Journal of Scientific Research and Development 2 (10): 112-116, 2015
Available online at www.jsrad.org
ISSN 1115-7569

Model generation using structural equation modeling

Dr Kamran Ahmed Siddiqui [1,]*, Manzoor Ali Mirani [2], Syed Muhammad Fahim [3]

[1]*University of Dammam, Dammam, Saudi Arabia*
[2]*Universiti Tun Abdul Razak, Malaysia*
[3]*DHA Suffa University, Karachi, Pakistan*

Abstract: The major objective of this paper is to share the experiences of using Structural Equation Modeling (SEM) in marketing research. It provides an abridge version of relevant literature with one live example of using SEM in social sciences. Structural Equation Modeling (SEM) is the youngest member of statistical modeling techniques, mainly used for cross-sectional Factor analyses, Path analyses and Regression analyses. The use of SEM is growing in marketing research for many reasons. This study uses a Model Generation Strategy, considering it to be the most suitable one. All SEM analyses were conducted using AMOS which is easy-to-use but powerful structural equation modeling (SEM) software. Most of the statistical techniques including SEM are sample size sensitive. Sample size plays an important role in the estimation and interpretation of SEM results. This paper also provides the two steps in SEM (a) Measurement Models and (b) Path Models. At the end it provides the choice of fit indices and cut-off criteria. This paper provides robust learning mechanism for young social scientist to learn the basics of SEM and with the help of one live example they can develop their own models using SEM and AMOS.

Key words: Structural equation modeling; Model building; Multivariate analysis; AMOS

1. Introduction

The paper aims to share the experiences of using Structural Equation Modeling (SEM) in social research. It provides an abridge version of relevant literature with one live example of using SEM in social sciences.

2. Literature review

Structural Equation Modeling (SEM) is the youngest member of statistical modeling techniques, mainly used for cross-sectional Factor analyses, Path analyses and Regression analyses (Byrne, 1998). It is more suitable to test a conceptual or theoretical model rather than exploratory studies. An informative presentation of SEM results should contain complete disclosure of the strategy, methods of estimation, and fit statistics with cut-off criteria, measurement models, and finally, path models (Hoyle & Panter, 1995).

The use of SEM is growing in marketing research for two reasons (a) separation of observational error from measurement of latent variables; (b) isolation of good indicators of the latent variables (Hancock, 2003).

SEM Strategy: Literature suggests three strategies for using Structural Equation Modeling (SEM) in any study (Jöreskog, 1993). Firstly, a strictly confirmatory strategy in which a model is tested using SEM goodness-of-fit tests to determine if the pattern of variances and covariance's in the data is consistent with a structural (path) model specified by the researcher. However, the other unexamined models may also fit the data or be better, and therefore an accepted model is normally regarded as a not-disconfirmed model. Secondly, the alternative models strategy in which more than one model is tested to determine 'the best fit model'. There are many goodness-of-fit measures, reflecting different aspects of fitness, and usually three or four are reported by the researcher. Although desirable in principle, this AM strategy runs into the real-world problem that in most specific research topic areas, the researcher does not find in the literature two well-developed alternative models to test. Thirdly, Jöreskog (1993) describes a model generation strategy in which a model is tested using SEM procedures, found to be deficient, and an alternative model is then tested based on changes suggested by SEM modification indexes. This is the most common approach found in the literature. The problem with this strategy is the confirmation of models in a manner which may not be stable (i.e. the confirmed model may not fit new data, having been created based on the uniqueness of an initial dataset). Researchers can attempt to overcome this problem by using a cross-validation strategy under which the model is developed using a calibration data sample and then confirmed using an independent validation sample (Jöreskog, 1993).

This study uses a Model Generation Strategy, considering it to be the most suitable one. All SEM analyses were conducted using AMOS which is easy-

* Corresponding Author.

to-use but powerful structural equation modeling (SEM) software.

