Fitting Problems: Evaluating Model Fit in Behavior Genetic Models

By

S. M. Garrison

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

August 7, 2020

Nashville, Tennessee

Approved

Joseph Lee Rodgers, PhD

Kristopher Preacher, PhD

Andrew Tomarken, PhD

David Zald, PhD

Dedication

For my grandfathers. Without their unconditional support, I never would have made it this far.

Walter Raymond Garrison, Jr.	(Jul 7, 1926 - Feb 24, 2019)
Warren H. Kellogg	(Jul 7, 1927 - Dec 17, 2018)

Table of Contents

Page

Chapter

List of Tables

List of Figures

Figure Page

Chapter 1

Introduction

Although many classic findings from behavior genetics replicate (Plomin et al., 2016), our field is not immune to the “replication crisis” (c.f. Lee & McGue, 2016; Turkheimer, 2016). Indeed, many once-classic findings (e.g., Caspi et al., 2003) have failed to replicate, such as candidate genes for intelligence (Chabris et al., 2012; Sniekers et al., 2017) and stress-by-gene interactions (Risch et al., 2009). Although the specific reasons for why those studies have failed to replicate are beyond the scope of this paper, their general causes are not. Many of the failed replications addressed methodological weaknesses from the original paper, often by following best practices. Accordingly, we should continue to implement methodological best practices in all behavior genetics studies, rather than just a select and anomalous few.

In behavior genetics, like any other area of psychology, researchers must decide whether their models adequately explain their data – whether their models “fit” at some satisfactory level. Well-fitting models are compelling, whereas poorly-fitting models are not (Rodgers & Rowe, 2002; cf. Roberts & Pashler, 2000). Oftentimes, researchers evaluate model fit by employing “universal” rules of thumb. These rules of thumb provide fit-index-specific criteria that allow researchers to declare a model well-fitting. Unsurprisingly, papers believed to provide such criteria are among the most highly-cited academic articles. For example, Hu and Bentler (1999) has been cited over 56,000 times. Although popular and listed in many of the “new statistics” guidelines (e.g., Appelbaum et al., 2018; Cumming, 2014; Hancock & Mueller, 2010; Mueller & Hancock, 2010), these rules of thumb should not be treated as universal.

Rather, these model-fit criteria should be viewed as model-specific, and heavily influenced by the measurement model (Hancock & Mueller, 2011; Kang et al., 2016). The cutoff criteria presented by Hu and Bentler (1999) cannot be generalized beyond the single-factor model. For example, if these criteria were applied to five-factor data, the best fitting model would have four or fewer factors (Garrison, 2017). Moreover, many works, including Savalei (2012), Millsap (2012), and Marsh, Hau, and Wen (2004) highlight that incorrect models can appear to be well-fitting when universal criteria are applied. Accordingly, cutoff criteria need to be generated for the specific model of interest. Otherwise, researchers risk rejecting well-fitting models or supporting poor-fitting models, and committing errors of the first and second kind, respectively.

My dissertation aims to address two major questions:

1) How are behavior geneticists evaluating model fit?
2) What are effective criteria to evaluate model fit for behavior genetic models?

In the next section, I provide an overview of behavior genetic models, focusing on matrix specification and model identification. Then, I review evaluating model fit for SEM, where I attend to best practices and evaluation of just-identified models. Following, I present a study of current practices in behavior genetic model fitting, followed by a simulation study. I conclude with summary recommendations.

Chapter 2

Background

Behavior Genetic Modeling

The aim of twin research and, more broadly, behavior genetics research is to "distinguish[…] between the effects of tendencies received at birth, and of those that were imposed by the circumstances of their after lives; in other words, between the effects of nature and of nurture" (Galton, 1876, p. 392). Such designs have been used to understand the genetic and environmental influences on virtually every construct on which people vary (Polderman et al., 2015). These individual constructs can range from the benign, such as height (Holzinger, 1929; Rodgers et al., 2019; Silventoinen et al., 2003) to potentially controversial, including intelligence, educational attainment, and economic success (Burks, 1938; Burt, 1966; Hernstein & Murray, 1994; Jensen, 1969; Murray, 1998; Thorndike, 1905). Researchers have used a variety of kinship groups in their biometrical comparisons, ranging from twins (Galton, 1876), different types of siblings (Thorndike, 1905), cousins (Fisher, 1919), and adoptees (Snygg, 1938); this review will focus on twin studies because they are the most popular within this discipline. Indeed, the "classic" behavior genetic study is a twin design (Liew et al., 2005; Rende et al., 1990), comparing correlations (or covariances) among monozygotic twins to correlations (or covariances) among dizygotic twins.

In behavior genetics, the proportion of variance in a trait is divided into genetic and environmental sources of variability. The proportion of variance attributable to genes is called "heritability," whereas the proportion of variance attributable to non-genetic sources, i.e., the environment is called "environmentality." Put another substantive way, heritability reflects the extent to which genetic differences contribute to observed individual differences (see Visscher, Hill, & Wray, 2008, for a broad review), whereas environmentality reflects the extent to which environmental differences contribute to observed individual differences.

These sources of variability can be further partitioned. Environmentality can be divided into shared- and non-shared environmental experiences. The shared-environment, in a biometrical sense, consists of experiences shared by siblings, whereas the non-shared-environment consists of experiences not shared by siblings. Broad heritability encompasses all genetic variance, including non-additive and nonlinear processes such as dominance, mitochondrial, and epistatic effects. In contrast, narrow heritability focuses exclusively on additive genetic variance. Additive genetic variance reflects the "sum of the average effects of individual alleles." Accordingly, early models of heritability in humans only incorporated the additive components. Methodological advancements have incorporated other genetic variance components, including dominance and epistatic interactions (Lykken et al., 1992). However, classic genetic models focus on narrow heritability, primarily[1] for methodological simplicity. For more on modeling heritability, see Neale and Maes (2004).

The "classic" way to estimate heritability from twin designs has evolved nearly in lockstep with methodological advancements (Jinks & Fulker, 1970; Rende et al., 1990). Before the development of the correlation (Galton, 1896; Pearson, 1909; Stigler, 1989) and the wider acceptance of twin types (Fisher, 1919; Newman, 1917; Siemens, 1924; Smith, 1856), the method was qualitative (Galton, 1876), with nebulous categories of nature and nurture. After those developments, the method became algebraic and model-based (Falconer, 1952; Holzinger,

---

[1] In non-human animal experiments, the majority of genetic variance has been found to be additive (Neale & Maes, 2004). Such findings lend additional justification for focusing on additive genetic variance.

1929), with important underlying assumptions and less nebulous categories of genetic and environmental influences. Other developments have led to incremental modeling improvements (see Jinks & Fulker, 1970 for added treatment on this topic).

Those incremental algebraic gains were upended with the insight that confirmatory factor analytic models could be repurposed for twin designs (Eaves & Gale, 1974; Loehlin & Vandenberg, 1968; Martin & Eaves, 1977). This insight has led behavior geneticists to embrace both factor analytic models and their more general extension, structural equation models (SEM). Structural equation modeling has facilitated great advancements in behavior genetic research, advances that were impossible under the assumptions of older designs. Indeed, some prior assumptions, such as no gene-by-environment correlations, have been transformed into their own research areas.

However, classic twin designs have pushed the limits of structural equation modeling, in terms of model identification. In general, a covariance-based model is identified when there exists a unique set of parameters that defines the maximum of the fit function (Bekker & Wansbeek, 2003). In less formal terms, a model is identified when free parameters cannot trade off with one another to produce the same covariance. I direct readers to Hunter, Garrison, Burt, and Rodgers (2020) for an analytic solution and further discussion of this issue. The classic ACE twin design produces a just-identified model, whereas the fuller ACDE model is under-identified (Hunter et al., 2020; Jinks & Fulker, 1970; Neale & Maes, 2004).[2] In the fuller ACDE model, the covariance structure for monozygotic twins, reared together is

$$Cov_{MZt} = \begin{bmatrix} a^2 + d^2 + c^2 + e^2 & a^2 + d^2 + c^2 \\ a^2 + d^2 + c^2 & a^2 + d^2 + c^2 + e^2 \end{bmatrix}$$

---

[2] I like Neale and Maes' definition (2004): "A model is under-identified if the number of free parameters is greater than the number of distinct statistics that it predicts".

where a^2 is additive genetic variance; d^2 is dominant genetic variance; c^2 is shared environmental variance; and e^2 is non-shared environmental variance. For dizygotic twins, reared together, the covariance structure is

$$Cov_{DZt} = \begin{bmatrix} a^2 + d^2 + c^2 + e^2 & \frac{1}{2}a^2 + \frac{1}{4}d^2 + c^2 \\ \frac{1}{2}a^2 + \frac{1}{4}d^2 + c^2 & a^2 + d^2 + c^2 + e^2 \end{bmatrix}$$

In this model, there are more unknowns (4; a, d, c, and e) than knowns (3; Cov MZt, Cov DZt, and phenotypic variance Vp[3]), resulting in an under-identified model. Without more information (i.e., knowns), these unknowns cannot be calculated. To address this problem, researchers either add information or reduce the number of parameters[4]. Additional information can be achieved by adding another group (such as twins reared apart) or use individual-level data. In the case in the classic ACE model, the number of parameters is reduced by removing dominance from the model. Then, the covariance structure for monozygotic twins is

$$Cov_{MZt} = \begin{bmatrix} a^2 + c^2 + e^2 & a^2 + c^2 \\ a^2 + c^2 & a^2 + c^2 + e^2 \end{bmatrix},$$

whereas the covariance structure for dizygotic twins is

$$Cov_{DZt} = \begin{bmatrix} a^2 + c^2 + e^2 & \frac{1}{2}a^2 + c^2 \\ \frac{1}{2}a^2 + c^2 & a^2 + c^2 + e^2 \end{bmatrix}.$$

Because we have assumed a lack of dominance in this model, the model now only contains three unknowns and three knowns. As a result, we have a just-identified model (Hunter et al., 2020). Although just-identified models can be used to estimate parameters, they cannot be used to evaluate model fit in the traditional sense. Traditional approaches to evaluating model fit rely on

---

[3] $Vp = a^2 + d^2 + c^2 + e^2$

[4] Often dominance is assumed to be zero.

comparing the model of interest to a null model, in the same manner that one would evaluate any null hypothesis. The default null model is often a saturated model, where as many parameters as possible can be defined from the data, and then reproduce perfectly the means and covariances (Widaman & Thompson, 2003). In other words, the default model is a just-identified model with a χ^2 value of zero and with zero residual degrees of freedom. Accordingly, if one were to report a χ^2 test comparing the ACE model to a saturated model, the test would indicate that the model fit perfectly. However, this perfect fit is a consequence of the classic ACE model being just identified. Regardless, nested model comparisons are possible for just-identified models, including classic ACE models. For example, one could compare statistically an AE model or a CE model (both of which are over-identified) to the full ACE model.

The classic covariance-based model (or other similar "two-kinship-category" models) is just-identified, but other models, such as twins-raised-apart-and-together or children-of-twins, are over-identified. In those cases, classic methods of evaluating fit ought to apply. Moreover, one can include additional information, such as group means or individual-level data, rather than just covariances. That additional information increases the total degrees of freedom and results in an over-identified model, with positive residual degrees of freedom. In these non-classic cases, traditional methods of evaluating model fit can and should be used. Many behavior genetic models are specific applications of structural equation models, and ought to be treated as such.

Model Fit for SEM

Best Practices for Evaluating Model Fit

Since the advent of the replication crisis (Open Science Collaboration, 2015; but see, Shrout & Rodgers, 2017; or Ioannidis, 2005), much of the behavioral and social sciences has become interested in improving research practices. Beyond an increased interest in replication, the replication crisis has encouraged social and behavioral scientists to follow best practices for their statistical analyses. These guidelines tend to be broad and flexible, aiming to give scholars enough information to evaluate the model being tested without excessively burdening the authors. Further, this flexibility recognizes that methodological advances, and guidelines for evaluating fit may change in accordance with those advances.

Within the context of structural equation modeling, the guidelines for best practice are broad (Appelbaum et al., 2018; Lance et al., 2016; Mueller & Hancock, 2010). For evaluating model fit, best practice recommends that multiple indices across classes be used in conjunction with the χ^2 statistic. The inclusion of the χ^2 statistic allows the reader to calculate most other fit indices – or at least those indices that are log-likelihood- or χ^2-based.

However, these guidelines are not without criticism. For example, Marsh, Hau, and Wen (2004) noted that best practices and recommendations of “golden rules” tended to deemphasize the numerous limitations of each metric and omit caveats discussed in the original papers. Although beyond the scope of this paper, I encourage readers to be mindful that every fit index discussed in the next section has limitations. Some of those limitations are more obvious (such as χ^2 being sensitive to sample size; Gerbing & Anderson, 1985), whereas other limitations are less obvious (such as, RMSEA being insensitive to omitted cross-loadings; Savalei, 2012).

Specific Methods for Evaluating Model Fit

Because there are numerous methods to evaluate model fit for SEM, I will focus on those recommended in current best practices (Appelbaum et al., 2018; Lance et al., 2016; Mueller & Hancock, 2010). For a classic (and more thorough) overview, see Bollen and Long (1993).

The two primary ways to evaluate model fit are through model comparison and goodness-of-fit indices. The classic model comparison is the χ^2 test (a.k.a. the likelihood ratio test), where the proposed model is compared to the saturated model. Significant values indicate less than perfect fit. Some methodologists consider that the χ^2 test is "overly strict" (Mueller & Hancock, 2010, p. 395) , "almost always statistically significant" (Kenny, n.d.), and overpowered (Bentler, 1990; Hu et al., 1992). Recent work by McNeish (2018) has indeed found that the χ^2 test's type I error rate is inflated at smaller samples.[5] Nested models can also be compared using the χ^2 test. However, these nested comparisons using the likelihood ratio test suffer the same inflated error rates for certain applications, such as classically-specified ACE models (Carey, 2005; Verhulst et al., 2019).[6] Although best practice remains to report the χ^2 test, many goodness-of-fit indices were developed to address the limitations of the χ^2 test (Mulaik et al., 1989).

These fit indices can be broadly classified into three classes: absolute, parsimonious, and incremental. The three broad classes prioritize different mixtures of fit with parsimony. Absolute fit indices describe the overall discrepancy between observed and model implied covariance values. The addition of more parameters improves fit without penalty. Examples include: SRMR (Standardized Root Mean Square Residual), where values below .08 are considered good fit (Hu & Bentler, 1999).

Although parsimonious fit indices also describe the overall discrepancy between observed and model implied covariance values, these indices penalize added model complexity. Model fit should improve with the addition of parameters only if the benefits outweigh the additional model complexity. Examples include: RMSEA (Root Mean Square Error of Approximation), where the lower bound of the 90% confidence interval falls below .05 (Hu &

---

[5] Inflated type I errors for ACE models might be the result of non-normal data, rather than an over-powerful test.

[6] The cause of these inflated rates again is likely more from mis-specified sampling distributions. However, the result is the same – an incorrectly performing test.

Bentler, 1999). Incremental (sometimes called relative) indices compare fits to a baseline model, typically the null model. Examples of these measures include TLI (Tucker-Lewis Index), and CFI (Comparative Fit Index), where values above .95 are considered good fit (Hu & Bentler, 1999; Marsh, 1995). Best practice recommends that multiple indices across classes be used in conjunction with the χ^2 statistic (Appelbaum et al., 2018; Lance et al., 2016; Mueller & Hancock, 2010).

Evaluating Just-identified Models

Many classic methods for evaluating model fit cannot be used on just-identified models, because the null comparison model is saturated and just-identified (Widaman & Thompson, 2003). However, some methods can be applied. I have already discussed nested model comparisons comparing the saturated model to the proposed model. However, that approach can be used for selecting between nested models, including the classic ACE model and models with fewer components, such as the AE or CE model. In addition to nested model comparisons, information criteria can be used, including AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), and aBIC (Adjusted BIC). These criteria are typically used to compare non-nested models, but they can be used in nested models, if they are fit using the same data.

Like the fit indices discussed earlier, these criteria aim to balance parsimony and fit, using the likelihood function rather than the χ^2. The fundamental idea is to minimize the amount of information lost between the data generation process and the model used to approximate it. Better fitting models lose less information. For example, AIC is a function of the number of parameters and maximum value of the likelihood function; models with smaller AIC values indicate relatively better fit as they have lost relatively less information using the same data. Some rules of thumb have been developed for evaluating whether the difference between these models is meaningful, such as differences larger than 10 (Burnham and Anderson, 2002). However, these rules are likely problematic within the context of SEM because of the large sampling variability for fit measures, including BIC (Preacher & Merkle, 2012). Nevertheless, non-nested models can be compared, even within the context of SEM (Merkle et al., 2016).

Chapter 3

Study 1: Current State of the Field

How are behavior geneticists evaluating model fit?

I examined how behavior geneticists are evaluating their model fits. Work by Jackson (2009) on general reporting practices for CFA revealed that most papers followed contemporary guidelines. However, a priori I suspected that this practice would not extend to the field of behavior genetics. In a recent review of the field's flagship journal, *Behavior Genetics*, I observed that many articles were using out-of-date software (Garrison, 2018). In 2016 and 2017, 22% of SEM-based articles in *Behavior Genetics* used Mx. Mx was last released in 2009, (Version 1.7.03; Neale, 2009); it is no longer compatible with most operating systems; and its developers have moved onto openMx (Boker et al., 2011; Neale et al., 2016). During the writing of that article, I observed that some articles omitted reporting standard fit statistics (e.g., χ^2), even in cases when the model was over-identified. In addition, I noticed that many of those papers tended to focus on AIC and a nested-model comparison. This method mirrors the last release of a Mx-companion behavior genetics book (Neale & Maes, 2004). That mirroring is suggestive that these papers used that book as a guide, without considering whether their own models were just- or over-identified.[7] Accordingly, this first study aims to quantify those observations.

Methods

I reviewed the two years of publications from the flagship journal *Behavior Genetics* (2016 and 2017), which I had previously identified in Garrison (2018). In that paper, I

---

[7] Neale and Maes (2004) has an excellent discussion on model identification for classic twin designs.

determined SEM programs were being employed, by reviewing the previous two years of publications in *Behavior Genetics.* The number of articles identified is summarized in Table 1. I identified 473 "publication units" (all publications) during those two years, using Publish or Perish (Harzing, 2018). I excluded published abstracts from the annual conference (n=310) as well as articles indexed across multiple databases (n=46). Further, I refined the remaining 117 articles to human participants (n=88), by excluding editorials (n=4; e.g., Ayorech et al., 2016), simulations (e.g., Verhulst, 2017) and non-human subjects (n=20), such as cows (Friedrich et al., 2017), flies (Garlapow et al., 2016), whales (Whitehead et al., 2017), and zebrafish (Gerlai et al., 2017). Lastly, I eliminated non-SEM analyses (n=48), such as co-twin control studies (e.g., Mosing et al., 2016) and various molecular genetic studies (e.g., van den Berg et al., 2016). The remaining 40 articles employed structural equation modeling for quantitative genetic modeling on humans.

Table 1 Number of Articles Identified by Stage

(#) Stage	Number of Articles
(0) Identified	473
(1) Abstracts	310
(2) Duplicates	46
(3) Non-human	29
(4) Non-SEM	48
(5) Remaining	40

Results

I identified which methods were being used to evaluate model fit in the flagship journal *Behavior Genetics*. The variability in practice of how much attention was paid to model fit was considerable, ranging from dedicated discussion subsections (e.g. Smolkina, Morley, Rijsdijk, et al, 2017) to the total absence of their usage (n=8; 20%). I have provided summary statistics in

Table 2. Given the variety of methods used and the broad guideline for best practice, I have classified each paper on their model-fitting practice. If a paper used more than one method to evaluate fit, I classified that paper as *acceptable*. If the paper used multiple classes of methods to evaluate fit, such as fit indices (e.g., absolute, parsimonious, incremental), I classified that paper as meeting the standard for *better* practice. Lastly, if the paper used multiple classes in addition to reporting either the χ^2 statistic or log likelihood, I classified that paper as meeting the standard for *best* practice. According to this classification, 5% met the standard for best practice (n=2). That percentage rose to 50% (n=20) for better practice, and 72.5% (n=29) for acceptable practice. Further, if nested model comparisons are included as a "class", these percentages increase to 42.5% (n = 17) for best, 72.5% (n=29) for better; and 75% (n=30) for acceptable.

For context, 80% (n=32) of articles reported a formal evaluation of model fitting. In those remaining 20% (n=8) of articles, models were selected for a variety of reasons, including "parsimony", visible inspection of covariances, significance of A or C parameters, and findings from previous literature. In general, if an article evaluated fit, it typically seemed to draw inspiration from Neale and Maes (2004). Further, if an article evaluated fit, it used the criteria correctly. Therefore, rather than further focusing on evaluating current practice, Study 2 will focus on developing clear criteria.

Table 2 Number of Papers per Fit Method

Method of Fit	Number of Papers
AIC/BIC	19
Nested Model Comparison	28
RMSEA/SRMR	7
CFI/TLI	2
Loglikelihood/χ^2	26
Any	32
Total	40

Brief Discussion

To summarize, I evaluated whether 2016-2017 *Behavior Genetic* papers met the standard for best practice (5%), better practice (50%), and acceptable practice (72.5%). Very few papers met the best standard (Appelbaum et al., 2018; Lance et al., 2016; Mueller & Hancock, 2010), which recommended that multiple indices across classes (parsimonious, relative, and absolute) be used in conjunction with the χ^2 statistic. In my operationalization of best practices, I believe that these recommendations are both generous and reasonable across all SEM models. Unfortunately, most papers published in *Behavior Genetics* did not meet this standard. They often did not report a second index from another class, such as an absolute measure such as RMSEA or a relative measure such as TLI. Instead papers were likely to include a nested model comparison instead of a second index. Accordingly, the "better practice" standard I developed allowed the nested model comparison to count as a class for evaluating fit. Even then, only half of papers met this standard. Frankly, the greatest concern is that 20% of papers did not evaluate fit at all. In those cases, it would be impossible to create a metric to include them as meeting some or any standard of practice.

Nevertheless, and a positive finding, if an article evaluated fit, that article typically used the criteria correctly. The optimistic interpretation of these findings is that most researchers recognize that evaluating model fit is part of the research process. Instead, rather than criticizing current practice, or focusing on the current weaknesses in the field, study 2 will focus on developing clear criteria that can be used to successfully evaluate model fit. Ideally, those criteria will be simple, universal, and meet the standards for best practice.

Chapter 4

Study 2: Evaluating Model Fit

Study 2 builds upon the findings of Study 1, by determining whether clear criteria for evaluating model fit can be developed, and whether those criteria differ from "classic" criteria for SEM in general. Specifically, I focused on developing custom fit criteria, in the style of Hu and Bentler (1999) for classic univariate ACE models, by fitting classic models to both correctly and incorrectly specified models using the classic twin design comparing MZ and DZ twin correlations/covariances.

Evaluation Plan

Methods of evaluating model fit included (when possible), RMSEA, SRMR, AIC, BIC, aBIC, χ^2, nested model comparison, CFI, TLI, as well as an additional method identified in the study census – the model was kept if it had significant A or C parameters. My evaluation plan emulated Hu and Bentler's (1999) approach by examining the Type I and Type II error rates for correctly and incorrectly specified models.[8] In addition, I examined sensitivity and specificity for these cut-off values. Sensitivity is a way to estimate the true positive rate, whereas specificity examines the true negative rate (Yerushalmy, 1947). Unacceptable rates were those that deviated substantially from 5% Type I error rates, 95% sensitivity, 20% Type II error rates, and 80% specificity. The ideal method will be the one that can achieve those 4 values, across all factors within the simulation designs. It may be unlikely that any measure/combination of measures will meet all these ideal values. Accordingly, I aimed to identify thresholds that were closest to

---

[8] Although Hu and Bentler (1999) is viewed as the classic article, I also recommend that researchers also consider using the approach laid out by Millsap (2012).

meeting these ideal values by minimizing Type I and II error rates, and maximizing sensitivity and specificity. Those optimal values were selected based on three optimizing metrics

Using the cutpointr package (Thiele, 2019), thresholds were selected, using three metrics. The first metric, the Youden's (1950) J statistic, is sensitivity + specificity - 1. This index ranges from 0 to 1, where 1 indicates perfect classification (i.e., no false negatives or positives) and 0 indicates that the criteria are non-informative. The other metrics conditionally maximize either specificity or sensitivity. The conditionally maximized specificity maximizes power at an acceptable Type I error rate ($\alpha \approx .05$)/sensitivity level of .95, whereas the , conditionally maximized sensitivity maximizes sensitivity at an acceptable level of power (specificity = 0.80).

Methods

All simulations were conducted in R version 3.6.1 (R Core Team, 2020) and mplus version 7.3 (Muthén & Muthén, 2017), using the R packages: MASS 7.3 (Venables and Ripley 2002), MplusAutomation (Hallquist & Wiley, 2018), and discord (Garrison & Ream, 2017).

I examined whether any of the classic cutoff values for model fit criteria are effective for classifying models correctly. In other words, are misspecified models rejected, and are correctly specified models accepted? In the spirit of Hu and Bentler (1999), I will be fitting a correctly specified model as well as many incorrectly specified models. Because I am focusing on classic univariate designs, I have produced a summary table of the conditions outlined below in Table 3.

Table 3 Simulation Conditions

	Condition	Values
Genetic Variance	Additive	0, 1, 2
Environmental Variance	Shared	0, 1, 2
	Non-Shared	1, 2
Design	Sample Size (total pairs)	200, 500, 1000, 2000
	Data	Raw, Covariance

Data Generation

I varied the following conditions with 5,000 replications (see Table 3):

Variance Components. I have included one genetic component (additive) and two environmental components (shared- and non-shared). Each component condition, except for non-shared, can be 0, 1, or 2. I have excluded the 0 non-shared variance condition because practically every measure has non-shared environmental variance (Polderman et al., 2015; Turkheimer & Waldron, 2000). These variance components can be converted into classic heritability estimates, by taking the proportion of variance components. For example, if the variance components were 1 additive genetic and 1 non-shared environmental, then that would result in $a^2 = .5$, $c^2 = 0$, and $e^2 = .5$.

Each variance component is generated separately and then combined to create a total score, using the discordsim function from my discord package (Garrison & Ream, 2017) in R. This function generates biometrically-informed data for up to two measures for each member of a kin pair, using multivariate normal functions for each variance component. These kin pairs can vary in their genetic and environmental similarity. Further, the two measures can have overlapping biometrical influences, that are either bidirectional (e.g., correlated factors model, direct symmetric model) or unidirectional (Cholesky model).

Sample Size. I have varied sample size ranging from 200 total pairs to 2000 total pairs (200, 500, 1000, 2000). These sample sizes are based on typical sample sizes from Polderman's et al. (2015)'s meta-analysis. Pairs will be balanced to 50% Monozygotic Twins and 50% Dizygotic twins, as is typical for classic studies (Polderman et al., 2015).

