Exploring Causal Path Directionality for a Marketing Model Using Cohen's Path Method[1].

William Callaghan.
Swinburne University, Faculty of Business and Enterprise,Melbourne, Australia.
strat@bigpond.com.au.
Bradley Wilson.
RMIT University, School of Applied Communication, Melbourne, Australia.
brad.wilson@rmit.edu.au
Christian M. Ringle
University of Hamburg, Institute of Industrial Management, Hamburg, Germany.
cringle@econ.uni-hamburg.de
Jörg Henseler.
Nijmegen School of Management, Radboud University, Nijmegen, Netherlands.
j.henseler@fm.ru.nl

Introduction

Researchers must frequently consider the directionality of relationships between variables when linking variables as well as when positing construct-to-construct relationships or when relations are specified at a higher order level of abstraction (Wilson, Callaghan & Stainforth, 2006). The psychometric literatures have been particularly mindful of these path directionality issues with measurement model specifications (item-to-construct directionality being either reflective or formative in orientation[2]).

Chin (1998b, p. IX) has recognised that, "a common and serious mistake often committed by research is used to inadvertently apply *formative* indicators in a (covariance-based) SEM analysis." This measurement model specification concern has also been empirically proven by Jarvis, Mackenzie and Podsakoff (2003) who highlighted the magnitude of the problem. They found that 29% of constructs were modeled incorrectly. In the majority of cases items should have been treated in a formative fashion but were analyzed as if they were reflective in nature.

The objective of this paper is to *revisit* the relatively simple path analysis procedure to help explore issues of causal direction between variables of interest. In view of his enormous contribution to this field we refer to it as "Cohen's path method" in this paper. We consider the

[1] We would like to acknowledge the inspiration and direction for this research through the excellent conference and journal submissions of Heshan Sun and Ping Zhang. They were also very helpful in email correspondence in clarifying matters of analytical process. We would also like to thank Paul Cohen for answers to our queries.

[2] We expect that the reader is familiar with basic measurement concepts such as reflective and formative measures and their unique characteristics. The interested reader is recommended to review Bollen & Lennox (1991); Chin (1998a), and Jarvis, MacKenzie & Podsakoff, (2003).

methodological issues using a simple path example. We believe that social scientists focus on the vast array of fit measures and predictive diagnostics that currently exist within CBSEM and PLS and perhaps do not consider directionality issues post hoc or the investigation of alternative models. In the theoretical development and model building stage assumptions are made about causal direction and are often not revisited. Not considering directionality issues with alternative models post hoc may be a small problem when the model is based on extremely well established theoretical underpinnings but this is often not the basis from which theorists are working from.

Cohen's Path Analysis Method and its Application

More sophisticated techniques to investigate path directionality exist, including Exploratory Tetrad Analysis (Glymour, Scheines, Spirtes, Kelly, 1987)] and Confirmatory Tetrad Analysis (CTA) (Bollen & Ting, 1993; Ting, 1995) and covariance-based structural equation modeling (CBSEM) techniques via nested chi-square tests analysis techniques. However, these approaches are complex and require computational approaches that are not yet widely used. The simplicity of Cohen's path analysis therefore offers advantages to a range of research contexts.

Sun & Zhang (2006) distinguish between the "connectedness" (which is addressed in SEM approaches) and the actual directionality issues in investigating causal relationships and highlight the limitations of currently used approaches. The consequences of failing to explore alternative causal representations in establishing the "best" underlying causal sequence is clearly very important in marketing where often the theoretical underpinnings are comparatively weak. It may also influence other decisions such as the choice of an appropriate data analysis method and the number of items that are necessary in the questionnaire representing a particular construct. Bollen and Lennox (1991) believe that if the measures are reflective, a small sample of measures from the population of measures of the construct is sufficient to represent the construct.