3. Methods

Most of the statistical techniques including SEM are sample size sensitive (Siddiqui, 2013) and sample size plays an important role in the estimation and interpretation of SEM results (Hair et al., 2006). There are various heuristics for sample size determination (a) sample size should be between 100 to 200 for SEM (Loehlin, 1992); (b) the sample size should be at least 50 more than eight times the number of variables in the model (Loehlin, 1992); (c) at least 15 cases per measured variable or predicator (Stevens, 2002) (c) minimum five cases per parameter estimate are required (Bentler & Chu, 1987); (d) use of 50 variables in any single model requires a sample size of about 450 (Marsh et al., 1988); (e) for smaller sample size or excessive kurtosis in SEM, the results should be reported in maximum likelihood (ML) estimation method (Hoyle & Panter, 1995). To summarize for SEM at least 15 cases per measured variable or indicator are needed (Siddiqui, 2013). Looking at these guidelines sample size of 220 was considered as adequate.

Method of Estimation: The choice of model generation strategy must be followed by selection of the method of estimation (Hu et al., 1992). It is also important to discuss this issue before the final presentation of results as different estimation methods and fit indices may lead to different inferential outcomes when evaluating structural equation models (Hu et al., 1992).

The default method for estimating free parameters in structural equation models is maximum likelihood (ML). Several studies indicate that ML performs reasonably well under a variety of less-than-optimal analytic conditions, such as small sample size and excessive kurtosis (Hoyle 1995). Because of the availability and the wide use of ML, Hoyle and Panter (1995) suggested that researchers should always report the results of ML estimation (Hoyle & Panter (1995). This study uses ML as the major method of estimation.

Fit Indices: Another issue concerning the presentation of SEM results is the choice of fit indices and cut-off criteria. Literature suggest that results should also provide information about which indices of overall fit will be used, and should describe the characteristics of these indices prior to reporting results of structural equation modeling analysis (Tanaka, 1993; Hoyle & Panter, 1995). Furthermore, it was also suggested that fit indices information must include (a) overall fit indices along with the justification for choosing those indices (Hoyle & Panter, 1995); (b) a clear conceptual definition of each index to be reported (Tanaka, 1993); and that (c) the ‘critical value’ of each index that indicates acceptable fit should be specified prior to reporting and interpreting the values of the indices (Hoyle & Panter, 1995).

SEM fit indices have been classified into two groups: (a) absolute fit concerns the degree to which the covariance’s implied by the proposed model match the observed covariance’s, and these indices typically gauge the ‘badness of fit’; and (b) incremental fit indices which concerns the degree to which the proposed model is superior to an alternative model (e.g., a model which specifies no covariance’s amongst the variables). In contrast to the absolute fit indices, incremental fit indices typically gauge ‘goodness of fit’, i.e., larger values indicate greater improvement of the proposed model over an alternative model (Hoyle and Panter, 1995).

There is no standard rule for reporting the fit indices for evaluating structural equation models, but researchers are encouraged to report multiple indices of overall fit (Bollen 1989a; Marsh, Balla & McDonald, 1988; Tanaka 1993). It is advised that the selection should be made from different groups of fit indices; one or two from the absolute fit group and one or two from the incremental fit indices. Based on the recommendation provided by Hu and Bentler (1995), Hoyle and Panter (1995), and incorporating more definitive and stringent criteria provided by Hu and Bentler (1999) the following indices were used to evaluate the results of the current study.

χ^2: The likelihood ratio or chi-square test is generally always reported. It is one of the few indices to provide a valid significance test and it forms the basis for many other indices. It is considered as one of the most recommended absolute fit indices. Despite numerous ambiguities associated with interpreting χ^2, the value of the statistic itself holds the most promise for the development of an index of fit for which the sampling distribution is known (Hoyle and Panter, 1995). It also forms the basis for nested model comparison and these values must be accompanied by the values of degrees of freedom, and sample size. However, χ^2 is sensitive to sample size; as the sample size increases (generally above 200), the χ^2 test statistic has a tendency to indicate a significant probability level. In contrast, as the sample size decreases (generally below 100), the χ^2 test statistic is prone to indicate non-significant probability levels (Schumacker, 1996).

