International Review of Social Sciences and Humanities
Vol. 11, No. 2 (2016), pp. 80-96
www.irssh.com
ISSN 2248-9010 (Online), ISSN 2250-0715 (Print)

Robust Estimation Methods in Confirmatory Factor Analysis of Likert Scales: A Simulation Study

Francisco Pablo Holgado-Tello
(Corresponding Author)
Universidad Nacional de Educación a Distancia (UNED)
Methodology of Behavioral Sciences Department
Juan del Rosal, 10, 28040, Madrid, Spain
E-mail: pfholgado@psi.uned.es

M. de los Ángeles Morata-Ramírez
Universidad Nacional de Educación a Distancia (UNED)
Methodology of Behavioral Sciences Department
Juan del Rosal, 10, 28040, Madrid, Spain
E-mail: mangelesmorata@yahoo.es

M. Isabel Barbero García
Universidad Nacional de Educación a Distancia (UNED)
Methodology of Behavioral Sciences Department
Juan del Rosal, 10, 28040, Madrid, Spain
E-mail: mbarbero@psi.uned.es

(Received: 23-5-16 / Accepted: 16-7-16)

Abstract

The measurement of psychological and sociological interventions often implies important random and systematic degrees of error which bias the substantive conclusions. Results can be affected by the mismatch between the statistical assumptions and the data characteristics. When using Confirmatory Factor Analysis to obtain construct validity evidences, it is usual to work with moderate or severe skewed ordinal variables. Although robust estimation methods are recommended (specifically, RML and RULS), there are no simulation studies enough to know in which conditions such methods are efficient. In this study we compare through Type-I error correctly specified and misspecified models analyzing the performance of RML and RULS in Likert scales according to χ^2, RMSEA and RMR. Four experimental factors were manipulated: the number of dimensions, the number of response categories, the degree of skewness, and the sample size. We advise using both estimation methods and give some practical recommendations in order to avoid systematic error.

Keywords: Confirmatory Factor Analysis, Likert scales, RML, RULS, Type I error.

Introduction

Construct validity assures that the observed behaviors included in a measurement instrument are true indicators of the construct to be measured. Structural Equation Modeling (SEM) consists in elaborating a model from the structure of relationships between the variables linked to the theoretical knowledge about a construct (Catena, Ramos & Trujillo, 2003). Along with Confirmatory Factor Analysis (CFA), SEM is considered a special case of covariance structure modeling and distinguishes between latent and observed variables (Catena et al., 2003; Cea, 2004). In case the measurement instrument is a Likert scale, one of the CFA application requirements, that is, the continuous nature of both latent and continuous variables (Mulaik, 1972) is not fulfilled. In this sense, it should be remembered that authors as Coenders and Saris (1995), DiStefano (2002) as well as Flora and Curran (2004) point out that the observed variables of Likert scales are measured according to an ordinal measurement scale. In addition, another application requirement, which implies that observed variables must follow a multivariate normal distribution, is not frequently satisfied in Likert scales. This is due to a certain degree of skewness, which can often be observed in the distribution of observed variables, and it means that a non-normal multivariate distribution occurs.

When running a CFA, the lack of fulfilment of the two mentioned requirements can lead to accept misspecified models and therefore to a wrong theoretical development of constructs or artifactual concepts because of a bad praxis in data analysis. More concretely, in the parameters estimation stage there are methods as Maximum Likelihood (ML), which is mostly used although it ignores the peculiarities of Likert scales (Brown, 2006). At the same time, the results of both goodness of fit indices and significance of parameters depend on the parameters estimation method chosen and in turn, they may influence the model reformulation stage from maladjusted starting criteria.

Considering the situation posed by Likert scales in relation to construct validation, it seems appropriate to analyze the behaviour of Robust Maximum Likelihood (RML) and Robust Unweighted Least Squares (RULS) estimation methods. Both ML and RML methods consider that the observed variables are continuous, but it is advisable to use RML when the multivariate normal distribution requirement is not fulfilled by the observed variables (Brown, 2006; Finney & DiStefano, 2013). For its part, RULS method, analyzed by authors as Yang-Wallentin, Jöreskog, and Luo (2010), is a variant of unweighted Least Squares (ULS) method, which is applied when the observed variables do not follow a multivariate normal distribution (Morata-Ramírez, Holgado-Tello, Barbero-García & Méndez, 2015). As far as we know, there are few simulation studies focused on Type I error of RML and RULS methods despite these topics are covered in articles as Lei (2009), Yang-Wallentin et al. (2010), Savalei and Rhemtulla (2013), Sass, Schmitt, and Marsh (2014) and Li (2015).

When the multivariate normal distribution requirement is not fulfilled, it is necessary to calculate de asymptotic covariance (AC) matrix (Jöreskog, 2004). This matrix is involved in the parameters estimation stage of CFA, when the generic expression $F = [\, S - \Sigma\ (p)\ ]'\ W\ [\, S - \Sigma\ (p)\ ]$ is minimized (Bollen, 1989) and the weighting matrix W is the inverse of the AC matrix (Batista & Coenders, 2000). The AC matrix weighs a polychoric correlation matrix previously obtained in order to achieve a free distribution matrix (Jöreskog & Sörbom, 1988). In RML method, AC matrix weighs a Pearson correlation matrix (Batista & Coenders, 2000).

