



# Using Amos for structural equation modeling in market research

by Lynd D. Bacon, Ph.D.,  
Lynd Bacon & Associates, Ltd.  
and SPSS Inc.

Structural equation models (SEMs) describe relationships between variables. They are similar to combining multiple regression and factor analysis. SEMs also offer some important, additional benefits over these techniques including an effective way to deal with multicollinearity, and methods for taking into account the unreliability of consumer response data.

This paper introduces you to SEMs as well as Amos, a software tool distributed by SPSS Inc. Amos stands for “Analysis of Moment Structures.” Amos has a unique graphical interface, and was specifically designed to make fitting SEMs easier.

This paper will:

- Review some SEM basics and compare SEMs to multiple regression and factor analysis models
- Cover sample size requirements and two important SEM issues: model identification and model equivalence
- Explain why using SEMs for multicollinearity and measurement reliability provides important advantages
- Give two examples using Amos for structural equation models

The first of the two examples compares the importance of “satisfaction drivers” across customer segments. The insight we get from Amos tells us the importance of the drivers probably shouldn’t be compared directly, because the data aren’t really measuring the same variables in the segments.

The second example is a twist on some recently reported modeling of “brand halos.” The goal was to find out how much halos may contribute to customers’ perceptions of products from two manufacturers. The observations show the brands have “halos” of different strengths, and their effects are different for different product attributes.

SEMs, regression models and factor analysis

SEMs can include two kinds of variables: observed and latent. Observed variables have data, like the numeric responses to a rating scale item on a questionnaire such as gender or height. Observed variables in SEMs are also usually continuous. Latent variables are not directly observed, but you still want to know about them. To observe latent variables, you must build models express latent variables in terms of observed variables.

The latent variables in SEMs are continuous variables and can, in theory, have an infinite number of values. Examples of latent variables in marketing include brand attitudes, customer satisfaction, perceived value, repurchase intentions and perceived quality. Household income might best be considered a latent variable in many applications.

Although various forms of SEMs have been developed, the vast majority express linear relationships between variables. This is also how variables are usually related in regression and factor analysis models — variables are expressed as weighted linear combinations of other variables.

Variables that depend on other variables are called “dependent.” Variables that do not depend on other variables in a model are called “independent.” When using SEMs, these variables are also called “endogenous” and “exogenous,” respectively.

Consider the linear relationship expressed by Equation 1. This equation says that perceived value for case “i” is the sum of the quality for “i” multiplied by the coefficient “a,” cost for “i” multiplied by the coefficient “b,” plus an “error.” The error term represents that part of perceived value for case “i” that is not captured by its linear dependence on quality and cost. When combined with some assumptions, the equation describes a model of value that may depend on quality and cost.

$$\text{value}_i = a * \text{quality}_i + b * \text{cost}_i + \text{error}_i$$

**Equation 1**

When we fit a model like the model in equation 1 to a data set, we’re trying to pick estimates for coefficients “a” and “b” that minimize some function of the errors across observations, given some assumptions about these errors. Note, our model assumes that all cases in the data set have the same values for “a” and “b.” They are fixed in the population.

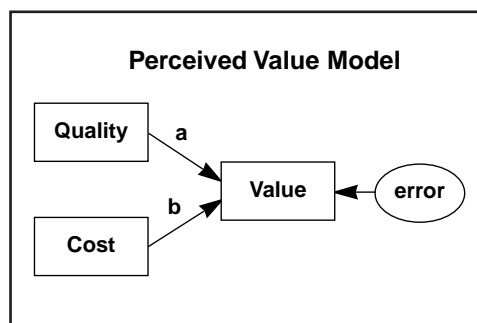
Equation 1 may look like a regression equation, without an intercept term on the right-hand side. The coefficients “a” and “b” represent the regression coefficients. “Value,” “quality” and “cost” are observed variables. “Error” is the difference between the observed and predicted values for each of the cases.

Or, you may see the equation as describing a factor model in which the observed variable called value “loads” on two factors called “quality” and “cost.” The error might be called “uniqueness,” instead. From this perspective, quality and cost are latent variables.

Introduction to SEM path diagrams

Figure 1 shows another way of describing the model in Equation 1. Figure 1 is called a path diagram. Path diagrams are a clear way of summarizing SEMs. You can draw path diagrams quite easily using the graphics tools in Amos. Amos also generates the necessary equation statements to fit the models you draw.

Observed variables are drawn as boxes, latent variables are drawn as circles or ellipses. Note, the error term in the path diagram is drawn as latent — errors are estimated, not measured directly. When one variable is believed to “cause” another variable, the relationship between the variables is shown as a directed or one-headed arrow, from cause to effect. Whether one variable “causes” another is an assumption that you make, not something the data can tell you.



**Figure 1. A path diagram for the value model.**

Sometimes covariation between two variables needs to be included in a SEM. This kind of undirected relationship is shown as a curved, two-headed arrow connecting the variables.

