

## Improper Solutions in SEM

In practice, testing structural equation models may not always go smoothly, and the solution that is printed in the output may be an *improper solution*. The term “improper solution” refers to several possible problems with model estimation, ranging from model nonconvergence to invalid values for path estimates. Perhaps the most common problem that people have when they begin testing models is an error message that says something like “sigma matrix is not positive definite” or “warning: negative psi matrix.” These error messages are indications that something serious has gone wrong with your results, and you should not trust the estimates printed in the output. Another improper solution involves “out of bounds” estimates, sometimes referred to as “Heywood cases,” which are standardized values over 1.0, negative measurement error variances, or negative disturbances. Heywood cases can occasionally be found in the output even without an error message, so be sure to check your output carefully.

There are several different terms related to improper solutions that may be used. The wording of error messages tend to vary by software package or the specific source of the problem, and statistical authors are not wholly consistent in their terminology. Generally, the warning in the output is not much help in identifying the exact source of the problem, so you will need to check your model specification thoroughly for mistakes if you have a problem with improper solutions.

### Causes of improper solutions

*Nonconvergence* occurs when the maximum likelihood process cannot find a minimum fit function. Usually the model does not converge because of very serious mistakes in how the model is specified—the modeler told the computer to estimate a model that is not correct or does not make sense—or because there are very serious problems with the data set (e.g., data reading error, mistakes in variable computations). If the model is *underidentified* (i.e., negative degrees of freedom), a solution cannot be attained, and the software package will usually print a message indicating that there are negative degrees of freedom or that the model is not identified. Underidentification typically occurs because there is a model constraint omitted, such as failure to set a scaling constraint for a latent variable by constraining a reference loading or factor variance to 1. Another common problem is *empirical underidentification*, usually indicated by a negative error variance but can also lead to nonconvergence. With empirical underidentification, the model overall has positive degrees of freedom, but there is insufficient covariance information in a portion of the model for the computer to generate valid estimates. One example is a factor with only two indicators in a larger model with positive degrees of freedom. Another example of empirical underidentification is when reciprocal paths between two variables are specified.

Standardized values over 1.0 can sometimes be valid (Joreskog, 1999), but may also indicate there is a correlation very near 1.0, unreasonable model constraints have been imposed, or there are other problems with the model.

There has been some statistical work devoted to understanding when improper solutions are most likely (e.g., Chen et al., 2001; Gagne & Hancock, 2006; Gerbing & Anderson, 1987). Improper solutions tend to occur with smaller samples and when a model has fewer structural constraints (i.e., a more saturated model that freely estimates nearer to all possible paths). Negative variances may occur with smaller samples, for example, because of random sampling variability when the population value is near zero. Another possible cause of improper solutions is a more basic issue with the data set, such as outliers or violations of regression assumptions (e.g., heteroscedasticity). Low factor loadings (poor scale reliability) will also make nonconvergence or negative error variance more likely (Gagne & Hancock, 2006). As you might expect, a combination of more than one of these conditions, such as low loadings and small sample size, are particularly likely to be problematic.

## Solutions to Improper solutions

More experienced modelers tend to have fewer improper solutions, and I suspect that is because most often the problems arise because of fairly fundamental mistakes in specifying a model (e.g., two few latent variable indicators, not imposing scaling constraints for a latent variable, empirical underidentification). So, understanding model identification is an important first step.

Assuming that basic identification conditions have been met and the model is specified the way the researcher really intends to specify it, there are some possible remedies to consider. One solution that is often discussed involves setting out of bounds estimates, such as a negative error variance, to a specific value or some positive value very near zero. A few software packages do this automatically, reporting that a parameter “has been constrained at the lower bound.” Whether done automatically by the software package or more intentionally by the researcher, most modeling experts probably agree that this is not a very sound solution, because the underlying cause of the problem has not been rectified. Without a good theoretical justification, setting a measurement error to zero, for example, is probably not a good choice from an empirical or a theoretical standpoint (Chen et al., 2001).

In some circumstances, another remedy might be to set equality constraints. For example, if nonconvergence or Heywood cases are due to a latent variable with only two indicators, the two loadings can be set to be equal. The equal loading constraint involves an assumption that the association between the loading and the factor is the same for both items (as well as assumption about the equality of variances). A case can be made that assuming equal loadings is preferable to using the sum or the mean of the two items instead of a latent variable, because some measurement error is estimated. Another solution that has been proposed for out of bounds estimates is to set *inequality* constraints in which the value is estimated but constrained to not go above or below a certain value, but not all software packages allow the user to make these constraints.

In general, the best remedy is prevention. Careful specification, larger samples, factors with three or more indicators, reliable measures (high loadings), and well-conditioned data will all reduce the chance of improper solutions.

## References

- Boomsma, A. (1985). Nonconvergence, improper solutions, and starting values in LISREL maximum likelihood estimation. *Psychometrika*, 50, 229-242.
- Chen, F., Bollen, K. A., Paxton, P., Curran, P. & Kirby, J. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods & Research*, 29, 468-508.
- Gagne, P. E., & Hancock, G. R. (2006). Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivariate Behavioral Research*, 41, 65-83.
- Gerbing, D.W., & Anderson, J.C. (1987). Improper solutions in the analysis of covariance structures: Their interpretability and a comparison of alternate respecifications. *Psychometrika*, 52, 99-111.
- Joreskog, K.G. (1999). How large can a standardized coefficient be. Unpublished report. SSI Central, Inc. <http://www.ssicentral.com/lisrel/techdocs/HowLargeCanAStandardizedCoefficientbe.pdf>.
- Kolenikov, S. & Bollen, K.A. (2007). *Testing Negative Error Variances: Is a Heywood Case a Symptom of Misspecification?* Unpublished paper available at: <http://web.missouri.edu/~kolenikovs/papers/heywood-4.pdf>
- Shapiro, A. (1985). Asymptotic distribution of test statistic in the analysis of moment structures under inequality constraints. *Biometrika* 72(1), 133-144.
- Stoel, R. D., Garre, F. G., Dolan, C. & van den Wittenboer, G. (2006). On the likelihood ratio test in structural equation modeling when parameters are subject to boundary constraints. *Psychological Methods* 11(4), 439-455.