Chapter 1
Assessment of Misspecification in CDMs using a Generalized Information Matrix Test

Reyhaneh Hosseinpourkhoshkbari [ID] and Richard M. Golden [ID]

Abstract If the probability model is correctly specified, then we can estimate the covariance matrix of the asymptotic maximum likelihood estimate distribution using either the first or second derivatives of the likelihood function. Therefore, if the determinants of these two different covariance matrix estimation formulas differ this indicates model misspecification. This misspecification detection strategy is the basis of the Determinant Information Matrix Test ($GIMT_{Det}$). To investigate the performance of the $GIMT_{Det}$, a Deterministic Input Noisy And gate (DINA) Cognitive Diagnostic Model (CDM) was fit to the Fraction-Subtraction dataset. Next, various misspecified versions of the original DINA CDM were fit to bootstrap data sets generated by sampling from the original fitted DINA CDM. The $GIMT_{Det}$ showed good discrimination performance for larger levels of misspecification. In addition, the $GIMT_{Det}$ did not detect model misspecification when model misspecification was not present and additionally did not detect model misspecification when the level of misspecification was very low. However, the $GIMT_{Det}$ discrimination performance was highly variable across different misspecification strategies when the misspecification level was moderately sized. The proposed new misspecification detection methodology is promising but additional empirical studies are required to further characterize its strengths and limitations.

Key words: misspecification, information matrix test, cognitive diagnostic model

Reyhaneh Hosseinpourkhoshkbari
Cognitive Informatics and Statistics Lab, School of Behavioral and Brain Sciences (GR4.302), University of Texas at Dallas, Richardson, TX. e-mail: reyhaneh.hosseinpour@utdallas.edu

Richard M. Golden
Cognitive Informatics and Statistics Lab, School of Behavioral and Brain Sciences (GR4.1), University of Texas at Dallas, Richardson, TX. e-mail: golden@utdallas.edu

1.1 Introduction

Cognitive diagnostic models (CDMs) are a family of restricted latent class psychometric models designed to assess examinee skill mastery by incorporating prior knowledge of the relationship between latent skills and student responses (de la Torre, 2009). In addition, a CDM outputs the attribute distribution or the probability that an examinee will have a particular set of skills given the examinee's exam performance. Characteristics of latent skills in CDMs are defined through a Q-matrix that specifies which specific skills are relevant for answering a particular exam question. Furthermore, Q-matrix misspecification can affect parameter estimates and respondent classification accuracy (Rupp and Templin, 2008). Therefore, careful validation of the Q-matrix is crucial to ensure that the model accurately represents the underlying relationships between the test items and the measured attributes. Methods have been developed to estimate and validate expert knowledge. However, despite best intentions, the possibility of CDM misspecification is always present.

The effects of model misspecification in CDMs have been investigated by directly comparing observed frequencies and predicted probabilities (Kunina-Habenicht et al., 2012), by comparing a correctly specified model to a nested model using Wald test methodology (de la Torre, 2011), and comparing selected first and second order observed and expected moments (Chen et al., 2018). These methodologies often face the challenge of test statistics with poor statistical power due to the difficulty of reliably estimating parameters in fully saturated (highly flexible models). For example, consider the problem of comparing the predicted probability of every pattern of responses to an exam with d items using a probability model with q free parameters. The degrees of freedom (and variance) for the Pearson Goodness-Of-Fit (GOF) chi-squared test statistic increases as an exponential function of the number of items d. The M2 statistic (Maydeu-Olivares and Joe, 2014) provides an improvement in statistical power by only examining the first and second moments resulting in a chi-squared test statistic whose degrees of freedom increase as a quadratic function of the number of items d.

Using a different approach, which detects model misspecification by comparing different covariance matrix estimators, White (1982) introduced the Information Matrix Test (IMT) which is a chi-squared test with $q(q-1)/2$ degrees of freedom. This test is based on comparing the inverse covariance matrix estimators: $\hat{\mathbf{A}}^{-1}$, derived from the Hessian matrix, and $\hat{\mathbf{B}}^{-1}$, derived from the outer product gradient (OPG). These estimators are calculated from the second and first derivatives of the log-likelihood function, respectively. The White (1982) IMT test statistic for misspecification detection has a chi-squared distribution whose degrees of freedom increase as a quadratic function of the number of parameters q in the model. Golden et al. (2013, 2016) proposed a generalization of the White (1982) IMT framework called the Generalized Information Matrix Test (GIMT) framework. However, no systematic studies have been conducted to specifically investigate the performance of the GIMT in the context of model misspecification detection in CDMs.

