Tricks and Traps in Structural Equation Modelling: a GEM Australia Example Using AMOS Graphics.

RACHEL ABRAMSON
rachela@ozemail.com.au

SHEIKH RAHMAN
srahman@swin.edu.au

PATRICIA BUCKLEY
pbuckley@groupwise.swin.edu.au

Australian Graduate School of Entrepreneurship, Faculty of Business and Enterprise, Swinburne University of Technology, Melbourne, Australia

Abstract

This paper demonstrates the use of structural equation modelling (SEM) with AMOS Graphics. SEM combines and goes beyond the power of multiple regression and factor analysis to test factorial structure and model the direction of relationships amongst a set of variables.

In the sections that follow, the general approach to SEM will be demonstrated, along with special issues to consider in undertaking SEM. An example will be interspersed throughout to demonstrate SEM's application in practice. Conclusions and implications will then be drawn.

Keywords: structural equation modelling, methodology, GEM Australia project.

It is generally known that multiple regression and factor analysis build a more complete picture of a research interest than would correlational or descriptive statistics alone. That picture will be further enriched when structural equation modelling (SEM) is used in addition to multiple regression or factor analysis.

Moreover, SEM combines and goes beyond the power of both factor analysis and multiple regression to test models about research interests: Factor analysis is an exploratory tool, reducing a set of variables to a smaller set of underlying factors and determining the variables that load on each factor. Multiple regression identifies the set of independent variables explaining a particular dependent variable and how much of its variance is explained by those variables (Tabachnick & Fidell, 2001). SEM goes beyond factor analysis to test expected relationships between a set of variables and the factors upon which they are expected to load. As such, it is considered to be a confirmatory tool (Byrne, 2001; Kline, 1998). SEM also goes beyond multiple regression to demonstrate how those independent variables contribute to explanation of the dependent variable. It models the direction of relationships within a multiple regression equation. Moreover, SEM models several multiple regression equations simultaneously, incorporating the use of moderators and mediators as necessary (Byrne, 2001; Kline, 1998). SEM also tests alternative model structures, relationships between sets of variables (Byrne, 2001; Kline, 1998; Ullman, 2001), whether the same model holds across groups (Kline, 1998; Ullman, 2001), as well as specifying reliability and error terms (Byrne, 2001; Ullman, 2001).

The goal of SEM is to identify a model that makes theoretical sense, is a good fit to the data (Kline, 1998) and is parsimonious (Arbuckle & Wothke, 1999; Ullman, 2001). The model developed should be theory-driven, or based on past research.

Purpose of This Paper

Perusal of published research using SEM suggests that it has been used primarily in psychological and health sciences research, and to a lesser extent, in marketing and I.T. Yet, the potential application of SEM for business research is much greater than the fields of marketing and I.T. alone, and extends to the full range of business research.

Thus, the purpose of this paper is to walk readers through the use of SEM for business research using AMOS Graphics. This paper, however, does not delve into SEM at an algebraic level, nor does it enter the debate of one approach to SEM over another. [Readers interested in an algebraic treatment of one or other aspect of SEM are directed to authors such as O'Dell and Cudeck (1995) or McDonald and Bolt (1998). Those interested in the latter, are directed to visit the SEMNET discussion group. SEMNET is managed by listserv@bama.ua.edu].

In the sections that follow, SEM language and AMOS Graphics are introduced, followed by three alternate approaches to SEM and five steps to conducting SEM. Special issues arising at each step are discussed at that step. An example using AMOS graphics is interspersed throughout to demonstrate the use of SEM in practice. Conclusions and implications follow. This paper is limited to models with unidirectional causal relationships for cross-sectional research.

SEM Language

SEM distinguishes between latent and manifest variables, as well as between endogenous and exogenous variables. Latent variables are theoretical constructions of manifest variables (also known as indicator, measured or observed variables). Latent variables equate to “factors” in factor analytic techniques (Arbuckle & Wothke, 1999; Byrne, 2001; Joreskog, 1977, 1993; Kline, 1998; Ullman, 2001). Unlike factor analytic techniques which label factors on the basis of items that load on those factors, the meaning accorded to latent variables determines the manifest variables selected to measure them.

Endogenous variables are those that are of interest and are explained within the constraints of the model being tested (Byrne, 2001; Kline, 1998). Endogenous variables equate with dependent variables (Byrne, 2001; Joreskog, 1993) in a multiple regression equation. By contrast, exogenous variables are those variables used to explain relationships within the model. However the explanation of exogenous variables are outside the scope of the model being tested (Byrne, 2001; Kline, 1998). Exogenous variables equate with independent variables (Byrne, 2001; Joreskog,

1993) in a multiple regression equation. If, for instance, prediction of *Investor Portfolio Wealth* was of interest, and it was assumed that both *Investor Profile* and *Share Market Factors* play a part in its prediction, then *Investor Portfolio Wealth* is an endogenous variable, whilst both *Investor Profile* and *Share Market Factors* are exogenous variables.

Figure 1 highlights how these constructs are represented graphically. As shown in figure 1, latent variables are typically denoted by a circle or oval, whereas manifest variables are typically denoted by a square or rectangle (Byrne, 2001; Ullman, 2001). Exogenous variables have arrows leading away, whilst endogenous variables have arrows leading towards them. Some endogenous variables also have arrows leading away (Arbuckle & Wothke, 1999) as is the case with *Confidence* in figure 2. The direction of the arrows are indicative of "causal" relationships between endogenous and exogenous variables (Ullman, 2001).

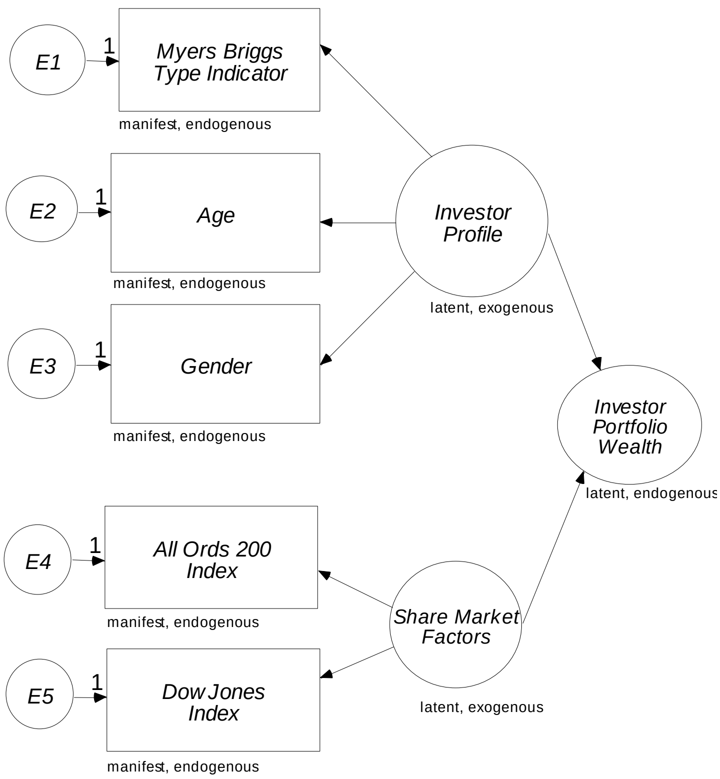

Figure 1: Distinction Between Manifest, Latent, Exogenous and Endogenous Variables in SEM.
Source: Author, 2005.

Whilst it initially seems counterintuitive, latent variables predict scores on manifest variables. SEM starts with the ‘factors’ rather than the variables used to measure them. So position on the latent variable determines scores on manifest variables used to measure them (Kline, 1998). The direction of arrows in the figure 1 model reflect this.

E1-E5 are also included in figure 1. They represent manifest variable residual (error) variance. Residual variance is a combination of random error in measuring the manifest variable as well as non-random error due to factors beyond that measure (Byrne, 2001; Joreskog, 1977, 1993; Kline, 1998). The “1” positioned over arrows leading from error term to manifest variable indicate that the variable’s full residual variance is represented in that error term. Whilst not shown in figure 1, error terms are also attached to latent variables. Latent variable error terms reflect errors in the model and random error (Byrne, 2001; Joreskog, 1977, 1993). Note that error terms for both manifest and latent variables are themselves latent constructs and hence represented as circular in structural equation models (Byrne, 2001).

Structural equation models are either recursive or nonrecursive: Nonrecursive models have bidirectional “causal” relationships, that is, feedback loops (Arbuckle & Wothke, 1999; Byrne, 2001; Kline, 1998; Ullman, 2001) correlated error terms, or both (Kline, 1998; Ullman, 2001). Recursive models have unidirectional “causal” relationships (Arbuckle & Wothke, 1999; Byrne, 2001; Kline, 1998; Ullman, 2001) and independent error terms (Kline, 1998; Ullman, 2001). Figure 1 is recursive as it demonstrates unidirectional causal relationships and independent error terms.

Just as there are correlated error terms, there are correlations between pairs of manifest variables where “causal” relationships are unspecified, (unanalysed associations). Unanalysed associations and correlated error terms are depicted by curved lines with arrow heads at each end (Arbuckle & Wothke, 1999; Byrne, 2001; Ullman, 2001). Figure 4b, shown later, depicts three such unanalysed associations.