The Root Mean Square Error of Approximation (RMSEA) is based on the non-centrality parameter. Conventionally, there is good model fit if the RMSEA is less than or equal to .05; and an adequate fit if it is less than or equal to .08 (Schumacker and Lomax, 2004). More recently, Hu and Bentler (1999) have suggested RMSEA <= .06 as the cut-off for a good model fit (Hu & Bentler, 1999). The RMSEA is a popular measure of fit, partly because it does not require comparison with a null model (Hair et al. 1998). It has a known distribution, related to the non-central chi-square distribution, and thus does not require bootstrapping to establish confidence intervals. It is one of the fit indices less affected by sample size, though for small sample sizes it over-estimates goodness of fit (Fan et al., 1999).

The Tucker and Lewis Index (TLI) are also called the Non-Normed Fit Index (NNFI, Bentler & Bonett, 1980). The TLI/NNFI works well when ML estimation is used, though it is significantly downwardly biased when using other estimation methods and in the context of relatively small sample sizes (Hu and Bentler, 1995). For the TLI, a value close to 0.90 reflects a good model fit (Schumacker, 1996). Hu and Bentler (1999) have suggested TLI >0.95 as the cut-off for a good model fit (Hu & Bentler, 1999).

The Comparative Fit Index (CFI) is an incremental fit index and its value falls between 0 and 1 (Hu & Bentler, 1995). The CFI is somewhat preferable among other incremental fit indices as its values fall within the familiar 'normed' range. For the CFI, Schumacker (1996) also proposed that a value close to .90 can reflect a good model fit, while Hu and Bentler (1999) have suggested CFI >0.95 as the cut-off for a good model fit (Hu & Bentler, 1999).

Measurement Models: As a necessary step in structural equation modeling, Confirmatory Factor Analysis (CFA) was conducted prior to the path diagrams, mainly for two reasons. Firstly, the structural portion of a full structural equation model only involves relations among latent constructs. Secondly, the major concern for a full structural equation model is to assess the extent to which these relations are valid. Therefore, it is critical to make sure that the measurements for different latent constructs are valid (Byrne, 1998). Most importantly, it is used to separate out measurement and structural problems.

CFA for the Measurement Model – Credit Card Usage: A factor model for credit card usage was constructed in the same manner, measuring the outcomes of customer loyalty (Zeithaml et al., 1996), customer satisfaction (Lee et al. 2001), and customer switching (Zeithaml et al., 1996).

CFA for the Measurement Models - Personality Factors: Separate second order CFA analyses were carried out for the measurement of the Five Factor Model, which has five distinct factors i.e., Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience, each of which is comprised of six subordinate dimensions known as facets (Costa & McCrae, 1992).

