3
Trait3 ~~ 0*Method1
Trait3 ~~ 0*Method2
Trait3 ~~ 0*Method3

Method1 ~~ 0*Method2
Method1 ~~ 0*Method3
Method2 ~~ 0*Method3"
```

```
MTMM_fit <- lavaan::cfa(MTMM_model, MTMM_data)
bifactorIndices(MTMM_fit)
```

With a two-tier model, only factor level indices are output. For the trait factors, these indices can be interpreted a though they were model level indices in a standard bifactor model.

```
$FactorLevelIndices
           ECV_SS     Omega   Omega_H
Trait1  0.7422100 0.9436118 0.8415608
Trait2  0.5967258 0.9215644 0.7490016
Trait3  0.6880337 0.9555450 0.8243821
Method1 0.3206363 0.9258991 0.5394873
Method2 0.3197150 0.9403045 0.5497289
Method3 0.4579639 0.8546333 0.6719250
```

2.3.4 Bifactor Indices Shiny App

For the convenience of Mplus users who may be unfamiliar with R, a Shiny-based application with graphical user interface was additionally developed and is freely available. The left panel of the interface contains fields for uploading Mplus .out files for confirmatory bifactor, unidimensional, or two-tier models. Results are calculated as soon as the confirmatory bifactor model is uploaded; the unidimensional model is only required for parameter bias indices. The right panel of the interface contains tabs holding the different categories of indices.

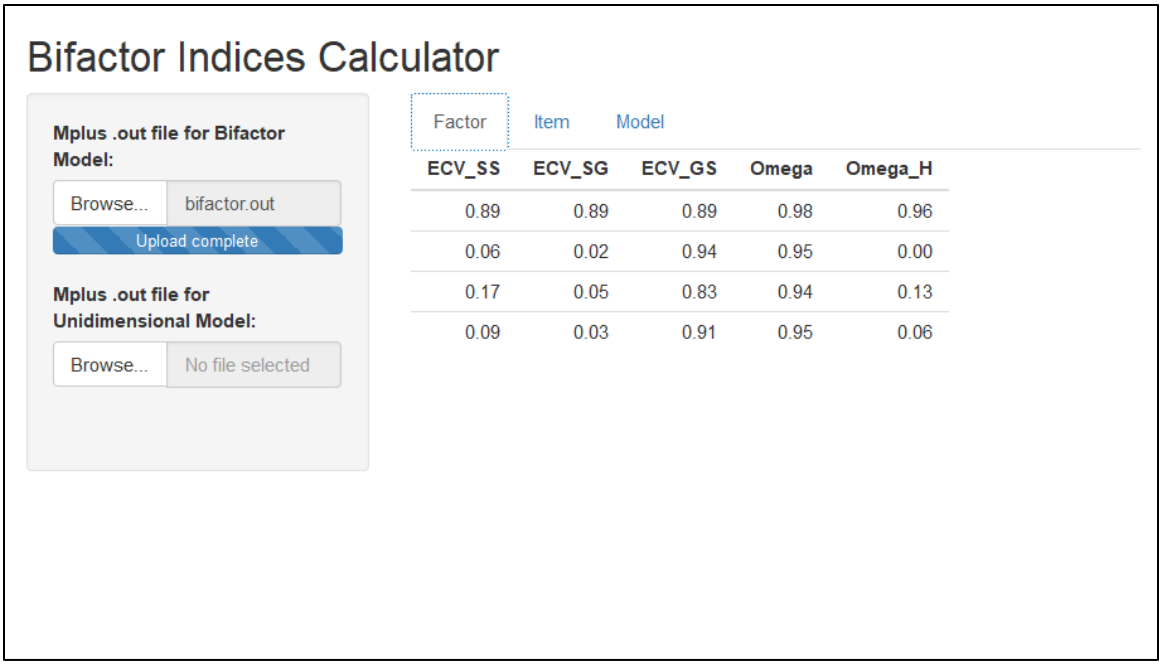

Figure 2.2 Bifactor Indices Calculator Shiny App

2.4 Discussion

This manuscript introduces an R package, BifactorIndicesCalculator, for computing auxiliary indices for both confirmatory and exploratory bifactor models. While these indices are not difficult to compute manually, the success of an earlier Excel-based version demonstrates demand for convenient calculators of these indices (Dueber, 2017). To that end, the package includes convenience functions for directly inputting output from various statistical programs that can estimate confirmatory and exploratory bifactor models. Additionally, a *Shiny*-based webapp has been provided for additional convenience to researchers unfamiliar with R.

In summary, we described computation of various indices to aid in the assessment of dimensionality, implemented these computations in an R package, and illustrated use of this package through several examples.

CHAPTER 3. STUDY TWO: BIFACTOR APPROACH TO SUBSCORE ANALYSIS

3.1 Introduction

Important social science constructs are often measured with multi-item instruments, and data collected from using these instruments often fail to satisfy the strict conditions of unidimensionality (Reise et al., 2013a). Instead, these data exhibit a multidimensional structure in which clusters of similar items measuring a facet or subdomain of the construct of interest comprise the dimensions. Using bifactor models and associated indices, methodologists have developed a framework for deciding when data can be interpreted unidimensionally, with a total score or unidimensional measurement model, despite the presence of some multidimensionality in item responses (Bonifay et al., 2015; Reise et al., 2013b; Rodriguez et al., 2016a).

However, it is also sometimes the case that researcher desire to interpret scores of the subdomains, termed subscores. This is particularly true in education testing contexts in which subscores provide diagnostic information about specific areas of strength and weakness (Monaghan, 2006; Wedman & Lyrén, 2015). In social science research, interpretation of subscores is particularly useful when different subdomains correlate differently with an external variable (Chen et al., 2012). In this case, use of only a total score can result in inappropriately nonspecific theories and recommendations (Hull et al., 1991). For example, Follberg et al. (2019) found that the dominance subdomain of a measure of agentic and communal goal orientations was correlated with career interest while other subdomains were not, whereas previous research had found no correlation between agentic and communal goal orientations total score and career interest (Diekman et al., 2010). Generally, when there is theoretical and psychometric evidence for the

appropriateness of their interpretation, the use of subscores may uncover relevant associations, group differences, or other results that would be masked or tempered when only a total score is interpreted.

3.1.1 Scoliosis Quality of Life Index

As an example, consider the Scoliosis Quality of Life Index (SQLI), which was designed to measure how an adolescent patient's idiopathic scoliosis affects their quality of life (Feise et al., 2005). The SQLI is comprised of 20 items, belonging to four subdomains each with five items: self-esteem, back pain, physical activity, and moods and feelings. Both the total SQLI score and subdomain scores have typically been interpreted (Feise et al., 2005; Rowe et al., 2006); however, no psychometric evidence concerning the dimensionality of data from the SQLI has been provided except for reliabilities, which are not useful indicators of dimensionality (McNeish, 2018). Doctors use SQLI total scores and subdomain scores to help provide a more holistic approach to the treatment of scoliosis. For example, a patient whose moods and feelings score drops severely may be referred to counseling, a patient whose back pain scores drop severely (indicating more back pain) may be prescribed medication, or a patient whose physical activity score drops severely following being fitted with a back brace may have their treatment plan reconsidered. Accordingly, interpretation of subdomain scores is clinically relevant.

Given this relevance, it is important to provide evidence that subdomain scores are of sufficient quality to be interpreted. After all, if data from the SQLI are truly unidimensional, then interpreting subdomain scores is always inappropriate (Bollen & Lennox, 1991; Sinharay et al., 2011). A common phenomenon for patients who are fitted with a brace is to show little to no change in overall SQLI score after receiving the brace,

but report substantial decrease in moods and feelings (compensated for by moderate improvement in back pain and physical activity). If SQLI subdomain scores do not possess interpretive value, then only the overall lack of change in SQLI total scores should be interpreted. In this case, the decrease in moods and feelings would be considered as measurement error and not clinically relevant; these patients' changes in subdomain scores would be the result of measurement error and interpreted as Type I error. However, if subdomain scores are interpretable, then these patients' decrease in moods and feelings is clinically relevant, and their doctors should feel confident in taking appropriate action.

3.1.2 Bifactor Models and Indices

When data are unidimensional, they may be modeled using a unidimensional confirmatory factor analysis model in which a single latent factor explains all covariances between items. When data are multidimensional, it is common to use a separate latent factor for each dimension. Items belonging to each dimension load on the corresponding factor, and factors are allowed to correlate; accordingly, these models are referred to as correlated traits models. However, when the dimensions of data are closely related, such as when comprising subdomains of a global construct, models can be used which reflect both a general factor and factors corresponding to each dimension. The most common of these models is a second-order factor model, in which a single latent factor explains covariances amongst the factors of a correlated traits model.

Another model which can be used for this purpose is the bifactor CFA model, which consists of a single general factor onto which all items load and orthogonal specific factors corresponding to the subdomains. Bifactor models and correlated traits models are statistically equivalent only under certain proportionality constraints; these constraints are

equivalent to the data satisfying a second-order CFA model structure (Yung et al., 1999). Even when multidimensional data does not satisfy this constraint, bifactor models tend to exhibit good fit to the data anyway (Cucina & Byle, 2017; Morgan et al., 2015). Bifactor CFA models are commonly fit to data for dimensionality assessment purposes even when a bifactor interpretation of the data is not intended (Rodriguez et al., 2016a); instead either a unidimensional (single total score) or correlated traits (subscores) interpretation will be used.

In a bifactor CFA model, items crossload onto both the general factor and a specific factor (although in some bifactor models not all items will crossload onto a specific factor); therefore, the variance of each item is split into three components: covariance with the general factor, covariance with the specific factor, and item specific variance, as depicted in Figure 3.1. Various statistical indices describing the partitioning of variance in a bifactor model have been devised and can be used for evaluating the appropriateness of making uni- or multi-dimensional interpretations of the data.

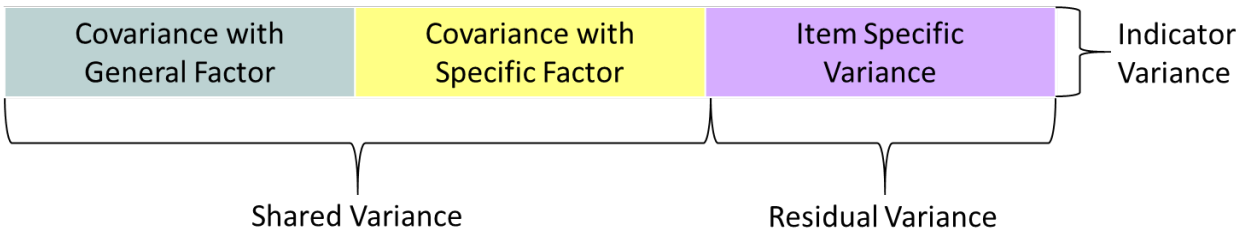

Figure 3.1 Partitioning of Item Variance in a Bifactor Model

Research concerning bifactor indices has primarily focused on indices for the general factor. These include omega, hierarchical omega, explained common variance, and the percent of uncontaminated correlations. Coefficient Omega (ω; McDonald, 1999) is a model-based estimate of composite reliability of total score. While Omega is typically used

with unidimensional models, Zinbarg et al. (2005, p. 126, Equation 8) imply the following formula for Omega based on bifactor model parameters:

$$\omega = \frac{\sum_i \left(\lambda_{i,G}\right)^2 + \sum_j \sum_i \left(\lambda_{i,S_j}\right)^2}{\sum_i \left(\lambda_{i,G}\right)^2 + \sum_j \sum_i \left(\lambda_{i,S_j}\right)^2 + \sum_i (1 - h_i^2)}, \tag{3.1}$$

where i varies over all items, j varies over all specific factors, $\lambda_{i,G}$ is the loading of item i onto the general factor, λ_{i,S_j} is the loading of item i onto specific factor j, and h_i^2 is the communality of item i. Since Omega is computed using the estimated parameters (namely factor loadings and residual variances) of a model, it estimates the proportion of variance in the total score explained by common variance (i.e., using all common factors) implied by the model (Bentler, 2009; Raykov, 1997; Revelle & Zinbarg, 2009).

Whereas Omega estimates the proportion of total score variance that can be explained within the bifactor model, hierarchical omega (OmegaH; ω_H; McDonald, 1999; Zinbarg et al., 2005) estimates the proportion of total score variance that can be explained by the general factor and is computed similarly as in Equation 1.6 except that only loadings from the general factor are considered in the numerator (Zinbarg et al., 2005, p. 126, Equation 8):

$$\omega_H = \frac{\sum_i \left(\lambda_{i,G}\right)^2}{\sum_i \left(\lambda_{i,G}\right)^2 + \sum_j \sum_i \left(\lambda_{i,S_j}\right)^2 + \sum_i (1 - h_i^2)}. \tag{3.2}$$

While OmegaH does not directly address the issue of unidimensionality (Reise et al., 2007), Gustafsson and Aberg-Bengtsson (2010) and McDonald (1999) argue that high OmegaH indicates that total scores primarily reflect a single dimension, since it may be interpreted as the squared correlation between observed total scale score and the latent

general factor. Rodriguez et al. (2016a) suggest that total scores can still be interpreted in the presence of some multidimensionality so long as OmegaH is high and give a cutoff of 0.8.

Explained common variance (ECV; Sijtsma, 2009; ten Berge & Socan, 2004) is the proportion of common variance across all items which is explained by the general factor,

$$ECV = \frac{\sum_i \lambda_{i,G}^2}{\sum_i \lambda_{i,G}^2 + \sum_j \sum_i \lambda_{i,S_j}^2}, \tag{3.3}$$

where i varies over all items, j varies over all specific factors, $\lambda_{i,G}$ is the loading of item i onto the general factor, and λ_{i,S_j} is the loading of item i onto specific factor j. Reise et al. (2013a) consider ECV to be a measure of the unidimensionality of the data from a modeling perspective. Reise et al. (2013a) and Bonifay et al. (2015) both found that ECV predicted the bias in model parameters when a unidimensional model is fit to multidimensional data. Finally, the percent of uncontaminated correlations (PUC) is the proportion of item covariances which are modeled only by the general factor (Bonifay et al., 2015, p. 4). As CFA is concerned with modeling covariances between items, a high PUC means that much of the information in the data is only relevant to the general factor of a bifactor model; the specific factors model only a small number of covariances. Reise et al. (2013a) and Bonifay et al. (2015) both found that as PUC increases, the role of ECV in predicting the bias model parameters when a unidimensional model is fit to multidimensional data diminishes. That is, when ECV is high or PUC is high and ECV is moderate, expected bias is low.

In the present study, however, where subdomain scores and therefore specific factors are of primary interest, bifactor indices relevant to specific factors are more relevant. An estimate of composite reliability of subscores (OmegaS; ω_S) can be computed

for each specific factor. The OmegaS index has a formula (Reise et al., 2013a, p. 134, Equation 5) similar to the formula for Omega, except that only items from a particular specific factor are included:

$$\omega_S = \frac{\sum_i (\lambda_{i,G})^2 + \sum_i (\lambda_{i,S})^2}{\sum_i (\lambda_{i,G})^2 + \sum_i (\lambda_{i,S})^2 + \sum_i (1 - h_i^2)}, \tag{3.4}$$

where i varies only over the items loading on specific factor S. While OmegaS does not generally represent a squared correlation between observed and true scores, in the case that specific factor loadings are proportional to general factor loadings, the items on the specific factor can be modeled using a unidimensional model (Yung et al., 1999) and OmegaS can be interpreted as a squared correlation between observed and true scores.

Hierarchical omega for a subdomain (OmegaHS; ω_{HS}) estimates the proportion of subscore variance that can be explained by the corresponding specific factor and is computed (Reise et al., 2013a, p. 134, Equation 6) as

$$\omega_{HS} = \frac{\sum_i (\lambda_{i,S})^2}{\sum_i (\lambda_{i,G})^2 + \sum_i (\lambda_{i,S})^2 + \sum_i (1 - h_i^2)}, \tag{3.5}$$

where i varies only over the items on the subscale S. As specific factors are interpreted as residuals after the general factor is accounted for, OmegaHS is not a reliability index. Instead, OmegaHS is sometimes interpreted as a measure of dimensional uniqueness for the subdomain (Gignac & Kretschmar, 2017; Reise et al., 2013b; Rodriguez et al., 2016a). While not providing any strict cutoffs, Reise et al. (2013b) and Gignac and Kretschmar (2017) both suggest using OmegaHS to aid in decision-making about interpreting subscores and suggest OmegaHS = .50 as a reasonable minimum for interpreting a subscore. A search of the PsychINFO database for peer-reviewed articles published in 2018

using the search term "bifactor" revealed 195 articles, of which 58 used OmegaHS to aid in decision-making about dimensionality. For the most part, authors of these studies conformed to the suggestion of not interpreting subscores when OmegaHS < .50, but a small number of authors recommended interpretation of subscores for smaller OmegaHS (e.g., Hukkelberg & Ogden, 2018; Stanton et al., 2018; Thompson et al., 2018).

Finally, while several ECV indices can be computed for specific factors, the most relevant to the current study is the explained common variance of the specific factor with respect to the items loading on that specific factor (ECV_{SS}, Dueber, 2017, 2019). The formula for ECV_{SS} is

$$ECV_{SS} = \frac{\sum_i \lambda_{i,S}^2}{\sum_i \lambda_{i,G}^2 + \sum_i \lambda_{i,S}^2}, \tag{3.6}$$

where sums are taken only over items loading on the specific factor of interest. ECV_{SS} is the complement of what Stucky and Edelen (2015, p. 201) refer to as "within-domain ECV," and can be considered as an indicator of dimensional uniqueness. In this way, ECV_{SS} and OmegaHS perform similar purposes, but from different perspectives: ECV_{SS} refers to item variance explained by a latent specific factor, while OmegaHS refers to subscore variance explained by a latent specific factor.

3.1.3 A Classical Test Theoretic Approach to Subscore Analysis

The central idea behind Haberman's (2005, 2008) subscore assessment technique is that if observed subscores (s) are to be useful, they must be able to predict true subscores (s_t) better than the observed total score (x) does. If a scale is truly unidimensional, then the total score will be a better predictor of s_t than the subscore because the total score is more reliable. On the other hand, if a scale is truly multidimensional, then the observed subscore will be a better predictor of s_t than the total score because the correlation between s_t and s

will be higher than the correlation between s_t and x. Therefore, if s is a better predictor of s_t than x, then the subscore will have added value, in the sense that interpreting s gives additional useful information about the subdomain above and beyond x.

The quality of prediction is measured using the proportional reduction in mean square error (PRMSE), which is equivalent to the coefficient of determination (R^2; Smith, 1977). When the observed subscore is used to predict the true subscore, the coefficient of determination PRMSE(s) is the squared correlation between the observed and true subscore, which is the reliability and can be estimated by an appropriate reliability coefficient such as Cronbach's alpha. When the observed total score is used to predict the true subscore, the coefficient of determination PRMSE(x) is the squared correlation between the observed total scores and true subscores. Computing PRMSE(x) can be accomplished by exploiting the bilinearity property of correlations as described in Reise et al. (2013); an implementation of this technique can be found in the 'subscore' R package (Dai et al., 2019).

In order to understand the behavior of PRMSE(s) and PRMSE(x), it is useful to write them as in Equations 3.7 and 3.8, which decompose the squared correlation into a squared correlation between true scores and a reliability (Sinharay et al., 2007; Sinharay et al., 2011):

