Acta Scientiarum

http://periodicos.uem.br/ojs
ISSN on-line: 1807-8664
Doi: 10.4025/actascitechnol.v43i1.49929

STATISTICS

Generalized ridge estimators adapted in structural equation models

Gislene Araújo Pereira[1], Mariana Resende[2] and Marcelo Ângelo Cirillo[2*]

[1]Instituto de Ciências Sociais Aplicadas, Universidade Federal de Alfenas, Varginha, Minas Gerais, Brazil. [2]Departamento de Estatística, Universidade Federal de Lavras, Aquenta Sol, 37200-900, Cx. Postal 3037, Lavras, Minas Gerais, Brazil. *Author for correspondence. E-mail: macufla@gmail.com

ABSTRACT. Multicollinearity is detected via regression models, where independent variables are strongly correlated. Since they entail linear relations between observed or latent variables, the structural equation models (SEM) are subject to the multicollinearity effect, whose numerous consequences include the singularity between the inverse matrices used in estimation methods. Given to this behavior, it is natural to understand that the suitability of these estimators to structural equation models show the same features, either in the simulation results that validate the estimators in different multicollinearity degrees, or in their application to real data. Due to the multicollinearity overview arose from the fact that the matrices inversion is impracticable, the usage of numerical procedures demanded by the maximum likelihood methods leads to numerical singularity problems. An alternative could be the use of the Partial Least Squares (PLS) method, however, it is demanded that the observed variables are built by assuming a positive correlation with the latent variable. Thus, theoretically, it is expected that the load signals are positive, however, there are no restrictions to these signals in the algorithms used in the PLS method. This fact implies in corrective areas, such as the observed variables removal or new formulations of the theoretical model. In view of this problem, this paper aimed to propose adaptations of six generalized ridge estimators as alternative methods to estimate SEM parameters. The conclusion is that the evaluated estimators presented the same performance in terms of accuracy, precision while considering the scenarios represented by model without specification error and model with specification error, different levels of multicollinearity and sample sizes.

Keywords: structural model; generalized ridge regression; multicollinearity.

Received on September 27, 2019.
Accepted on February 13, 2020.

Introduction

The existence of a strong correlation between the variables involved in the estimation of parameters of a model characterizes the multicollinearity problem, whose main consequence is the high estimates of standard coefficients and errors, compromising conclusions related to statistical inference (Mori & Suzuki, 2018). In view of this problem, numerous alternatives to detect and solve this problem are reported in literature. Tarka (2018), in a review article, mentions that despite the potentially severe consequences for statistical inference, the issue of multicollinearity, as well as the impact of omitting variables, has been little studied about the effects on the analysis of structural equation models (SEM) . Corroborating this statement, some studies on this problem are addressed. Yang and Yuan (2019) mention that the presence of multicollinearity involves obtaining operations of approximate or badly conditioned inverse matrices, causing a problem of numerical nature related to obtaining the estimates of maximum likelihood of a SEM.

Due to numerical convergence problems, Can, Schoot, and Hox (2015) proposed a study that includes the parameter estimates in multilevel structural equations models (MLSEM), obtained by the maximum likelihood and Bayesian methods specifying different degrees of correlation between and among the levels, so that, given a value above 0.80, multicollinearity was detected.

When considering the maximum likelihood approach, the solutions obtained as a result of numerical non-convergence were called as inadmissible solutions. In this way, both methods were compared, having as reference the Bayesian procedure, in which, all solutions were admissible. In view of the above, it was observed that the effect of multicollinearity has a greater impact on the estimates of the intraclass correlation coefficients in both methods, however, a greater amount of inadmissible solutions was found when multicollinearity was specified between the levels.

Leeflang (2011) mentions that, in general, the effect of multicollinearity observed in a SEM model, refers to high correlations between the latent exogenous variables and, as a consequence, specifically the validation of these models becomes complex. Can et al. (2015) point out that the effect of multicollinearity generates a correlation matrix between poorly conditioned predictors so that there is no single mathematical solution to estimate the model's coefficients, referring to the model identifiability problems.

Lan and Maguire (2012) emphasize that the interpretation of direct and indirect effects must be made with caution when the variables are multicolinear since these effects express a cause-and-effect relationship.

Among the solutions addressed and feasible to be applied, given the presence of multicollinearity in SEM, ridge modeling has been little explored. Nyrhinen and Leskinen (2014) studied two methodological procedures involving the ridge trace, namely method A and B. Method A consisted of assigning a constant k in all elements on the diagonal of the model correlation matrix. Method B characterized the assignment of this constant to the corresponding elements of the endogenous and exogenous variables in the model correlation matrix. In this context, in simulation studies, the authors concluded that in both methods, the coefficient estimates were the same; however, method B produced lower standard errors.

Yuan, Wu, and Bentler (2011) showed by means of empirical results that the Ridge procedure for SEM with ordinal data presents a better convergence rate, lower bias, lower mean squared error and better general model evaluation than the widely used procedure of maximum likelihood.

Given that the formation of an SEM model involves linear relations between independent and dependent variables, either observed or latent, with or without measuring errors (Marsh, Morin, Parker, & Kaur, 2014; Larina, 2015; Neelaveni & Manimaran, 2016), the adaptation of different ridge estimators becomes applicable, for example, the generalized ridge regression likelihood and other alternatives proposed by Kibria (2003).