Presnell and Boos (2004) developed and empirically evaluated a statistical test for comparing the "in-sample" (training data) log-likelihood model fit to the "out-

of-sample" (test data) log-likelihood model fit. Presnell and Boos (2004) referred to their test as the "in and out of sample" (IOS) test and showed the test statistic had a chi-squared distribution with 1 degree of freedom regardless of model or data complexity. The IOS test may be interpreted as a type of GIMT as described by Golden et al. (2013, 2016) which we call the *Determinant GIMT* ($GIMT_{Det}$) $GIMT_{Det} = (1/q) \log \det(\hat{\mathbf{A}}_n^{-1} \hat{\mathbf{B}}_n)$.

More recently, Liu et al. (2019) examined the performance of different methods in providing consistent standard errors (SEs) for item parameter estimates for situations where a CDM was misspecified or correctly specified. Liu et al. (2019) showed a difference among OPG, Hessian, and Robust standard errors when the Q-matrix is misspecified. Although Liu et al. (2019) had not intended to develop a method for misspecification detection, their empirical results nevertheless support a type of GIMT for misspecification detection as introduced by Golden et al. (2013, 2016).

In this paper, we describe another GIMT methodology that focuses on a single statistic for comparing covariance matrices rather than a comparison based upon their diagonal elements. First, we sketch the mathematical derivation of the asymptotic distribution of the $GIMT_{Det}$ statistic for CDMs using the methods described by Golden et al. (2013, 2016). Second, we empirically investigate the asymptotic behavior of the $GIMT_{Det}$ for CDMs in a series of simulation studies. The simulation studies simulate data sets from a known DINA CDM fit to the Tatsuoka (1984) Fraction-Subtraction data set. Next, the resulting bootstrapped simulated data sets are fit to the DINA CDM which generated the bootstrap data as well as different misspecified versions of the original DINA CDM. The discrimination performance of the $GIMT_{Det}$ is then reported to provide an empirical evaluation of the ability of the $GIMT_{Det}$ to detect model misspecification in a DINA CDM.

1.2 Mathematical Theory

1.2.1 Model Misspecification

In the statistical machine learning framework, the Data Generating Process (DGP) generates observable data directly from an underlying unobservable probability distribution, p_{DGP}. A probability model, M, is a collection of probability mass functions. If $p_{DGP} \in M$, then M is correctly specified; otherwise, M is misspecified with respect to p_{DGP}. Less formally, a model capable of representing the DGP is called a "correctly specified model."

1.2.2 Cognitive Diagnostic Model Specification

1.2.2.1 Data Set

Consider a scenario where N examinees are randomly selected to take a diagnostic test of J items. The outcomes of this test are recorded in a binary matrix $\mathbf{X}$, which has N rows and J columns. Each element, x_{ij}, represents the response of examinee i ($1 \le i \le N$) to item j ($1 \le i \le N$), with a value of 1 indicating a correct response and 0 indicating an incorrect response. The rows of the matrix $\mathbf{X}$, $\mathbf{x}_1, \ldots, \mathbf{x}_n$, are assumed to be a realization of a sequence of independent and identically distributed random vectors with a common probability mass function $p_{DGP}(x)$.

1.2.2.2 Evidence Model

Let $\alpha = (\alpha_1, \ldots, \alpha_k)$ represent the discrete attribute mastery profile for an examinee, where α_k is one if and only if that examinee demonstrates mastery of latent skill k. Let $\mathbf{q}_j = [q_{j1}, \ldots, q_{jk}]$ represent the jth row of the $\mathbf{Q}$-matrix, where each element, q_{jk} equals one if the kth skill is required to answer question j and is equal to zero otherwise. Let logistic sigmoidal function $\mathcal{S}(\phi)$ be defined such that $\mathcal{S}(\phi) = 1/(1 + \exp(-\phi))$. Let $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, .., \boldsymbol{\beta}_j]^T$ denote item parameters vector, where $\boldsymbol{\beta}_j = [\beta_{j1}, \beta_{j2}]^T$ and β_{j1} is the main and interaction effect parameters for question j, and β_{j2} is the intercept parameter. The probability of a correct response to question j, given the mastery profile, item parameters, and the skills required for question j, specified by $\mathbf{q}_j$, is calculated using the following formula:

