ABSTRACT

Title of Dissertation: CHARACTERIZING THE ADVENTITIOUS MODEL ERROR AS A RANDOM EFFECT IN ITEM-RESPONSE-THEORY MODELS

Shuangshuang Xu, Doctor of Philosophy, 2023

Dissertation Directed by: Associate Professor Yang Liu
Department of Human Development and Quantitative Methodology

When drawing conclusions from statistical inferences, researchers are usually concerned about two types of errors: sampling error and model error. The sampling error is caused by the discrepancy between the observed sample and the population from which the sample is drawn from (i.e., operational population). The model error refers to the discrepancy between the fitted model and the data-generating mechanism. Most item response theory (IRT) models assume that models are correctly specified in the population of interest; as a result, only sampling errors are characterized, not model errors. The model error can be treated either as fixed or random.

The proposed framework in this study treats the model error as a random effect (i.e., an adventitious error) and provides an alternative explanation for the model errors in IRT models that originate from unknown sources. A random, ideally small amount of discrepancy between the operational population and the fitted model is characterized using a Dirichlet-Multinomial

framework. A concentration/dispersion parameter is used in the Dirichlet-Multinomial framework to measure the amount of adventitious error between the operational population probability and the fitted model. In general, the study aims to: 1) build a Dirichlet-Multinomial framework for IRT models, 2) establish asymptotic results for estimating model parameters when the operational population probability is assumed known or unknown, 3) conduct numerical studies to investigate parameter recovery and the relationship between the concentration/dispersion parameter in the proposed framework and the Root Mean Square Error of Approximation (RMSEA), 4) correct bias in parameter estimates of the Dirichlet-Multinomial framework using asymptotic approximation methods, and 5) quantify the amount of model error in the framework and decide whether the model should be retained or rejected.

CHARACTERIZING THE ADVENTITIOUS MODEL ERROR AS A RANDOM EFFECT IN ITEM-RESPONSE-THEORY MODELS

by

Shuangshuang Xu

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2023

Advisory Committee:
Associate Professor Yang Liu, Chair/Advisor
Professor Gregory Hancock
Associate Professor Tracy Sweet
Associate Professor Ji Seung Yang
Associate Professor Xin He

Dedication

To my beloved family, whose unwavering love, encouragement, and support have been the guiding light throughout this academic journey.

Acknowledgments

I am sincerely grateful to my advisor, Dr. Yang Liu, for his guidance and support throughout my doctoral journey. His invaluable mentorship has played a pivotal role in shaping my research and academic growth, and I will forever cherish the knowledge and skills gained under his expert mentoring. I extend my heartfelt thanks to my esteemed committee members: Dr. Ji Seung Yang, whose generosity in providing opportunities to explore my specialty and collaborate with fellow researchers that enriched my academic experience; Drs. Gregory Hancock and Tracy Sweet, whose patience, encouragement, and unique perspectives introduced me to new dimensions of research; and Dr. Xin He, whose insights and assistance allowed me to view research challenges from fresh and diverse angles.

I want to thank my supervisors, Drs. Hong Jiao and Robert W. Lissitz, for providing me with invaluable opportunities to work with MARC. Collaborating with the Maryland State Department of Education broadened my perspective on quantitative research and its profound impact on future policy decisions for students. I am deeply thankful to Drs. Jeffery Harring, Laura Stapleton, and Peter Steiner for their constant support and encouragement. Their belief in my abilities has not only contributed to my academic success but has also made this journey truly enriching.

To the QMMS/EDMS program, I owe my profound appreciation. From a novice to an expert in the field, this program has transformed my knowledge and skills. The warm and welcom-

ing atmosphere from day one made me feel at home, and the program's comprehensive education has shaped the person I am today. I am forever grateful for the nurturing environment provided by our program.

I would also like to express my heartfelt appreciation to the remarkable students and colleagues I had the privilege of meeting in the program. Their warmth, diligence, and intelligence made our collective journey a rewarding and challenging one. Together, we provided support to each other, navigating the difficulties and emerging stronger. These shared experiences have forged enduring friendships that I will cherish for a lifetime.

Finally, I want to extend a special and heartfelt thank you to my family: my parents, grandparents, my beloved fiancé Alan Kong and his wonderful family. Your love, encouragement, and understanding have been the pillars of strength throughout my academic journey. Your support and belief in me have made all the difference, and I am deeply grateful for your presence in my life. Your constant encouragement and willingness to lend a helping hand have been a source of inspiration, and I am honored to be a part of this loving and supportive family. Thank you for being there for me every step of the way.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

2PL	Two-parameter Logistic Model
ADF	Asymptotically Distribution Free
AIC	Akaike Information Criterion
BC-MDMLE	Bias-Corrected Maximum Dirichlet-Multinomial Likelihood Estimator
BC-LL-DM	Bias-Corrected Log-Likelihood of the Dirichlet-Multinomial Distribution
CDF	Cumulative Distribution Function
CFA	Confirmatory Factor Analysis
CFI	Bentler's Comparative Fit Index
CI	Confidence Interval
CPES	Collaborative Psychiatric Epidemiological Surveys
CSM	Covariance Structure Model
CUU	Calibration Under Uncertainty
EM	Expectation-Maximization
FA	Factor Analysis
G-O-DM	General Objective Function of the Dirichlet-Multinomial Distribution
GFI	Goodness-of-Fit Index
GOF	Goodness-of-Fit
GRM	Graded Response Model
I.I.D.	Independent and Identically Distributed
IRF	Item Response Function
IRT	Item Response Theory
KL	Kullback-Leibler
LL-DM	Log-Likelihood of the Dirichlet-Multinomial Distribution
LR	Likelihood Ratio
LSAT7	Law School Admission Test Section 7

MDMLE	Maximum Dirichlet-Multinomial Likelihood Estimator
MLE	Maximum Likelihood Estimation
MMLE	Maximum Multinomial Likelihood Estimator
PDF	Probability Density Function
PMF	Probability Mass Function
RMSE	Root Mean Squared Error
RMSEA	Root Mean Square Error of Approximation
Science	The Consumer Protection and Perceptions of Science and Technology Section of the 1992 Euro-Barometer Survey
SE	Standard Error
SRMR	Squared Root Mean Residual
SRMSR	Standardized Root Mean Squared Residual
TLI	Tucker-Lewis Index
ULS	Unweighted Least Squares
UQ	Uncertainty Quantification
VGAM	An R Package for Vector Generalized Linear and Additive Models
WLS	Weighted Least Squares

Chapter 1: Introduction

1.1 Statement of the Problem

Researchers are interested in identifying the uncertainty of model predictions when making decisions based on statistical models. Major sources of uncertainties in model predictions include sampling error and model error. Sampling error happens when we look at a small group of examples (a sample) instead of the entire population we are interested in (Särndal, Swensson, & Wretman, 2003). It occurs because the sample might not perfectly represent the whole population. Sampling error is usually measured by the discrepancy between the observed sample and the population from which the sample is drawn from, and it can be captured by sampling distributions of parameter estimates and test statistics.

Model error is about the models themselves. When we create models, we try to make them match the real world as closely as possible, but they can still make mistakes. Model error occurs when the predictions made by our models do not match what actually happens in reality. In this particular study, model error refers to the discrepancy between a family of models to be fitted and the data generating mechanism (MacCallum & O'Hagan, 2015).

While sampling error decreases as we have a larger sample size, model error is about the models themselves and is not affected by the size of the sample. To assess model error, researchers can use either a fixed-effect or a random-effect approach, which are different methods

of analyzing and understanding the errors in the models. The fixed-effect and random-effect approaches differ in their assumptions about how samples are drawn from the population.

In the fixed-effect approach, we assume that all the samples used to build our model come from the same population. However, we believe that there may be a fixed amount of discrepancy between this population and the model we fitted. This fixed discrepancy, known as fixed-effect model error, can be quantified using effect size measures, which includes the absolute fit measures such as discrepancy functions and Root Mean Square Error of Approximation (RMSEA; Browne & Cudeck, 1992; Steiger, 1990, 2016), and incremental fit measures such as Bentler's Comparative Fit Index (CFI; Bentler, 1990) and Tucker-Lewis Index (TLI; Tucker & Lewis, 1973). They help us quantify the amount of discrepancy between the fitted models and the true population.

Alternatively, we can consider the error in the model as a random effect. Under this perspective, each sample is obtained from a particular population that differs slightly, in a random manner, from the general population of interest, where the hypothesized theory holds true. Adopting the random-effect approach means acknowledging that the data generation process in the particular population deviates from the hypothesized theory, acting as a form of "perturbation." The random-effect approach in model errors can be compared to the process of verifying a chemical law using results from different laboratories. For instance, consider a chemical law that can be verified in different laboratories, each with a slightly different experimental setup. In this scenario, we can consider the hypothesized theory as the chemical law we want to verify, while the laboratories represent particular populations and can be viewed as random because we can have infinite many laboratories to verify the chemical law. The result obtained from each lab can be seen as a sample obtained from its respective population.

Tucker, Koopman, and Linn (1969) was the first to acknowledge in factor analysis (FA)

models that the random-effect model error stems from pre-existing minor factors. This viewpoint was adopted in Briggs and MacCallum (2003) when they studied the parameter estimation in FA models. Wu and Browne (2015a) called the random-effect model error "an adventitious error" [1] and built an Inverted-Wishart-Beta framework to characterize it in covariance structure models (CSMs). As per Wu and Browne (2015a), the hypothesized theory (i.e., fitted model with certain true parameters) is assumed to always hold true for a general population of interest under a general measurement condition. When we collect samples to test our hypothesized theory, we obtain each sample from a single operational population, which is just one manifestation of the general population of interest. Tucker et al. (1969) argues that the difference between the general population of interest and each unique operational population, from which the each sample is obtained, is random in nature. They indicate that by aggregating samples from all operational populations, we can obtain a representative depiction of the general population of interest.

It is important to acknowledge that the model error, which refers to the difference between an operational population and the general population, is in theory distinct from sampling error. While sampling error occurs at the sample level, model error occurs at the population level. Despite both sampling error and model error being used to quantify model uncertainty, sampling error specifically addresses the fact that the samples we use to build models are not a flawless representation of the population we are interested in. By increasing the sample size, we can obtain more information about the target population, and resulting in more accurate model predictions. On the other hand, model error arises from discrepancy between the hypothesized theory and data-generating mechanism, neither of which are directly tied to the samples we obtain. However, since model error estimation still relies on sample data, it is common for model error and

[1] The names of random-effect model error and adventitious error are used interchangeably in this study.

sampling error to coexist.

1.2 Purpose of the Study

The current research is an extension of the adventitious error approach in Wu and Browne (2015a) and focuses on developing a stochastic framework to quantify the model errors in the analyses of contingency table data, or specifically, in the item response theory (IRT) models.

This study accomplishes four research objectives. Firstly, we propose an extension of Wu and Browne (2015a) by introducing a Dirichlet-Multinomial framework to model the adventitious error in IRT models. Secondly, we present asymptotic results for two maximum likelihood estimators: one for the case where the population probability is known using the Dirichlet likelihood, and another for the case where the population probability is unknown using the Dirichlet-Multinomial likelihood. To quantify the adventitious error within this framework, we estimate a concentration/dispersion parameter. Thirdly, we develop a bias-correction method to address the estimation bias in the concentration/dispersion parameter when the population probability is assumed to be unknown, which is a more practical scenario. Finally, we establish a connection between the corrected concentration/dispersion parameter and RMSEA, which is an existing model fit index in fixed-effect model error approaches.

1.3 Significance of the Study

Building on Wu and Browne (2015a) in CSMs, the proposed approach offers an alternative explanation for model errors originating from unknown sources within IRT models. The discrepancy between the specific operational population (i.e., the data-generating mechanism) and the

general population (i.e., the fitted model) is considered outside of the researchers' control and can be modeled as a random effect. This framework aims to demonstrate how inferences derived from observations, each obtained from a slightly different operational population, have implications for the general population of interest. A family of models are fitted to the sample data, and inferences made from the fitted model are about the parameters that describes the general population of interest.

The proposed framework allows researchers to quantify the amount of model error and make inference about the parameters in the general population simultaneously. In a fixed-effect model error approach, researchers need to complete a two-stage procedure, which requires them to first determine if the model fits the data reasonably well before obtaining parameter estimates to draw conclusions about the general population of interest. However, in the proposed framework, the model is allowed to be wrong to a certain extent, and the model error in the framework can be quantified by an additional component in the model. The magnitude of the model error and other model parameters can be jointly estimated by fitting the model to the data. It should be highlighted that our approach only considers inferences to be meaningful if there is a small discrepancy between the particular operational population and the general population.

Moreover, assuming the model error is random, the study examines the impact of this adventitious error under various conditions, such as conditions with different sample sizes, different test lengths, different number of categories contained in each item, and variable closeness of the operational population to the general population. In addition, the study also seeks to establish a connection between the concentration/dispersion parameter developed in the current framework and the model fit indices in the traditional fixed-effect IRT approaches.

1.4 Overview of the Chapters

Chapter 2 introduces the sources of uncertainty that researchers must take into account when drawing conclusions from statistical models and also explains why it is crucial to quantify the model errors. The fixed-effect and random-effect approaches to account for the model errors in CSMs and IRT models are then reviewed. A Dirichlet-Multinomial framework that treats the model error as a random effect in IRT models is developed in Chapter 3. The parameter estimation and asymptotic results for models when the population probability is assumed known or unknown are also discussed. Next, how the concentration/dispersion parameter in the proposed model framework relates to the RMSEA in the fixed-effect approaches is explained. Chapter 4 presents a simulation study that evaluates the performance of the Dirichlet-Multinomial framework in comparison to traditional multinomial submodels. The study focuses on aspects such as model convergence, point estimation, and interval estimation. In Chapter 5, empirical data analyses are conducted to apply the proposed framework to real datasets. In Chapter 6, we provide a summary of the Dirichlet-Multinomial framework developed in this study. Additionally, we discuss the results obtained from the simulation study and empirical data analyses, highlight the limitations of the current study, and explore potential promising directions for future research.

Chapter 2: Literature Review

2.1 Introduction to Model Error

Methods of uncertainty quantification (UQ) are developed to identify the many sources of uncertainty in model input variables and estimate their influence on model outcomes (Kennedy & O'Hagan, 2001; MacCallum & O'Hagan, 2015; Oakley & O'Hagan, 2002; O'Hagan et al., 1998). As according to MacCallum and O'Hagan (2015), in statistical models, the uncertainty in model input can be caused by "imperfect measures of inputs, imprecise values of parameters, errors in the code itself, etc.," and model outcomes usually refer to the results of "parameter estimates, measures of model fit, accuracy of predictions, generalizability, and cross-validity."

When making inferences based on statistical models, the major sources in UQ include but are not limited to sampling error and model error (MacCallum & O'Hagan, 2015). The sampling error refers to the discrepancy between the observed sample and the population from which the sample is drawn from, and can be quantified by confidence intervals (CIs) and test statistics. The model error is defined as the systematic error that describes the discrepancy between a family of models to be fitted and the data-generating mechanism. Model error is also known as model inadequacy, model discrepancy, and model misfit in previous literature (e.g., MacCallum & O'Hagan, 2015; Maydeu-Olivares, 2017; Maydeu-Olivares, Shi, & Rosseel, 2018; Wu & Browne, 2015a).

Making decisions based on statistical inferences from models that fail to account for model

errors might lead to severe consequences. Unlike the sampling error, the model error requires the knowledge of how it occurs between the hypothesized models and the data-generating mechanism. MacCallum and O'Hagan (2015) claimed that it is insufficient for researchers to only acknowledge that the used models are wrong to a certain degree and then proceed to explain the model results as if their models are correct in population. Neglecting model errors could possibly result in under-estimated uncertainty in model predictions and inaccurate explanations in parameter estimates (MacCallum & O'Hagan, 2015). Since models are only approximations to real-world phenomena, uncertainty in models is almost always inevitable.

The model error can be treated as fixed or random when being accounted for, and thus there are two approaches to characterize the impact of model errors — the fixed-effect approach and the random-effect approach. On the one hand, the fixed-effect approach assumes all samples are drawn from the same population, which differs from the model by a fixed amount. The model error in the fixed-effect approach is modeled as a fixed discrepancy between a family of fitted models and the data-generating mechanism. On the other hand, the model error can be considered random, with the assumption that each sample is collected from a specific population that randomly deviates from the fitted model by a small amount. In the random-effect approach, by expanding the existing model with additional concentration/dispersion parameters that describe the amount of model error, a full model can be fitted to data, and all parameters can be estimated simultaneously. If the model error is appropriately specified and included in the full model, the full model should have less corrupted parameter estimates, fit measures, predictions, and CIs than when the model error is ignored (MacCallum & O'Hagan, 2015; Wu & Browne, 2015a).

2.2 Model Error as a Fixed Effect

2.2.1 Effect Size Measures in Covariance Structure Models

After a model is fitted to data, it is necessary to assess the degree of the model error since inferences drawn from poorly fitting models can be misleading (Maydeu-Olivares et al., 2018). In general, the model error specifically refers the discrepancy between the data-generating process and the family of models being fitted. In CSMs, the model error can be assessed by the discrepancy between the true population covariance matrix and the fitted model.

CSMs model the covariance of the observed variables (Bentler, 1980). The model-implied covariance structure in a CSM is denoted $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\eta})$, in which $\boldsymbol{\eta}$ is a vector that contains model parameters. Figure 2.1 shows an example of a one-factor confirmatory factor analysis (CFA) model with three observed variables. Suppose the latent variable in the one-factor CFA is ξ_1 [1], the three observed variables are x_1, x_2, x_3, and the residuals in three observed variables are $\delta_1, \delta_2, \delta_3$. The loading of ξ_1 on x_3 is constrained to be 1, on x_1 is λ_{11}, on x_2 is λ_{21}. The variances of $\xi_1, \delta_1, \delta_2, \delta_3$ are $\phi_{11}, \omega_{11}, \omega_{22}, \omega_{33}$, respectively. The model-implied covariance matrix is

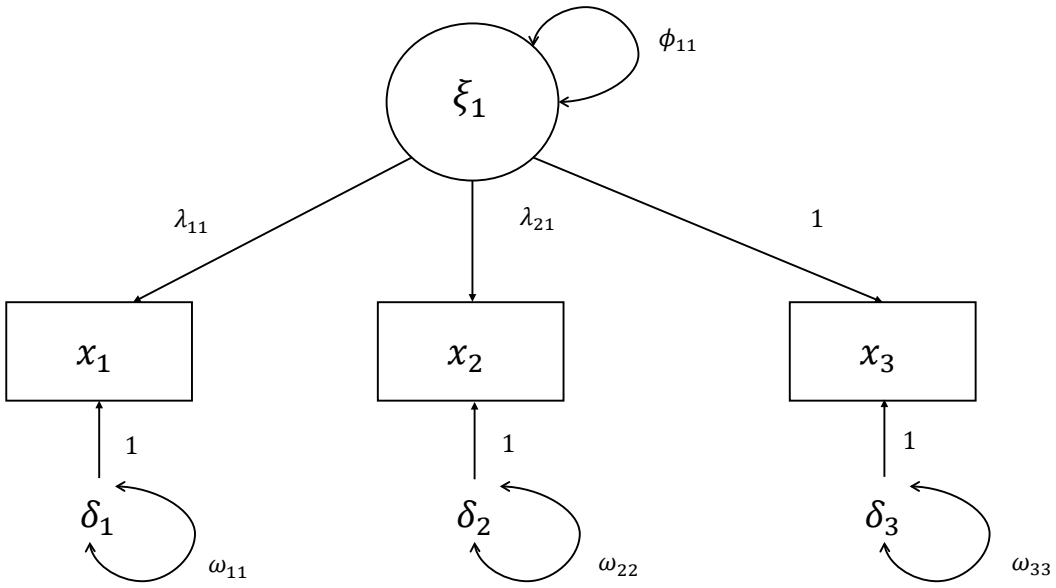
$$\boldsymbol{\Sigma}(\boldsymbol{\eta}) = \begin{bmatrix} \lambda_{11}^2\phi_{11} + \omega_{11} & & \\ \lambda_{11}\lambda_{21}\phi_{11} & \lambda_{21}^2\phi_{11} + \omega_{22} & \\ \lambda_{11}\phi_{11} & \lambda_{21}\phi_{11} & \phi_{11} + \omega_{33} \end{bmatrix}, \tag{2.1}$$

and the model parameters $\boldsymbol{\eta} = (\lambda_{11}, \lambda_{21}, \phi_{11}, \omega_{11}, \omega_{22}, \omega_{33})^T$.

[1] The exogenous latent factor is ξ, which should not be mistaken with the bold notation $\boldsymbol{\xi}$ for the vector of parameters used in this paper.

Figure 2.1

Path Diagram of a One-Factor Model With Three Observed Variables



The model fit can be evaluated by absolute and incremental fit effect size measures on the population level (Hu & Bentler, 1999). An absolute fit effect size measure evaluates how well a fitted model reproduces the data-generating model. The calculation of an absolute fit effect size measure may require a comparison to a saturated model that exactly reproduces the data-generating covariance matrix (Hu & Bentler, 1999). Popular absolute fit effect size measures include discrepancy functions associated with ML and Weighted Least Squares (WLS), Unweighted Least Squares (ULS), RMSEA, Squared Root Mean Residual (SRMR; Bentler, 1995), the expected log-penalty using Kullback-Leibler (KL) divergence (Kullback & Leibler, 1951), and the Goodness-of-Fit Index (GFI; Bollen & Stine, 1992). An incremental fit effect size evaluates the proportionate model fit improvement by comparing a fitted model with a more restricted nested baseline model. Examples of popular incremental fit effect size measures include CFI

(Bentler, 1990) and TLI (Tucker & Lewis, 1973).

2.2.1.1 Absolute Fit Measures

By fitting a CSM to data, the model parameters can be estimated by minimizing discrepancy functions with ML, WLS, ULS, and other estimation methods (Browne & Cudeck, 1992). Discrepancy functions associated with different types of estimators are absolute fit effect size measures: The smaller the discrepancies, the closer the model reproduces the data-generating model. Suppose that the discrepancy function F is the discrepancy between the true population covariance matrix $\boldsymbol{\Sigma}_0$ and the hypothesized model $\boldsymbol{\Sigma}(\boldsymbol{\eta})$. The Wishart ML (e.g., Lawley & Maxwell, 1962) discrepancy function is

$$F_{ML}(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}(\boldsymbol{\eta})) = \log|\boldsymbol{\Sigma}(\boldsymbol{\eta})| - \log|\boldsymbol{\Sigma}_0| + tr(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}(\boldsymbol{\eta})^{-1}) - p, \tag{2.2}$$

in which p denotes the number of variables being modeled. In addition, minimizing the discrepancy function F_{ML} is the same as minimizing the KL divergence (Arjovsky, Chintala, & Bottou, 2017).

KL divergence is another absolute fit measure that quantifies how one probability distribution is different from the probability distribution of the data-generating model (Kullback & Leibler, 1951). Suppose that there are two probability distributions g and f defined on the same probability space $\mathcal{X}$, the KL divergence quantifies the discrepancy between these two distribu-

tions by

$$D_{KL}(f,g) = \int_{x\in\mathcal{X}} \log\left(\frac{f(x)}{g(x)}\right) f(x)dx = \left[\int_{x\in\mathcal{X}} \log f(x)f(x)dx - \int_{x\in\mathcal{X}} \log g(x)f(x)dx\right]. \tag{2.3}$$

Typically f refers to the data-generating model with true parameter values, and g refers to a family of fitted models. It can be found that the first term on the right-hand side $\int_{x\in\mathcal{X}} \log f(x)f(x)dx$ does not change if the fitted model $g(x)$ changes because there is only one data-generating model. The expectation of the second term $E\int_{x\in\mathcal{X}} \log g(x)f(x)dx$ changes if the fitted model changes due to parameter optimization, and this expectation term is called the expected log-penalty. Because $\int_{x\in\mathcal{X}} \log f(x)f(x)dx$ does not change, minimizing the KL divergence is equivalent to minimizing the negative expected log-penalty, which is also the same as minimizing F_{ML} to find the ML estimate (MLE) of parameter $\boldsymbol{\eta}$. People also use this expected log-penalty as an absolute effect size measure when comparing statistical models.

Besides the ML discrepancy function, there are other discrepancy functions such as the WLS and the ULS discrepancy functions. The WLS discrepancy function is

$$F_{WLS}(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}(\boldsymbol{\eta})) = [vech(\boldsymbol{\Sigma}_0) - vech(\boldsymbol{\Sigma}(\boldsymbol{\eta}))]^T \boldsymbol{W}^{-1} [vech(\boldsymbol{\Sigma}_0) - vech(\boldsymbol{\Sigma}(\boldsymbol{\eta}))], \tag{2.4}$$

in which $vech(\boldsymbol{\Sigma}_0)$ denotes the vector of nonredundant elements (i.e., the lower triangular elements in the matrix plus all diagonal elements) in the data-generating covariance matrix $\boldsymbol{\Sigma}_0$, and $\boldsymbol{W}^{-1}$ is a $t \times t$ weight matrix, in which $t = p(p+1)/2$ is the amount of nonredundant information when the model has p variables (Henderson & Searle, 1979). The weight matrix $\boldsymbol{W}^{-1}$ is a consistent estimate of the covariance matrix of the elements of $vech(\boldsymbol{\Sigma}(\boldsymbol{\eta}))$ (Browne & Cudeck,

1992). Researchers often use the diagonal matrix $Diag(\boldsymbol{W}^{-1})$ in the WLS discrepancy function for computational simplicity and stability (Asparouhov & Muthen, 2007). ULS is a special case of WLS. When the identity matrix is chosen as the weight matrix in WLS, WLS reduces to ULS (Browne & Cudeck, 1992).

As an absolute fit effect size measure, RMSEA index on the population level is a measure of the model discrepancy per degree of freedom. The population RMSEA can be calculated by

$$\varepsilon = \sqrt{\frac{F_0}{df}}, \tag{2.5}$$

where ε is the RMSEA measure, df is the degree of freedom, F_0 is the minimum distance between the family of fitted models and data-generating model (Maydeu-Olivares et al., 2018). F_0 can be F_{ML}, F_{WLS}, F_{ULS} and other discrepancy functions. As an effect size measure, Browne and Cudeck (1992) recommended RMSEA $< .05$ as an indicator of close fit, $.05 - .08$ indicates fair fit, and $> .10$ indicates poor fit. Hu and Bentler (1999) suggests that an RMSEA cutoff value close to .06 indicates a good fit. It is also mentioned in previous literature that the RMSEA values have different meanings depending on the type and size of the CSM (Chen, Curran, Bollen, Kirby, & Paxton, 2008; Maydeu-Olivares et al., 2018).

SRMR is another widely used absolute fit effect size measure. SRMR is the standardized effect size of overall misfit for CSMs. The population SRMR can be computed by

$$SRMR = \sqrt{\frac{1}{t}\sum_{i \leq j} \frac{[(\boldsymbol{\Sigma}_0)_{ij} - (\boldsymbol{\Sigma}(\boldsymbol{\eta}))_{ij}]^2}{\sqrt{(\boldsymbol{\Sigma}_0)_{ii}(\boldsymbol{\Sigma}_0)_{jj}}}}, \tag{2.6}$$

in which $0 < i < p, 0 < i < p$, p is the number of variables being modeled, $t = p(p+1)/2$ is

the number of nonredundant entries in a population covariance matrix, $(\boldsymbol{\Sigma}_0)_{ij}$ is the element of variables i and j in the data-generating covariance matrix $\boldsymbol{\Sigma}_0$, and $(\boldsymbol{\Sigma}(\boldsymbol{\eta}))_{ij}$ denotes the element of variables i and j in the fitted model covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\eta})$. In general, SRMR$< .08$ indicates a good fit (Hu & Bentler, 1999).

The population GFI can be computed by

$$GFI = 1 - \frac{tr[(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1} - \boldsymbol{I}_p)^2]}{tr[(\boldsymbol{\Sigma}\boldsymbol{\Sigma}_0^{-1})^2]}, \tag{2.7}$$

in which $p = tr(\boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_0^{-1})$, $\boldsymbol{I}_p$ is the sum of the diagonal elements of a p by p identity matrix (Maiti & Mukherjee, 1990; Maydeu-Olivares et al., 2018). Equation 2.7 shows that the population GFI quantifies the discrepancy between the fitted model $\boldsymbol{\Sigma}$ and data-generating model $\boldsymbol{\Sigma}_0$. GFI is between 0 and 1 — the closer the GFI is to 1, the better the model fit. Hu and Bentler (1999) suggested that a cutoff value larger than .95 indicates good fit.

2.2.1.2 Incremental Fit Measures

Suppose the baseline model $\boldsymbol{\Sigma}_b$ refers to a more restricted model and is nested within the fitted model $\boldsymbol{\Sigma}(\boldsymbol{\eta})$. In comparison to a baseline model, the fitted model has a better model fit. The baseline model has the worst model fit when it is a null model. Popular incremental fit effect size measures include measures such as CFI and TLI. The population CFI is defined as relative effect sizes of overall misfit in Maydeu-Olivares et al. (2018). The formula to compute the population CFI is

$$CFI = 1 - \frac{F_f}{F_b}, \tag{2.8}$$

in which $F_f = F(\boldsymbol{\Sigma}(\boldsymbol{\eta}), \boldsymbol{\Sigma}_0)$ refers to the discrepancy between the fitted model and the data-generating model, and $F_b = F(\boldsymbol{\Sigma}_b, \boldsymbol{\Sigma}_0)$ is the discrepancy between the baseline model and the data-generating model (Bentler, 1990; Lai, 2019; Maydeu-Olivares et al., 2018). The discrepancy function F chosen here also depends on the normality of data and whether the asymptotically distribution free (ADF) is assumed. CFI ranges from 0 to 1, and similar to GFI, a cutoff value of .95 indicates a good model fit (Hu & Bentler, 1999). Another example for increment fit effect size measures is the TLI. TLI is non-normed and can be calculated as

$$TLI = \frac{\frac{F_b}{df_b} - \frac{F_f}{df_f}}{\frac{F_b}{df_b} - 1}, \tag{2.9}$$

in which df_b and df_f denote the degrees of freedom of the baseline model and fitted model, respectively (Bentler, 1990; Bentler & Bonett, 1980). TLI is non-normed and may not fall inside the $0 - 1$ range. A good fit model is often indicated by a cutoff value around .95 (Hu & Bentler, 1999).

2.2.2 Model Error as a Fixed Effect in IRT Models

Likert-type items with two or more categories can be calibrated using IRT models. Two-parameter logistic (2PL) model or Rasch model (Rasch, 1960) can be fitted to dichotomous items with two categories. Samejima's graded response model (GRM; Samejima, 1969) is a popular calibration method for polytomous items with more than two categories. GRM reduces to the 2PL model for dichotomous data (Maydeu-Olivares & Liu, 2015). Throughout the study, only the GRM family of models is considered.

When items are polytomous and each item is assumed to have C ordered response cate-

gories scored from 0 to $C-1$, the item response function (IRF) for item j, person i, and category c in a GRM model is

$$f_j(c|\theta) = Pr\{Y_{ij} = c|\theta_i = \theta\} = \frac{1}{1+\exp[-(a_j\theta + d_{j,c})]} - \frac{1}{1+\exp[-(a_j\theta + d_{j,c+1})]}. \quad (2.10)$$

θ_i denotes the latent trait for person i, $i = 1, 2, ..., n$. $\theta_i \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$. For item j, $j = 1, 2, ..., J$, a_j is the item slope parameter and $d_{j,c}$ is the item intercept for the cth category. The intercept parameters are in a strictly descending order; in particular, $d_{j,0} = \infty$ and $d_{j,C} = -\infty$. When items are dichotomous and responses are coded as 0 or 1, GRM is reduced to a 2PL model. In the 2PL model, the IRF is

$$f_j(c|\theta) = Pr\{Y_{i,j} = c|\theta_i = \theta\} = \frac{1}{1+\exp[(-1)^c(a_j\theta + d_{j,c})]}, \quad (2.11)$$

in which $c = 0, 1$ in the 2PL model.

Suppose there are J multinomial items in a test, the observations are from Y_1 to Y_J, each with C response categories, the total number of response patterns is $K = C^J$, and the probability in category c for item j can be calculated by IRFs in the aforementioned IRT models. Then, the probability vector $\boldsymbol{\pi}$ denotes the K-dimensional column vector of the response pattern probabilities in the population. We take into account a parametric model for the probability vector $\boldsymbol{\pi}$ that is dependent on an estimated q-dimensional vector of parameters $\boldsymbol{\xi}$, which contains item parameters estimated from the IRT models. This parametric model is denoted $\boldsymbol{\pi}(\boldsymbol{\xi})$ (Maydeu-Olivares & Joe, 2014).

After item calibration using the IRT models, the next question of interest is: to what de-

gree can the model approximate the unknown population probability. The solution can be found by quantifying the discrepancy between the fitted model and the true population probabilities (Browne & Cudeck, 1992; Maydeu-Olivares & Joe, 2014; Steiger, 1990). Maydeu-Olivares and Joe (2014) developed a family of RMSEA statistics in IRT models to quantify the discrepancy between the fitted model and the population probabilities when adjusting for model complexity. Under the assumption of exact fit, let $\boldsymbol{\pi}(\boldsymbol{\xi})$ be the fitted model and $\boldsymbol{\pi}_0$ be the data-generating model, and suppose $\boldsymbol{D}^{-1}$ is the diagonal matrix of fitted model probabilities $\boldsymbol{\pi}(\boldsymbol{\xi})$, a quadratic form of Pearson's X^2 can be computed by

$$D_n = (\boldsymbol{\pi}_0 - \boldsymbol{\pi}(\boldsymbol{\xi}))^T \boldsymbol{D}^{-1} (\boldsymbol{\pi}_0 - \boldsymbol{\pi}(\boldsymbol{\xi})), \tag{2.12}$$

in which D_n denotes the discrepancy function where up to n-way moments are used. The D_n is a discrepancy between a population probability vector $\boldsymbol{\pi}_0$ and the fitted model $\boldsymbol{\pi}(\boldsymbol{\xi})$ that is due to approximation (Maydeu-Olivares & Joe, 2014). In the likelihood ratio G^2 metric, the discrepancy can also be computed by

$$D_{n,G^2} = 2\boldsymbol{\pi}_0^T [\log(\boldsymbol{\pi}_0) - \log(\boldsymbol{\pi}(\boldsymbol{\xi}))], \tag{2.13}$$

in which $\boldsymbol{\pi}_0^T$ is the transpose of $\boldsymbol{\pi}_0$. The D_n and D_{n,G^2} in Equations 2.12 and 2.13 both utilize all information available in the data to test the model (Maydeu-Olivares & Joe, 2014). The full-information RMSEA$_n$ in IRT models uses D_n in the Pearson's X^2 metric and can be calculated as

$$\varepsilon_n = \sqrt{\frac{D_n}{df_n}}, \tag{2.14}$$

in which the df_n is the degrees of freedom of $K - q - 1$, K denotes the total number of response

categories in population probability vector $\boldsymbol{\pi}$, and q is the number of parameters to estimate in $\boldsymbol{\xi}$.

However, as is indicated in Maydeu-Olivares and Joe (2014), asymptotic methods only allow for good approximation of the sampling distribution of RMSEA$_n$ in small models when there is not much sparseness in contingency tables. For larger models with a lot of sparseness in contingency tables, Maydeu-Olivares and Joe (2005) developed classes of quadratic-form limited-information discrepancy measures that are based on multivariate moments or residuals of margins up to the order of r. The family of limited-information discrepancy function is

$$D_r = (\boldsymbol{\pi}_{0,r} - \boldsymbol{\pi}(\boldsymbol{\xi})_r)^T \boldsymbol{C}_r (\boldsymbol{\pi}_{0,r} - \boldsymbol{\pi}(\boldsymbol{\xi})_r), \tag{2.15}$$

in which $\boldsymbol{\pi}_{0,r}$ is the population probability using moments up to the order r, $\boldsymbol{\pi}(\boldsymbol{\xi})_r$ is the model-implied probability using moments up to the order r, $\boldsymbol{C}_r$ is the weight matrix for the fitted model and can be calculated using Equation (3) in Maydeu-Olivares and Joe (2014). D_2 uses univariate and bivariate moments and is recommended by Maydeu-Olivares and Joe (2005) — they demonstrated that as r increases, the accuracy of the asymptotic approximation to the sampling distribution of the statistic M_r decreases. M_r refers to the family of test statistics that uses residual moments, when $r = n$, M_r is the same as Pearson's X^2 (Maydeu-Olivares & Joe, 2014). The limited-information RMSEA$_2$ that uses univariate and bivariate moments can be calculated as

$$\varepsilon_2 = \sqrt{\frac{D_2}{df_2}}, \tag{2.16}$$

where df_2 is the available degrees of freedom when univariate and bivariate moments are used. $df_2 = s_2 - q$, in which s_2 denotes the total number of univariate and bivariate moments. $s_2 =$

$p(C-1)+\frac{p(p-1)}{2}(C-1)^2$, in which C is the number of categories in each item, and p is the total number of items.

Maydeu-Olivares and Joe (2014) also proposed the distance metric D_{ord} for ordinal response data. In this case, the goodness-of-fit (GOF) of the model can be evaluated by RMSEA$_{ord}$. RMSEA$_{ord}$ reduces to RMSEA$_2$ when the data is dichotomous. The quadratic form of D_{ord} is

$$D_{ord} = (\boldsymbol{\kappa}_0 - \boldsymbol{\kappa}(\boldsymbol{\xi}))^T \boldsymbol{C}_{ord} (\boldsymbol{\kappa}_0 - \boldsymbol{\kappa}(\boldsymbol{\xi})), \tag{2.17}$$

in which the $\boldsymbol{\kappa}_0$ and $\boldsymbol{\kappa}(\boldsymbol{\xi})$ are the vectors of means and cross-products of the true population probabilities and the fitted model, respectively, and $\boldsymbol{C}_{ord}$ is a weight matrix for the fitted model. The formulas to calculate $\boldsymbol{\kappa}$ and $\boldsymbol{C}_{ord}$ can be found in Maydeu-Olivares and Joe (2014) Equations (27) and (28), respectively. The corresponding RMSEA$_{ord}$ is computed as

$$\varepsilon_{ord} = \sqrt{\frac{D_{ord}}{df_{ord}}}, \tag{2.18}$$

in which $df_{ord} = p(p+1)/2 - q$.

In addition to RMSEA$_{ord}$, another effect size measure to evaluate the GOF of a large model with ordinal data is the Standardized Root Mean Squared Residual (SRMSR), which can be calculated by

$$SRMSR = \sqrt{\sum_{i<j} \frac{(\rho_{0,ij} - \rho(\boldsymbol{\xi})_{ij})^2}{p(p-1)/2}}, \tag{2.19}$$

in which the $\rho_{0,ij}$ and $\rho(\boldsymbol{\xi})_{ij}$ are the expected correlation for a pair of items i and j under the true population probabilities and the fitted model, respectively. Dividing the expected covariance by the expected standard deviations yields the expected correlation (Maydeu-Olivares & Joe, 2014).

2.2.3 Statistical Inferences for Model Discrepancy Measures

Model fit indices are the estimates of population effect size measures. On the one hand, in evaluating model misfit, effect size measures are the population parameters that can capture the discrepancy between the fitted models and the data-generating mechanism. On the other hand, model fit indices are sample estimates of effect sizes, without considering their sampling variability or referring to their corresponding population parameters (Maydeu-Olivares, 2017; Maydeu-Olivares et al., 2018). In most cases, there are direct connections between model fit indices and their effect size measures.

Akaike information criterion (AIC) is a penalized-likelihood information criteria index that uses log-likelihood functions with simple penalty terms. AIC is an estimated log-penalty evaluated at the MLEs when predicting from the current sample to a new sample that is drawn from the same population. The derivation of AIC is related to the effect size measure of the expected log-penalty. Suppose that there are two samples with equal sample size, a model is fitted to Sample 1 by ML, and the model fit is measured by the deviance G, which is -2 times the maximized log-likelihood. The deviance G^* can be calculated on Sample 2 with the MLEs obtained from Sample 1. Suppose G' denotes the deviance evaluated at the true parameter values, by Taylor-series expansions, the differences between the expected deviances EG^* and EG', and EG' and EG are both approximately equal to the number of estimated parameters q. Thus, the difference between EG^* and EG is approximately $2q$ (Akaike, 1998; Ripley, 2003). AIC is an unbiased estimator of EG^*, and can be calculated by

$$AIC = G + 2q. \tag{2.20}$$

The calculation of AIC is the same in both CSMs and IRT models (Kang & Cohen, 2007).

The likelihood ratio (LR) test statistic is a test statistic for model misfit and is related to effect size measures. The LR test statistic can be used to assess the discrepancy between the sample and fitted covariance matrices (Hu & Bentler, 1999). The widely used LR test statistic in CSM is derived from ML under the assumption of multivariate normality of variables (Hu & Bentler, 1998). Suppose that the discrepancy function used is a ML discrepancy as described in Equation 2.2, the estimation of the discrepancy is between a sample covariance matrix $\boldsymbol{S}$ and the model-implied covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\eta})$. An estimator $\hat{\boldsymbol{\eta}}$ can be obtained by minimizing this discrepancy function, and $\hat{F}_{ML}$ is the minimum discrepancy function computed by plugging in the estimator $\hat{\boldsymbol{\eta}}$. Under the normality assumptions, the LR test statistic can be calculated by

$$\hat{T}_{LR} = (n-1)\hat{F}_{ML}, \tag{2.21}$$

where n denotes the sample size. When the fitted model is correct, $\hat{T}_{LR}$ can be approximated asymptotically with a chi-square distribution. When the model is slightly incorrect, under Pitman drift assumption, the $\hat{T}_{LR}$ statistic asymptotically follows a noncentral chi-square distribution (Yuan, Hayashi, & Bentler, 2007).

In addition, two most commonly used statistics to assess the overall fit of the model are the Pearson's X^2 and the likelihood ratio statistic G^2. The Pearson's X^2 can be calculated by

$$X^2 = n\sum_{k=1}^{K}\frac{(p_k - \pi_k)^2}{\pi_k}, \tag{2.22}$$

and the likelihood ratio statistic G^2 is

$$G^2 = 2n \sum_{k=1}^{K} p_k \log \frac{p_k}{\pi_k}, \tag{2.23}$$

in which p_k is the observed probability and π_k is the expected probability according to the model for response pattern k. Equation 2.23 shows that the G^2 is calculated based on expected log penalty introduced in Section 2.2.1.1. If item parameters are estimated using ML method and the fitted model hold exactly in the population, X^2 and G^2 are approximately the same and can be approximated by a chi-square distribution with degrees of freedom of $K - q - 1$, in which K denotes the total number of response categories, and q is the number of parameters to estimate. X^2 and G^2 are both full information statistics since they test the model using all of the information in the data (Maydeu-Olivares & Joe, 2008, 2014; Maydeu-Olivares & Liu, 2015).

Researchers need to decide whether the approximation provided by the fitted model is good enough. People usually assess this GOF by testing whether the discrepancy between the population covariance matrix and the fitted model is less than or equal to some arbitrary value. The hypothesis testing for GOF can be either for exact fit or approximate fit (i.e., close fit). Point estimates and their corresponding CIs can be obtained from these hypothesis tests for statistical inferences.

Testing the exact model fit means assessing whether the population covariance matrix exactly matches the parametric model against the alternative that the model is incorrect (Li & Bentler, 2006). Taking the population RMSEA as an example. A typical exact fit test usually has a null hypothesis such as

$$H_0 : \varepsilon = 0, \tag{2.24}$$

and the alternative hypothesis is that population RMSEA is not 0. The null hypothesis will mostly be rejected if the sample size is sufficiently large. The exact fit tests tend to be over-powered and is likely to flag non-substantial model misfit as sample size goes up, even though the misfit is actually acceptable. The exact fit tests also encourage adding uninterpretable additional parameters to reduce the power of the test to avoid the rejection of the null hypothesis (Browne & Cudeck, 1992).

The likelihood ratio statistic G^2 and Pearson's X^2 statistic can be used to test the exact fit

$$H_0 : \boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\xi}), \qquad H_1 : \boldsymbol{\pi} \neq \boldsymbol{\pi}(\boldsymbol{\xi}), \tag{2.25}$$

in which $\boldsymbol{\xi}$ is the vector of model parameters to estimate. However, except for small models, the p-values obtained from these G^2 and X^2 statistics can be very inaccurate. In addition, in CSM, the p-values may also be inaccurate regardless of the sample size (Maydeu-Olivares & Joe, 2014).

Because of the disadvantages of exact fit tests, researchers are interested in testing the approximate model fit in practice. The approximate model fit tests whether the null model is a good enough approximation to the population covariance matrix. Browne and Cudeck (1992) used a less implausible interval hypothesis

$$H_0 : \varepsilon \leq 0.05 \tag{2.26}$$

to assess the approximate fit of the model. The choice of the upper or lower limit of the approximate test is based on researcher's experience with the estimates of the effect size measures.