Based on the assumption that the variables involved in structural equation modeling present different correlation levels, it is reasonable to assume the existence of multicollinearity. Given this motivation, this study aimed to incorporate the ridge estimators listed in Table 1 into structural equation modeling, as well as to evaluate them regarding the properties of accuracy and precision by means of the Monte Carlo simulation. Finally, an application to real data is presented, providing the script of the function used in the application and simulation of the parameter estimates.

Material and methods

The methodology proposed for the adaptation of ridge estimators in structural equation modeling is described in the following stages: i) Estimators of generalized ridge regression and alternatives; ii) specification of the structural equation model; iii) adaptation of ridge estimators to the structural equation model and iv) scenarios and parametric values used in the Monte Carlo procedure to validate the generalized ridge estimators in structural equation models.

Estimators of generalized ridge regression and its alternatives

Defining the linear regression model (Equation 1).

$$Y = X\beta + \varepsilon \qquad (1)$$

where:
$Y_{n\times 1}$ is the vector of independent observations, $\beta_{p\times 1}$ is the parametric vector of regression coefficients to be estimated, $X_{n\times p}$ is a known matrix of explanatory variables and $\varepsilon_{n\times 1}$ is the vector of errors, with each component being $\varepsilon_i \sim N(0,\sigma^2)$. Supposing the existence of an orthogonal matrix D, by means of decomposition $\Lambda = D^t C D$, where Λ is the diagonal matrix of eigenvalues of matrix $C(X^t X)$, model (1) is rewritten in canonical form as Equation 2.

$$Y = X^{*}\alpha + e \qquad (2)$$

where:
$X^{*} = XD$ and $\alpha = D'\beta$.

Note that α depends on β, according to Kibria (2003) recommendations , the mean square error (MSE) will be minimized when β is considered as the normalized eigenvector, corresponding to the highest eigenvalue of matrix C, respecting the restriction $\beta^t\beta = 1$. With these specifications, the generalized ridge estimators are obtained by Equation 3.

Table 1. Summary of some ridge regression estimators found in the literature.

Notation	Estimators
$\hat{k}_{HKB}$	$\hat{k} = \frac{p\hat{\sigma}^2}{\sum \hat{\alpha}_i^2} = \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}}$
$\hat{k}_{LW}$	$\hat{k} = \frac{p\hat{\sigma}^2}{\sum \lambda_i \hat{\alpha}_i^2}$
$\hat{k}_{HSL}$	$\hat{k} = \hat{\sigma}^2 \frac{\sum \lambda_i \hat{\alpha}_i^2}{\left(\sum \lambda_i \hat{\alpha}_i^2\right)^2}$
$\hat{k}_{AM}$	$\hat{k} = \frac{1}{p}\sum \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}$
$\hat{k}_{GM}$	$\hat{k} = \frac{\hat{\sigma}^2}{\left(\prod \hat{\alpha}_i^2\right)^{\frac{1}{p}}}$
$\hat{k}_{MED}$	$\hat{k} = Median\left\{\frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}\right\}$

Source: Kibria (2003).

$$\hat{\alpha}(k) = (X^{*\prime}X^{*} + K)^{-1}X^{*\prime}Y \tag{3}$$

where:

$K = diag(k_1, \ldots, k_p)$; $k_i > 0$; p equals the number of independent variables involved in the model, and $\hat{\alpha} = \Lambda^{-1}X^{*\prime}Y$ is the estimator of minimum squares of α. Estimators $\hat{\beta}$ and $\hat{\beta}(k)$ are obtained, respectively by the inverse transformation of $\hat{\alpha}$ and $\hat{\alpha}(k)$. Kibria (2003) mentions that an estimated value of k_i Equation 4, which minimizes the mean square error (MSE) Equation 5 of $\hat{\alpha}(k)$ is defined by:

$$k_i = \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}, \tag{4}$$

$$MSE(\hat{\alpha}(k)) = \hat{\sigma}^2 \sum \frac{\hat{\lambda}_i}{\left(\hat{\lambda}_i + \hat{k}_i\right)^2} + \sum \frac{\hat{k}_i^2 \hat{\alpha}_i^2}{\left(\hat{\lambda}_i + \hat{k}_i\right)^2} \tag{5}$$

error variance $\hat{\sigma}^2$ is estimated by the residual mean square; $\hat{\lambda}_i$ is the i^{th} eigenvalue of matrix $\boldsymbol{C}$ and $\hat{\alpha}_i$ the i^{th} element of $\hat{\boldsymbol{\alpha}}$. Following this methodology, several estimators were proposed in the literature by different authors, as illustrated by the summary described in Table 1.

Specification of the structural equation model

Consider the structural model (Bollen, 2012; Mai, Zhang, & Wen, 2018) Equation 6.

$$\eta = \theta\omega + \zeta \tag{6}$$

$\eta_{(r\times1)}$ referred to a vector of endogenous latent variables; $\theta_{(r\times(r+s))}$ is a partitioned matrix, containing the coefficients that correlate the r endogenous factors and relate the s exogenous factors to the r endogenous factors; $\omega_{((r+s)\times1)}$ represents a partitioned vector of endogenous latent variables and exogenous latent variables and $\zeta_{(r\times1)}$ is an error vector. As a matrix, the system is defined in Equation 7.

