MEM and SEM in the GME framework: Statistical Modelling of Perception and Satisfaction

Maurizio Carpita, Enrico Ciavolino

MEM and SEM in the GME framework: Statistical Modelling of Perception and Satisfaction[*]

Maurizio Carpita[†a] and Enrico Ciavolino[b]

[a]*Department of Economics and Management, University of Brescia, Italy*
[b]*Department of History, Society and Human Studies, University of Salento, Italy*

This paper presents a review of the original method recently developed by the authors with the Generalized Maximum Entropy (GME) estimator for the simple linear Measurement Error Model (MEM) and the Structural Equation Model (SEM). In socio-economic research, these two models often concern subjective or psychological variables (composite indicators), and represent relations between latent variables. In this review, two applications to the statistical modelling of economic perception and job satisfaction are presented.

keywords: Generalized Maximum Entropy Estimator; Measurement Error Model; Structural Equation Model; Item Analysis; Eurobarometer; Job Satisfaction.

1 Prologue

The measurement error is a well-know problem encountered in many socio-economic statistical analyses (Wansbeek and Meijer (2000)). In this research field, collected data are often based on questionnaires with Likert-type scales administer to many subjects,

used to obtain multiple indicators that are discrete variables; then, these variables are averaged to compute composite indicators, that are estimates of the underlying latent variables.

In this paper we presents a short review of the original method recently developed with the Generalized Maximum Entropy (GME) estimator for the simple linear Measurement Error Model (MEM) and the Structural Equation Model (SEM). These two models often concern subjective or psychological variables (the composite indicators) and represent relations between many latent variables. As well as to obtain an estimate of the parameters of these models considering the measurement errors (reliabilities) of the composite indicators, the GME approach can adjust them, allows to obtain estimates with higher correlation with the related latent variables. Hovever, unresolved drawback of this approach are the computational complexities for models with many parameters and errors, and its scalability for big data sets.

This paper is organised as follows. In paragraph 2, the GME estimation approach for the simple linear regression model is described. In paragraph 3 the GME estimation approach for the MEM with one composite indicator as explanatory variable is presented, together with a simulation and an example of economic perception from the Eurobarometer Survey. In paragraph 4 the GME estimator for the SEM with many composite indicators for many latent variable is explained, together with a simulation and an example in the case of job satisfaction from a survey on the workers of the Italian Social Cooperatives. Conclusions are given in paragraph 5.

2 Introducing the Generalized Maximum Entropy (GME) Estimator

The GME estimator was proposed by Golan et al. (1996) to estimate parameters in cases of ill-posed problems (e.g., short and fat matrices, multi-collinearity, etc) or when some prior-knowledge on the parameters are known. The formalization of the GME approach could be made by considering the following simple regression model for the ith unit of a sample with n observations:

$$y_i = \alpha + x_i\beta + \epsilon_i \qquad i = 1, \ldots, n \tag{1}$$

The idea underlying the GME estimator consists to rewrite the regression coefficients α and β as expected values of two random variables Z^α and Z^β:

$$\alpha = \sum_{k=1}^{K} z_k^\alpha p_k^\alpha \qquad\qquad \beta = \sum_{k=1}^{K} z_k^\beta p_k^\beta$$

where z_k^α and z_k^β are the supports, symmetric around zero, while p_k^α and p_k^β are the probabilities associated (usually $2 \leq K \leq 7$). The error terms ϵ_i can also be written in the same way, considering the random variables Z_i^ϵ:

$$\epsilon_i = \sum_{h=1}^{H} z_{ih}^{\epsilon} p_{ih}^{\epsilon} \qquad i = 1, \dots, n$$

where z_{ih}^{ϵ} are the supports, symmetric around zero, while p_{ih}^{ϵ} are the probabilities associated (usually $2 \leq H \leq 7$). The choice of the z_k^{α}, z_k^{β} and z_{ih}^{ϵ} have an important role in the probability estimation procedure. These vectors could be shaped starting from particular prior information or rather could be chosen ad-hoc: for example, by considering z_{ih}^{ϵ}, the choice could be the adoption of the three-sigma-rule (Pukelsheim (1994)). For the details, see the next paragraph.

