

# The asymptotic power of the Lagrange Multiplier tests for misspecified IRT models

Lucia Guastadisegni<sup>1</sup>, Silvia Cagnone<sup>2</sup>, Irini Moustaki<sup>3</sup>, and Vassilis Vasdekis<sup>4</sup>

<sup>1</sup> University of Bologna,

lucia.guastadisegni2@unibo.it

<sup>2</sup> University of Bologna,

silvia.cagnone@unibo.it

<sup>3</sup> London School of Economics and Political Science,

i.moustaki@lse.ac.uk

<sup>4</sup> Athens University of Economics and Business,

vasdekis@aueb.gr

**Abstract.** This article studies the power of the Lagrange Multiplier Test and the Generalized Lagrange Multiplier Test to detect measurement non-invariance in Item Response Theory (IRT) models for binary data. We study the performance of these two tests under correct model specification and incorrect distribution of the latent variable. The asymptotic distribution of each test under the alternative hypothesis depends on a noncentrality parameter that is used to compute the power. We present two different procedures to compute the noncentrality parameter and consequently the power of the tests. The performance of the two methods is evaluated through a simulation study. They turn out to be very similar to the classic empirical power but less time consuming. Moreover, the results highlight that the Lagrange Multiplier Test is more powerful than the Generalized Lagrange Multiplier Test to detect measurement non-invariance under all simulation conditions.

**Keywords:** Generalized Lagrange Multiplier test, MIMIC models, binary data

## 1 Introduction

The power of a test is usually estimated through Monte Carlo simulation methods. However, it can alternatively be computed asymptotically using the distribution of the test statistic under the alternative hypothesis that depends on a noncentrality parameter, often unknown or difficult to compute (Gudicha et al. 2017).

In this work we study the asymptotic power of two test statistics, the Lagrange Multiplier (LM) test and the Generalized Lagrange Multiplier (LM(S)) test, to detect measurement non-invariance under correct model specification and model misspecification.

An item is measurement non-invariant, or biased, if it measures different abilities for different group membership identified by an external variable (Mellenbergh 1982,1983). Group differences can be present only on the item intercept or simultaneously on the item intercept and slope.

The Lagrange Multiplier test is used in the IRT context to detect measurement non-invariance (Glas 1998, Fox and Glas 2005) and other types of model violations such as local dependence, incorrect specification of the item characteristic curve and a non normal distribution of the latent variables (Glas 1999, Glas and Falc3n 2003, Liu and Thissen 2012). Despite its extensive use in IRT models, only in a few studies the LM test has been applied in the case of model misspecification under the null and the alternative hypothesis (Glas and Falc3n 2003).

In order to take into account possible misspecification in the model, the LM test can be generalized obtaining the so-called Generalized Lagrange Multiplier test (LM(S)), whose expression involves the sandwich variance and covariance matrix (White 1982). In the IRT context the performance of the LM(S) test under model misspecification has been recently analyzed by Falk and Monroe (2018) through a elaborate simulation study.

The first objective of this paper is to present the theoretical computation of the asymptotic power of these tests using two different approximation methods to obtain the noncentrality parameter. The second objective is to compare the performance of the LM and LM(S) tests through a simulation study to detect measurement invariance under correct model specification and misspecification of the latent variable distribution in terms of asymptotic and empirical power. The model considered under the null and the alternative hypothesis is a classic Multiple Indicator Multiple Causes (MIMIC) model for binary data, based on the assumption of a normal distribution of the latent factor. The misspecification is introduced by assuming a non normal distribution of the latent factor in the data generating model.

The paper is organized as follows; in Sect. 2 we review the theory of the LM test and the procedures to estimate its asymptotic power, in Sect. 3 we describe the LM(S) test and the procedures to estimate its asymptotic power and in Sect. 4 we present a Monte Carlo simulation study. We conclude with some remarks in Sect. 5.