Data Input Method.[9] The generated data will be used directly (i.e., at the individual level) or indirectly, with estimated covariance matrices from those same generated data.

Each dataset was fit using the following models: ACE, AE, and CE, using the raw data and the covariance matrix. In each case, the data are fit to two incorrect models, and the correctly specified model. For example, if the correctly specified model is an AE model, then the incorrectly specified models fit will be ACE and CE. This simulation is a 3 x 3 x 2 x 4 x 2 (Genetic variance x Shared environmental variance x Non-shared environmental variance x sample size x covariance versus raw data; see Table 3) design, with 144 total cells. Data were generated using R. Models were estimated with Mplus, using syntax adapted from Garrison and Rodgers (2019). Select conditions were used to validate the data generation process by fitting ACE models to data generated from an E model[10].

Rejection Criteria

Rejection criteria were initially planned to be guided wholly by observed methods. However, given the small sample of those who used any given fit index, I selected criteria based on a mixture of those observed practices in the survey along with common practice and popular articles, books, and websites (Hooper et al., 2008; Kenny, n.d.; Kline, 2015). Those criteria are documented in Table 4. Please note that aBIC, AIC, and BIC as well as the nested model comparison are nested, and in that regard are compared to the saturated model (i.e., ACE) fit to those same data.

---

[9] An additional series of models were fit using correlation matrices, instead of the covariance or raw data. Unfortunately (or fortunately), Mplus does not allow for multi-group SEM models to be estimated with correlations. Preliminary attempts to work around this restriction by fitting the correlations as covariances with means of 0 led to inconclusive results.

[10] The E model describes any model where only non-shared environmental variance was generated and all other variance components (A & C) were 0.

Table 4 Model Rejection Criteria

Metric	Criteria	Nested
aBIC, AIC, BIC	Minimum value	Yes
ChiSqM Test	≤ .05	No
CFI, TLI	<.95	No
RMSEA Estimate	>.06	No
RMSEA 90% CI Lower Bound	>0	No
RMSEA 90% CI Upper Bound	>.06	No
p(RMSEA)<.05	<1	No
SRMR	>.08	No
Non-significant parameter	>.05	No
Nested Model Comparison	≤.05	Yes

Results

Given the number of conditions and methods of comparisons, I have focused on general trends and the three most promising methods for evaluating fit: TLI, χ^2 Test, and Nested Model Comparison. I direct readers to detailed tables in Appendix B and to the accompanying data files. Results did not differ meaningfully across data input condition (raw data or covariance-based summary statistics). Typically, values across levels of that condition did not differ until the third decimal place. Accordingly, I have narrowed my focus to the raw-data based method.

First, I conduct basic validity checks to determine whether the data generation and the model fitting process were performing correctly. Specifically, I examined whether true models performed “better” than untrue models in the expected direction for each fit statistic. I also examined how data generated from an E model – that is a model with only non-shared environmental variance – would fit an ACE model and if the descriptive statistics from those models performed as expected.

Second, I present descriptive statistics for models by whether the correct or incorrect model was fit as well as by whether the fitted model had three variance components (i.e., an ACE model) versus a model with only two variance components (i.e., AE or CE). Third, I present type I error rates and power for model selection, using conventional benchmarks. Fourth, I find optimized threshold values, using three common methods (Youden index; maximizing power, given an alpha level of ≈ .05; minimizing alpha, given a minimum power level of .80), implemented with cutpointr package (Thiele, 2019). After each subsubsection, I have provided brief discussions.

Validity Checks

In this subsubsection, I present a selection of analyses done to verify the simulation worked as intended with the discord package. Specifically, I examined for each fit statistic whether true models performed "better" than untrue models in the expected direction; and whether general summary statistics for ACE models fit to data generated from E models performed as expected. Tables 5- 12 contain descriptive statistics for each condition, and can be found in the appendix.

As expected, the following measures had lower median levels when the true model was fit relative to untrue models: aBIC, AIC, BIC, p(χ^2), χ^2, CFI, TLI, and p(RMSEA<.05). In contrast, the following measures had higher median levels when the true model was fit relative to untrue models: estimated RMSEA, RMSEA's upper bound of the 90% CI, and SRMR. All indices performed as expected, indicating that true models are at least better fitting than those that are not correctly specified.

Next, I examined descriptive statistics for data generated from an E-only model but fit to an ACE model. As expected across conditions, median standardized (and unstandardized) parameter estimates for A and C were 0 across all conditions. Further, median standardized parameter estimates for E ranged from 0.995 to 0.999 across conditions. The 95th percentile for standardized estimates of E ranged was 1 across conditions, whereas the 5th percentile ranged from 0.926 to 0.978. The p-values associated with these parameter estimates had median levels of 0 across conditions for E, and for A, they ranged from 0.998 to 1, and for C they ranged from 0.998 to 1. The 95th percentile for p-values associated with E was 0 across conditions, whereas the 5th percentile for A ranged from 0.008 to 0.016. The 5th percentiles for p values associated with C similarly ranged from 0.009 to 0.015. The 90th percentile for p-values associated with E was 0 across conditions, whereas the 10th percentile for A ranged from 0.082 to 0.141. The 10th percentiles for p-values associated with C similarly ranged from 0.093 to 0.111. These p-values are more extreme than one would hope; however, they are consistent with recent work by Verhulst and colleagues (2019) and reflect the fact that estimated values for C and A are at the boundary of 0 and are restricted to be positive. In addition, distributions of $p(\chi^2)$ performed as expected; specifically, median values ranged from 0.331 to 0.358. Their 5th percentile ranged from 0.02 to 0.026 across conditions.

Summary Statistics

In this subsubsection, I describe a handful of descriptive statistics for models by whether the correct or incorrect model was fit. These handful of descriptive statistics are TLI, the significance level of χ^2 test of model fit, and significance level for nested model comparisons, comparing the ACE model to the selected model of interest. Table 13 in the appendix provides median values for standardized measures of fit with 1%, 2.5%, 5%, 10%, 90%, 95%, 97.5%, and 99% quantiles for correctly-specified models. These values compare favorably to the majority of classic benchmarks. For example, approximately 95% of all true models reported $p(\chi^2)$ greater than .05. Unfortunately, the distributions for these measures (as well as non-standardized methods of fit) overlapped considerably with the distributions from incorrectly specified models (see Table 14 Appendix B).

Select Findings. Figures 1, 2, and 3 illustrate this general trend in overlapping distributions with box and whisker plots. For each figure, the whiskers span the 5% and 95% quantiles, and the Interquartile Range spans the 25% and 75% quantiles. As illustrated in Figure 1, median levels of TLI differed between true models (1.001) and untrue models (0.988). However, the 5% and 95% quantiles overlapped; for true models, the quantiles ranged from 0.98 to 1.018, and for untrue models from 0.913 to 1.015. Similarly, distributions overlapped for $p(\chi^2)$ where median levels were 0.495 for true models and for untrue models were 0.066 as illustrated in Figure 2. The quantiles ranged from 0.049 to 0.949 for true models and from 0 to 0.845 for untrue models. For nested model comparisons, the trend again was similar with differing median levels, but overlapping quantiles, for true (1, 90% QI [0.2, 1]) and untrue models (0.021 , 90% QI [0, 1]), as illustrated in Figure 3.

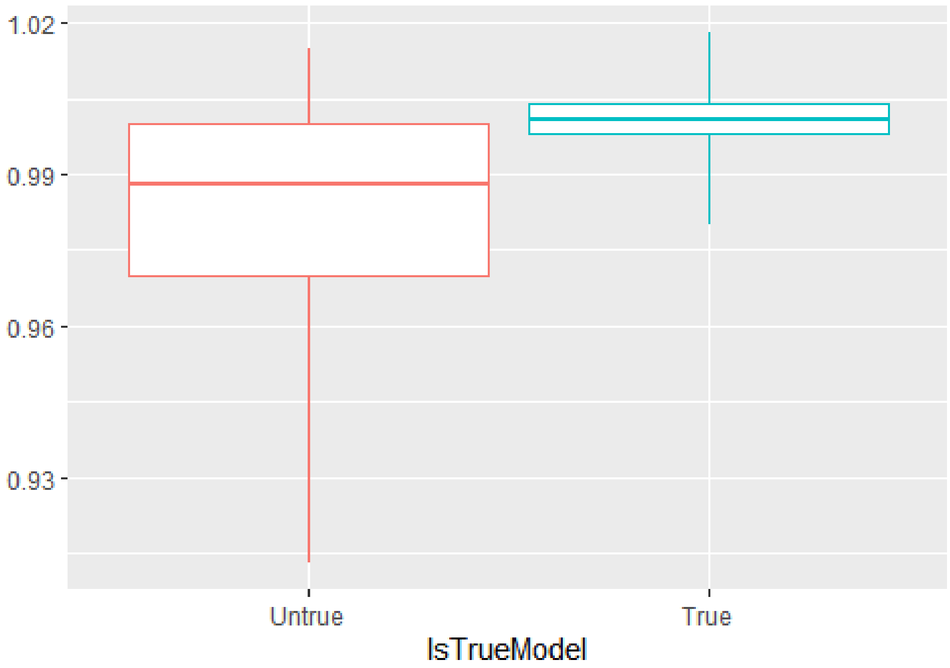

Figure 1 Box and whisker plot of TLI by Model Status

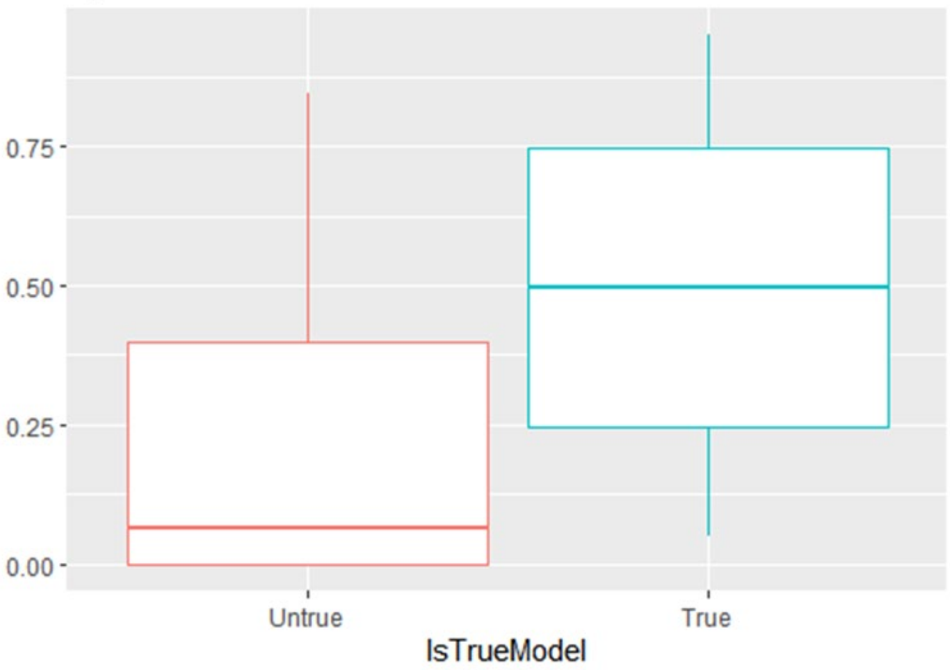

Figure 2 Box and whisker plot of p(χ²) by Model Status

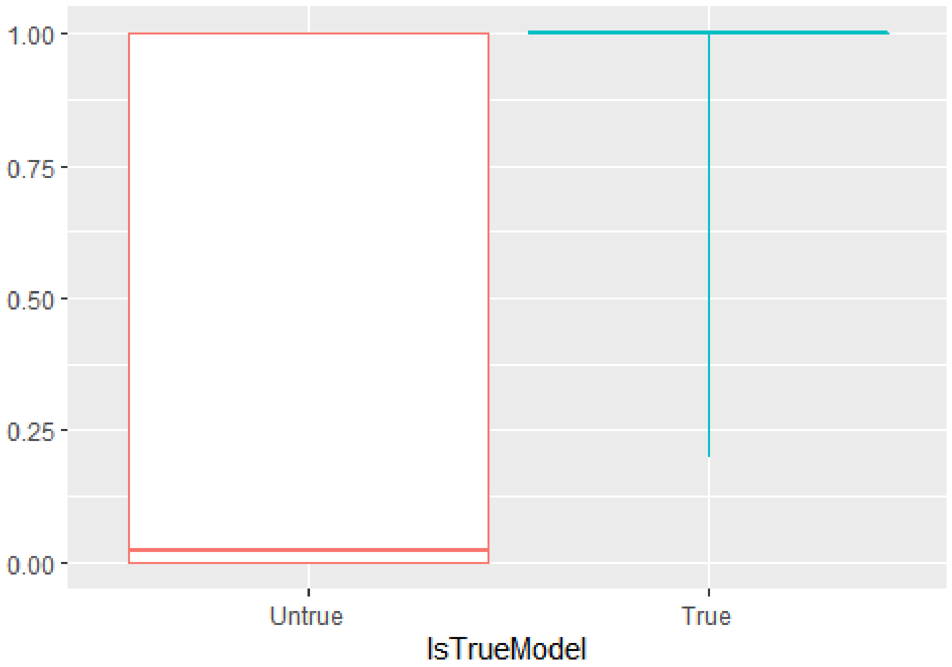

Figure 3 Box and whisker plot of p(Nested Model Comparison) by Model Status

Next, I distinguished by the type of model fit. Specifically, I compared models with all three variance components (i.e., ACE) to models with only two of the three variance components (AE and CE models). Median levels generally differed, but true and untrue models continued to have overlapping distributions. As expected, untrue ACE models tended to be indistinguishable from true models. Indeed, practically all measures showed this consistent pattern where median levels were indistinguishable for ACE models but differed for AE and CE models. For example, TLI levels were indistinguishable for ACE models (1, 90% QI [0.987, 1.011] for true vs 1, 90% QI [0.806, 1.028] for untrue; see Figure 4). However, median TLI values were lower when untrue two component models (0.983) models were fit compared to true two component which also had median TLI values of (1.001). A similar trend was observed for p(χ^2), where median levels of p(χ^2) for untrue ACE models (0.417) were indistinguishable from true ACE models (0.494). Similarly, median p(χ^2) values were lower when untrue two-component (0.008) models were fit compared to true two-component models which also had median p(χ^2) values of 0.495 (see Figure 5). Lastly, for nested comparisons, the trend again was similar with indistinguishable median values for ACE models (untrue 1; true 1), but lower values when two-component models were evaluated (untrue 0; true 1; see Figure 6). Nevertheless, even when distinguished by type, all measures overlapped in their distributions.

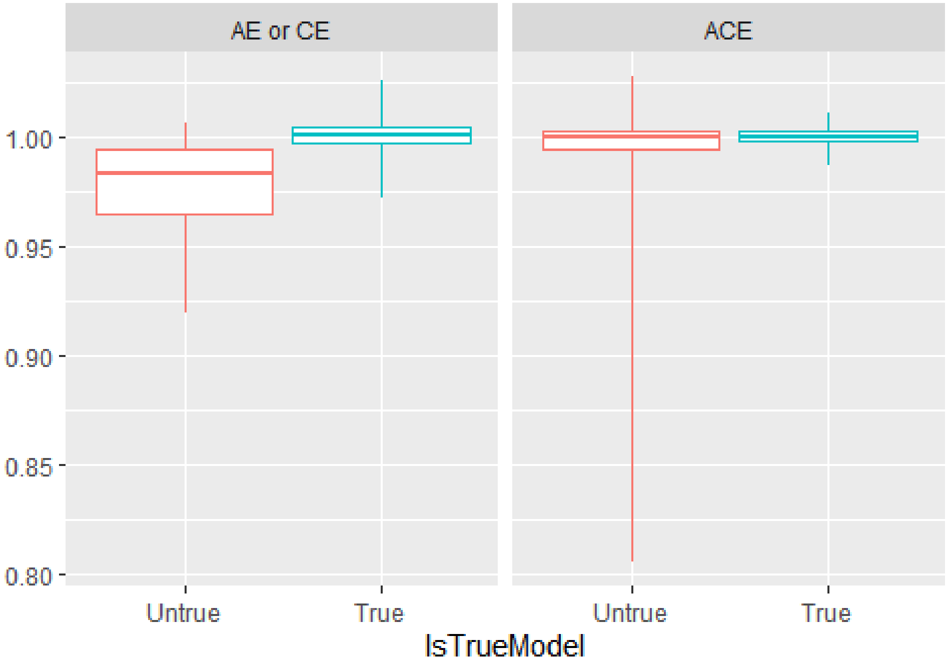

Figure 4 Box and whisker plot of TLI by Model Status and Type

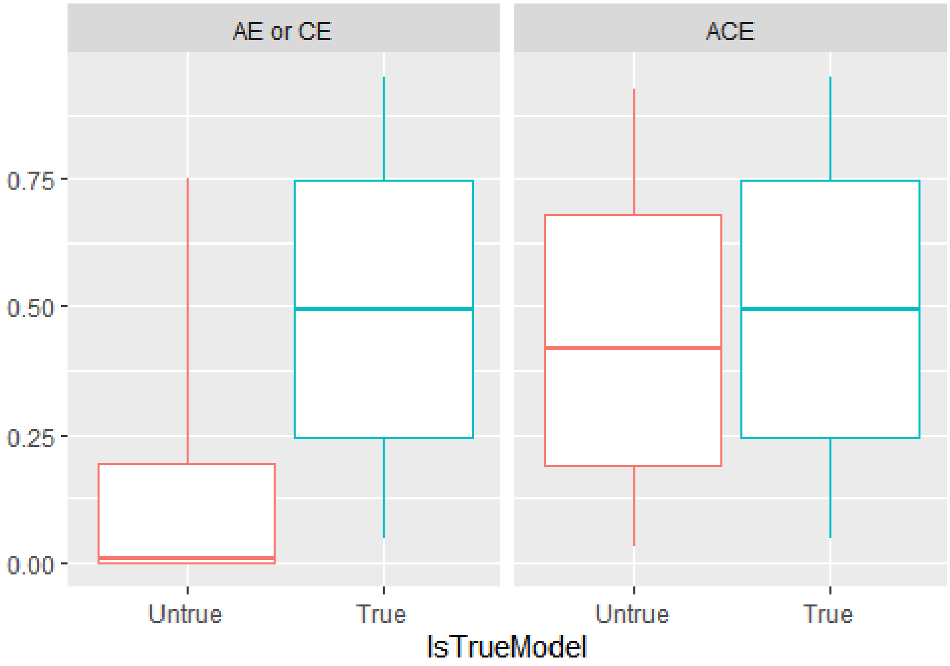

Figure 5 Box and whisker plot of p(χ²) by Model Status and Type

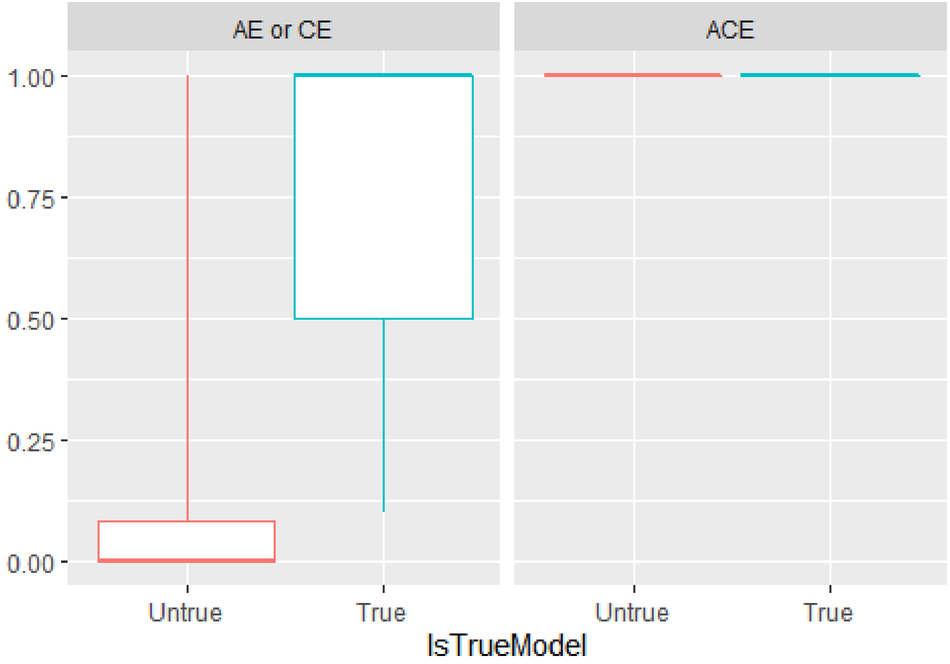

Figure 6 Box and whisker plot of p(Nested Model Comparison) by Model Status and Type

When distinguished by both type of model fit (ACE vs AE or CE) and sample size, distributions tended to narrow with larger sample sizes. The spread reduction was most drastic between 200 and 500, as one would expect. The same trends occurred as described previously. When fit to untrue models, ACE models still resembled true models, often with nearly indistinguishable distributions. As was the case without accounting for sample size, untrue AE and CE models had distinguishable median levels when compared to untrue ACE or true models in general.

I also examined whether these summary statistics differed across levels of total variance[11]. Total variance ranged from 1 to 6. In general, more variance did not impact model fits when true models were fit. However, a linear relationship was observed for most statistics when fit to untrue models, either in terms of decreasing variability (e.g., TLI; Figure 7) or in declining median levels. For example, without accounting for type of model fit, significance levels for the χ^2 test of overall fit did not vary as a function of total variance when the correct model was fit (e.g., median levels of significance were 0.496 when total variance was 2; 0.499 when total variance was 6, etc; see Figure 8). However, they did vary for untrue models. Increases in total variance were associated with decreases in p-value. Lastly, nested model comparisons behaved similarly (see Figure 9); however, this behavior seems to be more the product of total variance being partially confounded with type of model fit.

---

[11] Total variance was not randomized, and was not be balanced across true versus untrue models as well as across the type of model fit (ACE, AE, CE). For example, no true model had a total variance of 1.

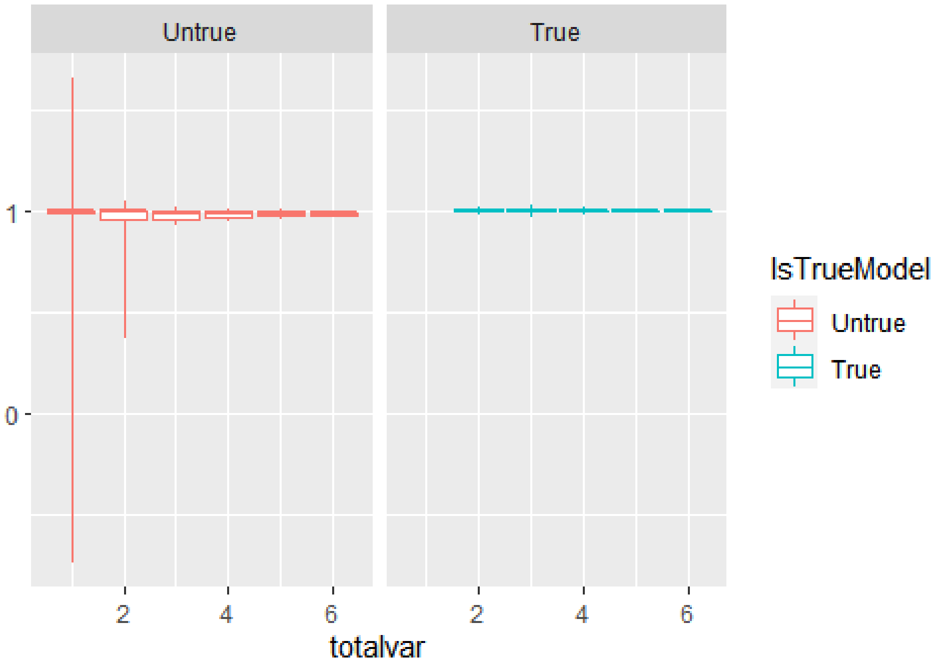

Figure 7 Box and whisker plots of TLI by Model Status and Total Variance

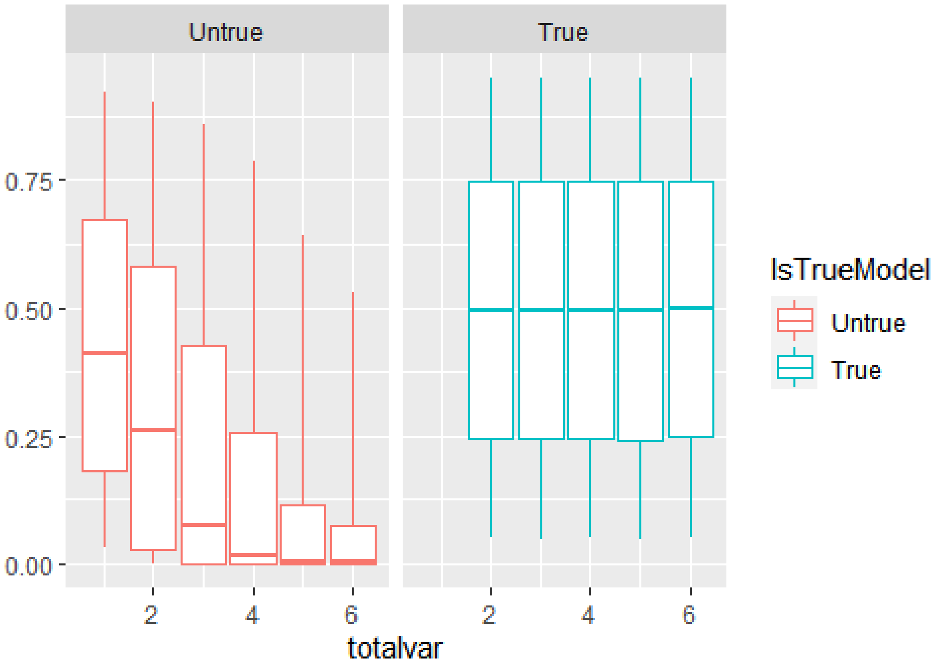

Figure 8 Box and whisker plot of p(χ²) by Model Status and Total Variance

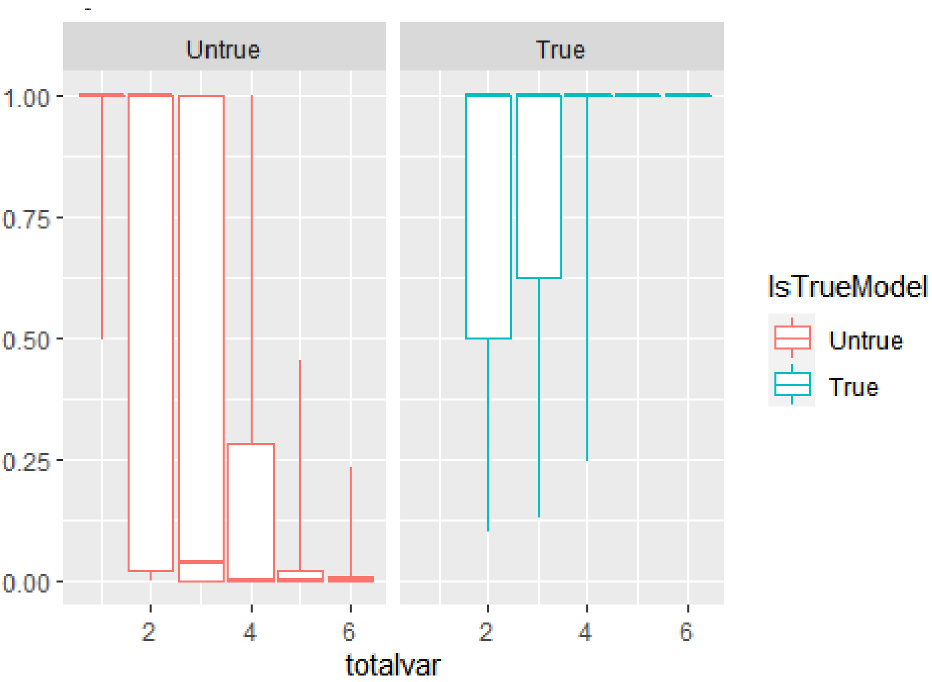

Figure 9 Box and whisker plot of p(Nested Model Comparison) by Model Status and Total Variance

The same relationships with total variance seemed to hold even when accounting for type of model fit (ACE vs AE or CE), with one expected exception. The ACE models fit to untrue models still tended to be indistinguishable from true ACE models. I have included Figures 10, 11, and 12 to illustrate the trends for TLI, significance levels for the χ^2 test of overall fit, and nested model comparisons. For example, significance levels for the χ^2 test of overall fit did not vary as a function of total variance when the correct model was fit (regardless of kind) or when the incorrect ACE model was fit. Those values did level off when fit to untrue AE or CE models, median significance leveled off quickly when fit to AE or CE, see Figure 11 (e.g., 0.444 when total variance was 1; 0.106 when total variance was 2; 0.001 when total variance was 3, etc).