$$\text{PRMSE}(s) = r^2(s_t, s) = r^2(s_t, s_t)\rho^2(s) \; = \; \rho^2(s) \tag{3.7}$$

and

$$\text{PRMSE}(x) = r^2(s_t, x) = r^2(s_t, x_t)\rho^2(x), \tag{3.8}$$

where r^2 is the squared correlation and ρ^2 is reliability. From equations 3.7 and 3.8 it is clear that PRMSE(s) is less than unity because of unreliability in s. On the other hand,

PRMSE(x) is less than one both because of the presence of multidimensionality (which lowers the correlation between s_t and x_t) and because of measurement error in x. However, since x typically has a higher reliability than s (Ling, 2012, p. 2), it is still possible for PRMSE(x) to be greater than PRMSE(s). The literature on PRMSE and VAR are silent as to the most appropriate ways to estimate reliability, but Chronbach's alpha is most commonly used (Dai et al., 2019; Sinharay, 2019).

Standard guidelines (Wedman & Lyren, 2015; Sinharay et al., 2011) are to only report subscores when PRMSE(s) is greater than PRMSE(x). However, Feinberg and Jurich (2017) advise performing a significance test using bootstrapping to see if PRMSE(s) is statistically significantly larger than PRMSE(x). Sinharay (2019) demonstrates a variety of ways to perform this significance test. Additionally, Feinberg and Jurich recommend only reporting subscores when PRMSE(s) is at least 10% greater than PRMSE(x).

Feinberg and Wainer (2014) introduced the value-added ratio (VAR) of a subscore, defined as the ratio of PRMSE(s) to PRMSE(x). If this ratio is greater than one, then PRMSE(s) exceeds PRMSE(x) and the subscore will have added value over the total score, meaning that interpretation of the subscore provides meaningful information above and beyond interpreting the total score. On the other hand, if VAR is less than one, then total scores provide a more accurate estimate of true subscores than the observed subscores do; thus, interpretation of the subscore does not contribute useful information. While Feinberg and Jurich (2017) indicate that there is no harm in interpreting a subscore so long as VAR > 1.0, their recommendation corresponds to reporting subscores when VAR > 1.1 to assure that the subscore explains a meaningful amount of true subscore variance above and beyond the total score.

3.1.4 Purpose

Applied researchers employ bifactor CFA models and indices based on model parameter estimates for dimensionality assessment purposes, including determining whether or not to interpret subscores. However, to date no rigorous guidelines for performing this determination have been developed. The present research endeavors to partially close this gap between methodology literature and research practice by aligning bifactor indices with PRMSE indices which are commonly used to assess the value added by interpretation of subscores in educational testing contexts (Sinharay, 2019). By providing this link between methods commonly used in testing contexts (PRMSE) and methods commonly used in psychological sciences research (bifactor models), it is hoped that the present research will provide results that aid psychology, education and, more generally, social science researchers in making rigorous decisions about whether to interpret subscores for use in research and in practical settings.

The purpose of this study is to use simulation techniques to devise a strategy for which bifactor indices can be used to determine whether a multidimensional interpretation is appropriate for a given data set. Specifically, cutoffs will be devised for a specific factor's bifactor indices, possibly conditioned upon general bifactor indices, such that exceeding these cutoffs indicates the subscore has added value over the total score. Use of these cutoffs will then be illustrated using data collected from the SQLI.

3.2 Method

A simulation study will be conducted using the R statistical computing environment (R Core Team, 2019) where bifactor indices will be related to PRMSE based indices with possible moderation of that relationship by general factor bifactor indices. The goal is to

use bifactor indices to develop cutoffs for determining when it is appropriate to interpret a subscore separately from the total score. A variety of factors will be manipulated to help probe those relationships and guide selection of cutoffs.

For each experimental condition, data will be generated from a second-order factor model, consistent with prior simulation work using bifactor models (Bonifay et al., 2015; Reise et al., 2013). As discussed in the introduction, second-order factor models are statistically equivalent to both correlated-trait models and bifactor models; as such, inferences about both the general factor and about the multidimensional structure of the scale can be made from a single model. Both second- and first-order factor scores as well as individual indicator (item) scores will be generated and recorded. The simulated factor scores will be treated as true scores in order to exactly compute PRMSE(s), PRMSE(x), and thereby VAR for the simulated data. A bifactor CFA model with specific factors corresponding to the first-order factors will be fit, and relevant bifactor indices computed for that model will then be compared to VAR.

Unlike typical simulation studies (Feinberg & Rubright, 2016), data will not be generated by repeatedly sampling from fixed population parameters for each condition. Rather, a single large sample (N = 100,000) will be generated for each condition so that population parameters can be estimated with great precision. This strategy is consistent with prior simulation studies involving bifactor indices (Bonifay et al., 2015; Reise et al., 2013) and with the finding of Ferrando and Lorenzo-Seva (2019) that PRMSE based indices are largely unaffected by sample size. In practice, sampling error can affect estimation of bifactor indices as well as PRMSE indices; thus, decisions based on sample statistics and sample indices will always have some degree of uncertainty to them. This

phenomenon is an unavoidable aspect of working with samples and in no way diminishes the value of understanding how population parameters function. In an effort to capture as much of the diverse range of relevant models as possible, simulation conditions were determined by sampling 1,000 possible population models from each of a number of model structures which each define a distribution of possible models. Sufficiency of this sample size (i.e., 1,000 per model structure) was verified by replicating the study with a different initial seed obtained from random.org for the random number generator.

Design of simulation conditions will be formulated to represent as much of the range of PRMSE and bifactor indices found in practice as possible. PRMSE(s) is equivalent to subscore reliability (Haberman, 2005) and should therefore be accurately estimated by OmegaS (Rodriguez et al., 2016b). On the other hand, PRMSE(x) is related both to total score theoretical reliability and to the correlation between subscore and total score (Reise et al., 2013). Noting that only reliabilities and scale-level correlations are of interest, the number of items per subdomain is not relevant. Instead, five items per subdomain will be used and first-order factor loadings will be chosen to match an OmegaS reliability index specified for that condition. Five items per subdomain is common in applied literature; in Rodriguez et al.'s (2016a) review of applied bifactor literature, five was the most common number of items in a specific factor. The correlation between subscore and total score is not directly expressed in a second-order factor model; however, second-order factor loadings represent correlations between latent subdomains (first-order factors) and the general second-order factor. Finally, the total score reliability is influenced by subscore reliability, correlations among subdomains, and the number of subdomains. Thus, the

number of factors, reliability of subscores, and second-order factor loadings will be manipulated in this simulation study.

3.2.1 Manipulated Variables

3.2.1.1 Number of Subdomains

Rodriguez et al. (2016a) surveyed 50 studies from the psychopathology, personality, and assessment literatures and found that scales on which bifactor models had been employed involved between two and seven specific factors. In order to capture the variability in number of dimensions found in research practice, the present study uses seven different conditions for the number of first-order factors: from two to eight.

3.2.1.1 Reliability of Subdomains

The standard recommendation is to not interpret or use (sub-)scores with a reliability lower than 0.7 or 0.8 depending on the purpose (Nunnally & Bernstein, 1994). Nevertheless, it is possible for low reliability and high reliability (defined below) subdomains to coexist in such a way that one or more high reliability subscore is interpretable separately from the total score, even though the low reliability subscores are not interpretable. As such, in the present study, subdomain reliabilities will be allowed to vary between .50 and .99. The choice of .99 for highest reliability was chosen as a practical maximum. The lowest reliability of .50 was chosen to correspond to the Omega reliability estimate of three items with standardized factor loadings of .5. In educational testing settings, subscores frequently have very low reliability, but subscore reliability tends to be higher in psychological and educational research settings in which measurement instruments use items with polytomous response options (Sinharay et al., 2011). The following three conditions will be used for subdomain reliabilities:

- All high reliabilities, sampled uniformly from the interval [.70, .99]
- Mixed reliabilities, sampled uniformly from the interval [.50, .99]
- Half low reliabilities (sampled uniformly from the interval [.50, .70]) and half high reliabilities (sampled uniformly from the interval [.70, .99])

Note, an "all low reliabilities" condition is not included since, in that case, subscores should not be interpreted even if they are sufficiently different from the total score.

3.2.1.2 Second-Order Factor Loadings

Prior simulation and theoretical literature concerning the assessment of uni- or multi-dimensionality using bifactor indices (Bonifay et al., 2015; Quinn, 2014; Reise et al., 2013) or PRMSE indices (Ferrando & Lorenzo-Seva, 2019; Quinn, 2014) have typically involved subdomains with identical theoretical reliabilities and identical correlations between subdomains in each condition. In this context, either all of the subscores can be interpreted separately from the total score or none of them can. In applied contexts, however, it is certainly possible for only a subset of the subscores to have added value (Reise et al., 2013a; Sinharay, 2011). Accordingly, in the present study, variability among correlations of subdomains as represented by first-order factors will be induced by selecting random second-order factor loadings from a specified distribution. Second-order factor loadings will range from low (0.50, corresponding to first-order factors correlating at 0.25) to high (0.99, which functions as a practical maximum). The magnitude of the lowest factor loading was chosen to correspond to the conditions in Bonifay et al. (2015) and Reise et al. (2013a) with the lowest ECV. The magnitude of the highest factor loading was set to .99 as that represents a practical maximum and corresponds to an ECV higher than any found in Bonifay et al. (2015) or Reise et al. (2013a). The cut between high and

low loadings was selected to be .80 based on a pilot study suggesting that this cutpoint would maximize variability in VAR of subscores across replications and conditions. The following four conditions will be set for second-order factor loadings:

- All high loadings, sampled uniformly from the interval [.80, .99]
- All low loadings, sampled uniformly from the interval [.50, .80]
- Mixed loadings, sampled uniformly from the interval [.50, .99]
- Half high and half low loadings

Noting that with only two first-order factors, the second-order factor model would be under-identified (Kline, 2016, p. 319). Thus, there are many different second-order models with two first-order factors that are statistically equivalent to each other. Namely, so long as the product of second-order factor loadings is the same, the models will be statistically equivalent; accordingly, in the case of only two first-order factors, the second-order loadings serve only to model the correlation between these two first order factors. Therefore, for the conditions with only two factors, both factors are assigned the same second-order factor loading.

3.2.1.3 Summary

The present study will employ 84 model structures determined by fully crossing 7 number of dimension conditions, 3 subdomain reliability conditions, and 4 second-order factor loading conditions. From each of these structures, 1,000 random population models will be drawn, resulting in a total of 84,000 population conditions being used. These conditions average 5 subdomains per condition, so a total of 420,000 subdomains will be evaluated.

3.2.2 Data Generation

Data generation will proceed in two steps: the second- and first-order factor scores will be computed first, and the item scores will be computed separately. The model used for generating data will be completely standardized; second-order factor scores, first-order factor scores, and item scores are all normally distributed with a mean of zero and a variance of one. Second-order factor scores will be generated by randomly sampling 100,000 numbers from the standard normal distribution. For each first-order factor, a second-order factor loading will be chosen from the distribution specified in the simulation condition and first-order factor scores will be computed from second-order scores with all factor means set to zero. For each first-order factor, a reliability will be chosen from the distribution specified in the simulation condition. This reliability was converted to an item factor loading by solving Equation 3.1 (in this unidimensional model, there are no specific factors) for the loading, assuming five items with equal loadings,

$$\lambda = \sqrt{\frac{\omega}{5 - 4 \times \omega}}, \tag{3.9}$$

where λ is the item's loading onto the first-order factor and ω is the desired reliability. Item scores are then computed from first-order factor scores with all item means set to zero. Thus, the elements of the second-order factor model are produced by creating two separate correlated factors models: one representing the second-order structure and one representing the first-order structure.

The population second order-factor model can be converted to a bifactor model using the Schmid-Leiman (1957) transformation. The Schmid-Leiman transformation

works by applying a rotation to the first-order factor loadings which adds a general factor and ensures that all factors are now orthogonal. The formula for this transformation is

$$\boldsymbol{\Lambda}_{\text{bifactor}} = \boldsymbol{\Lambda}_{\text{first-order}}\left(\boldsymbol{\Lambda}_{\text{second-order}}\middle|\sqrt{\boldsymbol{\Theta}_{\text{first-order}}}\right) \tag{3.10}$$

where $\boldsymbol{\Lambda}_{\text{bifactor}}$ is the factor loading matrix for the bifactor model, $\boldsymbol{\Lambda}_{\text{first-order}}$ is the factor loading matrix for the indicators onto the first order factors, $\boldsymbol{\Lambda}_{\text{second-order}}$ is the factor loading matrix for the first-order factors onto the second-order factors, $\sqrt{\boldsymbol{\Theta}_{\text{first-order}}}$ is the square matrix with the square root of first-order factor uniquenesses on the diagonal and zeros elsewhere, and (|) denotes a supermatrix. Use of the Schmid-Leiman transformation permits computation of the population bifactor model. However, it is unclear whether bifactor indices computed from this population model are more appropriate to use than ones computed from a bifactor model fit to the sample data given that sampling error affects (sub-)score properties and bifactor indices alike. After all, the sample will contain some idiosyncrasy which may be captured in a bifactor model estimated from the sample data.

A small pilot simulation study was conducted with 50 replications per condition in which bifactor indices were compared between the population bifactor model and the bifactor model estimated from generated sample data. When the number of first-order factors was greater than 2, the bias between population model bifactor indices and bifactor indices estimated from the sample was very small. Specifically, when the number of subdomains was greater than 2, average absolute bias, defined as the mean (across conditions, replications, and indices) of the absolute difference between the index computed from the population model and the index estimated from the sample, was less

than 0.002 and only 0.6% of indices exhibited an absolute bias greater than 0.01. While there is no body of literature examining the distribution and standard errors of bifactor indices to draw from, this level of error matches what the author expects given past personal experience with bifactor indices and the sample size of 100,000. Across all conditions, bias was most prevalent in factors with low reliability (< .70) and thus of minimal concern. After all, low reliability subscores ought not to be interpreted regardless of VAR. When the number of first-order factors was 2, however, bias was much greater. Specifically, average absolute bias was 0.099 and 78.9% of indices exhibited an absolute bias greater than 0.01. In all cases of large bias (i.e., absolute bias > .01) in OmegaS, OmegaS from the population model closely matched the squared correlation between observed subscores and first-order factor scores, while OmegaS from the estimated model over-estimated this value. Thus, the Schmid-Leiman transformation will be used to compute population bifactor models, as indices computed from these population models are either nearly equivalent (when number of dimensions is greater than 2) to indices estimated from the sample or more accurately measure what they are intended to measure (when number of dimensions is 2). These results suggest possible problems with using bifactor indices from models with two specific factors, as they may not accurately measure what they are intended to measure. Of note, a literature search has revealed no research concerning the accuracy and precision of estimated bifactor indices.

