

Hypothesis testing using factor score regression: A comparison of four methods

Ines Devlieger

Ghent University

Axel Mayer

Ghent University

Yves Rosseel

Ghent University

Author Note

Correspondence concerning this article should be addressed to Ines Devlieger,
Department of Data Analysis, Faculty of Psychology and Educational Sciences,
Ghent University, Henri Dunantlaan 1, 9000 Ghent, Belgium,
e-mail: Ines.Devlieger@ugent.be

Abstract

In this article, an overview is given of four methods to perform Factor Score Regression (FSR), namely regression FSR, Bartlett FSR, the bias avoiding method of Skrondal and Laake (2001) and the bias correcting method of Croon (2002). The bias correcting method is extended to include a reliable standard error. The four methods are compared to each other and to SEM by using analytic calculations and two Monte Carlo simulation studies to examine their finite sample characteristics. Several performance criteria are used, such as the bias using the unstandardized and standardized parameterization, efficiency, mean square error, standard error bias, type I-error rate and power. The results show that the bias correcting method, with the newly developed standard error, is the only suitable alternative for SEM. While it has a higher standard error bias than SEM, it has a comparable bias, efficiency, MSE, power and type I-error rate.

Keywords: Factor score regression, bias, standard error, standardized parameterization, unstandardized parameterization

Hypothesis testing using factor score regression: A comparison of four methods

Introduction

In the social and behavioral sciences, the aim of applied researchers is often to examine the relationships between latent variables. Latent variables are variables that are not directly observable, such as intelligence, skill or motivation. To measure these latent variables, observable indicators are used (Bollen & Hoyle, 2012). Structural Equation Modeling (SEM) can be used to simultaneously and consistently estimate both the measurement models and the structural relations between these latent variables (Jöreskog, 1973; Bentler & Chou, 1987). Despite the increasing popularity of SEM, many applied researchers prefer to use the Factor Score Regression (FSR) method, which is more intuitive and consists of two steps. In a first step, the scores on the latent variables are predicted using factor analysis (FA). In this paper, we will refer to these predicted scores as factor scores. In a second step, the factor scores are used in a linear regression (OLS) (Lu, Thomas, & Zumbo, 2005). Unfortunately, there are an infinite number of ways to compute these factor scores, all of which are consistent with the FA performed (Grice, 2001), meaning they are all equally viable. The two most commonly used predictors are the regression predictor (G. Thomson, 1934; Thurstone, 1935) and the Bartlett predictor (Bartlett, 1937; G. H. Thomson, 1938). The factor scores will be different depending on which predictor is used. This phenomenon is referred to as factor indeterminacy (Mulaik, 1972; Steiger, 1979; Maraun, 1996). The degree of indeterminacy is small if the relationship between the indicators and the latent variable is strong or if the number of indicators is high (Acito & Anderson, 1986). When there is a high degree of factor indeterminacy, it is even possible for an individual to score high according to the factor scores calculated by one method and score low according to the factor scores of another method (Grice, 2001). Lastovicka and Thamodaran (1991) conclude that indeterminacy implies that factor scores cannot be measured and researchers have to accept that factor analysis means it is impossible to obtain an unambiguous prediction or computation of the latent variable scores. In other words, a degree of uncertainty is inherent to factor scores, which is no longer accounted for when using factor scores in linear regression. This uncertainty causes the regression coefficient to be biased, as has been discussed extensively in the literature (e.g., Bollen,

1989; Lastovicka & Thamodaran, 1991; Shevlin, Miles, & Bunting, 1997; Lewis, 2005). Despite its obvious drawbacks, FSR remains a popular method among applied researchers (Lu & Thomas, 2008).

For this reason, improved methods to perform FSR have been developed, which result in an unbiased regression coefficient (Skrondal & Laake, 2001; Croon, 2002). Skrondal and Laake (2001) developed a method which avoids bias altogether, while Croon (2002) found a method to correct for the bias. We will refer to these methods as the bias avoiding and bias correcting method, respectively. Both methods indeed result in an unbiased parameter estimate, but are hardly ever used in practice. This is partly due to the highly technical and mathematical level of the papers describing the methods. Lu, Kwan, Thomas, and Cedzynski (2011) tried to remedy this by giving an overview of both methods. The statistical performance of both methods was also compared in a simulation study, with regard to accuracy and power. However, the methods remain unused by applied researchers. This could be due to some practical issues. First of all, it is only described how to obtain an unbiased estimate for the regression coefficient, but for the bias correcting method, there is no standard error available. This means that this method cannot be used yet to test hypotheses. Second, the results of the methods have only been described for the unstandardized parameterization. Neither Skrondal and Laake (2001), Croon (2002) or Lu et al. (2011) describe what happens when the standardized parameterization is used. In conclusion, the methods are not directly usable for the applied researchers.

The goal of this paper is to compare various methods for performing FSR, namely FSR using the regression predictor (regression FSR), FSR using the Bartlett predictor (Bartlett FSR), the method of Skrondal and Laake (2001) (bias avoiding method) and the method of Croon (2002) (bias correcting method). For each method we derive the bias analytically on the population level for both the standardized and unstandardized parameterization. Next, for the bias correcting method, a new standard error is developed, making it possible to use the bias correcting method to test hypotheses. Finally, two simulation studies are set up to compare the performance of the four methods in finite samples, using normal and non-normal data, respectively. The simulation studies

also allow us to evaluate the performance of the newly developed standard error. Since SEM is generally considered as the standard method to examine the regressions between latent variables, SEM is also included in the simulation studies. The aim is to be able to compare the methods on their overall statistical performance and formulate recommendations for the applied users.

Setting

To be able to compare the four methods, a simple regression model with one dependent and one independent variable is used. The simple regression model is used to reduce the notational complexity and enhance comprehensibility. However, it can easily be extended, to settings with more than one dependent variable and more than one independent variable. In fact, a more complex setting is used in the simulation study. Within this framework of a simple linear regression, we consider four possible scenarios, which are visualized in Figure ???. In the first scenario, both the dependent and independent variable are measured without error. They are considered observed variables. In the second scenario, factor scores are used for the independent variable, while the dependent variable is observed. In the third scenario, factor scores are used for the dependent variable, while the independent variable is observed. In the fourth and final scenario, factor scores are used for both variables. In all scenarios, the structural equation is:

$$\eta = \gamma \xi + \zeta, \quad (1)$$

where η is the dependent variable, ξ is the independent variable, γ is the regression coefficient and ζ is the residual error term. When η and ξ are latent, the following measurement models are used

$$\mathbf{x} = \Lambda_x \xi + \delta \quad (2)$$

$$\mathbf{y} = \Lambda_y \eta + \epsilon, \quad (3)$$

where $\mathbf{x} = (X_1, \dots, X_i, \dots, X_p)^T$ and $\mathbf{y} = (Y_1, \dots, Y_j, \dots, Y_q)^T$ are vectors of mean-centered observed indicators measuring ξ and η respectively, Λ_x and Λ_y are vectors of the factor loadings and δ and ϵ are the respective vectors of measurement error variables.

In the first step of factor score regression, we use these measurement models to perform a FA for each latent variable separately and to calculate the factor scores for ξ (F_ξ) and η (F_η). To be able to perform a FA, the metric scales of the latent variables ξ and η have to be fixed. This can be done in several ways, for example by fixing one factor loading per latent variable to 1 or by fixing the variance of the latent variable to 1. We will refer to the latter as the standardized parameterization, and the former as the unstandardized parameterization. In this paper, we will mainly use the unstandardized parameterization.

The factor scores are calculated by multiplying a factor score matrix \mathbf{A} with the observed indicators \mathbf{x} (or \mathbf{y}):

$$F_\eta = \mathbf{A}_\eta \mathbf{y} \quad (4)$$

$$F_\xi = \mathbf{A}_\xi \mathbf{x}. \quad (5)$$

The computation of the factor score matrices \mathbf{A}_η and \mathbf{A}_ξ depends on the method used for the prediction of the factor score. The different methods and their influence will be discussed in the next section.

In the second step of factor score regression, a linear regression is performed between the factor scores, resulting in a regression coefficient. In a simple linear regression, the true regression coefficient is defined as the true covariance between the dependent and the independent variable, divided by the variance of the independent variable:

$$\gamma = \frac{cov(\xi, \eta)}{var(\xi)}. \quad (6)$$

When performing the linear regression with factor scores, the regression coefficient becomes

$$\beta = \frac{cov(F_\xi, F_\eta)}{var(F_\xi)}, \quad (7)$$

which is not necessarily the same as the true regression coefficient. The relationship between γ and β can best be understood if we work out the covariance and variances of the factor scores. In Appendix A1, we derive the exact relationship between β and γ :

$$\beta = \frac{cov(F_\xi, F_\eta)}{var(F_\xi)} = \frac{\mathbf{A}_\xi \mathbf{\Lambda}_x cov(\xi, \eta) \mathbf{\Lambda}'_y \mathbf{A}'_\eta}{\mathbf{A}_\xi \mathbf{\Sigma}_x \mathbf{A}'_\xi} = \frac{\mathbf{A}_\xi \mathbf{\Lambda}_x var(\xi, \eta) \mathbf{\Lambda}'_y \mathbf{A}'_\eta}{\mathbf{A}_\xi \mathbf{\Sigma}_x \mathbf{A}'_\xi} \gamma. \quad (8)$$

From this, it is clear that in most cases the regression coefficient obtained with factor score regression will not be the same as the true regression coefficient. It is also possible to calculate the expected regression coefficients when factor scores are only used for one of the variables, dependent or independent. The interested reader can find the calculations in Appendices A2 and A3.

Methods to perform FSR

To perform FSR, several methods can be used, such as the regression FSR method, the Bartlett FSR method, the bias avoiding method (Skrondal & Laake, 2001) and the bias correcting method (Croon, 2002). In this section, these four methods will be discussed. In Table 1, an overview of this discussion is given.

Regression FSR method

The regression FSR method uses the regression predictor (G. Thomson, 1934; Thurstone, 1935) to compute the factor scores and then uses these factor scores in a linear regression. When using the regression predictor, the factor scoring matrices \mathbf{A}_ξ^R and \mathbf{A}_η^R are computed as follows:

$$\mathbf{A}_\xi^R = \Phi \mathbf{\Lambda}'_x \mathbf{\Sigma}_x^{-1} = var(\xi) \mathbf{\Lambda}'_x \mathbf{\Sigma}_x^{-1}, \quad (9)$$

and

$$\mathbf{A}_\eta^R = \text{var}(\eta)\Lambda_y'\Sigma_y^{-1}. \quad (10)$$

This means the formula for the variance of F_ξ can be simplified. In appendix B, this is done to show that the regression coefficient in factor score regression is not biased in all cases. When using the regression FSR method, there is only bias when factor scores are used for the dependent variable. With regard to bias, it is acceptable to use the regression FSR method if only the independent variables are factor scores. When factor scores are used for the dependent variable, one should not use regression FSR, since the regression parameter will be biased in most cases.

These results only apply when the unstandardized parameterization is used (Skrondal & Laake, 2001). The standardized regression coefficient γ_z is different from the unstandardized regression coefficient γ . The standardized regression coefficient can be calculated from the unstandardized regression coefficient and standard deviations:

$$\gamma_z = \frac{sd(\xi)}{sd(\eta)}\gamma. \quad (11)$$

Since $sd(F_\xi)$ and $sd(F_\eta)$ are biased, the standardized regression coefficient will be biased too if factor scores are used for any of the variables .

