



semPower: General power analysis for structural equation models

Morten Moshagen¹ · Martina Bader¹

Accepted: 22 September 2023
© The Author(s) 2023

Abstract

Structural equation modeling (SEM) is a widespread and commonly used approach to test substantive hypotheses in the social and behavioral sciences. When performing hypothesis tests, it is vital to rely on a sufficiently large sample size to achieve an adequate degree of statistical power to detect the hypothesized effect. However, applications of SEM rarely consider statistical power in informing sample size considerations or determine the statistical power for the focal hypothesis tests performed. One reason is the difficulty in translating substantive hypotheses into specific effect size values required to perform power analyses, as well as the lack of user-friendly software to automate this process. The present paper presents the second version of the R package *semPower* which includes comprehensive functionality for various types of power analyses in SEM. Specifically, *semPower 2* allows one to perform both analytical and simulated a priori, post hoc, and compromise power analysis for structural equation models with or without latent variables, and also supports multigroup settings and provides user-friendly convenience functions for many common model types (e.g., standard confirmatory factor analysis [CFA] models, regression models, autoregressive moving average [ARMA] models, cross-lagged panel models) to simplify power analyses when a model-based definition of the effect in terms of model parameters is desired.

Keywords Confirmatory factor analysis · Model evaluation · Sample size planning · Statistical power · Structural equation modeling

Structural equation modeling (SEM) is a commonly employed statistical technique to test various types of substantive hypotheses. These hypotheses may refer to global aspects of a model, such as whether the assumed structure provides an adequate description of the population mechanics, but also to more local aspects, such as whether a particular model parameter differs from zero, across groups, or across measurement occasions. Regardless of the particular hypothesis under scrutiny, it is vital to rely on a sufficiently large sample size to achieve an adequate degree of statistical power to detect the hypothesized effect.

Statistical power is a concept arising in the context of classical (frequentist) null-hypothesis significance testing, where a null hypothesis (H_0) is evaluated against an alternative hypothesis (H_1). In any hypothesis test, two types

of decision errors may occur: the alpha-error of incorrectly rejecting a true null hypothesis (and thus incorrectly accepting a false alternative hypothesis) and the beta-error of incorrectly retaining a false null hypothesis (and thus incorrectly rejecting a true alternative hypothesis). Statistical power is the complement of the beta-error and gives the probability of rejecting a null hypothesis if this hypothesis is factually wrong (and thus to correctly accept a true alternative hypothesis).

As an example, consider a simple confirmatory factor analysis (CFA) model involving two factors that are measured by three indicators each, and assume that the interest lies in the correlation between the factors. To test the null hypothesis that the factors are uncorrelated versus the alternative hypothesis that the correlation differs from zero, one can compare a model that freely estimates the factor correlation (reflecting the H_1) with a model that restricts the correlation between these factors to zero but is otherwise identical (reflecting the H_0). If the restricted model fits the data significantly worse than the unrestricted model, the null hypothesis of a zero correlation between the factors is rejected, in turn lending support

✉ Morten Moshagen
morten.moshagen@uni-ulm.de

¹ Psychological Research Methods, Institute of Psychology and Education, Ulm University, Albert-Einstein-Allee 4, 89081 Ulm, Germany

for the alternative hypothesis that the correlation between the factors differs from zero. Statistical power now gives the probability that the outcome of the statistical test contrasting the H0 and the H1 turns out significant on a certain alpha-error level with a certain sample size and given a certain discrepancy between the H0 and the H1 in the population. For instance, suppose that the true correlation between the factors in the population is $r = .20$ and that all nonzero primary loadings equal $.50$. Then, the statistical power to reject the factually incorrect null hypothesis (of $r = 0$) given $\alpha = .05$ and a sample size of $N = 125$ is just 20%. Stated differently, in four out of five samples (each of size $N = 125$) drawn randomly from the population, one will not detect that the factors are in fact correlated. Indeed, to obtain a more reasonable power of 80% in this scenario, a sample size of $N = 783$ is required.

As is evident from this example, hypotheses tests are only meaningful to the extent that statistical power is sufficiently high, because otherwise a nonsignificant test outcome provides little information regarding the veracity of the tested hypothesis. Conversely, even miniscule discrepancies between the H0 and the H1 that are of no practical relevance may turn out significant when statistical power is high. Statistical power is thus an integral part of sample size planning and statistical hypothesis testing more generally.

Unfortunately, statistical power analyses are often not reported at all in applications of SEM (Jackson et al., 2009), or are only performed with respect to a single hypothesis. However, power is always tied to a particular hypothesis and might vary strongly depending on which hypothesis is considered. For instance, in a simple two-wave cross-lagged panel model (CLPM; e.g., Newsom, 2015) involving four factors, each comprising three indicators loading by $.50$, $N = 252$ observations are required to yield 80% power on $\alpha = .05$ to detect global misspecifications corresponding to a root-mean-square error of approximation (RMSEA) $\geq .05$. In contrast, $N = 4629$ observations are required to detect a cross-lagged effect $\geq .10$ in the same model with an equal level of power. Correspondingly, power analyses should be performed for the focal hypothesis of interest, which is rarely done in practice. One reason for the notable absence of power considerations is arguably the lack of powerful software tools easily allowing for the assessment of power for both global and local hypotheses. The present paper therefore introduces the *semPower 2* package for the R environment, which supports various types of and approaches to power analyses covering both model-free and model-based definitions of the hypothesis of interest. In the remainder of this paper, we first provide the technical background underlying power analysis, and then illustrate how to use *semPower 2* based on illustrative examples.

Statistical background

Statistical power depends on (a) the degree to which the null hypothesis is wrong (the magnitude of effect), (b) the sample size (N), (c) the specified alpha-error, and (d) the degrees of freedom (df). Everything else being equal, power will be higher for a larger effect, a larger sample size, a higher alpha-error, and fewer df . In a power analysis, one of these quantities is computed as a function of the other quantities, giving rise to different types of power analyses (Faul et al., 2007). *A priori power analyses* determine the required number of observations to detect a certain effect with a desired power, given α and df , and are thus performed prior to data collection to inform sample size planning. *Post hoc power analyses* determine the achieved power to detect a certain effect with a given sample size, α , and df , and are thus performed after data collection to judge whether the given sample size yields a power that is sufficiently high for a meaningful test of a certain hypothesis. *Compromise power analyses* determine the alpha- and the beta-error, given the ratio between alpha and beta, the sample size, the effect, and the df , and are used to determine which decision rule to apply when the error probabilities should meet a desired ratio, such as being equal (see Moshagen & Erdfelder, 2016, for details).

To understand how a power analysis can be performed, a brief discussion of model testing is required. Generally, when testing a model representing the H0 against a more general alternative model representing the H1, a likelihood-ratio test (LRT) can be performed:

$$LRT = -2 \ln \left(\frac{L_{M0}}{L_{M1}} \right) \quad (1)$$

$$= (N - 1)(\hat{F}_{M0} - \hat{F}_{M1}) \quad (2)$$

$$= \chi_{M0}^2 - \chi_{M1}^2 \quad (3)$$

L_{M0} is the likelihood of the hypothesized (H0) model and L_{M1} is the likelihood of a more general (H1) model, where the H0 model must be statistically nested in the H1 model (i.e., the parameter space of the H0 model must be a subset of the parameter space of the H1 model; Bentler & Satorra, 2010). In SEM, maximizing the likelihood is equivalent to minimizing the maximum likelihood (ML) fitting function (Jöreskog, 1967), so the LRT can also be obtained by considering the sample-size-weighted difference in the minimized sample values of the fitting function for the H0 model (\hat{F}_{M0}) and the H1 model (\hat{F}_{M1}).

If the H0 model is correct, assuming multivariate normality, a sufficiently large sample size, and several mild regularity conditions (Browne, 1984; Yuan & Bentler, 2007),

the LRT asymptotically follows a central $\chi^2(df)$ distribution with $df = df_{M1} - df_{M0}$, so it can be computed in practice as the difference between the χ^2 -model test statistics associated with the H0 (χ^2_{M0}) and the H1 model (χ^2_{M1}). If the H0 model is invalid, however, the LRT—again assuming multivariate normality, a sufficiently large sample size, and several regularity conditions¹ (Satorra & Saris, 1985; Steiger et al., 1985)—asymptotically follows a noncentral $\chi^2(df, \lambda)$ distribution with non-centrality parameter λ and expected value $df + \lambda$, so that λ shifts the noncentral $\chi^2(df, \lambda)$ distribution to the right of the central χ^2 distribution. The non-centrality parameter is the sample-size-weighted difference in the population minima, F_0 , of the fitting function, $\lambda = (N - 1)(F_{0,M0} - F_{0,M1})$. Note that the minimized sample value of the fitting function, \hat{F} , is a biased estimate of F_0 (McDonald, 1989) with an expected value of $F_0 + df / (N - 1)$. The difference in the population minima between the H0 and the H1 expresses the extent to which the H0 is incorrect (with respect to the H1) and thereby defines the magnitude of effect.

As is evident from Eq. 1, an LRT always involves a comparison of the hypothesized H0 model against an alternative H1 model. The latter can either be explicitly specified or refer to the saturated model. Generally, a structural equation model in which the number of free parameters is identical to the number of observed means, variances, and covariances is saturated, so the saturated model typically just estimates each observed mean, variance, and covariance without any constraint. Consequently, a saturated H1 model can always describe the observed data perfectly and shows $F_{0,M1} = 0$ on zero df .