Where necessary, SEM includes moderators and mediators. Moderators perform a qualitatively different role from mediators. Moderators are those that interact with the relationship of one variable’s impact on another’s (Baron & Kenny, 1986). The variable labelled as moderator ultimately depends upon expectations arising from underlying theory and research interest. Mediators are those variables that affect the relationship between two other variables. Mediators come between two variables such that the first variable has an indirect effect on the second variable, through its direct effect on the mediator (Baron & Kenny, 1986). In figure 2, for example, *Age* is portrayed as moderating with *Time in Market* so as to have a joint impact on *Portfolio Value (in dollars).* Similarly, *Gender* is portrayed as having an indirect effect on *Portfolio Value (in dollars)* through its direct effect on *Confidence*, which in turn is portrayed to have a direct effect on *Portfolio Value (in dollars).*

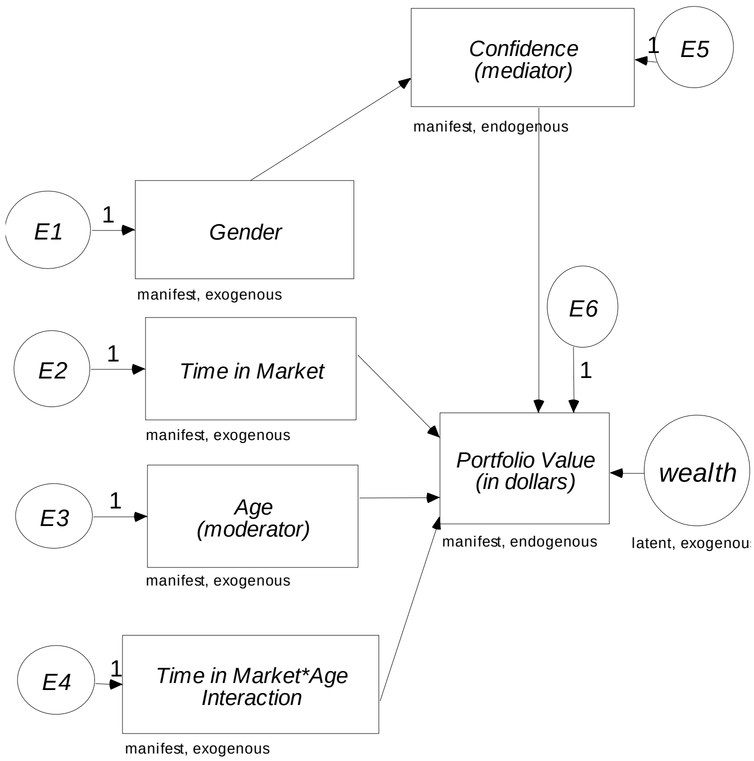

Figure 2 Distinction Between Moderators and Mediators in SEM.
Source: Author, 2005.

SEM examines structural models, measurement models, as well as full structural models. Structural models highlight relationships amongst latent variables. Measurement models highlight relationships between manifest variables and the latent variables they seek to measure, whilst full structural models combine both measurement and structural models (Arbuckle & Wothke, 1999; Byrne, 2001;

Joreskog, 1977, 1993; Ullman, 2001). Figure 2 represents a measurement model, whilst figure 1 represents a full structural model. Moreover, Figure 1's three latent variables alone represent a structural model. Similarly, figure 1's latent variable *Investor Profile*, along with its three measured manifest variables and their error terms reflect a measurement model. Replacing directional arrows into unanalysed associations amongst figure 1's three latent variables would turn the full structural model into a measurement model.

Finally, SEM estimates direct effects, means, intercepts, variances and covariances (Byrne, 2001). These parameters are either fixed to a set value, constrained as equal to other parameters, or freely estimated by SEM (Joreskog, 1977).

AMOS Graphics

AMOS (**A**nalysis of **MO**ments **S**tructures) is a statistical package specialising in structural equation modelling. It is purchased individually or as part of SPSS™. AMOS operates through one of two interfaces: AMOS Basic builds an analysis mathematically whilst AMOS Graphics does so pictorially. AMOS Graphics is user-friendly and contains all the features of AMOS Basic (Arbuckle & Wothke, 1999).

AMOS builds measurement, structural or full structural models. It tests, modifies and retests models. AMOS also tests alternate models, equivalence across groups or samples, as well as hypotheses about means and intercepts. It handles missing data using Maximum Likelihood (ML) estimation and provides bootstrapping procedures (Arbuckle & Wothke, 1999). Results obtained in AMOS are comparable to those obtained through other SEM packages (Ullman, 2001).

Apart from accepting SPSS™ datafiles, AMOS accepts data from spreadsheets, databases or standard text files (Arbuckle & Wothke, 1999). AMOS accepts correlation matrices with means and standard deviations, covariance matrices plus means, or the raw data (Arbuckle & Wothke, 1999) .

Because of its ease of use, AMOS Graphics was used to prepare the GEM Australia example.

GEM Australia Project Example

The GEM Australia Project is part of an international study exploring entrepreneurial activity. The project is sponsored by Westpac Banking Corporation. It is part of an international study to explore international differences in entrepreneurial activity, factors contributing to entrepreneurial activity as well as suggest policy to promote national entrepreneurship activities. The complete national survey includes a random telephone sample of Australian adults plus focus interviews (Hindle & O'Connor, 2005). However, this paper's example is limited to a subset of variables from the telephone survey data. Table 1 provides the correlation matrix. Sample sizes, means and standard deviations are also provided as appropriate.

Table 1

Correlations, Means and Standard Deviations for the GEM data.

	GENDER	AGE	MARKET OPPOR-TUNITY	STATE	INCOME	CUSTOMER SIZE	GROWTH EXPECTATIONS
GENDER	1						
AGE	.010	1					
MARKET OPPORTUNITY	-.064	-.073	1				
STATE	.020	.017	.127	1			
INCOME	-.031	-.008	.107	-.009	1		
CUSTOMER SIZE	.036	.056	-.014	.007	-.004	1	
GROWTH EXPECATIONS	-.103	-.084	.072	-.014	.038	.109	1
Sample size	*1991*	*1991*	*202*	*2000*	*1782*	*271*	*305*
Means		*49.570*	*1.629*		*2.330*	*30.210*	*4.200*
Standard deviations		*16.596*	*.795*		*1.445*	*34.203*	*32.877*

Alternate Approaches to SEM

Three approaches to SEM have been described: strictly confirmatory, exploration of alternative models, and model-generating models (Joreskog, 1993).

Strictly confirmatory approaches involve testing a theoretically-developed model, without modification, against a dataset. An alternative models approach involves testing several a priori, alternate, plausible models, with the goal of selecting the model that best fits theory and the data. The model-generating models approach uses the initial model to generate further models (Joreskog, 1993). In the latter case, the model selected is done so on the basis of "best fit" to theory and data, along with the goal of parsimony (Byrne, 2001; Joreskog, 1993; Kline, 1998).

However as the model-generating approach is exploratory, the generated model should be tested on a second dataset. An alternate practice is to randomly split a dataset into two (subsample A and subsample B), using subsample A for model development and subsample B for model testing (Breckler, 1990).

Whilst the three approaches to SEM are qualitatively different, the procedure described in the *Five Steps to SEM* is equally applicable. However, strictly confirmatory approaches end at the fourth step, whilst the remaining two approaches continue onto the final step. The next section describes the five steps to undertaking SEM.

Five Steps to SEM

The five steps to analysing a dataset using SEM are: model specification; model identifiability; measure selection, data collection, cleaning and preparation; model analysis and evaluation; and model respecification (Kline, 1998). Figure 3a represents the five steps pictorially. Each step is discussed below.

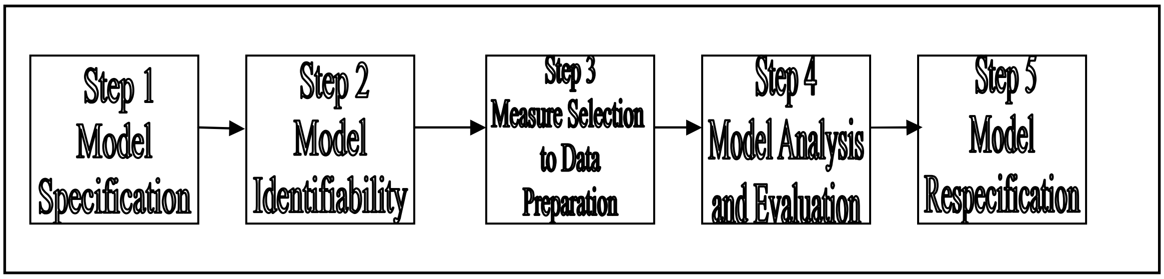

Figure 3a: Five Steps to SEM.
Source: Author, 2005

Step 1: Model Specification

Model specification involves mathematically or diagrammatically expressing hypothesized relationships amongst a set of variables (Kline, 1998). Figures 1 and 2 represent two such models.

The challenge at this step is to include all endogenous and exogenous variables, (including moderators and mediators), that are expected to contribute to central endogenous variables. Exclusion of important variables may result in the misestimation of endogenous variables. The extent of misestimation increases with the strength of the correlation between missing and endogenous variables. Whilst it is impossible to include all variables that contribute to the prediction of endogenous variables, it is possible to identify the main ones through careful examination of relevant theory and past research (Kline, 1998).

A second challenge is to determine the direction of relationships between pairs of variables in the SEM model. Actual direction is debatable, especially where manifest variables are measured at the same point in time. Thus, it becomes important to use theory or past research to determine direction of those relationships (Kline, 1998).

Figure 3 is a simple measurement model for the GEM Australia data, showing that growth expectations is a function of six other variables: current income, state, age, gender, customer size and market opportunity.