$$p_{ij} = p\left(x_{ij} = 1 | \alpha, \boldsymbol{\beta}_j, \boldsymbol{q}_j\right) = \mathcal{S}(\boldsymbol{\beta}_{j1}\psi(\alpha, \mathbf{q}_j) - \boldsymbol{\beta}_{j2}).$$

where larger values of $\psi(\alpha, \mathbf{q}_j)$ indicate an increased likelihood that item j is correctly answered given that latent skill pattern α. We have implemented the DINA CDM by defining $\psi(\boldsymbol{\alpha}, \boldsymbol{q}_j) = (\prod_{k=1}^{K} \alpha_k^{q_{j,k}}) \in \{0, 1\}$ to calculate the expected response to item j.
Let:
$$p\left(x_{ij} | \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{q}_j\right) = x_{ij} p_{ij} + (1 - x_{ij})(1 - p_{ij}).$$

The probability of all observed responses for ith examinee, x_i, given a specific pattern of latent skills, can be expressed as follows:

$$p(x_i | \boldsymbol{\alpha}, \boldsymbol{\beta}_j) = \prod_{j=1}^{J} p(x_{ij} | \boldsymbol{\alpha}, \boldsymbol{\beta}_j, \boldsymbol{q}_j).$$

1.2.2.3 Proficiency Model

After specifying the conditional distribution of student responses given latent skill profiles, the subsequent emphasis will be on investigating the joint distribution of attributes. The saturated joint attribute distribution model for all possible values that the k-dimensional binary vector $\boldsymbol{\alpha}$ can take on requires $2^k - 1$ parameters, so when the number of attributes is moderately large, a more constrained joint attribute model might be desired to represent the joint distribution of $\boldsymbol{\alpha}$. In this paper, a Bernoulli latent skill attribute probability model (e.g., Maris, 1999) is assumed where the latent skills are independent so that the probability that the kth latent skill, α_k, is present in the attribute pattern is given by the formula:

$$p(\alpha_k) = \alpha_k \mathcal{S}(-\eta_k) + (1 - \alpha_k)(1 - \mathcal{S}(-\eta_k))$$

where η_k can be a free parameter; however, in this paper, η_k is a constant chosen such that the guess probability $\mathcal{S}(-\boldsymbol{\eta})$ is 0.354. The probability, $p(\boldsymbol{\alpha}|\boldsymbol{\eta})$, of a skill attribute profile, $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_K]$, for an examinee given η is given by the formula:

$$p(\boldsymbol{\alpha}) = \prod_{k=1}^{K} p(\alpha_k).$$

1.2.3 Model Parameter Estimation

The parameter prior for the two-dimensional parameter vector $\boldsymbol{\beta}_j$, associated with the jth question, is represented by a bivariate Gaussian density, denoted as $p(\boldsymbol{\beta}_j)$. This density has a two-dimensional mean vector $\boldsymbol{\mu}_j$ and a two-dimensional covariance matrix $\mathbf{C}_\beta$ for $j = 1, \ldots, J$. It is assumed that the constants $\mu_1, \ldots, \mu_J$ are known, and $\mathbf{C}_\beta$ is a positive number. Let $\boldsymbol{\mu}_\beta = [\mu_1, \ldots, \mu_J]$. The joint distribution of $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J]$ is specified as the *parameter prior*:

$$p(\boldsymbol{\beta}) = \prod_{j=1}^{J} p(\boldsymbol{\beta}_j). \tag{1.1}$$

The *likelihood function* of the response vector of examinee i who is assumed to have attribute pattern $\boldsymbol{\alpha}$ is given by:

$$p(\mathbf{x}_i, \boldsymbol{\alpha}|\boldsymbol{\beta}) = p(\boldsymbol{\alpha})p(\mathbf{x}_i|\boldsymbol{\alpha}, \boldsymbol{\beta}) = p(\boldsymbol{\alpha}) \prod_{j=1}^{J} p(x_{ij}|\boldsymbol{\alpha}, \boldsymbol{\beta}_j, \mathbf{q}_j). \tag{1.2}$$

yielding the MAP empirical risk function:

$$\hat{\ell}_n(\boldsymbol{\beta}) = -(1/n)\log p(\boldsymbol{\beta}) + (1/n)\sum_{i=1}^{n} c(\mathbf{x}_i, \boldsymbol{\beta}). \tag{1.3}$$

where

$$c(\mathbf{x}_i; \boldsymbol{\beta}) = -\log p(\mathbf{x}_i|\boldsymbol{\beta}) = -\log \sum_{\alpha} p(\mathbf{x}_i|\boldsymbol{\alpha}, \boldsymbol{\beta}) p(\boldsymbol{\alpha}).$$