Returning to the introductory example concerning the correlation between two factors in a CFA model, the H0 model restricting the correlation to zero could thus be compared either against a less restricted model that freely estimates the correlation, but is otherwise identical to the H0 model (leading to $df = 1$), or against the saturated model freely estimating all variances and covariances (leading to $df = 9$). Although either approach is valid, the comparison model should generally be chosen such that it allows for an immediate test of the hypothesis articulated in the H0 model (Jöreskog, 1978). Thus, when a (local) hypothesis explicitly refers to a zero correlation between two factors, the suitable comparison model should only differ in this particular parameter. However, when a (global) hypothesis states that the model as a whole adequately describes the data, the saturated model is the proper comparison model.

Figure 1 exemplarily illustrates how the alpha- and beta-error arise in a hypothesis testing situation involving 50 df

and three different values for the non-centrality parameter. The null hypothesis is rejected when the p -value associated with the empirically observed LRT falls below the alpha-error. This is equivalent to evaluating whether the value of the LRT exceeds the critical χ^2_c -value corresponding to the chosen alpha-error (e.g., $\chi^2_c(df = 50) = 67.50$ for alpha = .05). Note that neither the alpha-error nor the implied critical χ^2_c -value change as a function of the non-centrality parameter. The area of the central $\chi^2(df)$ distribution to the right of the critical value corresponds to the alpha-error, whereas the beta-error is the area of the noncentral $\chi^2(df, \lambda)$ distribution to the left of the critical value, i.e.,

$$\beta = \int_0^{\chi^2_c} f_{\chi^2(df, \lambda)}(x) dx \quad (4)$$

Correspondingly, the beta-error (and thus statistical power, $1 - \beta$) can be computed given the non-centrality parameter, which depends on N and the magnitude of effect, and the critical value, which depends on the df and the alpha-error. Choosing a smaller alpha-error increases the beta-error by increasing the implied critical value χ^2_c . The beta-error (but not the alpha-error) is also a function of the overlap between the central $\chi^2(df)$ and the noncentral distributions $\chi^2(df, \lambda)$, which decreases with increasing non-centrality parameter λ , and thus with increasing sample size and with an increasing effect.

Analytical versus simulated power analysis

Power analyses can be performed either analytically drawing on asymptotic theory or based on a simulation approach. In SEM, analytical power analyses as described above require the definition of the effect of interest in terms of a measure that shows a direct relation to the non-centrality parameter, which is afforded by all non-centrality-based measures of fit such as the minimum of the employed fitting function itself or derived indices such as the RMSEA (because $F_0 = df \cdot RMSEA^2$; e.g., MacCallum et al., 1996). Analytical power analyses are thus inherently model-free in the sense that the achieved power to reject a H0 model exhibiting a specific degree of misspecification on a certain df and a defined alpha-error with a certain N is always the same, regardless of whether the model is a CFA model, a CLPM, a multigroup model, or any other SEM model. Beyond allowing for the direct comparison across different model types, reliance on a non-centrality-based effect size for power analyses is particularly useful when only the overall degree of misfit of a particular model is of interest, or when the effect of interest spreads across many parameters (such as in tests of measurement invariance or when testing a bifactor versus a higher-order structure). However,

¹ The regularity conditions include the so-called parameter-drift assumption, which states that the degree of misspecification may not be too large relative to the degree of sampling error.

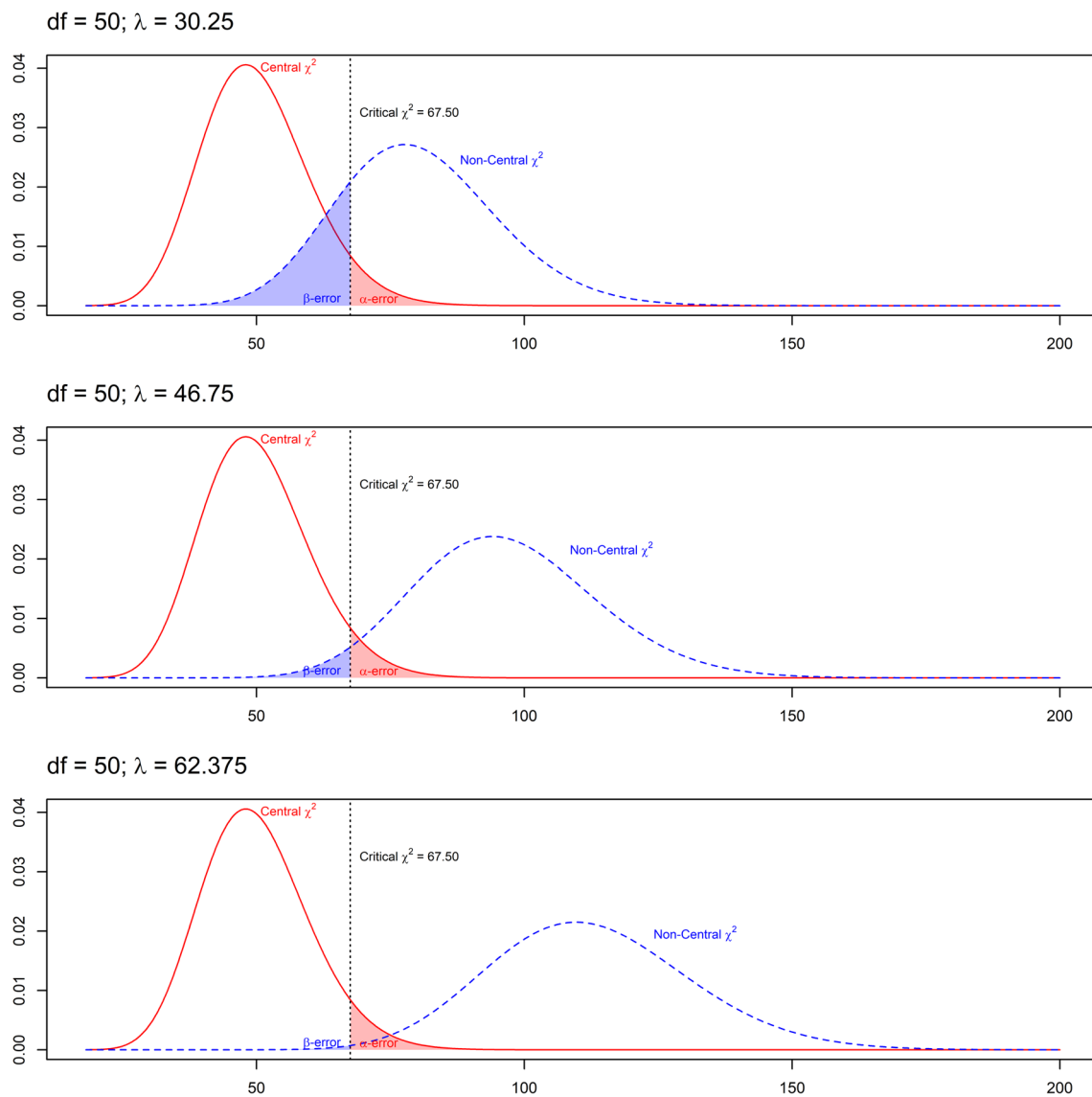


Fig. 1 Central and noncentral χ^2 distributions, critical χ^2 value, and associated decision errors. *Note.* Central $\chi^2(df)$ distribution (solid line) and noncentral $\chi^2(df, \lambda)$ distributions (dashed lines) with $df = 50$ and (from top to bottom) non-centrality parameters of $\lambda = 30.25$,

$\lambda = 46.75$, and $\lambda = 62.375$. The dotted vertical line indicates a critical value for χ^2_c of 67.50, which corresponds to $\alpha = .05$. The beta-errors are $\beta = .20$, $\beta = .03$, and $\beta = .003$, respectively

a common problem is that a focal hypothesis often refers to a specific model parameter (such as a hypothesized cross-lagged effect in a CLPM of a certain magnitude), but it is not straightforward to relate such hypotheses to a particular value of a non-centrality-based effect size.

Unlike analytical power analyses, simulated power analyses perform a simulation study to obtain an empirical estimate of power. Specifically, a model is defined that describes the true state of affairs in the population (in terms of the number of observed variables, the number of latent factors, the magnitude of loadings, (residual-)variances, covariances between latent variables, etc.), which in turn defines the population variance–covariance matrix and population means (if

these are part of the model). Based on these population characteristics, random data sets are repeatedly generated and analyzed (e.g., 1000 times). In addition, an analysis model is defined, which implements at least one restriction that is factually wrong (representing the H_0). When the H_1 does not refer to the saturated model, a more general, explicit H_1 comparison model can also be defined. Then, an empirical estimate of power is obtained by the proportion of significant outcomes of the LRT on a given alpha-error level.

As a consequence, simulated power analyses are directly tied to a particular model representing the H_0 and define the effect of interest in terms of the difference in a particular (set of) parameter(s) that are embedded in a certain model

structure with particular population values for all parameters (Satorra & Saris, 1983; Wang & Rhemtulla, 2021). Of note, the actual effect in terms of F_0 not only is a function of the parameters that differ between the H0 and the H1 model, but also depends on all other parameters in the model (e.g., Moshagen & Auerwald, 2018; Wang & Rhemtulla, 2021). Therefore, the level of statistical power is generally specific to the particular model under scrutiny and might change not only with the population difference of a particular parameter, but also as a function of the value of other parameters.