$$\begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_r \end{pmatrix} = \begin{pmatrix} 0 & \cdots & \beta_{r1} & \vdots & \gamma_{11} & \cdots & \gamma_{s1} \\ \beta_{12} & \cdots & \beta_{r2} & \vdots & \gamma_{12} & \cdots & \gamma_{s2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \beta_{1r} & \cdots & 0 & \vdots & \gamma_{1r} & \cdots & \gamma_{sr} \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_r \\ \cdots \\ \xi_1 \\ \xi_2 \\ \vdots \\ \xi_s \end{pmatrix} + \begin{pmatrix} \zeta_1 \\ \zeta_2 \\ \vdots \\ \zeta_r \end{pmatrix} \tag{7}$$

Consider subsystem Equation 8 referring to the exogenous variables.

$$X = \Lambda_x \xi + \delta \tag{8}$$

$X_{(q\times1)}$ corresponds to the observed exogenous variables vector; $\Lambda_{x_{(q\times s)}}$ is a matrix of regression coefficients that relates the s exogenous factors to each of q observable variable assigned to measure them; $\xi_{(s\times1)}$ represents a vector of exogenous latent variables and $\delta_{(q\times1)}$ is a vector of measurement errors in X. Following these specifications, the matrix representation is given in Equation 9.

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_q \end{pmatrix} = \begin{pmatrix} \lambda_{11}^x & \lambda_{21}^x & \cdots & \lambda_{s1}^x \\ \lambda_{12}^x & \lambda_{22}^x & \cdots & \lambda_{s2}^x \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_{1q}^x & \lambda_{2q}^x & \cdots & \lambda_{sq}^x \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_s \end{pmatrix} + \begin{pmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_q \end{pmatrix} \quad (9)$$

Adaptation of ridge estimators to the structural equation model

Following the structural model given in Equation 6, the adaptation of ridge estimators is initially made by obtaining an orthogonal matrix D, by means of0 the decomposition $\Psi = D'CD$, where Ψ is the matrix containing the eigenvalues of $C = \omega'\omega$. Therefore, the structural model was rewritten in canonical form as Equation 10:

$$\eta = \omega^*\gamma + \zeta \quad (10)$$

where:
$\omega^* = \omega D$ and $\gamma = D'\theta$.

In relation to the measurement model (8), representative of exogenous variables, the decomposition performed to obtain the orthogonal matrix D was done using equation $D'CD = \Phi$, where Φ is the matrix containing the eigenvalues of $C = \xi'\xi$. Therefore, the model of measurement in x rewritten in canonical form is given as Equation 11.

$$x = \xi^*\lambda^x + \delta \quad (11)$$

where:
$\xi^* = \xi D$ and $\lambda^x = D'\Lambda_x$.

Following the structural model Equation 10 and the measurement model Equation 11, the generalized ridge estimators were obtained by means of Equation 12 and 13.

$$\hat{\gamma}(k) = (\omega^{*\prime}\omega^* + K)^{-1}\omega^{*\prime}\eta \quad (12)$$

$$\widehat{\lambda^x}(k) = (\xi^{*\prime}\xi^* + K)^{-1}\xi^{*\prime}x \quad (13)$$

where:
K is a diagonal matrix of $(r + s)$ order for the structural model, and of (s) order for the measurement model in X, where each element $k_i > 0$. The estimations of ordinary minimum squares of γ and λ^x are given respectively by Equation 14 and 15. In comparison to Equation 4, the estimated value of ki is described in Equation 16 and 17.

$$\hat{\gamma} = \Psi^{-1}\omega^{*\prime}\eta \quad (14)$$

$$\widehat{\lambda^x} = \Phi^{-1}\xi^{*\prime}x \quad (15)$$

$$k_i^E = \frac{{\sigma_E}^2}{\gamma_i^2} \quad (16)$$

$$k_i^M = \frac{{\sigma_M}^2}{(\lambda_i^x)^2}. \quad (17)$$

As with the Equation 5, the mean square error is defined in Equation 18 and 19, respectively for each model and minimized with the estimation of k_i.

$$MSE(\hat{\gamma}(k_i)) = \hat{\sigma}_E^2 \sum \frac{\hat{\psi}_i}{(\hat{\psi}_i + \hat{k}_i)^2} + \sum \frac{\hat{k}_i^2\hat{\gamma}_i^2}{(\hat{\psi}_i + \hat{k}_i)^2} \quad (18)$$

$$MSE(\hat{\lambda}_x(k_i)) = \hat{\sigma}_M^2 \sum \frac{\hat{\Phi}_i}{(\hat{\Phi}_i + \hat{k}_i)^2} + \sum \frac{\hat{k}_i^2(\hat{\lambda}_i^x)^2}{(\hat{\Phi}_i + \hat{k}_i)^2} \quad (19)$$

where:

$\hat{\sigma}_E^2$ and $\hat{\sigma}_M^2$ refer to the estimations of error variances of Equation 10 and 11; $\hat{\lambda}_i^x$ the i^{th} estimation of λ^x; and $\hat{\gamma}_i$ the i^{th} estimation of γ. Therefore, following these specifications, the generalized ridge estimators listed in Table 1 and adapted to structural equation modeling, following the methodology proposed herein, are described in Table 2.

Concerning the measurement model related to the endogenous variables, the adaptations of the generalized ridge estimators are made as with the measurement model in X described in this study.

Scenarios and parametric values used in the Monte Carlo procedure to validate the generalized ridge estimators in structural equation models

After defining the generalized ridge estimators (Table 2), two thousand $(2,000)$ Monte Carlo simulations were used, based on the structural equation model, with the parametric values specified as illustrated in Figure 1.

