Workshop

November 17, 2023

Power Simulation for the Actor-Partner Interdependence Model and Cross-Lagged Models
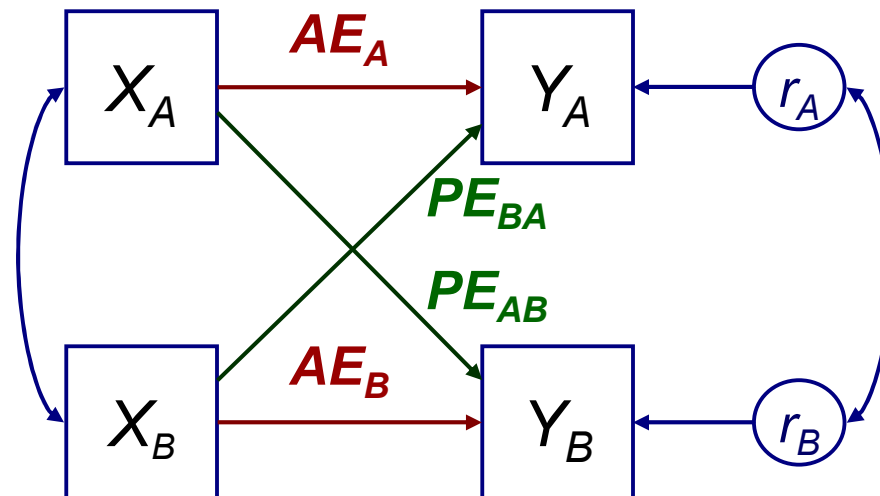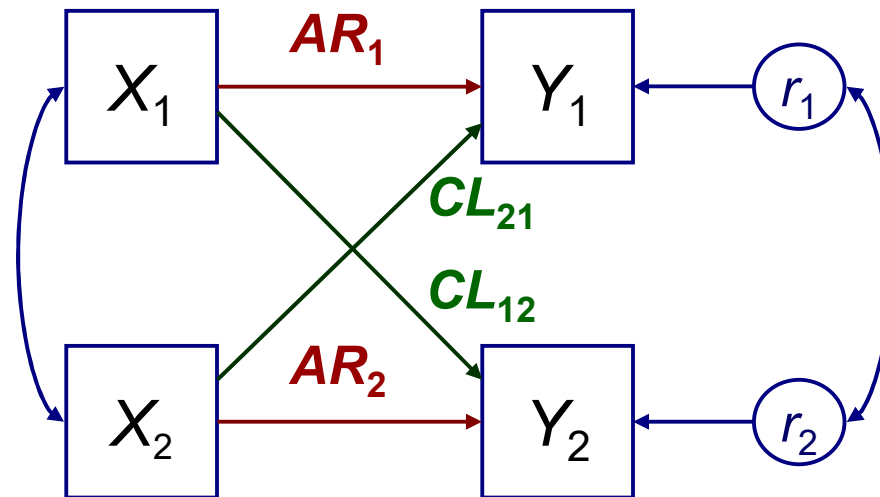
Tom Ledermann, Ph.D.

The Actor-Partner Interdependence Model (APIM)

The APIM (Kenny, 1996) is a model for data from dyads that has been designed to analyze actor (Intrapersonal) effects (AE) and partner (interpersonal) effects (PE).

The Cross-Lagged Model for 2 Waves

For 2 waves (time points), the cross-lagged model and the APIM are statistically identical.

In cross-lagged models,

- the actor effects are the autoregressive effects.

- the partner effects are the cross-lagged effects.

In designing a study, we commonly want to know what sample size is required to detect a specific effect or in the case the data collection is completed what is the power for a specific effect.