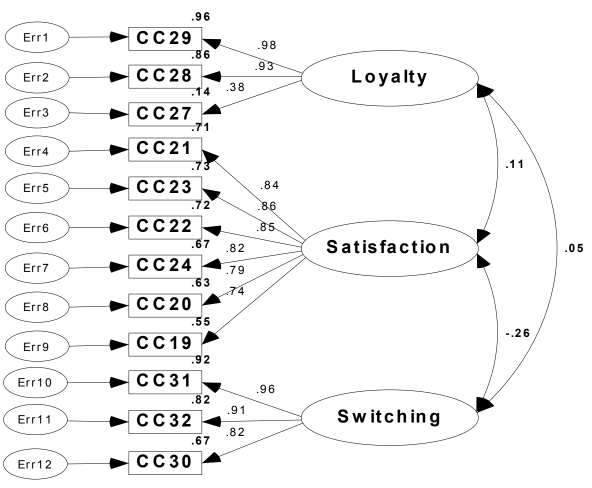

Fig. 1: Measurement model – credit card usage

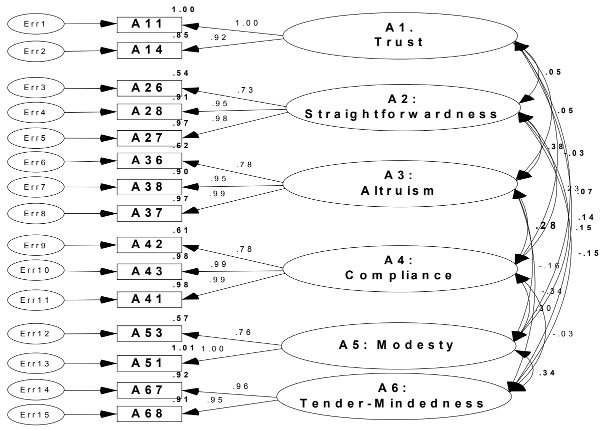

Fig. 2: Measurement model – agreeableness

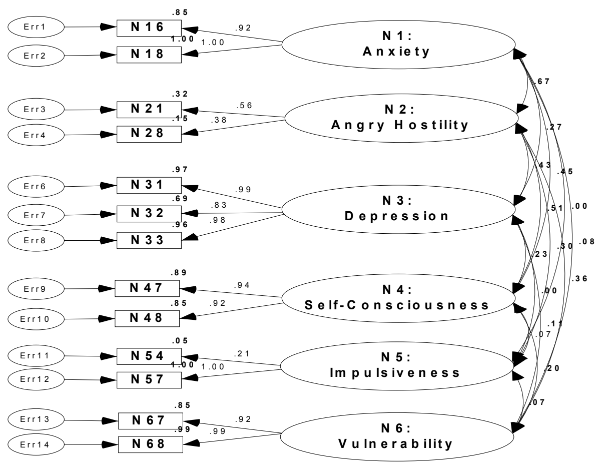

Fig. 3: Measurement model – neuroticism

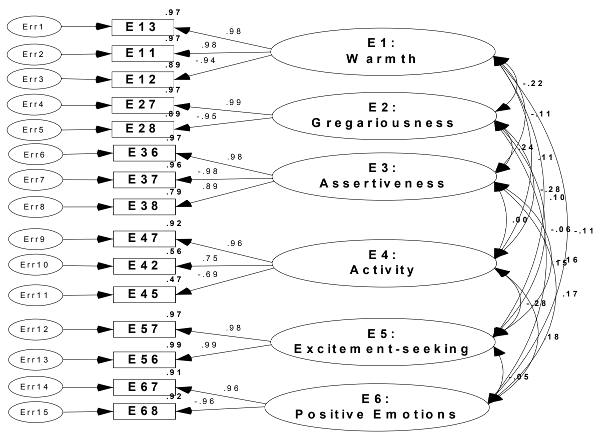

Fig. 4: Measurement model– extraversion

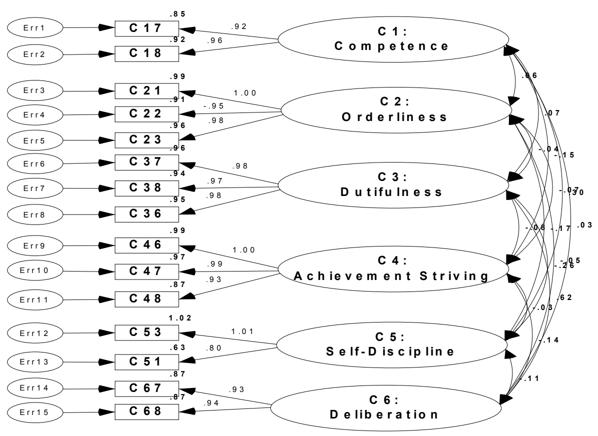

Fig. 5: Measurement model– conscientiousness

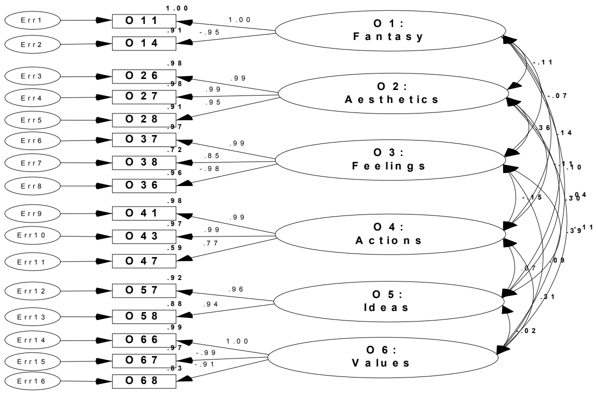

Fig. 6: Measurement model– openness

Table 1 presents the CFA for different measurement models. The results are evaluated by examining the feasibility of the parameter estimates, and local and overall model fit. Each measurement model is presented in graphical form for clarity (Fig. 1 –6).

