SEMinR: domain-specific language for building, estimating, and visualizing structural equation models in R

Soumya Ray, National Tsing Hua University Nicholas Danks, Trinity College-Dublin
André Calero Valdez, RWTH Aachen University

4/28/2021

Summary

SEMinR allows researchers to easily create, estimate, and visualize structural equation models (SEMs) for multiple estimation methods. SEMs are popular modeling techniques in social sciences and the life sciences, and can estimate relationships between concepts that need to be measured by multiple items. SEMinR can estimate SEMs using either covariance-based SEM (CBSEM, such as found in LISREL and Lavaan), or Partial Least Squares Path Modeling (PLS-PM, such as found in SmartPLS, semPLS, plspm, and csem). Moreover, SEMinR implements several advances in SEM methodologies not found elsewhere. And it also allows for visualization of all kinds of SEM models. SEMinR's model description syntax is plain-old-R-functions (PORF!), which allows users to extend and compose syntax in novel ways. Thus, SEMinR is a one-stop-shop for both SEM practitioners seeking to analyze empirical models and SEM methodologists seeking to automate and extend SEM methods. SEMinR is increasingly being used in universities, for both research and teaching needs, and companies world-wide.

Statement of Need

SEMinR seeks to bring the latest state-of-the-art advances in SEM methods to the R ecosystem. This package also seeks to make describing and analyzing SEMs easier for practitioners.

There have been several recent advances in the various branches of SEM that are often not reflected in existing R packages. For example, the PLS-PM approach requires adjustment in how models with interaction terms are estimated. PLS-PM methods have recently incorporated predictive methods such as plsPredict. Meanwhile, CB-SEM approach can avail ten Berge factor-score extraction that obtains construct scores with the same correlation patterns as the latent factors themselves. CB-SEM researchers should also consider VIF scores in their regression models. SEMinR incorporates these and other advancements.

Estimating an SEM using CB-SEM and PLS-PM requires different packages for the two estimation methods, which often requires researchers to wholly redescribe their models in different syntax. SEMinR allows researchers to describe their model once in a common syntax, and estimate the model using different estimation methods. SEMinr includes its own implementation of PLS-PM estimation that is tested against leading commercial applications to ensure comparable results. For CB-SEM estimation, SEMinR delegates the estimation to the popular Lavaan package. Regardless of which estimation method one uses, the results are structured in a similar way for reporting and visualization.

R packages for SEM often use a custom syntax that does not correspond to any programming language; nor does the syntax not reflect the terminology of SEM with which practitioners are familiar. SEMinR offers researchers a domain-specific language for modeling SEMs that uses function names that evoke major SEM

components: constructs, relationships, paths, reflective, composite, etc. As SEMinR's syntax is built using R functions, researchers can inject their own custom functions to extend the behavior of SEMinR.

SEMinR is the first package that allows researchers applying PLS-PM to visualize their graphical models and measurement qualities. Visualization of CB-SEM models is delegated to the semplot package. Moreover, SEMinR allows researchers to visualize models either before or after estimation.