Bartlett FSR method

The Bartlett FSR method uses the Bartlett predictor (Bartlett, 1937; G. H. Thomson, 1938) to compute the factor scores and then uses these factor scores in a linear regression. The factor scoring matrices \mathbf{A}_ξ^B and \mathbf{A}_η^B are calculated as follows:

$$\mathbf{A}_\xi^B = (\Lambda_x'\Theta_\delta^{-1}\Lambda_x)^{-1}\Lambda_x'\Theta_\delta^{-1} \quad (12)$$

$$\mathbf{A}_\eta^B = (\Lambda_y'\Theta_\epsilon^{-1}\Lambda_y)^{-1}\Lambda_y'\Theta_\epsilon^{-1}, \quad (13)$$

with Θ_δ and Θ_ϵ the covariance matrices of respectively, δ and ϵ .

The Bartlett predictor is less known than the regression predictor, but has the advantage that $\mathbf{A}_\xi^B \boldsymbol{\Lambda}_x = (\boldsymbol{\Lambda}_x' \boldsymbol{\Theta}_\delta^{-1} \boldsymbol{\Lambda}_x)^{-1} \boldsymbol{\Lambda}_x' \boldsymbol{\Theta}_\delta^{-1} \boldsymbol{\Lambda}_x = 1$ and $\mathbf{A}_\eta^B \boldsymbol{\Lambda}_y = 1$. This implies that the formulas for the covariances can be simplified. The formula for the variance of ξ stays the same. Combined, this gives the regression coefficients as in Appendix C and Table 1. When using Bartlett FSR, there is no longer bias when factor scores are used for the dependent variable. However, now there is bias when factor scores are used for the independent variable. There is also still bias when factor scores are used for both variables.

Again, these results only apply for the unstandardized parameterization. Since the standard deviations are also biased using this method, the standardized regression coefficient will be biased if factor scores are used for any of the variables.

Bias avoiding method

The bias avoiding method was developed by Skrondal and Laake (2001). Based on the results discussed in the previous section, they concluded that one should simply use the regression predictor to predict the factor scores of the independent variable, while one should use the Bartlett predictor to predict the factor scores of the dependent variable. It has already been proven that this works for the unstandardized parameterization when factor scores are used for only one of the variables, but Skrondal and Laake (2001) showed that this also works when factor scores are used for both variables. In Appendix D, it is proven that this method results in unbiased estimates for all settings, but it has some drawbacks. First of all, one has to determine in advance if a variable will be dependent or independent. Moreover, a variable can only be dependent or independent. Mediation relationships are not possible. Second, this method only works when the unstandardized parameterization is used. When using the standardized parameterization, the regression coefficient estimate will still be biased.

Bias correcting method

The bias correcting method was developed by Croon (2002). In this method, the factor scores are computed using either the regression predictor or the Bartlett predictor. After

computing the factor scores, their variances and covariances are calculated. Next, these variances and covariances of the factor scores are used to compute the variances and covariances of the true latent variable scores. Finally, these estimates are used to calculate the regression coefficient. In Appendix E, it is shown how the covariance and variance of the true latent variable scores can be computed. Once these computations have been made, the regression coefficient can be computed as $\beta = \frac{\text{cov}(\xi, \eta)}{\text{var}(\xi)}$. Since these variances and covariances are unbiased, this results in an unbiased regression coefficient estimate (see Table 1). While this process is more complex than the bias avoiding method, it does have some advantages over it. First of all, it works for both the Bartlett and the regression predictor. Secondly, since the variances are no longer biased, the standard deviations necessary to calculate the standardized regression coefficient are also unbiased. This means that the method of Croon (2002) also results in an unbiased regression coefficient when the standardized parameterization is used.

Standard errors

Skrondal and Laake (2001), Croon (2002) and Lu et al. (2011) only describe how to calculate the regression coefficient, just as in the previous section. To be able to use the methods for hypothesis testing, it is necessary to have a complementary significance test, which requires a standard error and a theoretical distribution. For regression FSR, Bartlett FSR and the bias avoiding method, this is no problem. All three methods perform a regular linear regression after calculating the factor scores. This means that the significance test from the linear regression can be used. This test uses the following standard error:

$$SE = \sqrt{\frac{S^2}{\widehat{\text{var}}(F_{\xi})(n-1)}},$$

where S^2 is defined as:

$$S^2 = \widehat{\text{var}}(F_{\eta})(1-r^2)\frac{n-1}{n-(p+1)},$$

with $r = \gamma_z = \frac{sd(\xi)}{sd(\eta)}\gamma$, n is the sample size and p is the number of independent variables. A t-statistic is calculated by dividing the regression coefficient by its standard error. Finally, a p-value is calculated by comparing this t-statistic to a t-distribution with $n-(p+1)$ degrees of freedom. Note, that when performing hypothesis tests, we are dealing with finite samples. For this reason, all population parameters are replaced with their corresponding sample estimates.

For the bias correcting method, the standard error is not so easy to calculate. If one would use the above standard error, then this would be the standard error that coincides with the original, uncorrected regression coefficient. In this case, using the corrected regression coefficient to calculate the t-value would result in an incorrect t-value (and p-value). On the other hand, using the uncorrected regression coefficient would just result in a significance test for the uncorrected regression coefficient and is again not adequate.

Another alternative would be to use the above formula for the standard error, but replace all variances and standard deviations with their corrected versions (Croon, 2002). Unfortunately, this approach implicitly assumes that the true latent scores are directly observable, resulting in an underestimation of the standard error. This suggests that the standard error consists of multiple parts, namely error resulting from the regression itself and error resulting from the factor scores. One way to calculate the error resulting from the factor scores, is to first calculate the prediction error in the factor scores. The prediction error in F_ξ and F_η will be denoted as var_{ε_x} and var_{ε_y} respectively. These prediction errors can be calculated for both the regression and Bartlett predictor. Here, we will only discuss the regression predictor. Skrondal and Rabe-Hesketh (2004, eq. 7.7) showed that the prediction errors can be calculated as follows when using the regression predictor:

$$\begin{aligned} var_{\varepsilon_x} &= \Phi - \Phi' \Lambda' \Sigma_x^{-1} \Lambda \Phi \\ &= var(\xi) - var(\eta). \end{aligned}$$

The last equation is derived from Equation 28 in Appendix B. The same derivations can be made

for the dependent variable η :

$$var_{\varepsilon_y} = var(\eta) - var(F_\eta).$$

The prediction error in the factor scores is thus the difference between the observed and the corrected variance. Since this is again a variance, the formula from the regular linear regression can be used to calculate the corresponding S^2 . From this, it can be derived that the total S^2 consists of three parts:

1. $S_{reg}^2 = var(\eta)(1 - r^2)\frac{n-1}{n-(p+1)}$
2. $S_y^2 = var_{\varepsilon_y}(1 - r^2)\frac{n-1}{n-(p+1)}$
3. $S_x^2 = var_{\varepsilon_x}r^2\frac{n-1}{n-(p+1)}$.

Note that for S_x^2 , r^2 is used instead of $1 - r^2$. This is because the prediction error in the independent variable has more influence on the standard error, as the relation between the variables increases. When the independent variable has no influence on the dependent variable, it also has no influence on the standard error. Now, the total S^2 can be calculated by summing up the three parts:

$$S_{total}^2 = S_{reg}^2 + S_y^2 + S_x^2. \tag{14}$$

Finally, a new adjusted standard error can be calculated as:

$$SE = \sqrt{\frac{S_{total}^2}{var(F_\xi)(n - 1)}}. \tag{15}$$

Using this approximate standard error and the corrected regression coefficient, an approximate t-statistic can be obtained and compared to the theoretical t-distribution with $n-(p+1)$ degrees of freedom. Using this newly developed standard error, the bias correcting method can now be used to perform hypothesis tests about the regression coefficient γ .

Simulation studies

Two simulation studies are conducted to examine the finite sample performance of these methods and significance tests. In a first study, the methods are studied using item responses that are normally distributed. In the second study, non-normal item responses are used. The results of these studies can be used as guidelines to determine which method to use, depending on the data and research questions.

First, an outline of how the data was simulated is given, followed by a description of the analyses performed on the simulated data and the results of both studies.

Data simulation

Study 1. Before the simulation of the data, a ground truth or population model is defined. The structural model consists of a multivariate regression between three latent independent variables, ξ_1 , ξ_2 and ξ_3 , and two latent dependent variables η_1 and η_2 , resulting in the structural equation:

$$\eta_1 = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_3 + \zeta_1, \quad (16)$$

$$\eta_2 = \gamma_3\xi_1 + \gamma_1\xi_2 + \gamma_2\xi_3 + \zeta_2. \quad (17)$$

The model is depicted in Figure 2.

The simulation of the data consists of two steps, which are carried out using R (R Development Core Team, 2008). In the first step, the true latent variable values of ξ_1 , ξ_2 , ξ_3 , η_1 and η_2 are generated. The variance of the independent variables ξ_1 , ξ_2 and ξ_3 is set at 100, while the residual variance of the dependent variables η_1 and η_2 is set at 400. The covariances between all latent variables are 0. To generate data that comply with these parameters, the true latent scores of ξ_1 , ξ_2 and ξ_3 are first generated, followed by the regression residuals ζ_1 and ζ_2 . Finally, using the structural equations $\eta_1 = \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_3 + \zeta_1$ and $\eta_2 = \gamma_3\xi_1 + \gamma_1\xi_2 + \gamma_2\xi_3 + \zeta_2$, the true latent scores on η_1 and η_2 are generated.

In the second step, data is generated for each observed item response x_{ij} and y_{ij} , using the true latent scores and the measurement models, with i referring to the latent variables and j to the items. The measurement models of the latent variables are $y_{ij} = \lambda_{y_{ij}}\eta_i + \epsilon_{ij}$ and $x_{ij} = \lambda_{x_{ij}}\xi_i + \delta_{ij}$. All $\lambda_{x_{ij}}$ and $\lambda_{y_{ij}}$ are set at 1. The residual variances $\Theta_{\epsilon_{ij}}$ of all y_{ij} are set at $\Theta_{\epsilon_{ij}} = \frac{\text{var}(\eta_i)(1-CD_{y_i})}{CD_{y_i}}$ and the residual variances $\Theta_{\delta_{ij}}$ of all x_{ij} are set at $\Theta_{\delta_{ij}} = \frac{\text{var}(\xi_i)(1-CD_{x_i})}{CD_{x_i}}$, with CD_{y_i} and CD_{x_i} the respective coefficients of determination for the measurement models. All CD_{x_i} and CD_{y_i} are equal and will thus be referred to as CD . To create item responses that are normally distributed (skewness = 0, kurtosis = 3), ϵ_{ij} and δ_{ij} are generated from a univariate normal distribution.

The coefficients of determination CD , the regression coefficient γ_i , the sample size n and the number of items I for the two latent variables are varied to create 216 experimental conditions. The sample size is varied to be able to determine the consistency of the methods, while the coefficient of determination is varied to account for the degree of factor score indeterminacy. A higher CD implies a lower degree of factor score indeterminacy (Acito & Anderson, 1986). The regression coefficient is varied to be able to determine the power and type I-error rate. The values used for these parameters can be found in Table 2.

