La versione definitiva è disponibile sul sito dell'editore Springer:

https://link.springer.com/article/10.1007/s11135-014-0061-4

Ciavolino, E., Carpita, M. The GME estimator for the regression model with a composite indicator as explanatory variable. Qual Quant 49, 955–965 (2015). https://doi.org/10.1007/s11135-014-0061-4

The GME estimator for the regression model with a composite indicator as explanatory variable

Enrico Ciavolino & Maurizio Carpita

Abstract

We use the generalized maximum entropy (GME) estimator to take into account the measurement error in the regression model with a composite indicator, Likert-type scales based, as explanatory variable. We show that, the reliability measure of the observed composite indicator can be used to define an estimator of the error variance and the supports required by the GME approach. As well as to obtain an estimate of the slope parameter of the model, that has statistical properties similar to the classical ordinary least squares adjusted for attenuation estimator, GME approach allows to estimate the measurement error that can be used to adjust the composite indicator of the latent explanatory variable. An extensive simulation and two case studies show the usefulness of this approach.

Introduction

Psychological and social researches are often based on questionnaires with Likert-type or multi-item scales, used to obtain multiple indicators that are discrete variables. These multiple indicators are assumed to be fallible parallel measurements, and their average is a composite indicator that represents an estimate of the underlying latent variable, as it tends to cancel out the measurement errors (Carpita and Manisera 2012). In this paper we use the Generalized Maximum Entropy (GME) estimation approach to take into account the measurement error in the regression model due to the explanatory variable that is a composite indicator, based on a set of discrete indicators. We show that the reliability measure obtained from the observed composite indicator can be profitably used to define an estimator of the error variance and supports required by the GME approach. A Monte Carlo simulation study aimed at comparing the approach proposed with the Ordinary Least Squares Adjusted for attenuation (OLSA) was performed and the performance are compared in terms of the root mean squared errors (MSE), standard error, and estimation accuracy. Finally, two case studies, about the wage satisfaction and the customer satisfaction are presented, in order to give some comparative empirical results.

The paper is organised as follows: in Sect. 2, the regression model with a composite indicator as explanatory variable and the GME estimation approach are described; Sect. 3 shows the simulation study and draws some conclusions on the performance of the two approaches presented; Two empirical applications are reported in Sect. 4; conclusions and remarks are given in Sect. 5.

The regression model with a composite indicator as explanatory variable

Consider two continuous latent random variables ξξ and ηη satisfying a linear deterministic structural relationship (Al-Nasser 2005):

$$\eta = \alpha + \beta \cdot \xi \tag{1}$$

where α and β are unknown structural parameters. Usually, the slope β is of primary interest for the analysis.

The measurement model is different for the two latent variables η and ξ. For η, only one continuous variable with error is observed; instead, for ξ one can obtain from a Likert-type scale only JJ discrete indicators. To formalize this, let consider 1+J continuous random variables, with additive errors εì and δj, respectively, that are uncorrelated between them and with ξ:

$$Y = \eta + \varepsilon \qquad X_j^* = \xi + \delta_j \qquad j = 1, \ldots, J \tag{2}$$

Then, to complete the measurement model, we consider the standard assumption that from each continuous variable X∗j only a discrete variable Xj on a Likert-type scale is observed:

$$X_j = \begin{cases} 1 & \text{if } X_j^* \le \tau_1 \\ 2 & \text{if } \tau_1 < X_j^* \le \tau_2 \\ \vdots & \\ c-1 & \text{if } \tau_{c-2} < X_j^* \le \tau_{c-1} \\ c & \text{if } \tau_{c-1} < X_j^* \end{cases} \tag{3}$$

were τk are (c−1) cuts on the domain of X∗j that transform this continuous variable in a discrete one Xj, assuming that the psychological distances between the cc integers are equals (Wakita et al. 2012). Formulas (1)–(3) define the regression model with no error in the equation (Fuller 2009, par. 2.3), extended to the case of multiple discrete indicators observed for the explanatory latent variable. In the standard approach, ξ is estimated by the average of the JJ multiple indicators:

$$\hat{\xi} = \frac{1}{J} \cdot \sum_{j=1}^{J} X_j \tag{4}$$

Our goal is to obtain good estimates of (i) the unknown structural parameter β and (ii) the latent variable ξ.