## 2 The Lagrange Multiplier Test

Consider a sample  $\mathbf{y}_1, \dots, \mathbf{y}_n$  from a model  $f(\mathbf{y}, \boldsymbol{\theta})$ . Let  $\boldsymbol{\theta}_0$  denote the true parameter vector, that can be divided in two subvectors  $\boldsymbol{\theta}'_0 = (\boldsymbol{\theta}'_{01}, \boldsymbol{\theta}'_{02})$ . The hypotheses  $H_0$  and  $H_1$  can be formalized as follows:

$$H_0 : \boldsymbol{\theta}'_{02} = \mathbf{c} \quad vs \quad H_1 : \boldsymbol{\theta}'_{02} \neq \mathbf{c}, \quad (1)$$

where  $\mathbf{c}$  is a vector of constants. The LM statistic is (Engle 1984):

$$LM = \frac{1}{n} S_2(\tilde{\boldsymbol{\theta}}) A^{22}(\tilde{\boldsymbol{\theta}})^{-1} S_2(\tilde{\boldsymbol{\theta}}), \quad (2)$$

where  $\tilde{\boldsymbol{\theta}}' = (\tilde{\boldsymbol{\theta}}'_1, \mathbf{c})$  denotes the restricted maximum likelihood estimates of the parameters  $\boldsymbol{\theta}$ ,  $S_2$  is the subset of the vector of score functions  $S = \frac{\partial \ln l(\mathbf{y}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$  corresponding to the parameters  $\boldsymbol{\theta}_{02}$  evaluated at  $\tilde{\boldsymbol{\theta}}$ . The matrix  $A^{22}$  is the block of the partitioned Fisher information matrix  $A = -E \left[ \frac{1}{n} \frac{\partial^2 l(\mathbf{y}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]$  defined as:

$$A^{22} = A_{22} - A_{21} A_{11}^{-1} A_{12}, \quad (3)$$

evaluated at  $\tilde{\boldsymbol{\theta}}$ . The partition of  $A$  into  $A_{22}, A_{21}, A_{11}, A_{12}$  is derived from the partition of  $\boldsymbol{\theta}'_0$  into  $(\boldsymbol{\theta}'_{01}, \boldsymbol{\theta}'_{02})$ . In this study, we consider the LM test computed with the observed Hessian approach, where the Fisher information matrix in formula (2) is replaced by the corresponding observed Hessian matrix

$$\hat{A}(\boldsymbol{\theta}) = -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 l_i(\mathbf{y}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \quad (4)$$

Under a correct specified likelihood and under  $H_0$ , the LM statistic is asymptotically distributed as a  $\chi_r^2$ , where  $r$  are the degrees of freedom ( $df$ ) equal to the dimension of  $\boldsymbol{\theta}_{02}$  (Silvey 1959). When the alternative hypothesis is true but the null is tested, the LM test statistic has an asymptotic noncentral chi-square distribution that depends on two parameters, the  $df$  and a noncentrality parameter (Bollen 1989). To compute the local asymptotic power of the LM test, a standard approach is to consider a set of local alternatives that are close to the null value for large  $n$ ,  $H_1 : \boldsymbol{\theta}_{02} = \mathbf{c} + \frac{\boldsymbol{\xi}}{\sqrt{n}}$ , where  $\boldsymbol{\xi}$  is an arbitrary vector with the same dimension of  $\boldsymbol{\theta}_{02}$  (Boos and Stefanski 2013). Under  $H_1$ , the test statistic LM converges in distribution to a  $\chi_r^2(\lambda)$  with noncentrality parameter  $\lambda$  equal to (Cox and Hinkley 1979):

$$\lambda = \boldsymbol{\xi}' A^{22}(\boldsymbol{\theta}^0) \boldsymbol{\xi}, \quad (5)$$

where  $\boldsymbol{\theta}^0 = (\boldsymbol{\theta}_{01}, \mathbf{c})$ .

The asymptotic local power is computed as  $P(\chi_r^2(\lambda) > \chi_r^2(\lambda, 1 - \alpha))$ .

## 2.1 Approximation procedures for the asymptotic power

The asymptotic distribution of the LM test under the alternative hypothesis as a non-central chi-square with noncentrality parameter (5) holds when the model defined under the set of local alternatives is true, i.e. when the model under the null hypothesis is barely incorrect for large  $n$  (see Agresti 2002, Reiser 2008). In practice, it is often reasonable to adopt an alternative hypothesis for fixed and finite  $n$  (Agresti 2002), as  $H_1 : \boldsymbol{\theta}_{02} = \mathbf{c} + \boldsymbol{\xi}$ , or to use hypotheses as (1) (Gudicha et al. 2017). We present here two different approximation procedures for the computation of the noncentrality parameter.