For each arrow, there may be an estimated loading or weight, like the coefficients “a” and “b” in Figure 1. Weights constrained to a particular value are often labeled with that value. Although it’s not shown in Figure 1, a coefficient of “1.0” would be specified for the effect of error on value. This constraint means the error has the same units of measurement scale as the brand rating. In a regression model, the residuals have the same scale of measurement as the dependent variable.

SEMs may also include one or more linear regression equations that describe how some variables depend on others. These are called structural equations. The collection of them is sometimes called the structural equations model, or the structural model in an SEM. The coefficients describing how dependent variables depend on independent variables are sometimes called path coefficients.

You can use latent variables in structural equations as dependent or independent variables. When you use a latent variable in a SEM, it is usually modeled using two or more observed variables called “indicator” variables.

For example, you want to model brand loyalty as a latent variable. You ask customers to make quantitative judgments about their use of a brand, their intentions to continue using the brand, and their willingness to recommend the brand to others. You could then use the responses on these indicator variables to model loyalty as a single latent variable. How each indicator variable related to loyalty would be expressed as a factor loading.

As you might imagine, once you’ve put together some structural equations and some measurement models, you can get a much more complicated model than what’s shown in Figure 1. Remember, these models are built using simple parts, even though they can get complex. You’ll find path diagrams are an effective way of summarizing even very extensive SEMs.

### Exploring confirmatory models

There is an important difference between the factor analysis models that most marketing research analysts use and the kind estimated for a SEM. Most applications of factor analysis are “exploratory,” meaning that the goal is to reveal the relationships underlying a set of variables. Sometimes, the objective is to reduce a set of variables to a smaller, more manageable, number. You can use either exploratory factor analysis (EFA), or principal components analysis (PCA) for this.

In the case of either EFA or PCA, the loadings of any observed variable on any factor can assume any value. That is, which variables load on which factors is not fixed, or constrained, in any way. What is constrained is the number of factors, and often the correlations between the factors are constrained to zero. But the observed variables are allowed to load on any and all factors.

When using SEMs, you take an approach like confirmatory factor analysis (CFA). You specify which loadings and path coefficients are free to vary, and which are to be fixed at particular values. You also specify whether variables are independent of each other, or whether they co-vary. See Bollen (1989) for a more detailed discussion of the various differences between exploratory and confirmatory factor analysis.

### Getting fit

The procedure for estimating and assessing the fit of SEMs is similar to what you do with other statistical models. First, examine your data and check to see if the necessary distributional assumptions are reasonable, and do what you can about them. The most common estimation method for SEMs is maximum likelihood (ML) estimation. A key assumption for this method is multivariate normality for the exogenous variables.

Next, describe one or more models to Amos, indicating the estimation method along with other options. Amos attempts to fit your model to the data. Amos's goal is to provide the best estimates of the freely varying parameters based on minimizing a function that indexes how well the model fits, and subject to the constraints you have defined. Amos gives you goodness-of-fit measures to help you evaluate your model's fit.

After inspecting the results, you can adjust particular models and try to improve the fit. You can also compare competing models to find out which is better. In general, it's wise to consider more than one model for any given data set. Amos provides extensive model fit diagnostics, including a large number of the fit indices used in the SEM literature. See the Amos reference manual Arbuckle (1997), Bollen (1989) and Tanaka (1993) for descriptions of these fit measures.

Comparing SEM models is the basic method for testing all but the simplest of hypotheses. Differences between nested models are usually evaluated using the difference between their chi square ( $\chi^2$ ) statistics relative to the difference in their degrees of freedom. One model is nested within another when it is a simplification of the other due to one or more additional constraints.

Here is an example of nested models. Assume that Equation 1 represents a model that has been fit. It's basically a regression model with no intercept. Equation 1 would be easy to fit using Amos.

Now, consider the model represented by Equation 2. Equation 2 is the same as Equation 1, except the coefficient of the cost variable "b" has been constrained equal to zero.

$$\text{value}_i = a * \text{quality}_i + \text{error}_i$$

### Equation 2

Equation 2 is nested within the Equation 1 model. Equation 2 uses one less degree of freedom than the Equation 1 model. If you had fit the two models using least squares regression, you could do an F test for the difference in the model  $R^2$ s. But, using Amos you would compare their  $\chi^2$ s.

You can make nested models using other kinds of constraints. For example, if model A lets Y and X be correlated, and model B requires their correlation to be 0.50, then B is nested within Y.

Comparing models that aren't nested, isn't as easy. This is usually done using measures of fit for which there isn't a convenient statistical theory, such as the  $\chi^2$  statistic. The inconvenience of comparing non-nested models is not specific to SEMs.

### Sample size and power

An important consideration for any market researcher is how large a sample is needed. Observations are expensive, and you don't want to incur unnecessary costs. On the other hand, you want to have enough data so that important differences or relationships can be observed, should they exist.

Methods exist for doing power analysis for SEMs. They focus on single parameters, sets of parameters or overall model fit. Like any inferential method, the sample size you choose should depend on the size of the differences or associations important for observation.

The following examples use unusually large sample sizes. Published SEM applications typically use 200-400 cases to fit models that have from 10-15 observed variables. An initial guess of 300 isn't a bad starting point for fitting simple SEMs. As with other techniques, it's often worthwhile to do some power analysis before collecting data. SamplePower, from SPSS Inc., facilitates power analysis and sample size determination.