Mantaining the usual assumptions of the structural model, where the expectations of error vectors and latent variables equal zero, ζ and ξ_i (i = 1, 2, 3) are not correlated; ε_j (j = 1, 2, 3, 4) are not correlated to η, ξ_i and δ_i; and δ_i (i = 1, 2, 3) are nor correlated to ξ_i, η and ε_j. The indicators of exogenous latent variables will be considered multicollinear in different levels, being generated by $2,000$ Monte Carlo simulations, according to the procedure proposed by Pereira, Milani, and Cirillo (2014) [Equation 20].

Table 2. Generalized ridge estimators adapted to structural equation models.

Structural Model	Measurement Model X
$\hat{k}_{HKB} = \frac{(r+s)\hat{\sigma}^2}{\sum \hat{\gamma}_i^2}$	$\hat{k}_{HKB} = \frac{s\hat{\sigma}^2}{\sum (\hat{\lambda}_i^x)^2}$
$\hat{k}_{LW} = \frac{\hat{\sigma}^2}{\sum \psi_i \hat{\gamma}_i^2}$	$\hat{k}_{LW} = \frac{\hat{\sigma}^2}{\sum \phi_i (\hat{\lambda}_i^x)^2}$
$\hat{k}_{HSL} = \hat{\sigma}^2 \frac{\sum \psi_i \hat{\gamma}_i^2}{(\sum \psi_i \hat{\gamma}_i^2)^2}$	$\hat{k}_{HSL} = \hat{\sigma}^2 \frac{\sum \phi_i (\hat{\lambda}_i^x)^2}{\left(\sum \phi_i (\hat{\lambda}_i^x)^2\right)^2}$
$\hat{k}_{AM} = \frac{1}{(r+s)} \sum \frac{\hat{\sigma}^2}{\hat{\gamma}_i^2}$	$\hat{k}_{AM} = \frac{1}{s} \sum \frac{\hat{\sigma}^2}{(\hat{\lambda}_i^x)^2}$
$\hat{k}_{GM} = \frac{\hat{\sigma}^2}{(\prod \hat{\gamma}_i^2)^{\frac{1}{(r+s)}}}$	$\hat{k}_{GM} = \frac{\hat{\sigma}^2}{(\prod (\hat{\lambda}_i^x)^2)^{\frac{1}{s}}}$
$\hat{k}_{MED} = Median\left\{\frac{\hat{\sigma}^2}{\hat{\gamma}_i^2}\right\}$	$\hat{k}_{MED} = Median\left\{\frac{\hat{\sigma}^2}{(\hat{\lambda}_i^x)^2}\right\}$

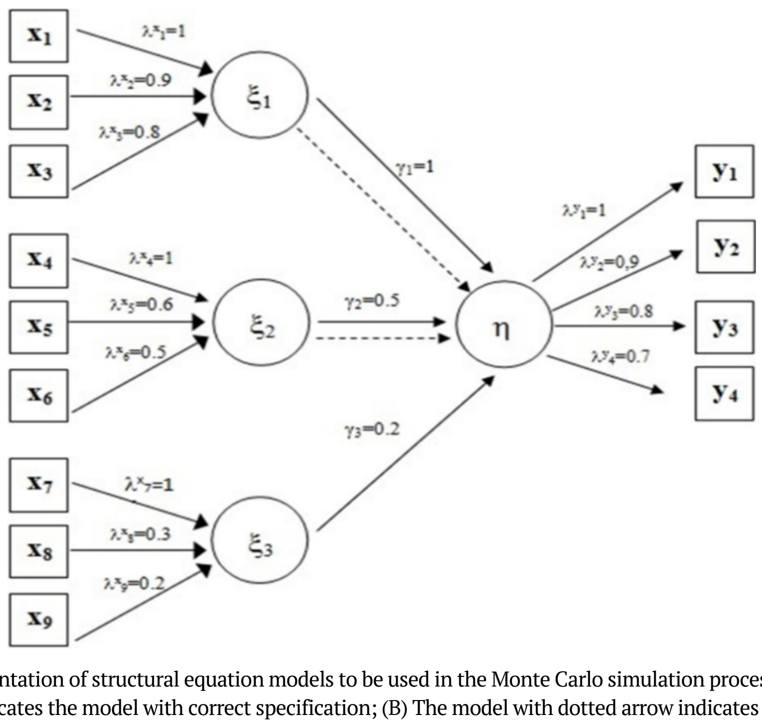

Figure 1. Graphical representation of structural equation models to be used in the Monte Carlo simulation process: (A) The structural model with continuous arrows indicates the model with correct specification; (B) The model with dotted arrow indicates a model with a specification error in the joint omission of $_1$ and $_2$. Source: Adapted from Cirillo and Barroso (2012) and Maydeu-Olivares, Shi, & Rosseel (2019).

$$x_{ij} = (1-\rho^2)^{1/2} z_{ij} + \rho z_{i(j+1)}, (i = 1, \dots n); (j = 1, \dots q) \quad (20)$$

where:
$z_{ij} \sim N(0,1)$, ρ is specified so that the correlation between both explanatory variables is given by ρ^2; n is the sample size; and q is the number of observed exogenous variables to be generated. Thus, the multicollinearity level between variables, controlled and classified in levels, is: weak ($\rho^2 = 0.09$), moderate ($\rho^2 = 0.49$ and 0.64) and strong ($\rho^2 = 0.81$ and 0.98), for sample sizes evaluated in $n = 100$, 200 and $1,000$, therefore, a identifiability of the model due to the number of parameters required as a function of sample size (Cirillo & Barroso, 2017) for each estimator (Table 2).