Following computation of a bifactor model, the BifactorIndicesCalculator package (Dueber, 2019) for the R statistical computing environment will be used to compute bifactor indices for a given model. Specifically, the ECV, Omega, OmegaH, and all OmegaS, OmegaHS, and ECV_{SS} indices will be computed. Furthermore, the PRMSE

indices PRMSE(s), PRMSE(x), and VAR will be computed by correlating observed subscores and total scores with first-order factor scores, which function here as true scores. For each replication, the simulation condition information will be stored, as well as the following information for each first-order factor: second-order factor loading, reliability, first-order factor loading, bifactor indices, and PRMSE indices.

3.2.3 Analyzing Results of Simulation

The process of analyzing the data collected to decide upon a decision rule for interpreting a subscore was largely exploratory with a goal of creating a model of VAR using bifactor indices as predictors. Then, using that model, bifactor index cutoffs were created for both VAR > 1 and VAR > 1.1, as those are the common PRMSE based cutoffs (Feinberg & Jurich, 2017; Sinharay et al, 2011; Wedman & Lyren, 2015). Independent variables considered as predictors of VAR included OmegaS, OmegaHS, ECV_{SS}, ECV, Omega, and OmegaH. The factors comprising the simulation conditions were also considered as covariates, especially the number of factors, as well as their interactions with other predictors. Of note, each first-order factor will be separately assessed for having added value, so the total number of subdomains to be assessed is higher than the total number of replications.

Determination of cutoffs or other decision rules for when a subscore has added value will proceed in a similar manner as the analyses performed by Reise et al. (2013b). Modeling of VAR will be conducted using multiple regression in an exploratory fashion by sequentially adding bifactor indices to the model as long as their inclusion substantially improves predictive accuracy ($\Delta R^2 > .02$). In this fashion, it will be determined which bifactor indices are influential and whether significant interactions exist. Once relevant

predictors and interactions are identified, levels of those variables will be determined to create a decision rule or set of decision rules with accuracy as high as possible. In the case of Reise et al. (2013b), they were able to give a cutoff for ECV when PUC was not large (< .80) and assert that when PUC was large (> .80), their criterion would be met regardless of ECV. As several authors (Gignac & Kretzschmar, 2017; Gignac & Watkins, 2013; Reise, 2012; Reise et al., 2013a; Rodriguez et al., 2016b) have suggested OmegaHS is an indicator of dimensional uniqueness, it is hoped that a decision rule based upon OmegaHS can be found.

Adequacy of the resulting decision rules were evaluated by computing their sensitivity and specificity. While no fixed rules exist for what levels of sensitivity and specificity are adequate, very high levels (> 90%) are desired here as it is expected that practitioners may decide whether or not to interpret subscores based on these decision rules. Furthermore, every effort was made to generate the simplest set of decision rules while maintaining a high level of accuracy.

A set of replication samples was simulated using the same conditions as the original set of samples. The replicated samples were used to check for consistency of results with the original simulated samples. Specifically, the replication samples were used to check for consistency of parameters in the linear models and also for consistency of sensitivity and specificity of the decision rules.

3.3 Results

The two bifactor indices which best predicted VAR individually were OmegaHS and ECV_{SS}, which explained 84.8% and 73.6% of variance in VAR, respectively. The estimated regression equation using OmegaHS as a predictor of VAR is

$$VAR = 0.679 + 2.039 \times \omega_{SH} + e \quad (3.11)$$

where *e* is an error term with a variance of .022. Using Equation 3.11, VAR has an expected value of 1.0 for OmegaHS = .157, and VAR has an expected value of 1.1 for OmegaHS = .206. Using these values of OmegaHS as cutoffs results in acceptable levels of sensitivity (.909 for VAR = 1.0 and .919 for VAR = 1.1) but unacceptable levels of specificity (.724 for VAR = 1.0 and .840 for VAR = 1.1). The estimated regression equation using ECV_{SS} as a predictor of VAR is

$$VAR = 0.674 + 1.595 \times ECV_{SS} + e \quad (3.12)$$

where *e* is an error term with a variance of .038. Using Equation 3.12, VAR has an expected value of 1.0 for ECV_{SS} = .205, and VAR has an expected value of 1.1 for ECV_{SS} = .267. Using these values of ECV_{SS} as cutoffs results in nearly acceptable levels of sensitivity (.886 for VAR = 1.0 and .884 for VAR = 1.1) but unacceptable levels of specificity (.570 for VAR = 1.0 and .708 for VAR = 1.1).

Due to the unacceptably low levels of specificity for the above cutoffs, additional predictors will be included to better predict VAR. The remainder of the results are presented in two sections: one which builds a model for VAR starting with OmegaHS and the other starting with ECV_{SS}.

3.3.1 Building a Model for VAR Starting with OmegaHS

Inclusion of other bifactor indices in the regression model naturally increased predictive accuracy; the index which most increased variance explained was OmegaS, when its interaction with OmegaHS was also included. This model explained 87.8% of variability in VAR. Examination of residuals from this model revealed heteroscedasticity

and nonlinearity in the relationship between OmegaHS and VAR after accounting for OmegaS, as seen in Figure 3.2.

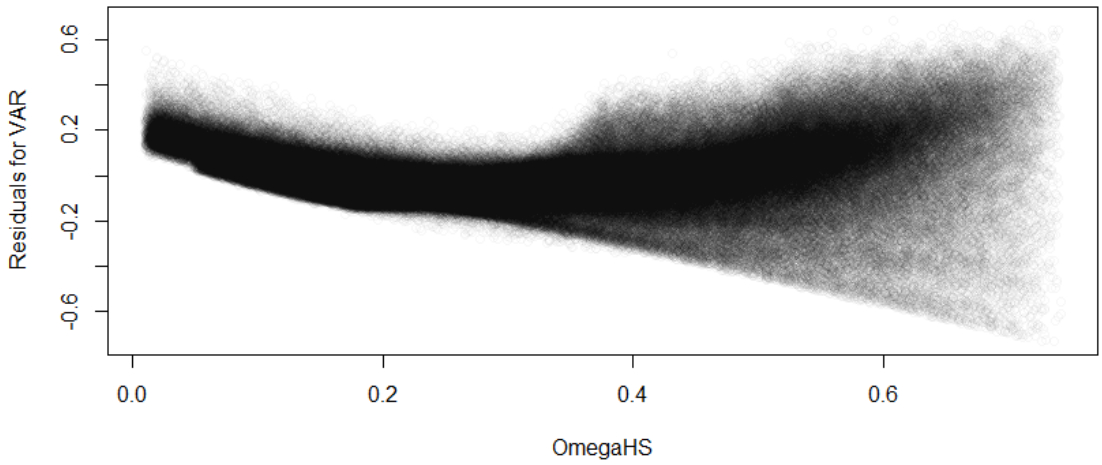

Figure 3.2 Residuals for VAR Regressed on OmegaS and OmegaHS

To account for nonlinearity in the relationship between OmegaHS and VAR, a quadratic term for OmegaHS (i.e., OmegaHS^2) was included in the model for VAR; this resulted in improved prediction of VAR (R^2 = .909) and substantially decreases the apparent nonlinearity (Figure 3.3). Note that there is still substantial heteroscedasticity, as the variance of residuals notably increases as OmegaHS increases.

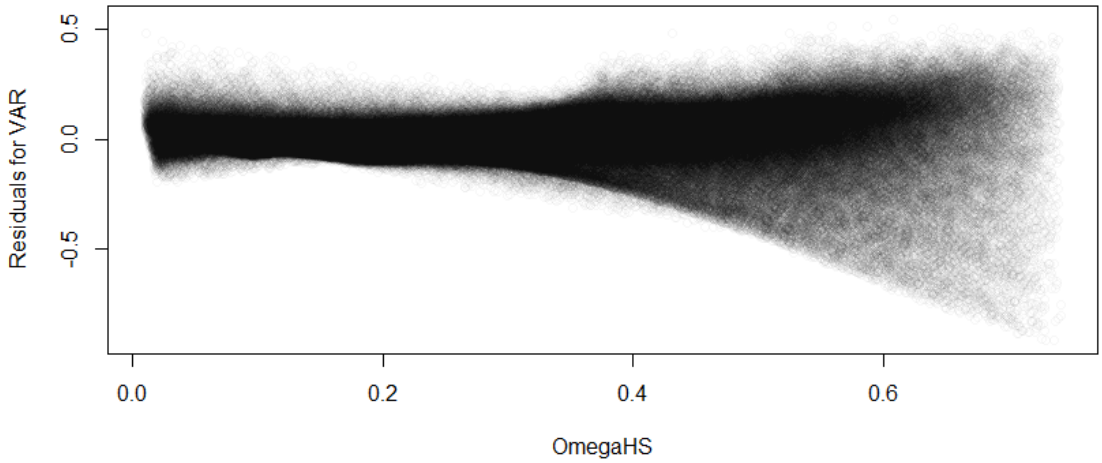

Figure 3.3 Residuals for VAR After Including a Quadratic Term for OmegaHS

The estimated regression equation using OmegaHS and OmegaS as well as their interaction and a quadratic term for OmegaHS as predictors of VAR is

$$VAR = -0.260 + 4.053 \times \omega_{HS} + 1.513 \times \omega_S - 4.858 \times \omega_S \times \omega_{HS} + 2.798\omega_{HS}^2 + e, \quad (3.13)$$

where e is an error term with a variance of .013. Using Equation 3.13, VAR has different OmegaHS cutoffs for different levels of OmegaS, as listed in Table 3.1. For example, for a specific factor with OmegaS = .75, OmegaHS of at least .151 is necessary for expected VAR to exceed 1.0 and OmegaHS of at least .220 is necessary for expected VAR to exceed 1.1.

Table 3.1 OmegaHS Cutoffs for Varying Levels of OmegaS

OmegaS	VAR = 1.0	VAR = 1.1
.50	.224	.258
.55	.216	.253
.60	.206	.257
.65	.193	.241
.70	.176	.232
.75	.151	.220
.80	.108	.204
.85	.000	.178
.90	.000	.111
.95	.000	.000

Using these values of OmegaHS as cutoffs results in excellent levels of sensitivity (.984 for VAR = 1.0 and .979 for VAR = 1.1) but unacceptable levels of specificity (.852 for VAR = 1.0 and .877 for VAR = 1.1).

To further improve predictive accuracy, all the bifactor indices and number of factors were again checked for their incremental predictive value. The number of factors was found to have the most value, when included with all interactions (including the three-way interaction between OmegaHS, OmegaS, and number of factors). Because of the strength of interaction terms involving number of factors, separate regression models were fit for each number of factors condition using OmegaS, OmegaHS, their interaction, and $OmegaH^2$ as predictors of VAR. Variance in VAR explained by these models varied by

number of factors, ranging from R^2 = .879 for three factors to R^2 = .984 for eight factors. Cutoffs for OmegaHS at different levels of Omega for VAR = 1.0 are found in Table 3.2, and cutoffs for VAR = 1.1 are found in Table 3.3.

Table 3.2 OmegaHS Cutoffs for Varying Levels of OmegaS and Number of Factors for VAR = 1.0

	Number of Factors						
OmegaS	2	3	4	5	6	7	8
.50	.216	*.214*	.218	.222	.224	.227	.228
.55	.210	*.203*	.208	.214	.216	.219	.221
.60	.204	*.189*	.196	.203	.206	.210	.212
.65	.195	***.172***	*.181*	.189	.194	.198	.200
.70	.183	***.148***	*.161*	.171	.177	.182	.185
.75	*.166*	***.113***	***.131***	.145	.152	.158	*.163*
.80	***.140***	***.048***	***.075***	*.097*	.108	*.118*	*.124*
.85	***.090***	.000	.000	.000	.000	.000	.000
.90	.000	.000	.000	.000	.000	.000	.000
.95	.000	.000	.000	.000	.000	.000	.000

Note. *Italicized* numbers differ from the cutoffs in Table 3.1 by more than .010. ***Bold italicized*** numbers differ from the cutoffs in Table 3.1 by more than .020.

Table 3.3 OmegaHS Cutoffs for Varying Levels of OmegaS and Number of Factors for VAR = 1.1

	Number of Factors						
OmegaS	2	3	4	5	6	7	8
.50	.260	.260	.256	.257	.257	.258	.258
.55	.260	.254	.251	.252	.252	.253	.253
.60	*.259*	.247	.244	.245	.246	.247	.248
.65	*.258*	.238	.236	.238	.239	.241	.242
.70	***.256***	.227	.226	.228	.230	.232	.234
.75	***.255***	.211	.212	.216	.218	.221	.223
.80	***.252***	*.188*	*.192*	.197	.202	.205	.208
.85	***.248***	***.148***	*.158*	*.167*	.174	.180	.185
.90	***.241***	***.000***	***.000***	***.078***	*.096*	.115	*.128*
.95	***.223***	.000	.000	.000	.000	.000	.000

Note. *Italicized* numbers differ from the cutoffs in Table 3.1 by more than .010. ***Bold italicized*** numbers differ from the cutoffs in Table 3.1 by more than .020.

Graphs displaying the relationship between OmegaHS and VAR for different levels of OmegaS and different numbers of factors can be found in Figures 3.4 through 3.9. Graphs are only shown for 2, 3, and 6 factors for the sake of brevity. Also featured on the graphs are a quadratic curve of best fit and a demarcation of the cutoff for a specified level of VAR.

Sensitivity and specificity from using the cutoffs listed in Table 3.2 and Table 3.3 can be found in Table 3.4. Sensitivity is excellent across all factors, but specificity is not acceptable, particularly for the 3, 4, and 5 factor conditions.

Table 3.4 Sensitivity and Specificity for Cutoffs in Table 3.2 and 3.3

	VAR = 1.0		VAR = 1.1	
Number of Factors	Sensitivity	Specificity	Sensitivity	Specificity
2	.977	.870	.971	.953
3	.970	.724	.952	.789
4	.978	.797	.970	.840
5	.984	.840	.978	.866
6	.989	.865	.984	.890
7	.990	.873	.987	.897
8	.992	.888	.988	.910

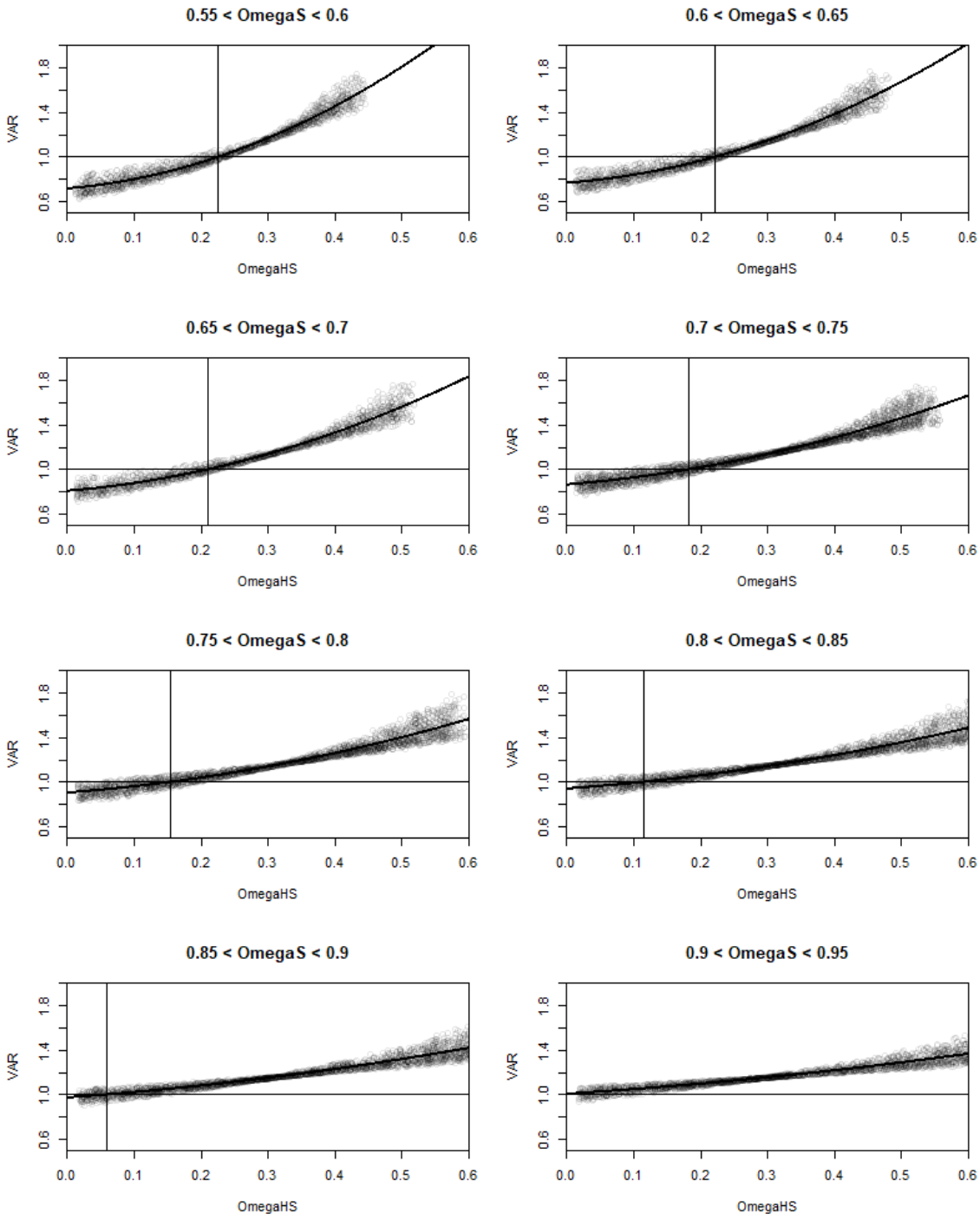

Figure 3.4 VAR versus OmegaHS for Two Factors, Showing Cutoffs for VAR = 1.0

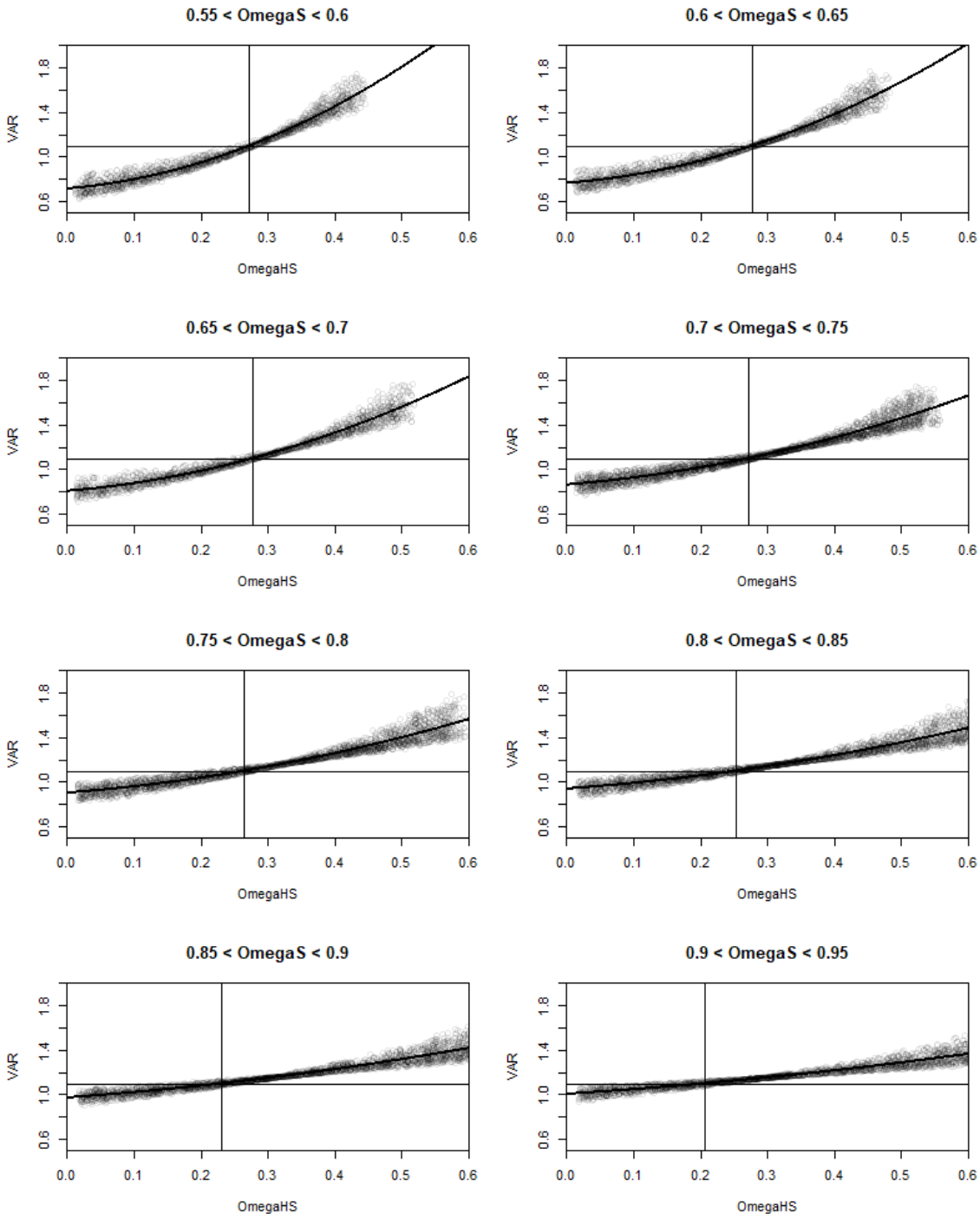

Figure 3.5 VAR versus OmegaHS for Two Factors, Showing Cutoffs for VAR = 1.1

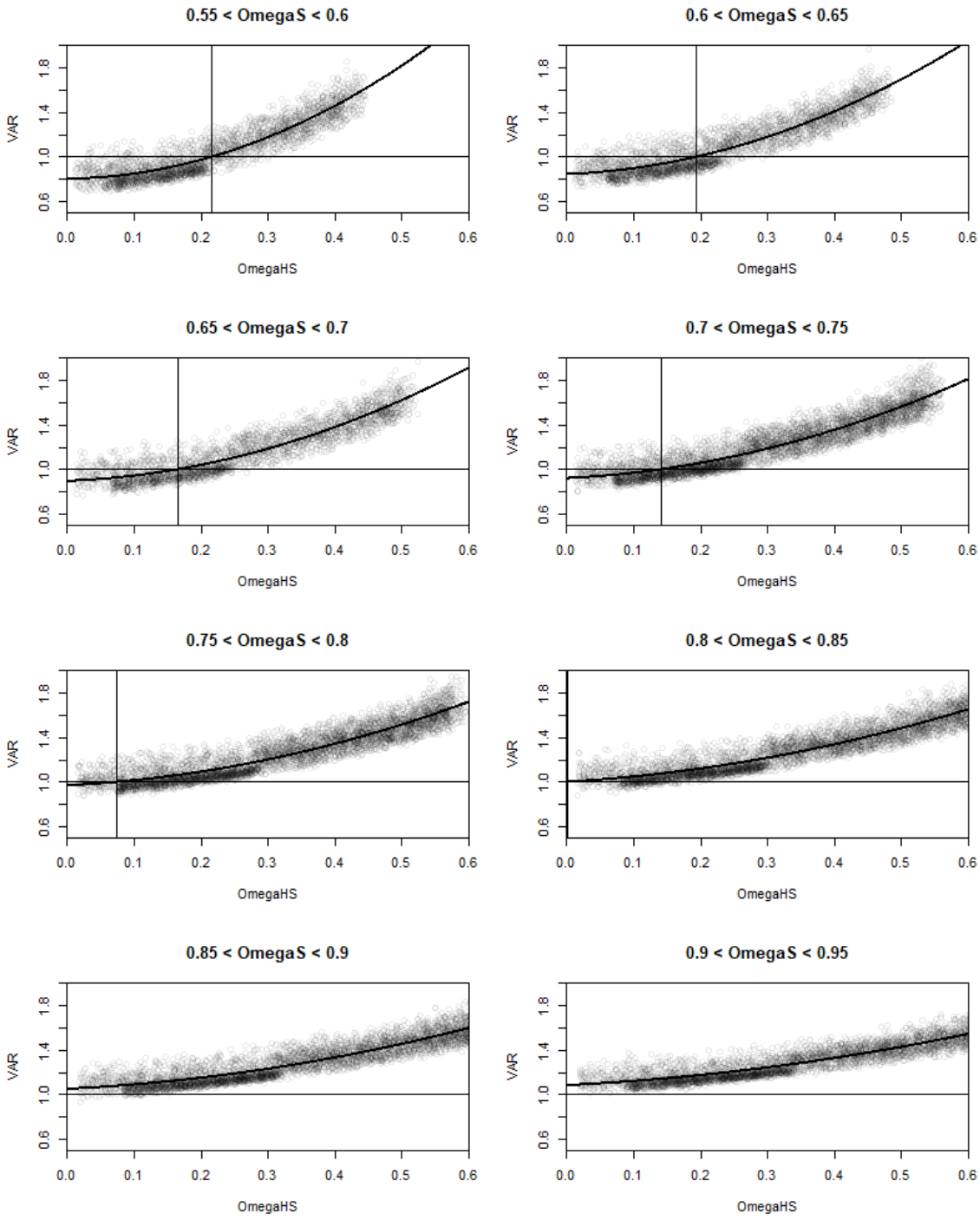

Figure 3.6 VAR versus OmegaHS for Three Factors, Showing Cutoff for VAR = 1.0

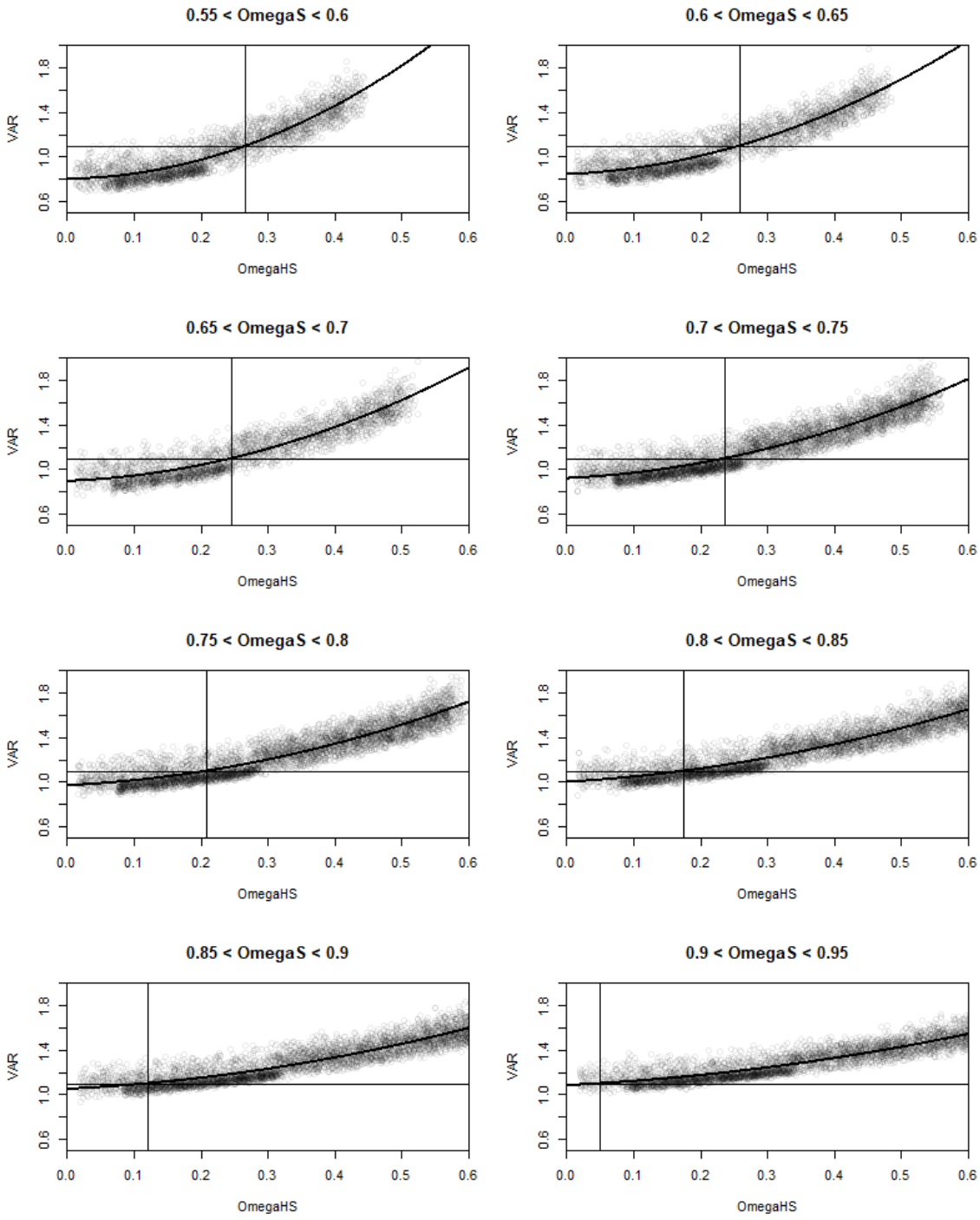

Figure 3.7 VAR versus OmegaHS for Three Factors, Showing Cutoffs for VAR = 1.1

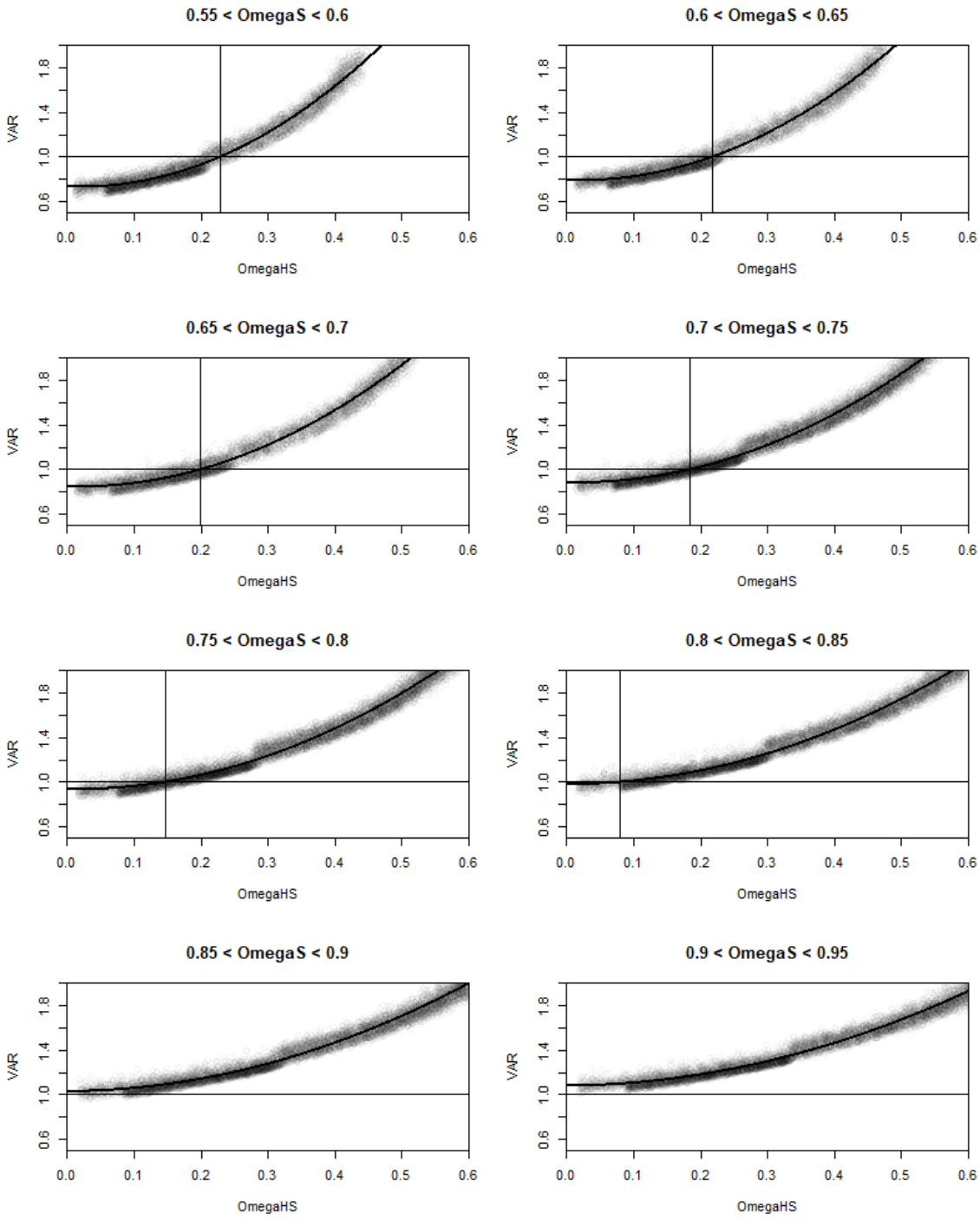

Figure 3.8 VAR versus OmegaHS for Six Factors, Showing Cutoffs for VAR = 1.0

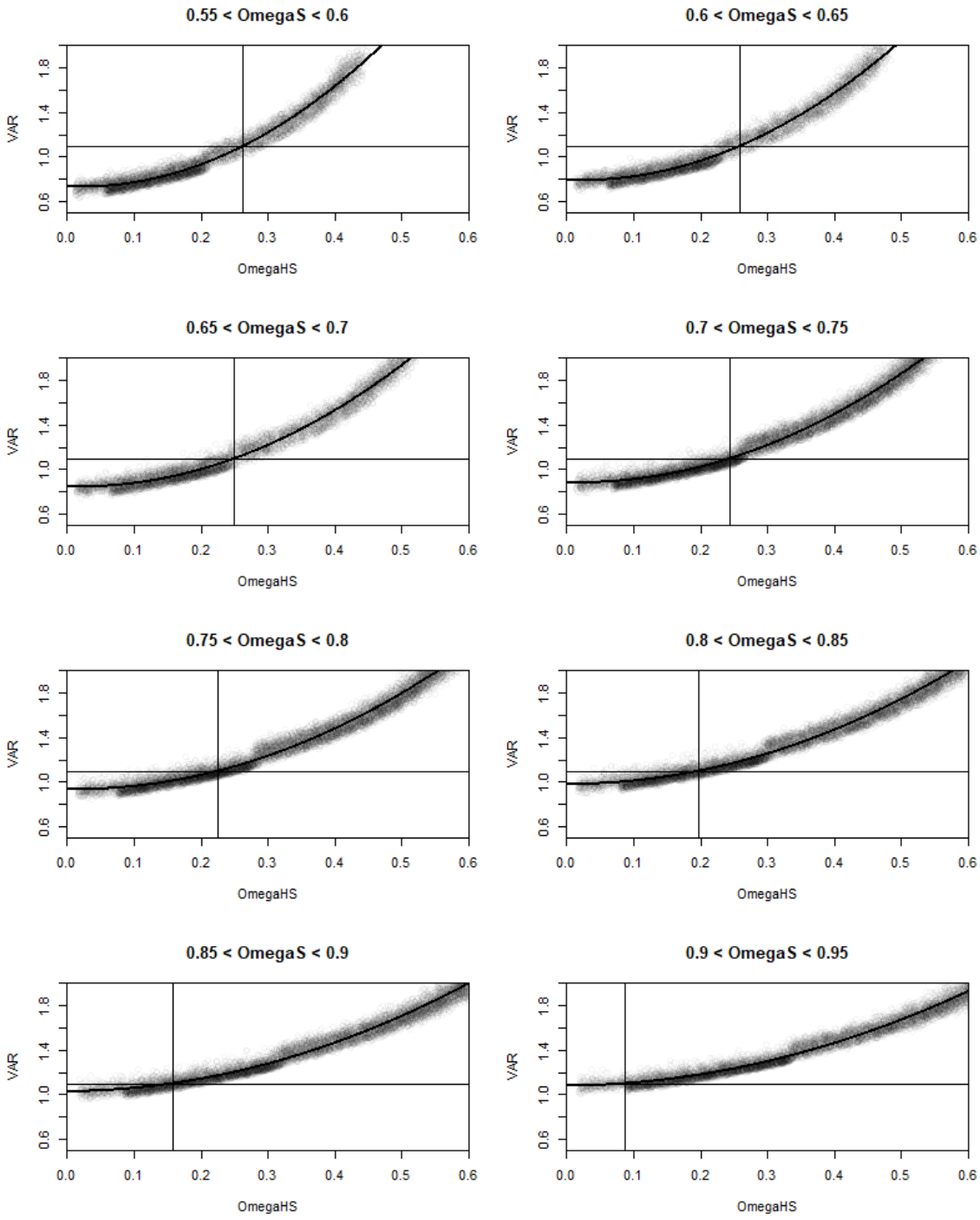

Figure 3.9 VAR versus OmegaHS for Six Factors, Showing Cutoffs for VAR = 1.1

3.3.2 Building a Model for VAR Starting with ECV_{SS}

Inclusion of other bifactor indices in addition to ECV_{SS} as predictors in the regression model naturally increased predictive accuracy; the index which most increased variance explained was OmegaS. The regression model with ECV_{SS} and OmegaS as predictors explained 87.5% of variability in VAR. Including the interaction between OmegaS and ECV_{SS} as a predictor did not increase explained variance, so no interaction term was included. Examination of residuals from this model revealed heteroscedasticity and nonlinearity in the relationship between ECV_{SS} and VAR after accounting for OmegaS, as seen in Figure 3.10.

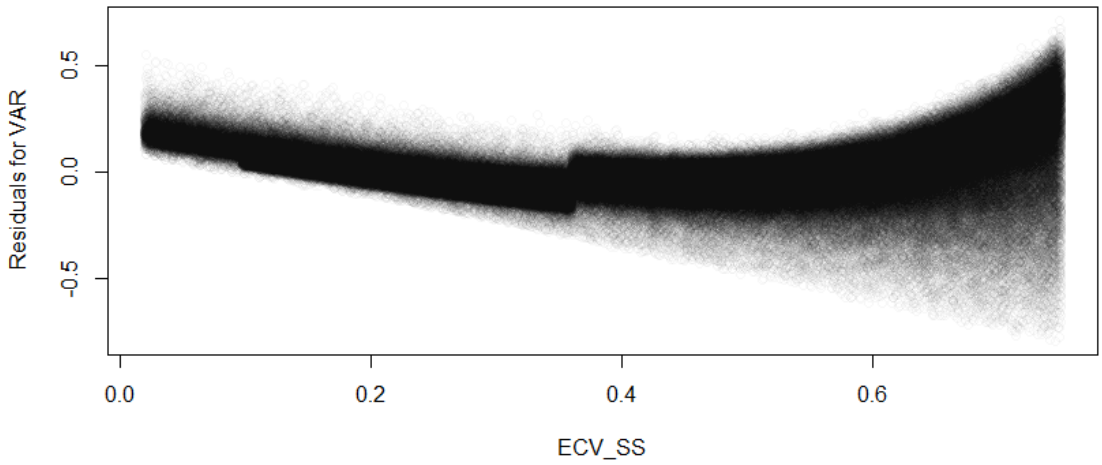

Figure 3.10 Residuals for VAR regressed on OmegaS and ECV

To account for nonlinearity in the relationship between ECV_{SS} and residuals for VAR, a quadratic term for ECV_{SS} (i.e., ECV_{SS}^2) was included in the model for VAR; this resulted in improved prediction of VAR (R^2 = .911) and substantially decreases the apparent nonlinearity (Figure 3.11). Note that there is still substantial heteroscedasticity,

as the variance of residuals notably increases as ECV_{SS} increases.

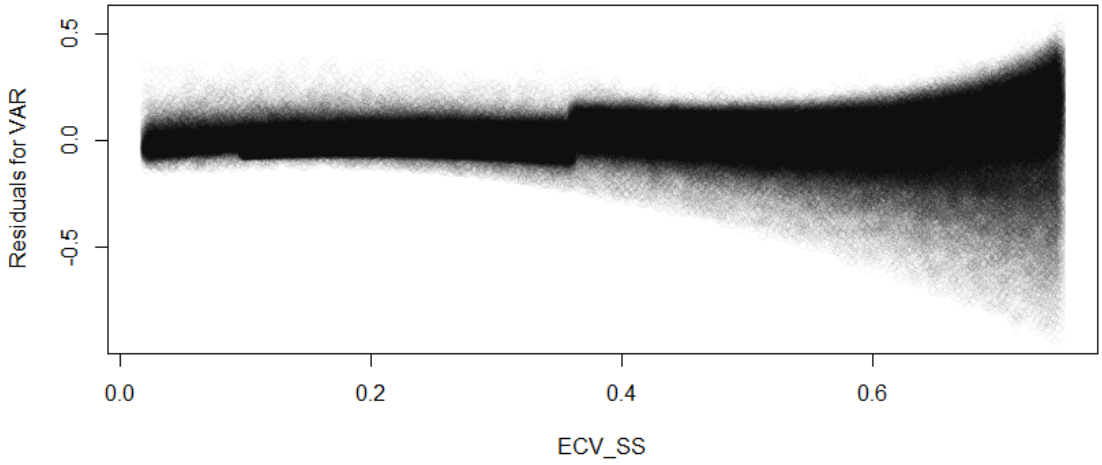

Figure 3.11 Residuals for VAR after including a quadratic term for ECV_{SS}

The estimated regression equation using ECV_{SS} and Omega as well as a quadratic term for ECV_{SS} as predictors of VAR is

$$VAR = 0.120 + 1.055 \times \omega_S - 0.460 \times ECV_{SS} + 1.961 \times ECV_{SS}^2 + e \qquad (3.14)$$

where e is an error term with a variance of .013. Using Equation 3.14, VAR has different OmegaS cutoffs for different levels of ECV_{SS}, as listed in Table 3.5. For example, for a specific factor with ECV_{SS} = .20, an OmegaS of at least .768 is required for expected VAR to exceed 1.0 and an OmegaS of at least .863 is required for expected VAR to exceed 1.1. Using these values of OmegaS as cutoffs results in excellent levels of sensitivity (.976 for VAR = 1.0 and .968 for VAR = 1.1) and adequate levels of specificity (.910 for VAR = 1.0 and .923 for VAR = 1.1).

Table 3.5 OmegaS Cutoffs for varying levels of ECV_{SS}

ECV_{SS}	VAR = 1.0	VAR = 1.1
.05	.832	.927
.10	.820	.915
.15	.799	.894
.20	.768	.863
.25	.729	.824
.30	.680	.775
.35	.622	.716
.40	.554	.649
.45	.477	.572
.50	.391	.486

To further improve predictive accuracy, all the bifactor indices and number of factors were checked for their incremental predictive value. The number of factors was found to have the most value, when included with all interactions. Because of the strength of interaction terms involving number of factors, separate regression models were fit for each number of factors condition using ECV_{SS}, OmegaS, and $OmegaS^2$ as predictors of VAR. Variance in VAR explained by these models varied by number of factors, ranging from R^2 = .867 for three factors to R^2 = .985 for eight factors. Cutoffs for OmegaS at different levels of ECV_{SS} and number of factors for VAR = 1.0 are found in Table 3.6, and cuttoffs for VAR = 1.1 are found in Table 3.7.

Table 3.6 OmegaS Cutoffs for Varying Levels of ECV_{SS} and Number of Factors for VAR = 1.0

	Number of Factors						
ECV_{SS}	2	3	4	5	6	7	8
.05	***.998***	*.813*	.824	.830	.829	.828	.826
.10	***.947***	***.788***	*.806*	.815	.818	.820	.821
.15	***.888***	***.755***	***.779***	.792	.798	.803	.806
.20	***.820***	***.716***	***.744***	.760	.769	.776	*.780*
.25	*.744*	***.670***	***.700***	.719	.730	.738	*.744*
.30	*.660*	***.618***	***.649***	*.669*	.682	*.691*	*.698*
.35	***.567***	***.559***	***.590***	*.611*	.624	*.634*	*.641*
.40	***.467***	***.493***	***.523***	*.544*	.557	*.567*	***.574***
.45	***.358***	***.421***	***.448***	.468	.480	*.490*	*.497*
.50	***.241***	***.341***	***.364***	.383	.394	*.403*	*.409*

Note. *Italicized* numbers differ from the cutoffs in Table 3.1 by more than .010. ***Bold italicized*** numbers differ from the cutoffs in Table 3.1 by more than .020.

Graphs displaying the relationship between OmegaS and VAR for different levels of ECV_{SS} and different numbers of factors can be found in Figures 3.12 through 3.17. Graphs are only shown for 2, 3, and 6 factors for the sake of brevity. Also featured on the graphs are a line of best fit and a demarcation of the cutoff from Table 3.6 and Table 3.7 for a specified level of VAR.

Table 3.7 OmegaS Cutoffs for Varying Levels of ECV_{SS} and Number of Factors for VAR = 1.1

	Number of Factors						
ECV_{SS}	2	3	4	5	6	7	8
.05	***N/A***	***.958***	*.939*	.928	.918	*.911*	***.905***
.10	***N/A***	*.932*	.920	.913	.908	*.903*	*.900*
.15	***N/A***	.899	.893	.890	.888	.886	.885
.20	***N/A***	.860	.858	.858	.858	.859	.859
.25	***.985***	.815	.815	.817	.820	.821	.823
.30	***.900***	*.762*	*.764*	.768	.771	.774	.777
.35	***.808***	*.703*	*.705*	.709	.714	.717	.720
.40	***.707***	*.637*	*.638*	.642	.646	.650	.653
.45	***.598***	.565	.562	.566	.570	.573	.576
.50	.481	.486	.479	.481	.483	.486	.488

Note. *Italicized* numbers differ from the cutoffs in Table 3.1 by more than .010. ***Bold italicized*** numbers differ from the cutoffs in Table 3.1 by more than .020. N/A indicates that a reliability greater than one is required to achieve VAR = 1.1.

Sensitivity and specificity from using the cutoffs listed in Table 3.6 and Table 3.7 can be found in Table 3.8. Sensitivity is adequate across all numbers of factors, but specificity is not acceptable for 3 and 4 factors.

Table 3.8 Sensitivity and Specificity for Cutoffs in Tables 3.6 and 3.7

	VAR = 1.0		VAR = 1.1	
Number of Factors	Sensitivity	Specificity	Sensitivity	Specificity
2	.927	.917	.918	.987
3	.969	.731	.929	.867
4	.973	.848	.952	.905
5	.977	.894	.965	.919
6	.980	.921	.974	.932
7	.981	.931	.978	.928
8	.982	.941.	.980	.932

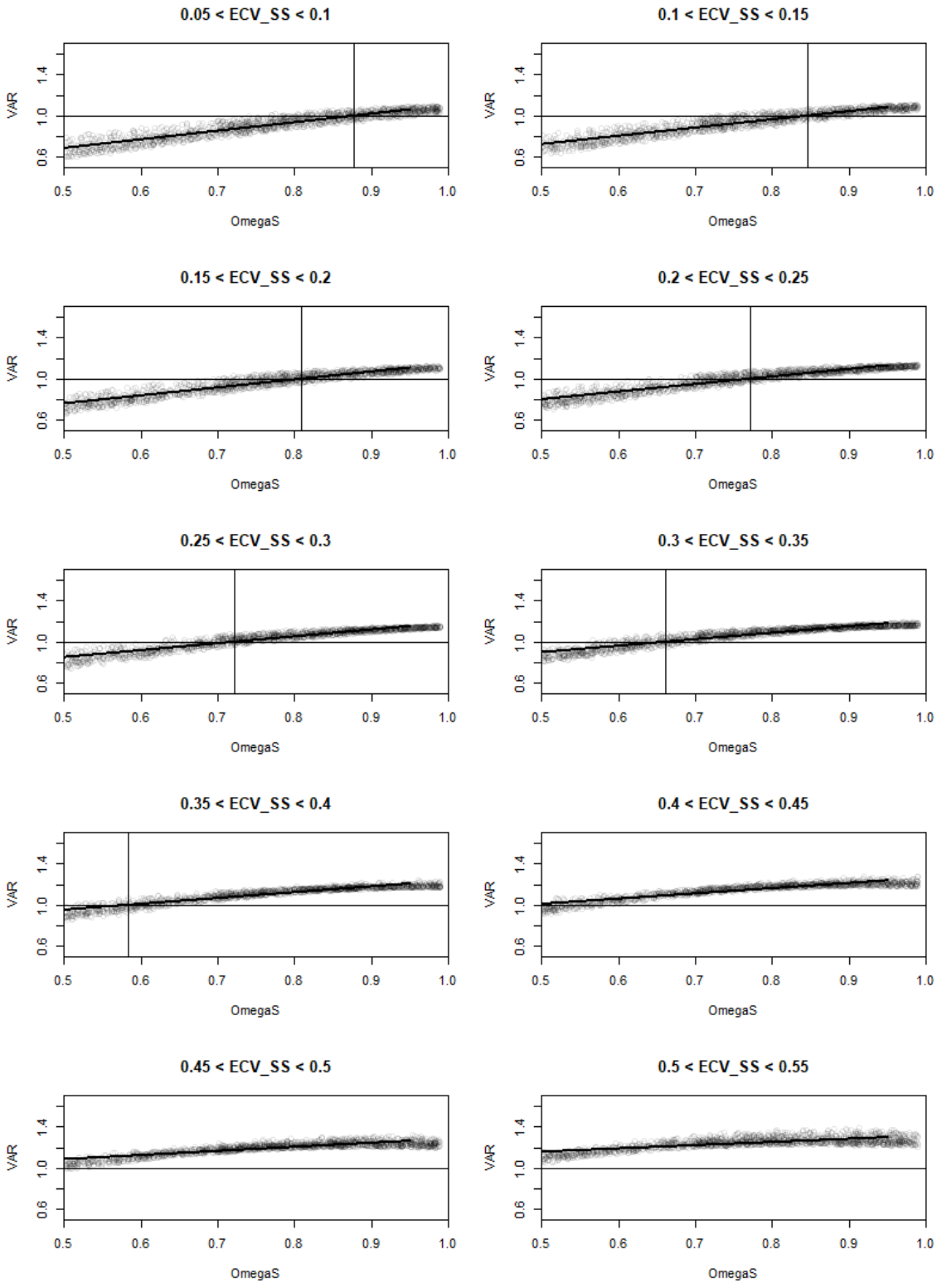

Figure 3.12 VAR versus OmegaS for Two Factors, Showing Cutoffs for VAR = 1.0

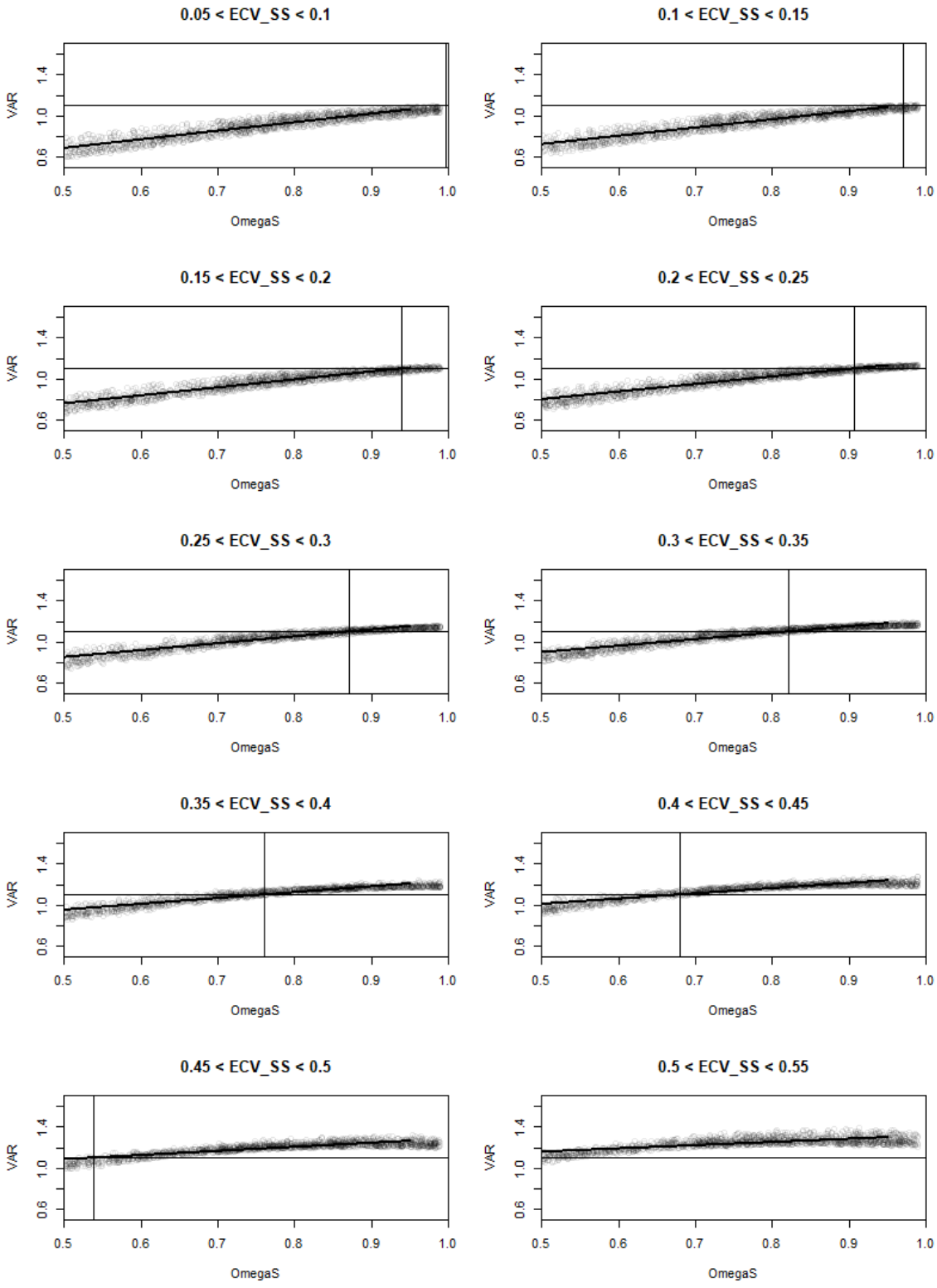

Figure 3.13 VAR versus OmegaS for Two Factors, Showing Cutoff for VAR = 1.1

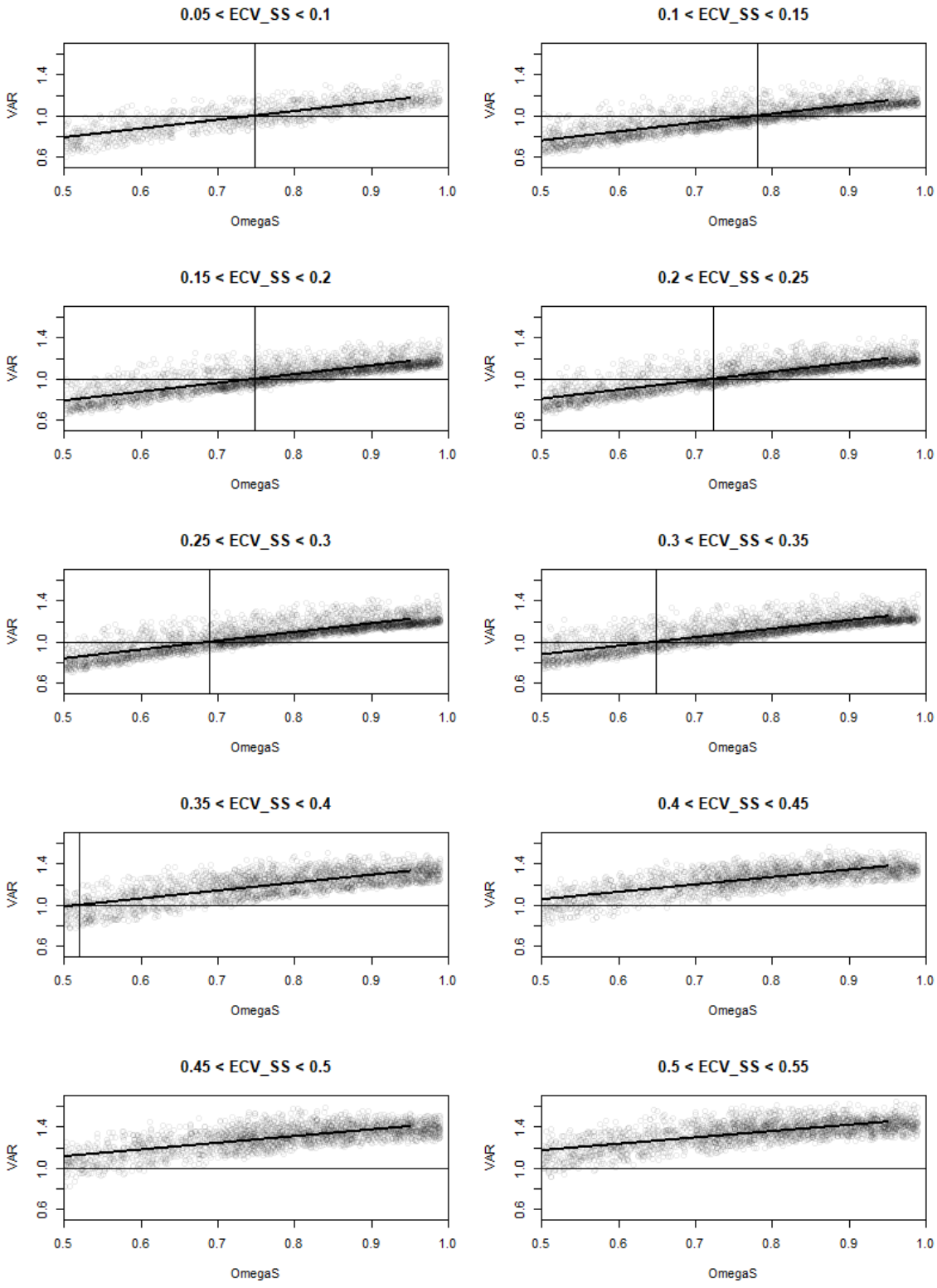

Figure 3.14 VAR versus OmegaS for Three Factors, Showing Cutoffs for VAR = 1.0

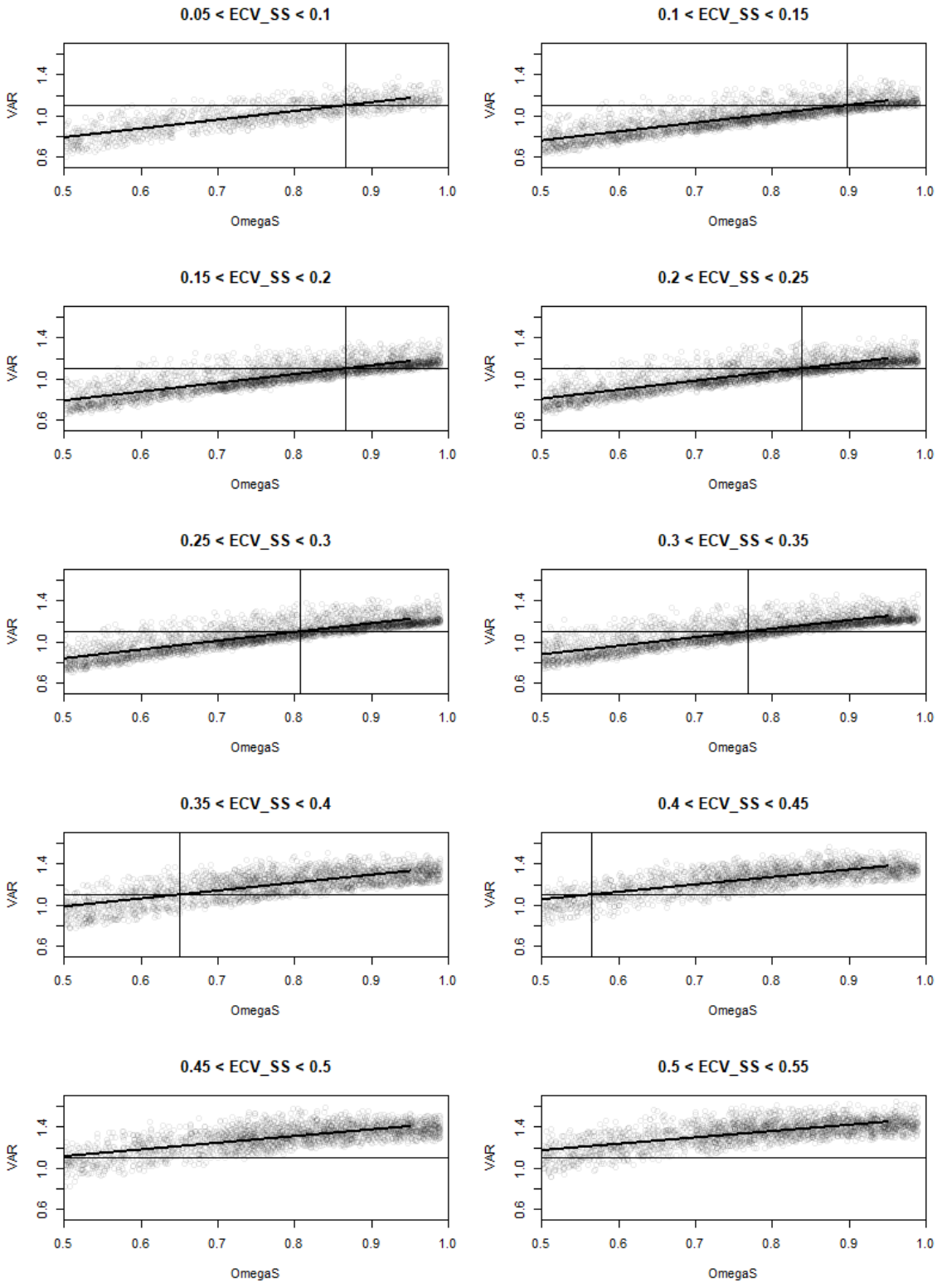

Figure 3.15 VAR versus OmegaS for Three Factors, Showing Cutoffs for VAR = 1.1

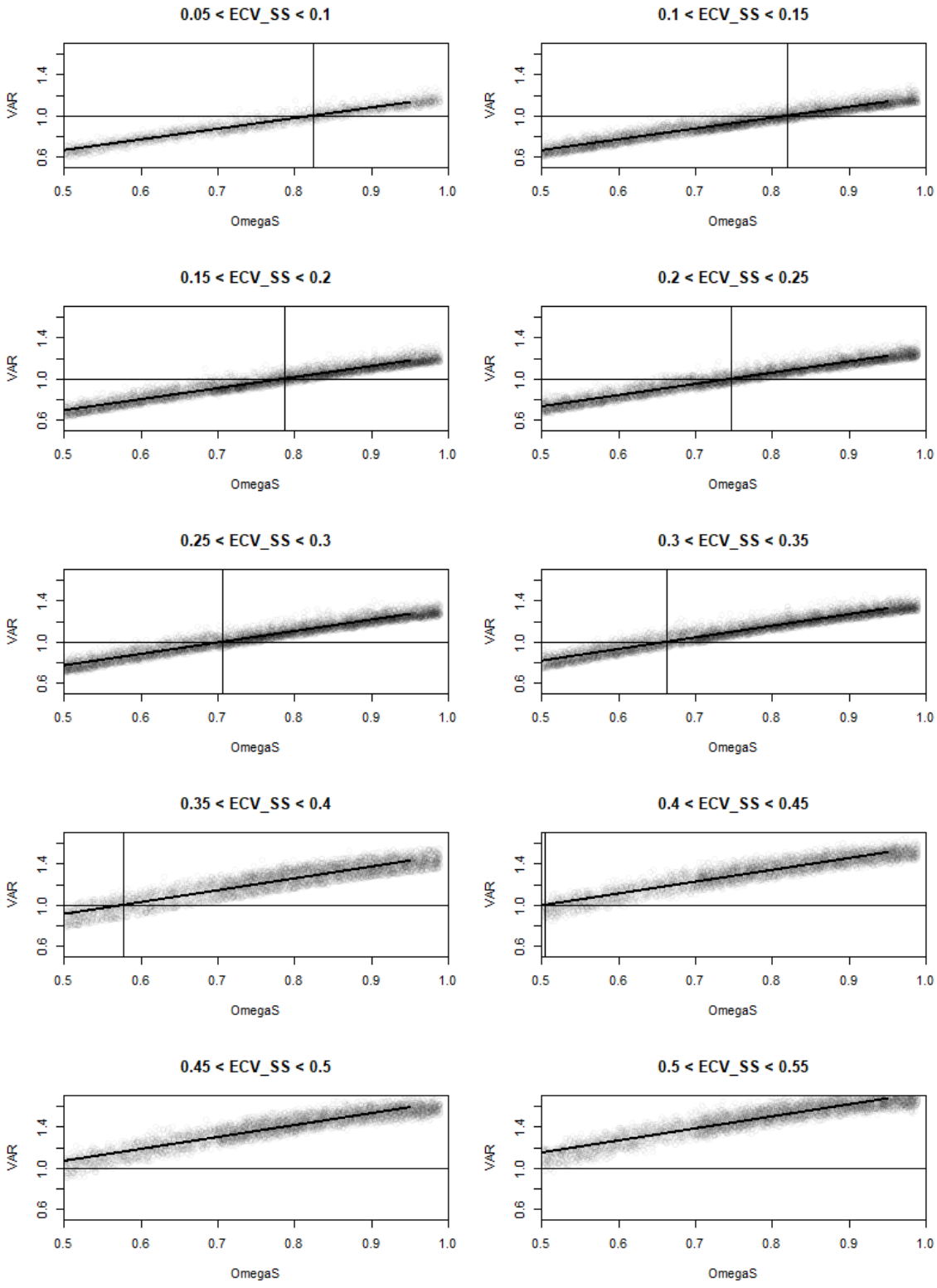

Figure 3.16 VAR versus OmegaS for Six Factors, Showing Cutoffs for VAR = 1.0

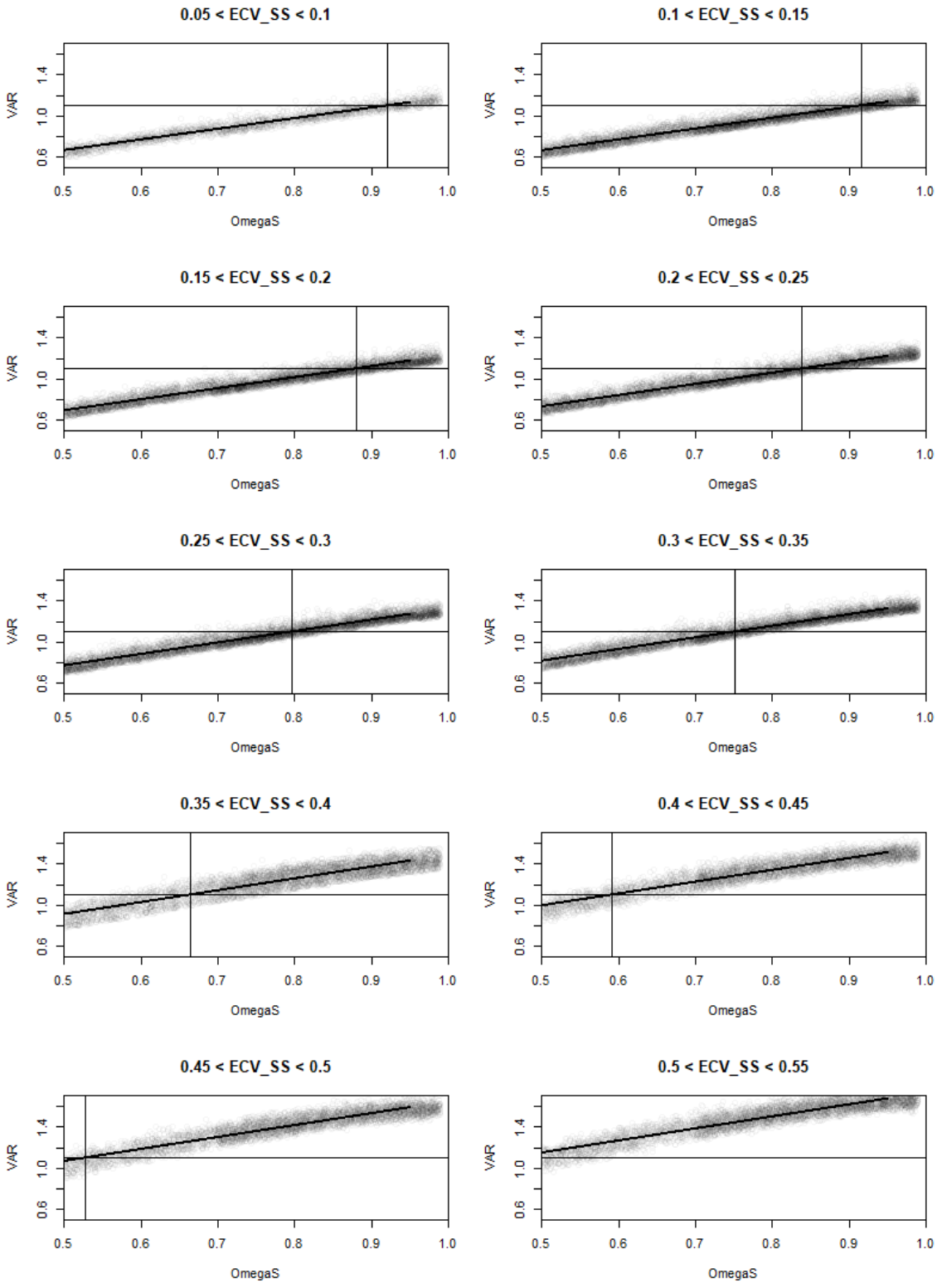

Figure 3.17 VAR versus OmegaS for Six Factors, Showing Cutoffs for VAR = 1.1

3.3.3 Testing of Models with Replication Sample

Regression coefficients, coefficients of determination, OmegaHS cutoffs, OmegaS cutoffs, and all sensitivities and specificities were re-estimated using a second set of simulated data. No estimate differed by more than .02 between the original and replication simulated data; in fact, only a small number differed by more than .01.