In this study, the MAP empirical risk function (3) involves summing over all possible latent skill attribute patterns. Assume $\hat{\ell}_n$ is a twice-continuously differentiable objective function. Once a critical point is reached, then the Hessian of $\hat{\ell}_n$ can be evaluated at that point to check if the critical point is a strict local minimizer (e.g., Golden, 2020). Here we assume that a parameter estimate $\hat{\boldsymbol{\beta}}_n$ has been obtained which is a strict local minimizer of $\hat{\ell}_n(\boldsymbol{\beta})$ in some (possibly very small) closed, bounded, and convex region of the parameter space Ω for all sufficiently large n. Furthermore, we assume that the expected value of $\hat{\ell}_n(\boldsymbol{\beta})$, $\ell(\boldsymbol{\beta})$, has a strict global minimizer, $\boldsymbol{\beta}^*$, in the interior of Ω. This setup of the problem thus allows for situations where ℓ has multiple minimizers, maximizers, and saddlepoints over the entire unrestricted parameter space. Given these assumptions, it can be shown (e.g., Golden, 2020; White, 1982) that $\hat{\boldsymbol{\beta}}_n$ in this case is a consistent estimator of $\boldsymbol{\beta}^*$. In addition, for large sample sizes, the effects of the parameter prior becomes negligible and the MAP estimate has the same asymptotic distribution as the maximum likelihood estimate.

1.2.4 Information Matrix Test Methods for Detection of Model Misspecification

1.2.4.1 Determinant Generalized Information Matrix Test Statistical Theory

In this section, we present explicit details regarding the derivation of the asymptotic distribution of the $GIMT_{Det}$. Let $\mathbf{g}(\tilde{\mathbf{x}}; \boldsymbol{\beta}) \equiv -\nabla \log p(\tilde{\mathbf{x}}; \boldsymbol{\beta})$. Let $\mathbf{A}^*$ denote the Hessian of ℓ, evaluated at β^*, $\mathbf{A}^* = \mathbf{A}(\boldsymbol{\beta}^*) = -\nabla^2 E\{\log p((\tilde{\mathbf{x}}_i; \boldsymbol{\beta}))\}$. Let $\mathbf{B}^*$ denote $\mathbf{B}^* = \mathbf{B}(\boldsymbol{\beta}^*) = E\{\mathbf{g}(\tilde{\mathbf{x}}; \boldsymbol{\beta})\mathbf{g}(\tilde{\mathbf{x}}; \boldsymbol{\beta})^T\}$ evaluated at the point $\boldsymbol{\beta}^*$. It is well known (e.g., Golden et al., 2013, 2016; White, 1982) that if $\mathbf{A}^*$ and $\mathbf{B}^*$ are positive definite, then the asymptotic distribution of $\hat{\boldsymbol{\beta}}_n$ is a multivariate Gaussian with mean $\boldsymbol{\beta}^*$ and covariance matrix $(1/n)\mathbf{C}^* \equiv (1/n)(\mathbf{A}^*)^{-1}\mathbf{B}^*(\mathbf{A}^*)^{-1}$ as $n \to \infty$. In the special case where the model M is correctly specified in the sense that the observed data is i.i.d. with a common probability mass function such that: $p(\mathbf{x}|\boldsymbol{\beta}^*) = p_{DGP}(\mathbf{x})$, then the covariance matrix $\mathbf{C}^*$ can be computed using either $(\mathbf{A}^*)^{-1}$ or $(\mathbf{B}^*)^{-1}$. It then follows that if M is correctly specified with respect to p_{DGP}, then $\mathbf{A}^* = \mathbf{B}^*$. Consequently, if $\mathbf{A}^* \neq \mathbf{B}^*$, then M is misspecified with respect to p_{DGP}.