However, although the model-free versus model-based definition of the effect of interest is indeed a key difference between analytical and simulated power analysis, analytical power analyses can also be performed in a model-based manner. To this end, analytical power analysis requires that an effect defined in terms of model parameters is transformed into an effect based on non-centrality (Saris & Satorra, 1993). This can be achieved by evaluating the discrepancy between the population variance–covariance matrix (Σ) and the model-implied variance–covariance matrix ($\hat{\Sigma}$) (and mean vectors μ and $\hat{\mu}$, if means are part of model) as measured via a suitable fitting function, such as weighted least squares (Browne, 1984) or ML. By default, *semPower 2* relies on the ML fitting function assuming continuously distributed data:

$$F_0(\Sigma, \hat{\Sigma}) = \ln|\hat{\Sigma}| - \ln|\Sigma| + \text{tr}(\Sigma\hat{\Sigma}^{-1}) - p + (\mu - \hat{\mu})\hat{\Sigma}^{-1}(\mu - \hat{\mu}) \quad (5)$$

In more practical terms, the model representing the H0 is fitted to a population covariance matrix Σ and the population means μ . The resulting parameter estimates are used to construct the model-implied covariance matrix $\hat{\Sigma}$ and means $\hat{\mu}$ associated with the H0. Finally, Eq. 5 is used to obtain the population minimum of the ML fitting function F_0 . Effectively, a model-based power analysis can be transformed into a model-free power analysis using F_0 as an effect size metric. In this way, analytical power can always be performed for both model-based and model-free definitions of the effect. In contrast, simulated power analyses are necessarily model-based.

Beyond their greater versatility in terms of the definition of the effect of interest, analytical power analyses are generally preferable over simulated power analyses to the extent that the assumptions underlying the asymptotically expected $\chi^2(df)$ distributions under the null and the alternative hypothesis are met. This is because computation is much faster (instantly); in particular, when an a priori power analysis is performed, there is no random error resulting from the sampling process, and there are no additional intricacies associated with model estimation, such as non-convergence. However, when the assumptions underlying analytical power analysis are violated, simulated power analysis may offer a more realistic assessment, provided that these violations are embedded in the simulation and that the defined population structure and data-generating

process can be considered representative of empirical reality. Simulated power analysis further offers the advantage that it may provide information beyond statistical power, such as rates of proper convergence and parameter recovery, which helps to inform sample size requirements more generally. In sum, both analytical and simulated power analyses have their advantages and are best used complementarily.

Software for power analysis in SEM

Despite the importance of statistical power in sample size planning and hypothesis testing, available software programs and packages directed towards this goal are subject to various shortcomings. To our knowledge, existing tools rely on either an analytical or a simulated approach to power analysis, but do not provide both options. General power analysis software programs such as *G*Power* (Faul et al., 2007) do not offer effect sizes relevant for SEM. Existing tools providing analytical power analyses specifically for SEM (e.g., MacCallum et al., 1996, 2006; Preacher & Coffman, 2006) require provision of a non-centrality-based measure of effect such as the RMSEA and are thus limited to a model-free approach. Although simulation approaches to statistical power necessarily allow for a model-based definition of the effect of interest, most existing tools implement post hoc power analyses only and are typically limited to particular model types (Kievit et al., 2018; Mulder, 2022; Schoemann et al., 2017; Zhang & Liu, 2018). A number of options providing either analytical (Jak et al., 2021; Miles, 2003; Z. Zhang & Yuan, 2018) or simulated power analysis (Muthén & Muthén, 2002; Wang & Rhemtulla, 2021) allow for a generic model-based definition of the effect, but these require the manual definition of the values of all parameters of the population model. Such a manual definition is cumbersome and potentially error-prone, in particular when more complex models are considered and it is desired to define the model in terms of standardized parameters, which requires rather sophisticated knowledge concerning the proper definition of the residual variances (see e.g., Bader & Moshagen, 2022).

As a remedy, the present paper introduces the second version of the *semPower* package—*semPower 2*—for the R environment, which can be obtained from CRAN (<https://cran.r-project.org/package=semPower>) or from GitHub (<https://github.com/moshagen/semPower>), which also hosts its source code. The *semPower 2* package is a massive improvement and extension over the first version of *semPower* (Moshagen & Erdfelder, 2016), which only provided analytical power analysis for common effect size metrics or based on a covariance matrix input (thereby offering model-based power analysis via rather cumbersome procedures; see Jobst et al., 2023). Specifically, *semPower 2* provides a priori, post hoc, and compromise power analysis, both

analytical and simulated, for structural equation models with or without latent variables, supports multigroup settings, and offers highly abstracted input for many common model types to simplify the model specification when a model-based definition of the effect in terms of model parameters is desired.

The remainder of this paper presents an overview of semPower 2. We illustrate various types of analytical power analyses based on both model-free and model-based definitions of the effect, and also briefly describe how simulated power analyses can be performed.

semPower 2 overview

In the following, an introductory overview of the core features of semPower 2 is presented. First, we provide various examples of how to perform analytical model-free power analyses comparing the H0 model either against the saturated model or against a less restrictive (but not saturated) competing model. We then illustrate analytical model-based power analyses based on the convenience functions of semPower

2 using several examples. Finally, we move to simulation approaches to model-based power analysis. Readers are referred to the extensive online documentation available at <https://moshagen.github.io/semPower/> for a complete and more detailed description of the capabilities of semPower 2.

Analytical model-free power analysis

We first consider analytical model-free power analyses comparing the H0 model against the saturated H1 comparison model. Any model-free power analysis requires provision of the measure (effect.measure) and magnitude (effect) of effect that is to be detected as well as the *df*. In an a priori power analysis, the desired power also needs to be specified, so that a model-free a priori power analysis can be performed by calling the semPower function setting type = 'a-priori' and corresponding further arguments. For instance, the following code requests the required sample size to detect misspecifications of a model involving 50 *df* corresponding to RMSEA $\geq .05$ on alpha = .05 with a power of 80%:

```
ap <- semPower(type = 'a-priori', effect = .05, effect.measure = 'RMSEA',
              alpha = .05, power = .80, df = 50)
summary(ap)

##  semPower: A priori power analysis
##
##  F0                0.125000
##  RMSEA              0.050000
##  Mc                 0.939413
##
##  df                 50
##  Required Num Observations 243
##
##  Critical Chi-Square  67.50480
##  NCP                 30.25000
##  Alpha               0.050000
##  Beta                0.199142
##  Power (1 - Beta)    0.800858
##  Implied Alpha/Beta Ratio 0.251077
```

The output first provides several non-centrality measures of effect, here showing that an RMSEA of .05 (on 50 *df*) is equivalent to $F_0 = .125$ and to McDonald's (1989) index of non-centrality² (Mc) of .939. The output further

² McDonald's (1989) index of non-centrality is computed from the minimum of the fit function by $Mc = \exp(-.5 \cdot F_0)$, so that the Mc ranges from 0 to 1, with higher values indicating better fit.

indicates a required sample size of $N = 243$ to detect this effect with a power of 80% on alpha = .05.

A post hoc power analysis computes the achieved power with a given sample size, so the semPower function with type = 'post-hoc' expects the N argument. For instance, the following code computes the achieved power with a sample size of $N = 500$ in the same situation as above:

```
ph <- semPower(type = 'post-hoc', effect = .05, effect.measure = 'RMSEA',
              alpha = .05, N = 500, df = 50)
summary(ph)
```

This yields an output structured identically to the one of an a priori power analysis and shows that with $N = 500$, the achieved power to detect misspecifications of a model involving 50 df amounting to $\text{RMSEA} \geq .05$ on $\alpha = .05$ is 99.7%. This implies that committing an alpha-error is about 15 times as likely as committing a beta-error (the ratio of alpha to beta is $.05 / .00335 = 14.9$).

The result of any (model-free or model-based) a priori or post hoc power analysis can also be plugged into the `semPower.powerPlot` function to produce a figure showing how power for the specified effect varies as a function of sample size. Thus, a plot akin to Fig. 2 is produced by calling

```
ph <- semPower(type = 'post-hoc', effect = .05, effect.measure = 'RMSEA',
              alpha = .05, N = 500, df = 50)
semPower.powerPlot(ph)
```

If the goal is to determine a decision rule (i.e., the critical value χ_c^2 of the χ^2 test statistic) that balances the error probabilities, a compromise power analysis can be used. The `semPower` function with `type = 'compromise'` expects the

`abratio` argument defining the desired ratio between alpha and beta. For instance, the following code determines χ_c^2 and the associated error probabilities, such that the alpha- and beta-error are equal:

```
cp <- semPower(type = 'compromise', effect = .05, effect.measure = 'RMSEA',
              alpha = .05, N = 500, df = 50, abratio = 1)
summary(cp)
```

The output is again structured identically to the one shown above and indicates that choosing a critical value of $\chi_c^2 = 74.64$ is associated with equal error probabilities, $\alpha = \beta = .013$.