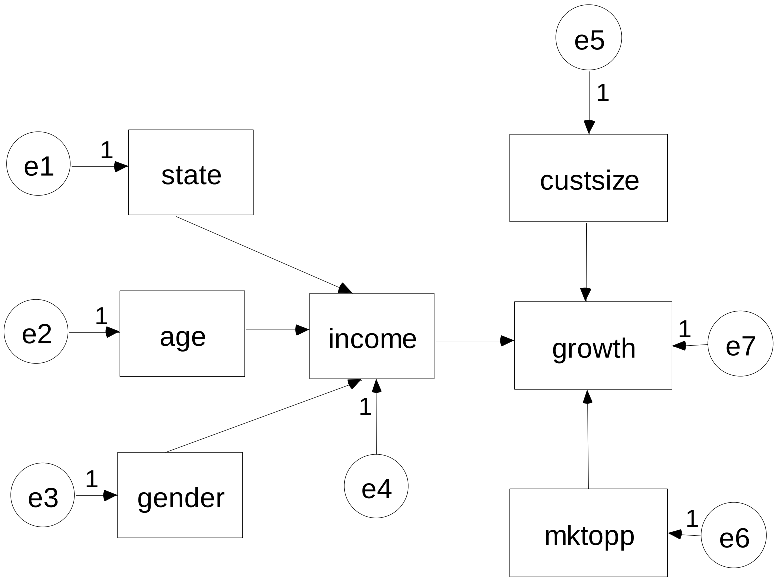

Figure 3: An overidentified measurement model using the GEM Australia data.
Source: Author, 2005.

Step 2: Model Identifiability

Specified models need to be checked for identifiability. A model is theoretically identifiable if there is a unique solution possible for it and each of its parameters. If a model is not identifiable, then it has no unique solution and SEM software will fail to converge. Such models need to be respecified to be identifiable (Kline, 1998; Ullman, 2001).

Recursive models are considered theoretically identifiable. However, multicollinearity or poor starting values (ie., initial estimates of model parameters) result in an unidentifiable model in practice (Kline, 1998). The issue of multicollinearity is discussed under *Step 3d - Data Preparation* whilst the issue of poor starting values is discussed under *Step 4*.

In practical terms, there is a limit to the number of parameters fitted in SEM. When attempting to fit more parameters than there are unique variances and covariances in the underlying covariance matrix, there is mathematically insufficient information from which to develop a unique solution. Under these conditions, SEM software cannot converge (Kline, 1998). Such models are underidentified (Kline, 1998) and need to be respecified. [Indeed, AMOS will advise that too many parameters were attempted to be fitted. It will also advise how many additional constraints are required to assess the model].

The maximum number of parameters that can be specified in the model is equivalent to the number of unique variances and covariances that can be found in its underlying covariance matrix (Kline, 1998). If, for example, there are four variables (say: *A*, *B*, *C*, and *D*), a covariance matrix has four unique variances (one for each variable) along with six unique covariances (*AB*, *AC*, *AD*, B*C*, *BD* and *CD*), giving a total of ten unique parameters. (See figure 4a). Thus, any model developed using these four variables would have a maximum of ten parameters available. One such model is given in figure 4b. Models containing the maximum number of parameters possible are said to be just-identified, or saturated models. Just-identified models fully reproduce the data (Arbuckle & Wothke, 1999; Kline, 1998).

Based on theory or past research, certain parameters would be expected to add little to the prediction of endogenous variables. Dropping such parameters from the model would provide a more parsimonious solution. If, for example, the association between *A* and *B* is negligible, including their covariance would provide no additional information. Thus, a more parsimonious solution would be to drop it from the model. This has been done in figure 4c. Models containing fewer parameters than variances and covariances in the underlying covariance matrix are said to be overidentified. Overidentified models need certain constraints to be set in order to be mathematically calculable, and with those constraints in place, imperfectly reproduce the data (Kline, 1998).

	A	*B*	*C*	*D*
A	Var(*A*)			
B	Cov(*AB*)	Var(*B*)		
C	Cov(*AC*)	Cov(*BC*)	Var(*C*)	
D	Cov(*AD*)	Cov(*BD*)	Cov(*CD*)	Var(*D*)

Figure 4a: A Covariance Matrix With Four Variables, *A*, *B*, *C* and *D*.
Note: For four variables, there are four unique variances and six unique covariances, giving a maximum of ten parameters estimable with SEM.
Source: Author, 2005.

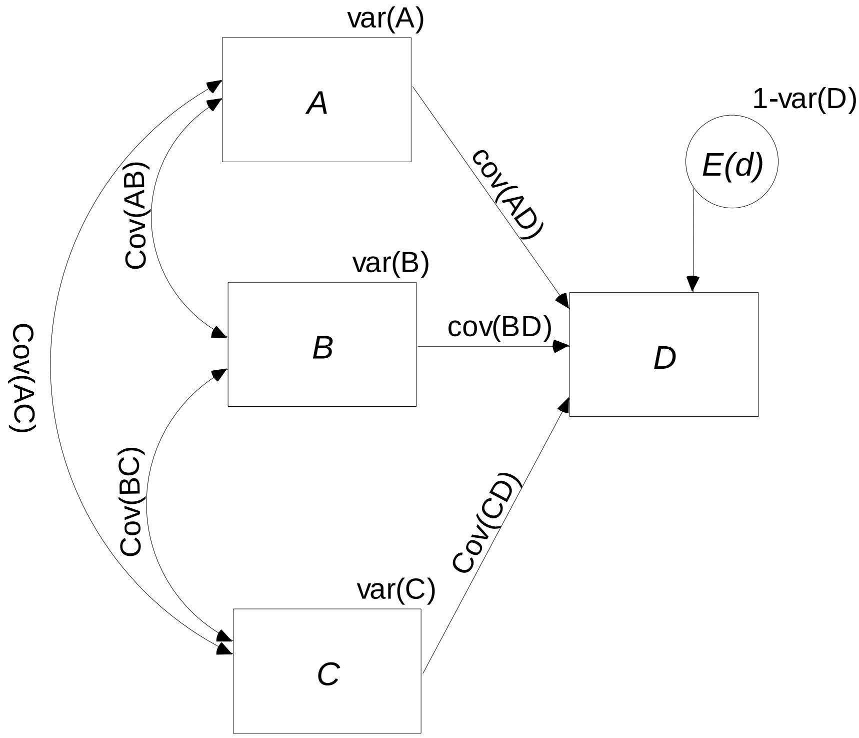

Figure 4b: A Just-identified Model With Four Variables.
Source: Author, 2005.

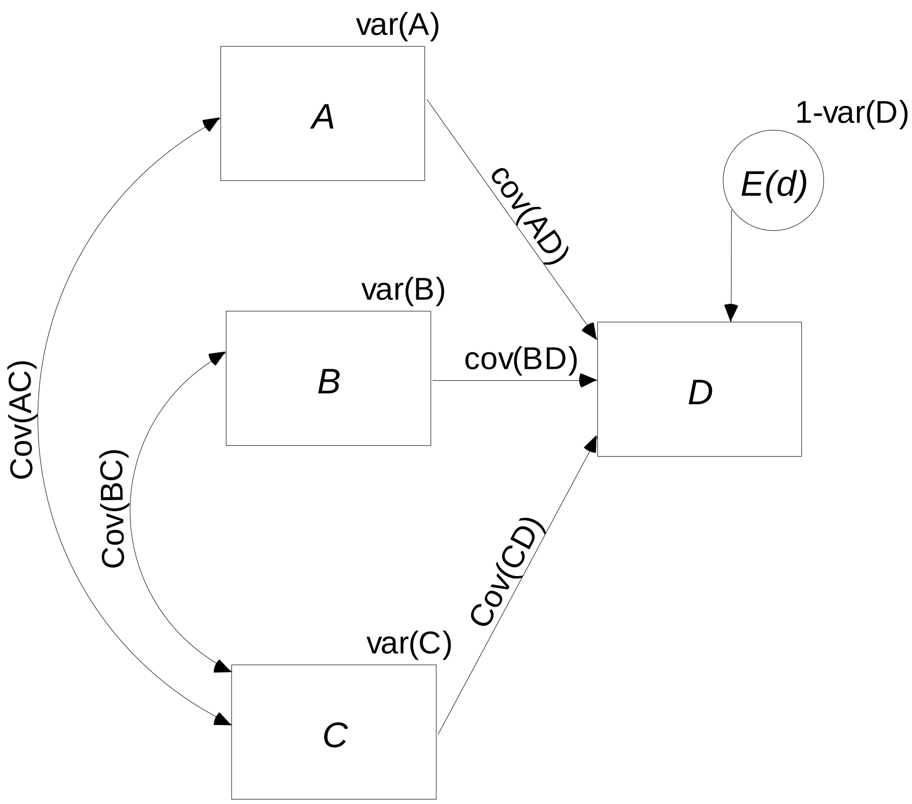

Figure 4c: An Over-identified Model With Four Variables.
Source: Author, 2005.

Figure 3 contains thirteen parameters (six hypothesized relationships and seven error terms). As twenty-eight parameters are possible for a seven-variable model, this model is overidentified. For it to become just-identified, fifteen additional directional relationships or unanalysed associations could be included.

Step 3: Measure Selection, Data Collection, Cleaning and Preparation

Step 3 has four substeps: measure selection, data collection, data cleaning and data preparation. The latter substep includes checking statistical assumptions.

Step 3a - Measure Selection

Manifest variables are estimates of the underlying latent constructs they purport to measure. It is therefore recommended that each latent construct be measured by at least two manifest variables (Joreskog, 1977). Where only one manifest variable is available, it is recommended that the measure's internal reliability coefficient be included in the model (Kline, 1998).

Moreover, measures selected need to demonstrate good psychometric properties. That is, they need to be both "reliable" and "valid" measures of the latent constructs they seek to address. A measure is considered reliable when it gives consistent, or repeatable, results. It is considered valid when it measures what it says it measures. When measures have poor reliability and/or validity properties, ML estimates become statistically biased (Kline, 1998).

