Modelling with structural equation modelling – application and issues

Aleksandar Krstev*, Dejan Krstev**, Radmil Polenakovik**
*Faculty of Computer Science, Goce Delcev University, Stip, Macedonia
**Faculty of Mechanical Engineering, Ss. Cyril and Methodius University, Skopje, Macedonia – PhD Student
aleksandar.krstev@ugd.edu.mk, krstev.deni@gmail.com

Abstract: Structural equation modeling (SEM) is a comprehensive statistical modeling tool for analyzing multivariate data involving complex relationships between and among variables. SEM surpasses traditional regression models by including multiple independent and dependent variables to test associated hypothesizes about relationships among observed and latent variables. SEM explain why results occur while reducing misleading results by submitting all variables in the model to measurement error or uncontrolled variation of the measured variables. SEM provides a way to test the specified set of relationships among observed and latent variables as a whole, and allow theory testing even when experiments are not possible. Structural Equation Modeling (SEM) is a powerful collection of multivariate analysis techniques, which specifies the relationships between variables through the use of two main sets of equations: Measurement equations and structural equations. Measurement equations test the accuracy of proposed measurements by assessing relationships between latent variables and their respective indicators. The structural equations drive the assessment of the hypothesized relationships between the latent variables, which allow testing the statistical hypotheses for the study. Additionally, SEM considers the modeling of interactions, nonlinearities, correlated independents, measurement error, correlated error terms, and multiple latent independents each measured by multiple indicators.

In this paper will be presented application of relationship between reverse logistics and circular economy using some SEM fit indexes. The process of validating the measurement model requires testing each cluster of observed variables separately to fit the hypothesized CFA model. The statistical test uses the most popular procedures of evaluating the measurement model: Chi-square CMIN (χ2), Goodness-of-Fit Index (GFI), and Percent Variance Explained.

Keywords: SEM, statistics, social sciences, reverse logistic, circular economy

I. INTRODUCTION

Structural Equation Modelling (SEM) is an extension of the general linear model (GLM) that allows the researcher to simultaneously test a set of regression equations. In other words, the purpose of SEM is to examine the set of relationships between one or more exogenous variables (independent variables) and one or more endogenous variables (dependent variables). SEM software can test traditional models, but also allows the examination of more complex relationships and models, such as confirmation of factor analysis and time series analysis. In addition, SEM structural relationships can be graphically modeled to provide a clear understanding of the theory under consideration. Compared to the old multi-variant procedures, several advantages can be noted when using SEM. It performs a verified, rather than a research approach to data analysis (research approach can also be implemented through SEM): - SEM evaluates the parameters of the error variance, but the traditional multi-variant procedures are not able to estimate the measurement error; - SEM can include both observed and latent variables, while previous methods are based only on observed measurements; - The browser can get a unifying frame that corresponds to a number of linear models; - SEM programs provide overall model tests and individual parameter assessment tests simultaneously.

Coefficients, reactions, and variants of the reaction can be compared simultaneously, even in different groups. You can handle long-term data, databases with automatic error correction structures (time series analysis), databases with unusually distributed variables, and incomplete data. Due to these advantages of SEM, it has become a popular methodology in non-experimental research.

II. COMMON TERMS IN SEM

Especially in behavior and social sciences, researchers are often interested in studying two types of theoretical constructions, namely observed, observed (manifested) and latent variables. However, research often has to do with latent variables that cannot be directly measured, such as personality, perception, buying behavior, and so on. Research has used observed, variable to measure latent variables. Observation may include, for example, answers to self-reported attitudes, coded answers to interview questions, answers to surveys or questionnaires, and so on. These measured grades, or in other words observed (observed) or manifested variables, are used to measure latent variables.

A. Exogenous and endogenous latent variables (variable)

Exogenous latent variables are synonymous with independent variables and endogenous latent variables are synonymous with dependent variables. Endogenous variables are influenced by exogenous variables directly or indirectly.