The simple principle involved in Cohen's path analysis (Cohen et al., 1993) is that estimated correlations based on path analysis should be as close as possible to the actual correlations. Cohen et al. (1993) describe it as a generalization of multiple linear regression that builds models with causal interpretations. The first stage in the analysis procedure is to calculate the actual correlation coefficients between all model variables (see Table 1 for a presentation of the specific actual correlations for the linkages in Figure 1). These actual correlation estimates are then compared with estimates derived from calculations involving the original (Model 1) and the alternative model (Model 2) where the one alternative causal link is involved. The substantive theoretical detail of the model is not germane to this paper. The data is from a commercial marketing research study on meat consumption where the dependent variable was number of servings of red meat consumed per week with X_1 to X_5 considered variables that might impact on this and included perceived health benefits, top of mind awareness, price perceptions and advertising recall. The nodes represent these variables or constructs. PLS path model estimation uses the SmartPLS 2.0 (Ringle et al. 2005) software application.

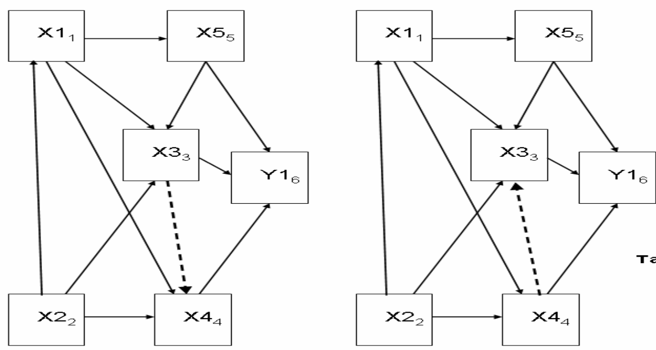

Fig 1a. Original Model Fig 1b. Alternative Model. Node 3 and 4 Reversed.

Link	Correlation
1-2	0.07
1-3	0.26
1-4	0.15
1-5	0.37
2-3	0.16
2-4	0.27
3-4	0.22
3-6	0.43
4-6	0.34
5-3	-0.12
5-6	0.41

Table 1: Node and Link Correlations

Cohen, Carlson, Ballesteros & St Amant (1993) observed that when examining paths between independent (X) and dependent (Y) variables in a path model, analysts can utilize two basic rules extending Sewall Wrights' original path analysis rules. These rules enable the researcher investigate all the relevant paths between variables in the model. These are stated as:

1. A path cannot go through a node twice.
2. Once a node has been entered by an arrowhead, no node can be left by an arrowhead.

The estimated correlations for each model and between any two variables linked by various paths can be estimated by considering all direct and indirect relationships. The total effects are compared to the actual correlations (Appendix 1) for each model. The actual coefficients may derive from SEM or related procedures that fit the overall model.

In this case model total squared errors are 0.035 for Model 1 and 0.082 for Model 2. The error changes from Model 1 to Model 2 indicated that the TSE changed by 130% (0.082-0.035/0.035). To employ another Cohen contribution the effect size is calculated using Cohen's d formula (Cohen, 1988) (see equation 1). Cohen (1988) indicated an effect size of 0.2 is small, 0.5 as medium and greater than 0.8 as large. The Cohen d value estimated for the difference between these models is 1.65 indicating that Model 1 is to be strongly preferred:

$$d = \mathrm{TSE}_2 - \mathrm{TSE}_1 / \sigma \tag{1}$$

where σ is the pooled standard deviation of the TSE values.

Discussion and Conclusions

Cohen's path method needs to be regarded as exploratory in establishing causal relationships rather than definitive. Cohen's path method is appealing when theory or the literature in an area do not offer clear guidelines on causal directions.

The approach is not without limitations. One of the most important limitations is the use of correlations that do not take into account the error within the measures and by default constructs (if investigating structural models). This error becomes more imbedded within a structural model when utilising PLS modeling approaches. We therefore recommend that this exploratory technique be used when item and construct reliabilities are relatively high. The consistency at large assumption with PLS may overcome this to a degree (Dijkstra, 1983). Issues of attenuation may also need to be explored further as they affect the underlying correlations. Secondly, the decision rule in accepting the model result with the lowest total squared error is sensitive to variable "noise." Cohen et al. (1993; 1994) suggest that other decision heuristics be evaluated to supplement the total squared error (TSE) estimates. We urge further simulation research in this area to ascertain TSE robustness.