3.4 Empirical Example

A dataset of 1,074 adolescent idiopathic scoliosis patients with SQLI data measured at doctor visits between 2010 and 2017 was kindly provided for use (Anonymous, 2019) as an example in this study; patients averaged 2.16 visits. In order to avoid person dependencies in the data, a single timepoint was randomly chosen for each participant. The resulting dataset consisted of complete responses to the 20 SQLI items for 1,074 adolescents with idiopathic scoliosis. As both total SQLI scores and subdomain scores are commonly interpreted, unidimensional, four-factor, and bifactor with four specific factors models were estimated using the *lavaan* package for R (Rosseel & Jorgensen, 2019) using the DWLS estimator. Model fit information can be found in Table 3.9, and was compared to common fit index cutoff to judge quality of fit. Specifically, a non-significant chi-square test indicates exact fit, RMSEA < .06, CFI > .95, and SRMR < .08 indicate close fit, RMSEA < .08, CFI > .95, and SRMR < .08 indicate acceptable fit, and RMSEA < .10 and CFI > .92 indicate marginal fit (Byrne, 2008; Hu & Bentler, 1999; MacCallum et al., 1996; West et al., 2012). The unidimensional model exhibits poor fit, while the four factor model exhibits marginal fit and the bifactor model exhibits acceptable fit.

Table 3.9 Model Fit Information

Model	Chi-Square	RMSEA	CFI	SRMR
Unidimensional	$\chi^2(170) = 4390, p < .001$	.152	.808	.116
Four Factor	$\chi^2(164) = 1301, p < .001$	.080	.932	.071
Bifactor	$\chi^2(150) = 802, p < .001$	.064	.970	.058

Note. In the four factor model, latent inter-factor correlations ranged from .521 to .759.

As the fit of the bifactor model is acceptable, the model and its indices may be interpreted. Note that the bifactor model is being proposed as a supplemental model to assess dimensionality (Rodriguez et al., 2016a) rather than as a "true" model. That is, the bifactor model serves here only as a tool to examine the partitioning of variance into general and specific sources to aid in decision-making about whether unidimensional and/or multidimensional (i.e., correlated traits model) interpretations are appropriate. Parameter estimates and bifactor indices for the bifactor model can be found in Table 3.10.

With overall OmegaH = .849, observed total scores can safely be interpreted (Rodriguez et al, 2016b). To determine whether subscores can also be interpreted, bifactor indices are compared to the criteria listed in Tables 3.3 and 3.7Table 3.7. According to these decision rules, all four subscores have VAR > 1.1, and therefore their interpretation adds value. Additionally, Equations 3.13 and 3.14 were used to estimate VAR, which was also computed using the *subscore* package for R (Dai et al., 2019). Estimates of VAR can be found in Table 3.11. Differences amongst estimates are approximately what would be expected given the variance of the error terms in Equations 3.13 and 3.14. Some of that difference may be explained by the number of factors; it was noted in the simulation results

Table 3.10 Standardized Factor Loadings and Bifactor Indices for Bifactor Model of SQLI

Item	General Factor	Self-Esteem	Back Pain	Physical Activity	Moods & Feelings
SQLI 1	.536	.270			
SQLI 2	.449	.629			
SQLI 3	.525	.654			
SQLI 4	.429	.738			
SQLI 5	.681	.101			
SQLI 6	.725		.640		
SQLI 7	.737		.453		
SQLI 8	.679		.442		
SQLI 9	.629		.306		
SQLI 10	.753		.097		
SQLI 11	.694			.316	
SQLI 12	.708			-.010	
SQLI 13	.726			.590	
SQLI 14	.656			.508	
SQLI 15	.719			.494	
SQLI 16	.610				.496
SQLI 17	.443				.576
SQLI 18	.577				.536
SQLI 19	.540				.433
SQLI 20	.556				.562
ECV/ ECV_{SS}	.627	.507	.268	.279	.477
Omega/ OmegaS	.959	.855	.910	.909	.870
OmegaH / OmegaHS	.849	.389	.211	.206	.415

that the number of factors additionally influenced VAR above and beyond the predictors of Equations 3.13 and 3.14. Notably, using ECV_{SS} and OmegaS resulted in estimates closer to the *subscore* estimates than estimates using OmegaHS and OmegaS when VAR was large (> 1.20), while estimates were more consistent when VAR was not large. Finally, it should also be noted that estimates provided by *subscore* use coefficient alpha for reliability; inaccuracies in reliability estimation due to the multidimensionality of SQLI and violation of tau-equivalence in the data (McNeish, 2018) will bias estimation of VAR.

Table 3.11 Estimates of VAR for SQLI subdomains

Estimation Method	Self Esteem	Back Pain	Physical Activity	Moods & Feelings
subscore	1.275	1.136	1.194	1.310
Equation 3.13	1.417	1.164	1.159	1.466
Equation 3.14	1.292	1.098	1.103	1.264

In the example dataset, approximately 10% of patients who are fitted with a brace show little to no change in overall SQLI score after receiving the brace, but report substantial decrease in moods and feelings (compensated for by moderate improvement in back pain and physical activity). As subdomain scores for the SQLI indeed provide added value over the total score, the subdomains are indeed individually interpretable, and doctors should feel confident making clinical decisions based on these differences.

Additionally, while SQLI total scores are interpretable, a unidimensional model for SQLI should not be used in structural equation modeling contexts, as multidimensionality in the data may lead to significant bias in estimated coefficients. Instead, a bifactor measurement model should be used if a general SQLI factor is to be interpreted in a

structural equation model (Bonifay et al., 2015; Reise et al., 2013a). This data supports both unidimensional and multidimensional interpretations of the SQLI.

3.5 Discussion

3.5.1 Estimating VAR using OmegaHS and OmegaS

While OmegaHS has been touted as an indicator of dimensional uniqueness and as being sufficient to justify the interpretability, or lack thereof, of subscores (Gignac & Kretschmar, 2017; Reise et al., 2013), the results of this study illustrate a more nuanced picture. Furthermore, while Reise et al. (2013) dismiss OmegaHS estimates of .11 and .22 as being insufficient for separately interpreting subdomains, the results of Table 3.3 and Table 3.7 suggest that both of these subdomains have added value and VAR > 1.1.

While OmegaHS was found to be insufficiently predictive of VAR on its own, accuracy was much improved by also including OmegaS as a predictor. Interestingly, after including OmegaS as a predictor, nonlinearity in the relationship between VAR and OmegaHS became apparent, necessitating the use of a quadratic OmegaHS term. A possible explanation for this phenomenon is that, as OmegaHS increases, the influence of the specific factor increases while the influence of the general factor decreases. These two influences feed off each other, resulting in growth which is faster than linear. Using both OmegaHS and OmegaS to predict VAR was found to have excellent sensitivity but relatively poor specificity. The large difference between sensitivity and specificity can be explained by the shape of the relationship between OmegaHS and VAR as pictured in Figure 3.4 though Figure 3.9. As VAR curves around the cutoff point, there are naturally more points above and to the left of the cutoff (false negatives) than there are below and to the right of the cutoff (false positives). Examination of the response operator characteristic

(ROC) curves reveals cutoffs with lower sensitivity and higher specificity. For example, by slightly increasing the cutoffs in Table 3.2 and Table 3.3, specificity and sensitivity of greater than .90 can be obtained for all number of factor conditions. However, modeling VAR was seen as a more important goal than producing strict cutoffs, so cutoffs based on expected value of VAR from regression equations were used.

In an effort to further improve predictive capabilities, separate regression models were fit for each different number of factors. This revealed a generally increasing level of accuracy as the number of factors increased but also unexpected results when the number of factors was two. Recall that special restrictions on the population model had to be imposed for two factors because second-order factor models with only two first-order factors are not identified. Accordingly, for two-factor models, both factors are equally correlated with the second-order factor. This symmetry between the two factors explains why results for two dimensions do not match results for more dimensions. As for increasing accuracy with increasing number of factors, it seems likely that this is mostly due to the way sampling was performed in the simulation.

As the number of factors increases, the average of factor-level properties (reliability, second-order factor loading) becomes more stable, reducing overall variability. As a result, sensitivity and specificity increase because of lowered overall variability. With such a wide range of factor-level properties and such highly variable sampling approach (i.e., using uniform distributions), this simulation study likely generates more highly variable models than are found in practice. While regression model parameters, and thereby cutoff recommendations are not expected to be unduly influenced by the specific sampling choices in simulation design, sensitivity and specificity are influenced by sampling

variability in the simulation design. Accordingly, it is expected that sensitivity and specificity across the population of models fit in the course of empirical research will not exactly match what has been found herein. Finally, it should be noted that, per Table 3.2 and Table 3.3, the number of factors makes little difference in OmegaHS cutoffs so long as the number of factors is at least four. Given that accuracy of estimates of bifactor indices is also lesser for two and three factors, this observation raises questions about the appropriateness of making overall recommendations concerning use of bifactor indices. Instead, the cases of two- and three- specific factor models should be treated separately and with great care.

The relationship between OmegaHS and VAR conditional upon OmegaS can clearly be seen in Figure 3.4 through Figure 3.9. As OmegaS increases, the level of OmegaHS required to create a certain expected level of VAR decreases. In each of Figure 3.4 through Figure 3.9, the cutoff point for OmegaHS moves to the left as OmegaS increases. This situation can be explained by considering Equations 3.11 and 3.12. As OmegaS increase, the numerator of VAR naturally increases; on the other hand, while OmegaHS increases, the denominator of VAR decreases. So, as OmegaS increases, VAR will naturally increase independent of OmegaHS. Thus, less dimensional uniqueness (OmegaHS) is required for reaching a certain level of VAR. Indeed, for very high levels of OmegaS, subscores almost always have added value, regardless of OmegaHS. This fact is much more relevant in psychological and related literature where subscores often have high reliability than it is in education testing settings, where subscores more often have low reliability (Sinharay et al., 2011).

While the overall accuracy of using OmegaHS and OmegaS to create cutoffs for VAR = 1.0 and VAR = 1.1 did not reach the desired levels of sensitivity and specificity, OmegaHS was shown to be in indicator of dimensional uniqueness as claimed by others (Gignac & Kretschmar, 2017; Reise et al., 2013). Furthermore, OmegaS was found to also have a significant contribution to predicting VAR; indeed when OmegaS is high, OmegaHS becomes less relevant to determining if a subscore has added value.

Despite not recommending strict cutoffs, improvement on the suggestions of other researchers (Gignac & Watkins, 2013; Reise et al., 2013a) can still be made. Table 3.2 and Table 3.3 make clear that a cutoff of .50 for OmegaHS is inappropriately high. For low subscore reliability (OmegaS = .60), OmegaHS = .25 is sufficient that the subscore has a good chance of having added value (VAR > 1.1) above and beyond the total score. For moderate reliability (OmegaS = .80), OmegaHS = .20 is sufficient, and the role of OmegaHS diminishes as OmegaS increases further. It is important to note that a subscore having added value does not necessitate its interpretation. Instead, when subscores are desired to be interpreted, high OmegaHS can be considered as evidence that such an interpretation is statistically appropriate. When only interpreting a total score is desired, high OmegaHS is not necessarily problematic so long as the total score has adequate psychometric properties (e.g., OmegaH > .80). For very large values of OmegaHS, inclusion of items on that subdomain may degrade measurement of the general construct; in this case, performance of a sensitivity analysis is recommended: analyses can be conducted with a total score and then again with the subscore removed from the total score. An example of using OmegaHS to support a recommendation to remove a subscore from a total score can be found in Mészáros et al. (2014), who suggest that the Maslach Burnout

Inventory – Human Services Survey (Maslach & Jackson, 1981) be scored using separate personal accomplishment and burnout scores, with burnout consisting of emotional exhaustion and depersonalization items.

3.5.2 Estimating VAR using ECV_{SS} and OmegaS

ECV_{SS} is a natural indicator of dimensional uniqueness from a latent variable modeling perspective, as it represents the proportion of common variance of items loading on a specific factor which is independent of the general factor. While ECV_{SS} was found to be insufficiently predictive of VAR on its own, accuracy was much improved by also including OmegaS as a predictor. Interestingly, after including OmegaS as a predictor, non-linearlity in the relationship between VAR and ECV_{SS} became apparent, necessitating the use of a quadratic ECV_{SS} term. A possible explanation for this phenomenon is that, as ECV_{SS} increases, the influence of the specific factor increases while the influence of the general factor decreases. These two influences feed off each other, resulting in growth which is faster than linear. Using both ECV_{SS} and OmegaS to predict VAR was found to have excellent sensitivity and adequate specificity. Examination of the response operator characteristic (ROC) curves would no doubt reveal cutoffs with lower sensitivity and higher specificity. However, modeling VAR was seen as a more important goal than producing strict cutoffs, so cutoffs based on expected value of VAR from regression equations were used.

In an effort to further improve predictive capabilities, separate regression models were fit for each different number of factors. This revealed a generally increasing level of accuracy as the number of factors increased but also unexpected results when the number of factors was two. This generally matches the results from using OmegaHS and OmegaS

to predict VAR. Specifically, cutoffs for VAR = 1.0 and VAR = 1.1 were found to be relatively consistent for five or more factors, while the conditions with two or three factors diverged strongly. Similarly, predictive accuracy was acceptable for five or more factors, marginal for four factors, and poor for three factors. As with using OmegaS and OmegaHS to predict VAR, specificity and sensitivity were very high for two factors, likely as a result of the low variability amongst models with only two specific factors in the simulation design.

The relationship between ECV_{SS} and VAR conditional upon OmegaS can clearly be seen in Figure 3.12 through Figure 3.17. As ECV_{SS} increases, the level of OmegaS required to create a certain expected level of VAR decreases. In each of Figure 3.12 through Figure 3.17, the cutoff point for OmegaS moves to the left as ECV_{SS} increases. This situation can be explained by considering Equations 3.7 and 3.8. As ECV_{SS} increase, the denominator of VAR naturally increases; on the other hand, as OmegaS increases, the numerator of VAR increases. So, as ECV_{SS} increases, VAR will naturally increase. Thus, less subscore reliability (OmegaS) is required for reaching a certain level of VAR. Indeed, for very high levels of ECV_{SS}, subscores can have added value even when OmegaS is low enough that interpreting that subscore is not recommended.

When subscores are found to not have added value, a common response is to add items to the subdomain in order to improve its reliability. In fact, Brennan (2012) proposed a utility index, $\widetilde{U}$, which uses the Spearman-Brown prophecy formula to estimate the number of additional parallel items which would need to be added to a subdomain to ensure than the subscores had added value. Brennan's (2012) technique for estimating the number of parallel items needed to be added is statistically equivalent to using the prophecy

formula with Haberman's (2008) indices, assuming that PRMSE(x) will not change as a result of adding these items. Unfortunately, adding items to a subdomain also tends to increase reliability of the total score; thus, using $\widetilde{U}$ is likely to result in an underestimate of the number of items which need to be added in order for a subscore to have added value. The method of using ECV_{SS} and OmegaS to predict VAR does not have this drawback. While bifactor indices such as Omega, OmegaH, ECV, OmegaS, and OmegaHS will necessarily change as additional items are added to a subdomain, ECV_{SS} depends only on the ratio of common variance explained by the general and specific factors for the items in the subdomain. Thus, so long as new items load similarly on both general and specific factors as the old items do, ECV_{SS} will not be affected. Accordingly, it is possible to use Equation 3.14 with ECV_{SS} from a fitted bifactor model and a desired level of VAR to find a required level of OmegaS to achieve that VAR. Then, the prophecy formula can be used to determine the number of additional items which need to be added to achieve that level of OmegaS. As an example, consider a specific factor with 8 items from a bifactor model with five specific factors with ECV_{SS} = .300 and OmegaS = 0.720. Per Table 3.7, OmegaS = .768 would be required to achieve VAR = 1.1. The prophecy formula is given by

$$\rho_{new}^2 = \frac{n \times \rho_{old}^2}{1 + (n-1)\rho_{old}^2}, \tag{3.15}$$

where n is the ratio of the length of the old subdomain to the new subdomain, ρ_{old}^2 is the reliability of the old subdomain, and ρ_{new}^2 is the reliability of the new subdomain with the added items. In this example, therefore, we have