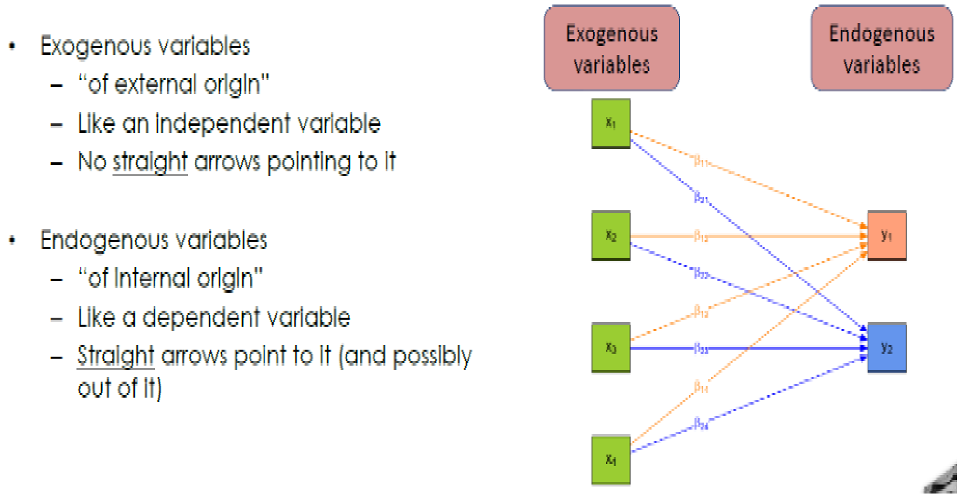

Figure 1. Exogenous and endogenous variables

B. Exploratory Factor Analysis (EFA) and Factor Analysis Certificate (CFA)

Factor analysis was performed to examine the relationships between the sets of observed and latent variables. If the relationships between the excited and latent variables are unknown or uncertain, research factor analysis is performed. Research factor analysis is performed to determine how and to what extent the observed variables are related to their underlined factors. Confirmation of factor analysis is appropriate when the researcher has some understanding (through theory, empirical research, or both) of the latent variable structure.

C. The path diagram

The road diagram is a visual representation of the relationships between variables that are assumed to be present in the study. Basically four geometric symbols are used in path diagrams; circles or ellipses (○) represent unprotected latent variables, squares or rectangles represent (□) observed variables, single-headed arrows (→) represent the effect of one variable on another variable, and double-headed arrows (↔) represent covariance or correlation . between two variables. Figure 2. is a simple model used to explain the meanings of symbols on a path diagram.

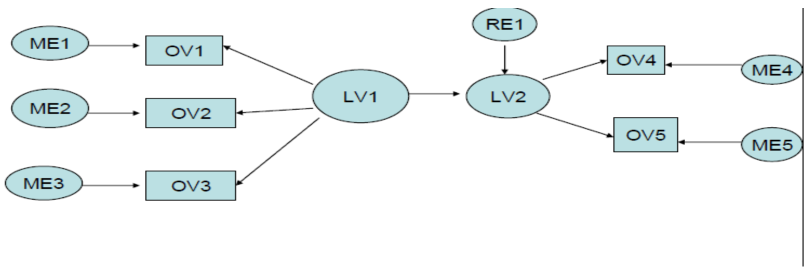

Figure 2. - Simple path diagram

In the above model, there are two latent variables (LV1 is exogenous variable and LV2 is endogenous variable) and five observed variables; three are used to measure LV1 and two are used to measure LV2. In addition, there are five measurement errors (ME 1- ME5); associated with each observed variable and one residual error associated with the predicted factor (LV2). Measurement error of the underlined factor or latent variable through the observed variable is reflected by a measurement error. The remaining error is an error in predicting an endogenous factor from an exogenous factor. For example, the residual error shown in Figure (RE1) is an error in predicting the endogenous factor (LV2) of the exogenous factor (LV1). The general SEM can be divided into two sub-models; measuring models and structural models. The measurement model shows the relationship between the observed and latent variables. In other words, it is a CFA model, specifying the model by which each measure carries a load on a particular factor. But the structural model shows the relationship between latent variables. There are two measurement models and one structural model discussed earlier. Figure shows an example of a measurement model and shows an example of a structural model and both models are sub-models derived from the model.