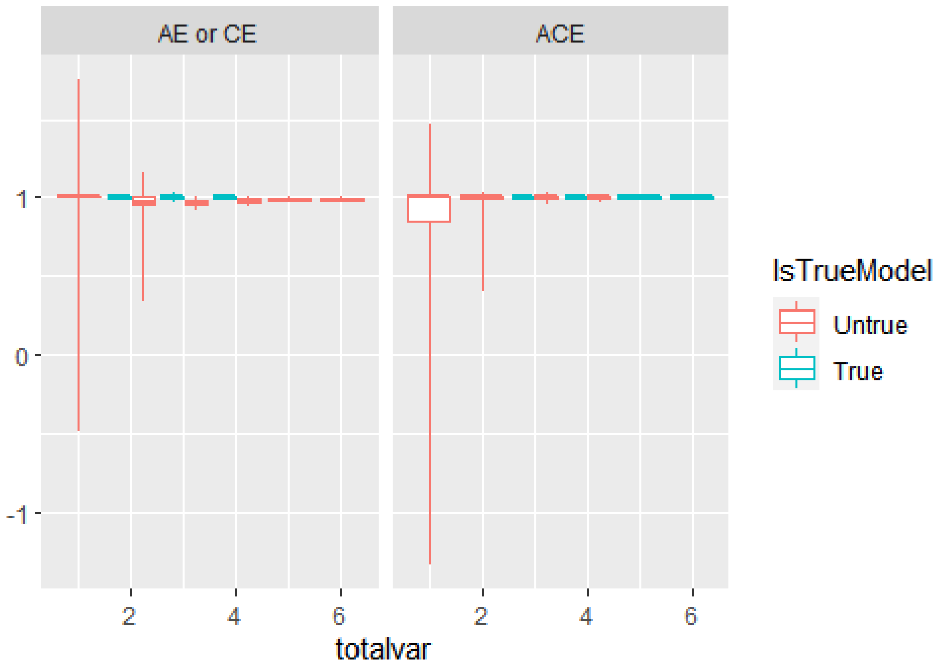

Figure 10 Box and whisker plots of TLI by Model Status, Model Type, and Total Variance

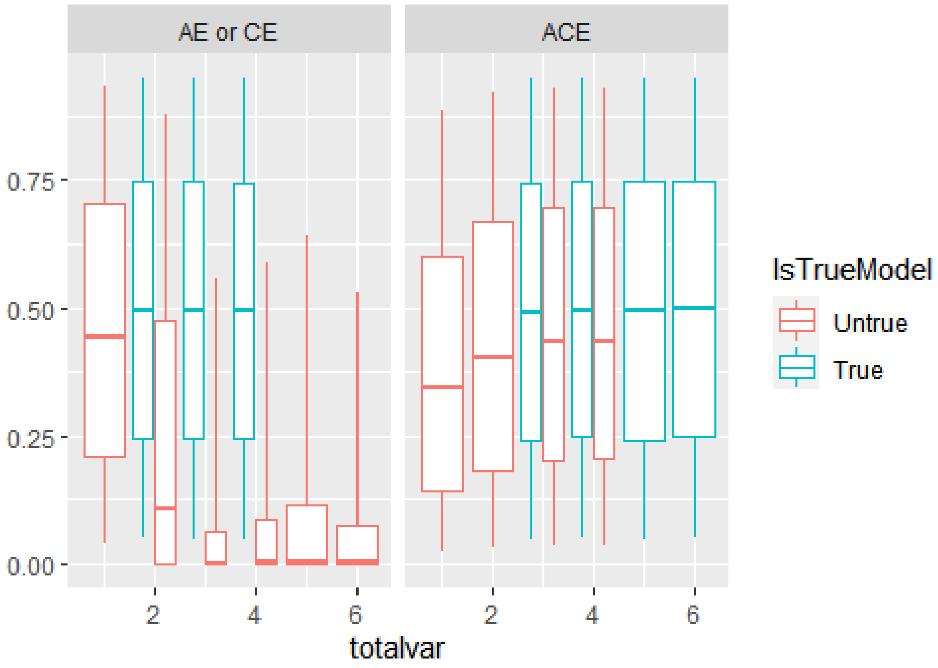

Figure 11 Box and whisker plots of $p(\chi^2)$ by Model Status, Model Type, and Total Variance

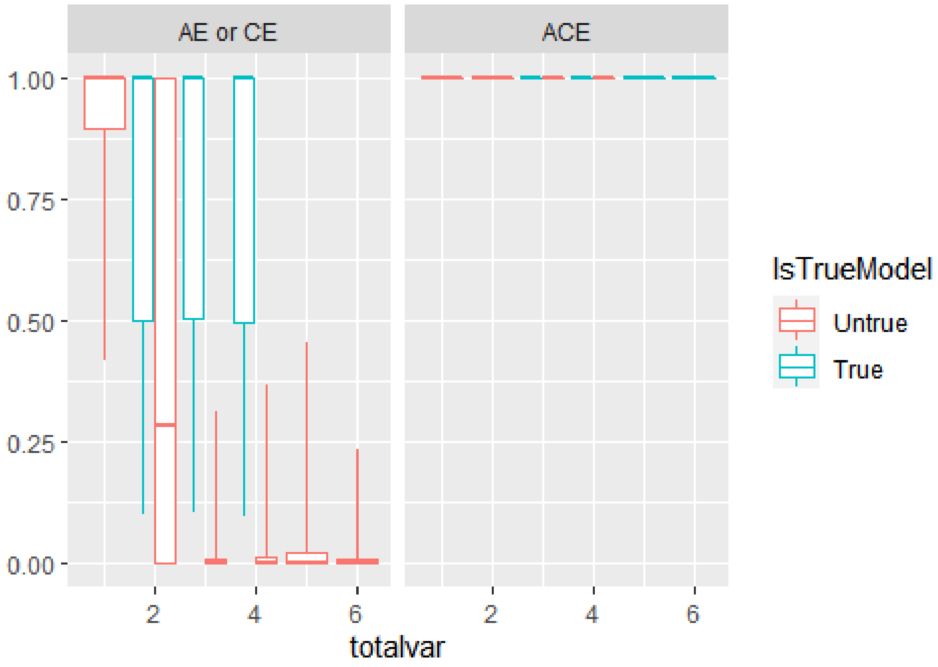

Figure 12 Box and whisker plots of p(Nested Model Comparison) by Model Status, Type, & Total Variance

Brief Discussion. The main finding from the descriptive statistics revealed that although indices for correctly-fit models tended to compare favorably to classic benchmarks, there was considerable overlap in distributions between correctly and incorrectly specified models. These distribution overlaps were present even when compared across type of model fit (ACE, CE, AE), sample size, and total variance. These overlaps were most problematic when comparing untrue ACE models to true models, where the distributions were practically indistinguishable. However, those indistinguishable distributions are likely the product of the ACE model being saturated.

Power and Type I Error

Next, I examined model rejection rates, using classic benchmarks and nested model comparisons. Type I errors were defined as rejecting a correctly specified model. Power was defined as rejecting an incorrectly specified model. Type II error rates were defined as β (1-

power), or incorrectly specified models that were not rejected. These rates were calculated by condition. For the sake of brevity, I will again focus on the three most promising metrics: TLI, p(χ^2), and the nested model comparison. I direct readers to Table 15 in the appendix for these rejections as well as all the other tested metrics by condition.

Select Findings. For correctly-specified models, the median rejection rate across conditions for true models (i.e., Type I errors) ranged from 0 (significance of model parameters, TLI) to 0.992 for p(RMSEA)<1. For the sake of brevity, I focus on the three most promising metrics: TLI, p(χ^2), and the nested model comparison. The χ^2 test had median rejection rates near 0.05 (0.051; min 0.043, max 0.06). In contrast, TLI and nested model comparisons had rejection rates well below .05. For TLI, the median rejection rate was 0 (min 0, max 0.222), and for nested χ^2 model comparisons it was 0.009 (min 0, max 0.03). Figures 13 - 15 illustrate variability and differences in rejection rates with box and whisker plots. In each figure, power and type I error rates are benchmarked with horizontal lines at 0.80 and 0.05.

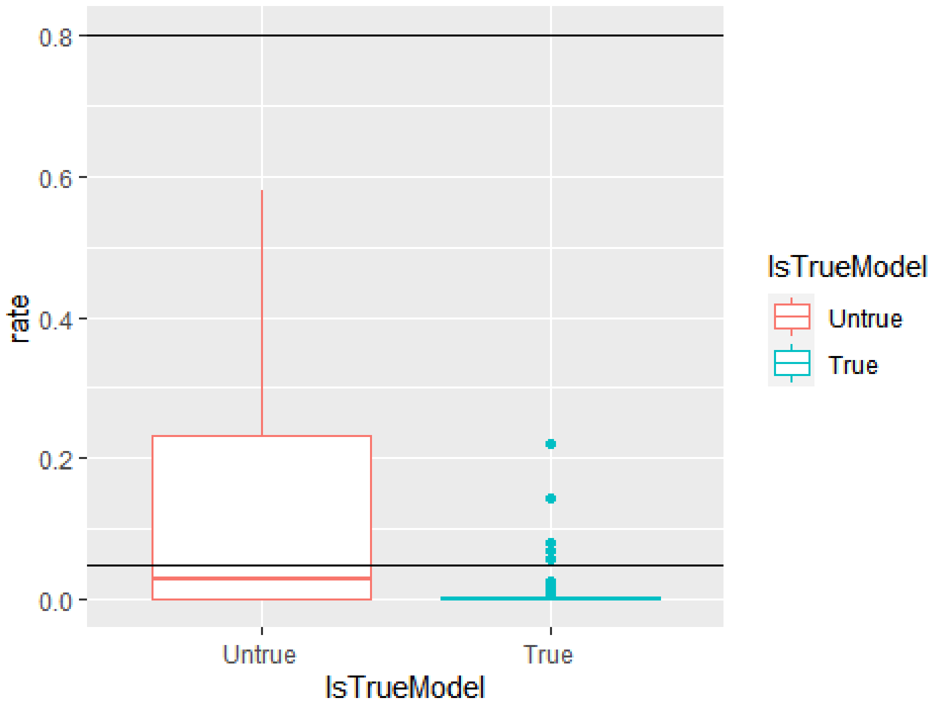

Figure 13 Distribution of model rejection rates using TLI across conditions

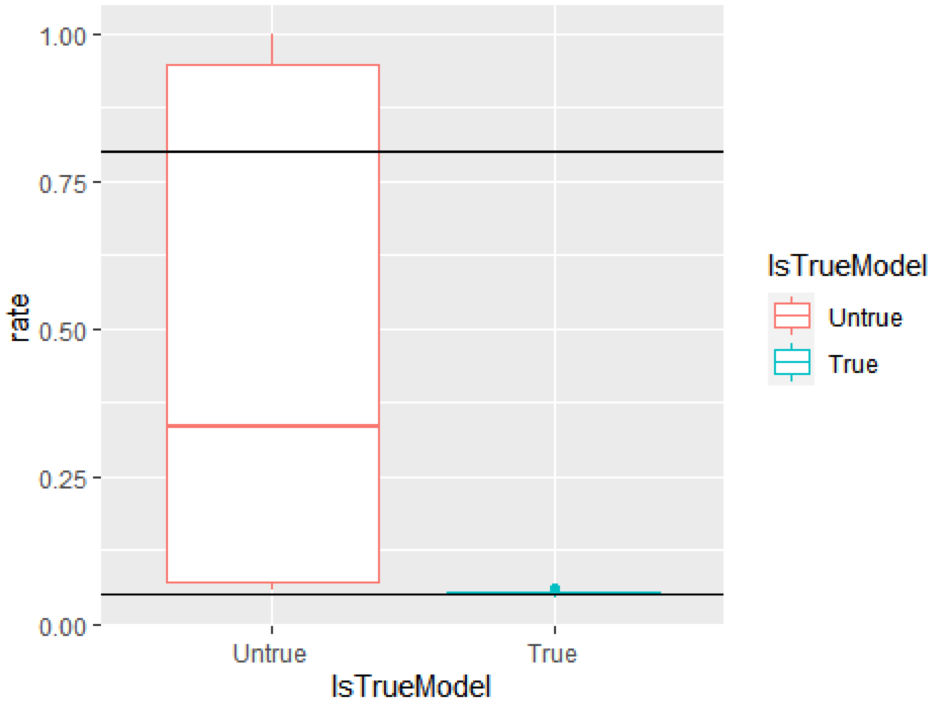

Figure 14 Distribution of model rejection rates using p(χ^2) across conditions

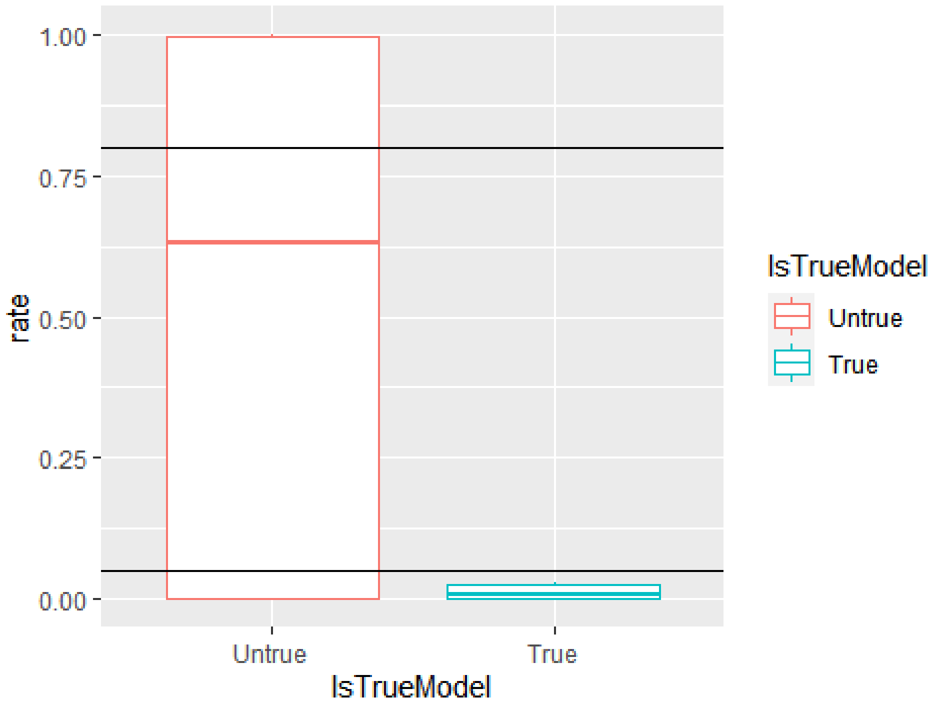

Figure 15 Distribution of model rejection rates using nested model comparisons across conditions

For incorrectly specified models, the median rejection rate across conditions for untrue models (i.e., power) ranged from 0 (significant model parameters) to 1 for p(RMSEA)<1. All three of the metrics on which I focused had median rejection rates well below .80: TLI (0.028; min 0, max 0.578; see Figure 13), $p(\chi^2)<.05$ (0.333; min 0.057, max 1; see Figure 14); and nested χ^2 model comparisons (0.631; min 0, max 1; see Figure 15).

Given that distributions of fit statistics seemed to differ as a function of sample size as well as total variance, I examined whether rejection rates were associated with these values. All rates for correctly specified models were negatively associated with sample size. Virtually all correlations were large (>.4). These correlations are provided in Table 16 and ranged from -0.972 for RMSEA <.06 to -0.024 for χ^2 nested model comparisons. Many rates for incorrectly specified models were positively associated with sample size, but this trend was not universal.

p(χ^2) and nested model comparisons were positively associated with sample size, while TLI was negatively associated.

In addition, many rates were associated with total variance for both correctly and incorrectly specified models, but not as universally as sample size. p(χ^2) was positively associated with total variance for both model specifications. TLI was negatively associated for both. However, for nested model comparisons the association with total variance was positive for incorrectly specified models and negative for correctly specified models.

Like the descriptive statistics, rejection rates varied across type of model fit (ACE, CE, and AE) – primarily ACE compared to either CE or AE. Figures 16, 17 and 18 illustrate these differences using box and whisker plots. These plots show the distribution of rejection rates by condition. An ideal metric would have low variability across conditions and exceed the minimum standards of power>.80.

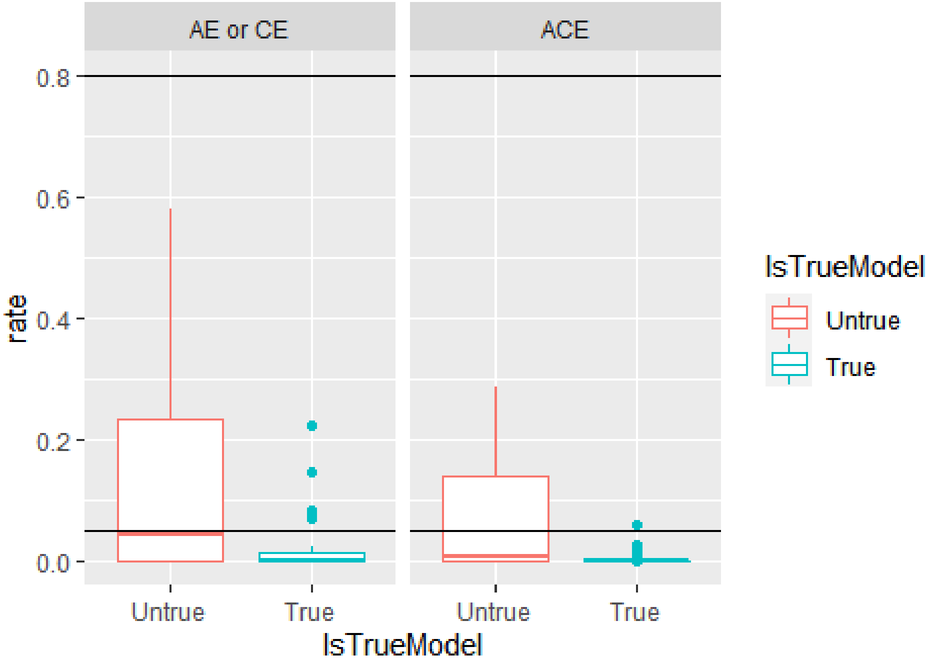

Figure 16 Distribution of model rejection rates using TLI across conditions by model type

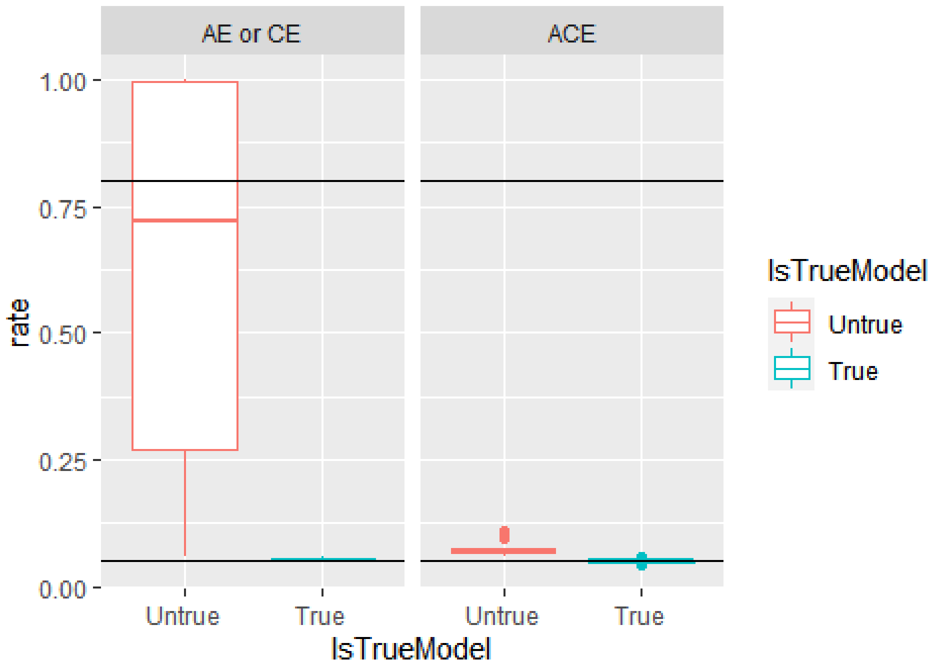

Figure 17 Distribution of model rejection rates using p(χ²) across conditions by model type

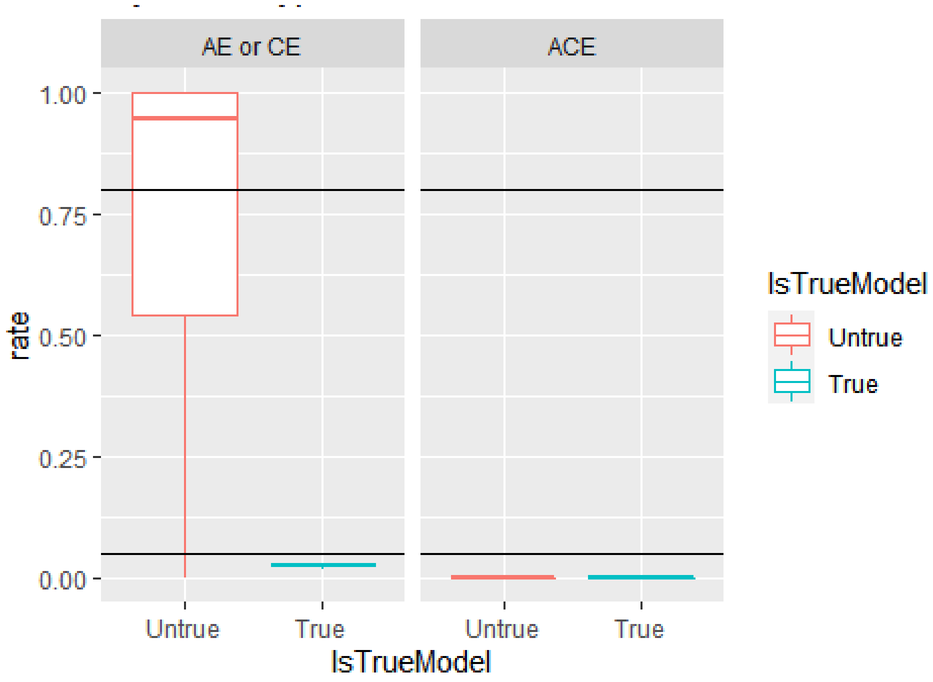

Figure 18 Distribution of model rejection rates using nested model comparisons across conditions by model type

For correctly specified models, median rejection rates ranged across type of model fit. For true ACE models, the median rejections rate ranged from 0 (TLI, nested χ^2 model comparisons) to 0.992 for p(RMSEA)<1. For our three metrics of focus: p(χ^2)<.05 had median rejection rates around .05, specifically 0.051 (min 0.043, max 0.06). Our other two metrics had rejection rates well below .05: TLI (0; min 0, max 0.058), and nested χ^2 model comparisons (0; min 0, max 0). For correctly specified two-parameter (AE or CE) models, the median rejection rate ranged from 0 (significant model parameters) to 0.99 for p(RMSEA)<1. p(χ^2)<.05 median rejection rates around .05: (0.051; min 0.046, max 0.06), while TLI's median rejection rate was well above .05: TLI (0; min 0, max 0.222). Lastly, nested χ^2 model comparisons had rejection rates well below .05 (0.025; min 0.018, max 0.03).

For incorrectly specified models, the median rejection rate ranged across conditions (i.e., power). Untrue ACE models ranged from 0 (significant model parameters) to 1 for p(RMSEA)<1. All three metrics had median rejection rates below .80: TLI (0.005; min 0, max 0.286), nested χ^2 model comparisons (0; min 0, max 0), and p(χ^2)<.05 (0.07; min 0.059, max 0.112). For incorrectly specified AE or CE models, the rejection rates ranged from 0 (significant model parameters) to 1 for p(RMSEA>.05). For our three metrics, nested χ^2 model comparisons had median rejection rates of at least .80: (0.943; min 0, max 1). However, TLI and p(χ^2)<.05 rejection rates were below .80: TLI (0.045; min 0, max 0.578); and p(χ^2)<.05 (0.719; min 0.057, max 1) and had considerable variability.

Brief Discussion. This stage examined power and type I error rates for distinguishing between correctly specified and incorrectly specified models using classic criterion. These examinations revealed that many classic criteria had Type I error rates that deviated from .05, drastically. Moreover, within each metric, there was considerable variability in Type I error rates across conditions, even when grouped by type of model fit, sample size, and total variance. All metrics had inflated Type I error rates in at least one condition: except for nested χ^2 comparisons and RMSEA's Lower Bound > 0.

Similarly, the power to reject an incorrect model using classic criteria deviated considerably across conditions ranging from 0 to 1, even when grouped by type of model fit (ACE versus CE or AE), sample size, and total variance. The metrics had decreased power rates (<.8) in at least one condition, with two exceptions. Only p(RMSEA>.05) < 1 and Lower Bound of RMSEA > 0 were consistently and sufficiently powered.