Study 2. In the second simulation study, the aim is to compare the methods when the observed item responses are not normally distributed. This means the simulation of the data is done in the same way as in study 1, except for the generation of the ϵ_{ij} and δ_{ij} . To create non-normal item responses, the generation of ϵ_{ij} and δ_{ij} is done in two different ways. To create item responses that are symmetrical, but have a large kurtosis, ϵ_{ij} and δ_{ij} are generated from a t-distribution with 3 degrees of freedom, multiplied by $\sqrt{\Theta_{\epsilon_{ij}}}$ or $\sqrt{\Theta_{\delta_{ij}}}$, respectively. On average, this results in data with an almost-zero skewness (=0.001) and a kurtosis of about 10.057. To create item responses that are skewed and have a non-normal kurtosis, ϵ_{ij} and δ_{ij} are generated from a χ^2 -distributions with 1 degree of freedom, multiplied by $\sqrt{\Theta_{\epsilon_{ij}}}$ or $\sqrt{\Theta_{\delta_{ij}}}$, respectively. On average, this results in data with a skewness of 0.693 and a kurtosis of 4.785. Note that the way the data are simulated implies that the degree of non-normality reduces as the coefficient of determination increases. This is due to the fact that ϵ_{ij} and δ_{ij} have less influence when the factor

loadings become stronger.

Analyses

The analysis performed on the data is the same in both studies. Five methods of analyses are performed on the simulated data, namely regression FSR, Bartlett FSR, the bias avoiding method (Skrondal & Laake, 2001), the bias correcting method (Croon, 2002) and SEM.

For the SEM analysis, a correctly specified SEM-model is constructed, meaning that it corresponds with the population model used to generate the data. This model is estimated using the simulated dataset, with a "maximum likelihood" estimator. From the results, the regression coefficient $\hat{\gamma}$, and its standard error SE and p-value \hat{p} are obtained, as well as the standardized regression coefficient $\hat{\gamma}_z$. This is done for a 1000 simulated datasets for every simulation condition. Based on all 1000 replications, several performance criteria are calculated, namely the mean regression coefficient estimations $\bar{\gamma}$ and $\bar{\gamma}_z$, the bias using the unstandardized and the bias using standardized parameterization, the empirical standard deviation, the mean square error, the mean standard error and the standard error bias for the regression coefficient and the power of the statistical test. For the conditions with a regression coefficient of 0, the type I-error rate is calculated instead of the power.

For the four FSR methods, a FA is performed for all latent variables. Hereafter, factor scores are calculated, using the Bartlett and the regression predictor. Next, the regression factor scores are used in two linear regressions for the regression FSR and the Bartlett factor scores are used for the Bartlett FSR. For the bias avoiding method, two linear regressions are performed, using the Bartlett factor scores for the dependent variables and the regression factor scores for the independent variables. Finally, for the bias correcting method, the regression coefficient and standard errors are calculated using the formulas described above. Again, the regression coefficients $\hat{\gamma}$ and $\hat{\gamma}_z$ and the standard error SE and p-value \hat{p} for $\hat{\gamma}$ were retained. This was repeated 1000 times and the same performance criteria were calculated. The criteria are summarized in Table 3.

Analysis of the results

Study 1. To compare the five methods with regard to the bias using the unstandardized and standardized parameterization, efficiency, MSE, standard error bias, type I-error rate and power, an ANOVA was performed for each of these performance measures. The independent variables were the design factors, namely the sample size, coefficient of determination, method, number of items and the value of gamma. All possible two-way interactions were also included in the analyses. This resulted in 15 predictors for each model. The results of this ANOVAs can be found in Table 4. Due to space constraints, only the informative effects are discussed in the results section.

Study 2. The same ANOVAs were performed when the data were not normally distributed. However, one extra independent variable was added, namely the degree of non-normality. This resulted in 21 predictors. The results can be found in Table 6.

Results Study 1

The results are discussed per statistical performance criterion. The proportion of successful replications, the bias and efficiency results are discussed first, followed by the MSE, standard error bias, type I-error rate and power. In Table 5, a comparison between the methods for all performance criteria is given.

Proportion of successful replications. The first performance measure that was considered was the proportion of successful replications for each method. The proportion of successful replications is very high for all methods and conditions, namely 0.989 or higher. The proportion of successful replications of SEM is even higher than the proportions of the other four methods, which all have the same proportions.

Bias. As can be seen in Figure 3a, only the regression FSR and the Bartlett FSR method are biased, while the three remaining methods are not. More specifically, both methods underestimate the regression coefficient and the Bartlett FSR method is more severely biased than the regression FSR method. Moreover, the bias is not influenced by the sample size, implying that

both the regression and Bartlett FSR methods are also inconsistent. The bias does disappear with an increasing factor loading and when γ is equal to 0. When the number of items goes up, the bias also declines, but has not disappeared completely when the number of items reaches 10. The three other methods, namely the bias avoiding, bias correcting and SEM method, exhibit, as expected, very little bias. The three methods perform very similar.

Bias using standardized parameterization. The patterns change slightly when the standardized parameterization is used (see Figure 4a). Now, the bias avoiding method also underestimates the regression coefficient and is inconsistent. There is also no longer a difference between the Bartlett FSR, regression FSR and bias avoiding methods. All other effects can be interpreted in the same way as in the unstandardized parameterization.

Efficiency. While regression FSR and Bartlett FSR are the most biased methods, they are also the most efficient methods. The Bartlett FSR method is even slightly more efficient than the regression FSR method. The three other methods have very similar standard errors. However, the bias avoiding method is slightly more efficient than the other two when the coefficient of determination is low. It is also important to note that the differences between the methods disappear as the coefficient of determination increases.

Mean Square Error. The regression and Bartlett FSR methods have a high MSE, as compared to the other three methods, with the Bartlett FSR having the worst MSE. The MSE of the other three methods is very similar to each other. The differences between the five methods disappear when the CD approaches 1 or when the value of γ approaches 0.

Standard error bias. With regard to the standard error bias, there is a large difference between the methods. In Figure 5a, it can be seen that all five methods show bias in the standard error. However, the SEM method clearly has the smallest standard error bias. The four alternative methods are more severely biased with regard to the standard error. On average, the Bartlett FSR method gives the second best estimation of the standard error, followed by the bias correcting and regression FSR methods.

Over all methods, a higher sample size, a higher CD or more items, lowers the standard

error bias, while a larger value of γ increases the standard error bias. When the CD approaches 1, all differences between the methods disappear and when the value of γ increases, the differences between the methods also increases. It is important to note that when there is no effect ($\gamma = 0$), the standard error bias almost completely disappears for all five methods. This means the standard error bias will have little effect on the type I-error rate.

Type I-error. The type I-error rates for the regression FSR, Bartlett FSR and bias avoiding method are very similar, while SEM has a slightly higher type I-error rate and the bias correcting method has a slightly lower type I-error rate. While these differences are statistically significant, for all five methods the type-I error rates are around the expected value of 0.05 . In practice, this means these differences are not really relevant.

Power. All five methods have approximately the same power. There are only small differences when the coefficient of determination is very weak. In that case, SEM has the lowest power, followed by the bias correcting method. The other methods have the same power.

Results Study 2

The results of the second study are described in relation to the first study. The impact of the kind of non-normality is also discussed.

Proportion of successful replications. When the data are not normally distributed, the proportion of successful replications is a lot lower and the differences between the methods are larger. Especially the proportion of successful replications of SEM is very low as compared to the other four methods. This is mostly due to the conditions where the coefficient of determination is very low (CD=0.3) or the number of items is low (I=3). The proportion of the bias correcting method is the second lowest, followed by the bias avoiding method. The regression FSR method has the highest proportion of successful replications. All these effects are largely due to the conditions where a t-distribution was used to simulate the item responses. The χ^2 -distribution seems to have less effect. This could be due to the fact that the kurtosis is larger when using the t-distribution than when using the χ^2 -distribution.

Bias. With regard to the bias, there are two main shifts in the patterns (Figure 3b). Firstly, now, there is a difference between SEM, the bias avoiding and bias correcting method. SEM still shows almost no bias, while the bias correcting and bias avoiding method do show a little bias. The bias avoiding method is more biased than the bias correcting method. The regression FSR and Bartlett FSR still show a large amount of bias. Secondly, the sample size does have an influence for SEM, the bias avoiding and bias correcting method. Because of the larger bias in these methods, it can now be seen that these methods are consistent. Again, these patterns are largely caused by the conditions using the t-distribution.

Bias using standardized parameterization. The patterns with regard to the bias using standardized parameterization change in two ways (see Figure 4b). Firstly, both SEM and the bias correcting method show a small bias in the standardized regression parameter, with SEM having the smallest bias. Secondly, the sample size does have an influence on the bias of the bias correcting and SEM method when the data is not normally distributed. When the sample size increases, the bias decreases.

Efficiency. The efficiency is a lot lower when the data are not normal, especially when the t-distribution was used to simulate the data.

Mean Square Error. While the MSE of SEM, bias correcting and bias avoiding method were very similar to each other when the data was normally distributed, now there are small differences. SEM has the smallest MSE, closely followed by the bias correcting method. This is again due to the high kurtosis of the conditions simulated with the t-distribution. All other effects remain the same.

Standard error bias. When the data are not normally distributed, the standard error bias of the bias correcting method is the worst in some conditions(see Figure 5b), especially when the CD or the value of γ is very low.

Type I-error. When the data are not normally distributed, especially when there is a high kurtosis, the type I-error rate is larger than the expected value of 0.05 for all methods. Especially the bias correcting method seems to have a large type I-error rate, when the coefficient of

determination is very low, namely 0.30.

Power. As can be expected, the power seems to be lower when the data are not normally distributed. Table 6 shows that the method does not have a significant influence on the power, when the data are not normally distributed.

Discussion

In this paper, an overview was given of four methods for factor score regression, namely regression FSR, Bartlett FSR, the bias avoiding method (Skron dal & Laake, 2001) and the bias correcting method (Croon, 2002). The four methods were described and their statistical properties were discussed on the population level. Since there was no adequate standard error available for the bias correcting method, a new standard error was developed. To be able to determine the statistical properties of the methods in finite samples and to evaluate the performance of the newly developed standard error, two Monte Carlo simulation studies were performed.

The simulation studies showed that the regression FSR method does not perform well. This confirms the general expectations found in the literature (e.g., Bollen, 1989; Lastovicka & Thamodaran, 1991; Shevlin et al., 1997; Skron dal & Laake, 2001; Croon, 2002; Lewis, 2005). It also complies with the results of Lu et al. (2011). The method is biased for both the standardized and unstandardized parameterization and is inconsistent. The method does have a high efficiency, but it also has a high MSE. This means that the high efficiency cannot compensate for the high bias. Moreover, the estimates of the standard error have the second highest bias observed.

The Bartlett FSR method performs even worse than the regression FSR method, with a comparable bias using the standardized parameterization, but a higher unstandardized bias and MSE. It does have a lower standard error bias. It is also the most efficient method, but at the same time it has the highest bias and MSE of all methods.

The first corrected method, namely the bias avoiding method of Skron dal and Laake (2001), only performs slightly better than the regression FSR and Bartlett FSR methods. It is unbiased when using the unstandardized parameterization, but it is still biased when using the standardized

parameterization. This result highlights the fact that standardized and unstandardized regression coefficients do not always behave in the same way (Kim & Mueller, 1976). The standard error bias is the largest of all methods and it has the same power and type I-error rate as regression FSR and Bartlett FSR. It can be concluded that this method only outperforms regression FSR and Bartlett FSR with regard to the unstandardized bias and is outperformed by SEM and the bias correcting method.