### Two important SEM Issues

With the power and comprehensiveness of SEMs come two methodological issues. The first is whether the parameters you want to estimate are identified. That is, can you obtain a unique estimate for the parameter. When all parameters of a model are identified, the model is said to be identified.

The identification problem is much like high school algebra, trying to figure out whether there were enough independent equations to solve for the X's, Y's and Z's. Sometimes, solving for X, Y and Z is not easy to figure out.

Ideally, you want all your model parameters to be identified. There are heuristics available to help you determine whether a SEM is identified. Amos can detect and notify you of a range of identification problems. Amos also offers suggested remedies. You can sometimes fix an identification problem by using more constraints.

The second issue is model equivalence. Two SEMs are equivalent if they predict the same values from the same data. What you analyze when fitting SEMs is either a covariance matrix or a correlation matrix. Sometimes the observed means are used as well, when intercepts or factor means are estimated. Any two SEMs that predict the same moments (i.e., covariances, means, and so on), are equivalent.

At present there is no known comprehensive procedure for enumerating all possible equivalent models for any SEM you might specify. To deal with this, you often need to rely on information beyond your data to help choose the “best” models. This information may be from prior research, knowledge about the circumstances of data collection, managerial beliefs or your intuition. Equivalence isn’t much different from knowing which variables are dependent on others, and which are not. Most of the time your data can’t tell you; the knowledge has to come from experience.

What do I gain from using SEMs?

At this point you may have concluded that using SEMs might be painful, so why bother? Why not use easier and possibly less time consuming methods instead? There are three particularly important reasons, especially for market researchers.

First, many important marketing variables are latent. Market researchers often try to estimate latent variables with only a single observed measurement, and the reliability of this measure is usually unknown. The lower the measurement reliability, the less likely you are to observe relationships between the latent variable and other variables. By using SEMs with multiple indicator variables, you can model important latent variables while taking into account the unreliability of the indicators.

Second, customer evaluation, perception or behavior measures may (also) have low reliability, and failing to take this into account can hurt you. Assume that your dependent measure is a perceived value rating for a product, and you want to predict the value rating using a least squares multiple regression model. Your predictors are variables such as customer perceptions of the product across various attributes.

Now, even if your perceived value measure is perfectly reliable, unreliability in your predictor variables can still seriously mislead you. If only one predictor is unreliable, then its true regression coefficient will be underestimated. Unfortunately, the coefficients for the completely reliable predictors will also be biased. The bias can be either upward or downward (Maddala 1977).

This is a well-known problem in econometrics and in structural equation modeling. If one or more of your predictor variables are unreliable, not only can the size of your coefficients be wrong, but their signs can be incorrect also. That is, you may find that predictors you’d expect to be positively related to your dependent variable end up with negative coefficients, or vice versa. If you have analyzed any satisfaction data by regressing ratings on other ratings, and have gotten coefficients with unexpected signs, unreliability in your predictor variables might be one culprit. Model mis-specification may be another.

It’s worth noting that separate, bivariate regressions may mitigate this problem a bit, but they will not get rid of the difficulties caused by unreliable measures. As an example, consider using bivariate correlations to do “driver analysis” of customer satisfaction. This analysis computes correlation coefficients between an observed satisfaction measure and different attribute evaluation ratings. The coefficients are then used to rank the attributes in terms of their importance. This helps managers decide where to focus their marketing efforts.

Now, consider that all these measures have unknown reliability and their reliability might differ quite a bit. If this is the case, the ranking produced by a “driver analysis” may be false. And, how incorrect you don’t know. A better idea is to try modeling unreliability using SEMs, and thereby take unreliability into account explicitly.

Third, SEMs can be a powerful method for dealing with multicollinearity in sets of predictor variables. Multicollinearity is nothing more than when two or more variables are not independent. It’s a matter of degree and is diagnosable. When the variables are used as predictors, and their interdependence is strong enough, model results are poor and misleading.

Three different approaches exist to deal with multicollinearity: (1) ignore multicollinearity; (2) remove multicollinearity by using data reduction methods like principal components analysis; (3) model multicollinearity. The first two approaches can create serious problems, modeling is the best choice.

Analysts have modeled multicollinearity in different ways. One view taken is that multicollinearity in survey data results from asking the same respondents for multiple answers in the same context. Given this perspective, multicollinearity can be modeled as a “method factor” using SEMs. It’s a kind of context effect.

Another perspective is that multicollinearity can be due to a “halo” around, or general impression of, the objects we ask our respondents to evaluate. For example, brand halos. Brand halos are a latent construct that most marketers believe exist for branded products. A brand halo can be thought of as a component of brand equity and can be modeled using SEMs (Dillon et al. 1996).

### SEM examples using Amos

The first example examines the relative importance of service and product quality to overall satisfaction. The second example models the effects of “brand halos” on brand ratings. Hopefully, they can show you some interesting ways to use SEMs in marketing research. The examples are not intended to be prototypes of how to fit and evaluate SEMs. There are many good references on SEM fitting and evaluation, several of which are listed in the Amos manual (Arbuckle 1997).