Given a random sample of n observations:

$$(x_{ji}, y_i) \quad j = 1, \ldots, J \quad i = 1, \ldots, n \tag{5}$$

without consider the measurement errors in equation (2), the empirical version of the regression (1) is:

$$y_i = \alpha + \beta \cdot \hat{\xi}_i + \nu_i \quad i = 1, \ldots, n \tag{6}$$

For this model, the usual Ordinary Least Squares (OLS) estimator of the structural parameter β is obtained as:

$$\hat{\beta}_{OLS} = \frac{Cov(\hat{\xi}, y)}{Var(\hat{\xi})} \tag{7}$$

It's well known that β^OLS has a downward bias that depends on the size of the measurement error of the explanatory variable (Fuller 2009), in particular for the model (1)–(3) on the reliability of ξ^.

The ordinary least squares adjusted for attenuation (OLSA) estimator

A standard solution to estimate the parameters of the linear regression model with measurement error in the explanatory variable, is the Ordinary Least Squares Adjusted for attenuation (OLSA) estimator.

If the explanatory variable is the composite indicator ξ^ in the equation (4), its true reliability index is:

$$\kappa_\xi = \frac{Var(\xi)}{Var(\hat{\xi})} = \frac{J \cdot \rho_\xi^2}{1 + (J - 1) \cdot \rho_\xi^2} \tag{8}$$

with ρξ>0 the true linear correlation coefficient between Xj and ξ. The index κξ takes values in the interval between 0 (no reliability) and 1 (max reliability). Using the sample in (5), we can compute the mean of indicator correlations r¯x (as estimate of ρ2ξ), and obtain the estimated reliability index:

$$\hat{\kappa}_\xi = \frac{J \cdot \bar{r}_x}{1 + (J - 1) \cdot \bar{r}_x} \tag{9}$$

that is the general form of the Spearman-Brown prophecy formula, related to the Cronbach's Alpha of the classical item analysis (Bernstein 1994; formula 6.18 and 6.26). With the index κ^ξ we obtain the unbiased estimate of β using the OLS Adjusted for attenuation (OLSA) estimator (Fuller 2009):

$$\beta_{OLSA} = \frac{\ddot{\beta}_{OLS}}{\hat{\kappa}_{\xi}} \qquad (10)$$

The generalized maximum entropy (GME) estimator

When the data exist in terms of noisy observations, the GME approach proposed by Golan and Judge (1996) allows to override the distribution and additional assumptions that are made in the traditional methods. In the regression context, this approach allows to estimate at the same time all the parameters and the errors terms of the model (Ciavolino and Al-Nasser 2009; Al-Nasser 2005). In this study we have developed the GME estimator for the regression model with a composite indicator as explanatory variable described in the previous paragraph. Considering the measurement model (2)–(3), for the composite indicator (4), we have:

$$\hat{\xi} = \xi + \delta \qquad \delta = \frac{1}{J} \cdot \sum^{J} \delta_j \qquad (11)$$

Substituting η in (2) with (1) and considering (11), the following model specification is obtained:

$$y_i = \alpha + \beta \cdot (\hat{\xi}_i - \delta_i) + \epsilon_i \quad i = 1, \ldots, n \qquad (12)$$

For the model (12) the GME estimator is outlined by the reformulation of the structural parameters and the two error terms as expected values of some discrete random variables Zα, Zβ, Zδ and Zϵ:

$$y_i = \sum_{k=1}^{K} z_k^{\alpha} p_k^{\alpha} + \sum_{k=1}^{K} z_k^{\beta} p_k^{\beta} \cdot \left(\hat{\xi}_i - \sum_{h=1}^{H} z_h^{\delta} p_{ih}^{\delta} \right) + \sum_{h=1}^{H} z_h^{\epsilon} p_{ih}^{\epsilon} \qquad (13)$$

The discrete random variables are usually composed by three or five support points, symmetric around zero, with the associated probability distributions p=(pα,pβ,pδ,pϵ) which assume value in the interval (0,1)(0,1) and respect the following normalization constraints:

$$\sum_{k=1}^{K} p_k^{\alpha} = \sum_{k=1}^{K} p_k^{\beta} = 1 \qquad \sum_{h=1}^{H} p_{ih}^{\delta} = \sum_{h=1}^{H} p_{ih}^{\epsilon} = 1; \quad i = 1, \ldots, n \qquad (14)$$