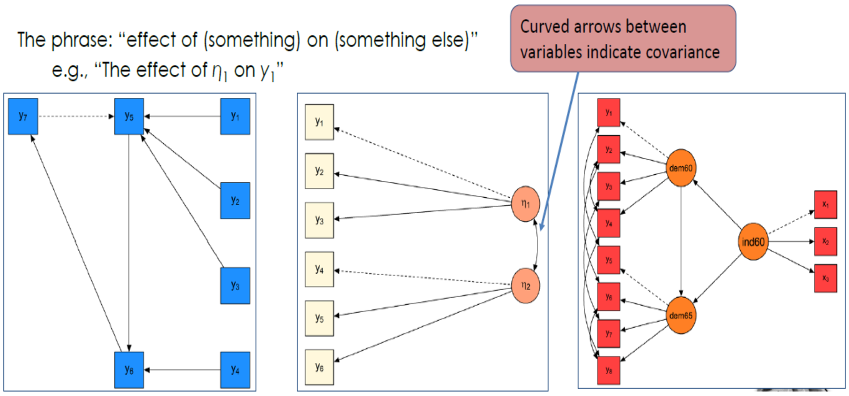

Figure 3. Structural and measurement models

III. DATA PROCESSING ACCORDING TO PART OF THE HYPOTHETICAL RESEARCH FRAMEWORK

The data processing with the SEM-Amos software will process the hypothetical frames (mix of questions) for the results of the Questionnaire on the Impact of the Reversible Logistics of the Circular Economy. The scope of the research emphasizes the interaction of the research components for which the research can be clearly understood. It also illustrates how reversible logistics relates to the Circular Economy and other concepts for achieving the research goal. The circular economy operates according to 3R (Reduce, Reuse & Recycle) - the approach of "Reduction, Reuse (Reuse) and Recycling". Recycling and Reuse (reuse) of waste or already used and obsolete or damaged products is the first major step in changing the ways of thinking of businessmen, but it also represents the overall cultural change in society. Remarkable recommendations of other managers regarding the implementation of the principles of Circular Economy and Reversible Logistics. (Indicator: answer questions no. Q1, Q5, Q6, Q10, Q11, Q 15.). Management of waste collection systems is the notion of zero waste in the company, and the Circular Economy and Reversible Logistics in general have had a positive impact on the process. (Indicator: answer questions no. Q3, Q7, Q10, Q11, Q12). According to experience, recommendations will be given to other managers regarding the implementation of the principles of Circular Economy and Reversible Logistics that would enable zero waste and environmental processes, as well as the relevance of Reversible Logistics in organizations. (Indicator: answer questions no. Q1, Q5, Q8, Q13, Q14, Q15).