The data are from two surveys of off-road motorcyclists. Both studies were commissioned by the DirtyScooter Motorcycle Company<sup>1</sup> (DMC) in support of their strategic marketing plan. The surveyed cyclists had all purchased a new motorcycle within 12 months of receiving the mailed questionnaire.

In the first survey, Survey A, 6,000 questionnaires were mailed. The response rate was approximately 50 percent. For Survey B, 5,500 surveys were mailed to owners of DirtyScooters and to owners of TrailBomber motorcycles, DMC’s main competitor. The Survey B response rate was about 76 percent.

<sup>1</sup> Not the real name of the firm, of course. The real product was a consumer durable.



### Example 1: Quality, value and fruit

Hauser (1991) cautions against analyzing satisfaction data that combine customers who have bought different products. He points out that the importance of satisfaction “drivers” could vary substantially across such segments. Combining data from the segments may smear these differences and mislead decision makers into making poor resource allocations.

Clearly, customers aren’t randomly assigned to the products they buy. What they buy is indicative of differences between them. Combining and then analyzing their data to find out what’s important to them isn’t so much like comparing apples and oranges as it is like making a fruit cocktail. A simple summary, like a single spoonful, is unlikely to be very informative about all the ingredients.

You can make related mistakes by assuming that your measures truly measure similar feelings in customer segments that you want to compare. If your data don’t truly measure the same dimensions in your segments, then trying comparing segments on those dimensions doesn’t make much sense. Just because a rating scale produces numerical responses from two different customer segments doesn’t guarantee that the responses mean the same thing.

DMC had to address this issue in order to find out if the effect of product quality and service quality on perceived value depends on the brand of motorcycle purchased. DMC’s question was: do product and service quality have the same or different effects on perceived value for buyers of DirtyScooters and TrailBombers? To validly answer this question, DMC had to ensure that the as quality, performance and value measurements were the same across segments.

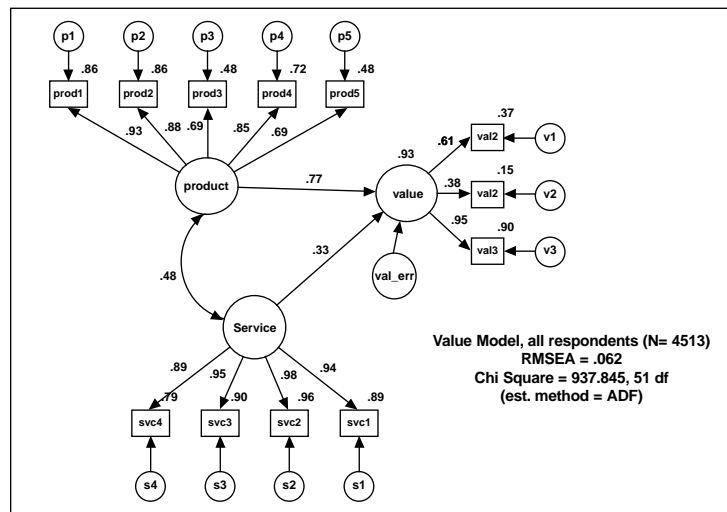
DMC had previously developed some survey items to measure these latent variables. In an earlier study, DMC used SPSS Factor Analysis and Reliability to reduce larger sets of candidate indicator variables for each of the latent variables. These items were included in Survey B, resulting in 4,513 usable responses. These consisted of 1,624 DirtyScooter owners and 2,889 owners of TrailBombers<sup>2</sup>

DMC specified a rather simple SEM reflecting the relationships thought to exist between the latent variables. They first fit this model, and variations on it, to the data of both segments combined to see how well the model performed across the entire sample. The input data was the variance/covariance matrix for the indicator variables.

Amos’ diagnostics did not encourage DMC to believe that the sample came from a multivariate normal population. So DMC used the “asymptotically distribution free” (ADF) estimation method described by Browne (1984), instead of maximum likelihood. Another approach would have been to use Amos’s built-in bootstrapping capabilities.

Figure 2 shows Amos’s path diagram for the final model. You can see that the product quality variable had five indicators, value had three and service quality had four indicators. The latent variable “val\_err” is the residual for the latent variable value’s structural equation. This equation includes the latent variables product and service as predictors. It’s similar to multiple regression equation with two predictors, only the covariation between the predictors is also modeled.

<sup>2</sup> Observations that had missing values weren’t included in the analyses. Amos has facilities for imputing missing values under different assumptions, but these weren’t used in this example.



**Figure 2. Path diagram for value model. Standardized results are shown. Quantities close to variables are their squared multiple correlations. Quantities near paths are standardized loadings or correlations. See text for details.**

The Root Mean Square Error of Approximation (RMSEA) fit statistic for the model was 0.062. A heuristic for using RMSEA was proposed by Browne & Cudeck (1993): values of 0.08 or smaller indicate acceptable fits. The value reported as  $\chi^2$  is the minimum value of the fitting function, and it is large compared to its df.  $\chi^2$  is quite sensitive to sample size, and rejects for smaller and smaller discrepancies as the sample gets bigger.  $\chi^2$  is probably less useful as an indicator than other model fit measures for this model based on samples as large as this.