The point estimates of population values of the effect size measures may vary from one

sample to another, and thus CIs are needed because interval estimates lose some precision in the estimate as compared to the point estimates, but gain confidence that the claim is correct (Casella & Berger, 2021). For example, a point estimate of the population RMSEA is

$$\hat{\varepsilon} = \sqrt{\frac{\hat{F}_0}{df}}. \tag{2.27}$$

F_0 is not directly estimable but can be estimated through the sample discrepancy function $\hat{F} = F(\boldsymbol{S}, \hat{\boldsymbol{\Sigma}})$, in which $\boldsymbol{S}$ is the sample covariance matrix, and $\hat{\boldsymbol{\Sigma}}$ is the minimizer of the discrepancy function. As aforementioned in the previous sections, under the Pitman drift assumption, $(n-1)\hat{F}$ follows a noncentral chi-square distribution (Yuan et al., 2007). The expected $\hat{F}$ can be computed by

$$E\hat{F} \approx F_0 + (n-1)^{-1}df. \tag{2.28}$$

$\hat{F}$ is a biased estimator of F_0, and after correcting this bias, a point estimator of F_0 can be determined by

$$\hat{F}_0 = \max\{\hat{F} - (n-1)^{-1}df, 0\}. \tag{2.29}$$

In real data scenarios, we can estimate the full-information RMSEA$_n$ using Pearson's X^2 statistic and the corresponding df as outlined in Maydeu-Olivares and Joe (2014). The estimated $\hat{\varepsilon}_n$ was calculated using the formula:

$$\hat{\varepsilon}_n = \sqrt{\text{Max}\left(\frac{\hat{X}^2 - df}{n \times df}, 0\right)}. \tag{2.30}$$

A 90% CI for the estimated $\hat{\varepsilon}_n$ is

$$(\hat{\varepsilon}_L, \hat{\varepsilon}_U) = \left(\sqrt{\frac{\hat{\lambda}_L}{n \times df}}, \sqrt{\frac{\hat{\lambda}_U}{n \times df}} \right), \tag{2.31}$$

in which $\hat{\varepsilon}_L$ and $\hat{\varepsilon}_U$ are the lower and upper limit of the CI for the RMSEA. Suppose $G(x|\lambda, df)$ is the cumulative distribution function (CDF) of the noncentral chi-square distribution with a noncentrality parameter λ and df degrees of freedom, $\hat{\lambda}_L$ and $\hat{\lambda}_U$ are solutions for λ in $G(x|\lambda, df) = 0.95$ and $G(x|\lambda, df) = 0.05$, respectively. It should be noted that if $G(x|0, df) < 0.95$ for the lower bound, $\hat{\lambda}_L$ is set to 0. Similarly, if $G(x|0, df) < 0.05$ for the upper bound, $\hat{\lambda}_U$ is set to 0 (Browne & Cudeck, 1992).

It is noticeable that when using LR test statistic to measure the model misfit, the noncentral chi-square approximation only holds under the Pitman drift assumption (Yuan et al., 2007). The Pitman drift assumption is also called the local alternatives. The LR test statistic asymptotically follows a noncentral chi-square distribution under a sequence of local alternative hypotheses. When under a fixed alternative, the noncentral chi-square distribution does not hold. The Pitman drift assumes that the true population covariance matrix becomes closer to the model as the sample size becomes larger, however, this is not plausible since the population covariance matrix should not be impacted by the sample size (Wu & Browne, 2015a). Nevertheless, Yuan et al. (2007) indicated that if the effect size is small, the noncentral chi-square distribution under the Pitman drift appropriately describes the asymptotic behavior of the LR test statistic.

As an alternative to the noncentral chi-square approximation, a normal approximation developed by Yuan et al. (2007) does not rely on the Pitman drift assumption. Under some regularity conditions and a fixed alternative hypothesis, the normal approximation assumes that

$\sqrt{n}(\frac{T_{LR}}{n} - M)$ converges to a normal distribution, in which M is a quantity that can be computed in Equation (15) in Yuan et al. (2007). They found that 1) when data are normally distributed and the sample size is large, the normal distribution describes the behavior of T_{LR} better than the noncentral chi-square distribution when RMSEA is greater than 0.05 or 0.07 or the effect size of the latent variable is above 0.03 in a one-factor model, 2) when data are nonnormally distributed, the normal distribution is also preferred when the sample size is not very small, and 3) when data are nonnormally distributed and the sample size is small, none of the procedure works well.

2.2.4 Limitations of Current Effect Size Measures

The effect size measures to quantify the fixed model error is widely used in CSMs, but there are some limitations within these measures. First of all, the cutoff values for some measures are arbitrary. For example, Browne and Cudeck (1992) and Steiger (1990) recommended using .05 as the close fit cutoff value for RMSEA, and RMSEA larger than .10 as an indicator for poor fit. MacCallum, Browne, and Sugawara (1996) considered RMSEA values between .08 and .10 as mediocre fit. It is hard to judge whether mediocre fit is acceptable or not, and people become lenient with the cutoffs values when selecting their preferred models. Moreover, Hu and Bentler (1999) suggested a cutoff value greater than .95 for ML-based CFI indicate good fit, while others considered .90 to be good enough (Hooper, Coughlan, & Mullen, 2008).

Secondly, the cutoff values for a single model fit index may not work equally well under different conditions of sample sizes, estimators, or distributions (Hu & Bentler, 1997). For instance, ML-based SRMR is most sensitive to models with misspecified factor covariances or latent structures, while RMSEA is more sensitive to models with misspecified factor loadings.

Thus, it is suggested that various combinations of cutoff values for multiple indices should be used to evaluate the model (Hu & Bentler, 1999).

Thirdly, some model fit indices are not correctly implemented in software programs, and researchers need to be very careful when calculating these fit indices using these software programs. The sample SRMR values calculated in CSM software are biased estimate of population SRMR in finite samples (Maydeu-Olivares et al., 2018), and thus it may lead to poor model fit interpretations. Better model fit measures for fixed-effect model errors are needed in future research.

2.3 Model Error as a Random Effect

2.3.1 The Framework of Calibration Under Uncertainty

Another way to incorporate the effect of model error is to model the model error (Cudeck & Henly, 1991; MacCallum & O'Hagan, 2015; Wu & Browne, 2015a). By expanding the existing model with model discrepancy and additional parameters, a full model can be fit to data, and all parameters can be estimated simultaneously. The approach of modeling the model error is referred to as the Calibration Under Uncertainty (CUU) in UQ. In statistical modeling under the CUU framework, the occurrence of a model discrepancy is seen as affecting the model estimation (MacCallum & O'Hagan, 2015). The approach in Wu and Browne (2015a) to characterize the model error as a random effect is a method for CUU in CSMs (MacCallum & O'Hagan, 2015). In Wu and Browne (2015a), the random-effect model error is referred to as the adventitious error, which is defined as the model error that emerges by chance from unidentified sources outside of the researchers' control of the study.

The model error is a result of the fact that there is a discrepancy between all observations obtained in one study and the hypothesized model by theory, and this deviation can be random from study to study under different conditions even if there is only one population of interest. The presence of this type of error is eventually due to the existence of differences in "two populations." In a CSM, suppose the general population of interest has the covariance matrix $\boldsymbol{\Omega}$, and the observed samples are drawn from an operation population with the covariance matrix $\boldsymbol{\Sigma}$. Researchers are interested in establishing a theory for the general population under a general measurement condition, but observations are usually from a specific group of people under a specific condition. The deviation between sample covariance matrix $\boldsymbol{S}$ and the operational population covariance $\boldsymbol{\Sigma}$ can be quantified by sampling error. The sampling error can be reduced by methods such as increasing sample size, a fully Bayesian approach or a multiple imputation approach (Yang, Hansen, & Cai, 2012). Researchers also wonder how to model the model discrepancy, which is specifically the deviation between $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ in CSMs. This is a deviation that cannot be reduced with a larger sample size.

The traditional approach does not model the deviation between $\boldsymbol{\Sigma}$ and $\boldsymbol{\Omega}$ but treat it as a fixed discrepancy (see Section 2.2). Nevertheless, Wu and Browne (2015a) claimed that it is meaningful to think of each $\boldsymbol{\Sigma}$ as a realization of the theoretical population but with slightly different operationalization. The substantive model $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\boldsymbol{\eta})$ is assumed to hold exactly in the general population, and $\boldsymbol{\Sigma}$ follows a distribution centered at the model $\boldsymbol{\Omega}(\boldsymbol{\eta})$ (MacCallum, Browne, & Cai, 2006; Wu & Browne, 2015a). Wu and Browne's approach falls within the category of CUU, in which people have developed different methods to quantify the random model errors. For example, Gaussian processes rather than constant bias terms were used to model the model errors in Kennedy and O'Hagan (2001) and Trucano, Swiler, Igusa, Oberkampf, and Pilch

(2006). The discrepancy between fitted models and the data-generating model can be modeled as Gaussian processes with simple parameterization in a Bayesian approach.

2.3.2 Adventitious Error Framework in Covariance Structure Models

Statistical formulas are developed to model the adventitious error in a random-effect approach in CSMs. In the statistical modeling in Wu and Browne (2015a), three distributions are specified: the distribution for the sampling covariance matrix $(\boldsymbol{S}|\boldsymbol{\Sigma})$, the distribution for the population covariance matrix $(\boldsymbol{\Sigma}|\boldsymbol{\Omega})$, and the marginal distribution of the sample covariance matrix $(\boldsymbol{S}|\boldsymbol{\Omega})$ where $\boldsymbol{\Sigma}$ is integrated out. The distribution of the sampling error $\boldsymbol{S} - \boldsymbol{\Sigma}$ can be specified by $(\boldsymbol{S}|\boldsymbol{\Sigma})$ under the normality assumption, and the distribution of the model error $\boldsymbol{\Sigma} - \boldsymbol{\Omega}(\boldsymbol{\xi})$ can be specified by $(\boldsymbol{\Sigma}|\boldsymbol{\Omega})$. Because the true population covariance matrix $\boldsymbol{\Sigma}$ is usually unknown, the parameters can be estimated by maximizing the likelihood of the marginal distribution $(\boldsymbol{S}|\boldsymbol{\Omega})$ by integrating out the $\boldsymbol{\Sigma}$. The distributions for these quantities can be assumed in a realistic form based on researchers' experience.

The adventitious error approach in Wu and Browne (2015a) used conjugate distributions for mathematical convenience. They assumed the distribution of $(\boldsymbol{S}|\boldsymbol{\Sigma})$ follows a Wishart distribution, and the distribution of $(\boldsymbol{\Sigma}|\boldsymbol{\Omega})$ follows an inverted Wishart distribution. The marginal distribution $(\boldsymbol{S}|\boldsymbol{\Omega})$, when $\boldsymbol{\Sigma}$ is integrated out, follows a Type II matrix-variate Beta distribution. Within this model specification, the model parameters $\boldsymbol{\eta}$ and a concentration parameter m can be estimated by maximizing the likelihood (or minimizing the discrepancy function) of the Type II matrix-variate Beta distribution. The concentration parameter m gives a measure of the extent of model misfit. When the sample size goes to infinity, $\boldsymbol{S}$ converges to $\boldsymbol{\Sigma}$ and no sampling error is

present. Similarly, as m goes to infinity, $\boldsymbol{\Sigma}$ converges to $\boldsymbol{\Omega}$ and no model error is present.

The estimate of the concentration parameter m can be biased in the Type II matrix-variate Beta distribution estimation method, and thus a bias correction procedure is needed to improve the performance of the adventitious error approach. Suppose that the true population covariance matrix $\boldsymbol{\Sigma}$ is known, it is possible to estimate the model parameters $\boldsymbol{\eta}$ and the concentration parameter m by minimizing the discrepancy function in the inverted Wishart distribution. It was found that in this estimation procedure, m was over-estimated. Wu and Browne (2015a) added a tuning parameter in the modified discrepancy function in the inverted Wishart distribution and obtained asymptotically unbiased estimate of the concentration parameter. Furthermore, the same modification can be applied to the discrepancy function in MBL to correct the bias in the concentration parameter.

2.3.3 Connection to Root Mean Square Error of Approximation

In Wu and Browne (2015a) approach, the concentration parameter was found to have an asymptotic connection with the population RMSEA. Based on the modified estimating equation in the inverted Wishart distribution, it was found that the estimate of the inverse of the concentration parameter $\hat{\nu}^{IW}$ (i.e., dispersion parameter) can be computed by $\frac{\hat{F}^{IW}}{df}$ plus some small error term. Furthermore, because the population RMSEA can be estimated by the square root of the discrepancy function per degree of freedom (see Equation 2.5), and the inverted Wishart distribution and Wishart distribution discrepancy functions are asymptotically equivalent in the form

$$F^{IW}(\boldsymbol{\Sigma}, \boldsymbol{\Omega}) = F^{W}(\boldsymbol{\Omega}, \boldsymbol{\Sigma}) = F^{W}(\boldsymbol{\Sigma}^{-1}, \boldsymbol{\Omega}^{-1}), \tag{2.32}$$

the $\hat{\nu}^{IW}$ was found to be asymptotically equal to the squared RMSEA plus some small error term.

There are three important implications in the finding of the connection between the dispersion parameter ν and the fixed-effect model error measure RMSEA. First, it links the metrics of the model error in the fixed-effect approach and the random-effect approach. Since the fixed-effect method has established RMSEA cutoff values for good fit, the assessment of the adventitious error may be based on comparable standards. The criteria for an acceptable level of model error were provided by Wu and Browne (2015a). The theoretical covariance structure will be rejected if the model error is too large (i.e., above this criteria), which suggests that the current operational population is insufficient to accurately reflect the theoretically hypothesized model. Secondly, the connection can be applied to point estimates, but in the random-effect method, the associated CIs for the square root of ν are wider than the CIs for the RMSEA in the fixed-effect approach. This is due to the addition of the concentration/dispersion parameter in the random-effect approach that captures the extra source of randomness. Thirdly, it is easier to retain a close fit hypothesis since the CIs for the square root of ν are wider. The 90% CI on the square root of ν can be used to determine whether to reject or keep the model based on the criterion that the lower bound of 90% CI less than 0.05 suggests a small amount of model error and the model can be retained in this case (Wu & Browne, 2015b).

2.3.4 Limitations of Random-Effect Approach

There is a controversy in the stochastic nature of the random-effect model error approach in CSMs. Several discussion papers on Wu and Browne (2015a) pointed out that the major concern is in defining the "ideal" general population. Steyer, Sengewald, and Hahn (2015) hold the opin-

ion that there is no such "general population," and statistical inferences should not be made on the abstract "general group of people" under a "general measurement condition." They believed that the model error stated in Wu and Browne's method is essentially a systematic error caused by the biased and nonrandom sample, or, in other words, "a specific group of people," after taking into account the heterogeneity of subpopulations and various measurement conditions. Thus, instead of using Wu and Browne's approach, Steyer et al. (2015) recommended concentrating on understanding the target population, enhancing the sampling design, or adequately weighting the subsamples.

In addition, Shapiro (2015) challenged Wu and Browne (2015a) in the changes of the general population. In Wu and Browne (2015a), the example of a general population is the "adults in the U.S.," and a corresponding operational population is the "adults from Chicago in summer." Shapiro (2015) indicated that the existence of "adults in the U.S." is questionable because this population may change in time. To clarify the issues mentioned above, Wu and Browne (2015b) commented using the opinion in MacCallum and O'Hagan (2015) that the general population does not need an empirical description. In Wu and Browne's approach, the model rather than abstract empirical nature defines the general population. In addition, MacCallum and O'Hagan (2015) considered that the two population framework should be extended. They considered that the discrepancy functions in Wu and Browne's approach should include both the deviation between the operational population and the general population as well as the general population's deviation from the ideal population in which the model is hypothesized because if the model only holds in an abstract ideal population, it may not hold in a real-world general population.

Moreover, one of the concerns mentioned in both Satorra (2015) and Shapiro (2015) is that Wu and Browne's approach only used one concentration/dispersion parameter to quantify

the random-effect model error. It was noted in Shapiro (2015) that the concentration/dispersion parameter estimate in Wu and Browne's technique alone determines whether the model should be accepted or rejected, which makes it challenging to assess the stability and interpretability of the parameter. The method should thus be "viewed as a guidance rather than a final judgement" because of this. In a similar vein, Satorra (2015) emphasized that models should not be evaluated just on the basis of one particular number. Satorra (2015) noted that fixed-effect model error techniques have advantages over adventitious error approaches since many effect size measures, modification indices, and parameter change indicators have been developed for these traditional fixed-effect approaches. These multiple fixed-effect measurements aid in cross-validating or improving the model. More indices can be created using Wu and Browne's methodology, and more linkages between these indices and the fixed-effect effect size measures may be discovered, according to a suggestion in Satorra (2015). Wu and Browne (2015b) responded to these recommendations by pointing out that most indices in the current fixed-effect model error methods can be built upon their adventitious error framework — this is one of the future paths down the line.

Chapter 3: Theory and Method

In this chapter, we derive a Dirichlet-Multinomial framework to characterize the model errors as random effects under a general family of multinomial submodels, which includes the family of GRMs characterized by Equations 2.10 and 2.11 as special cases. In this framework, the magnitude of adventitious error is quantified by an additional concentration/dispersion parameter. This parameter can be associated with RMSEA, a commonly used index for evaluating the model discrepancy per degree of freedom. We also discuss how to correct the bias in the estimated parameters in the framework.

3.1 Three Quantities in the Framework

The framework incorporating the random effect in multinomial submodels requires specifying probability/proportion vectors at three levels: the general population of interest, the operational population, and the sample. Suppose the total number of response pattern categories is K. A general population of interest refers to the population to which the theory applies. Let $\boldsymbol{\xi}$ denote the submodel parameters, and the model-implied probability vector for the response pattern categories be

$$\boldsymbol{\tau}(\boldsymbol{\xi}) = (\tau_1(\boldsymbol{\xi}), ..., \tau_i(\boldsymbol{\xi}), ..., \tau_K(\boldsymbol{\xi}))^T.$$

For the response pattern category i, $\tau_i(\boldsymbol{\xi})$ is determined by the item parameters $\boldsymbol{\xi}$. Assume that the dimension of $\boldsymbol{\xi}$ is q. The term "multinomial submodel" is used to describe models in which q is smaller than K. This includes various models such as IRT models, which are the primary focus of the current thesis, and log-linear models. The data-generating probability vector $\boldsymbol{\pi}$ in the operational population represents the population vector from which the observations are collected, and is denoted

$$\boldsymbol{\pi} = (\pi_1, ..., \pi_i, ..., \pi_K)^T.$$

$\boldsymbol{\pi}$ is random and centered at $\boldsymbol{\tau}$. Sample data are drawn from the operational population and are summarized by the observed counts as

$$\boldsymbol{x} = (x_1, ..., x_i, ..., x_K)^T.$$

$\boldsymbol{x}$ follows a multinomial distribution, with sample size n and probabilities $\boldsymbol{\pi}$. $\boldsymbol{x}$ can be converted into proportions in response pattern categories: i.e., $\boldsymbol{x} = n\boldsymbol{p}$, in which $\boldsymbol{p}$ denotes the vector of sample proportions

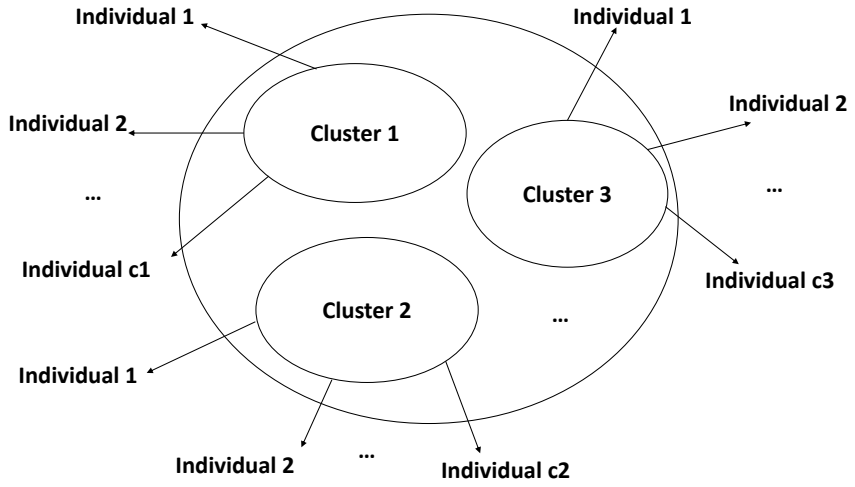
$$\boldsymbol{p} = (p_1, ..., p_i, ..., p_K)^T.$$

The connection between the aforementioned three quantities can be explained using an analogy to a multilevel model as illustrated in Figure 3.1. The analogy can be described as follows: In a multilevel model, we choose a cluster from a population of clusters, which is similar to our general population of interest. The selected cluster in the multilevel model can be regarded as the operational population, and we obtain data by sampling individuals from that cluster, similar to how we sample individuals from the operational population. As illustrated in Figure 3.1,

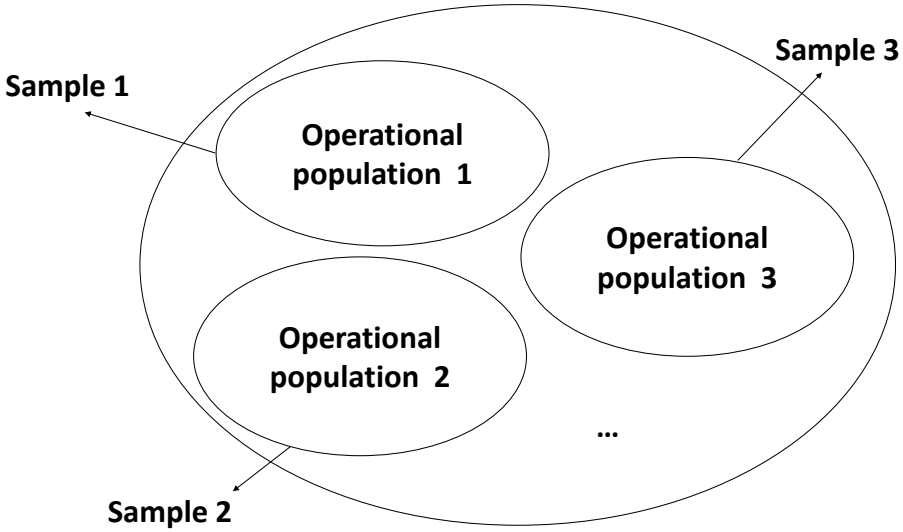
despite the operational population in our framework being comparable to the clusters in a multilevel model, it should be noted our approach only draws a single sample from each operational population. In multilevel models, however, we typically sample multiple clusters. Attempting to estimate the dispersion/concentration parameter, which characterizes the variability of the operational population around the general population, poses challenges when relying solely on data from one single operational population.

Figure 3.1

Comparing the Multilevel Model and the Adventitious Error Approach



(a) Population of clusters in a multilevel model



(b) General population of interest in the adventitious error approach.

Note. In (a), there are $c1, c2, c3$ individuals in Clusters $1, 2, 3$, respectively.

3.1.1 Statistical Formulation

Now we introduce the statistical formulations of modeling the adventitious error in multinomial submodels. The sample observed counts $\boldsymbol{x}$ conditional on data-generating probability vector $\boldsymbol{\pi}$ follows a multinomial distribution (e.g., Agresti, 2003). The multinomial probability mass function (pmf) is

$$f(\boldsymbol{x}|\boldsymbol{\pi}, n) = \left[\frac{n!}{(x_1)!\cdots(x_K)!}\right] \pi_1^{x_1} \cdots \pi_K^{x_K}. \tag{3.1}$$

In this case, $\boldsymbol{p}$ is the sample proportion vector calculated as the sample observed counts $\boldsymbol{x}$ divided by the sample size n. $\boldsymbol{p}$ is an unbiased estimator of the data-generating probability vector $\boldsymbol{\pi}$.

Moreover, $\boldsymbol{\pi}$ conditional on the model-implied proportion $\boldsymbol{\tau}$ is assumed to follow a Dirichlet distribution with a K-dimensional vector $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_K)^T$ (e.g., Bickel & Doksum, 2015). The probability density function (pdf) of a Dirichlet distribution is:

$$f(\boldsymbol{\pi}|\boldsymbol{\tau}, \alpha) = \frac{\Gamma(\sum_{i=1}^K \alpha_i)}{\prod_{i=1}^K \Gamma(\alpha_i)} \prod_{i=1}^K \pi_i^{\alpha_i - 1}, \tag{3.2}$$

in which $\boldsymbol{\alpha}$ is the concentration parameter vector. For simplicity, we further assume that the concentration parameters are proportional to the general population probabilities. That is $\boldsymbol{\alpha} = \alpha\boldsymbol{\tau} = \alpha(\tau_1, \tau_2, \cdots, \tau_K)^T$, $\alpha\tau_i > 0, 1 < i < K, 0 < \pi_i < 1$, and $\sum_{i=1}^K \pi_i = 1$. By the definition of the Dirichlet distribution, the mean of $\boldsymbol{\pi}$ is $\boldsymbol{\tau}$. In the sequel, the scalar parameter α is referred to as the concentration parameter, and its reciprocal, denoted v, is referred as the dispersion parameter. The concentration parameter α quantifies the extent to which the operational population is to the

general population of interest.

When the data-generating probability $\boldsymbol{\pi}$ that characterizes the operational population is unknown and random, model parameters can be estimated by maximizing the likelihood function of the marginal distribution of the sample response patterns $(\boldsymbol{p}|\boldsymbol{\tau}, \alpha, n)$ after integrating out $\boldsymbol{\pi}$. The marginal distribution of $(\boldsymbol{p}|\boldsymbol{\tau}, \alpha, n)$ follows a Dirichlet-Multinomial distribution (e.g., Minka, 2000) with pmf

$$f(\boldsymbol{x}|\boldsymbol{\tau}, \alpha, n) = \frac{\Gamma(n+1)\Gamma(\sum_{i=1}^{K}\alpha_i)}{\Gamma(n+\sum_{i=1}^{K}\alpha_i)}\prod_{i=1}^{K}\frac{\Gamma(x_i+\alpha_i)}{\Gamma(x_i+1)\Gamma(\alpha_i)}. \tag{3.3}$$

3.1.2 Asymptotic Behavior

Similar to Wu and Browne (2015a), the three pmf/pdf of $f(\boldsymbol{p}|\boldsymbol{\pi}, n)$, $f(\boldsymbol{\pi}|\boldsymbol{\tau}, \alpha)$, $f(\boldsymbol{p}|\boldsymbol{\tau}, \alpha, n)$ in the current work are related asymptotically as n and/or α tends to infinity. Theoretically, as $\alpha \to \infty$ while keeping n fixed, the model error diminishes. $\boldsymbol{\pi}$ converges in probability to $\boldsymbol{\tau}$, and the pmf $f(\boldsymbol{p}|\boldsymbol{\tau}, \alpha, n)$ reduces to the multinomial distribution with pmf $f(\boldsymbol{p}|\boldsymbol{\tau}, n)$. Additionally, as $n \to \infty$ with α fixed, the sampling error diminishes. $\boldsymbol{p}$ converges in probability to $\boldsymbol{\pi}$, and the pmf $f(\boldsymbol{p}|\boldsymbol{\tau}, \alpha, n)$ reduces to the Dirichlet distribution with pdf $f(\boldsymbol{\pi}|\boldsymbol{\tau}, \alpha)$. The following proposition summarizes the asymptotic behaviors of these three distributions:

Proposition 1. *The marginal Dirichlet-Multinomial distribution with pmf $f(\boldsymbol{p}|\boldsymbol{\tau}, \alpha, n)$ converges to the multinomial distribution with pmf $f(\boldsymbol{p}|\boldsymbol{\pi}, n)$ when n is fixed and $\alpha \to \infty$, and to the Dirichlet distribution with pdf $f(\boldsymbol{\pi}|\boldsymbol{\tau}, \alpha)$ as α is fixed and $n \to \infty$.*

Additionally, the following proposition can be established when both α and n tends to ∞.

Proposition 2. *Assuming the models are presented by Equations 3.1 and 3.2, as $\alpha \to \infty$ and*

$n \to \infty$,

$$\sqrt{\frac{n\alpha}{n+\alpha}}(\boldsymbol{p} - \boldsymbol{\tau}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}), \tag{3.4}$$

in which $\boldsymbol{\Gamma}$ [1] *is a* $K \times K$ *matrix*

$$\boldsymbol{\Gamma} = Diag(\boldsymbol{\tau}_0) - \boldsymbol{\tau}_0\boldsymbol{\tau}_0^T, \tag{3.5}$$

where the $\boldsymbol{\tau}_0$ *denotes the true model-implied proportion vector,* $Diag(\boldsymbol{\tau}_0)$ *denotes the diagonal matrix having* $\boldsymbol{\tau}_0$ *on the diagonal.*

Proposition 2 indicates that when both the concentration parameter α and the sample size n are large enough, the marginal distribution of $\boldsymbol{p}$ can be approximated by $\mathcal{N}(\boldsymbol{\tau}_0, (\frac{1}{\alpha} + \frac{1}{n})\boldsymbol{\Gamma})$. It is worth noting that the matrix $\boldsymbol{\Gamma}$ has rank $K - 1$.

3.2 Estimating Concentration and Submodel Parameters

In our Dirichlet-Multinomial framework, all model parameters including the submodel parameters (i.e., item parameters) and the concentration/dispersion parameter are simultaneously estimated. The estimation problem is presented in two scenarios. In the first, ideal scenario, it is assumed that the population probability vector $\boldsymbol{\pi}$ is known, and the model parameters are estimated by maximizing the Dirichlet log-likelihood. In the second, practical scenario, the population probability vector is unknown, and the model parameters are estimated by maximizing the Dirichlet-Multinomial log-likelihood. Because only one sample from one operational population is obtained from the general population, it is expected that the estimated concentration/dispersion

[1] In the study, the matrix is denoted as the bold symbol $\boldsymbol{\Gamma}$ to distinguish it from the unbold Γ representing the Gamma function.

parameter is severely biased. We then propose a bias correction procedure by modifying the estimating equation when the population is unknown.

3.2.1 Model for a Known Population

3.2.1.1 Asymptotic Result

The pdf of $\boldsymbol{\pi}$ is specified in Equation 3.2. From this point forward, the log-likelihood functions will include the dispersion parameter v instead of the concentration parameter α for computational purposes. The log-likelihood of the Dirichlet distribution is

$$\log f_D(\boldsymbol{\pi}|\boldsymbol{\tau}, v) = \log\Gamma(\frac{1}{v}) - \sum_{i=1}^{K}\log\Gamma[\frac{1}{v}\tau_i(\boldsymbol{\xi})] + \sum_{i=1}^{K}[\frac{1}{v}\tau_i(\boldsymbol{\xi}) - 1]\log\pi_i, \tag{3.6}$$

in which $\boldsymbol{\tau}(\boldsymbol{\xi})$ can be obtained by calculating the IRFs in multinomial submodels discussed in Section 2.2.2. By maximizing the log-likelihood in Equation 3.6, two sets of parameters can be estimated simultaneously: the submodel parameters $\boldsymbol{\xi}$ and the dispersion parameter v. Profile likelihood is used when two sets of parameters are estimated from the same likelihood function (Murphy & van der Vaart, 2000; Sprott, 2008). I first express an approximate solution to the estimating equation for $\boldsymbol{\xi}$ as a function of v in Equation 3.8. The solution is then plugged into the estimating equation for v in Equation 3.7, from which the estimate for v is obtained.

We can take partial derivatives on Equation 3.6 with respect to v and $\boldsymbol{\xi}$

$$\frac{\partial}{\partial v}\log f_D = -\psi(\frac{1}{v})\frac{1}{v^2} + \sum_{i=1}^{K}\psi[(\frac{1}{v})\tau_i(\boldsymbol{\xi})]\tau_i(\boldsymbol{\xi})\frac{1}{v^2} - \sum_{i=1}^{K}\tau_i(\boldsymbol{\xi})\log\pi_i(\frac{1}{v^2}), \tag{3.7}$$

$$\frac{\partial}{\partial \boldsymbol{\xi}}\log f_D = -\sum_{i=1}^{K}\psi[(\frac{1}{v})\tau_i(\boldsymbol{\xi})](\frac{1}{v})\frac{\partial\tau_i(\boldsymbol{\xi})}{\partial\boldsymbol{\xi}} + \sum_{i=1}^{K}(\log\pi_i)(\frac{1}{v})\frac{\partial\tau_i(\boldsymbol{\xi})}{\partial\boldsymbol{\xi}}. \tag{3.8}$$

Suppose v is fixed and is close enough to 0 (i.e., α is very large), and let $\frac{\partial}{\partial \boldsymbol{\xi}} \log f_D = 0$, the estimator of submodel parameters $\boldsymbol{\xi}$ can be obtained using approximation methods (e.g., Stirling's formula and Taylor series expansion; Abramowitz & Stegun, 1964), and the following proposition can be established

Proposition 3.

$$\sqrt{\alpha}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, (\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1} \boldsymbol{\Delta}_0)^{-1}), \tag{3.9}$$

where $\boldsymbol{\xi}_0$ denotes the true submodel parameters. $\boldsymbol{D}_0^{-1}$ is a $K \times K$ diagonal matrix,

$$\boldsymbol{D}_0^{-1} = \begin{pmatrix} \frac{1}{\tau_1(\boldsymbol{\xi}_0)} & 0 & \cdots & 0 \\ 0 & \ddots & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\tau_K(\boldsymbol{\xi}_0)} \end{pmatrix}. \tag{3.10}$$

$\boldsymbol{\Delta}_0$ is a $K \times q$ vector of partial derivatives

$$\boldsymbol{\Delta}_0 = \begin{pmatrix} \left(\frac{\partial}{\partial \boldsymbol{\xi}} \tau_1(\boldsymbol{\xi}_0)\right)^T \\ \vdots \\ \left(\frac{\partial}{\partial \boldsymbol{\xi}} \tau_K(\boldsymbol{\xi}_0)\right)^T \end{pmatrix}. \tag{3.11}$$

3.2.1.2 Bias in Dispersion Parameter Estimate

After $\hat{\boldsymbol{\xi}}$ is estimated, $\hat{v}$ can also be obtained by plugging in $\hat{\boldsymbol{\xi}}$ into the partial derivative of $\frac{\partial}{\partial v} \log f_D = 0$ in Equation 3.7. It is found that

$$\hat{v} = \frac{1}{K-1} \sum_{i=1}^{K} \frac{(\pi_i - \tau_i(\hat{\boldsymbol{\xi}}))^2}{\tau_i(\hat{\boldsymbol{\xi}})} + o_p(v_0). \tag{3.12}$$

In the given context, v_0 represents the true dispersion parameter, while $\hat{v}$ is an estimator of v_0. Asymptotically, as $\alpha \to \infty$, it is observed that

$$\frac{1}{v_0} \sum_{i=1}^{K} \frac{(\pi_i - \tau_i(\hat{\boldsymbol{\xi}}))^2}{\tau_i(\hat{\boldsymbol{\xi}})} = \frac{1}{v_0} (\boldsymbol{\pi} - \boldsymbol{\tau}(\hat{\boldsymbol{\xi}}))^T \boldsymbol{D}_0^{-1} (\boldsymbol{\pi} - \boldsymbol{\tau}(\hat{\boldsymbol{\xi}})) \xrightarrow{d} \chi^2_{K-q-1}. \tag{3.13}$$

Based on the findings presented in Equations 3.12 and 3.13, we can derive the asymptotic result as follows:

Proposition 4.

$$\frac{\hat{v}}{v_0} \xrightarrow{d} \frac{\chi^2_{K-q-1}}{K-1}. \tag{3.14}$$

Equation 3.14 shows the approximation of the estimated parameter $\hat{v}$ over the true parameter v_0 in the Dirichlet model. In the specified chi-square distribution, where the random variable $X \sim \chi^2_{K-q-1}$, the expected mean is equal to the df, which is given by $K-q-1$ in that distribution. In this scenario,

$$E(\frac{\hat{v}}{v_0}) \approx \frac{K-q-1}{K-1} < 1. \tag{3.15}$$

Consequently, $\hat{v}$ underestimates the true v_0.

3.2.1.3 Connection Between the Concentration Parameter and RMSEA$_n$

The proposed framework aims to provide a support to why inferences made from one single operational population has implications in the general population of interest. The research found a way to quantify whether the operational population is representative of the general population of interest, and to what extent the model can be used to make inferences on the general population of interest. Under the assumption that the adventitious error is a random effect, the study builds a connection between v and RMSEA$_n$, which is an effect size measure that quantifies model discrepancies.

In a fixed-effect approach in Maydeu-Olivares and Joe (2014), suppose $\boldsymbol{\pi}$ is the true population (data-generating) probability, and $\boldsymbol{\tau}(\boldsymbol{\xi}_0)$ is the closest approximation to $\boldsymbol{\pi}$ in KL divergence within the fitted family of models. The RMSEA$_n$, denoted ε_n, can be computed as follows:

$$\varepsilon_n = \sqrt{\frac{D_n}{df}} = \sqrt{\frac{(\boldsymbol{\pi} - \boldsymbol{\tau}(\boldsymbol{\xi}_0))^T \boldsymbol{D}_0^{-1} (\boldsymbol{\pi} - \boldsymbol{\tau}(\boldsymbol{\xi}_0))}{K - q - 1}}. \tag{3.16}$$

From Equation 3.12, it is established that

$$\hat{v} = \frac{1}{K-1} (\boldsymbol{\pi} - \boldsymbol{\tau}(\hat{\boldsymbol{\xi}}))^T \boldsymbol{D}_0^{-1} (\boldsymbol{\pi} - \boldsymbol{\tau}(\hat{\boldsymbol{\xi}})) + o_p(v_0). \tag{3.17}$$

As $\alpha \to \infty$, from Lemma 8 in Appendix A, it is found that $\boldsymbol{\tau}(\hat{\boldsymbol{\xi}}) \xrightarrow{p} \boldsymbol{\tau}(\boldsymbol{\xi}_0)$. Taking expectation on both sides of Equation 3.17 and using the results in Equation 3.16 yields

$$E(\hat{v}) \approx \frac{1}{K-1} \varepsilon_n^2 (K - q - 1) \tag{3.18}$$

From the Equations 3.15 and 3.18, it can be concluded that the connection between v_0 and ε_n is

$$v_0 \approx \varepsilon_n^2. \tag{3.19}$$

Cutoff suggestions for the indices can be used to quantify the largest adventitious error allowed to retain the model, such that we know how good our model is fitted to data in the sense of a broader general population of interest. According to Maydeu-Olivares and Joe (2014), a stringent close-fit cutoff value of 0.03 for RMSEA_n is recommended for unidimensional IRT models with binary data.

3.2.2 Model for an Unknown Population

The pmf of a Dirichlet-Multinomial distribution is specified in Equation 3.3. The general objective function of the Dirichlet-Multinomial distribution (G-O-DM) is

$$\begin{aligned}&\log f^*_{DM}(\boldsymbol{p}|\boldsymbol{\xi}, v, n)\\ &= \frac{1}{m}\log\Gamma(m\frac{1}{v}) - \frac{1}{m}\log\Gamma(m(n+\frac{1}{v})) + \sum_{i=1}^{K}[\log\Gamma(np_i + \frac{1}{v}\tau_i(\boldsymbol{\xi})) - \log\Gamma(\frac{1}{v}\tau_i(\boldsymbol{\xi}))].\end{aligned} \tag{3.20}$$

To estimate the model parameters v and $\boldsymbol{\xi}$, we take partial derivatives with respect to these two sets of parameters.

$$\begin{aligned}\frac{\partial}{\partial v}\log f_{DM}^{*} &= \psi(m\frac{1}{v})(-v^{-2}) - \psi[m(n+\frac{1}{v})](-v^{-2}) \\ &\quad + \sum_{i=1}^{K}\tau_i(\boldsymbol{\xi})(-v^{-2})[\psi(np_i+\frac{1}{v}\tau_i(\boldsymbol{\xi})) - \psi(\frac{1}{v}\tau_i(\boldsymbol{\xi}))], \qquad (3.21)\end{aligned}$$

$$\frac{\partial}{\partial \boldsymbol{\xi}}\log f_{DM}^{*} = \sum_{i=1}^{K}\frac{1}{v}\frac{\partial\tau_i(\boldsymbol{\xi})}{\partial\boldsymbol{\xi}}[\psi(np_i+\frac{1}{v}\tau_i(\boldsymbol{\xi})) - \psi(\frac{1}{v}\tau_i(\boldsymbol{\xi}))]. \qquad (3.22)$$

We can obtain the $\hat{\boldsymbol{\xi}}^{*}$ and $\hat{v}^{*}$ by setting $\frac{\partial}{\partial\boldsymbol{\xi}}\log f_{DM}^{*}=0$ and $\frac{\partial}{\partial v}\log f_{DM}^{*}=0$.

As $n\to\infty$ and $\alpha\to\infty$ the following proposition shows the asymptotic behavior of the estimator $\hat{\boldsymbol{\xi}}^{*}$:

Proposition 5.

$$\sqrt{\frac{n\alpha}{n+\alpha}}(\hat{\boldsymbol{\xi}}^{*}-\boldsymbol{\xi}_0)\xrightarrow{d}\mathcal{N}(\mathbf{0},(\boldsymbol{\Delta}_0^{T}\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}). \qquad (3.23)$$

The proof to Equation 3.23 follows from Lemma 1 in Appendix A. In addition, it is found that

$$\frac{n\alpha}{n+\alpha}(\boldsymbol{p}-\boldsymbol{\tau}(\hat{\boldsymbol{\xi}}^{*}))^{T}\boldsymbol{D}_0^{-1}(\boldsymbol{p}-\boldsymbol{\tau}(\hat{\boldsymbol{\xi}}^{*}))\xrightarrow{d}\chi^2_{K-q-1}. \qquad (3.24)$$

The proof to Equation 3.24 follows from Lemma 5 in Appendix A. Moreover, the following proposition shows the asymptotic behavior of the estimator for the dispersion parameter

Proposition 6.

$$(\frac{\hat{v}^{*}+\epsilon}{v_0+\epsilon})\xrightarrow{d}\frac{\chi^2_{K-q-1}}{\frac{mK-1}{m}}. \qquad (3.25)$$

The proof to Equation 3.25 follows from Lemma 7 in Appendix A.

In Equation 3.20, $m > 0$ is an additional tuning parameter, which is introduced in order to correct the bias in the estimation of dispersion parameter v. When $m = 1$, the term $\frac{mK-1}{m}$ in Equation 3.25 becomes $K-1$. In this scenario, Equation 3.20 reduces to the log-likelihood of the Dirichlet-Multinomial distribution (LL-DM), where the maximizer of the objective function is the maximum Dirichlet-Multinomial likelihood estimator (MDMLE). Detailed information regarding this scenario is provided in Section 3.2.2.1. Alternatively, as will be established in Section 3.2.2.2, setting $m = \frac{1}{1+q}$ yields a bias-corrected estimator of v, and the term $\frac{mK-1}{m}$ in Equation 3.25 becomes $K-q-1$. In this case, Equation 3.20 reduces to the bias-corrected log-likelihood of the Dirichlet-Multinomial distribution (BC-LL-DM), and the maximizer of the objective function is the bias-corrected maximum Dirichlet-Multinomial likelihood estimator (BC-MDMLE).

It is noted that Wu and Browne (2015a) performed bias correction to their dispersion parameter in the adventitious error approach for covariance structure models. Similar to the m tuning parameter used in our framework, they also introduced a tuning parameter m_1 to define the discrepancy functions for their model. In our case, setting m to $\frac{1}{1+q}$ helps in bias correction in the estimation of the dispersion parameter. In Wu and Browne (2015a), utilizing $m_1 = \frac{df}{q(q+1)/2}$ achieves the bias correction.

3.2.2.1 Asymptotic Results for MDMLE

When two parameter sets are estimated from the same likelihood function while assuming an unknown population, we employ the profile likelihood approach to simultaneously estimate both $\boldsymbol{\xi}$ and v. This approach is similar to the Dirichlet case in Section 3.2.1.1, where the population is assumed to be known. When $m = 1$, maximizing the LL-DM gives the MDMLE for

the model parameters $\boldsymbol{\xi}$ and v. The estimating equation for the item parameters $\hat{\boldsymbol{\xi}}$ is presented in Equation 3.31, while the estimating equation for the dispersion parameter $\hat{v}_{DM}$ can be found in Equation 3.32.

The asymptotic results for $\hat{\boldsymbol{\xi}}$ is specified in Equation 3.23 in Proposition 5. For the asymptotic behavior of the dispersion parameter v, when $\hat{v}_{DM} = 0$, the true sampling distribution of $\hat{v}_{DM}$ has a point mass $P(\chi^2_{K-q-1} < (K-1)\frac{\epsilon}{v_0+\epsilon})$, in which $\epsilon = \frac{1}{n}$[2] and v_0 denotes the true dispersion parameter. The asymptotic approximation in Equation 3.24 shows that $\frac{1}{v_0+\epsilon}(\boldsymbol{p} - \boldsymbol{\tau}(\hat{\boldsymbol{\xi}}))\boldsymbol{D}_0^{-1}(\boldsymbol{p}-\boldsymbol{\tau}(\hat{\boldsymbol{\xi}})) \xrightarrow{d} \chi^2_{K-q-1}$. And thus, if we let $\hat{v}_{DM} = \frac{1}{K-1}(\boldsymbol{p}-\boldsymbol{\tau}(\hat{\boldsymbol{\xi}}))^T\boldsymbol{D}_0^{-1}(\boldsymbol{p}-\boldsymbol{\tau}(\hat{\boldsymbol{\xi}}))-\epsilon$, then

$$\begin{aligned}\frac{\hat{v}_{DM}+\epsilon}{v_0+\epsilon}(K-1) &= (K-1)\frac{1}{v_0+\epsilon}[\frac{1}{K-1}(\boldsymbol{p}-\boldsymbol{\tau}(\hat{\boldsymbol{\xi}}))^T\boldsymbol{D}_0^{-1}(\boldsymbol{p}-\boldsymbol{\tau}(\hat{\boldsymbol{\xi}})) - \epsilon + \epsilon] \\ &\xrightarrow{d} \chi^2_{K-q-1}. \end{aligned} \tag{3.26}$$

Then we can get the estimator $\hat{v}$ by applying the result in Equation 3.26

$$\hat{v} = \begin{cases} \hat{v}_{DM} & \text{if } \hat{v}_{DM} > 0 \\ \frac{1}{K-q-1}(\boldsymbol{p}-\boldsymbol{\tau}(\hat{\boldsymbol{\xi}}))^T\boldsymbol{D}_0^{-1}(\boldsymbol{p}-\boldsymbol{\tau}(\hat{\boldsymbol{\xi}})) - \epsilon & \text{if } \hat{v}_{DM} = 0. \end{cases} \tag{3.27}$$

Based on the aforementioned information, we can conclude that the MDMLE for model parameters includes the estimates of item parameters $\hat{\boldsymbol{\xi}}$ and the dispersion parameter $\hat{v}$.