For all estimators, and for each configuration between sample size and multicollinearity level, the mean square error Equation 18 and 19 was computed considering the mean estimations obtained in $2,000$ Monte Carlo iterations. For this purpose, a function will be built and implemented in the software R (R Core Team, 2018).

Result and discussion

Accuracy and precision of generalized ridge estimators considering a structural equation model without specification error

According to the methodology, in view of the evaluated scenarios (Section iii), given the approaches of estimations of mean square errors due to Monte Carlo oscillations, the results described in Table 3 show that all estimators were accurate and precise for all levels of multicollinearity evaluated.

Diamantopoulos, Riefler and Roth (2008) mention that the presence of multicollinear variables in models with formative indicators may induce the researcher to exclude insignificant indicators, thus altering the construct definition. Therefore, considering that results are similar, generalized ridge estimators can be recommended as an alternative method for estimation of the model parameters in relation to their competitors, once their accuracy and precision have been confirmed through the low values of the mean square error. Concerning the effect of sample size, the accuracy and precision of the generalized ridge estimators (Table 3) were in accordance with the studies conducted by Cassel, Hackl, and Westlund (1999) by using the PLS estimation method, where it was concluded that the bias estimations were not affected by the increase in sample size. However, comparing to the results obtained by Jung (2013), where the ridge regression incorporated to the square minimum method in two stages was considered, the authors concluded that, for small sample sizes, the estimations obtained by a ridge method were more stable and precise, but associated with the greatest biases.

Table 3. Estimation of Mean Square Errors (MSE) of the structural model without any specification error, considering different sample sizes (n) and different multicollinearity levels (ρ).

				$\rho^2 = 0.09$			
n	OLS	$\hat{k}_{HKB}$	$\hat{k}_{LW}$	$\hat{k}_{HSL}$	$\hat{k}_{AM}$	$\hat{k}_{GM}$	$\hat{k}_{MED}$
100	1.3182	0.0106	0.0115	0.0113	0.0113	0.0106	0.011
200	1.3008	0.0057	0.0059	0.0058	0.0058	0.0059	0.0058
1000	1.2978	0.0012	0.0012	0.0012	0.0012	0.0013	0.0012
				$\rho^2 = 0.49$			
n	OLS	$\hat{k}_{HKB}$	$\hat{k}_{LW}$	$\hat{k}_{HSL}$	$\hat{k}_{AM}$	$\hat{k}_{GM}$	$\hat{k}_{MED}$
100	1.3296	0.0107	0.0116	0.0114	0.0114	0.0107	0.0111
200	1.2939	0.0056	0.0058	0.0058	0.0058	0.006	0.0057
1000	1.2961	0.0012	0.0012	0.0012	0.0012	0.0014	0.0012
				$\rho^2 = 0.64$			
n	OLS	$\hat{k}_{HKB}$	$\hat{k}_{LW}$	$\hat{k}_{HSL}$	$\hat{k}_{AM}$	$\hat{k}_{GM}$	$\hat{k}_{MED}$
100	1.3352	0.0108	0.0116	0.0114	0.0115	0.0108	0.0112
200	1.2848	0.0056	0.0058	0.0057	0.0058	0.0059	0.0057
1000	1.3038	0.0012	0.0012	0.0012	0.0012	0.0012	0.0012
				$\rho^2 = 0.81$			
n	OLS	$\hat{k}_{HKB}$	$\hat{k}_{LW}$	$\hat{k}_{HSL}$	$\hat{k}_{AM}$	$\hat{k}_{GM}$	$\hat{k}_{MED}$
100	1.3203	0.0108	0.0117	0.0115	0.0115	0.0108	0.0112
200	1.3025	0.0057	0.0059	0.0058	0.0058	0.006	0.0058
1000	1.2872	0.0012	0.0012	0.0012	0.0012	0.0013	0.0012
				$\rho^2 = 0.98$			
n	OLS	$\hat{k}_{HKB}$	$\hat{k}_{LW}$	$\hat{k}_{HSL}$	$\hat{k}_{AM}$	$\hat{k}_{GM}$	$\hat{k}_{MED}$
100	1.3366	0.0108	0.0117	0.0114	0.0115	0.0108	0.0112
200	1.3362	0.0058	0.0060	0.0060	0.0060	0.0062	0.0059
1000	1.2968	0.0012	0.0012	0.0012	0.0012	0.0013	0.0012

Given that the residues were generated by assuming normality, there are statistical evidences to state that accuracy and precision are reached by generalized ridge estimators. The equality between estimations of mean square errors was certainly influenced by the fact that residues were generated by a symmetric distribution, in this case, normal distribution. This statement is confirmed through studies obtained by Cirillo and Barroso (2012) by considering robust estimators LMS and LTS for the same model (Figure 1), but generated with symmetric and asymmetric errors. The results related to biases were quite discrepant. In relation to the accuracy of estimations, for all sample sizes, the LMS method showed a trend of overestimating parametric values; and the LTS method, a trend of underestimating them.

Given the same scenarios evaluated by the Monte Carlo simulation, the results described in Table 4 showed that, by considering the specification error, omitting ξ_1 and ξ_2 simultaneously, mean square error estimations were accurate and precise, for all generalized ridge estimators with small oscillations due to the Monte Carlo error. It is important to note that the results obtained for these estimators are coherent with studies conducted by Maydeu-Olivares et al. (2019), who compared a structural model without specification error to a model considering the omission of two causal relations simultaneously. In this context, the authors concluded that there are no significant differences between models with or without specification error in relation to the adjustment quality.