The idea underling the GME method is to estimate the unknown parameters and the error terms, by maximizing the Shannon's entropy function:

$$\max H(p) = -\sum_{k=1}^{K} p_k^{\alpha} \log p_k^{\alpha} - \sum_{k=1}^{K} p_k^{\beta} \log p_k^{\beta} - \sum_{i=1}^{n}\sum_{h=1}^{H} p_{ih}^{\delta} - \sum_{i=1}^{n}\sum_{h=1}^{H} p_{ih}^{\epsilon} \qquad (15)$$

subject to the data constraints, which are represented by the re-written model in equation (13), called consistency constraint, and the normalization constraints, given by the equation (14). For more details see Golan and Judge (1996).

Given the estimated probability distributions, it is possible to derive the estimate of the parameters of the model (12) as expected values:

$$\hat{\alpha}_{GME} = \sum_{k=1}^{K} z_k^{\alpha} \hat{p}_k^{\alpha} \qquad \hat{\beta}_{GME} = \sum_{k=1}^{K} z_k^{\beta} \hat{p}_k^{\beta} \qquad (16)$$

and for the two error terms:

$$\hat{\delta}_i^{GME} = \sum_{h=1}^{H} z_h^{\delta} \hat{p}_{ih}^{\delta} \qquad \hat{\epsilon}_i^{GME} = \sum_{h=1}^{H} z_h^{\epsilon} \hat{p}_{ih}^{\epsilon} \quad i = 1, \ldots, n \qquad (17)$$

From the first estimated error term in (17) we compute the GME adjusted composite indicator as follow:

$$\hat{\xi}_i^{GME} = \hat{\xi}_i - \hat{\delta}_i^{GME} \qquad i = 1, \ldots, n \qquad (18)$$

Simulation study

With this simulation we compare the performance of different designs of the regression model (1)–(3), using different combinations of discrete variables and levels of reliability of the composite indicator used as explanatory variable. The method we have used to construct homogeneous data with a one-dimensional latent trait underlying JJ discrete variables used by Carpita and Manisera (2012) and is based on the discretization of JJ continuous variables following a multivariate standard normal distribution with equal correlations.

Discretization procedure

The JJ standard normal variables in (2) were discretized by mapping continuous intervals into cc equally spaced integer numbers using (c−1)(c−1) cuts τkτk as in (3). We considered three discretization procedures resulting from non-linear monotonic transformations and providing three distributional forms for the indicators: the one is the optimal discrete probability distribution O, which resembles the original normal distribution rather closely; the other two discretization procedures distort the normal distribution resulting in right-skewed discrete probability distribution (R, with positive skewness) and left-skewed discrete probability distribution (L, with negative skewness). Following Carpita and Manisera (2012), we chose k=5k=5 for each discrete variable in (3) with corresponding probabilities (0.11; 0.24; 0.30; 0.24; 0.11) for the O distribution. For the skewed variables we considered the following frequencies (0.45; 0.25; 0.15; 0.10; 0.05) for the R and (0.05; 0.10; 0.15; 0.25; 0.45) the LL distributions. Moreover, when the distributions of the analysed discrete variables are very different, these can be thought to be not linearly related. The adopted procedure of discretization of the standard normal distribution in the three discrete probability distributions OO, LL and RR, is represented in the Fig. 1