Power for effect-size differences

The examples above were concerned with the situation that an H0 model is compared against a saturated H1 model. A common scenario is to test two competing models against each other, where the more restrictive H0 model (involving

more df) is compared against a less restrictive H1 model (involving fewer df), so that power is to be determined concerning the difference between these models in terms of overall fit. This can be achieved by providing a vector of effect sizes to the effect argument. For instance, the following determines the required sample size to test a more general H1 model exhibiting an RMSEA of .04 on 44 df against a nested H0 model exhibiting an RMSEA of .05 on 41 df which shows that $N = 247$ observations are necessary to detect a difference of $\text{RMSEA} = .01$ between these models with a power of 80%.

```
ap <- semPower(type = 'a-priori', effect = c(.04, .05), effect.measure =
              'RMSEA', alpha = .05, power = .80, df = c(41, 44))
summary(ap)
```

The same approach can also be used in multiple-group contexts, when sufficient power is desired to detect whether certain cross-group constraints on the model parameters (such as equal loadings across groups) are tenable. The general syntax is the same as above, but now the N argument also needs to be set, which gives the number of observations per group in compromise and post hoc power analyses and the group weights in a priori power analyses. For example, it has been

```
ap <- semPower(type = 'a-priori', effect = c(.99, .97), effect.measure = 'Mc',
              alpha = .05, power = .80, df = c(57, 69), N = c(1, 1, 1))
summary(ap)
```

The results indicate that $N = 426$ ($n = 142$ by group) observations are required to yield a power of 80%.

Analytical model-based power analysis

semPower 2 provides a number of convenience functions (see Table 1 for an overview) to perform (both analytical and simulated) a priori, post hoc, and compromise power analyses using a model-based definition of the effect of interest in terms of particular values for the model parameters. These convenience functions offer a high level of abstraction to simplify the definition of the model and the effect of interest. In addition, semPower 2 includes a more generic function (semPower.powerLav), which allows one to perform model-based power analyses for models and hypotheses not covered by any more specific convenience function dedicated to a particular model structure and thereby provides a high degree of flexibility.

To perform analytical model-based power analyses, a population model and a corresponding analysis model are transformed into the population and model-implied covariance matrices (and mean vectors if these are part of the model), assuming continuously distributed data. In the context of CFA models, the population variance–covariance matrix (Σ) is given by

$$\Sigma = \Lambda\Phi\Lambda' + \Theta, \quad (6)$$

where Λ is the loading matrix, Φ is the variance–covariance matrix of the factors, and Θ is the variance–covariance matrix of the indicators. The population means (μ) are

$$\mu = \tau + \Lambda\alpha \quad (7)$$

where τ is a vector of indicator intercepts and α denotes a vector containing the latent factor means. In more general SEM models, Σ can be computed by

suggested that a decline of $\geq .02$ in the Mc associated with a metric in comparison to a configural invariance model may be considered a meaningful departure from measurement invariance (Cheung & Rensvold, 2002). Accordingly, the following asks for the required sample size to detect that a model exhibiting an Mc of .99 on 57 df differs from a model exhibiting an Mc of .97 on 69 df in a three-group model, where all groups are of equal size ($N = c(1, 1, 1)$):

$$\Sigma = \Lambda(I - B)^{-1}\Psi[(I - B)^{-1}]'\Lambda' + \Theta, \quad (8)$$

where B contains the regression coefficients between the factors and Ψ is the (residual) variance–covariance matrix between the latent factors. The means are then given by

$$\mu = \tau + \Lambda(I - B)^{-1}\alpha. \quad (9)$$

In any power analysis, the relevant model matrices are defined first, so that Σ and μ in the population can be obtained. In analytical model-based power analyses, the model implementing the null hypothesis of interest is then fitted to the population variance–covariance matrix using the lavaan package (Rosseel, 2012), so that the population minimum F_0 along with the resulting model-implied variance covariance matrix ($\hat{\Sigma}$) and mean vector ($\hat{\mu}$) are obtained based on the estimated parameters. In simulated power

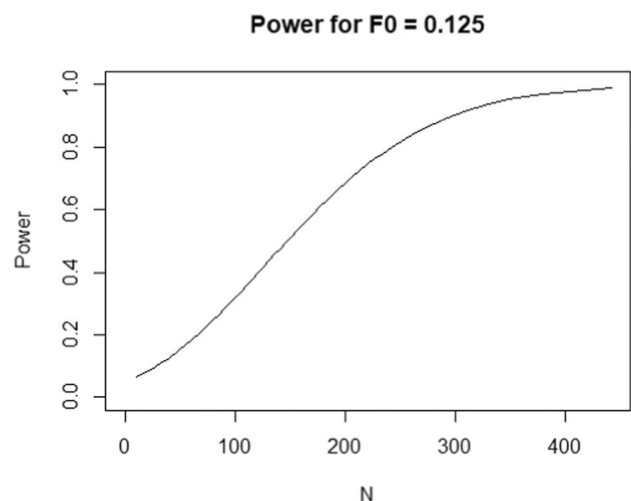


Fig. 2 Statistical power as a function of sample size to detect $RMSEA = .05$ on 50 df ($F_0 = 0.125$)

Table 1 semPower 2 functions providing model-based power analyses

Function	Model	Hypotheses
semPower.power-CFA	CFA	$\Lambda_{ij} = 0$ $\phi_{ij} = 0$ $\phi_{ij} = \phi_{k,l}$ $\phi_{ij,m} = \phi_{i,j,n}$
semPower.powerBifactor	Bifactor	$\phi_{ij} = 0$ $\phi_{ij} = \phi_{k,l}$ $\phi_{ij,m} = \phi_{k,l,n}$
semPower.power-Regression	Single linear regression	$B_{ij} = 0$ $B_{ij} = B_{i,k}$ $B_{ij,m} = B_{i,j,n}$
semPower.powerMediation	Mediation	$B_{ij} \cdot B_{k,l} = 0$ $B_{i,j,m} \cdot B_{k,l,m} = B_{i,j,n} \cdot B_{k,l,n}$
semPower.powerPath	General regression	$B_{ij} = 0$ $B_{ij} = B_{k,l}$ $B_{ij,m} = B_{i,j,n}$
semPower.powerAutoreg	Autoregression	$B_{ij} = 0$ $B_{ij} = B_{k,l}$ $B_{i,j,m} = B_{i,j,n}$ $\psi_{ij} = 0$ $\psi_{ij} = \psi_{k,l}$ $\psi_{i,j,m} = \psi_{i,j,n}$ $\psi_{i,j,k} = \psi_{k,l,m}$ $\alpha_i = 0$ $\alpha_i = \alpha_j$ $\alpha_{im} = \alpha_{in}$
semPower.power-ARMA	ARMA	$B_{ij} = 0$ $B_{ij} = B_{k,l}$ $B_{i,j,m} = B_{i,j,n}$ $\psi_{ij} = 0$ $\psi_{ij} = \psi_{k,l}$ $\psi_{i,j,n} = \psi_{i,j,m}$ $\alpha_i = 0$ $\alpha_i = \alpha_j$ $\alpha_{im} = \alpha_{in}$
semPower.power-CLPM	CLPM	$B_{ij} = 0$ $B_{ij} = B_{k,l}$ $B_{i,j,m} = B_{i,j,n}$ $\psi_{ij} = 0$ $\psi_{ij} = \psi_{k,l}$ $\psi_{i,j,k} = \psi_{k,l,m}$
semPower.power-RI-CLPM	RI-CLPM	$B_{ij} = 0$ $B_{ij} = B_{k,l}$ $B_{i,j,m} = B_{i,j,n}$ $\psi_{ij} = 0$ $\psi_{ij} = \psi_{k,l}$ $\psi_{i,j,k} = \psi_{i,j,m}$
semPower.powerMI	Multigroup invariance	$\Lambda_m = \Lambda_n$ $\tau_m = \tau_n$ $\Theta_m = \Theta_n$ $\phi_m = \phi_n$ $\alpha_m = \alpha_n$
semPower.powerLI	Longitudinal invariance	$\Lambda_m = \Lambda_n$ $\tau_m = \tau_n$ $\Theta_m = \Theta_n$ $\phi_m = \phi_n$ $\alpha_m = \alpha_n$
semPower.powerLav	Generic	Any

Note. CFA = confirmatory factor analysis. ARMA = autoregressive moving average. CLPM = cross-lagged panel model. RI-CLPM = random intercept cross-lagged panel model. The model matrices are defined in Eqs. 6-9. Indices *i, j, k, l* refer to a particular element of a given vector or matrix; indices *m* and *n* denote different groups.

analyses, the analysis model is fitted repeatedly to random data generated from a population with Σ and μ .

Regardless of which particular function is used to perform a model-based power analysis, a number of decisions always needs to be made. First, and obviously, the type of power analysis needs to be defined. Second, one needs to set the proper comparison model via the comparison argument. This can be either the saturated model (comparison = 'saturated') or a less restricted model that merely differs from the analysis model in the absence of the constraint(s) defining the null hypothesis (comparison = 'restricted'). Finally, the measurement model for the factors (i.e., Λ) needs to be defined. semPower 2 provides three ways to do this: by providing the complete loadings matrix via the Lambda argument, by providing the nonzero primary loadings on each factor via the loadings argument, or by providing a single loading magnitude to apply to each indicator of a factor or to all indicators (loadM) along with the number of indicators by factor (nIndicators). Given any definition of Λ , semPower 2 defines the variance-covariance matrix of the residuals Θ as a diagonal matrix with elements such that the variances of the indicators are equal to 1, implying that the loadings are in a standardized metric.

