Behavior Research Methods (2024) 56:8105–8131
https://doi.org/10.3758/s13428-024-02435-y

NOTES/COMMENTS/REPLY MANUSCRIPT

Enabling analytical power calculations for multilevel models with autocorrelated errors through deriving and approximating the precision matrix

Ginette Lafit[1] · **Richard Artner**[2] · **Eva Ceulemans**[2]

Accepted: 22 April 2024 / Published online: 15 July 2024
© The Psychonomic Society, Inc. 2024

Abstract
To unravel how within-person psychological processes fluctuate in daily life, and how these processes differ between persons, intensive longitudinal (IL) designs in which participants are repeatedly measured, have become popular. Commonly used statistical models for those designs are multilevel models with autocorrelated errors. Substantive hypotheses of interest are then typically investigated via statistical hypotheses tests for model parameters of interest. An important question in the design of such IL studies concerns the determination of the number of participants and the number of measurements per person needed to achieve sufficient statistical power for those statistical tests. Recent advances in computational methods and software have enabled the computation of statistical power using Monte Carlo simulations. However, this approach is computationally intensive and therefore quite restrictive. To ease power computations, we derive simple-to-use analytical formulas for multilevel models with AR(1) within-person errors. Analytic expressions for a model family are obtained via asymptotic approximations of all sample statistics in the precision matrix of the fixed effects. To validate this analytical approach to power computation, we compare it to the simulation-based approach via a series of Monte Carlo simulations. We find comparable performances making the analytic approach a useful tool for researchers that can drastically save them time and resources.

Keywords Statistical power · Intensive longitudinal designs · Linear mixed effect models · AR(1) errors

Introduction

In recent years, many researchers have conducted intensive longitudinal (IL) studies (see e.g., Ariens et al. 2020; Eisele et al. 2022; Mestdagh and Dejonckheere 2021; Myin-Germeys et al. 2009; Myin-Germeys et al. 2018). These studies are characterized by repeated measurement occasions nested within individuals. Such datasets are commonly used in investigations of a wide variety of theories and research hypotheses related to how within-person psychological processes fluctuate in daily life, and how these processes differ between persons. In particular, researchers often investigate inter-group or inter-personal differences in the mean level of a time-varying target outcome, within-person effects of time-varying predictors, and interaction effects between time-varying and invariant predictors (see e.g., Heininga et al. 2019; Dejonckheere et al. 2018; Myin-Germeys et al. 2003). Typically, these research questions are investigated by fitting multilevel (or mixed-effects) regression models to the IL data and by conducting statistical hypotheses tests for the model parameters (see Raudenbush and Bryk (2002) for an introduction to multilevel modeling). These models account for the nested structure of the data by allowing some or all regression coefficients (i.e., intercepts and slopes) to vary across persons. Thus, this model estimates fixed effects (i.e., the average intercept or slope across persons) and random effects, which represent how person-specific intercepts or slopes deviate from the corresponding fixed effect.

As previously stated, IL data are characterized by repeated measurements that are closer in time, implying some form of dependence between measurements. Thus, when modeling IL data, neglecting the serial dependence by using traditional regression models may introduce misspecification

✉ Ginette Lafit
ginette.lafit@kuleuven.be

[1] Methodology of Educational Sciences, KU Leuven, Leuven, Belgium

[2] Research Group of Quantitative Psychology and Individual Differences, KU Leuven, Leuven, Belgium

biases (Ariens et al., 2023). For multilevel models applied to IL data, a common way of handling the serial dependency between the repeated measurements is to assume the within-person errors to be serially correlated through an AR(1) process (Goldstein et al., 1994). Failing to account for the serial dependency by ignoring that within-person errors are serially correlated can lead to overestimation of the variance of the random effects and standard errors of the fixed effects estimates (see e.g., Ferron et al. 2002; Kwok et al. 2007). Therefore, in this paper, we centered our attention on multilevel models with autocorrelated errors.

When designing an IL study, it is important to have an idea about statistical power for the to-be-tested statistical hypotheses, to inform sample size decisions (Lakens, 2022). In IL studies, the sample size is determined by both the number of participants and the number of measurements per participant. Statistical power is then defined as the probability of correctly rejecting the null hypothesis when the alternative hypothesis is true in the population under study (Cohen, 2013). The statistical power to detect a hypothesized effect is determined by the size of the effect in the population, the predetermined type I error rate (i.e., the significance level), and the standard error of the test statistic used. Power is higher if the population effect is larger, the type I error rate is higher, and the standard error of the test statistic is smaller. The standard error is defined as the standard deviation of the sampling distribution of the test statistic and it is a function of both the true values of the parameters of the model considered and the sample size of the study. For many non-IL models, a priori (i.e., before collecting the data) power computations can be done by making use of analytical formulas for computing the standard error of the test statistic, which form the backbone of widely used software packages such as G*Power (Faul et al., 2007).

Even though a priori power analysis is the most common approach for deciding on sample size (Lakens, 2022), most of the published empirical studies that use IL designs do not explicitly report having done one. For example, a recent review found that only 2% of the articles published between 2012 and 2018 in the field of daily life research on psychopathology explicitly reported a power analysis (Trull and Ebner-Priemer, 2020). This may be explained by the fact that analytical formulas (see e.g., Hedeker et al. 1999; Moerbeek et al. 2000; Moerbeek et al. 2001; Moerbeek and Maas 2005; Raudenbush 1997; 2001 Raudenbush and Liu; Snijders and Bosker 1993; Wang et al. 2015) for the computation of power are model-specific and have not yet been derived for models that explicitly account for the serial dependency that characterizes IL data. These issues are discussed in detail in Arend and Schäfer (2019), Lafit et al. (2021), and Murayama et al. (2022). Furthermore, ignoring the serial dependency in the within-person errors can yield underpowered IL studies due to the overestimation of the standard errors of the fixed effects estimates which lower statistical power (Kwok et al., 2007).

The so-called simulation-based approach provides a work around. This approach gained popularity because it does not require the derivation of complex formulas for the standard error(s) of the used test statistic(s) (see Arend and Schäfer (2019) for an overview). To compute statistical power for a fixed effect of interest, a researcher can use the simulation-based approach as follows. First, the researcher selects a multilevel model and plausible values of its parameters (see Lafit et al. 2021, 2022, Lane and Hannes (2018) for a discussion on how to determine plausible values for the parameters of multilevel models for computing statistical power) and the number of participants and measurements per participant. Using these values, a large number of data sets are simulated. Each of these data sets is then used to fit the multilevel model and test the hypothesis of interest. Finally, statistical power is estimated as the percentage of generated datasets in which the null hypothesis of interest was rejected at the selected alpha level.

The simulation-based approach has the advantage that it can always be applied no matter how complicated the parametric statistical model of interest is. Researchers can easily implement this approach due to the availability of step-by-step tutorials and open-source applications (Arend and Schäfer, 2019; Astivia et al., 2019; Bolger, 2011; Browne et al., 2009; Green and MacLeod, 2016; Lafit et al., 2021; Lafit et al., 2022). Although these resources are a crucial addition to the toolkit of IL researchers, they have the downside of being computationally intensive. When the goal is to conduct a power analysis to select the number of persons, for a fixed number of repeated measurement occasions, the computational time can increase considerably since statistical power is computed by conducting Monte Carlo simulations varying the number of persons. This issue is exacerbated when varying both the number of persons and the number of repeated measurements per person. For example, in the illustrations included in Lafit et al. (2022), the computational time for obtaining statistical power using the simulation-based approach for the simplest proposed model (i.e., the longitudinal actor-partner interdependence model with fixed intercepts only) took over three hours when using an Intel Core i7-7820HQ processor. This computational burden poses limitations to the implementation of the simulation-based approach, especially since usually multiple potential sample sizes have to be evaluated. For example, for complicated models, this approach can only realistically be applied via large-scale computers (see e.g., Strubell et al. 2019 for a

discussion on resources used by complex computational approaches).

In this article, we provide a way out by returning to the analytic approach to power calculations for multilevel models with autocorrelated errors. Specifically, we focus on deriving analytical formulas for the standard error estimates that can be used for computing the statistical power of Wald-type tests for testing single fixed effects for five types of multilevel models in which repeated measurements (level 1) are nested within persons (level 2) where within-person errors follow an AR(1) process. These models allow investigating: (i) the effect of a continuous time-varying level 1 predictor on the mean level of a time-varying outcome; (ii) differences in the mean level of a time-varying outcome between two groups of persons; (iii) the effect of a level 1 continuous predictor on a time-varying outcome; (iv) interaction effects between a level 1 and a level 2 continuous predictors; and (v) differences between two groups of persons with respect to the effect of a continuous level 1 predictor on a time-varying outcome. For these models, we show that when the number of persons and the number of repeated measurement occasions per person is sufficiently large, we can obtain a closed-form expression of the standard error of the fixed effect estimates by deriving and approximating the information matrix of the fixed effect estimates of two-level models with autocorrelated errors. Using this closed-form expression, we use the analytic approach to compute statistical power for fixed effects in multilevel models with AR(1) errors, without having to fit the model to large amounts of simulated datasets.

In the remainder of the article, we first recap the analytic and simulation-based approach for conducting power analysis. Next, we analytically derive the precision matrix of five multilevel models that assume within-person errors are autocorrelated, and we validate the analytical approach by comparing its performance with the simulation-based approach. We conclude this paper with a general discussion.

Recap of the approaches for the computation of statistical power for multilevel models without AR(1) errors

In this section, we are going to juxtapose two approaches for conducting power analysis via an empirical example. The first approach is the *simulation-based approach*, which is currently implemented in the Shiny app PowerAnalysisIL (Lafit et al., 2021). The second one is the *analytical approach* which, prior to this paper, has only been developed for multilevel models without AR(1) errors. In this example, it will be shown in detail how to conduct an a priori power analysis for the statistical test of a fixed effect of interest in a simple two-level regression model via the analytical and via the simulation-based approach.

The data and the simple model

To demonstrate the two approaches for the computation of statistical power, we are going to use the popularity data set included in Hox et al. (2017). The popularity data set includes data from 2000 pupils in 100 classes. The main variables of interest are pupil popularity (measured on a scale from one to ten) and teacher experience (measured in years) (see Hox et al. 2017). This study showed that between-class differences in pupil popularity are associated with differences in teacher experience, suggesting that more experienced teachers facilitate stronger classroom ties. But are those findings replicable? Researcher Jane wants to conduct a new study to conceptually replicate this finding. To help Jane with her power analysis, we first need to know how she intends to test the substantive hypothesis that teacher experience is associated with pupil popularity once the data have been gathered.

Jane models the popularity score y_{ij} of pupil $j = (1, \ldots, n_i)$ in class $i = (1, \ldots, N)$ as

$$y_{ij} = \beta_{00} + \beta_{01} w_i + \nu_{0i} + \epsilon_{ij} \tag{1}$$

with w_i representing the experience of the teacher of class i. In Eq. 1, β_{00} denotes the fixed intercept and represents the average popularity level in a class where the teacher has no experience. The parameter β_{01} represents the influence of the class-level predictor, teacher experience[1], such that if β_{01} is positive, the average popularity is larger in classes with more experienced teachers. Conversely, if β_{01} is negative, the average popularity is lower in classes with more experienced teachers. To allow for additional between-class differences in the average level of popularity, we include the random variable ν_{0i}, which represents the random intercept of class i. If ν_{0i} is positive (negative), the average popularity in class i is larger (smaller) than the average popularity level in the population. The random intercept is assumed to be normally distributed with mean zero and variance $\sigma^2_{\nu_0}$. The random variable ϵ_{ji} represents the level 1 or within-class errors of pupil j in class i. In this model, the level 1 errors ϵ_{ij} are independent and identically distributed with mean zero and variance σ^2_{ϵ}.

To ease the subsequent derivations, we represent the model that Jane is going to fit using matrix notation as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \nu_{0i} + \boldsymbol{\epsilon}_i \tag{2}$$

[1] In this paper, the terms 'effect' and 'association' are not supposed to imply that regression coefficients quantify causal effects.

where $\mathbf{y}_i$ contains the $n_i = 20$ observations of pupil popularity in class i. The 20×2 design matrix for the fixed part $\mathbf{X}_i$ is equal to

$$\mathbf{X}_i = \begin{bmatrix} 1 & w_i \\ \vdots & \vdots \\ 1 & w_i \end{bmatrix}$$

and the (20×1) design matrix for the random part $\mathbf{Z}_i$ is equal to

$$\mathbf{Z}_i = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}.$$

The vector of fixed regression coefficients is denoted by $\boldsymbol{\beta} = (\beta_{00}, \beta_{01})^T$. Finally, the 20×1 matrix $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{i20})^T$ includes the $n_i = 20$ level 1 errors of class i. Since Jane's model states that the level 1 errors are independent and identically distributed, their covariance matrix $\boldsymbol{\Sigma}_\epsilon$ equals $\sigma^2_\epsilon \mathbf{I}_{20}$, with $\mathbf{I}_{20}$ being the identity matrix of size 20. In this model, the covariance matrix of $\mathbf{y}_i$ conditioned on $\mathbf{X}_i$ and $\mathbf{Z}_i$ is defined as $\mathbf{V}_i$ and equal to[2]

$$\mathbf{V}_i = \sigma^2_{\nu_0} \mathbf{Z}_i \mathbf{Z}_i^T + \boldsymbol{\Sigma}_\epsilon.$$

Testing the statistical hypothesis of interest

To test the hypothesis that the average popularity is larger in classes with more experienced teachers, Jane is going to test $H_0 : \beta_{01} \leq 0$ vs. $H_1 : \beta_{01} > 0$. The test statistic equals the ratio of the fixed effect estimate $\hat{\beta}_{01}$ and an estimate of the standard error $\widehat{\text{SE}}(\hat{\beta}_{01})$:

$$T = \frac{\hat{\beta}_{01}}{\widehat{\text{SE}}(\hat{\beta}_{01})} \tag{3}$$

Assuming no severe violations of the modeling assumptions, this t-ratio approximately follows a t-distribution with $N - q - 1$ degrees of freedom, where q denotes the number of level 2 predictors in the model (Raudenbush and Bryk, 2002). Since teacher experience is the only level 2 predictor in Jane's study, we have that $q = 1$.

To compute the t-ratio, Jane needs to estimate the four unknown model parameters in Eq. 1, that is the two fixed effects β_{00} and β_{01}, the level 2 variance $\sigma^2_{\nu_0}$, and the level 1 variance σ^2_ϵ. Arguably the best way of doing this is to estimate the two unknown variances via restricted maximum likelihood (REML) and subsequently β_{00} and β_{01} via maximum likelihood (ML) with the likelihood function being Gaussian. On the OSF repository of this paper, we show how to estimate and extract these parameter estimates using nlme version 3.1-155 (Pinheiro et al., 2019) and R version 4.1.2 (R Core Team, 2013). Table 1 shows the estimated parameter values when the level 2 predictor is grand-mean centered.