The GME method, therefore, allows to estimate the regression coefficients and the error terms, by recovering the probability distribution of a discrete random variables set. The parameter estimates are obtained by the maximization of the following Shannon's entropy function (Shannon et al. (1949)):

$$H(\mathbf{p}^{\alpha}, \mathbf{p}^{\beta}, \mathbf{p}^{\epsilon}) = -\sum_{k=1}^{K} p_k^{\alpha} \ln p_k^{\alpha} - \sum_{k=1}^{K} p_k^{\beta} \ln p_k^{\beta} - \sum_{i=1}^{n} \sum_{h=1}^{H} p_{ih}^{\epsilon} \ln p_{ih}^{\epsilon}$$

subject to the *consistency constraints* that represent the re-parametrization of the regression model (1):

$$y_i = \sum_{k=1}^{K} z_k^{\alpha} p_k^{\alpha} + x_i \sum_{k=1}^{K} z_k^{\beta} p_k^{\beta} + \sum_{h=1}^{H} z_{ih}^{\epsilon} p_{ih}^{\epsilon} \qquad i = 1, \dots, n$$

and the following *normalization constraints*:

$$\sum_{k=1}^{K} p_k^{\alpha} = 1 \qquad \sum_{k=1}^{K} p_k^{\beta} = 1 \qquad \sum_{h=1}^{H} p_{ih}^{\epsilon} = 1 \qquad i = 1, \dots, n$$

The main advantages of using GME estimation method are its desirable properties, which can be briefly summarized in the following points (Golan et al. (1996); Golan (2008)):

- The GME approach uses all the data points and does not require restrictive moments or distributional error assumptions;

- Thus, unlike other estimators, the GME is robust for a general class of error distributions;

- The GME estimator may be used when the sample is small, where there are many covariates, and when the covariates are highly correlated;

- Moreover, using the GME method, it is easy to impose nonlinear and inequality constraints.

Therefore the GME works well in case of ill-behaved data, where the estimators, like for instance the Maximum Likelihood Estimation, cannot proceed. However, some drawbacks can be considered: in particular, the GME estimator is cumbersome with many parameters and errors and it's not very suitable for big data problems.

3 The GME estimator for the Measurement Error Model (MEM)

The measurement error that is present in subjective or psychological variables is one of the most frequent (sometimes ignored) problems in statistical analyses. Let consider the following *linear deterministic structural relationship*:

$$\eta_i = \alpha + \xi_i\beta \qquad i = 1, \dots, n \tag{2}$$

where η and ξ are two latent random variables and α and β are unknown structural parameters. To estimate these parameters it is possible to use multiple indicators, i.e. observe the realization of $1 + J$ random variables, with additive errors ϵ and δ_j respectively, that are uncorrelated between them and with the latent variable ξ:

$$y_i = \eta_i + \epsilon_i \qquad x_{ij} = \xi_i + \delta_{ij} \qquad j = 1, \dots, J \qquad i = 1, \dots, n \tag{3}$$

These $1 + J$ reflective relations are represented in the figure 1. In the simpler case, the latent variable ξ can be estimated with the average of $X's$ as follow:

$$\hat{\xi}_i = \sum_{j=1}^{J} x_{ij}/J \qquad i = 1, \dots, n \tag{4}$$

From the equation (3) on the right and the average (4), we can write:

$$\xi_i = \hat{\xi}_i - \delta_i \qquad i = 1, \dots, n \tag{5}$$

with $\delta_i = \sum_j \delta_{ij}/J$. Considering the ith unit of a sample with n observations and by replacing the equations (2) and (5) in the equation (3) on the left, we can obtain the following specification of the MEM:

$$y_i = \eta_i + \epsilon_i = \alpha + \xi_i\beta + \epsilon_i = \alpha + (\hat{\xi}_i - \delta_i)\beta + \epsilon_i \qquad i = 1, \dots, n$$

A standard solution to the parameter estimation problem is the Ordinary Least Square Adjusted for attenuation (OLSA) estimator (Fuller (1987)):

$$\hat{\beta}_{OLSA} = \hat{\beta}_{OLS}/\hat{\kappa}_\xi$$

where $\hat{\kappa}_\xi$ is the estimated reliability index, defined as follow:

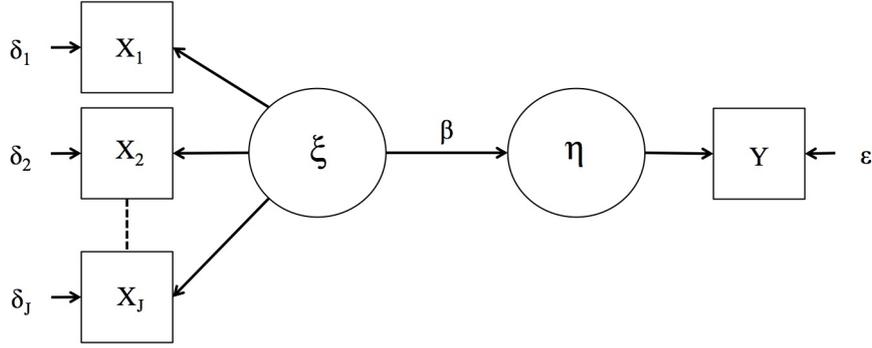$$\hat{\kappa}_\xi = \frac{J\bar{r}_x}{1 + (J-1)\bar{r}_x}$$

Figure 1: Path diagram of the Measurement Error Model (MEM)

with $\bar{r}_x$ the average correlation of $X's$. The GME estimator is outlined by the reformulation of the intercept and slope coefficients and the two error terms in form of expected values of the discrete random variables Z^α, Z^β, Z^δ, Z_i^ϵ:

$$y_i = \sum_{k=1}^{K} z_k^\alpha p_k^\alpha + \left(\hat{\xi}_i - \sum_{h=1}^{H} z_h^\delta p_{ih}^\delta \right) \sum_{k=1}^{K} z_k^\beta p_k^\beta + \sum_{h=1}^{H} z_h^\epsilon p_{ih}^\epsilon \qquad i = 1, \ldots, n \qquad (6)$$

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

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

subject to the consistency constraints, which are represented by the re-written model in equation (6), and adding up the normalization constraints. Note that, with the MEM-GME approach we obtain an error estimate $\hat{\delta}^{GME}$, and therefore the estimate $\hat{\xi}^{GME} = \hat{\xi} - \hat{\delta}^{GME}$ of the latent variable ξ.

3.1 A simulation for the MEM-GME estimator in the case of four optimal discrete variables

Since the analysis of the perception or satisfaction are usually made by using questionnaire based on Likert-scale, in this section we compare the performance of the estimator based on the GME, contrasted with OLSA, using different levels of reliability for the composite indicator used as explanatory variable.

The method we have used to construct homogeneous data with a one-dimensional latent trait underlying J discrete variables is based on the discretization of J continuous variables following a multivariate standard normal distribution with equal correlations. The approach followed for the discretization procedure is called *optimal discrete probability distribution*, which resembles the original Normal distribution, and it is indicated

with the letter O. Following Carpita and Manisera (2012), we chose $k = 5$ for each discrete variable with corresponding probabilities (0.11; 0.24; 0.30; 0.24; 0.11). The latent probability distribution and the discrete variable optimal O are represented in the figure 2.
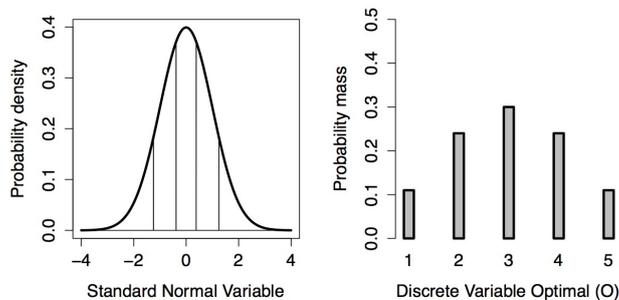

Figure 2: From the Standard Normal probability distribution to the discrete variable optimal O

To compare the performance of GME and OLSA estimators for the model (2)-(3), we fixed the structural parameters to $\alpha = 0$ and $\beta = 0.5$, $J=4$, $n=30$, $\sigma_\epsilon^2=0.2$ and $\kappa_\xi = (0.7, 0.8, 0.9)$.

Table 1: Simulation results (average of 2,000 replications) of the three estimators (MEM: $\alpha = 0$, $\beta = 0.5$, $J=4$, $n = 30$, $\sigma_\epsilon^2=0.2$)

	$\kappa_\xi = 0.7$	$\kappa_\xi = 0.8$	$\kappa_\xi = 0.9$
$\hat{\beta}_{GME}$	0.452	0.442	0.446
$\hat{\beta}_{OLSA}$	0.465	0.444	0.440
$\hat{\beta}_{OLS}$	0.291	0.332	0.379
SE_{GME}	0.104	0.075	0.057
SE_{OLSA}	0.134	0.083	0.060
SE_{OLS}	0.052	0.053	0.049
$RMSE_{GME}$	0.114	0.095	0.079
$RMSE_{OLSA}$	0.138	0.100	0.085
$RMSE_{OLS}$	0.215	0.177	0.131