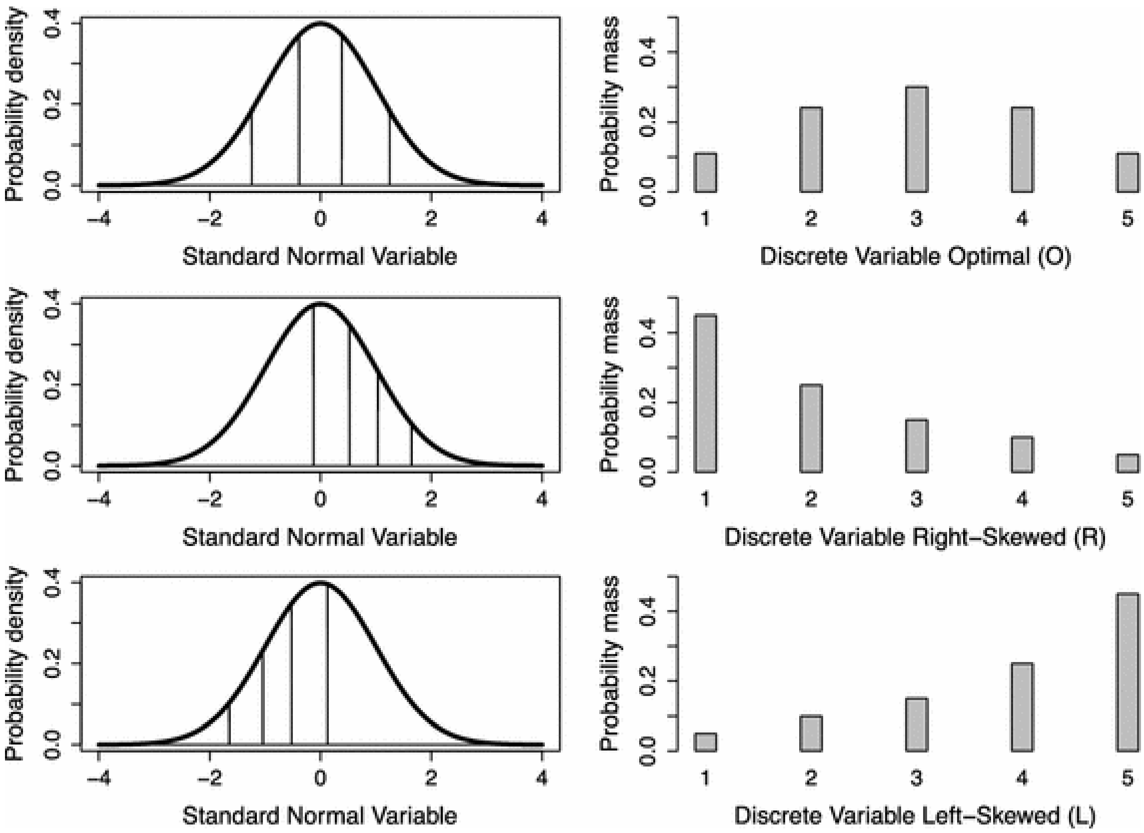

Fig. 1 Probability distributions for the Optimal(O), Right(R) and Left(L) skewed discrete variables

Simulation scenario

To compare the performance of GME and OLSA estimators for the model (1)–(3), we fixed the structural parameters to α=0 and β=0.5 and the reliability index in (8) as κξ=(0.7,0.8,0.9).

The variance of the error δj in (2) is computed as $\sigma_\delta^2 = 1 - \rho_\xi^2,$ where:

$$\rho_\xi^2 = \frac{\kappa_\xi}{J - (J-1)\cdot\kappa_\xi}$$

The variance of the error ϵ in (2) is computed as $\sigma_\epsilon^2 = \beta^2\rho_\xi^2(1/\rho_\eta^2 - 1)$ where $\rho_\eta^2 = 0.5$ is the true linear correlation coefficient between y and η. To evaluate the interaction of discrete variables with different probability distributions, we combined optimal (O), right (R) and left (L) skewed variables as in Carpita and Manisera (2012). We consider four useful scenarios of interaction (Carpita and Manisera 2012): OOOO, LLRR, LLRO and LROO. Then we choose the sample size n=100 and selected 2,000 random samples for each of these combinations.

Simulation results

The first part of Table 1 reports the simulation results for three estimators (GME, OLSA and OLS) of ββ (averages of the 2,000 replications), their standard errors (SE), the root mean of squared errors (RMSE). Results are divided according to the GME, OLSA and OLS estimators, reported in the rows, and according to different levels of reliability, in the columns and with four scenarios of interaction: OOOO, LLRR, LLRO and LROO. In the Table 2, we reported the correlation between the true Latent Varibale (LV) and the estimated LV with both the methods, in way to evaluate the accuracy, that means the ability of the GME to recover the measurement error.