For this paper, it is important to show how the numerator and denominator of the t-ratio in Eq. 3 are computed via the REML estimates of the population variances. Using the REML estimates $\hat{\sigma}^2_{\nu_0}$ and $\hat{\sigma}^2_\epsilon$ we estimate the covariance matrix of $\mathbf{y}_i$ conditional on the design matrices for the fixed and random parts as:

$$\widehat{\mathbf{V}}_i = \hat{\sigma}^2_{\nu_0} \mathbf{Z}_i \mathbf{Z}_i^T + \hat{\sigma}^2_\epsilon I_{20}. \tag{4}$$

The fixed regression coefficients are then estimated via ML as:

$$\hat{\boldsymbol{\beta}} | \mathbf{X}_i, \mathbf{Z}_i = \left(\sum_{i=1}^{N} \mathbf{X}_i^T \widehat{\mathbf{V}}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^{N} \mathbf{X}_i^T \widehat{\mathbf{V}}_i^{-1} \mathbf{y}_i. \tag{5}$$

Subsequently, we compute the standard error in the denominator of the t-ratio in Eq. 3. To do this, we first estimate the (2×2) covariance matrix of the ML estimate $\hat{\boldsymbol{\beta}}$ as

$$\widehat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\beta}}} = \left(\sum_{i=1}^{N} \mathbf{X}_i^T \widehat{\mathbf{V}}_i^{-1} \mathbf{X}_i \right)^{-1}. \tag{6}$$

The estimate of the standard error $\widehat{\text{SE}}(\hat{\beta}_{01})$ then is the square root of the second diagonal element of Eq. 6. We also note that the estimated covariance matrix of the maximum likelihood estimate $\hat{\boldsymbol{\beta}}$ is the inverse of the estimated precision

[2] To derive $\mathbf{V}_i$, we first derive the conditional expectation of $\mathbf{y}_i$:

$$\begin{aligned} \mathbb{E}(\mathbf{y}_i | \mathbf{X}_i, \mathbf{Z}_i) &= \mathbb{E}(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \nu_{0i} + \boldsymbol{\epsilon}_i | \mathbf{X}_i, \mathbf{Z}_i) \\ &= \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbb{E}(\nu_{0i} | \mathbf{X}_i, \mathbf{Z}_i) + \mathbb{E}(\boldsymbol{\epsilon}_i | \mathbf{X}_i, \mathbf{Z}_i) \\ &= \mathbf{X}_i \boldsymbol{\beta} \end{aligned}$$

Using the above derivation we find that

$$\begin{aligned} \mathbf{V}_i \overset{def}{=} \boldsymbol{\Sigma}_{\mathbf{y}_i | \mathbf{X}_i, \mathbf{Z}_i} &= \mathbb{E}\left((\mathbf{y}_i - \mathbb{E}(\mathbf{y}_i | \mathbf{X}_i, \mathbf{Z}_i)(\mathbf{y}_i - \mathbb{E}(\mathbf{y}_i | \mathbf{X}_i, \mathbf{Z}_i))^T | \mathbf{X}_i, \mathbf{Z}_i \right) \\ &= \mathbb{E}\left((\mathbf{Z}_i \nu_{0i} + \boldsymbol{\epsilon}_i)(\mathbf{Z}_i \nu_{0i} + \boldsymbol{\epsilon}_i)^T | \mathbf{X}_i, \mathbf{Z}_i \right) \\ &= \mathbb{E}\left(\nu_{0i}^2 \mathbf{Z}_i \mathbf{Z}_i^T + \mathbf{Z}_i \nu_{0i} \boldsymbol{\epsilon}_i^T + \nu_{0i} \boldsymbol{\epsilon}_i \mathbf{Z}_i^T + \boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^T | \mathbf{X}_i, \mathbf{Z}_i \right). \end{aligned}$$

Since the level 1 errors and the random effects are independent, we can further simplify $\mathbf{V}_i$ to

$$\begin{aligned} \mathbf{V}_i = \boldsymbol{\Sigma}_{\mathbf{y}_i | \mathbf{X}_i, \mathbf{Z}_i} &= \mathbb{E}\left(\nu_{0i}^2 | \mathbf{X}_i, \mathbf{Z}_i \right) \mathbf{Z}_i \mathbf{Z}_i^T + \mathbb{E}\left(\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i^T | \mathbf{X}_i, \mathbf{Z}_i \right) \\ &= \sigma^2_{\nu_0} \mathbf{Z}_i \mathbf{Z}_i^T + \boldsymbol{\Sigma}_\epsilon. \end{aligned}$$

If we further assume that $\boldsymbol{\epsilon}_i$ follows a multivariate normal distribution and that ν_{0i} follows a normal distribution, $\mathbf{y}_i$ follows a multivariate normal distribution with mean $\mathbf{X}_i \boldsymbol{\beta}$ and covariance matrix $\sigma^2_{\nu_0} \mathbf{Z}_i \mathbf{Z}_i^T + \Sigma_\epsilon$.

Table 1 Sample characteristics and estimated model parameters of the simple two-level regression model when using the popularity data set

	Statistic	Value
Number of schools	N	100
Number of pupils per class	n_i	20
Sample mean of teacher experience in years	$\hat{\mu}_w$	14.30
Standard deviation of teacher experience in years	$\hat{\sigma}_w$	6.61
Fixed intercept estimate	$\hat{\beta}_{00}$	4.20
Estimated effect of the level 2 variable on the level 1 intercept	$\hat{\beta}_{01}$	0.06
Estimate of the standard deviation of the level 1 error	$\hat{\sigma}_\epsilon$	1.11
Estimate of the standard deviation of the random intercept	$\hat{\sigma}_{v_0}$	0.74

matrix of $\hat{\boldsymbol{\beta}}$:

$$\widehat{\mathbf{P}}_{\hat{\boldsymbol{\beta}}} = \left(\sum_{i=1}^{N} \mathbf{X}_i^T \widehat{\mathbf{V}}_i^{-1} \mathbf{X}_i \right) \tag{7}$$

This matrix thus tells us something about the estimation accuracy of the ML estimates of the fixed effects which is why its derivation is pivotal for the analytical approach of computing statistical power (see e.g., Ly et al. 2017).

Finally, Jane can compute the t-ratio in Eq. 3 by dividing $\hat{\beta}_{01}$ by $\widehat{\text{SE}}(\hat{\beta}_{01})$ and reject H_0 if it is larger than the $1-\alpha$ quantile of the t-distribution with $N-2$ degrees of freedom[3]. In the case of the original study (Hox et al., 2017), $\hat{\beta}_{01}$ is 0.06, $\widehat{\text{SE}}(\hat{\beta}_{01})$ is 0.01, and the t-ratio is equal to 5.21 which is larger than the 95% quantile of the t-distribution with 98 degrees of freedom (i.e., 1.66). Therefore, the effect of a teacher's experience on the average popularity level in a class was statistically significant at α-level 0.05 in the original study.

Conducting an a priori power analysis

Remember that Jane is interested in replicating the effect of teachers' experience on the average popularity level in a class in a new study. Now that we know how Jane intends to test the statistical hypothesis $H_0 : \beta_{01} \leq 0$ vs. $H_1 : \beta_{01} > 0$ in model (2), we can help her compute the statistical power for this hypothesis, given the significance level α of her choice. To this end, she has to specify how large she expects the true effect β_{01}^* to be. To select the true effect, Jane considers the findings of previous research. She is not aware of any other data besides the popularity data set and she has no strong prior beliefs about realistic effect sizes. Because of this, she decides to simply use the point estimate from the popularity data, which is displayed in Table 1, to compute statistical power[4].

Jane will first compute statistical power using the *analytical approach*. The statistical power of Jane's one-sided hypothesis test is defined as

$$\mathcal{P}(\alpha, \beta^*) := \Pr\left(\frac{\hat{\beta}_{01}}{\widehat{\text{SE}}(\hat{\beta}_{01})} \geq \mathcal{T}_r^{-1}(1-\alpha) \middle| \alpha, \beta_{01} = \beta_{01}^* \right) \tag{8}$$

with $\mathcal{T}_r^{-1}(\cdot)$ denoting the quantile function of a Student's t-distribution with $r = N-2$ degrees of freedom. Assuming that the number of pupils is equal across all classes (i.e., $n_i = n$) we make use of an asymptotic result (see Snijders and Bosker, 1993)[5]:

$$\mathcal{P}(\alpha, \beta^*) \approx 1 - \mathcal{T}_r\left(\mathcal{T}_r^{-1}(1-\alpha) - \frac{\beta_{10}}{\text{SE}(\hat{\beta}_{10})} \right) \tag{9}$$

where $\text{SE}(\hat{\beta}_{10})$ is the standard error of $\hat{\beta}_{01}$. To use Eq. 9, Jane needs to know $\text{SE}(\hat{\beta}_{01})$. Since the covariance matrix of $\hat{\boldsymbol{\beta}}$ is the inverse of the precision matrix of the vector of fixed effects $\mathbf{P}_{\hat{\boldsymbol{\beta}}}$, she could use algebra to calculate $\text{SE}(\hat{\beta}_{01})$ via $\mathbf{P}_{\hat{\boldsymbol{\beta}}}$. However, there are two obstacles. Firstly, $\mathbf{P}_{\hat{\boldsymbol{\beta}}}$ depends on two model parameters: σ_ϵ^2, $\sigma_{v_0}^2$ (see Appendix A). Secondly, this matrix depends on two statistics namely the sample mean and the sample variance of the level 2 predictor. So, what is Jane to do?

To tackle the first problem, Jane can do what is typically done when conducting power analysis which is replacing

[3] At this point, it seems important to stress that this hypothesis test does not always respect the nominal Type-I risk since it is based on asymptotic approximations. Generally speaking, the larger the number of pupils per class and the larger the number of classes, the smaller the approximation error (McNeish, 2017).

[4] Although standard practice, this approach to power analysis is problematic. The reason is the so-called 'statistical significance filter' - statistically significant estimates in the literature are positively biased. Hence, even non-Bayesians should try to gauge what realistic and non-realistic effects look like. Unfortunately, this is very hard in multilevel models because variables cannot just be standardized to yield scale-free and easily interpretable coefficients (Rights and Sterba, 2019)

[5] The statistical power for the two-side hypothesis test is estimated as $1 - \mathcal{T}_r\left(\mathcal{T}_r^{-1}(1-\alpha/2) - \frac{\beta_{10}}{\text{SE}(\hat{\beta}_{10})}\right) + \mathcal{T}_r\left(\mathcal{T}_r^{-1}(\alpha/2) - \frac{\beta_{10}}{\text{SE}(\hat{\beta}_{10})}\right)$.

unknown parameters with educated guesses. To tackle the latter, Jane can respectively replace the sample mean with the population mean and the sample variance with the population variance, knowing that the sample moments converge almost surely to population moments. Therefore, Jane needs to provide values for four unknown quantities: σ_ϵ^2, $\sigma_{v_0}^2$, μ_w, and σ_w^2. Assuming for simplicity that $n_i = n \ \forall \ i$, Jane will arrive at the following matrix (see Snijders and Bosker, 1993, for a proof):

$$\mathbf{P}^*_{\hat{\boldsymbol{\beta}}} := \frac{n}{\sigma_\epsilon^2 + \sigma_{v_0}^2 n} \begin{bmatrix} N & N\mu_w \\ N\mu_w & (N-1)\sigma_w^2 + N\mu_w^2. \end{bmatrix} \tag{10}$$

By determining the sample sizes of the replication study as $N = 150$ and $n = 20$ and by replacing the unknown parameters in (10) with their point estimates from the popularity data set (see Table 1), $\mathbf{P}^*_{\hat{\boldsymbol{\beta}}}$ becomes a known, real-valued (2×2) matrix. Jane can then finally compute $\text{SE}(\hat{\beta}_{01}) = 0.01$ by taking the square root of the second diagonal element of the inverse of $\mathbf{P}^*_{\hat{\boldsymbol{\beta}}}$. Plugging it into Eq. 9 then yields a statistical power of 1.00 for $\alpha = 0.05$.

Unsure of whether this power estimate equal to one is plausible, Jane goes on to also compute statistical power using the *simulation-based approach*. She sets the number of classes to $N = 150$ and the number of pupils per class to $n = 20$. Just like in the analytic approach, Jane needs to assume values for the four unknown quantities σ_ϵ^2, $\sigma_{v_0}^2$, μ_w, and σ_w^2. As mentioned before, she assumes that those values are equal to the point estimates of the popularity data set (see Table 1). Having done so, she generates 1000 random data sets via Monte Carlo simulation. Afterward, model (1) is fitted to each generated dataset using REML. The resulting estimate of the fixed effect $\hat{\beta}_{01}$ and its standard error estimate $\widehat{\text{SE}}(\hat{\beta}_{01})$ are then used to compute the t-ratio and the p-value (under H_0, the t-ratio follows a t-distribution with $N-2 = 148$ degrees of freedom). Finally, the number of generated datasets for which the null hypothesis is rejected is divided by 1000 to yield an estimate of the statistical power. The simulation-based approach yields an estimate of statistical power equal to 1.00. The average standard error of $\hat{\beta}_{01}$ across the 1000 Monte Carlo replicates is 0.01. The total computational time in seconds of the simulation-based approach is 28.79[6].

Note that in Jane's example, the simulation and the analytical approach yield the same power estimate (when rounded to the third decimal). However, power estimates will not always be that similar. The analytic approach assumes values for the unknown model parameters and builds on asymptotic assumptions. In contrast, the simulation-based approach also assumes values for the unknown model parameters but instead of relying on asymptotic theory, it estimates power based on the estimated model parameters for the generated data sets whose sampling distributions depend on the sample size. Given that enough data samples are used in the simulation-based approach, it will therefore yield a more accurate power estimate than the analytical approach. Whereas the simulation power estimate is unbiased, the analytical estimate is not. Ceteris paribus, the smaller N or n, the larger the bias of the analytical estimate. Why would Jane ever want to use the analytical approach then? Although the analytical approach is biased, its bias is negligible when N and n are sufficiently large. Up to a certain number of sampled datasets in the simulation-based approach, the analytical estimate is therefore more accurate[7]. As will be shown, the number of simulated datasets required for the simulation estimate to be more accurate than the analytical one can be quite large which is why the computational burden of the simulation-based approach can be non-negligible. In the next section, we investigate the differences between the two approaches for various sample sizes.

Precision matrix for two-level models with autocorrelated errors

In the previous section, we showed that the standard error of the fixed effect estimates is a crucial element in the analytical computation of statistical power. We also demonstrated that to compute the standard error estimate of the fixed effect, we first need to estimate the precision matrix. Therefore, in this section, we showed that given certain assumptions, we can derive an analytical expression of the precision matrix for a set of models that are popular in IL designs. For simplicity of description, these algebraic derivations of the precision matrix are included in Appendix A.

Concretely, we focus on five two-level models with autocorrelated errors, which are widely used to study individual differences in IL studies and properly account for the temporal dependency (see e.g., Lafit et al. 2021). The first model includes a level 2 or time-invariant continuous predictor that

[6] All analyses were carried out on an AMD Ryzen 7 PRO 5850U with Radeon Graphics 1.90 GHz processor.

[7] When measuring accuracy via the mean-squared error of estimation (MSE) we prefer the analytical approach to the simulation-based approach if

$$\left|\mu_{\widehat{\mathcal{P}}_{(ana)}} - \mathcal{P}\right| < \sigma_{\widehat{\mathcal{P}}_{(sim)}}$$

since $\mu_{\widehat{\mathcal{P}}_{(sim)}} = \mathcal{P}$ and $\sigma_{\widehat{\mathcal{P}}_{(ana)}} = 0$. That is, we prefer the analytical approach (in the MSE sense) if the absolute value of the bias of this power estimator is smaller than the standard deviation of the simulation power estimator. Note thereby that $\sigma_{\widehat{\mathcal{P}}_{(sim)}}$ is a function of the number of Monte Carlo samples.

allows investigating changes in the mean level of an outcome due to changes in the person-level predictor. As such, this model only differs from the simple two-level model that Jane used in that the within-person errors follow an AR(1) process. The second model includes a level 2 binary predictor to estimate differences in the mean level of the outcome between two groups of persons. The third model estimates the fixed within-person association between a level 1 or time-varying predictor and the outcome. The fourth model extends the third model by including a level 2 continuous predictor to investigate cross-level interaction effects between the time-varying predictor and the time-invariant predictor. The last model extends the third model by including a binary level 2 predictor to estimate group differences in the within-persons associations between the time-varying predictor and the outcome. A summary of the models is included in Table 2.

To illustrate how we can use the analytical derivations to compute statistical power we use the clinical data set reported on by Heininga et al. (2019). This data set includes 38 individuals diagnosed with major depressive disorder (MDD) and 40 control subjects. Participants took part in an ESM study of 7 days. The design included 70 measurement occasions per participant. At each measurement occasion, participants rated 27 items measuring various constructs, including negative affect (NA) and anhedonia, on a visual analog scale ranging from "not at all" on the left to "very much" on the right, resulting in a score between 0 to 100. Depressive symptom severity was measured before the ESM testing period by summing the scores on the items of the Quick Inventory of Depressive Symptomatology (i.e., QIDS; Rush et al. 2003). The average compliance was 0.88 (SD=0.32). When estimating the models we handled missing data using listwise deletion. The data are available upon request on the EMOTE database (Kalokerinos et al., 2022): https://emotedatabase.com/datasets/3. To validate the proposed approach, we compare its performance with the simulation-based approach, as implemented by Lafit et al. (2021).

Effect of a level 2 continuous predictor on the mean level

To begin with, we focus on a two-level model that allows us to investigate the effect of depression (Dep), which is a grand-mean centered level 2 or time-invariant continuous predictor on the person-specific mean levels of NA in persons diagnosed with MDD. The regression equation is:

$$\mathrm{NA}_{it} = \beta_{00} + \beta_{01}\mathrm{Dep}_i + \nu_{0i} + \epsilon_{it} \tag{11}$$

where $i = (1, \ldots, N)$ indicates the person and $t = (1, \ldots, T_i)$ the measurement occasion. Note that T_i denotes the total number of measurement occasions for the i-th person. The coefficient β_{00} represents the mean NA level of a person whose depression level equals the average level in the population. β_{01} represents the influence of depression level: if β_{01} is positive, the average NA is larger for persons with higher depression levels. Further inter-individual differences in mean NA are modeled by the random effect ν_{0i}, which is normally distributed with mean zero and standard deviation σ_{ν_0}.

In this model, we account for the serial dependency that characterizes IL designs by assuming that the level 1 errors follow an autoregressive (AR(1)) process (Goldstein et al., 1994):

$$\epsilon_{it} = \rho_\epsilon \epsilon_{it-1} + \varepsilon_{it} \tag{12}$$

where the correlation between two consecutive errors is denoted by ρ_ϵ, and ε_{it} is a Gaussian error with mean zero and variance σ^2_ε. Under this model, the variance of the level 1 errors is given by $\sigma^2_\epsilon = \sigma^2_\varepsilon/(1 - \rho^2_\epsilon)$. To guarantee that this model is stationary, the absolute value of ρ_ϵ needs to be smaller than 1 (Hamilton, 1994).

