Summary of Identification Rules

Note: All the rules listed here assume that errors are independent of exogenous variables, and all variables have expected value zero.

- *Counting Rule:* There should be at least as many identifying equations as parameter values. This applies to all models.
- *Recursion Rule*: The Observed Variable Model $\mathbf{Y} = \boldsymbol{\beta}\mathbf{Y} + \boldsymbol{\Gamma}\mathbf{X} + \boldsymbol{\zeta}$ is identified if the model is recursive and $V(\boldsymbol{\zeta}) = \boldsymbol{\Psi}$ has a particular block diagonal structure [1].

The two-variable and three-variable rules apply to the *Factor Analysis Model*: $\mathbf{X} = \boldsymbol{\Lambda}\mathbf{F} + \mathbf{e}$. These rules assume that all errors are uncorrelated, and each observed variable is caused by only one factor. If a model includes variables that are caused by more than one factor, it may be possible to add them to the model later using the Expansion Rule below.

- *Three-Indicator Rule for Standardized Variables*: For a factor analysis model with standardized observed variables (classical), the model will be identified if
  - The variance of each factor equals one.
  - There are at least 3 variables with non-zero loadings per factor.
  - The sign of one non-zero loading is known.
- *Three-Indicator Rule for Unstandardized Variables*: If a factor causes at least three observed variables and
  - The scale is fixed, meaning one factor loading equals one, and
  - At least two additional observed variables have loadings that do not equal zero,

  then the variance of the factor, the variances of the error terms and the factor loadings are all identified.
- *Two-Indicator Rule for Unstandardized Variables*: If a factor causes two observed variables and
  - The scale is fixed, meaning one factor loading equals one,
  - The other factor loading is non-zero,
  - The model contains at least one other factor having a non-zero correlation with this factor, and

[1] Organize the variables into sets. Set 0 consists of the exogenous variables. For $j = 1, \ldots, k$, each variable in set j is caused by at least one variable in set $j-1$, and also possibly by variables in earlier sets. Error terms for the variables in a set may have non-zero covariances. All other error terms have zero covariance.

- The scale of the other factor is fixed,

then the variance of the factor, the variances of the error term, and the one factor loading are all identified.

- *Double Measurement Rule*: The double measurement model is identified. Correlated measurement errors are allowed within sets of measurements, but not between sets.
- *Combination Rule*: Suppose that in a factor analysis model for unstandardized variables, some factors follow the Three-Variable Rule, others follow the Two-Variable rule, and still others follow the Double Measurement Rule. In this case, model parameters are identified in stages.
  - Apply the Double Measurement Rule to all relevant factors, identifying their variances and covariances, as well as the variances and possibly covariances of sets of error terms.
  - Apply the two and three-variable rules to each remaining factor, one at a time. This will identify the factor loadings and the variances of the factors.
  - If the scales are fixed for two factors, their covariance is identified provided the error terms of the two measurements are independent.

The last two rules apply to general Structural Equation Models (such as the *LISREL Model*) that have both latent and observed variables.

- *Two-Step Rule*

  1: Consider the latent variable model as a model for observed variables. Check identification (usually using the Counting Rule and the Recursive Rule).

  2: Consider the measurement model as a factor analysis model, ignoring the structure of $V(\mathbf{F})$. Check identification.

  If both identification checks are successful, the whole model is identified.

- *Expansion Rule* Suppose the measurement component of a structural equation model is identified. A vector of observed variables may be added to the latent component of the model without losing the identification of the measurement component, [2] provided
  - The additional variables are independent of the error terms in the original measurement model.
  - In the original measurement model, each latent variable has at least one observed variable that is a function of that latent variable, and of no other latent variable except for error terms. This will automatically be true if the rules above are used to establish identification of the measurement model.

[2] It might even help, but this needs to be proved on a case-by-case basis.