Application to real data

Based on data related to the profile description of coffee consumers in relation to brand and quality, a questionnaire was applied using a 5-point Likert scale involving demographic and economic questions in a sample of 80 individuals (Table 5).

As these are categorical and ordinal questions, the authors chose to transform the data into a continuous scale between 0 and 1. For such, the transformation given in Equation 21 was used, referring to the i^{th} response to the k^{th} question.

$$x_{ik}^* = \frac{x_{ik} - x_{Lk}}{x_{Uk} - x_{Lk}} \quad (21)$$

where:

x_{Lk} and x_{Uk} are, respectively, the lowest and highest responses. Aiming to identify latent variables ξ and η, an exploratory factorial analysis was performed, using the data transformed in continuous scale (Equation 21), justifying the Pearson correlation matrix used in the analysis. Results are listed below in Table 6.

Table 4. Estimation of Mean Squared Error (MSE) of the structural model with a specification error in joint omission of ξ_1 and ξ_1, considering different sample sizes (n) and different degrees of multicollinearity (ρ).

				$\rho^2 = 0.09$			
n	OLS	$\hat{k}_{HKB}$	$\hat{k}_{LW}$	$\hat{k}_{HSL}$	$\hat{k}_{AM}$	$\hat{k}_{GM}$	$\hat{k}_{MED}$
100	1.1560	0.0112	0.0112	0.0137	0.0112	0.0112	0.0112
200	1.1491	0.0058	0.0058	0.0065	0.0058	0.0058	0.0058
1000	1.1408	0.0011	0.0011	0.0012	0.0012	0.0011	0.0011
				$\rho^2 = 0.49$			
n	OLS	$\hat{k}_{HKB}$	$\hat{k}_{LW}$	$\hat{k}_{HSL}$	$\hat{k}_{AM}$	$\hat{k}_{GM}$	$\hat{k}_{MED}$
100	1.1495	0.0111	0.0111	0.0135	0.0111	0.0111	0.0111
200	1.1546	0.0057	0.0057	0.0065	0.0057	0.0057	0.0057
1000	1.1404	0.0012	0.0012	0.0012	0.0012	0.0012	0.0012
				$\rho^2 = 0.64$			
n	OLS	$\hat{k}_{HKB}$	$\hat{k}_{LW}$	$\hat{k}_{HSL}$	$\hat{k}_{AM}$	$\hat{k}_{GM}$	$\hat{k}_{MED}$
100	1.1497	0.0111	0.0111	0.0135	0.0111	0.0111	0.0111
200	1.1551	0.0058	0.0058	0.0065	0.0058	0.0058	0.0058
1000	1.1479	0.0012	0.0012	0.0012	0.0012	0.0012	0.0012
				$\rho^2 = 0.81$			
n	OLS	$\hat{k}_{HKB}$	$\hat{k}_{LW}$	$\hat{k}_{HSL}$	$\hat{k}_{AM}$	$\hat{k}_{GM}$	$\hat{k}_{MED}$
100	1.1523	0.0111	0.0111	0.0135	0.0111	0.0111	0.0111
200	1.1494	0.0057	0.0057	0.0064	0.0057	0.0057	0.0057
1000	1.1435	0.0012	0.0012	0.0012	0.0012	0.0012	0.0012
				$\rho^2 = 0.98$			
n	OLS	$\hat{k}_{HKB}$	$\hat{k}_{LW}$	$\hat{k}_{HSL}$	$\hat{k}_{AM}$	$\hat{k}_{GM}$	$\hat{k}_{MED}$
100	1.1639	0.0114	0.0114	0.0138	0.0114	0.0114	0.0114
200	1.1521	0.0058	0.0058	0.0065	0.0058	0.0058	0.0058
1000	1.1420	0.0011	0.0011	0.0012	0.0011	0.0011	0.0011

The selection of the variables to be used in constructing each construct was determined according to the variables that had higher estimations of factorial loads (Table 6). Therefore, the set of equations constituting the structural model is represented by Equation 22.

$$\begin{aligned} \xi_1 &= \lambda_1^x x_1^* + \lambda_2^x x_2^* + \delta_1 \\ \xi_2 &= \lambda_6^x x_6^* + \lambda_7^x x_7^* + \lambda_8^x x_8^* + \delta_2 \\ \xi_3 &= \lambda_3^x x_3^* + \lambda_9^x x_9^* + \delta_3 \\ \xi_4 &= \lambda_4^x x_4^* + \lambda_5^x x_5^* + \delta_4 \\ \eta &= \gamma_1\xi_1 + \gamma_2\xi_2 + \gamma_3\xi_3 + \gamma_4\xi_4 + \zeta. \end{aligned} \quad (22)$$

The results in Table 7 validate the model adjustment because they have low values for the mean square error of generalized ridge estimators in comparison to square minimum estimations.

The generalized ridge estimators showed inferior results to those of the MSE obtained by the square minimum method. Another important result is the proximity of estimations obtained in each method, corroborating the results obtained in simulation. Subsequently, the estimations of the model parameters for each evaluated method are described in Table 8, along with the coefficient of determination R^2.