The first method extends the approximation procedure for the asymptotic power derived by Gudicha et al. (2017) for the Likelihood-Ratio and the Wald tests to the LM test. It can be summarized in the following steps:

1. From the model defined under the alternative hypothesis, create a large data set (e.g.  $N = 10000$  observations).
2. Fit the model under  $H_0$  to the data.
3. Take the value of the LM statistic as the estimate of the noncentrality parameter  $\lambda$  (Satorra 1989, Bollen 1989).
4. Compute the noncentrality parameter for a sample of size 1 equal to  $\lambda_1 = \frac{\lambda}{N}$ .
5. The noncentrality parameter for a sample of size  $n$  is  $\lambda_n = n\lambda_1$ .

The power of the LM test can be determined by comparing the  $\lambda_n$  obtained in step 5 with the tabled values of the noncentral chi-square with  $df$  corresponding to the number of parameters constrained under  $H_0$  and significance level  $\alpha$  (Bollen 1989).

We propose a second method, that is also based on some of the steps of the procedure proposed by Gudicha et al. 2017, but the noncentrality parameter is computed according to formula (5). The procedure can be summarized as follows:

1. From the model defined under the alternative hypothesis, create a large data set (e.g.  $N = 10000$  observations).
2. Fit the model under  $H_0$  to the data.
3. Compute  $\boldsymbol{\xi} = \sqrt{N}(\boldsymbol{\theta}_{02} - \mathbf{c})$ , where  $\boldsymbol{\theta}_{02}$  is the vector of the data generating values (values under  $H_1$ ) of the constrained parameters and  $\mathbf{c}$  is the vector of constants under the null hypothesis (Reiser 2008).
4. Compute the noncentrality parameter according to formula (5) where  $A^{22}(\boldsymbol{\theta}^0)$  can be consistently estimated by the corresponding matrix  $\hat{A}$ , evaluated at  $\tilde{\boldsymbol{\theta}}$ .
5. Compute the noncentrality parameter for a sample of size 1 as  $\lambda_1 = \frac{\lambda}{N}$ .
6. The non centrality parameter for a sample of size  $n$  is  $\lambda_n = n\lambda_1$

The power is computed as before, using the noncentrality parameter computed at point 5.

### 3 The Generalized Lagrange Multiplier Test

Consider a sample  $\mathbf{y}_1, \dots, \mathbf{y}_n$  from a model with true density  $g(\mathbf{y})$ . The model  $f(\mathbf{y}; \boldsymbol{\theta})$  is assumed to be true one for the data and differs from  $g(\mathbf{y})$ . Under the assumptions given in White 1982 the vector of parameter  $\hat{\boldsymbol{\theta}}_n$ , that maximizes the log-likelihood function based on model  $f(\mathbf{y}; \boldsymbol{\theta})$  (Quasi-ML estimator, White 1982), converges in probability to  $\boldsymbol{\theta}_*$ , the parameter vector that minimizes the Kullback-Leibler information criterion. Moreover the variance and covariance matrix of the Quasi-LM estimator is the sandwich variance and covariance matrix  $\hat{C}(\hat{\boldsymbol{\theta}}_n) = \hat{A}^{-1}(\hat{\boldsymbol{\theta}}_n) \hat{B}(\hat{\boldsymbol{\theta}}_n) \hat{A}^{-1}(\hat{\boldsymbol{\theta}}_n)$ , where the matrix  $\hat{A}$  is defined in formula (4) and  $\hat{B} = \frac{1}{n} \sum_{i=1}^n \frac{\partial l_i(\mathbf{y}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial l_i(\mathbf{y}_i; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$  is the observed cross-product matrix (White 1982).

Under model misspecification, the null and the alternative hypotheses are posed in terms of  $\boldsymbol{\theta}_*$ . Let  $\boldsymbol{\theta}_*$  be divided in two subvectors  $\boldsymbol{\theta}'_* = (\boldsymbol{\theta}'_{*1}, \boldsymbol{\theta}'_{*2})$ . The hypotheses  $H_0$  and  $H_1$  can be formalized as follows:

$$H_0 : \boldsymbol{\theta}'_{*2} = \mathbf{c} \quad vs \quad H_1 : \boldsymbol{\theta}'_{*2} \neq \mathbf{c}, \quad (6)$$

where  $\mathbf{c}$  is a vector of constants.

