HomeWork on Estimation

Chi-Lin Yu (R05227101)

The present work aimed to demonstrate whether a given model is an identified one through the results of statistical softwares. As shown in the given figure, the model had one exogenous variable, two endogenous variables, and six indicators (Figure 1). They were respectively named as ξ_1, η_1, η_2, and y_1 to y_6. In the model, the latent variables ξ_1, η_1, and η_2 were scaled by fixing one of their loadings to 1. According to Figure 1, we understood that the degree of freedom (df) of this model is 7, shown by $p(p+1)/2 - q$, where p is the number of indicators and q is the number of distinct free parameter.

We used the `lavaan` package, a free open source package in `R` that provide a good platform for latent variable analyses (i.e., structural equation modeling; SEM), to analyze the data. The data was a covariance matrix with the sample size of 200 (N = 200). The maximum likelihood estimator of SEM, which assumes multivariate normality, was used in the present work to estimate each parameter and obtain the overall model fitting. The optimization procedure used here was the PORT routines with Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, one of the quasi-Newton methods with Hessian approximation, in the `nlminb` package in `R`. Note that all the code, results, and figures would made available upon request.

The results demonstrated that we failed to discomfit the present model ($\chi^2(7, N = 200) = 4.080$, $p = 0.771$), suggesting that the model implied covariance matrix is somehow close to the given sample covariance matrix. Other fit indexes also suggested fair overall model fitting: comparative fit index (CFI) = 1; tucker-lewis index (TLI) = 1 ; root mean square error of approximation (RMSEA, 90% confidence interval) = 0 (0-.059); standardized root mean square residual (SRMR) = .024.

The Estimates of parameter are shown in Figure 2 & Table 1. Although the point estimates could be computed, the standard error (SE) were not obtainable. To be more specific, neither the SE of loadings nor the SE of variances were not obtainable. It was an important sign, implying that the model or the data in the fitting has some problems. In this case, since the SE estimates were not obtainable, we suggested that the model used here was not identified. In another viewpoint, if we only considered the higher order factor structure part of this model (or the structural part), the model could be regarded as a two indicator model (one factor with two indicators). It was obvious that a two indicator model without relationships with other factors (e.g., no other higher order factors in this case) would not be identified. Likewise, if we applied two step methods to analyze the data, we could obtain a 2×2 covariance matrix with 3 unique data points, implying the parameter in the higher order (structural) part could not be appropriately estimated. Furthermore, we could also take $\sum(\hat{\theta})$ as an input matrix or take another sets of starting values to see whether the results remain the same. Although these are not required in the present study due to the obvious under-identified situation, it could still serve as useful empirical checks.

In brief, even if the degree of freedom of the model was bigger than 0, the model fit indexes could be computed, and the point estimates of parameters could be calculated, we could still conclude that this model was an unidentified one from at least two perspectives, including the graphical (i.e., theoretical/mathematical) one and the numerical one (i.e., the SE estimates were not obtainable.)