Main Article, R Code, GitHub Webpage

Ledermann, T., Rudaz, M., Wu, Q., & Cui, M. (2022). Determine power and sample size for the simple and mediation Actor–Partner Interdependence Model. *Family Relations, 71*(4), 1452-1469. http://dx.doi.org/10.1111/fare.12644
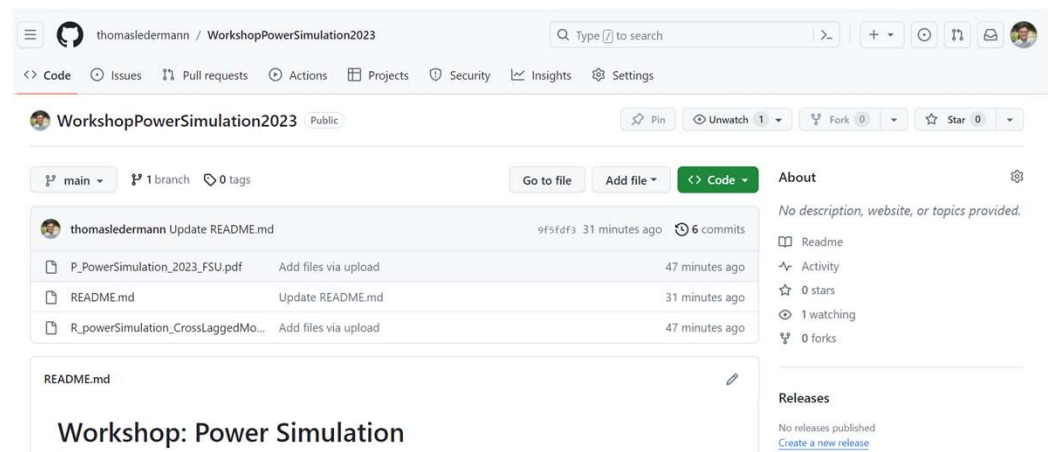
R Code on my GitHub account:
https://github.com/thomasledermann/powerAPIM

GitHub Webpage:
https://github.com/thomasledermann/WorkshopPowerSimulation2023

Overview

1. Power

2. Existing Methods to Compute Power

3. Monte Carlo Simulation

4. MC Simulation for the Cross-Lagged Model

5. Power Estimation for the Cross-Lagged Model

6. Evaluation of the Simulation

7. Sample Size Determination

8. Non-Normality

9. Conclusion, Limitations, and Take-Home Message

Power

Power ($1 - \beta$) is the probability of correctly rejecting the Null Hypothesis and a function of:

- Sample size ($\uparrow$)

- Effect size ($\uparrow$)

- Statistical method and test

- Significance level ($\uparrow$)

Power depends also on

- Amount of missing data

- Distribution of the data

- Reliability of the variables

Monte Carlo Simulation

Monte Carlo (MC) Simulation

-   is a method that allows for the determination of power for specific parameters (correlations, regression weights, variances, means).

-   uses many random samples based on a statistical model with predefined population parameters.

Monte Carlo Simulation

General steps

1. Generate many random samples (e.g., 10,000) for specified sample sizes using the statistical model with predefined population parameters and distribution function.

2. Estimate the model of interest for each generated dataset and obtain the p value for the effect of interest.

3. Determine for each generated dataset whether the effect is statistically significant based on the p value and the alpha level.

4. Calculate the power of a specific effect by dividing the number of samples for which the effect is statistically significant (reject the null hypothesis) by the total number of samples generated.

MC Simulation for the Cross-Lagged Model

The Solution for the APIM or Cross-Lagged Model

- uses MC simulation in R and Structural Equation Modeling (SEM) techniques.

- requires the

    - correlations among the continuous variables (*X*1, *X*2, *Y*1, and *Y*2).

    - sample size for which power is to be determined.

    - whether dyad members are distinguishable or indistinguishable (not relevant for cross-lagged models).

    - significance level.

MC Simulation for the Cross-Lagged Model

Of note,

- the variances of the observed variables are set to 1 by
  default.
  Using the actual variances (e.g., from previous research),
  the parameters of the effects based on the MC simulation
  resemble the actual parameters of the model.

- the power estimates for the parameters are neither
  affected by the means nor by the variances.

- the default alpha level is .05 and the default random
  samples is 10,000.

Power Estimation for the Cross-Lagged Model

Steps

1. Generate a variance-covariance matrix for the model variables from the correlations among these variables.

2. Estimate the parameters of the model based on the variance-covariance matrix and a sample size of 100,000 using the function `sem()` of the package `lavaan` (Rosseel, 2012).
    These parameter estimates serve as population parameter values for the generation of the random samples.