Using the asymptotic results, the following shows the asymptotic behavior of the estimated dispersion parameter $\hat{v}$:

[2]The ϵ (reciprocal of sample size) is different from ε (RMSEA measure).

Proposition 7.

$$\frac{\hat{v}+\epsilon}{v_0+\epsilon} \xrightarrow{d} \frac{\chi^2_{K-q-1}}{K-1}, \tag{3.28}$$

The proof of the approximation in Equation 3.28 follows from Lemma 6 in Appendix A. A connection can be established with Proposition 4, which states that $\frac{\hat{v}}{v_0}$ converges in distribution to the same ratio when the population $\boldsymbol{\pi}$ is assumed to be known. The only distinction between Propositions 4 and 7 is that when the population $\boldsymbol{\pi}$ is assumed to be unknown, the asymptotic relationship includes an additional term, which is ϵ.

Given that the expected value of χ^2_{K-q-1} is $K-q-1$, taking the expectation on both sides of Proposition 7, we can conclude that

$$E(\hat{v}) \approx \frac{K-q-1}{K-1}v_0 - \frac{q}{K-1}\epsilon < \frac{K-q-1}{K-1}v_0 < \frac{K-1}{K-1}v_0 = v_0, \tag{3.29}$$

which means $\hat{v}$ tends to underestimate the true v_0.

As mentioned in Section 3.1.2, when both $\alpha \to \infty$ and $n \to \infty$, the Dirichlet-Multinomial model converges to the multinomial model. However, directly evaluating the log-likelihood in Equation 3.20 using the log Gamma function has two issues: 1) computationally, evaluating log Gamma functions and their derivatives (digamma functions) may encounter problems when the values involved are very small, 2) the log Gamma function is not defined for $v = 0$, which occurs when the Dirichlet-Multinomial likelihood reduces to a multinomial likelihood. To ensure computational stability and a smooth transition from the Dirichlet-Multinomial model to the multinomial model, when $m = 1$, an alternative formulation of the log-likelihood in Equation 3.20 was proposed in the R package known as Vector Generalized Linear and Additive Models

(VGAM) (Yee, 2010, 2012; Yee & Wild, 1996; Yu & Shaw, 2014). For the rest of this study, we will use the term "VGAM-formulated LL-DM" to refer to this alternative formulation of the LL-DM.

This VGAM-formulated LL-DM can be expressed as follows:

$$\begin{aligned}&\log f_{DM}(\boldsymbol{p}|\boldsymbol{\xi}, v, n)\\&= \sum_{i=1}^{K}\sum_{r=1}^{x_i}\log[\tau_i(\boldsymbol{\xi})(\frac{1}{1+v}) + (r-1)(\frac{v}{1+v})] - \sum_{r=1}^{n}\log[\frac{1}{1+v} + (r-1)\frac{v}{1+v}].\end{aligned} \tag{3.30}$$

Again, to estimate the model parameters v and $\boldsymbol{\xi}$, we take partial derivatives with respect to these two sets of parameters.

$$\frac{\partial}{\partial v}\log f_{DM} = \sum_{i=1}^{K}\sum_{r=1}^{x_i}\{\frac{-\tau_i(\boldsymbol{\xi}) + r - 1}{[\tau_i(\boldsymbol{\xi}) + (r-1)v](1+v)}\} - \sum_{r=1}^{n}\{\frac{r-2}{[1+(r-1)v](1+v)}\}, \tag{3.31}$$

$$\frac{\partial}{\partial \boldsymbol{\xi}}\log f_{DM} = \sum_{i=1}^{K}\sum_{r=1}^{x_i}\{\frac{1}{\tau_i(\boldsymbol{\xi}) + (r-1)v}\frac{\partial \tau_i(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}}\}. \tag{3.32}$$

Following a similar approach as in the Dirichlet case, we can determine the $\hat{\boldsymbol{\xi}}$ and $\hat{v}_{DM}$ by setting $\frac{\partial}{\partial \boldsymbol{\xi}}\log f_{DM} = 0$ and $\frac{\partial}{\partial v}\log f_{DM} = 0$.

3.2.2.2 Asymptotic Results for BC-MDMLE

From the proof in Lemma 7 in Appendix A, we found that letting $m = \frac{1}{1+q}$ in the G-O-DM in Equation 3.20 helps us to correct the bias in the estimated dispersion parameter. When $m = \frac{1}{1+q}$, maximizing the BC-LL-DM gives the BC-MDMLE for the model parameters $\boldsymbol{\xi}$ and v. The estimating equation for the item parameters $\hat{\boldsymbol{\xi}}$ is presented in Equation 3.37, while the estimating equation for the dispersion parameter $\hat{v}_{BC-DM}$ can be found in Equation 3.38.

Similar to the asymptotic results for MDMLE in Section 3.2.2.1, the asymptotic results for $\hat{\boldsymbol{\xi}}$ is also specified in Equation 3.23 in Proposition 5. As for the asymptotic behavior of the dispersion parameter v, we found out that when the estimator $\hat{v}_{BC-DM} = 0$, the true sampling distribution of $\hat{v}_{BC-DM}$ has a point mass $P(\chi^2_{K-q-1} < (K-q-1)\frac{\epsilon}{v_0+\epsilon})$. The asymptotic approximation in Equation 3.24 shows that $\frac{1}{v_0+\epsilon}(\boldsymbol{p}-\boldsymbol{\tau}(\hat{\boldsymbol{\xi}}))\boldsymbol{D}_0^{-1}(\boldsymbol{p}-\boldsymbol{\tau}(\hat{\boldsymbol{\xi}})) \xrightarrow{d} \chi^2_{K-q-1}$. And thus, if we let $\hat{v}_{BC-DM} = \frac{1}{K-q-1}(\boldsymbol{p}-\boldsymbol{\tau}(\hat{\boldsymbol{\xi}}))^T\boldsymbol{D}_0^{-1}(\boldsymbol{p}-\boldsymbol{\tau}(\hat{\boldsymbol{\xi}})) - \epsilon$, then

$$\begin{aligned}\frac{\hat{v}_{BC-DM}+\epsilon}{v_0+\epsilon}(K-q-1) &= (K-q-1)\frac{1}{v_0+\epsilon}[\frac{1}{K-q-1}(\boldsymbol{p}-\boldsymbol{\tau}(\hat{\boldsymbol{\xi}}))^T\boldsymbol{D}_0^{-1}(\boldsymbol{p}-\boldsymbol{\tau}(\hat{\boldsymbol{\xi}})) - \epsilon + \epsilon] \\ &\xrightarrow{d} \chi^2_{K-q-1}. \end{aligned} \tag{3.33}$$

We can get the unbiased estimator $\hat{v}_0$ by applying the result in Equation 3.33

$$\hat{v}_0 = \begin{cases} \hat{v}_{BC-DM} & \text{if } \hat{v}_{BC-DM} > 0 \\ \frac{1}{K-q-1}(\boldsymbol{p}-\boldsymbol{\tau}(\hat{\boldsymbol{\xi}}))^T\boldsymbol{D}_0^{-1}(\boldsymbol{p}-\boldsymbol{\tau}(\hat{\boldsymbol{\xi}})) - \epsilon & \text{if } \hat{v}_{BC-DM} = 0. \end{cases} \tag{3.34}$$

Based on the aforementioned information, we can conclude that the BC-MDMLE for model parameters includes the estimates of item parameters $\hat{\boldsymbol{\xi}}$ and the dispersion parameter $\hat{v}_0$.

Using the asymptotic results, the following shows the asymptotic behavior of the estimated dispersion parameter $\hat{v}_0$

Proposition 8.

$$\frac{\hat{v}_0+\epsilon}{v_0+\epsilon} \xrightarrow{d} \frac{\chi^2_{K-q-1}}{K-q-1}. \tag{3.35}$$

in which $\hat{v}_0$ exhibits asymptotic unbiasedness.

Similar to the LL-DM, when $v = 0$, it indicates a smooth transition from a Dirichlet-

Multinomial model to a multinomial model. However, the VGAM-formulated LL-DM specified in Equation 3.30 to enable this transition is not directly applicable to the BC-LL-DM. This is because a VGAM-formulated BC-LL-DM requires the ratio $\frac{n}{m}$ to be an integer, which may not always hold true. Instead, by utilizing Stirling's formula, it is discovered that

$$\begin{aligned}
&\log f_{DM}^{\#}(\boldsymbol{p}|\boldsymbol{\xi}, v, n) \\
&\approx \sum_{i=1}^{K}\sum_{r=1}^{x_i}\log[\tau_i(\boldsymbol{\xi})(\frac{1}{1+v}) + (r-1)(\frac{v}{1+v})] - \sum_{r=1}^{n}\log[\frac{1}{1+v} + (r-1)\frac{v}{1+v}] \\
&+ (\frac{1}{2m} - \frac{1}{2})\log(nv+1). \qquad (3.36)
\end{aligned}$$

Taking partial derivatives with respect to v and $\boldsymbol{\xi}$ shows that

$$\begin{aligned}
\frac{\partial}{\partial v}\log f_{DM}^{\#} &\approx \sum_{i=1}^{K}\sum_{r=1}^{x_i}\{\frac{-\tau_i(\boldsymbol{\xi}) + r - 1}{[\tau_i(\boldsymbol{\xi}) + (r-1)v](1+v)}\} - \sum_{r=1}^{n}\{\frac{r-2}{[1+(r-1)v](1+v)}\} \\
&\quad + (\frac{1}{2m} - \frac{1}{2})\frac{n}{nv+1}, \qquad (3.37) \\
\frac{\partial}{\partial \boldsymbol{\xi}}\log f_{DM}^{\#} &\approx \sum_{i=1}^{K}\sum_{r=1}^{x_i}\{\frac{1}{\tau_i(\boldsymbol{\xi}) + (r-1)v}\frac{\partial \tau_i(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}}\}. \qquad (3.38)
\end{aligned}$$

Then, we can obtain the $\hat{\boldsymbol{\xi}}$ and $\hat{v}_{BC-DM}$ by setting $\frac{\partial}{\partial \boldsymbol{\xi}}\log f_{DM}^{\#} = 0$ and $\frac{\partial}{\partial v}\log f_{DM}^{\#} = 0$.

After establishing the asymptotic distributions for $\hat{\boldsymbol{\xi}}$ in Equation 3.23 and $\hat{v}_0$ in Equation 3.35, two-sided Wald-type equal-tailed CIs can be constructed to quantify the uncertainty for the item parameters. By consistency established in Lemma 8 in Appendix A, we can replace the unknown true parameters with their corresponding estimates. From Proposition 5, by replacing $(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}$ with $(\hat{\boldsymbol{\Delta}}^T\hat{\boldsymbol{D}}^{-1}\hat{\boldsymbol{\Delta}})^{-1}$, a Wald-type 95% CI for the item parameters in response

category i can be constructed as

$$\hat{\xi}_i \pm z_{(97.5\%)} \sqrt{(\hat{v}_0 + \epsilon)[\hat{\boldsymbol{\Delta}}^T \hat{\boldsymbol{D}}^{-1} \hat{\boldsymbol{\Delta}}]^{ii}}. \tag{3.39}$$

In the given context, $\hat{\xi}_i$ represents the i-th component of $\hat{\boldsymbol{\xi}}$, and $z_{(97.5\%)}$ denotes the 97.5-th percentile of the standard normal distribution $N(0, 1)$. Additionally, $[\hat{\boldsymbol{\Delta}}^T \hat{\boldsymbol{D}}^{-1} \hat{\boldsymbol{\Delta}}]^{ii}$ refers to the diagonal elements located in the i-th row and i-th column of the inverse matrix of $(\hat{\boldsymbol{\Delta}}^T \hat{\boldsymbol{D}}^{-1} \hat{\boldsymbol{\Delta}})$. From the asymptotic relationship in Equation 3.35, a 95% CI for the dispersion parameter can be constructed as

$$\frac{(K - q - 1)(\hat{v}_0 + \epsilon)}{\chi^2_{K-q-1}(2.5\%)} - \epsilon \quad \text{and} \quad \frac{(K - q - 1)(\hat{v}_0 + \epsilon)}{\chi^2_{K-q-1}(97.5\%)} - \epsilon, \tag{3.40}$$

in which the $\chi^2_{K-q-1}(2.5\%)$ and $\chi^2_{K-q-1}(97.5\%)$ are at the 2.5% and 97.5% quantiles of the chi-square distribution with a $df = K - q - 1$.

Chapter 4: Monte Carlo Simulations

4.1 Purpose

This chapter presents the results of a Monte Carlo simulation study that compares the use of our random-effect Dirichlet-Multinomial framework with traditional multinomial submodels, considering both dichotomous and polytomous item responses. The study investigates various aspects including model convergence, point estimation, and interval estimation. It is anticipated that the conventional maximum multinomial likelihood estimation in the traditional multinomial submodels performs poorly when the magnitude of adventitious model error is non-negligible. We expect that the proposed Dirichlet-Multinomial model with the BC-MDMLE, will outperform both the Dirichlet-Multinomial model with the MDMLE and the maximum multinomial likelihood estimator (MMLE) from the traditional multinomial submodels.

The chapter is structured as follows. Firstly, the data generating process is described to provide an understanding of the simulation study conditions. This includes information on how the item parameters and dispersion parameters are generated, as well as the true parameters utilized in the simulation study. Secondly, the method section presents the computational details and evaluation criteria employed in the study. Finally, the results section presents the findings. This section examines the convergence rates, bias, and Root Mean Squared Error (RMSE) values observed when using the BC-MDMLE, MDMLE, and the MMLE. Additionally, it evaluates the

empirical coverage of the 95% CIs for the model parameters associated with these estimators.

4.2 Data Generating Process

The following four factors are manipulated in the simulation study:

(1) Population RMSEA$_n$: It is the full-information RMSEA$_n$ defined in Equation 2.14. The values considered for this factor are 0.01, 0.03, 0.06, and 0.09. The chosen values are based on the suggested cutoffs in Maydeu-Olivares and Joe (2014).

(2) Sample size (n): Two values are examined for this factor, namely 500 and 5000.

(3) Number of response categories (C) per item: Two options are explored, specifically two and four. Two for dichotomous items, and four for polytomous items with four response categories.

(4) Total number of response patterns (K): For dichotomous items, the test length J can be either six or 12, whereas for polytomous items with four response categories, J can be three or six. There are two conditions in this factor: $K = 64$ when $C = 2, J = 6$ and $C = 4, J = 3$, or $K = 4096$ when when $C = 2, J = 12$ and $C = 4, J = 6$.

Our study examines the effects of four manipulated factors, which are organized into different conditions. Our study follows a four-way fully crossed design, which includes all possible combinations of the four manipulated variables. The total number of response patterns K has two conditions, corresponding to either two or four response categories C per item. Consequently, our study encompasses a total of 32 simulation conditions, derived from the calculation

4 (population RMSEA$_n$) $\times$ 2 (sample sizes) $\times$ 2 (number of response categories)

$\times$ 2 (total number of response patterns) $= 32$.

The population RMSEA$_n$ values are selected based on the recommended cutoffs outlined in Maydeu-Olivares and Joe (2014). In their study, they suggested a cutoff value of 0.03 for RMSEA$_n$ as a reasonable indicator of close-fit for unidimensional IRT models with binary data. Starting from this value of 0.03, our investigation aimed to explore conditions where the model fit could be even better or worse. The sample size in this study varies from a relatively small size of 500 to a large size of 5000. Additionally, we chose to have two and four response categories per item, as these are commonly encountered in real testing scenarios. The total number of response patterns K are determined to reflect the degree of sparsity, which means that the dfs remain the same for scenarios where $C = 2, J = 6, K = 64$ and $C = 4, J = 3, K = 64$. Similarly, the df matches for scenarios involving $C = 2, J = 12, K = 4096$ and $C = 4, J = 6, K = 4096$.

We use the dispersion parameter, denoted as v, instead of the concentration parameter (represented as both $\log(\alpha)$ and α) for computational convenience. The IRT models employed in this study include the 2PL model for items with two response categories and the GRM for items with four response categories. True item parameters for dichtomous items were generated according to the following procedure: For each item j, the item slope parameter a_j was selected from a uniform distribution $\mathcal{U}(0.75, 2)$, while the item difficulty parameter b_j was chosen independently from a uniform distribution $\mathcal{U}(-2, 2)$ (Liu, Yang, & Maydeu-Olivares, 2019; Rupp, 2013). The item intercept parameter d_j was calculated as $d_j = -a_j b_j$. The true item parameters for the six and 12 dichotomous items can be found in Table 4.1.

Table 4.1

The True Item Parameters for Dichotomous Items

Test Length	Item	Slope a_j	Intercept d_j
6	1	1.109	-0.125
	2	1.735	-2.724
	3	1.261	-0.259
	4	1.854	0.322
	5	1.926	-3.519
	6	0.807	0.151
12	1	0.874	1.030
	2	1.612	-2.306
	3	0.934	1.084
	4	1.650	-0.966
	5	1.605	2.301
	6	1.614	2.376
	7	0.877	0.157
	8	0.822	0.353
	9	1.013	-0.512
	10	0.813	1.556
	11	1.527	0.053
	12	1.469	-1.790

In generating item parameter values for polytomous items with four response categories, we used the following procedure: For each item j, the item slope parameter a_j was randomly sampled from a uniform distribution $\mathcal{U}(0.75, 2)$. The item difficulty parameters for each category c (denoted as b_{jc}, where c takes values from one to three) was generated from a uniform distribution $\mathcal{U}(-2, 2)$. The three difficulty parameters (b_{j1}, b_{j2}, and b_{j3}) were sorted in an increasing order. In addition, the difference between any two consecutive difficulty parameters was required to be larger than 0.7 for computation convenience. Similar to the data generation process for dichotomous items with two response categories, the slope and difficulty parameters in this case were also generated independently. For each category c, the item intercept parameter was calculated as $d_{jc} = -a_j b_{jc}$. The item intercept parameters were in a decreasing order from category

one to category three. The true item parameters for the three and six polytomous items with four response categories can be found in Table 4.2.

Table 4.2

The True Item Parameters for the Polytomous Items With Four Response Categories

Test Length	Item	Slope a_j	Intercept d_{j1}	Intercept d_{j2}	Intercept d_{j3}
3	1	1.266	1.464	0.030	-1.011
	2	1.694	2.604	0.358	-2.923
	3	1.495	2.277	-0.899	-3.043
6	1	1.472	1.910	-1.345	-2.713
	2	1.782	2.801	1.294	-2.070
	3	1.672	2.713	-0.737	-3.014
	4	1.636	2.217	0.697	-2.966
	5	1.321	2.607	-0.061	-2.201
	6	1.885	1.265	-1.344	-3.650

The data generation process is as follows. In each simulation condition, the population probability vectors $\boldsymbol{\pi}$ are initially generated using a Dirichlet distribution with the true dispersion parameter v_0 and true item parameters $\boldsymbol{\xi}_0$ (see Equation 3.2). The true values of v_0 are derived from the corresponding population RMSEA$_n$ (see the connection in Equation 3.19). $\boldsymbol{\tau}(\boldsymbol{\xi}_0)$ can be directly computed based on either the 2PL or the GRM (see Equations 2.11 and 2.10). To approximate the intractable integral, which cannot be solved analytically using standard mathematical techniques, numerical quadrature was used. Subsequently, the counts $\boldsymbol{x}$ are generated by a multinomial distribution with a sample size of n and the generated population probability vector $\boldsymbol{\pi}$. The resulting population counts $\boldsymbol{x}$ represent our generated data when the population probability is assumed to be unknown.

4.3 Method

4.3.1 Computation Details

R (R Core Team, 2022) was used for all the computation in the simulation study. To estimate the item parameters in traditional multinomial submodel approaches, the mirt function from the R package mirt (Chalmers et al., 2015) was employed. The expectation-maximization (EM) algorithm (Bock & Aitkin, 1981; Dempster, Laird, & Rubin, 1977) was used for estimating item parameters. For precise parameter estimation, a convergence criterion of 10^{-6} was set in the mirt function. Additionally, to enhance accuracy, the maximum number of EM cycles was set to 5000. The number of replications across all conditions was set to 500.

To obtain the BC-MDMLE and MDMLE, the nlminb function from the R *stats* was used for optimization. In the nlminb function, we developed analytical functions that served as the objective function and gradients, aiming to achieve more accurate model parameter estimates and improve computation efficiency. The objective function for the BC-MDMLE is defined in Equation 3.36, while the objective function for MDMLE is specified in Equation 3.30. The gradients for BC-MDMLE are provided in Equations 3.37 and 3.38, while the gradients for MDMLE are given in Equations 3.31 and 3.32.

Regarding other argument parameters used in the nlminb function, for items with two response categories, the starting values for the slope parameters were set to 1, while the intercept parameters were initialized at 0. For items with four response categories, the slope parameters also began at 1, and the three intercept parameters were initialized in decreasing order as -2, 0, and 2. A constraint was placed on the dispersion parameter estimate by setting its lower bound

to 0, ensuring that the estimate could not go below this value. No upper bound was enforced on the dispersion parameter estimate. In contrast, there were no limitations imposed on the item parameter estimates, allowing them to take any real value. The lower bound was denoted as '-Inf' (negative infinity), while the upper bound was represented as 'Inf' (infinity). To ensure better convergence, the maximum number of allowed iterations was set to 2000, allowing for a larger number of iterations to be performed during the analysis.

4.3.2 Evaluation Criteria

4.3.2.1 Convergence

Multiple criteria were established to determine the convergence of parameter estimation algorithms. When fitting a traditional multinomial submodel using the EM algorithm, we establish a more stringent convergence criterion. In the mirt package, this criterion is set to a threshold of 10^{-6}. This means that the algorithm continues iterating until the change in estimated parameters is below this threshold, indicating a high level of precision and convergence in the model estimation process. By employing such a stringent convergence threshold, we strive to obtain accurate and reliable results when applying the EM algorithm to fit 2PL or GRM models within the mirt framework.

During the fitting process, when comparing the BC-MDMLE and the MDMLE, we applied a more stringent parameter convergence tolerance ('x.tol' in the nlminb function in R). 'x.tol' sets the tolerance for relative changes in the parameter estimates. The algorithm continues to iterate until the relative changes in the parameter estimates are smaller than the specified tolerance value. Specifically, we used a tolerance of 10^{-10} instead of the default value of 1.5×10^{-8} to obtain more

reliable parameter estimates.

In addition to the convergence criteria, we also excluded problematic replications in which the Fisher information for item parameters exhibits a large condition number. The condition number quantifies the sensitivity of the estimated parameters to small changes in the data and indicates the potential for numerical instability. A small condition number of the information matrix implies that the parameter estimates reach a stable and relatively accurate solution. Conversely, a large condition number of the Fisher information matrix implies that the matrix is close to being singular, which can lead to inaccurate results and significant numerical errors (e.g., Belsley, Kuh, & Welsch, 2005; Higham, 1995). The condition number is the ratio of the largest to the smallest eigenvalue of the estimated information matrix.

There is no definitive threshold for determining what constitutes a large condition number for the Fisher information matrix, as it depends on various factors such as the complexity of the model and the scale of the estimated parameters. In our study, we set a cutoff of 5000. Replications with a Fisher information condition number exceeding 5000 were considered as bad replications and excluded from the calculation of results. We established this threshold by examining the boxplot for outliers in the condition numbers across all simulation conditions.

4.3.2.2 Point Estimates

To evaluate the accuracy of item parameter estimation, bias was calculated for the BC-MDMLE, MDMLE, and the MMLE of the item parameters. The bias values were determined by calculating the difference between the estimated item parameters and the true item parameters.

We also calculated the bias for both BC-MDMLE and the MDMLE of the dispersion pa-

rameter. Due to numerical scale considerations, we used the relative bias, which was expressed as the MDMLE $\hat{v}$ or the BC-MDMLE $\hat{v}_0$ divided by the true dispersion parameter v_0.

To further assess the accuracy of our parameter estimates, we computed the RMSE across replications:

$$RMSE = \sqrt{\frac{1}{H}\sum_{h=1}^{H}(\hat{\boldsymbol{\theta}}^{(h)} - \boldsymbol{\theta}_0)^2}, \tag{4.1}$$

in which H is the total number of replications in the simulation study, $\hat{\boldsymbol{\theta}}^{(h)}$ represents the estimated parameters for the h-th replication, while $\boldsymbol{\theta}_0$ represents the true parameters. The model parameters in our study include both the item parameters and the dispersion parameter, and the RMSE was investigated for both sets of parameters. Similar to the analysis of bias, we calculated RMSEs for the MMLE, as well as for the BC-MDMLE and MDMLE of the item parameters. Additionally, we calculated RMSEs for the BC-MDMLE and MDMLE of the dispersion parameter.

4.3.2.3 Confidence Intervals

We computed 95% CIs for the item parameters associated with three estimators: the MMLE, the BC-MDMLE, and the MDMLE. We then calculated the empirical coverage for these 95% CIs to determine the proportion of instances where the intervals contained the true item parameters (Dodge, Cox, & Commenges, 2003). For the MMLE, we computed the standard errors (SEs) in order to create the 95% CIs for the item parameters.

The SEs were derived using two methods: the expected-information method (SE.type = "Fisher" in mirt) and the sandwich-form information method (SE.type = "sandwich.Louis" in mirt). The sandwich-form SEs are derived using a sandwich-type covariance estimate based on

the cross-product-form information (Yuan, Cheng, & Patton, 2014), along with the exact computation of the Hessian-form information matrix as proposed by Louis (1982). The expected-information SEs are obtained from the expected Fisher information, which is calculated by the negative second derivative of the log-likelihood function with respect to the item parameters (Chalmers et al., 2015). Yuan et al. (2014) has shown that the sandwich-type SEs is robust to model misspecification as long as the likelihood remains to be independent and identically distributed (i.i.d.). Consequently, the CIs calculated using sandwich-type SEs generally tend to be wider compared to those calculated using expected-information SEs. In the present study, we refer to these sandwich-form SEs as the robust SEs, and the expected-information SEs as the non-robust SEs.

The interval defined in Equation 3.39 was used to construct the 95% wald-type CIs for the item parameters associated with the estimators BC-MDMLE and MDMLE. In our study, we utilized the z-distribution to construct these CIs. However, it's worth mentioning that in a similar parameter recovery study conducted by Wu and Browne (2015a), they used the t-distribution for constructing their CIs. The t-distribution approximates the standard normal distribution as the degrees of freedom increase. In our simulation conditions, the smallest degrees of freedom exceeded 50, indicating that the t-distribution closely aligns with the z-distribution. It is important to note that the 95% CI formula in the Equation 3.39 differs from the 95% CI formula used in the traditional multinomial submodels only by inclusion of the term $\hat{v}_0$ within the square root function.

We also calculate the 95% CIs for the dispersion parameter associated with the estimators BC-MDMLE and MDMLE, using the interval specified in Equation 3.40. It is important to note that in practical applications, the empirical coverage of the constructed CIs for the dispersion

parameter may not always be reliable due to sparseness issues. This theory is further supported by the simulation results discussed in Sections 4.4.4.2 and 4.4.4.3, providing more detailed explanation into this issue.

A CI is deemed accurate if the empirical coverage is closer to the nominal level, which equals 95%. If the actual empirical coverage exceeds the nominal level, the intervals are considered conservative. Conversely, if the actual empirical coverage falls below the nominal coverage probability, the intervals are termed "liberal".

4.4 Results

In general, obtaining MMLE is faster compared to MDMLE and BC-MDMLE. The major difference lies within the estimation algorithms used. In our simulation study, we only consider the unidimensional underlying latent trait. The MMLE in the traditional multinomial submodels are obtained by maximizing the likelihood of observed item response patterns using EM algorithms. On the other hand, the optimization used in the study for obtaining MDMLE and BC-MDMLE is the nlminb function in R, which is a nonlinear optimization technique with box constraints. Mathematically speaking, the computation complexity is the same for MDMLE and BC-MDMLE. The loop runs through all the observed response patterns from $i = 1$ to the total number of response patterns K. During the estimation process, response patterns with an observed count equal to 0 are skipped.

4.4.1 Convergence Rate

According to the convergence rate information presented in Figures 4.1, 4.2, 4.3, and 4.4, the convergence rate is higher when the value of RMSEA$_n$ is equal to or less than 0.03 for both the traditional multinomial submodel (2PL or GRM) approaches (see 'mirt' results in the figures), as well as in the Dirichlet-Multinomial framework (see 'BC-MDMLE' and 'MDMLE' results in the figures). Additionally, the convergence rate tends to improve with larger sample sizes. Holding sample size n constant, as the RMSEA$_n$ increases (indicating a larger model error), the convergence rate worsens. However, when the number of response categories is four and the test length is six (see Figure 4.4), the convergence rate reaches 100% across all RMSEA$_n$ and sample size conditions. Comparing the BC-MDMLE/MDMLE to the MMLE, the convergence rate is generally better in the former across all number of response categories and total number of response patterns conditions. This is particularly true when the RMSEA$_n$ is large (in this case, 0.09). The convergence rates of obtaining the BC-MDMLE and MDMLE from the Dirichlet-Multinomial framework are very similar.

Figure 4.1

The Convergence Rate for Six Dichtomous Items

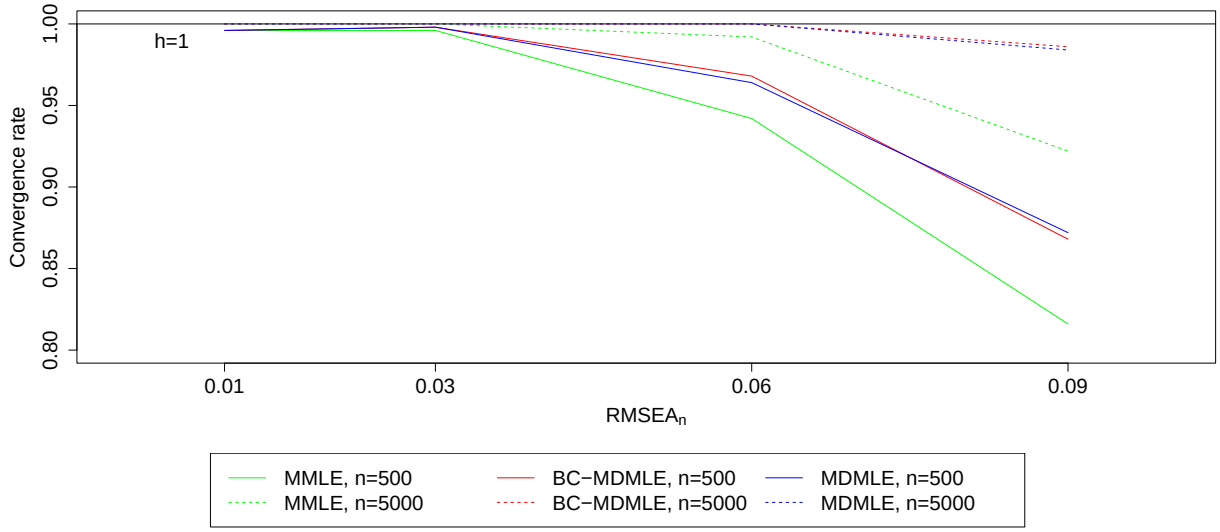


Figure 4.2

The Convergence Rate for 12 Dichtomous Items

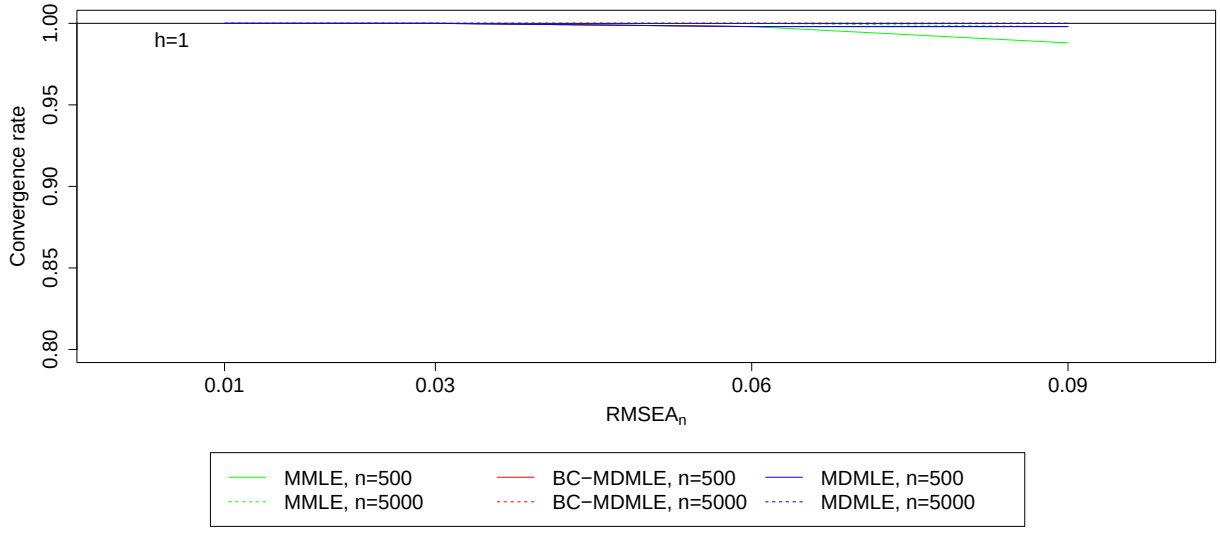


Figure 4.3

The Convergence Rate for Three Polytomous Items With Four Categories

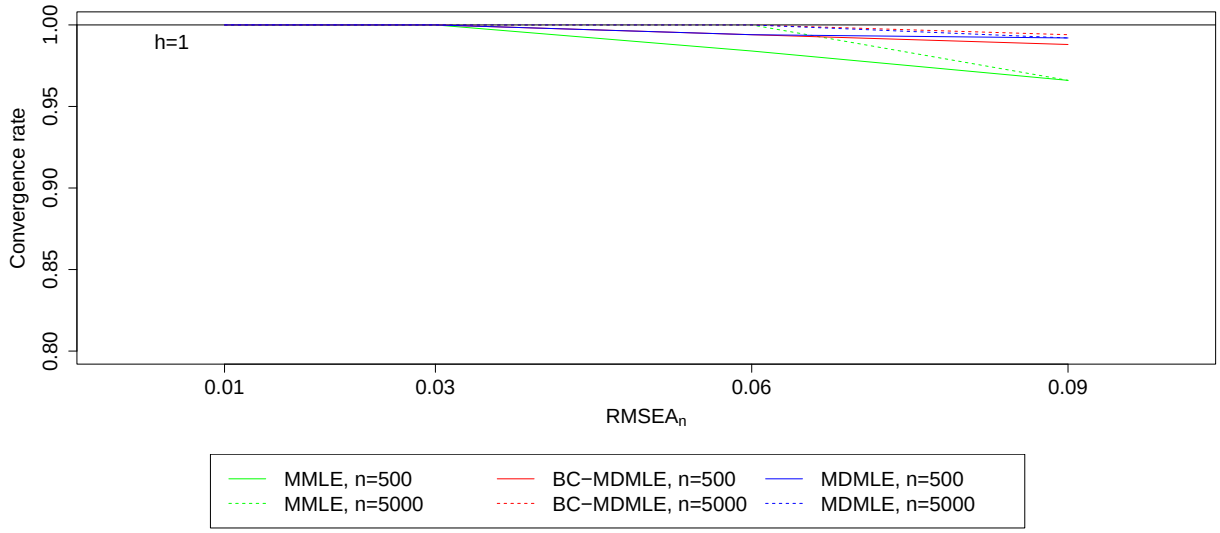
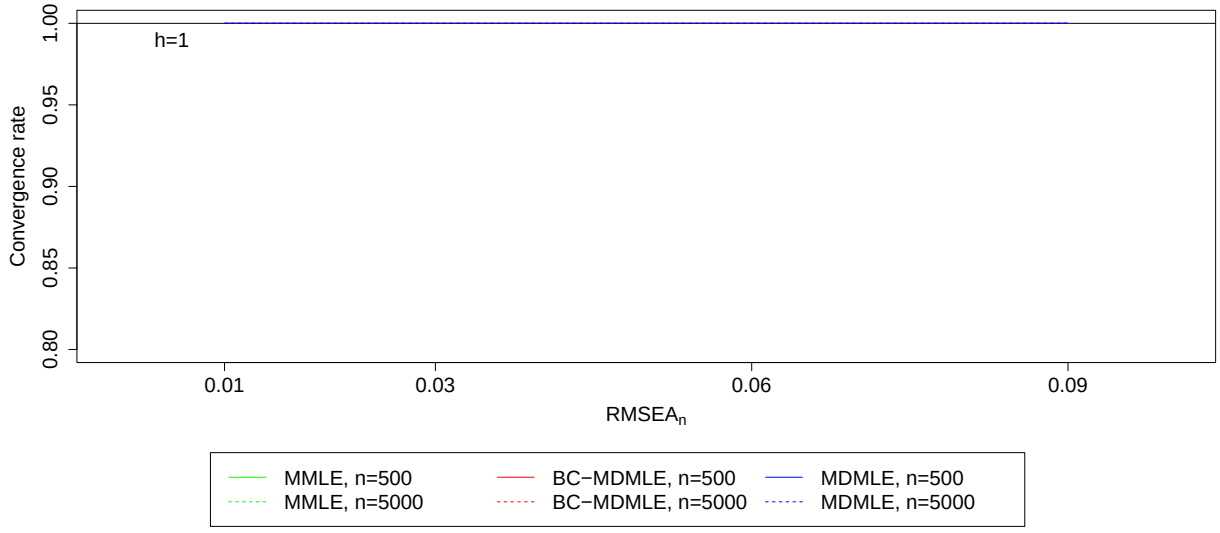


Figure 4.4

The Convergence Rate for Six Polytomous Items With Four Categories



4.4.2 Bias for Parameter Estimates

The bias in both item parameter estimates and dispersion parameter estimates is calculated to evaluate its recovery. For item parameter estimates, the bias is computed by subtracting the estimated parameter from the true parameter. As for dispersion parameter estimates, the relative bias is calculated by dividing the estimated parameter by the true parameter. This approach is chosen because some of the dispersion parameters are extremely close to 0, making the ratio of the estimated dispersion parameters over the true dispersion parameters more informative. In Section 4.4.2.1, the bias in the item parameter estimates is presented. Graphical representations are used to illustrate the bias patterns in the MMLE, as well as in the BC-MDMLE and the MDMLE. Section 4.4.2.2 focuses on the relative bias for the BC-MDMLE and MDMLE of the dispersion parameter.

4.4.2.1 Item Parameters

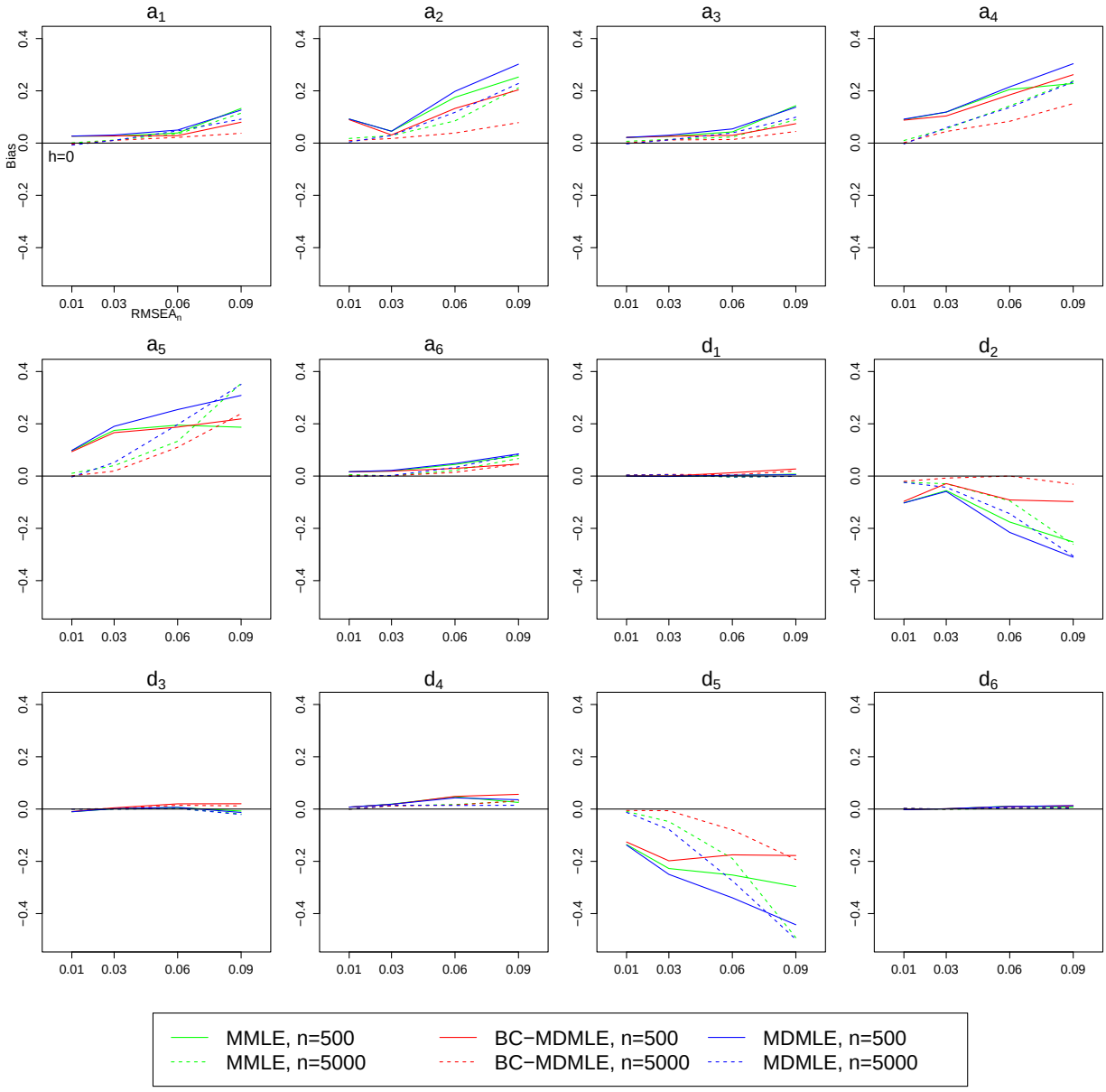
Figures 4.5 to 4.10 illustrate the bias observed in item parameter estimates across various simulation conditions. In these figures, the solid lines represent the parameter bias when the sample size is 500, while the dashed lines represent the bias for a sample size of 5000. The color scheme used is as follows: green corresponds to bias in the MMLE, red represents bias in the BC-MDMLE, and blue represents bias in the MDMLE. On the x-axis, you can find values such as 0.01, 0.03, 0.06, and 0.09, which indicate the levels of population RMSEA$_n$ used in the current study. Higher values of RMSEA$_n$ indicate larger model errors. The y-axis displays the bias between the estimated item parameter and the true item parameter. Each graph includes black lines that represent a bias of 0, indicating no bias on those lines. The closer the bias value

is to this black line, the smaller the parameter bias observed. These figures allow for a visual understanding of the bias patterns in item parameter estimates across the considered simulation conditions.

Figure 4.5 depicts the bias observed in item parameter estimates for a set of six items with two response categories. In the figure, the titles contain 'a's to denote the slope parameters and 'd's to represent the intercept parameters. For instance, a_1 refers to the slope parameter of Item 1, while d_2 represents the intercept parameter of Item 2. Considering there are six items with two response categories, a total of $6 \times 2 = 12$ item parameters are involved. Moving on, Figures 4.6 and 4.7 demonstrate the bias in item parameter estimates for a collection of 12 items with two response categories. With 12 items of this nature, there are $12 \times 2 = 24$ item parameters in total. Figure 4.6 shows the results for the 12 slope parameters, while Figure 4.7 focuses on the 12 intercept parameters. Figure 4.8 presents the bias results for three items with four response categories. In this scenario, encompassing three items of such nature, there are $3 \times 4 = 12$ item parameters in total. The notation for intercept parameters differs when the response categories equal four compared to when the item responses are dichotomous. For instance, the parameter d_{12} denotes the second intercept parameter for Item 1. Continuing further, Figures 4.9 and 4.10 exhibit the bias in item parameter estimates for six items with four response categories. Here, a total of $6 \times 4 = 24$ item parameters exist. Figure 4.9 illustrates the bias for the first 12 item parameter estimates, ranging from a_1 to d_{23}. On the other hand, Figure 4.10 showcases the bias for the remaining 12 item parameter estimates, spanning from d_{31} to d_{63}.