The second corrected method, the bias correcting method, performs better than the bias avoiding method. It is unbiased for both the standardized and unstandardized parameterization and has the highest power. When the observed item responses are not normally distributed, it does show a slightly larger bias than the SEM method, but its proportion of successful replications is also much higher. With regard to the standard error bias, only SEM and Bartlett FSR do better. This result shows that the newly developed standard error is reliable and even performs better than the regular standard error used in the bias avoiding method. However, it is important to note that when the data are not normally distributed and the factor loadings are very weak ($CD=0.3$), the standard error bias goes up, resulting in a higher type I-error rate. In all other conditions, the estimate of the standard error is reliable. The method does have the lowest efficiency, but also has the lowest MSE, meaning that the low efficiency does not have much influence.

The SEM method performs very similar to the bias correcting method. When the data are normally distributed, SEM has the same bias when using the unstandardized and standardized parameterization, efficiency, MSE, power and type I-error rate. When the data are not normally distributed, SEM has a lower bias and MSE, but also has a very low proportion of successful replications. This means that, although SEM gives less biased regression coefficients, the chance that the model will not converge, is also a much larger for SEM. On the other hand, there is almost no standard error bias when using the SEM method.

Overall, it can be concluded that only the bias correcting method is a suitable alternative for SEM. The method performs similar to the SEM method with regard to bias, efficiency, MSE,

power and type I-error rate. It does have more standard error bias than SEM, but it has the second lowest standard error bias of the four FSR methods. It also has more successful replications than SEM when the data are not normally distributed. The method does have some drawbacks in comparison to SEM, because of its two-step nature. First, at this moment, there are no overall fit indices available. Second, the method cannot be used for all possible structural models. For example, the method cannot handle mediational relationships and is not applicable to non-recursive methods. However, it is the intention of the authors to extend the method to be applicable to the full SEM-model and to develop a set of fit indices for this method.

A second conclusion that can be made is that factor score indeterminacy plays a great role in deciding which method to use. The simulation study showed that if the factor score indeterminacy is low (i.e. a coefficient of determination of 0.99) the differences between the methods disappear completely, on all performance criteria. As a result, there is no longer a problem with performing a conventional factor score regression. This implies that it is important to first determine the factor score indeterminacies by use of indeterminacy indices. If these indices suggest that the indeterminacies are very low, any of the methods can be used. If these indices suggest the indeterminacies are moderate or high, one should use SEM or the bias correcting method.

In this paper it was shown that the bias correcting method of Croon (2002) is a reliable and unbiased method to perform a factor score regression. A new and reliable standard error was also developed, meaning that the bias correcting method can now be used by applied users to perform significance tests. However, performing the method is a rather complex and technical process. For this reason, software to perform the method will be developed and made available for the applied users in the near future.

References

- Acito, F., & Anderson, R. D. (1986). A simulation study of factor score indeterminacy. *Journal of Marketing Research*, 23(2), 111–118. doi: 10.2307/3151658
- Bartlett, M. S. (1937, July). The statistical conception of mental factors. *British Journal of Psychology. General Section*, 28(1), 97–104. doi: 10.1111/j.2044-8295.1937.tb00863.x
- Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, 16, 78–117. doi: 10.1177/0049124187016001004
- Bollen, K. (1989). *Structural equations with latent variables*. New York: NY: Wiley.
- Bollen, K., & Hoyle, R. H. (2012). Latent variables in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 59–67). the Guilford Press.
- Croon, M. (2002). Using predicted latent scores in general latent structure models. In G. Marcoulides & I. Moustaki (Eds.), *Latent variable and latent structure modeling* (pp. 195–223). Mahwah: Lawrence Erlbaum Associates, Inc.
- Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, 6(4), 430–450. doi: 10.1037//1082-989X.6.4.430
- Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. Goldberger & O. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85–112). New York: Seminar Press. doi: 10.1002/j.2333-8504.1970.tb00783.x
- Kim, J.-O., & Mueller, C. W. (1976). Standardized and unstandardized coefficients in causal analysis: An expository note. *Sociological Methods & Research*, 4(4), 423–438. doi: 10.1177/004912417600400402
- Lastovicka, J. L., & Thamodaran, K. (1991). Common factor score estimates in multiple regression problems. *Journal of Marketing Research*, 28(1), 105–112. doi: 10.2307/3172730
- Lewis, J. B. (2005, July). Estimating regression models in which the dependent variable is based on estimates. *Political Analysis*, 13(4), 345–364. doi: 10.1093/pan/mpi026

- Lu, I. R., Kwan, E., Thomas, D. R., & Cedzynski, M. (2011, September). Two new methods for estimating structural equation models: An illustration and a comparison with two established methods. *International Journal of Research in Marketing*, 28(3), 258–268. doi: 10.1016/j.ijresmar.2011.03.006
- Lu, I. R., & Thomas, D. R. (2008, July). Avoiding and correcting bias in score-based latent variable regression with discrete manifest items. *Structural Equation Modeling: A Multidisciplinary Journal*, 15(3), 462–490. doi: 10.1080/10705510802154323
- Lu, I. R., Thomas, D. R., & Zumbo, B. D. (2005). Embedding IRT in structural equation models : A comparison with regression based on IRT scores. *Structural equation modeling: A multidisciplinary Journal*, 12:2(January 2014), 263–277. doi: 10.1207/s15328007sem1202
- Maraun, M. D. (1996). Metaphor taken as math: Indeterminacy in the factor analysis model. *Multivariate Behavioral Research*, 31(4), 517–538. doi: 10.1207/s15327906mbr3104_6
- Mulaik, S. A. (1972). Factor scores and factor indeterminacy. In S. A. Mulaik (Ed.), *Foundations of factor analysis*. New York: McGraw-Hill Book Company.
- R Development Core Team. (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical computing.
- Shevlin, M., Miles, J., & Bunting, B. (1997, October). Summated rating scales. A Monte Carlo investigation of the effects of reliability and collinearity in regression models. *Personality and Individual Differences*, 23(4), 665–676. doi: 10.1016/S0191-8869(97)00088-3
- Skrondal, A., & Laake, P. (2001). Regression among factor scores. *Psychometrika*, 66(4), 563–575. doi: 10.1007/BF02296196
- Steiger, J. H. (1979). Factor indeterminacy in the 1930's and the 1970's some interesting parallels. *Psychometrika*, 44(2), 157–167. doi: 10.1007/BF02293967
- Thomson, G. (1934). The meaning of i in the estimate of g . *British Journal of Psychology*, 25, 92–99. doi: 10.1111/j.2044-8295.1934.tb00728.x
- Thomson, G. H. (1938, February). Methods of Estimating Mental Factors. *Nature*, 141(3562),

246–246. doi: 10.1038/141246a0

Thurstone, L. (1935). *The vectors of mind*. Chicago, IL: University of Chicago Press.

Table 1
The regression parameter β in relationship to γ when using FSR

$\beta =$	both variables observed	independent variable factor scores, dependent variable observed	independent variable observed, dependent variable factor scores	both variables factor scores
The general case	γ	$\frac{\mathbf{A}_x \Lambda_x \text{var}(\xi)}{\mathbf{A}_x \Sigma_x \mathbf{A}_x'} \gamma$	$\Lambda_y' \mathbf{A}_y' \gamma$	$\frac{\mathbf{A}_x \Lambda_x \text{var}(\xi) \Lambda_y' \mathbf{A}_y'}{\mathbf{A}_x \Sigma_x \mathbf{A}_x'} \gamma$
Regression FSR	γ	γ	$\mathbf{A}_\eta^R \Lambda_y \gamma$	$\mathbf{A}_\eta^R \Lambda_y \gamma$
Bartlett FSR	γ	$\frac{\text{var}(\xi)}{\Lambda_\xi^B \Sigma_x \Lambda_\xi^{B'}} \gamma$	γ	$\frac{\text{var}(\xi)}{\Lambda_\xi^B \Sigma_x \Lambda_\xi^{B'}} \gamma$
Bias avoiding method	γ	γ	γ	γ
Bias correcting method	γ	γ	γ	γ

Note: γ is the true population regression parameter, β is the regression parameter that is obtained when using factor score regression. When $\beta = \gamma$, there is no bias. If it is not equal, there is bias.

Table 2

Summary of the model parameters used in simulation

Model parameter	value
Regression coefficients γ_1	0
γ_2	1.5
γ_3	0.41
coefficient of determination CD	0.3, 0.6 0.7, 0.9 0.95, 0.99
The sample size n	300, 500, 800, 2000
The number of items I	3, 5, 10

Table 3

Summary of the performance criteria, with R the number of successful replications.

Criteria	Formula
$\bar{\gamma}$	$\frac{1}{R} \sum_{i=1}^R \hat{\gamma}_i$
$\bar{\gamma}_z$	$\frac{1}{R} \sum_{i=1}^R \hat{\gamma}_{z_i}$
Bias	$\frac{1}{R} \sum_{i=1}^R (\hat{\gamma}_i - \gamma)$
Relative bias	$\frac{\text{Bias}}{\gamma}$
Empirical standard deviation (ESD)	$\sqrt{\frac{1}{R-1} \sum_{i=1}^R (\hat{\gamma}_i - \bar{\gamma})^2}$
Mean square error (MSE)	$\frac{1}{R} \sum_{i=1}^R (\hat{\gamma}_i - \gamma)^2$
Mean standard error (MSTE)	$\frac{1}{R} \sum_{i=1}^R \text{SE}$
Standard error bias (SEB)	$\frac{1}{R} \sum_{i=1}^R (\text{MSTE} - \text{ESD})$
Power/Type I-error	$\frac{1}{R} \sum_{i=1}^R (\hat{p}_i < 0.05)$

Table 4

ANOVA-models, using normally distributed data (*** < 0.000, **0.000-0.001, *0.001-0.01, +0.01-0.05).

