Testing the Difference between Structural Models -- EQS

1. Degrees of Freedom

The data points of EQS are the variances and covariances generated by a specified set of variables. The number of data points is $n(n+1)/2$ where n is the number of variables. In the example I use below there are 8 variables, which therefore generate 36 data points. We lose degrees of freedom when we use these data points to estimate parameters. In EQS parameters are of three main types a) Regression coefficients, variances of the independent variables, and covariances of the independent variables. It should be noted that error terms have variances to be estimates (although as they are assumed to be uncorrelated with the independent variables they do not share covariances, and as they are themselves unmeasured varaibles we have to fix their scale therefore no estimate is made of B_e (it is fixed at 0). Given the number of parameters available it over-identification of the model is relatively easy -- we can try to estimate more parameters than are there are available data points. One way to overcome this problem is, in terms of covariances, to treat models in a similar fashion to normal regression, that is covariances are assumed to be fixed at 0. In EQS this occurs unless otherwise specified.

2. Models

In the example I present below I have used data for Thai women aged 25-34 in 1980. I estimate two models. The input for both models are the same -- a variance/covariance matrix of CEB, E480, W780, CHSCL80, CHWORK80, IM80, MIG, and AGE. The first model is the standard regression model and there the parameters to be estimated are:

1. The regression coefficients of E480, W780, CHSL80, CHWORK80, IM80, MIG, AND

AGE. Seven coefficients.

2. The variances for each of the seven independent variables plus the variance of the error term. Eight variances.

Combined this gives me a total of 15 parameters to estimate and, therefore a total of 21 degrees of freedom (as I have 36 data points).

The second model is the latent variable model in which E480 and W780 are used to define an unmeasured latent variable referred to here as F1 and where CHSCL80 and CHWORK80 are used to define a latent variable called F2. As it is the unoberseved variable which generates the observed indicators each of the four variables, E480, W780, CHSCL80, and CHWORK80, become dependent variables, with each equation containing a regression coefficient to be estimated (and a variance of the coefficient - which we however, set to 1 in order to determine the scale of the latent variable) and an error term which has a variance to be estimated. Therefore the parameters to be estimated are:

1. A total of 9 regression coefficients in 4 equations.

2. The variances IM80, MIG, and AGE, in equation 1. The other two independent variables (which appear a total of 6 times combined) F1, and F2, have their variances fixed at 1. The are also 5 error variances to be estimated. This results in a total of 8 variances to be estimated.

Combined this gives 17 parameters to be estimated from the 36 data points, therefore there remain 19 degrees of freedom.

3. Testing the Equivalency of Models

The relative power of one model compared to another can be tested when the models are *nested*. The simplest definition that I could find of what constitutes nested models comes from Bishop et al (1975: 524) where they state that models are nested when the parameters of one model are a subset of the parameters of a second model. Bentler provides a similar definition on p28 of the EQS manual "In the simplest and most typical application, two models would differ in that one model would contain extra parameters beyond those provide by the other model; all other parameters would be the same". Under such a situation a simple difference of Chi-Squared test, described by Bentler on page 28, can evaluate the relative strengths of the models.

It is clear that the two models I have presented here are not nested. Further, I think that in a situation in which one model contains independent variables that in a second model become dependent variables (through the construction of latent variables) it is not possible to have nested models. As pointed out by Bentler and Chou (1987: 94) "there is little agreement on methods for evaluating the relative merits of models that are not nested". A non-formal test (ie. significance values for differences between models have not yet been developed) is available to evaluate the fit of non-nested models.

Table 1

Standardized Solution of EQS Models

	Regression Model	Latent Variable Model
Women's Roles:		
Proportion in Grade 4	-0.038	
		-0.074
Proportion Working	-0.012	
Children's Roles		
Proportion at School	0.009	
		-0.054
Proportion Working	-0.025	
Infant Mortality	0.047	0.046
Age	-0.024	-0.024
Migration	0.374	0.374
Chi-Squared	26,729	22,187
DF	21	19
Bentler-Bonnet Fit	0.994	0.995