Reliability is assessed through alternate form, internal consistency/split-half, interrater, or test-retest correlation coefficients. The resulting coefficient indicates the measure's degree of, content-, homogeneity-, rater-related- or time- repeatability respectively. Coefficients of 0.8 or above suggest good reliability, whilst those in the range of 0.7 to 0.8 suggest adequacy. Coefficients below 0.5 should be avoided (Kline, 1998) or improved before use in research.

Validity is assessed by examining its content, criterion-related, convergent or discriminant validities. Content validity exists when experts agree that the measure is tapping into the relevant domain. Criterion-related validity assesses whether a measure taps into a particular domain, as assessed against some set criteria. That criteria is assessed either simultaneously (concurrent validity) or after the measure of interest (predictive validity). Convergent validity exists when measures that purport to measure the same construct have moderate to high correlations. Similarly, discriminant validity exists when measures that purport to measure different constructs have low to moderate correlations (Kline, 1998).

Step 3b - Data Collection

A sufficiently large sample needs to be drawn in order to analyse the model specified at Step 1. The sample drawn should be ten times the number of model parameters to be estimated, with a minimum of 200 cases (Curran, Bollen, Paxton, Kirby, & Chen, 2002; Kline, 1998). If planning to divide the sample in two for model development and testing purposes, then each half sample needs to be sufficiently large. Moreover, expected response rates should be factored into consideration when drawing the sample.

Step 3c - Data "Cleaning"

The acronym GIGO (Garbage In, Garbage Out) highlights the importance of checking the veracity and integrity of data entry. In statistical terms, doing so ensures that data is "clean" before proceeding further. The efforts taken to ensure data integrity early can save on potential heartache down the track (Kline, 1998).

Checking each datapoint of a large dataset may be tedious. However, it is possible to check (and correct) the first five or ten cases and extrapolating their accuracy rate to the remaining cases in the dataset. If accuracy is less than, say, 95%, the data could be reentered using a double entry method. Alternatively, the remaining cases could be checked and corrected. If accuracy is higher than 95%, the use of descriptive statistics, histograms, and scatterplots would suffice to check, and correct, unusual datapoints. Through such simple steps, out-of-range datapoints will come to light and can easily be identified and corrected (Kline, 1998; Tabachnick & Fidell, 2001).

Step 3d - Data Preparation

Micceri (1989) found that approximately 30% of distributions approached a normal distribution. However, the practice of assessing assumptions and redressing any identified violations does not appear to be widely observed. Indeed, Breckler (1990) found approximately 10% of SEM articles reported doing so. Both papers combined, highlight the importance of checking assumptions, as well as redressing any violations of those assumptions before proceeding with data analysis.

Maximum Likelihood (ML) estimation is the preferred estimation procedure for SEM. It is also the AMOS default estimation procedure (Arbuckle & Wothke, 1999). ML is described under *Step 4: Model Analysis and Evaluation*. ML assumes multivariate normality that asymptotically approximates a χ^2 distribution; asymptotically large samples; a correctly specified model; linearity and homoscedasticity of variables; manifest variables measured on an interval scale; and, random sampling.

Multivariate normality, asymptotically large samples and correctly specified models: Under conditions of multivariate normality, ML estimates are statistically unbiased, asymptotically efficient and consistent (Curran et al., 2002; West, Finch, & Curran, 1995), as well as least biased in the presence of missing data (Byrne, 2001). ML parameter estimates are relatively robust from deviations from multivariate normality (Kline, 1998).

An asymptotically large sample has a minimum of 10 cases per model parameter, with the ideal ratio of 20:1, and a minimum of 200 cases (Curran et al., 2002; Kline, 1998). When a sample is asymptotically large, and the model has been correctly specified, ML estimates approximate a χ^2 distribution with a mean of df and a variance of twice the df (Curran et al., 2002). Thus, having an asymptotically large sample, and a correctly specified model, means the probability of rejecting the null hypothesis when it is true approaches the exact critical level of significance as sample size increases (Arbuckle & Wothke, 1999).

ML estimates still follow a χ^2 distribution when there is moderate model misspecification. However that distribution becomes noncentral and requires the addition of a constant. When the model has been grossly misspecified, or sample size is less than 200, ML estimates follow some other (unknown) distribution. The impact of test statistics under these conditions is unknown (Curran et al., 2002).

Whilst it is not possible to know when a model has been misspecified, (and if so, the degree of misspecification), an alternative models or model-generating models approach, may provide some protection against unknown model misspecification. Moreover, the model that appears to best fit the data should be tested on a second, asymptotically large sample. Alternatively, as described in *Alternate Approaches to SEM*, the researcher can randomly divide the original sample in two, (Subsamples A and B), develop the best-fitting model on Subsample A and test it on Subsample B (as per Breckler, 1990).

Multivariate normality is assessed by examining the distributions of each manifest variable, as well as the multivariate normal probability plot for the dataset (Tabachnick & Fidell, 2001). Distributions should also be examined for the presence of both univariate and multivariate outliers, skewness and kurtosis (Tabachnick & Fidell, 2001; Ullman, 2001; West et al., 1995). Outliers due to data entry errors can be corrected (Tabachnick & Fidell, 2001). Remaining outliers are handled by either removing the outlying case, transforming the variable containing the outlier (Kline, 1998), redefining the population of interest or redefining the model itself (West et al., 1995).

Data transformations redress deviations from multivariate normality not due to outliers (Tabachnick & Fidell, 2001). ML estimates are then obtained on the transformed data. However, transformed data may still fail to meet the requirements of multivariate normality, or may be more difficult to "interpret" than the original untransformed data (Micceri, 1989; Tabachnick & Fidell, 2001; Ullman, 2001; West et al., 1995). A viable alternative to transforming the data in the case of large samples involves combining a bootstrapping procedure with ML estimation. [For nonnormal samples of at least 200 cases, bootstrapping, combined with ML estimation, is considered to give better results than ML alone (Byrne, 2001)]. However, raw data, without missing values, is required to run the bootstrapping procedure (Arbuckle & Wothke, 1999). There is no alternative to transforming the data for small samples.

Linearity and homoscedasticity: A component of the multivariate normality assumption is that of linearity and homoscedasticity between manifest variables

(Kline, 1998). Furthermore, the ML estimation procedure makes use of the Pearson's correlation matrix, or its covariance counterpart. Pearson's r assumes relationships between pairs of variables are linear. To the extent that this assumption is met, the true extent of the relationship between pairs of variables will be recognised. If the relationship is nonlinear, the relationship between pairs of variables may be underestimated (Tabachnick & Fidell, 2001). Homoscedasticity occurs when there is equal spread of scores for one variable across levels of a second (Tabachnick & Fidell, 2001). Examination of scatterplots facilitate assessment of both the linearity and homoscedasticity assumptions. However, this can be cumbersome when a dataset contains many variables. With large datasets, randomly drawn pairs of variables can be assessed for linearity and homoscedasticity. Alternatively, those pairs of variables more likely to depart from either of these assumptions (as assessed through skewness coefficients) can be assessed for linearity and homoscedasticity (Tabachnick & Fidell, 2001).

Deviations from linearity can be redressed through power transformation (Tabachnick & Fidell, 2001; Ullman, 2001). Deviations from homoscedasticity may be remedied by transformations (Kline, 1998; Tabachnick & Fidell, 2001) such as reciprocal, log or inverse-power transformations.

Table 1 provided the cleaned GEM Australia data, prior to any assumption checks. Examination of the means and standard deviations suggest that at least two of the variables are likely to be non-normal (growth expectations and customer size). Univariate normal probability plots confirmed this. Moreover, linear relationships appear to exist between pairs of variables other than these two variables. Analysis of the scattergrams suggested that growth expectations contains two outlying observations. Attempts at transforming both variables were not successful. However, recoding the two outlying cases had the impact of marginally improving the linearity of relationships between growth expectations and the remaining variables. Variable distributions are available from the author upon request. In the absence of missing data, the bootstrapping procedure would have been appropriate.

For a thirteen parameter model, a minimum of 130 cases would have been required. Given that this number is less than the SEM minimum requirement of 200 cases, 200 cases become the minimum required to undertake SEM. Each variable in the GEM Australia dataset has at least 200 cases. Moreover, to protect against the problem of model-misspecification, a model-generating models approach was utilised.

Interval scaled measures: SEM assumes that manifest variables are measured on an interval scale (Ullman, 2001). The more a variable deviates from an interval scale to that of an ordinal or nominal scale, and the fewer the response alternatives, the greater the impact on ML test statistics (West et al., 1995). Manifest variables using a nominal scale are categorical, with each category being qualitatively different. They can thus be treated as grouping variables (Ullman, 2001). Each category becomes a separate subsample: The first subsample (category) is then used to develop the model and remaining subsamples test the model (as per Breckler, 1990).

Manifest variables using an ordinal scale involve ranked or ordered data. The impact on ML test statistics becomes minimal for ordinal data the greater the number of response alternatives and when inputting the covariance matrix into SEM programs (Byrne, 2001). By default, SEM programs use the pearson's correlation (through correlation or covariance matrices) as input. However, alternatives to the pearson's correlation should be used when ordinal measures have two or three response alternatives: Polyserial correlations should be used where one of two measures is ordinal and the other interval. Polychoric correlations should be used when both variables are ordinal (Ullman, 2001).

Of the GEM Australia dataset, age, customer size and growth expectations are continuous variables. Moreover income and market opportunity are ordinal in nature. Both these variables have at least four levels. Thus, using the covariance matrix as input into SEM will minimise any impact on ML statistics. However, both gender and state have nominal scales. With the sample size constraints of several variables, it would not be possible to divide the sample in two to develop the model on one group and test it on remaining groups. Thus, it was decided to remove both variables from the model. Later GEM studies are invited to test gender and state effects on the model. Figure 5 provides the reduced model ready for testing.