Finally, it is important to acknowledge that there are now other quantitative techniques that assist when solving path direction questions with non-experimental designs. For instance, Exploratory TETRAD analysis, (Glymour et al., 1987), Confirmatory[3] TETRAD analysis and CBSEM nested tests can be considered. In addition Jarvis et al. (2003) provide a comprehensive series of qualitative decision rules for examining the formative versus reflective issue in modeling. Cohen et al. (1993) and Gregory & Cohen (1994) have also integrated an algorithm with Cohen's path method that finds the most plausible causal path given the data at hand. The Cohen's path model algorithm runs through all variable path combinations[4] (not just the one path that we have illustrated here). This work and the algorithm offers much potential for future research and subsequent integration within PLS software packages. We believe researchers can probably benefit from using a combination of both exploratory and confirmatory approaches in most research instances. Future validation studies utilizing experimental designs is the ultimate goal.

---

[3] We realise that directionality can never be confirmed "per se" and do not have the space required to discuss the varying opinions on the philosophy of causation. Experimental design methods are ideal but often not practical in complex SEM models to implement. This point is more in reference to the method being "confirmatory" in that is fixes a specific path for investigation. E.g., Confirmatory Tetrad Analysis (Bollen & Ting, 1993).

[4] 2^{N2} is the total number of models explored by the algorithm. Where N = no. of variables.

Bibliography

Bollen, K. A., & Lennox, R. (1991). "Conventional Wisdom on Measurement: A Structural Equation Perspective." *Psychological Bulletin*, 110 (2), 305-314.

Bollen, K. A. & Ting, K. F. (1993). "Confirmatory tetrad analysis." In P. Marsden (ed.), *Sociological Methodology*. Washington, DC: American Sociological Society, 147-175.

Chin W. W. (1998a). "The Partial Least Squares Approach to Structural Equation Modeling." in Marcoulides, George A. (Ed.) *Modern Methods for Business Research.* Lawrence Erlbaum, Mahwah, NJ, 295-336.

Chin, Wynne W (1998b). "Issues and Opinion on Structural Equation Modeling." *MIS Quarterly.* (March), VII-XVI.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences.* Hillsdale, NJ: Lawrence Erlbaum.

Cohen, P. R., Carlsson, A., Ballesteros, L., & Amant R. S. (1993). "Automating path analysis for building causal models from data. in *Proceedings of the International Workshop on Machine Learning,* 57-64.

Cohen, Paul R., Hart, David R., St. Amant, R., Ballesteros, L. A., & Carlson, A. (1994). "Path Analysis Models of an Autonomous Agent in a Complex Environment." In *Selecting Models from Data: AI and Statistics IV,* Peter Cheeseman and R. W. Oldford, Eds. Springer-Verlag, Technical Report 94-33, Dept. of Computer Science, University of Massachusetts/Amherst, 243-251.

Dijkstra, Theo. (1983). "Some Comments on Maximum Likelihood and Partial Least Squares Methods." Journal of Econometrics. 22, 67-90.

Glymour, C., Scheines, R., Spirites, P., & Kelly, K. (1987). *Discovering Causal Structure.* Academic Press.

Gregory, D. E., & Cohen, P.R. (1994). "Two Algorithms for Inducing Causal Models from Data." In *Knowledge Discovery in Databases Workshop*, Twelfth National Conference on Artificial Intelligence, Technical Report 94-42. Department of Computer Science, University of Massachusetts/Amherst. pp. 73-84.

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). "A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research." *Journal of Consumer Research.* 30 (September), 199-218.

Ringle, C. M., Wende, S. & Will, A. (2005). *SmartPLS 2.0 (beta)*. Hamburg: www.smartpls.de

Sun, H. & Zhang, P. (2006). "Causal Relationships between Perceived Enjoyment and Perceived Ease of Use: An Alternative Approach." *Journal of the Association for Information Systems (JAIS),* 7 (9), 618-645.