Optimized Thresholds

In this final subsubsection, I identify alternative thresholds, by using three common methods (Youden index, maximizing specificity while maintaining at least .80 sensitivity, and maximizing sensitivity while maintaining at least .95 for specificity), implemented with cutpointr package (Thiele, 2019). Conditionally maximizing specificity is the equivalent of minimizing alpha, given a minimum power level of .8. Conditionally maximizing specificity is the equivalent of maximizing power, given an alpha level of ≈ .05. Because the thresholds aim to be universal, I have focused on thresholds developed using the full data set with the exception of examining whether these thresholds differ when an ACE model is fit versus a model with fewer parameters (i.e., AE or CE).

Select Findings. Detailed results are provided in Tables 17, 18, and 19 of the appendix. Each table provides the optimized cutpoint, the maximized metric, the accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), true negative rate (TNR), true positive rate (TPR), false negative rate (FNR), false positive rate (FPR) for the full sample as well as by the parameter subgroups.

When the Youden index is maximized, most metrics varied by whether subgroups were accounted for (see Table 17). The TLI yielded similar thresholds regardless of grouping (along with a few other metrics); in contrast, the p(χ^2) and nested χ^2 model comparisons metric yielded results that differed by grouping. For a predictable subset of metrics, specifically those where a 3-parameter model was compared to a saturated model, thresholds were incalculable, resulting in values of negative or positive infinity. In those cases, nested χ^2 model comparisons, and differences in AIC, BIC, etc, yielded sensitivity scores of 0 and specificity scores of 1. When the sensitivity and specificity are examined for those Youden maximized thresholds, there was considerable variability. Sensitivity values ranged from 0.193 to 0.999; where specificity values ranged from 0.076 to 0.94.

It was not always possible to maximize specificity while maintaining at least .80 sensitivity (see Table 18, Appendix B). For example, a threshold for nested χ^2 model comparisons could not be made when the model had all three parameters. Indeed, most nested metrics could only be generated when the model was not saturated, such as the difference in AIC, aBIC, or BIC. Nevertheless, when all three groupings generated thresholds meeting the sensitivity minimum, some thresholds differed by grouping. However, TLI, and p(χ^2) thresholds did not differ by grouping. Even with minimum sensitivity levels, specificity often differed, ranging from 0.076 to 0.828.

Further and similarly, it was not always possible to maximizing sensitivity while maintaining at least .95 for specificity. When it was possible to generate thresholds for all groupings, TLI and RMSEA <.05 yielded similar results regardless of grouping; in contrast the remaining metrics yielded results that differed by grouping, including p(χ^2) and nested χ^2 model comparisons. Even with minimum specificity levels, sensitivity often differed, ranging from 0.005 to 0.321.

When generated thresholds are compared across the multiple optimizing methods, only TLI was consistent, whereas all others differed considerably. TLI's thresholds ranged from 0.972 to 1.006. Its' sensitivity ranged from 0.131 to 0.952 and specificity from 0.161 to 0.834. All other metrics either differed across optimization methods or by subgroup. For our other two metrics of interest, the nested model comparison was only useable for two-parameter models (AE or CE). In contrast, although $p(\chi^2)$ could be used on three-parameter models, its optimized thresholds varied considerably from 0.049 to 0.737.

Brief Discussion. Universal thresholds that had both power greater than .8 and type I error rates below .05 did not exist. The closest metric that could perhaps be universally applied was TLI; and yet TLI still had considerable variability in sensitivity and specificity across conditions. Like all metrics, TLI struggled to identified misspecified ACE models. Furthermore, many of the "optimized" thresholds for TLI were right at the boundary of what was possible ($TLI \geq 1$). In more pragmatic terms using this metric would require virtually zero model misspecification and even then, would only be helpful in identifying poorly fitting models. A perfect TLI score would tell little about whether a model fit well, only that it might not fit badly.

Chapter 5

Broad Discussion

Study 1 asked "How are behavior geneticists evaluating model fit?" Specifically, it sought to understand how behavior geneticists were evaluating model fit by examining two years of recent publications in the flagship journal *Behavior Genetics*. I evaluated whether these papers met the standard for best practice (5%), better practice (50%), and acceptable practice (72.5%). Very few papers met the best standard (Appelbaum et al., 2018; Lance et al., 2016; Mueller & Hancock, 2010), while 20% of papers did not evaluate fit at all.

Study 2 asked: "What are effective criteria to evaluate model fit for behavior genetic models?". It addressed this question by conducting a simulation study. Data were generated from a 3 x 3 x 2 x 4 x 2 design, where variance components (A, C, and E), sample size, and estimation method varied. These data were then fit to both incorrectly and correctly specified ACE, AE, and CE models in a standard twin design. The analyses were broken in three stages. Stage 1 examined descriptive statistics for correct and incorrect models. Stage 2 examined model rejection rates, using classic benchmarks as well as nested model comparisons. Stage 3 identified alternative thresholds, by using three common methods, with the intent of developing a universal metric.

Stage-specific results are discussed at the end of each sub-subsection in the Results section. Broadly, the answer is no; there are not universal criteria that can consistently distinguish between data fit to true versus untrue models with anything close to good sensitivity and specificity. This status will be discussed separately for the ACE models, and the AE and CE models.

The fundamental problem seems to be that for data fit to ACE models, it is borderline impossible to distinguish correct models from incorrect models. In the twin design used here, the ACE model is saturated, leaving no degrees of freedom to test model fit, which creates the challenge identified in this study. Further, power and type I error rates revealed that untrue ACE models were difficult to distinguish from true models of any type, using classic benchmarks. The ACE models were harder to correctly identify and thus, disproportionately contributed to the reduced power and inflated type I error rates. Accordingly, these two stages indicated that criteria would likely have greatest difficulty with detecting model misspecification in saturated models, i.e., ACE models.

From the study's onset, it was clear that just-identified models like ACE models were going to be challenging for fit statistics to provide useful diagnostic information. Indeed, the fact that Neale and Maes (2004) dedicated an entire section to model identification and fit, indicates that this challenge has been long apparent – and most likely previously discovered and just accepted as an inherent headache for modeling in the field of behavior genetics. Indeed, those metrics most used by Neale and Maes (2004) performed relatively better than those metrics not included in their text. This pattern supports that they likely recognized the problems identified within the current study. However, there is no discussion[12] in the behavior genetic literature to support this earlier recognition.

It is possible that some combination of indices that included both a highly specific metric and a highly sensitive metric might be created to jointly produce model selection rules that achieve both power greater than .8 and type I error rates at or below .05. Such a combination of indices would be consistent with best practices for evaluating fit for SEM models and would

---

[12] To the best of my knowledge.

have the chance of providing universal selection criteria – a series of golden rules rather than a single golden rule. Specifying such a combination was not a goal of the current study but is a useful suggestion for fruitful future research.

Nevertheless, if I had to select a single metric, it would have to be TLI. TLI was the closest metric that could be universally applied, and yet TLI still had considerable variability in sensitivity and specificity. Furthermore, many of the “optimized” thresholds were right at the boundary of what was possible: TLI ≥ 1. Accordingly, although this metric could be used universally, I still would not recommend it, as it is impractical, might encourage HARKing (hypothesizing after the results are known; Kerr, 1998; Rubin, 2017), and does not require that the researcher think critically about the model they are fitting.

Fit statistics thresholds at the boundary are impractical because they set an impossible expectation. Perfectly-fitting models using real data can only be achieved through overfitting or selectively reporting fit indices. Researchers could be incentivized to tinker with their models to meet those thresholds. The resulting exploratory model might be presented and/or interpreted as a confirmatory one. It is just an impossible expectation that does more harm than good. And frankly, any model that can perfectly reproduce its covariance matrix is not useful. A model is a simplification of reality, not a reproduction (Rodgers, 2010).

Chapter 6

Concluding Thoughts

I was tempted to provide criteria that differed depending on whether a saturated model (ACE) model was being used, or if a fully identified model (AE or CE) was being used. However, after considering the long-reaching implications and the confidence that declaring a well-fitting model brings, I decided against giving specific criteria. Readers who insist on custom criteria can find them in Tables 17-19 in the appendix. However, I honestly caution against eagerly accepting these benchmarks, especially for the ACE model and any saturated models. Even so, if you are evaluating a saturated model, in this paper's case the ACE model, I encourage you to be ruthlessly strict when evaluating your fit statistics. Further, I recommend that you be skeptical of any model whose fit statistics noticeably deviate from perfection. Even if your model "fits," I encourage you to fit additional models to their data that are fully identified, and not merely just-identified. Researchers can either use a design that has more than two kinship categories, add covariates, or at least fit a restricted model like an AE or CE model. If your model does indeed well fit, it will survive this skepticism, and indicate that it fits better relative the other models.

While drafting this paper, I reflected upon my own behavior genetic work, mulling over implications. Primarily, I considered my master's thesis (Garrison, 2016) and the papers that resulted from it (Garrison & Rodgers, 2017, 2019). It was my first behavior genetic project, with an ambitious aim to "untilt the SES-health gradient." I took a meticulous approach to evaluating my models; in part, this approach was driven by an eagerness to demonstrate my abilities to my department. But it was also motived by a skepticism in my own abilities. As a consequence of that zeal and skepticism, I am still confident that the models I selected were the best I could have arrived at. I am confident not because of the fit statistics, but because I eliminated every alternative model. I further tested my results in sensitivity analyses by adding kin groups and using different operationalizations of the constructs. I even replicated my analyses by using subsets of the data. Had I not taken this meticulous approach. I would be much less confident in my findings from that project.

More broadly, I think that combining that zealous skepticism with the general best practices of using multiple metrics from across classes of indices, should improve our ability to select the better fitting model – ideally the best identifiable model. Although the optimized metrics are using individual indices, future work should examine whether a combination of these metrics could achieve sufficient power and acceptable alpha levels. Indeed, if only one lesson is to be learned from these findings, it is the following: behavior genetic researchers should evaluate models beyond a "well-fitting" ACE model. A "well-fitting" ACE model does not indicate that that model is actually "well-fitting," even that the model is likely to be correct. In general, behavior genetics researchers should be ruthlessly skeptical of model fitting results, especially if that model is close to being just-identified.

References

Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, *73*(1), 3–25. https://doi.org/10.1037/amp0000191

Ayorech, Z., Selzam, S., Smith-Woolley, E., Knopik, V. S., Neiderhiser, J. M., DeFries, J. C., & Plomin, R. (2016). Publication Trends Over 55 Years of Behavioral Genetic Research. *Behavior Genetics*, *46*(5), 603–607. https://doi.org/10.1007/s10519-016-9786-2

Bekker, P., & Wansbeek, T. (2003). Identification in Parametric Models. In B. H. Baltagi (Ed.), *A Companion to Theoretical Econometrics* (pp. 144–161). Blackwell Publishing Ltd. https://doi.org/10.1002/9780470996249.ch8

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. https://doi.org/10/dbj

Boker, S., Neale, M., Maes, H., Wilde, M., Spiegel, M., Brick, T., Spies, J., Estabrook, R., Kenny, S., Bates, T. C., Mehta, P., & Fox, J. (2011). OpenMx: An Open Source Extended Structural Equation Modeling Framework. *Psychometrika*, *76*(2), 306–317. https://doi.org/10/fvqbph

Bollen, K. A., & Long, J. S. (Eds.). (1993). *Testing structural equation models*. Sage Publications, Inc.

Burks, B. S. (1938). On the Relative Contributions of Nature and Nurture to Average Group Differences in Intelligence. *Proceedings of the National Academy of Sciences of the United States of America*, *24*(7), 276–282. https://doi.org/10/c6nx97

Burt, C. (1966). The genetic determination of differences in intelligence: A study of monozygotic twins reared together and apart. *British Journal of Psychology*, *57*(1-2), 137–153. https://doi.org/10/bfcmfp

Carey, G. (2005). Cholesky problems. *Behavior Genetics*, *35*(5), 653–665. https://doi.org/10.1007/s10519-005-5355-9

Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., Mcclay, J., Mill, J., Martin, J., Braithwaite, A., & Poulton, R. (2003). Influence of Life Stress on Depression: Moderation by a Polymorphism in the 5-HTT Gene. *Science*, *301*(5631), 386–389. https://doi.org/10/bjfxw5

Chabris, C. F., Hebert, B. M., Benjamin, D. J., Beauchamp, J., Cesarini, D., van der Loos, M., Johannesson, M., Magnusson, P. K. E., Lichtenstein, P., Atwood, C. S., Freese, J., Hauser, T. S., Hauser, R. M., Christakis, N., & Laibson, D. (2012). Most Reported Genetic Associations With General Intelligence Are Probably False Positives. *Psychological Science*, *23*(11), 1314–1323. https://doi.org/10/f25m66

Cumming, G. (2014). The New Statistics: Why and How. *Psychological Science*, *25*(1), 7–29. https://doi.org/10.1177/0956797613504966

Eaves, L. J., & Gale, J. S. (1974). A method for analyzing the genetic basis of covariation. *Behavior Genetics*, *4*(3), 253–267. https://doi.org/10.1007/BF01074158

Falconer, D. S. (1952). The Problem of Environment and Selection. *The American Naturalist*, *86*(830), 293–298. https://doi.org/10/d3dtjx

Fisher, R. A. (1919a). The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Transactions of the Royal Society of Edinburgh*, *52*(2), 399–433. https://doi.org/10.1017/S0080456800012163

Fisher, R. A. (1919b). The Genesis of Twins. *Genetics*, *4*(5), 489–499.

Friedrich, J., Brand, B., Graunke, K. L., Langbein, J., Schwerin, M., & Ponsuksili, S. (2017). Adrenocortical Expression Profiling of Cattle with Distinct Juvenile Temperament Types. *Behavior Genetics*, *47*(1), 102–113. https://doi.org/10.1007/s10519-016-9816-0

Galton, F. (1876). The History of Twins, as a Criterion of the Relative Powers of Nature and Nurture. *The Journal of the Anthropological Institute of Great Britain and Ireland*, *5*, 391–406. JSTOR. https://doi.org/10/d8g7tp

Galton, F. (1896). Note to the Memoir by Professor Karl Pearson, F.R.S., on Spurious Correlation. *Proceedings of the Royal Society of London*, *60*, 498–502. JSTOR. http://www.jstor.org/stable/115880

Garlapow, M. E., Everett, L. J., Zhou, S., Gearhart, A. W., Fay, K. A., Huang, W., Morozova, T. V., Arya, G. H., Turlapati, L., St. Armour, G., Hussain, Y. N., McAdams, S. E., Fochler, S., & Mackay, T. F. C. (2016). Genetic and Genomic Response to Selection for Food Consumption in Drosophila melanogaster. *Behavior Genetics*, 1–17. https://doi.org/10.1007/s10519-016-9819-x

Garrison, S. M. (2016). Decomposing the Causes of the SES-Health Gradient with Biometrical Modeling [Master's Thesis, Vanderbilt University]. In *Multivariate Behavioral Research*. https://doi.org/10.1080/00273171.2016.1265434

Garrison, S. M. (2018). Popular Structural Equation Modeling Programs for Behavior Genetics. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(6), 972–977. https://doi.org/10.1080/10705511.2018.1493385

Garrison, S. M., & Ream, C. (2017). *Discord R Package*.

Garrison, S. M., & Rodgers, J. L. (2017). Decomposing the Causes of the Socioeconomic Status–Health Gradient with Biometrical Modeling. *Multivariate Behavioral Research*, *52*(1), 118–119. https://doi.org/10/gft74z

Garrison, S. M., & Rodgers, J. L. (2019). Decomposing the causes of the socioeconomic status-health gradient with biometrical modeling. *Journal of Personality and Social Psychology*, *116*(6), 1030–1047. https://doi.org/10.1037/pspp0000226

Gerbing, D. W., & Anderson, J. C. (1985). The effects of sampling error and model characteristics on parameter estimation for maximum likelihood confirmatory factor analysis. *Multivariate Behavioral Research*, *20*(3), 255–271.

Gerlai, R., Poshusta, T. L., Rampersad, M., Fernandes, Y., Greenwood, T. M., Cousin, M. A., Klee, E. W., & Clark, K. J. (2017). Forward Genetic Screening Using Behavioral Tests in Zebrafish: A Proof of Concept Analysis of Mutants. *Behavior Genetics*, *47*(1), 125–139. https://doi.org/10.1007/s10519-016-9818-y

Hallquist, M. N., & Wiley, J. F. (2018). MplusAutomation: An R Package for Facilitating Large-Scale Latent Variable Analyses in Mplus. *Structural Equation Modeling*, 1–18. https://doi.org/10.1080/10705511.2017.1402334

Hancock, G. R., & Mueller, R. O. (2010). *The Reviewer's Guide to Quantitative Methods*.

Hancock, G. R., & Mueller, R. O. (2011). The Reliability Paradox in Assessing Structural Relations Within Covariance Structure Models. *Educational and Psychological Measurement*, *71*(2), 306–324. https://doi.org/10.1177/0013164410384856

Harzing, A. W. (2018). *Publish or perish*.

Hernstein, R. J., & Murray, C. (1994). *The bell curve: The reshaping of American life by differences in intelligence* (1st Free Press pbk. ed). Free Press.

Holzinger, K. J. (1929). The relative effect of nature and nurture influences on twin differences. *Journal of Educational Psychology*, *20*(4), 241–248. https://doi.org/10.1037/h0072484

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, *6*(1), 53–60. https://doi.org/10.1037/1082-989X.12.1.58

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

Hu, L.-T., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, *112*(2), 351–362. https://doi.org/10.1037/0033-2909.112.2.351

Hunter, M. D., Garrison, S. M., & Burt, S. A. (2020). *A note on the analytic identification of variance component models common to behavior genetics*.

Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, *2*(8), 0696–0701. https://doi.org/10.1371/JOURNAL.PMED.0020124

Jackson, D. L., Gillaspy, J. A., & Purc-Stephenson, R. (2009). Reporting Practices in Confirmatory Factor Analysis: An Overview and Some Recommendations. In *Psychological Methods* (Vol. 14, Issue 1, pp. 6–23). AMER PSYCHOLOGICAL ASSOC. https://doi.org/10/cbd6zd

Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement. *Harvard Educational Review*, *39*(1), 1–123. https://doi.org/10/gdkxt4

Jinks, J. L., & Fulker, D. W. (1970). Comparison of the biometrical genetical, MAVA, and classical approaches to the analysis of human behavior. *Psychological Bulletin*, *73*(5), 311–349. https://doi.org/10/bmfrvb

Kang, Y., McNeish, D. M., & Hancock, G. R. (2016). The Role of Measurement Quality on Practical Guidelines for Assessing Measurement and Structural Invariance. *Educational and Psychological Measurement*, *76*(4), 533–561. https://doi.org/10/ggfnx3

Kenny, D. (n.d.). *SEM: Fit*. Retrieved April 11, 2019, from http://davidakenny.net/cm/fit.htm

Kerr, N. L. (1998). HARKing: Hypothesizing After the Results are Known. *Personality and Social Psychology Review*, *2*(3), 196–217. https://doi.org/10/dnqm8w

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

Lance, C. E., Beck, S. S., Fan, Y., & Carter, N. T. (2016). A taxonomy of path-related goodness-of-fit indices and recommended criterion values. *Psychological Methods*, *21*(3), 388–404. https://doi.org/10/gcz6zx

Lee, J. J., & McGue, M. (2016). Why Behavioral Genetics Matters. *Perspectives on Psychological Science*, *11*(1), 29–30. https://doi.org/10/gf8cw4

Liew, S. H. M., Elsner, H., Spector, T. D., & Hammond, C. J. (2005). The first “classical” twin study? Analysis of refractive error using monozygotic and dizygotic twins published in 1922. *Twin Research and Human Genetics*, *8*(3), 198–200. https://doi.org/10.1375/1832427054253158

Loehlin, J. C., & Vandenberg, S. G. (1968). Genetic and environmental components in the covariation of cognitive abilities: An additive model. In S. G. Vandenberg (Ed.), *Progress in Human Behavior Genetics*. Johns Hopkins University Press.

Lykken, D. T., McGue, M., Tellegen, A., & Bouchard, T. J., Jr. (1992). Emergenesis. Genetic traits that may not run in families. *American Psychologist*, *47*(12), 1565–1577. https://doi.org/10.1037/0003-066X.47.12.1565

Marsh, H., Hau, K.-T., & Wen, Z. (2004). In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler’s (1999) Findings. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2

Martin, N. G., & Eaves, L. J. (1977). The genetical analysis of covariance structure. *Heredity*, *38*(1), 79–95. https://doi.org/10.1038/hdy.1977.9

McNeish, D. (2018). Should We Use *F* -Tests for Model Fit Instead of Chi-Square in Overidentified Structural Equation Models? *Organizational Research Methods*, 109442811880949. https://doi.org/10.1177/1094428118809495

Merkle, E. C., You, D., & Preacher, K. J. (2016). Testing nonnested structural equation models. *Psychological Methods*, *21*(2), 151–163. https://doi.org/10/f8thxp

Millsap, R. E. (2012). A simulation paradigm for evaluating approximate fit. In M. C. Edwards & R. C. MacCallum (Eds.), *Current topics in the theory and application of latent variable models* (pp. 165–182). Routledge.

Mosing, M. A., Cnattingius, S., Gatz, M., Neiderhiser, J. M., & Pedersen, N. L. (2016). Associations Between Fetal Growth and Self-Perceived Health Throughout Adulthood: A Co-twin Control Study. *Behavior Genetics*, *46*(3), 457–466. https://doi.org/10.1007/s10519-015-9776-9

Mueller, R. O., & Hancock, G. R. (2010). Structural Equation Modeling. In *The Reviewer's Guide to Quantitative Methods*.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, *105*(3), 430–445. https://doi.org/10.1037/0033-2909.105.3.430

Murray, C. (1998). *Income inequality and IQ.* American Enterprise Inst. for Public Policy Research.

Muthén, L. K., & Muthén, B. O. (2017). *Mplus: Statistical analyses with latent variables. User's guide version 8* (Vol. 3).

Neale, M. C. (2009). *Mx Graphical User Interface*. https://vipbg.vcu.edu/resources/statistical-software/mxgui/#RASmfIwc

Neale, M. C., Hunter, M. D., Pritikin, J. N., Zahery, M., Brick, T. R., Kirkpatrick, R. M., Estabrook, R., Bates, T. C., Maes, H. H., & Boker, S. M. (2016). OpenMx 2.0: Extended Structural Equation and Statistical Modeling. *Psychometrika*, *81*(2), 535–549. https://doi.org/10/f8rfrg

Neale, M. C., & Maes, H. H. M. (2004). *Methodology for Genetic Studies of Twins and Families* (Vol. 48). Kluwer Academic Publishers B.V. https://doi.org/10.1136/jmg.30.9.800-a

Newman, H. H. (1917). *The biology of twins (mammals)*. University of Chicago Press.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716–aac4716. https://doi.org/10.1126/SCIENCE.AAC4716

Pearson, K. (1909). Determination of the coefficient of correlation. *Science*, *30*(757), 23–25. JSTOR. https://doi.org/10.1126/SCIENCE.30.757.23

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2016). Top 10 Replicated Findings From Behavioral Genetics. *Perspectives on Psychological Science*, *11*(1), 3–23. https://doi.org/10.1177/1745691615617439

Polderman, T. J. C., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, *47*(7), 702–709. https://doi.org/10/f3nbfg

Preacher, K. J., & Merkle, E. C. (2012). The problem of model selection uncertainty in structural equation modeling. *Psychological Methods*, *17*(1), 1–14. https://doi.org/10/fzbzph

R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rende, R. D., Plomin, R., & Vandenberg, S. G. (1990). Who discovered the twin method? *Behavior Genetics*, *20*(2), 277–285. https://doi.org/10.1007/BF01067795

Risch, N., Herrell, R., Lehner, T., Liang, K.-Y., Eaves, L., Hoh, J., Griem, A., Kovacs, M., Ott, J., & Merikangas, K. R. (2009). Interaction Between the Serotonin Transporter Gene (5-HTTLPR), Stressful Life Events, and Risk of Depression: A Meta-analysis. *JAMA*, *301*(23), 2462–2471. https://doi.org/10.1001/jama.2009.878

Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, *107*(2), 358–367. https://doi.org/10.1037/0033-295X.107.2.358

Rodgers, J. L. (2010). The epistemology of mathematical and statistical modeling: A quiet methodological revolution. *American Psychologist*, *65*(1), 1–12. https://doi.org/10/cc8fhq

Rodgers, J. L., Garrison, S. M., O’Keefe, P., Bard, D. E., Hunter, M. D., Beasley, W. H., & Oord, E. J. C. G. van den. (2019). Responding to a 100-Year-Old Challenge from Fisher: A Biometrical Analysis of Adult Height in the NLSY Data Using Only Cousin Pairs. *Behavior Genetics*. https://doi.org/10.1007/s10519-019-09967-6

Rodgers, J. L., & Rowe, D. C. (2002). Theory development should begin (but not end) with good empirical fits: A comment on Roberts and Pashler (2000). *Psychological Review*, *109*(3), 599–604; discussion 605-607. https://doi.org/10.1037/0033-295x.109.3.599

Rubin, M. (2017). When Does HARKing Hurt? Identifying When Different Types of Undisclosed Post Hoc Hypothesizing Harm Scientific Progress: *Review of General Psychology*. https://doi.org/10.1037/gpr0000128

Savalei, V. (2012). The Relationship Between Root Mean Square Error of Approximation and Model Misspecification in Confirmatory Factor Analysis Models. *Educational and Psychological Measurement*, *72*(6), 910–932. https://doi.org/10.1177/0013164412452564

Shrout, P. E., & Rodgers, J. L. (2017). Psychology, Science, and Knowledge Construction: Broadening Perspectives from the Replication Crisis. *Annual Review of Psychology*. https://doi.org/10.1146/ANNUREV-PSYCH-122216-011845

Siemens, H. W. (1924). *Die zwillingspathologie; ihre bedeutung, ihre methodik, ihre bisherigen ergebnisse*. Springer.