Later, in the goodness-of-fit evaluation stage of CFA, the χ^2 Likelihood Ratio Test is the index par excellence. It is the unique goodness-of-fit index which implies a statistical test where the null hypothesis stands when the theoretical model adjusts to the empirical model. In case of a significant χ^2 value ($p< .05$), the null hypothesis is rejected and it is considered that there is no adjustment between theoretical and empirical model (Schumacker & Lomax, 1996; Cea, 2004; Brown, 2006).

According to Jöreskog (2004), LISREL program includes AC matrix depending on the estimation method and it displays a series of χ^2 indices in the output. As a result, when RML method is used, there are four χ^2 indices obtained: C1 ("Minimum Fit Function Chi-Square"), C2 ("Normal Theory Weighted Least Squares Chi-Square"), C3 ("Satorra-Bentler Scaled Chi-Square") and C4 ("Chi-Square Corrected for Non-Normality"). If RULS method is applied, the χ^2 indices obtained are C2, C3 and C4. It must be highlighted that C3 index is a correction of typical χ^2 for non-normality conditions (Brown, 2006; Finney & DiStefano, 2013). It was proposed by Satorra and Bentler (1990) in order to test the hypothesis about goodness of fit when the normality assumption is not fulfilled (Cea, 2004).

Given that χ^2 index has limitations derived from its connection to χ^2 central distribution and its sensitivity to sample size, it is recommended to obtain additional results with another goodness-of-fit indices (Bollen, 1989; Byrne, 1998; Cea, 2004; Brown, 2006). In the present simulation study, those indices are RMSEA and RMR.

RMSEA (Root Mean Square Error of Approximation) is a goodness of fit index, which measures the level of discrepancy between the theoretical and the empirical model according to the degrees of freedom. When its value is smaller than .05, it is considered that there is a good adjustment of the theoretical model, whereas the adjustment or approximation error is reasonable when RMSEA values are between .05 and .08 (Browne & Cudeck, 1993). RMSEA is sensitive to the number of parameters of the theoretical model, that is, the complexity of the model. In this sense, the simpler the model is the greater degree of freedom it has, so there is a greater probability to reject the model (Byrne, 1998; Batista & Coenders, 2000).

RMR (Root Mean Square Residual) is a goodness of fit index which represents the average residual value which results of the differences between the variance-covariance theoretical and observed matrices (Byrne, 1998). Although a RMR value equal to zero indicates a perfect adjustment of the theoretical model (Kline, 2011), there is a good adjustment if RMR is smaller than .05 (Sörbom & Jöreskog, 1982; Cole, 1987).

The objective is to compare through Type I error the performance of RML and RULS in Likert scales according to χ^2, RMSEA and RMR indices.

Method

Four experimental factors were manipulated: a) number of factors, or latent variables, b) number of response categories, c) degree of skewness, and d) sample size. The number of factors had five experimental levels (2, 3, 4, 5, 6). For each factor were simulated three items. The factor loadings of the items were always the same in all factors, namely .9, .8 and .7 for the first, second and third item of each factor. Items were generated according to a normal distribution *N (0, 1)*. Then these answers were categorized according to 3, 4, 5, and 6 point Likert scales, that is, the number of response categories was established with four experimental levels. Likert scales were categorized so that: a) the responses to all items remained symmetrical, b) all items had moderate skewness, or c) all items had severe skewness. Thus, the degree of skewness had three experimental levels: skewness = 0; skewness = 1, or moderate; skewness = 2, or severe.

To categorize the Likert scales, as stated by Bollen and Barb (1981), the continuum was divided into equal intervals from z = -3 to z = 3 in order to calculate the thresholds of the condition in which the response distribution to all items is symmetrical (skewness = 0). For skewed distributions, the thresholds were calculated, according to Muthén and Kaplan (1985), in such a way that observations were accumulated in one of the extreme categories as the degree of skewness increased. Half of the variables of each factor were categorized with the same positive skewness and the rest of variables with the same negative skewness with the purpose of simulating difficulty factors.

As regards the fourth experimental factor, the sample size had six experimental values (100, 150, 250, 450, 650 and 850 subjects).

The combination of number of factors, number of response categories, degree of skewness and sample size produced 360 experimental conditions (5x4x3x6) which were replicated 500 times. These replications were performed using R version 2.12.0 (R Development Core Team, 2010), which invoked successively PRELIS version 2.0 (Jöreskog & Sörbom, 1996b) to generate the corresponding data matrices according to the specifications resulting from the combination of the experimental conditions. Thus, for each generated matrix, Pearson and polychoric correlations matrices as well as asymptotic covariance matrices were obtained.

After obtaining correlations matrices and asymptotic covariance matrices for each data matrix generated under the concrete specifications of the experimental factors, it was performed the corresponding Confirmatory Factor Analysis (CFA) successively, that is, until 500 times (one CFA for each replication). As in the previous case, these replications were performed using R version 2.12.0, which invoked successively LISREL version 8.80 (Jöreskog & Sörbom, 1996a). There were two types of theoretical models which were tested through LISREL:

1. Correctly Specified Models: This means that the theoretical model which has been specified using LISREL syntax corresponds exactly with the model which has enabled data generation. So, for example, if we are working with a 5 factors model, in the syntax it is specified that the model to be tested has 5 factors, where the factor loading of each item depends on its theoretical factor.

2. Misspecified Models: In this case, the theoretical model which has been specified using LISREL syntax does not correspond exactly with the model which has enabled data generation. To do this, in the syntax it is specified a model so that an item of each factor saturates incorrectly in another one. For example, if the model generated by the data had 2 factors, an item of each factor is saturated incorrectly in the syntax by including it in a factor which is different from the one that would correspond theoretically with the item.