$$.768 = \frac{n \times .720}{1 + (n - 1).720},$$

so that n = 1.29, and three more items are required for the subdomain to have a reliability above .768 and a VAR above 1.1.

The overall accuracy of using ECV_{SS} and OmegaS to create cutoffs for VAR = 1.0 and VAR = 1.1 reached the desired levels of sensitivity and specificity, and researchers should feel confident interpreting ECV_{SS} as an indicator of dimensional uniqueness. The higher specificity of cutoffs created using ECV_{SS} and OmegaS is partially the result of the linear shape of relationship between OmegaS and VAR, compared to the quadratic relationship between OmegaHS and VAR. Furthermore, predicting VAR with ECV_{SS} and OmegaS permits use of the prophecy formula to unbiasedly estimate VAR upon addition of parallel items to the subdomain.

Despite not recommending strict cutoffs, some suggestions for interpreting specific values of ECV_{SS} can still be made. Table 3.6 and Table 3.7 suggest that ECV_{SS} = .45 is probably sufficient to warrant interpretation of a subscore when that subscore's reliability is low (OmegaS = .60). Furthermore, for moderate reliability (OmegaS = .80), ECV_{SS} = .30 is probably sufficient to warrant interpretation of a subscore, and the importance of ECV_{SS} diminishes as OmegaS increases further. As with recommendations for OmegaHS, high ECV_{SS} should be considered as evidence for interpreting a subscore, not as evidence for not interpreting a total score.

3.5.3 Comparison with Prior Bifactor Simulation Research

As with prior simulation research involving bifactor models and dimensionality (Bonifay et al., 2015; Reise et al., 2013a), this study only considered second-order factor models (technically, both of those other studies considered bifactor models which are statistically equivalent to second-order models). Compared to these studies, however, the

present research considered a far greater diversity of models and a far greater number of models. While previous research involved bifactor models each of which had constant loadings across all specific factors, loadings were allowed to vary across specific factors in this study. As the goal of the present study was to evaluate specific factors, the specific factors were required to be different from each other. However, since prior bifactor dimensionality research only considered the general factor, no attention was paid to the specific factors. Also, Reise et al. (2013a) considered only 120 different models and Bonifay et al. (2015) considered only 300 different models, compared to the 84,000 models considered herein. Both the additional complexity and sheer volume of models considered contribute to the complexity and difficulty of interpreting the results. Therefore, the results recorded herein raise the question of whether the simplicity and clarity of Bonifay et al.'s (2015) and Reise et al.'s (2013) conclusions are misleading.

Some other simulation research involving bifactor models, not specific to dimensionality, is similarly afflicted by overly simplistic model choices. For example, Green et al. (2018) simulated from both correlated traits and bifactor models in their comparison of parallel analysis methods; however, their bifactor models were structured such that they were nearly identical to some of the correlated traits modes, albeit with lower inter-factor correlations. However, research concerning fit index bias in favor of bifactor models (Green et al., 2019; Morgan et al., 2015) have necessarily involved more sophisticated and varied models from which data were simulated, including ones for which neither a second-order nor a bifactor model are a perfect fit in the population.

3.6 For the Applied Researcher

The primary takeaway of this study for the applied researcher is that OmegaHS and ECV_{SS} are both useful indicators of dimensional uniqueness. Their precise interpretation requires also considering OmegaS. Levels of OmegaHS required for appropriately interpreting subscores are much lower than claimed by prior methodological literature (Gignac & Kretschmar, 2013; Reise et al., 2013a). For low subscore reliability (OmegaS = .60), OmegaHS = .25 or ECV_{SS} = .45 is sufficient that the subscore has a good chance of having added value (VAR > 1.1) above and beyond the total score. For moderate reliability (OmegaS = .80), OmegaHS = .20 or ECV_{SS} = .30 is sufficient, and the role of OmegaHS and ECV_{SS} diminish as OmegaS increases further. Importantly, a subscore having added value does not necessitate its interpretation. Instead, when subscores are desired to be interpreted, high OmegaHS or ECV_{SS} can be considered as evidence that such an interpretation is statistically appropriate. When only interpreting a total score is desired, high OmegaHS is not necessarily problematic so long as the total score has adequate psychometric properties (e.g., OmegaH > .80).

These recommendations are tempered by limitations of this study. When the number of subdomains is low, slightly higher values of OmegaHS and ECV_{SS} may be required. Furthermore, when the number of subdomains is two, the extent of error in estimating population bifactor indices from a fitted model may be severe. Also, as with all simulation studies, factors not incorporated into study design may alter interpretation of OmegaHS and ECV_{SS}. Specifically, this study did not consider sampling error, misfit in the second-order CFA model, or very low reliabilities in some subscores.

3.7 Limitations and Conclusion

3.7.1 Limitations

A major limitation of any simulation study is imperfect coverage of data structures found in real-world research. While a wide range of second-order models were generated and analyzed in this study, it is possible that researchers may encounter data which is not well-described by any of these models. For example, models including specific factors with very low reliability or very low second-order factor loadings are not represented in this study. More importantly, multidimensional data often does not satisfy a second-order structure (Chen et al., 2006; Chen et al., 2012). When the true model for data is other than second-order, a bifactor model will not fit the data perfectly, so any conclusions drawn from such a model should be tempered with uncertainty due to misfit. It is a subject for future research to analyze the usefulness of bifactor indices with correlated traits models which do not satisfy the conditions of a second-order model.

As with other dimensionality simulation research examining multidimensionality (Bonifay et al., 2015; Reise et al., 2013b), sampling was not considered in this study. When fitting bifactor models based on samples, parameter estimates and estimates of bifactor indices will be subject to sampling error. The influence of sampling error on estimation of bifactor indices has not yet been studied. Furthermore, as discovered in the pilot study, bifactor models fit to data may result in biased estimates of bifactor indices, particularly when the number of dimensions is small. This phenomenon has also not been subject to research.

Finally, the level of error and complexity of Table 3.2, Table 3.3, Table 3.6, and Table 3.7 prohibit them from functioning as perfect cutoffs for decision-making regarding

whether subscores have added value. However, it is hoped that these tables may serve as general guidelines, to be interpreted with the degree of possible error in mind.

3.7.2 Conclusions

When the number of specific factors is at least four, Equations 3.13 and 3.14 and Table 3.2, Table 3.3, Table 3.6, and Table 3.7 provide reasonably accurate information about VAR to aid in decision-making about interpretation of subscores. OmegaHS and ECV_{SS} both serve as effective indicators of dimensional uniqueness and, conditional upon OmegaS, as effective indicators of the value added by a subscore.

Additionally, this study raises awareness of several gaps in the literature on bifactor models. First, the sampling distribution of bifactor indices is completely unknown. Second, the ability of bifactor models fit to empirical data to yield accurate and meaningful bifactor indices is suspect when the number of specific factors is small, and is unknown in the case of true correlated traits models not satisfying the restrictions of a second-order factor model. Finally, the relative accuracy of VAR as traditionally computed (say, using the *subscore* package) compared to the accuracy of Equations 3.13 and 3.14 is unknown.

While previous research concerning bifactor indices were largely concerned with using them to evaluate whether data could be considered essentially unidimensional (Bonifay et al., 2015; Reise et al., 2013a; Rodriguez et al., 2016a), this study was concerned with whether data was multidimensional enough for subscores to be interpreted. When combined with prior bifactor research, this work extends a framework (Rodriguez et al., 2016a, 2016b) of using confirmatory bifactor models for dimensionality assessment.

CHAPTER 4. CONCLUSION

Bifactor models and their indices have been widely used to aid in decision-making about dimensionality. While recommendations for using bifactor indices to evaluate the appropriateness of interpreting a total score or unidimensional measurement model are well establish (Rodriguez et al., 2016a, 2016b), recommendations for using bifactor models to evaluate the appropriateness of subscores have previously taken the form of unsubstantiated options (Gignac & Kretschmar, 2017; Gignac & Watkins, 2013; Reise et al., 2013a). Furthermore, computation of bifactor indices has typically been done manually or using one of several programs (Dueber, 2017; Revelle, 2020; Watkins, 2013) which lack in convenience features, fail to provide all relevant indices, are limited to being used with a narrow range of models, or require the user to employ a specific program for estimation of factor models.

The BifactorIndicesCalculator package for the R statistical computing environment provides a user-friendly platform for computing bifactor indices. Estimated exploratory and confirmatory models can be input directly into BifactorIndicesCalculator functions, or the user can directly input a matrix of factor loadings. The range of model types supported make the BifactorIndicesCalculator a convenient tool for use both in empirical research and in simulation research. Researchers who do not wish to use R can access BifactorIndicesCalculator as a Shiny app and directly load Mplus .out files in order to obtain bifactor indices for the estimated model.

In Study Two, the utility of bifactor indices in predicting whether subscores will have added value over the total score was investigated. OmegaHS and ECV_{SS} were each found to substantially predict VAR, especially when OmegaS was included as a predictor. Results showed some variation for models with two and three subscales, but where fairly robust for larger models. Suggestions for when OmegaHS or ECV_{SS} are large enough that a subscore may safely be interpreted separately from the total score were made; required levels of OmegaHS were much lower than previous researchers have suggested (Gignac & Watkins, 2013; Reise et al., 2013a). The relationship between ECV_{SS}, OmegaS, and VAR was used in tandem with the Spearman-Brown prophecy formula to estimate the number of additional items a subdomain would need to have added value. Finally, a demonstration was made using data from the SQLI to show a concrete example of data for which a total score is interpretable and subscores are also interpretable.

4.1 Implications for Applied Researchers

The BifactorIndicesCalculator package for the R statistical computing environment and associated Shiny app, provide applied researchers with a convenient way to compute bifactor indices.

Study Two provides a more rigorous set of guidelines around interpreting OmegaHS (and ECV_{SS}) than was previously available. While very low levels of OmegaHS indicate that only a total score should be interpreted, levels of OmegaHS previously considered small or moderate are nevertheless associated with subscores having added value. Researchers arguing that only a total score should be interpreted ought not to consider moderate levels of OmegaHS as evidence in support of their claim. On the other hand, researchers who desire to interpret subscores need not meet the stringent criteria

suggested by Gignac and Watkins (2013) or Reise et al. (2013a); much lower levels of OmegaHS are sufficient to provide evidence for the suitability of interpreting subscores. If subscores do not have added value but researchers desire to interpret those subscores, a mechanism was provided using ECV_{SS}, OmegaS, and the Spearman-Brown prophecy formula by which the number of additional items which should be added to the subdomain may be estimated.

4.2 Implications for Future Research

It has been said that good research raises more questions than it answers; by this metric, at least, the present research has been a success. The most significant question raised through this research is how well bifactor indices measure what they are purported to measure, particularly in non-ideal situations. As discussed in the methodology of Study Two, bifactor indices computed from estimated models showed severe bias compared to the theoretical population values when the number of specific factors was two. A bifactor model with only two specific factors was estimated in 26% (39 out of 149) of studies reporting an estimated bifactor model from the search of PsychINFO reported in earlier chapters, highlighting the importance of this issue. Simulation research could be performed to more fully examine the extent of this bias and determine to what extent the conclusions of current and prior bifactor simulation research (Bonifay et al., 2015; Reise et al., 2013b) are valid when the number of specific factors is small.

The issues of recovery and interpretation of population bifactor indices using estimated models extends far beyond the case of a small number of specific factors. As the current research and prior bifactor simulation research (Bonifay et al., 2015; Reise et al., 2013b) have employed only second-order models, there is a lingering question of how well

indices computed from actual estimated models measure what they are purported to measure. Are the recommendations of Study Two and of prior bifactor simulation research applicable when the true model is a correlated traits model which does not conform to a second-order pattern. As Chen et al. (2006, 2012) observed, well-fitting second-order models are rare in applied research? Even more generally, how well do these recommendations hold up when the correlated traits model exhibits misfit? Additionally, prior bifactor simulation research (Bonifay et al., 2015; Reise et al., 2013b) utilized a narrow range of models, which may relative simplicity of results in these studies as compared to Study Two. How would conclusion of these studies differ if they were to consider a more representative sample of models? Furthermore, bifactor simulation research has thus far been restricted to population models (Reise et al., 2013b) or very large samples (Bonifay et al., 2015). While both the delta method and bootstrapping show some promise for estimating confidence intervals of reliability coefficients (Kelley & Pornprasertmanit, 2016), little is known about the sampling distribution of bifactor indices. Understanding the expected level of sampling error would help researchers be more confident (or more tentative) in the conclusions they draw based on bifactor indices.

Continuing the theme of accurate recovery of population bifactor indices, it was noted earlier that some preliminary work has suggested that bifactor indices computed from exploratory models may be interpreted in a similar way as indices computed from a confirmatory model (Murray et al., 2019). However, Perreira et al. (2018) raise concerns about using exploratory models for computing bifactor indices as well as concerns about the formula used for indices. In recent years, the variety of types of hierarchical models with general factors being fit to data has exploded; in addition to bifactor CFA models,

bifactor exploratory factor analysis models (Jennrich & Bentler, 2011), bifactor exploratory structural equation models (Morin et al., 2016), incomplete bifactor models (Eid et al., 2016), trifactor models (Rijmen, 2011), and two-tier models (Cai, 2010) are becoming common in applied literature. While the various bifactor indices make intuitive sense in all of these contexts, information about how well interpretations and recommended cutoffs transfer to these other analytic contexts has not been developed.

Finally, while Study Two addresses the question of whether a subscore can appropriately be interpreted, it does not address the question of whether that subscore ought to be interpreted instead of a total score. Incorporating a subscore which is sufficiently uncorrelated with the construct of interest into the total score will degrade the quality of measurement of that construct, as its inclusion adds error variance but not construct relevant variance. Thus, at some level of dimensional uniqueness, a statistical recommendation can be made to remove a subscore from the total score so as to improve measurement. This type of claim has been observed in empirical research (Mészáros et al., 2014). While the decision to include specific subdomains in a scale is a mostly theoretical one, the ability to use bifactor indices to determine when inclusion of a subdomain degrades measurement of the general factor would help provide statistical evidence for the appropriateness of that theoretical decision.

4.3 Final Conclusions

Bifactor models and indices computed based on parameter estimates from bifactor models are useful tools for providing evidence about the appropriateness of different dimensional interpretations of data. This dissertation adds to prior research in that it provides a mechanism for using bifactor indices for subscore assessment. Abundant

questions remain about uses and appropriateness of bifactor models and indices, particularly when the number of specific factors is low.

APPENDICES

APPENDIX A

Code for Simulating Data