Considering the summary in table 1 (bootstrap estimates, standard errors - SE and root mean square errors - RMSE of the estimators of β), the simulation shows that, on the contrary to the classic OLS estimator which has a bias very high (between 22% and 40%, increasing with the measurement error), the bias of the GME estimator is

Table 2: Simulation results (average of 2,000 replications) for the correlation of $\hat{\xi}_{GME}$
and $\hat{\xi}$ with ξ (MEM: $\alpha = 0$, $\beta = 0.5$, $J=4$, $n = 30$, $\sigma_\epsilon^2=0.2$)

	$\kappa_\xi = 0.7$	$\kappa_\xi = 0.8$	$\kappa_\xi = 0.9$
$Corr(\hat{\xi}_{GME}, \xi)$	0.884	0.899	0.922
SE	0.033	0.027	0.016
$Corr(\hat{\xi}, \xi)$	0.790	0.845	0.901
SE	0.060	0.042	0.022

relatively low and constant (about 5%) similar to the OLSA estimator. The RMSE of the estimators GME and OLSA are practically the same. Table 2 shows that the estimated $\hat{\xi}_{GME}$ is more correlated with the latent variable ξ with respect to the simple average $\hat{\xi}$. Because the GME correction, the value of the correlation coefficient increases from 2.4% (with an error of rate of 10%) to 12.7% (with a measurement error of 30%).

3.2 An application for the MEM-GME estimator in the case of the Eurobarometer perception

The empirical evidence proposed in this section concerns the study of the public opinion in the European Union on four socio-economic aspects and how these perceptions could be related to the future Gross Domestic Product. The data refers to 27 European Union (EU) Countries on four socio-economic questions (the $X's$ of the survey Eurobarometer Standard (2012)), and sample of Gross domestic product per capita for the year 2013 (the Y provided by European Commission (2013)). The respondents judge the current situation in each of his/her Country, for the following aspects:

- X_1 - The situation of the (NATIONALITY) economy

- X_2 - The situation of the European economy

- X_3 - The situation of the economy in the world

- X_4 - Your personal job situation

The response scale is: *1=very bad, 2=rather bad, 3=rather good, 4=very good*, so that, for each of 27 Countries, there is a frequency distribution on four categories. To obtain four optimal discrete variables as in figure 2, for each variable X we sorted the country means, and discretized them on 1-5 points using the absolute frequency bins (3; 6; 9; 6; 3) that close resemble the optimal relative frequencies (0.11; 0.24; 0.30; 0.24; 0.11). Finally, with these four indicators we have compute the composite indicator (4).

Figure 3 sums up the estimation results, obtained after the standardization of the dependent and indipendent variable of the MEM. In the first part is reported the correlation matrix of the $X's$ variables and also the correlation with the Gross Domestic

Correlation Matrix					
	Y	X_1	X_2	X_3	X_4
Y	1.000				
X_1	0.399	1.000			
X_2	0.663	0.361	1.000		
X_3	0.706	0.472	0.778	1.000	
X_4	0.358	0.806	0.167	0.417	1.000
$\bar{r}_x = 0.654$			$\hat{\kappa}_\xi = 0.777$		

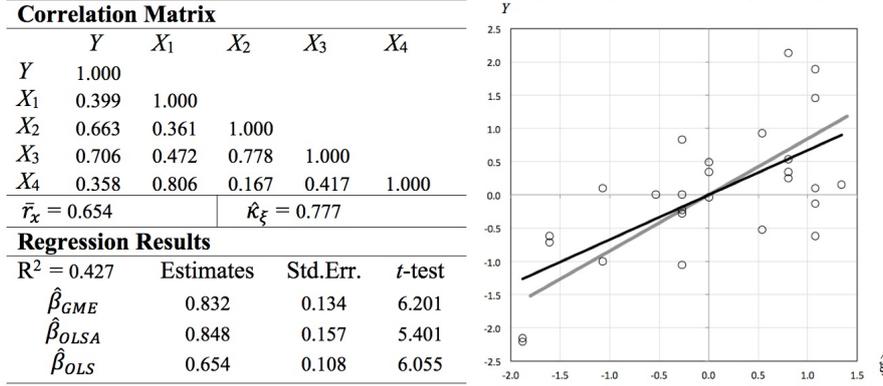| Regression Results			
$R^2 = 0.427$	Estimates	Std.Err.	*t*-test
$\hat{\beta}_{GME}$	0.832	0.134	6.201
$\hat{\beta}_{OLSA}$	0.848	0.157	5.401
$\hat{\beta}_{OLS}$	0.654	0.108	6.055