Let $s : \mathcal{R}^{q \times q} \times \mathcal{R}^{q \times q} \to \mathcal{R}$ be a continuously differentiable function and assume ∇s evaluated at $\mathbf{A}^*$ and $\mathbf{B}^*$ has full row rank. The function s is called a GIMT Hypothesis Function that has the property that if $\mathbf{A} = \mathbf{B}$, then $s(\mathbf{A}, \mathbf{B}) = \mathbf{0}$ for every symmetric positive definite matrix $\mathbf{A}$ and for every symmetric positive definite matrix $\mathbf{B}$. Let $s^* \equiv s(\mathbf{A}^*, \mathbf{B}^*)$ following Golden et al. (2013, 2016).

Let $\hat{\mathbf{A}}_n$ denote the Hessian of $\hat{\ell}_n$ evaluated at $\hat{\boldsymbol{\beta}}_n$. Let

$$\hat{\mathbf{B}}_n = (1/n) \sum_{i=1}^{n} \mathbf{g}(\tilde{\mathbf{x}}_i, \hat{\boldsymbol{\beta}}_n) \mathbf{g}(\tilde{\mathbf{x}}_i, \hat{\boldsymbol{\beta}}_n)^T.$$

A GIMT is defined as a test statistic $\hat{s}_n \equiv s(\hat{\mathbf{A}}_n, \hat{\mathbf{B}}_n)$ that evaluates the null hypothesis:

$$H_0 : s^* = 0.$$

The $GIMT_{Det}$ (Golden et al., 2013; also see Golden, 2020; Golden et al., 2016) is specified by the GIMT hypothesis function s is defined such that: $s(\mathbf{A}^*, \mathbf{B}^*) = \frac{1}{q} \log \det((\mathbf{A}^*)^{-1} \mathbf{B}^*)$. Let the Wald test statistic:

$$\mathcal{W}_n = n(\hat{s}_n)^T \hat{\mathbf{C}}_{n,s}^{-1} (\hat{s}_n).$$

where $\hat{s}_n = \frac{1}{q} \log \det((\hat{\mathbf{A}})^{-1} \hat{\mathbf{B}})$, If the null hypothesis $H_0 : s^* = 0$ holds, then $\hat{\mathcal{W}}_n$ converges in distribution to a chi-squared random variable with one degree of freedom as $n \to \infty$ (see Theorem 7 for Golden et al., 2016). If the null hypothesis fails, $\hat{\mathcal{W}}_n \to \infty$ as $n \to \infty$ with a probability of one (see Theorem 7 for Golden et al., 2016). Here the estimator $\hat{\mathbf{C}}_{n,s}^{-1}$ is an estimate of the asymptotic covariance matrix of $n^{1/2}(\hat{s}_n - s^*)$, $(\mathbf{C}_s^*)^{-1}$. Golden et al., 2013, 2016 shows how this asymptotic covariance matrix can be estimated using the first, second, and third derivatives of the log-likelihood function. Notice that since the degrees of freedom of the Wald statistic for the $GIMT_{Det}$ is equal to 1, the test statistic has a distribution that is the square of a normalized Gaussian random variable.

1.3 Simulation Study

1.3.1 Dataset

The simulation studies used CDMs with parameters estimated from the Tatsuoka (1984) fraction-subtraction data set. The dataset consists of a set of math problems involving fractions subtraction, designed to assess students' ability to solve problems involving fractions. In this case, we utilized the 'Fraction.1' dataset from the 'CDM' R package (George et al., 2016) containing the dichotomous responses of 536 middle school students over 15 fraction subtraction test items. The **Q**-matrix specifying which of five distinct skills were required to answer a particular test item was based upon the **Q**-matrix used by Tatsuoka (1984) and provided in George et al. (2016).

1.3.2 Methods

First, the described DINA CDM was fit to the data set with a sample size of $n = 536$ using the previously described MAP estimation method (see Equation 1.3). A multivariate Gaussian prior with mean vector $\boldsymbol{\mu}_{\beta}$ and covariance matrix $\mathbf{C}_{\beta}$ was used for all items (see Equation 1.1). To ensure guess (g) and slip (s) probabilities both equaled 0.354 for a specific item, the Gaussian Parameter Prior mean vector for that item was set as $\boldsymbol{\mu}_{\beta} = \begin{bmatrix} 1.2 & 0.6 \end{bmatrix}^{T}$. An uninformative Gaussian Parameter Prior covariance matrix $\mathbf{C}_{\beta}$ was chosen to be $\sigma^2 \mathbf{I}_2$, where $\sigma^2 = 4500$. The fitted DINA CDM will be considered as the Data Generating Process CDM (DGP_{CDM}). To create a scenario in which the CDM is correctly specified, we refitted the initial original DINA CDM by sampling from the original fitted DINA CDM. Five different versions of the original DINA CDM, corresponding to "misspecified models," were used. These versions were created by flipping Tatsuoka's $\mathbf{Q}$-matrix elements with different probabilities (0%, 1%, 5%, 10%, 15%, and 20%). The 0% misspecification level corresponds to the correctly specified case. For each level of $\mathbf{Q}$-matrix misspecification, we generate 5 different misspecified Q matrices at that level, allowing for variations in which elements are altered. We refer to a specific way of misspecifying the $\mathbf{Q}$-matrix at a given level as a "replication." Then, we fit each of the 5 replications of the $\mathbf{Q}$-matrix at a particular misspecification level to 50 bootstrap-generated datasets from the original DINA CDM.