Table 2 Overview of the models of interest

Model	Description
1	Effect of a continuous time-invariant predictor on the mean level: estimates whether a Level 2 variable predicts differences across persons in the mean level of the outcome.
2	Group differences in mean level: estimate differences between two groups of persons in the mean of the outcome.
3	Effect of a level 1 continuous predictor: estimates whether a level 1 variable predicts the outcome.
4	Cross-level interaction between two continuous predictors: estimates whether the effect of a level 1 predictor on the outcome is moderated by a continuous level 2 predictor.
5	Group differences in the effect of a level 1 continuous predictor: estimates differences between two groups of persons concerning the effect of a level 1 predictor on the outcome.

Asymptotic approximation of the precision matrix

To derive the precision matrix of the fixed effect estimates of model (11), we once again use general matrix notation to represent the model for person i as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\nu_{0i} + \boldsymbol{\epsilon}_i \tag{13}$$

with the outcome vector $\mathbf{y}_i$ containing the T_i NA observations of person i, the $(T_i \times 2)$ design matrix for the fixed part being equal to

$$\mathbf{X}_i = \begin{bmatrix} 1 & \text{Dep}_i \\ \vdots & \vdots \\ 1 & \text{Dep}_i \end{bmatrix},$$

the $(T_i \times 1)$ design matrix for the random part being equal to

$$\mathbf{Z}_i = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix},$$

the vector of fixed regression coefficients being $\boldsymbol{\beta} = (\beta_{00}, \beta_{01})^T$, and the level 1 error vector of person i being $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{iT_i})^T$. The covariance matrix of the level 1 error vector $\boldsymbol{\epsilon}_i$ is

$$\boldsymbol{\Sigma}_\epsilon = \sigma_\epsilon^2 \begin{bmatrix} 1 & \rho_\epsilon & \rho_\epsilon^2 & \ldots & \rho_\epsilon^{T_i-1} \\ \rho_\epsilon & 1 & \rho_\epsilon & \ldots & \rho_\epsilon^{T_i-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_\epsilon^{T_i-1} & \rho_\epsilon^{T_i-2} & \rho_\epsilon^{T_i-3} & \ldots & 1 \end{bmatrix}. \tag{14}$$

Conditioning on the design matrices $\mathbf{X}_i$ and $\mathbf{Z}_i$, the conditional covariance matrix of the outcome vector $\mathbf{y}_i$ for person i is

$$\mathbf{V}_i = \sigma_{\nu_0}^2 \mathbf{Z}_i\mathbf{Z}_i^T + \boldsymbol{\Sigma}_\epsilon. \tag{15}$$

Assuming that the number of persons N is sufficiently large, and the number of measurement occasions is large and equal across all persons (i.e., $T_i = T$), it follows from Eqs. 14 and 15 that the asymptotic approximation of the precision matrix for this model is equal to

$$\mathbf{P}^*_{\hat{\boldsymbol{\beta}}} = \frac{(T-2)(1-\rho_\epsilon)^2 + 2(1-\rho_\epsilon)}{\sigma_\epsilon^2(1-\rho_\epsilon^2) + \sigma_{\nu_0}^2((T-2)(1-\rho_\epsilon)^2 + 2(1-\rho_\epsilon))} \begin{bmatrix} N & N\mu_{\text{Dep}} \\ N\mu_{\text{Dep}} & (N-1)\sigma_{\text{Dep}}^2 + N\mu_{\text{Dep}}^2 \end{bmatrix} \tag{16}$$

where μ_{Dep} and σ_{Dep}^2 denote the mean and variance of the level 2 variable depression (see Appendix A.2 for a step-by-step derivation). We note that $\mathbf{P}^*_{\hat{\boldsymbol{\beta}}}$ for model (1) is a special case of Eq. 16 where ρ_ϵ is equal to zero.

Illustration

To compare the results of the analytical and simulation-based approaches, we used the data set from the clinical study to obtain plausible values of the parameters of the model in Eq. 11:

$$\text{NA}_{it} = 43.01 + 1.5 \times \text{Dep}_i + \nu_{0i} + \epsilon_{it}$$

with $\sigma_{\nu_0} = 12.90$, $\sigma_\epsilon = 12.62$, $\rho_\epsilon = 0.46$, $\mu_{\text{Dep}} = 15.70$, and $\sigma_{\text{Dep}} = 5.00$.

Next, we compute the power to test the hypothesis that persons who reported higher levels of depression have on average larger mean NA levels (i.e., $H_1 : \beta_{01} > 0$). To test this hypothesis, we set the value of α to 5%, the number of persons to 20, 30, 60, 80, 100, 200, and 300, and the number of (equal) measurement occasions per person to 20, 70, and 140. In addition, to investigate how much statistical power varied as a function of the value of the main effect of interest, we conducted a sensitivity analysis varying the value of β^*_{01}. First, we conducted the analysis using the estimated value from the clinical data set (i.e., $\beta^*_{01} = 1.5$). Since statistical power is positively associated with the value of β^*_{01}, we investigated the impact of smaller effect sizes by also running the analysis with $\beta^*_{01} = 0.15$ which corresponds to a 90% decrease in effect size. In the simulation-based approach, we follow the approach described in Lafit et al. (2021), and we generate 1000 replicates for each combination of the number of persons and the number of measurement occasions. On the GitLab repository of the project, we include the simulation scripts to run the analyses reported in this article using R (see https://gitlab.kuleuven.be/ppw-okpiv/researchers/u0119584/ApproxPowerIL).

Figure 1 displays the obtained estimates of statistical power for the main effect of interest (i.e., the effect of the level 2 predictor depression on the mean NA level) for each combination of N, T, and β^*_{01}. Figure 2 displays the standard error of $\hat{\beta}_{01}$ for each sample size and value of β^*_{01} [8]. We observe that the two approaches produce similar results for statistical power and standard error estimates across the

[8] In these illustrations we focus on the main effect of interest for each of the five models, however, additional results such as the statistical power of single tests for each of the fixed effects, the standard error of the fixed effects estimates, the estimation performance of the fixed and random effects, and the computational time in seconds for the simulation-based approach as a function of N are included in the supplementary material.

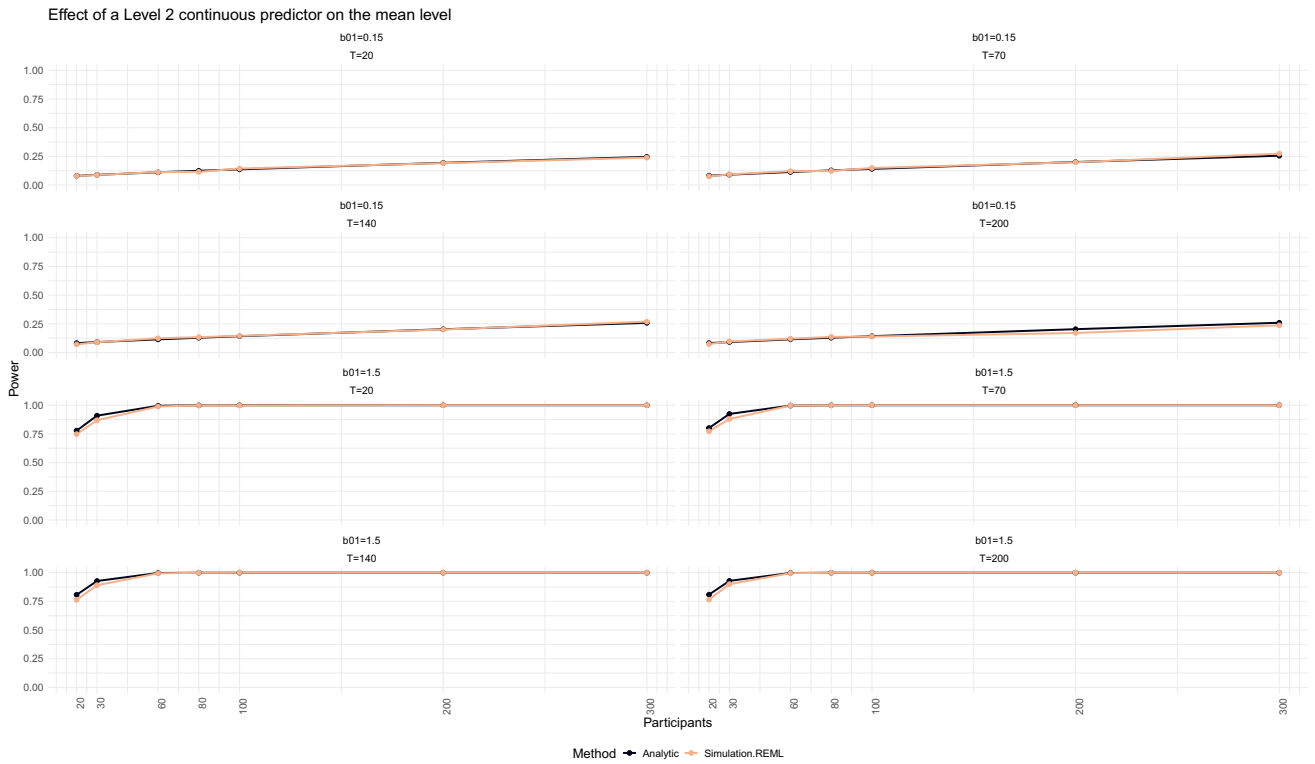


Fig. 1 Illustration 1. Effect of depression on the person-specific mean level of negative affect. The figures show the statistical power for estimating the effect of interest (β_{01}) of the analytic and simulation-based approaches

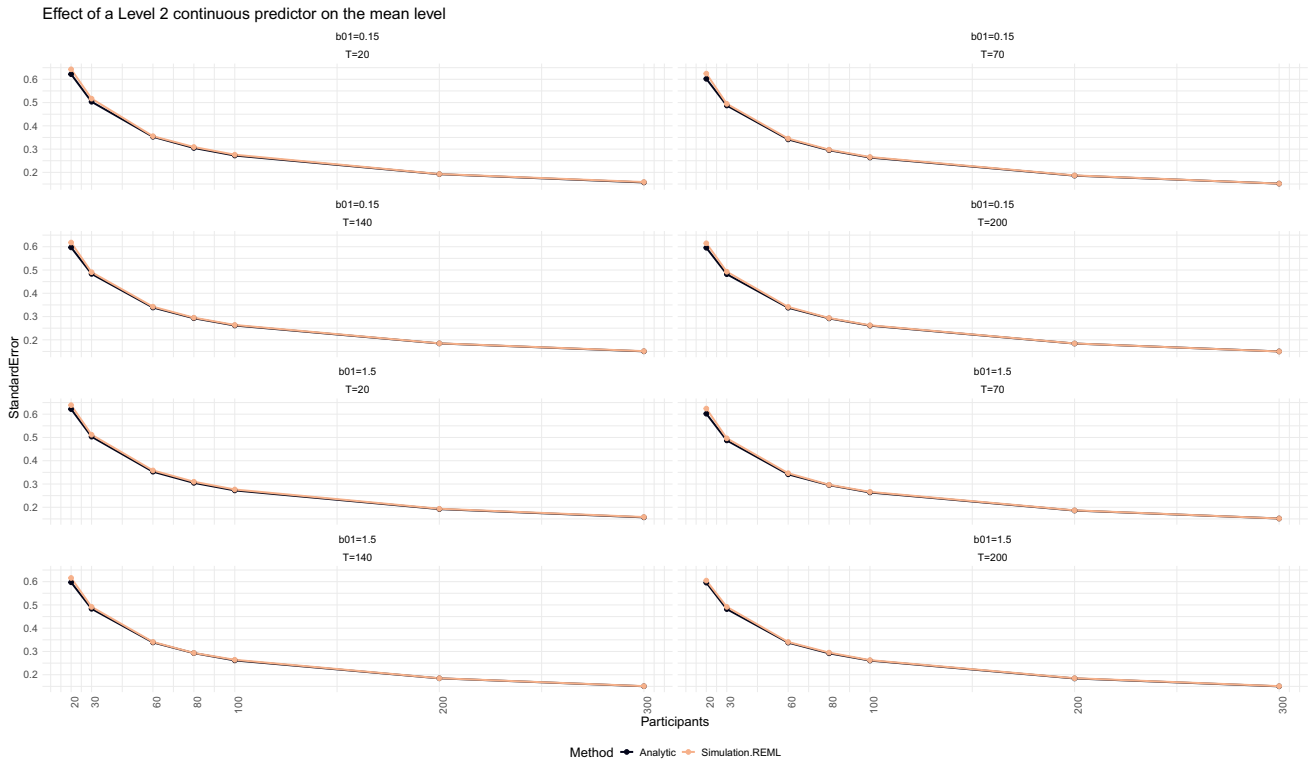


Fig. 2 Illustration 1. Effect of depression on the person-specific mean level of negative affect. The figures show the standard error of the fixed effect estimate $\hat{\beta}_{01}$ of the analytic and simulation-based approaches

different values of T and β_{01}^*, with the analytic approach producing slightly larger values of statistical power when $\beta_{01}^* = 1.5$ and N is below 60.

Effect of a level 2 binary predictor on the mean level

Next, we obtain $\mathbf{P}^*_{\hat{\boldsymbol{\beta}}}$ for a model that allows estimating group differences in the mean NA level of persons diagnosed with MDD and healthy controls:

$$\text{NA}_{it} = \beta_{00} + \beta_{01}\text{Diag}_i + \nu_{0i} + \epsilon_{it} \tag{17}$$

This model includes a binary level 2 predictor (Diag_i) that is equal to one if person i was diagnosed with MDD and zero otherwise. The coefficient β_{00} thus represents the mean NA level of persons included in the control group, and $\beta_{00} + \beta_{01}$ thus represents the mean NA level of persons diagnosed with MDD. Our main effect of interest in this model shall be β_{01}, representing the average difference in NA levels between persons diagnosed with MDD and people not diagnosed with MDD. Within both groups, inter-individual differences in NA levels are modeled by the random effect ν_{0i}, which is normally distributed with mean zero and standard deviation σ_{ν_0}. To capture the temporal dependency in NA, the level 1 errors are assumed to follow an AR(1) process as in Eq. 12.

Asymptotic approximation of the precision matrix

The asymptotic approximation of the precision matrix of the fixed effects β_{00} and β_{01} is given by

$$\mathbf{P}^*_{\hat{\boldsymbol{\beta}}} = \frac{(T-2)(1-\rho_\epsilon)^2 + 2(1-\rho_\epsilon)}{\sigma_\epsilon^2(1-\rho_\epsilon^2) + \sigma_{\nu_0}^2((T-2)(1-\rho_\epsilon)^2 + 2(1-\rho_\epsilon))} \begin{bmatrix} N & N_1 \\ N_1 & (N-1)\frac{N_1 N_0}{N^2} + \frac{N_1^2}{N} \end{bmatrix} \tag{18}$$

where N_0 indicates the number of persons in the control group, N_1 is the number of persons in the other group, and $N = N_0 + N_1$ (see Appendix A for a step-by-step derivation).

Illustration

Once again, we compare the results of the two approaches for computing statistical power to test our hypothesis of interest (i.e., $H_1 : \beta_{01} \neq 0$). Based on the data from the clinical study, the parameters in model (17) are set to: $\beta_{00} = 10.40$, $\beta_{01} \in \{3.261, 32.61\}$, $\sigma_{\nu_0} = 11.15$, $\sigma_\epsilon = 10.16$, $\rho_\epsilon = 0.42$, and $\alpha = 0.05$. Furthermore, $N \in \{20, 30, 60, 80, 100, 200, 300\}$ and $T \in \{20, 70, 140\}$.

Figures 3 and 4 respectively show the estimated power and the estimated standard error of $\hat{\beta}_{01}$ as a function of the number of participants in each group. The two procedures produce comparable results.