Table 1 Simulation results for the GME, OLSA and OLS estimators of $\beta = 0.5$ (average of 2,000 random samples with sample size $n = 100$)

	$\kappa_\xi = 0.7$	$\kappa_\xi = 0.8$	$\kappa_\xi = 0.9$	$\kappa_\xi = 0.7$	$\kappa_\xi = 0.8$	$\kappa_\xi = 0.9$
	Scenario OOOO			Scenario LLRR		
$\hat{\beta}_{GME}$	0.449	0.449	0.445	0.470	0.467	0.469
$\hat{\beta}_{OLSA}$	0.452	0.449	0.442	0.476	0.469	0.468
$\hat{\beta}_{OLS}$	0.299	0.344	0.386	0.290	0.336	0.386
SE_{GME}	0.053	0.040	0.032	0.065	0.047	0.036
SE_{OLSA}	0.056	0.042	0.032	0.068	0.049	0.037
SE_{OLS}	0.029	0.028	0.028	0.030	0.029	0.029
$RMSE_{GME}$	0.074	0.065	0.063	0.072	0.057	0.047
$RMSE_{OLSA}$	0.074	0.066	0.066	0.072	0.058	0.049
$RMSE_{OLS}$	0.203	0.158	0.117	0.212	0.166	0.118
	Scenario RRLO			Scenario LROO		
$\hat{\beta}_{GME}$	0.463	0.463	0.461	0.459	0.456	0.458
$\hat{\beta}_{OLSA}$	0.468	0.468	0.459	0.464	0.458	0.456
$\hat{\beta}_{OLS}$	0.291	0.291	0.385	0.294	0.340	0.388
SE_{GME}	0.061	0.061	0.035	0.060	0.042	0.033
SE_{OLSA}	0.064	0.064	0.036	0.063	0.044	0.034
SE_{OLS}	0.029	0.029	0.029	0.029	0.029	0.028
$RMSE_{GME}$	0.072	0.072	0.052	0.073	0.061	0.053
$RMSE_{OLSA}$	0.072	0.072	0.055	0.073	0.061	0.056
$RMSE_{OLS}$	0.211	0.211	0.119	0.208	0.163	0.116

The OLS estimator is strongly negatively biased (about 40 %), whereas the OLSA and the GME estimators are only slightly negatively biased (no more then 10 %). The SE of the GME estimator is slightly lower (for κξ=0.7 and 0.8) or equal (for κξ=0.9) of the OLSA estimator one, so that the RMSE of these two estimators are roughly the same. Note that the RMSE of the OLS estimator is much higher of the previous two, due to the higher negative bias.

Table 2 shows remarkable differences of the GME for all the scenarios in its ability to re-construct the latent variable. The simulation results bring us to the conclusion that, in term of RMSE and standard error, there is no big differences between GME and OLSA, but the advantages can be appreciate in term of accuracy in the estimation of the latent variable.

Table 2 Correlation between the estimates $\hat{\xi}_{GME}$ and $\hat{\xi}$ and the latent explanatory variable ξ (average of 2,000 random samples with sample size $n = 100$)

	$\kappa_\xi = 0.7$	$\kappa_\xi = 0.8$	$\kappa_\xi = 0.9$	$\kappa_\xi = 0.7$	$\kappa_\xi = 0.8$	$\kappa_\xi = 0.9$
	Scenario OOOO			Scenario LLRR		
$Corr(\hat{\xi}_{GME}, \xi)$	0.912	0.927	0.945	0.908	0.921	0.940
SE	0.016	0.013	0.008	0.017	0.014	0.010
$Corr(\hat{\xi}, \xi)$	0.809	0.868	0.922	0.786	0.849	0.909
SE	0.033	0.023	0.012	0.037	0.025	0.014
	Scenario RRLO			Scenario LROO		
$Corr(\hat{\xi}_{GME}, \xi)$	0.908	0.908	0.940	0.910	0.924	0.943
SE	0.017	0.017	0.009	0.017	0.013	0.009
$Corr(\hat{\xi}, \xi)$	0.790	0.790	0.910	0.797	0.858	0.916
SE	0.035	0.035	0.014	0.034	0.024	0.013

Empirical evidences

In the following two sections the results obtaining with the proposed approach for two real case studies are presented, with different regression coefficients and levels of reliability. The first case study refers to a survey on Italian social cooperatives to evaluate the quality of work; the second one proposes the study of the several features of the products and service offered to adults and children of the Italian McDonald's restaurants. We discuss the results obtained on the real data set, and relate them to the conclusions drawn via simulation study.