Figure 3: The Eurobarometer perception results obtained with the MEM

Product. The mean of the correlation is 0.654 with a corresponding Cronbach's Alfa equal to 0.777, so that the measurement error is 22.3%. The regression coefficient β and the standard errors are obtained via bootstrap with 2000 replications and samples size equal to 27. The regression coefficients estimated with the two methods show there is no difference between the GME and OLSA, but in terms of standard errors, GME has smaller standard error. OLS and GME regression lines are reported in the figure respectively in black and in grey colors.

4 The GME estimator for the Structural Equation Model (SEM)

The SEM can be seen as a generalization of the MEM by obtaining a more complex system of relation among the variables and relative measurement errors. The SEM consists of two main parts (Bollen (1998)):

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\tau}$$
$$\mathbf{y} = \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\epsilon}$$
$$\mathbf{x} = \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta}$$

The first part is the Structural Model and represents the linear relationships among the latent variables, where $\boldsymbol{\eta}$ is the vector of the m endogenous latent variables, $\boldsymbol{\xi}$ is the vector of the n exogenous latent variables, $\mathbf{x}$ and $\mathbf{y}$, are respectively the vectors of the q and p manifest exogenous and endogenous variables; the coefficient matrices $\mathbf{B}$ and $\boldsymbol{\Gamma}$ contain the path coefficients respectively of the effects of the endogenous on the endogenous variables and the effects of exogenous on the endogenous variables. The second part is the Measurement Model and represents the relationships between the manifest and latent variables, endogenous and exogenous; the coefficient matrices $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Lambda}_x$ measure the causal relationships of the manifest from the latent variables, both

endogenous and exogenous, respectively. The vectors $\boldsymbol{\tau}$, $\boldsymbol{\epsilon}$ and $\boldsymbol{\delta}$ are the structural and measurement error vectors. As showed for the MEM, the GME estimator is defined by the re-parameterization of the unknown parameters and the disturbance terms as a convex combination of expected values of discrete random variables as follow (Ciavolino and Al-Nasser (2009)):

$$\mathbf{B} = \mathbf{Z}^B \mathbf{P}^B, \boldsymbol{\Gamma} = \mathbf{Z}^\Gamma \mathbf{P}^\Gamma, \boldsymbol{\Lambda}_y = \mathbf{Z}^{\Lambda_y} \mathbf{P}^{\Lambda_y}, \boldsymbol{\Lambda}_x = \mathbf{Z}^{\Lambda_x} \mathbf{P}^{\Lambda_x}, \boldsymbol{\tau} = \mathbf{Z}^\tau \mathbf{P}^\tau, \boldsymbol{\epsilon} = \mathbf{Z}^\epsilon \mathbf{P}^\epsilon, \boldsymbol{\delta} = \mathbf{Z}^\delta \mathbf{P}^\delta$$

The parameters and the error terms estimation is obtained by the maximization of the following entropy function:

$$\max\{H(\mathbf{p}, \mathbf{p}_e)\} = -\mathbf{p}' \ln \mathbf{p} - \mathbf{p}'_e \ln \mathbf{p}_e$$

where $\mathbf{p}' = [\mathbf{p}'_B, \mathbf{p}'_\Gamma, \mathbf{p}'_{\Lambda_y}, \mathbf{p}'_{\Lambda_x}]$ and $\mathbf{p}'_e = [\mathbf{p}'_\tau, \mathbf{p}'_\epsilon, \mathbf{p}'_\delta]$ are vectors containing the *vec* operations of the coefficients and error terms matrices. The maximization is subjected to the consistency constraint, obtained as unique SEM function:

$$\boldsymbol{y} = \boldsymbol{\Lambda}_y \left(\boldsymbol{I} - \boldsymbol{B}\right)^{-s} \left[\boldsymbol{\Gamma}\boldsymbol{\Lambda}_x^{-1}\left(\boldsymbol{x} - \boldsymbol{\delta}\right) + \boldsymbol{\tau}\right] + \boldsymbol{\epsilon} \tag{7}$$

Adding up normalization constraints which guarantee the sum of each coefficient and the error terms probability vector equal to 1. A problem for the GME estimation is to define the support point of the SEM errors.