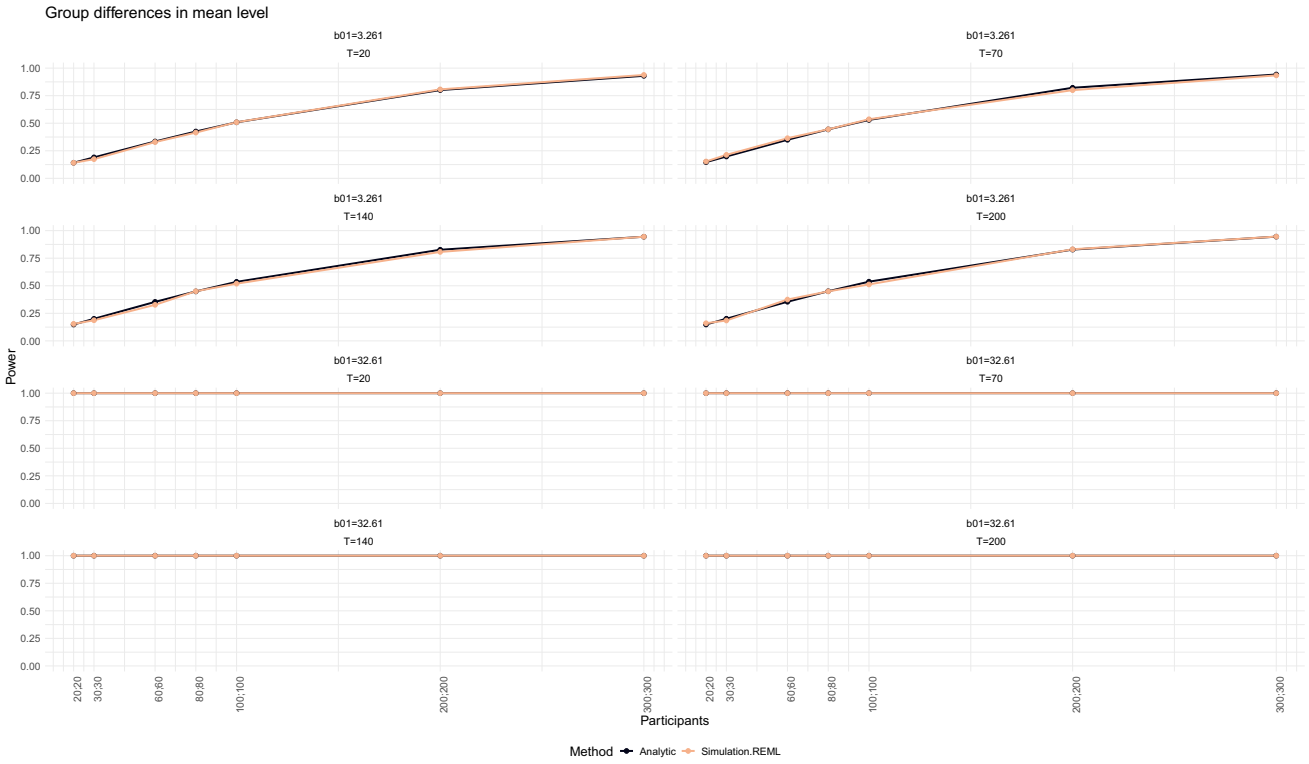


Fig. 3 Illustration 2. Differences between groups in the mean of negative affect. The figures show the statistical power for estimating the effect of interest (β_{01}) of the analytic and simulation-based approaches

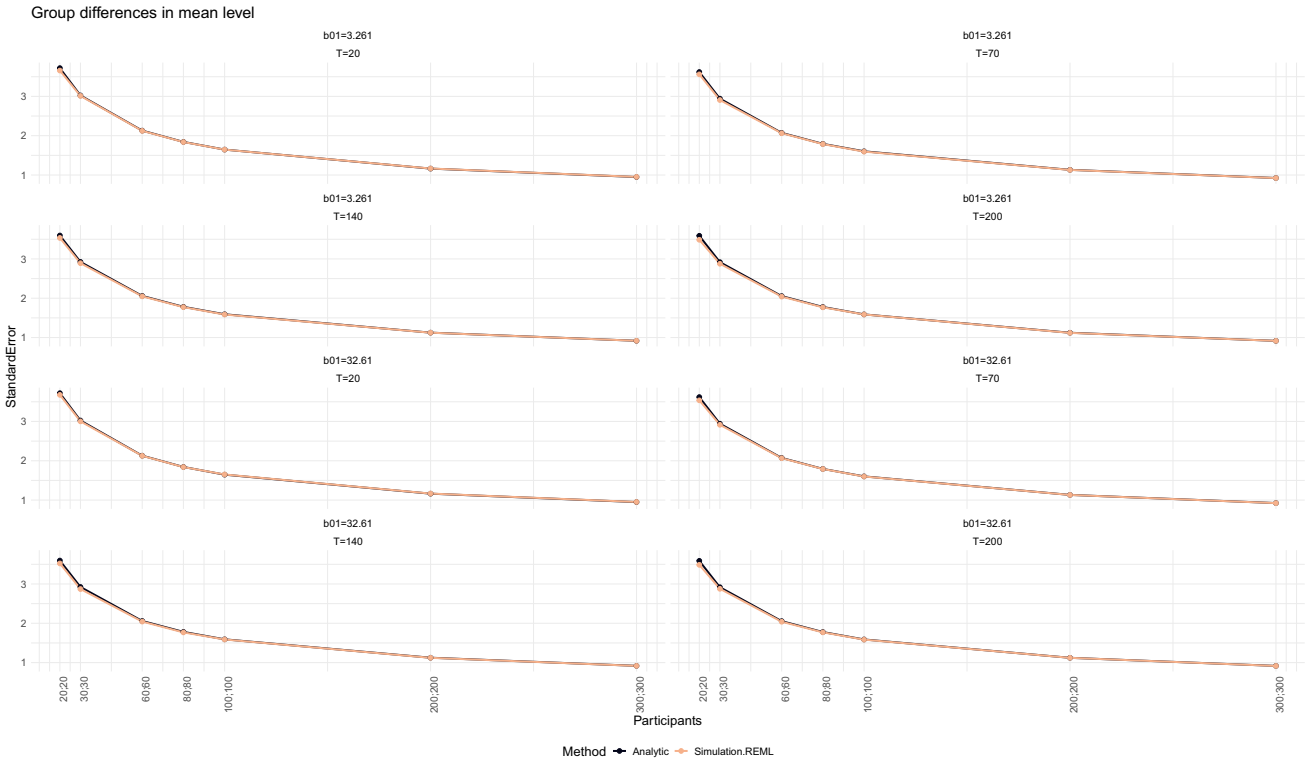


Fig. 4 Illustration 2. Differences between groups in the mean of negative affect. The figures show the standard error of the fixed effect estimate $\hat{\beta}_{01}$ of the analytic and simulation-based approaches

Effect of a level 1 continuous predictor

We now consider a two-level model for persons diagnosed with MDD in which the NA scores are related to a continuous, person-mean-centered and time-varying predictor, namely anhedonia ($\tilde{\mathrm{A}}_{it}$), while imposing an AR(1) process on the level 1 errors. The regression equation is

$$\mathrm{NA}_{it} = \beta_{00} + \beta_{10}\tilde{\mathrm{A}}_{it} + \upsilon_{0i} + \upsilon_{1i}\tilde{\mathrm{A}}_{it} + \epsilon_{it} \tag{19}$$

where β_{00} is the fixed intercept and β_{10} is the fixed slope representing the overall (i.e., across all persons) effect of anhedonia on NA. Between-person differences in the relation between Anhedonia and NA are captured by including a random intercept υ_{0i} and random slope υ_{1i}. These random effects are assumed to follow a bivariate normal distribution with mean vector $(0, 0)^T$ and covariance matrix $\boldsymbol{\Sigma}_{\upsilon}$:

$$\boldsymbol{\Sigma}_{\upsilon} = \begin{bmatrix} \sigma^2_{\upsilon_0} & \sigma_{\upsilon_{01}} \\ \sigma_{\upsilon_{01}} & \sigma^2_{\upsilon_1} \end{bmatrix}$$

For our purposes, we also have to specify a model for the between-person differences in the level 1 predictor Anhedonia:

$$\mathrm{A}_{it} = \mu_{\mathrm{A}} + \upsilon_i + \tau_{it} \tag{20}$$

In model (20), μ_{A} represents the mean of Anhedonia across all persons and all time points, υ_i represents the deviation of each person's mean anhedonia level from μ_{A}, and τ_{it} denotes the within-person error of A_{it}, which is assumed to follow an AR(1) process with autocorrelation ρ_τ and variance σ^2_τ.

In this model, the population mean of anhedonia for the i-th person is $\mu_{\mathrm{A}} + \upsilon_i$. Under the assumption that T is sufficiently large, the sample mean of person i converges in probability to the population mean (see Hamilton, 1994, Chapter 3, p.46). Thus, when T is sufficiently large, the fixed slope in Eq. 20 approximately represents the average of the within-person slopes.

Asymptotic approximation of the precision matrix

To derive the precision matrix of the fixed effect estimates of model (19), we again represent the model for person i as:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{\nu}_i + \boldsymbol{\epsilon}_i \tag{21}$$

where $\mathbf{y}_i$ contains the T_i observations of the outcome variable for person i. $\mathbf{X}_i$ is the $(T_i \times 2)$ design matrix for the fixed

part given by

$$\mathbf{X}_i = \begin{bmatrix} 1 & \tilde{\text{A}}_{i1} \\ \vdots & \vdots \\ 1 & \tilde{\text{A}}_{iT_i} \end{bmatrix}.$$

$\mathbf{Z}_i$ is the $(T_i \times 2)$ design vector for the random part and equal to $\mathbf{X}_i$ for every person i. The vector of fixed regression coefficients is denoted by $\boldsymbol{\beta} = (\beta_{00}, \beta_{10})^T$. The (1×2) vector $\boldsymbol{\nu}_i = (\nu_{0i}, \nu_{1i})^T$ consists of the random effects for the person-specific intercept and slopes. The $(T_i \times 1)$ vector $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{iT_i})^T$ includes the level 1 errors of person i. The covariance matrix of the level 1 errors is given by Eq. 14.

Conditioning on the design matrices $\mathbf{X}_i$ and $\mathbf{Z}_i$, the covariance matrix $\mathbf{V}_i$ of outcome vector $\mathbf{y}_i$ for person i is equal to:

$$\mathbf{V}_i = \mathbf{Z}_i \boldsymbol{\Sigma}_\nu \mathbf{Z}_i^T + \boldsymbol{\Sigma}_\epsilon$$

When assuming that $T_i = T$ for all persons i and that the N and T are sufficiently large, we obtain the following closed-form approximation of the precision matrix (see Appendix A for a step-by-step derivation)[9]:

$$\mathbf{P}^*_{\hat{\boldsymbol{\beta}}} = N(\mathbf{C}^* - \mathbf{C}^*(\mathbf{I}_2 + \boldsymbol{\Sigma}_\nu \mathbf{C}^*)^{-1}\mathbf{C}^*)$$

where $\mathbf{C}^*$ is a (2×2) matrix with elements

$$\begin{aligned} c^*_{11} &= \frac{(T-2)(1-\rho_\epsilon^2) + 2(1-\rho_\epsilon)}{\sigma_\epsilon^2(1-\rho_\epsilon^2)} \\ c^*_{12} &= c^*_{21} = 0 \\ c^*_{22} &= \frac{(1+\rho_\epsilon^2)(T-1)(\sigma_\upsilon^2 + \sigma_\tau^2/(1-\rho_\tau^2)) - 2\rho_\epsilon(T-1)(\sigma_\upsilon^2 + \rho_\tau\sigma_\tau^2/(1-\rho_\tau^2))}{\sigma_\epsilon^2(1-\rho_\epsilon)^2} \end{aligned} \tag{22}$$

[9] We note that if we do not assume between-person differences in the distribution of the level 1 predictor and that level 1 errors are independent and identically distributed with variance σ_τ^2, Eq. 20 reduces to

$$\text{A}_{it} = \mu_\text{A} + \tau_{it}$$

which simplifies the elements of matrix $\mathbf{C}^*$ to:

$$\begin{aligned} c^*_{11} &= \frac{(T-2)(1-\rho_\epsilon^2) + 2(1-\rho_\epsilon)}{\sigma_\epsilon^2(1-\rho_\epsilon^2)} \\ c^*_{12} &= c^*_{21} = 0 \\ c^*_{22} &= \frac{\sigma_\tau^2(1+\rho_\epsilon^2)(T-1)}{\sigma_\epsilon^2(1-\rho_\epsilon)^2} \end{aligned}$$

Illustration

To show how (22) can be used to estimate power, we use the clinical study data set to obtain plausible parameter values for model (19):

$$\text{NA}_{it} = 42.98 + 0.14\tilde{\text{A}}_{it} + \nu_{0i} + \nu_{1i}\tilde{\text{A}}_{it} + \epsilon_{it}$$

with $\sigma_{\nu_0} = 14.78$, $\sigma_{\nu_1} = 0.12$, $\sigma_{\nu_{01}} = 0.005$, $\rho_\epsilon = 0.43$, $\sigma_\epsilon = 11.92$. The values of the parameters for model (20) are:

$$\tilde{\text{A}}_{it} = 51.74 + \upsilon_i + \tau_{it} \tag{23}$$

with $\sigma_\upsilon = 13.63$, $\rho_\tau = 0.19$, and $\sigma_\tau = 19.61$.

To compare the two approaches we set β^*_{10} to 0.014 and 0.14, the number of participants to 20, 30, 60, 80, 100, 200, and 300, and the number of measurement occasions per person to 20, 70, and 140. For the simulation-based approach, data from the level 1 predictor was simulated using model (23). Figures 5 and 6 report the statistical power estimates for testing the hypothesis that β_{10} is positive and the standard error of $\hat{\beta}_{10}$, respectively. We observe that the analytic approach yields estimates comparable to the simulation-based ones.

Cross-level interaction effects

Next, we focus on persons without MDD and investigate if depression moderates the effect of Anhedonia on NA. To this end, we consider a model that incorporates an interaction effect between the continuous level 2 predictor Dep_i and the continuous person-mean centered level 1 predictor $\tilde{\text{A}}_{it}$. The regression equation is:

$$\begin{aligned} \text{NA}_{it} =& \beta_{00} + \beta_{01}\text{Dep}_i + \beta_{10}\tilde{\text{A}}_{it} + \beta_{11}\text{Dep}_i\tilde{\text{A}}_{it} \\ &+ \nu_{0i} + \nu_{1i}\tilde{\text{A}}_{it} + \epsilon_{it} \end{aligned} \tag{24}$$

where β_{00} is the fixed intercept, β_{01} represents the main effect of Depression, β_{10} is the fixed slope of Anhedonia, and β_{11} represents the cross-level interaction effect between Anhedonia and Depression which assesses whether Depression moderates the effect of Anhedonia on NA. Just like in the previous model, we model between-person differences in the distribution of Anhedonia using (20).

Asymptotic approximation of the precision matrix

For this model, the approximation of the precision matrix of the fixed effects is given by the following (4×4) block

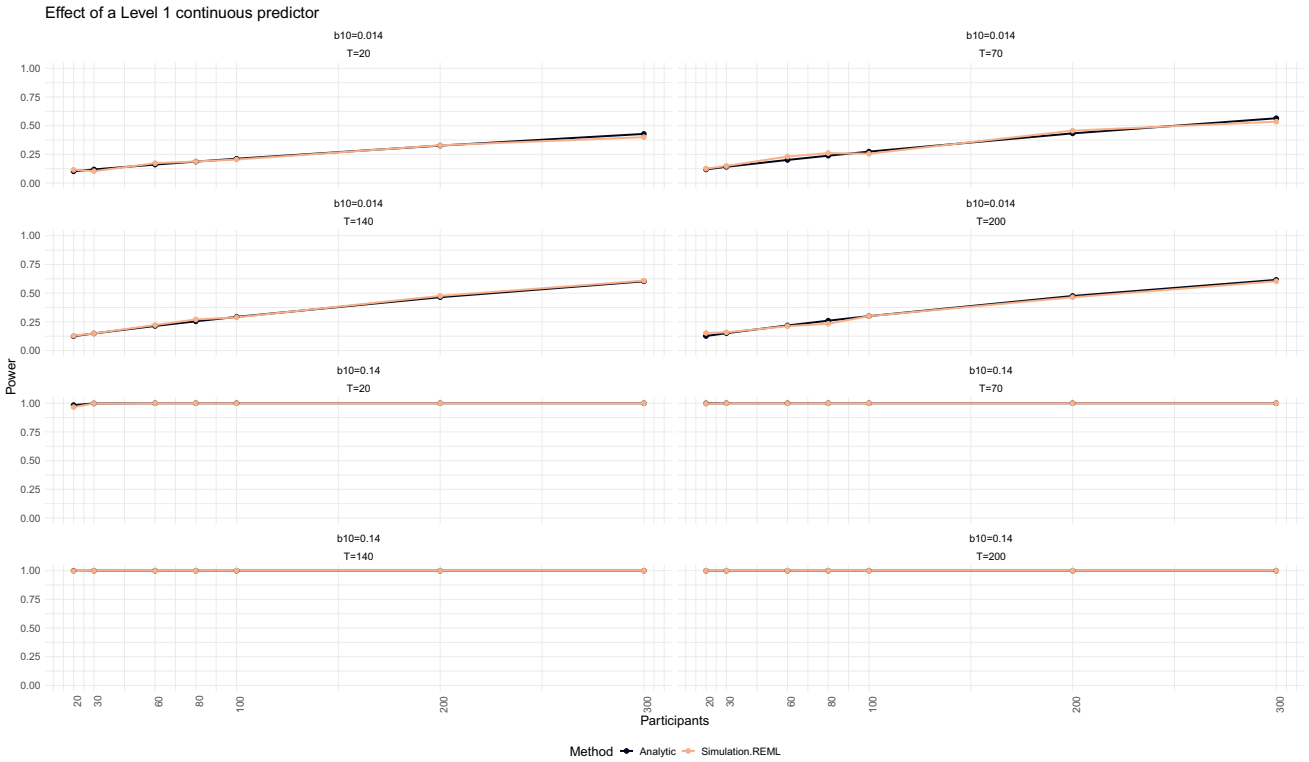


Fig. 5 Illustration 3. Effect of anhedonia on negative affect. The figures show the statistical power for estimating the effect of interest (β_{10}) of the analytic and simulation-based approaches