In order to facilitate the management and compilation of fit indices files, a specific program was generated in JAVA language.

Data Analysis

Type I error was calculated by the percentage of rejections of the null hypothesis in correctly specified and misspecified models. All these calculations were made for each of the χ2 values reported in the LISREL output files obtained for each estimation method used in the present study. As we mentioned before, those χ^2 values are known as C1 (“Minimum Fit Function Chi-Square”), C2 (“Normal Theory Weighted Least Squares Chi-Square”), C3 (“Satorra-Bentler Scaled Chi-Square”) and C4 (“Chi-Square Corrected for Non-Normality”) (Jöreskog, 2004). Thus, the χ^2 values obtained for each estimation method depending on the fulfillment of the multivariable normality assumption were the following ones: for RML method, C1, C2, C3 and C4; for RULS method, C2, C3 and C4.

To raise the null hypothesis for χ^2 Likelihood Ratio Test a 5% nominal value was considered, which in practice means a probability equal to or greater than 0.05. As far as RMSEA index is concerned, the theoretical models are accepted when its value is lower than 0.08, so models with a reasonable adjustment (Browne & Cudeck, 1993) were included. For its part, according to RMR index, models with values lower than .05 were also accepted (Sörbom & Jöreskog, 1982; Cole, 1987).

Results

Number of Factors

The results related to percentage of Type I error for correctly specified and misspecified models of each of RML and RULS estimation methods depending on the number of factors of the theoretical models examined are presented in Table 1.

Table 1: Number of factors

		Correctly Specified Models		Misspecified Models	
	Factors	% Type I error		% Type I error	
		RML	RULS	RML	RULS
C1	2	23.4		81	
	3	21.3		95.7	
	4	22.9		97.1	
	5	25.1		98.6	
	6	12.7		100	
C2	2	24.5	56	81.8	98
	3	20.7	76	96.1	100
	4	21.8	85.2	96.8	100
	5	22.9	89.8	98.1	100
	6	9.8	81.5	100	100
C3	2	8.8	12	60.5	83.7
	3	0.8	5.6	76.1	98
	4	0.3	5.9	72.6	98.8
	5	0.1	7.1	70.6	99.3
	6	0.1	5.3	99.3	100
C4	2	10	14.2	55.7	82.6
	3	3.6	15.4	73.8	98.7
	4	11.9	36.8	87.4	100
	5	32.7	58	96.8	100
	6	45.5	74.9	100	100
RMSEA	2	6	8.7	35.3	68
	3	0.1	0.5	47.2	91.6
	4	0	0.1	28.9	74.9
	5	0	0.1	24.2	68.1

	6	0	0	64.3	96.9
RMR	2	16.9	20.3	67.2	94.8
	3	19.4	27	98.9	100
	4	25.9	32.7	99.6	100
	5	29.8	34.2	99.7	100
	6	10.2	15.8	99.7	100

In correctly specified models, as regards Type I error made with RML method, according to C1 and C2 the percentage of correct rejected models remains stable by 21-25% as the number of factors increases. However, for 6 factors models it decreases to 12.7% (C1) and 9.8% (C2). For C3 there are no models rejected except for 2 factors models (8.8%). For its part, C4 shows an increasing in the percentage of rejected correct models as the number of factors is greater, although the smallest percentage of rejections is found for 3 factors models (3.6%).

With respect to Type I error with RULS, the highest percentage of rejected models is achieved with C2 for 5 factors models (89.8%) from a great increasing observed for 2 and 3 factors models. That percentage remains stable according to C3 (approximately 5-7%) for models with more than 3 factors, since for 2 factors models it reaches 12%. When observing C4 results, there are remarkable upward trends in the percentages from 3 factors on. In this sense, more than one third of correct models is rejected when they have 4 factors, whereas more than half of the models with 5 factors is rejected and two thirds of 6 factors models are rejected.

As far as Type I error made with RMSEA index is concerned, either of the estimation methods show no rejected models except for 2 factors models, where the percentage of rejection is 6% for RML and 8.7% for RULS.

With respect to Type I error related to RMR index, in both estimation methods the percentage of rejected models go up as the number of factors is greater. However, for 6 factors models the percentages are 10.2% for RML and 15.8% for RULS, and therefore they are smaller than the ones obtained for models with 2 factors (16.9% for RML and 20.3% for RULS).

In misspecified models C1 and C2 show similar Type I error percentages when using RML method, which implies the rejection of most of all the misspecified models with 3, 4 or 5 factors. Those percentages are lower for 2 factors models (around 81%), whereas all the models with 6 factors are rejected. For its part, C3 rejects over half of 2 factors models (60.5%) and this percentage increases up to 70-76% for the rest of the models except for 6 factors models, since almost all of them are rejected (99.3%). A similar trend is observed for C4 but the increasing percentage of rejected incorrect models is more gradual.

As regards RULS method, almost all 2 factors models (98%) are rejected with C2, whereas this percentage is slightly lower with C3 and C4 (83.7% and 82.6%, respectively). All misspecified models with 3 or more factors are rejected with C2, whereas C3 and C4 reject all the models with 6 and 4 factors, respectively.

With respect to Type I error related to RMSEA index, there is no clear upward or downward trend as the number of factors increase with RML and RULS methods. However, both of them show the highest percentages of Type I error for 6 factors models (64.3% and 96.9%, respectively).