The Generalized Lagrange Multiplier Test is defined as:

$$LM(S) = \frac{1}{n} S_2(\tilde{\boldsymbol{\theta}}_n)' \hat{A}^{22}(\tilde{\boldsymbol{\theta}}_n)^{-1} \hat{C}_{22}(\tilde{\boldsymbol{\theta}}_n)^{-1} \hat{A}^{22}(\tilde{\boldsymbol{\theta}}_n)^{-1} S_2(\tilde{\boldsymbol{\theta}}_n), \quad (7)$$

where  $\tilde{\boldsymbol{\theta}}_n$  is the constrained quasi-ML estimator,  $\hat{A}^{22}$  is the block of the partitioned observed Hessian matrix computed as in formula (3), evaluated at  $\tilde{\boldsymbol{\theta}}_n$  and  $\hat{C}_{22}$  is the block of the matrix  $\hat{C}$  corresponding to  $\boldsymbol{\theta}'_{*2}$ , evaluated at  $\tilde{\boldsymbol{\theta}}_n$ . Under  $H_0$  the statistic

LM(S) is distributed as a  $\chi_r^2$ , where  $r$  are the  $df$  equal to the dimension of  $\boldsymbol{\theta}_{*2}$  (White 1982). To compute the local asymptotic power of the LM(S) test, a standard approach is to consider a set of local alternatives  $H_1 : \boldsymbol{\theta}_{*2} = \boldsymbol{c} + \frac{\boldsymbol{\xi}}{\sqrt{n}}$ , where  $\boldsymbol{\xi}$  is an arbitrary vector of dimension  $\boldsymbol{\theta}_{*2}$ . Under  $H_1$ , the test statistic  $LM(S)$  converges in distribution to a  $\chi_r^2(\lambda)$ , where  $r$  are the  $df$  equal to the dimension of  $\boldsymbol{\theta}_{*2}$  and  $\lambda$  is the noncentrality parameter given by (Bera et al. 2020):

$$\lambda = \boldsymbol{\xi}' A^{22'} (B_{22} - A_{21} A_{11}^{-1} B_{12} - B_{21} A_{11}^{-1} A_{12} + A_{21} A_{11}^{-1} B_{11} A_{11}^{-1} A_{12})^{-1} A^{22} \boldsymbol{\xi} \quad (8)$$

where  $A$  is the Fisher information matrix and  $B$  is the expected cross-product matrix, evaluated at  $\boldsymbol{\theta}^*$ .

If the model is correctly specified, the LM(S) coincides with LM test (White 1982).

### 3.1 Estimation procedure for the noncentrality parameter

The estimation method described in Sect. 2.1 to compute the asymptotic power is used here to estimate the asymptotic power for the LM(S) test, with some differences.

In step 3 of the first method, the LM(S) statistic is taken as the estimate of the noncentrality parameter (the proof of this result can be found in Satorra 1989).

In step 4 of the second method, the noncentrality parameter is computed according to formula (8), where the matrices  $A(\boldsymbol{\theta}^*)$  and  $B(\boldsymbol{\theta}^*)$  are consistently estimated by  $\hat{A}$  and  $\hat{B}$ , evaluated at  $\hat{\boldsymbol{\theta}}_n$ .

Moreover, the model fitted under  $H_0$  at step 2 is assumed to be misspecified. Under correct model specification the LM(S) and the LM tests have the same noncentrality parameter and, consequently, the same asymptotic power.

## 4 Simulation Study

### 4.1 Simulation design

The aim of this section is to compare the different procedures described above to estimate the asymptotic and the empirical power of the LM and LM(S) tests to detect measurement non-invariance by means of a simulation study. A MIMIC model for binary data is considered. Both under correct and model misspecification, we consider a binary group variable  $x$  because we study measurement non-invariance only in two subgroups of population. Given  $n$  individuals and  $p$  items, under correct model specification data are generated from the following model, where measurement non-invariance is introduced on the intercept of the last item  $p$  through the parameter  $\gamma_1$  and the group variable  $x$ :

$$\begin{aligned} \text{logit}(\pi_{ij}) &= \alpha_{0j} + \alpha_{1j} z_i & i = 1, \dots, n & \quad j = 1, \dots, p-1 \\ \text{logit}(\pi_{ip}) &= \alpha_{0p} + \alpha_{1p} z_i + \gamma_1 x_i & & \\ z &\sim N(0, 1) & & \end{aligned} \quad (9)$$