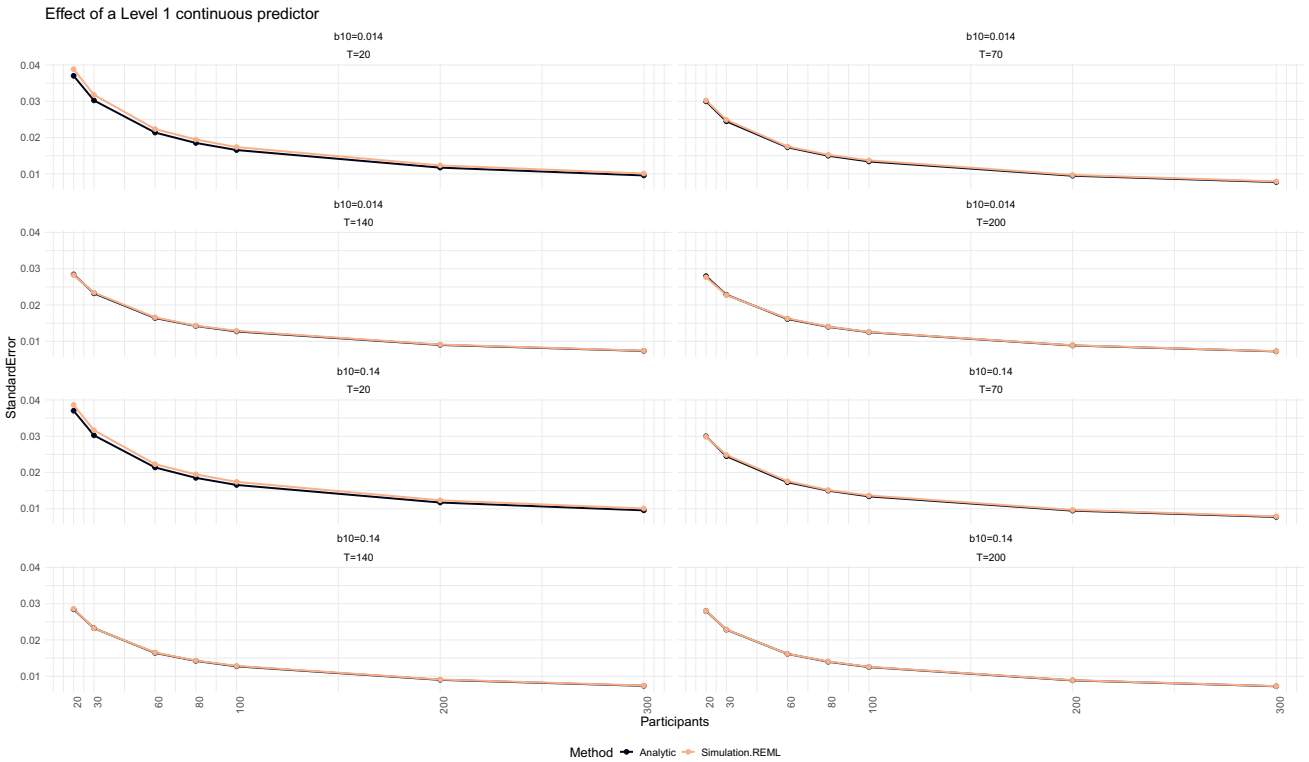


Fig. 6 Illustration 3. Effect of anhedonia on negative affect. The figures show the standard error of the fixed effect estimate $\hat{\beta}_{10}$ of the analytic and simulation-based approaches

matrix (see Appendix A for a step-by-step derivation)

$$\mathbf{P}^*_{\hat{\boldsymbol{\beta}}} = \begin{bmatrix} N\mathbf{E}^* & N\mu_{\text{Dep}}\mathbf{E}^* \\ N\mu_{\text{Dep}}\mathbf{E}^* & ((N-1)\sigma^2_{\text{Dep}} + N\mu^2_{\text{Dep}})\mathbf{E}^* \end{bmatrix}$$

with $\mathbf{E}^* = (\mathbf{C}^* - \mathbf{C}^*(\mathbf{I}_2 + \boldsymbol{\Sigma}_\nu \mathbf{C}^*)^{-1}\boldsymbol{\Sigma}_\nu \mathbf{C}^*)$ and $\mathbf{C}^*$ as in Eq. 22.

Illustration

Using the clinical data set we arrive at the following data-generating model:

$$\begin{aligned} \text{NA}_{it} = \quad & 10.37 + 1.10\text{Dep}_i + 0.16\tilde{\text{A}}_{it} + 0.02\text{Dep}_i\tilde{\text{A}}_{it} \\ & + \nu_{0i} + \nu_{1i}\tilde{\text{A}}_{it} + \epsilon_{it} \end{aligned}$$

with $\sigma_{\nu_0} = 5.45$, $\sigma_{\nu_1} = 0.10$, $\sigma_{\nu_{01}} = 0.125$, $\rho_\epsilon = 0.23$, $\sigma_\epsilon = 6.49$, $\mu_{\text{Dep}} = 3.33$, and $\sigma_{\text{Dep}} = 2.53$. Further, we have

$$\text{A}_{it} = 17.22 + \upsilon_i + \tau_{it} \tag{25}$$

with $\sigma_\upsilon = 12.48$, $\rho_\tau = 0.23$, and $\sigma_\tau = 16.79$.

Figure 7 shows the estimates of the statistical power for testing whether $H_1 : \beta_{11} > 0$ and Fig. 8 displays the standard error of $\hat{\beta}_{11}$ when varying the number of persons, the number of repeated measurements, and the value of the main effect of interest β^*_{11}. In line with the previous results, the simulation-based approach and the analytic approach yield similar results, with the statistical power for the analytic approach being slightly larger than the simulation-based approach when β^*_{11} is 0.02, N lower than 100, and T equal to 20.

Group differences in the effect of a level 1 continuous predictor

The final model estimates differences in the association between NA and Anhedonia between people that got diagnosed with MDD and the control group under the assumption of AR(1) errors:

$$\begin{aligned} y_{it} = & \beta_{00} + \beta_{01}\text{Diag}_i + \beta_{10}\tilde{\text{A}}_{it} + \beta_{11}\text{Diag}_i\tilde{\text{A}}_{it} \\ & + \nu_{0i} + \nu_{1i}\tilde{\text{A}}_{it} + \epsilon_{it} \end{aligned} \tag{26}$$

In this model, the coefficient β_{10} represents the within-person effect of anhedonia on NA for the control group, and $\beta_{10} + \beta_{11}$ the within-person effect of anhedonia on NA for persons diagnosed with MDD. In this model, the main effect of interest is the coefficient β_{11} as it represents group differences in the fixed slope.

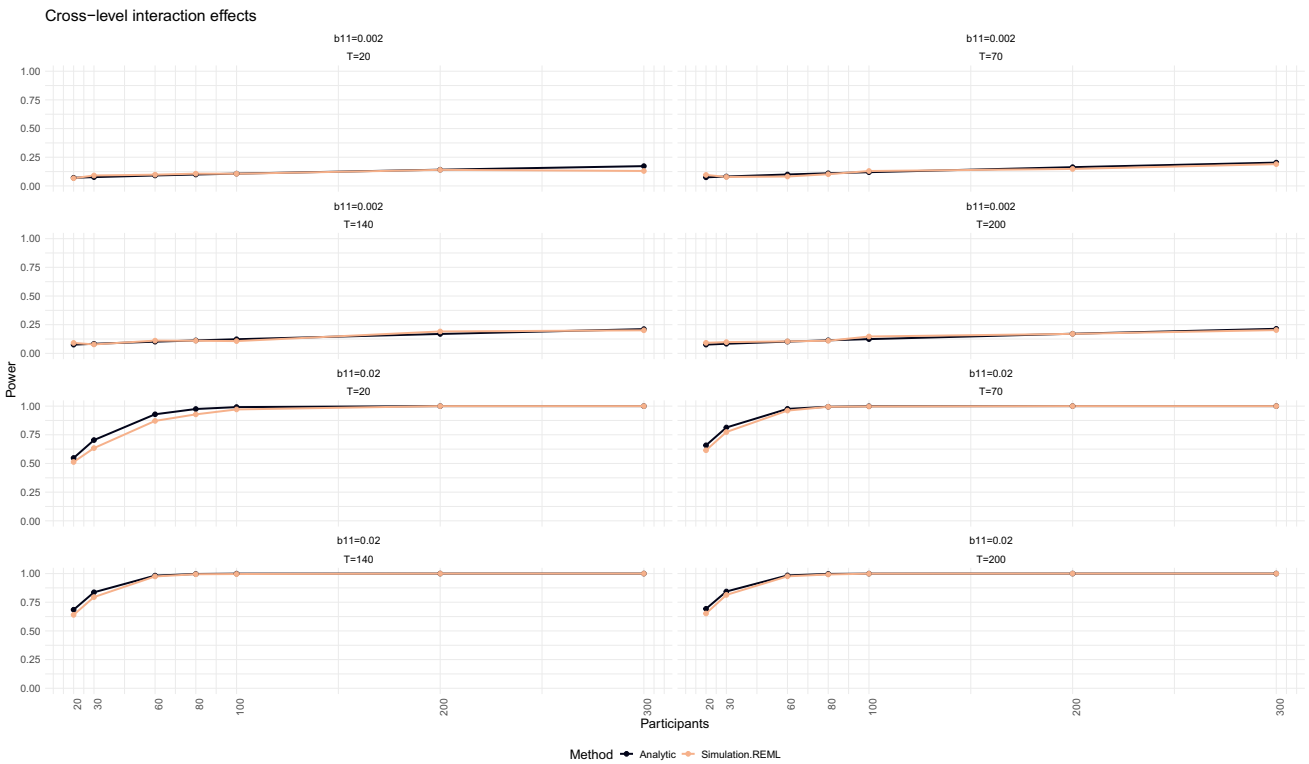


Fig. 7 Illustration 4. Cross-level interaction effect between depression and anhedonia on negative affect. The figures show the statistical power for estimating the effect of interest (β_{11}) of the analytic and simulation-based approaches

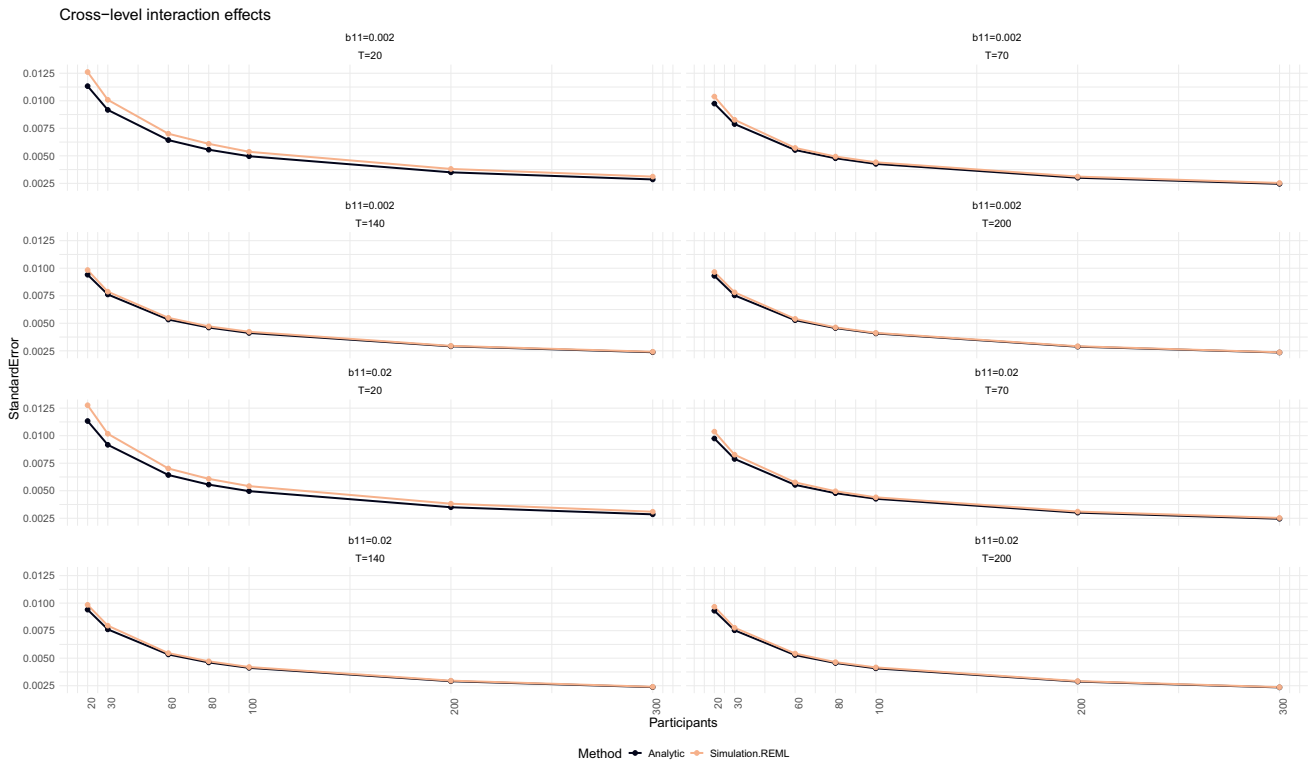


Fig. 8 Illustration 4. Cross-level interaction effect between depression and anhedonia on negative affect. The figures show the standard error of the fixed effect estimate $\hat{\beta}_{11}$ of the analytic and simulation-based approaches

In addition, we assume the following model for the level 1 predictor for participants in the k-th group (for $k \in \{0, 1\}$):

$$\mathrm{A}_{kit} = \mu_{\mathrm{A}_k} + \upsilon_{ki} + \tau_{kit}$$

where μ_{A_k} is the mean of Anhedonia across all persons in group k, υ_{ki} is the deviation of each person's mean anhedonia level from μ_{A_k} with standard deviation σ_{υ_k}, and τ_{kit} denotes the within-person error of A_{kit}, which is assumed to follow an AR(1) process with autocorrelation ρ_{τ_k} and variance $\sigma^2_{\tau_k}$.

Asymptotic approximation of the precision matrix

The asymptotic approximation of the precision matrix of the fixed effects $\boldsymbol{\beta} = (\beta_{00}, \beta_{01}, \beta_{10}, \beta_{11})^T$ is given by the following expression:

$$\mathbf{P}^*_{\hat{\boldsymbol{\beta}}} = \begin{bmatrix} N_0\mathbf{E}^*_0 + N_1\mathbf{E}^*_1 & N_1\mathbf{E}^*_1 \\ N_1\mathbf{E}^*_1 & N_1\mathbf{E}^*_1 \end{bmatrix}$$

where $\mathbf{E}^*_0 = (\mathbf{C}^*_0 - \mathbf{C}^*_0(\mathbf{I}_2 + \boldsymbol{\Sigma}_\nu\mathbf{C}^*_0)^{-1}\boldsymbol{\Sigma}_\nu\mathbf{C}^*_0)$, with $\mathbf{C}^*_0$ defined by Eq. 22 for participants in the control group, while $\mathbf{E}^*_1 = (\mathbf{C}^*_1 - \mathbf{C}^*_1(\mathbf{I}_2 + \boldsymbol{\Sigma}_\nu\mathbf{C}^*_1)^{-1}\boldsymbol{\Sigma}_\nu\mathbf{C}^*_1)$ is defined replacing the corresponding values in Eq. 22 with those for participants with MDD.

Illustration

Using the Clinical study we obtain the following values of the parameters of model (26): $\beta_{00} = 10.39$, $\beta_{01} = 32.59$, $\beta_{10} = 0.15$, $\beta_{11} = 0.05$, $\sigma_{\nu_0} = 11.20$, $\sigma_{\nu_1} = 0.11$, $\sigma_{\nu_{01}} = 0.033$, $\rho_\epsilon = 0.38$, and $\sigma_\epsilon = 9.40$. For persons diagnosed with MDD and controls, the values of the parameters for model (20) are reported in Eqs. 23 and 25, respectively.

Figures 9 and 10 show the estimates of statistical power and the standard error of $\hat{\beta}_{11}$. In general, we observe that results are comparable between the two approaches. Specifically, we observe that the statistical power to test group differences in the effect of anhedonia on NA is slightly higher for the analytic approach when the number of persons in each group is below 200 and the number of repeated measurements is 20. In these settings, we also observe that the standard error of $\hat{\beta}_{11}$ computed with the analytic approach is smaller than that of the simulation-based approach.