3. Run a MC simulation for the desired number of simulation using the function `sim()` of the package `simsem` (Pornprasertmanit et al., 2020).

Power Estimation for the Cross-Lagged Model

Key Results

The `summaryParam()` function returns for each parameter

- the average of parameter estimates.
- the *SD* of parameter estimates across all random samples.
- power.

Illustration

Population parameters

Correlations

Variable	X_1	X_2	Y_1
X_1	-		
X_2	.10	-	
Y_1	.30	.20	-
Y_2	.20	.30	.30

Standard Deviations = 1.0

Means = 0.0

Distribution of the data = normal

Number of Replications = 10,000

Illustration Population Parameter Estimates

Estimates (population values) and standard errors

Effect	Estimate	SE
Autoregression 1	0.283	0.003
Autoregression 2	0.283	0.003
Cross-lagged 2-1	0.172	0.003
Cross-lagged 1-2	0.172	0.003
$\mathrm{Cov}(X_1, X_2)$	0.100	0.003
$\mathrm{Cov}(E_1, E_2)$	0.192	0.003
$\mathrm{Var}(X_1)$	1.000	0.004
$\mathrm{Var}(X_2)$	1.000	0.004
$\mathrm{Var}(E_1)$	0.881	0.004
$\mathrm{Var}(E_2)$	0.881	0.004

Illustration Power Estimation

Summary of the results of the power analysis for $N = 100$

Effect	Average Est.	Average SE	Power
Autoregression 1	0.284	0.094	.847
Autoregression 2	0.285	0.094	.849
Cross-lagged 2-1	0.172	0.094	.449
Cross-lagged 1-2	0.171	0.094	.444
$Cov(X_1, X_2)$	0.099	0.099	.159
$Cov(E_1, E_2)$	0.185	0.087	.579

Evaluation of the Simulation

To assess the quality of the simulation, Muthén and Muthén (2002) suggest to look at

- the relative parameter estimate bias.

$$\frac{\text{Average Parameter Estimate} - \text{Parameter Value}}{\text{Parameter Value}}$$

- the relative standard error (SE) bias.

$$\frac{\text{Average } SE \text{ across all replications} - SD \text{ across all replications}}{SD \text{ across all replications}}$$

- Coverage (the proportion of replications for which the confidence interval contains the population parameter value).

Evaluation of the Simulation

Evaluation Criteria (Muthén & Muthén, 2002)

- **The relative parameter estimate and relative *SE* bias** should not exceed 10% for any parameter in the model.

- **The relative *SE* bias** for the parameters for which power is assessed should not exceed 5%.

- **Coverage** should be between 0.91 and 0.98.

The function `summaryParam()` provides these statistics.

Illustration Evaluation of the Simulation

Summary of the results of the power analysis for $N = 100$

Effect	Average Est.	Average SE	Power	Relative Bias Parameter	SE	Coverage
Autoregression 1	0.284	0.094	.847	.004	-.015	.943
Autoregression 2	0.285	0.094	.849	.007	-.023	.941
Cross-lagged 2-1	0.172	0.094	.449	.000	-.030	.940
Cross-lagged 1-2	0.171	0.094	.444	-.005	-.020	.944
Cov(X_1, X_2)	0.099	0.099	.159	-.009	-.003	.952
Cov(E_1, E_2)	0.185	0.087	.579	-.035	-.014	.946

Sample Size Determination

The package `simsem` allows the determination of the required sample size for each effect given power and the alpha level in three steps.

Steps

1. Specify a sequence of sample sizes (e.g., 100, 150, 200). This can be done by the function `seq()`.

2. Run a MC simulation for the number of desired replications (e.g., 1,000) using the function `sim()` of the package `simsem`.