Silventoinen, K., Sammalisto, S., Perola, M., Boomsma, D. I., Cornes, B. K., Davis, C., Dunkel, L., de Lange, M., Harris, J. R., Hjelmborg, J. V. B. B., Luciano, M., Martin, N. G., Mortensen, J., Nisticò, L., Pedersen, N. L., Skytthe, A., Spector, T. D., Stazi, M. A., Willemsen, G., & Kaprio, J. (2003). Heritability of Adult Body Height: A Comparative Study of Twin Cohorts in Eight Countries. *Twin Research*, *6*(05), 399–408. https://doi.org/10.1375/TWIN.6.5.399

Smith, W. T. (1856). A Course of Lectures ON THE THEORY AND PRACTICE OF OBSTETRICS. *The Lancet*, *67*(1703), 419–423. https://doi.org/10.1016/S0140-6736(02)55452-0

Sniekers, S., Stringer, S., Watanabe, K., Jansen, P. R., Coleman, J. R. I., Krapohl, E., Taskesen, E., Hammerschlag, A. R., Okbay, A., Zabaneh, D., Amin, N., Breen, G., Cesarini, D., Chabris, C. F., Iacono, W. G., Ikram, M. A., Johannesson, M., Koellinger, P., Lee, J. J., … Posthuma, D. (2017). Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nature Genetics*, *49*(7), 1107–1112. https://doi.org/10/b7gm

Snygg, D. (1938). The Relation Between the Intelligence of Mothers and of Their Children Living in Foster Homes. *The Pedagogical Seminary and Journal of Genetic Psychology*, *52*(2), 401–406. https://doi.org/10.1080/08856559.1938.10534325

Stigler, S. M. (1989). Francis Galton's Account of the Invention of Correlation. *Statistical Science*, *4*(2), 73–79. https://doi.org/10.1214/ss/1177012580

Thiele, C. (2019). *cutpointr: Determine and Evaluate Optimal Cutpoints in Binary Classification Tasks* (1.0.1) [R]. https://CRAN.R-project.org/package=cutpointr

Thorndike, E. L. (1905). Measurement of Twins. *The Journal of Philosophy, Psychology and Scientific Methods*, *2*(20), 547–553. JSTOR. https://doi.org/10/c8c7xk

Turkheimer, E. (2016). Weak Genetic Explanation 20 Years Later: Reply to Plomin et al. (2016). *Perspectives on Psychological Science*, *11*(1), 24–28. https://doi.org/10.1177/1745691615617442

Turkheimer, E., & Waldron, M. (2000). Nonshared environment: A theoretical, methodological, and quantitative review. *Psychological Bulletin*, *126*(1), 78–108. https://doi.org/10/c6ndk8

van den Berg, S. M., de Moor, M. H. M., Verweij, K. J. H. H., Krueger, R. F., Luciano, M., Arias-Vásquez, A., Matteson, L. K., Derringer, J., Esko, T., Amin, N., Gordon, S. D., Hansell, N. K., Hart, A. B., Seppälä, I., Huffman, J. E., Konte, B., Lahti, J., Lee, M., Miller, M., … ... (2016). Meta-analysis of Genome-Wide Association Studies for Extraversion: Findings from the Genetics of Personality Consortium. *Behavior Genetics*, *46*(2), 170–182. https://doi.org/10.1007/s10519-015-9735-5

Verhulst, B. (2017). A Power Calculator for the Classical Twin Design. *Behavior Genetics*, *47*(2), 255–261. https://doi.org/10.1007/s10519-016-9828-9

Verhulst, B., Prom-Wormley, E., Keller, M., Medland, S., & Neale, M. C. (2019). Type I Error Rates and Parameter Bias in Multivariate Behavioral Genetic Models. *Behavior Genetics*, *49*(1), 99–111. https://doi.org/10.1007/S10519-018-9942-Y

Whitehead, H., Vachon, F., & Frasier, T. R. (2017). Cultural Hitchhiking in the Matrilineal Whales. *Behavior Genetics*, *47*(3), 324–334. https://doi.org/10.1007/s10519-017-9840-8

Widaman, K. F., & Thompson, J. S. (2003). On Specifying the Null Model for Incremental Fit Indices in Structural Equation Modeling. *Psychological Methods*, *8*(1), 16–37. https://doi.org/10/dpps6n

Yerushalmy, J. (1947). Statistical problems in assessing methods of medical diagnosis, with special reference to X-ray techniques. *Public Health Reports (Washington, D.C.: 1896)*, *62*(40), 1432–1449.

Appendix A. Code

R Function to Simulate Biometrically informed Multivariate Data

```
#' Simulate Biometrically informed Multivariate Data
#' @description Generate paired multivariate data, given ACE parameters.
#' @importFrom stats rnorm sd
#' @param r_all Levels of relatedness; default is MZ and DZ twins c(1,.5).
#' @param npg_all Sample size per group; default is 500.
#' @param npergroup_all Vector of sample sizes by group; default repeats
\code{npg_all} for all groups
#' @param variables Number of variables to generate; default is 2. Current max 2
#' @param mu_all Mean for each generated variable; default is 0.
#' @param mu_list List of means by variable; default repeats \code{mu_all} for all
variables
#' @param r_vector Alternative, give vector of r coefficients for entire sample.
#' @param ace_all Vector of variance components for each generated variable; default
is c(1,1,1).
#' @param ace_list Matrix of ACE variance components by variable, where each row is
its own variable; default is to repeat \code{ace_all} for each variable.
#' @param ... Optional pass on additional inputs.
#' @param reliability_all Optional Additional error. Indicates correlation between
True Y and Measured Y. Default is 1
#' @param prop_var_explained_all Optional Additional error. Indicates proportion of
variance explained in True Y from Measured Y. Default is \code{reliability_all}^2
#' @param reliability_list Vector of Reliabilities for each generated variable;
default is to repeat \code{reliability_all} for each variable
#' @param prop_var_explained_list Vector of R^2 for each generated variable; default
is to repeat \code{prop_var_explained_all} for each variable
#'@param cov_a Shared variance for additive genetics (a); default is 1
#'@param cov_c Shared variance for shared-environment (c); default is 1
#'@param cov_e shared variance for non-shared-environment (e); default is 1
#'@param model Model type. Default is correlated factors model "Correlated";
alterative specification as a "Cholesky" model, where variable 1 accounts for
variance in variable 2, is currently disabled.

#' @return Returns \code{data.frame} with the following:
#' \item{Ai_1}{genetic component for variable i for kin1}
#' \item{Ai_2}{genetic component for variable i for kin2}
#' \item{Ci_1}{shared-environmental component for variable i for kin1}
```

```
#' \item{Ci_2}{shared-environmental component for variable i for kin2}
#' \item{Ei_1}{non-shared-environmental component for variable i for kin1}
#' \item{Ei_2}{non-shared-environmental component for variable i for kin2}
#' \item{yi_1}{generated variable i for kin1}
#' \item{yi_2}{generated variable i for kin2}
#' \item{r}{level of relatedness for the kin pair}
#' \item{id}{id}

kinsim_multi <- function(
  r_all=c(1,.5),
  npg_all=500,
  npergroup_all=rep(npg_all,length(r_all)),
  mu_all=0,
  variables=2,
  mu_list=rep(mu_all,variables),
  reliability_list=NULL,
  r_vector=NULL, # alternative specification, give vector of rs
  ace_all=c(1,1,1), # variance default
  ace_list=matrix(rep(ace_all,variables),byrow=TRUE,nrow=variables),
  cov_a=0, #default shared variance for genetics
  cov_c=0, #default shared variance for c
  cov_e=0, #default shared variance for e
  ...){
  mu=NULL
  sA <- ace_list[,1]^0.5; sC <- ace_list[,2]^0.5; sE <- ace_list[,3]^0.5
  S2 <- diag(4)*-1+1

  datalist <- list()
  if(variables==1){
    data_v<-kinsim1(r=r_all,
                    npergroup=npergroup_all,  #
                    mu=mu_list[1],                   #intercept
                    ace= ace_list[[1]],r_vector=r_vector
    )
    data_v$A1_u<-data_v$A1
    data_v$A2_u<-data_v$A2
```

```
    data_v$C1_u<-data_v$C1
    data_v$C2_u<-data_v$C2
    data_v$E1_u<-data_v$E1
    data_v$E2_u<-data_v$E2
    data_v$y1_u<-data_v$y1
    data_v$y2_u<-data_v$y2

    merged.data.frame =data_v
    names(merged.data.frame)[c(1,10)]<-c("id","r")
  }
  if(variables>2){
    stop("You have tried to generate data beyond the current limitations of this
program. Maximum variables 2.")
  }
    if(is.null(r_vector)){
      id=1:sum(npergroup_all)
      for(i in 1:length(r_all)){
        n = npergroup_all[i]

        # Genetic Covariance
        sigma_a<-diag(4)+S2*r_all[i]
        sigma_a[1,3]<-cov_a;
        sigma_a[3,1]<-cov_a;sigma_a[2,4]<-cov_a;sigma_a[4,2]<-cov_a
        sigma_a[1,4]<-cov_a*r_all[i];sigma_a[4,1]<-cov_a*r_all[i];sigma_a[3,2]<-
cov_a*r_all[i];sigma_a[2,3]<-cov_a*r_all[i]
        A.r <- rmvn(n,sigma=sigma_a)
        A.r[,1:2]<- A.r[,1:2]*sA[1]; A.r[,3:4]<- A.r[,3:4]*sA[2]

        # Shared C Covariance
        sigma_c<-diag(4)+S2*1
        sigma_c[1,3]<-cov_c;sigma_c[3,1]<-cov_c;sigma_c[2,4]<-cov_c;sigma_c[4,2]<-
cov_c
        sigma_c[1,4]<-cov_c*1;sigma_c[4,1]<-cov_c*1;sigma_c[3,2]<-
cov_c*1;sigma_c[2,3]<-cov_c*1
        C.r <- rmvn(n,sigma=sigma_c)
        C.r[,1:2]<- C.r[,1:2]*sC[1]; C.r[,3:4]<- C.r[,3:4]*sC[2]

        # Shared E Covariance
```

```
        sigma_e<-diag(4)+S2*0
        sigma_e[1,3]<-cov_e;sigma_e[3,1]<-cov_e;sigma_e[2,4]<-cov_e;sigma_e[4,2]<-cov_e
        E.r <- rmvn(n,sigma=sigma_e)
        E.r[,1:2]<- E.r[,1:2]*sE[1]; E.r[,3:4]<- E.r[,3:4]*sE[2]

        y.r <-  A.r + C.r + E.r
        y.r[,1:2]<-y.r[,1:2]+mu_list[1]
        y.r[,3:4]<-y.r[,3:4]+mu_list[2]
        r_ <- rep(r_all[i],n)
        data.r<-data.frame(A.r,C.r,E.r,y.r,r_)
        names(data.r)<-c("A1_1","A1_2","A2_1","A2_2","C1_1","C1_2","C2_1","C2_2","E1_1","E1_2","E2_1","E2_2","y1_1","y1_2","y2_1","y2_2","r")

        datalist[[i]] <- data.r
        names(datalist)[i]<-paste0("datar",r_all[i])
        print(r_all[i])
      }
      merged.data.frame = Reduce(function(...) merge(..., all=T), datalist)
      merged.data.frame$id<-id
    }else{
      id=1:length(r_vector)
data_vector=data.frame(id,r_vector,matrix(rep(as.numeric(NA),length(id)*4),nrow=length(id),ncol=4))
      names(data_vector)<-c("id","r","A1_1","A1_2","A2_1","A2_2")
      unique_r= matrix(unique(r_vector))
      for(i in 1:length(unique_r)){
        n=length(r_vector[r_vector==unique_r[i]])
        # Genetic Covariance
        sigma_a<-diag(4)+S2*unique_r[i]
        sigma_a[1,3]<-cov_a;
        sigma_a[3,1]<-cov_a;sigma_a[2,4]<-cov_a;sigma_a[4,2]<-cov_a
        sigma_a[1,4]<-cov_a*unique_r[i];sigma_a[4,1]<-cov_a*unique_r[i];sigma_a[3,2]<-cov_a*unique_r[i];sigma_a[2,3]<-cov_a*unique_r[i]
        A.r <- rmvn(n,sigma=sigma_a)
        data_vector$A1_1[data_vector$r_vector==unique_r[i]] <- A.r[,1]*sA[1]
        data_vector$A1_2[data_vector$r_vector==unique_r[i]] <- A.r[,2]*sA[1]
```

```
          data_vector$A2_1[data_vector$r_vector==unique_r[i]] <- A.r[,3]*sA[2]
          data_vector$A2_2[data_vector$r_vector==unique_r[i]] <- A.r[,4]*sA[2]
          A.r[,1:2]<- A.r[,1:2]; A.r[,3:4]<- A.r[,3:4]*sA[2]}
        n=length(r_vector)
        A.r<-
matrix(c(data_vector$A1_1,data_vector$A1_2,data_vector$A2_1,data_vector$A2_2),ncol=4,
nrow=n)
        # Shared C Covariance
        sigma_c<-diag(4)+S2*1
        sigma_c[1,3]<-cov_c;sigma_c[3,1]<-cov_c;sigma_c[2,4]<-cov_c;sigma_c[4,2]<-cov_c
        sigma_c[1,4]<-cov_c*1;sigma_c[4,1]<-cov_c*1;sigma_c[3,2]<-
cov_c*1;sigma_c[2,3]<-cov_c*1
        C.r <- rmvn(n,sigma=sigma_c)
        C.r[,1:2]<- C.r[,1:2]*sC[1]; C.r[,3:4]<- C.r[,3:4]*sC[2]
        # Shared E Covariance
        sigma_e<-diag(4)+S2*0
        sigma_e[1,3]<-cov_e;sigma_e[3,1]<-cov_e;sigma_e[2,4]<-cov_e;sigma_e[4,2]<-cov_e
        E.r <- rmvn(n,sigma=sigma_e)
        E.r[,1:2]<- E.r[,1:2]*sE[1]; E.r[,3:4]<- E.r[,3:4]*sE[2]


        y.r <- A.r
        y.r[,1:2]<-A.r[,1:2]*ace_list[1,1] + C.r[,1:2]*ace_list[1,2] +
E.r[,1:2]*ace_list[1,3]
        y.r[,3:4]<-A.r[,3:4]*ace_list[2,1] + C.r[,3:4]*ace_list[2,2] +
E.r[,3:4]*ace_list[2,3]
        y.r[,1:2]<-y.r[,1:2]+mu_list[1]
        y.r[,3:4]<-y.r[,3:4]+mu_list[2]
        y.r <- mu + A.r + C.r + E.r
        data.r<-data.frame(A.r,C.r,E.r,y.r,r_vector,id)
        names(data.r)<-
c("A1_1","A1_2","A2_1","A2_2","C1_1","C1_2","C2_1","C2_2","E1_1","E1_2","E2_1","E2_2"
,"y1_1","y1_2","y2_1","y2_2","r","id")
      datalist[[i]] <- data.r
      names(datalist)[i]<-paste0("datar",r_all[i])
      merged.data.frame = data.r}
    return(merged.data.frame)}
```

Sample Mplus Automation Code

```
[[init]]
iterators = conditions sample;
sample = 1:5000;
conditions = 1:72;
filename = "uni_cov_ace_[[conditions]]_[[sample]].inp";
outputDirectory = "P:\Dissertation\covariance\[[conditions]]";
size#conditions= 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100 100
100 250 250 250 250 250 250 250 250 250 250 250 250 250 250 250 250 250 250 500 500
500 500 500 500 500 500 500 500 500 500 500 500 500 500 500 500 1000 1000 1000 1000
1000 1000 1000 1000 1000 1000 1000 1000 1000 1000 1000 1000 1000 1000;
[[/init]]
TITLE: ACE model cov condition [[conditions]] sample [[sample]]
DATA: FILE = "cov_ds_[[conditions]]_[[sample]].dat";
TYPE IS MEANS COVARIANCE;
NGROUPS = 2;
NOBSERVATIONS = [[size#conditions]] [[size#conditions]];
VARIABLE:
 NAMES =
Y_1
Y_2;
missing=.;
USEVAR are
  Y_1 Y_2;
ANALYSIS:
MODEL=NOCOVARIANCES;
ITERATIONS = 25000;

MODEL: !set up values for all groups
Y_1@0; Y_2@0; !fix residual variances to zero
A1 BY Y_1*.5 (a1); A2 BY Y_2*.5 (a1); !additive genetic loadings
C1 BY Y_1*.5 (12); C2 BY Y_2*.5 (12); !common envt loadings
E1 BY Y_1*.5 (13); E2 BY Y_2*.5 (13); !specific envt loadings

 !fix latent variable means=0
[A1@0 A2@0];
[C1@0 C2@0];
```

```
[E1@0 E2@0];

!fix latent variable vars
A1@1 A2@1 ;
C1@1 C2@1 ;
E1@1 E2@1;

!latent variable corrs
A1 WITH C1-C2@0; A2 WITH C1-C2@0;
A1 WITH E1-E2@0; A2 WITH E1-E2@0;
C1 WITH C2@1; C1 WITH E1-E2@0; C2 WITH E1-E2@0;
E1 WITH E2@0;

!MODEL G1:
[Y_1 Y_2] (m); !means
A1 WITH A2@1;

MODEL G2:
[Y_1 Y_2] (m); !means
A1 WITH A2@0.5;

OUTPUT: STDyx TECH1 TECH3 TECH4 CINTERVAL;
```

Appendice B. Tables

Table 5 Descriptive Statistics for E data fit to ACE models (Condition 1)

Quantile	1	3	5	10	25	50	75	90	95	98	99
ChiSqM_Value	1.323	1.804	2.283	3.072	4.586	6.701	9.506	12.645	14.820	17.013	19.672
ChiSqM_PValue	0.003	0.009	0.022	0.049	0.147	0.349	0.598	0.800	0.892	0.937	0.970
LL	-598.672	-593.788	-589.537	-584.664	-576.251	-566.466	-556.926	-547.671	-542.844	-537.935	-533.143
UnrestrictedLL	-595.746	-590.411	-585.845	-580.984	-572.651	-562.719	-553.170	-544.051	-538.559	-534.021	-529.321
CFI	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000
TLI	-9.206	-3.119	-1.125	0.028	0.878	1.000	1.000	1.005	1.470	2.327	5.072
AIC	1074.286	1083.870	1093.688	1103.341	1121.852	1140.931	1160.502	1177.328	1187.075	1195.575	1205.345
BIC	1087.480	1097.063	1106.881	1116.535	1135.046	1154.125	1173.695	1190.522	1200.268	1208.769	1218.538
aBIC	1074.807	1084.391	1094.209	1103.862	1122.373	1141.452	1161.022	1177.849	1187.595	1196.096	1205.865
RMSEA_Estimate	0.000	0.000	0.000	0.000	0.000	0.034	0.076	0.105	0.121	0.135	0.151
RMSEA_90CI_LB	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.007	0.043	0.062	0.080
RMSEA_90CI_UB	0.000	0.032	0.059	0.083	0.111	0.138	0.164	0.187	0.201	0.213	0.228
RMSEA_pLT05	0.014	0.032	0.062	0.117	0.269	0.503	0.728	0.878	0.938	0.965	0.984
SRMR	0.042	0.049	0.056	0.064	0.080	0.098	0.119	0.141	0.155	0.168	0.190
AICC	1074.491	1084.075	1093.893	1103.546	1122.057	1141.136	1160.707	1177.533	1187.280	1195.780	1205.550
a_uns_est	-0.001	0.000	0.000	0.000	0.000	0.000	0.137	0.292	0.348	0.390	0.432
a_uns_est_se	-0.001	-0.001	0.000	0.000	0.000	0.000	0.293	1.728	2.625	3.291	4.103
a_uns_pval	0.000	0.001	0.009	0.084	0.770	1.000	1.000	1.000	1.000	1.000	1.000
c_uns_est	-0.001	0.000	0.000	0.000	0.000	0.000	0.119	0.250	0.302	0.342	0.387
c_uns_est_se	-0.001	-0.001	0.000	0.000	0.000	0.000	0.280	1.588	2.400	3.118	4.028
c_uns_pval	0.000	0.002	0.017	0.112	0.780	1.000	1.000	1.000	1.000	1.000	1.000
e_uns_est	0.872	0.889	0.904	0.922	0.951	0.981	1.009	1.035	1.048	1.061	1.076
e_uns_est_se	15.137	15.716	16.260	16.749	17.586	20.001	28.284	28.284	28.284	28.285	28.285
e_uns_pval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
a_stz_est	-0.001	0.000	0.000	0.000	0.000	0.000	0.137	0.292	0.348	0.387	0.434
a_stz_est_se	-0.001	-0.001	0.000	0.000	0.000	0.000	0.293	1.739	2.660	3.360	4.238
a_stz_pval	0.000	0.001	0.008	0.082	0.770	1.000	1.000	1.000	1.000	1.000	1.000
c_stz_est	-0.001	0.000	0.000	0.000	0.000	0.000	0.119	0.252	0.303	0.343	0.384
c_stz_est_se	-0.001	-0.001	0.000	0.000	0.000	0.000	0.280	1.595	2.426	3.177	4.156
c_stz_pval	0.000	0.001	0.015	0.111	0.780	1.000	1.000	1.000	1.000	1.000	1.000
e_stz_est	0.897	0.912	0.927	0.942	0.968	0.995	1.000	1.000	1.000	1.000	1.000
e_stz_pval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
totalvar	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
aBIC_sat	1074.807	1084.391	1094.209	1103.862	1122.373	1141.452	1161.022	1177.849	1187.595	1196.096	1205.865
AIC_sat	1074.286	1083.870	1093.688	1103.341	1121.852	1140.931	1160.502	1177.328	1187.075	1195.575	1205.345
BIC_sat	1087.480	1097.063	1106.881	1116.535	1135.046	1154.125	1173.695	1190.522	1200.268	1208.769	1218.538
ChiSqM_Value_sat	1.323	1.804	2.283	3.072	4.586	6.701	9.506	12.645	14.820	17.013	19.672
ChiSqM_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
nested_sig	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
AIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
aBIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 6 Descriptive Statistics for E data fit to ACE models (Condition 10)

Quantile	1	2.5	5	10	25	50	75	90	95	97.5	99
ChiSqM_Value	1.391	1.928	2.428	3.105	4.653	6.820	9.715	12.941	14.914	16.973	19.878
ChiSqM_PValue	0.003	0.009	0.021	0.044	0.137	0.338	0.589	0.796	0.876	0.926	0.966
LL	-737.782	-731.646	-727.712	-722.461	-714.332	-705.091	-695.271	-686.476	-681.126	-676.607	-672.264
UnrestrictedLL	-734.339	-728.363	-723.827	-718.948	-710.753	-701.368	-691.696	-682.523	-677.137	-672.913	-668.373
CFI	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000
TLI	-8.806	-3.026	-1.179	0.008	0.875	1.000	1.000	1.000	1.441	2.162	4.167
AIC	1352.528	1361.214	1370.253	1380.951	1398.542	1418.181	1436.663	1452.922	1463.423	1471.292	1483.565
BIC	1365.721	1374.407	1383.446	1394.144	1411.735	1431.375	1449.856	1466.115	1476.617	1484.485	1496.758
aBIC	1353.048	1361.735	1370.774	1381.472	1399.063	1418.702	1437.184	1453.443	1463.944	1471.813	1484.085
RMSEA_Estimate	0.000	0.000	0.000	0.000	0.000	0.037	0.079	0.108	0.122	0.135	0.152
RMSEA_90CI_LB	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.044	0.062	0.081
RMSEA_90CI_UB	0.000	0.041	0.065	0.084	0.112	0.139	0.165	0.189	0.201	0.213	0.229
RMSEA_pLT05	0.013	0.032	0.061	0.108	0.255	0.491	0.720	0.875	0.928	0.959	0.982
SRMR	0.043	0.050	0.057	0.065	0.080	0.098	0.120	0.142	0.158	0.173	0.190
AICC	1352.733	1361.420	1370.458	1381.156	1398.747	1418.386	1436.868	1453.128	1463.628	1471.497	1483.770
a_uns_est	-0.001	0.000	0.000	0.000	0.000	0.000	0.157	0.400	0.483	0.540	0.601
a_uns_est_se	-0.002	-0.001	-0.001	0.000	0.000	0.000	0.194	1.526	2.462	3.211	4.081
a_uns_pval	0.000	0.001	0.014	0.127	0.847	1.000	1.000	1.000	1.000	1.000	1.000
c_uns_est	-0.001	-0.001	0.000	0.000	0.000	0.000	0.167	0.359	0.439	0.493	0.555
c_uns_est_se	-0.002	-0.002	-0.001	0.000	0.000	0.000	0.284	1.642	2.511	3.347	4.198
c_uns_pval	0.000	0.001	0.012	0.100	0.777	1.000	1.000	1.000	1.000	1.000	1.000
e_uns_est	1.227	1.254	1.277	1.304	1.346	1.388	1.428	1.462	1.481	1.497	1.519
e_uns_est_se	15.072	15.719	16.246	16.823	17.706	20.002	28.284	28.284	28.284	28.285	28.285
e_uns_pval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
a_stz_est	-0.001	0.000	0.000	0.000	0.000	0.000	0.112	0.283	0.339	0.385	0.426
a_stz_est_se	-0.002	-0.001	-0.001	0.000	0.000	0.000	0.194	1.533	2.491	3.276	4.211
a_stz_pval	0.000	0.001	0.013	0.125	0.847	1.000	1.000	1.000	1.000	1.000	1.000
c_stz_est	-0.001	0.000	0.000	0.000	0.000	0.000	0.117	0.254	0.313	0.351	0.394
c_stz_est_se	-0.002	-0.002	-0.001	0.000	0.000	0.000	0.284	1.649	2.541	3.419	4.343
c_stz_pval	0.000	0.001	0.011	0.099	0.777	1.000	1.000	1.000	1.000	1.000	1.000
e_stz_est	0.895	0.911	0.926	0.942	0.969	0.996	1.000	1.000	1.000	1.000	1.000
e_stz_pval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
totalvar	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000
aBIC_sat	1353.048	1361.735	1370.774	1381.472	1399.063	1418.702	1437.184	1453.443	1463.944	1471.813	1484.085
AIC_sat	1352.528	1361.214	1370.253	1380.951	1398.542	1418.181	1436.663	1452.922	1463.423	1471.292	1483.565
BIC_sat	1365.721	1374.407	1383.446	1394.144	1411.735	1431.375	1449.856	1466.115	1476.617	1484.485	1496.758
ChiSqM_Value_sat	1.391	1.928	2.428	3.105	4.653	6.820	9.715	12.941	14.914	16.973	19.878
ChiSqM_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
nested_sig	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
AIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
aBIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 7 Descriptive Statistics for E data fit to ACE models (Condition 19)