4.1 A simulation for the SEM-GME estimator in the case of five latent variables

Several simulation studies have been conducted to evaluate the performance of the GME estimator for SEM, on different experimental conditions (Ciavolino and Al-Nasser (2009)), or as extension to the spatio-temporal case (Papalia and Ciavolino (2011)), for the simultaneous equation model (Ciavolino and Dahlgaard (2009)) and for fuzzy measures models (Ciavolino and Calcagnì (2013)). The simulation reported in this section is presented in detail on (Ciavolino and Al-Nasser (2009)) and the structural model is defined by the path diagram of Figure 4, where there are three endogenous variables (Perceived Value, Customer Satisfaction and Loyalty), only two exogenous variables (Hardware and Software) and the relationships among the latent variables are shown with arrows.

The simulation study is conducted under the following assumptions:

1. Generate 500 random samples each of size 10, 20 and 40 from the given model;

2. For the random variable $X \sim N(2, 2)$, and Y is computed based on re-formulation Equation (7);

3. The error terms are generated from symmetric distributions, $\tau \sim N(0, 1)$; $\delta \sim U(0, 1)$ and $\epsilon \sim Beta(6, 6)$, the observed values are then scaled such that the error terms have expectation value zero. The parameters are initialised as follow: $\beta_1 = 0.8$; $\beta_2 = 0.7$, $\gamma_1 = 0.35$, $\gamma_2 = 0.25$, $\gamma_3 = 0.6$, $\gamma_4 = 0.45$ λ_y=all are set to 0.3 and λ_x=all are set to 0.6;
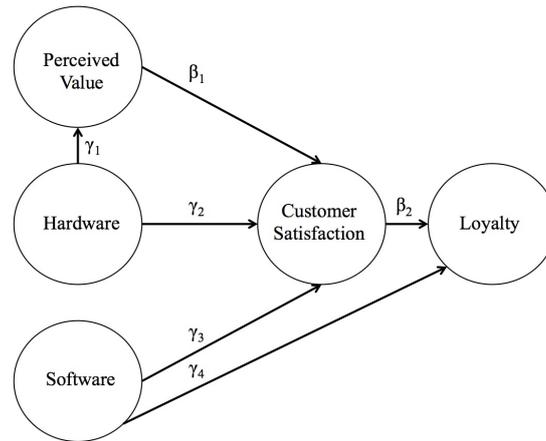
Figure 4: Path diagram for the SEM simulation study

4. The support values of the parameters (β, γ, λ_y and λ_x) in GME formulations are initialised by three data points in the interval $[100, 0, 100]$;

5. The support values of the residual terms are initialised by three data points according to the three sigma rule.

Sample size	10		20		40	
	GME	PLS	GME	PLS	GME	PLS
MSE($\hat{\boldsymbol{\beta}}_1$)	0.0639	0.4231	0.09001	0.2549	0.0489	0.0925
MSE($\hat{\boldsymbol{\beta}}_2$)	0.0400	0.3764	0.04399	0.3542	0.0338	0.1678
MSE($\hat{\boldsymbol{\gamma}}_{11}$)	0.2887	0.5492	0.10630	0.4872	0.3998	0.2154
MSE($\hat{\boldsymbol{\gamma}}_{12}$)	0.3607	0.5204	0.20336	0.4279	0.3670	0.2235
MSE($\hat{\boldsymbol{\gamma}}_{22}$)	0.0639	0.1623	0.49070	0.1257	0.0553	0.0882
MSE($\hat{\boldsymbol{\gamma}}_{32}$)	0.6114	0.7612	0.43407	0.5531	0.0226	0.3940
MSE($\hat{\boldsymbol{\lambda}}_y$)	0.1026	0.3214	0.02783	0.2167	0.0142	0.1764
MSE($\hat{\boldsymbol{\gamma}}_x$)	0.2916	0.3546	0.25493	0.2843	0.1516	0.1877

Table 3: MSE comparisons between GME and PLS estimators

The results of the simulations are showed in the table 3, where the rows show the MSE of the estimated coefficients and the columns the sample sizes, with the GME and the Partial Least Squares (PLS). The Monte Carlo results are obtained under the above assumptions by the generation of 500 random samples each of size 10, 20 and 40. For this case. Based on the simulation assumptions, the results indicate that the GME outperforms the PLS in estimating the parameters of the given model, with small sample size ($n = 10$) for all parameters estimated. Improvements with PLS appear with

some coefficients for moderate ($n = 20$) or large sample sizes ($n = 40$). In general, the GME method has better performance than the PLS in terms of MSE.