As far as Type I error made with RMR index is concerned, when using RML almost all the models with 3 or more factors (namely, from 98.9% to 99.7% of them) are rejected, while the percentage of 2 factors models which are rejected is a lower (67.2%). In contrast, RULS method rejects 94.8% of 2 factors models as well as the rest of models with a greater number of factors.

Number of Categories

Table 2 shows the percentage of Type I error for correctly specified and misspecified models of each of RML and RULS estimation methods depending on the number of response categories of the models.

Table 2: Number of categories

		Correctly Specified Models		Misspecified Models	
	Categories	% Type I error		% Type I error	
		RML	RULS	RML	RULS
C1	3	19		91.2	
	4	26.9		95.2	
	5	21.7		93.7	
	6	23.1		94.1	
C2	3	17.8	90.4	91.4	99.7
	4	25.8	82.6	95.2	99.6
	5	20.7	71.6	93.8	99.5
	6	22.2	64.3	94.3	99.4
C3	3	0.1	6.8	60	93.9
	4	6.9	12	74.3	96
	5	0.9	5.8	74.7	95.7
	6	1.4	5.4	78.6	95.4
C4	3	10.3	32.4	68.8	94.1
	4	20.3	38.7	83.3	96.9
	5	16.6	32.6	82.1	95.9
	6	19.4	32.4	85.2	95.5
RMSEA	3	0	0.7	13.8	66.1
	4	5.2	6.6	39.8	77.7
	5	0.2	0.8	43.6	80.7
	6	0.2	0.7	46.7	83.2

RMR	3	21.5	31.9	88.7	98.8
	4	27.6	32.6	94.3	99.1
	5	20	24.6	91.9	98.6
	6	19.4	21.9	92.6	98.6

With respect to Type I error in correctly specified models using RML, the percentages of correct models rejected according to C1 and C2 show an upward trend, although the models with 4 categories have a higher probability of being rejected than the rest: 26.9% (C1) and 25.8% (C2). There are hardly rejected models in C3 index, except for 4 categories models (6.9%). In accordance with C4, there is an increasing tendency in the percentage of Type I error but the models with 4 categories show again a greater percentage of rejected correct models (specifically, 20.3%).

When RULS method is applied, C2 presents a downward tendency in Type I error as the number of categories is greater. In both C3 and C4 indices the percentages of accepting correct models remain stable, except for 4 categories models, where those percentages are higher (12% and 38.7%, respectively).

As regards RMSEA, both RML and RULS methods show that the correct models with 4 categories are the ones with more probabilities of being rejected (5.2% and 6.6%, respectively).

According to RMR index, with RML the percentage of Type I error goes around 20%, while for 4 categories models it increases until 27.6%. When using RULS method, the percentage of rejected models remains stable for 3 and 4 factors models, and then it goes down. Once again the models with 4 categories present the higher percentage of Type I error (namely, 32.6%).

Working with misspecified models, as far as χ^2 index is concerned, with RML estimation method C1 and C2 show a slight upward trend as the number or categories grows, but the highest Type I error percentages correspond to 4 categories models (95.2%). According to C3 and C4, there is a greater percentage of Type I error as the number of categories increases from 3 to 4 categories. Then the percentages are relatively stable, but the percentage of rejected models with 6 categories in C3 is smaller than in C4 (78.6% and 85.2%, respectively). When RULS method is used, C2 rejects almost all the incorrect models (about 99%). C3 and C4 present quite stable percentages of Type I error, but the models with 4 categories show the highest ones (96% and 96.9%, respectively).

With respect to RMSEA, both estimation methods show an increasing in the percentages of Type I error up to 46.7% (RML) and 83.2% (RULS) as the models have more categories. That increasing is more pronounced from 3 to 4 categories. In addition, the percentages are greater with RULS than with RML, particularly for 3 categories models. In this case, the percentage shown by RML is 13.8%, whereas the percentage of RULS is 66.1%.

As regards RMR index, RML shows a slight upward trend as the number of categories is greater, except for 4 categories models, which has a higher percentage (namely, 94.3%). For its part, RULS show stable percentages around 98%, but the models with 4 categories have also the greatest one (that is, 99.1%).

Degree of Skewness

In Table 3 we present the results about percentage of Type I error for correctly specified and misspecified models obtained with RML and RULS estimation methods depending on the degree of skewness.

Table 3: Skewness

		Correctly Specified Models		Misspecified Models	
	Skewness	% Type I error		% Type I error	
		RML	RULS	RML	RULS
C1	0	7.1		97.5	
	1	17.9		94.2	
	2	46.1		88.2	
C2	0	5.4	56.4	97.7	99.2
	1	16.3	87.6	94.4	99.6
	2	46.3	96.7	88.1	99.8
C3	0	0.6	5.4	94.1	97.7
	1	3.4	9.2	73.7	95.4
	2	3.3	8.6	43.6	91.9
C4	0	23.2	39.3	94.5	97.1
	1	14.2	33.4	81.9	95.5
	2	11.4	26.2	60.3	93.7
RMSEA	0	0.3	0.8	77.4	96.1
	1	2.9	3.8	21.2	84.2
	2	1.2	2.2	1	44.9
RMR	0	14.2	18.8	96.9	98.6
	1	22.8	35.2	91.2	98.7
	2	30.8	32	86.6	99

In correctly specified models the smallest percentage of Type I error when using RML method is achieved by C3 index. As regards C1 and C2, it can be noticed a similar upward trend in the percentage of rejected correct models as the skewness grows, which reaches in both of them about 46% when the skewness is severe. Moreover, according to C3, almost none of the correct models are rejected when the response distribution is symmetric, whereas only about 3% of correct models with moderate and severe skewed distributions are rejected. C4 shows a slight downward trend from 23.2% to 11.4% in the percentage of rejected correct models as their degree of skewness is greater.