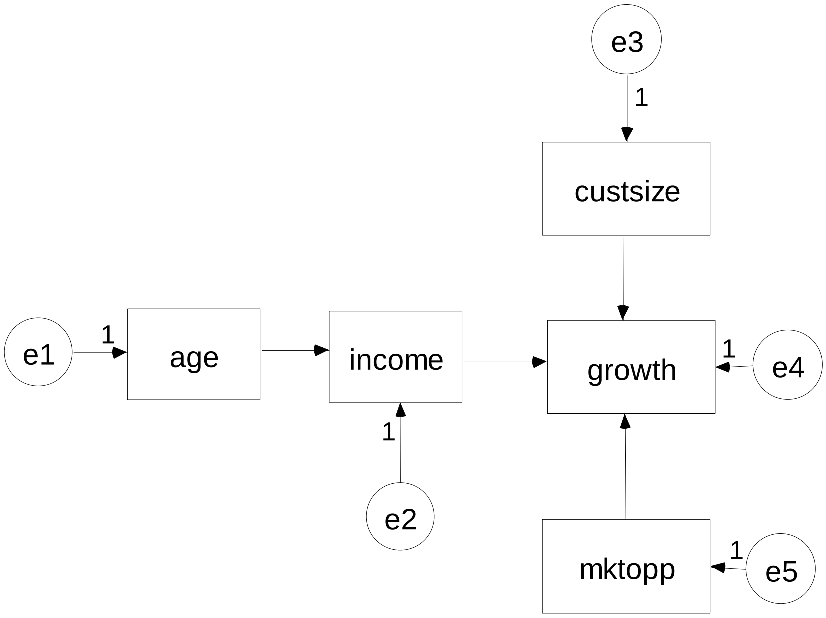

Figure 5: A reduced overidentified measurement model using the GEM Australia data.
Source: Author, 2005.

Random sampling: Random sampling is a statistical assumption underlying the broader research process. Research participants are expected to be randomly drawn from a population of interest (Joreskog, 1993). In survey research, respondents are randomly selected to participate in the study. In experimental research, respondents are randomly assigned to treatment conditions, holding all other factors constant (Arbuckle & Wothke, 1999; Tabachnick & Fidell, 2001). If the sampling process has not been random, interpretation of research results is limited to the results of the sample itself, rather than generalisable to other members of the same population (Tabachnick & Fidell, 2001). Thus, at the design and data-collection phases, it is important to ensure that data is collected using random sampling or random allocation procedures. The GEM Australia data was collected using simple random sampling techniques. Hence, results are generalisable to other Australian adults.

Other data preparation matters – multicollinearity and singularity: Multicollinearity occurs when variables are so highly correlated that they almost measure the same concept. Multicollinearity is present when pairs of variables have correlations of 0.9 or more, squared multiple correlations (SMC) approach 0.99. or tolerance statistics (the inverse of SMCs) approach 0.01 (Tabachnick & Fidell, 2001).

Similarly, singularity occurs when attempts are made to analyse a variable, along with each of its composite measures. Singularity would thus occur when attempting to analyse a full test, together with each of its subtests. Moreover, the closer the SMC coefficient is to ‘1’, (and hence, the tolerance statistic approaching zero) the more likely singularity is to be present (Tabachnick & Fidell, 2001).

Datasets containing multicollinearity or singularity include redundant variables in that there is not as much information available as the dataset’s listed number of variables suggest. Moreover, SEM programs will not run if singularity is present. They also become unstable when multicollinearity is present. It is therefore important to screen for these problems and redress any discovered problems (Tabachnick & Fidell, 2001).

Combining variables into a composite measure, and using the composite measure alone, would redress problems of multicollinearity. Similarly, removal of one of the variables contributing to singularity would redress this problem (Tabachnick & Fidell, 2001; Ullman, 2001). In the former case, multiple regression analysis or factor analysis would help determine the best combination of variables for that composite measure. The low correlations in table 1 suggest that problems of multicollinearity and singularity are unlikely for the GEM Australia data.

Other data preparation matters – missing data: Three forms of missing data have been identified: Missing completely at random (MCAR), missing at random (MAR) and nonignorable missing data. MCAR occurs when the presence of missing data is unrelated to the values that would have been obtained if the data was not missing

(Little & Rubin, 1987). MAR occurs when the presence of missing data is unrelated to the values that would have been obtained on the basis of some other variable in the dataset (Little & Rubin, 1987). MAR is the weakest condition under which it is acceptable to ignore the underlying cause of the missing data (Little & Rubin, 1987; Rubin, 1976). Nonignorable missing data occurs when the missingness is related to the values that would have been obtained had they been available. In other words, the missingness of data is systematic or biased in some way. Under these conditions, the nature of the missingness cannot be ignored (Little & Rubin, 1987). The source of bias and its implications for interpretation of results need to be identified (Kline, 1998).

AMOS uses the *Full Information Maximum Likelihood* (FIML) method to estimate the values of missing data. FIML is one of several algorithms available in the literature. Irrespective of algorithm used, ML generates the most consistent and statistically efficient solutions for datasets with data MAR or MCAR (Arbuckle & Wothke, 1999; Little & Rubin, 1987). Moreover, in the case of nonignorable missing data, ML is the least statistically biased (Arbuckle & Wothke, 1999).

Whilst there was missing data in the GEM Australia dataset, the missingness was a function of research design. Hence, it is ignorable, and either MCAR or MAR.

FIML requires input of the full dataset. However, the presence of ordinal data requires input of the covariance matrix. As the missingness was a function of research design, it was decided to input the covariance matrix into SEM. [FIML results using the full dataset into SEM were also obtained. The statistical results differed. However, these differences could be attributed to the FIML requirement of estimating means and intercepts in addition to model testing. Actual parameter coefficients differed little across the two initial solutions, suggesting that the decision to use the covariance matrix would have little impact on final solutions selected. The FIML initial solution is available from the author on request]. Table 2 shows how the covariance matrix needs to be prepared for input into SEM.

Table 2
The GEM Australia Covariance Matrix input into AMOS Graphics

		age	mktopp	income	custsize	growth
n		1991.000	202.0000	1782.000	271.0000	305.0000
cov	age	275.4210	.	.	.	.
cov	mktopp	-.6650	.6330	.	.	.
cov	income	-.1850	.1210	2.0890	.	.
cov	custsize	22.1480	-.3980	-.2180	1169.833	.
cov	growth	-15.4020	1.0710	.7490	43.1460	153.7450
mean		49.5700	1.6287	2.3300	30.2103	4.2000

Step 4: Model Analysis and Evaluation

This step has two substeps: analysis and evaluation. It is also the step which researchers find exciting because this is the step where actual results are generated.

Step 4a – Model Analysis

Model analysis involves the use of an estimation procedure to fit the model with the data provided (Kline, 1998). As mentioned above, ML is the preferred estimation procedure for SEM (Kline, 1998) because it gives statistically robust results with complete data, or data that is at least MAR, irrespective of the distribution of the data (Little & Rubin, 1987). ML estimation is an iterative process that estimates all model parameters simultaneously. In the case of just-identified, recursive models, the solution generated with ML techniques is identical to that generated using multiple regression techniques (Kline, 1998).

Parameters estimated by ML include unanalysed associations between exogenous variables, direct effects on endogenous variables, variance (and conversely, error variance) of all variables (Kline, 1998).

The standardized output report model parameters in the same “metric”, whilst unstandardized solutions does so in each variable’s own “metrics”, leaving a lack of comparability across variables Indeed, in standardized output, regression beta weights with absolute values of .10, .30 or .50 can be regarded as having a “small”, “moderate” or “large” effect respectively (Kline, 1998). For this reason, standardized solutions may be easier to interpret. However, standardized output provides the endogenous variable variances (through the use of the squared multiple correlation coefficient) whilst the unstandardized output provides individual exogenous variable variances directly on the model itself (Arbuckle & Wothke, 1999). So, when looking at model output, it may be of value to look at both the standardized and unstandardized output.

Table 3 summarizes the parameter estimates obtained through ML, as well as how AMOS reports these estimates in standardized and unstandardized form:

Table 3

Parameter Estimates, and their Standardized or Unstandardized Output

Parameter Estimate	Standardized Output	Unstandardized Output
Unanalyzed associations between exogenous variables	Pearson’s correlations	Covariance coefficients[#]
Direct effects on endogenous variables	Regression beta-weights	Unstandardized regression coefficients
Variances endogenous variables (and hence their converse, error variances)	Squared multiple correlations (ie., R^2).	Unreported
Variances of exogenous variables (and hence their converse, error variances)	Unreported	Variances

Adapted from Arbuckle and Wothke (1999) and Kline (1998)

[#] The covariance is a correlation between two variables, multiplied by each variable’s standard deviation.

Figures 6 and 7 provide the unstandardized and standardized output for the GEM Australia data.