Table 1: CFA for the Measurement Models

Models/ Results	χ^2	df	TLI	CFI	RMSEA
Credit Card -FIT	128	51	0.96	0.96	0.07
Agreeableness-FIT	619	76	0.95	0.96	0.08
Neuroticism--FIT	281	52	0.97	0.96	0.06
Extraversion-- FIT.	510	75	0.96	0.97	0.07
Conscientious-FIT	552	75	0.97	0.97	0.08
Openness - FIT	1062	90	0.95	0.96	0.08

p = <.01

Path Models: A number of path model analyses were carried out in order to explore the relationship between personality variables and customer switching variables and these analyses were assumed to be of a confirmatory nature as suggested by Byrne (1998). In other words, using a path model, the postulated causal relations (which are supposed to be grounded in theory, empirical research, or both) among the constructs in the proposed conceptual model can be verified based on collected data.

Personality predicting Customer Switching: Path models were developed using personality facets predicting customer switching behavior. The SEM for personality facets predicting credit card switching revealed an adequate fit to the data as per the overall fit criterion. The adjusted R^2 was 0.21 for Credit Card Switching. Hence the proportion of variance in Credit Card Switching predicted by the personality facets was 21.0%. (Fig.7)

Table 2: CFA for the path models

Models/ Results	χ^2	df	TLI	CFI	RMSEA
Personality predicting credit card switching [FIT]	289	91	0.96	0.94	0.08

p = <.01

4. Conclusion

Finally, Structural Equation Modelling was used to identify the relationships between personality and customer switching. With all the assumptions concerning the conduct of SEM being satisfied, the results of the CFA for the measurement models in this study were presented. Models tested revealed moderate to close fit of models depicting the personality facets impact on customer switching variables with effect size of 21%.

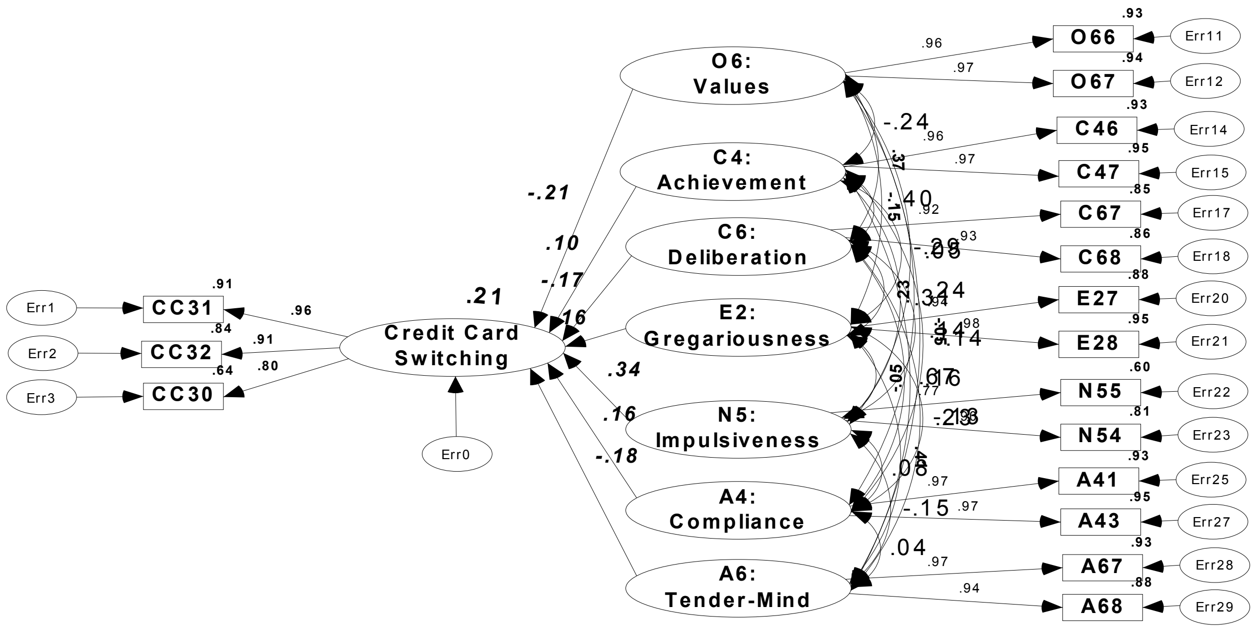

Fig. 7: Path Model: Personality Facets Predicting Credit Card Switching

References

Bandalos, D.L., & Finney, S.J., (2001). "Item parceling issues in structural equation modeling", Marcoulides, G.A., Schumacker, R.E., New Developments and Techniques in Structural Equation Modeling, Lawrence Erlbaum Associates, London, 269-96.