Other fit statistics from Amos suggested the model's fit could be improved. In particular, one modification index statistic indicated that by allowing the errors p5 and v1 to co-vary, the fit would be improved considerably. It's always tempting to make a change like this. In the present case it wasn't clear how the changes could be justified based on how DMC's scales were developed. So, the change was not made.

You can see in Figure 2 that the path coefficient estimates for product quality and service quality are different. To run statistical tests of this difference, fit a version of the model in Figure 2 that constrained the path coefficients to be equal. This model is nested within the unconstrained model shown in Figure 2. The difference between the  $\chi^2$  statistics of these two models was 200.01, while the degrees of freedom for the difference was 1. Because of the highly significant  $\chi^2$  value, product quality has a bigger impact than service quality.

Up to this point, the model that has been fit to the whole sample. Recall that DMC's question had to do with whether the path coefficients for product quality or service quality differ across the customer segments. To address this question, fit the models for the two segments separately, but constrain the corresponding coefficients to be equal. Then, compare results by fitting the two segments' data without constraining their coefficients. Amos allows you to fit models for two or more groups simultaneously, and summarizes how well everything fits with a  $\chi^2$  statistic. Since the constrained version of the model is nested within the unconstrained version, it is possible to test whether the segments' path coefficients differ.

There is a problem with proceeding in this way, however. This plan does not allow for the possibility that the segments' models may differ in other ways, too. What if the indicator variables are related to their latent variables differently across the segments? What would a difference in the path coefficients mean? What if the covariance between product and service is different?

Under what conditions does comparing coefficients make sense? There are no hard and fast rules for doing this. A good rule requires the measurement models be the same across all groups. For this example, assume that the same model form is appropriate for both groups; this assumption should be given more thought in a real application.

But, what does it mean when measurement models are the same? Several definitions have been proposed. The most liberal definition is similar to what Jöreskog (1971) called a co-generic test — two co-generic tests have the same loadings for the same measured variables, but the error variances can differ.

This is a reasonable definition for the present example. You can use Amos to do simultaneous tests seeing if the co-generic test applies to your data. To do this, run two, two-group analyses. In one analysis, constrain the loading for each indicator variable, so the loadings are the same in both groups. In the second, allow the loadings to differ. Then, compare the results of these nested models.

Going through this procedure, the hypothesis that the loadings are the same in the two owner segments ( $\chi^2 = 43.88, 9 \text{ df}$ ) gets rejected. Therefore, comparing the path coefficients of the two segments may not make sense because the segments probably measure linear relationships between different latent variables.

It's worth comparing the magnitudes of the loadings across the groups to see how large the absolute differences are. Since the two samples used here are large, tests of their differences are very powerful. They may be rejecting the null hypothesis based on numerical differences that are small, substantively speaking.

If you decided to use a different definition of "same" for the measurement models, your end result may be different. You could, for example, also require that the errors in the measurement model have the same variances across groups. You might even require that the covariances between product and service be the same. What criteria you use, and how you go about evaluating those criteria, should depend on what you believe makes for meaningful comparisons. Bollen (1989) provides some strategies for evaluating model consistency across groups for an assortment of SEM types.

But where does this leave DMC? Separate models can still be estimated for the owner segments, and the coefficients within these models can be compared to each other. DMC found out that the owner segments differ in terms of how DMC's scales work. This merits further investigation. The differences may have strategic implications.

This first analysis helped DMC avoid making decisions based on possible invalid comparisons. If instead DMC had simply correlated scores computed from their three scales, and then compared these correlations across the segments, they could have

been comparing apples and oranges. This is an important benefit of using Amos to fit SEMs to the data: alternative methods would have led DMC to comparing things that weren't comparable.

### Example 2: Brand halos and brand evaluations

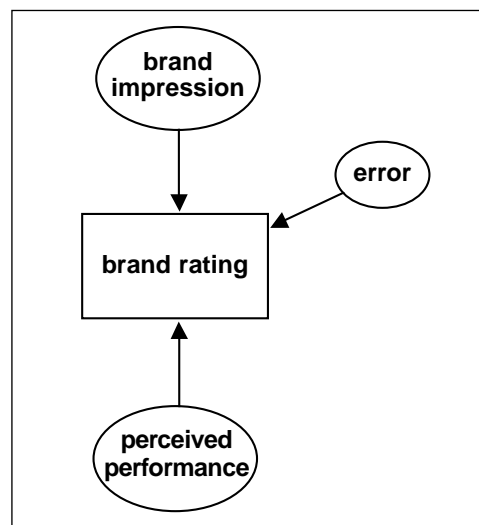
Dillon, Kumar, Walker & White (1996) and others have pointed out that brand attribute ratings may reflect more than just objective assessments of how different brands perform on different attributes. In addition to reflecting perceived performance, brand attribute ratings may also include the influence of a "brand halo" or "general impression." The idea that a brand halo can contaminate product evaluations makes it difficult for product managers to determine if customers' product perceptions reflect performance. And, how much of the evaluation is general "brand attitude."