Under misspecification of the latent variable distribution data are generated from the following model, where measurement non-invariance is introduced as before on the intercept of the last item  $p$  through the parameter  $\gamma_1$  and the group variable  $x$ :

$$\begin{aligned} \text{logit}(\pi_{ij}) &= \alpha_{0j} + \alpha_{1j}z_i & i = 1, \dots, n & \quad j = 1, \dots, p-1 \\ \text{logit}(\pi_{ip}) &= \alpha_{0p} + \alpha_{1p}z_i + \gamma_1x_i & & \\ z &\sim SN(\kappa) \end{aligned} \quad (10)$$

In this case, the latent variable  $z$  is generated from a skew-normal (SN) with skewness parameter  $\kappa$ , with the following probability density function (Azzalini 1985):

$$\phi(\epsilon; \kappa) = 2\phi(\epsilon)\Phi(\epsilon; \kappa)$$

where  $\phi$  and  $\Phi$  are the standard normal density and distribution function, respectively. The parameter  $\kappa$  can take values from  $-\infty$  to  $+\infty$ : when it is equal to 0, the skew-normal reduces to a standard normal distribution. In the simulations, we consider two values of  $\kappa$ , 3 and 5. When  $\kappa = 3$  the mean and the variance of the latent variable are 0.76 and 0.43, respectively, and when  $\kappa = 5$ , the mean and the variance of the latent variable are 0.78 and 0.39, respectively. In both models (9) and (10) we consider two possible effect sizes, equal to 0.2 and 0.5, for the parameter  $\gamma_1$ . Moreover, in both cases, the values  $x$ s are generated from a Bernoulli distribution with success probability 0.7, the intercepts from a normal distribution with 0 mean and Standard Deviation (SD) 0.1 and the slopes from a normal distribution with 0 mean and SD 0.5.

The following set of hypotheses is being tested:

$$H_0 : \gamma_{1p} = 0 \quad vs \quad H_1 : \gamma_{1p} \neq 0,$$

that implies that the last item is tested for measurement invariance.

Model (9) is fitted to the data with  $\gamma_{1p}$  fixed to 0. When data are generated from model (10) we are working under model misspecification. Indeed, as mentioned before, the true latent variable has mean and variance around 0.7 and 0.4, respectively, and its skewed. Since model (9) is fitted to the data, the misfit is in the mean, assumed to be 0, in the variance, assumed to be 1 and in the distribution of the latent variable, assumed to be symmetric. The following simulation conditions are considered: number of items ( $p = 10$ )  $\times$  sample size ( $n = 200, 500, 1000, 5000, 10000$ )  $\times$  Test statistic ( $LM, LM(S)$ ). Due to the time complexity, the empirical power is computed only for  $n = 200, 500, 1000$ . 200 replications are considered for each condition of the study. The empirical power  $\hat{p}$  is computed as  $\hat{p} = \sum_{l=1}^{N_v} \frac{I(T_l \geq c)}{N_v}$ , where  $N_v$  is the number of valid statistics out of the number of replications,  $I$  is the indicator function,  $T_l$  is the value of the test statistic evaluated in the  $l$ -th replication and  $c$  is the theoretical asymptotic critical value corresponding to the 95-th percentile of the  $\chi_{df}^2$  distribution, with  $df$  equal to the number of constrained parameter under  $H_0$ . If non valid statistics occur, they are excluded from the analysis. The asymptotic power is computed through methods 1 and 2 described in Sect. 2.1 and 3.1. The nominal level  $\alpha$  is equal to 0.05 in all simulations. ML estimates of the parameters are obtained with direct maximization of the likelihood function using 21 Gauss-Hermite quadrature

points. Numerical derivatives are used to compute the Hessian and cross-product matrices.

## 4.2 Results

Table 1 presents the results for the LM and LM(S) tests under correct model specification when  $\gamma_1$  is equal to 0.2 and 0.5 in the data generating model,  $p = 10$ ,  $n = 200, 500, 1000, 5000, 10000$ . We can notice that, in general, the differences between the asymptotic and empirical power are small and method 1 is slightly closer to the empirical power than method 2. For what concerns the power to detect measurement non-invariance, the LM test has a slightly higher power compared to the LM(S) tests under all conditions, with the exception of the case  $\gamma_1 = 0.5$  and for large sample sizes ( $n = 5000, 10000$ ), where the two tests reach the same power, as expected from the theory.