Figure 4.5

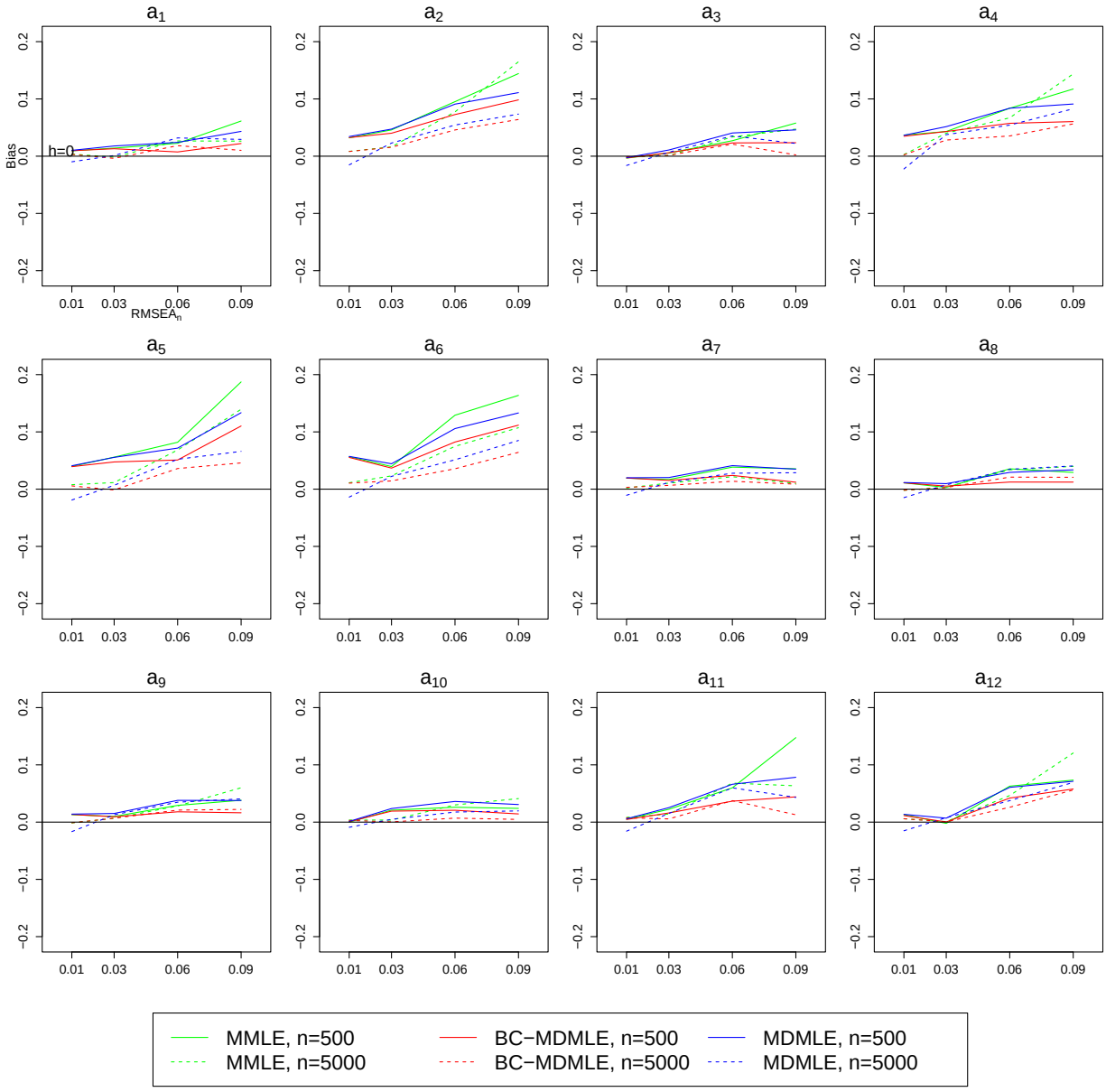
The Bias for Item Parameter Estimates for Six Dichtomous Items



Note. The a denote the slope parameters, and d denotes the intercept parameters.

Figure 4.6

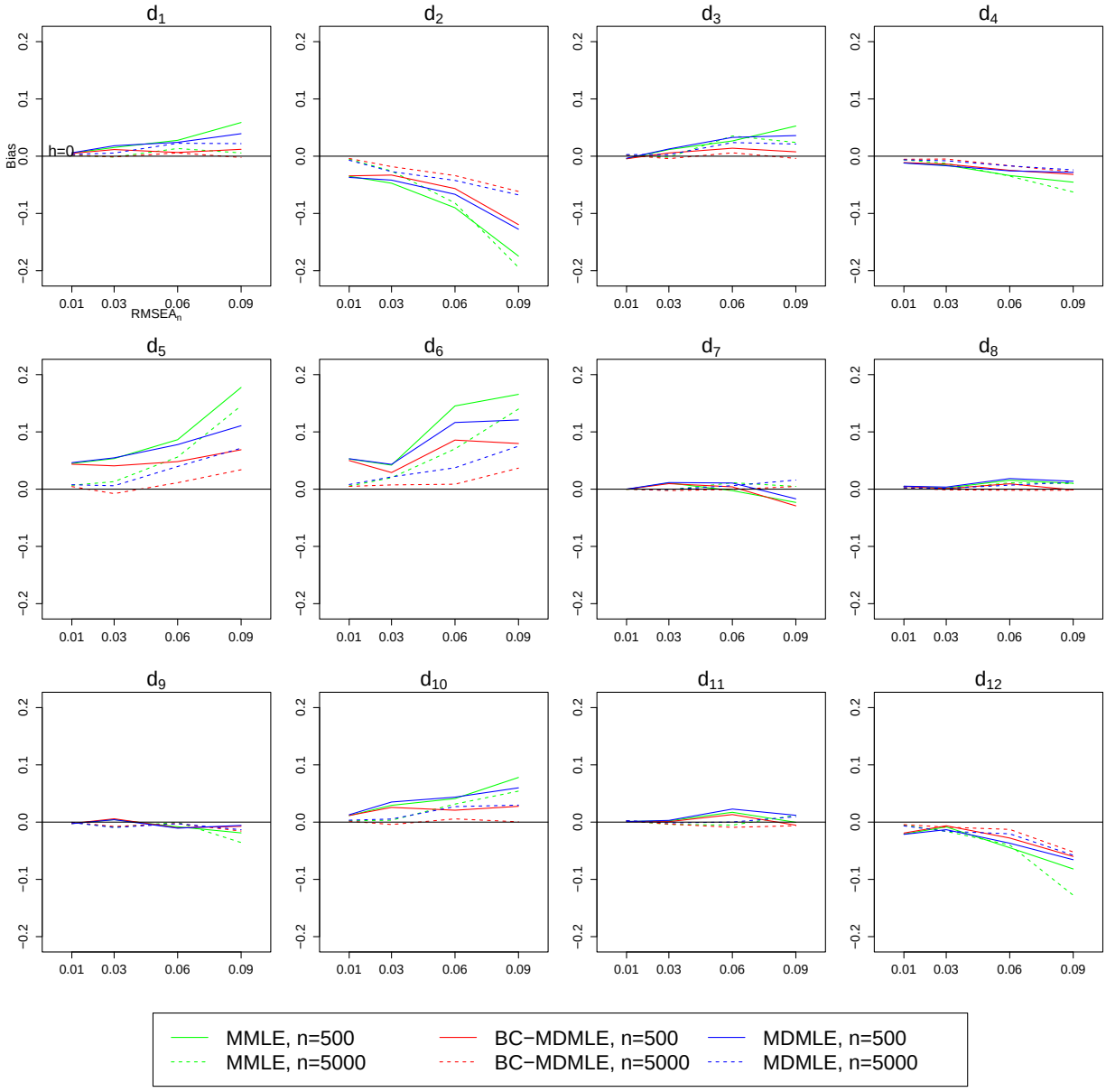
The Bias for Item Parameter Estimates for 12 Dichtomous Items - Only Slopes



Note. The a denote the slope parameters.

Figure 4.7

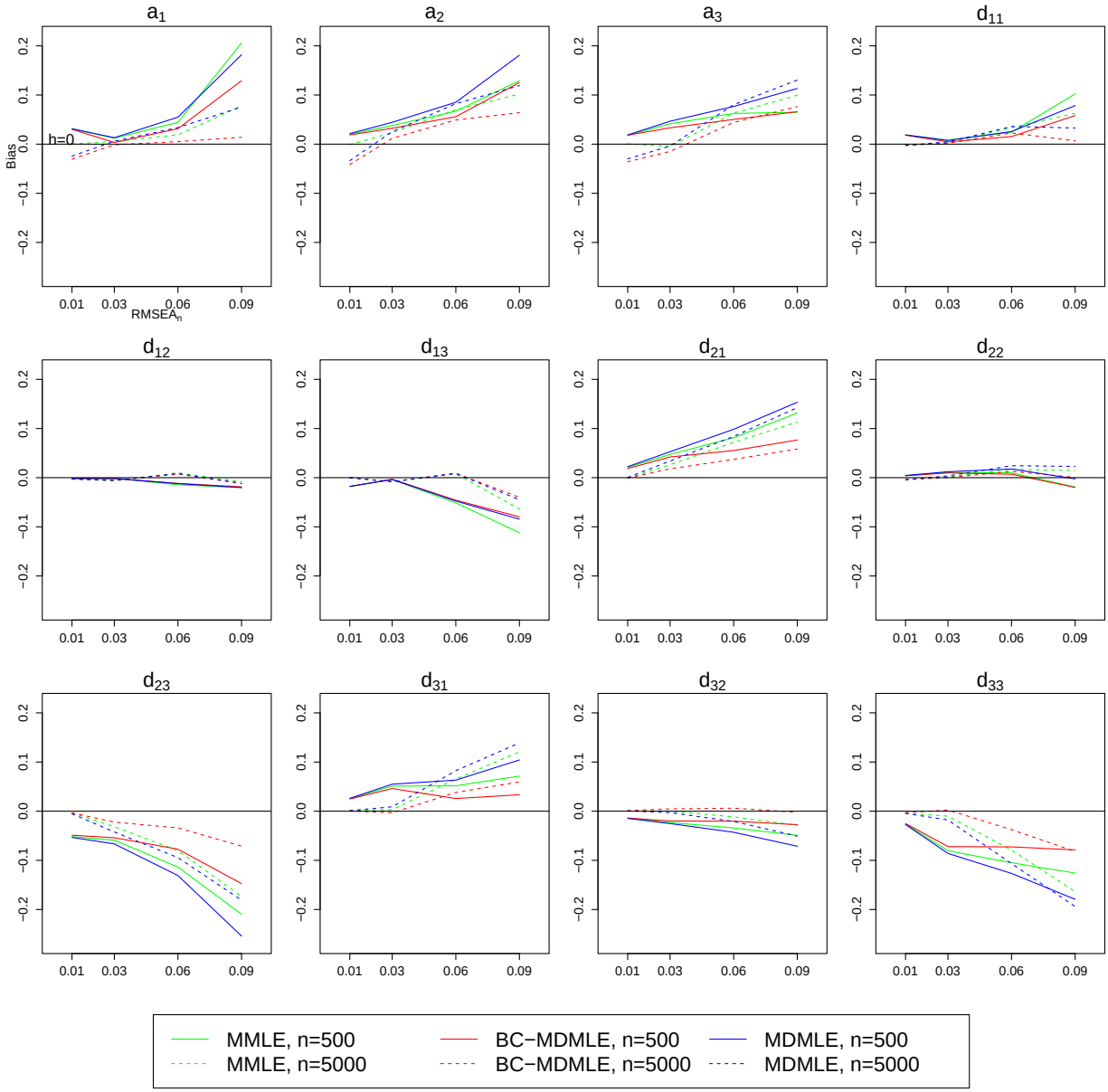
The Bias for Item Parameter Estimates for 12 Dichtomous Items - Only Intercepts



Note. The d denote the intercept parameters.

Figure 4.8

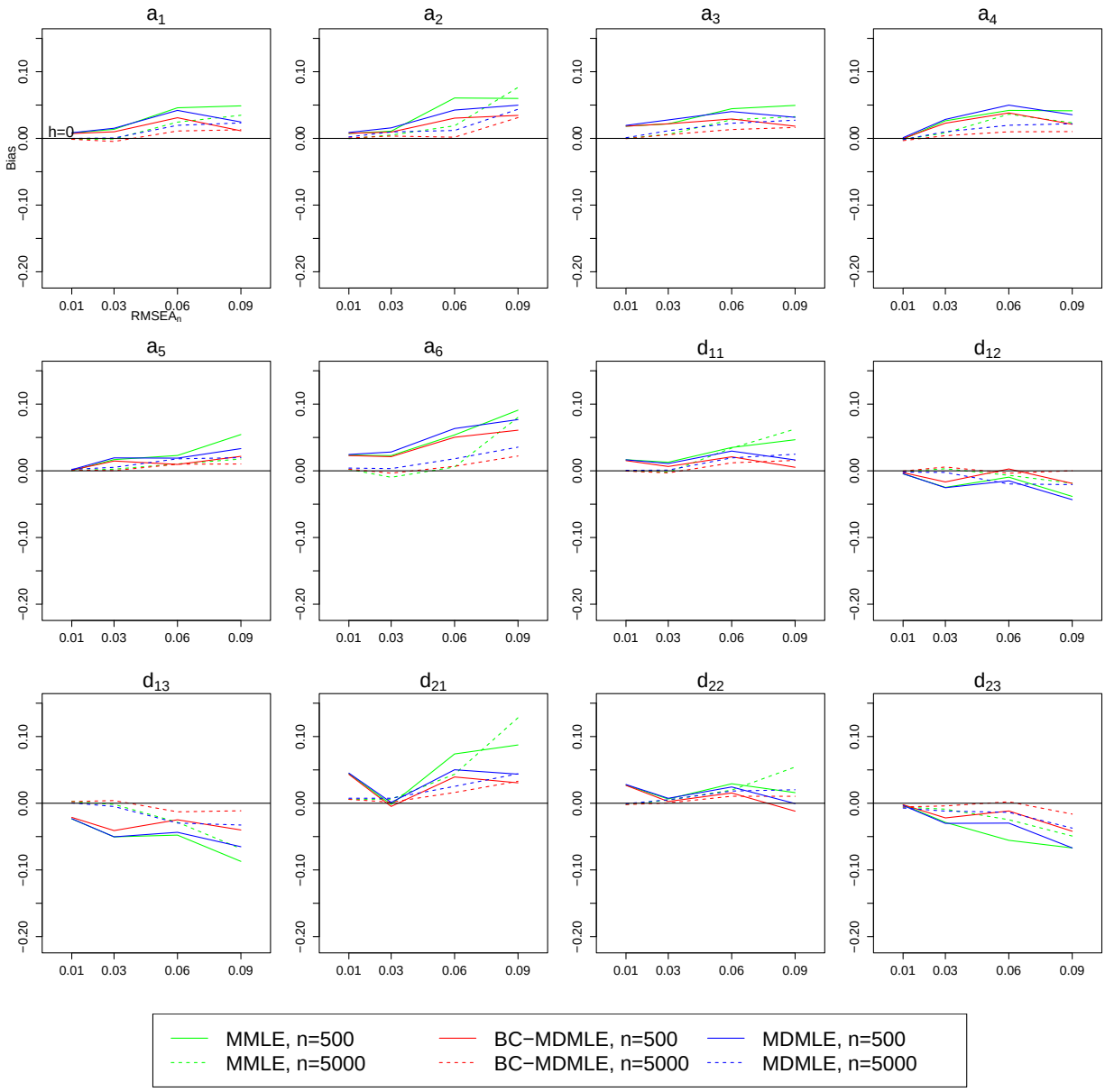
The Bias for Item Parameter Estimates for Three Polytomous Items With Four Categories



Note. The a denote the slope parameters, and d denotes the intercept parameters. d_{12} denotes the second intercept parameter for Item 1.

Figure 4.9

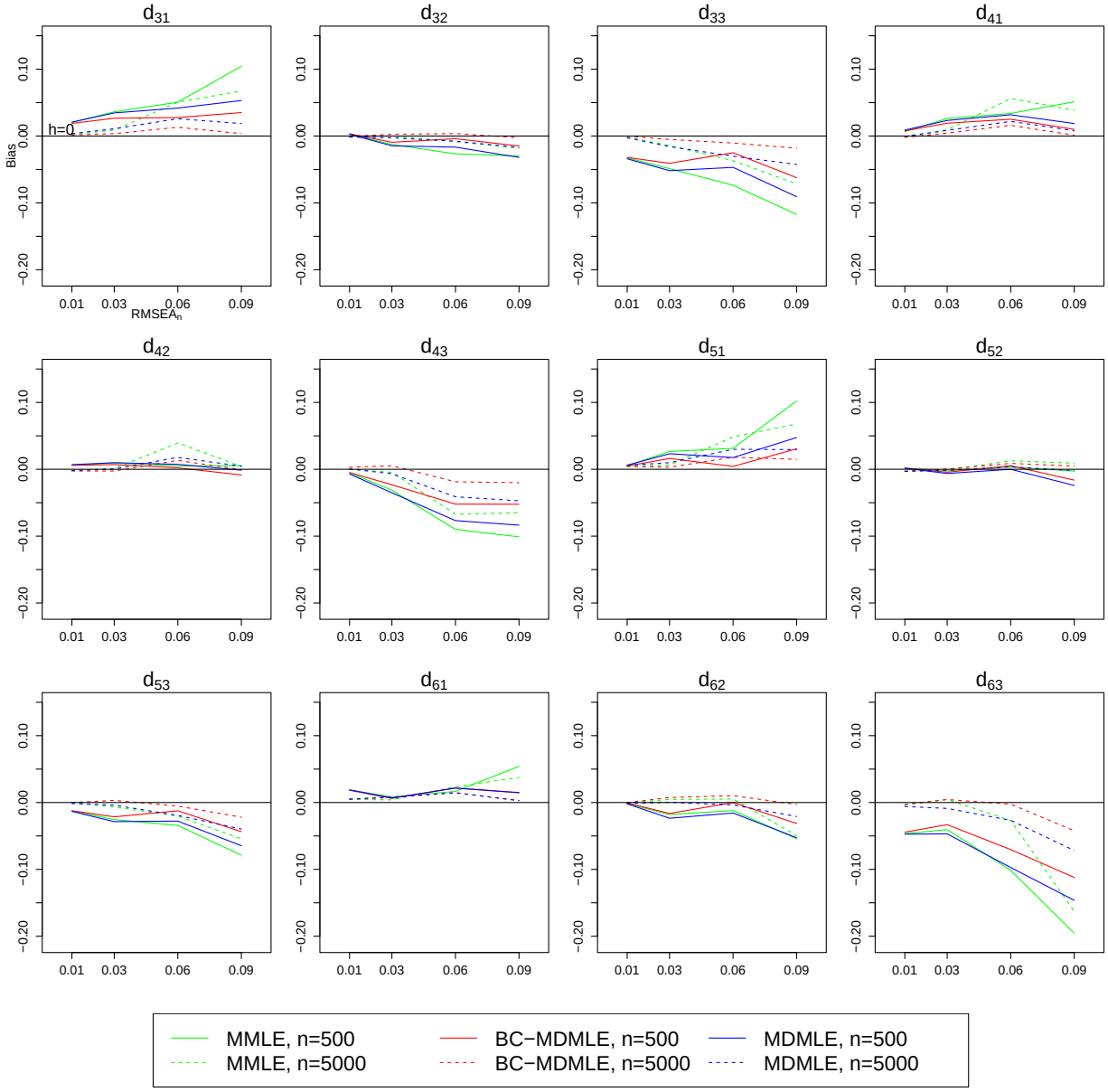
The Bias for Item Parameter Estimates for Six Polytomous Items With Four Categories - First 12 parameters



Note. The a denote the slope parameters, and d denotes the intercept parameters. d_{12} denotes the second intercept parameter for Item 1.

Figure 4.10

The Bias for Item Parameter Estimates for Six Polytomous Items With Four Categories - Other 12 parameters



Note. The a denote the slope parameters, and d denotes the intercept parameters. d_{12} denotes the second intercept parameter for Item 1.

The analysis of Figures 4.5 to 4.10 reveals a consistent pattern indicating that the bias in the BC-MDMLE is in general smaller compared to the MDMLE and the maximum multinomial likelihood estimates for 2PL and GRM models, across all simulation conditions. Further elaboration on these findings is provided below.

It has been observed that increasing the test length leads to improved performance of the BC-MDMLE, the MDMLE, as well as the MMLE, in terms of item parameter recovery. In Figure 4.5, for dichotomous items and a sample size of 500, it can be seen that the BC-MDMLE yields the smallest bias values for item parameters $a_1, a_2, a_3, a_4, a_5, a_6, d_2, d_5$ when compared to the MDMLE and the MMLE, as indicated by the solid lines. However, for the remaining item parameters, the BC-MDMLE sometimes performs similarly or even worse than the MDMLE and the MMLE. It is worth mentioning that the true values for the remaining item parameters are relatively close enough to 0 (see Table 4.1). This trend persists even when the sample size is increased to 5000. Nevertheless, the overall magnitude of bias tends to decrease with larger sample sizes.

By increasing the test length from six to 12, as depicted in Figures 4.6 and 4.7, it becomes evident that the bias in the BC-MDMLE is consistently smaller compared to the bias observed with the MDMLE and the bias in the MMLE. Similar patterns emerge when the items have polytomous responses with four categories. Regardless of test length, the BC-MDMLE consistently yields the smallest item parameter bias compared to the MDMLE and the MMLE.

Furthermore, a notable pattern is observed where larger sample sizes correspond to smaller bias. Whether the items are dichotomous or polytomous with response categories greater than two, the dashed lines (representing a sample size of 5000) tend to be closer to the 0 bias lines compared to their corresponding solid lines (representing a sample size of 500). In some instances,

particularly when the RMSEA$_n$ value is high (e.g., RMSEA$_n = 0.09$), it is possible for the bias to be smaller with smaller sample sizes. For instance, in Figure 4.5, when RMSEA$_n = 0.09$, the bias for the slope parameter of Item 5 is 0.22 with a sample size of 500, whereas with a sample size of 5000, the bias is 0.24.

In addition, when the RMSEA$_n$ is extremely small (i.e., 0.01), the bias among the three estimators is nearly identical. Overall, the bias remains small when the RMSEA$_n \leq 0.03$. It is observed that although there may be slight variations in bias in the BC-MDMLE, MDMLE, and in the MMLE, these differences become negligible when the RMSEA$_n$ is 0.01. Furthermore, the bias for all item parameters across all simulation conditions at RMSEA$_n = 0.01$ is very close to 0, indicating that when the model error is minimal, accounting for adventitious error becomes less critical as the bias for item parameters remains small. There are instances where the bias is smaller at RMSEA$_n = 0.03$ compared to RMSEA$_n = 0.01$, such as the bias for d_{21} in Figure 4.9. This pattern aligns with the recommendation proposed in Maydeu-Olivares and Joe (2014) that a close-fit cutoff value of 0.03 is suitable for unidimensional IRT models with binary data. Based on the findings of this study, it can be suggested that a close-fit cutoff value of 0.03 is also appropriate for unidimensional IRT models with polytomous data under the simulated conditions.

4.4.2.2 Dispersion Parameter

Figures 4.11 to 4.14 demonstrate the bias of dispersion parameter estimates under different simulation conditions. The conditions include six dichotomous items, 12 dichotomous items, three polytomous items with four categories, and six polytomous items with four categories, respectively. The relative bias is calculated as the ratio of estimated dispersion parameter to

the true dispersion parameter. A relative bias close to 1 indicates smaller bias in the dispersion parameter estimates.

In these figures, solid lines represent sample size of 500, while dashed lines represent a sample size of 5000. The red lines indicate the relative bias for the BC-MDMLE of the dispersion parameter, and the blue lines represent the relative bias for the MDMLE. The x-axis of these figures represents the RMSEA$_n$ conditions ranging from 0.01 to 0.09, while the y-axis represents the relative bias.

The findings support the theory that the MDMLE tends to underestimate the dispersion parameter; while in the meantime, the BC-MDMLE tends to slightly overestimate the dispersion parameter. The relative bias in the estimation of dispersion parameter is most prominent when the true dispersion parameter value, denoted as v_0, is small. For instance, when RMSEA$_n = 0.01$, the true value of v_0 is approximately 0.0001. In Figure 4.11, which corresponds to the case of six dichotomous items, with a sample size of 500, the BC-MDMLE is 2.94 times the true dispersion parameter, while in the MDMLE, it is 0.34 times the true value. However, as the true RMSEA$_n$ increases from 0.01 to 0.03, the relative bias in the BC-MDMLE decreases significantly. At a sample size of 500 when the RMSEA$_n = 0.03$, the BC-MDMLE is 1.16 times the true value and the MDMLE is 0.36 times the true value. Furthermore, it appears that the BC-MDMLE performs reasonably well in all simulation conditions except for when the sample size is 500 and RMSEA$_n = 0.01$.

A smaller RMSEA$_n$ suggests that there is a smaller model error, and the estimated parameters should be more accurate. However, this is true with respect to the estimated item parameters, but not necessarily the estimated dispersion parameters. It is due to the fact that when the population RMSEA$_n$ is smaller, by applying the connection between the RMSEA$_n$ and the dispersion

parameter v, the MDMLE and the BC-MDMLE of the dispersion parameter are also smaller. Extremely small dispersion parameters pose challenges for parameter recovery due to the sensitivity of small changes and numerical instability near the optimization boundary for the dispersion parameter (which is 0 in this case). The difficulty in recovering the dispersion parameter decreases as the true RMSEA$_n$ increases from 0.01 to 0.03, leading to a greater distance of the true dispersion parameter from the boundary (i.e., an increase in v_0 from 0.0001 to 0.0009).

Furthermore, it is evident that increasing the sample size leads to a reduction in bias, with the relative bias approaching 1. This trend can be observed in Figures 4.11 to 4.14, where a comparison is made between the solid lines (sample size of 500) and the corresponding dashed lines (sample size of 5000). In both the BC-MDMLE and MDMLE, a larger sample size consistently results in a smaller relative bias in the dispersion parameter estimates. This effect is particularly noticeable in the BC-MDMLE when the true RMSEA$_n$ is 0.01. Taking the bias for six dichotomous items in Figure 4.11 as an example, at RMSEA$_n = 0.01$, the estimated dispersion parameter is 2.94 times the true value when the sample size is 500. However, with other factors fixed, when the sample size is increased from 500 to 5000, the estimated dispersion parameter reduces to 1.05 times the true value. In the MDMLE, the pattern is not as evident as in the BC-MDMLE. This pattern, indicating that the increased sample size leads to a reduction in bias, is observed across other test length and number of response category conditions as well. Moreover, when the true RMSEA$_n$ is extremely small, the impact of increasing the sample size is more effective in reducing bias in the BC-MDMLE compared to the MDMLE.

Lastly, the test length plays a crucial role in the recovery of the dispersion parameter, particularly in the case of MDMLE. When considering a test with six dichotomous items (see Figure 4.11), for a sample size of 500, the relative bias for the MDMLE of dispersion param-

eter at RMSEA$_n$ $= 0.01, 0.03, 0.06, 0.09$ are $0.34, 0.36, 0.60, 0.65$ respectively. As the number of items increases to 12 dichotomous items (see Figure 4.12), while maintaining a sample size of 500, the corresponding relative bias for the dispersion parameter estimates improves to $0.75, 0.91, 0.94, 0.96$, which are closer to the desired value of 1. On the other hand, the test length does not significantly impact the parameter recovery for the BC-MDMLE. For instance, when the sample size is 500 and there are six dichotomous items in the test, the relative bias for the BC-MDMLE are $2.94, 1.16, 1.18, 1.21$. When the number of items is increased to 12, the corresponding relative bias become $2.07, 1.08, 1.10, 1.13$. These results suggest that the BC-MDMLE is less sensitive to changes in test length compared to the MDMLE in terms of the recovery of the dispersion parameter.

Figure 4.11

The Bias for Dispersion Parameter Estimates for Six Dichtomous Items

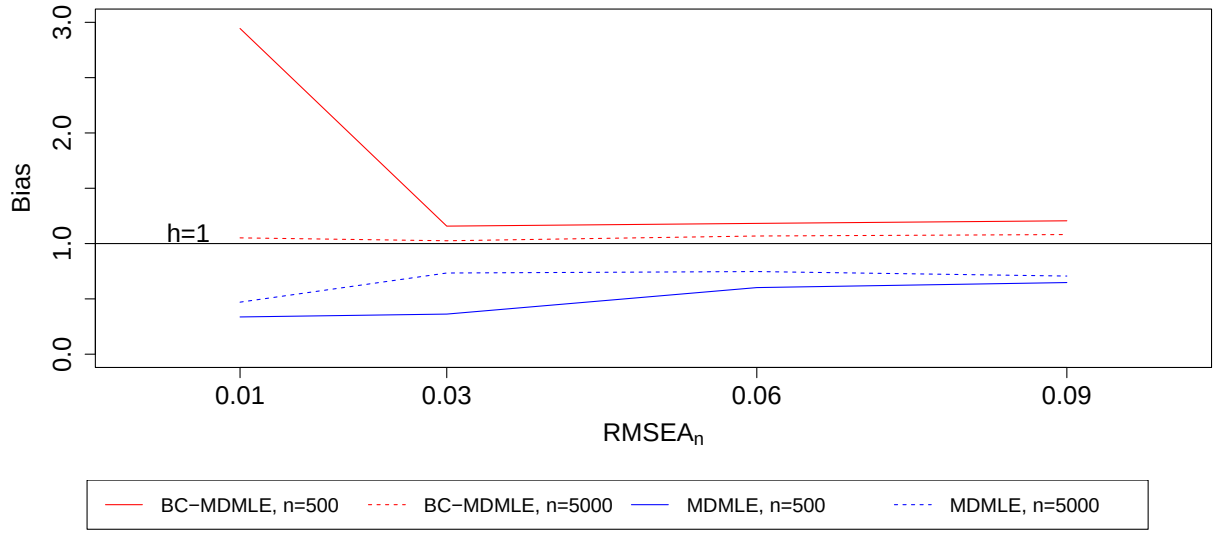


Figure 4.12

The Bias for Dispersion Parameter Estimates for 12 Dichtomous Items

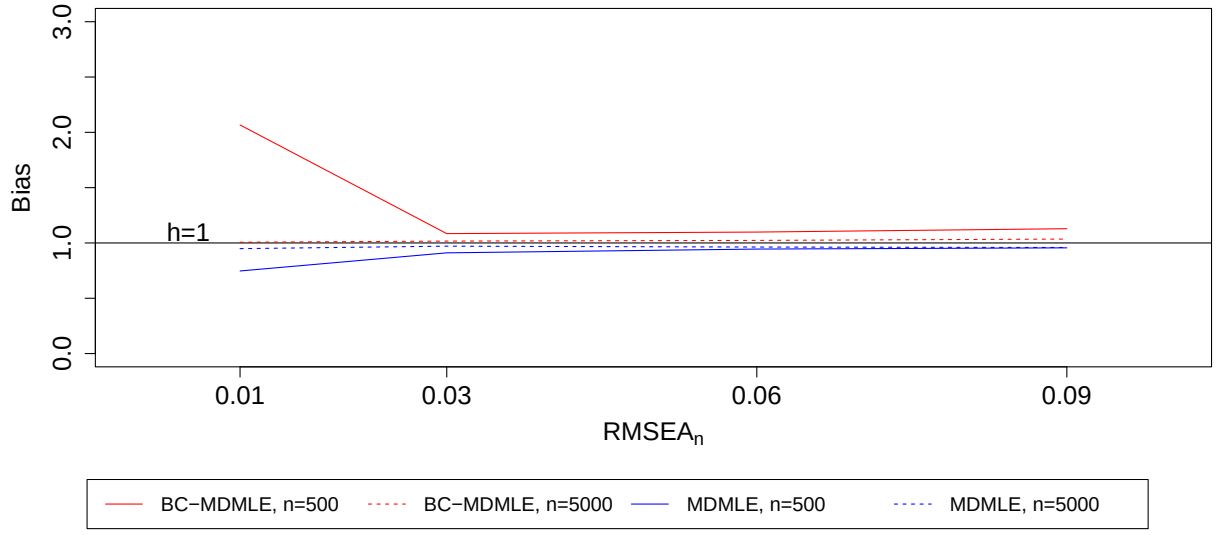


Figure 4.13

The Bias for Dispersion Parameter Estimates for Three Polytomous Items With Four Categories

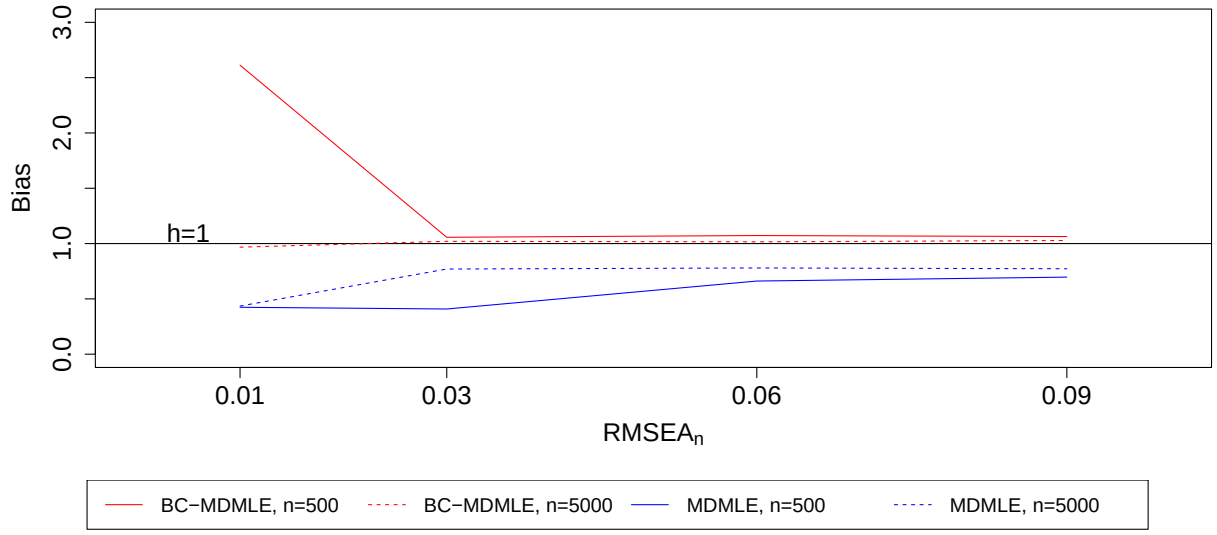
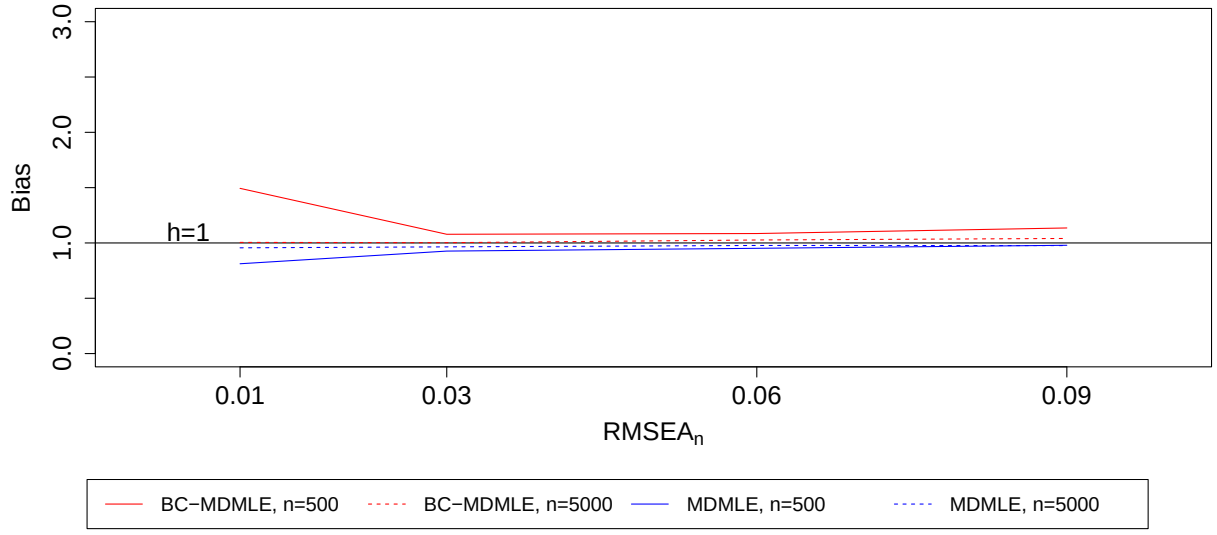


Figure 4.14

The Bias for Dispersion Parameter Estimates for Six Polytomous Items With Four Categories



4.4.3 RMSE for Parameter Estimates

To assess the accuracy of the estimated parameters compared to their true values, RMSE values are computed using the Equation 4.1. These RMSE values are calculated separately for the item parameter estimates and the dispersion parameter estimates. The RMSE serves as a quantitative measure to evaluate how closely the estimated parameters align with their true values. By calculating RMSE, we gain insights into the accuracy of the parameter estimates and assess the overall quality of the estimation process.

4.4.3.1 Item Parameters

Figures 4.15 to 4.20 present graphical representations of the RMSE observed in the estimation of item parameters under various simulation conditions. The solid lines in the figures depict the RMSE values when the sample size is 500, whereas the dashed lines represent the RMSE for a sample size of 5000. The color scheme used is as follows: green indicates the RMSE from the MMLE, red denotes the RMSE from the BC-MDMLE, and blue represents the RMSE from the MDMLE. The x-axis of these figures displays values such as 0.01, 0.03, 0.06, and 0.09, representing the levels of population RMSEA$_n$ utilized in the current study. Higher RMSEA$_n$ values indicate larger model errors. On the y-axis, the RMSE between the estimated item parameter and the true item parameter is depicted. These figures provide a visual representation of the patterns in RMSE observed in the estimation of item parameter estimates across the simulated conditions considered in this study.

There are three patterns observed in the results. Firstly, across all conditions, the RMSE from the MMLE is consistently larger compared to the RMSE from the BC-MDMLE and MDMLE.

When the test length is shorter (as depicted in Figures 4.15 and 4.18), for most items, the difference in RMSE values among these three estimators is not significant. However, as the test length increases (as observed in Figures 4.16, 4.17, 4.19, and 4.20), the difference becomes more evident. In general, the larger the population RMSEA$_n$, the greater the disparity in the RMSE values between the MMLE and the BC-MDMLE/MDMLE.

Secondly, regardless of the type of the estimator, there is a general trend of decreasing RMSE as the test length increases. Figure 4.15 reveals that when the test length is six, the RMSE can reach approximately 0.9 for certain parameters such as 'd_5'. However, when examining Figures 4.16 and 4.17, where the number of items increased from six to 12, the largest RMSE value approach is less than 0.6. It is essential to highlight that this observation only reflects a general pattern. For certain items, regardless of the estimator type, the RMSE values are already sufficiently low even with shorter tests.

Lastly, across all simulation conditions, there is a general trend of decreasing RMSE as the sample size increases. In other words, larger sample sizes tend to result in smaller RMSE values. This is true when the RMSEA$_n$ is not as large as 0.09. Across most items depicted in Figures 4.15 to 4.20, regardless of the type of estimator used, it was observed that the RMSE values were smaller when the sample size increased to 5000. However, certain patterns were also noticed where the RMSE values were larger with a sample size of 5000 compared to 500. For instance, this pattern was evident in items 'a_5' and 'd_5' at RMSEA$_n = 0.09$ in Figure 4.15.