ICSI example

The dataset used derives from the ICSI2007 (Indagine sulle Cooperative Sociali Italiane, 2007), a survey about social cooperatives sampled from the Istat-Census 2003 database of the Italian National Institute of Statistics, with paid workers that in 2007 answered the questionnaire designed by academic experts in economic and organisation fields with the aim to investigate the objective and subjective quality of work in the non-profit sector (Carpita and Golia 2012; Ciavolino and Nitti 2013). For this application we regress the Y = Wage Satisfaction, measured on a 7 points scale, on the composite indicator (ξ^), obtained as average of J=4 items, measured through a 5 points Likert-type scale: X1 = Responsibility;

X2 = Effort; X3 = Stress; X4 = Loyalty. The random sample of workers is equal to 100 and it is selected by considering only the graduated employed women. The frequency distributions of the 4 items are reported in the Fig. 2.

Fig. 2 Frequency distributions of the four items (X_j) used to calculate the composite indicator for the ICSI example

Table 3 ICSI example results

Correlation matrix	X_1	X_2	X_3	X_4
X_1	1			
X_2	0.510	1		
X_3	0.509	0,782	1	
X_4	0.404	0.679	0.712	1
Y	0.441	0.553	0.528	0.556
Mean of correlations $(\bar{r}_x) = 0.567$		Reliability $(\hat{\kappa}_\xi) = 0.870$		
Regression results				
$R^2_{GME} = 0.403$; $R^2_{OLSA} = 0.344$		Estimate	SE	t Stat
$\hat{\beta}_{GME}$		0.988	0.161	6.136
$\hat{\beta}_{OLSA}$		1.008	0.150	6.720
$\hat{\beta}_{OLS}$		0.878	0.131	6.702
$\hat{Corr}(\hat{\xi}_{GME}, \xi)$		0.937	0.010	
$\hat{Corr}(\hat{\xi}, \xi)$		0.902	0.015	

In Table 3 are reported the correlation matrix and the estimation results for the ICSI example. The correlations between the four items range from 0.404 to 0.783, the mean of indicator correlations between the four items (r¯x) used to compute the composite indicator (ξ^), is equal to 0.567, so that the estimated reliability (κ^ξ) defined in (9) is 0.870.

The regression model for the ICSI data has the coefficient of determinations $R^2_{GME}=0.403$ and $R^2_{OLSA}=0.344$ (obtained with the ratio 0.3/0.87 = 0.344, as we correct for attenuation due to the measurement error). The GME and OLSA estimates are both near to 1 and with a similar standard errors and t-statistics, obtained by bootstraping the sample. The OLS estimates is lower about of 12 %. We compute the correlation between the composite indicator ($\hat{\xi}$) in (4) and the GME adjusted composite indicator ($\hat{\xi}^{GME}$) in (18), obtaining a positive value equal to 0.893.

To obtain an estimate of the correlations of the composite indicators $\hat{\xi}$ and $\hat{\xi}^{GME}$ with the true explanatory latent variable ξ for this example, we use a model-based bootstrap procedure: first of all, we recover the cuts τ_k in (3), by using the quantiles of the four discrete empirical distributions in Fig. 2 then, using the model (1)-(3) we replicate 2,000 times a random sample of n=100 observations from the standard normal distribution of ξ with reliability $\kappa_\xi=0.87$ (i.e. equal to the estimated reliability for this example), $\sigma^2_\epsilon=1.545$ (i.e. the estimated variance of the GME errors), and $\beta=0.988$ (i.e. equal to the GME estimate of slope the coefficient for this example). The obtained results are reported at the bottom of Table 3: under these assumptions (in particular the measurement error equal to 13 %), we estimate that the true latent variable has correlation with the GME adjusted composite indicator that is 3.5 % points greater then the correlation with the unadjusted composite indicator (0.937 vs. 0.902).