**Table 1.** Asymptotic and empirical power of the LM and LM(S) tests under correct model specification,  $\gamma_1 = 0.2, 0.5$ ,  $p = 10$ ,  $n = 200, 500, 1000, 5000, 10000$ .

$p$	$\gamma_1$	$n$	Method 1		Method 2		Empirical	
			LM	LM(S)	LM	LM(S)	LM	LM(S)
10	0.2	200	0.086	0.085	0.080	0.079	0.08	0.06
		500	0.144	0.140	0.126	0.122	0.185	0.17
		1000	0.241	0.234	0.204	0.198	0.26	0.25
		5000	0.802	0.785	0.714	0.696	-	-
		10000	0.978	0.973	0.947	0.938	-	-
10	0.5	200	0.240	0.222	0.229	0.211	0.285	0.235
		500	0.508	0.468	0.484	0.445	0.54	0.5
		1000	0.799	0.758	0.775	0.732	0.8	0.78
		5000	1	1	1	1	-	-
		10000	1	1	1	1	-	-

Table 2 shows the results for the LM and LM(S) tests computed under misspecification of the latent variable distribution when  $\gamma_1$  is equal to 0.2 and 0.5 in the data generating model,  $p = 10$ ,  $n = 200, 500, 1000, 5000, 10000$ . Also in this case the differences between the asymptotic and empirical power are small. For what concerns the power to detect measurement non-invariance under model misspecification, despite the fact that the LM(S) test is derived under model misspecification, the LM test has the highest power under all conditions. The two tests reach the same power only when  $\gamma_1 = 0.5$  and  $n = 10000$ . In both Tables and for both tests, the power increases with the sample size and the effect size of the parameter  $\gamma_1$  and decreases when the model is misspecified.

**Table 2.** Asymptotic power of the LM and LM(S) tests under incorrect distribution of the latent variable,  $\gamma_1 = 0.2, 0.5$ ,  $p = 10$ ,  $n = 200, 500, 1000, 5000, 10000$ .

$p$	ES	$\alpha$	$n$	Method 1		Method 2		Empirical																									
				LM	LM(S)	LM	LM(S)	LM	LM(S)																								
10	0.2	3	200	0.066	0.065	0.071	0.070	0.085	0.04																								
			500	0.091	0.089	0.104	0.101	0.11	0.075																								
			1000	0.133	0.129	0.159	0.154	0.185	0.14																								
			5000	0.464	0.447	0.569	0.550	-	-																								
			10000	0.753	0.734	0.854	0.839	-	-																								
			200	0.069	0.068	0.071	0.070	0.07	0.055																								
			500	0.010	0.097	0.102	0.010	0.135	0.085																								
			1000	0.151	0.146	0.157	0.151	0.145	0.135																								
			5000	0.538	0.517	0.561	0.540	-	-																								
			10000	0.828	0.809	0.848	0.829	-	-																								
			10	0.5	3	200	0.158	0.145	0.170	0.155	0.202	0.13																					
						500	0.325	0.292	0.353	0.317	0.41	0.34																					
						1000	0.567	0.514	0.609	0.555	0.625	0.585																					
						5000	0.997	0.994	0.998	0.997	-	-																					
						10000	1	1	1	1	-	-																					
5	200	0.163				0.148	0.168	0.153	0.21	0.15	0.337	0.301	0.347	0.310	0.425	0.345																	
																	1000	0.585	0.529	0.601	0.544	0.61	0.57										
																								5000	0.998	0.995	0.999	0.996					
																													10000	1	1	1	1



## 5 Conclusion

In this paper we presented two methods to compute the power of the LM and LM(S) tests, based on their asymptotic distributions under the alternative hypothesis. Moreover, we assessed the performance of these two tests to detect measurement non-invariance under correct model specification and misspecification of the latent variable distribution. The simulation study highlighted that the asymptotic power, computed through the two different approximation methods for the noncentrality parameter, is very close to the empirical power, also under model misspecification. Small differences between the empirical and asymptotic power have been found also by Gudicha et al. (2017) for the Likelihood-Ratio and Wald tests and by Saris et al. (1987) for the score test.

To compute the noncentrality parameter of the LM and LM(S) tests, we have generated data from the model under the alternative hypothesis considering 10000 observations. Increasing this number could reduce the differences between the empirical and asymptotic power, but it would increase the time burden to obtain the parameter estimates and the numerical derivatives used in the noncentrality parameter approximation procedures.