Effect	Df	Successful replications	Bias	Standardized bias	Efficiency
I	2	1253.97 ***	43.37 ***	105.00 ***	271.22 ***
γ	2	0.00	264.20 ***	722.37 ***	1192.35 ***
CD	5	839.31 ***	154.10 ***	412.64 ***	3131.66 ***
Method	4	48.24 ***	265.33 ***	230.36 ***	310.78 ***
n	3	1.41	0.16	0.18	15202.77 ***
Method*CD	20	43.35 ***	74.21 ***	67.21 ***	127.58 ***
Method* γ	8	0.00	131.22 ***	113.74 ***	3.92 **
Method*I	8	40.24 ***	17.73 ***	16.52 ***	32.96 ***
Method*n	12	2.22 *	0.02	0.012	12.15 ***
CD* γ	10	0.00	74.66 ***	202.22 ***	389.46 ***
CD*I	10	800.24 ***	10.60 ***	26.21 ***	204.36 ***
CD*n	15	7.21 ***	0.03	0.04	181.63 ***
γ *I	4	0.00	20.55 ***	50.20 ***	0.88
γ *n	6	0.00	0.70	0.33	50.27 ***
I*n	6	3.11 *	0.08	0.08	70.98 ***

Effect	Df	MSE	Standard error bias	Type I- error rate	Power
I	2	38.98 ***	14.13 ***	47.11 ***	55.84 ***
γ	2	144.13 ***	1122.45 ***	/	921.72 ***
CD	5	164.28 ***	531.54 ***	2.92 +	177.54 ***
Method	4	40.55 ***	62.15 ***	9.33 ***	0.22
n	3	65.76 ***	79.20 ***	38.13 ***	429.60 ***
Method*CD	20	25.57 ***	32.43 ***	1.42	0.17
Method* γ	8	44.50 ***	65.25 ***	/	0.22
Method*I	8	8.14 ***	1.64	0.16	0.05
Method*n	12	0.37	1.35	0.22	0.04
CD* γ	10	83.48 ***	355.61 ***	/	177.66 ***
CD*I	10	23.48 ***	12.22 ***	15.40 ***	24.47 ***
CD*n	15	3.33 ***	28.90 ***	3.41 ***	48.33 ***
γ *I	4	18.90 ***	0.97	/	56.01 ***
γ *n	6	1.01	50.96 ***	/	429.17 ***
I*n	6	0.65	0.99	5.89 ***	13.17 ***
Residuals	2044			634	1339

Table 5

Comparison between the five methods per performance criteria

	regression FSR	Bartlett FSR	Bias avoiding	Bias correcting	SEM
Number of Successful replications	2	2	2	2	1
Bias	4	5	1	1	1
Standardized bias	3	3	3	1	1
Efficiency	2	1	3	4	4
MSE	4	5	1	1	1
Standard error bias	4	2	5	3	1
Type I-error	1	1	1	1	1
Power	1	1	1	1	1

Note: The numbers indicate the performance of the methods in relation to each other for each performance criteria. A score of 1 means the method performed best on that particular performance criteria, while a score of 5 means the methods performance the worst.

Table 6

ANOVA-models, using non-normal data (*** < 0.000, **0.000-0.001, *0.001-0.01, +0.01-0.05).

Effect	Df	Successful replications	Bias	Standardized bias	Efficiency
I	2	124.49 ***	168.59 ***	270.76 ***	240.69 ***
γ	2	0.00	1215.43 ***	2449.80 ***	2974.83 ***
CD	5	1219.83 ***	686.33 ***	1266.31 ***	6770.51 ***
Method	4	183.97 ***	528.00 ***	525.92 ***	190.16 ***
n	3	79.03 ***	4.31 *	4.24 *	2657.27 ***
Non-normality (NN)	1	2100.28 ***	299.53 ***	342.32 ***	8044.76 ***
Method*CD	20	121.49 ***	110.43 ***	122.62 ***	15.08 ***
Method* γ	8	0.00	262.68 ***	264.49 ***	3.55 **
Method*I	8	28.19 ***	22.15 ***	25.99 ***	3.15 *
Method*n	12	1.83 +	0.55	0.20	0.49
Method*NN	4	153.36 ***	5.82 **	7.09 ***	3.15 *
CD* γ	10	0.00	340.41 ***	615.95 ***	644.05 ***
CD*I	10	36.24 ***	33.25 ***	45.76 ***	32.18 ***
CD*n	15	36.33 ***	2.64 **	1.77	62.45 ***
CD*NN	5	1087.06 ***	111.44 ***	92.32 ***	1874.78 ***
γ *I	4	0.00	77.43 ***	126.66 ***	4.54 *
γ *n	6	0.00	1.88	1.33	2.57 +
γ *NN	2	0.00	152.23 ***	165.91 ***	1060.31 ***
I*n	6	3.39 *	1.07	0.42	4.04 **
I*NN	2	61.01 ***	22.12 ***	14.55 ***	8.72 **
n*NN	3	54.25 ***	7.35 ***	2.89 +	2.53
Effect	Df	MSE	Standard error bias	Type I-error rate	Power
I	2	96.50 ***	4.57 +	3.83 +	321.30 ***
γ	2	824.44 ***	2473.21 ***	/	3549.41 ***
CD	5	935.59 ***	2733.13 ***	378.18 ***	1307.85 ***
Method	4	48.17 ***	95.91 ***	35.50 ***	0.72
n	3	70.00 ***	4.67 *	31.634 ***	810.00 ***
Non-normality (NN)	1	657.45 ***	6300.45 ***	1804.36 ***	365.06 ***
Method*CD	20	20.40 ***	35.43 ***	31.72 ***	1.49
Method* γ	8	64.50 ***	70.14 ***	/	0.99
Method*I	8	3.88 **	2.19 +	1.27	0.57
Method*n	12	0.48	0.68	1.21	0.05
Method*NN	4	5.28 **	22.85 ***	33.79 ***	2.11
CD* γ	10	408.33 ***	524.64 ***	/	926.40 ***
CD*I	10	38.39 ***	0.74	15.70 ***	111.62 ***
CD*n	15	16.93 ***	1.48	17.76 ***	114.89 ***
CD*NN	5	308.80 ***	1449.50 ***	340.35 ***	126.93
γ *I	4	55.55 ***	23.48 ***	/	198.07 ***
γ *n	6	4.79	9.81 ***	/	662.82 ***
γ *NN	2	265.38 ***	1037.45 ***	/	158.74 ***
I*n	6	1.13	0.64	1.36	22.92 ***
I*NN	2	3.88 **	8.01 **	0.31	26.26 ***
n*NN	3	11.56 ***	0.71	74.15 ***	13.08
Residuals	4187			1339	2763