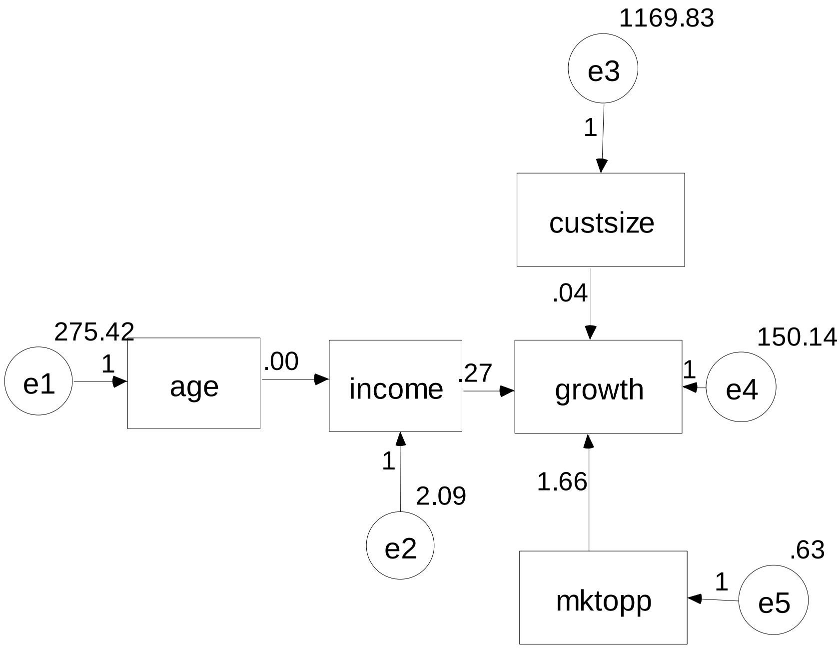

Figure 6: A reduced overidentified measurement model using the GEM Australia data: Unstandardized estimates
Source: Author, 2005.

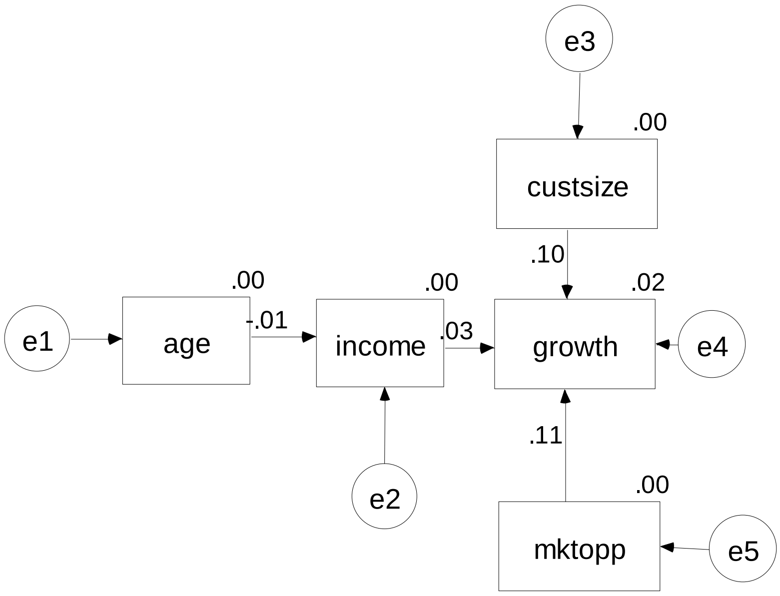

Figure 7: A reduced overidentified measurement model using the GEM Australia data: Standardized estimates
Source: Author, 2005.

Because of the iterative nature of the ML estimation procedure, computer programs often set starting values to start the estimation process. Good starting values enable ML to converge on a solution more quickly, whilst poor starting values may result in a solution that fails to converge. However, when starting values turn out to be poor, the researcher can rerun the estimation procedure with different starting values (Kline, 1998).

Starting values may be set for direct effects or error variances (of endogenous variables). If, for example, it is expected that the direct effect of one variable on another is moderate, then a standardized coefficient of .3 may be substituted as its starting value. Similarly, if it is expected 75% of the variance in a manifest, endogenous variable can be explained by the model, then a standardized coefficient of .25 may be substituted as a starting value for its error variance (Kline, 1998).

It was not necessary to set starting values for the GEM Australia data. Indeed, it took three iterations and .22 seconds for AMOS Graphics to produce the initial solution.

Step 4b – Model Evaluation

Model evaluation involves the use of significance tests to assess the adequacy of model fit. However, evaluation of the full structural model takes place over two mini steps: The measurement component of the model is examined first, followed by examination of the full structural model (Byrne, 2001; Joreskog, 1993; Kline, 1998).

Model evaluation is thus accomplished by initially turning direct effects amongst latent variables into unanalysed associations (and hence, a measurement model) and assessing the results. If the results indicate the model has a reasonably good fit to the data, then those direct effects are restored and the full structural model is analysed. If not, the measurement model is modified and reassessed until good fit is obtained before continuing on. In this way, any source of poor fit is easily identified and rectified (Kline, 1998).

The GEM Australia model is a measurement model so only one step is required for its evaluation.

The commonly reported test statistics, summarised in table 4 and described below, enable evaluation of model fit at both the measurement level and full structural model.

Table 4
Commonly Reported Test Statistics Used to Evaluate Model Fit

Test Statistics	Critical value	Interpretation
Chi-squared Tests 1. Chi-squared goodness of fit test 2. Normed chi-squared test	 Chi-squared = n.s. Chi-squared/df <=3	 Good fit to the just-identified model. Good fit to the just-identified model.
Test Statistics Using Covariance Matrix 1. Goodness of fit index (GFI) 2. Adjusted goodness of fit index (AGFI) 3. Standardized root mean squared residual (SRMR)	 0.9 < GFI < 1 0.9 < AGFI < 1 0 < SRMR < 0.05	 Good fit to the just-identified model. Good fit to the just-identified model. Good model fit.
Comparisons with Independence Models 1. Normed fit index (NFI) 2. Non-normed fit index (NNFI) [aka the Tucker-Lewis Index] 3. Comparative fit index (CFI)	 0.9 < NFI < 1 0.9 < NNFI < 1 0.9 < CFI < 1	 Percent improvement over null model Percent improvement over null model Percent improvement over null model
Root mean square error of approximation (RMSEA)	0 < RMSEA < .08	Good model fit.

(Adapted from Arbuckle & Wothke, 1999; Byrne, 2001; Kline, 1998).

The most commonly reported test statistic is the chi-squared goodness of fit test. Its degrees of freedoms are calculated by subtracting the number of parameters in the just-identified model against that of the overidentified model. As such, this test assesses the adequacy of the overidentified model in relation to the just identified model. If significant, it implies that the overidentified model is significantly different from the just-identified model. Moreover, a significant result suggests that the model would have been a better fit to the data should it contain more of the additional parameters available in the just-identified alternative. Conversely, a non-significant result implies that the overidentified model conveys just as much information as the just-identified model (Kline, 1998). Thus, the goal is to find a non-significant goodness of fit result.

The chi-squared goodness of fit test is affected by sample size, with larger samples finding significant differences more often than warranted and smaller samples finding nonsignificant differences more often than warranted (Arbuckle & Wothke, 1999; Kline, 1998; Ullman, 2001). For this reason, the normed chi-squared test, which divides the goodness of fit chi-squared test by its degrees of freedom, tends to be the second most commonly reported statistic. A normed chi-squared test result of 3 or less is non-significant, suggesting that the overidientified model conveys the same information as the just-identified model (Arbuckle & Wothke, 1999; Kline, 1998).

Table 5 below shows the chi-squared statistics for the GEM Australia Model. This table has been excerpted from AMOS output. As can be seen from this table, the default model has nine parameters, of a possible 15. [The maximum number of parameters can be confirmed by counting the number of variances and covariances in Table 2. Similarly, examination of figures 5, 6 or 7 will confirm that the model required estimation of four pathways and five variances (or error variances)]. The chi-squared goodness of fit statistic for this model is 41.484. With six degrees of freedom, this model has produced a significant result. The normed chi squared goodness of fit of 6.914 is still significant. Both statistics suggest that more parameters are required to adequately explain the relationship amongst the GEM Australia variables.

Table 5

Chi Squared Statistics for the GEM Australia Reduced Overidentified Model

Model	NPAR	CMIN	DF	P	CMIN/DF
Default model	9	41.484	6	.000	6.914
Saturated model	15	.000	0		
Independence model	5	88.753	10	.000	8.875

The goodness of fit index (GFI) is a marker of the proportion of covariances explained by the model's constructed covariance matrix. The GFI ranges from 0 to 1, with high scores reflecting a good fit to the data (Arbuckle & Wothke, 1999; Byrne, 2001; Kline, 1998). Similarly, the adjusted goodness of fit index (AGFI) adjusts the GFI to account for model complexity. It is interpreted in the same way as the GFI (Arbuckle & Wothke, 1999; Byrne, 2001; Kline, 1998; Ullman, 2001).

The standardized root mean squared residual (SRMR) also compares the observed covariance matrix against that constructed through the model, where the greater the residual differences between both matrices, the higher the SRMR scores (Arbuckle & Wothke, 1999; Byrne, 2001; Kline, 1998; Ullman, 2001). Thus, the objective in this case, is to have a SRMR score approaching zero (Byrne, 2001; Ullman, 2001).

Table 6 below provides a further excerpt from the AMOS output. As can be seen from this table, the GFI and the AGFI statistics are both greater than .9. The reproduced covariance matrix adequately fits the original covariance matrix.

Table 6

Covariance Statistics for the GEM Australia Reduced Overidentified Model

Model	RMR	GFI	AGFI	PGFI
Default model	6.964	.992	.979	.397
Saturated model	.000	1.000		
Independence model	13.145	.982	.974	.655

The normed fit index (NFI), the non-normed fit index (NNFI) [also known as the Tucker-Lewis index] and the comparative fit index (CFI) compare the fitted model to a 'null', or independence, model where all manifest variables are assumed to be uncorrelated. These three test statistics thus consider whether the model is a significant improvement over the null model. High scores indicate a significant improvement over an independence model, whilst low scores indicate no difference from an independence model (Arbuckle & Wothke, 1999; Byrne, 2001; Kline, 1998). Of the three test statistics, the CFI is less affected by sample size, whilst the NNFI adjusts the index for model complexity (Byrne, 2001; Kline, 1998).