4.2 An application for the SEM-GME estimator in the case of the job satisfaction

The job satisfaction is the degree to which workers like their work, and its importance in socio-economic sciences has been recognized, as high levels of it are related to well-being, extra work effort and performance (Spector (1997); Taylor (2006); Millán et al. (2013)). Using data from a survey on the Italian Social Cooperatives carried out on about 4,000 workers over more than 300 social cooperatives (known as $ICSI^{2007}$; Carpita and Golia (2012)), we have study with the SEM-GME approach the links between two dimension of the job satisfaction (intrinsic and extrinsic) with others work quality latent factors as fairness (distributive and procedural) and motivation, that together constitute the psychological contract signed between employees and their organization (Borzaga and Depedri (2005); Borzaga and Tortia (2006)). For the estimation of the parameters of this complex model with items, we have adopting a two steps procedure (Oberski and Satorra (2013)):

1. In the first step a reliability study is conduct to obtain multidimensional measures of the subjective quality of work, starting with Nonlinear Principal Component Analysis and then using the Rating Scale Model (NPCA-RSM; Carpita and Vezzoli (2012)).

2. In the second step a substantive research is develop for the SEM with 9 composite indicators, considering their reliabilities via the GME estimator (Ciavolino et al. (2012)).

Figure 5 shows the $ICSI^{2007}$ Model with the estimates and standard errors of the path coefficients obtained using the SEM-GME approach with a random sample of $n = 360$ workers: fairness (in particular procedural fairness) has the most significant effect on both the dimensions of the job satisfaction considered, whereas motivations are not so important, especially for intrinsic job satisfaction (see Ciavolino et al. (2012) for other details).

Also in this application, to define the supports for the error terms we adopt the three-sigma-rule of Pukelsheim (1994). In particular, we use the estimate the variances of the measurement errors ϵ and δ obtained in the first step of the procedure from the reliability indices of the 9 composite indicators: table 4 shows that the estimates of these variances in the two analyses are not so different, but generally with the SEM-GME we have lower estimates of those with the NPCA-RSM.

In this study we appreciate the flexibility and more thoughtful analysis offered by the combined use of two different statistical approaches as NPCA-RSM and SEM-GME.
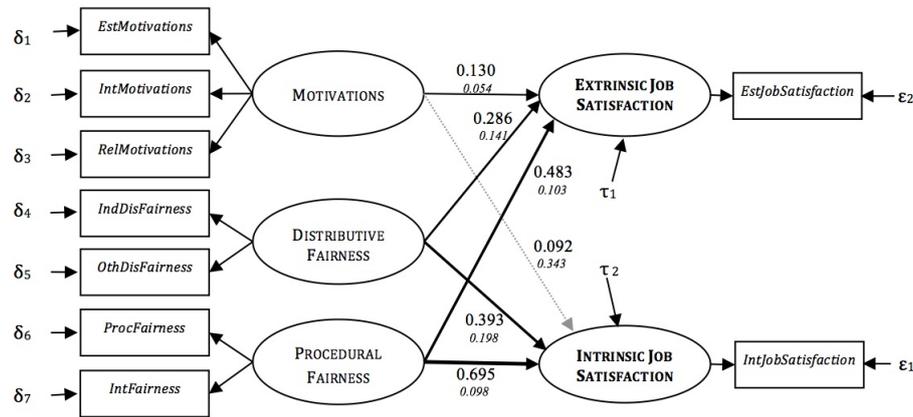
Figure 5: The $ICSI^{2007}$ path model with the structural parameter estimates obtained with the SEM-GME approach

Table 4: Estimates of the measurement error variances obtained with NPCA-RSM and SEM-GME approaches

Measurement error variances of	δ_1	δ_2	δ_3	δ_4	δ_5	δ_6	δ_7	ϵ_1	ϵ_2
NPCA-RSM	0.38	0.34	0.52	0.11	0.26	0.20	0.21	0.26	0.13
SEM-GME	0.33	0.28	0.45	0.13	0.25	0.22	0.20	0.21	0.12

5 Epilogue

In this paper we have presented a synthesis of some simulation and application results about our methodological proposal, recently developed with the Generalized Maximum Entropy (GME) estimator for the simple linear Measurement Error Model (MEM) and the Structural Equation Model (SEM). In socio-economic research, these two models represent relations between latent variables and make often use of composite indicators, obtained from subjective or psychological (discrete) data sets.