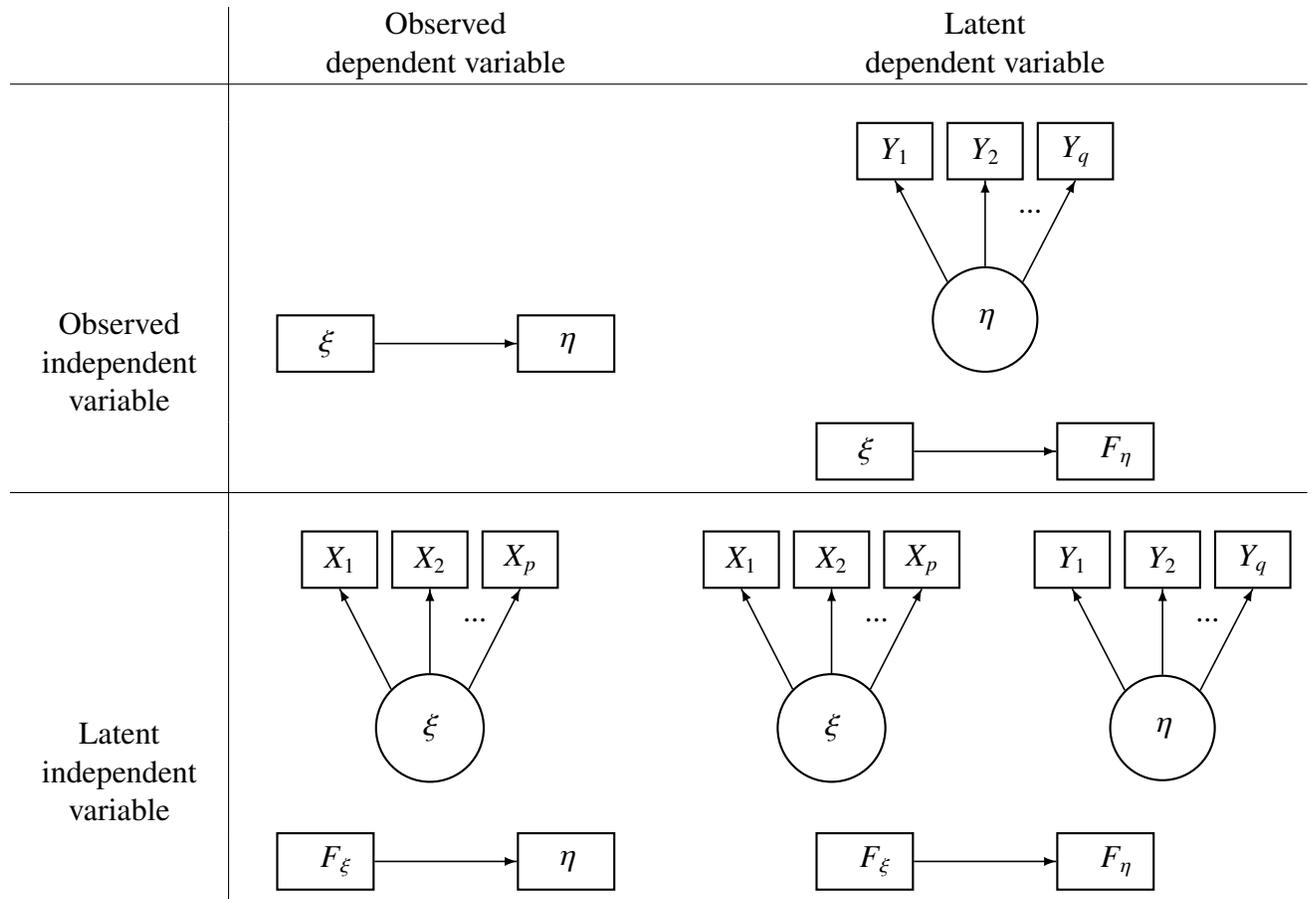


Figure 1. The four scenarios considered

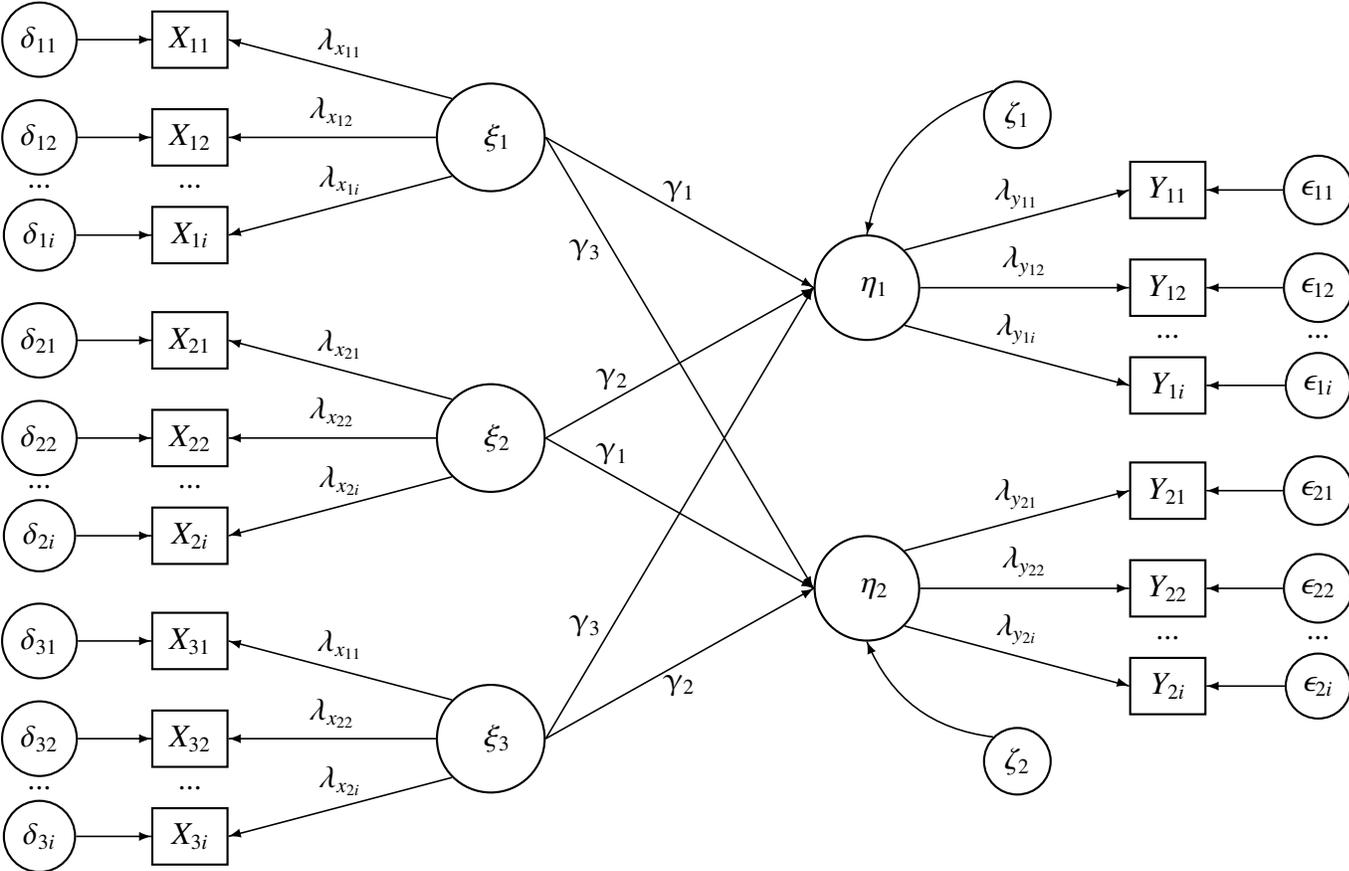
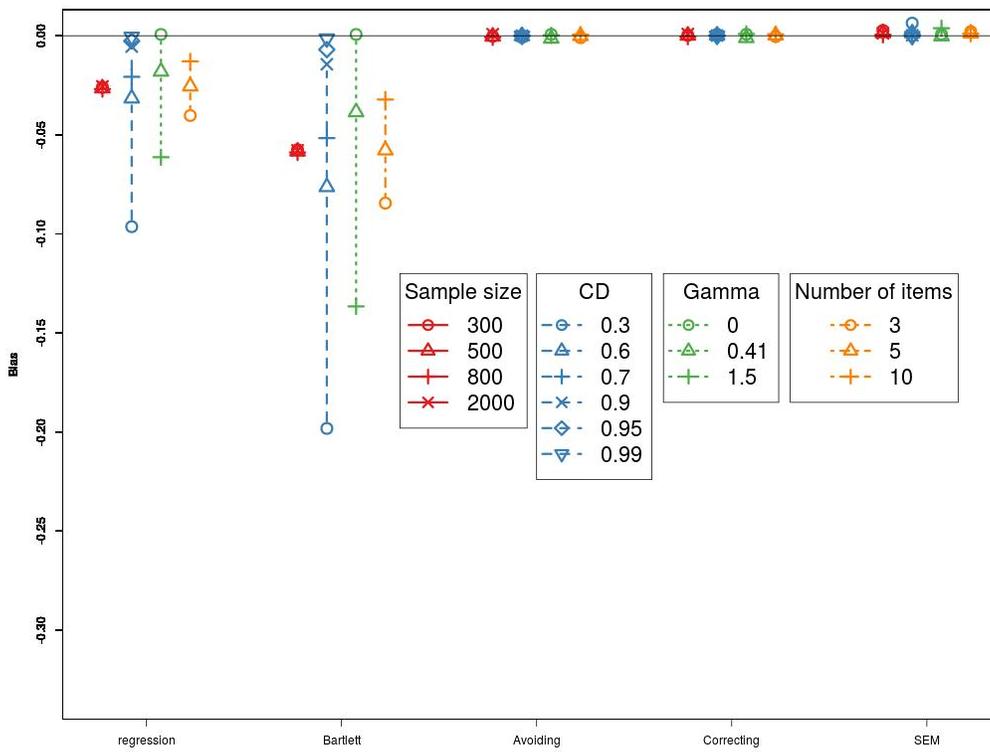
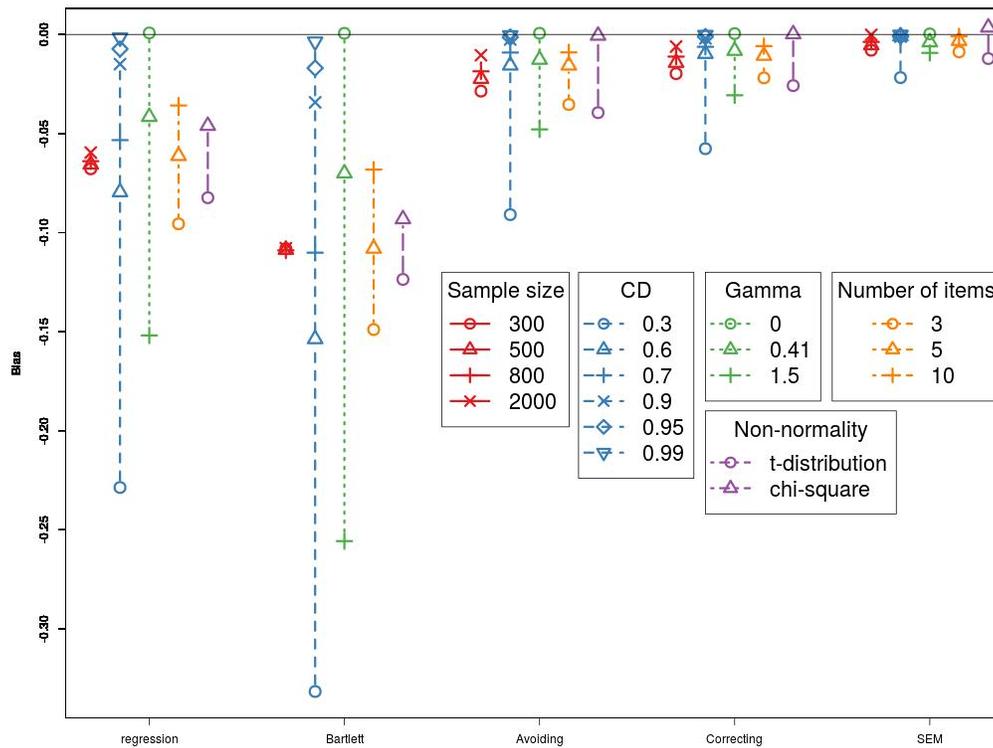


Figure 2. The population model

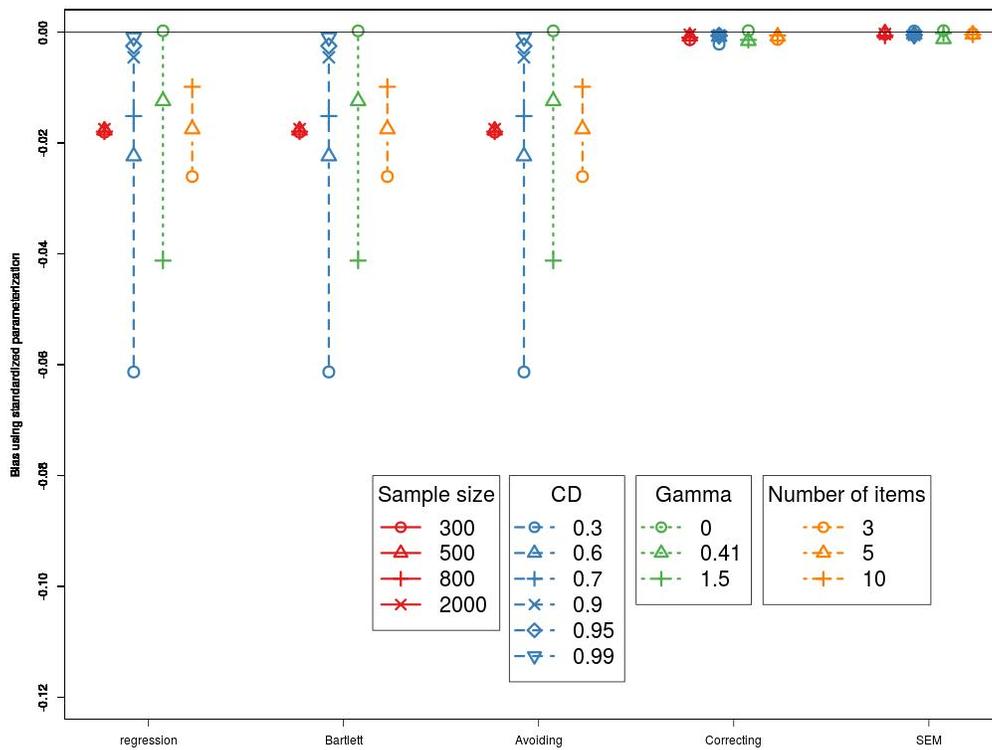


(a) The influence of sample size, coefficient of determination, number of items and the value of γ on the bias, in interaction with the method.

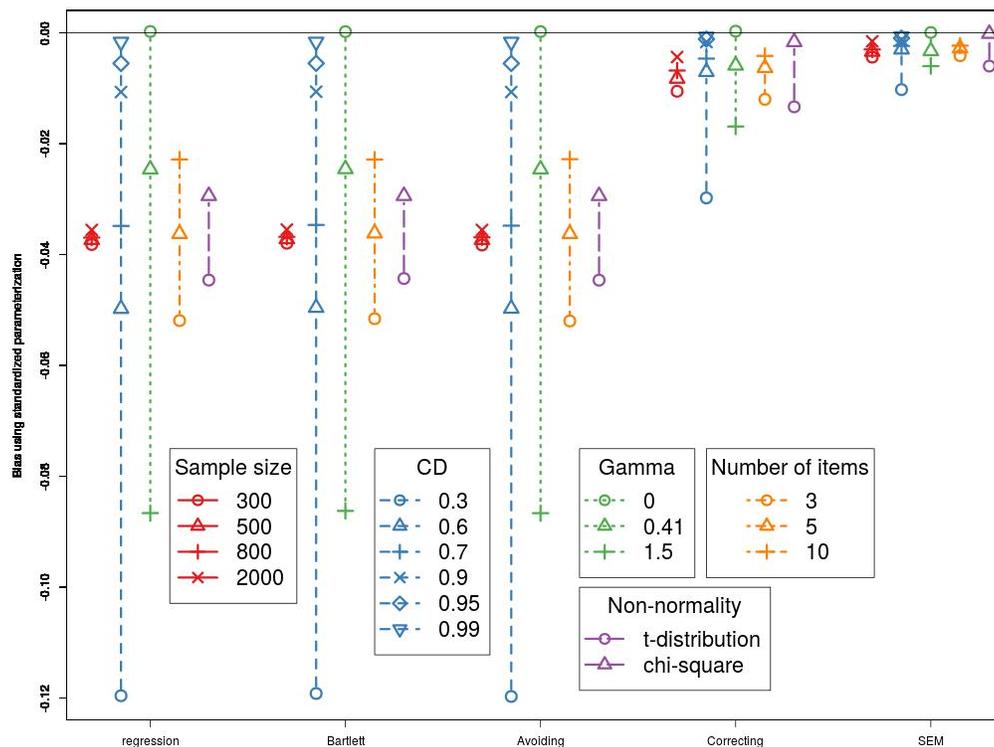


(b) The influence of sample size, coefficient of determination, number of items and the value of γ on the bias, in interaction with the method, when the data are not normally distributed.