When using RULS method, C3 is also the index which shows the smallest Type I error. For its part, C2 presents an upward trend in the percentage of rejected correct models. In this sense, almost half of the models (56.4%) with symmetrical distributions is rejected, whereas almost all the correct models (96.7%) with severe skewed distributions are rejected. In contrast, C3 shows a small percentage of correct rejected models for symmetrical distribution (5.4%) in comparison with the similar percentages obtained for both moderate and severe skewed distributions (9.2% and 8.6%, respectively). Regarding C4 index, there is a slight downward tendency as the response distributions of correct models have a greater degree of skewness, falling from 39.3% for symmetrical distributions to 26.2% for severe skewed distributions.

With respect to RMSEA, Type I error percentages shown by RML and RULS methods are alike. As regards correct models with symmetric response distributions, only 0.3% of them are rejected with RML and 0.8% with RULS. The percentages of rejected models are slightly greater for skewed distributions, but for moderate skewness they are bigger than for severe skewness.

Type I error in RMR shows an upward tendency for RML method in the percentage of rejected correct models as skewness is greater, going from 14.2% to 30.8%. As regards RULS, there are more rejected correct models as long as their response distribution is skewed instead of symmetrical, but for moderate skewed distributions the percentage of rejection is higher (35.2%) than for severe skewed distributions (32%).

In misspecified models, regarding χ^2 index when using RML, a huge percentage of misspecified models is rejected in accordance with C1 and C2. In this sense, despite the slight downward tendency as the skewness grows, for severe skewed response distributions that percentage is about 90%. C3 and C4 are similar in percentages of Type I error, but there are less rejected models when C3 is considered, particularly for severe skewed distributions. Namely, 43.6% with C3 and 60.3% with C4.

As far as RULS is concerned, 99% of incorrect models are rejected with C2 whatever the degree of skewness. In addition, the percentages of C3 and C4 are over 90%, but they are lower as the degree of skewness increases.

According to RMSEA index, with RML there are great differences between Type I error percentages depending on the degree of skewness. In other words, 77% of misspecified models with symmetric distribution is rejected, whereas this percentage goes down to 21.2% for models with moderate skewness and to only 1% for severe skewness. There is also a decreasing trend with RULS method, but it is more pronounced when comparing moderate to severe skewness and, moreover, the amount of rejected incorrect models is greater than the one found for RML. Hence the percentages go from 96.1% for symmetric distributions to 44.9% for severe skewness.

As regards RMR, most of the misspecified models are rejected with RML. In this sense, Type I error percentages go down from 96.9% to 86.6% for symetric and severe skewed distributions, respectively. As RULS method is used, those percentages are higher and close to 100%, especially when the skewness is severe.

Sample size

In Table 4 we present the data related to sample size with respect to percentages of Type I error for correctly specified and misspecified models obtained with RML and RULS estimation methods depending on the sample size.

Table 4: Sample size

		Correctly Specified Models		Misspecified Models	
	Sample size	% Type I error		% Type I error	
		RML	RULS	RML	RULS
C1	100	10.7		74.7	
	150	14.7		89.1	
	250	17.4		96.5	
	450	24.6		99.3	
	650	30.1		99.8	
	850	37.6		99.9	
C2	100	6.8	70.5	74.3	97.7
	150	12	75.9	89.8	99.2
	250	16.2	79	96.7	99.9
	450	24.8	78.8	99.4	100
	650	30.8	77.6	99.9	99.9
	850	37.8	77.4	99.9	100
C3	100	0.3	6.1	31.1	82.8
	150	2.8	9.5	50.5	90.7
	250	3.9	9.5	67.4	96.7
	450	4	9	86.5	99.2
	650	0.8	5.5	95.2	99.8
	850	2	5.1	96.6	100
C4	100	39.1	59.5	61.4	86.1
	150	32.2	53.9	66.7	91.1
	250	16.8	44.9	72.4	95.9
	450	8	28.3	86.5	99
	650	2.7	17.6	93.2	99.7
	850	3.3	13.3	97	99.9
RMSEA	100	0.6	3.9	25.2	74.3
	150	2.7	3.4	32.7	74.6
	250	3.1	3.3	37.3	76.7
	450	2	3.3	39.9	79.1
	650	0	0	39.2	78.4
	850	0	0	40.5	78.8

RMR	100	75.9	82.4	95.6	98.9
	150	47.3	65.1	95	98.7
	250	10	36.1	93.4	98.7
	450	4	9.5	91.7	98.8
	650	0.1	0.4	88.6	98.8
	850	0	0	87.5	98.8

Type I error in correctly specified models when using RML is very similar for C1 and C2, since they show an upward tendency as the sample is bigger. In this sense, the percentage of correct models with 850 cases rejected is approximately 37%. When C3 is considered, the tendency in the percentage of correct rejected models is not clear. In other words, the smallest percentages of rejection correspond to models with 100 and 650 cases (0.3% and 0.8%, respectively), while the highest ones belong to models with 250 and 450 cases (about 4%). According to C4, Type I error percentages diminish as models have a greater number of cases. Namely, they fall from 39.1% (models with 100 cases) to 3.3% (models with 850 cases), although the percentage for 650 cases is a bit smaller (2.7%).