In what follows, power analyses for hypotheses arising in a CFA model, a CLPM model, and a generic path model are described to provide illustrative examples. In all cases, an a priori power analysis (type = 'a-priori') to yield a power of 80% (power = .80) on alpha = .05 (alpha = .05) in reference to a less restricted alternative model (comparison = 'restricted') is requested.

Power analysis for hypotheses arising in a CFA model

semPower 2 provides prebuilt methods to perform power analyses concerning the following hypotheses arising in a standard CFA model: (a) detect that a loading differs from zero, (b) detect that a correlation differs from zero, (c) detect that a correlation differs across groups, and (d) detect that two correlations differ from each other. Here we restrict ourselves to examples covering the latter three hypotheses.

Consider the introductory example of a CFA model involving two factors that are measured by three indicators each, and suppose one is interested in the sample size required to detect that the correlation between the factors is $r \geq .20$. This type of hypothesis is set via the nulleffect = 'cor = 0' argument. The Phi argument defines the correlation between the factors in the population to be .20. The measurement model is defined using the loadings argument, which is a list of vectors giving the (nonzero primary) loadings on each factor. In line with the

introductory example, assume that all loadings are equal to .50. Running `semPower.powerCFA` with these arguments provides an output structured identically to the one of a

corresponding model-free power analysis and indicates that 783 observations are required to detect this effect with a power of 80% on $\alpha = .05$.

```
powerCFA <- semPower.powerCFA(
  # define power analysis
  type = 'a-priori', alpha = .05, power = .80, comparison = 'restricted',
  # define model
  loadings = list(
    c(.5, .5, .5), # factor 1
    c(.5, .5, .5) # factor 2
  ),
  Phi = .2,
  # define hypothesis
  nullEffect = 'cor = 0'
)
summary(powerCFA)
```

The results variable (`powerCFA`) also contains additional information such as the population and model-implied variance–covariance matrices (`Sigma`, `SigmaHat`) and means (`mu`, `muHat`; if these are part of the model) as well as `lavaan`

model strings representing the H_0 (`modelH0`) and the H_1 (`modelH1`) models. These model strings can be plugged into `lavaan` to verify that the population was defined as intended:

```
library(lavaan)
lavres <- sem(model = powerCFA$modelH1, sample.cov = powerCFA$Sigma,
             sample.cov.rescale = FALSE, sample.nobs = 1000)
summary(lavres, standardized = TRUE)
```

It should be noted that the actual degree of discrepancy between the H_0 and the H_1 models (in terms of F_0) depends not only on the magnitude of the population correlation between the factors, but also on other parameters in the model. Indeed, power generally varies strongly with the number of indicators and the loading magnitude. When repeating the analysis above assuming five indicators for each factor with loadings of (.70, .80, .90, .70, .90) and (.80, .60, .90, .50, .70), respectively, the required N drops to 237. It is crucial to take care in defining appropriate factor loadings, which should be guided by previous empirical results.

`semPower.powerCFA` can also be used to perform power analyses regarding whether a correlation between factors differs across two or more groups by setting `nullEffect = 'corA = corB'`. Generally, `semPower 2` uses the “A” / “B” suffix (e.g., `corA = corB`) to refer to equality constraints across multiple groups, whereas the “X” / “Z” suffix (e.g., `corX = corZ`) is used to refer to equality constraints across different parameters in a single group (see below). The syntax for power analyses in a multiple group setting is highly similar to the previous example, except that the factor correlation must now be provided for each group, so the `Phi` argument becomes a list

containing the correlation in the first and second groups (of .00 and .20, respectively). Furthermore, any type of power analysis now expects the *N* argument, which gives the group weights in the case of an a priori power analysis. For instance, *N* = list(1, 1) specifies that both groups are of equal size. Given that cross-group comparisons of

factor correlations are only meaningful when both loadings and factor variances are equal across groups, semPower 2 implements these invariance constraints when determining power and therefore requires that the measurement model is defined only once (which then applies equally to all groups).

```
powerCFA <- semPower.powerCFA(
  # define power analysis
  type = 'a-priori', alpha = .05, power = .80, N = list(1, 1),
  comparison = 'restricted',
  # define model
  loadings = list(
    c(.5, .5, .5), # factor 1
    c(.5, .5, .5) # factor 2
  ),
  Phi = list(0, .2), # cor in group 1 and group 2
  # define hypothesis
  nullEffect = 'corA = corB'
)
summary(powerCFA)
```

Running the power analysis indicates that $N = 3116$ observations (1558 per group) are required to detect that the specified factor correlation differs across groups with a power of 80%. Note again that this power estimate is only valid for this particular measurement model and the particular values assumed for the correlations. For instance, when replacing *Phi* above by *Phi* = list(.6, .8), the required sample size is only $N = 2286$, although the difference between the correlations across groups remains .20.

semPower.powerCFA also supports hypotheses regarding the equality of correlations in a single-group design using nullEffect = 'corX = corZ'. Consider a CFA model involving two factors measured by three indicators each and an additional observed covariate. The factors are assumed to correlate by .50, and the correlations of the factors to the observed covariate are assumed to be .20 and .30, respectively. *Phi* now becomes a correlation matrix, where the first two columns refer to the factors,

and the final column refers to the observed covariate. The loadings argument defines the loadings on the first factor to equal .5, .6, and .7, and those on the second factor to equal .8, .4, and .8, and defines the observed covariate by specifying a single loading of 1. Note that if a single loading equal to 1 is provided for all factors, semPower.powerCFA reduces to the special case of power analyses for correlation coefficients between observed variables (Olkin & Finn, 1995). Suppose the interest lies in detecting that the two factors correlate differently to the observed covariate. The nullWhich argument is a list comprising two vectors, jointly defining which elements in *Phi* to set to equality, so in the present example nullWhich = list(c(1, 3), c(2, 3)) targets the correlation between the first factor and the covariate as well as the correlation between the second factor and the covariate. Based on these input values, the required sample size is $N = 1329$.

```

# define factor correlation matrix
Phi <- matrix(c(
  # f1 f2 cov
  c(1, .5, .2),
  c(.5, 1, .3),
  c(.2, .3, 1)),
  ncol = 3, byrow = TRUE)

powerCFA <- semPower.powerCFA(
  # define power analysis
  'a-priori', alpha = .05, power = .80, N = list(1, 1), comparison = 'restricted',
  # define model
  loadings = list(
    c(.5, .6, .7), # factor 1
    c(.8, .4, .8), # factor 2
    c(1)          # observed covariate
  ),
  Phi = Phi,
  # define hypothesis
  nullEffect = 'corX = corZ',
  nullWhich = list(c(1, 3), c(2, 3))
)
summary(powerCFA)

```

Power analysis for hypotheses arising in a CLPM

The `semPower.powerCLPM` function can be used to perform power analyses concerning different hypotheses arising in a CLPM. In the present example, the CLPM involves two factors X and Y, each of which are measured at three waves. This allows for various relevant hypotheses: (a) detect that an autoregressive effect differs from zero, (b) detect that a cross-lagged effect differs from zero, (c) detect that the autoregressive effect of X differs from the autoregressive effect of Y, (d) detect that the cross-lagged effect of X differs from the cross-lagged effect of Y, (f) detect that an autoregressive effect differs from wave to wave, (g) detect that a cross-lagged effect differs from wave to wave, (h) detect that the synchronous residual correlations between X and Y differ across waves, (i) detect that an autoregressive effect differs across groups, and (j) detect that a cross-lagged effect differs across groups. In the following examples, we exemplarily consider power analyses to detect that a cross-lagged effect differs from zero, that the cross-lagged effect of X differs from the one of Y, and that the cross-lagged effect of X varies across waves.

While there are different ways for specifying a CLPM with three waves ($nWaves = 3$) (e.g., Newsom, 2015), a common approach is to place equality constraints on the autoregressive and cross-lagged effects on waves, so that

there is only one autoregressive and one cross-lagged effect each for X and Y. This specification can be achieved by setting the `waveEqual` argument accordingly (`waveEqual = c('autoregX', 'autoregY', 'crossedX', 'crossedY')`). These equality constraints then apply to both the H0 model implementing the hypothesized effect and the H1 model omitting the restriction associated with the hypothesized effect. Similarly, when X and Y are latent factors (as in the present example), metric invariance over waves and autocorrelated indicator residuals are usually assumed. This is implemented by setting the arguments `metricInvariance = TRUE` and `autocorResiduals = TRUE`, which also applies to both models. The population values for the autoregressive effects for X and Y are provided in the `autoregEffects` argument, and take values of .60 for X and .70 for Y (both equal across waves). The cross-lagged effects (equal across waves) are defined in the `crossedEffects` argument, specifying a cross-lagged effect of X on Y of .10 and a cross-lagged effect of Y on X of .20. The synchronous (residual) correlations between X and Y are defined to be .30, .20, and .10 at the first, second, and third waves, respectively ($r_{XY} = c(.3, .1, .1)$). Because the standardized argument is set to `TRUE`, all coefficients are treated as completely standardized, which implies that `semPower 2` generates the (residual) variance–covariance matrix of the factors (Ψ) and the variance–covariance matrix of the indicators (Θ) such