Table 7 provides the statistics for model comparison against an independence model. As can be seen from the NFI, TLI and CFI statistics, there is some improvement over the independence model. However, this improvement has not reached the critical range (of .9 or greater).

Table 7

Statistics to Compare the GEM Australia Model Against a Baseline

Model	NFI Delta1	RFI rho1	IFI Delta2	TLI rho2	CFI
Default model	.533	.221	.571	.249	.549
Saturated model	1.000		1.000		1.000
Independence model	.000	.000	.000	.000	.000

The root mean square error of approximation (RMSEA) examines model fit, adjusted for model complexity with lower values indicating better model fit (Arbuckle

& Wothke, 1999; Byrne, 2001). AMOS also provides 90% confidence intervals and probability values for RMSEA, providing further evidence of model fit adequacy. The tighter the confidence intervals around the RMSEA value, the greater the confidence in adequacy of model fit. RMSEA confidence intervals are affected by sample size and model complexity and thus needed to be considered with caution. Probability values of .5 or greater suggest adequacy of model fit (Byrne, 2001).

Table 8 provides the RMSEA statistics, along with confidence intervals. As can be seen from this table, the RMSEA for this model is .055, with a true range of .04 to .07. From this alone, the conclusion would be drawn that there is adequate model fit. However, the probability of .29 is less than .5, suggesting that there is a need to treat this statistic with caution.

Table 8

RMSEA Statistics for the GEM Australia Reduced Overidentified Model

Model	RMSEA	LO 90	HI 90	PCLOSE
Default model	.055	.040	.071	.290
Independence model	.063	.051	.075	.035

Irrespective of the results of the test statistics chosen, it is possible to have a model that is statistically acceptable (because the tests indicate a good fit), but has a poor fit in different parts of the model, little predictive power, or poor theoretical value. It is thus common practice to report a number of test statistics, with the greater the number of tests supporting the model's fit, the greater the confidence with the chosen model (Kline, 1998). It is also advisable to assess whether model relationships are in the expected direction (Bollen & Long, 1993). Indeed, in the GEM Australia example, the interpretation of results have been inconsistent, with some statistics suggesting adequacy of model fit (eg, the GFI and AGFI, and possibly the RMSEA) while others suggest inadequacy (the chi-squared goodness of fit, normed chi-squared goodness of fit, NFI, TLI and CFI).

It is also of value to examine the residuals between the reproduced and original correlation matrices. Residuals are considered negligible if they have a value of 0.1 or less. Where this is the case, the model is considered to have a good fit to the data

(Kline, 1998). However, where this is not the case, the model would benefit from respecification. Similarly, standardized covariance residuals of 2.58 suggest the need for model respecification (Byrne, 2001). This topic is discussed in *Step 5: Model Respecification* below.

Table 9 shows the standardized covariance residuals for the GEM Australia reduced overidentified model. As can be seen, there are two coefficients with values exceeding 2.58. These suggest a need for model respecification.

Table 9
Standardized Covariance Residuals for the GEM Australia Model

	age	mktopp	custsize	income	growth
age	.000				
mktopp	-2.247	.000			
custsize	1.741	-.652	.000		
income	.000	4.694	-.197	.000	
growth	-3.329	.078	-.075	.481	.011

Step 5: Model Respecification

Model respecification builds on the work at step 4 and can be either theory-driven or exploratory. The alternate models approach is more likely to be theory-driven whilst the model-generating models approach is more likely to be exploratory by nature. The GEM Australia example is using a model-generating models approach. Thus, whilst the theory-driven/alternate models approach has been described, examples were provided only for the alternate models/exploratory approach.

Models are "trimmed" or "built" by removing or adding direct effects (Kline, 1998). Models are also modified by reconfiguring the direction of direct effects. However, in exploratory model respecification, there remains the decision as to which direct effects to modify. In both exploratory and theory-driven model changes, there is also a need to identify which changes have contributed to model improvement. Table 10 summarises the tools that are used in both theory-driven and exploratory approaches to model respecification. The remainder of this section describes six tools that are used to facilitate such decisions.

Table 10

Tools for Theory Driven and Exploratory Approaches to Model Respecification

Test Statistics	Critical Value	Interpretation
Theory Driven Model Respecification 1. Chi-squared goodness of fit test	Significant Nonsignificant	Accept model with more parameters Accept model with fewer parameters
Exploratory Approaches to Model Respecification 2. Residual matrices 3. Modification indices 4. Wald statistics or z-statistics 5. AIC 6. CAIC	RMAT resids > |.1| CVAR resids>|3.58| Largest sig. values n.s. lower value lower value	] Insert associated] variables] into model Drop parameter from model] Accept model with] lower value

(Adapted from Arbuckle & Wothke, 1999; Byrne, 2001; Joreskog, 1977; Kline, 1998)

In a theory-driven approach, the chi-squared difference test is used to assess whether additional or eliminated direct effects have improved the model. This test compares the initial chi-squared goodness of fit against the modified model's chi-squared goodness of fit test. Degrees of freedom for the chi-squared difference test is the difference between the degrees of freedom of each of the two chi-squared goodness of fit tests (Arbuckle & Wothke, 1999; Kline, 1998; Ullman, 2001). Interpretation of the chi-squared difference test differs on the basis of whether a direct effect was removed or added. In the case of model trimming, a significant chi-squared goodness of fit test suggests that the revised model has a poorer fit to the data than the original model, and hence, it should be reinstated. In the case of model building, however, a significant chi-squared goodness of fit test suggests that the revised model (with the additional direct effect) is an improvement over the existing model, and hence, should be retained (Joreskog, 1977; Kline, 1998). In other words, where the chi-squared difference test is significant, the model with the greater number of direct effects reflect the better fit to the data. Conversely, where the chi-squared difference test is non-significant, the model with fewer direct effects has a better, more parsimonious, fit to the data.

Whilst more than one direct effect can be added or deleted simultaneously, the chi-squared result reflects the composite of changes made to the model (Kline, 1998), and hence, it may not be clear as to which of these changes is positively contributing to the overall chi-squared result. However, undertaking model building or trimming changes one direct effect at a time may increase the risk of a Type I error. Thus, for theory-driven model changes, it is recommended that the composite of changes be evaluated as a whole. Then the researcher can return to the initial model and introduce each change one direct effect at a time, using the chi-squared difference test, at each step, to assess which modifications made a positive contribution to the model.

When model respecification is exploratory, five tools are used to aid model building or trimming decisions. The first involves the use of residual matrices (between the reproduced and original correlation matrix, or between the reproduced and original covariance matrix). Correlation residuals, with absolute values greater than 0.1, suggest the need for a direct effect between the pair of variables behind those high residuals. Similarly, large covariance residuals suggest the same need for an additional direct effect. However, as covariances retain the metric of the original variables, what is considered 'large' will vary from variable to variable (Kline, 1998). However, standardized covariance residuals of 2.58 or more may be considered large (Byrne, 2001).

The second tool is the modification index (MI). The MI is used to determine which direct effect, if included in the model, is likely to contribute to the explanation of the data. The larger the MI value, the greater the contribution of that direct effect to model improvement. Probability values are given for each MI so that the statistical significance of the contribution can also be assessed (Arbuckle & Wothke, 1999; Byrne, 2001; Kline, 1998).

The third tool is the Wald W statistic. This statistic is a model trimming tool in that it examines which parameter could be dropped, leaving no change to the goodness of model fit. Parameters are dropped when their associated Wald W statistic is nonsignificant (Kline, 1998; Ullman, 2001). Similarly, the z-statistic also indicates which pathways can be dropped from the model. It is interpreted in the same way as the Wald statistic (Arbuckle & Wothke, 1999).

Both the MI and Wald W statistics are affected by sample size and it is recommended that the absolute magnitude of the statistic be considered, along with the test's significance level (Kline, 1998).

Moreover, for purposes of clarity, it is recommended that all additional parameters be introduced into the model one parameter at a time. Then any parameters identified for deletion be introduced one at a time (Ullman, 2001)

The final tools are that of the Akaike Information Criterion (AIC) and Consistent Akaike Information Criterion (CAIC). AIC is used to compare different models. It is equivalent to the chi-squared goodness of fit test, adjusted for model complexity. In each case, the model with the lower AIC score is preferred. The CAIC adjusts for both sample size and model complexity and is interpreted in the same way as the AIC (Arbuckle & Wothke, 1999; Byrne, 2001; Kline, 1998; Ullman, 2001).

Tables 11 and 12 show the results from the modification indices and z-tests. The results from the standardized residuals were provided in Table 9. Table 9 suggests that including unanalysed associations between age and growth, and between market opportunities and income would improve the model.

Table 11 shows that introducing two unanalysed associations (e2 with e5, and e1 with e5) would significantly improve the model. However, there needs to be strong theoretical reasons for inclusion of unanalysed associations between error terms. Thus, this line of exploration was not continued. Table 11 also confirms the associations between age and growth, but suggests that a direct effect from age to growth would provide a small improvement to the model. Moreover, the association between market opportunity and income has also been confirmed (in both directions). And, whilst both provide a small improvement to the model, the direction from market opportunity to income would be greater. However, the direction from income to market opportunity makes greater sense. If, for instance, level of income was unsatisfactory, market opportunities might be sought elsewhere.

Table 12 indicates that the direct effects from age to income, and from income to growth were not significant and could be dropped from the model.