We computed $GIMT_{Det}$ statistics as a function of misspecification level. Then we evaluated the performance of the $GIMT_{Det}$ in classifying correctly specified and misspecified models using the Receiver Operating Characteristic (ROC) curve. The ROC curve plots the true positive rate against the false positive rate at different decision thresholds, allowing for an investigation of the discrimination performance of the $GIMT_{Det}$. Additionally, the area under the ROC curve (AUROC) can provide a quantitative measure of the $GIMT_{Det}$ discrimination, with a high AUROC value indicating that the $GIMT_{Det}$ effectively distinguishes between correctly specified and misspecified models. Conversely, an AUROC close to 0.5 indicates no discrimination performance.

1.3.3 Results and Discussion

Table 1.1 displays the AUROC values under various misspecification levels which shows that AUROC values increase as the misspecification level increases. The results of the simulation study are also presented in Figure 1.1, which displays how the ROC curves evolve at various misspecification levels (0%, 1%, 5%, 10%, 15%, and 20%) with a constant sample size of n=536. At the 20% misspecification level, the ROC curve demonstrates good discrimination performance, closely approaching the upper-left corner.
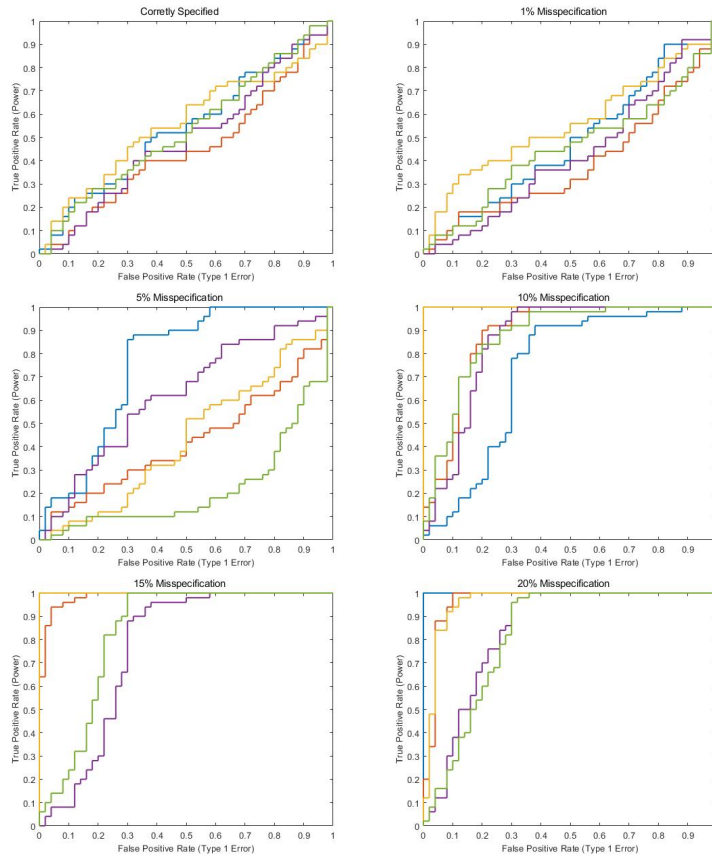
Fig. 1.1 Influence of Misspecification Levels on ROC Curves: Level of misspecification: 0%, 1%, 5%, 10%, 15%, and 20%; Sample Size n = 536; Conducted with five replications for each level. The discrimination performance of $GIMT_{Det}$ was effective in the 10%, 15%, and 20% cases, while there was no indication of discrimination performance in the 0% and 1% misspecification cases.