Quantile	1	2.5	5	10	25	50	75	90	95	97.5	99
ChiSqM_Value	1.265	1.854	2.401	3.160	4.550	6.797	9.654	12.887	14.805	16.656	19.231
ChiSqM_PValue	0.004	0.011	0.022	0.045	0.140	0.340	0.603	0.789	0.879	0.933	0.974
LL	-1469.745	-1460.947	-1453.793	-1446.166	-1432.225	-1417.880	-1402.326	-1387.864	-1379.419	-1372.824	-1365.625
UnrestrictedLL	-1465.891	-1457.880	-1450.060	-1442.412	-1428.685	-1413.644	-1398.633	-1384.321	-1375.429	-1368.959	-1362.002
CFI	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000
TLI	-12.260	-4.677	-1.536	-0.155	0.823	1.000	1.000	1.000	1.380	2.100	3.718
AIC	2739.249	2753.648	2766.839	2783.728	2812.651	2843.760	2872.450	2900.331	2915.586	2929.893	2947.489
BIC	2756.108	2770.506	2783.697	2800.587	2829.509	2860.618	2889.309	2917.190	2932.444	2946.751	2964.348
aBIC	2743.411	2757.810	2771.001	2787.891	2816.813	2847.922	2876.613	2904.494	2919.748	2934.055	2951.652
RMSEA_Estimate	0.000	0.000	0.000	0.000	0.000	0.023	0.049	0.068	0.077	0.084	0.094
RMSEA_90CI_LB	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.010	0.027	0.037	0.049
RMSEA_90CI_UB	0.000	0.023	0.040	0.054	0.070	0.088	0.104	0.119	0.127	0.134	0.143
RMSEA_pLT05	0.054	0.102	0.157	0.239	0.441	0.676	0.855	0.939	0.969	0.985	0.995
SRMR	0.026	0.032	0.036	0.041	0.050	0.062	0.075	0.088	0.095	0.101	0.111
AICC	2739.330	2753.729	2766.919	2783.809	2812.731	2843.841	2872.531	2900.412	2915.667	2929.974	2947.570
a_uns_est	-0.001	0.000	0.000	0.000	0.000	0.000	0.111	0.232	0.280	0.311	0.346
a_uns_est_se	-0.002	-0.001	-0.001	0.000	0.000	0.000	0.327	1.704	2.605	3.337	4.041
a_uns_pval	0.000	0.001	0.009	0.088	0.744	1.000	1.000	1.000	1.000	1.000	1.000
c_uns_est	-0.001	0.000	0.000	0.000	0.000	0.000	0.087	0.204	0.243	0.273	0.305
c_uns_est_se	-0.004	-0.002	-0.001	0.000	0.000	0.000	0.247	1.614	2.490	3.248	4.049
c_uns_pval	0.000	0.001	0.013	0.107	0.806	1.000	1.000	1.000	1.000	1.000	1.000
e_uns_est	0.920	0.930	0.941	0.951	0.969	0.988	1.006	1.021	1.030	1.037	1.047
e_uns_est_se	24.349	25.143	26.170	26.930	27.804	31.625	44.721	44.721	44.722	44.722	44.722
e_uns_pval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
a_stz_est	-0.001	0.000	0.000	0.000	0.000	0.000	0.112	0.232	0.280	0.311	0.343
a_stz_est_se	-0.002	-0.001	-0.001	0.000	0.000	0.000	0.327	1.708	2.619	3.364	4.090
a_stz_pval	0.000	0.001	0.009	0.088	0.744	1.000	1.000	1.000	1.000	1.000	1.000
c_stz_est	-0.001	0.000	0.000	0.000	0.000	0.000	0.088	0.203	0.243	0.274	0.304
c_stz_est_se	-0.004	-0.002	-0.001	0.000	0.000	0.000	0.247	1.618	2.502	3.274	4.099
c_stz_pval	0.000	0.001	0.012	0.106	0.806	1.000	1.000	1.000	1.000	1.000	1.000
e_stz_est	0.936	0.946	0.953	0.963	0.979	0.997	1.000	1.000	1.000	1.000	1.000
e_stz_pval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
totalvar	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
aBIC_sat	2743.411	2757.810	2771.001	2787.891	2816.813	2847.922	2876.613	2904.494	2919.748	2934.055	2951.652
AIC_sat	2739.249	2753.648	2766.839	2783.728	2812.651	2843.760	2872.450	2900.331	2915.586	2929.893	2947.489
BIC_sat	2756.108	2770.506	2783.697	2800.587	2829.509	2860.618	2889.309	2917.190	2932.444	2946.751	2964.348
ChiSqM_Value_sat	1.265	1.854	2.401	3.160	4.550	6.797	9.654	12.887	14.805	16.656	19.231
ChiSqM_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
nested_sig	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
AIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
aBIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 8 Descriptive Statistics for E data fit to ACE models (Condition 28)

Quantile	1	2.5	5	10	25	50	75	90	95	97.5	99
ChiSqM_Value	1.237	1.744	2.288	3.034	4.481	6.800	9.602	12.706	14.970	16.990	19.351
ChiSqM_PValue	0.004	0.009	0.020	0.048	0.142	0.340	0.612	0.805	0.891	0.942	0.975
LL	-1814.650	-1807.031	-1800.373	-1792.561	-1779.898	-1764.499	-1749.110	-1734.976	-1726.622	-1719.358	-1709.748
UnrestrictedLL	-1811.428	-1803.348	-1797.389	-1789.191	-1776.389	-1760.486	-1745.256	-1731.137	-1722.666	-1715.666	-1706.094
CFI	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000
TLI	-10.298	-3.817	-1.363	-0.064	0.843	1.000	1.000	1.000	1.425	2.353	5.062
AIC	3427.496	3446.716	3461.245	3477.952	3506.220	3536.998	3567.796	3593.123	3608.746	3622.062	3637.299
BIC	3444.354	3463.574	3478.103	3494.811	3523.078	3553.856	3584.655	3609.981	3625.604	3638.920	3654.157
aBIC	3431.658	3450.878	3465.407	3482.114	3510.382	3541.160	3571.959	3597.285	3612.908	3626.224	3641.461
RMSEA_Estimate	0.000	0.000	0.000	0.000	0.000	0.023	0.049	0.067	0.077	0.086	0.094
RMSEA_90CI_LB	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.028	0.039	0.049
RMSEA_90CI_UB	0.000	0.016	0.037	0.052	0.069	0.088	0.104	0.118	0.127	0.135	0.143
RMSEA_pLT05	0.052	0.094	0.152	0.247	0.445	0.676	0.860	0.944	0.973	0.987	0.995
SRMR	0.025	0.030	0.035	0.041	0.050	0.062	0.074	0.087	0.096	0.102	0.112
AICC	3427.576	3446.797	3461.326	3478.033	3506.301	3537.079	3567.876	3593.204	3608.827	3622.142	3637.379
a_uns_est	-0.002	-0.001	0.000	0.000	0.000	0.000	0.153	0.327	0.391	0.441	0.491
a_uns_est_se	-0.004	-0.002	-0.001	0.000	0.000	0.000	0.286	1.714	2.595	3.257	4.070
a_uns_pval	0.000	0.001	0.009	0.086	0.775	0.999	1.000	1.000	1.000	1.000	1.000
c_uns_est	-0.003	-0.001	-0.001	0.000	0.000	0.000	0.120	0.287	0.348	0.388	0.429
c_uns_est_se	-0.005	-0.003	-0.002	-0.001	0.000	0.000	0.230	1.669	2.564	3.269	4.029
c_uns_pval	0.000	0.001	0.010	0.095	0.819	1.000	1.000	1.000	1.000	1.000	1.000
e_uns_est	1.298	1.315	1.328	1.346	1.372	1.398	1.423	1.444	1.456	1.468	1.480
e_uns_est_se	24.723	25.432	26.240	26.869	27.816	31.625	44.721	44.721	44.722	44.722	44.722
e_uns_pval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
a_stz_est	-0.001	-0.001	0.000	0.000	0.000	0.000	0.108	0.230	0.277	0.310	0.344
a_stz_est_se	-0.004	-0.002	-0.001	0.000	0.000	0.000	0.286	1.718	2.608	3.284	4.121
a_stz_pval	0.000	0.001	0.009	0.086	0.775	0.999	1.000	1.000	1.000	1.000	1.000
c_stz_est	-0.002	-0.001	0.000	0.000	0.000	0.000	0.086	0.203	0.246	0.275	0.303
c_stz_est_se	-0.005	-0.003	-0.002	-0.001	0.000	0.000	0.230	1.673	2.577	3.295	4.078
c_stz_pval	0.000	0.001	0.010	0.094	0.819	1.000	1.000	1.000	1.000	1.000	1.000
e_stz_est	0.936	0.945	0.954	0.964	0.980	0.997	1.000	1.000	1.000	1.000	1.000
e_stz_pval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
totalvar	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000
aBIC_sat	3431.658	3450.878	3465.407	3482.114	3510.382	3541.160	3571.959	3597.285	3612.908	3626.224	3641.461
AIC_sat	3427.496	3446.716	3461.245	3477.952	3506.220	3536.998	3567.796	3593.123	3608.746	3622.062	3637.299
BIC_sat	3444.354	3463.574	3478.103	3494.811	3523.078	3553.856	3584.655	3609.981	3625.604	3638.920	3654.157
ChiSqM_Value_sat	1.237	1.744	2.288	3.034	4.481	6.800	9.602	12.706	14.970	16.990	19.351
ChiSqM_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
nested_sig	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
AIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
aBIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 9 Descriptive Statistics for E data fit to ACE models (Condition 37)

Quantile	1	2.5	5	10	25	50	75	90	95	97.5	99
ChiSqM_Value	1.337	1.829	2.327	3.119	4.593	6.889	9.672	12.801	14.979	17.158	19.789
ChiSqM_PValue	0.003	0.009	0.020	0.046	0.139	0.331	0.597	0.794	0.887	0.935	0.970
LL	-2908.721	-2899.032	-2888.496	-2877.178	-2857.992	-2837.138	-2815.141	-2795.189	-2784.242	-2773.964	-2763.925
UnrestrictedLL	-2905.236	-2895.045	-2885.037	-2873.736	-2854.031	-2833.375	-2811.490	-2791.021	-2780.251	-2770.587	-2760.000
CFI	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000
TLI	-11.925	-3.803	-1.297	-0.065	0.823	1.000	1.000	1.000	1.461	2.468	5.782
AIC	5535.851	5555.927	5576.485	5598.379	5638.282	5682.276	5723.984	5762.355	5784.992	5806.065	5825.443
BIC	5555.482	5575.558	5596.116	5618.010	5657.913	5701.907	5743.615	5781.986	5804.623	5825.696	5845.074
aBIC	5542.778	5562.854	5583.411	5605.306	5645.209	5689.203	5730.910	5769.282	5791.919	5812.991	5832.369
RMSEA_Estimate	0.000	0.000	0.000	0.000	0.000	0.017	0.035	0.048	0.055	0.061	0.068
RMSEA_90CI_LB	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.006	0.020	0.028	0.036
RMSEA_90CI_UB	0.000	0.015	0.027	0.038	0.050	0.062	0.074	0.084	0.090	0.096	0.102
RMSEA_pLT05	0.160	0.255	0.360	0.487	0.691	0.857	0.951	0.984	0.993	0.997	0.999
SRMR	0.019	0.022	0.025	0.029	0.036	0.044	0.052	0.061	0.066	0.071	0.078
AICC	5535.891	5555.967	5576.525	5598.420	5638.322	5682.316	5724.024	5762.396	5785.032	5806.105	5825.483
a_uns_est	-0.001	0.000	0.000	0.000	0.000	0.000	0.090	0.193	0.233	0.262	0.292
a_uns_est_se	-0.003	-0.002	-0.001	-0.001	0.000	0.000	0.265	1.671	2.567	3.327	4.225
a_uns_pval	0.000	0.001	0.010	0.095	0.791	0.999	1.000	1.000	1.000	1.000	1.000
c_uns_est	-0.003	-0.001	0.000	0.000	0.000	0.000	0.076	0.173	0.208	0.230	0.259
c_uns_est_se	-0.008	-0.004	-0.001	-0.001	0.000	0.000	0.247	1.677	2.580	3.322	4.188
c_uns_pval	0.000	0.001	0.010	0.093	0.805	1.000	1.000	1.000	1.000	1.000	1.000
e_uns_est	0.940	0.948	0.958	0.966	0.979	0.992	1.005	1.015	1.022	1.027	1.034
e_uns_est_se	35.277	36.046	37.523	38.416	39.340	44.726	63.245	63.246	63.246	63.246	63.246
e_uns_pval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
a_stz_est	-0.001	0.000	0.000	0.000	0.000	0.000	0.090	0.194	0.234	0.261	0.292
a_stz_est_se	-0.003	-0.002	-0.001	-0.001	0.000	0.000	0.265	1.672	2.573	3.341	4.252
a_stz_pval	0.000	0.001	0.010	0.094	0.791	0.999	1.000	1.000	1.000	1.000	1.000
c_stz_est	-0.003	-0.001	0.000	0.000	0.000	0.000	0.076	0.173	0.207	0.232	0.258
c_stz_est_se	-0.008	-0.004	-0.001	-0.001	0.000	0.000	0.247	1.678	2.586	3.336	4.216
c_stz_pval	0.000	0.001	0.010	0.093	0.805	1.000	1.000	1.000	1.000	1.000	1.000
e_stz_est	0.954	0.961	0.967	0.975	0.986	0.998	1.000	1.000	1.000	1.000	1.000
e_stz_pval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
totalvar	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
aBIC_sat	5542.778	5562.854	5583.411	5605.306	5645.209	5689.203	5730.910	5769.282	5791.919	5812.991	5832.369
AIC_sat	5535.851	5555.927	5576.485	5598.379	5638.282	5682.276	5723.984	5762.355	5784.992	5806.065	5825.443
BIC_sat	5555.482	5575.558	5596.116	5618.010	5657.913	5701.907	5743.615	5781.986	5804.623	5825.696	5845.074
ChiSqM_Value_sat	1.337	1.829	2.327	3.119	4.593	6.889	9.672	12.801	14.979	17.158	19.789
ChiSqM_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
nested_sig	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
AIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
aBIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 10 Descriptive Statistics for E data fit to ACE models (Condition 46)

Quantile	1	2.5	5	10	25	50	75	90	95	97.5	99
ChiSqM_Value	1.359	1.811	2.282	3.024	4.578	6.864	9.626	12.756	15.076	16.925	19.059
ChiSqM_PValue	0.004	0.010	0.020	0.047	0.141	0.334	0.599	0.806	0.892	0.936	0.968
LL	-3601.643	-3591.719	-3581.648	-3570.532	-3551.669	-3530.682	-3508.572	-3489.476	-3477.375	-3467.891	-3454.108
UnrestrictedLL	-3598.088	-3587.791	-3578.523	-3566.869	-3547.878	-3526.780	-3504.959	-3485.453	-3473.688	-3463.984	-3450.106
CFI	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000
TLI	-12.687	-4.485	-1.466	-0.091	0.843	1.000	1.000	1.000	1.468	2.336	4.936
AIC	6916.215	6943.782	6962.750	6986.951	7025.144	7069.365	7111.338	7149.063	7171.297	7191.439	7211.285
BIC	6935.846	6963.413	6982.381	7006.582	7044.775	7088.996	7130.969	7168.694	7190.928	7211.070	7230.916
aBIC	6923.142	6950.709	6969.677	6993.878	7032.071	7076.292	7118.265	7155.990	7178.224	7198.366	7218.212
RMSEA_Estimate	0.000	0.000	0.000	0.000	0.000	0.017	0.035	0.047	0.055	0.060	0.066
RMSEA_90CI_LB	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.020	0.027	0.034
RMSEA_90CI_UB	0.000	0.014	0.026	0.037	0.050	0.062	0.074	0.084	0.090	0.095	0.100
RMSEA_pLT05	0.183	0.265	0.354	0.489	0.694	0.858	0.951	0.985	0.994	0.997	0.999
SRMR	0.019	0.022	0.025	0.029	0.036	0.044	0.052	0.061	0.067	0.072	0.081
AICC	6916.256	6943.822	6962.791	6986.991	7025.184	7069.405	7111.378	7149.103	7171.338	7191.479	7211.325
a_uns_est	-0.002	-0.001	-0.001	0.000	0.000	0.000	0.109	0.267	0.326	0.364	0.405
a_uns_est_se	-0.006	-0.004	-0.002	-0.001	0.000	0.000	0.168	1.548	2.491	3.158	3.932
a_uns_pval	0.000	0.002	0.013	0.121	0.867	0.999	1.000	1.000	1.000	1.000	1.000
c_uns_est	-0.005	-0.002	-0.001	0.000	0.000	0.000	0.112	0.243	0.293	0.330	0.361
c_uns_est_se	-0.011	-0.007	-0.004	-0.001	0.000	0.000	0.291	1.636	2.564	3.311	4.019
c_uns_pval	0.000	0.001	0.010	0.102	0.771	0.999	1.000	1.000	1.000	1.000	1.000
e_uns_est	1.333	1.345	1.355	1.366	1.385	1.404	1.422	1.436	1.445	1.452	1.461
e_uns_est_se	35.093	35.916	37.067	38.384	39.470	44.732	63.245	63.246	63.246	63.246	63.246
e_uns_pval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
a_stz_est	-0.001	-0.001	0.000	0.000	0.000	0.000	0.077	0.189	0.231	0.256	0.287
a_stz_est_se	-0.006	-0.004	-0.002	-0.001	0.000	0.000	0.168	1.549	2.497	3.169	3.955
a_stz_pval	0.000	0.002	0.013	0.121	0.867	0.999	1.000	1.000	1.000	1.000	1.000
c_stz_est	-0.004	-0.001	-0.001	0.000	0.000	0.000	0.080	0.172	0.207	0.233	0.255
c_stz_est_se	-0.011	-0.007	-0.004	-0.001	0.000	0.000	0.291	1.638	2.570	3.325	4.044
c_stz_pval	0.000	0.001	0.010	0.101	0.771	0.999	1.000	1.000	1.000	1.000	1.000
e_stz_est	0.955	0.963	0.968	0.974	0.986	0.998	1.000	1.000	1.000	1.000	1.000
e_stz_pval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
totalvar	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000
aBIC_sat	6923.142	6950.709	6969.677	6993.878	7032.071	7076.292	7118.265	7155.990	7178.224	7198.366	7218.212
AIC_sat	6916.215	6943.782	6962.750	6986.951	7025.144	7069.365	7111.338	7149.063	7171.297	7191.439	7211.285
BIC_sat	6935.846	6963.413	6982.381	7006.582	7044.775	7088.996	7130.969	7168.694	7190.928	7211.070	7230.916
ChiSqM_Value_sat	1.359	1.811	2.282	3.024	4.578	6.864	9.626	12.756	15.076	16.925	19.059
ChiSqM_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
nested_sig	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
AIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
aBIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 11 Descriptive Statistics for E data fit to ACE models (Condition 55)

Quantile	1	2.5	5	10	25	50	75	90	95	97.5	99
ChiSqM_Value	1.441	1.858	2.353	3.037	4.489	6.611	9.470	12.365	14.359	16.024	18.377
ChiSqM_PValue	0.005	0.014	0.026	0.054	0.149	0.358	0.611	0.804	0.885	0.932	0.963
LL	-5777.444	-5763.601	-5748.621	-5731.227	-5703.575	-5674.285	-5643.565	-5616.681	-5600.236	-5585.050	-5567.891
UnrestrictedLL	-5773.342	-5759.725	-5744.466	-5727.394	-5700.212	-5670.569	-5639.881	-5612.754	-5596.561	-5580.800	-5564.512
CFI	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000
TLI	-10.501	-3.386	-1.202	-0.033	0.903	1.000	1.000	1.022	1.575	2.464	5.340
AIC	11143.782	11178.101	11208.474	11241.362	11295.130	11356.570	11415.151	11470.454	11505.242	11535.201	11562.888
BIC	11166.186	11200.504	11230.877	11263.766	11317.534	11378.973	11437.554	11492.858	11527.645	11557.605	11585.291
aBIC	11153.478	11187.796	11218.169	11251.058	11304.825	11366.265	11424.846	11480.150	11514.937	11544.897	11572.583
RMSEA_Estimate	0.000	0.000	0.000	0.000	0.000	0.010	0.024	0.033	0.037	0.041	0.045
RMSEA_90CI_LB	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.017	0.023
RMSEA_90CI_UB	0.000	0.011	0.020	0.026	0.035	0.043	0.052	0.058	0.063	0.066	0.070
RMSEA_pLT05	0.581	0.695	0.771	0.851	0.937	0.982	0.996	0.999	1.000	1.000	1.000
SRMR	0.014	0.016	0.018	0.020	0.025	0.030	0.037	0.043	0.046	0.049	0.052
AICC	11143.802	11178.121	11208.494	11241.382	11295.150	11356.590	11415.171	11470.474	11505.262	11535.221	11562.908
a_uns_est	-0.001	-0.001	0.000	0.000	0.000	0.000	0.082	0.161	0.196	0.217	0.240
a_uns_est_se	-0.007	-0.003	-0.002	-0.001	0.000	0.001	0.284	1.649	2.493	3.202	3.981
a_uns_pval	0.000	0.001	0.013	0.099	0.776	0.999	1.000	1.000	1.000	1.000	1.000
c_uns_est	-0.007	-0.002	-0.001	0.000	0.000	0.000	0.069	0.143	0.174	0.193	0.216
c_uns_est_se	-0.018	-0.010	-0.004	-0.001	0.000	0.000	0.299	1.654	2.619	3.263	4.127
c_uns_pval	0.000	0.001	0.009	0.098	0.765	0.999	1.000	1.000	1.000	1.000	1.000
e_uns_est	0.957	0.964	0.970	0.976	0.985	0.994	1.003	1.011	1.015	1.019	1.024
e_uns_est_se	50.233	51.023	52.355	54.614	55.507	63.253	89.442	89.443	89.443	89.443	89.444
e_uns_pval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
a_stz_est	-0.001	-0.001	0.000	0.000	0.000	0.000	0.082	0.160	0.196	0.217	0.241
a_stz_est_se	-0.007	-0.003	-0.002	-0.001	0.000	0.001	0.284	1.650	2.496	3.208	3.992
a_stz_pval	0.000	0.001	0.013	0.099	0.776	0.999	1.000	1.000	1.000	1.000	1.000
c_stz_est	-0.007	-0.002	-0.001	0.000	0.000	0.000	0.069	0.144	0.174	0.193	0.216
c_stz_est_se	-0.018	-0.010	-0.004	-0.001	0.000	0.000	0.299	1.655	2.622	3.269	4.140
c_stz_pval	0.000	0.001	0.009	0.098	0.765	0.999	1.000	1.000	1.000	1.000	1.000
e_stz_est	0.968	0.973	0.978	0.982	0.990	0.998	1.000	1.000	1.000	1.000	1.000
e_stz_pval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
totalvar	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
aBIC_sat	11153.478	11187.796	11218.169	11251.058	11304.825	11366.265	11424.846	11480.150	11514.937	11544.897	11572.583
AIC_sat	11143.782	11178.101	11208.474	11241.362	11295.130	11356.570	11415.151	11470.454	11505.242	11535.201	11562.888
BIC_sat	11166.186	11200.504	11230.877	11263.766	11317.534	11378.973	11437.554	11492.858	11527.645	11557.605	11585.291
ChiSqM_Value_sat	1.441	1.858	2.353	3.037	4.489	6.611	9.470	12.365	14.359	16.024	18.377
ChiSqM_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
nested_sig	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
AIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
aBIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 12 Descriptive Statistics for E data fit to ACE models (Condition 64)

Quantile	1	2.5	5	10	25	50	75	90	95	97.5	99
ChiSqM_Value	1.374	1.834	2.333	3.065	4.542	6.676	9.463	12.626	14.592	17.118	19.226
ChiSqM_PValue	0.004	0.009	0.024	0.049	0.149	0.352	0.604	0.801	0.887	0.934	0.967
LL	-7161.869	-7145.724	-7133.398	-7118.081	-7090.536	-7062.070	-7030.751	-7004.464	-6987.298	-6973.816	-6958.263
UnrestrictedLL	-7157.416	-7142.716	-7130.051	-7114.413	-7087.047	-7058.501	-7027.104	-7000.343	-6984.103	-6970.012	-6953.260
CFI	0.000	0.000	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000
TLI	-13.177	-4.688	-1.543	-0.079	0.916	1.000	1.000	1.069	1.563	2.475	4.620
AIC	13924.526	13955.632	13982.597	14016.928	14069.502	14132.139	14189.071	14244.164	14274.797	14299.448	14331.738
BIC	13946.930	13978.036	14005.001	14039.332	14091.905	14154.543	14211.474	14266.568	14297.200	14321.851	14354.141
aBIC	13934.221	13965.327	13992.293	14026.623	14079.197	14141.834	14198.766	14253.859	14284.492	14309.143	14341.433
RMSEA_Estimate	0.000	0.000	0.000	0.000	0.000	0.011	0.024	0.033	0.038	0.043	0.047
RMSEA_90CI_LB	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.013	0.020	0.024
RMSEA_90CI_UB	0.000	0.011	0.019	0.026	0.035	0.043	0.052	0.059	0.063	0.068	0.071
RMSEA_pLT05	0.539	0.643	0.761	0.842	0.937	0.982	0.996	0.999	1.000	1.000	1.000
SRMR	0.014	0.016	0.018	0.020	0.025	0.030	0.037	0.043	0.047	0.050	0.055
AICC	13924.546	13955.652	13982.617	14016.948	14069.522	14132.159	14189.091	14244.184	14274.817	14299.468	14331.758
a_uns_est	-0.002	-0.001	-0.001	0.000	0.000	0.000	0.097	0.223	0.269	0.304	0.342
a_uns_est_se	-0.009	-0.006	-0.003	-0.001	0.000	0.001	0.205	1.472	2.414	3.139	4.030
a_uns_pval	0.000	0.002	0.016	0.141	0.838	0.998	1.000	1.000	1.000	1.000	1.000
c_uns_est	-0.009	-0.005	-0.002	-0.001	0.000	0.000	0.101	0.207	0.247	0.277	0.309
c_uns_est_se	-0.021	-0.015	-0.009	-0.002	0.000	0.000	0.325	1.678	2.618	3.282	4.155
c_uns_pval	0.000	0.001	0.009	0.093	0.746	0.998	1.000	1.000	1.000	1.000	1.000
e_uns_est	1.358	1.365	1.373	1.381	1.394	1.407	1.419	1.429	1.436	1.441	1.447
e_uns_est_se	49.982	50.868	51.972	54.466	55.592	63.261	89.442	89.443	89.443	89.443	89.444
e_uns_pval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
a_stz_est	-0.002	-0.001	-0.001	0.000	0.000	0.000	0.069	0.157	0.190	0.214	0.241
a_stz_est_se	-0.009	-0.006	-0.003	-0.001	0.000	0.001	0.205	1.473	2.417	3.145	4.042
a_stz_pval	0.000	0.002	0.016	0.141	0.838	0.998	1.000	1.000	1.000	1.000	1.000
c_stz_est	-0.006	-0.003	-0.001	0.000	0.000	0.000	0.072	0.147	0.174	0.195	0.217
c_stz_est_se	-0.021	-0.015	-0.009	-0.002	0.000	0.000	0.325	1.679	2.621	3.288	4.168
c_stz_pval	0.000	0.001	0.009	0.093	0.746	0.998	1.000	1.000	1.000	1.000	1.000
e_stz_est	0.968	0.974	0.978	0.982	0.990	0.999	1.000	1.000	1.000	1.000	1.000
e_stz_pval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
totalvar	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000	2.000
aBIC_sat	13934.221	13965.327	13992.293	14026.623	14079.197	14141.834	14198.766	14253.859	14284.492	14309.143	14341.433
AIC_sat	13924.526	13955.632	13982.597	14016.928	14069.502	14132.139	14189.071	14244.164	14274.797	14299.448	14331.738
BIC_sat	13946.930	13978.036	14005.001	14039.332	14091.905	14154.543	14211.474	14266.568	14297.200	14321.851	14354.141
ChiSqM_Value_sat	1.374	1.834	2.333	3.065	4.542	6.676	9.463	12.626	14.592	17.118	19.226
ChiSqM_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
nested_sig	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
AIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
BIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
aBIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 13 Quantiles for correctly-specified models