For what concerns the power of the two tests to detect measurement non-invariance, the LM test has a slightly higher power compared to the LM(S) test under most simulation conditions. The two tests reach the same power only for large sample sizes. A similar behaviour of the power of the LM and LM(S) tests has been found also by Falk and Monroe (2018), under correct model specification and misspecification due to an omitted cross-loading.

From this study we can conclude that the asymptotic power can be a valid alternative to obtain the power of a test, both under the correct model and a model with a misspecified distribution of the latent variable since it allows us to reduced the time complexity compared to the empirical power. Although not shown here, the asymptotic power can be used also to find sample sizes necessary to reach a certain power (Boos and Stefanski 2013, Gudicha et al. 2017). However, the asymptotic power can be computed only for certain test statistics with known noncentrality parameter.

This work was limited only to one type of misspecification. Further analysis should be carried out on the LM(S) test to evaluate if there might be an improvement in its performance considering different types of model misspecification and different estimation methods.

## Bibliography

- Agresti A. (2002). *Categorical data analysis*. New York: Wiley.
- Azzalini A. (1985). A class of distributions which includes the normal ones. *Scandinavian journal of statistics*. pp. 171-178.
- Bera A. K., Doğan O., Biliş Y., Yoon M. J. and Taşpınar S. (2020). Adjustments of Rao's Score Test for Distributional and Local Parametric Misspecifications. *Journal of Econometric Methods*. 9(1), pp 1-29.
- Bollen K. A. (1989). *Structural Equations with Latent Variables*. John Wiley & Sons.
- Boos D.D. and Stefanski L.A. (2013). *Essential statistical inference*. Springer.
- Cox D.R. and Hinkley D.V. (1979). *Theoretical statistics*. CRC Press.
- Engle R. F. (1984). Chapter 13 Wald, likelihood ratio, and Lagrange multiplier tests in econometrics. *Handbook of Econometrics*. 2, pp. 775 - 826. Elsevier.
- Falk C. F. and Monroe S. (2018). On Lagrange multiplier tests in multidimensional item response theory: Information matrices and model misspecification. *Educational and psychological measurement*. 78(4), pp. 653-678.
- Fox J.-P. and Glas C. A. (2005). Bayesian modification indices for IRT models. *Statistica Neerlandica*. 59(1), pp. 95-106.
- Glas C. A. (1998). Detection of differential item functioning using Lagrange multiplier tests. *Statistica Sinica*. pp. 647-667.
- Glas C. A. (1999). Modification indices for the 2-PL and the nominal response model. *Psychometrika*. 64(3), pp. 273-294.
- Glas C. A. and Falcón J. C. S. (2003). A comparison of item-fit statistics for the three-parameter logistic model. *Applied Psychological Measurement*. 27(2), pp. 87-106.
- Gudicha D. W., Schmittmann V. D. and Vermunt J. K. (2017). Statistical power of likelihood ratio and Wald tests in latent class models with covariates. *Behavior research methods*. 49(5), pp. 1824-1837.
- Liu Y. and Thissen D. (2012). Identifying local dependence with a score test statistic based on the bifactor logistic model. *Applied Psychological Measurement*. 36(8), pp. 670-688.
- Mellenbergh G. J. (1982). Contingency table models for assessing item bias. *Journal of educational statistics*. 7(2), pp. 105-118.
- Mellenbergh G. J. (1983). Contingency table models for assessing item bias. In: Irvine S.H., Berry J.W. (eds) *Human Assessment and Cultural Factors*. NATO Conference Series (III Human Factors). 21. Springer, Boston, MA.
- Reiser M. (2008). Goodness-of-fit testing using components based on marginal frequencies of multinomial data. *British Journal of Mathematical and Statistical Psychology*. pp. 331-360.
- Saris W., Satorra A. and Sorbom D. (1987). The Detection and Correction of Specification Errors in Structural Equation Models. *Sociological Methodology*. 17, pp. 105.
- Satorra A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika*. 54(1), pp. 131-151.
- Silvey S. D. (1959). The Lagrangian multiplier test. *The Annals of Mathematical Statistics*. 30(2), pp. 389-407.

White H. (1982). Maximum Likelihood Estimation of Misspecified Models. *Econometrica* . 50(1), pp. 1-25.