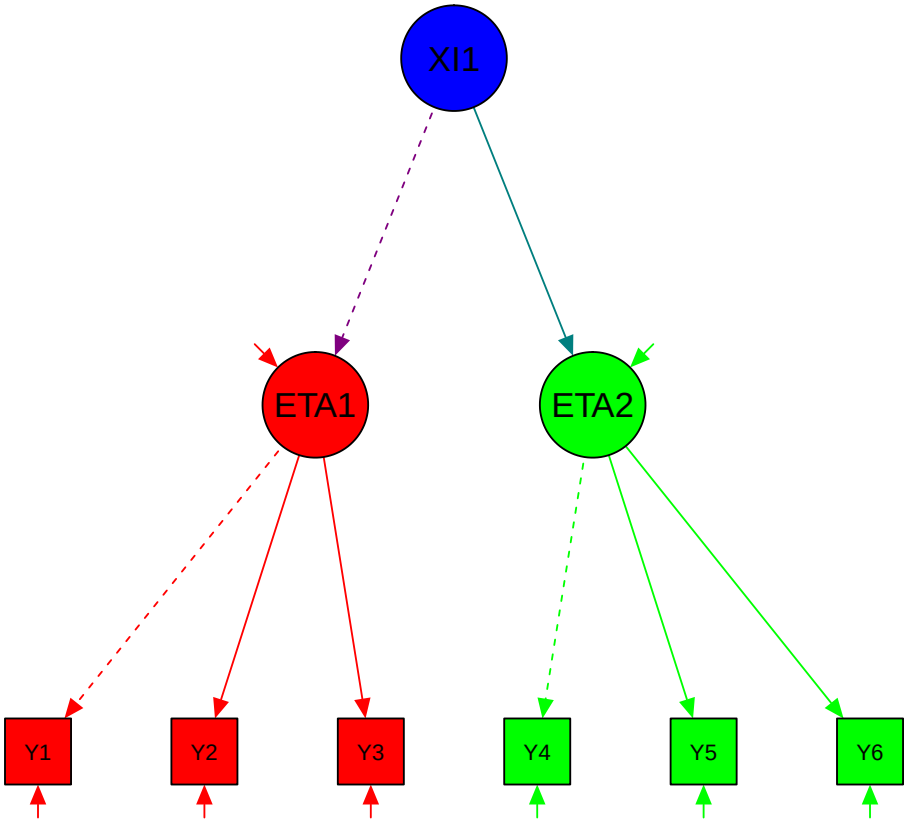

Figure 1: The given model. XI1 indicates ξ_1. ETA1 and ETA2 indicates η_1 and η_2. The colors specify different latent variables and its indicators.

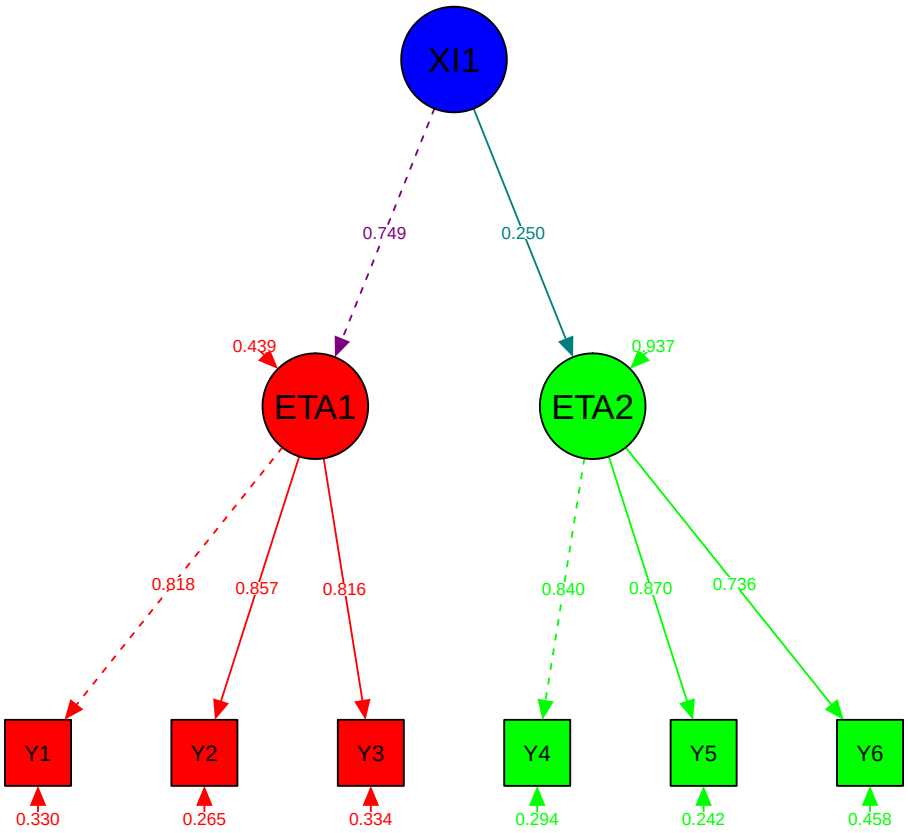

Figure 2: The point estimates of the parameter in the specified model.

Table 1. *Parameter Estimates*

Latent Variable					
	Estimate	SE	*Z*	*P*	Std
ETA1					
Y1	1				0.818
Y2	1.163	NA	NA	NA	0.857
Y3	0.875	NA	NA	NA	0.816
ETA2					
Y4	1				0.84
Y5	1.215	NA	NA	NA	0.87
Y6	0.793	NA	NA	NA	0.736
XI1					
ETA1	1				0.749
ETA2	0.336	NA	NA	NA	0.25
Variances					
	Estimate	SE	*Z*	*P*	Std
Y1	0.508	NA	NA	NA	0.33
Y2	0.503	NA	NA	NA	0.265
Y3	0.396	NA	NA	NA	0.334
Y4	0.435	NA	NA	NA	0.294
Y5	0.493	NA	NA	NA	0.242
Y6	0.554	NA	NA	NA	0.458
ETA1	0.452	NA	NA	NA	0.439
ETA2	0.978	NA	NA	NA	0.937
XI1	0.578	NA	NA	NA	1

Note. SE: standard error of estimates; Z: Z value; P: P value; Std: standrardized results.