Figure 4.15

The RMSE for Item Parameter With Six Dichotomous Items

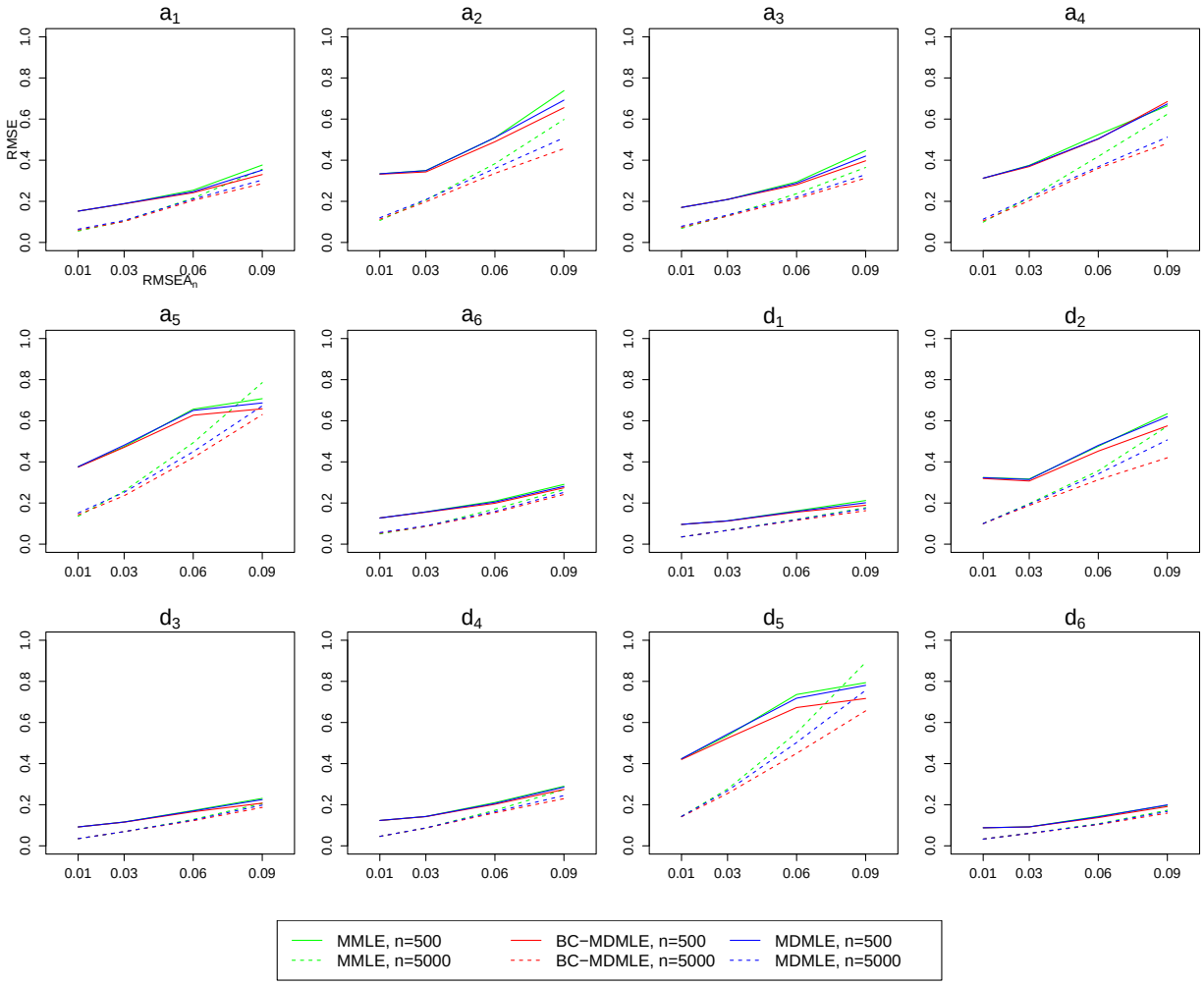


Figure 4.16

The RMSE for Item Parameter With 12 Dichotomous Items - Only Slopes

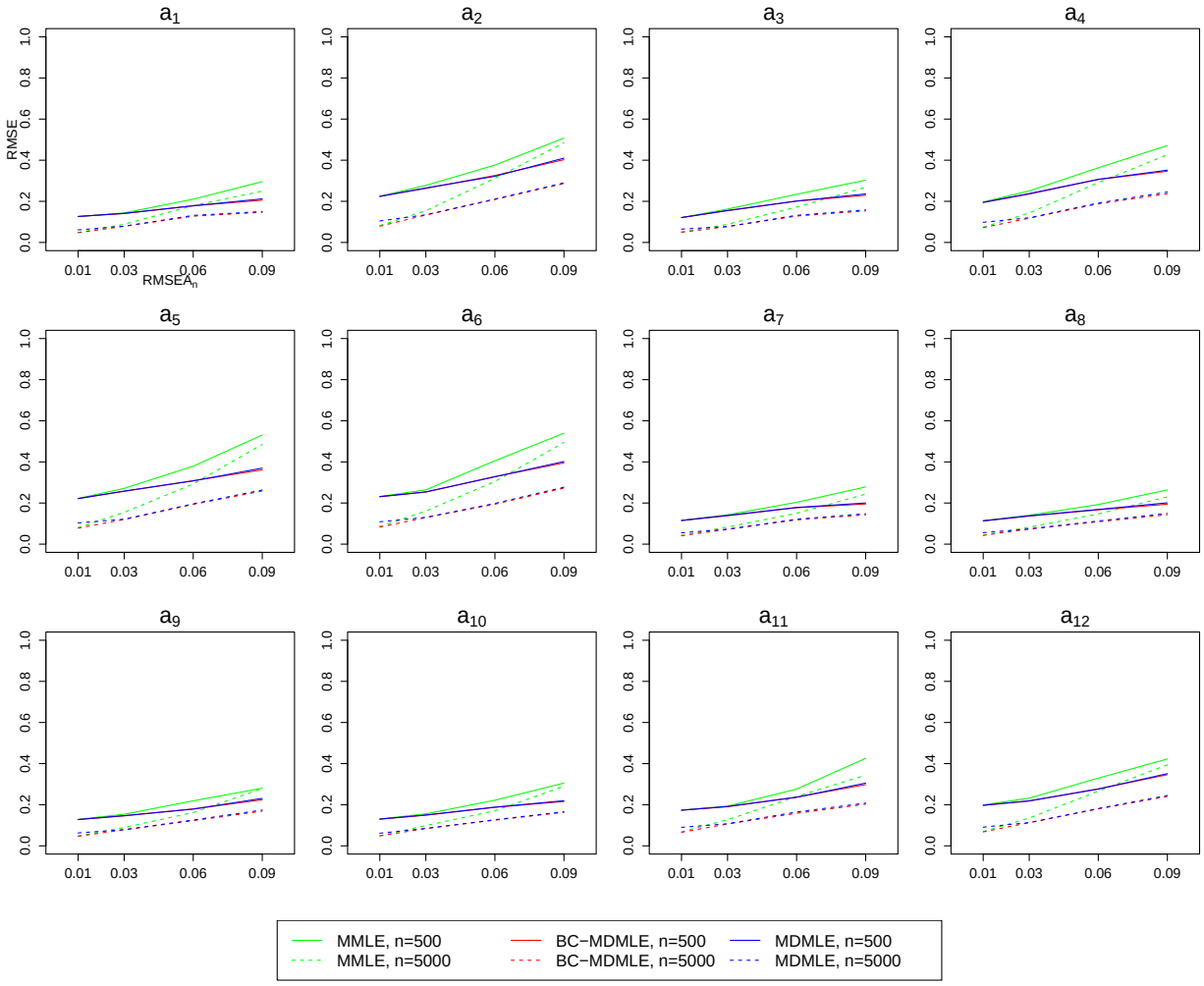


Figure 4.17

The RMSE for Item Parameter With 12 Dichotomous Items - Only Intercepts

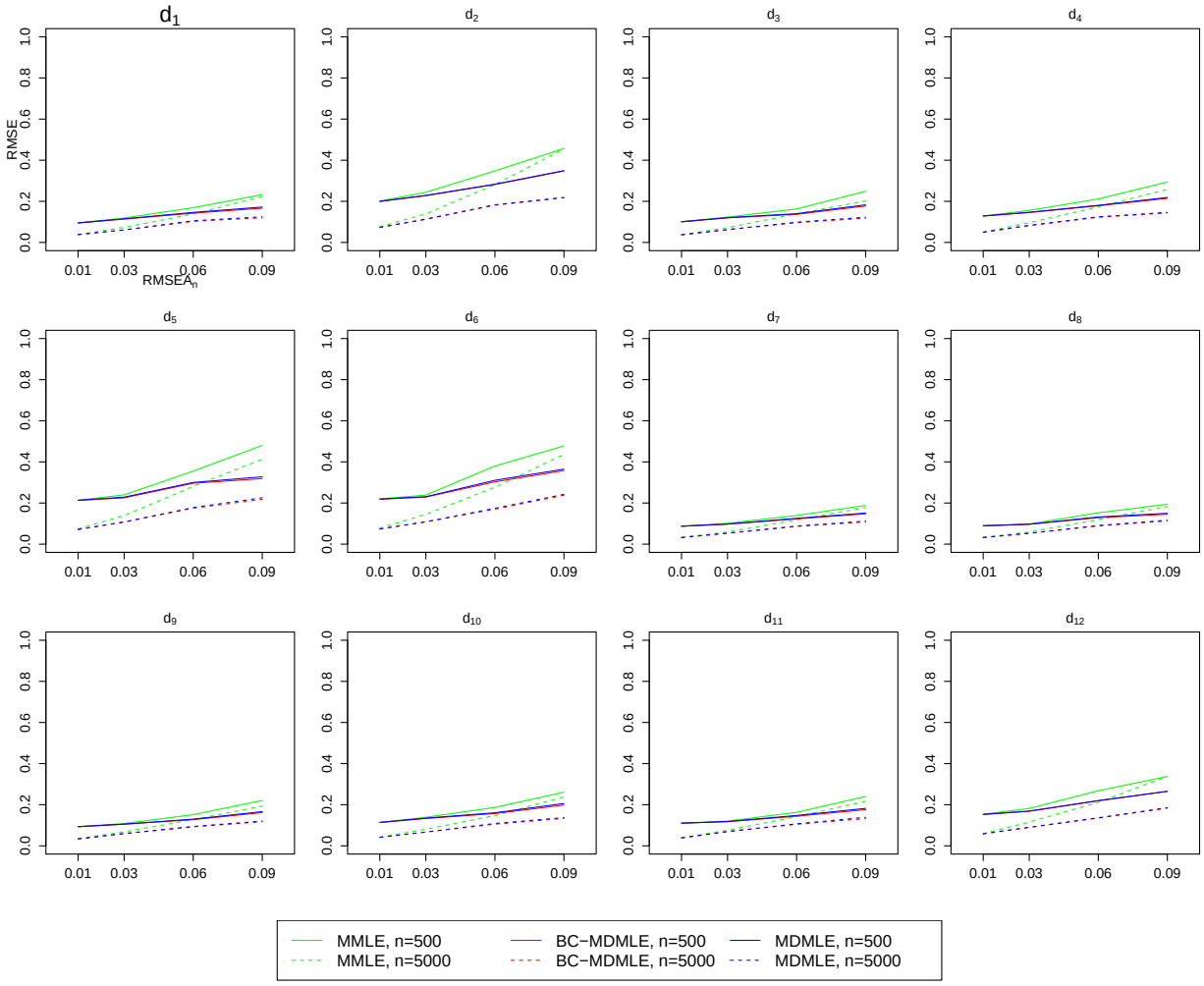


Figure 4.18

The RMSE for Item Parameter for Three Polytomous Items With Four Categories

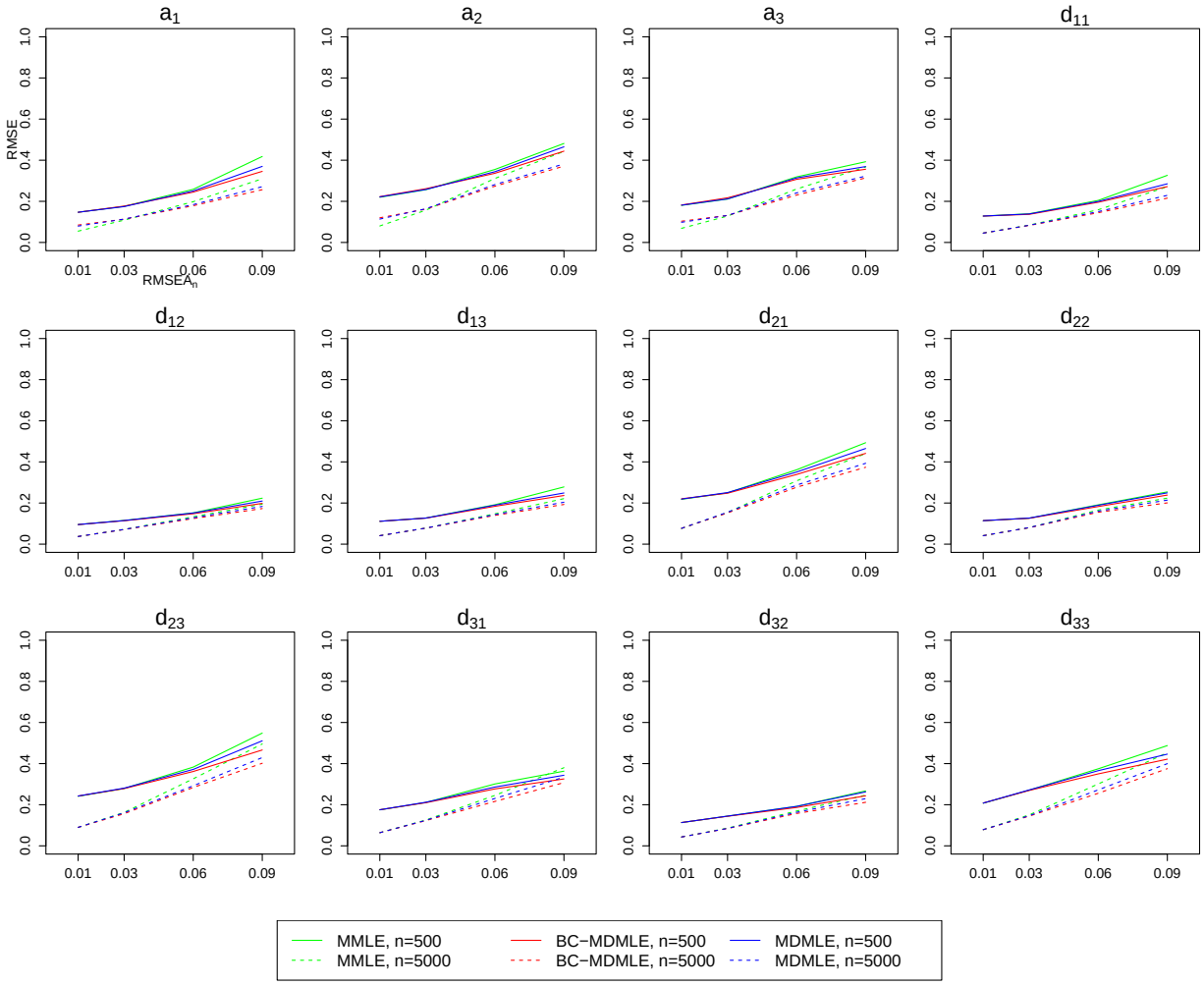


Figure 4.19

The RMSE for Item Parameter for Six Polytomous Items With Four Categories - First 12 parameters

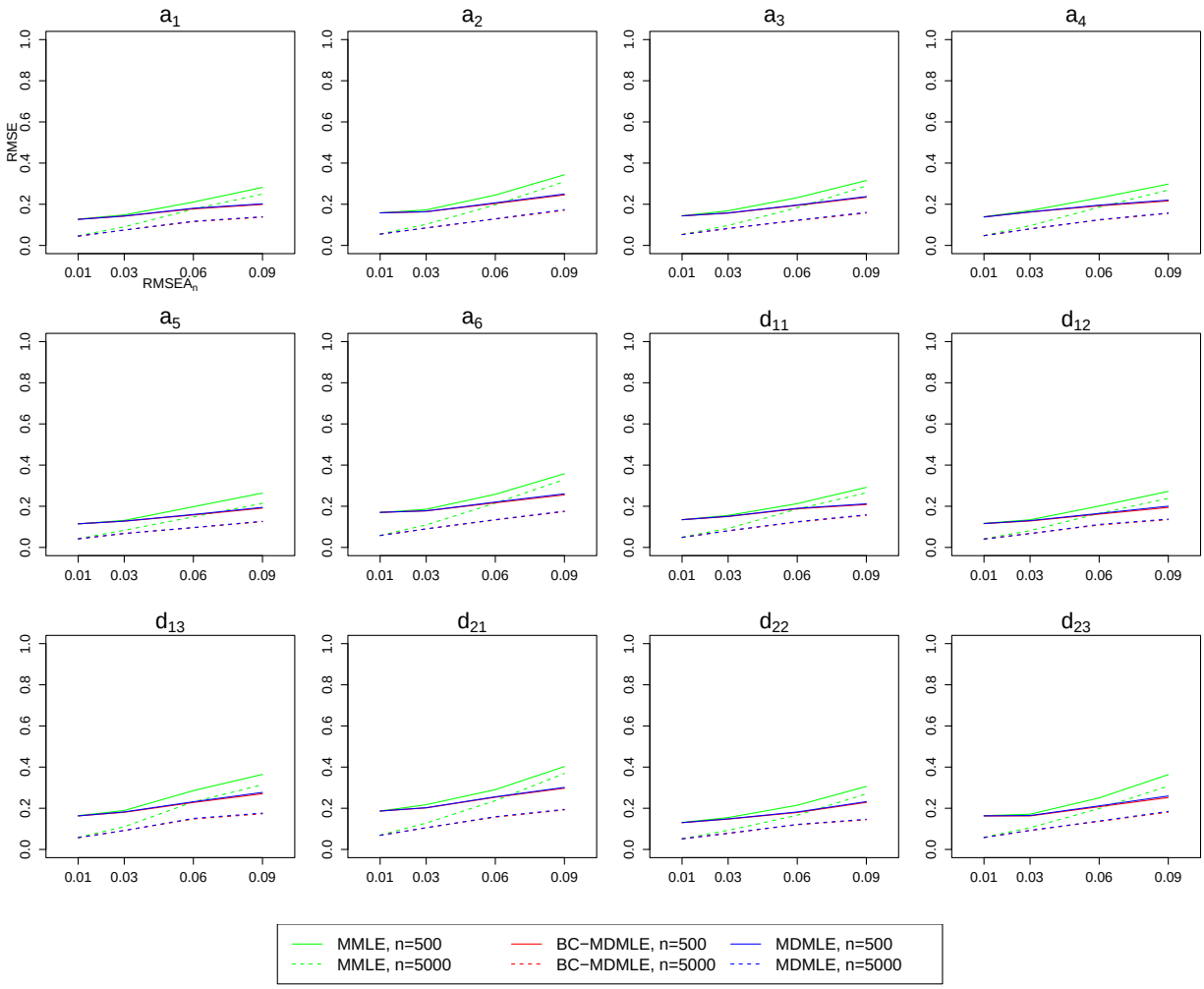
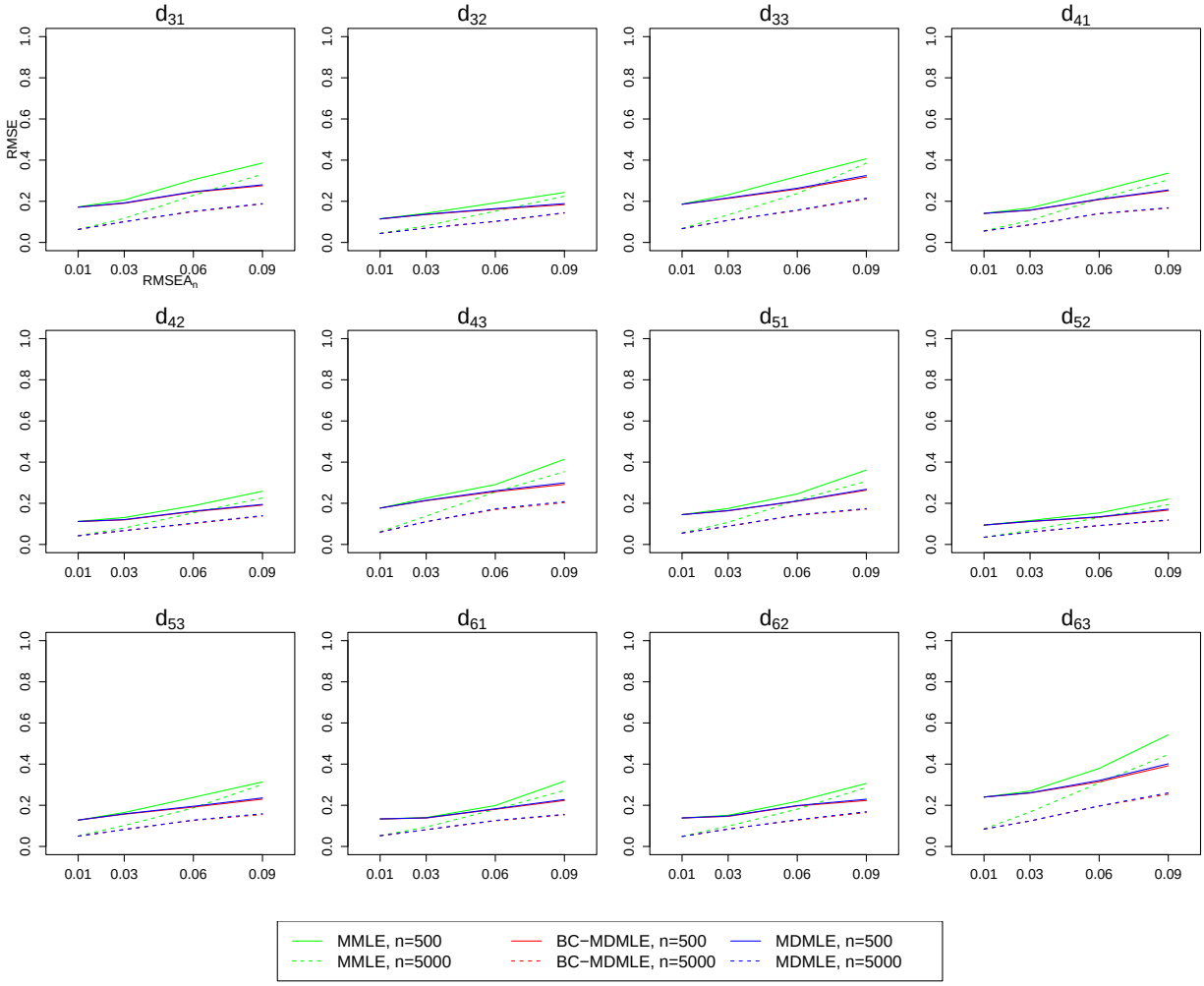


Figure 4.20

The RMSE for Item Parameter for Six Polytomous Items With Four Categories - Other 12 parameters



4.4.3.2 Dispersion Parameter

The RMSE analysis for the dispersion parameters is also conducted under different simulation conditions. These conditions include six dichotomous items, 12 dichotomous items, three polytomous items with four categories, and six polytomous items with four categories, respectively. Figures 4.21 to 4.24 depict these results, where solid lines correspond to a sample size of 500, while dashed lines represent a sample size of 5000. The red lines indicate the RMSE values from the BC-MDMLE of the dispersion parameter, while the blue lines represent the RMSE values from the MDMLE. The x-axis of these figures illustrates the range of RMSEA_n conditions, varying from 0.01 to 0.09, while the y-axis represents the corresponding RMSE values.

The analysis reveals that, across different levels of RMSEA_n, the RMSE values for the dispersion parameters do not consistently favor the BC-MDMLE over the MDMLE. When the sample size is 500, the RMSE from the BC-MDMLE is smaller than that from the MDMLE only when the test consists of three polytomous items with four categories (see Figure 4.23). In general, with a sample size of 500, the RMSE from the BC-MDMLE is slightly larger than that from the MDMLE. Conversely, when the sample size is 5000, the RMSE from the BC-MDMLE is mostly similar or slightly smaller than that from the MDMLE.

However, it is important to note that across all conditions, the RMSE values are very small, mostly below 0.004. This is due to the fact that the dispersion parameter itself is very small in the simulation study settings, with v_0 values of 0.0001, 0.0009, 0.0036, and 0.0081 corresponding to RMSEA_n values of 0.01, 0.03, 0.06, and 0.09, respectively.

Figure 4.21

The RMSE for Dispersion Parameter With Six Dichotomous Items

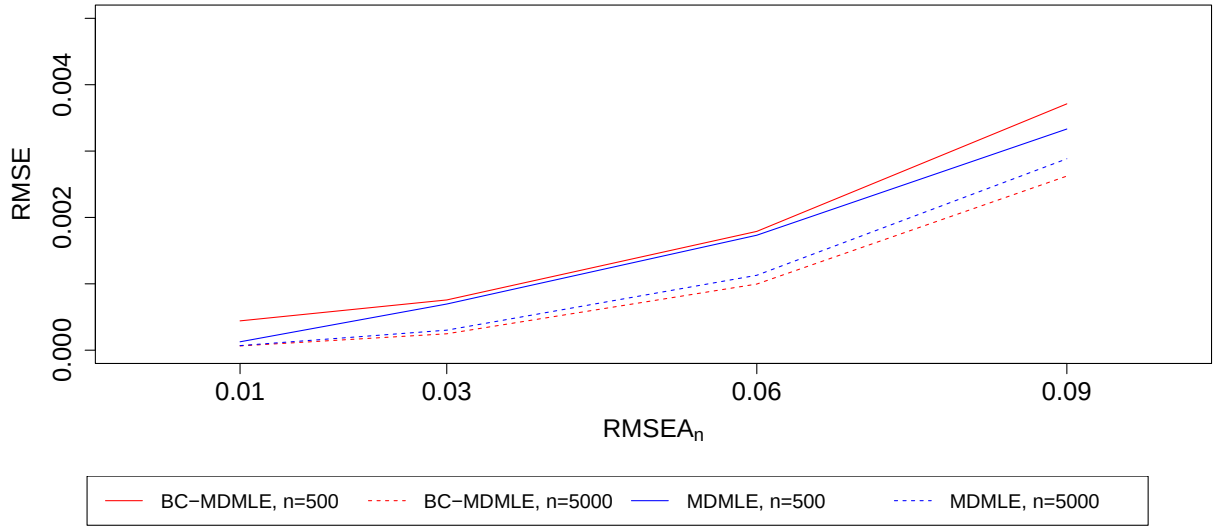


Figure 4.22

The RMSE for Dispersion Parameter With 12 Dichotomous Items

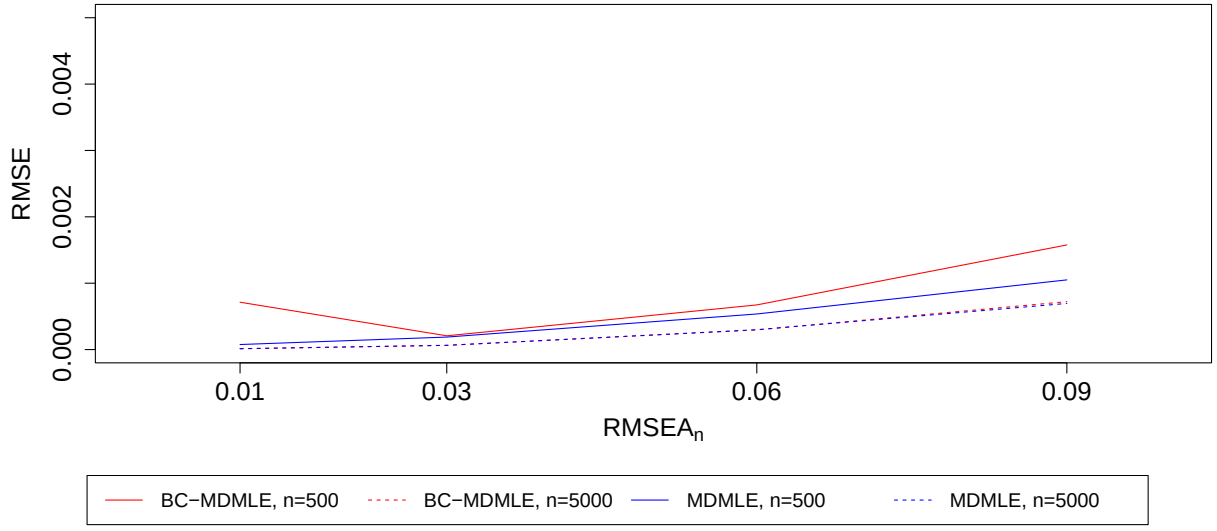


Figure 4.23

The RMSE for Dispersion Parameter for Three Polytomous Items With Four Categories

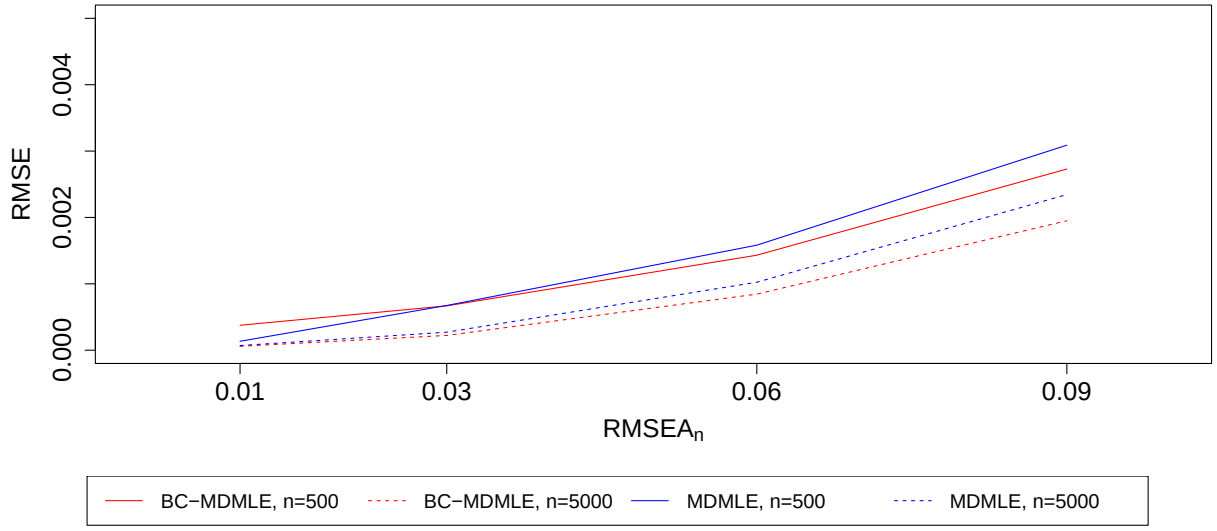
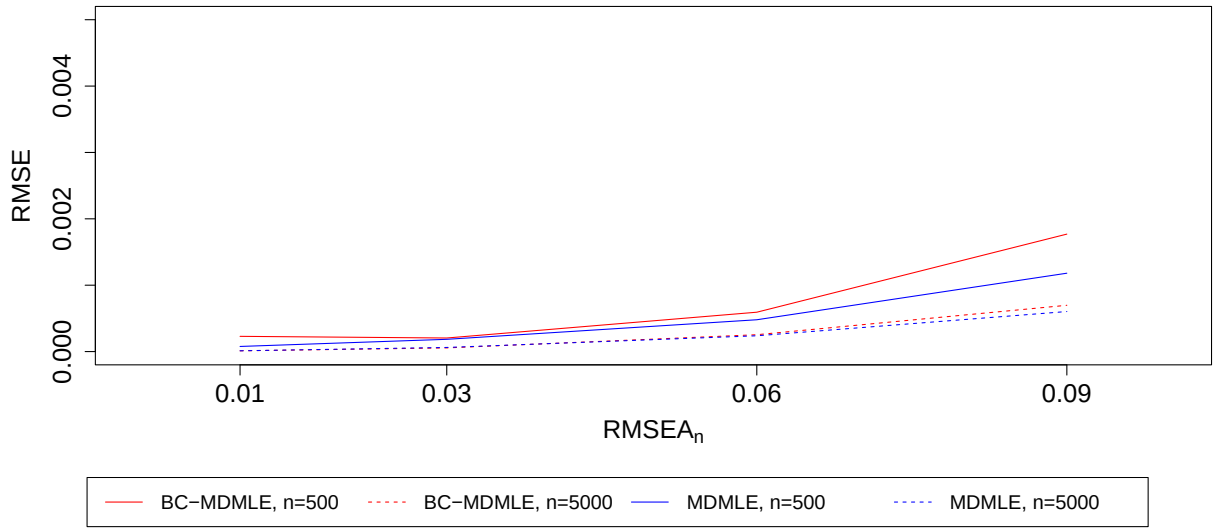


Figure 4.24

The RMSE for Dispersion Parameter for Six Polytomous Items With Four Categories



4.4.4 Empirical Coverage of the 95% Confidence Intervals

This section presents the empirical coverage of the 95% CI of both the item parameters and the dispersion parameter. The empirical coverage is determined by calculating the ratio of the number of times the 95% CI includes the true parameter over the total number of valid replications as indicated in Section 4.4.1.

4.4.4.1 Item Parameters

The empirical coverages of the 95% CIs are depicted in Figures 4.25 to 4.30. These figures illustrate different scenarios: six dichotomous items, 12 dichotomous items, three polytomous items with four categories, and six polytomous items with four categories. The solid lines represent a sample size of 500, while the dashed lines represent a sample size of 5000. To interpret the figures, the red lines indicate the empirical coverage of the 95% CI associated with the BC-MDMLE of item parameters, while the blue lines represent the MDMLE. Additionally, the orange lines depict the empirical coverage of the 95% CI associated with the MMLE with robust SEs, and the green lines correspond to the empirical coverage of the 95% CI associated with the MMLE with non-robust SEs. The x-axis of these figures represents the RMSEA_n ranging from 0.01 to 0.09. The y-axis represents the empirical coverage of the 95% CI for the item parameters.

Across various conditions involving different combinations of response categories and test lengths, the empirical coverage of the 95% CIs associated with the BC-MDMLE and MDMLE tend to be closer to 0.95, compared to the MMLE with either the robust or non-robust SEs. This trend holds true across different levels of RMSEA_n, as indicated by the proximity of the red and blue lines to the nominal level of 0.95 compared to the orange and green lines in the graphs.

As an example, for the BC-MDMLE, when the sample size is 500 in Figure 4.29, the empirical coverage of the 95% CI for the slope parameter of the first item (a_1) ranges from 0.960 to 0.994 as the RMSEA$_n$ increases from 0.01 to 0.09. With a sample size of 5000, the corresponding empirical coverage expands from 0.956 to 0.996. Similarly, for the MDMLE, the empirical coverage ranges from 0.958 to 0.992 with a sample size of 500, and from 0.956 to 0.996 with a sample size of 5000.

However, when examining the MMLE, the empirical coverage of the 95% CI becomes considerably less stable across different levels of RMSEA$_n$. The empirical coverage performs well at RMSEA$_n = 0.01$ but deteriorates rapidly as RMSEA$_n$ increases. Take the parameter a_1 in Figure 4.29 for example, when using the robust SEs, across the RMSEA$_n$ levels, the empirical coverage ranges from 0.952 to 0.634 for a sample size of 500, and from 0.886 to 0.242 for a sample size of 5000. Using the non-robust SEs yields an empirical coverage ranging from 0.946 to 0.622 for a sample size of 500, and from 0.888 to 0.238 for a sample size of 5000.

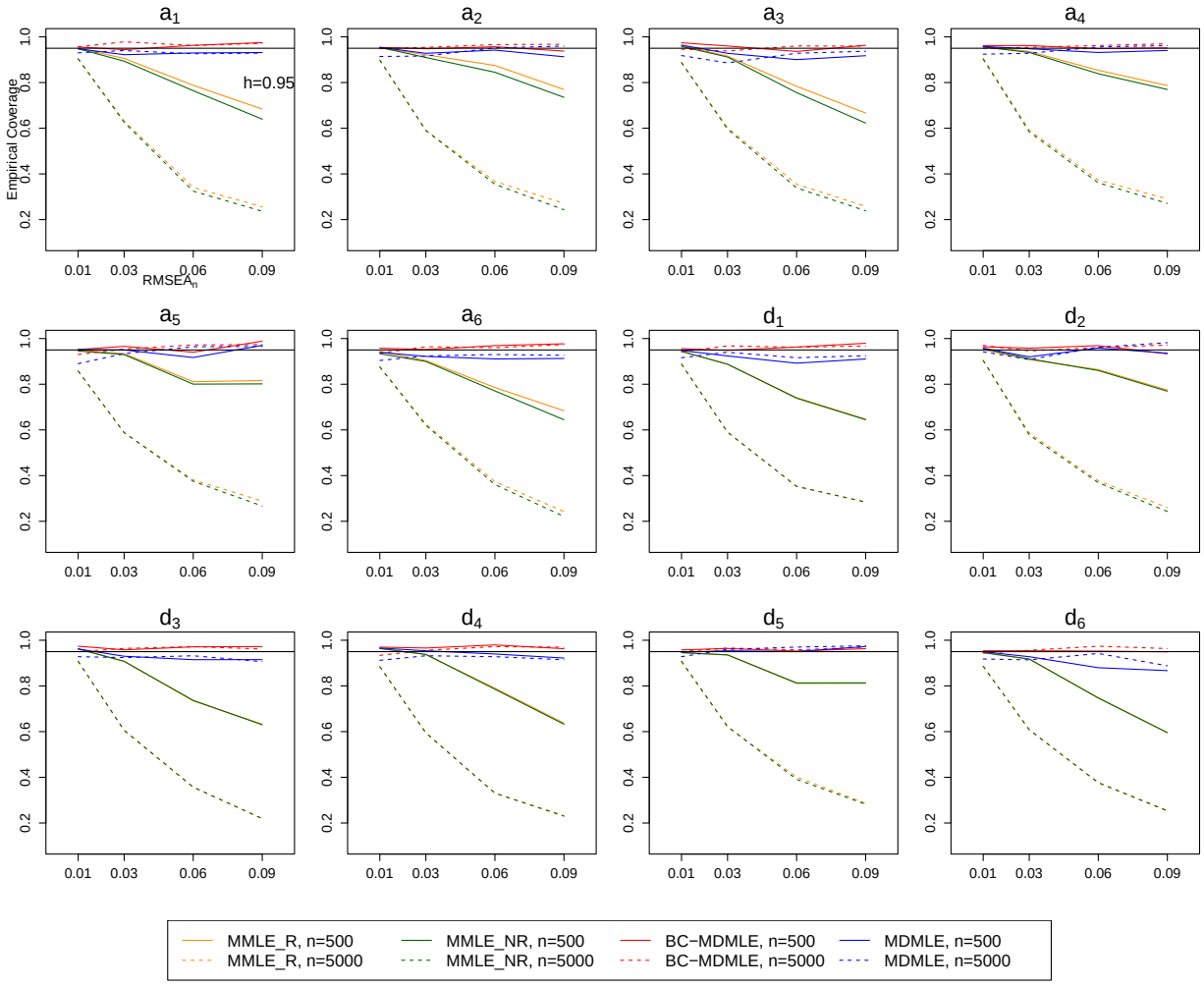
Furthermore, it is observed that a larger sample size leads to a poorer empirical coverage of the 95% CI associated with maximum multinomial likelihood estimation. For multinomial submodels, the empirical coverage is closer to 0.95 when the sample size is 500 compared to 5000. The disparity in the empirical coverage in smaller and larger sample sizes becomes more evident as RMSEA$_n$ grows from 0.01 to 0.09. Among all the examined conditions, the multinomial submodels utilizing the robust and non-robust SEs yield the worst empirical coverage when the sample size is 5000. However, regardless of the sample size, the empirical coverage results do not differ significantly between using the robust and non-robust SEs.

Moreover, another important observation is the relationship between the magnitude of the RMSEA$_n$ values and the empirical coverage. The empirical coverage gradually shifts away from

the desired 0.95 as the RMSEA_n values increase for longer tests (see Figures 4.26, 4.27, 4.29, and 4.30). When considering the BC-MDMLE and MDMLE, the empirical coverage shows a progressive increase from 0.95 towards 1. Notably, the empirical coverage is closest to 0.95 when the RMSEA_n value is at its lowest (i.e., 0.01). This indicates that the BC-MDMLE and MDMLE are more conservative in terms of achieving a empirical coverage closer to the desired level of 0.95, particularly when confronted with larger magnitudes of model error. In contrast, the MMLE, using either the robust or non-robust SEs, demonstrate a drastic decrease in the empirical coverage from 0.95 to approximately 0.2 as the RMSEA_n values increase from 0.01 to 0.09. This suggests that the traditional approaches are more liberal in terms of providing narrower confidence intervals that capture the true item parameters with a lower level of confidence.

Figure 4.25

The Item Parameter Empirical Coverage for Six Dichotomous Items



Note. Due to spacing constraints, the legends' text has been abbreviated. In the legends, 'MMLE_R' represents when the robust SEs were used in the multinomial submodels, while 'MMLE_NR' represents when the non-robust SEs were used in the multinomial submodels.

Figure 4.26

The Item Parameter Empirical Coverage for 12 Dichotomous Items - Only Slopes

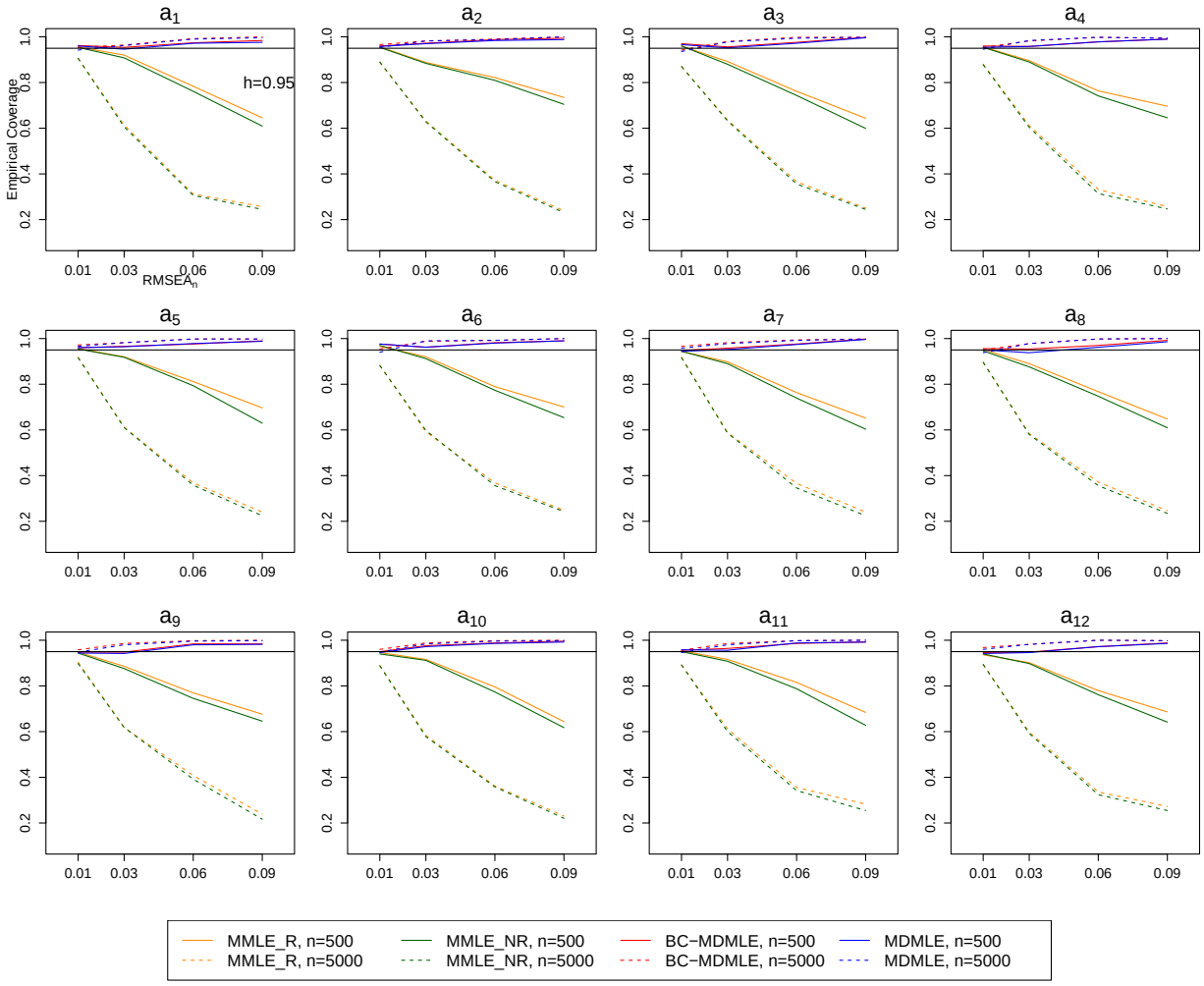


Figure 4.27

The Item Parameter Empirical Coverage for 12 Dichotomous Items - Only Intercepts

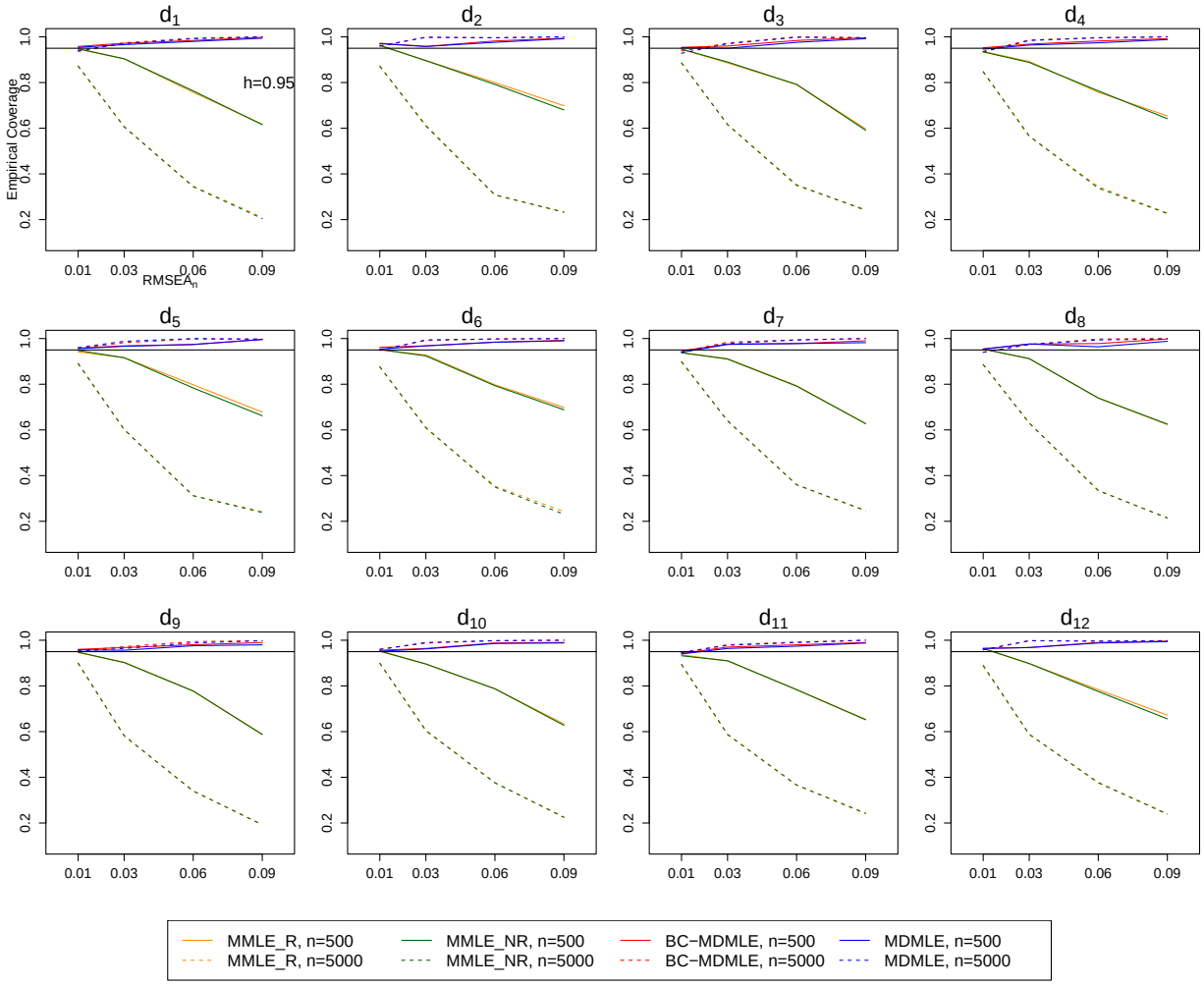


Figure 4.28

The Item Parameter Empirical Coverage for Three Polytomous Items With Four Categories

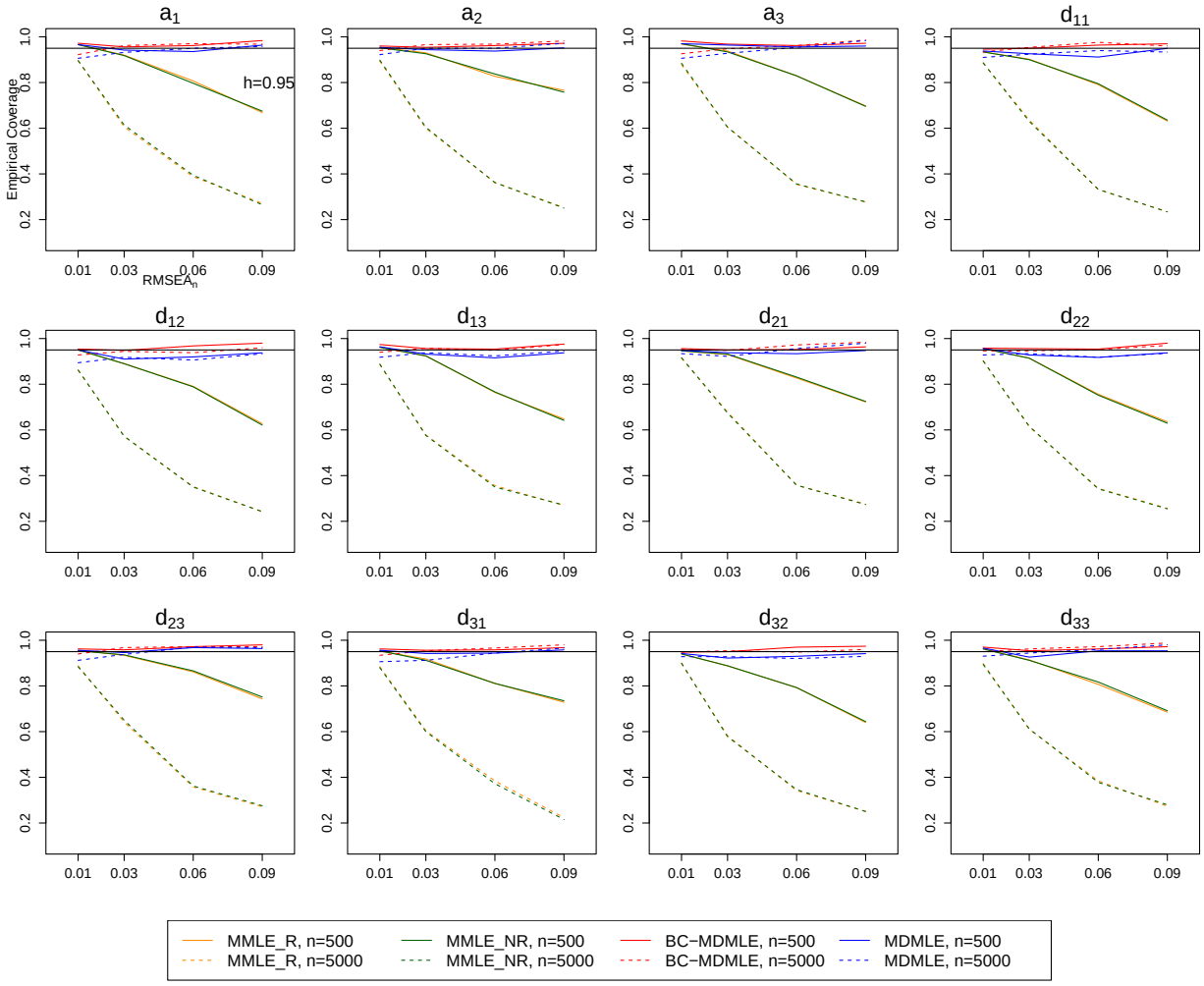


Figure 4.29

The Item Parameter Empirical Coverage for Six Polytomous Items With Four Categories - First 12 parameters