Shiny application

To complement the results presented in this section, we developed the Shiny application *ApproxPowerIL*. This app allows to analytically compute power for the five models presented in this section using the asymptotic approximation of the precision matrix of the fixed effects. The Shiny application was

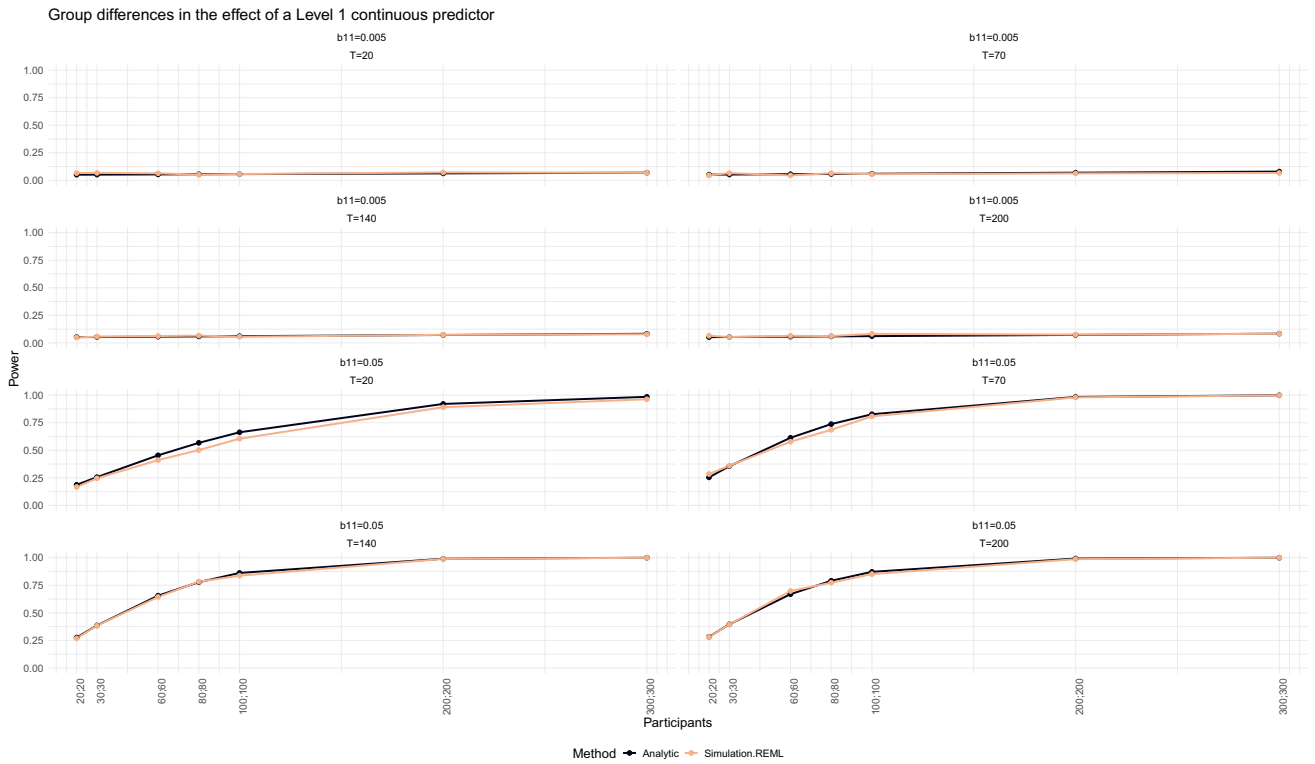


Fig. 9 Illustration 5. Group differences in the effect of anhedonia on negative affect. The figures show the statistical power for estimating the effect of interest (β_{11}) of the analytic and simulation-based approaches

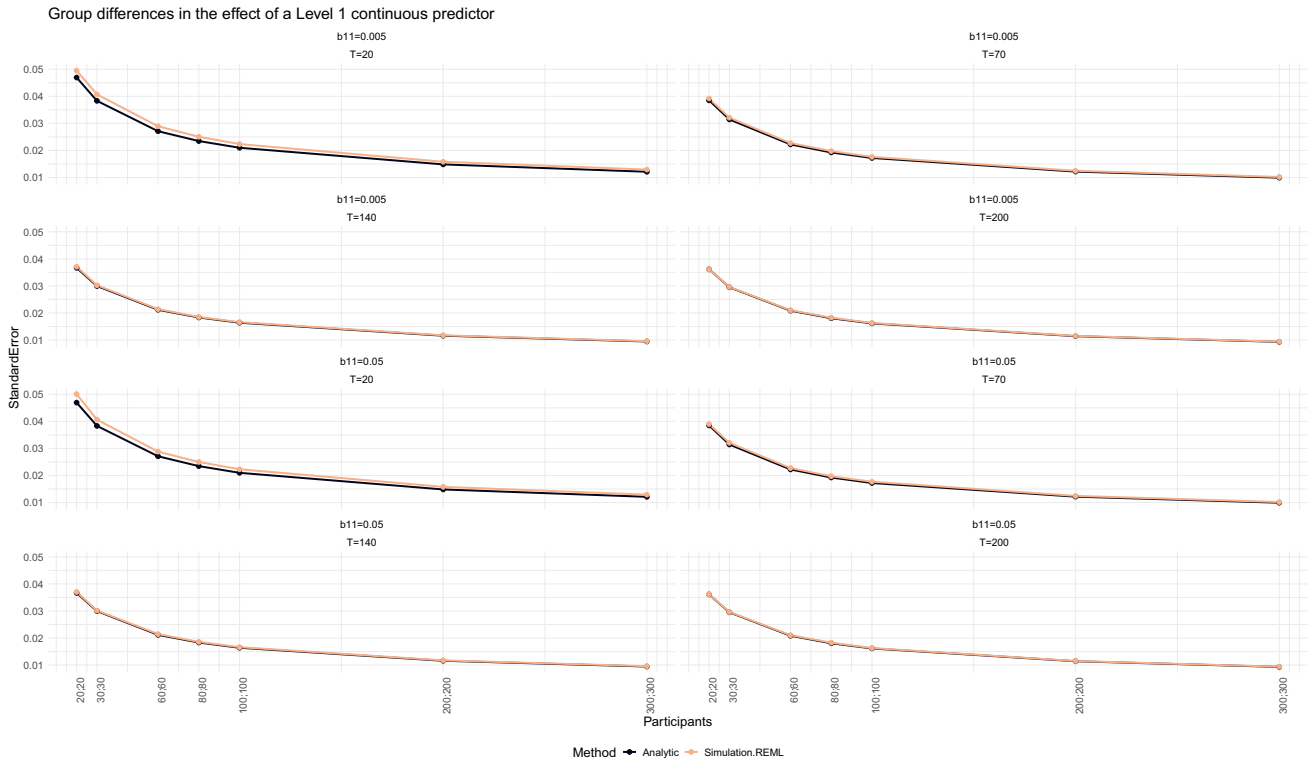


Fig. 10 Illustration 5. Group differences in the effect of anhedonia on negative affect. The figures show the standard error of the fixed effect estimate $\hat{\beta}_{11}$ of the analytic and simulation-based approaches

created using the R package Shiny version 1.5.0 (Chang et al., 2019). The application is freely available on a Git repository hosted on GitLab at https://gitlab.kuleuven.be/ppw-okpiv/researchers/u0119584/ApproxPowerIL.

Discussion

Properly powered studies mitigate the chances of false negative findings and gross over- and underestimation of effect sizes. Therefore, running power analyses when designing a study contributes to the replicability and generalizability of scientific findings (see Ioannidis, 2005; Szucs and Ioannidis, 2017). To aid researchers working with IL designs, we derived closed-form expressions for the precision matrix of the fixed effects of five different multilevel models that handle the serial dependence in the data. Furthermore, we derived asymptotic approximations of those matrices that do not depend on sample statistics under the assumption that the number of measurement occasions is equal across persons. We then demonstrated how these formulas allow researchers to conduct a power analysis for NHST of the fixed effects. To validate this approach to power analysis, we compared it to the simulation-based approach as done in Lafit et al. (2021) using Monte-Carlo simulations. We found no or only minor differences in power estimates showcasing that the analytical formulas provided are useful as they free researchers from the computational burden of the simulation-based approach.

A first direction for future research is related to the assumption that the number of measurement occasions and persons is sufficiently large. As we showed in the illustrations, when the number of persons or the number of repeated measurements are small the differences between the analytic and the simulation-based approach increase. How large the sample size has to be for the differences between the two approaches to be small likely depends on the model under study and the true values of the model parameters, therefore, it is not possible to provide general rules of thumb in this regard.

Secondly, we note that in this paper only focused on five commonly used models in IL research. Although we did not investigate the potential of the analytical derivations for other types of models, future research could focus on deriving analytical formulas for the precision matrix of more complex models. Note that (27) shows the precision matrix of a model that includes P level 1 predictors and multiple cross-level interaction effects. This result can be the starting point for the derivation of analytical formulas that can be used to compute statistical power for more complex models.

A third research direction pertains to the impact of having different numbers of measurement occasions across persons and having unequal time intervals between measurements (such as night breaks). Concerning the first issue, future research could investigate under which conditions it is still possible to obtain a closed-form expression of the precision matrix of the fixed effects. For example, if there are subgroups of individuals with the same number of measurement occasions, the precision matrix of the fixed effects can be computed as the average of the precision matrices in each of the subgroups. Concerning the second issue, the multilevel models presented in this paper cannot accommodate unequal time intervals in the data. In the $N = 1$ case, Adolf et al. (2021) derived analytical expressions for the covariance matrix for the continuous-time autoregressive model estimators to obtain the optimal sampling rate. This framework could be extended to the multilevel case to accommodate unequal time intervals.

A fourth research direction concerns which null hypothesis is tested. In this paper, we only focused on null hypothesis tests of fixed effects of interest. Future research could use the analytical approximations of the precision matrix derived in this paper to obtain statistical power for variance component testing (see e.g., Berkhof and Snijders, 2001).

A final direction for future research concerns the uncertainty in the parameter estimates used for conducting power analysis. For multilevel models, power analysis is often conducted using pilot data or data from previous studies to obtain the values of the model parameters (Lane and Hennes, 2018). This can result in uncertainty about sample size recommendations in case of large uncertainty in the parameter estimates based on these data (e.g., when using data from a small pilot study). Future research could investigate the impact of uncertainty about the true values of the model parameters by conducting sensitivity analyses (see e.g., Lafit et al. 2022; Lane and Hennes 2018; Lane and Hennes 2019). Given the low computational burden of the analytic approach, such sensitivity analyses can be easily implemented. We also note that assumptions about the distribution of the level 1 and level 2 predictors are needed to compute statistical power with the simulation-based or the analytic approach. Violations of these assumptions can severely impact sample size recommendations. Thus, sensitivity analyses can also be performed to investigate the impact of violations of the distributional assumptions about the model's predictors on statistical power.

Conclusion

The current study introduced asymptotic approximations of the precision matrix for multilevel models with autocorrelated errors. These approximations can be used to conduct statistical power analyses without any computational burden.

A Appendix

A.1 Multilevel linear model with autocorrelated errors

To begin with, we focus on a two-level model including a level 1 continuous outcome y_{it} for person $i = (1, \ldots, N)$ and measurement occasion $t = (1, \ldots, T_i)$, where T_i denotes the total number of measurement occasions for the i-th individual. Specifically, we are interested in investigating the effect of a level 2, or time-invariant, predictor w_i and P continuous level 1, or time-varying, person-mean centered predictors $\tilde{x}_{pit} = x_{pit} - \bar{x}_{pi}$ for $p = (1, \ldots, P)$, where $\bar{x}_{pi}$ denotes the mean of each of the p-th level 1 predictor of person i. We restrict our attention to the low-dimensional case where the number of predictors P is much smaller than the number of measurement occasions (i.e., $\min_{i=1,\ldots,N} T_i \gg P$). Furthermore, we assume that the level 1 predictors are independent. The model also includes interaction effects between the level 2 predictor and the level 1 predictors. The level 1 regression equation is:

$$y_{it} = \gamma_{0i} + \sum_{p=1}^{P} \gamma_{pi} \tilde{x}_{pit} + \epsilon_{it}$$

where γ_{0i} is the intercept for person i, γ_{pi} is the slope of the p-th level 1 predictor of person i, and ϵ_{it} is the level 1 or within-person error of person i at time-point t. We assume that the level 1 errors follow an autoregressive (AR(1)) process (Goldstein et al., 1994):

$$\epsilon_{it} = \rho_{\epsilon} \epsilon_{it-1} + \varepsilon_{it}$$

where the correlation between two consecutive errors is denoted by ρ_{ϵ}, and ε_{it} is a Gaussian error with mean zero and variance σ^2_{ε}. Under this model, the variance of the level 1 errors are given by $\sigma^2_{\epsilon} = \sigma^2_{\varepsilon}/(1 - \rho^2_{\epsilon})$. To guarantee that this model is stationary, we assume that the absolute value of the autocorrelation ρ_{ϵ} is strictly smaller than 1 (Hamilton, 1994).

The level 2 models for the random intercept and the random slopes are as follows:

$$\gamma_{0i} = \beta_{00} + \beta_{01} w_i + \nu_{0i}$$
$$\gamma_{pi} = \beta_{p0} + \beta_{p1} w_i + \nu_{pi} \quad \text{for } p = (1, \ldots, P)$$

The level 2 errors follow a multivariate normal distribution with mean zero and covariance matrix $\boldsymbol{\Sigma}_{\nu}$:

$$\boldsymbol{\Sigma}_{\nu} = \begin{bmatrix} \sigma^2_{\nu_0} & \sigma_{\nu_{01}} & \ldots & \sigma_{\nu_{0P}} \\ \sigma_{\nu_{01}} & \sigma^2_{\nu_1} & \ldots & \sigma_{\nu_{1P}} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\nu_{0P}} & \sigma_{\nu_{1P}} & \ldots & \sigma^2_{\nu_P} \end{bmatrix}$$

Substituting the level 2 models into the level 1 equation we obtain the so-called reduced model:

$$y_{it} = \beta_{00} + \sum_{p=1}^{P} \beta_{p0} \tilde{x}_{pit} + \beta_{01} w_i + \sum_{p=1}^{P} \beta_{p1} w_i \tilde{x}_{pit} + \nu_{0i} + \sum_{p=1}^{P} \nu_{pi} \tilde{x}_{pit} + \epsilon_{it}$$

Concerning the level 2 predictor w_i we restrict our attention to two settings. In the first scenario, the time-invariant predictor is a dummy variable indicating group membership of participant i. Thus, the coefficients β_{p1} represent the difference in the fixed slopes between two groups of participants. In the second scenario, we consider w_i to be a continuous variable. The coefficients then β_{p1} represent cross-level interaction effects between the level 2 and the p-th level 1 predictors, allowing the researcher to investigate whether the time-invariant predictor moderates the influence of the level 1 predictors.

To model between-persons differences in the distribution of the p-th level 1 predictor x_{pit}, we specify the following model for x_{pit}:

$$x_{pit} = \mu_{x_p} + \upsilon_{pi} + \tau_{pit}$$

where μ_{x_p} is the mean of the level 1 predictor across all persons, υ_{pi} is the deviation in the mean of each person from μ_{x_p} with variance denoted by $\sigma^2_{\upsilon_p}$, and τ_{pit} denotes the within-person error of x_{pit}, which is assumed to follow an AR(1) process with autocorrelation denoted by ρ_{τ_p} and variance $\sigma^2_{\tau_p}$. In this model, the mean of x_{pit} for each person is $\mu_{x_{pi}} = \mu_{x_p} + \upsilon_{pi}$, the variance of x_{pit} is $\sigma^2_{\upsilon_p} + \sigma^2_{\tau_p}/(1-\rho^2_{\tau_p})$, and the correlation between two consecutive measures of the level 1 predictor is $\sigma^2_{\upsilon_p} + \rho_{\tau_p}\sigma^2_{\tau_p}/(1 - \rho^2_{\tau_p})$.

A.1.1 Maximum likelihood estimator of the fixed effects

Several approaches (i.e., method of moments, maximum likelihood (ML), restricted maximum likelihood (REML); Raudenbush (1997)) have been proposed to estimate the fixed effects, variance, and covariance estimates of the random effects, and the variance and autocorrelation of the level 1 errors. In this section, we focus on the Maximum likelihood (ML) estimator of the fixed effects.