A halo effect doesn't have to be the same for competing brands. That's why halos are an interesting and important problem for marketers. How useful is a perceptual map showing how brands are positioned, when you can't ensure that a brand's position is due to performance, or a general impression.

Dillon et al. (1996) described different kinds of models for halo effects. This example examines a variation on their additive model. In this model, each brand rating is a weighted linear combination of a general brand factor score, an attribute factor score and a measurement error. The weights are the factor loadings. The goal is to uncover how much variation in an observed brand rating can be attributed to the perception of that brand's performance on the attribute, and how much is due to the brand's halo.

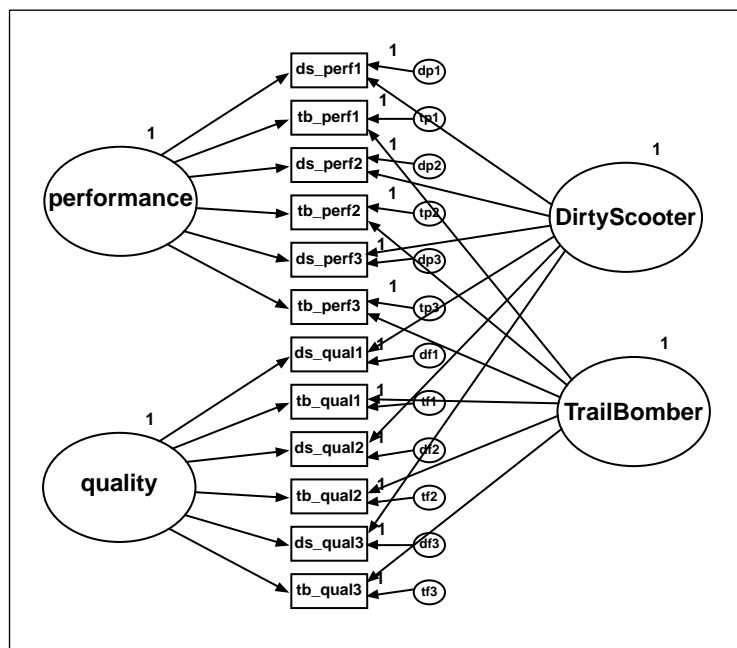
Figure 3 is a path diagram for a single observed rating in the model. Notice that each brand rating combines the effects of a brand impression, perceived performance, and error. "x" and "y" represent loadings that need estimation. Note, the scale of the error term has been set equal to the scale for the rating. That's what the "1" does on the path from the error to the brand rating.

Twelve questions in Survey A asked respondents to rate two brands: DirtyScooter and TrailBomber. Six of the 12 questions dealt with motorcycle performance, and six with "fit and finish" quality, such as assembly and paint. The importance of these two attributes was indicated prior to DMC research.



**Figure 3. Path diagram for a single brand rating.**

Figure 4 shows the basic model, estimated within Amos using confirmatory factor analysis. The halo, attribute and error effects are in ellipses and circles. The brand ratings are named to indicate both brand name and attribute. For example, “ds\_perf1” is the DirtyScooter rating on the first performance item. “tb\_qual3” is the TrailBomber rating on the third quality attribute.



**Figure 4. Path diagram for initial brand halo model. See text for details.**

Note that in the model in Figure 4, the variances of the brand and attribute factors have been constrained to 1.0. The errors, or uniqueness, like “dr,” “tr,” and so on, aren’t allowed to be correlated. Brand factors and attribute factors aren’t allowed to co-vary, either.

It might have occurred to you that you could fit an additive model using only a single rating for each attribute and brand. There would be twice as many factors as ratings, and each rating would depend on a unique pair of brand and attribute factors. Dillon, et al. (1996) give an example of this model.

A halo model of this sort is very much like a psychometric “multi-trait/multi-method” (MTMM) model. MTMM models are usually fit to correlation matrices. They are known to be prone to identification problems (Wothke 1996). These problems usually show themselves in solutions that produce negative values for predicted variances, or with the fitting program not converging to a solution, even after a large number of iterations.

The model in Figure 4 has another kind of under-identification. The loading signs depend on what is constrained (loadings or variances) to set the scales of the factors. The model has a proneness to under-identification in common with many factor models that have more than one factor influencing each observed measure (Wothke 1996). Here, the size of the loadings is interesting and not their signs, so under-identification isn’t a debilitating problem.

Several variations on the Figure 4 model were examined using the ADF estimation method. Amos’s output indicated the loadings of TrailBomber’s first and second quality rating didn’t differ from zero. Based on an  $\chi^2$  test of a nested model difference, constraining the ratings to zero didn’t make the model’s fit worse.

The modification indices in Amos indicated the model's fit could be improved by allowing a correlation of the brand factors, and by correlating the attribute factors. Correlating the brand factors makes some sense. There may be individual differences in how much consumers rely on brand impressions, generally. How influential a brand's impression is for any particular consumer may be correlated with the influence of other brands' impressions. The estimated correlation between the brand factors was 0.193.