```
library(BifactorIndicesCalculator)
library(parallel)

# Generates all relevant data:
#   Second order factor scores
#   Second order factor loadings
#   First order scores
#   First order factor loadings
#   Indicator scores
GenerateData <- function (N, nfac, load, rel, nitems) {
  # Generates first order factor scores
  GenerateFirstOrder <- function (N, nfac, loads, scores) {
    sapply(1:nfac, function(x) {
      scores*loads[x] + rnorm(N, 0, sqrt(1-loads[x]^2))
    })
  }

  # Generates item scores
  GenerateItems <- function (N, rel, scores, nitems) {
    load <- sqrt(rel/(nitems - (nitems-1)*rel))
    sapply(1:nitems, function(x) {
      scores*load + rnorm(N, 0, sqrt(1-load^2))
    })
  }

  # Now lets start generating the data
  SecondOrderScores <- rnorm(N, 0, 1)

  # Second order factor loadings according to the
  # loading distribution type
  # If nfac is 2, then we need to make the loadings the same
  if (nfac == 2) {
    if (load == "high") {
      SecondOrderLoadings <- rep(runif(1, .80, .99), 2)
    } else if (load == "low") {
      SecondOrderLoadings <- rep(runif(1, .50, .80), 2)
    } else if (load == "mixed") {
      SecondOrderLoadings <- rep(runif(1, .50, .99), 2)
    } else { #load == "half"
      SecondOrderLoadings <- rep(runif(1, .50, .99), 2)
    }
  } else {
    if (load == "high") {
      SecondOrderLoadings <- runif(nfac, .80, .99)
    } else if (load == "low") {
      SecondOrderLoadings <- runif(nfac, .50, .80)
    } else if (load == "mixed") {
      SecondOrderLoadings <- runif(nfac, .50, .99)
```

```
    } else { #load == "half"
      SecondOrderLoadings <- c(runif(nfac/2, .50, .80),
                               runif(nfac/2, .80, .99))
      if (nfac %% 2 == 1) {
        SecondOrderLoadings <- c(SecondOrderLoadings,
                                 runif(1, .5, .99))
      }
    }
  }

  FirstOrderScores <- GenerateFirstOrder(N, nfac,
                            SecondOrderLoadings, SecondOrderScores)

  # First order reliabilities according to the reliability
  # distribution type
  if (rel == "high") {
    FirstOrderReliability <- runif(nfac, .7, .99)
  } else if (rel == "mixed") {
    FirstOrderReliability <- runif(nfac, .5, .99)
  } else { # rel == "half"
    FirstOrderReliability <- c(runif(nfac/2, .7, .99),
                               runif(nfac/2, .5, .7))
    if (nfac %% 2 == 1) {
      FirstOrderReliability <- c(FirstOrderReliability,
                                 runif(1, .5, .99))
    }
  }

  ItemScores <- do.call(cbind,
                        lapply(1:nfac, function(x)
                           GenerateItems(N, FirstOrderReliability[x],
                           FirstOrderScores[,x], nitems)))

  return(list(
    SecondOrderScores     = SecondOrderScores,
    SecondOrderLoadings   = SecondOrderLoadings,
    FirstOrderScores      = FirstOrderScores,
    FirstOrderReliability = FirstOrderReliability,
    ItemScores            = ItemScores
  ))
}

# Schmid-Leiman Transformation for bifactor parameters
# from second-order parameters
SchmidLeimanTrans <- function (SimData, nfac) {

  # A function to do the actual transformation after
  # we do some wrangling
  SLT <- function (SecondOrderLoadings, FirstOrderLoadings) {
    U2 <- diag(sqrt(1-SecondOrderLoadings^2))
    B2 <- cbind(SecondOrderLoadings, U2)
    FirstOrderLoadings %*% B2
  }

  # Data wrangling to generate loadings matrices
  sload_matrix <- SimData$SecondOrderLoadings
  # This euqation solves the Omega equation for loading, basically
  floads <- sqrt(SimData$FirstOrderReliability/(nitems - (nitems-
                                  1)*SimData$FirstOrderReliability))

  # Create first order matrix. Vectorizing this would take
  # more time than it would save.
  for (i in 1:nfac) {
    if (i == 1) {
```

```
      load_vector <- rep(floads[i], nitems)
    } else {
      # fill in zeros, then next dimension of factor loadings
      load_vector <- c(load_vector, rep(0, nitems*nfac), rep(floads[i],
                                                             nitems))
    }
  }

  # Now matrix-ify it
  fload_matrix <- matrix(load_vector, ncol = nfac, byrow = FALSE)

  return ( SLT(sload_matrix, fload_matrix) )
}

# PRMSE Indices (theoretical)
PRMSE_Indices <- function(RawData, FirstOrderScores, nfac) {
  Tot <- rowSums(RawData)

  # Let's make subscores
  for (i in 1:nfac) {
    if (i == 1) {
      # First subscores
      Subscores <- rowSums(RawData[,1:nitems])
    } else {
      # add one dimension at a time
      m <- (i-1)*nitems + 1
      n <- i*nitems
      Subscores <- cbind(Subscores, rowSums(RawData[,m:n]))
    }
  }


  PRMSES <- sapply(1:ncol(Subscores), function(x)
                           cor(FirstOrderScores[,x], Subscores[,x])^2)
  PRMSET <- sapply(1:ncol(Subscores), function(x)
                           cor(FirstOrderScores[,x], Tot)^2)
  VAR    <- PRMSES/PRMSET

  return(do.call(cbind, list(PRMSES = PRMSES,
                             PRMSET = PRMSET,
                             VAR    = VAR)))
}

# Put it all together and do the simulation
sim_analysis <- function (c, N, nitems, numreps) {
  # set simulation conditions
  nfac <- conditions$numfactors[c]
  load <- conditions$loadings[c]
  rel  <- conditions$reliabilities[c]

  run_reps <- function(i, N, nfac, load, rel, nitems) {
    # Generate Data
    SimData <- GenerateData(N, nfac, load, rel, nitems)

    # Create bifactor model (S-L style)
    BiModel_SL <- SchmidLeimanTrans(SimData, nfac)

    # Bifactor Indices
    SL_bindices <- bifactorIndices(BiModel_SL)

    # PRMSE Indices
    PRMSE <- PRMSE_Indices(SimData$ItemScores,
                           SimData$FirstOrderScores, nfac)
```

```
    # Glue it all together into a vector. Ugly but easy to follow
    RepResultMat <- SL_bindices$FactorLevelIndices[-1,]
    RepResultMat <- cbind(rep(SL_bindices$FactorLevelIndices$ECV_SS[1],
                              nfac), RepResultMat)
    colnames(RepResultMat)[1] <- "ECVG"
    RepResultMat <- cbind(
             rep(SL_bindices$FactorLevelIndices$Omega_H[1], nfac),
             RepResultMat)
    colnames(RepResultMat)[1] <- "OmegaGH"
    RepResultMat <- cbind(
             rep(SL_bindices$FactorLevelIndices$Omega[1], nfac),
             RepResultMat)
    colnames(RepResultMat)[1] <- "OmegaG"
    RepResultMat <- cbind(PRMSE, RepResultMat)
    RepResultMat <- cbind(SimData$SecondOrderLoadings, RepResultMat)
    colnames(RepResultMat)[1] <- "SecondOrderLoadings"
    RepResultMat <- cbind(SimData$FirstOrderReliability, RepResultMat)
    colnames(RepResultMat)[1] <- "FirstOrderReliability"
    RepResultMat <- cbind(rep(rel, nfac), RepResultMat)
    colnames(RepResultMat)[1] <- "rel"
    RepResultMat <- cbind(rep(load, nfac), RepResultMat)
    colnames(RepResultMat)[1] <- "load"
    RepResultMat <- cbind(rep(nfac, nfac), RepResultMat)
    colnames(RepResultMat)[1] <- "nfac"
    RepResultMat <- cbind(rep((c-1)*numreps+i, nfac), RepResultMat)
    colnames(RepResultMat)[1] <- "IterationNumber"

    return(RepResultMat)

  }

  # Let's get replicating
  rep_results <- lapply (1:numreps, run_reps, N, nfac,
                                                 load, rel, nitems)
  rep_results <- as.data.frame(do.call(rbind, rep_results))
  return (rep_results)
}


# Let's get some constants out of the way
{
  N <- 100000
  nitems <- 5
  numreps <- 1000
  loadings <- c("high", "low", "mixed", "half")
  reliabilities <- c("high", "mixed", "half")
  numfactors <- c(2, 3, 4, 5, 6, 7, 8)
  conditions <- expand.grid(loadings, reliabilities, numfactors)
  colnames(conditions) <- c("loadings", "reliabilities", "numfactors")
  num_conditions <- nrow(conditions)
}

# set up clusters
{
  cl <- makeCluster(detectCores(logical = FALSE))
  clusterEvalQ(cl, library(lavaan))
  clusterEvalQ(cl, library(BifactorIndicesCalculator))
  clusterExport(cl, "conditions")
  clusterExport(cl, "GenerateData")
  clusterExport(cl, "SchmidLeimanTrans")
  clusterExport(cl, "PRMSE_Indices")
  clusterExport(cl, "N")
  clusterExport(cl, "nitems")
  clusterExport(cl, "numreps")
```

```
  clusterExport(cl, "sim_analysis")

  # Asked random.org for a random number between 1 and a billion
  clusterSetRNGStream(cl, iseed = 12495640)
  # This next one is for replication samples
  #clusterSetRNGStream(cl, iseed = 204059826)
}

# Now let's do the work!

system.time(results <- parLapply(cl, 1:num_conditions, sim_analysis,
                                                    N, nitems, numreps))
# It's a massive list, so let's make it into a dataframe
results <- as.data.frame(do.call(rbind, results))

stopCluster(cl)

write.csv(results, "DissResults.csv", row.names = FALSE)
# This next one is for replication samples
#write.csv(results, "DissResultsReplication.csv", row.names = FALSE)
```

APPENDIX B

Code for Analysis of Simulated Data Sets

```
# Source the Auxiliary Functions file first!!