Considering the estimates of model parameters, the results in Table 8 showed that all generalized regression ridge methods have similar estimations, including results related to the mean square error (Table 6). Therefore, there is statistical evidence to state that the inferential procedures, which can be applied to model adjustment, will provide similar results to those of the estimators evaluated in this study.

Table 5. Categorical Questions and observed variables (X) used in the research.

	Social Class	
	Questions	Description of categories
X_1	1- Scholarly	Scholarly levels: 1; 2; 3; 4; 5.
X_2	2-Family income	Wage income levels: 1; 2; 3; 4; 5.
	Importance assigned	
	Questions	Description of categories
X_3	3- How important do you attribute the brand when choosing coffee?	Degree of importance: 1; 2; 3; 4; 5
X_4	4- The coffee quality depends more on the roaster firm and not the farmers.	Score for quality: 1; 2; 3; 4; 5
X_5	5- The coffee producer directly affects the quality of coffee.	Degree of importance of producer: 1; 2; 3; 4; 5
X_6	6- The region where coffee is produced interferes with the quality.	Degree of importance to region: 1; 2; 3; 4; 5
X_7	7- When I buy coffee, I look for information on the production region.	Degree of importance to region: 1; 2; 3; 4 ; 5
X_8	8 - I prefer coffee with identification of the origin, even if they are more expensive.	Degree of importance for the preference: 1; 2; 3; 4 ; 5
X_9	9- I prefer coffee of superior quality, even if they are more expensive.	Degree of importance for the price: 1; 2; 3; 4; 5

Table 6. Factorial loads obtained in the factorial analysis through the *varimax criterion*.

Constructs	Variable	Factor1	Factor2	Factor3	Factor4
Social Class (ξ_1)	x_1^*	0.106	0.963	0.174	0.164
	x_2^*		0.543	0.160	
Origin/Region (ξ_2)	x_6^*	0.684	0.208		0.322
	x_7^*	0.805			0.443
	x_8^*	0.955	0.113	0.263	
Price/Brand (ξ_3)	x_3^*		0.267	0.638	
	x_9^*	0.258	0.189	0.844	
Production (ξ_4)	x_4^*	-0.110	-0.233	0.354	0.489
	x_5^*		-0.135	0.139	0.575

Table 7. Mean Squared Error (MSE) estimations for the structural model.

Estimator	MSE
OLS	1.416441
$\hat{k}_{HKB}$	0.024535
$\hat{k}_{LW}$	0.023308
$\hat{k}_{HSL}$	0.022168
$\hat{k}_{AM}$	0.022947
$\hat{k}_{GM}$	0.033156
$\hat{k}_{MED}$	0.022153

Table 8. Parameter estimations for the structural model given in Equation 22.

			ξ_1 (R^2 = 0.5484)				
			Estimates				
λ^X	OLS	$\hat{k}_{HKB}$	$\hat{k}_{LW}$	$\hat{k}_{HSL}$	$\hat{k}_{AM}$	$\hat{k}_{GM}$	$\hat{k}_{MED}$
λ_1^x	0.4104	0.4033	0.4103	0.4066	0.4087	0.3834	0.4069
λ_2^x	0.5424	0.5239	0.5421	0.5321	0.5376	0.4813	0.5329
			ξ_2(R^2 = 0.6277)				
			Estimates				
λ^X	OLS	$\hat{k}_{HKB}$	$\hat{k}_{LW}$	$\hat{k}_{HSL}$	$\hat{k}_{AM}$	$\hat{k}_{GM}$	$\hat{k}_{MED}$
λ_6^x	-0.3945	-0.3058	-0.3930	-0.3623	-0.3806	-0.3877	-0.3571
λ_7^x	-0.2914	-0.2707	-0.2917	-0.2907	-0.2925	-0.2885	-0.2897
λ_8^x	-0.2657	-0.2264	-0.2649	-0.2507	-0.2588	-0.2587	-0.2485
			ξ_3(R^2 = 0.5768)				
			Estimates				
λ^X	OLS	$\hat{k}_{HKB}$	$\hat{k}_{LW}$	$\hat{k}_{HSL}$	$\hat{k}_{AM}$	$\hat{k}_{GM}$	$\hat{k}_{MED}$
λ_3^x	0.4118	0.4035	0.4117	0.4074	0.4097	0.3691	0.4077
λ_9^x	0.5029	0.4846	0.5026	0.4930	0.4981	0.4252	0.4935
			ξ_4(R^2 = 0.6447)				
			Estimates				
λ^X	OLS	$\hat{k}_{HKB}$	$\hat{k}_{LW}$	$\hat{k}_{HSL}$	$\hat{k}_{AM}$	$\hat{k}_{GM}$	$\hat{k}_{MED}$
λ_4^x	-0.5097	-0.4314	-0.5079	-0.4682	-0.4876	-0.4696	-0.4673
λ_5^x	0.5246	0.4435	0.5227	0.4816	0.5016	0.4712	0.4806
			η(R^2 = 0.6475)				
			Estimates				
γ	OLS	$\hat{k}_{HKB}$	$\hat{k}_{LW}$	$\hat{k}_{HSL}$	$\hat{k}_{AM}$	$\hat{k}_{GM}$	$\hat{k}_{MED}$
γ_1	0.1928	0.1705	0.1924	0.1872	0.1914	0.1823	0.1871
γ_2	-0.4253	-0.3521	-0.4238	-0.4050	-0.4199	-0.3825	-0.4043
γ_3	0.2713	0.2231	0.2703	0.2577	0.2677	0.2470	0.2573
γ_4	0.4173	0.3408	0.4157	0.3956	0.4114	0.4013	0.3950

Conclusion

Our results support the conclusion that the generalized ridge estimators adapted to structural equation models can be applied to real situations, including problems involving a strong multicollinearity between the observed variables. The generalized ridge estimators showed the same performance in relation to accuracy, precision, model specification error, multicollinearity level and sample size.