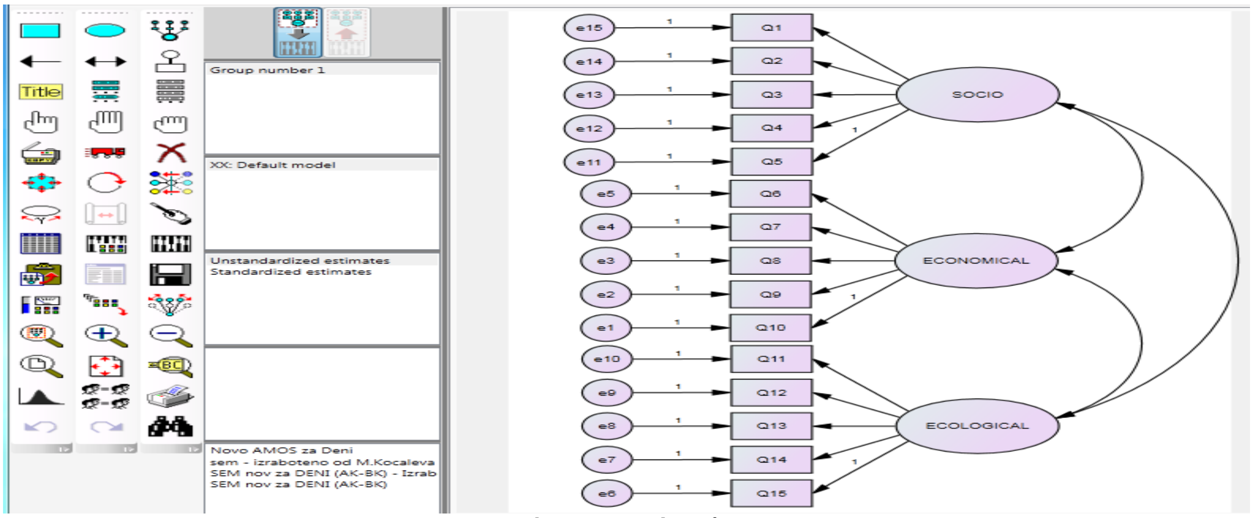

Figure 4. Schematic procedure of SEM-Amos

Table 1. Sum of answers of respondents for mix of questions

TOTAL	300	300	300	300	300	300	300	300	300	300	300
5	50/40	9/8	26/34	70/58	45/25	30/35	40/45	55/40	70/60	6/12	45/35
4	60/30	5/8	20/20	62/50	25/25	30/25	45/50	45/40	62/48	6/6	35/35
3	15/15	42/48	72/68	16/14	50/40	50/58	56/34	40/50	12/18	40/50	60/60
2	30/20	50/62	16/20	9/6	30/20	20/20	10/8	8/10	8/7	55/40	8/8
1	24/16	46/22	10/14	9/6	20/20	14/18	6/6	6/6	8/7	45/40	8/6
	Q1	**Q3**	**Q5**	**Q6**	**Q7**	**Q8**	**Q10**	**Q11**	**Q12**	**Q13**	**Q15**
5	90	17	60	128	70	65	85	95	130	18	80
4	90	13	40	112	50	55	95	85	110	12	70
3	30	90	140	30	90	108	90	90	30	90	120
2	50	112	36	15	50	40	18	18	15	95	16
1	40	68	24	15	40	32	12	12	15	85	14
Total	300	300	300	300	300	300	300	300	300	300	300

Table 2. Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	**N of Items**
.951	**.974**	**16**

Table 3. Implied Covariances (Group number 1 - Default model)

	Q15	**Q13**	**Q12**	**Q11**	**Q10**	**Q8**	**Q7**	**Q6**	**Q5**	**Q3**	**Q1**
Q15	1,268										
Q13	-,399	1,191									
Q12	,401	-,014	1,179								
Q11	,373	-,013	,013	1,181							
Q10	-,299	,011	-,011	-,010	1,164						
Q8	-,251	,009	-,009	-,008	,366	1,522					
Q7	-,024	,001	-,001	-,001	,036	,030	1,729				
Q6	-,224	,008	-,008	-,007	,327	,274	,027	1,158			
Q5	-,028	,001	-,001	-,001	,415	,348	,034	,311	1,309		
Q3	-,027	,001	-,001	-,001	,413	,347	,034	,309	,549	1,101	
Q1	-,028	,001	-,001	-,001	,418	,351	,034	,313	,555	,552	1,982

Table 4. Implied Correlations (Group number 1 - Default model)

	Q15	**Q13**	**Q12**	**Q11**	**Q10**	**Q8**	**Q7**	**Q6**	**Q5**	**Q3**	**Q1**
Q15	1,000										
Q13	-,325	1,000									
Q12	,328	-,012	1,000								
Q11	,305	-,011	,011	1,000							
Q10	-,246	,009	-,009	-,008	1,000						
Q8	-,181	,007	-,007	-,006	,275	1,000					
Q7	-,016	,001	-,001	-,001	,025	,018	1,000				

Q6	-,185	,007	-,007	-,006	,281	,206	,019	1,000			
Q5	-,021	,001	-,001	-,001	,336	,247	,022	,252	1,000		
Q3	-,023	,001	-,001	-,001	,365	,268	,024	,274	,457	1,000	
Q1	-,017	,001	-,001	-,001	,275	,202	,018	,207	,345	,374	1,000

Table 5 Residual Covariances (Group number 1 - Default model)

	Q15	Q13	Q12	Q11	Q10	Q8	Q7	Q6	Q5	Q3	Q1
Q15	-,091										
Q13	,100	,000									
Q12	-,161	,426	,000								
Q11	,021	-,089	,076	-,014							
Q10	-,035	,087	,197	-,014	,000						
Q8	-,057	-,096	-,142	,367	,006	,000					
Q7	,584	-,097	,304	,360	-,074	,526	,000				
Q6	,316	,321	,731	-,084	,068	-,195	,419	,000			
Q5	,042	,203	,352	,111	-,110	-,119	,112	,216	,000		
Q3	,082	,264	,412	,138	-,027	,138	,198	,068	-,032	,000	
Q1	,121	,312	,447	,005	-,029	-,425	-,422	,168	,153	-,053	,000