McDonald's example

We consider a market survey promoted with the involvement of the Corporate Relations Manager of McDonald's Italia. It was carried out in 2011 on a random sample of restaurants, stratified according to their geographical location, with the aim of evaluating several features of the products and service offered to adults and children. In this study, we consider the answers to the customer experience management in fast food survey, conducted by administering, in a given week, a specific questionnaire to a random sample of adults (Dancelli et al. 2013).

Among the several questions in the questionnaire, we focus on Y == overall satisfaction, measured on a 10 scale points, on the composite indicator $\hat{\xi}$, measured through a 5 points Likert-type scale. In particular, respondents were asked to rating from 1 == Very Bad, 2 == Bad, 3 == Equal, 4 == Good and 5 == Very Good the McDonald with respect to other fast foods, considering four aspects: X1 == Products Variety; X2 == Food Taste; X3 == Quality Ingredients; X4 == Nutritional Quality.

From this survey, we extract a random sample of 100 customers: the frequency distributions of the 4 items are reported in the Fig. 3.

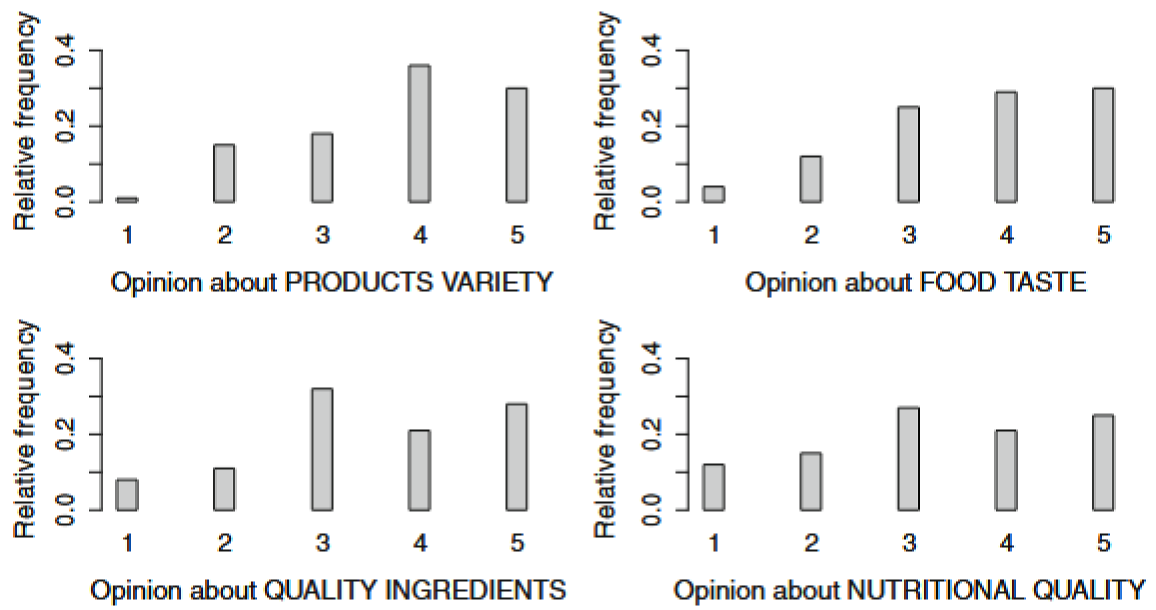

Fig. 3 Frequency distributions of the four items (X_j) used to calculate the composite indicator for the McDonald's example

Table 4 McDonald's example results

Correlation matrix	X_1	X_2	X_3	X_4
X_1	1			
X_2	0.513	1		
X_3	0.524	0,733	1	
X_4	0.416	0.633	0.540	1
Y	0.503	0.506	0.560	0.691
Mean of correlations ($\bar{r}_x$) = 0.562		Reliability ($\hat{\kappa}_\xi$) = 0.859		
Regression results				
$R^2_{GME} = 0.464$; $R^2_{OLSA} = 0.391$		Estimate	SE	t Stat
$\hat{\beta}_{GME}$		0.868	0.109	7.963
$\hat{\beta}_{OLSA}$		0.851	0.111	7.666
$\hat{\beta}_{OLS}$		0.732	0.095	7.705
$\hat{Corr}(\hat{\xi}_{GME}, \xi)$		0.932	0.011	
$\hat{Corr}(\hat{\xi}, \xi)$		0.888	0.018	

In Table 4 are reported the correlation matrix and the estimation results for the McDonald's Example. The correlations between the four items range from 0.416 to 0.733, the mean of correlations between the four items (r¯x) used to compute the composite indicator (ξ^) is equal to 0.562, so that the estimated reliability (κ^ξ), defined in (9) is 0.859.