3. The function `findPower()` provides the required sample size estimates for each effect and given power.

Illustration Sample Size Determination

Summary of the results of the power analysis for $N = 100$

Effect	Average Est.	Average SE	Power	N required
Autoregression 1	0.284	0.094	.847	92
Autoregression 2	0.285	0.094	.849	90
Cross-lagged 2-1	0.172	0.094	.449	249
Cross-lagged 1-2	0.171	0.094	.444	249
$Cov(X_1, X_2)$	0.099	0.099	.159	728
$Cov(E_1, E_2)$	0.185	0.087	.579	171

Non-Normality
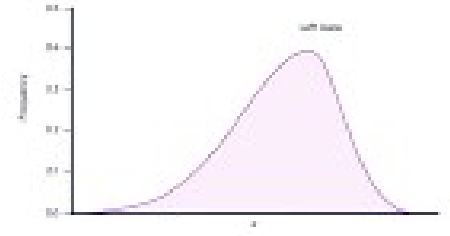
The variables can be skewed or kurtotic distributed.

Skewness and kurtosis can be introduced using the argument `indDist` in the `sim()` function.

Using the function `bindDist()`, skewness and kurtosis can be specified simultaneously for each variable separately.

Example

For four variables, the first one being normally distributed and the others being skewed or kurtotic:

```
bindDist(skewness = c(0, 1, -2, 2),
         kurtosis = c(0, 6, 3, 4))
```

Illustration Non-Normality

Skewness = 3, kurtosis = 21 for the two outcomes. $N = 100$.

Effect	Average Est.	Average SE	Power	N required
Normal data				
Autoregressive	0.284	0.066	.988	45
Cross-lagged	0.171	0.066	.736	121
$Cov(X_1, X_2)$	0.099	0.100	.153	731
$Cov(E_1, E_2)$	0.187	0.089	.571	172
Skewed data				
Autoregressive	0.150	0.049	.856	89
Cross-lagged	0.095	0.049	.498	219
$Cov(X_1, X_2)$	0.099	0.195	.082	3053
$Cov(E_1, E_2)$	0.236	0.095	.726	126
Kurtotic data				
Autoregressive	0.296	0.067	.991	42
Cross-lagged	0.178	0.067	.735	119
$Cov(X_1, X_2)$	0.100	0.100	.148	748
$Cov(E_1, E_2)$	0.187	0.089	.570	174

All pairwise effects were set equal (pooled).

Conclusion

- Power simulations provide a useful and flexible means to estimate power or to determine the sample size for parameters in a cross-lagged model or the simple and mediation APIM.

- The `simsem` package allows also the specification of different missing patterns, including missing completely at random (MCAR), missing at random (MAR), attrition, and planned missing data (see Schoemann, Miller, Pornprasertmanit, & Wu, 2014, for details).

Conclusion

The solution presented in here can be employed to other models that can be estimated using SEM:

- Random-Intercept Cross-Lagged Panel Models

- Growth Curve Models

- Latent Change Score Models

- Multiple Group Analysis

- Confirmatory Factor Analysis

Limitations

- The correlations among the variables must be known, something that is not required using G*Power or PASS 20.

- Some R skills are required if a researcher wants to estimate power for a model for which no code has been provided.

- The power analysis is currently limited to models that can be analyzed by SEM.

- Inclusion of dichotomous variables is challenging.

- Power simulation is based on multiple replications that make the power analysis slow.

Take-Home Message

- R provides a user-friendly and flexible solution for determining power and sample sizes for a wide range of models, including the cross-lagged models or the simple and mediation APIM.

- The illustrations demonstrate that skewed data tend to require larger sample sizes than normal data.

- The solution presented can also be used for pilot data and post hoc power analysis.

References

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Muthén, L. K., & Muthén, B. O. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling, 9*, 599-620.

nQuery (2017). *Sample size and power calculation*. Statsols (Statistical Solutions Ltd), Cork, Ireland.

PASS 2020 Power Analysis and Sample Size Software (2020). NCSS, LLC. Kaysville, Utah, USA, ncss.com/software/pass.

Pornprasertmanit, S., Miller, P., Schoemann, A., & Jorgensen, T. D. (2020). simsem: SIMulated Structural Equation Modeling [Computer software]. Retrieved from https://CRAN.R-project.org/package=simsem

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36.

Schoemann, A. M., Miller, P., Pornprasertmanit, S., & Wu, W. (2014). Using Monte Carlo simulations to determine power and sample size for planned missing designs. *International Journal of Behavioral Development, 38*, 471-479.

Thank you!

Email: tledermann@fsu.edu

GitHub: https://github.com/thomasledermann/powerAPIM