Figure 3. Bias

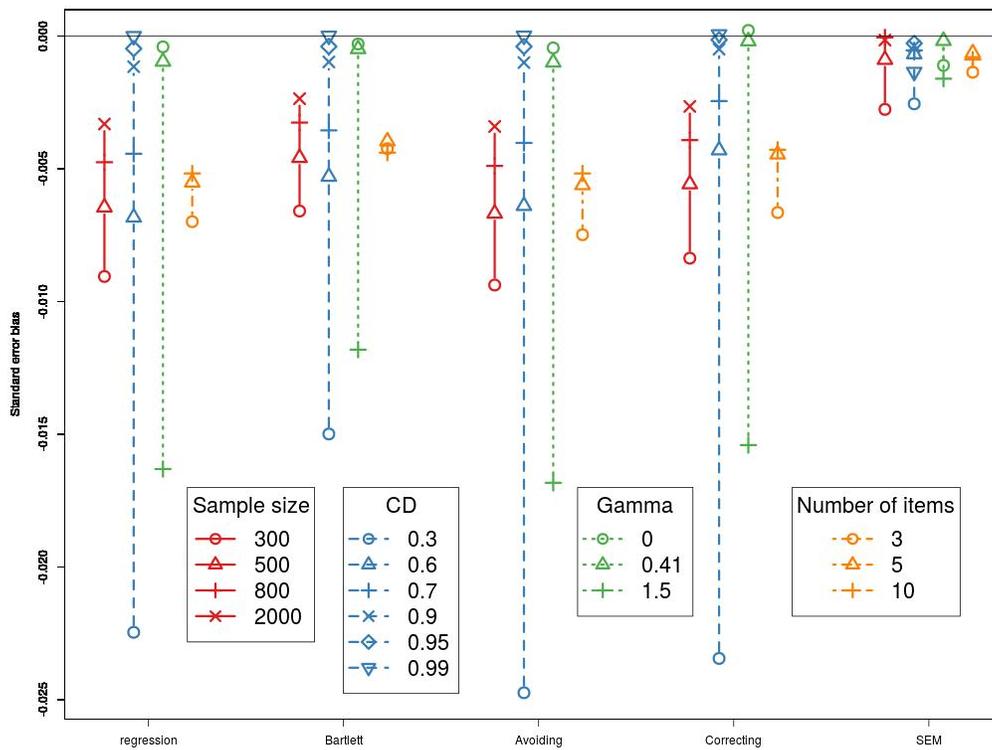


(a) The influence of sample size, CD, number of items and the value of γ on the bias using the standardized parameterization, in interaction with the method.

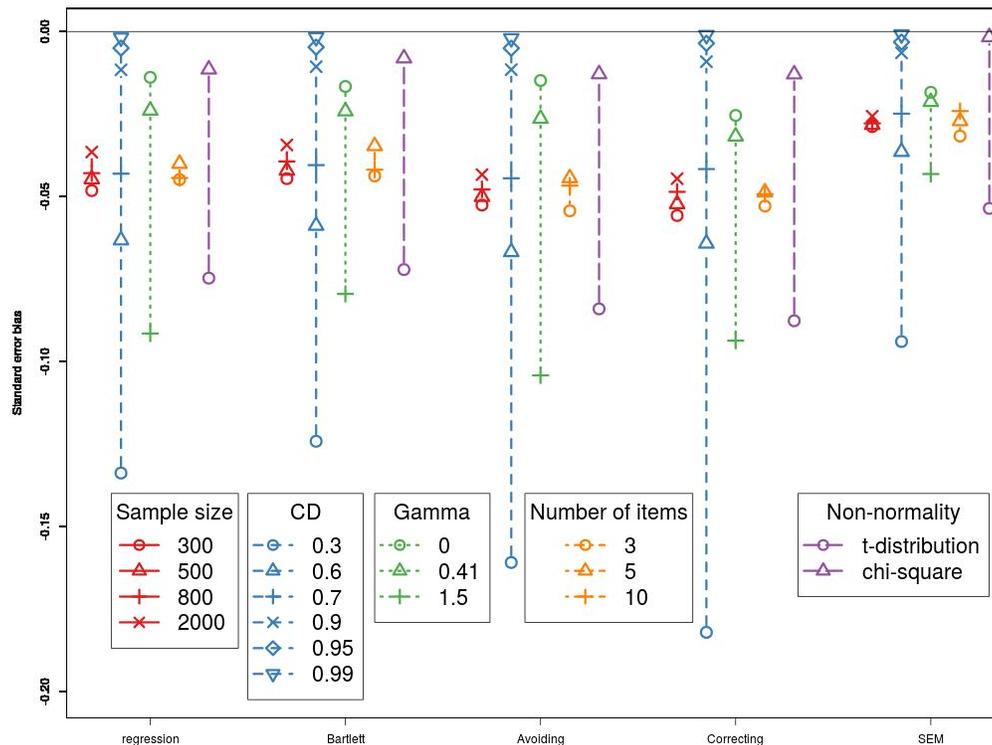


(b) The influence of sample size, CD, number of items and the value of γ on the bias using the standardized parameterization, in interaction with the method, when the data are not normally distributed.

Figure 4. Bias using the standardized parameterization



(a) The influence of sample size, coefficient of determination, number of items and the value of γ on the standard error bias, in interaction with the method.



(b) The influence of sample size, coefficient of determination, number of items and the value of γ on the standard error bias, in interaction with the method, when the data are not normally distributed. Note that the scale of the y-axis is not the same as in panel (a).

Appendix A

Regression coefficient when performing FSR, in the general case

In this appendix, we derive the relationship between β and γ in the general case. We make a distinction between three scenarios:

- 1) both the independent and dependent variable are latent variables
- 2) only the independent variable is a latent variable
- 3) only the dependent variable is a latent variable"

A1. Independent and dependent latent variable

When performing the linear regression with factor scores, the regression coefficient becomes $\beta = \frac{cov(F_\xi, F_\eta)}{var(F_\xi)}$, which is not necessarily the same as the true regression coefficient. The relationship between γ and β can best be understood if we work out the covariance and variances of the factor scores. The covariance can be written as :

$$\begin{aligned}
 cov(F_\xi, F_\eta) &= cov(\mathbf{A}_\xi \mathbf{x}, \mathbf{A}_\eta \mathbf{y}) \\
 &= \mathbf{A}_\xi cov(\mathbf{x}, \mathbf{y}) \mathbf{A}'_\eta \\
 &= \mathbf{A}_\xi cov(\Lambda_x \xi + \delta, \Lambda_y \eta + \epsilon) \mathbf{A}'_\eta \\
 &= \mathbf{A}_\xi \Lambda_x cov(\xi + \delta, \eta + \epsilon) \Lambda'_y \mathbf{A}'_\eta \\
 &= \mathbf{A}_\xi \Lambda_x [cov(\xi, \eta) + cov(\xi, \epsilon) + \\
 &\quad cov(\eta, \epsilon) + cov(\delta, \epsilon)] \Lambda'_y \mathbf{A}'_\eta \\
 &= \mathbf{A}_\xi \Lambda_x cov(\xi, \eta) \Lambda'_y \mathbf{A}'_\eta.
 \end{aligned} \tag{18}$$

The variance can be written as:

$$\begin{aligned}
 var(F_\xi) &= var(\mathbf{A}_\xi \mathbf{x}) \\
 &= \mathbf{A}_\xi var(\mathbf{x}) \mathbf{A}'_\xi \\
 &= \mathbf{A}_\xi \Sigma_x \mathbf{A}'_\xi,
 \end{aligned} \tag{19}$$

where Σ_x is the variance of \mathbf{x} . Based on these calculations, the regression coefficient becomes:

$$\beta = \frac{cov(F_\xi, F_\eta)}{var(F_\xi)} = \frac{\mathbf{A}_\xi \Lambda_x cov(\xi, \eta) \Lambda'_y \mathbf{A}'_\eta}{\mathbf{A}_\xi \Sigma_x \mathbf{A}'_\xi} = \frac{\mathbf{A}_\xi \Lambda_x var(\xi, \eta) \Lambda'_y \mathbf{A}'_\eta}{\mathbf{A}_\xi \Sigma_x \mathbf{A}'_\xi} \gamma. \quad (20)$$

A2. Independent latent variable

When factor scores are only used for the independent variable, the regression coefficient becomes $\beta = \frac{cov(F_\xi, \eta)}{var(F_\xi)}$. Again, we work out the covariance and variances of the factor scores:

$$var(F_\xi) = \mathbf{A}_\xi \Sigma_x \mathbf{A}'_\xi \quad (21)$$

$$\begin{aligned} cov(F_\xi, \eta) &= cov(\mathbf{A}_\xi \mathbf{x}, \eta) \\ &= \mathbf{A}_\xi cov(\mathbf{x}, \eta) \\ &= \mathbf{A}_\xi cov(\Lambda_x \xi, \eta) \\ &= \mathbf{A}_\xi \Lambda_x cov(\xi, \eta) \end{aligned} \quad (22)$$

Based on these calculations, the estimated regression coefficient becomes:

$$\beta = \frac{cov(F_\xi, \eta)}{var(F_\xi)} = \frac{\mathbf{A}_\xi \Lambda_x cov(\xi, \eta)}{\mathbf{A}_\xi \Sigma_x \mathbf{A}'_\xi} = \frac{\mathbf{A}_\xi \Lambda_x var(\xi, \eta)}{\mathbf{A}_\xi \Sigma_x \mathbf{A}'_\xi} \gamma. \quad (23)$$

A3. Dependent latent variable

When only the dependent variable consists of factor scores, the regression coefficient becomes $\beta = \frac{cov(\xi, F_\eta)}{var(\xi)}$. Again, we work out the covariance and variances of the factor scores:

$$\begin{aligned} cov(\xi, F_\eta) &= cov(\xi, \mathbf{A}_\eta \mathbf{y}) \\ &= cov(\xi, \mathbf{y}) \mathbf{A}_\eta \\ &= cov(\xi, \Lambda_y \eta) \mathbf{A}'_\eta \\ &= cov(\xi, \eta) \Lambda'_y \mathbf{A}'_\eta \end{aligned} \quad (24)$$

Based on these calculations, the estimated regression coefficient becomes:

$$\beta = \frac{\text{cov}(\xi, F_\eta)}{\text{var}(\xi)} = \frac{\mathbf{A}_\xi \mathbf{\Lambda}_x \text{cov}(\xi, \eta)}{\text{var}(\xi)} = \mathbf{A}_\xi \mathbf{\Lambda}_x \gamma. \quad (25)$$