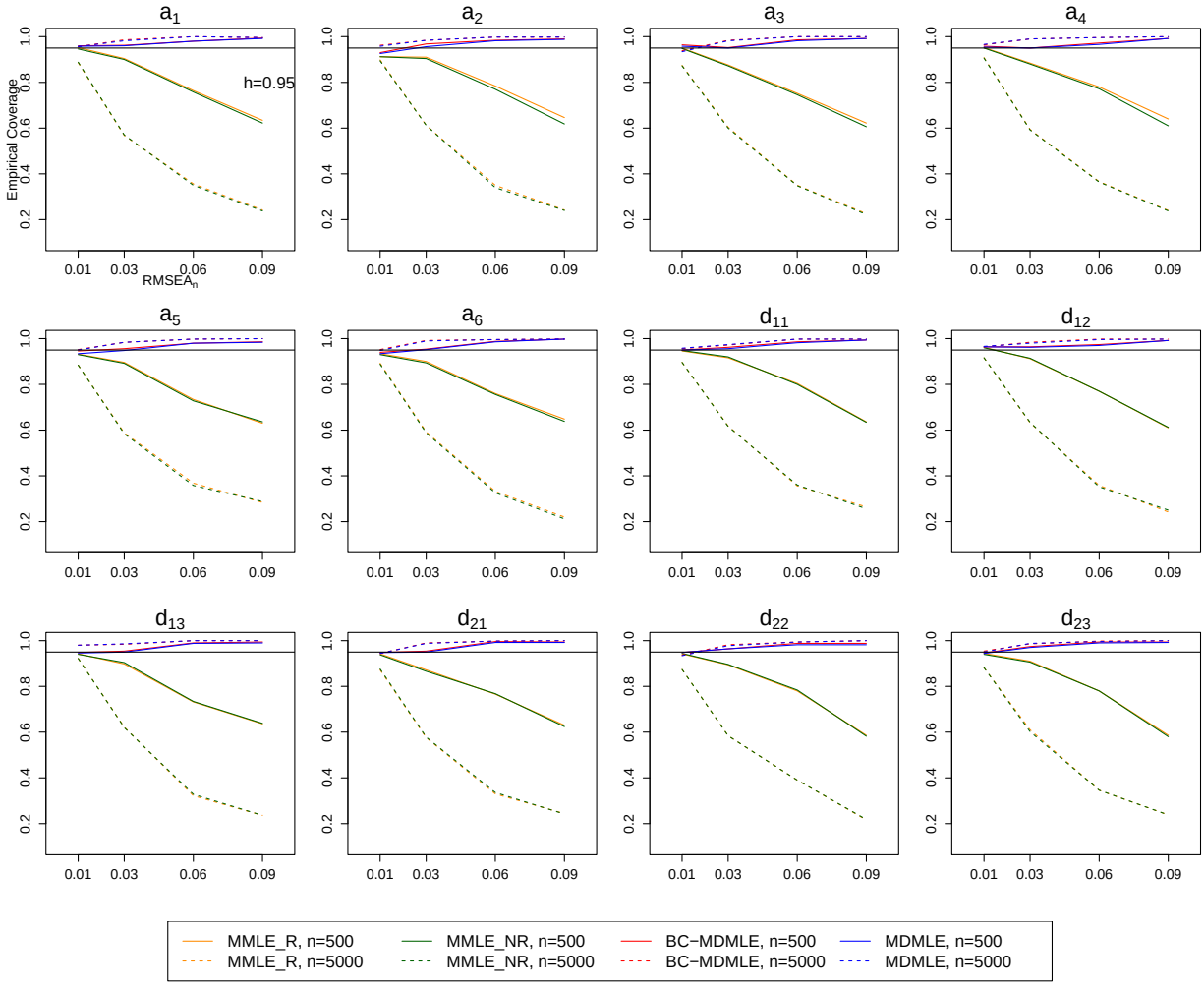
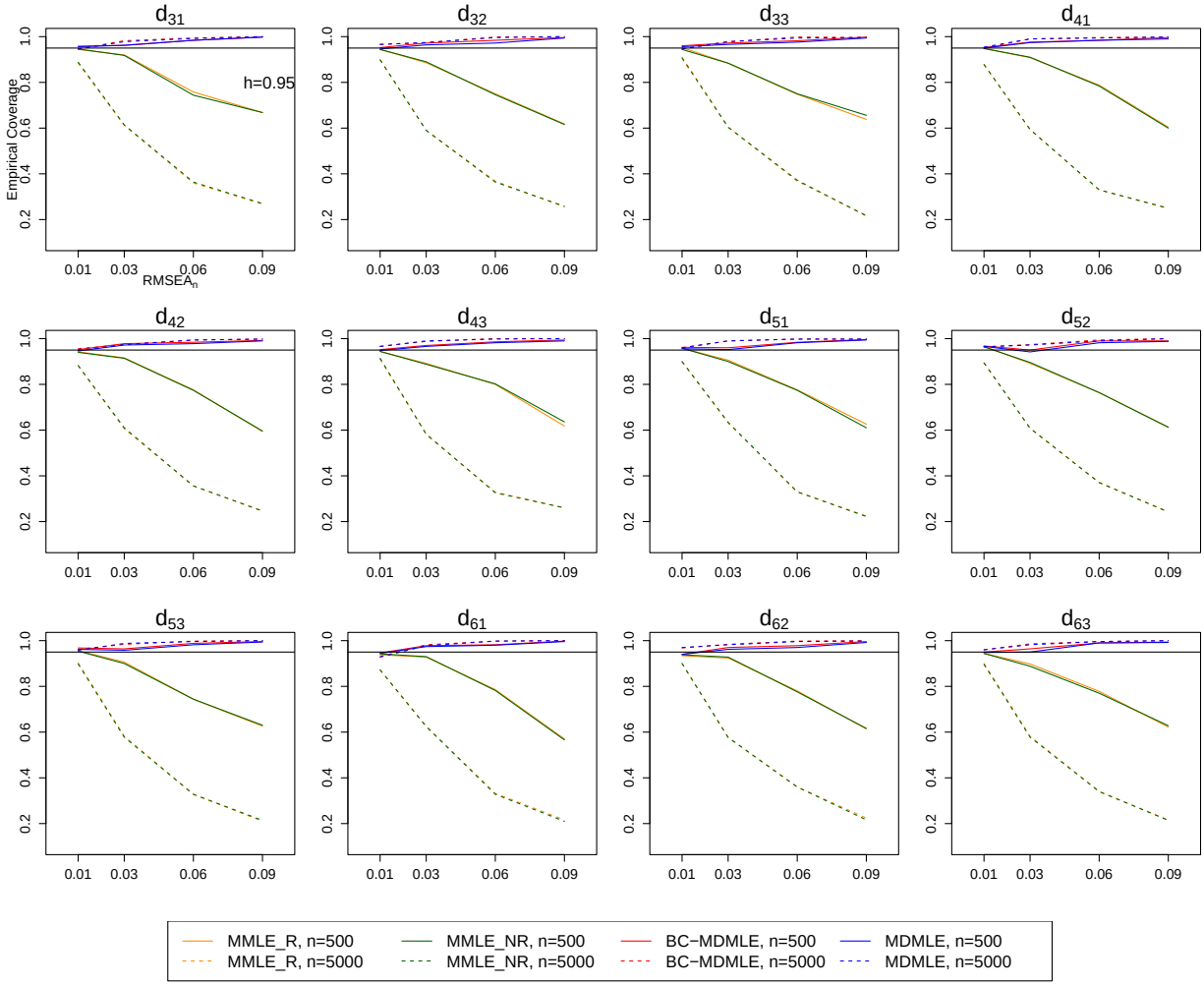


Figure 4.30

The Item Parameter Empirical Coverage for Six Polytomous Items With Four Categories - Other 12 parameters



4.4.4.2 Dispersion Parameter

The empirical coverage of the 95% CI for the dispersion parameter, associated with the BC-MDMLE and the MDMLE, is computed using different constant parameters. In the MDMLE, the approximation in Equation 3.28 reveals a bias when estimating $\hat{v}$ without the bias-correction. Consequently, the 95% CI for the dispersion parameter estimated using MDMLE is bounded by $\frac{(K-1)(\hat{v}+\epsilon)}{\chi^2_{K-q-1}(2.5\%)} - \epsilon$ and $\frac{(K-1)(\hat{v}+\epsilon)}{\chi^2_{K-q-1}(97.5\%)} - \epsilon$, with $K-1$ as the constant. On the other hand, in the BC-MDMLE, the bias in estimating the dispersion parameter is corrected (see Equation 3.35), leading to a 95% CI of the dispersion parameter bounded by the interval presented in Equation 3.40, with $K-q-1$ as the constant. In this context, $\hat{v}$ represents the MDMLE of the dispersion parameter, while $\hat{v}_0$ represents the BC-MDMLE of the dispersion parameter.

In general, the 95% CI for the dispersion parameter demonstrates a better empirical coverage when using the BC-MDMLE compared to the MDMLE, particularly when the test length is short. For instance, Figure 4.33 illustrates this trend for a test comprising three polytomous items with four categories. When the sample size is 500, the empirical coverages associated with the BC-MDMLE (indicated by the red solid line) are 0.958, 0.942, 0.889, and 0.856, while the empirical coverages associated with the MDMLE (represented by the blue solid line) are 0.998, 0.998, 0.696, and 0.631 for RMSEA$_n$ levels of 0.01, 0.03, 0.06, and 0.09, respectively. Similarly, when the sample size is 5000, the empirical coverages associated with the BC-MDMLE (indicated by the red dashed line) are 0.980, 0.948, 0.920, and 0.911, and the empirical coverages associated with the MDMLE (represented by the blue dashed line) are 0.766, 0.808, 0.712, and 0.661 for RMSEA$_n$ levels of 0.01, 0.03, 0.06, and 0.09, respectively.

However, this observed pattern does not hold when the test becomes longer. When the test

length is increased from three to six, examining Figure 4.34 reveals a noticeable trend: with a larger sample size of 5000, the 95% CI calculated from both the BC-MDMLE and MDMLE exhibit higher empirical coverages compared to a sample size of 500. While both the BC-MDMLE and MDMLE show better empirical coverages with larger sample sizes, the improvement is not necessarily substantial. Regardless of the sample size, the empirical coverages deviate significantly from the target of 0.95 when the test becomes longer. This pattern is also observed in the cases of tests with dichotomous items, as depicted in Figures 4.31 and 4.32.

The simulation study reveals that the empirical coverage of the 95% CI for the dispersion parameter does not perform as well as the empirical coverage of the 95% CI for the item parameters, associated with both the BC-MDMLE and the MDMLE. The primary issue lies in the sparseness of the contingency table. As discussed in Maydeu-Olivares and Joe (2005), asymptotic methods can only provide accurate approximations of the sampling distribution of RMSEA_n in small models when there is minimal sparseness in contingency tables. The large-sample CI for the dispersion parameter relies on a chi-square approximation with $df = K - q - 1$. A closer examination of the proof of Equation 3.24 reveals that the chi-square approximation is essentially the same as that for the Pearson's X^2 statistic. When the test is long, the total number of response patterns K tends to be large: for example, in a simulation scenario with two response categories for 12 items, the corresponding $df = 2^{12} - 2 \times 12 - 1 = 4071$. Even with the largest sample size in our simulations being 5000, it becomes evident that a significant number of cells in the contingency table would have expected counts lower than five. In the condition where the test length $J = 3$ and the number of response categories per item $C = 4$, the proportion of cells with expected counts below five is 0.55 for a sample size of 500, and 0.09 for a sample size of 5000. However, when $J = 6$ and $C = 4$, the proportion of cells with expected counts below five is 1

for a sample size of 500, and 0.94 for a sample size of 5000.

In order to validate the theory regarding the sparseness issue, an exceptionally large sample size of 500,000 is considered. The empirical coverage of the 95% CI for the dispersion parameter is then computed for both the BC-MDMLE and MDMLE using this increased sample size. We expect that the empirical coverage should be closer to the nominal level of 0.95 when employing this larger sample size.

Figure 4.31

The Empirical Coverage for Dispersion Parameter With Six Dichotomous Items

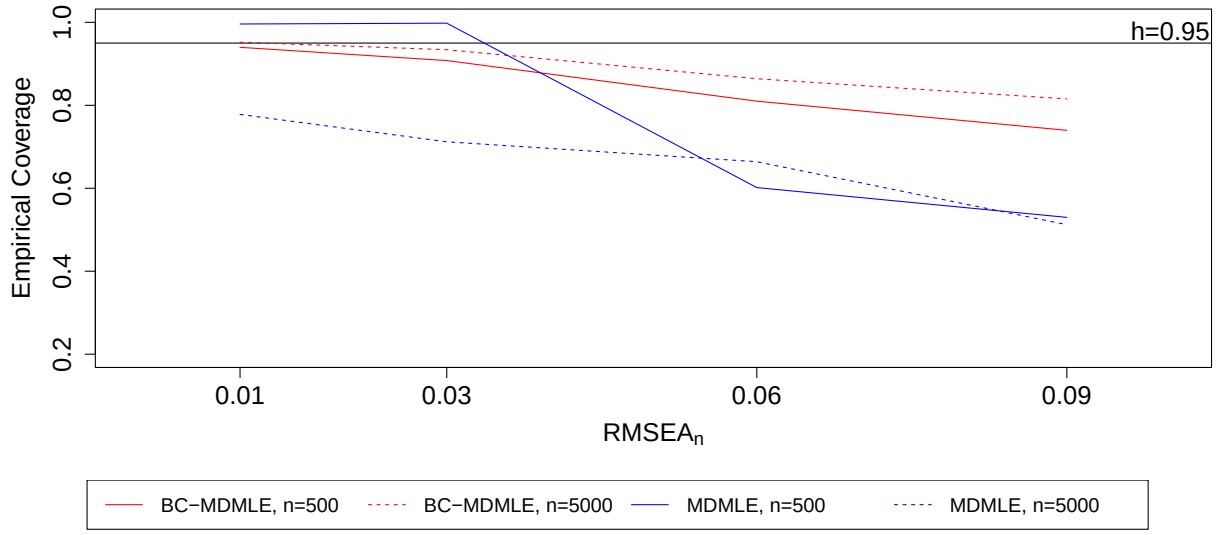


Figure 4.32

The Empirical Coverage for Dispersion Parameter With 12 Dichotomous Items

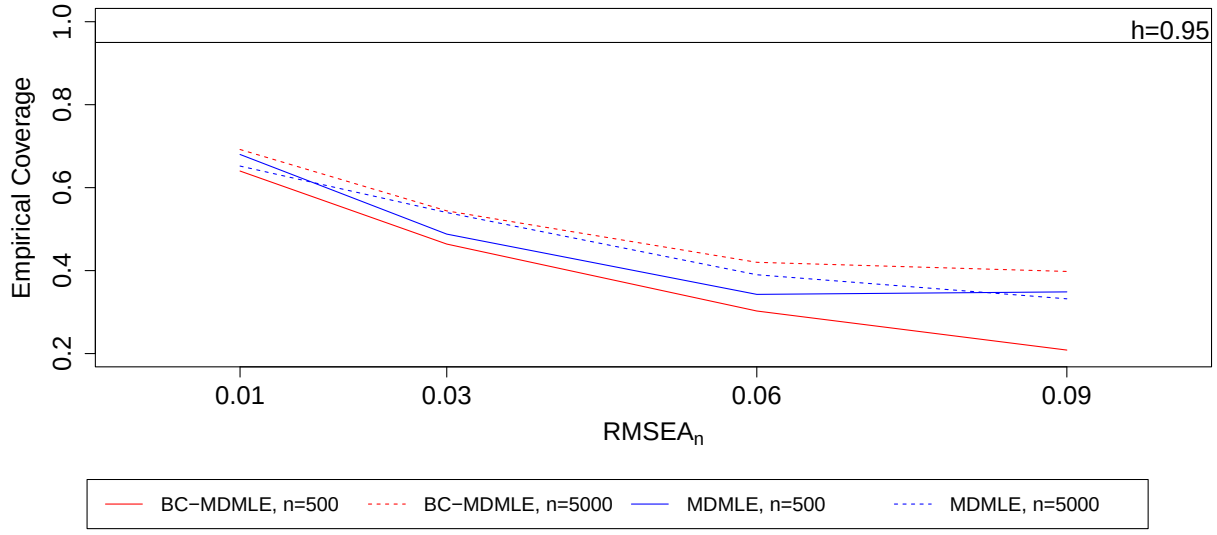


Figure 4.33

The Empirical Coverage for Dispersion Parameter for Three Polytomous Items With Four Categories

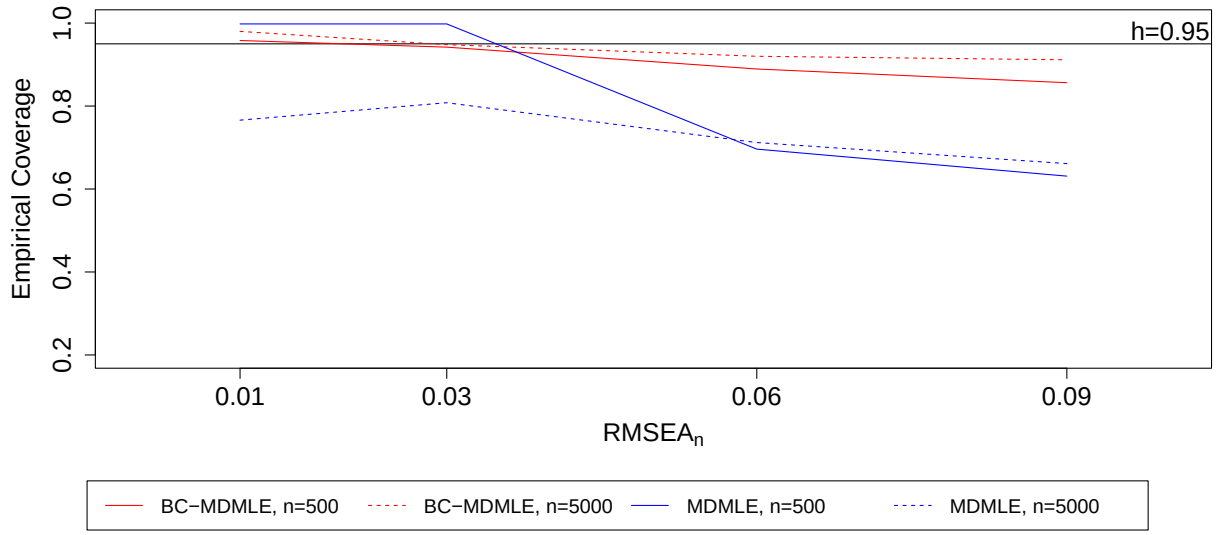
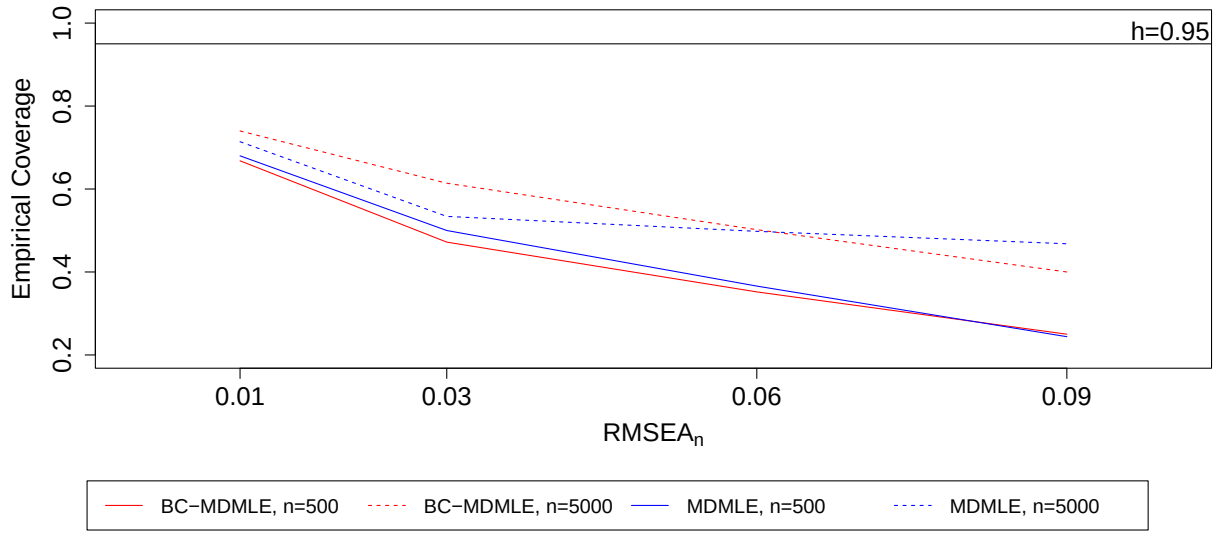


Figure 4.34

The Empirical Coverage for Dispersion Parameter for Six Polytomous Items With Four Categories



4.4.4.3 Dispersion Parameter Empirical Coverage With a Sample Size of 500,000

The empirical coverages of the 95% CIs for the dispersion parameter are shown in Figures 4.35 and 4.36. Figure 4.35 represents the test with three polytomous items, each having four categories. Figure 4.36, on the other hand, represents the test with six polytomous items, also with four categories. In both figures, the red lines represent the empirical coverages associated with the BC-MDMLE, while the blue lines represent the empirical coverages associated with the MDMLE. The solid lines correspond to a sample size of 500, the dashed lines represent a sample size of 5000, and the dotted lines represent a sample size of 500,000.

Observing Figure 4.35, it can be noted that when the test length is shorter, increasing the sample size to a very large number does not yield a substantial improvement in the empirical cov-

erage associated with both the BC-MDMLE and MDMLE. Generally, regardless of the sample size, the empirical coverages associated with the BC-MDMLE tend to be closer to 0.95 compared to those obtained from the MDMLE. Furthermore, as the test length increases from three to six, as we can find in Figure 4.36, the improvement in the empirical coverage associated with both the BC-MDMLE and the MDMLE becomes more significant when the sample size reaches 500,000. In the case of a test comprising three polytomous items with four categories, the proportion of cells with expected counts below five is 0 for a sample size of $n = 500,000$ (compared to 0.55 when $n = 500$ and 0.09 when $n = 5000$). However, when considering a test consisting of six polytomous items with four categories, the proportion of cells with expected counts below five is 0.33 for a sample size of $n = 500,000$ (compared to 1 when $n = 500$ and 0.94 when $n = 5000$). However, even with this notable improvement, the empirical coverages associated with both the estimators BC-MDMLE and MDMLE, across the various levels of RMSEA_n, still fall significantly short of the desired value of 0.95.

To summarize, increasing the sample size to a large value has a notable effect on the empirical coverage associated with both the BC-MDMLE and MDMLE when the test length is longer. However, despite the observed improvements in the empirical coverage, it remains significantly below the desired value of 0.95. The empirical coverage performs best when the test length is shorter and the sample size is large.

Figure 4.35

The Three Empirical Coverage for υ for Three Polytomous Items With Four Categories

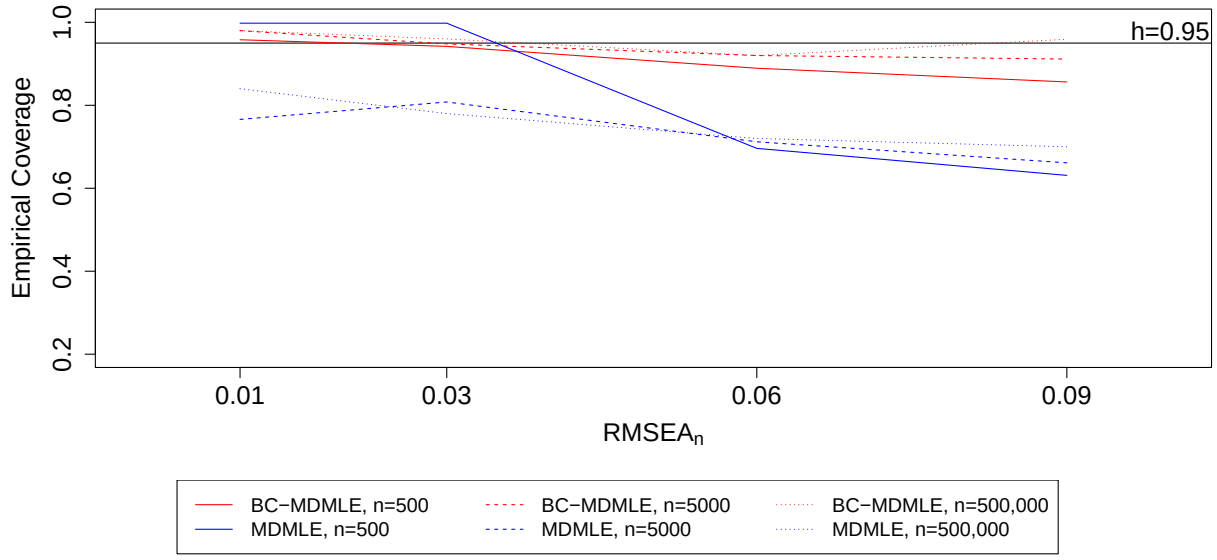
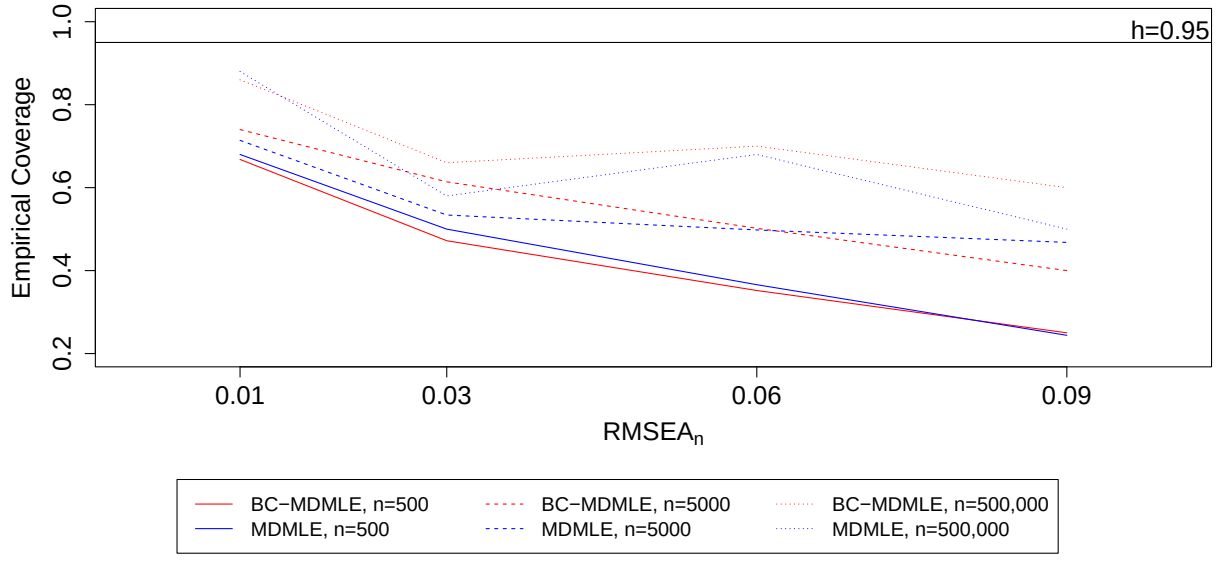


Figure 4.36

The Three Empirical Coverage for υ for Six Polytomous Items With Four Categories



Chapter 5: Empirical Example

In this chapter, we present empirical illustrations for the proposed Dirichlet-Multinomial model with two datasets: the Law School Admission Test Section 7 (LSAT7) dataset (Bock & Lieberman, 1970; Thissen, 1982) and the Consumer Protection and Perceptions of Science and Technology section of the 1992 Euro-Barometer Survey (Science) dataset (Reif & Melich, 1995; Rizopoulos, 2007; Rizopoulos & Rizopoulos, 2018). The proposed Dirichlet-Multinomial model, as described in Section 3.2.2.2, refers to the bias-corrected form of the Dirichlet-Multinomial model. Maximizing the bias-corrected form BC-LL-DM enables the derivation of the BC-MDMLE. For each dataset, we fit a standard IRT model assuming the multinomial likelihood (i.e., 2PL for the LSAT7 data and GRM for the Science data) as well as the corresponding proposed Dirichlet-Multinomial model. Parameter estimates (i.e., MMLE and BC-MDMLE) and the 95% CIs associated with those estimators are compared.

To assess model fit, we compared the full-information-based RMSEA_n (i.e., $\hat{\varepsilon}_n$) estimates from the multinomial submodels with the square root of the BC-MDMLE of the dispersion parameter (see connection specified in Equation 3.19). The decision to retain or reject the proposed Dirichlet-Multinomial model was determined by evaluating the magnitude of the adventitious error, which is reflected by the estimated $\hat{\varepsilon}_n$ and the corresponding 95% CI. Drawing upon the methodology described in Wu and Browne (2015a), if the lower bound of the 95% CI associated

with the estimated RMSEA_n is larger than the predefined threshold in Maydeu-Olivares and Joe (2014) for maximum allowable model error (i.e., 0.03), the model is rejected.

5.1 Data

The empirical analysis contains data analysis using two datasets: LSAT7 and Science. Both datasets are available in mirt package (Chalmers, 2012; Chalmers et al., 2015) in R. Science dataset is also available in ltm package (Rizopoulos, 2007; Rizopoulos & Rizopoulos, 2018) in R.

The LSAT7 dataset consists of five dichotomous items related to Debate, as described by Bock and Lieberman (1970). These items specifically correspond to Item 11 through 15 of Section seven in the LSAT test from the 1970s. The dataset comprises 1000 responses, and no missing values were identified within the dataset.

The Science dataset is derived from a survey sample conducted in Great Britain, as documented by Rizopoulos (2007) and Rizopoulos and Rizopoulos (2018). This dataset includes responses from 392 individuals and focuses on people's attitudes toward science and technology. It comprises four items that ask about the following aspects:

- Item 1: Comfort - This item assesses agreement with the statement "Science and technology are making our lives healthier, easier, and more comfortable."
- Item 2: Work - This item assesses agreement with the statement "The application of science and new technology will make work more interesting."
- Item 3: Future - This item assesses agreement with the statement "Thanks to science and technology, there will be more opportunities for future generations."

- Item 4: Benefit - This item assesses agreement with the statement "The benefits of science are greater than any harmful effect it may have."

Each of the four items features four response categories, ranging from strongly disagree (1) to strongly agree (4). There are also no missing values within the dataset.

5.2 Analysis

To evaluate model fit, the multinomial submodels utilized the $\hat{\varepsilon}_n$ estimates based on Pearson's X^2 statistic and the corresponding df as outlined in Maydeu-Olivares and Joe (2014). The estimated $\hat{\varepsilon}_n$ was calculated using the formula specified in Equation 2.30, where the estimated $\hat{X}^2$ can be computed using the formula specified in Equation 2.22. In addition, the upper and lower bounds of RMSEA$_n$ in the 95% CI were calculated using the formula specified in Equation 2.31. It should be noted that if the CI is 95% instead of 90%, the corresponding solutions $\hat{\lambda}_L$ and $\hat{\lambda}_U$ are for λ in $G(x|\lambda, df) = 0.975$ and $G(x|\lambda, df) = 0.025$, respectively.

Within the proposed Dirichlet-Multinomial framework, the dispersion parameters $\hat{v}_0$ were estimated alongside the item parameters. The upper and lower bounds of the dispersion parameter were calculated using the formula specified in Equation 3.40 in Section 3.2.2.2. We obtained the estimated RMSEA$_n$ along with the upper and lower bounds in the 95% CI by calculating the square root of the corresponding dispersion parameter estimates using the approximation specified in Equation 3.19.

5.3 Results

The findings presented in this section provide the parameter estimates, as well as the lower and upper bounds of the 95% CIs for: 1) item parameters, 2) dispersion parameter (only applicable to the BC-MDMLE), and 3) RMSEA$_n$. Additionally, the interval lengths were computed to illustrate the width of the CIs for each parameter across various models.

Table 5.1 presents the parameter estimates and CI results for the LSAT7 dataset with dichotomous responses. There are a total of 10 item parameters in the dataset, including five item slopes (a_1 to a_5) and five item intercepts (d_1 to d_5). The MMLE in the 2PL model (with both the non-robust SEs and the robust SEs), as well as the BC-MDMLE, are found to be very similar (see the first three columns in Table 5.1). Furthermore, it was observed that the lower bounds of item parameters are smallest in the BC-MDMLE, followed by the MMLE with robust SEs, and finally the MMLE with the non-robust SEs. Regarding the upper bounds of item parameters, the BC-MDMLE shows the highest upper bounds, followed by the MMLE with the robust SEs, and then the MMLE with the non-robust SEs.

The CI intervals for the item parameters can be found in Figure 5.1. The figure reveals that the MMLE with robust SEs yields wider intervals compared to those with non-robust SEs. This difference can be attributed to the fact that robust SEs remain valid even when the model is misspecified, provided that the likelihood remains i.i.d. (Yuan et al., 2014). Additionally, in the BC-MDMLE, the CI intervals for item parameters are approximately 1.16 times wider compared to those in the MMLE with the robust SEs, and 1.21 times wider compared to those in the MMLE with the non-robust SEs.

Furthermore, it can be found from Table 5.1 that both the MMLE and the BC-MDMLE

yield $\hat{\varepsilon}_n$ values of approximately 0.02, indicating a good fit to the LSAT7 data. The CI interval for the RMSEA$_n$ is slightly larger in the BC-MDMLE compared to the MMLE with both the robust and non-robust SEs. Since the lower bound of the BC-MDMLE's RMSEA$_n$ is smaller than the cutoff value of 0.03, it is suggested to retain the proposed Dirichlet-Multinomial model.

Table 5.1

Parameter Estimates and CIs in LSAT7 Data

Parameters	Estimates			Lower Bound			Upper Bound		
	NR[a]	R[b]	BC-MDMLE[c]	NR	R	BC-MDMLE	NR	R	BC-MDMLE
a_1	0.988	0.988	0.976	0.655	0.615	0.571	1.320	1.360	1.381
a_2	1.081	1.081	1.097	0.743	0.733	0.675	1.419	1.428	1.520
a_3	1.707	1.707	1.670	1.087	1.061	0.932	2.328	2.354	2.407
a_4	0.765	0.765	0.762	0.509	0.493	0.448	1.021	1.037	1.075
a_5	0.736	0.736	0.717	0.443	0.429	0.361	1.029	1.042	1.073
d_1	1.856	1.856	1.850	1.603	1.592	1.543	2.108	2.120	2.156
d_2	0.808	0.808	0.810	0.628	0.628	0.589	0.988	0.988	1.032
d_3	1.805	1.805	1.790	1.408	1.395	1.318	2.203	2.216	2.263
d_4	0.486	0.486	0.488	0.339	0.339	0.309	0.633	0.633	0.667
d_5	1.854	1.854	1.844	1.631	1.629	1.574	2.078	2.080	2.113
$\hat{v}_0$	N/A*	N/A	0.0005	N/A	N/A	0.000	N/A	N/A	0.002
$\hat{\varepsilon}_n$	0.023	0.023	0.022	0.000	0.000	0.000	0.041	0.041	0.045

Note. In the context provided, the symbol as represent the parameter slopes, a_1 denotes slope parameter for the first item, and ds refer to the parameter intercepts. Due to the page width limitation, the model names in the study have been abbreviated in the table.

[a] The "NR" represents the MMLE with non-robust SEs.

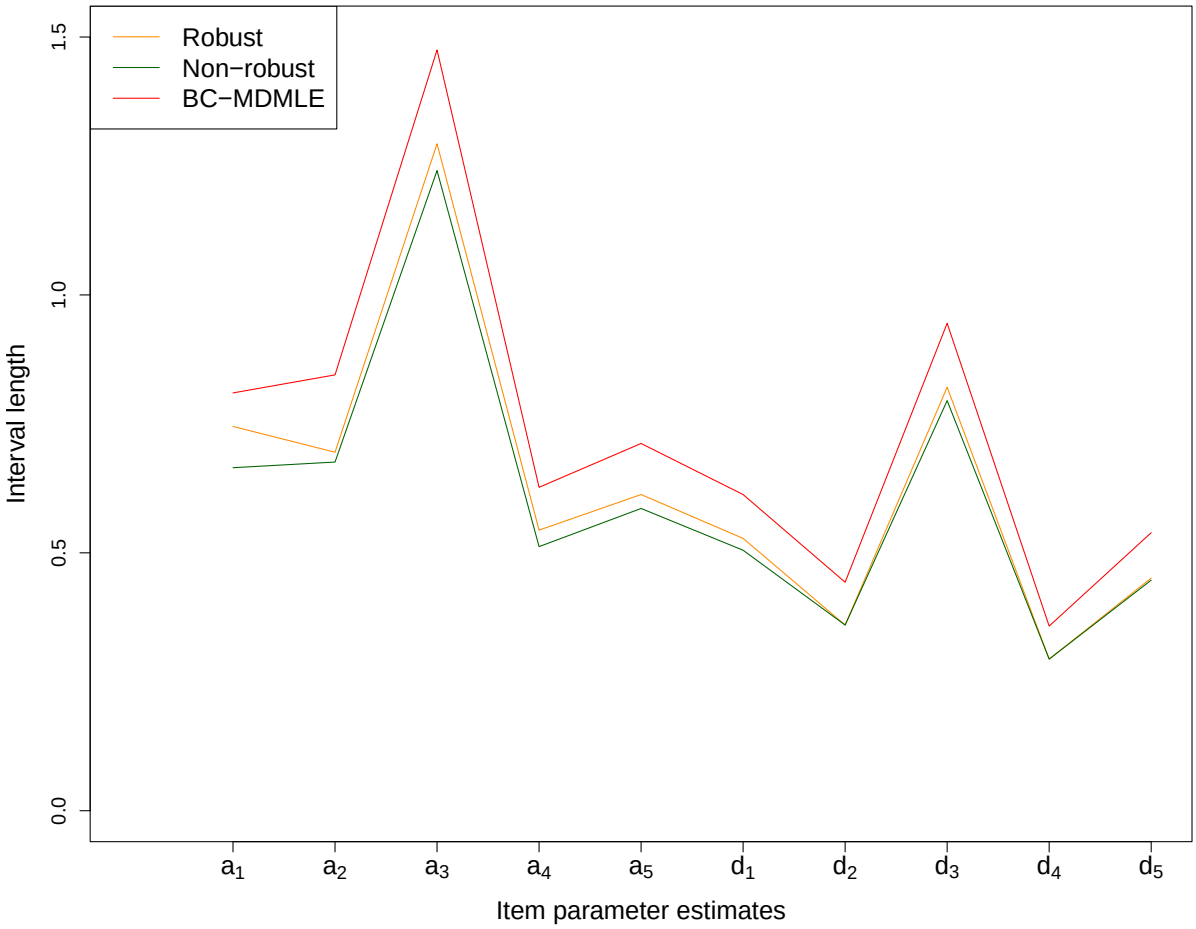
[b] The "R" corresponds to MMLE with robust SEs.

[c] The "BC-MDMLE" refers to the bias-corrected maximum Dirichlet-Multinomial likelihood estimator.

* The reason for "N/A" is that there is no estimate for $\hat{v}_0$ in the traditional 2PL models.

Figure 5.1

Interval Lengths of Item Parameters in LSAT7 Data Analysis



Note. The a denote the slope parameters, and d denotes the intercept parameters. The terms 'Robust', 'Non-robust', and 'BC-MDMLE' represent the interval lengths for the MMLE using robust and non-robust SEs, as well as for the BC-MDMLE of item parameters, respectively.

Table 5.2 presents the parameter estimates and CIs obtained from the Science dataset when the polytomous items have four response categories. It is noted that the item parameter estimates are similar, in both the MMLE with non-robust SEs and robust SEs, and in the BC-MDMLE. In contrast to the results in the LSAT7 analysis, the BC-MDMLE does not yield the largest difference between upper and lower bound when compared to the MMLE in the Science analysis.

Figure 5.2 illustrates the interval lengths of item parameter estimates. In the figure, the red line represents the interval lengths obtained from the BC-MDMLE. Additionally, orange line corresponds to the interval lengths from the MMLE with robust SEs, and green line represents the interval lengths from the MMLE with non-robust SEs. The following patterns can be found in Figure 5.2. Firstly, similar to the 2PL case, it is observed that the MMLE with robust SEs gives wider intervals compared to the MMLE with non-robust SEs. Secondly, unlike the 2PL case, the CI intervals associated with the BC-MDMLE are not consistently wider than those associated with the MMLE with robust SEs. In Table 5.2, when comparing the BC-MDMLE with the MMLE with robust SEs, the smallest difference in interval lengths occurs for the parameter d_{31}. For the BC-MDMLE, the interval is calculated as $6.647 - 3.641 = 3.006$, while for the MMLE with robust SEs, the interval is $6.822 - 3.646 = 3.176$. Thus, the smallest difference between the intervals is $3.006 - 3.176 = -0.170$. On the other hand, the largest difference between intervals is $2.114 - 1.833 = 0.281$ for d_{11}. Overall, the item parameter interval lengths in both the BC-MDMLE and the MMLE with robust SEs are similar and larger compared to the interval lengths in the MMLE with non-robust SEs.

The estimated values for $\hat{\varepsilon}_n$ are as follows: 0.069 in the MMLE with both robust and non-robust SEs, and 0.022 in the BC-MDMLE. The GRM model assumes a relatively large amount of model error, resulting in a larger estimated $\hat{\varepsilon}_n$. Conversely, the proposed Dirichlet-Multinomial model assumes less adventitious error, leading to smaller estimated $\hat{\varepsilon}_n$ values. However, it is important to note that the MMLE, whether with robust or non-robust SEs, yields a narrower CI interval for RMSEA$_n$ compared to the BC-MDMLE. In the MMLE with robust and non-robust SEs, the CI interval is calculated as $0.077 - 0.062 = 0.015$, while in the BC-MDMLE, the CI interval is $0.034 - 0.005 = 0.029$. It is worth mentioning that the lower bound of RMSEA$_n$

associated with BC-MDMLE is smaller than 0.03 (see Table 5.2). Based on the same rationale as in the 2PL case, the proposed Dirichlet-Multinomial model should be retained.

Table 5.2

Parameter Estimates and CIs in Science Data

Parameters	Estimates			Lower Bound			Upper Bound		
	NR[a]	R[b]	BC-MDMLE[c]	NR	R	BC-MDMLE	NR	R	BC-MDMLE
a_1	1.042	1.042	1.038	0.695	0.575	0.655	1.389	1.509	1.422
a_2	1.226	1.226	1.181	0.860	0.850	0.787	1.592	1.602	1.574
a_3	2.293	2.293	2.237	1.378	1.270	1.254	3.209	3.317	3.220
a_4	1.095	1.095	1.082	0.762	0.650	0.716	1.428	1.540	1.448
d_{11}	4.864	4.864	4.910	3.919	3.948	3.853	5.810	5.781	5.967
d_{12}	2.640	2.640	2.620	2.208	2.157	2.148	3.072	3.123	3.091
d_{13}	-1.466	-1.466	-1.482	-1.774	-1.802	-1.821	-1.158	-1.130	-1.143
d_{21}	2.924	2.924	2.912	2.445	2.441	2.394	3.403	3.407	3.429
d_{22}	0.901	0.901	0.892	0.618	0.615	0.587	1.184	1.187	1.197
d_{23}	-2.267	-2.267	-2.217	-2.666	-2.656	-2.643	-1.868	-1.877	-1.791
d_{31}	5.234	5.234	5.144	3.824	3.646	3.641	6.644	6.822	6.647
d_{32}	2.214	2.214	2.187	1.528	1.480	1.451	2.900	2.947	2.923
d_{33}	-1.964	-1.964	-1.910	-2.588	-2.626	-2.572	-1.340	-1.302	-1.247
d_{41}	3.348	3.348	3.300	2.820	2.776	2.730	3.876	3.920	3.871
d_{42}	0.992	0.992	0.989	0.717	0.708	0.689	1.266	1.276	1.289
d_{43}	-1.688	-1.688	-1.675	-2.012	-2.041	-2.028	-1.365	-1.335	-1.323
$\hat{v}_0$	N/A*	N/A	0.0005	N/A	N/A	0.000	N/A	N/A	0.001
$\hat{\varepsilon}_n$	0.069	0.069	0.022	0.062	0.062	0.005	0.077	0.077	0.034

Note. In the context provided, the symbol as represent the parameter slopes, a_1 denotes slope parameter for the first item, and ds refer to the parameter intercepts. The symbol d_{21} denotes the first intercept parameter for the second item. Due to the page width limitation, the model names in the study have been abbreviated in the table.