Correlating the performance and quality factors is sensible, too. The correlation simply allows for customers' perceptions of quality and performance to be related. The estimated correlation between these factors was only 0.179, so it can't be easily argued that they are really measuring the same thing.

With these changes, Amos's fit statistics indicated a reasonable fit. The RMSEA index was 0.047, with a probability of a "close" model fit of 0.846. The ratio of  $\chi^2$  to its degree of freedom was about 6.988, which isn't too large given the sample size. The Amos results did show a few large residuals for the covariances, suggesting improving the model's fit.

An interesting and important question is whether the brand and attribute effects differ. This question can be examined by constraining the appropriate loadings so they are equal between brands, refitting the model, and then comparing the refitted model to the unconstrained version. That procedure is left for you to do on your own.

It's probably more useful to know what proportion of variation in each rating the model indicates is attributed to brand halos, than what the loadings are. The question is, how much does each brand's halo contribute to customers' ratings of the brand?

This question is easy to answer using the model results. Amos provides the squared multiple correlation ( $R^2$ ) for each observed variable's rating. Each  $R^2$  can be interpreted as the proportion of the total variance in an observed rating that is explained by all variables that the rating directly depends on. In the final halo model, each observed brand rating depends on two (or fewer) variables.

Squaring each standardized factor loading for a rating in this model gives us the proportion of that rating's variance that is due to that factor. Recall that brand factors and attribute factors aren't correlated in the present model. If they had been correlated, computing the rating variance percentages would have been a bit more complicated. The calculations of the percentages would need to take the factor covariances into account.

Table 1 shows the results of these calculations. It gives the breakdown of total variance for each brand rating. Notice that performance ratings for DirtyScooter are influenced more by the brand effect than are the performance ratings of TrailBomber. The opposite seems to be true for the quality ratings. Both conclusions assume the model is correct. The marketing implications for DirtyScooter include the possibility that making objective improvements in performance may have little effect on perceptions, since the latter are determined mostly by the halo effect. The question remains as to what actually causes the halo. This is outside the scope of the study.

**Table 1. Sources of Brand rating variance according to the additive model**

Brand rating	Brand evaluated	Evaluation attribute	Item R squared	Squared loadings:	
				Brand	Attribute
ds_perf1	DS	perform#1	0.76	0.71	0.04
tb_perf1	TB	perform#1	0.80	0.40	0.39
ds_perf2	DS	perform#2	0.76	0.74	0.02
tb_perf2	TB	perform#2	0.70	0.41	0.30
ds_perf3	DS	perform#3	0.75	0.72	0.03
tb_perf3	TB	perform#3	0.70	0.48	0.22
ds_qual1	DS	quality #1	0.77	0.51	0.26
tb_qual1	TB	quality #1	0.78	0.78	0*
ds_qual2	DS	quality #2	0.88	0.56	0.32
tb_qual2	TB	quality #2	0.91	0.91	0*
ds_qual3	DS	quality #3	0.64	0.47	0.17
tb_qual3	TB	quality #3	0.54	0.54	0.00

**Notes: 0\* indicates the loading was constrained to be zero. The sum of the squared loadings for each item equals R<sup>2</sup>, except for rounding.**

There are three closing comments for this example. First, there are other model forms that could be used for this kind of application. Dillon et al. (1996) also describe a multiplicative, rather than additive, version. That is, the effects of brand and attribute multiply, rather than sum. Other model forms are suggested by the literature on MTMM models. See Wothke (1996) for examples.

Second, the construct validity of our model's brand and attribute factors is still a long way from being established. The issue is: whether the factors fit into the model really do measure brand impressions and evaluations. Assessing the factors' construct validity would require using other data to strengthen the argument that the factors are what you think they are. Clearly, it is very important to assess the construct validity before any model results are used for decision making. Dillon, et al. (1996) describe how they investigated the convergent validity of their brand factors. The issue of validity isn't specific to SEMs. Validity should be addressed whenever latent variables are being used, which in market research, is most of the time.

Finally, in the DMC survey, every respondent evaluated every brand on every attribute. This contrasts with the common practice of having respondents evaluate subsets of brands. Researchers often allow respondents to judge only a subset of brands, because of interview length limitations, self-reported lack of knowledge of some brands or other considerations.

When respondents only evaluate subsets of brands, however, the resulting data may be useless for many kinds of modeling. One reason is the large amount of missing data.

Also, what's missing is usually not "missing at random," meaning missing data is often determined by what brands customers recognize, what they say they have experience with, or some other factors that are not independent of their responses.

Given either of these conditions, it is unlikely that the missing values could be imputed in any meaningful way, even using Amos. This isn't a limitation of SEMs, per se. The data would be of limited value whether you were using EFA, discriminant analysis or another multivariate technique.

Where to go from here

This paper has only touched on some of the interesting and useful characteristics of the SEM methodology. Another aspect that should be mentioned is that you can model means on latent variables or intercepts in equations, and also test for differences between them across groups.