Quantile	1%	2.50%	5%	10%	25%	50%	75%	90%	95%	97.50%	99%
ChiSqM_Value	1.012	1.426	1.868	2.492	3.856	5.882	8.521	11.446	13.470	15.427	17.904
ChiSqM_PValue	0.010	0.024	0.049	0.098	0.246	0.495	0.747	0.899	0.949	0.974	0.990
CFI	0.802	0.889	0.934	0.967	0.992	1.000	1.000	1.000	1.000	1.000	1.000
TLI	0.942	0.967	0.980	0.990	0.998	1.001	1.004	1.010	1.018	1.028	1.048
AIC	1355.198	1394.357	1477.424	1554.823	2894.152	5555.657	10088.528	16274.805	16631.965	17174.510	17615.241
BIC	1365.093	1404.252	1488.688	1566.111	2906.933	5571.448	10107.452	16294.408	16649.787	17196.914	17637.645
aBIC	1355.589	1394.747	1477.905	1555.270	2896.623	5560.335	10095.541	16283.182	16639.688	17184.205	17624.937
RMSEA_Estimate	0.000	0.000	0.000	0.000	0.000	0.000	0.028	0.053	0.072	0.090	0.109
RMSEA_90CI_LB	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.021	0.038
RMSEA_90CI_UB	0.000	0.000	0.004	0.024	0.040	0.060	0.092	0.131	0.153	0.170	0.187
RMSEA_pLT05	0.081	0.159	0.256	0.403	0.676	0.900	0.984	0.998	0.999	1.000	1.000
SRMR	0.013	0.016	0.019	0.022	0.031	0.045	0.068	0.097	0.115	0.132	0.152
AICC	1355.320	1394.479	1477.608	1554.987	2894.240	5555.709	10088.561	16274.817	16631.978	17174.530	17615.261
a_stz_pval	0.000	0.000	0.000	0.000	0.000	0.000	250.500	999.000	999.000	999.000	999.000
c_stz_pval	0.000	0.000	0.000	0.000	0.000	0.000	250.500	999.000	999.000	999.000	999.000
e_stz_pval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
totalvar	2.000	2.000	2.000	2.000	3.000	4.000	4.250	5.000	6.000	6.000	6.000
aBIC_sat	1357.109	1396.446	1478.942	1556.102	2555.807	6569.818	11139.579	16285.430	16642.299	17184.206	17624.937
AIC_sat	1356.588	1395.925	1478.421	1555.581	2553.466	6562.891	11131.268	16275.735	16632.603	17174.511	17615.241
BIC_sat	1369.781	1409.118	1491.615	1568.774	2568.492	6582.522	11152.285	16298.138	16655.007	17196.915	17637.645
ChiSqM_Value_sat	0.934	1.327	1.751	2.341	3.658	5.633	8.222	11.091	13.092	15.010	17.449
ChiSqM_diff	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.711	1.645	2.710	4.226
nested_sig	0.040	0.100	0.200	0.399	1.000	1.000	1.000	1.000	1.000	1.000	1.000
AIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	2.000	2.000	2.000	2.000	2.226
BIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	6.214	6.908	7.601	7.601	7.601
aBIC_diff	0.000	0.000	0.000	0.000	0.000	0.000	3.040	3.822	4.424	4.424	4.424

Table 14 Quantiles for incorrectly-specified models

Quantile	1%	2.50%	5%	10%	25%	50%	75%	90%	95%	97.50%	99%
ChiSqM_Value	1.570	2.232	2.958	4.039	6.751	12.908	29.676	60.392	89.219	122.478	155.032
ChiSqM_PValue	0.000	0.000	0.000	0.000	0.000	0.066	0.398	0.717	0.845	0.918	0.965
CFI	0.000	0.000	0.000	0.743	0.877	0.952	0.998	1.000	1.000	1.000	1.000
TLI	-0.354	0.605	0.913	0.946	0.970	0.988	1.000	1.004	1.015	1.042	1.445
AIC	1140.239	1346.904	1406.824	1520.776	1841.182	4527.757	8996.880	15840.921	16563.049	17162.774	17625.544
BIC	1151.197	1358.392	1418.042	1531.820	1851.077	4540.401	9011.603	15857.822	16581.166	17179.577	17642.347
aBIC	1140.682	1347.368	1407.258	1521.227	1841.573	4530.878	9002.075	15848.210	16570.874	17170.046	17632.816
RMSEA_Estimate	0.000	0.000	0.000	0.000	0.010	0.057	0.095	0.130	0.149	0.167	0.183
RMSEA_90CI_LB	0.000	0.000	0.000	0.000	0.000	0.000	0.058	0.094	0.114	0.129	0.148
RMSEA_90CI_UB	0.000	0.021	0.032	0.044	0.070	0.105	0.144	0.179	0.200	0.217	0.239
RMSEA_pLT05	0.000	0.000	0.000	0.000	0.023	0.351	0.815	0.974	0.994	0.998	1.000
SRMR	0.018	0.022	0.026	0.032	0.046	0.065	0.088	0.112	0.128	0.143	0.161
AICC	1140.405	1347.071	1406.968	1520.938	1841.305	4527.805	8996.904	15840.933	16563.061	17162.786	17625.556
a_stz_pval	0.000	0.000	0.000	0.000	0.000	0.732	999.000	999.000	999.000	999.000	999.000
c_stz_pval	0.000	0.000	0.000	0.000	0.000	0.704	999.000	999.000	999.000	999.000	999.000
e_stz_pval	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
totalvar	1.000	1.000	1.000	2.000	2.000	3.000	4.000	5.000	6.000	6.000	6.000
aBIC_sat	1142.201	1344.967	1406.625	1516.781	1829.210	4502.791	8960.887	15781.766	16543.389	17141.096	17594.724
AIC_sat	1141.680	1344.446	1406.105	1516.260	1828.690	4498.629	8953.960	15772.070	16533.694	17131.400	17585.028
BIC_sat	1154.874	1357.639	1419.298	1529.453	1841.883	4515.487	8973.591	15794.474	16556.098	17153.804	17607.432
ChiSqM_Value_sat	0.969	1.378	1.816	2.429	3.780	5.801	8.447	11.382	13.422	15.375	17.879
ChiSqM_diff	0.000	0.000	0.000	0.000	0.000	5.349	23.102	53.980	82.890	116.073	148.222
nested_sig	0.000	0.000	0.000	0.000	0.000	0.021	1.000	1.000	1.000	1.000	1.000
AIC_diff	0.000	0.000	0.000	0.000	0.000	3.348	21.102	51.980	80.890	114.074	146.222
BIC_diff	0.000	0.000	0.000	0.000	0.000	5.254	16.437	46.790	75.554	108.638	140.632
aBIC_diff	0.000	0.000	0.000	0.000	0.000	3.732	19.611	49.967	78.731	111.815	143.809

Table 15 Model Rejection by Condition and Fit Model

condition	fitmodel	TrueModel	# Models converged	aBIC	AIC	BIC	ChiSqM PValue	CFI	TLI	RMSEA Estimate	RMSEA 90CI_LB	RMSEA 90CI_UB	RMSEA pLT05	SRMR	a_uns_pval	c_uns_pval	a_stz_pval	c_stz_pval	nested_sig
1	ACE	E	4991	4991	4991	4991	509	2854	1310	1778	509	4724	4991	3721	4557	4622	4552	4619	0
1	AE	E	4998	1348	1343	1342	329	2363	1128	1264	329	4535	4996	3767	4186	0	4176	0	61
1	CE	E	5000	3650	3655	3656	331	2381	1139	1275	331	4528	4996	3765	0	4221	0	4211	84
2	ACE	AE	5000	4980	4973	5000	368	1615	610	1448	368	4632	4997	3178	564	4085	561	4073	0
2	AE	AE	5000	689	692	678	280	1443	408	1152	280	4474	4989	3346	0	0	0	0	61
2	CE	AE	5000	4331	4335	4322	1563	3384	1840	3044	1562	4886	5000	4203	0	0	0	0	5
3	ACE	AE	5000	4721	4681	4995	382	1015	125	1425	382	4561	4998	3011	24	4103	24	4086	0
3	AE	AE	5000	355	387	154	286	907	59	1179	285	4397	4992	3134	0	0	0	0	81
3	CE	AE	5000	4924	4932	4851	3785	4512	2476	4629	3785	4994	5000	4609	0	0	0	0	1
4	ACE	CE	5000	4842	4816	5000	331	1186	181	1439	331	4627	4998	3025	4167	140	4164	135	0
4	AE	CE	5000	4707	4714	4643	2584	3964	1633	3984	2584	4966	5000	4614	0	0	0	0	2
4	CE	CE	5000	451	470	357	261	1060	123	1160	261	4498	4994	3117	0	0	0	0	95
5	ACE	ACE	5000	2699	2558	4862	295	618	29	1202	295	4479	4993	2742	695	620	693	601	0
5	AE	ACE	5000	3673	3740	2563	1345	2175	141	2856	1345	4871	4998	3957	0	0	0	0	7
5	CE	ACE	5000	3628	3702	2575	1389	2180	235	2837	1389	4858	4999	3446	0	0	0	0	12
6	ACE	ACE	5000	2375	2283	4307	246	378	4	1183	246	4425	4993	2624	24	1290	24	1267	0
6	AE	ACE	5000	2720	2804	1145	773	1086	9	2137	772	4765	5000	3442	0	0	0	0	20
6	CE	ACE	5000	4905	4913	4548	3712	4058	902	4598	3712	4992	5000	3985	0	0	0	0	0
7	ACE	CE	5000	4635	4600	4949	365	464	11	1366	364	4578	4994	2522	4189	0	4188	0	0
7	AE	CE	5000	5000	5000	4981	4778	4890	2351	4967	4777	5000	5000	4976	0	0	0	0	0
7	CE	CE	5000	365	400	70	300	418	2	1103	300	4404	4995	2574	0	0	0	0	94
8	ACE	ACE	5000	1451	1364	3490	269	270	2	1152	269	4460	4997	2464	696	15	697	14	0
8	AE	ACE	5000	4915	4923	4477	3565	3787	213	4566	3565	4994	5000	4726	0	0	0	0	0
8	CE	ACE	5000	3634	3713	2033	1363	1487	18	2823	1363	4884	5000	2916	0	0	0	0	12
9	ACE	ACE	5000	512	468	2155	263	210	0	1149	263	4463	4996	2408	28	144	28	136	0
9	AE	ACE	5000	4590	4624	3421	2400	2272	23	3830	2395	4960	4999	4284	0	0	0	0	2
9	CE	ACE	5000	4898	4908	4424	3611	3550	260	4555	3611	4995	5000	3409	0	0	0	0	0
10	ACE	E	4995	4995	4995	4995	561	2946	1322	1797	561	4775	4993	3716	4622	4599	4620	4598	0
10	AE	E	4997	1345	1343	1346	351	2445	1153	1339	350	4537	4993	3738	4208	0	4199	0	67
10	CE	E	5000	3652	3654	3651	354	2430	1138	1338	354	4546	4996	3758	0	4190	0	4184	89
11	ACE	AE	5000	5000	5000	5000	353	2073	1428	1411	352	4606	4997	3285	1807	4130	1802	4122	0
11	AE	AE	5000	1253	1252	1252	271	1825	1110	1075	271	4408	4995	3439	12	0	12	0	75
11	CE	AE	5000	3747	3748	3748	645	2620	1841	1853	644	4694	4999	3881	0	31	0	30	32
12	ACE	AE	5000	4982	4977	5000	357	1552	617	1401	357	4577	4999	3155	562	4153	557	4144	0
12	AE	AE	5000	657	661	649	291	1401	397	1116	291	4406	4995	3299	0	0	0	0	87
12	CE	AE	5000	4361	4362	4351	1554	3347	1805	3013	1554	4896	5000	4188	0	0	0	0	8
13	ACE	CE	5000	4999	4998	5000	341	1838	972	1363	341	4591	4997	3159	4148	1242	4145	1238	0
13	AE	CE	5000	3987	3987	3986	867	2809	1655	2215	867	4768	5000	4056	0	0	0	0	20
13	CE	CE	5000	1014	1015	1014	254	1601	724	1110	254	4399	4993	3286	0	0	0	0	69

condition	fitmodel	TrueModel	# Models converged	aBIC	AIC	BIC	ChiSqM PValue	CFI	TLI	RMSEA Estimate	RMSEA 90CI_LB	RMSEA 90CI_UB	RMSEA pLT05	SRMR	a_uns_pval	c_uns_pval	a_stz_pval	c_stz_pval	nested_sig
14	ACE	ACE	5000	4870	4835	5000	255	1237	292	1223	255	4492	4993	2989	2104	1688	2097	1672	0
14	AE	ACE	5000	2736	2754	2669	663	2060	525	1929	662	4680	4995	3755	0	0	0	0	34
14	CE	ACE	5000	2394	2411	2331	556	1866	480	1715	555	4674	4994	3463	0	0	0	0	34
15	ACE	ACE	5000	4204	4087	5000	253	908	128	1183	253	4484	4994	2818	640	2096	639	2075	0
15	AE	ACE	5000	1987	2059	1553	505	1441	139	1656	503	4664	4991	3440	0	0	0	0	43
15	CE	ACE	5000	3809	3854	3447	1378	2733	645	2899	1376	4867	5000	3770	0	0	0	0	6
16	ACE	CE	5000	4860	4829	5000	366	1171	219	1428	365	4579	4997	2966	4185	125	4179	121	0
16	AE	CE	5000	4723	4734	4667	2588	3922	1622	3958	2585	4963	5000	4593	0	0	0	0	1
16	CE	CE	5000	417	437	333	270	1050	117	1139	270	4433	4995	3059	0	0	0	0	79
17	ACE	ACE	5000	3876	3761	4983	301	763	84	1204	301	4474	4996	2752	2172	368	2162	362	0
17	AE	ACE	5000	4192	4239	3726	1775	2942	527	3349	1771	4921	5000	4254	0	0	0	0	6
17	CE	ACE	5000	1932	2000	1291	595	1341	115	1720	594	4641	4998	3135	0	0	0	0	30
18	ACE	ACE	5000	2615	2453	4850	244	553	30	1119	244	4443	4994	2642	680	605	680	592	0
18	AE	ACE	5000	3678	3762	2536	1249	2095	126	2805	1247	4869	5000	3911	0	0	0	0	8
18	CE	ACE	5000	3707	3785	2614	1369	2154	204	2821	1365	4863	5000	3364	0	0	0	0	3
19	ACE	E	4979	4979	4979	4979	538	2899	1350	747	536	4221	4976	836	4559	4569	4558	4568	0
19	AE	E	4992	1348	1339	1347	347	2421	1180	423	347	3651	4977	853	4171	0	4165	0	76
19	CE	E	5000	3644	3653	3645	342	2424	1187	427	342	3671	4985	861	0	4172	0	4169	87
20	ACE	AE	4999	4910	4673	4999	333	711	54	486	333	3832	4993	569	28	4121	26	4117	0
20	AE	AE	5000	239	413	180	254	621	25	312	254	3418	4980	610	0	0	0	0	87
20	CE	AE	5000	4850	4913	4820	3585	4245	1530	3755	3584	4965	5000	2429	0	0	0	0	0
21	ACE	AE	5000	4806	4625	4978	349	183	0	485	349	3879	4995	543	0	4190	0	4172	0
21	AE	AE	5000	194	375	30	274	154	0	332	273	3485	4992	570	0	0	0	0	82
21	CE	AE	5000	5000	5000	4992	4974	4961	2645	4980	4974	5000	5000	3987	0	0	0	0	0
22	ACE	CE	5000	4818	4612	4989	320	251	1	462	319	3839	4995	478	4170	0	4168	0	0
22	AE	CE	5000	4993	5000	4970	4719	4748	1189	4775	4718	5000	5000	3851	0	0	0	0	0
22	CE	CE	5000	189	388	41	243	219	0	297	243	3421	4979	499	0	0	0	0	81
23	ACE	ACE	5000	583	289	2188	249	47	0	377	248	3649	4986	459	63	45	63	44	0
23	AE	ACE	5000	4706	4867	3871	3186	2062	2	3359	3186	4937	5000	1798	0	0	0	0	2
23	CE	ACE	5000	4711	4844	3941	3327	2278	7	3503	3327	4947	5000	978	0	0	0	0	3
24	ACE	ACE	5000	1091	683	2459	262	17	0	382	262	3616	4988	425	0	300	0	296	0
24	AE	ACE	5000	3909	4317	2546	1980	510	0	2180	1980	4755	4999	1052	0	0	0	0	5
24	CE	ACE	5000	5000	5000	4995	4958	4747	346	4967	4958	5000	5000	2029	0	0	0	0	0
25	ACE	CE	5000	4803	4633	4970	333	12	0	478	331	3834	4987	435	4215	0	4218	0	0
25	AE	CE	5000	5000	5000	5000	5000	4999	2385	5000	5000	5000	5000	4993	0	0	0	0	0
25	CE	CE	5000	197	367	30	234	10	0	298	233	3422	4984	452	0	0	0	0	89
26	ACE	ACE	5000	304	145	1082	233	5	0	359	233	3632	4980	397	54	0	54	0	0
26	AE	ACE	5000	5000	5000	4995	4979	4546	20	4988	4979	5000	5000	4431	0	0	0	0	0
26	CE	ACE	5000	4696	4855	3923	3281	965	0	3444	3280	4945	4999	616	0	0	0	0	0
27	ACE	ACE	5000	18	5	127	252	2	0	377	252	3676	4985	394	0	1	0	1	0
27	AE	ACE	5000	4982	4995	4875	4626	2418	0	4698	4626	4999	5000	3067	0	0	0	0	0

condition	fitmodel	TrueModel	# Models converged	aBIC	AIC	BIC	ChiSqM PValue	CFI	TLI	RMSEA Estimate	RMSEA 90CI_LB	RMSEA 90CI_UB	RMSEA pLT05	SRMR	a_uns_pval	c_uns_pval	a_stz_pval	c_stz_pval	nested_sig
27	CE	ACE	5000	5000	5000	4998	4971	4285	17	4979	4971	5000	5000	1124	0	0	0	0	0
28	ACE	E	4974	4974	4974	4974	514	2875	1331	723	513	4193	4968	807	4557	4570	4555	4567	0
28	AE	E	4994	1324	1328	1322	344	2415	1156	427	344	3657	4971	823	4186	0	4178	0	134
28	CE	E	5000	3670	3666	3672	349	2439	1179	419	348	3654	4976	823	0	4198	0	4193	138
29	ACE	AE	5000	5000	4996	5000	320	1540	532	467	320	3870	4990	580	722	4189	718	4181	0
29	AE	AE	5000	747	749	747	245	1356	343	299	245	3452	4985	639	0	0	0	0	75
29	CE	AE	5000	4253	4255	4253	1339	3136	1513	1509	1337	4533	4996	1394	0	0	0	0	16
30	ACE	AE	5000	4915	4667	5000	364	761	40	520	363	3852	4993	618	34	4132	34	4125	0
30	AE	AE	5000	252	425	195	273	655	16	348	273	3522	4988	655	0	0	0	0	81
30	CE	AE	5000	4833	4908	4805	3565	4234	1541	3727	3565	4956	5000	2416	0	0	0	0	1
31	ACE	CE	5000	4996	4927	5000	338	1130	183	492	338	3873	4991	579	4174	208	4170	207	0
31	AE	CE	5000	4559	4588	4556	2200	3607	1307	2418	2198	4793	5000	1936	0	0	0	0	1
31	CE	CE	5000	445	485	444	262	979	114	316	261	3434	4981	622	0	0	0	0	81
32	ACE	ACE	5000	3924	2746	4993	244	380	8	368	244	3632	4989	479	898	590	892	584	0
32	AE	ACE	5000	3253	3836	2732	1378	1801	67	1581	1378	4581	4996	1132	0	0	0	0	12
32	CE	ACE	5000	2823	3418	2275	1113	1515	48	1272	1113	4447	4997	834	0	0	0	0	13
33	ACE	ACE	5000	2707	1925	4612	261	175	1	377	260	3626	4986	463	53	1004	53	992	0
33	AE	ACE	5000	2561	3208	1228	1021	824	1	1181	1018	4355	4996	843	0	0	0	0	15
33	CE	ACE	5000	4732	4867	4160	3445	3271	146	3621	3445	4950	5000	1457	0	0	0	0	0
34	ACE	CE	5000	4815	4601	4988	369	292	0	522	368	3844	4991	485	4155	1	4152	1	0
34	AE	CE	5000	4988	4998	4952	4697	4730	1284	4757	4696	4997	5000	3862	0	0	0	0	0
34	CE	CE	5000	197	401	60	273	255	0	340	273	3424	4977	500	0	0	0	0	78
35	ACE	ACE	5000	2386	1720	4145	284	115	0	404	284	3650	4979	461	976	8	976	8	0
35	AE	ACE	5000	4919	4969	4564	3960	3524	60	4124	3960	4988	5000	2587	0	0	0	0	0
35	CE	ACE	5000	2695	3311	1291	1074	679	2	1242	1072	4395	4998	669	0	0	0	0	13
36	ACE	ACE	5000	625	305	2200	254	53	0	373	254	3662	4986	427	57	47	57	45	0
36	AE	ACE	5000	4682	4842	3860	3199	2051	1	3395	3198	4941	4999	1808	0	0	0	0	0
36	CE	ACE	5000	4693	4853	3940	3260	2237	10	3443	3260	4942	5000	1016	0	0	0	0	0
37	ACE	E	4962	4962	4962	4962	532	2931	1370	136	532	2704	4941	41	4558	4562	4558	4561	0
37	AE	E	4990	1303	1305	1306	352	2461	1194	52	352	1859	4910	43	4201	0	4201	0	97
37	CE	E	5000	3687	3685	3684	350	2475	1209	53	350	1862	4922	44	0	4201	0	4200	107
38	ACE	AE	5000	4876	4620	4994	323	150	0	50	322	2114	4945	15	0	4188	0	4185	0
38	AE	AE	5000	127	380	30	230	128	0	26	230	1504	4890	15	0	0	0	0	89
38	CE	AE	5000	4997	5000	4976	4865	4780	1187	4352	4865	4993	5000	876	0	0	0	0	0
39	ACE	AE	5000	4876	4638	4971	306	6	0	68	306	2156	4954	26	0	4185	0	4178	0
39	AE	AE	5000	124	362	29	235	4	0	32	235	1492	4904	26	0	0	0	0	100
39	CE	AE	5000	5000	5000	5000	5000	4999	2788	4999	5000	5000	5000	3188	0	0	0	0	0
40	ACE	CE	5000	4845	4608	4978	341	21	0	57	341	2154	4952	16	4199	0	4198	0	0
40	AE	CE	5000	5000	5000	5000	4995	4971	874	4976	4995	4999	5000	3041	0	0	0	0	0
40	CE	CE	5000	155	392	22	252	14	0	30	252	1536	4901	17	0	0	0	0	105
41	ACE	ACE	5000	34	6	224	260	0	0	38	259	1889	4942	24	1	1	1	1	0

condition	fitmodel	TrueModel	# Models converged	aBIC	AIC	BIC	ChiSqM PValue	CFI	TLI	RMSEA Estimate	RMSEA 90CI_LB	RMSEA 90CI_UB	RMSEA pLT05	SRMR	a_uns_pval	c_uns_pval	a_stz_pval	c_stz_pval	nested_sig
41	AE	ACE	5000	4984	4998	4876	4707	2003	0	3954	4707	4975	5000	428	0	0	0	0	0
41	CE	ACE	5000	4982	4996	4900	4753	2293	0	4105	4753	4972	5000	103	0	0	0	0	0
42	ACE	ACE	5000	243	88	850	245	0	0	43	244	1816	4912	14	0	28	0	27	0
42	AE	ACE	5000	4757	4912	4150	3682	143	0	2187	3681	4745	5000	118	0	0	0	0	0
42	CE	ACE	5000	5000	5000	5000	5000	4956	85	5000	5000	5000	5000	555	0	0	0	0	0
43	ACE	CE	5000	4884	4625	4980	351	0	0	75	351	2229	4962	18	4232	0	4231	0	0
43	AE	CE	5000	5000	5000	5000	5000	5000	2393	5000	5000	5000	5000	5000	0	0	0	0	0
43	CE	CE	5000	116	375	20	250	0	0	36	249	1574	4919	19	0	0	0	0	106
44	ACE	ACE	5000	18	4	102	247	0	0	44	247	1850	4930	23	0	0	0	0	0
44	AE	ACE	5000	5000	5000	5000	5000	4890	0	5000	5000	5000	5000	4187	0	0	0	0	0
44	CE	ACE	5000	4982	4996	4898	4717	561	0	3961	4716	4974	5000	44	0	0	0	0	0
45	ACE	ACE	5000	0	0	1	264	0	0	49	264	1913	4923	23	0	0	0	0	0
45	AE	ACE	5000	5000	5000	4999	4996	2503	0	4971	4996	5000	5000	1777	0	0	0	0	0
45	CE	ACE	5000	5000	5000	5000	5000	4718	2	4999	5000	5000	5000	164	0	0	0	0	0
46	ACE	E	4945	4945	4945	4945	531	2951	1331	121	529	2682	4929	50	4580	4539	4580	4539	0
46	AE	E	4990	1403	1395	1390	347	2451	1165	43	346	1824	4923	50	4199	0	4196	0	208
46	CE	E	5000	3587	3595	3600	342	2463	1164	43	342	1823	4928	50	0	4194	0	4193	178
47	ACE	AE	5000	4998	4875	5000	340	872	98	65	340	2172	4950	24	133	4189	133	4186	0
47	AE	AE	5000	343	413	342	250	778	50	28	250	1503	4891	25	0	0	0	0	82
47	CE	AE	5000	4659	4712	4658	2648	3735	1191	1204	2646	4338	4997	187	0	0	0	0	0
48	ACE	AE	5000	4862	4577	4993	329	153	1	60	328	2138	4963	16	0	4136	0	4132	0
48	AE	AE	5000	144	425	36	240	137	0	31	240	1549	4909	19	0	0	0	0	79
48	CE	AE	5000	4994	4998	4971	4863	4767	1198	4334	4863	4989	5000	861	0	0	0	0	0
49	ACE	CE	5000	4948	4639	5000	367	445	9	56	366	2162	4939	16	4134	14	4133	14	0
49	AE	CE	5000	4867	4946	4852	3880	4190	901	2502	3880	4796	4999	442	0	0	0	0	0
49	CE	CE	5000	185	415	148	250	391	5	33	250	1574	4889	16	0	0	0	0	82
50	ACE	ACE	5000	1760	744	4013	261	56	0	57	261	1941	4928	21	211	82	210	80	0
50	AE	ACE	5000	4357	4745	3292	2826	1695	0	1307	2825	4422	4998	114	0	0	0	0	2
50	CE	ACE	5000	3883	4511	2695	2296	1247	1	934	2295	4126	4993	63	0	0	0	0	2
51	ACE	ACE	5000	1364	659	2796	291	13	0	63	291	1885	4922	20	0	294	0	291	0
51	AE	ACE	5000	3646	4341	2281	2042	393	0	743	2041	3871	4993	67	0	0	0	0	3
51	CE	ACE	5000	4990	5000	4923	4793	3741	17	4193	4793	4981	5000	244	0	0	0	0	0
52	ACE	CE	5000	4881	4602	4977	322	27	0	67	322	2185	4959	26	4165	0	4164	0	0
52	AE	CE	5000	5000	5000	5000	5000	4971	849	4978	5000	5000	5000	3025	0	0	0	0	0
52	CE	CE	5000	119	398	23	236	23	0	41	236	1532	4912	25	0	0	0	0	84
53	ACE	ACE	5000	1308	609	2632	250	2	0	52	249	1831	4934	21	268	0	268	0	0
53	AE	ACE	5000	4999	5000	4988	4943	4162	6	4682	4943	5000	5000	1082	0	0	0	0	0
53	CE	ACE	5000	3693	4391	2380	2054	254	0	767	2053	3963	4990	38	0	0	0	0	2
54	ACE	ACE	5000	29	4	206	255	0	0	52	255	1869	4933	20	0	0	0	0	0
54	AE	ACE	5000	4982	4998	4891	4726	1983	0	3967	4725	4972	5000	399	0	0	0	0	0
54	CE	ACE	5000	4989	4998	4903	4752	2264	0	4070	4751	4989	5000	99	0	0	0	0	0