Using general matrix-based notation (Moerbeek, 2008) our model for person i is equal to

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \boldsymbol{\nu}_i + \boldsymbol{\epsilon}_i$$

where $\mathbf{y}_i$ contains the T_i observations of the outcome variable for person i. $\mathbf{X}_i$ is the $(T_i \times (2 + 2p))$ design matrix for the

fixed part given by

$$\mathbf{X}_i = \begin{bmatrix} 1 & \tilde{x}_{1i1} & \dots & \tilde{x}_{pi1} & w_i & w_i\tilde{x}_{1i1} & \dots & w_i\tilde{x}_{pi1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \tilde{x}_{1iT_i} & \dots & \tilde{x}_{piT_i} & w_i & w_i\tilde{x}_{1iT_i} & \dots & w_i\tilde{x}_{piT_i} \end{bmatrix}$$

$\mathbf{Z}_i$ is the $(T_i \times (1+p))$ design vector for the random part

$$\mathbf{Z}_i = \begin{bmatrix} 1 & \tilde{x}_{1i1} & \dots & \tilde{x}_{pi1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \tilde{x}_{1iT_i} & \dots & \tilde{x}_{piT_i} \end{bmatrix}$$

The vector of fixed regression coefficients is denoted by

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_{00} \\ \beta_{10} \\ \vdots \\ \beta_{P0} \\ \beta_{01} \\ \beta_{11} \\ \vdots \\ \beta_{P1} \end{bmatrix}$$

with β_{00} representing the fixed intercept, β_{p0} the fixed slope for each of the P level 1 predictors, β_{01} the fixed effect of the level 2 predictor, and β_{p1} the fixed interaction effect between the level 2 predictor and p-th level 1 predictor. The $(1 \times (1+P))$ vector $\boldsymbol{\nu}_i = (\nu_{0i}, \nu_{1i}, \dots, \nu_{Pi})^T$ consists of the random effects for the person-specific intercept and slopes.

The $(T_i \times 1)$ vector $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \dots, \epsilon_{iT_i})^T$ includes the level 1 errors of person i. The $(T_i \times T_i)$ covariance matrix of the level 1 errors is given by:

$$\boldsymbol{\Sigma}_{\epsilon_i} = \sigma_\epsilon^2 \boldsymbol{\Gamma}_{\epsilon_i} = \sigma_\epsilon^2 \begin{bmatrix} 1 & \rho_\epsilon & \rho_\epsilon^2 & \dots & \rho_\epsilon^{T_i-1} \\ \rho_\epsilon & 1 & \rho_\epsilon & \dots & \rho_\epsilon^{T_i-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_\epsilon^{T_i-1} & \rho_\epsilon^{T_i-2} & \rho_\epsilon^{T_i-3} & \dots & 1 \end{bmatrix}$$

Conditioning on the design matrices $\mathbf{X}_i$ and $\mathbf{Z}_i$, the $(T_i \times T_i)$ covariance matrix $(\mathbf{V}_i)$ of outcome vector $\mathbf{y}_i$ for person i is equal to:

$$\mathbf{V}_i := \text{cov}(\mathbf{y}_i | \mathbf{X}_i, \mathbf{Z}_i) = \mathbf{Z}_i \boldsymbol{\Sigma}_\nu \mathbf{Z}_i^T + \boldsymbol{\Sigma}_{\epsilon_i}$$

The maximum likelihood (ML) estimator of the fixed effects (i.e., $\boldsymbol{\beta}$) can be represented as generalized least squares (GLS) problem (see Laird and Ware, 1982) where the vector of fixed effects can be estimated as

$$\hat{\boldsymbol{\beta}} | \mathbf{X}_i, \mathbf{Z}_i = \left(\sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{y}_i.$$

Its covariance matrix is equal to

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\beta}}} = \left(\sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \right)^{-1}$$

with the square root of the diagonal elements of the covariance matrix $\text{cov}(\hat{\boldsymbol{\beta}})$ representing the standard errors of the fixed effects estimates.

The covariance matrix of $\hat{\boldsymbol{\beta}}$ is the inverse of the precision matrix of the vector of fixed effects $\boldsymbol{\beta}$:

$$\mathbf{P}_{\hat{\boldsymbol{\beta}}} = \left(\sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \right)$$

A.1.2 Algebraic Derivation of the Precision Matrix

In this section, we show a step-by-step derivation of the precision matrix of the fixed effect estimates.

Via the Woodbury matrix identity (Hager, 1989) we can write the precision matrix as

$$\begin{aligned} \mathbf{P}_{\hat{\boldsymbol{\beta}}} &= \sum_{i=1}^N \mathbf{X}_i^T \mathbf{V}_i^{-1} \mathbf{X}_i \\ &= \sum_{i=1}^N \mathbf{X}_i^T \left(\boldsymbol{\Sigma}_{\epsilon_i} + \mathbf{Z}_i \boldsymbol{\Sigma}_\nu \mathbf{Z}_i^T \right)^{-1} \mathbf{X}_i \\ &= \sum_{i=1}^N \mathbf{X}_i^T \left(\boldsymbol{\Sigma}_{\epsilon_i}^{-1} - \Sigma_{\epsilon_i}^{-1} \mathbf{Z}_i \left(\boldsymbol{\Sigma}_\nu^{-1} + \mathbf{Z}_i^T \boldsymbol{\Sigma}_{\epsilon_i}^{-1} \mathbf{Z}_i \right)^{-1} \mathbf{Z}_i^T \boldsymbol{\Sigma}_{\epsilon_i}^{-1} \right) \mathbf{X}_i \\ &= \sum_{i=1}^N \underbrace{\mathbf{X}_i^T \boldsymbol{\Sigma}_{\epsilon_i}^{-1} \mathbf{X}_i}_{=:\boldsymbol{A_i}} - \underbrace{(\mathbf{X}_i^T \boldsymbol{\Sigma}_{\epsilon_i}^{-1} \mathbf{Z}_i)}_{=:\boldsymbol{B_i}} \underbrace{(\boldsymbol{\Sigma}_\nu^{-1} + \overbrace{\mathbf{Z}_i^T \boldsymbol{\Sigma}_{\epsilon_i}^{-1} \mathbf{Z}_i}^{=:\boldsymbol{C_i}})^{-1}}_{=:\boldsymbol{D_i}} \underbrace{(\mathbf{Z}_i^T \boldsymbol{\Sigma}_{\epsilon_i}^{-1} \mathbf{X}_i)}_{=\boldsymbol{B_i}^T} \end{aligned}$$

To simplify the above formula, we first define the sample cross-regressive covariance of LAG 1 for predictor variables x_{kit} and x_{lit} of person i as $\phi_i^{kl} := \frac{1}{T_i - 1} \sum_{j=1}^{T_i - 1} \tilde{x}_{ki,T_i+1} \tilde{x}_{li,T_i}$. Further, let $S_i^{kk} := \frac{1}{T_i - 1} \sum_{j=1}^{T_i} \tilde{x}_{kij}^2$ be the sample variance of predictor x_{kit} of person i and $S_i^{kl} := \frac{1}{T_i - 1} \sum_{j=1}^{T_i} \tilde{x}_{kij} \tilde{x}_{lij}$ be the sample covariance of predictors x_{kit} and x_{jit} of person i with $\tilde{x}_{kij} := x_{kij} - \frac{1}{T_i} \sum_j^T x_{kij}$ representing the centralized predictor variable x_{kij} at time point j. Using *Theorem 2.1* in Sutradhar and Kumar (2003), we can write the inverse of $\boldsymbol{\Gamma}_{\epsilon_i}$ as

$$\frac{1}{1 - \rho_\epsilon^2} \begin{bmatrix} 1 & -\rho_\epsilon & 0 & 0 & \dots & 0 & 0 \\ -\rho_\epsilon & 1 + \rho_\epsilon^2 & -\rho_\epsilon & 0 & \dots & 0 & 0 \\ 0 & -\rho_\epsilon & 1 + \rho_\epsilon^2 & -\rho_\epsilon & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 + \rho_\epsilon^2 & -\rho_\epsilon \\ 0 & 0 & 0 & 0 & \dots & -\rho_\epsilon & 1 \end{bmatrix}$$

and note that this matrix has dimension $(T_i \times T_i)$.

Let us start with $\boldsymbol{C_i} = \boldsymbol{C_i}^T = (c_{kl})_i \in \mathbb{R}^{(p+1)\times(p+1)}$ which is equal to

$$\begin{bmatrix} 1 & \dots & 1 \\ x_{1i1} & \dots & x_{1iT} \\ \vdots & \vdots & \vdots \\ x_{pi1} & \dots & x_{piT} \end{bmatrix} \frac{1}{\sigma_\epsilon^2(1-\rho_\epsilon^2)} \begin{bmatrix} 1 & -\rho_\epsilon & 0 & 0 & \dots & 0 & 0 \\ -\rho_\epsilon & 1+\rho_\epsilon^2 & -\rho_\epsilon & 0 & \dots & 0 & 0 \\ 0 & -\rho_\epsilon & 1+\rho_\epsilon^2 & -\rho_\epsilon & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1+\rho_\epsilon^2 & -\rho_\epsilon \\ 0 & 0 & 0 & 0 & \dots & -\rho_\epsilon & 1 \end{bmatrix} \begin{bmatrix} 1 & x_{1i1} & \dots & x_{pi1} \\ \vdots & \vdots & \vdots & \\ 1 & x_{1iT} & \dots & x_{piT} \end{bmatrix}$$

Hence, we have:

$$\forall i : c_{11} = \frac{(T_i-2)(1-\rho_\epsilon)^2 + 2(1-\rho_\epsilon)}{\sigma_\epsilon^2(1-\rho_\epsilon^2)}$$

$$\begin{aligned} \forall i, \forall k \geq 2 : c_{1k} = c_{k1} &= \tilde{x}_{ki1} + (1+\rho_\epsilon^2)\sum_{j=2}^{T_i-1}\tilde{x}_{kij} - \rho_\epsilon\sum_{j=1}^{T_i}\tilde{x}_{kij} - \rho_\epsilon\sum_{j=2}^{T_i-1}\tilde{x}_{kij} + \tilde{x}_{kiT_i} \\ &= (1+\rho_\epsilon)(\tilde{x}_{ki1} + \tilde{x}_{kiT_i}) - 2\rho_\epsilon\underbrace{\sum_{j=1}^{T_i}\tilde{x}_{kij}}_{=0} - (1+\rho_\epsilon^2)(\tilde{x}_{ki1} + \tilde{x}_{kiT_i}) + (1+\rho_\epsilon^2)\underbrace{\sum_{j=1}^{T_i}\tilde{x}_{kij}}_{=0} \\ &= \frac{\rho_\epsilon(1-\rho_\epsilon)(\tilde{x}_{ki1} + \tilde{x}_{kiT_i})}{\sigma_\epsilon^2(1-\rho_\epsilon^2)} \end{aligned}$$

$\forall i, \forall k, l \geq 2 :$

$$\begin{aligned} c_{kl} = c_{lk} &= \frac{\tilde{x}_{ki1}\tilde{x}_{li1} - \rho_\epsilon\sum_{j=1}^{T_i-1}\tilde{x}_{ki,j+1}\tilde{x}_{li,j} - \rho_\epsilon\sum_{j=1}^{T_i-1}\tilde{x}_{kij}\tilde{x}_{li,j+1} + (1+\rho_\epsilon)^2\sum_{k=2}^{T_i-2}\tilde{x}_{kij}\tilde{x}_{lij} + \tilde{x}_{kiT_i}\tilde{x}_{liT_i}}{\sigma_\epsilon^2(1-\rho_\epsilon^2)} \\ &= \frac{(1+\rho_\epsilon)^2(T_i-1)S_i^{kl} - \rho_\epsilon(T_i-1)(\phi_i^{kl}+\phi_i^{lk}) - \rho_\epsilon^2(\tilde{x}_{ki1}\tilde{x}_{li1} + \tilde{x}_{kiT_i}\tilde{x}_{liT_i})}{\sigma_\epsilon^2(1-\rho_\epsilon^2)} \\ &= \frac{(1+\rho_\epsilon)^2(T_i-1)S_i^{kk} - 2\rho_\epsilon(T_i-1)\phi_i^{kk} - \rho_\epsilon^2(\tilde{x}_{ki1}^2 + \tilde{x}_{kiT_i}^2)}{\sigma_\epsilon^2(1-\rho_\epsilon^2)} \text{ if } k = l \end{aligned}$$

Noting that

$$\mathbf{X}_i = \begin{bmatrix} 1 & \tilde{x}_{1i1} & \dots & \tilde{x}_{pi1} & w_i & w_i\tilde{x}_{1i1} & \dots & w_i\tilde{x}_{pi1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \tilde{x}_{1iT_i} & \dots & \tilde{x}_{piT_i} & w_i & w_i\tilde{x}_{1iT_i} & \dots & w_i\tilde{x}_{piT_i} \end{bmatrix} = \begin{bmatrix} \boldsymbol{Z_i} & w_i\boldsymbol{Z_i} \end{bmatrix}$$

we further find that $\boldsymbol{A_i} = \boldsymbol{A_i}^T$ is equal to

and that $\boldsymbol{B_i}$ is equal to

$$\mathbf{X}_i^T\boldsymbol{\Sigma}_{\epsilon_i}^{-1}\mathbf{Z}_i = \begin{bmatrix} \boldsymbol{Z_i}^T \\ w_i\boldsymbol{Z_i}^T \end{bmatrix}\boldsymbol{\Sigma}_{\epsilon_i}^{-1}\boldsymbol{Z_i} = \begin{bmatrix} \boldsymbol{Z_i}^T\boldsymbol{\Sigma}_{\epsilon_i}^{-1}\boldsymbol{Z_i} \\ w_i\boldsymbol{Z_i}^T\boldsymbol{\Sigma}_{\epsilon_i}^{-1}\boldsymbol{Z_i} \end{bmatrix} = \begin{bmatrix} \boldsymbol{C_i} \\ w_i\boldsymbol{C_i} \end{bmatrix}.$$

Following the results in Miller (1981), the inverse of the matrix $\mathbf{D}_i$ can be written as

$$\mathbf{D}_i^{-1} = (\mathbf{I} + \boldsymbol{\Sigma}_v\mathbf{C}_i)^{-1}\boldsymbol{\Sigma}_v$$

$$\mathbf{X}_i^T\boldsymbol{\Sigma}_{\epsilon_i}^{-1}\mathbf{X}_i = \begin{bmatrix} \boldsymbol{Z_i}^T \\ w_i\boldsymbol{Z_i}^T \end{bmatrix}\boldsymbol{\Sigma}_{\epsilon_i}^{-1}\begin{bmatrix} \boldsymbol{Z_i} & w_i\boldsymbol{Z_i} \end{bmatrix} = \left[\begin{array}{c|c} \boldsymbol{Z_i}^T\boldsymbol{\Sigma}_{\epsilon_i}^{-1}\boldsymbol{Z_i} & \boldsymbol{Z_i}^T\boldsymbol{\Sigma}_{\epsilon_i}^{-1}w_i\boldsymbol{Z_i} \\ \hline w_i\boldsymbol{Z_i}^T\boldsymbol{\Sigma}_{\epsilon_i}^{-1}\boldsymbol{Z_i} & w_i\boldsymbol{Z_i}^T\boldsymbol{\Sigma}_{\epsilon_i}^{-1}w_i\boldsymbol{Z_i} \end{array}\right] = \left[\begin{array}{c|c} \boldsymbol{C_i} & w_i\boldsymbol{C_i} \\ \hline w_i\boldsymbol{C_i} & w_i^2\boldsymbol{C_i} \end{array}\right]$$

and, consequently, the precision matrix of the fixed effects estimator is:

$$\boldsymbol{I}(\boldsymbol{\beta}) = \sum_{i=1}^{N} \mathbf{A}_i - \mathbf{B}_i(\mathbf{I} + \boldsymbol{\Sigma}_v \mathbf{C}_i)^{-1} \boldsymbol{\Sigma}_v \mathbf{B}_i^T$$

Putting the above results together, we find that the precision matrix is equal to

$$\begin{aligned}
\mathbf{P}_{\hat{\boldsymbol{\beta}}} &= \sum_{i=1}^{N} \begin{bmatrix} \mathbf{C}_i & w_i\mathbf{C}_i \\ w_i\mathbf{C}_i & w_i^2\mathbf{C}_i \end{bmatrix} - \begin{bmatrix} \mathbf{C}_i \\ w_i\mathbf{C}_i \end{bmatrix} (\mathbf{I} + \boldsymbol{\Sigma}_v\mathbf{C}_i)^{-1}\boldsymbol{\Sigma}_v \begin{bmatrix} \mathbf{C}_i^T & w_i\mathbf{C}_i^T \end{bmatrix} \\
&= \sum_{i=1}^{N} \begin{bmatrix} \mathbf{C}_i & w_i\mathbf{C}_i \\ w_i\mathbf{C}_i & w_i^2\mathbf{C}_i \end{bmatrix} - \begin{bmatrix} \mathbf{C}_i(I + \boldsymbol{\Sigma}_v\mathbf{C}_i)^{-1}\boldsymbol{\Sigma}_v\mathbf{C}_i^T & w_i\mathbf{C}_i(I + \boldsymbol{\Sigma}_v\mathbf{C}_i)^{-1}\boldsymbol{\Sigma}_v\mathbf{C}_i^T \\ w_i\mathbf{C}_i(I + \boldsymbol{\Sigma}_v\mathbf{C}_i)^{-1}\boldsymbol{\Sigma}_v\mathbf{C}_i^T & w_i^2\mathbf{C}_i(I + \boldsymbol{\Sigma}_v\mathbf{C}_i)^{-1}\boldsymbol{\Sigma}_v\mathbf{C}_i^T \end{bmatrix}
\end{aligned}$$