SimResults <- read.csv("DissResults.csv")

# Which variables are most predictive?
summary(lm(VAR ~ Omega_H, SimResults))
summary(lm(VAR ~ Omega, SimResults))
summary(lm(VAR ~ ECV_SG, SimResults))
summary(lm(VAR ~ ECV_SS, SimResults))
summary(lm(VAR ~ nfac, SimResults))
summary(lm(VAR ~ OmegaG, SimResults))
summary(lm(VAR ~ OmegaGH, SimResults))
summary(lm(VAR ~ ECVG, SimResults))

# OmegaHS and ECV_SS are clearly the best.
# First, build a model starting from OmegaHS
# Then build a model starting from ECV_SS


# Starting from OmegaHS, build a model for VAR
OH_fit <- lm(VAR ~ Omega_H, SimResults)
summary(OH_fit) #R2 = .848
plot(x = SimResults$Omega_H,
     y = OH_fit$residuals,
     col = rgb(red = 0, green = 0, blue = 0, alpha = 0.03),
     xlab = "OmegaHS",
     ylab = "Residuals for VAR")
plot(x = SimResults$VAR,
     y = OH_fit$residuals,
     col = rgb(red = 0, green = 0, blue = 0, alpha = 0.03),
     xlab = "VAR",
     ylab = "Residuals for VAR")
SensSpec(OH_fit, SimResults)

# That heteroscedasticity is a little sketchy.
# Sensitivity is a bit low, too.
# Let's add another variable and see how it looks
# Try all of them and see which is best

summary(lm(VAR ~ Omega_H*Omega,   SimResults))
summary(lm(VAR ~ Omega_H*ECV_SG,  SimResults))
summary(lm(VAR ~ Omega_H*ECV_SS,  SimResults))
summary(lm(VAR ~ Omega_H*nfac,    SimResults))
summary(lm(VAR ~ Omega_H*OmegaG,  SimResults))
summary(lm(VAR ~ Omega_H*OmegaGH, SimResults))
summary(lm(VAR ~ Omega_H*ECVG,    SimResults))

# OmegaS, ECV_SG, and nfac are clearly the winners
# nfac should only be included if absolutely
# necessary because of complexity of interpretation

# Further analyses with ECV_SG are not included
# because OmegaS worked out better in the end

# Add OmegaS to the model and see where that takes me

OOH_fit <- lm(VAR ~ Omega_H*Omega, SimResults)
```

```
summary(OOH_fit) #R2 = .878
plot(x = SimResults$Omega_H,
     y = OOH_fit$residuals,
     col = rgb(red = 0, green = 0, blue = 0, alpha = 0.03),
     xlab = "OmegaHS",
     ylab = "Residuals for VAR")
plot(x = SimResults$Omega,
     y = OOH_fit$residuals,
     col = rgb(red = 0, green = 0, blue = 0, alpha = 0.03),
     xlab = "OmegaS",
     ylab = "Residuals for VAR")
plot(x = SimResults$VAR,
     y = OOH_fit$residuals,
     col = rgb(red = 0, green = 0, blue = 0, alpha = 0.03),
     xlab = "VAR",
     ylab = "Residuals for VAR")

# That heteroscedasticity is nasty looking
# Let's add a quadratic OmegaHS term and see if it helps

OOH2_fit <- lm(VAR ~ Omega_H*Omega + I(Omega_H^2), SimResults)
summary(OOH2_fit) #R2 = .909
plot(x = SimResults$Omega_H,
     y = OOH2_fit$residuals,
     col = rgb(red = 0, green = 0, blue = 0, alpha = 0.03),
     xlab = "OmegaHS",
     ylab = "Residuals for VAR")
plot(x = SimResults$Omega,
     y = OOH2_fit$residuals,
     col = rgb(red = 0, green = 0, blue = 0, alpha = 0.03),
     xlab = "OmegaS",
     ylab = "Residuals for VAR")
plot(x = SimResults$VAR,
     y = OOH2_fit$residuals,
     col = rgb(red = 0, green = 0, blue = 0, alpha = 0.03),
     xlab = "VAR",
     ylab = "Residuals for VAR")
SensSpec(OOH2_fit, SimResults)

# Residuals look a lot better. Some wonky heteroscedasticity
# with VAR. There are some datasets for which VAR is MAJORLY
# overestimated. This will be a problem for specificity.
# Sensitivity is ok, but specificity is a little low.
# Let's make some cutoffs

OmegaVals <- c(.5, .55, .6, .65, .7, .75, .8, .85, .9, .95)
CutoffGen_OmegaOmegaH2(OOH2_fit, OmegaVals)

# Let's see if we can improve things better by adding another variable

summary(lm(VAR ~ Omega_H*Omega*ECV_SG  + I(Omega_H^2),  SimResults))
summary(lm(VAR ~ Omega_H*Omega*ECV_SS  + I(Omega_H^2),  SimResults))
summary(lm(VAR ~ Omega_H*Omega*ECVG    + I(Omega_H^2),  SimResults))
summary(lm(VAR ~ Omega_H*Omega*OmegaG  + I(Omega_H^2),  SimResults))
summary(lm(VAR ~ Omega_H*Omega*OmegaGH + I(Omega_H^2),  SimResults))
summary(lm(VAR ~ Omega_H*Omega*nfac    + I(Omega_H^2),  SimResults))

# ECV_SG and nfac are the winners. Again, analyses with ECV_SG
# are not included here because they didn't turn out as well.
# Instead, I'll pursue nfac. Since nfac is categorical, I am going
# to fit models for each dimension separately.

# I want a table of cutoffs indexed by nfac and OmegaS
# with the overall numbers at the top
```

```
OOH2_cuts <- unlist(CutoffGen_OmegaOmegaH2(OOH2_fit, OmegaVals))[11:30]
OOH2_cuts <- matrix(OOH2_cuts, ncol = 20)
colnames(OOH2_cuts) <- c(paste0("V1_", OmegaVals),
                         paste0("V11_", OmegaVals))

# Now let's go through the dimensions one at a time...
OOH2_cuts <- OmegaS_Cuts_Increment(2, SimResults, OOH2_cuts)
OOH2_cuts <- OmegaS_Cuts_Increment(3, SimResults, OOH2_cuts)
OOH2_cuts <- OmegaS_Cuts_Increment(4, SimResults, OOH2_cuts)
OOH2_cuts <- OmegaS_Cuts_Increment(5, SimResults, OOH2_cuts)
OOH2_cuts <- OmegaS_Cuts_Increment(6, SimResults, OOH2_cuts)
OOH2_cuts <- OmegaS_Cuts_Increment(7, SimResults, OOH2_cuts)
OOH2_cuts <- OmegaS_Cuts_Increment(8, SimResults, OOH2_cuts)

# Merge them all together and write to file
OOH2_cuts <- cbind(c("Overall", "2fac", "3fac", "4fac", "5fac",
                                "6fac", "7fac", "8fac"), OOH2_cuts)
colnames(OOH2_cuts)[1] <- "nfac"
write.csv(OOH2_cuts, "OmegaOmegaHCutoffs.csv")

# Next make the plots of OmegaHS vs VAR for different levels of OmegaS
VarOmegaHSPlots(SimResults[SimResults$nfac == 2, ], 1, .1)
VarOmegaHSPlots(SimResults[SimResults$nfac == 2, ], 1.1, .1)
VarOmegaHSPlots(SimResults[SimResults$nfac == 3, ], 1, .085)
VarOmegaHSPlots(SimResults[SimResults$nfac == 3, ], 1.1, .085)
VarOmegaHSPlots(SimResults[SimResults$nfac == 4, ], 1, .085)
VarOmegaHSPlots(SimResults[SimResults$nfac == 4, ], 1.1, .085)
VarOmegaHSPlots(SimResults[SimResults$nfac == 5, ], 1, .07)
VarOmegaHSPlots(SimResults[SimResults$nfac == 5, ], 1.1, .07)
VarOmegaHSPlots(SimResults[SimResults$nfac == 6, ], 1, .04)
VarOmegaHSPlots(SimResults[SimResults$nfac == 6, ], 1.1, .04)
VarOmegaHSPlots(SimResults[SimResults$nfac == 7, ], 1, .04)
VarOmegaHSPlots(SimResults[SimResults$nfac == 7, ], 1.1, .04)
VarOmegaHSPlots(SimResults[SimResults$nfac == 8, ], 1, .03)
VarOmegaHSPlots(SimResults[SimResults$nfac == 8, ], 1.1, .03)


# Now let's do it all again with ECV_SS!!

ECV_fit <- lm(VAR ~ ECV_SS, SimResults)
summary(ECV_fit) #R2 = .736
plot(x = SimResults$ECV_SS,
     y = ECV_fit$residuals,
     col = rgb(red = 0, green = 0, blue = 0, alpha = 0.03),
     xlab = "ECV_SS",
     ylab = "Residuals for VAR")
plot(x = SimResults$VAR,
     y = ECV_fit$residuals,
     col = rgb(red = 0, green = 0, blue = 0, alpha = 0.03),
     xlab = "VAR",
     ylab = "Residuals for VAR")
SensSpec(ECV_fit, SimResults)

# That heteroscedasticity is a little sketchy.
# Sensitivity is very low, too.
# Let's add another variable and see how it looks
# Try all of them and see which is best

summary(lm(VAR ~ ECV_SS*Omega_H, SimResults))
summary(lm(VAR ~ ECV_SS*Omega,   SimResults))
summary(lm(VAR ~ ECV_SS*ECV_SG,  SimResults))
summary(lm(VAR ~ ECV_SS*nfac,    SimResults))
summary(lm(VAR ~ ECV_SS*OmegaG,  SimResults))
```

```
summary(lm(VAR ~ ECV_SS*OmegaGH, SimResults))
summary(lm(VAR ~ ECV_SS*ECVG,    SimResults))

# OmegaS and nfac are the winners
# nfac should only be included if absolutely
# necessary because of complexity of interpretation

# Add OmegaS to the model and see where that takes me
oecv_fit <- lm(VAR ~ Omega*ECV_SS, SimResults)
summary(oecv_fit) #R2 = .875
plot(x = SimResults$ECV_SS,
     y = oecv_fit$residuals,
     col = rgb(red = 0, green = 0, blue = 0, alpha = 0.03),
     xlab = "ECV_SS",
     ylab = "Residuals for VAR")
plot(x = SimResults$Omega,
     y = oecv_fit$residuals,
     col = rgb(red = 0, green = 0, blue = 0, alpha = 0.03),
     xlab = "OmegaS",
     ylab = "Residuals for VAR")
plot(x = SimResults$VAR,
     y = oecv_fit$residuals,
     col = rgb(red = 0, green = 0, blue = 0, alpha = 0.03),
     xlab = "VAR",
     ylab = "Residuals for VAR")

# That heteroscedasticity is nasty looking
# Let's add a quadratic ECV_SS term and see if it helps

oecv2_int_fit <- lm(VAR ~ Omega*ECV_SS + I(ECV_SS^2), SimResults)
summary(oecv2_int_fit)   #R2 = .911

# That coefficient for the interaction term is TINY
# Let's see if it matters
oecv2_fit <- lm(VAR ~ Omega + ECV_SS + I(ECV_SS^2), SimResults)
summary(oecv2_fit)   #R2 = .911

# The interaction really doesn't matter, so we will omit it
plot(x = SimResults$ECV_SS,
     y = oecv2_fit$residuals,
     col = rgb(red = 0, green = 0, blue = 0, alpha = 0.03),
     xlab = "ECV_SS",
     ylab = "Residuals for VAR")
plot(x = SimResults$Omega,
     y = oecv2_fit$residuals,
     col = rgb(red = 0, green = 0, blue = 0, alpha = 0.03),
     xlab = "OmegaS",
     ylab = "Residuals for VAR")
plot(x = SimResults$VAR,
     y = oecv2_fit$residuals,
     col = rgb(red = 0, green = 0, blue = 0, alpha = 0.03),
     xlab = "VAR",
     ylab = "Residuals for VAR")
SensSpec(oecv2_fit, SimResults)

# Reiduals look much better, but some wonky heteroscedasticity
# with VAR. There are some datasets for which VAR is MAJORLY
# overestimated. This will be a problem for specificity.
# Sensitivity and specificity are ok.
# Let's make some cutoffs

ECVals <- c(.05, .10, .15, .20, .25, .30, .35, .40, .45, .50)
CutoffGen_ECV2Omega(oecv2_fit, ECVals)
```

```
# Let's see if we can improve things better by adding another variable
summary(lm(VAR ~ (Omega + ECV_SS)*Omega_H + I(ECV_SS^2), SimResults))
summary(lm(VAR ~ (Omega + ECV_SS)*ECV_SG  + I(ECV_SS^2), SimResults))
summary(lm(VAR ~ (Omega + ECV_SS)*OmegaG  + I(ECV_SS^2), SimResults))
summary(lm(VAR ~ (Omega + ECV_SS)*OmegaGH + I(ECV_SS^2), SimResults))
summary(lm(VAR ~ (Omega + ECV_SS)*ECVG    + I(ECV_SS^2), SimResults))
summary(lm(VAR ~ (Omega + ECV_SS)*nfac    + I(ECV_SS^2), SimResults))

# nfac is the winner here. Since nfac is categorical, I am going
# to fit models for each dimension separately.

# I want a table of cutoffs indexed by nfac and OmegaS
# with the overall numbers at the top
ECV2O_cuts <- unlist(CutoffGen_ECV2Omega(oecv2_fit, ECVals))[11:30]
ECV2O_cuts <- matrix(ECV2O_cuts, ncol = 20)
colnames(ECV2O_cuts) <- c(paste0("V1_",  ECVals),
                          paste0("V11_", ECVals))

# Now let's go through the dimensions one at a time...
ECV2O_cuts <- ECV_Cuts_Increment(2, SimResults, ECV2O_cuts) #R2 = .891
ECV2O_cuts <- ECV_Cuts_Increment(3, SimResults, ECV2O_cuts) #R2 = .867
ECV2O_cuts <- ECV_Cuts_Increment(4, SimResults, ECV2O_cuts) #R2 = .941
ECV2O_cuts <- ECV_Cuts_Increment(5, SimResults, ECV2O_cuts) #R2 = .967
ECV2O_cuts <- ECV_Cuts_Increment(6, SimResults, ECV2O_cuts) #R2 = .978
ECV2O_cuts <- ECV_Cuts_Increment(7, SimResults, ECV2O_cuts) #R2 = .983
ECV2O_cuts <- ECV_Cuts_Increment(8, SimResults, ECV2O_cuts) #R2 = .985

# Let's write these cutofs to file
write.csv(ECV2O_cuts, "OmegaECVCutoffs.csv")


# Next make the plots of OmegaS vs VAR for different levels of ECV_SS
VarOmegaPlots(SimResults[SimResults$nfac == 2, ], 1, .1)
VarOmegaPlots(SimResults[SimResults$nfac == 2, ], 1.1, .1)
VarOmegaPlots(SimResults[SimResults$nfac == 3, ], 1, .085)
VarOmegaPlots(SimResults[SimResults$nfac == 3, ], 1.1, .085)
VarOmegaPlots(SimResults[SimResults$nfac == 4, ], 1, .085)
VarOmegaPlots(SimResults[SimResults$nfac == 4, ], 1.1, .085)
VarOmegaPlots(SimResults[SimResults$nfac == 5, ], 1, .07)
VarOmegaPlots(SimResults[SimResults$nfac == 5, ], 1.1, .07)
VarOmegaPlots(SimResults[SimResults$nfac == 6, ], 1, .04)
VarOmegaPlots(SimResults[SimResults$nfac == 6, ], 1.1, .04)
VarOmegaPlots(SimResults[SimResults$nfac == 7, ], 1, .04)
VarOmegaPlots(SimResults[SimResults$nfac == 7, ], 1.1, .04)
VarOmegaPlots(SimResults[SimResults$nfac == 8, ], 1, .03)
VarOmegaPlots(SimResults[SimResults$nfac == 8, ], 1.1, .03)
```

APPENDIX C

Code for Auxiliary Functions Used in Analysis of Simulated Data

```
# source this file before doing analyses

options(digits=5)

# computes sensitivity and specificity for a given regression
# model of VAR
SensSpec <- function(fit, data) {
  PredVar <- predict.lm(fit)

  a <- sum(PredVar > 1.0 & data$VAR > 1.0) / sum(data$VAR > 1.0)
  b <- sum(PredVar < 1.0 & data$VAR < 1.0) / sum(data$VAR < 1.0)
  c <- sum(PredVar > 1.1 & data$VAR > 1.1) / sum(data$VAR > 1.1)
  d <- sum(PredVar < 1.1 & data$VAR < 1.1) / sum(data$VAR < 1.1)

  sens_spec <- matrix(c(1, a, b, 1.1, c, d), nrow = 2, byrow = TRUE)
  sens_spec <- as.data.frame(sens_spec)
  colnames(sens_spec) <- c("VAR", "SENS", "SPEC")

  sens_spec
}


# Creates OmegaHS cutoffs for given OmegaS values
# based on a model with OmegaHS, OmegaS, their
# interaction, and OmegaHS^2 as predictors
CutoffGen_OmegaOmegaH2 <- function(fit, OmegaVals) {
  QF <- function (a, b, c) {
    suppressWarnings((-b + sqrt(b*b-4*a*c))/(2*a))
  }

  int <- fit$coefficients[1]
  b_omegah <- fit$coefficients["Omega_H"]
  b_omega <- fit$coefficients["Omega"]
  b_omegah2 <- fit$coefficients["I(Omega_H^2)"]
  b_interact <- fit$coefficients["Omega_H:Omega"]

  a <- b_omegah2
  b <- b_omegah + b_interact*OmegaVals
  c <- int + b_omega*OmegaVals

  Var1 <- QF(a, b, c-1)
  Var11 <- QF(a, b, c-1.1)

  res <- data.frame(Omega = OmegaVals, Var1 = Var1, Var11 = Var11)
  res[apply(res,2,is.nan)] <- 0

  return(res)
}

# Create OmegaHS cutoffs using a model with OmegaHS
# and OmegaHS as predictors. Used for models with
# (nearly) constant OmegaS
CutoffGen_OmegaH2 <- function(fit) {
```

```
  QF <- function (a, b, c) {
    suppressWarnings((-b + sqrt(b*b-4*a*c))/(2*a))
  }

  int <- fit$coefficients[1]
  b_omegah <- fit$coefficients["Omega_H"]
  b_omegah2 <- fit$coefficients["I(Omega_H^2)"]

  a <- b_omegah2
  b <- b_omegah
  c <- int

  Var1 <- QF(a, b, c-1)
  Var11 <- QF(a, b, c-1.1)

  res <- data.frame(Var1 = Var1, Var11 = Var11)
  res[apply(res,2,is.nan)] <- 0

  return(res)
}

# Generates cutoffs for OmegaHS based on OmegaS
# for a given number of specific factors
# and adds them on to OOH2_cuts
OmegaS_Cuts_Increment <- function(nfac, SimResults, OOH2_cuts) {
  OOH2_fit_n <- lm(VAR ~ Omega_H*Omega + I(Omega_H^2),
                                  SimResults[SimResults$nfac == nfac, ])
  print(summary(OOH2_fit_n))
  # Sensitivity and Specificity
  print(SensSpec(OOH2_fit_n, SimResults[SimResults$nfac == nfac, ]))
  # cutoffs
  OmegaS_cuts_n <- CutoffGen_OmegaOmegaH2(OOH2_fit_n, OmegaVals)
  OOH2_cuts <- rbind(OOH2_cuts, unlist(OmegaS_cuts_n)[11:30])
  OOH2_cuts
}


# Generates matrix of plots for OmegaHS vs VAR
VarOmegaHSPlots <- function(data, VAR, alpha) {
  par(mfrow=c(4,2))
  minOs <- .5+.05*1:8
  for (x in minOs) {
    minOmega <- x
    maxOmega <- x + .05
    temp_data <- data[data$Omega > minOmega & data$Omega < maxOmega,]
    func_mod <- lm(VAR ~ Omega_H + I(Omega_H^2), temp_data)
    xvals <- seq(0, .6, len = 1000)
    yvals <- predict.lm(func_mod, newdata = data.frame(Omega_H = xvals))
    if (VAR == 1) {
      plot(temp_data[,"Omega_H"],
           temp_data[,"VAR"],
           col = rgb(red = 0, green = 0, blue = 0, alpha = alpha),
           main = paste0(minOmega, " < OmegaS < ", maxOmega),
           xlim = c(0, .6),
           ylim = c(0.5, 2.0),
           yaxs = "i", xaxs = "i",
           xlab = "OmegaHS",
           ylab = "VAR")
      abline(h = 1, lwd = 1)
      abline(v = CutoffGen_OmegaH2(func_mod)[1], lwd = 1)
```

```
      lines(xvals, yvals, lwd = 2)
    } else {
      plot(temp_data[,"Omega_H"],
           temp_data[,"VAR"],
           col = rgb(red = 0, green = 0, blue = 0, alpha = alpha),
           main = paste0(minOmega, " < OmegaS < ", maxOmega),
           xlim = c(0, .6),
           ylim = c(0.5, 2.0),
           yaxs = "i", xaxs = "i",
           xlab = "OmegaHS",
           ylab = "VAR")
      abline(h = 1.1, lwd = 1)
      abline(v = CutoffGen_OmegaH2(func_mod)[2], lwd = 1)
      lines(xvals, yvals, lwd = 2)
    }
  }
}