As far as RULS method is concerned, Type I error for C2 presents an upward trend, going from 70% to approximately 77%. However, the percentage of rejection of correct models increases until about 79% for 250 and 450 cases. For its part, for C3 the percentages of rejection are small (around 5-6%) for models with 100, 650 and 850 cases, whereas for models from 150 to 450 cases the percentage is higher (about 9%). There is a downward trend for C4 related to the percentages of rejection as the sample size is bigger, falling from 59.5% to 13.3%.

When using RML method, a few correct models are rejected according to RMSEA index. In other words, only 2-3% of the models with a sample size between 150 and 450 cases and none of the models with 100, 650 or 850 cases are rejected. This situation is similar with RULS method, but in this case there is a clear downward trend when the number of cases is greater, as for models with 100 cases samples the percentage is 3.9% and none or the models with 650 and 850 cases samples is rejected.

Type I error for both RML and RULS methods show a downward tendency, reaching percentages close or equal to 0% for models with 650 and 850 cases. However, with RML there is a noticeable decrease when the models with 100, 150 and 250 cases are compared (namely, the percentage of rejected models falls from 75.9% to 10%), whereas when RULS is used that reduction is more progressive.

In misspecified models, with respect to C1 and C2 indices when using RML method, about 75% of the misspecified models with 100 subjects are rejected. This percentage increases until 99% for models with 450 subjects or more. There are similarities between C3 and C4 values. In this sense, the percentage of rejected incorrect models is higher as the sample size grows, reaching values above 90% for samples with 650 and 850 subjects. However, C3 and C4 percentages are different for 100 cases, since one third of the models are rejected with C3 and over half of them with C4.

As regards RULS method, the percentages of rejected misspecified models with C2 are close or equal to 100% whatever the sample size. C3 shows high values with an upward trend in Type I error percentage as the sample size grows but they are only very close or equal to 100% in models with 450 cases or more. For its part, C4 values are very similar to C3.

As far as RMSEA is concerned, Type I error percentages obtained with RML method are lower than the ones with RULS method, despite both of them show a very slight upward tendency. In this sense, one fourth of misspecified models with 100 subjects are rejected with RML and two thirds of them with RULS, whereas Type I error percentages for models with 450 to 850 subjects raise to about 40% with RML and about 79% of them with RULS.

Concerning RMR index, most of misspecified models are rejected with both RML and RULS methods. However, as the sample size grows, Type I error percentages go down when RML is used, whereas those percentages remain stable at approximately 98% with RULS.

Discussion

Number of Factors

As regards the number of factors and Type I error obtained in correctly specified models from χ^2, RML shows a lower percentage of rejected models than RULS, particularly in C3 values. If RMSEA is considered, both estimation methods reject a minority of correct models with 2 factors and accept those with 3 or more factors. For its part, in accordance with RMR index there is a higher percentage of accepted models when using RML instead of RULS method, especially with 3 factors models.

When misspecified models are considered regarding χ^2, the percentages of rejection are higher with RULS. In this sense, the greater differences between both estimation methods come from C3 values (except for 6 factors models) as well as from C4 values up to 4 factors models. Regarding RMSEA index, the amount of misspecified rejected models is greater when RULS is used (particularly with models with 3 or 6 factors) despite the percentages of rejection for those specific models are also high with RML. Related to RMR there are no noticeable differences between the estimation methods except for 2 factors models, where RULS permits the rejection of most of misspecified models.

Number of Categories

With respect to the number of categories and Type I error obtained in correctly specified models, as χ^2 is taken into consideration there are fewer rejected models when using RML method, particularly if we pay attention to C3 values. As RMSEA is concerned, both estimation methods accept all correct models but the percentage shown by RML is lower than RULS, even with 4 categories models, where the percentage of rejected correct models is a bit higher than the rest. RMR index shows lower percentages of rejection when RML is used, even though the differences between both methods are greater as the number of categories decreases.

Regarding misspecified models, the results from χ^2 show that RULS method reject a bigger amount of models, particularly C3 values. For its part, RMSEA index shows a higher percentage of rejected misspecified models when using RULS instead of RML, mainly if they have 3 response categories. When RMR is considered, almost all the incorrect models are rejected with RULS method, despite the large amount of models which RML method rejects.

Degree of Skewness

Taking into account the degree of skewness in correctly specified models, as regards χ^2 the percentage of Type I error is greater with RULS. Consequently, RML rejects fewer correct models than RULS, particularly when C3 is considered. As regards RMSEA index, there is a small amount of rejected models but RULS method shows higher rejection percentages than RML, especially for skewed distributions. The results obtained for RMR index are very similar in both estimation methods, but the percentages of rejected correct models are higher when RULS is used, particularly with moderate skewed response distributions.

For misspecified models, the percentage of rejected models according to χ^2 is greater with RULS method. It is worthy to note that C3 shows the biggest differences between the estimation methods, especially with severe skewed response distributions. For its part, RMSEA index shows higher percentages of rejected misspecified models with RULS rather than RML. In this sense, the differences between these estimation methods are evident with severe skewed distributions. As regards RMR, when using RULS method all the incorrect models are rejected despite the degree of skewness, whereas the high percentages of Type I error with RML decrease as skewness grows, particularly from moderate to severe asymmetry.