Regarding the situations in which the model was generated considering the specification error of the latent variables, the ridge estimators presented accurate and precise results, showing robustness in relation to the omission of the latent variables.

Acknowledgements

The authors thank the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq), for the financial support though the research productivity project no. 304974/2015-3 and the project n° 140242/2019-8.

References

Bollen, K. A. (2012). Instrumental variables in sociology and the social sciences. *Annual Review of Sociology, 38*(1), 37-72. doi: 10.1146/annurev-soc-081309-150141

Can, S., Schoot, R. V., & Hox, J. (2015). Collinear latent variables in multilevel confirmatory factor analysis: a comparison of maximum likelihood and Bayesian estimations. *Educational and Psychological Measurement, 75*(3), 406-427. doi: 10.1177/0013164414547959

Cassel, C., Hackl, P., & Westlund, A. H. (1999). Robustness of partial least-squares method for estimating latent variable quality structures. *Journal of Applied Statistics, 26*(4), 435-446. doi: 10.1080/02664769922322

Cirillo, M. Â., & Barroso, L. P. (2012). Robust regression estimates in the prediction of latent variables in structural equation models. *JMASM - Journal of Modern Applied Statistical Methods, 11*(1), 42-53. doi: 10.22237/jmasm/1335844980

Cirillo, M. Â., & Barroso, L. P. (2017). Effect of outliers on the GFI quality adjustment index in structural equation model and proposal of alternative indices. *Communications in Statistics - Simulation and Computation, 46*(3), 1895-1905. doi: 10.1080/03610918.2015.1018998

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research, 61*(12), 1203-1218. doi: 10.1016/j.jbusres.2008.01.009

Jung, S. (2013). Structural equation modeling with small sample sizes using two-stage ridge least-squares estimation. *Behavior Research Methods, 45*(1), 75-81. doi: 10.3758/s13428-012-0297-7

Kibria, B. M. G. (2003). Performance of some new ridge regression estimators. *Communications in Statistics - Simulation and Computation, 32*(2), 419-435. doi: 10.1081/SAC-120017499

Lam, T. Y., & Maguire, D. A. (2012). Structural equation modeling: theory and applications in forest management. *International Journal of Forestry Research, 2012*, 1-16. doi: 10.1155/2012/263953

Larina, L. N. (2015). Practical application of total quality management system to education of international students. *Procedia - Social and Behavioral Sciences, 215*, 9-13. doi: 10.1016/j.sbspro.2015.11.566

Leeflang, P. (2011). Paving the way for "distinguished marketing". *International Journal of Research in Marketing, 28*(2), 76-88. doi: 10.1016/j.ijresmar.2011.02.004

Mai, Y., Zhang, Z., & Wen, Z. (2018). Comparing exploratory structural equation modeling and existing approaches for multiple regression with latent variables. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(5), 737-749. doi: 10.1080/10705511.2018.1444993

Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology, 10*(1), 85-110. doi: 10.1146/annurev-clinpsy-032813-153700

Maydeu-Olivares, A., Shi, D., & Rosseel, Y. (2019). Instrumental variables two-stage least squares (2SLS) vs. maximum likelihood structural equation modeling of causal effects in linear regression models. *Structural Equation Modeling: A Multidisciplinary Journal, 26*(6), 876-892. doi: 10.1080/10705511.2019.1607740

Mori, Y., & Suzuki, T. (2018). Generalized ridge estimator and model selection criteria in multivariate linear regression. *Journal of Multivariate Analysis, 165*, 243-261. doi: 10.1016/j.jmva.2017.12.006

Neelaveni, C., & Manimaran, S. (2016). Structural equation modeling for analysing the impact of quality of administrative practices in higher educational institutions. *Quality & Quantity, 50*(4), 1663-1674. doi: 10.1007/s11135-015-0227-8

Niemelä-Nyrhinen, J., & Leskinen, E. (2014). Multicollinearity in marketing models: notes on the application of ridge trace estimation in structural equation modelling. *The Electronic Journal of Business Research Methods, 12*(1), 3-15.

Pereira, G. A., Milani, L. L., & Cirillo, M. A. (2014). Uso de alguns estimadores ridge na análise estatística de experimentos em entomologia. *Revista Ceres, 61*(3), 338-342. doi: 10.1590/S0034-737X2014000300006

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, AT: R Foundation for Statistical Computing.

Tarka, P. (2018). An overview of structural equation modeling: its beginnings, historical development, usefulness and controversies in the social sciences. *Quality & Quantity, 52*(1), 313-354. doi: 10.1007/s11135-017-0469-8

Yang, M., & Yuan, K.-H. (2019). Optimizing ridge generalized least squares for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 26*(1), 24-38. doi: 10.1080/10705511.2018.1479853

Yuan, K.-H., Wu, R., & Bentler, P. M. (2011). Ridge structural equation modeling with correlation matrices for ordinal and continuous data. *British Journal of Mathematical and Statistical Psychology, 64*(Pt. 1), 107-133. doi: 10.1348/000711010X497442