# Creates OmegaS cutoffs for given ECV_SS values
# based on a model with ECV_SS, OmegaS, and ECV_SS^2
# as predictors
CutoffGen_ECV2Omega <- function(fit, ECV_Vals) {
  int <- fit$coefficients[1]
  b_omega <- fit$coefficients["Omega"]
  b_ECV <- fit$coefficients["ECV_SS"]
  b_ECV2 <- fit$coefficients["I(ECV_SS^2)"]

  Var1 <- (1 - int - b_ECV*ECV_Vals - b_ECV2*ECV_Vals^2) / (b_omega)
  Var11 <- (1.1 - int - b_ECV*ECV_Vals - b_ECV2*ECV_Vals^2) / (b_omega)

  return(data.frame(ECV = ECV_Vals, Var1 = Var1, Var11 = Var11))
}

# Create OmegaS cutoffs using a model with
# OmegaS as the only predictor. Used when
# ECV_SS is (mostly) constant
CutoffGen_Omega <- function(fit) {
  int <- fit$coefficients[1]
  b_omega <- fit$coefficients["Omega"]


  Var1  <- (1   - int) / b_omega
  Var11 <- (1.1 - int) / b_omega

  return(data.frame(Var1 = Var1, Var11 = Var11))
}

# Generates cutoffs for OmegaS based on ECV
# for a given number of specific factors
# and adds them on to ECV2O_cuts
ECV_Cuts_Increment <- function(nfac, SimResults, ECV2O_cuts) {
  ECV2O_fit <- lm(VAR ~ ECV_SS + Omega + I(ECV_SS^2),
                                SimResults[SimResults$nfac == nfac, ])
  print(summary(ECV2O_fit))
  # Sensitivity and Specificity
  print(SensSpec(ECV2O_fit, SimResults[SimResults$nfac == nfac, ]))
  # cutoffs
  ECV2O_cuts_n <- CutoffGen_ECV2Omega(ECV2O_fit, ECVals)
  ECV2O_cuts <- rbind(ECV2O_cuts, unlist(ECV2O_cuts_n)[11:30])
```

```
  ECV2O_cuts
}


# Generates matrix of plots for OmegaS vs VAR
VarOmegaPlots <- function(data, VAR, alpha) {
  par(mfrow=c(5,2))
  minECVs <- .05*1:10
  for (x in minECVs) {
    minECV <- x
    maxECV <- x + .05
    temp_data <- data[data$ECV_SS > minECV & data$ECV_SS < maxECV,]
    func_mod <- lm(VAR ~ Omega, temp_data)
    xvals <- seq(0.5, .95, len = 1000)
    yvals <- predict.lm(func_mod, newdata = data.frame(Omega = xvals))
    if (VAR == 1) {
      plot(temp_data[,"Omega"],
           temp_data[,"VAR"],
           col = rgb(red = 0, green = 0, blue = 0, alpha = alpha),
           main = paste0(minECV, " < ECV_SS < ", maxECV),
           xlim = c(.5, 1.00),
           ylim = c(0.5, 1.7),
           yaxs = "i", xaxs = "i",
           xlab = "OmegaS",
           ylab = "VAR")
      abline(h = 1, lwd = 1)
      abline(v = CutoffGen_Omega(func_mod)[1], lwd = 1)
      lines(xvals, yvals, lwd = 2)
    } else { # VAR = 1.1
      plot(temp_data[,"Omega"],
           temp_data[,"VAR"],
           col = rgb(red = 0, green = 0, blue = 0, alpha = alpha),
           main = paste0(minECV, " < ECV_SS < ", maxECV),
           xlim = c(0.5, 1.00),
           ylim = c(0.5, 1.7),
           yaxs = "i", xaxs = "i",
           xlab = "OmegaS",
           ylab = "VAR")
      abline(h = 1.1, lwd = 1)
      abline(v = CutoffGen_Omega(func_mod)[2], lwd = 1)
      lines(xvals, yvals, lwd = 2)
    }
  }
}
```

REFERENCES

Anonymous. (2019). SQLI data from adolescent idiopathic scoliosis patients. Received via e-mail from anonymous source.

Asher, M. A., Lai, S. M., Glattes, R. C., Burton, D. C., Alanay, A., & Bago, J. (2006). Refinement of the SRS-22 health-related quality of life questionnaire function domain. *Spine, 31*(5), 593-597. https://doi.org/10.1097/01.brs.0000201331.50597.ea

Bentler, P. M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika, 74*(1), 137-143. https://doi.org/10.1007/s11336-008-9100-1

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110*(2), 305-314.

Bonifay, W. E., Reise, S. P., Scheines, R., & Meijer, R. R. (2015). When are multidimensional data unidimensional enough for structural equation modeling? An evaluation of the DETECT multidimensionality index. *Structural Equation Modeling: A Multidisciplinary Journal, 22*(4), 504-516. https://doi.org/10.1080/10705511.2014.938596

Brennan, R. L. (2012). *Utility indexes for decisions about subscores*. Center for Advanced Studies in Measurement and Assessment.

Byrne, B. M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema, 20*(4), 872-882.

Cai, L. (2010). A two-tier full-information item factor analysis model with applications. *Psychometrika, 75*(4), 581-612. https://doi.org/10.1007/s11336-010-9178-0

Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J. P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality, 80*(1), 219-251. https://doi.org/ 10.1111/j.1467-6494.2011.00739.x

Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, *41*(2), 189-225. https://doi.org/10.1207/s15327906mbr4102_5

Cucina, J., & Byle, K. (2017). The bifactor model fits better than the higher-order model in more than 90% of comparisons for mental abilities test batteries. *Journal of Intelligence, 5*(3), 27. https://doi.org/10.3390/jintelligence5030027

Dagnall, N., Denovan, A., Parker, A., Drinkwater, K., & Walsh, R. S. (2018). Confirmatory factor analysis of the inventory of personality organization-reality testing subscale. *Frontiers in Psychology, 9*, 1116. https://doi.org/ 10.3389/fpsyg.2018.01116

Dai, S., Wang, X., & Svetina, D. (2019) *subscore: Computing subscores in classical test theory and item response theory*. R package version 3.1.

DeMars, C. E. (2013). A tutorial on interpreting bifactor model scores. *International Journal of Testing, 13*(4), 354-378. https://doi.org/10.1080/15305058.2013.799067

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2018). An alternative conceptualization of the theoretical structure of the Woodcock-Johnson IV Tests of Cognitive Abilities at school age: A confirmatory factor analytic investigation. *Archives of Scientific Psychology, 6*(1), 1-13. https://doi.org/10.1037/arc0000039

Drinkwater, K., Denovan, A., Dagnall, N., & Parker, A. (2018). The Australian sheep-goat scale: an evaluation of factor structure and convergent validity. *Frontiers in Psychology, 9*, 1594. https://doi.org/10.3389/fpsyg.2018.01594

Dueber, D. M. (2017). Bifactor Indices Calculator: A Microsoft Excel-based tool to calculate various indices relevant to bifactor CFA models. https://doi.org/10.13023/edp.tool.01

Dueber, D. M. (2019) *BifactorIndicesCalculator: Bifactor indices calculator*. R package version 0.1.0.

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods, 22*(3), 541-562. https://doi.org/10.1037/met0000083

Feinberg, R. A., & Jurich, D. P. (2017). Guidelines for interpreting and reporting subscores. *Educational Measurement: Issues and Practice*, *36*(1), 5-13. https://doi.org/ 10.1111/emip.12142

Feinberg, R. A., & Rubright, J. D. (2016). Conducting simulation studies in psychometrics. *Educational Measurement: Issues and Practice, 35*(2), 36-49. https://doi.org/10.1111/emip.12111

Feinberg, R. A., & Wainer, H. (2014). A simple equation to predict a subscore's value. *Educational Measurement: Issues and Practice*, *33*(3), 55-56. https://doi.org/10.1111/emip.12035

Feise, R. J., Donaldson, S., Crowther, E. R., Menke, J. M., & Wright, J. G. (2005). Construction and validation of the scoliosis quality of life index in adolescent idiopathic scoliosis. *Spine, 30*(11), 1310-1315. https://doi.org/10.1097/01.brs.0000163885.12834.ca

Fenollar-Cortés, J., López-Pinar, C., & Watkins, M. W. (2019). Structural validity of the Spanish Wechsler Intelligence Scale for Children–Fourth Edition in a large sample of Spanish children with attention-deficit hyperactivity disorder. *International Journal of School & Educational Psychology, 7*(sup1), 2-14. https://doi.org/10.1080/21683603.2018.1474820

Ferrando, P. J., & Lorenzo-Seva, U. (2019). On the added value of multiple factor score estimates in essentially unidimensional models. *Educational and Psychological Measurement, 79*(2), 249-271. https://doi.org/10.1177/0013164418773851

Finch, W. H., & French, B. F. (2015). *Latent variable modeling with R*. Routledge.

Folberg, A. M., Kercher, K., & Ryan, C. S. (2019). *Sex Roles*. Advanced online publication. https://doi.org/10.1007/s11199-019-01104-1

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*(3), 286-299. https://doi.org/10.1037/1040-3590.7.3.286

Gignac, G. E., & Kretzschmar, A. (2017). Evaluating dimensional distinctness with correlated-factor models: Limitations and suggestions. *Intelligence, 62*, 138-147. https://doi.org/10.1016/j.intell.2017.04.001

Gignac, G. E., & Watkins, M. W. (2013). Bifactor modeling and the estimation of model-based reliability in the WAIS-IV. *Multivariate Behavioral Research, 48*(5), 639-662. https://doi.org/10.1080/00273171.2013.804398

Gonzalez, O., & MacKinnon, D. P. (2018). A bifactor approach to model multifaceted constructs in statistical mediation analysis. *Educational and Psychological Measurement, 78*(1), 5-31. https://doi.org/10.1177/0013164416673689

Green, S., Xu, Y., & Thompson, M. S. (2018). Relative accuracy of two modified parallel analysis methods that use the proper reference distribution. *Educational and Psychological Measurement, 78*(4), 589-604. https://doi.org/10.1177/0013164417718610

Greene, A. L., Eaton, N. R., Li, K., Forbes, M. K., Krueger, R. F., Markon, K. E., Markon, K., Waldman, I., Cicero, D. C., Conway, C. C., Docherty, A. R., Fried, E. I., Ivanova, M. Y., Jonas, K. G., Latzan, R. D., Patrick, C. J., Reininghaus, U., Tackett, J. L., Wright, A. G. C., & Kotov, R. (2019). Are fit indices used to test psychopathology structure biased? A simulation study. *Journal of Abnormal Psychology, 128*(7), 740. https://doi.org/10.1037/abn0000434

Gustafsson, J.E., & Åberg-Bengtsson, L. (2010). Unidimensionality and the interpretability of psychological instruments. In S.E. Embretson (Ed.), *Measuring psychological constructs* (pp. 97–121). Washington, DC: American Psychological Association.

Haberman, S. J. (2005). *When can subscores have value?* (ETS Research Report RR-05-08). Princeton, NJ: Educational Testing Service.

Haberman, S. J. (2008). *Subscores and validity* (ETS Research Report RR-08–64). Princeton, NJ: Educational Testing Service.

Haberman, S. J., & Sinharay, S. (2010). Reporting of subscores using multidimensional item response theory. *Psychometrika, 75*(2), 209-227. https://doi.org/10.1007/s11336-010-9158-4

Hammer, J. H., McDermott, R. C., Levant, R. F., & McKelvey, D. K. (2018). Dimensionality, reliability, and validity of the Gender-Role Conflict Scale–Short Form (GRCS-SF). *Psychology of Men & Masculinity, 19*(4), 570. https://doi.org/10.1037/men0000131

Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika, 2*(1), 41-54. https://doi.org/10.1007/BF02287965

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6(*1), 1-55. https://doi.org/ 10.1080/ 10705519909540118

Hukkelberg, S. S., & Ogden, T. (2018). Dimensionality of oppositional defiant disorder. *Child and Adolescent Mental Health, 23*(2), 121-129. https://doi.org/ 10.1111/camh.12248

Hull, J. G., Lehn, D. A., & Tedlie, J. C. (1991). A general approach to testing multifaceted personality constructs. *Journal of Personality and Social Psychology, 61*(6), 932–945.https://doi.org/ 10.1037/0022-3514.61.6.932

Isoard-Gautheur, S., Martinent, G., Guillet-Descas, E., Trouilloud, D., Cece, V., & Mette, A. (2018). Development and evaluation of the psychometric properties of a new measure of athlete burnout: The Athlete Burnout Scale. *International Journal of Stress Management, 25*(S1), 108. https://doi.org/10.1037/str0000083

Jennrich, R. I., & Bentler, P. M. (2011). Exploratory bi-factor analysis. *Psychometrika, 76*(4), 537-549. https://doi.org/10.1007/s11336-011-9218-4

Kelley, K., & Pornprasertmanit, S. (2016). Confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. *Psychological Methods, 21*(1), 69-92. https://doi.org/10.1037/a0040086

Keum, B. T., Miller, M. J., Lee, M., & Chen, G. A. (2018). Color-Blind Racial Attitudes Scale for Asian Americans: Testing the factor structure and measurement invariance across generational status. *Asian American Journal of Psychology, 9*(2), 149-157. https://doi.org/10.1037/aap0000100

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Ling, G. (2012). Why the major field test in business does not report subscores: Reliability and construct validity evidence. *ETS Research Report Series, 2012*(1), 1-26. https://doi.org/10.1002/j.2333-8504.2012.tb02293.x

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130. https://doi.org/ 10.1037/1082-989X.1.2.130

Marshall, A. J., Evanovich, E. K., David, S. J., & Mumma, G. H. (2018). Separating Common from Unique Variance Within Emotional Distress: An Examination of Reliability and Relations to Worry. *Behavioural and Cognitive Psychotherapy, 46(*5), 633-638. https://doi.org/10.1017/S1352465817000777

Maslach, C., & Jackson, S. E. (1981). The measurement of experienced burnout. *Journal of Organizational Behavior, 2*(2), 99-113. https://doi.org/10.1002/job.4030020205

McDonald, R. P. (1999). *Test theory: A unified approach*. Erlbaum.

McGill, R. J., & Dombrowski, S. C. (2018). What does the WRAML2 core battery measure? Utilizing exploratory and confirmatory techniques to disclose higher order structure. *Assessment, 25*(6), 729-743. http://doi.org/ 10.1177/ 1073191116677799

McKay, M. T., Morgan, G. B., van Exel, N. J., & Worrell, F. C. (2015). Back to "the future": Evidence of a bifactor solution for scores on the Consideration of Future Consequences Scale. *Journal of Personality Assessment, 97*, 395–402. http://doi.org/10.1080/00223891.2014.999338.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*(3), 412-433. https://doi.org/10.1037/met0000144

Mészáros, V., Ádám, S., Szabó, M., Szigeti, R., & Urbán, R. (2014). The Bifactor Model of the Maslach Burnout Inventory–Human Services Survey (MBI-HSS)—An Alternative Measurement Model of Burnout. *Stress and Health, 30*(1), 82-88. https://doi.org/10.1002/smi.2481

Monaghan, W. (2006). *The facts about subscores* (Report No. RDC-04). Educational Testing Service.

Morgan, G. B., Hodge, K. J., Wells, K. E., & Watkins, M. W. (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research?: A comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations. *Journal of Intelligence, 3*(1), 2-20. https://doi.org/10.3390/jintelligence3010002

Morin, A. J., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(1), 116-139. https://doi.org/10.1080/10705511. 2014.961800

Muthén, L. K., & Muthén, B. O. (1998-2019). *Mplus user's guide* (8th ed.). Muthén & Muthén.

Murray, A. L., Booth, T., Eisner, M., Obsuth, I., & Ribeaud, D. (2019). Quantifying the strength of general factors in psychopathology: A comparison of CFA with maximum likelihood estimation, BSEM, and ESEM/EFA bifactor approaches. *Journal of Personality Assessment, 101*(6), 631-643. https://doi.org/10.1080/ 00223891.2018.1468338

Naser, S., Hitti, A., & Overstreet, S. (2018). The Behavioral and Emotional Screening System Student Form: Is There Evidence of a Global At-Risk Factor in a Sample of Predominantly African American Youth? *Journal of Psychoeducational Assessment, 36*(5), 446-463. https://doi.org/10.1177/0734282916686015

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Perreira, T. A., Morin, A. J., Hebert, M., Gillet, N., Houle, S. A., & Berta, W. (2018). The short form of the Workplace Affective Commitment Multidimensional Questionnaire (WACMQ-S): A bifactor-ESEM approach among healthcare

professionals. *Journal of Vocational Behavior, 106*, 62-83. https://doi.org/10.1016/j.jvb.2017.12.004

Quinn, H. O. C. (2014). *Bifactor models, explained common variance (ECV), and the usefulness of scores from unidimensional item response theory analyses* (Unpublished master's thesis). University of North Carolina.

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Raykov, T. (1997). Scale reliability, Cronbach's coefficient alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research, 32*(4), 329-353. https://doi.org/10.1207/s15327906mbr3204_2

Raykov, T., & Marcoulides, G. A. (2011). *Introduction to psychometric theory*. Routledge.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667-696. https://doi.org/10.1080/00273171.2012.715555

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013a). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, *95*(2), 129-140. https://doi.org/10.1080/00223891.2012.725437

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*(6), 544-559. https://doi.org/ 10.1080/00223891.2010.496477

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research, 16*, 19–31. https://doi.org/10.1080/00223891.2010.496477

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013b). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement, 73*, 5–26. https://doi.org/10.1177/0013164412449831

Revelle, W. (2020). *psych*. R package version 1.9.12.31.

Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, *74*(1), 145. https://doi.org/10.1007/s11336-008-9102-z

Rijmen, F. (2011). Hierarchical factor item response theory models for PIRLS: Capturing clustering effects at multiple levels. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments, 4*, 59-74.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016a). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*(3), 223-237. https://doi.org/10.1080/00223891.2015.1089249

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016b). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137-150. https://doi.org/10.1037/met0000045

Rosseel, Y., & Jorgensen, T. (2019) *lavaan: Latent variable analysis*. R package version 0.6-5.

Rowe, D. E., Feise, R. J., Crowther, E. R., Grod, J. P., Menke, J. M., Goldsmith, C. H., Stoline, M. R., Souza, T. A., & Kambach, B. (2006). Chiropractic manipulation in adolescent idiopathic scoliosis: a pilot study. *Chiropractic & Osteopathy, 14*(1), 15. https://doi.org/10.1186/1746-1340-14-15

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22*(1), 53-61. https://doi.org/10.1007/BF02289209

Sellbom, M., & Tellegen, A. (2019). Factor analysis in psychological assessment research: Common pitfalls and recommendations. *Psychological Assessment, 31*(2), 1428-1441. https://doi.org/10.1037/pas0000623

Sinharay, S. (2011). How often do subscores have added value? Results from operational and simulated data. *Journal of Educational Measurement*, *47*(2), 150-174. https://doi.org/10.1111/j.1745-3984.2010.00106.x

Sinharay, S. (2019). Added value of subscores and hypothesis testing. *Journal of Educational and Behavioral Statistics, 44*(1), 25-44. https://doi.org/10.3102/1076998618788862

Sinharay, S., Haberman, S., & Puhan, G. (2007). Subscores based on classical test theory: To report or not to report. *Educational Measurement: Issues and Practice*, *26*(4), 21-28. https://doi.org/10.1111/j.1745-3992.2007.00105.x

Sinharay, S., Puhan, G., & Haberman, S. J. (2011). An NCME instructional module on subscores. *Educational Measurement: Issues and Practice*, *30*(3), 29-40. https://doi.org/10.1111/j.1745-3992.2011.00208.x

Shihata, S., McEvoy, P. M., & Mullan, B. A. (2018). A bifactor model of intolerance of uncertainty in undergraduate and clinical samples: Do we need to reconsider the two-factor model? *Psychological Assessment, 30*(7), 893-903. https://doi.org/10.1037/pas0000540

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*, 107–120. https://doi.org/10.1007/s11336-008-9101-0

Smith, R. B. (1977). Proportional reduction in error interpretations for the squared generalized multiple, partial, and multiple-partial correlation coefficients and their special Ccses. *Social Forces, 56*(2), 688-702. https://doi.org/10.1093/sf/56.2.688

Stanton, K., Forbes, M. K., & Zimmerman, M. (2018). Distinct dimensions defining the Adult ADHD Self-Report Scale: Implications for assessing inattentive and hyperactive/impulsive symptoms. *Psychological Assessment, 30*(12), 1549-1559. https://doi.org/10.1037/pas0000604

Stucky, B. D., & Edelen, M. O. (2015). Using hierarchical IRT models to create unidimensional measures from multidimensional data. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of item response theory modeling: Applications to typical performance assessment* (pp. 183–206). Routledge/Taylor & Francis Group.

Stucky, B. D., Thissen, D., & Orlando Edelen, M. (2013). Using logistic approximations of marginal trace lines to develop short assessments. *Applied Psychological Measurement, 37*(1), 41-57. https://doi.org/10.1177/0146621612462759

Ten Berge, J. M., & Sočan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika*, *69*(4), 613-625. https://doi.org/10.1007/BF02289858

Thompson, J. K., Schaefer, L. M., & Dedrick, R. F. (2018). On the measurement of thin-ideal internalization: Implications for interpretation of risk factors and treatment outcome in eating disorders research. *International Journal of Eating Disorders, 51*(4), 363-367. https://doi.org/10.1002/eat.22839

Watkins, M. W. (2013). Omega [computer software]. *Ed & Psych Associates*.

Wedman, J., & Lyrén, P. E. (2015). Methods for examining the psychometric quality of subscores: A review and application. *Practical Assessment, Research & Evaluation*, *20*, 21. https://doi.org/10.7275/ng3q-0d19

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling*, (pp. 209-231). Guilford Press.

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. https://doi.org/10.1007/BF02294531

Zinbarg, R. E., Revelle, W., Yovel, I., & Li, W. (2005). Cronbach's α, Revelle's β, and McDonald's ωH: Their relations with each other and two alternative concepttualizations of reliability. *Psychometrika, 70*(1), 123-133. https://doi.org/10.1007/s11336-003-0974-7

VITA

David Michael Dueber

Education:

May 2006 Master of Arts, Mathematics, University of Wisconsin, Madison

May 2003 Bachelor of Science, Mathematics and Chemical Engineering, University of Illinois, Urbana-Champaign

Publications:

Toland, M. D., Bowen, P., & Dueber, D. M. (In press). Psychometric analysis of work-related, after-hours contact and associated psychosocial measurement scales for construction professionals in South Africa. *Journal of Construction Engineering and Management*. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001845

Thomas, J. N. & Dueber, D. M. (2020). Thinking with our hands: The power of gesture in mathematical moments. *Mathematics Teacher: Learning and Teaching PK-12, 113*(1), 69-73. https://doi.org/ 10.5951/MTLT.2019.0094

Thomas, J. N., Dueber, D. M., Fisher, M. H., Jong, C., & Schack, E. O. (2019). Professional noticing into practice: An examination of inservice teachers' conceptions and enactment. *Investigations in Mathematics Learning*. Advance online publication. https://doi.org/ 10.1080/19477503.2019.1681834

Rosenkrantz, D., Rostosky, S., Toland, M. D., & Dueber, D. M. (2019). Factors impacting parental acceptance of an LGBT child. *Psychology of Sexual Orientation and Gender Diversity*. Advance online publication. https://doi.org/10.1037/sgd0000355

Dueber, D. M., Love, A. M. A., & Toland, M. D., (2019). Comparison of single-stimulus format and forced-choice format instruments using Thurstonian item response theory. *Educational and Psychological Measurement, 79*(1), 108-128. https://doi.org/10.1177/0013164417752782

Fisher, M. H., Thomas, J. N., Jong, C., Schack, E. O, & Dueber, D. M. (2019). Comparing preservice teachers' professional noticing skills in elementary mathematics classrooms. *School Science and Mathematics Journal, 119*(3), 142-149. https://doi.org/10.1111/ssm.12324