condition	fitmodel	TrueModel	# Models converged	aBIC	AIC	BIC	ChiSqM PValue	CFI	TLI	RMSEA Estimate	RMSEA 90CI_LB	RMSEA 90CI_UB	RMSEA pLT05	SRMR	a_uns_pval	c_uns_pval	a_stz_pval	c_stz_pval	nested_sig
55	ACE	E	4933	4933	4933	4933	465	2805	1262	1	465	366	4628	0	4553	4547	4553	4547	0
55	AE	E	4983	1354	1355	1355	294	2336	1093	1	293	97	4168	0	4205	0	4205	0	214
55	CE	E	5000	3629	3628	3628	287	2349	1110	1	286	104	4182	0	0	4230	0	4230	209
56	ACE	AE	5000	4931	4608	4989	345	8	0	0	343	263	4443	0	0	4158	0	4157	0
56	AE	AE	5000	69	392	11	255	7	0	0	255	96	3984	0	0	0	0	0	108
56	CE	AE	5000	5000	5000	5000	4999	4984	831	4830	4999	4999	5000	126	0	0	0	0	0
57	ACE	AE	5000	4902	4605	4988	362	0	0	1	362	286	4497	0	0	4140	0	4134	0
57	AE	AE	5000	98	395	12	279	0	0	0	278	120	4097	0	0	0	0	0	121
57	CE	AE	5000	5000	5000	5000	5000	5000	2890	5000	5000	5000	5000	2213	0	0	0	0	0
58	ACE	CE	5000	4919	4654	4983	328	0	0	0	328	257	4484	0	4244	0	4244	0	0
58	AE	CE	5000	5000	5000	5000	5000	5000	456	5000	5000	5000	5000	2066	0	0	0	0	0
58	CE	CE	5000	81	346	17	247	0	0	0	247	104	4031	0	0	0	0	0	95
59	ACE	ACE	5000	0	0	0	237	0	0	1	236	187	4329	0	0	0	0	0	0
59	AE	ACE	5000	5000	5000	5000	4999	1932	0	4503	4999	4995	5000	21	0	0	0	0	0
59	CE	ACE	5000	5000	5000	5000	4999	2305	0	4582	4999	4994	5000	2	0	0	0	0	0
60	ACE	ACE	5000	5	0	46	240	0	0	1	240	178	4343	0	0	0	0	0	0
60	AE	ACE	5000	4995	5000	4954	4904	19	0	2274	4904	4764	5000	3	0	0	0	0	0
60	CE	ACE	5000	5000	5000	5000	5000	4999	4	5000	5000	5000	5000	54	0	0	0	0	0
61	ACE	CE	5000	4916	4622	4984	373	0	0	1	373	274	4457	0	4181	0	4181	0	0
61	AE	CE	5000	5000	5000	5000	5000	5000	2365	5000	5000	5000	5000	5000	0	0	0	0	0
61	CE	CE	5000	84	378	16	256	0	0	0	256	94	4012	0	0	0	0	0	109
62	ACE	ACE	5000	0	0	1	216	0	0	0	215	155	4283	1	0	0	0	0	0
62	AE	ACE	5000	5000	5000	5000	5000	4997	0	5000	5000	5000	5000	4036	0	0	0	0	0
62	CE	ACE	5000	5000	5000	4999	5000	190	0	4524	5000	4993	5000	1	0	0	0	0	0
63	ACE	ACE	5000	0	0	0	268	0	0	1	268	200	4323	0	0	0	0	0	0
63	AE	ACE	5000	5000	5000	5000	5000	2757	0	5000	5000	5000	5000	621	0	0	0	0	0
63	CE	ACE	5000	5000	5000	5000	5000	4964	0	5000	5000	5000	5000	5	0	0	0	0	0
64	ACE	E	4922	4922	4922	4922	497	2827	1261	0	496	395	4617	0	4565	4509	4565	4509	0
64	AE	E	4990	1408	1402	1414	323	2358	1100	0	322	152	4205	0	4189	0	4188	0	399
64	CE	E	5000	3582	3588	3576	317	2348	1095	0	316	147	4218	0	0	4171	0	4170	348
65	ACE	AE	5000	4980	4632	5000	338	267	1	0	337	257	4448	0	5	4176	5	4176	0
65	AE	AE	5000	93	382	77	241	233	0	0	241	96	4023	0	0	0	0	0	163
65	CE	AE	5000	4927	4986	4923	4331	4323	766	800	4331	3849	4997	2	0	0	0	0	0
66	ACE	AE	5000	4910	4598	4984	296	7	0	0	295	236	4524	0	0	4186	0	4183	0
66	AE	AE	5000	90	402	16	230	5	0	0	230	93	4066	0	0	0	0	0	109
66	CE	AE	5000	5000	5000	5000	5000	4979	886	4787	5000	4998	5000	115	0	0	0	0	0
67	ACE	CE	5000	4915	4618	4997	326	50	0	1	325	265	4494	0	4184	0	4184	0	0
67	AE	CE	5000	4997	5000	4990	4932	4763	484	2738	4932	4845	5000	17	0	0	0	0	0
67	CE	CE	5000	88	382	13	247	44	0	0	247	98	4008	0	0	0	0	0	128
68	ACE	ACE	5000	301	57	1109	243	0	0	0	242	195	4330	0	14	1	14	1	0
68	AE	ACE	5000	4926	4992	4692	4503	1465	0	955	4502	4129	4998	0	0	0	0	0	0

condition	fitmodel	TrueModel	# Models converged	aBIC	AIC	BIC	ChiSqM PValue	CFI	TLI	RMSEA Estimate	RMSEA 90CI_LB	RMSEA 90CI_UB	RMSEA pLT05	SRMR	a_uns_pval	c_uns_pval	a_stz_pval	c_stz_pval	nested_sig
68	CE	ACE	5000	4773	4951	4199	3975	828	0	490	3974	3392	4990	0	0	0	0	0	0
69	ACE	ACE	5000	329	74	1032	223	0	0	0	221	174	4316	0	0	20	0	20	0
69	AE	ACE	5000	4671	4926	3968	3736	88	0	340	3736	3095	4981	2	0	0	0	0	0
69	CE	ACE	5000	5000	5000	5000	4999	4224	0	4677	4999	4999	5000	10	0	0	0	0	0
70	ACE	CE	5000	4903	4568	4985	325	0	0	0	323	262	4479	0	4169	0	4169	0	0
70	AE	CE	5000	5000	5000	5000	5000	4999	466	5000	5000	5000	5000	1986	0	0	0	0	0
70	CE	CE	5000	97	432	15	249	0	0	0	249	92	4021	0	0	0	0	0	106
71	ACE	ACE	5000	297	47	887	256	0	0	0	256	217	4307	0	14	0	14	0	0
71	AE	ACE	5000	5000	5000	5000	5000	4648	0	4958	5000	5000	5000	159	0	0	0	0	0
71	CE	ACE	5000	4703	4953	4113	3883	42	0	452	3881	3241	4994	1	0	0	0	0	0
72	ACE	ACE	5000	0	0	1	234	0	0	1	234	177	4339	0	0	0	0	0	0
72	AE	ACE	5000	5000	5000	4999	4993	1888	0	4514	4993	4988	5000	25	0	0	0	0	0
72	CE	ACE	5000	5000	5000	5000	5000	2304	0	4590	5000	4996	5000	1	0	0	0	0	0

Table 16 Correctly-specified model rejection rate correlations with total variance and n

Rejection Rate	n	Total Variance
aBIC	-0.730	0.166
AIC	-0.654	0.003
BIC	-0.632	0.346
ChiSqM_PValue	-0.484	0.032
CFI	-0.803	-0.235
TLI	-0.720	-0.162
RMSEA_Estimate	-0.972	0.129
RMSEA_90CI_LB	-0.487	0.033
RMSEA_90CI_UB	-0.968	0.126
RMSEA_pLT05	-0.967	0.122
SRMR	-0.975	-0.052
a_uns_pval	-0.509	0.300
c_uns_pval	-0.373	0.451
a_stz_pval	-0.401	0.394
c_stz_pval	-0.373	0.451
nested_crit	-0.024	-0.626
nested_sig	-0.024	-0.626

Table 17 Optimized cutpoint using the Youden index

predictor	subgroup	optimal cutpoint	metric value	acc	sensitivity	specificity	AUC	PPV	NPV	TNR	FNR	TPR	FPR
a_stz_pval	All	0.032	0.274	0.605	0.716	0.558	0.622	0.406	0.824	0.558	0.284	0.716	0.442
a_stz_pval	p2	0.000	0.068	0.553	0.499	0.568	0.517	0.249	0.799	0.568	0.501	0.499	0.432
a_stz_pval	p3	0.004	0.489	0.728	0.897	0.592	0.767	0.638	0.878	0.592	0.103	0.897	0.408
a_uns_pval	All	0.030	0.274	0.605	0.715	0.559	0.622	0.406	0.823	0.559	0.285	0.715	0.441
a_uns_pval	p2	0.000	0.068	0.553	0.499	0.569	0.517	0.249	0.799	0.569	0.501	0.499	0.431
a_uns_pval	p3	0.004	0.490	0.728	0.897	0.593	0.768	0.638	0.878	0.593	0.103	0.897	0.407
aBIC_diff	All	4.424	0.458	0.620	0.998	0.460	0.715	0.438	0.998	0.460	0.002	0.998	0.540
aBIC_diff	p2	4.425	0.620	0.707	0.996	0.624	0.761	0.431	0.998	0.624	0.004	0.996	0.376
aBIC_diff	p3	-Inf	0.000	0.555	0.000	1.000	0.500	NA	0.555	1.000	1.000	0.000	0.000
AIC_diff	All	2.001	0.519	0.667	0.987	0.531	0.740	0.470	0.990	0.531	0.013	0.987	0.469
AIC_diff	p2	2.001	0.696	0.776	0.977	0.719	0.829	0.499	0.991	0.719	0.023	0.977	0.281
AIC_diff	p3	-Inf	0.000	0.555	0.000	1.000	0.500	NA	0.555	1.000	1.000	0.000	0.000
AICC	All	14352.745	0.034	0.639	0.219	0.816	0.514	0.333	0.712	0.816	0.781	0.219	0.184
AICC	p2	4898.233	0.000	0.500	0.500	0.500	0.466	0.222	0.778	0.500	0.500	0.500	0.500
AICC	p3	15644.788	0.133	0.608	0.193	0.940	0.586	0.721	0.592	0.940	0.807	0.193	0.060
BIC_diff	All	7.601	0.349	0.542	0.999	0.349	0.657	0.393	0.999	0.349	0.001	0.999	0.651
BIC_diff	p2	7.602	0.473	0.591	0.999	0.474	0.604	0.352	1.000	0.474	0.001	0.999	0.526
BIC_diff	p3	-Inf	0.000	0.555	0.000	1.000	0.500	NA	0.555	1.000	1.000	0.000	0.000
c_stz_pval	All	0.025	0.273	0.604	0.716	0.557	0.621	0.405	0.823	0.557	0.284	0.716	0.443
c_stz_pval	p2	0.000	0.069	0.553	0.500	0.569	0.517	0.249	0.799	0.569	0.500	0.500	0.431
c_stz_pval	p3	0.005	0.477	0.720	0.908	0.569	0.758	0.628	0.885	0.569	0.092	0.908	0.431
c_uns_pval	All	0.030	0.272	0.603	0.717	0.555	0.620	0.405	0.823	0.555	0.283	0.717	0.445
c_uns_pval	p2	0.000	0.069	0.554	0.500	0.569	0.517	0.249	0.799	0.569	0.500	0.500	0.431
c_uns_pval	p3	0.008	0.476	0.719	0.912	0.564	0.758	0.626	0.889	0.564	0.088	0.912	0.436
CFI	All	0.981	0.499	0.709	0.849	0.650	0.775	0.505	0.911	0.650	0.151	0.849	0.350
CFI	p2	0.981	0.583	0.784	0.805	0.777	0.826	0.508	0.933	0.777	0.195	0.805	0.223
CFI	p3	0.984	0.185	0.561	0.876	0.309	0.590	0.504	0.757	0.309	0.124	0.876	0.691
ChiSqM_diff	All	2.122	0.537	0.689	0.964	0.573	0.775	0.488	0.974	0.573	0.036	0.964	0.427
ChiSqM_diff	p2	2.590	0.709	0.804	0.946	0.763	0.887	0.533	0.980	0.763	0.054	0.946	0.237
ChiSqM_diff	p3	-Inf	0.000	0.555	0.000	1.000	0.500	NA	0.555	1.000	1.000	0.000	0.000
ChiSqM_PValue	All	0.103	0.437	0.647	0.895	0.542	0.774	0.452	0.925	0.542	0.105	0.895	0.458
ChiSqM_PValue	p2	0.098	0.579	0.729	0.900	0.680	0.852	0.445	0.960	0.680	0.100	0.900	0.320
ChiSqM_PValue	p3	0.417	0.077	0.534	0.578	0.500	0.555	0.481	0.596	0.500	0.422	0.578	0.500
nested_sig	All	0.146	0.537	0.689	0.964	0.573	0.775	0.488	0.974	0.573	0.036	0.964	0.427
nested_sig	p2	0.108	0.709	0.804	0.946	0.763	0.887	0.533	0.980	0.763	0.054	0.946	0.237
nested_sig	p3	Inf	0.000	0.555	0.000	1.000	0.500	NA	0.555	1.000	1.000	0.000	0.000
RMSEA_90CI_LB	All	0.000	0.426	0.617	0.949	0.477	0.722	0.433	0.957	0.477	0.051	0.949	0.523
RMSEA_90CI_LB	p2	0.000	0.569	0.694	0.949	0.621	0.796	0.417	0.977	0.621	0.051	0.949	0.379
RMSEA_90CI_LB	p3	0.000	0.025	0.464	0.949	0.076	0.512	0.451	0.649	0.076	0.051	0.949	0.924
RMSEA_90CI_UB	All	0.082	0.358	0.674	0.691	0.667	0.728	0.467	0.837	0.667	0.309	0.691	0.333

predictor	subgroup	optimal cutpoint	metric value	acc	sensitivity	specificity	AUC	PPV	NPV	TNR	FNR	TPR	FPR
RMSEA_90CI_UB	p2	0.082	0.482	0.761	0.705	0.777	0.802	0.474	0.902	0.777	0.295	0.705	0.223
RMSEA_90CI_UB	p3	0.047	0.053	0.547	0.340	0.713	0.536	0.487	0.574	0.713	0.660	0.340	0.287
RMSEA_Estimate	All	0.039	0.442	0.675	0.835	0.607	0.761	0.472	0.897	0.607	0.165	0.835	0.393
RMSEA_Estimate	p2	0.041	0.585	0.761	0.850	0.736	0.840	0.479	0.945	0.736	0.150	0.850	0.264
RMSEA_Estimate	p3	0.004	0.077	0.534	0.578	0.500	0.544	0.481	0.596	0.500	0.422	0.578	0.500
RMSEA_pLT05	All	0.502	0.432	0.659	0.856	0.576	0.778	0.460	0.905	0.576	0.144	0.856	0.424
RMSEA_pLT05	p2	0.488	0.575	0.743	0.868	0.707	0.863	0.459	0.949	0.707	0.132	0.868	0.293
RMSEA_pLT05	p3	0.877	0.057	0.529	0.525	0.532	0.541	0.474	0.583	0.532	0.475	0.525	0.468
SRMR	All	0.048	0.272	0.670	0.551	0.720	0.668	0.454	0.792	0.720	0.449	0.551	0.280
SRMR	p2	0.049	0.335	0.738	0.540	0.795	0.704	0.429	0.858	0.795	0.460	0.540	0.205
SRMR	p3	0.038	0.057	0.540	0.428	0.629	0.541	0.480	0.579	0.629	0.572	0.428	0.371
TLI	All	0.995	0.453	0.681	0.840	0.613	0.757	0.478	0.901	0.613	0.160	0.840	0.387
TLI	p2	0.995	0.548	0.757	0.804	0.744	0.818	0.473	0.930	0.744	0.196	0.804	0.256
TLI	p3	0.995	0.123	0.527	0.876	0.247	0.542	0.483	0.714	0.247	0.124	0.876	0.753

Table 18 Optimized cutpoint when specificity was conditionally maximized

predictor	subgroup	optimal_cutpoint	metric_value	acc	sensitivity	specificity	AUC	PPV	NPV	TNR	FNR	TPR	FPR
a_stz_pval	p3	0.000	0.622	0.726	0.856	0.622	0.767	0.644	0.843	0.622	0.144	0.856	0.378
a_uns_pval	p3	0.000	0.623	0.726	0.855	0.623	0.768	0.645	0.843	0.623	0.145	0.855	0.377
aBIC_diff	All	3.363	0.532	0.612	0.800	0.532	0.715	0.419	0.863	0.532	0.200	0.800	0.468
aBIC_diff	p2	3.823	0.674	0.702	0.800	0.674	0.761	0.412	0.922	0.674	0.200	0.800	0.326
AIC_diff	All	2.000	0.531	0.667	0.987	0.531	0.740	0.470	0.990	0.531	0.013	0.987	0.469
AIC_diff	p2	2.000	0.721	0.777	0.974	0.721	0.829	0.500	0.990	0.721	0.026	0.974	0.279
AICC	All	1681.133	0.206	0.382	0.800	0.206	0.514	0.298	0.710	0.206	0.200	0.800	0.794
AICC	p2	1621.613	0.149	0.294	0.800	0.149	0.466	0.212	0.723	0.149	0.200	0.800	0.851
AICC	p3	1727.584	0.250	0.495	0.800	0.250	0.586	0.461	0.610	0.250	0.200	0.800	0.750
BIC_diff	All	6.479	0.415	0.529	0.800	0.415	0.657	0.366	0.831	0.415	0.200	0.800	0.585
BIC_diff	p2	6.908	0.521	0.583	0.800	0.521	0.604	0.323	0.901	0.521	0.200	0.800	0.479
c_stz_pval	p3	0.000	0.596	0.718	0.870	0.596	0.758	0.633	0.851	0.596	0.130	0.870	0.404
c_uns_pval	p3	0.000	0.597	0.716	0.865	0.597	0.758	0.632	0.846	0.597	0.135	0.865	0.403
CFI	All	0.988	0.688	0.721	0.801	0.688	0.775	0.519	0.891	0.688	0.199	0.801	0.312
CFI	p2	0.981	0.777	0.784	0.805	0.777	0.826	0.508	0.933	0.777	0.195	0.805	0.223
CFI	p3	0.992	0.374	0.565	0.803	0.374	0.590	0.507	0.703	0.374	0.197	0.803	0.626
ChiSqM_diff	All	0.065	0.643	0.690	0.800	0.643	0.775	0.486	0.884	0.643	0.200	0.800	0.357
ChiSqM_diff	p2	0.711	0.828	0.821	0.800	0.828	0.887	0.570	0.935	0.828	0.200	0.800	0.172
ChiSqM_PValue	All	0.196	0.622	0.675	0.800	0.622	0.774	0.472	0.881	0.622	0.200	0.800	0.378
ChiSqM_PValue	p2	0.196	0.752	0.763	0.800	0.752	0.852	0.480	0.929	0.752	0.200	0.800	0.248
ChiSqM_PValue	p3	0.196	0.259	0.499	0.800	0.259	0.555	0.464	0.618	0.259	0.200	0.800	0.741
nested_sig	All	0.799	0.643	0.690	0.800	0.643	0.775	0.486	0.884	0.643	0.200	0.800	0.357
nested_sig	p2	0.399	0.828	0.821	0.800	0.828	0.887	0.570	0.935	0.828	0.200	0.800	0.172
RMSEA_90CI_LB	All	0.000	0.477	0.617	0.949	0.477	0.722	0.433	0.957	0.477	0.051	0.949	0.523
RMSEA_90CI_LB	p2	0.000	0.621	0.694	0.949	0.621	0.796	0.417	0.977	0.621	0.051	0.949	0.379
RMSEA_90CI_LB	p3	0.000	0.076	0.464	0.949	0.076	0.512	0.451	0.649	0.076	0.051	0.949	0.924
RMSEA_90CI_UB	All	0.102	0.522	0.604	0.801	0.522	0.728	0.414	0.861	0.522	0.199	0.801	0.478
RMSEA_90CI_UB	p2	0.100	0.639	0.675	0.804	0.639	0.802	0.389	0.919	0.639	0.196	0.804	0.361
RMSEA_90CI_UB	p3	0.105	0.227	0.483	0.802	0.227	0.536	0.454	0.589	0.227	0.198	0.802	0.773
RMSEA_Estimate	All	0.034	0.635	0.685	0.802	0.635	0.761	0.481	0.884	0.635	0.198	0.802	0.365
RMSEA_Estimate	p2	0.034	0.771	0.779	0.805	0.771	0.840	0.501	0.933	0.771	0.195	0.805	0.229
RMSEA_Estimate	p3	0.035	0.246	0.495	0.806	0.246	0.544	0.462	0.614	0.246	0.194	0.806	0.754
RMSEA_pLT05	All	0.603	0.627	0.678	0.800	0.627	0.778	0.475	0.882	0.627	0.200	0.800	0.373
RMSEA_pLT05	p2	0.614	0.765	0.773	0.800	0.765	0.863	0.493	0.931	0.765	0.200	0.800	0.235
RMSEA_pLT05	p3	0.592	0.247	0.493	0.800	0.247	0.541	0.460	0.607	0.247	0.200	0.800	0.753
SRMR	All	0.076	0.364	0.495	0.805	0.364	0.668	0.348	0.816	0.364	0.195	0.805	0.636
SRMR	p2	0.078	0.392	0.483	0.801	0.392	0.704	0.274	0.873	0.392	0.199	0.801	0.608
SRMR	p3	0.073	0.233	0.487	0.803	0.233	0.541	0.456	0.597	0.233	0.197	0.803	0.767
TLI	All	0.996	0.632	0.687	0.818	0.632	0.757	0.484	0.892	0.632	0.182	0.818	0.368
TLI	p2	0.995	0.744	0.757	0.804	0.744	0.818	0.473	0.930	0.744	0.196	0.804	0.256
TLI	p3	0.997	0.291	0.530	0.827	0.291	0.542	0.483	0.678	0.291	0.173	0.827	0.709

Table 19 Optimized cutpoint when sensitivity was conditionally maximized

predictor	subgroup	optimal_cutpoint	metric_value	acc	sensitivity	specificity	AUC	PPV	NPV	TNR	FNR	TPR	FPR
aBIC_diff	p2	0.746	0.039	0.747	0.039	0.950	0.761	0.184	0.776	0.950	0.961	0.039	0.050
AIC_diff	p2	0.769	0.086	0.758	0.086	0.950	0.829	0.330	0.784	0.950	0.914	0.086	0.050
AICC	All	16563.071	0.056	0.685	0.056	0.950	0.514	0.319	0.705	0.950	0.944	0.056	0.050
AICC	p3	16089.869	0.156	0.597	0.156	0.950	0.586	0.714	0.584	0.950	0.844	0.156	0.050
BIC_diff	p2	0.821	0.006	0.740	0.006	0.950	0.604	0.031	0.770	0.950	0.994	0.006	0.050
ChiSqM_PValue	All	0.845	0.152	0.714	0.152	0.950	0.774	0.562	0.727	0.950	0.848	0.152	0.050
ChiSqM_PValue	p2	0.749	0.248	0.794	0.248	0.950	0.852	0.586	0.815	0.950	0.752	0.248	0.050
ChiSqM_PValue	p3	0.925	0.074	0.561	0.074	0.950	0.555	0.545	0.562	0.950	0.926	0.074	0.050
RMSEA_90CI_UB	All	0.031	0.157	0.717	0.157	0.953	0.728	0.584	0.728	0.953	0.843	0.157	0.047
RMSEA_90CI_UB	p2	0.043	0.317	0.809	0.317	0.950	0.802	0.645	0.830	0.950	0.683	0.317	0.050
RMSEA_90CI_UB	p3	0.018	0.074	0.560	0.074	0.950	0.536	0.542	0.561	0.950	0.926	0.074	0.050
RMSEA_pLT05	All	0.995	0.158	0.717	0.158	0.953	0.778	0.585	0.729	0.953	0.842	0.158	0.047
RMSEA_pLT05	p2	0.974	0.321	0.810	0.321	0.950	0.863	0.648	0.831	0.950	0.679	0.321	0.050
RMSEA_pLT05	p3	1.000	0.039	0.558	0.039	0.973	0.541	0.538	0.558	0.973	0.961	0.039	0.027
SRMR	All	0.025	0.154	0.717	0.154	0.955	0.668	0.588	0.728	0.955	0.846	0.154	0.045
SRMR	p2	0.033	0.281	0.802	0.281	0.951	0.704	0.619	0.822	0.951	0.719	0.281	0.049
SRMR	p3	0.019	0.076	0.562	0.076	0.952	0.541	0.560	0.563	0.952	0.924	0.076	0.048
TLI	All	1.016	0.060	0.688	0.060	0.952	0.757	0.346	0.706	0.952	0.940	0.060	0.048
TLI	p2	1.008	0.186	0.781	0.186	0.951	0.818	0.518	0.803	0.951	0.814	0.186	0.049
TLI	p3	1.029	0.005	0.530	0.005	0.950	0.542	0.073	0.544	0.950	0.995	0.005	0.050