Appendix B

Regression coefficient when performing FSR, using regression FSR

In this appendix, we derive the relationship between β and γ when we use the regression predictor to calculate the factor scores. When using the regression predictor, the factor scoring matrices \mathbf{A}_ξ^R and \mathbf{A}_η^R are computed as follows:

$$\mathbf{A}_\xi^R = \Phi \Lambda'_x \Sigma_x^{-1} = \text{var}(\xi) \Lambda'_x \Sigma_x^{-1}, \quad (26)$$

$$\mathbf{A}_\eta^R = \text{var}(\eta) \Lambda'_y \Sigma_y^{-1}. \quad (27)$$

This means the formula for the variance of F_ξ can be simplified:

$$\begin{aligned} \text{var}(F_\xi^R) &= \mathbf{A}_\xi^R \Sigma_x \mathbf{A}_\xi^{R'} \\ &= \text{var}(\xi) \Lambda'_x \Sigma_x^{-1} \Sigma_x \Lambda_\xi^{R'} \\ &= \text{var}(\xi) \Lambda'_x \mathbf{I} \Lambda_\xi^{R'} \\ &= \text{var}(\xi) \Lambda'_x \Lambda_\xi^{R'} = \Phi' \Lambda' \Sigma_x^{-1} \Lambda \Phi \\ &= \mathbf{A}_\xi^R \Lambda_x \text{var}(\xi). \end{aligned} \quad (28)$$

The formulas for the covariances stay unchanged. Based on these calculations, the regression coefficient can be recalculated. When both variables are latent, the regression coefficient becomes:

$$\beta = \frac{\text{cov}(F_\xi, F_\eta)}{\text{var}(F_\xi)} = \frac{\mathbf{A}_\xi \Lambda_x \text{cov}(\xi, \eta) \Lambda'_y \mathbf{A}'_\eta}{\mathbf{A}_\xi^R \Lambda_x \text{var}(\xi)} = \frac{\text{cov}(\xi, \eta) \Lambda'_y \mathbf{A}'_\eta}{\text{var}(\xi)} = \Lambda'_y \mathbf{A}'_\eta \gamma. \quad (29)$$

When only the independent variable is latent, the regression coefficient becomes:

$$\beta = \frac{\text{cov}(F_\xi, \eta)}{\text{var}(F_\xi)} = \frac{\mathbf{A}_\xi \Lambda_x \text{cov}(\xi, \eta)}{\mathbf{A}_\xi^R \Lambda_x \text{var}(\xi)} = \frac{\text{cov}(\xi, \eta)}{\text{var}(\xi)} = \gamma. \quad (30)$$

When only the dependent variable is latent, the regression coefficient becomes:

$$\beta = \frac{\text{cov}(\xi, F_\eta)}{\text{var}(\xi)} = \frac{\mathbf{A}_\xi \mathbf{\Lambda}_x \text{cov}(\xi, \eta)}{\text{var}(\xi)} = \mathbf{A}_\xi \mathbf{\Lambda}_x \gamma. \quad (31)$$

Appendix C

Regression coefficient when performing FSR, using Bartlett FSR

In this appendix, we derive the relationship between β and γ when we use the Bartlett predictor to calculate the factor scores. The Bartlett predictor has the advantage that $\mathbf{A}_\xi^B \mathbf{\Lambda}_x = 1$ and $\mathbf{A}_\eta^B \mathbf{\Lambda}_y = 1$. This implies that the formulas for the covariances can be simplified:

$$\begin{aligned} \text{cov}(F_\xi, F_\eta) &= \mathbf{A}_\xi \mathbf{\Lambda}_x \text{cov}(\xi, \eta) \mathbf{\Lambda}_y' \mathbf{A}_\eta' = \text{cov}(\xi, \eta) \\ \text{cov}(F_\xi, \eta) &= \mathbf{A}_\xi \mathbf{\Lambda}_x \text{cov}(\xi, \eta) = \text{cov}(\xi, \eta) \\ \text{cov}(\xi, F_\eta) &= \text{cov}(\xi, \eta) \mathbf{\Lambda}_y' \mathbf{A}_\eta' = \text{cov}(\xi, \eta) \end{aligned} \quad (32)$$

The formula for the variance of ξ stays the same. Combined, this gives the following regression coefficient, when both variables are latent:

$$\beta = \frac{\text{cov}(F_\xi, F_\eta)}{\text{var}(F_\xi)} = \frac{\text{cov}(\xi, \eta)}{\mathbf{A}_\xi \mathbf{\Sigma}_x \mathbf{A}_\xi'} = \frac{\text{var}(\xi)}{\mathbf{A}_\xi \mathbf{\Sigma}_x \mathbf{A}_\xi'} \gamma. \quad (33)$$

When only the independent variable is latent, the regression coefficient becomes:

$$\beta = \frac{\text{cov}(F_\xi, \eta)}{\text{var}(F_\xi)} = \frac{\text{cov}(\xi, \eta)}{\mathbf{A}_\xi \mathbf{\Sigma}_x \mathbf{A}_\xi'} = \frac{\text{var}(\xi)}{\mathbf{A}_\xi \mathbf{\Sigma}_x \mathbf{A}_\xi'} \gamma. \quad (34)$$

When only the dependent variable is latent, the regression coefficient becomes:

$$\beta = \frac{\text{cov}(\xi, F_\eta)}{\text{var}(\xi)} = \frac{\text{cov}(\xi, \eta)}{\text{var}(\xi)} = \gamma. \quad (35)$$

Appendix D

Regression coefficient when performing FSR, using the bias avoiding method

In this appendix, we derive the relationship between β and γ when we use the bias avoiding method, which uses the regression predictor to predict the factor scores of the independent variable and the Bartlett predictor to predict the factor scores of the dependent variable. The covariances between the two variables then becomes:

$$\begin{aligned}
 cov(F_{\xi}^R, F_{\eta}^B) &= cov(\mathbf{A}_{\xi}^R \mathbf{x}, \mathbf{A}_{\eta}^B \mathbf{y}) \\
 &= \mathbf{A}_{\xi}^R cov(\mathbf{x}, \mathbf{y}) \mathbf{A}_{\eta}^{B'} \\
 &= \mathbf{A}_{\xi}^R cov(\Lambda_x \xi + \delta, \Lambda_y \eta + \epsilon) \mathbf{A}_{\eta}^{B'} \\
 &= \mathbf{A}_{\xi}^R \Lambda_x cov(\xi + \delta, \eta + \epsilon) \Lambda_y' \mathbf{A}_{\eta}^{B'} \\
 &= \mathbf{A}_{\xi}^R \Lambda_x [cov(\xi, \eta) + cov(\xi, \epsilon) + cov(\eta, \epsilon) + cov(\delta, \epsilon)] \Lambda_y' \mathbf{A}_{\eta}^{B'} \\
 &= \mathbf{A}_{\xi}^R \Lambda_x cov(\xi, \eta) \Lambda_y' \mathbf{A}_{\eta}^{B'} \\
 &= \mathbf{A}_{\xi}^R \Lambda_x cov(\xi, \eta)
 \end{aligned} \tag{36}$$

$$cov(F_{\xi}^R, \eta) = \mathbf{A}_{\xi}^R \Lambda_x cov(\xi, \eta) \tag{37}$$

$$cov(\xi, F_{\eta}^B) = cov(\xi, \eta) \Lambda_y' \mathbf{A}_{\eta}^{B'} = cov(\xi, \eta), \tag{38}$$

while the variance can be written as:

$$var(F_{\xi}^R) = \mathbf{A}_{\xi}^R \Lambda_x var(\xi). \tag{39}$$

Combined, this gives the following regression coefficient, when both variables are latent:

$$\beta = \frac{cov(F_{\xi}^R, F_{\eta}^B)}{var(F_{\xi}^R)} = \frac{\mathbf{A}_{\xi}^R \Lambda_x cov(\xi, \eta)}{\mathbf{A}_{\xi}^R \Lambda_x var(\xi)} = \frac{cov(\xi, \eta)}{var(\xi)} = \gamma. \tag{40}$$

When only the independent variable is latent, the regression coefficient becomes:

$$\beta = \frac{\text{cov}(F_{\xi}^R, \eta)}{\text{var}(F_{\xi}^R)} = \frac{\mathbf{A}_{\xi}^R \Lambda_x \text{cov}(\xi, \eta)}{\mathbf{A}_{\xi}^R \Lambda_x \text{var}(\xi)} = \frac{\text{cov}(\xi, \eta)}{\text{var}(\xi)} = \gamma. \quad (41)$$

When only the dependent variable is latent, the regression coefficient becomes:

$$\beta = \frac{\text{cov}(\xi, F_{\eta}^B)}{\text{var}(\xi)} = \frac{\text{cov}(\xi, \eta)}{\text{var}(\xi)} = \gamma. \quad (42)$$

Appendix E

Regression coefficient when performing FSR, using the bias correcting method

In this appendix, we derive the relationship between β and γ when we use the bias correcting method. To be able to perform the bias correcting method, the covariance between the true latent variable scores and the variance of the independent true latent variable scores needs to be calculated. The computation of the covariance is based on the formula given in equation (18):

$$\text{cov}(\xi, \eta) = \frac{\text{cov}(F_\xi, F_\eta)}{\mathbf{A}_\xi \mathbf{\Lambda}_x \mathbf{\Lambda}'_y \mathbf{A}'_\eta}. \quad (43)$$

When calculating factor scores, $\mathbf{\Lambda}s$ and the $\mathbf{A}s$ matrices are readily available and the $\text{cov}(F_\xi, F_\eta)$ can be calculated. This means that it is possible to compute the true covariance. Similar calculations can be made for $\text{cov}(F_\xi, \eta)$ and $\text{cov}(\xi, F_\eta)$. The calculations for the variance of ξ is slightly more complex. First, we calculate the variance of the factor scores:

$$\begin{aligned} \text{var}(F_\xi) &= \mathbf{A}_\xi \text{var}(\mathbf{x}) \mathbf{A}'_\xi \\ &= \mathbf{A}_\xi \text{var}(\mathbf{\Lambda}_x \xi + \boldsymbol{\delta}) \mathbf{A}'_\xi \\ &= \mathbf{A}_\xi [\text{var}(\mathbf{\Lambda}_x \xi) + \text{var}(\boldsymbol{\delta})] \mathbf{A}'_\xi \\ &= \mathbf{A}_\xi [\mathbf{\Lambda}_x \text{var}(\xi) \mathbf{\Lambda}'_x + \boldsymbol{\Theta}_\delta] \mathbf{A}'_\xi. \end{aligned} \quad (44)$$

Based on this formula, we can derive a formula for the variance of ξ :

$$\begin{aligned} \mathbf{\Lambda}_x \text{var}(\xi) \mathbf{\Lambda}'_x + \boldsymbol{\Theta}_\delta &= \text{var}(F_\xi) (\mathbf{A}_\xi \mathbf{A}'_\xi)^{-1} \\ \mathbf{\Lambda}_x \text{var}(\xi) \mathbf{\Lambda}'_x &= \text{var}(F_\xi) (\mathbf{A}_\xi \mathbf{A}'_\xi)^{-1} - \boldsymbol{\Theta}_\delta \\ \mathbf{\Lambda}_x \text{var}(\xi) \mathbf{\Lambda}'_x &= (\text{var}(F_\xi) - \mathbf{A}_\xi \boldsymbol{\Theta}_\delta \mathbf{A}'_\xi) (\mathbf{A}_\xi \mathbf{A}'_\xi)^{-1} \\ \text{var}(\xi) &= (\text{var}(F_\xi) - \mathbf{A}_\xi \boldsymbol{\Theta}_\delta \mathbf{A}'_\xi) (\mathbf{A}_\xi \mathbf{\Lambda}_x \mathbf{A}'_\xi \mathbf{\Lambda}'_x)^{-1}. \end{aligned} \quad (45)$$

Once the covariance and variance of the true latent variable scores are computed, the regression

coefficient can be computed as $\beta = \frac{\text{cov}(\xi, \eta)}{\text{var}(\xi)}$.