that all variances are equal to 1. If `standardized = FALSE`, `semPower` sets all diagonal elements of Ψ equal to 1, thus leading to unstandardized synchronous correlations, autoregressive, and cross-lagged effects. Finally, the hypothesis

of interest is defined in the `nullEffect` argument. Here, the required sample size is requested to detect that the cross-lagged effect of X on Y (`nullEffect = 'crossedX = 0'`) is $\geq .10$ with a power of 80% on $\alpha = .05$, which yields $N = 958$.

```
powerCLPM <- semPower.powerCLPM(
  # define power analysis
  type = 'a-priori', alpha = .05, power = .80, comparison = 'restricted',
  # define model
  nWaves = 3,
  autoregEffects = c(.6, .7),      # (X1 -> X2 & X2 -> X3), (Y1 -> Y2 & Y2 -> Y3)
  crossedEffects = c(.1, .2),    # (X1 -> Y2 & X2 -> Y3), (Y1 -> X2 & Y2 -> X3)
  rXY = c(.3, .1, .1),          # X1 <-> Y1, X2 <-> Y2, X3 <-> Y3
  waveEqual = c('autoregX', 'autoregY', 'crossedX', 'crossedY'),
  loadings = list(
    c(.4, .5, .6, .4), # X1
    c(.5, .7, .5),    # Y1
    c(.4, .5, .6, .4), # X2
    c(.5, .7, .5),    # Y2
    c(.4, .5, .6, .4), # X3
    c(.5, .7, .5)     # Y3
  ),
  standardized = TRUE,
  metricInvariance = TRUE,
  autocorResiduals = TRUE,
  # define hypothesis
  nullEffect = 'crossedX = 0'
)
summary(powerCLPM)
```

If the required sample size to detect that the cross-lagged effect of X (of .10) differs from the cross-lagged effect of Y (of .20) is to be determined, the only change refers to the `nullEffect` argument. Setting `nullEffect = 'crossedX = crossedY'` and rerunning `semPower.powerCLPM` gives a required sample of $N = 2208$ for a power of 80% on $\alpha = .05$.

If the CLPM is based on observed variables (rather than latent factors), the measurement model needs to be defined such that there is a single loading of 1 for each factor, which is most easily done by providing an identity matrix to the `Lambda` argument. For instance, when replacing the `loadings` argument by `Lambda = diag(6)` in the previous example, the required sample size to detect that the cross-lagged effects of X and Y differ with a power of 80% is N

$= 449$. The `metricInvariance` and `autocorResiduals` arguments are ignored for the observed variables.

The previous examples implemented the CLPM assuming constant autoregressive and cross-lagged effects across waves. Wave-dependent effects can be specified by dropping the respective entries in the `waveEqual` argument and providing lists to the `autoregEffects` and `crossedEffects` arguments. In the following example, the autoregressive effects are still constant across waves, whereas the cross-lagged effect of X on Y is .10 from wave 1 to 2, but .30 from wave 2 to 3. Then, the required sample size to detect that the autoregressive effects of X vary across waves (`nullEffect = crossedX`) is requested, which yields $N = 1348$.


```

powerCLPM <- semPower.powerCLPM(
  # define power analysis
  type = 'a-priori', alpha = .05, power = .80, comparison = 'restricted',
  # define population model
  nWaves = 3,
  autoregEffects = list(
    c(.7, .7),          # X1 -> X2, X2 -> X3
    c(.6, .6)          # Y1 -> Y2, Y2 -> Y3
  ),
  crossedEffects = list(
    c(.1, .3),        # X1 -> Y2, X2 -> Y3
    c(.2, .2)        # Y1 -> X2, Y2 -> X3
  ),
  rXY = c(.3, .1, .1), # X1 <-> Y1, X2 <-> Y2, X3 <-> Y3
  loadings = list(
    c(.4, .5, .6, .4), # X1
    c(.5, .7, .5),    # Y1
    c(.4, .5, .6, .4), # X2
    c(.5, .7, .5),    # Y2
    c(.4, .5, .6, .4), # X3
    c(.5, .7, .5)     # Y3
  ),
  standardized = TRUE,
  metricInvariance = TRUE,
  autocorResiduals = TRUE,
  # define hypothesis
  nullEffect = 'crossedX'
)
summary(powerCLPM)

```

Power analysis for hypotheses arising in a generic path model

Although `semPower 2` contains several functions for special regression problems (such as a single linear regression, mediation structures, and CLPMs; see Table 1), some regression-related hypotheses are not covered. Therefore, `semPower 2` also provides a more generic function to perform model-based power analyses arising in path models that requires specification of \mathbf{B} and, optionally, Ψ .

As an example, consider a model involving two correlated factors and two correlated observed outcomes. A substantive hypothesis might be that the slope of the first factor in the prediction of the first outcome does not differ from the slope of the second factor in the prediction of the second outcome. The Beta matrix defined in the code below sets up the regression relationships in the population and implies a slope for the first and second factors of .20 and .10, respectively, in the prediction of the first outcome, and a slope for the first and second

factors of .10 and .40, respectively, in the prediction of the second outcome. Further, the Psi matrix defines a correlation between the factors of .30, and a residual correlation between the outcomes of .20. If the standardized argument is set to TRUE, all slopes are interpreted in a standardized metric and the matrix provided to Psi is interpreted as the matrix of (residual-)correlations, so that the actual Ψ is computed such that all variances are equal to 1. If `standardized = FALSE`, the coefficients in Beta and Psi are directly used to define \mathbf{B} and Ψ , thus (in general) implying unstandardized coefficients. Finally, the hypothesis that the slope of the first factor in the prediction of the first outcome is equal to the slope of the second factor in the prediction of the second outcome is defined in `nullEffect = 'betaX = betaZ'` along with the `nullWhich` argument, which is a list containing vectors referring to which elements in Beta to constrain to equality. Running `semPower.powerPath` with these arguments leads to a required sample size of $N = 840$ to detect the specified effect with a power of 80%.

```

Beta <- matrix(c(
  c(0.0, 0.0, 0.0, 0.0), # F1
  c(0.0, 0.0, 0.0, 0.0), # F2
  c(0.2, 0.1, 0.0, 0.0), # Y1 = .2*F1 + .1*F2
  c(0.1, 0.4, 0.0, 0.0) # Y2 = .1*F1 + .4*F2
), ncol = 4, byrow = TRUE)

Psi <- matrix(c(
  c(1.0, 0.3, 0.0, 0.0), # F1
  c(0.3, 1.0, 0.0, 0.0), # F2
  c(0.0, 0.0, 1.0, 0.2), # Y1
  c(0.0, 0.0, 0.2, 1.0) # Y2
), ncol = 4, byrow = TRUE)

powerPath <- semPower.powerPath(
  # define power analysis
  type = 'a-priori', alpha = .05, power = .80, comparison = 'restricted',
  # define population model
  Beta = Beta,
  Psi = Psi,
  loadings = list(
    c(.6, .5, .7), # F1
    c(.6, .5, .7), # F2
    c(1), # Y1
    c(1) # Y2
  ),
  standardized = TRUE,
  nullEffect = 'betaX = betaZ',
  nullWhich = list(c(3, 1), c(4, 2))
)
summary(powerPath)

```

Generic model-based power analysis

semPower 2 also provides a generic function to perform model-based power analyses for scenarios not covered in any more specific function. Consider the situation in which one is interested in determining whether the observed responses on eight indicators reflect two separate (but correlated) factors or can be described by assuming just a single factor, and that a correlation of $r \geq .90$ is considered to imply that a single factor is sufficient. A suitable H0 model to test this hypothesis would specify two factors and constrain their

correlation to 1. Although the model conforms to a standard CFA model, this particular hypothesis is not covered by the `semPower.powerCFA` function, so the more general `semPower.powerLav` function needs to be employed.

`semPower.powerLav` requires a lavaan model string defining the H0 model (and optionally a model string defining the H1 model), and either a lavaan model string defining the population model or a population variance–covariance matrix (and mean vector, if necessary). A useful utility function supporting the definition of the population variance–covariance matrix is `semPower.genSigma`, which

expects the model matrices (see Eqs. 6-9) as input, but accepts the same shortcuts as any other function performing model-based power analysis. For instance, the following generates the implied covariance matrix from a population model involving two factors that are correlated by .90 and are measured by four indicators each.

```
generated <- semPower.genSigma(
  Phi = .90,
  loadings = list(
    c(.7, .7, .5, .5),
    c(.8, .6, .6, .4))
)
```

Beyond the implied covariance matrix (Sigma), the results variable returned by `semPower.genSigma` also includes a lavaan model string (`modelTrue`) created such that the model perfectly describes the population. In the present example, this is a simple two-factor model that freely estimates the factor correlation, and thereby represents a suitable H1 model. The model string representing the H0 must restrict the factor correlation to be equal to 1:

```
modelH0 <- '
  f1 =~ NA*x1 + x2 + x3 + x4
  f2 =~ NA*x5 + x6 + x7 + x8
  f1 ~~ 1*f1
  f2 ~~ 1*f2
  # restrict correlation to 1
  f1 ~~ 1*f2
'
```