Table 11

Modification Indices for GEM Australia Reduced Overidentified Model

cov.			M.I.	Par Change
e5	<-->	e1	5.048	-.665
e2	<-->	e5	21.872	.121
e4	<-->	e1	10.938	-15.076

var.			M.I.	Par Change
mktopp	<---	age	5.048	-.002
mktopp	<---	income	22.033	.058
income	<---	mktopp	21.872	.190
growth	<---	age	10.938	-.055

Table 12

z-statistics for the GEM Australia Reduced Overidentified Model

			Estimate	S.E.	C.R.	P	Label
income	<---	age	-.001	.002	-.344	.731	par_2
growth	<---	income	.266	.190	1.400	.162	par_1
growth	<---	custsize	.037	.008	4.669	***	par_3
growth	<---	mktopp	1.665	.345	4.822	***	par_4

Taken together, these tables suggest that the model can be improved by building two direct effects (income to market opportunity, and age to growth). It can also be improved by trimming two direct effects (age to income, and income to growth). As recommended, model building was undertaken one direct effect at a time. Then, model trimming was undertaken one direct effect at a time. The impact of each step was assessed at the end of each step. This could be done by examination of the same statistics, as well as examination of the AIC and CAIC statistics. Table 13 reports the initial AIC and CAIC statistics. These were used to compare model solutions.

Table 13

AIC statistics for the GEM Australia Reduced Overidentified Model

Model	AIC	BCC	BIC	CAIC
Default model	59.484	59.538	109.852	118.852
Saturated model	30.000	30.091	113.946	128.946
Independence model	98.753	98.784	126.735	131.735

Table 14 provides summary AIC statistics for each of the four steps undertaken to improve the GEM Australia model. Whilst only the AIC statistics have been reported, similar improvements have been found across the board with the other exploratory tools, as well as with model fit statistics.

Table 14
AIC Fit Statistics at Each Step of Model Respecification, GEM Australia Data

Step	AIC Statistic
1. Build Direct Effect: Income to Market Opportunity	39.328
2. Build Direct Effect: Age to Growth	30.315
3. Trim Direct Effect: Age to Income	28.434
4. Trim Direct Effect: Income to Growth	28.360

After the fourth step, there were no large modification indices, suggesting no further additional direct effects to include. There were also no insignificant z-statistics, suggesting no further direct effects to remove. Moreover, the standardized residual covariance matrix had no new high entries, suggesting that the estimated covariance matrix was now a good fit to the original covariance matrix. All the fit statistics except one (NFI) were now in the expected direction, indicating that the model was a good fit to the data. This would suggest that no further modifications to the model are necessary. Figure 8 provides the standardized coefficients for the GEM Australia model.

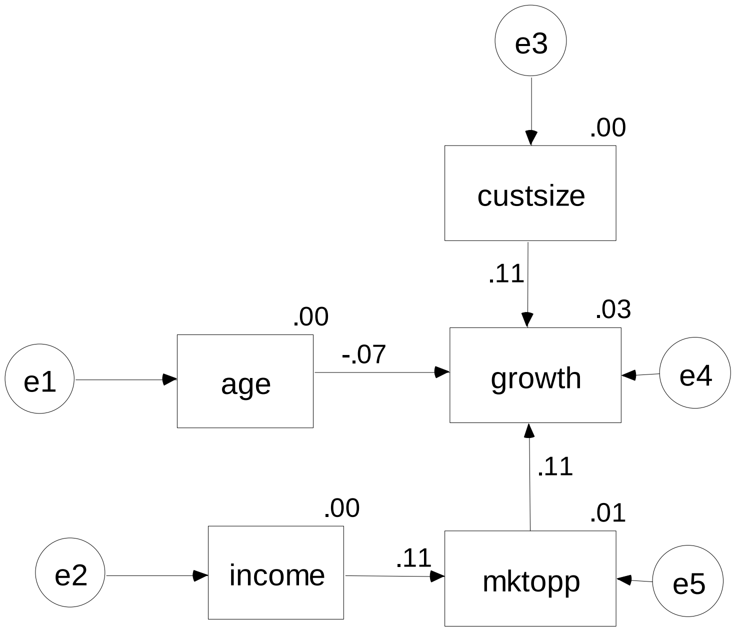

Figure 8: Final overidentified measurement model using the GEM Australia data: Standardized estimates
Chi-square=10.360, df= 6, p= .110.
Source: Author, 2005.

A final word of warning: Use of exploratory tools does not ensure the identification of the “true” model. Moreover, with the potential to test many model changes one step at a time, the risk of making a Type I error is increased. Thus, even when taking an exploratory approach, it is advisable to limit model changes to those expected on the basis of theory or past research (Arbuckle & Wothke, 1999; Kline, 1998).

It is also recommended that the final model be tested on a second sample (Ullman, 2001). Indeed, in the case of the GEM Australia data, it is worthwhile repeating the study on GEM data collected in future years. Moreover, if sample size permits, it may be worthwhile assessing whether the same model holds across gender and states.

Conclusions and Implications

SEM is a powerful tool. It enables confirmation of factor structure amongst a set of variables. It also goes beyond identification of the set of variables that contribute to the prediction of another variable. It identifies the direction of relationships amongst that set of variables. It can also accommodate moderator (interaction terms) within the model. Within the constraints of sample size, SEM can also model mediators. In the absence of missing data, bootstrapping procedures can also be performed.

Whilst SEM is relatively robust from deviations from statistical assumptions, these assumptions need to be assessed. Difficulties will also arise when data does not meet assumptions, such as when the sample collected is too small to test the hypothesized model. Decisions would need to be made when the requirements to redress one assumption conflicts with requirements for another. This was the case in the GEM Australia example: FIML estimation required the use of the original dataset, whilst handling ordinal data required the use of a covariance matrix as input. As is the case with all research, the rationale for decisions made need to be recorded and acknowledged.

With the powerfulness of the tool, comes complexity. Decisions made can become a matter of judgement. For this reason, it has become standard practice to report sufficient results so that findings can be reproduced (Kline, 1998), as well as facilitate exploration of previously unexplored avenues. As such, reported results need to include the correlation matrix, means and standard deviations (Kline, 1998), or a covariance matrix with means. It is also advisable to include a copy of the research model(s) described (Kline, 1998).

Acknowledgements

Papers are never developed in isolation and, this paper is no exception. Discussions with a number of fellow PhD candidates within the AGSE have contributed to this final product, along with those experienced with statistics but not necessarily with SEM. I would like to express my appreciation for discussions and reflections shared with Dr. Michael Abramson, Loris Gillin, Dr. Russell Langley, Raveendra Nayak, Daniel Oertquist, Susan Rushworth, Professor Chris Selverajah, Janusz Tanas and Dr. Shahid Yamin.

Finally, our heartfelt thanks are given to Alan O'Connor for access to the GEM Australia Project data; without which, this paper could not have been completed.

Abbreviations and Acronyms Used

AGFI	Adjusted Goodness of Fit Index
AIC	Akaike Information Criterion
AMOS	**A**nalysis of **MO**ments **S**tructure
CAIC	Consistent version of Akaike Information Criterion
CFI	Comparative Fit Index
df	degrees of freedom
FIML	Full Information Maximum Likelihood
GIGO	Garbage In, Garbage Out
GFI	Goodness of Fit Index
MAR	Missing at Random
MCAR	Missing Completely at Random
MI	Modification Index
ML	Maximum Likelihood
NFI	Normed Fit Index
RMSEA	Root Mean Square Error of Approximation
NNFI	Non-normed Fit Index (also known as the Tucker-Lewis Index)
SMC	Squared Multiple Correlation
SRMR	Standardized Root Mean Squared Residual
SEM	Structural Equation Modelling

References

Arbuckle, J. L., & Wothke, W. (1999). *Amos 4.0 User's Guide*. Chicago, IL: SPSS Inc.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182.

Bollen, K. A., & Long, J. S. (1993). Introduction. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (Vol. 154, pp. 1-9). Newbury Park: SAGE Publications.

Breckler, S. J. (1990). Applications of covariance structure modeling in psychology: Cause for concern? *Psychological Bulletin, 107*(2), 260-273.

Byrne, B. M. (2001). *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming*. Mahway, New Jersey: Lawrence Erlbaum Associates, Publishers.

Curran, P. J., Bollen, K. A., Paxton, P., Kirby, J., & Chen, F. (2002). The noncentral chi-square distribution in misspecified structural equation models: Finite sample results from a monte carlo simulation. *Multivariate Behavioral Research*, *37*(1), 1-36.

Hindle, K., & O'Connor, A. (2005). *Westpac GEM Australia: A Study of Australian Entrepreneurship in 2004. Australian Graduate School of Entrepreneurship Research Report Series* (No. 2(1)). Hawthorn, Melbourne: Swinburne University of Technology.

Joreskog, K., G. (1977). *Structural equation models in the social sciences: Specification, estimation and testing.* Paper presented at the Applications of Statistics: Proceedings of the Symposium Held at Wright State University, Dayton, Ohio, 14-18 June 1976.

Joreskog, K., G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (Vol. 154, pp. 294-316). Newbury Park: SAGE Publications.

Kline, R. B. (1998). *Principles and Practice of Structural Equation Modeling.* New York: The Guilford Press.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical Analysis With Missing Data*. New York: John Wiley & Sons.

McDonald, R. P., & Bolt, D. M. (1998). The determinacy of variables in structural equation models. *Multivariate Behavioral Research, 33*(3), 385-401.

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin, 105*(1), 156-166.

O'Dell, L. L., & Cudeck, R. (1995). Relationships among measurement models for dichotomous variables and associated composites. *Multivariate Behavioral Research, 30*(1), 77-81.

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*(3), 581-592.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using Multivariate Statistics* (4th ed.). Boston: Allyn and Bacon.

Ullman, J. B. (2001). Structural equation modelling. In B. G. Tabachnick & L. S. Fidell (Eds.), *Using Multivariate Statistics* (4th ed.). Boston: Allyn and Bacon.

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural Equation Modeling: Concepts, Issues, and Applications* (pp. 56-75). Thousand Oaks: SAGE Publications.