The regression model based on the McDonald's data has the coefficient of determinations $R^2_{GME} = 0.464$ and $R^2_{OLSA} = 0.391$ (obtained with the ratio 0.336/0.859 = 0.391, as we correct for attenuation due to the measurement error). The GME and OLSA estimates are both near to 1 and with a similar standard errors and t-statistics obtained by bootstraping the sample. The OLS estimates is lower about of 12 %. We compute the correlation between the composite indicator (ξ^) in (4) and the GME adjusted composite indicator (ξ^GME) in (18), obtaining a positive value equal to 0.869: this evidence indirectly confirms the simulation results, where the true LV are correctly estimated by using the GME.

The estimate of the correlations of the composite indicators ξ^ and ξ^GME with the true explanatory latent variable ξ for the McDonald's example are obtained with the same bootstrap procedure reported in the above section: we recover the cuts τk in (3), by using the quantiles of the four discrete empirical distributions in Fig. 3 then, using the model (1)–(3) we replicate 2,000 times a random sample of n=100 observations from a standard normal distribution with κξ=0.859, σ2ϵ=0.858 and β=0.868, estimated from the sample as explained in the ICSI example. The final results are reported at the bottom of Table 4: under these assumptions (measurement error equal to 14 %), we estimate that the latent variable has correlation with the GME adjusted composite indicator that is 3.4 % points greater then the correlation with the unadjusted composite indicator (0.932 vs. 0.888).

Conclusion and further remarks

The purpose of this paper was to extend the regression model with a composite indicator and affected by measurement errors by adopting the GME estimator for the case of discrete data (Likert-type scales). The idea was to incorporate external information about the reliability of the composite indicator by the definition of the GME errors structure. By means of a Monte Carlo simulation study the different approaches have been compared in terms of standard error, root mean square errors and estimation accuracy of the latent variable.

The four scenarios of interaction and two case studies show there are no significant differences between the GME and OLSA in term of RMSE and standard error, although GME gives better results for all the simulation conditions. The main differences can be found in the reconstruction of the latent variable, where the GME gives better results also with a reliability of the composite indicator equal to 0.9.

References

- Al-Nasser, A.D.: Entropy type estimator to simple linear measurement error models. Aust J Stat 34(3), 283–294 (2005)
- Bernstein, I.: Psychometric Theory 3E. McGraw-Hill Series in Psychology, New York (1994)
- Carpita, M., Golia, S.: Measuring the quality of work: the case of the Italian social cooperatives. Quality Quantity 46(6), 1659–1685 (2012)
- Carpita, M., Manisera, M.: Constructing indicators of unobservable variables from parallel measurements. Electron. J. Appl. Stat. Anal. 5(3) (2012). http://siba-ese.unisalento.it/index.php/ejasa/article/view/12146

- Ciavolino, E., Al-Nasser, A.: Comparing generalised maximum entropy and partial least squares methods for structural equation models. J. Monparametr. Stat. 21(8), 1017–1036 (2009)
- Ciavolino, E., Nitti, M.: Simulation study for pls path modelling with high-order construct: a job satisfaction model evidence. In: Advanced Dynamic Modeling of Economic and Social Systems, pp. 185–207. Springer, Berlin (2013)
- Dancelli, L., Manisera, M., Vezzoli, M.: Interpreting clusters and their bipolar means: a case study. Statistica & Applicazioni 11(1), 49–62 (2013)
- Fuller, W.A.: Measurement Error Models, vol. 305. Wiley, New York (2009)
- Golan, A., Judge, G.: Maximum Entropy Econometrics: Robust Estimation with Limited Data. Wiley, New York (1996)
- Wakita, T., Ueshima, N., Noguchi, H.: Psychological distance between categories in the likert scale comparing different numbers of options. Educ. Psychol. Meas. 72(4), 533–546 (2012)