Finally, everything is plugged into `semPower.powerLav`, in the present example requesting an a priori power analysis, which shows that 323 observations are required to detect a factor correlation $< .90$ with a power of 80%.

```
powerLav <- semPower.powerLav(
  # define power analysis
  type = 'a-priori', alpha = .05, power = .80,
  # define population Sigma
  Sigma = generated$Sigma,
  # define analysis models
  modelH0 = modelH0,
  modelH1 = generated$modelTrue
)
summary(powerLav)
```

Simulated power analysis

All model-based power analyses provided by `semPower 2` can also be performed as a simulated, rather than analytical, power analysis by adding the argument `simulatedPower = TRUE` to the respective function call. In addition, a list with options controlling the simulation process can be provided via the `simOptions` argument, allowing for changing the number of replications, parallel processing, and the generation of non-normal data and/or missingness. For instance, the following code performs a simulated a priori power analysis concerning the correlation in a two-factor model using the default settings (i.e., 500 replications, multivariate normal and complete data, ML estimation), and additionally requests parallel processing using four cores (`simOptions = list(nCores = 4)`):

```

set.seed(1234)
powerCFA <- semPower.powerCFA(
  # define power analysis
  type = 'a-priori', alpha = .05, power = .80, comparison = 'restricted',
  # define model
  loadings = list(
    c(.7, .8, .6, .6, .7), # F1
    c(.8, .8, .5, .6, .5) # F2
  ),
  Phi = .2,
  # define hypothesis
  nullEffect = 'cor = 0',
  # request simulated power
  simulatedPower = TRUE,
  # set simulation options
  simOptions = list(
    nCores = 4
  )
)
summary(powerCFA)

```

Given that simulated power analyses are arguably most useful when the data generation process implements violations of the assumptions underlying the chosen estimator (and as a consequence also the assumptions underlying analytical power analysis), *semPower 2* includes options to generate non-normal data and data with missing values. Generating multivariate non-normal data with known covariance structure is an intricate issue and active area of research. Therefore, *semPower 2* interfaces four different data generation methods, namely Foldnes and Olsson (2016), Qu et al. (2020), Ruscio and Kacetow (2008), and Vale and Maurelli (1983), each of which is associated with different multivariate distributions that may affect the empirically observed χ^2 -distributions differently (e.g., Auerswald & Moshagen, 2015; Foldnes &

Grønneberg, 2015; Jobst et al., 2022) and requires different input. As an example, the following code uses the Vale-Maurelli approach (`type = 'VM'`) specifying symmetric (skewness), but moderately to strongly kurtotic (kurtosis) marginal distributions. Moreover, there are also missing data on the first and second indicator of the first factor as well as the final two indicators of the second factor (`missingVars = c(1, 2, 9, 10)`), generated based on a missing-at-random mechanism (`missingMechanism = 'MAR'`) with a missing proportion of 25% (`missingProp = .25`). For MAR, *semPower 2* implements the percentile method generating the maximum number of missing patterns (Zhang, 2021). When missing data are generated, *semPower 2* resorts to full information maximum likelihood estimation.

```

set.seed(1234)
powerCFA <- semPower.powerCFA(
  # define power analysis
  type = 'a-priori', alpha = .05, power = .80, comparison = 'restricted',
  # define model
  loadings = list(
    c(.7, .8, .6, .6, .7), # factor 1
    c(.8, .8, .5, .6, .5) # factor 2
  ),
  Phi = .2,
  # define hypothesis
  nullEffect = 'cor = 0',
  # request simulated power
  simulatedPower = TRUE,
  # set simulation options
  simOptions = list(
    nReplications = 500,
    nCores = 4,
    type = 'VM',
    skewness = c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0),
    kurtosis = c(10, 10, 10, 10, 10, 15, 15, 15, 15, 15),
    # missingness
    missingVars = c(1, 2, 9, 10),
    missingMechanism = 'MAR',
    missingProp = .25
  )
)
summary(powerCFA)

## semPower: A priori power analysis
##
## Simulated power based on 500 successful replications.
## Note that simulated a-priori power analyses are only approximate,
## unless the number of replications is large.
##
## Analytical Simulated
##
## F0 0.027787 0.028790
## RMSEA 0.166695 0.169675
## Mc 0.986202 0.985708
##
## df 1 1
## Required Num Observations 284 291
##
## Critical Chi-Square 3.841459 3.841459
## NCP 7.863790 8.349022
## Alpha 0.050000 0.050000
## Beta 0.199255 0.184000
## Power (1 - Beta) 0.800745 0.816000
## Implied Alpha/Beta Ratio 0.250934 0.271739
##
## Simulation Results:
##
## Convergence Rate (%) of the H0 model 100.00
##
## Chi-Square Bias (%)
## H0 Model 17.77
## H1 Model 20.98
## H0-H1 Difference 5.44
##
## Chi-Square KS-Distance
## H0 Model 0.173592
## H1 Model 0.198021
## H0-H1 Difference 0.014572
##
## Rejection Rate (%)
## H0 Model 46.60
## H1 Model 22.40
##
## Average Parameter Bias (%) in the H1 Model:
## Loadings -1.59
## Variances/Covariances -0.39

```


The output now contains an additional column containing the simulation results. Note that the measures of effect refer to the estimates of the population values (rather than being averaged sample values), so these can be immediately compared across the different approaches to power analysis. Of note (and perhaps surprisingly, but consistent with Steiger et al., 1985), the simulation results very closely match the analytical results in terms of the sample size required to detect the hypothesized effect despite the non-normal and incomplete data.

The output further reports on the simulation results in greater detail. In particular, the convergence rate (of the H0 model) and the average relative parameter bias by type of model parameter is provided (i.e., $RB = 100 \cdot (\hat{\theta} - \theta)/\theta$), jointly giving an indication of whether the sample size is sufficient to support successful parameter estimation of the model under scrutiny. In addition, the empirically observed central and noncentral χ^2 distributions are compared against the respective asymptotically expected reference distribution using three measures. First, the relative bias evaluates the departure of the empirical mean from the expected value ($RB = 100 \cdot (\bar{\chi}^2 - df)/df$ and $RB = 100 \cdot (\bar{\chi}^2 - (df + \lambda))/(df + \lambda)$), respectively). Second, the average Kolmogorov–Smirnov (KS) distance measures the discrepancy between the empirical and the reference cumulative distribution functions akin to the Kolmogorov–Smirnov test statistic, but relies on the average, rather than the maximum, absolute distance, $KS = M(|F_e(x) - F_r(x)|)$ (see Yuan et al., 2007). The average KS distance thus ranges from 0 (for no difference) to .50 (for completely nonoverlapping distributions). Finally, the empirical rejection rate given the specified alpha-error is provided as a measure of how strongly the tail of the empirical distribution differs from the asymptotic reference distribution. When the model representing the H0 is

compared against the saturated model, this is the same as the empirical power estimate.

In the example above, the relative bias and the KS-distance indicate strong departures of the empirical distributions from the respective asymptotic distribution, which is also mirrored by a rejection rate of the (correctly specified) H1 model that clearly exceeds the nominal alpha-error of 5%. However, the difference distribution closely follows the theoretically expected distribution, so there is virtually no difference (beyond sampling error) between the analytical and the simulated power estimate. Indeed, it is our experience that simulated power analyses for the comparison against an explicit H1 model usually show only minor departures from corresponding analytical power analyses. When the comparison is performed against the saturated model, however, simulated power is often higher as a consequence of the well-documented positive bias of the model test when the assumptions are violated (e.g., Curran et al., 1996; Moshagen, 2012).

By default, `semPower 2` performs simulated power analyses employing ML estimation with the uncorrected model test statistic. However, `semPower 2` also allows one to change the estimator and test statistic employed in a simulated power analysis by setting the `lavOptions` argument, which is passed to `lavaan` and thus conforms to the standard `lavaan` conventions. For instance, `lavOptions = list(estimator = 'mlr')` leads to a simulated power estimate relying on a corrected test statistic that is asymptotically equivalent to the Yuan & Bentler (2000) statistic, whereas the analytic power estimate is still based on the asymptotically expected distributions based on ML. Repeating the previous example with this additional argument yields less biased distributions of the corrected test statistic under both the H0 and the H1 as compared to the results obtained above using the uncorrected test statistic. Note that the simulated power estimate is still very close to the analytical power estimate:

```

##
##           Analytical Simulated
##
## F0           0.027787  0.028610
## RMSEA        0.166695  0.169146
## Mc           0.986202  0.985797
##
## df           1         1
## Required Num Observations 284     284
##
## Critical Chi-Square    3.841459  3.841459
## NCP                7.863790  8.096706
## Alpha              0.050000  0.050000
## Beta               0.199255  0.206000
## Power (1 - Beta)    0.800745  0.794000
## Implied Alpha/Beta Ratio 0.250934  0.242718
##
##
## Simulation Results:
##
## Convergence Rate (%) of the H0 model      100.00
##
## Chi-Square Bias (%)
## H0 Model                                  15.08
## H1 Model                                  19.04
## H0-H1 Difference                          6.03
##
## Chi-Square KS-Distance
## H0 Model                                  0.151380
## H1 Model                                  0.171136
## H0-H1 Difference                          0.025160
##
## Rejection Rate (%)
## H0 Model                                  44.80
## H1 Model                                  20.80
##
## Average Parameter Bias (%) in the H1 Model:
## Loadings                                  -1.76
## Variances/Covariances                     1.00

```

Conclusion

The present paper provided an introductory overview of the `semPower 2` package for the R environment, which provides both analytical and simulated a priori, post hoc, and compromise power analysis based on either a model-free or a model-based definition of the effect and provides high-level functions to simplify the definition of meaningful hypotheses arising in popular types of models (which are to be expanded to cover additional models in future versions).