Sample Size

Considering the sample size and Type I error in correctly specified models, the results related to χ^2 show that there is a lower percentage of rejected models with RML method instead of RULS method, particularly with C3 values. Regarding RMSEA, the amount of correct models rejected by RML and RULS is very similar, but there are noticeable differences between the estimation methods in samples of 450 subjects and, above all, in 100 subjects. It is remarkable that neither RML nor RULS reject correctly specified models with 650 or 850 subjects. For its part, RMR shows a decreasing trend in the percentages of rejected correct models as the sample size grows for RML and RULS methods, but they are greater when RULS is used. However, all the correct models with 650 or 850 subjects are accepted by both estimation methods.

As far as misspecified models are concerned regarding sample size, the results from χ^2 indicate that Type I error percentages are higher with RULS method, especially C3 and C4 values. In this sense, there are more incorrect models which are rejected by RULS when the sample size is between 100 and 250 subjects. It is also true with 450 subjects, but the differences between the estimation methods are not so pronounced. RMSEA index shows a bigger amount of rejected incorrect models when RULS is used, particularly with 100 subjects, despite the upward trend in both estimation methods as the simple size grows. With respect to RMR, despite the high Type I error percentages which RML present, almost all misspecified models are rejected with RULS whatever the number of subjects.

General Guidelines

Comparing RML and RULS estimation methods in relation to correctly specified models, the percentages of Type I error obtained through χ^2 indices are mostly higher with RULS instead of RML whatever the experimental factor (number of factors, number of categories, degree of skewness and sample size). In other words, given a specific experimental level, C2, C3 and C4 values when using RULS method reject more correct models than RML, despite the lower levels which C3 exhibits. For its part, RMSEA shows rejection percentages of correct models with RML quite similar to RULS, except for 2 factors models, for 4 categories models, for models whose response distribution has moderate or severe skewness and for models from 100 to 450 subjects. In those cases, RULS method rejects clearly more correct models than RML. When RMR index is considered, the rejection percentages are mostly higher if RULS is used, particularly with 3 or 4 factors models, with 3 or 4 categories models, with models whose response distribution has a moderate skewness and models with 150 or 250 subjects.

According to the results obtained in the four experimental factors taken into account, RULS method rejects most of the times a greater amount of misspecified models than RML. That superiority of RULS is more evident in C3 values of χ^2 for models up to 5 factors as well as in

C4 values for models up to 4 factors. Besides, the higher reject percentages are achieved particularly with C3 for all number of categories. As regards the degree of skewness, C3 values reject especially misspecified models with severe skewed distributions. In addition, C3 and C4 values with RULS are higher than the ones with RML for incorrect models, mainly with sample sizes up to 250 subjects (although the rejection percentages with C3 are higher than C4 compared to RML method, particularly with 100 subjects models). As regards RMSEA index when misspecified models are analyzed, RULS method permits the rejection of more models than RML, particularly with 4 factors models, with 3 categories models, with models which have moderate and, above all, severe skewed response distributions and, especially, with 100 subjects models. For its part, RMR shows higher Type I error percentages for RULS method rather than RML method, particularly with 2 factors models, with 3 categories models, with models with severe skewed response distribution and models with 650 or 850 subjects.

Based on the above comments, it seems that, compared to RULS, RML method is more effective when the models are correctly specified, since it rejects a fewer amount of them than RULS. In contrast, RULS method is preferable to RML in order to reject more misspecified models. However, as Jöreskog and Sörbom (1989) defend, polychoric correlations are better than Pearson correlations when the measurement scale of observed variables is ordinal. In this sense, using Pearson correlations (as RML does) result in incorrect $\chi 2$ indices (Morata-Ramírez & Holgado-Tello, 2013).

In addition, the results might have been influenced by the fact that the experimental factors are considered in isolation. In this sense, the lack of information about how the interaction between number of factors, number of categories, degree of skewness and sample size affect Type I error is one of the limitations of the present study.

In this study it has been possible to know the behaviour of C3 value (or Satorra-Bentler χ^2 scaled) with robust estimation methods since the observed variables do not fulfill the CFA normality assumption. As results show, it sticks out from the rest of χ^2 values. According to Cea (2004), kurtosis is involved when calculating C3. For this reason, it might be interesting to take kurtosis into consideration in future studies. It would also supply more information related to skewness.

Moreover, the RMSEA index depends on sample size and degrees of freedom. In others words, the higher the sample size and degrees of freedom, the lower the RMSEA values (Kline, 2011; MacCallum, Browne & Sugawara, 1996). In this sense, following Byrne (1998), another topic to discuss in future investigations when analyzing RMSEA results is the complexity of models related to their degree of freedom.

Finally, as Morris, Bergan, and Fulginiti (1993) point out, model misspecification relies on the mismatch between correlation and causation, so studies need to be cautiously designed in a way that rival hypotheses are also included in them.

References

[1] J.M. Batista and G. Coenders, *Modelos de Ecuaciones Estructurales (Structural Equation Modeling)*, (2000), Madrid: La Muralla.
[2] K.A. Bollen, *Structural Equations with Latent Variables*, (1989), New York: Wiley.
[3] K.A. Bollen and K.H. Barb, Pearson's r and coarsely categorized measures, *American Sociological Review*, 46(2) (1981), 232-239.
[4] T.A. Brown, *Confirmatory Factor Analysis for Applied Research*, (2006), New York: Guildford Press.