Ting, K. F. (1995). "Confirmatory tetrad analysis is SAS." *Structural Equation Modeling.* 2, 163-171.

Wilson, B., Callaghan, W., & Stainforth, G. (2006). "An Investigation of Path Directionality Issues in a Branding Structural Model: An Application of Vanishing Tetrads Analysis." Third International Business Research Conference, Melbourne, November 20-26th.

Zhang, P., Li, N., & Sun, H. (2006). "Affective quality and cognitive absorption: Extending technology acceptance research", Proceedings of the 39th Hawaii International Conference on System Sciences (HICSS 2006). January 4-7, 2006, Kauai, Hawaii.

Model 1							Model 2						
Path	Direct	indirect	Estimated correlation for indirect	Total estimated correlations (including direct)	Actual correlation	Squared error	Path	Direct	indirect	Estimated correlation for indirect	Total estimated correlations (including direct)	Actual correlation	Squared error
1-6	none	1-5-6	0.152	0.315	0.330	0.00023	1-6	none	1-5-6	0.152	0.288	0.330	0.00179
		1-5-3-6	-0.019						1-5-3-6	-0.019			
		1-3-6	0.112						1-3-6	0.112			
		1-4-6	0.051						1-4-6	0.051			
		1-3-4-6	0.019						1-5-3-6	-0.008			
2-6	none	2-1-5-6	0.011	0.207	0.150	0.00326	2-6	none	2-1-5-6	0.011	0.207	0.150	0.00329
		2-1-3-6	0.008						2-1-3-6	0.008			
		2-3-4-6	0.012										
		2-1-4-6	0.004						2-1-4-6	0.004			
		2-3-6	0.069						2-3-6	0.069			
		2-4-6	0.092						2-4-6	0.092			
		2-3-4-6	0.012						2-4-3-6	0.026			
		2-1-3-4-6	0.001										
		2-1-5-3-4-6	0.000						2-1-5-3-6	0.000			
		2-1-5-3-6	-0.001						2-1-5-3-6	-0.001			
3-6	3-6	3-4-6	0.075	0.505	0.620	0.01327	3-6	3-6	none		0.430	0.620	0.03610
4-6	4-6	none		0.340	0.290	0.00250	4-6	4-6	4-3-6	0.095	0.435	0.290	0.02091
5-6	5-6	5-3-4-6	-0.009	0.349	0.280	0.00482	5-6	5-6	5-3-6	-0.052	0.358	0.280	0.00615
		5-3-6	-0.052										
1-2	2-1	none		0.070	0.043	0.00073	1-2	2-1	none		0.070	0.043	0.00073
1-3	1-3	1-5-3	-0.044	0.216	0.265	0.00244	1-3	1-3	1-5-3	-0.044	0.216	0.265	0.00244
1-4	1-4	1-3-4	0.057	0.197	0.170	0.00075	1-4	1-4	1-3-4	0.057	0.207	0.170	0.00138
		1-5-3-4	-0.010										
1-5	1-5	none		0.370	0.400	0.00090	1-5	1-5	none		0.370	0.400	0.00090
2-3	2-3	2-1-3	0.018	0.175	0.150	0.00063	2-3	2-3	2-1-3	0.018	0.175	0.150	0.00063
		2-1-5-3	-0.003						2-1-5-3	-0.003			
2-4	2-4	2-3-4	0.035	0.305	0.280	0.00064	2-4	2-4	none		0.270	0.280	0.00010
3-4	3-4	none		0.220	0.260	0.00160	3-4	3-4	none		0.220	0.260	0.00160
3-5	5-3	none		-0.120	-0.180	0.00360	3-5	5-3	none		-0.120	-0.180	0.00360
4-5	none				-0.07		4-5	none	5-3-4	-0.0264	-0.026	-0.07	0.00190
Total Squared error (TSE)						0.03537	**Total Squared error (TSE)**						0.08152

Appendix 1. The Results of Path Analysis for Models 1 and 2.