Table 6. Standardized Residual Covariances (Group number 1 - Default model)

	Q15	Q13	Q12	Q11	Q10	Q8	Q7	Q6	Q5	Q3	Q1
Q15	-,874										
Q13	1,339	,001									
Q12	-2,166	6,216	,001								
Q11	,278	-1,304	1,115	-,143							
Q10	-,485	1,275	2,913	-,203	,000						
Q8	-,700	-1,231	-1,833	4,741	,078	,000					
Q7	6,822	-1,173	3,687	4,361	-,896	5,602	,000				
Q6	4,437	4,720	10,818	-1,237	,969	-2,491	5,124	,000			
Q5	,565	2,812	4,904	1,541	-1,460	-1,419	1,287	2,937	,000		
Q3	1,194	3,993	6,258	2,092	-,389	1,777	2,485	,999	-,421	,000	
Q1	1,325	3,510	5,058	,058	-,313	-4,146	-3,939	1,882	1,555	-,581	,000

Table 7. Factor Score Weights (Group number 1 - Default model)

	Q15	Q13	Q12	Q11	Q10	Q8	Q7	Q6	Q5	Q3	Q1
Ecological	,467	,141	-,143	-,132	,104	,052	,004	,062	-,030	-,040	-,016
Economical	-,172	-,052	,052	,049	,105	,053	,004	,063	,096	,130	,051
Social	,068	,020	-,021	-,019	,134	,067	,005	,080	,193	,262	,103

Table 8. Appropriate indices (Group number 1 - Default model)

CMIN **Model**	**NPAR**		**CMIN**	**CMIN/DF**
Default model	25		30,53	3,053
Model	RMR	GFI	AGFI	PGFI
Default model	,233	,745		
Bootstrap:	,000			
Total:	,527			

All tables in the above text stems and are generated through the processing from using the data of SEM-Amos software. In the above tables are presented the following data: Reliability Statistics, Implied Covariances and Correlations, Residual and Standardized Covariances, Factor Score Weights for Social, Ecological and Economical aspects and Appropriate indices.

IV. CONCLUSION

Modeling is an integral part of the thought process. People think within certain standards and rely on them. It can be said that modeling is a rational, systematic, complex procedure of properly presenting the important features of processes, phenomena or their representations as certain units. Modeling is a systematic research procedure through which real or thought models are made, ie models of sketches, objects, mathematical formulas, etc. The structure of modeling is made up of four factors: passive objective factors-subject to modeling, active subjective factors, means or tools, position in objective reality and conditions. Various software solutions will be used to explain the circular economy, especially reversible logistics (SEM – Sequential Equation Modeling). Finally, based on the data obtained in the case of the research and the model of reversible logistics, a comparative analysis of the processes for design and development of new innovative products will be made. The respondents-respondents 300 in the survey with 15 questions, declared in accordance with each statement or question with answers "5-Excellent", "4-Very Good", "3-Good", "2-Not Enough" and "1- Very little". Of course, the reversible logistics and circular economy model will be a function of future and possible development of innovative activity. SEM can simultaneously test a complex set of regressive equations. Furthermore, SEM may include both observed and latent variables, while previous methods are based only on observed measurements. However, the most commonly applicable or acceptable values, but not always or rarely achievable in research, for fitting (Appropriate indices and Tolerance of appropriate indices) are the following with their limit measures: CMIN / DF = 2 <CMIN / DF≤ 3. Reliability statistics: Cronbach’s alpha=.951

REFERENCES

[1] Byrne, B. M. (2010). *Structural Equation Modeling with AMOS.* New York: Routledge Taylor &Francis Group.
[2] Hair, J., Hult, G., Ringle, C., & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling PLS-SEM.* Los Angeles: SAGE.
[3] Hershberger, S.L. (2003), “The growth of structural equation modeling”, Structural Equation Modeling, Vol. 10 No. 1, p. 35-46.
[4] Kline, R. (2011). *Principles and practice of structural equation modeling* (3nd Ed. b.). New York: Guilford.
[5] Raykov, T., & Marcoulides, G. (2006). *A First Course in Structural Equation Modeling.* Mahwah:awrence
Erlbaum Associates.
[6] Rigdon, E. (2012). Rethinking partial least squares path modelling: In praise of simple methods. *Long Range Planning, 45*, 341-358.