`semPower 2` thus offers a high degree of both accessibility and flexibility in performing power analysis, and thereby overcomes several limitations of previous tools providing either analytical or simulated power analysis (but not both), thus complicating a direct comparison across these approaches. In addition, most existing tools offering simulated power analysis provide post hoc power analyses only, often consider particular model types, require the cumbersome and error-prone definition of the values of all population parameters, and do not offer easily accessible options to

change the estimator or to embed non-normally distributed data and or different patterns of missingness in the simulation process.

Although the examples provided in the present paper focused on power analysis concerning models that include latent variables, semPower also provides power analysis of correlation and regression problems involving only observed variables as a special case. As such, we hope that semPower 2 may promote the use of power analysis in determining the required sample size and evaluating hypothesis tests in a meaningful way more generally.

Limitations of the current version of semPower 2 include that multilevel models and models involving ordered categorical data are not yet supported. Also, semPower 2 is operated entirely through the R language. Although a Shiny app providing model-free power analyses is available at <https://sempower.shinyapps.io/sempower>, there is currently no graphical user interface covering the complete functionality of semPower 2. Readers who are more comfortable with a graphical user interface are referred to Jak et al. (2021) for analytical and to Wang and Rhemtulla (2021) for simulated power analyses.

Finally, it must be stressed that the prime purpose of semPower 2 is to perform power analyses concerning a particular hypothesis. Although the simulation option provides some indication of whether the required sample size yielding a particular power is also sufficient to support successful estimation of the model parameters, the question of whether a certain sample size guarantees proper convergence and good parameter recovery is not the focus of semPower 2. Statistical power is only one of several aspects to consider in determining sample size requirements.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Auerswald, M., & Moshagen, M. (2015). Generating correlated, non-normally distributed data using a non-linear structural model. *Psychometrika*, *80*, 920–937.
- Bader, M., & Moshagen, M. (2022). No probifactor model fit index bias, but a propensity toward selecting the best model. *Journal of Psychopathology and Clinical Science*, *131*(6), 689–695. <https://doi.org/10.1037/abn0000685>
- Bentler, P. M., & Satorra, A. (2010). Testing model nesting and equivalence. *Psychological Methods*, *15*(2), 111–123. <https://doi.org/10.1037/a0019625>
- Browne, M. W. (1984). Asymptotic distribution free methods in the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 127–141.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255.
- Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, *1*, 16–29.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Foldnes, N., & Grønneberg, S. (2015). How general is the Vale-maurelli simulation approach? *Psychometrika*, *80*(4), 1066–1083. <https://doi.org/10.1007/s11336-014-9414-0>
- Foldnes, N., & Olsson, U. H. (2016). A simple simulation technique for nonnormal data with prespecified skewness, kurtosis, and covariance matrix. *Multivariate Behavioral Research*, *51*(2–3), 207–219. <https://doi.org/10.1080/00273171.2015.1133274>
- Jackson, D. L., Gillaspay, J. A., Jr., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, *14*, 6–18.
- Jak, S., Jorgensen, T. D., Verdam, M. G. E., Oort, F. J., & Elffers, L. (2021). Analytical power calculations for structural equation modeling: A tutorial and shiny app. *Behavior Research Methods*, *53*(4), 1385–1406. <https://doi.org/10.3758/s13428-020-01479-0>
- Jobst, L. J., Auerswald, M., & Moshagen, M. (2022). The effect of latent and error non-normality on measures of fit in structural equation modeling. *Educational and Psychological Measurement*, *82*(5), 911–937. <https://doi.org/10.1177/00131644211046201>
- Jobst, L. J., Bader, M., & Moshagen, M. (2023). A tutorial on assessing statistical power and determining sample size for structural equation models. *Psychological Methods*, *28*(1), 207–221. <https://doi.org/10.1037/met0000423>
- Jöreskog, K. G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, *32*(4), 443–482. <https://doi.org/10.1007/BF02289658>
- Jöreskog, K. G. (1978). Structural analysis of covariance and correlation matrices. *Psychometrika*, *43*(4), 443–477. <https://doi.org/10.1007/BF02293808>
- Kievit, R. A., Brandmaier, A. M., Ziegler, G., van Harmelen, A.-L., de Mooij, S. M., Moutoussis, M., Goodyer, I. M., Bullmore, E., Jones, P. B., & Fonagy, P. (2018). Developmental cognitive neuroscience using latent change score models: A tutorial and applications. *Developmental Cognitive Neuroscience*, *33*, 99–117.
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130–149.
- MacCallum, R. C., Browne, M. W., & Cai, L. (2006). Testing differences between nested covariance structure models: Power analysis and null hypotheses. *Psychological Methods*, *11*, 19–35.
- McDonald, R. P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of Classification*, *6*, 97–103.
- Miles, J. (2003). A framework for power analysis using a structural equation modelling procedure. *BMC Medical Research Methodology*, *3*(1), 27. <https://doi.org/10.1186/1471-2288-3-27>

- Moshagen, M. (2012). The model size effect in SEM: Inflated goodness-of-fit statistics are due to the size of the covariance matrix. *Structural Equation Modeling*, *19*, 86–98.
- Moshagen, M., & Auerswald, M. (2018). On congruence and incongruence of measures of fit in structural equation modeling. *Psychological Methods*, *23*, 318–336.
- Moshagen, M., & Erdfelder, E. (2016). A new strategy for testing structural equation models. *Structural Equation Modeling*, *23*, 54–60.
- Mulder, J. D. (2022). Power Analysis for the Random Intercept Cross-Lagged Panel Model Using the powRICLPM R-Package. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–14. <https://doi.org/10.1080/10705511.2022.2122467>
- Muthén, L. K., & Muthén, B. (2002). How to use a Monte Carlo study to decide on sample size and determine power. *Structural Equation Modeling*, *9*, 599–620.
- Newsom, J. T. (2015). *Longitudinal structural equation modeling: A comprehensive introduction*. Routledge.
- Olkin, I., & Finn, J. D. (1995). Correlations redux. *Psychological Bulletin*, *118*(1), 155–164. <https://doi.org/10.1037/0033-2909.118.1.155>
- Preacher, K. J., & Coffman, D. L. (2006). *Computing power and minimum sample size for RMSEA* [Computer software]. <http://quantpsy.org>. Accessed 3/11/2023.
- Qu, W., Liu, H., & Zhang, Z. (2020). A method of generating multivariate non-normal random numbers with desired multivariate skewness and kurtosis. *Behavior Research Methods*, *52*, 939–946.
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36.
- Ruscio, J., & Kaczetow, W. (2008). Simulating multivariate nonnormal data using an iterative algorithm. *Multivariate Behavioral Research*, *43*(3), 355–381. <https://doi.org/10.1080/00273170802285693>
- Saris, W. E., & Satorra, A. (1993). Power evaluations in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (154th ed., pp. 181–204). Sage.
- Satorra, A., & Saris, W. E. (1983). The accuracy of a procedure for calculating the power of the likelihood ratio test as used within the LISREL framework. In C. O. Middendorp (Ed.), *Sociometric research* (pp. 129–190). Sociometric Research Foundation.
- Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika*, *50*, 83–90.
- Schoemann, A. M., Boulton, A. J., & Short, S. D. (2017). Determining power and sample size for simple and complex mediation models. *Social Psychological and Personality Science*, *8*(4), 379–386. <https://doi.org/10.1177/1948550617715068>
- Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika*, *50*, 253–263.
- Vale, C. D., & Maurelli, V. A. (1983). Simulating multivariate non-normal distributions. *Psychometrika*, *48*, 465–471.
- Wang, Y. A., & Rhemtulla, M. (2021). Power analysis for parameter estimation in structural equation modeling: a discussion and tutorial. *Advances in Methods and Practices in Psychological Science*, *4*(1), 251524592091825. <https://doi.org/10.1177/2515245920918253>
- Yuan, K. H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, *30*, 165–200.
- Yuan, K. H., & Bentler, P. M. (2007). Robust procedures in structural equation modeling. In S.-Y. Li (Ed.), *Handbook of latent variable and related models* (pp. 367–397). North-Holland.
- Yuan, K. H., Hayashi, K., & Bentler, P. M. (2007). Normal theory likelihood ratio statistic for mean and covariance structure analysis under alternative hypotheses. *Journal of Multivariate Analysis*, *98*, 1262–1282.
- Zhang, Z., & Liu, H. (2018). Sample Size and Measurement Occasion Planning for Latent Change Score Models through Monte Carlo Simulation 1. In E. Ferrer, S. M. Boker, & K. J. Grimm (Eds.), *Longitudinal Multivariate Psychology* (1st ed., pp. 189–211). Routledge. <https://doi.org/10.4324/9781315160542-10>
- Zhang, Z., & Yuan, K.-H. (2018). *Practical statistical power analysis using Webpower and R*. Isdsa Press.
- Zhang, X. (2021). *Tutorial: How to Generate Missing Data For Simulation Studies* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/rq6yb>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.