[5] M.W. Browne and R. Cudeck, Alternative ways of assessing model fit, In K.A. Bollen and J.S. Long (Eds.), *Testing Structural Equation Models*, (1993), 136-162, Newbury Park, CA: Sage.
[6] B.M. Byrne, *Structural Equation Modelling with LISREL, PRELIS, and SIMPLIS: Basic Concepts, Applications and Programming*, (1998), London: Lawrence Erlbaum Associates.
[7] A. Catena, M.M. Ramos and H.M. Trujillo, *Análisis Multivariado: Un Manual para Investigadores (Multivariate Analysis: A Handbook for Researchers)*, (2003), Madrid: Biblioteca Nueva.
[8] M.A. Cea, *Análisis Multivariable: Teoría y Práctica en La Investigación Social (Multivariate Analysis: Theory and Practice in Social Research)*, (2004), Madrid: Síntesis.
[9] G. Coenders and W.E. Saris, Categorization and measurement quality: The choice between pearson and polychoric correlations, In W.E. Saris and Á. Münnich (Eds.), *The Multitrait-Multimethod Approach to Evaluate Measurement Instruments*, (1995), 125-144, Budapest: Eötvös University Press.
[10] D.A. Cole, Utility of confirmatory factor analysis in test validation research, *Journal of Consulting and Clinical Psychology*, 55(4) (1987), 584-594.
[11] C. DiStefano, The impact of categorization with confirmatory factor analysis, *Structural Equation Modelling*, 9(3) (2002), 327-346.
[12] S.J. Finney and C. DiStefano, Non-normal and categorical data in structural equation modelling, In G.R. Hancock and R.O. Mueller (Eds.), *Structural Equation Modelling: A Second Course*, (2013), 269-314, Information Age Publishing: Greenwich.
[13] D.B. Flora and P.J. Curran, An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data, *Psychological Methods*, 9(4) (2004), 466-491.
[14] K.G. Jöreskog, On Chi-Squares for the independence model and fit measures in LISREL, (2004), Accessed February 20 (2016), http://www.ssicentral.com/lisrel/techdocs/ftb.pdf.
[15] K.G. Jöreskog and D. Sörbom, *LISREL 7: A Guide to the Program and Applications*, (1988), Chicago (IL): SPSS Inc.
[16] K.G. Jöreskog and D. Sörbom, *LISREL 7: User's Reference Guide*, (1989), Chicago: Scientific Software International.
[17] K.G. Jöreskog and D. Sörbom, *LISREL 8: User's Reference Guide*, (1996a), Chicago: Scientific Software International.
[18] K.G. Jöreskog and D. Sörbom, *PRELIS 2: User's Reference Guide*, (1996b), Chicago: Scientific Software International.
[19] R.B. Kline, *Principles and Practice of Structural Equation Modeling (3rd ed.)*, (2011), New York: The Guilford Press.
[20] P.W. Lei, Evaluating estimation methods for ordinal data in structural equation modeling, *Quality & Quantity*, 43(3) (2009), 495-507.
[21] C.H. Li, Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares, *Behavioral Research Methods*, (2015), (Epub ahead of print).
[22] R.C. McCallum, M.W. Browne and H.M. Sugawara, Power analysis and determination of sample size for covariance structure modeling, *Psychological Methods*, 1(2) (1996), 130-149.
[23] M.A. Morata-Ramírez and F.P. Holgado-Tello, Construct validity of likert scales through confirmatory factor analysis: A simulation study comparing different methods of estimation based on pearson and polychoric correlations, *International Journal of Social Science Studies*, 1(1) (2013), 54-61.
[24] M.A. Morata-Ramírez, F.P. Holgado-Tello, I. Barbero-García and G. Méndez, Confirmatory factor analysis: Recommendations for unweighted least squares method related to Chi-square and RMSEA type I error, *Acción Psicológica*, 12(1) (2015), 79-90.

[25] R.J. Morris, J.R. Bergan and J.V. Fulginiti, Structural equation modelling in clinical assessment research with children, In A.E. Kazdin (Ed.), *Methodological Issues & Strategies in Clinical Research*, (1993), 367-386, Washington, D.C.: American Psychological Association.

[26] S.A. Mulaik, *The Foundations of Factor Analysis (Vol. 88)*, (1972), New York: McGraw-Hill.

[27] B. Muthén and D. Kaplan, A comparison of some methodologies for the factor analysis of non-normal likert variables, *British Journal of Mathematical and Statistical Psychology*, 38(2) (1985), 171-189.

[28] R Development Core Team, *R: A Language and Environment for Statistical Computing 2.12.0*, (2010), Vienna: R Foundation for Statistical Computing.

[29] D.A. Sass, T.A. Schmitt and H.W. Marsh, Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators, *Structural Equation Modeling*, 21(2) (2014), 167-180.

[30] A. Satorra and P.M. Bentler, Model conditions for asymptotic robustness in the analysis of linear relations, *Computational Statistics & Data Analysis*, 10(3) (1990), 235-249.

[31] V. Savalei and M. Rhemtulla, The performance of robust test statistics with categorical data, *British Journal of Mathematical and Statistical Psychology*, 66(2) (2013), 201-233.

[32] R.E. Schumacker and R.G. Lomax, *A Beginner's Guide to Structural Equation Modelling*, (1996), New Jersey: Lawrence Erlbaum Associates.

[33] D. Sörbom and K.G. Jöreskog, The use of structural equation models in evaluation research. In C. Fornell (Ed.), *A Second Generation of Multivariate Analysis: Vol. 2: Measurement and Evaluation*, (1982), 381-418, New York: Praeger.

[34] F. Yang-Wallentin, K.G. Jöreskog and H. Luo, Confirmatory factor analysis of ordinal variables with misspecified models, *Structural Equation Modeling*, 17(3) (2010), 392-423.