Misspecification level	AUROC Mean	Lower CI	Upper CI
Correctly Specified	0.51	0.46	0.57
1% Misspecification	0.46	0.38	0.54
5% Misspecification	0.54	0.23	0.85
10% Misspecification	0.87	0.75	0.99
15% Misspecification	0.91	0.78	1.00
20% Misspecification	0.92	0.82	1.00

Table 1.1 Average AUROC and Confidence Intervals at Different Levels of Model Misspecification, Conducted with five versions for Each Level. The average AUROC increases with higher levels of misspecification. Notably, the 5% misspecification case exhibits the most variation in the "version of misspecification," while reasonably good discrimination performance is consistently achieved for the 10%, 15%, and 20% cases.

When the misspecification level was greater than 5%, the GIMT showed good discrimination performance regardless of the manner in which the Q matrix was decimated. In addition, the GIMT did not detect model misspecification for the correctly specified case (0%) and did not detect model misspecification for the slightly misspecified case (1%). For moderate misspecification levels (5%), however, only two of the five replications showed effective discrimination performance. There was also substantial variability across replications for the 5% case. Currently, we view the 5% level as a "transition region" and we plan to investigate this further using additional replications. In summary, the new GIMT method for misspecification detection in CDMs appears to be promising. Further research will continue to explore the capabilities and limitations of this approach.

References

1. Chen, F., Liu, Y., Xin, T. & Cui, Y. Applying the M2 Statistic to Evaluate the Fit of Diagnostic Classification Models in the Presence of Attribute Hierarchies. *Frontiers In Psychology*. **9** (2018,10), https://www.frontiersin.org/article/10.3389/fpsyg.2018.01875/full
2. de la Torre, J. DINA Model and Parameter Estimation: A Didactic. *Journal Of Educational And Behavioral Statistics*. **34**, 115-130 (2009,3), http://journals.sagepub.com/doi/10.3102/1076998607309474
3. de la Torre, J. The Generalized DINA Model Framework. *Psychometrika*. **76** (2011)
4. George, A., Robitzsch, A., Kiefer, T., Groß, J. & Ünlü, A. The R Package CDM for Cognitive Diagnosis Models. *Journal Of Statistical Software*. **74**, 24 (2016), https://www.jstatsoft.org/v074/i02
5. Golden, R. Statistical Machine Learning: A unified framework. (CRC Press,2020)
6. Golden, R., Henley, S., White, H. & Kashner, T. New Directions in Information Matrix Testing: Eigenspectrum Tests. *Recent Advances And Future Directions In Causality, Prediction, And Specification Analysis: Essays In Honor Of Halbert L. White Jr*. pp. 145-177 (2013)
7. Golden, R., Henley, S., White, H. & Kashner, T. Generalized Information Matrix Tests for Detecting Model Misspecification. *Econometrics*. **4**, 46 (2016)
8. Kunina-Habenicht, O., Rupp, A., Wilhelm, O. The Impact of Model Misspecification on Parameter Estimation and Item-Fit Assessment in Log-Linear Diagnostic Classification Models. *Journal Of Educational Measurement*. **49**, 59-81 (2012,3), https://onlinelibrary.wiley.com/doi/10.1111/j.1745-3984.2011.00160.x
9. Liu, Y., Xin, T., Andersson, B. & Tian, W. Information matrix estimation procedures for cognitive diagnostic models. *Br J Math Stat Psychol*. **72**, 18-37
10. Maris, E. Estimating multiple classification latent class models. *Psychometrika*. **64** pp. 187-212 (1999)
11. Maydeu-Olivares, A. & Joe, H. Assessing Approximate Fit in Categorical Data Analysis. *Multivariate Behavioral Research*. **49**, 305-328 (2014)
12. Presnell, B. & Boos, D. The IOS Test for Model Misspecification. *Journal Of The American Statistical Association*. **99**, 216-227 (2004)
13. Rupp, A. & Templin, J. The Effects of Q-Matrix Misspecification on Parameter Estimates and Classification Accuracy in the DINA Model. *Educational And Psychological Measurement*. **68**, 78-96 (2008,2), http://journals.sagepub.com/doi/10.1177/0013164407301545
14. Tatsuoka, K. Analysis of Errors in Fraction Addition and Subtraction Problems. Final Report for NIE-G-81-0002. (1984)
15. White, H. Maximum Likelihood Estimation of Misspecified Models. *Econometrica*. **50**, 1-25 (1982), http://www.jstor.org/stable/1912526