[a] The "NR" represents the MMLE with non-robust SEs.

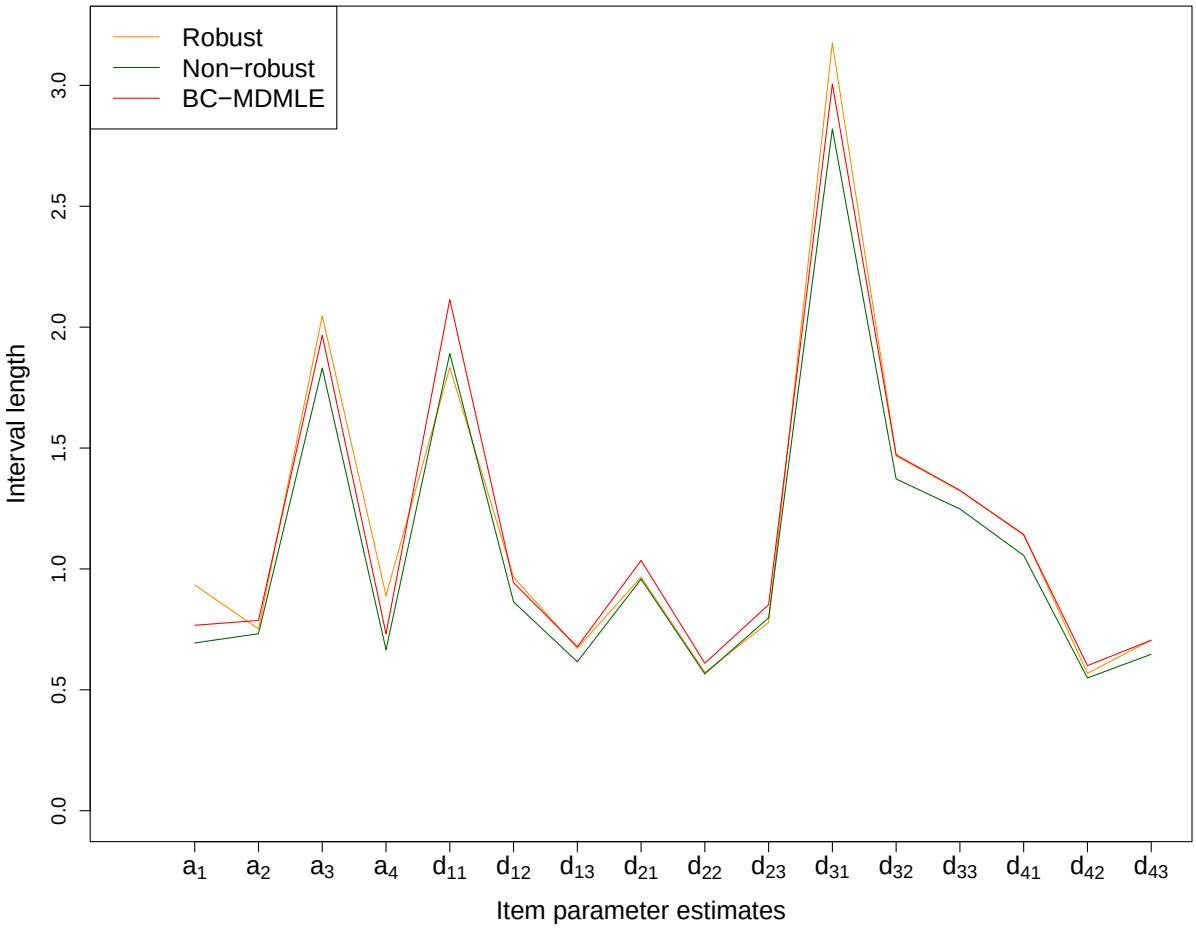
[b] The "R" corresponds to the MMLE with robust SEs.

[c] The "BC-MDMLE" refers to the bias-corrected maximum Dirichlet-Multinomial likelihood estimator.

* The reason for "N/A" is that there is no estimate for $\hat{v}_0$ in the traditional GRM models.

Figure 5.2

Interval Lengths of Item Parameters in Science Data Analysis



Note. The a denote the slope parameters, and d denotes the intercept parameters. d_{12} denotes the second intercept parameter for Item 1. The terms 'robust', 'non-robust', and 'BC-MDMLE' represent the interval lengths for the MMLE using robust and non-robust SEs, as well as for the BC-MDMLE of item parameters, respectively.

5.3.1 Further Investigation

It is worth highlighting that the Science dataset has a relatively small sample size of 392 observations, which may lead to potential limitations in accurately estimating the true population RMSEA$_n$. Additionally, there exists a discrepancy between the estimated amount of model error

obtained from the GRM model and the proposed Dirichlet-Multinomial model. The results from the MMLE with both the robust and non-robust SEs indicate an estimated $\hat{\varepsilon}_n$ of 0.069, suggesting a relatively large model error, while the BC-MDMLE suggests a smaller $\hat{\varepsilon}_n$ of 0.022, indicating that the model error may not necessarily be large. This discrepancy could potentially due to the combination of the small sample size and the presence of a relatively larger amount of model error.

We further investigated the discrepancy in the estimated $\hat{\varepsilon}_n$ values between the multinomial submodels and the proposed Dirichlet-Multinomial model. Firstly, we examined the impact of altering the parameter starting values during the optimization process. The results indicated that modifying the starting values did not necessarily result in the optimization algorithm converging at different solutions in terms of maximized log-likelihood. We experimented by switching from the 'nlminb' optimization function to the 'optim' function from R *stats*, or the 'nloptr' function in the 'nloptr' package in R, and all optimization processes led to very similar solutions.

Furthermore, we explored another dataset from Magnus and Liu (2022), which contains 14 polytomous items with four response categories from the Collaborative Psychiatric Epidemiological Surveys (CPES). The dataset consisted of 3000 observations without missing values, which is substantially larger than the Science data. Due to computational limitations, we randomly selected six out of the 14 items and obtained the MMLE (with robust and non-robust SEs), as well as the BC-MDMLE. In this case, we observed a similar pattern to what we found in the LSAT7 results presented in this study. Remarkably, the estimated $\hat{\varepsilon}_n$ values from both the MMLE and the BC-MDMLE are approximately 0.02. Moreover, the CI intervals for item parameters were wider in the BC-MDMLE compared to the MMLE with both robust and non-robust SEs, aligning with the previous findings from the LSAT7 analysis. Based on this indirect evidence, it appears that

the sample size, or the combination of sample size and model errors, may have an impact on the estimated $\hat{\varepsilon}_n$ values and their corresponding CI intervals.

Chapter 6: Summary and Discussion

Quantifying model errors is of great interest to researchers when utilizing statistical models for decision-making purposes (MacCallum & O'Hagan, 2015). Primary contributors to this uncertainty include sampling error and model error. Sampling error denotes the discrepancy between observed values within a sample and the true values of the entire population. On the other hand, model error characterizes the discrepancy between the predictions or estimates made by a statistical model and the true values in the real world. While the magnitude of sampling error diminishes with an increased sample size, model error occurs at the population level and remains unaffected by variations in the sample size.

In statistical modeling, there are two approaches to deal with model error: the fixed-effect approach and the random-effect approach. The fixed-effect approach assumes that all samples are drawn from a specific population, which has a fixed amount of difference from the fitted model. On the other hand, the random-effect approach assumes that the samples come from multiple populations, each with its own random difference from the fitted model. In our study, we adopt the random-effect approach, which is referred to as "adventitious error" approach according to Wu and Browne (2015a). Ignoring model error can lead to underestimated uncertainty in model predictions and inaccurate explanations in parameter estimates (MacCallum & O'Hagan, 2015). Consequently, it is crucial to appropriately account for model error before making any inferences

or decisions based on a statistical model.

Both the fixed-effect and random-effect approaches have advantages when addressing model errors. In the fixed-effect approach, by assuming a fixed discrepancy between the fitted model and the population, well-established effect size measures such as RMSEA, SRMR, CFI, and GFI can be utilized to assess how accurately the fitted model represents the population. These measures are widely accepted and commonly used in numerous studies for evaluating model fit. However, it should be noted that the cutoff values recommended in papers like Browne and Cudeck (1992), Steiger (1990), and Hu and Bentler (1999) may be specific to the simulation conditions in those papers and may not be universally applicable to other situations. In practice, opting for the fixed-effect approach in IRT models enables us to obtain model fit indices, providing insights into the level of model error when fitting models to the data.

On the other hand, the random-effect approach considers multiple operational populations as realizations of the fitted model, each with slight variations in operationalization. This approach allows for quantifying the extent of model error and simultaneously making inferences about the model parameters. The concentration/dispersion parameter estimated in the random-effect approach can be approximately translated to fixed-effect approach effect size measures such as RMSEA. However, there is ongoing controversy regarding the stochastic nature of random-effect model error approaches, including the definition of the three layers of populations and the stability and interpretability of the concentration/dispersion parameters. In real scenarios, by fitting the proposed Dirichlet-Multinomial framework, we obtain an estimate of the RMSEA along with the estimated item parameters. This estimated RMSEA provides valuable information about the extent of model error in the fitted model and helps us determine the reliability of the item parameter estimates derived from the model.

The main contribution of this research is extending the adventitious error approach proposed by Wu and Browne (2015a). In particular, we developed a stochastic Dirichlet-Multinomial framework to quantify model errors in IRT analyses. This framework allows for the measurement of model errors and the estimation of item parameters simultaneously. Within the Dirichlet-Multinomial framework, we proposed a bias correction technique and derived asymptotic properties for the resulting estimator. To evaluate the performance of our proposal, we assess model convergence, point estimation (including bias and RMSEs), and interval estimation. In comparison to standard multinomial submodels like the 2PL and GRM, our findings indicate that, in general, the BC-MDMLE from the proposed Dirichlet-Multinomial model exhibits lower parameter estimation bias and provide better empirical coverage of confidence intervals. These confidence intervals capture the uncertainty from both sampling errors and adventitious errors.

6.1 Interpretation and Analysis of Findings

In summary, the results from the simulation study suggest that achieving empirical coverage close to the nominal level is more likely when the RMSEA_n value is 0.03 or lower for all models. Additionally, larger sample sizes are associated with improved empirical coverage. The findings are consistent with the theoretical asymptotic results presented in Chapter 3. This can be explained by the fact that a smaller RMSEA_n indicates a better model fit, leading to improved empirical coverage. Moreover, larger sample sizes provide more information about the population characteristics, resulting in more accurate parameter estimates and, on average, better empirical coverage.

Regarding item parameter estimates, it is observed that the BC-MDMLE from the pro-

posed Dirichlet-Multinomial model exhibits consistently smaller bias and RMSE compared to the MMLE, particularly when the population RMSEA$_n$ is larger. While the traditional multinomial submodels only account for sampling error when estimating item parameters, the proposed Dirichlet-Multinomial model addresses both the effects of sampling error and adventitious error. Moreover, the proposed Dirichlet-Multinomial model provides more accurate item parameter estimates by removing the asymptotic bias of the dispersion parameter.

Regarding the relative bias in dispersion parameter estimates, it is found that the MDMLE from the Dirichlet-Multinomial model without bias-correction tends to underestimate the dispersion parameter, aligning with the theoretical conclusions. On the other hand, under certain conditions, it is observed that the BC-MDMLE from the proposed Dirichlet-Multinomial model has a tendency to overestimate the dispersion parameter. In addition, when considering the RMSE for the dispersion parameter, the BC-MDMLE is not always preferred over the MDMLE.

To provide an explanation for the overestimated BC-MDMLE of the dispersion parameter under specific conditions, it can be understood in the following way. Generally, a smaller RMSEA$_n$ indicates a smaller model error, which suggests more accurate parameter estimates. However, in cases where the population RMSEA$_n$ is smaller, the BC-MDMLE of the dispersion parameter also tends to be smaller, as it is closely related to RMSEA$_n$ under conditions of large sample size and concentration parameter. However, extremely small dispersion parameters pose challenges for parameter recovery due to their sensitivity to minor changes and numerical instability near the optimization boundary (which is 0 in this case). Nevertheless, as the true RMSEA$_n$ and sample size increase, the difficulty in recovering the dispersion parameter diminishes.

Furthermore, a significant finding regarding the empirical coverage of the CI for item parameters is that, in the case of the MMLE, as the value of RMSEA$_n$ increases, the empirical

coverage declines rapidly, regardless of whether the non-robust or the robust SEs is used. Conversely, the BC-MDMLE exhibits a stable empirical coverage of approximately the nominal level across various levels of population RMSEA$_n$ and different simulation conditions.

As the RMSEA$_n$ increases, the departure of the empirical coverage of CI for item parameters in traditional multinomial submodels from the nominal level can be attributed to the violation of the assumption that the data is i.i.d. In the study by Yuan et al. (2014), it is indicated that the robust SEs are expected to accurately capture the sampling variability of the MLEs of model parameters when the model is misspecified (i.e., when model error is present). However, their study assumes that the individual response data are i.i.d. In contrast, our study involves a true data-generating model known as the Dirichlet-Multinomial model, with Dirichlet priors on the operational population probability $\boldsymbol{\pi}$. This implies that there is underlying dependency among the individual responses. As a result, the data were not generated from an i.i.d. multinomial distribution, which is the assumption underlying the fitting of traditional multinomial submodels. Therefore, the theory of robust SEs does not apply to our specific condition, given the presence of dependency among the data.

It is evident that the empirical coverage of CI for the dispersion parameter is affected when the population RMSEA$_n$ is larger and the test length is longer. A larger population RMSEA$_n$ indicates a higher degree of model error, which naturally leads to poorer parameter estimates and CIs. However, the main issue lies in the sparseness of the contingency table, which has as many cells as the number of response patterns. As discussed in Section 4.4.4.2, asymptotic methods can accurately approximate the sampling distribution of RMSEA$_n$ in small models with minimal sparseness in contingency tables. However, when the number of cells in a contingency table is large, regardless of the sample size, there can be a substantial number of cells with expected

frequencies less than five. As explained in Section 4.4.4.3, in general, increasing the sample size to a very large number helps in improving the empirical coverage of CIs for the dispersion parameter.

It is important to highlight that, despite the imperfect empirical coverage of CI in the dispersion parameter in certain scenarios, we still recommend using the BC-MDMLE in practical applications. The BC-MDMLE addresses the bias in estimating the dispersion parameter, making it preferable over the MDMLE. Moreover, in scenarios with large model error, BC-MDMLE demonstrates significant advantages over MMLE. BC-MDMLE exhibits smaller bias and RMSE in item parameter estimates, and the empirical coverage of the 95% CIs for item parameters is closer to the nominal level when using BC-MDMLE. Given these benefits, it is highly recommended to use BC-MDMLE for estimating the item parameters in cases where the model error is substantial.

Furthermore, it should be noted that in our approach, the primary focus is not on assessing model fit. Instead, the proposed Dirichlet-Multinomial framework incorporates both sampling error and model error while making inferences based on the fitted models. By accounting for model uncertainty arising from both sources, our BC-MDMLE adjusts for point estimates and empirical coverage of confidence intervals compared to the MMLE obtained from traditional multinomial submodels. This is the major contribution of our proposed framework. It enables more accurate and robust inferences, considering the complexities introduced by both sampling and model errors.

Our analysis of the proposed Dirichlet-Multinomial model, as well as the traditional multinomial submodels, using the LSAT7 and Science datasets revealed several key findings. Firstly, the item parameter estimates were consistently similar across both models for both datasets.

Secondly, the CI intervals associated with the BC-MDMLE were not always wider than those associated with the MMLE, regardless of whether robust or non-robust standard errors were used. We observed a discrepancy in the estimated RMSEA$_n$ between the MMLE and the BC-MDMLE. This may be attributed to a combination of factors, including a small sample size and the presence of a relatively large amount of model error. Future research should focus on conducting further explorations within the proposed Dirichlet-Multinomial framework to advance our understanding in this area.

6.2 Limitations and Future Directions

While this paper primarily focuses on the full-information RMSEA$_n$, it is important to acknowledge that the proposed method can also be applied to other limited-information RMSEA measures, such as RMSEA$_2$ and RMSEA$_{ord}$. Findings from the simulation studies and empirical data analyses in this thesis have revealed that the proposed Dirichlet-Multinomial framework and the RMSEA$_n$ used in the study rely on full-information discrepancy functions and non-sparse contingency tables. In practical applications, data frequently exhibit sparseness, and consequently, the CIs obtained using asymptotic methods tend to underestimate the true variability of the estimated RMSEA$_n$ (Maydeu-Olivares & Joe, 2014). To address this limitation, future research will explore a limited-information-based Dirichlet-Multinomial framework that incorporates limited-information RMSEA measures. This future approach aims to effectively address the limitations within full-information-based Dirichlet-Multinomial model and obtain more reliable results, particularly in the presence of potential adventitious errors.

Another promising direction for future research is to conduct a simulation study that ex-

plores the relationship between the SEs associated with the MMLE and the SEs calculated by incorporating an additional $\hat{v}_0$ parameter (as demonstrated in Equation 3.39). Our simulation study has revealed that when the model error is small, the MMLE closely align with the BC-MDMLE. In such instances, the empirical coverage of CIs in item parameters associated with the MMLE may not necessarily be compromised when we plug-in an extra $\hat{v}_0$ parameter in the SE calculations. However, it is worth noting that if we intend to obtain this additional parameter estimate $\hat{v}_0$ while constructing CIs using the maximum multinomial likelihood estimation, it might still be necessary to fit the proposed Dirichlet-Multinomial model in addition to the traditional multinomial model.

In future research, a more systematic exploration can be conducted to thoroughly investigate whether the estimated dispersion parameter from the proposed Dirichlet-Multinomial framework matches the approximate RMSEA value obtained from traditional multinomial submodel. Previous investigations revealed that when the sample size is very large (e.g., 1,000,000), the estimated RMSEA_n values obtained from the traditional multinomial submodels align closely with the true RMSEA_n values. Furthermore, in our examination of a test comprising three items with four response categories and a sample size of 500,000, both MDMLE and BC-MDMLE provided RMSEA estimates that closely matched their corresponding true RMSEA values, although the MDMLE of the RMSEA slightly underestimated the true value. These explorations primarily aimed to verify the accuracy of the R code. For future research intending to validate Equation 3.19, a more thorough and rigorous study can be conducted.

In addition to the aforementioned considerations, it is important to address concerns raised by Satorra (2015) and Shapiro (2015) regarding the reliance on a single concentration/dispersion parameter to quantify adventitious error. They argue that relying solely on the RMSEA value

to determine model acceptance or rejection is somewhat arbitrary. They express concerns that such an approach may be too restrictive to capture the potential complexity of adventitious error. Moreover, relying on arbitrary cutoffs while using a particular effect size measure can be risky, especially when these cutoff values depend on the model and true parameter values. It is crucial to avoid such pitfalls in the analysis. In the discussions in Wu and Browne (2015b), they indicated that to enhance the robustness of our model selection decisions and enable a more comprehensive evaluation of whether to retain or reject a model, an important direction for future research is to explore the incorporation of other model fit indices within the framework.

It is important to mention that the Dirichlet-Multinomial framework proposed in this study is applicable primarily to cases where the test length is not excessively long. For example, in our simulation, the longest test consisted of 12 dichotomous items or six polytomous items with four response categories. However, in real testing scenarios, tests tend to be longer, and polytomous items might have more response categories. With longer tests, the computation time significantly increases. For instance, if we have 20 dichotomous items in the test, the total number of response patterns would be $2^{20} = 1,048,576$. In the proposed Dirichlet-Multinomial framework, considering the objective function and gradients, even when integrating over all the observed response patterns (excluding those with count equal to 0), the number of observed response patterns could become large and beyond our control. We must recognize that our investigation of the Dirichlet-Multinomial framework within IRT models is still in its early stages. To enhance the practicality and broad applicability of this framework, it is essential to improve computational efficiency, consider other estimation algorithms, and ensure more stable parameter estimates.

It is worth noting that while we utilized R for model fitting, other software options are available as well. For fitting traditional multinomial submodels, the 'pyirt' library in Python can

be employed. Alternatively, Mplus offers the capability to fit various item response theory IRT models using the MODEL command. In Python, the 'scipy.optimize.minimize' function from the 'scipy' library serves as a valuable tool for numerical optimization, enabling the search for the minimum or maximum of a given function. Notably, in the context of Mplus, optimization is automatically performed when utilizing the MODEL command, with various algorithms like ML or Bayesian methods used to estimate model parameters that best align with the observed data and the specified model.

Finally, similar to Wu and Browne (2015a), the choice of distributions within the current Dirichlet-Multinomial framework is primarily based on mathematical convenience and conjugacy in Bayesian inference. In future research, alternative distributions based on empirical observations or fully Bayesian approaches could be considered.

Appendix A: Lemmas and Proofs

Lemma 1. *The log-likelihood function for the Dirichlet-Multinomial distribution is represented by Equation 3.20. In order to simplify the estimation of the item parameter $\boldsymbol{\xi}$, we consider the case where $m = 1$ and focus on the LL-DM. Using the parameterization with the concentration parameter α, the log-likelihood function is a sum of two parts,*

$$\log f_{DM}(\boldsymbol{p}|\boldsymbol{\xi}, \alpha, n) = \ell_1(\alpha) + \ell_2(\alpha, \boldsymbol{\xi}),$$

in which

$$\ell_1(\alpha) = \log\Gamma(\alpha) - \log\Gamma(n + \alpha)$$

$$\ell_2(\alpha, \boldsymbol{\xi}) = \sum_{i=1}^{K} \log\Gamma(np_i + \alpha\tau_i(\boldsymbol{\xi})) - \log\Gamma(\alpha\tau_i(\boldsymbol{\xi})).$$

For the purpose of estimating $\boldsymbol{\xi}$, the estimating function reduces to the $\boldsymbol{\xi}$-gradient of the $\ell_2(\alpha, \boldsymbol{\xi})$,

$$\frac{\partial\ell_2(\alpha, \boldsymbol{\xi})}{\partial\boldsymbol{\xi}} = \alpha \sum_{i=1}^{K} \frac{\partial}{\partial\boldsymbol{\xi}}\tau_i(\boldsymbol{\xi})[\psi(np_i + \alpha\tau_i(\boldsymbol{\xi})) - \psi(\alpha\tau_i(\boldsymbol{\xi}))].$$

An alternative method, outlined in the Abramowitz and Stegun (1964), is to utilize the asymptotic approximation using Stirling's formula. It is indicated that as $x \to \infty$, $\psi(x) = \log(x) - \frac{1}{2x} +$

$O(x^{-2})$. *Then, as* $n \to \infty$ *and* $\alpha \to \infty$, *we have*

$$\psi(np_i + \alpha\tau_i(\boldsymbol{\xi})) = \log(np_i + \alpha\tau_i(\boldsymbol{\xi})) - \frac{1}{2}\frac{1}{np_i + \alpha\tau_i(\boldsymbol{\xi})} + O_p(\frac{1}{\alpha})$$
$$\psi(\alpha\tau_i(\boldsymbol{\xi})) = \log(\alpha\tau_i(\boldsymbol{\xi})) - \frac{1}{2}\frac{1}{\alpha\tau_i(\boldsymbol{\xi})} + O_p(\frac{1}{\alpha}).$$

Here we assume $np_i + \alpha\tau_i(\boldsymbol{\xi})$ *and* $\alpha\tau_i(\boldsymbol{\xi})$ *are very large because from Lemma 8, we know* $\boldsymbol{\tau}(\hat{\boldsymbol{\xi}}) \xrightarrow{p} \boldsymbol{\tau}(\boldsymbol{\xi}_0)$, *assume that* $\boldsymbol{\xi}_0$ *is not on the boundary within the vicinity and all numbers are positive, a positive number times* α *or* n *also goes to infinity. Put these approximations back to the derivative, now we have*

$$\begin{aligned}\frac{\partial \ell_2(\alpha, \boldsymbol{\xi})}{\partial \boldsymbol{\xi}} &= \alpha \sum_{i=1}^{K} \frac{\partial}{\partial \boldsymbol{\xi}} \tau_i(\boldsymbol{\xi}) [\log(np_i + \alpha\tau_i(\boldsymbol{\xi})) - \log(\alpha\tau_i(\boldsymbol{\xi}))] \\ &+ \frac{1}{2} \sum_{i=1}^{K} \frac{\partial}{\partial \boldsymbol{\xi}} \tau_i(\boldsymbol{\xi}) \frac{np_i}{\tau_i(\boldsymbol{\xi})(np_i + \alpha\tau_i(\boldsymbol{\xi}))} + O_p(\frac{1}{\alpha}). \end{aligned} \tag{IA.1}$$

An approach to obtaining the estimated item parameter $\hat{\boldsymbol{\xi}}$ *is by solving the Equation IA.1 such that it equals zero. Since* $\sum_{i=1}^{K} \frac{\partial}{\partial \boldsymbol{\xi}} \tau_i(\boldsymbol{\xi}) \log(n + \alpha) = 0$ *and* $\sum_{i=1}^{K} \frac{\partial}{\partial \boldsymbol{\xi}} \tau_i(\boldsymbol{\xi}) \log(\alpha) = 0$, *the derivative in Equation IA.1 is then*

$$\begin{aligned}\frac{\partial \ell_2(\alpha, \boldsymbol{\xi})}{\partial \boldsymbol{\xi}} &= \underbrace{\alpha \sum_{i=1}^{K} \frac{\partial}{\partial \boldsymbol{\xi}} \tau_i(\boldsymbol{\xi}) [\log(\frac{np_i + \alpha\tau_i(\boldsymbol{\xi})}{n + \alpha}) - \log(\tau_i(\boldsymbol{\xi}))]}_{\textit{Part (I)}} \\ &+ \underbrace{\frac{1}{2} \sum_{i=1}^{K} \frac{\partial \tau_i(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}} \frac{1}{\tau_i(\boldsymbol{\xi})} \frac{np_i}{np_i + \alpha\tau_i(\boldsymbol{\xi})}}_{\textit{Part (II)}} + O_p(\frac{1}{\alpha}). \end{aligned} \tag{IA.2}$$

***Part (I)**. Let us assume that $\hat{\boldsymbol{\tau}}$ is the same as $\boldsymbol{\tau}(\hat{\boldsymbol{\xi}})$. $\frac{np_i+\alpha\hat{\tau}_i}{n+\alpha}$ is the posterior mode of the Dirichlet distribution, and $\hat{\tau}_i$ is the MLE of τ_{0i}. Both the posterior mode and MLE are close to $\boldsymbol{\tau}_0$ as $n \to \infty$ and $\alpha \to \infty$. By plus and minus $\log(\boldsymbol{\tau}_0)$, Part (I) can be rewritten as*

$$\begin{aligned}
&\alpha \sum_{i=1}^{K} \frac{\partial \hat{\tau}_i}{\partial \boldsymbol{\xi}} [\log(\frac{np_i + \alpha\hat{\tau}_i}{n+\alpha}) - \log(\hat{\tau}_i)] \\
&= \alpha \sum_{i=1}^{K} \frac{\partial \hat{\tau}_i}{\partial \boldsymbol{\xi}} \{[\log(\frac{np_i + \alpha\hat{\tau}_i}{n+\alpha}) - \log(\tau_{0i})] + [\log(\tau_{0i}) - \log(\hat{\tau}_i)]\}. \qquad \text{(IA.3)}
\end{aligned}$$

By Taylor expansion, $\log(y) = \log(x) + \frac{y-x}{z}$, where $z = tx + (1-t)y$ for some $t \in [0,1]$, which means z is between x and y. Suppose $\overline{\tau}_i$ is between $\frac{np_i+\alpha\hat{\tau}_i}{n+\alpha}$ and τ_{0i}, then

$$\begin{aligned}
&\log(\frac{np_i + \alpha\hat{\tau}_i}{n+\alpha}) - \log(\tau_{0i}) \\
&= [\frac{np_i + \alpha\hat{\tau}_i}{n+\alpha} - \tau_{0i}] \frac{1}{\overline{\tau}_i} \\
&= \frac{n[p_i - \tau_{0i}] + \alpha[\hat{\tau}_i - \tau_{0i}]}{(n+\alpha)\overline{\tau}_i}. \qquad \text{(IA.4)}
\end{aligned}$$

Suppose $\tilde{\tau}_i$ is between the MLE $\hat{\tau}_i$ and τ_{0i},

$$\log(\hat{\tau}_i) - \log(\tau_{0i}) = \frac{\hat{\tau}_i - \tau_{0i}}{\tilde{\tau}_i}. \qquad \text{(IA.5)}$$

The Equations IA.4 minus IA.5 results in the bracketed term in Part (I) in Equation IA.2. Then we know

$$\begin{aligned}
&\frac{n[p_i - \tau_{0i}] + \alpha[\hat{\tau}_i - \tau_{0i}]}{(n+\alpha)\overline{\tau}_i} - \frac{\hat{\tau}_i - \tau_{0i}}{\tilde{\tau}_i} \\
&= \frac{n[p_i - \tau_{0i}] + \alpha[\hat{\tau}_i - \tau_{0i}]}{(n+\alpha)\overline{\tau}_i} - \frac{n(\hat{\tau}_i - \tau_{0i})}{(n+\alpha)\overline{\tau}_i} + \frac{n(\hat{\tau}_i - \tau_{0i})}{(n+\alpha)\overline{\tau}_i} - \frac{\hat{\tau}_i - \tau_{0i}}{\tilde{\tau}_i} \\
&= \frac{n[(p_i - \tau_{0i}) - (\hat{\tau}_i - \tau_{0i})]}{(n+\alpha)\overline{\tau}_i} + (\frac{1}{\overline{\tau}_i} - \frac{1}{\tilde{\tau}_i})(\hat{\tau}_i - \tau_{0i}).
\end{aligned} \tag{IA.6}$$

Suppose the remainder term $R = (\frac{1}{\overline{\tau}_i} - \frac{1}{\tilde{\tau}_i})(\hat{\tau}_i - \tau_{0i})$. *By Triangle inequality,* $|a-b| \leq |a+b| \leq |a|+|b|$. *Since* $\tilde{\tau}_i \in (0,1)$ *and* $\overline{\tau}_i \in (0,1)$, *both* $\tilde{\tau}_i$ *and* $\overline{\tau}_i$ *are probabilities,* $|\tilde{\tau}_i - \overline{\tau}_i| \leq |\tilde{\tau}_i| + |\overline{\tau}_i| \leq 2$. *And thus,* $|R| \leq \frac{2}{min\{\overline{\tau}_i \tilde{\tau}_i\}} |\overline{\tau}_i - \tilde{\tau}_i| = O|\overline{\tau}_i - \tilde{\tau}_i|$ *provided that* $\boldsymbol{\xi}_0$ *is in the interior of the parameter space, and* $\tau_i, i = 1, ..., K$ *is continuous and positive for* $\boldsymbol{\xi}$ *in the vicinity of* $\boldsymbol{\xi}_0$.

Assuming consistency, $\overline{\tau}_i - \tilde{\tau}_i$ *is small, which means* $\hat{\tau}_i - \tau_{0i} \to 0$ *as* $\alpha \to \infty$ *and* $n \to \infty$. *By utilizing matrix notation, Part (I) in Equation IA.2 can be expressed in an alternative form as follows:*

$$Part(I) = \frac{\alpha n}{(n+\alpha)} \hat{\boldsymbol{\Delta}}^T \bar{\boldsymbol{D}}^{-1} (\boldsymbol{p} - \boldsymbol{\tau}(\boldsymbol{\xi}_0)) - \frac{\alpha n}{(n+\alpha)} \hat{\boldsymbol{\Delta}}^T \bar{\boldsymbol{D}}^{-1} \boldsymbol{\Delta}_0 (\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0) + O_p(\| \hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0 \|). \tag{IA.7}$$

In Equation IA.7, there are several important considerations to keep in mind.

1. *We present the residual term as* $O_p(\| \hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0 \|)$ *due to our assumption of a smooth function in the Delta method. The function has to be differenciable.*

2. *Within Equation IA.6, when considering the R term, we already know that* $(\frac{1}{\overline{\tau}_i} - \frac{1}{\tilde{\tau}_i})$ *is bounded, and* $\boldsymbol{\tau}(\hat{\boldsymbol{\xi}}) - \boldsymbol{\tau}(\boldsymbol{\xi}_0) = \boldsymbol{\Delta}_0(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0)$, $\boldsymbol{\Delta}_0$ *is the derivative and is bounded in the*

vicinity.

3. *As demonstrated in Lemma 8 in Equation IA.45, $\hat{\boldsymbol{\xi}}$ is consistent since it converges in probability to the true $\boldsymbol{\xi}_0$, which means $O_p(\| \hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0 \|) = o_p(1)$. The O_p term in Equation IA.7 is thus negligible.*

Suppose we multiply the estimating equation in Equation IA.2 by a factor of $\sqrt{\frac{n+\alpha}{n\alpha}}$, then the estimating equation becomes

$$\begin{aligned}\frac{\partial \ell_2(\alpha, \boldsymbol{\xi})}{\partial \boldsymbol{\xi}} = &\sqrt{\frac{n\alpha}{n+\alpha}}\hat{\boldsymbol{\Delta}}^T \bar{\boldsymbol{D}}^{-1}(\boldsymbol{p} - \boldsymbol{\tau}(\boldsymbol{\xi}_0)) - \sqrt{\frac{n\alpha}{n+\alpha}}\hat{\boldsymbol{\Delta}}^T \bar{\boldsymbol{D}}^{-1}\boldsymbol{\Delta}_0(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0) \\ &+ \sqrt{\frac{n+\alpha}{n\alpha}}o_p(1) + \sqrt{\frac{n+\alpha}{n\alpha}}Part(II) + \sqrt{\frac{n+\alpha}{n\alpha}}O_p(\frac{1}{\alpha}).\end{aligned} \tag{IA.8}$$

***Part (II)**. The term $\sqrt{\frac{n+\alpha}{n\alpha}}$ times Part (II) in Equation IA.8 is further expanded as*

$$\sqrt{\frac{n+\alpha}{n\alpha}}Part(II) = \frac{1}{2}\sqrt{\frac{1}{n} + \frac{1}{\alpha}}\sum_{i=1}^{K}\frac{\partial \tau_i(\hat{\boldsymbol{\xi}})}{\partial \boldsymbol{\xi}}\frac{1}{\tau_i(\hat{\boldsymbol{\xi}})}\frac{np_i}{np_i + \alpha\tau_i(\hat{\boldsymbol{\xi}})}. \tag{IA.9}$$

In Equation IA.9, $\sqrt{\frac{1}{n} + \frac{1}{\alpha}}$ goes to 0 as $n \to \infty$ and $\alpha \to \infty$. Both $\frac{\partial \tau_i(\hat{\boldsymbol{\xi}})}{\partial \boldsymbol{\xi}}$ and $\frac{1}{\tau_i(\hat{\boldsymbol{\xi}})}$ are bounded. $\frac{np_i}{np_i + \alpha\tau_i(\hat{\boldsymbol{\xi}})} \in (0, 1)$. It can thus be inferred that

$$\sqrt{\frac{n+\alpha}{n\alpha}}Part(II) = o_p(1). \tag{IA.10}$$

It is known that in asymptotic methods, $a_n = o(b_n)$ if the ratio of $|\frac{a_n}{b_n}|$ converges to 0 (e.g., Bishop, Fienberg, & Holland, 2007). In our case, $|\sqrt{\frac{1}{n} + \frac{1}{\alpha}}| \to 0$ as $n \to \infty$ and $\alpha \to \infty$. And thus $\sqrt{\frac{1}{n} + \frac{1}{\alpha}} = o_p(1)$. In addition, we know that $O(a_n)o(b_n) = o(a_n b_n)$, in that case $\sqrt{\frac{n+\alpha}{n\alpha}}O_p(\frac{1}{\alpha}) = O_p(\frac{1}{\alpha})o_p(1) = o_p(\frac{1}{\alpha})$. As $n \to \infty$ and $\alpha \to \infty$, $o_p(\frac{1}{\alpha})$ is small enough and can

be ignored. Furthermore, $\sqrt{\frac{n+\alpha}{n\alpha}}o_p(1) = o_p(1)o_p(1) = o_p(1)$*. Thus, let the esimating equation in Equation IA.8 equal to 0, it can be simplified to*

$$\sqrt{\frac{n\alpha}{n+\alpha}}\hat{\boldsymbol{\Delta}}^T\bar{\boldsymbol{D}}^{-1}\boldsymbol{\Delta}_0(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi}_0) = \sqrt{\frac{n\alpha}{n+\alpha}}\hat{\boldsymbol{\Delta}}^T\bar{\boldsymbol{D}}^{-1}(\boldsymbol{p}-\boldsymbol{\tau}(\boldsymbol{\xi}_0)) + o_p(\frac{1}{\alpha}). \tag{IA.11}$$

Further, it can be deduced that

$$\sqrt{\frac{n\alpha}{n+\alpha}}(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi}_0) \simeq \sqrt{\frac{n\alpha}{n+\alpha}}(\hat{\boldsymbol{\Delta}}^T\bar{\boldsymbol{D}}^{-1}\boldsymbol{\Delta}_0)^{-1}\hat{\boldsymbol{\Delta}}^T\bar{\boldsymbol{D}}^{-1}(\boldsymbol{p}-\boldsymbol{\tau}(\boldsymbol{\xi}_0)). \tag{IA.12}$$

As demonstrated by the consistency results in Lemma 8, $\hat{\boldsymbol{\Delta}}^T\bar{\boldsymbol{D}}^{-1}$ *can be replaced by* $\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}$ *evaluated at* $\boldsymbol{\xi}_0$*. From Proposition 2, we know that* $\sqrt{\frac{n\alpha}{n+\alpha}}(\boldsymbol{p}-\boldsymbol{\tau}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0},\boldsymbol{\Gamma})$*. By rearranging some terms in the approximation in Equation IA.12 we get*

$\sqrt{\frac{n\alpha}{n+\alpha}}(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi}_0) \simeq (\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\sqrt{\frac{n\alpha}{n+\alpha}}(\boldsymbol{p}-\boldsymbol{\tau}(\boldsymbol{\xi}_0))$*.* $(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}$ *is a constant matrix, and* $(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}$ *is invertible. In that case, we get*

$$\sqrt{\frac{n\alpha}{n+\alpha}}(\hat{\boldsymbol{\xi}}-\boldsymbol{\xi}_0) \xrightarrow{d} \mathcal{N}(\mathbf{0},\boldsymbol{Q}\boldsymbol{\Gamma}\boldsymbol{Q}^T). \tag{IA.13}$$

Using the results in Lemma 2, Proposition 5 is demonstrated.

Lemma 2. $\boldsymbol{Q}\boldsymbol{\Gamma}\boldsymbol{Q}^T$ *is equal to* $(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}$*.*

Proof.

$$\boldsymbol{Q}\boldsymbol{\Gamma}\boldsymbol{Q}^T = (\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Gamma}\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}.$$

$\boldsymbol{D}_0^{-1}\boldsymbol{D}_0 = \boldsymbol{I}, \boldsymbol{I}$ *is the identity matrix.* $\boldsymbol{D}_0^{-1}\boldsymbol{\tau}_0 = \mathbb{1}$*,* $\mathbb{1}$ *is the vector of all ones.* $\boldsymbol{\Delta}_0^T\mathbb{1} = \mathbf{0}$*,* $\mathbf{0}$ *is the*

vector of all zeros. Then we know

$$\begin{aligned}\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Gamma} &= \boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}(\boldsymbol{D}_0 - \boldsymbol{\tau}_0\boldsymbol{\tau}_0^T) \\ &= \boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{D}_0 - \boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\tau}_0\boldsymbol{\tau}_0^T \\ &= \boldsymbol{\Delta}_0^T.\end{aligned}$$

Thus,

$$\begin{aligned}\boldsymbol{Q}\boldsymbol{\Gamma}\boldsymbol{Q}^T &= (\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Gamma}\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1} \\ &= (\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1} \\ &= (\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}.\end{aligned}$$

Lemma 3. $\sqrt{\frac{n\alpha}{n+\alpha}}(\boldsymbol{\tau}(\hat{\boldsymbol{\xi}}) - \boldsymbol{\tau}(\boldsymbol{\xi}_0)) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0^T)$.

Proof.

By the 1st order Taylor series expansion, suppose $\boldsymbol{\tau}(\hat{\boldsymbol{\xi}}) = \boldsymbol{\tau}(\boldsymbol{\xi}_0) + \boldsymbol{\tau}'(\tilde{\boldsymbol{\xi}})(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0)$*, in which* $\tilde{\boldsymbol{\xi}}$ *lies between* $\hat{\boldsymbol{\xi}}$ *and* $\boldsymbol{\xi}_0$*. By applying the consistency results in Lemma 8,* $\hat{\boldsymbol{\xi}} \xrightarrow{p} \boldsymbol{\xi}_0$*,* $|\tilde{\boldsymbol{\xi}} - \boldsymbol{\xi}_0| < |\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0|$*, it must be that*

$$\tilde{\boldsymbol{\xi}} \xrightarrow{p} \boldsymbol{\xi}_0. \tag{IA.14}$$

Since $\boldsymbol{\tau}$ *is a continuous function, applying the continuous mapping theorem yields*

$$\boldsymbol{\tau}'(\tilde{\boldsymbol{\xi}}) \xrightarrow{p} \boldsymbol{\tau}'(\boldsymbol{\xi}_0) = \frac{\partial\boldsymbol{\tau}(\boldsymbol{\xi}_0)}{\partial\boldsymbol{\xi}} = \boldsymbol{\Delta}_0. \tag{IA.15}$$

Given this convergence in probability relationship and the convergence in distribution relationship in Proposition 5, by applying Slutsky's theorem,

$$\sqrt{\frac{n\alpha}{n+\alpha}}(\boldsymbol{\tau}(\hat{\boldsymbol{\xi}}) - \boldsymbol{\tau}(\boldsymbol{\xi}_0)) = \boldsymbol{\Delta}_0\sqrt{\frac{n\alpha}{n+\alpha}}(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0) \quad \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0^T). \quad \text{(IA.16)}$$

Lemma 4. $\sqrt{\frac{n\alpha}{n+\alpha}}(\boldsymbol{p} - \boldsymbol{\tau}(\hat{\boldsymbol{\xi}})) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$.

Proof.

On the left-hand side,

$$\sqrt{\frac{n\alpha}{n+\alpha}}(\boldsymbol{p} - \boldsymbol{\tau}(\hat{\boldsymbol{\xi}})) = \sqrt{\frac{n\alpha}{n+\alpha}}[(\boldsymbol{p} - \boldsymbol{\tau}(\boldsymbol{\xi}_0)) - (\boldsymbol{\tau}(\hat{\boldsymbol{\xi}}) - \boldsymbol{\tau}(\boldsymbol{\xi}_0))]. \quad \text{(IA.17)}$$

$\boldsymbol{p} - \boldsymbol{\tau}(\boldsymbol{\xi}_0)$ *and* $\boldsymbol{\tau}(\hat{\boldsymbol{\xi}}) - \boldsymbol{\tau}(\boldsymbol{\xi}_0)$ *are not independent. From the approximation in Equation IA.12, since* $\sqrt{\frac{n\alpha}{n+\alpha}} \neq 0$*,*

$$\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0 \simeq (\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}(\boldsymbol{p} - \boldsymbol{\tau}(\boldsymbol{\xi}_0)). \quad \text{(IA.18)}$$

Because $\boldsymbol{\tau}(\hat{\boldsymbol{\xi}}) - \boldsymbol{\tau}(\boldsymbol{\xi}_0) = \boldsymbol{\Delta}_0(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0)$*, it is inferred that* $\boldsymbol{\tau}(\hat{\boldsymbol{\xi}}) - \boldsymbol{\tau}(\boldsymbol{\xi}_0) \simeq \boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}(\boldsymbol{p} - \boldsymbol{\tau}(\boldsymbol{\xi}_0))$*, and*

$$\begin{aligned}
&\sqrt{\frac{n\alpha}{n+\alpha}}[(\boldsymbol{p} - \boldsymbol{\tau}(\boldsymbol{\xi}_0)) - (\boldsymbol{\tau}(\hat{\boldsymbol{\xi}}) - \boldsymbol{\tau}(\boldsymbol{\xi}_0))] \\
&= \sqrt{\frac{n\alpha}{n+\alpha}}(\boldsymbol{p} - \boldsymbol{\tau}(\boldsymbol{\xi}_0))[\boldsymbol{I} - \boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}]. \quad \text{(IA.19)}
\end{aligned}$$

In the quadratic form, define

$$\boldsymbol{\Sigma} \triangleq (\boldsymbol{I} - \boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1})\boldsymbol{\Gamma}(\boldsymbol{I} - \boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1})^T \tag{IA.20}$$

It is proved in Lemma 4.1 that $\boldsymbol{\Sigma} = \boldsymbol{\Gamma} - \boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0^T$. *It is also proved in Lemma 4.2 that* $\boldsymbol{\Sigma}\boldsymbol{D}_0^{-1}\boldsymbol{\Sigma} = \boldsymbol{\Sigma}$, $\boldsymbol{D}_0^{-1}$ *is the generalized inverse. By applying Proposition 2, we confirmed this asymptotic result.*

Lemma 4.1. $\boldsymbol{\Sigma} = \boldsymbol{\Gamma} - \boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0^T$.