Consider again the perceived value example. Value was modeled as a latent variable. Assuming the same measurement model is appropriate for both owner segments, you could use SEMs to find out if the segments' perceived-value factor means were the same or not. You'd be answering the question "Are DirtyScooter owners and TrailBomber owners different in terms of the average value they receive from the motorcycle they've recently bought?"

If you are excited to begin using SEMs in your research, here's a recommendation for you. The best way to learn any modeling technique is by practice. Try using Amos to fit the kinds of regression or factor analysis models you are familiar with. Start with the simplest possible model specifications, and elaborate on them as your data and your judgment allow. Starting simply almost always makes diagnosing problems easier.

There are other products you could use to fit SEMs, of course. I think if you look at Amos, however, you'll find its graphical interface easy and convenient to use. It lets you draw the models you want to fit, and Amos can make publication-quality path diagrams for your reports and presentations. You'll also find that Amos's design is intuitive enough that you can easily remember how to use it, even when you haven't done so for a while. In my opinion, this is an important characteristic for software that research practitioners use.



## References

- Arbuckle, J. (1997), *Amos Users' Guide Version 3.6*, Chicago IL: Smallwaters Corporation.
- Bollen, K.A. (1989), *Structural Equations with Latent Variables*, New York: Wiley.
- Browne, M.W. (1984), "Asymptotically distribution-free methods for the analysis of covariance structures," *British Journal of Mathematical and Statistical Psychology*, 37, 62-83.
- Browne, M.W. and R. Cudeck (1993), "Alternative ways of assessing model fit," in *Testing Structural Equation Models*, ed. K.A. Bollen and Long. S., Newbury Park CA: Sage, 136-162.
- Dillon, W.R, A. Kumar, B. Walker and J. White (1996), "What's in a Brand-Attribute Rating: Halo or Equity or...?" *Advanced Research Techniques Forum*, Beaver Creek CO: American Marketing Association, June.
- Hauser, J. (1991), *Comparison of importance measurement methodologies and their relationship to consumer satisfaction*, M.I.T. Marketing Center Working Paper, vol. 91-1, Cambridge MA: Sloan School of Management, Massachusetts Institute of Technology.
- Jöreskog, K.G. (1971), "Statistical analysis of sets of cogeneric tests," *Psychometrika*, 36, 109-136.
- Maddala, G. (1977), *Econometrics*, New York: McGraw-Hill.
- Tanaka, J. (1993), "Multifaceted concepts of fit in structural equation models," in *Testing Structural Equation Models*, ed. K. Bollen and S. Long, Newberry Park CA: Sage Publications, 10-39.
- Wothke, W. (1996), "Models for Multitrait-Multimethod Matrix Analysis," in *Advances Structural Equation Modeling*, ed. G. Marcoulides and R. Schumacker, Mahwah NJ: Erlbaum, 7-56.

**About SPSS**

SPSS Inc. is a multinational software products company that provides statistical product and service solutions. The company's mission is to drive the widespread use of statistics. SPSS products and services are used worldwide in corporate, academic and government settings for all types of research and data analysis.

The company's four lines of business are: business analysis (including survey research, marketing and sales analysis and data mining); scientific research; quality improvement; and process management. Headquartered in Chicago, SPSS has more than 30 offices and 60 distributors serving countries around the world.

**Contacting SPSS**

To place an order or to get more information, call your nearest SPSS office or visit our World Wide Web site at <http://www.spss.com>

SPSS Inc. United States and Canada	+1.312.329.2400 Toll-free: 1.800.543.2185	SPSS Ireland	+353.1.66.13788
SPSS Bay Area	+1.415.453.6700	SPSS Israel Ltd.	+972.9.526700
SPSS Federal Systems (U.S.)	+1.703.527.6777	SPSS Italia srl	+39.51.252573
SPSS Argentina srl.	+541.816.4086	SPSS Japan Inc.	+81.3.5466.5511
SPSS Asia Pacific Pte. Ltd.	+65.3922.738	SPSS Korea	+82.2.552.9415
SPSS Australasia Pty. Ltd.	+61.2.9954.5660 Toll-free: 1800.024.836	SPSS Latin America	+1.312.494.3226
SPSS Belgium	+32.162.389.82	SPSS Malaysia Sdn Bhd	+603.704.5877
SPSS Benelux	+31.183.636711	SPSS Mexico Sa de CV	+52.5.575.3091
SPSS Central and Eastern Europe	+44.(0)1483.719200	SPSS Middle East and Southeast Asia	+971.4.525536
SPSS East Mediterranean and Africa	+972.9.526700	SPSS Newton	+1.617.965.6755
SPSS France SARL	+33.1.4699.9670	SPSS Scandinavia AB	+46.8.102610
SPSS Germany	+49.89.4890740	SPSS Schweiz AG	+41.1.266.90.30
SPSS Hellas SA	+30.1.7251925	SPSS Singapore Pte.	+65.2991238
SPSS Hispano-portuguesa S.L.	+34.1.447.37.00	SPSS Taiwan Corp.	+886.2.5771100
		SPSS UK Ltd.	+44.1483.719200

SPSS is a registered trademark and the other SPSS products named are trademarks of SPSS Inc. All other names are trademarks of their respective owners.