The simulation point out that the GME estimator: (1) for the MEM performs as well as the standard OLSA estimator with relatively small samples and allows a more accurate estimate of the latent variable of interest; (2) for the SEM has some advantages with respect to Partial Least Squares (PLS) estimator with small samples.

The two applications (one with economic perception data from the Eurobarometer survey, and one with job satisfaction data from the $ICSI^{2007}$ survey) confirm the simulation results: (1) the GME estimator for MEM and relatively small dataset has a lower RMSE of the OLSA estimator; (2) the GME estimator for SEM with complex structure allows to separate the preliminary reliability analysis of the composite indicators from more substantive research.

As any other estimator, also the GME estimator has some drawbacks. In particular, computational problems arise with big data sets, and for models with many parameters and errors. Our actual research on the GME estimator, as well as to extend the simulation scenario, is facing with these important application issues.

Acknowledgements

References

Bollen, K. A. (1998). *Structural equation models*. Wiley Online Library.

Borzaga, C. and Depedri, S. (2005). Interpersonal relations and job satisfaction: Some empirical results in social and community care services. *Economics and social interaction: Accounting for interpersonal relations*, pages 132–153.

Borzaga, C. and Tortia, E. (2006). Worker motivations, job satisfaction, and loyalty in public and nonprofit social services. *Nonprofit and voluntary sector quarterly*, 35(2):225–248.

Carpita, M. and Golia, S. (2012). Measuring the quality of work: the case of the italian social cooperatives. *Quality & Quantity*, 46(6):1659–1685.

Carpita, M. and Manisera, M. (2012). Constructing indicators of unobservable variables from parallel measurements. *Electronic Journal of Applied Statistical Analysis*, 5(3):320–326.

Carpita, M. and Vezzoli, M. (2012). Statistical evidence of the subjective work quality: the fairness drivers of the job satisfaction. *Electronic Journal of Applied Statistical Analysis*, 5(1):89–107.

Ciavolino, E. and Al-Nasser, A. (2009). Comparing generalised maximum entropy and partial least squares methods for structural equation models. *Journal of nonparametric statistics*, 21(8):1017–1036.

Ciavolino, E. and Calcagnì, A. (2013). A generalized maximum entropy (gme) approach for crisp-input/fuzzy-output regression model. *Quality & Quantity*, page Online First.

Ciavolino, E., Carpita, M., and Al-Nasser, A. (2012). A job satisfaction structural equation model obtained combining rasch analysis and generalized maximum entropy estimation. *Available at SSRN 1993102*.

Ciavolino, E. and Dahlgaard, J. (2009). Simultaneous equation model based on the generalized maximum entropy for studying the effect of management factors on enterprise performance. *Journal of applied statistics*, 36(7):801–815.

Eurobarometer Standard, . (2012). Public opinion in the european union. *European Economy*, 1:2012.

European Commission, . (2013). European economic forecast winter 2013. *European Economy*, 1:2013.

Fuller, W. A. (1987). *Measurement error models*. John Wiley & Sons, Inc.

Golan, A. (2008). *Information and entropy econometrics-a review and synthesis*, volume 2. Now Pub.

Golan, A., Judge, G., and Karp, L. (1996). A maximum entropy approach to estimation and inference in dynamic models or counting fish in the sea using maximum entropy. *Journal of Economic Dynamics and Control*, 20(4):559–582.

Millán, J. M., Hessels, J., Thurik, R., and Aguado, R. (2013). Determinants of job satisfaction: a european comparison of self-employed and paid employees. *Small Business Economics*, 40(3):651–670.

Oberski, D. and Satorra, A. (2013). Measurement error models with uncertainty about the error variance. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(3):409–428.

Papalia, R. B. and Ciavolino, E. (2011). Gme estimation of spatial structural equations models. *Journal of classification*, 28(1):126–141.

Pukelsheim, F. (1994). The three sigma rule. *The American Statistician*, 48(2):88–91.

Shannon, C., Weaver, W., Blahut, R., and Hajek, B. (1949). *The mathematical theory of communication*, volume 117. University of Illinois press Urbana.

Spector, P. E. (1997). *Job satisfaction: Application, assessment, causes, and consequences*, volume 3. Sage.

Taylor, M. P. (2006). Tell me why i don't like mondays: investigating day of the week effects on job satisfaction and psychological well-being. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169(1):127–142.

Wansbeek, T. J. and Meijer, E. (2000). *Measurement error and latent variables in econometrics*, volume 37. Elsevier Amsterdam.