Proof.

By definition,

$$\begin{aligned}\boldsymbol{\Sigma} &\triangleq (\boldsymbol{I} - \boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T \boldsymbol{D}_0{}^{-1})\boldsymbol{\Gamma}(\boldsymbol{I} - \boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T \boldsymbol{D}_0{}^{-1})^T \\ &= (\boldsymbol{\Gamma} - \boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T \boldsymbol{D}_0{}^{-1}\boldsymbol{\Gamma})(\boldsymbol{I} - \boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T \boldsymbol{D}_0{}^{-1})^T\end{aligned} \tag{IA.21}$$

In the equation,

$$(\boldsymbol{I} - \boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T \boldsymbol{D}_0{}^{-1})^T$$

$$= \boldsymbol{I} - \boldsymbol{D}_0{}^{-1}\boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T$$

Put this into the Equation IA.21, now the equation becomes

$$\begin{aligned}\boldsymbol{\Sigma} = &\underbrace{\boldsymbol{\Gamma}}_{\textit{Part 1}} - \underbrace{\boldsymbol{\Gamma}\boldsymbol{D}_0{}^{-1}\boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T}_{\textit{Part 2}} \\ &- \underbrace{\boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T \boldsymbol{D}_0{}^{-1}\boldsymbol{\Gamma}}_{\textit{Part 3}} \\ &+ \underbrace{\boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T \boldsymbol{D}_0{}^{-1}\boldsymbol{\Gamma}\boldsymbol{D}_0{}^{-1}\boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T}_{\textit{Part 4}}.\end{aligned} \tag{IA.22}$$

- *In Part 2,*

$$\begin{aligned}
&\boldsymbol{\Gamma}\boldsymbol{D}_0{}^{-1}\boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T \\
&= (\boldsymbol{D}_0 - \boldsymbol{\tau}_0\boldsymbol{\tau}_0^T)\boldsymbol{D}_0{}^{-1}\boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T \\
&= \boldsymbol{D}_0\boldsymbol{D}_0{}^{-1}\boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T - \boldsymbol{\tau}_0\boldsymbol{\tau}_0^T\boldsymbol{D}_0{}^{-1}\boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T.
\end{aligned}$$

Since $\boldsymbol{D}_0\boldsymbol{D}_0{}^{-1} = \boldsymbol{I}$, $\boldsymbol{\tau}_0^T\boldsymbol{D}_0{}^{-1}\boldsymbol{\Delta}_0 = 0$, *Part 2 becomes* $\boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T$.

- *In Part 3,*

$$\begin{aligned}
&\boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T\boldsymbol{D}_0{}^{-1}(\boldsymbol{D}_0 - \boldsymbol{\tau}\boldsymbol{\tau}^T) \\
&= \boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T\boldsymbol{D}_0{}^{-1}\boldsymbol{D}_0 - \boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T\boldsymbol{D}_0{}^{-1}\boldsymbol{\tau}_0\boldsymbol{\tau}_0^T.
\end{aligned}$$

Since $\boldsymbol{\Delta}_0{}^T\boldsymbol{D}_0{}^{-1}\boldsymbol{\tau}_0 = 0$, *Part 3 is also* $\boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T$.

- *In Part 4,*

$$\begin{aligned}
&\boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T\boldsymbol{D}_0{}^{-1}(\boldsymbol{D}_0 - \boldsymbol{\tau}_0\boldsymbol{\tau}_0^T)\boldsymbol{D}_0{}^{-1}\boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T \\
&= \boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T\boldsymbol{D}_0{}^{-1}\boldsymbol{D}_0\boldsymbol{D}_0{}^{-1}\boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T \\
&- \boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T\boldsymbol{D}_0{}^{-1}\boldsymbol{\tau}_0\boldsymbol{\tau}_0^T\boldsymbol{D}_0{}^{-1}\boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T.
\end{aligned}$$

Since $\boldsymbol{D}_0\boldsymbol{D}_0{}^{-1} = \boldsymbol{I}, \boldsymbol{\Delta}_0{}^T\boldsymbol{D}_0{}^{-1}\boldsymbol{\tau}_0 = 0$,

$$\textit{Part 4} = \boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}(\boldsymbol{\Delta}_0{}^T\boldsymbol{D}_0{}^{-1}\boldsymbol{\Delta}_0)(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T.$$

Because $(\boldsymbol{\Delta}_0{}^T \boldsymbol{D}_0{}^{-1} \boldsymbol{\Delta}_0)(\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1} \boldsymbol{\Delta}_0)^{-1} = \boldsymbol{I}$,

Part 4 is $\boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1} \boldsymbol{\Delta}_0)^{-1} \boldsymbol{\Delta}_0{}^T$.

Therefore, we have successfully demonstrated that $\boldsymbol{\Sigma} = \boldsymbol{\Gamma} - \boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1} \boldsymbol{\Delta}_0)^{-1} \boldsymbol{\Delta}_0^T$.

Lemma 4.2. $\boldsymbol{D}_0^{-1}$ *is the generalized inverse, and that* $\boldsymbol{\Sigma} \boldsymbol{D}_0^{-1} \boldsymbol{\Sigma} = \boldsymbol{\Sigma}$.

Proof.

$$
\begin{aligned}
&\boldsymbol{\Sigma} \boldsymbol{D}_0^{-1} \boldsymbol{\Sigma} \\
&= \underbrace{\boldsymbol{\Gamma} \boldsymbol{D}_0^{-1} \boldsymbol{\Gamma}}_{\textit{Part 1}} - \underbrace{\boldsymbol{\Gamma} \boldsymbol{D}_0^{-1} \boldsymbol{\Delta}_0 (\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1} \boldsymbol{\Delta}_0)^{-1} \boldsymbol{\Delta}_0{}^T}_{\textit{Part 2}} \\
&- \underbrace{\boldsymbol{\Delta}_0 (\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1} \boldsymbol{\Delta}_0)^{-1} \boldsymbol{\Delta}_0{}^T \boldsymbol{D}_0^{-1} \boldsymbol{\Gamma}}_{\textit{Part 3}} \\
&+ \underbrace{\boldsymbol{\Delta}_0 (\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1} \boldsymbol{\Delta}_0)^{-1} \boldsymbol{\Delta}_0{}^T \boldsymbol{D}_0^{-1} \boldsymbol{\Delta}_0 (\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1} \boldsymbol{\Delta}_0)^{-1} \boldsymbol{\Delta}_0{}^T}_{\textit{Part 4}}
\end{aligned}
$$

- *In Part 1,*

$$\boldsymbol{\Gamma} \boldsymbol{D}_0^{-1} \boldsymbol{\Gamma} = \boldsymbol{\Gamma} \boldsymbol{D}_0^{-1} (\boldsymbol{D}_0 - \boldsymbol{\tau}_0 \boldsymbol{\tau}_0^T)$$

$$= \boldsymbol{\Gamma} (\boldsymbol{I} - \boldsymbol{D}_0^{-1} \boldsymbol{\tau}_0 \boldsymbol{\tau}_0^T)$$

Since $\boldsymbol{D}_0^{-1} \boldsymbol{\tau}_0 = \mathbb{1}_{K \times 1}$, *Part 1 is equal to*

$$\boldsymbol{\Gamma}(I - \mathbb{1} \boldsymbol{\tau}^T).$$

In this expression,

$$\boldsymbol{\Gamma} \mathbb{1} = (D - \boldsymbol{\tau}_0 \boldsymbol{\tau}_0^T) \mathbb{1} = D \mathbb{1} - \boldsymbol{\tau}_0 \boldsymbol{\tau}_0^T \mathbb{1}.$$

$D\mathbb{1} = \boldsymbol{\tau}_0, \boldsymbol{\tau}_0^T\mathbb{1} = \sum_{i=1}\tau_i = 1(scalar)$.

Thus, Part 1 is $\boldsymbol{\Gamma} - (\boldsymbol{\tau}_0 - \boldsymbol{\tau}_0) = \boldsymbol{\Gamma}$.

- *Part 2 and Part 3 are the same as Part 2 and 3 in Equation IA.22. Part 2 and Part 3 both equal to* $\boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T$.

- *Part 4 is also the same as part 4 in Equation IA.22. Part 4 is* $\boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T$.

Hence, it is possible to prove that

$$\boldsymbol{\Sigma}\boldsymbol{D}_0^{-1}\boldsymbol{\Sigma} = \boldsymbol{\Gamma} - 2\boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T + \boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T$$

$$= \boldsymbol{\Gamma} - \boldsymbol{\Delta}_0(\boldsymbol{\Delta}_0^T\boldsymbol{D}_0^{-1}\boldsymbol{\Delta}_0)^{-1}\boldsymbol{\Delta}_0{}^T = \boldsymbol{\Sigma},$$

and that $\boldsymbol{D}_0^{-1}$ *is the generalized inverse.*

Lemma 5. *Since we proved Lemma 4 and demonstrated that* $\boldsymbol{D}_0^{-1}$ *is the generalized inverse in Lemma 4.2, by applying the theorem in Lemma 1b in Moore (1977), it can then be inferred that the approximation in Equation 3.24 is true.*

Lemma 6. *In order to obtain an estimate of the concentration parameter* α *using the original parameterization of the log-likelihood for the Dirichlet-Multinomial distribution, we differentiate the LL-DM with respect to* α *and set it equal to zero.*

$$\frac{\partial \ell(\alpha, \boldsymbol{\xi})}{\partial \alpha} = 0$$
$$\Rightarrow \psi(n + \hat{\alpha}) - \psi(\hat{\alpha}) = \sum_{i=1}^{K} \hat{\tau}_i[\psi(np_i + \hat{\alpha}\hat{\tau}_i) - \psi(\hat{\alpha}\hat{\tau}_i)] \tag{IA.23}$$

By Stirling's formula, when both n and α are very large,

$$\begin{cases} \psi(\hat{\alpha}) = \log(\hat{\alpha}) - \frac{1}{2\hat{\alpha}} + O_p(\frac{1}{\hat{\alpha}^2}) \\ \psi(n+\hat{\alpha}) = \log(n+\hat{\alpha}) - \frac{1}{2(n+\hat{\alpha})} + O_p(\frac{1}{(n+\hat{\alpha})^2}) \\ \psi(np_i+\hat{\alpha}\hat{\tau}_i) = \log(np_i+\hat{\alpha}\hat{\tau}_i) - \frac{1}{2(np_i+\hat{\alpha}\hat{\tau}_i)} + O_p(\frac{1}{(np_i+\hat{\alpha}\hat{\tau}_i)^2}) \\ \psi(\hat{\alpha}\hat{\tau}_i) = \log(\hat{\alpha}\hat{\tau}_i) - \frac{1}{2(\hat{\alpha}\hat{\tau}_i)} + O_p(\frac{1}{(\hat{\alpha}\hat{\tau}_i)^2}). \end{cases} \tag{IA.24}$$

Because the O_p term contains only the highest power of $\hat{\alpha}$, among the four O_p terms $O_p(\frac{1}{\hat{\alpha}^2}), O_p(\frac{1}{(n+\hat{\alpha})^2}), O_p(\frac{1}{(np_i+\hat{\alpha}\hat{\tau}_i)^2}), O_p(\frac{1}{(\hat{\alpha}\hat{\tau}_i)^2})$, the highest power is $\hat{\alpha}^2$. The sum of these four O_p terms is $O_p(\frac{1}{\hat{\alpha}^2})$.

Plugging the approximations in Equation IA.24 back to the estimating equation in Equation IA.23, and let $\hat{v}^2 = \frac{1}{\hat{\alpha}^2}$ it is found that

$$\begin{aligned} &\log(n+\hat{\alpha}) - \frac{1}{2(n+\hat{\alpha})} - \log(\hat{\alpha}) + \frac{1}{2\hat{\alpha}} \\ &= \sum_{i=1}^{K} \hat{\tau}_i [\log(np_i+\hat{\alpha}\hat{\tau}_i) - \frac{1}{2(np_i+\hat{\alpha}\hat{\tau}_i)} - \log(\hat{\alpha}\hat{\tau}_i) + \frac{1}{2(\hat{\alpha}\hat{\tau}_i)}] + O_p(\hat{v}^2). \end{aligned} \tag{IA.25}$$

On the right-hand side of Equation IA.25, $\log(\hat{\alpha}\hat{\tau}_i) = \log\hat{\alpha} + \log\hat{\tau}_i$. We also know that $\sum_{i=1}^{K} \hat{\tau}_i \log\hat{\alpha} = \log\hat{\alpha}$, $\sum_{i=1}^{K} \hat{\tau}_i \frac{1}{2\hat{\alpha}\hat{\tau}_i} = \frac{K}{2\hat{\alpha}}$, $\sum_{i=1}^{K} \hat{\tau}_i \log(n+\hat{\alpha}) = \log(n+\hat{\alpha})$. Taking these back to Equation IA.25, now we have

$$\frac{K-1}{2\hat{\alpha}} + \frac{1}{2(n+\hat{\alpha})} - \sum_{i=1}^{K} \frac{\hat{\tau}_i}{2(np_i+\hat{\alpha}\hat{\tau}_i)} + \sum_{i=1}^{K} \hat{\tau}_i [\log(\frac{np_i+\hat{\alpha}\hat{\tau}_i}{n+\hat{\alpha}}) - \log\hat{\tau}_i] + O_p(\hat{v}^2) = 0. \tag{IA.26}$$

In the Equation IA.26, $\frac{np_i+\hat{\alpha}\hat{\tau}_i}{n+\hat{\alpha}}$ is very close to $\hat{\tau}_i$, therefore, by Taylor series expansion

$$\log(\frac{np_i+\hat{\alpha}\hat{\tau}_i}{n+\hat{\alpha}}) = \log\hat{\tau}_i + \frac{1}{\hat{\tau}_i}(\frac{np_i+\hat{\alpha}\hat{\tau}_i}{n+\hat{\alpha}} - \hat{\tau}_i) - \frac{1}{2\bar{\bar{\tau}}_i^2}(\frac{np_i+\hat{\alpha}\hat{\tau}_i}{n+\hat{\alpha}} - \hat{\tau}_i)^2, \tag{IA.27}$$

in which $\bar{\bar{\tau}}_i^2$ is between $\frac{np_i+\hat{\alpha}\hat{\tau}_i}{n+\hat{\alpha}}$ and $\hat{\tau}_i$. Thus, we have

$$\log(\frac{np_i+\hat{\alpha}\hat{\tau}_i}{n+\hat{\alpha}}) - \log\hat{\tau}_i = \frac{1}{\hat{\tau}_i}\frac{n}{n+\hat{\alpha}}(p_i-\hat{\tau}_i) - \frac{1}{2\bar{\bar{\tau}}_i^2}[\frac{n}{n+\hat{\alpha}}(p_i-\hat{\tau}_i)]^2 \tag{IA.28}$$

Returning to Equation IA.26, we already have knowledge that $\sum_{i=1}^{K}\hat{\tau}_i[\frac{1}{\hat{\tau}_i}\frac{n}{n+\hat{\alpha}}(p_i-\hat{\tau}_i)] = 0$. Additionally, based on the consistency results stated in Lemma 8, we know that $\hat{\boldsymbol{\xi}}$ converges in probability to $\boldsymbol{\xi}_0$, $\boldsymbol{\tau}(\hat{\boldsymbol{\xi}})$ converges in probability to $\boldsymbol{\tau}(\boldsymbol{\xi}_0)$ as $n\to\infty$ and $\alpha\to\infty$. Moreover, for large n and α, $\bar{\bar{\tau}}_i \xrightarrow{p} \boldsymbol{\tau}(\boldsymbol{\xi}_0)$. Consequently, we can conclude that

$$\frac{K-1}{\hat{\alpha}} + \frac{1}{n+\hat{\alpha}} - \sum_{i=1}^{K}\frac{\hat{\tau}_i}{np_i+\hat{\alpha}\hat{\tau}_i} + O_p(\hat{v}^2) = \sum_{i=1}^{K}\frac{1}{\tau_{0i}}\frac{n^2}{(n+\hat{\alpha})^2}(p_i-\hat{\tau}_i)]^2. \tag{IA.29}$$

If we multiply $\frac{(n+\hat{\alpha})^2}{n^2}\frac{\alpha_0 n}{\alpha_0+n}$ to both left- and right-hand sides of Equation IA.29, we get

$$\begin{aligned}
&\frac{K-1}{\hat{\alpha}}\frac{(n+\hat{\alpha})^2}{n^2}\frac{\alpha_0 n}{\alpha_0+n} + \frac{1}{n+\hat{\alpha}}\frac{(n+\hat{\alpha})^2}{n^2}\frac{\alpha_0 n}{\alpha_0+n} \\
&- \frac{(n+\hat{\alpha})^2}{n^2}\frac{\alpha_0 n}{\alpha_0+n}\sum_{i=1}^{K}\frac{\hat{\tau}_i}{np_i+\hat{\alpha}\hat{\tau}_i} + O_p(\hat{v}^2\frac{(n+\hat{\alpha})^2}{n^2}\frac{\alpha_0 n}{\alpha_0+n}) \\
&= \sum_{i=1}^{K}\frac{1}{\tau_{0i}}\frac{\alpha_0 n}{\alpha_0+n}(p_i-\hat{\tau}_i)^2 \to \chi^2_{K-q-1}.
\end{aligned} \tag{IA.30}$$

Since $\sum_{i=1}^{K} \frac{(n+\hat{\alpha})\hat{\tau}_i}{np_i+\hat{\alpha}\hat{\tau}_i} \xrightarrow{p} K$ *when* $n \to \infty$ *and* $\alpha \to \infty$*, Equation IA.30 can be simplified to*

$$\frac{\hat{v}+\epsilon}{v_0+\epsilon}(K-1) + O_p[(\hat{v}+\epsilon)(\frac{\hat{v}+\epsilon}{v_0+\epsilon})] \to \chi^2_{K-q-1}. \tag{IA.31}$$

And thus,

$$\frac{\hat{v}+\epsilon}{v_0+\epsilon}[(K-1) + O_p(\hat{v}+\epsilon)] \to \chi^2_{K-q-1}. \tag{IA.32}$$

Given that $O_p(\hat{v}+\epsilon) = o_p(\hat{v}+\epsilon) = o_p(\frac{1}{\hat{\alpha}}+\frac{1}{n}) = o_p(1)$*, we can conclude as stated in Proposition 7.*

Lemma 7. *In order to correct the bias to obtain an unbiased estimator for dispersion parameter, using the concentration parameter* α *in the model parameterization, we let Equation 3.21 equal to 0, which is to say*

$$\psi(m(n+\hat{\alpha})) - \psi(m\hat{\alpha}) = \sum_{i=1}^{K} \hat{\tau}_i[\psi(np_i+\hat{\alpha}\hat{\tau}_i) - \psi(\hat{\alpha}\hat{\tau}_i)]. \tag{IA.33}$$

By Stirling's formula, when both n *and* α *are very large,*

$$\begin{cases} \psi(m\hat{\alpha}) = \log(m\hat{\alpha}) - \frac{1}{2m\hat{\alpha}} + O_p(\frac{1}{(m\hat{\alpha})^2}) \\ \psi(m(n+\hat{\alpha})) = \log(m(n+\hat{\alpha})) - \frac{1}{2m(n+\hat{\alpha})} + O_p(\frac{1}{m^2(n+\hat{\alpha})^2}) \\ \psi(np_i+\hat{\alpha}\hat{\tau}_i) = \log(np_i+\hat{\alpha}\hat{\tau}_i) - \frac{1}{2(np_i+\hat{\alpha}\hat{\tau}_i)} + O_p(\frac{1}{(np_i+\hat{\alpha}\hat{\tau}_i)^2}) \\ \psi(\hat{\alpha}\hat{\tau}_i) = \log(\hat{\alpha}\hat{\tau}_i) - \frac{1}{2(\hat{\alpha}\hat{\tau}_i)} + O_p(\frac{1}{(\hat{\alpha}\hat{\tau}_i)^2}). \end{cases} \tag{IA.34}$$

Because the O_p *term contains only the highest power of* $\hat{\alpha}$*, among the four* O_p *terms*

$O_p(\frac{1}{(m\hat{\alpha})^2}), O_p(\frac{1}{m^2(n+\hat{\alpha})^2}), O_p(\frac{1}{(np_i+\hat{\alpha}\hat{\tau}_i)^2}), O_p(\frac{1}{(\hat{\alpha}\hat{\tau}_i)^2})$, *the highest power is* $\hat{\alpha}^2$. *The sum of these four* O_p *terms is* $O_p(\frac{1}{\hat{\alpha}^2})$.

Plugging the approximations in Equation IA.34 back to the estimating equation in Equation IA.33, and let $\hat{v}^2 = \frac{1}{\hat{\alpha}^2}$ *it is found that*

$$\begin{aligned}
&\log(m(n+\hat{\alpha})) - \frac{1}{2m(n+\hat{\alpha})} - \log(m\hat{\alpha}) + \frac{1}{2m\hat{\alpha}} \\
&= \sum_{i=1}^{K} \hat{\tau}_i [\log(np_i + \hat{\alpha}\hat{\tau}_i) - \frac{1}{2(np_i+\hat{\alpha}\hat{\tau}_i)} - \log(\hat{\alpha}\hat{\tau}_i) + \frac{1}{2(\hat{\alpha}\hat{\tau}_i)}] + O_p(\hat{v}^2). \qquad \text{(IA.35)}
\end{aligned}$$

On the right-hand side of Equation IA.35, $\log(\hat{\alpha}\hat{\tau}_i) = \log\hat{\alpha} + \log\hat{\tau}_i$. *We also know that* $\sum_{i=1}^{K} \hat{\tau}_i \log\hat{\alpha} = \log\hat{\alpha}$, $\sum_{i=1}^{K} \hat{\tau}_i \frac{1}{2\hat{\alpha}\hat{\tau}_i} = \frac{K}{2\hat{\alpha}}$, $\sum_{i=1}^{K} \hat{\tau}_i \log(n+\hat{\alpha}) = \log(n+\hat{\alpha})$. *Taking these back to Equation IA.35, now we have*

$$\frac{Km-1}{2m\hat{\alpha}} + \frac{1}{2m(n+\hat{\alpha})} - \sum_{i=1}^{K} \frac{\hat{\tau}_i}{2(np_i+\hat{\alpha}\hat{\tau}_i)} + \sum_{i=1}^{K} \hat{\tau}_i [\log(\frac{np_i+\hat{\alpha}\hat{\tau}_i}{n+\hat{\alpha}}) - \log\hat{\tau}_i] + O_p(\hat{v}^2) = 0. \qquad \text{(IA.36)}$$

Using the results in Equations IA.27 and IA.28, returning to Equation IA.36, we already have knowledge that $\sum_{i=1}^{K} \hat{\tau}_i [\frac{1}{\hat{\tau}_i} \frac{n}{n+\hat{\alpha}} (p_i - \hat{\tau}_i)] = 0$. *Additionally, based on the consistency results stated in Lemma 8, we know that* $\hat{\boldsymbol{\xi}}$ *converges in probability to* $\boldsymbol{\xi}_0$, $\boldsymbol{\tau}(\hat{\boldsymbol{\xi}})$ *converges in probability to* $\boldsymbol{\tau}(\boldsymbol{\xi}_0)$ *as* $n \to \infty$ *and* $\alpha \to \infty$. *Moreover, for large* n *and* α, $\bar{\hat{\tau}}_i \xrightarrow{p} \boldsymbol{\tau}(\boldsymbol{\xi}_0)$. *Consequently, we can conclude that*

$$\frac{Km-1}{m\hat{\alpha}} + \frac{1}{m(n+\hat{\alpha})} - \sum_{i=1}^{K} \frac{\hat{\tau}_i}{np_i+\hat{\alpha}\hat{\tau}_i} + O_p(\hat{v}^2) = \sum_{i=1}^{K} \frac{1}{\tau_{0i}} \frac{n^2}{(n+\hat{\alpha})^2} (p_i - \hat{\tau}_i)^2. \qquad \text{(IA.37)}$$

If we multiply $\frac{(n+\hat{\alpha})^2}{n^2}\frac{\alpha_0 n}{\alpha_0+n}$ to both left- and right-hand sides of Equation IA.37, we get

$$\begin{aligned}
&\frac{Km-1}{m\hat{\alpha}}\frac{(n+\hat{\alpha})^2}{n^2}\frac{\alpha_0 n}{\alpha_0+n}+\frac{1}{m(n+\hat{\alpha})}\frac{(n+\hat{\alpha})^2}{n^2}\frac{\alpha_0 n}{\alpha_0+n}\\
&-\frac{(n+\hat{\alpha})^2}{n^2}\frac{\alpha_0 n}{\alpha_0+n}\sum_{i=1}^{K}\frac{\hat{\tau}_i}{np_i+\hat{\alpha}\hat{\tau}_i}+O_p(\hat{v}^2\frac{(n+\hat{\alpha})^2}{n^2}\frac{\alpha_0 n}{\alpha_0+n})\\
&=\frac{\alpha_0 n}{\alpha_0+n}\sum_{i=1}^{K}\frac{1}{\tau_{0i}}(p_i-\hat{\tau}_i)^2. \qquad \text{(IA.38)}
\end{aligned}$$

Suppose $\sum_{i=1}^{K}\frac{(n+\hat{\alpha})\hat{\tau}_i}{np_i+\hat{\alpha}\hat{\tau}_i}=K_{n,\alpha}$. Then the Equation IA.38 can be simplified to

$$\begin{aligned}
&\frac{(mK-1)\alpha_0(n+\hat{\alpha})n}{mn(\alpha_0+n)\hat{\alpha}}+\frac{(mK-1)\alpha_0(n+\hat{\alpha})\hat{\alpha}}{mn(\alpha_0+n)\hat{\alpha}}+\frac{(1-mK_{n,\alpha})\alpha_0\hat{\alpha}(n+\hat{\alpha})}{mn(\alpha_0+n)\hat{\alpha}}\\
&\to\chi^2_{K-q-1}. \qquad \text{(IA.39)}
\end{aligned}$$

The first term in Equation IA.39 can be simplified as

$$\frac{(mK-1)\alpha_0(n+\hat{\alpha})n}{mn(\alpha_0+n)\hat{\alpha}}=\frac{mK-1}{m}\frac{\hat{v}+\epsilon}{v_0+\epsilon}. \qquad \text{(IA.40)}$$

The second and third terms in Equation IA.39 can be simplified as

$$\frac{(mK-1)\alpha_0(n+\hat{\alpha})\hat{\alpha}}{mn(\alpha_0+n)\hat{\alpha}}+\frac{(1-mK_{n,\alpha})\alpha_0\hat{\alpha}(n+\hat{\alpha})}{mn(\alpha_0+n)\hat{\alpha}}=\frac{\hat{\alpha}}{n}\frac{\hat{v}+\epsilon}{v_0+\epsilon}(K-K_{n,\alpha}). \qquad \text{(IA.41)}$$

Then, in Equation IA.41, it is found that

$$\begin{aligned}
&\frac{\hat{\alpha}}{n}|K - K_{n,\alpha}| \\
&= \frac{\hat{\alpha}}{n}|\sum_{i=1}^{K}[\mathbb{1} - \frac{(n+\hat{\alpha})\hat{\tau}_i}{np_i + \hat{\alpha}\hat{\tau}_i}]| \\
&\textit{By Jensen's inequality,} \\
&\leq \frac{\hat{\alpha}}{n}\sum_{i=1}^{K}|\mathbb{1} - \frac{(n+\hat{\alpha})\hat{\tau}_i}{np_i + \hat{\alpha}\hat{\tau}_i}| \\
&= \sum_{i=1}^{K}\frac{1}{\hat{\tau}_i}|p_i - \hat{\tau}_i| \\
&\because \frac{1}{\hat{\tau}_i} \xrightarrow{p} \frac{1}{\hat{\tau}_0} \textit{ is bounded, and } |p_i - \hat{\tau}_i| \xrightarrow{p} 0 \\
&\therefore \frac{\hat{\alpha}}{n}|K - K_{n,\alpha}| = o_p(1).
\end{aligned} \tag{IA.42}$$

The Equation IA.38 is now

$$\begin{aligned}
&[\frac{mK-1}{m} + o_p(1)](\frac{\hat{v}+\epsilon}{v_0+\epsilon}) \to \chi^2_{K-q-1} \\
&\Rightarrow (\frac{\hat{v}+\epsilon}{v_0+\epsilon}) \xrightarrow{d} \frac{\chi^2_{K-q-1}}{\frac{mK-1}{m}}.
\end{aligned} \tag{IA.43}$$

Because we want to correct the bias when estimating the dispersion parameter, we need to let

$$\begin{aligned}
&E[\frac{m(K-q-1)}{mK-1}] = 1 \\
&\Rightarrow m = \frac{1}{1+q}.
\end{aligned} \tag{IA.44}$$

Thus we claim that letting $m = \frac{1}{1+q}$ can correct the bias when estimating the dispersion parameter.

Lemma 8.

$$\hat{\boldsymbol{\xi}} \xrightarrow{p} \boldsymbol{\xi}_0 \text{ as } n \to \infty \text{ and } \alpha \to \infty. \tag{IA.45}$$

$$\boldsymbol{\tau}(\hat{\boldsymbol{\xi}}) \xrightarrow{p} \boldsymbol{\tau}(\boldsymbol{\xi}_0) \text{ as } n \to \infty \text{ and } \alpha \to \infty. \tag{IA.46}$$

$$\frac{\partial \tau(\hat{\boldsymbol{\xi}})}{\partial \boldsymbol{\xi}} \xrightarrow{p} \frac{\partial \tau(\boldsymbol{\xi}_0)}{\partial \boldsymbol{\xi}} \text{ as } n \to \infty \text{ and } \alpha \to \infty. \tag{IA.47}$$

Proof.

1. *In order to demonstrate the convergence in probability of* $\hat{\boldsymbol{\xi}}$ *to* $\boldsymbol{\xi}_0$ *in Equation IA.45, we need to establish that for any given* $\eta > 0$*, the probability* $P(|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0| > \eta)$ *approaches 0 as* $n \to \infty$ *and* $\alpha \to \infty$*. In other words, we aim to show that* $\lim_{\substack{n\to\infty \\ \alpha\to\infty}} P(|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0| > \eta) = 0$*. From Proposition 5 we know that* $\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0$ *converges in distribution to a normal distribution, given that* $\frac{\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0}{\sqrt{(v_0+\epsilon)[\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1} \boldsymbol{\Delta}_0]^{ii}}} \xrightarrow{d} N(0,1)$*, we can conclude that*

$$\lim_{\substack{n\to\infty \\ \alpha\to\infty}} P\left(\frac{|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0|}{\sqrt{(v_0+\epsilon)[\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1} \boldsymbol{\Delta}_0]^{ii}}} > t\right) = 1 - \phi(t),$$

in which ϕ *is the cumulative distribution function (CDF) of a standard normal distribution. The notation* $[]^{ii}$ *refers to the element located in the* i*-th row and* i*-th column within the diagonal of the inverse of a matrix. And thus,*

$$P(|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0| > \eta) = P\left(\frac{|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0|}{\sqrt{(v_0+\epsilon)[\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1} \boldsymbol{\Delta}_0]^{ii}}} > \frac{\eta}{\sqrt{(v_0+\epsilon)[\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1} \boldsymbol{\Delta}_0]^{ii}}}\right).$$

Let $\delta = \frac{\eta}{\sqrt{(v_0+\epsilon)[\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1} \boldsymbol{\Delta}_0]^{ii}}}$*, we have*

$$P(|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0| > \eta) = P(\frac{|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0|}{\sqrt{(v_0+\epsilon)[\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1} \boldsymbol{\Delta}_0]^{ii}}} > \delta),$$

and

$$\lim_{\substack{n\to\infty \\ \alpha\to\infty}} P(\frac{|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0|}{\sqrt{(v_0+\epsilon)[\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1} \boldsymbol{\Delta}_0]^{ii}}} > \delta) = 1 - \phi(\delta).$$

Since ϕ *is a continuous function and* $\phi(t) \to 1$ *as* $t \to \infty$*,*

$$\lim_{\substack{n\to\infty \\ \alpha\to\infty}} P(\frac{|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0|}{\sqrt{(v_0+\epsilon)[\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1} \boldsymbol{\Delta}_0]^{ii}}} > \delta) = 1 - \phi(\delta) \to 0 \text{ as } \delta \to \infty.$$

Therefore,

$$\lim_{\substack{n\to\infty \\ \alpha\to\infty}} P(|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0| > \eta) = \lim_{\delta\to\infty} P(\frac{|\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}_0|}{\sqrt{(v_0+\epsilon)[\boldsymbol{\Delta}_0^T \boldsymbol{D}_0^{-1} \boldsymbol{\Delta}_0]^{ii}}} > \delta) = 0,$$

such that $\hat{\boldsymbol{\xi}} \xrightarrow{p} \boldsymbol{\xi}_0$ *as* $n \to \infty$ *and* $\alpha \to \infty$*.*

2. *The convergence described in Equation IA.46 can be demonstrated using a similar procedure as in Equation IA.45, employing the results presented in Lemma 3.*

3. *We assume the* $\boldsymbol{\tau}$ *function is continuously differentiable. It is shown that* $\boldsymbol{\tau}(\hat{\boldsymbol{\xi}}) \xrightarrow{p} \boldsymbol{\tau}(\boldsymbol{\xi}_0)$ *as* $n \to \infty$ *and* $\alpha \to \infty$*. The continuous mapping theorem states that if* g *is a continuous function and* Y_n *converges in probability to* Y*, the* $g(Y_n)$ *converges in probability to* $g(Y)$*. By applying the continuous mapping theorem, we can establish Equation IA.47 and conclude that* $\frac{\partial \tau(\hat{\boldsymbol{\xi}})}{\partial \boldsymbol{\xi}} \xrightarrow{p} \frac{\partial \tau(\boldsymbol{\xi}_0)}{\partial \boldsymbol{\xi}}$ *as* $n \to \infty$ *and* $\alpha \to \infty$*.*

Reference

Abramowitz, M., & Stegun, I. (1964). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. U.S. Government Printing Office.

Agresti, A. (2003). *Categorical data analysis*. Wiley.

Akaike, H. (1998). Factor analysis and aic. In E. Parzen, K. Tanabe, & G. Kitagawa (Eds.), *Selected papers of hirotugu akaike* (pp. 371–386). New York, NY: Springer New York. doi: 10.1007/978-1-4612-1694-0_29

Arjovsky, M., Chintala, S., & Bottou, L. (2017, 06–11 Aug). Wasserstein generative adversarial networks. In D. Precup & Y. W. Teh (Eds.), *Proceedings of the 34th international conference on machine learning* (Vol. 70, pp. 214–223). PMLR.

Asparouhov, T., & Muthen, B. (2007). Computationally efficient estimation of multilevel high-dimensional latent variable models. In *proceedings of the 2007 jsm meeting in salt lake city, utah, section on statistics in epidemiology* (pp. 2531–2535).

Belsley, D. A., Kuh, E., & Welsch, R. E. (2005). *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons.

Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual review of psychology*, *31*(1), 419–456.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238.

Bentler, P. M. (1995). *Eqs structural equations program manual* (Vol. 6). Multivariate software

Encino, CA.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, *88*(3), 588.

Bickel, P. J., & Doksum, K. A. (2015). *Mathematical statistics: basic ideas and selected topics, volumes i-ii package*. Chapman and Hall/CRC.

Bishop, Y. M., Fienberg, S. E., & Holland, P. W. (2007). *Discrete multivariate analysis: theory and practice*. Springer Science & Business Media.

Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an em algorithm. *Psychometrika*, *46*(4), 443–459.

Bock, R. D., & Lieberman, M. (1970). Fitting a response model forn dichotomously scored items. *Psychometrika*, *35*(2), 179–197.

Bollen, K. A., & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods & Research*, *21*(2), 205-229. doi: 10.1177/0049124192021002004

Briggs, N. E., & MacCallum, R. C. (2003). Recovery of weak common factors by maximum likelihood and ordinary least squares estimation. *Multivariate Behavioral Research*, *38*(1), 25-56. doi: 10.1207/S15327906MBR3801_2

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, *21*(2), 230-258. doi: 10.1177/0049124192021002005

Casella, G., & Berger, R. (2021). *Statistical inference*. Cengage Learning.

Chalmers, P. (2012). mirt: A multidimensional item response theory package for the r environ-

ment. *Journal of statistical Software*, *48*, 1–29.

Chalmers, P., Pritikin, J., Robitzsch, A., Zoltak, M., KwonHyun, K., Falk, C., & Meade, A. (2015). Package 'mirt'. *Zugriff am*, *21*(01), 2016.

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in rmsea test statistic in structural equation models. *Sociological Methods & Research*, *36*(4), 462-494. doi: 10.1177/0049124108314720

Cudeck, R., & Henly, S. J. (1991). Model selection in covariance structures analysis and the "problem" of sample size: a clarification. *Psychological Bulletin*, *109*(3), 512.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1-22. doi: https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

Dodge, Y., Cox, D., & Commenges, D. (2003). *The oxford dictionary of statistical terms*. Oxford University Press on Demand.

Henderson, H. V., & Searle, S. R. (1979). Vec and vech operators for matrices, with some uses in jacobians and multivariate statistics. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, *7*(1), 65–81.

Higham, D. J. (1995). Condition numbers and their condition numbers. *Linear Algebra and its Applications*, *214*, 193–213.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic journal of business research methods*, *6*(1), pp53–60.

Hu, L.-t., & Bentler, P. M. (1997). Selecting cutoff criteria for fit indexes for model evaluation:

Conventional versus new alternatives. *Santa Cruz, CA: University of California, Santa Cruz.*.

Hu, L.-t., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological methods*, *3*(4), 424.

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

Kang, T., & Cohen, A. S. (2007). Irt model selection methods for dichotomous items. *Applied Psychological Measurement*, *31*(4), 331-358. doi: 10.1177/0146621606292213

Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(3), 425-464. doi: https://doi.org/10.1111/1467-9868.00294

Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79–86.

Lai, K. (2019). A simple analytic confidence interval for cfi given nonnormal data. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(5), 757-777. doi: 10.1080/10705511.2018.1562351

Lawley, D. N., & Maxwell, A. E. (1962). Factor analysis as a statistical method. *Journal of the Royal Statistical Society. Series D (The Statistician)*, *12*(3), 209–229.

Li, L., & Bentler, P. (2006). Robust statistical tests for evaluating the hypothesis of close fit of misspecified mean and covariance structural models.

Liu, Y., Yang, J. S., & Maydeu-Olivares, A. (2019). Restricted recalibration of item response theory models. *psychometrika*, *84*(2), 529–553.

Louis, T. A. (1982). Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *44*(2), 226–233.

MacCallum, R. C., Browne, M. W., & Cai, L. (2006). Testing differences between nested covariance structure models: Power analysis and null hypotheses. *Psychological methods*, *11*(1), 19.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological methods*, *1*(2), 130.

MacCallum, R. C., & O'Hagan, A. (2015). Advances in modeling model discrepancy: Comment on wu and browne (2015). *psychometrika*, *80*(3), 601–607.

Magnus, B. E., & Liu, Y. (2022). Symptom presence and symptom severity as unique indicators of psychopathology: An application of multidimensional zero-inflated and hurdle graded response models. *Educational and Psychological Measurement*, *82*(5), 938–966.

Maiti, S. S., & Mukherjee, B. N. (1990). A note on distributional properties of the jöreskog-sörbom fit indices. *Psychometrika*, *55*(4), 721–726.

Maydeu-Olivares, A. (2017). Assessing the size of model misfit in structural equation models. *Psychometrika*, *82*(3), 533–558.

Maydeu-Olivares, A., & Joe, H. (2005). Limited-and full-information estimation and goodness-of-fit testing in 2n contingency tables: A unified framework. *Journal of the American Statistical Association*, *100*(471), 1009–1020.

Maydeu-Olivares, A., & Joe, H. (2008). An overview of limited information goodness-of-fit testing in multidimensional contingency tables. *New trends in psychometrics*, 253–262.

Maydeu-Olivares, A., & Joe, H. (2014). Assessing approximate fit in categorical data analysis. *Multivariate Behavioral Research*, *49*(4), 305-328. doi: 10.1080/00273171.2014.911075

Maydeu-Olivares, A., & Liu, Y. (2015). Item diagnostics in multivariate discrete data. *Psychological Methods*, *20*(2), 276.

Maydeu-Olivares, A., Shi, D., & Rosseel, Y. (2018). Assessing fit in structural equation models: A monte-carlo evaluation of rmsea versus srmr confidence intervals and tests of close fit. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(3), 389-402. doi: 10.1080/10705511.2017.1389611

Minka, T. P. (2000). *Estimating a dirichlet distribution* (Tech. Rep.).

Moore, D. S. (1977). Generalized inverses, wald's method, and the construction of chi-squared tests of fit. *Journal of the American Statistical Association*, *72*(357), 131–137.

Murphy, S. A., & van der Vaart, A. W. (2000). On profile likelihood. *Journal of the American Statistical Association*, *95*(450), 449-465. doi: 10.1080/01621459.2000.10474219

Oakley, J., & O'Hagan, A. (2002, 12). Bayesian inference for the uncertainty distribution of computer model outputs. *Biometrika*, *89*(4), 769-784. doi: 10.1093/biomet/89.4.769

O'Hagan, A., Bernardo, J. M., Berger, J. O., Dawid, A. P., (eds, A. F. M. S., Kennedy, M. C., & Oakley, J. E. (1998). *Uncertainty analysis and other inference tools for complex computer codes.*

R Core Team. (2022). R: A language and environment for statistical computing [Computer

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

Rasch, G. (1960). Studies in mathematical psychology: I. probabilistic models for some intelligence and attainment tests.

Reif, K., & Melich, A. (1995). *Euro-barometer 38.1: Consumer protection and perceptions of science and technology, november 1992*. Inter-university Consortium for Political and Social Research.

Ripley, B. D. (2003). Model selection in complex classes of models. In *Statistical learning, amsi summer school meeting at unsw (université de nouvelle-galles du sud).*

Rizopoulos, D. (2007). ltm: An r package for latent variable modeling and item response analysis. *Journal of statistical software*, *17*, 1–25.

Rizopoulos, D., & Rizopoulos, M. D. (2018). Package 'ltm'. *URL http://wiki. r-project. org/rwiki/doku. php*.

Rupp, A. A. (2013). A systematic review of the methodology for person fit research in item response theory: Lessons about generalizability of inferences from the design of simulation studies. *Psychological Test and Assessment Modeling*, *55*(1), 3.

Samejima, F. (1969, Mar 01). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, *34*(1), 1-97. doi: 10.1007/BF03372160

Särndal, C.-E., Swensson, B., & Wretman, J. (2003). *Model assisted survey sampling*. Springer Science & Business Media.

Satorra, A. (2015). A comment on a paper by h. wu and mw browne (2014). *psychometrika*, *80*(3), 613–618.

Shapiro, A. (2015). Comments on "quantifying adventitious error in a covariance structure as a random effect" by hao wu and michael browne. *psychometrika*, *80*(3), 611–612.

Sprott, D. A. (2008). *Statistical inference in science*. Springer Science & Business Media.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*(2), 173-180. doi: 10.1207/s15327906mbr2502_4

Steiger, J. H. (2016). Notes on the steiger–lind (1980) handout. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(6), 777-781. doi: 10.1080/10705511.2016.1217487

Steyer, R., Sengewald, E., & Hahn, S. (2015). Some comments on wu and browne. *psychometrika*, *80*(3), 608–610.

Thissen, D. (1982). Marginal maximum likelihood estimation for the one-parameter logistic model. *Psychometrika*, *47*(2), 175–186.

Trucano, T., Swiler, L., Igusa, T., Oberkampf, W., & Pilch, M. (2006). Calibration, validation, and sensitivity analysis: What's what. *Reliability Engineering & System Safety*, *91*(10), 1331–1357. doi: https://doi.org/10.1016/j.ress.2005.11.031

Tucker, L. R., Koopman, R. F., & Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, *34*(4), 421–459.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*(1), 1–10.

Wu, H., & Browne, M. W. (2015a). Quantifying adventitious error in a covariance structure as a random effect. *Psychometrika*, *80*(3), 571–600.

Wu, H., & Browne, M. W. (2015b). Random model discrepancy: Interpretations and technicalities (a rejoinder). *psychometrika*, *80*(3), 619–624.

Yang, J. S., Hansen, M., & Cai, L. (2012). Characterizing sources of uncertainty in item response theory scale scores. *Educational and psychological measurement*, *72*(2), 264–290.

Yee, T. W. (2010). The vgam package for categorical data analysis. *Journal of Statistical Software*, *32*, 1–34.

Yee, T. W. (2012). *Vgam: Vector generalized linear and additive models. r package.*

Yee, T. W., & Wild, C. (1996). Vector generalized additive models. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(3), 481–493.

Yu, P., & Shaw, C. A. (2014). An efficient algorithm for accurate computation of the dirichlet-multinomial log-likelihood function. *Bioinformatics*, *30*(11), 1547–1554.

Yuan, K.-H., Cheng, Y., & Patton, J. (2014). Information matrices and standard errors for mles of item parameters in irt. *Psychometrika*, *79*, 232–254.

Yuan, K.-H., Hayashi, K., & Bentler, P. M. (2007). Normal theory likelihood ratio statistic for mean and covariance structure analysis under alternative hypotheses. *Journal of Multivariate Analysis*, *98*(6), 1262-1282. doi: https://doi.org/10.1016/j.jmva.2006.08.005