Bentler, P.M., & Chu, C.P., (1987) "Practical issues in structural modeling", Sociological Methods and Research, 16, 78-117

Bollen, K.A., & Lennox R. (1991) "Conventional Wisdom on Measurement: A Structural Equation Perspective." Psychological Bulletin, 110:305-14.

Bollen, K.A., (1989a). "Structural Equation Modeling with Latent Variables", New York: John Willey and Sons.

Bollen, K.A., (1989b). "A New Incremental Fit Index for General Structural Models," Sociological Methods and Research, 17 (3), 303-316.

Byrne, B.M., (1998), "Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming" (Multivariate Applications Series), London: Lawrence Eribaum Associates, Inc.

Byrne, B.M., (2001). "Structural Equation Modeling with AMOS". Rahwah, NJ: Lawrence Erlbaum Associates

Fan, X., Thompson, B., & Wang, L., (1999). "Effects of sample size, estimation method, and model specification on structural equation modeling fit indexes". Structural Equation Modeling (6): 56-83.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R.L., (2006) "Multivariate Data Analysis", 6th Ed., Prentice Hall, New Jersey.

Hancock, G.R., (2003) "Fortune Cookies, Measurement Error, And Experimental Design," Journal of Modern Applied Statistical Methods, 2(2), 293-305

Hoyle, R.H. & Panter, A.T., (1995). "Writing About Structural Equation Models." In Hoyle, R.H. (Ed.) (1995). Structural Equation Modeling. Concepts, Issues, and Applications, Thousand Oaks, CA: Sage Publications.

Hoyle, R.H. (Ed.) (1995), "Structural Equation Modeling. Concepts, Issues, and Applications", Thousand Oaks, CA: Sage Publications.

Hu, L. & Bentler, P.M., (1999). "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives." Structural Equation Modeling 6(1): 1-55.

Hu, L., & Bentler, P.M., (1995), "Evaluating Model Fit," In Hoyle, R.H. (Ed.) (1995) Structural Equation Modeling: Concepts, Issues, and Applications, Thousand Oaks, CA: Sage Publications.

Jöreskog, K.G., (1993), "Testing Structural Equation Models," in Testing Structural Equation Models, Kenneth A. Bollen and J. Scott Long (Eds.), Newbury Park, CA: Sage Publications.

Loehlin, J.C., (1992) "Latent Variable Models: An introduction to factor, path and structural analysis", 2nd Ed, Hillsdale, New Jersey

Marsh, H.W., Balla, J.R., and Hau, K.T. (1996). "An evaluation of incremental fit indexes: A clarification of mathematical and empirical properties." pp. 315-353 in G. A. Marcoulides and R. E. Schumacker, eds, Advanced structural equation modeling techniques. Mahwah, NJ : Lawrence Erlbaum.

Marsh, H.W., Balla, J.R., and McDonald, R.P. (1988). "Goodness of fit indexes in confirmatory factor analysis: The effect of sample size". Psychological Bulletin, 103, 391-410.

Schumacker, R.E., & Lomax, R.G., (2004). "A beginner's guide to Structural Equation Modeling", 2nd edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Schumacker, R.E., (1996), "Editor's Note," Structural Equation Modeling. 3 (1), 1-3.

Siddiqui, K., (2013). "Heuristics for Sample Size Determination in Multivariate Statistical Techniques". World Applied Sciences Journal, 27(2), 285-287.

Stevens, J.P., (2002). "Applied Multivariate Statistics for Social Science," 4th Ed. Hillsdale, New Jersey

Tanaka, J.S., (1993), "Multifaceted Conceptions of Fit in Structural Equation Models," in Testing Structural Equation Models, Kenneth A. Bollen and J. Scott Long (Eds.), Newbury Park: Sage Publications.