Defining the matrix $\mathbf{E}_i$ as

$$\mathbf{E}_i = \mathbf{C}_i(\mathbf{I} - (\mathbf{I} + \boldsymbol{\Sigma}_v\mathbf{C}_i)^{-1}\Sigma_v\mathbf{C}_i^T)$$

We can represent the precision matrix as

$$I(\boldsymbol{\beta}) = \sum_{i=1}^{N} \begin{bmatrix} \mathbf{E}_i & w_i\mathbf{E}_i \\ w_i\mathbf{E}_i & w_i^2\mathbf{E}_i \end{bmatrix} \tag{27}$$

A.2 Precision matrix for models 1 and 2

Via the Woodbury matrix identity, we have that

$$\mathbf{P}_{\hat{\boldsymbol{\beta}}} = \sum_{i=1}^{N} \underbrace{\mathbf{X}_i^T\boldsymbol{\Sigma}_{\epsilon_i}^{-1}\mathbf{X}_i}_{=:\boldsymbol{A_i}} - \underbrace{(\mathbf{X}_i^T\boldsymbol{\Sigma}_{\epsilon_i}^{-1}\mathbf{Z}_i)}_{=:\boldsymbol{B_i}} \underbrace{(\boldsymbol{\Sigma}_v^{-1} + \overbrace{\mathbf{Z}_i^T\boldsymbol{\Sigma}_{\epsilon_i}{}^{-1}\mathbf{Z}_i}^{=:\boldsymbol{C_i}})^{-1}}_{=:\boldsymbol{D_i}} \underbrace{(\mathbf{Z}_i^T\boldsymbol{\Sigma}_{\epsilon_i}^{-1}\mathbf{X}_i)}_{=\boldsymbol{B_i}^T}$$

Since we assume that $T_i = T \ \forall\ i \in \{1, \ldots, N\}$ we have that $\boldsymbol{\Sigma}_{\epsilon_i}^{-1} = \boldsymbol{\Sigma}_{\epsilon}^{-1} = \sigma_\epsilon^2\boldsymbol{\Gamma}_\epsilon \ \forall\ i \in \{1, \ldots, N\}$. Further we have $\boldsymbol{\Sigma}_v^{-1} = \sigma_{v_0}^{-2}\mathbf{I}_T$. Using *Theorem 2.1* in Sutradhar and Kumar (2003) to invert $\boldsymbol{\Gamma}_\epsilon$ we can thus write $\mathbf{A}_i$, $\mathbf{B}_i$, and $\mathbf{D}_i$ respectively as

$$\begin{aligned}
\mathbf{A}_i &= \frac{1}{\sigma_\epsilon^2(1-\rho_\epsilon^2)} \begin{bmatrix} 1 & \ldots & 1 \\ w_i & \ldots & w_i \end{bmatrix} \begin{bmatrix} 1 & -\rho_\epsilon & 0 & 0 & \ldots & 0 & 0 \\ -\rho_\epsilon & 1+\rho_\epsilon^2 & -\rho_\epsilon & 0 & \ldots & 0 & 0 \\ 0 & -\rho_\epsilon & 1+\rho_\epsilon^2 & -\rho_\epsilon & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1+\rho_\epsilon^2 & -\rho_\epsilon \\ 0 & 0 & 0 & 0 & \ldots & -\rho_\epsilon & 1 \end{bmatrix} \begin{bmatrix} 1 & w_i \\ \vdots & \vdots \\ 1 & w_i \end{bmatrix} \\
&= \frac{1}{\sigma_\epsilon^2(1-\rho_\epsilon^2)} \begin{bmatrix} 1-\rho_\epsilon & (1-\rho_\epsilon)^2 & (1-\rho_\epsilon)^2 & \ldots & 1-\rho_\epsilon \\ w_i(1-\rho_\epsilon) & w_i(1-\rho_\epsilon)^2 & w_i(1-\rho_\epsilon)^2 & \ldots & w_i(1-\rho_\epsilon) \end{bmatrix} \begin{bmatrix} 1 & w_i \\ \vdots & \vdots \\ 1 & w_i \end{bmatrix} \\
&= \frac{1}{\sigma_\epsilon^2(1-\rho_\epsilon^2)} \begin{bmatrix} (T-2)(1-\rho_\epsilon)^2 + 2(1-\rho_\epsilon) & w_i((T-2)(1-\rho_\epsilon)^2 + 2(1-\rho_\epsilon)) \\ w_i((T-2)(1-\rho_\epsilon)^2 + 2(1-\rho_\epsilon)) & w_i^2((T-2)(1-\rho_\epsilon)^2 + 2(1-\rho_\epsilon)) \end{bmatrix} \\
&= \underbrace{\frac{(T-2)(1-\rho)^2 + 2(1-\rho_\epsilon)}{\sigma_\epsilon^2(1-\rho_\epsilon^2)}}_{=:a} \begin{bmatrix} 1 & w_i \\ w_i & w_i^2 \end{bmatrix},
\end{aligned}$$

$$\begin{aligned}
\mathbf{B}_i &= \frac{1}{\sigma_\epsilon^2(1-\rho_\epsilon^2)} \begin{bmatrix} 1-\rho_\epsilon & (1-\rho_\epsilon)^2 & (1-\rho_\epsilon)^2 & \ldots & 1-\rho_\epsilon \\ w_i(1-\rho_\epsilon) & w_i(1-\rho_\epsilon)^2 & w_i(1-\rho_\epsilon)^2 & \ldots & w_i(1-\rho_\epsilon) \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \\
&= a \begin{bmatrix} 1 \\ w_i \end{bmatrix},
\end{aligned}$$

$$\begin{aligned}
\mathbf{D}_i &= \boldsymbol{\Sigma}_{v_i}^{-1} + \frac{1}{\sigma_\epsilon^2(1-\rho_\epsilon^2)} \begin{bmatrix} 1 & \ldots & 1 \end{bmatrix} \\
&\quad \begin{bmatrix} 1 & -\rho_\epsilon & 0 & 0 & \ldots & 0 & 0 \\ -\rho_\epsilon & 1+\rho_\epsilon^2 & -\rho_\epsilon & 0 & \ldots & 0 & 0 \\ 0 & -\rho_\epsilon & 1+\rho_\epsilon^2 & -\rho_\epsilon & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1+\rho_\epsilon^2 & -\rho_\epsilon \\ 0 & 0 & 0 & 0 & \ldots & -\rho_\epsilon & 1 \end{bmatrix} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \\
&= \frac{1}{\sigma_{v_0}^2} + a.
\end{aligned}$$

Further, we find:

$$\begin{aligned}
\mathbf{A}_i - \mathbf{B}_i\mathbf{D}_i^{-1}\mathbf{B}_i^T &= \mathbf{A}_i - \mathbf{D}_i^{-1}\mathbf{B}_i\mathbf{B}_i^T \\
&= \mathbf{A}_i - \left(\frac{\sigma_{v_0}^2}{1+\sigma_{v_0}^2 a}\right)\mathbf{B}_i\mathbf{B}_i^T \\
&= \mathbf{A}_i - \left(\frac{\sigma_{v_0}^2}{1+\sigma_{v_0}^2 a}\right)a^2\begin{bmatrix} 1 & w_i \\ w_i & w_i^2 \end{bmatrix} \\
&= \mathbf{A}_i - \left(\frac{\sigma_{v_0}^2}{1+\sigma_{v_0}^2 a}\right)a\mathbf{A}_i \\
&= \left(1 - \frac{\sigma_{v_0}^2 a}{1+\sigma_{v_0}^2 a}\right)\mathbf{A}_i \\
&= \left(\frac{1}{1+\sigma_{v_0}^2 a}\right)\mathbf{A}_i \\
&= \frac{a}{1+\sigma_{v_0}^2 a}\begin{bmatrix} 1 & w_i \\ w_i & w_i^2 \end{bmatrix}
\end{aligned}$$

Summing over the N participants and re-substituting a, then yields the precision matrix:

$$\begin{aligned}
\sum_{i=1}^{N}\mathbf{A}_i - \mathbf{B}_i\mathbf{D}_i^{-1}\mathbf{B}_i^T &= \sum_{i=1}^{N}\frac{a}{1+\sigma_{v_0}^2 a}\begin{bmatrix} 1 & w_i \\ w_i & w_i^2. \end{bmatrix} \\
&= \frac{a}{1+\sigma_{v_0}^2 a}\sum_{i=1}^{N}\begin{bmatrix} 1 & w_i \\ w_i & w_i^2 \end{bmatrix} \\
&= \frac{a}{1+\sigma_{v_0}^2 a}\begin{bmatrix} N & \sum_{i=1}^N w_i \\ \sum_{i=1}^N w_i & \sum_{i=1}^N w_i^2 \end{bmatrix} \\
&= \frac{a}{1+\sigma_{v_0}^2 a}\begin{bmatrix} N & N\bar{w} \\ N\bar{w} & (N-1)S_w^2 + N\bar{w}^2 \end{bmatrix} \\
&= \frac{\frac{(T-2)(1-\rho_\epsilon)^2+2(1-\rho_\epsilon)}{\sigma_\epsilon^2(1-\rho_\epsilon^2)}}{1+\frac{\sigma_{v_0}^2((T-2)(1-\rho_\epsilon)^2+2(1-\rho_\epsilon))}{\sigma_\epsilon^2(1-\rho_\epsilon^2)}}\begin{bmatrix} N & N\bar{w} \\ N\bar{w} & (N-1)S_w^2 + N\bar{w}^2 \end{bmatrix} \\
&= \frac{(T-2)(1-\rho_\epsilon)^2+2(1-\rho_\epsilon)}{\sigma_\epsilon^2(1-\rho_\epsilon^2)+\sigma_{v_0}^2((T-2)(1-\rho_\epsilon)^2+2(1-\rho_\epsilon))}\begin{bmatrix} N & N\bar{w} \\ N\bar{w} & (N-1)S_w^2 + N\bar{w}^2 \end{bmatrix}
\end{aligned}$$

where S_w^2 is the sample variance of the level 2 variable w_i.

A.3 Asymptotic approximation of the precision matrix for Model 1

Assuming that the number of persons N is sufficiently large $\bar{w}$ converges in probability to μ_w, and S_w^2 converges in probability to σ_w^2. The asymptotic approximation of the precision matrix therefore is:

$$\mathbf{P}_{\hat{\boldsymbol{\beta}}}^{*} = \frac{(T-2)(1-\rho_\epsilon)^2+2(1-\rho_\epsilon)}{\sigma_\epsilon^2(1-\rho_\epsilon^2)+\sigma_{v_0}^2((T-2)(1-\rho_\epsilon)^2+2(1-\rho_\epsilon))}\begin{bmatrix} N & N\mu_w \\ N\mu_w & (N-1)\sigma_w^2 + N\mu_w^2 \end{bmatrix}$$

A.3.1 Asymptotic approximation of the precision matrix for Model 2

The total number of participants N is equal to $N_0 + N_1$. Assuming that the variable w_i is a dummy variable indicating the group membership $\bar{w} = N_1/N$ converges in probability to the proportion of people in the population that are in the non-reference group and $S_w^2 = \left(\frac{N_1}{N}\right)\left(\frac{N_0}{N}\right)$ converges in probability to the product of the proportion of people in the non-reference and the proportion of people in the reference group.

$$\mathbf{P}_{\hat{\boldsymbol{\beta}}}^{*} = \frac{(T-2)(1-\rho_\epsilon)^2+2(1-\rho_\epsilon)}{\sigma_\epsilon^2(1-\rho_\epsilon^2)+\sigma_{v_0}^2((T-2)(1-\rho_\epsilon)^2+2(1-\rho_\epsilon))}\begin{bmatrix} N & N_1 \\ N_1 & (N-1)\frac{N_1 N_0}{N^2} + \frac{N_1^2}{N} \end{bmatrix}$$

A.4 Precision matrix for Model 3

Via the Woodburry matrix identity and because $\mathbf{X}_i = \mathbf{Z}_i$

$$\begin{aligned}
\mathbf{P}_{\hat{\boldsymbol{\beta}}} &= \sum_{i=1}^{N}\underbrace{\mathbf{X}_i^T\boldsymbol{\Sigma}_{\epsilon_i}^{-1}\mathbf{X}_i}_{=:\mathbf{C}_i} - (\mathbf{X}_i^T\boldsymbol{\Sigma}_{\epsilon_i}^{-1}\mathbf{X}_i)(\boldsymbol{\Sigma}_v^{-1} + \mathbf{X}_i^T\boldsymbol{\Sigma}_{\epsilon_i}^{-1}\mathbf{X}_i)^{-1}(\mathbf{X}_i^T\boldsymbol{\Sigma}_{\epsilon_i}^{-1}\mathbf{X}_i) \\
&\overset{def}{=} \sum_{i=1}^{N}\mathbf{C}_i - \mathbf{C}_i(\boldsymbol{\Sigma}_v^{-1} + \mathbf{C}_i)^{-1}\mathbf{C}_i
\end{aligned}$$

Since we assume that $T_i = T \ \forall \ i \in \{1, \ldots, N\}$ we have that $\mathbf{\Sigma}_{\epsilon_i}^{-1} = \mathbf{\Sigma}_{\epsilon}^{-1} = \sigma_{\epsilon}^2 \mathbf{\Gamma}_{\epsilon} \ \forall \ i \in \{1, \ldots, N\}$. Using *Theorem 2.1* in Sutradhar and Kumar (2003) to invert $\mathbf{\Gamma}_{\epsilon}$ we can thus write $\mathbf{C}_i$ as

$$
\begin{aligned}
\mathbf{C}_i &= \frac{1}{\sigma_{\epsilon}^2(1-\rho_{\epsilon}^2)} \begin{bmatrix} 1 & 1 & \ldots & 1 \\ \tilde{x}_{i1} & \tilde{x}_{i2} & \ldots & \tilde{x}_{iT} \end{bmatrix} \begin{bmatrix} 1 & -\rho_{\epsilon} & 0 & 0 & \ldots & 0 & 0 \\ -\rho_{\epsilon} & 1+\rho_{\epsilon}^2 & -\rho_{\epsilon} & 0 & \ldots & 0 & 0 \\ 0 & -\rho_{\epsilon} & 1+\rho_{\epsilon}^2 & -\rho_{\epsilon} & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1+\rho_{\epsilon}^2 & -\rho_{\epsilon} \\ 0 & 0 & 0 & 0 & \ldots & -\rho_{\epsilon} & 1 \end{bmatrix} \begin{bmatrix} 1 & \tilde{x}_{i1} \\ 1 & \tilde{x}_{i2} \\ \vdots & \vdots \\ 1 & \tilde{x}_{iT} \end{bmatrix} \\
&= \frac{1}{\sigma_{\epsilon}^2(1-\rho_{\epsilon}^2)} \begin{bmatrix} (T-2)(1-\rho_{\epsilon})^2 + 2(1-\rho_{\epsilon}) & \rho(1-\rho_{\epsilon})(\tilde{x}_{i1}+\tilde{x}_{iT}) \\ \rho_{\epsilon}(1-\rho_{\epsilon})(\tilde{x}_{i1}+\tilde{x}_{iT}) & (1+\rho_{\epsilon})^2 \sum_{t=1}^{T} \tilde{x}_{it}^2 - 2\rho_{\epsilon} \sum_{t=1}^{T-1} \tilde{x}_{it}\tilde{x}_{it+1} - \rho_{\epsilon}^2(\tilde{x}_{i1}^2 + \tilde{x}_{iT}^2) \end{bmatrix} \\
&= \frac{1}{\sigma_{\epsilon}^2(1-\rho_{\epsilon}^2)} \begin{bmatrix} (T-2)(1-\rho_{\epsilon})^2 + 2(1-\rho_{\epsilon}) & \rho_{\epsilon}(1-\rho_{\epsilon})(\tilde{x}_{i1}+\tilde{x}_{iT}) \\ \rho_{\epsilon}(1-\rho_{\epsilon})(\tilde{x}_{i1}+\tilde{x}_{iT}) & (1+\rho_{\epsilon})^2(T-1)S_x^2 - 2\rho_{\epsilon}(T-1)\phi_x - \rho_{\epsilon}^2(\tilde{x}_{i1}^2 + \tilde{x}_{iT}^2) \end{bmatrix}
\end{aligned}
$$

with S_x^2 be sample variance of the level 1 variable x_{it}.

A.4.1 Asymptotic Approximation of the Precision Matrix for Model 3

Assuming that both N and T are sufficiently large, S_x^2 converges in probability to $\sigma_{\upsilon}^2 + \sigma_{\tau}^2/(1-\rho_{\tau}^2)$, and ϕ_x converges in probability to $\sigma_{\upsilon}^2 + \rho_{\tau}\sigma_{\tau}^2/(1-\rho_{\tau}^2)$. To show that $\mathbf{C}_i$ converges to a diagonal matrix we write it as $T\left(\frac{1}{T}\mathbf{C}_i\right)$. The bottom-left element of $\left(\frac{1}{T}\mathbf{C}_i\right)$ converges almost surely to zero since

$$
\begin{aligned}
\frac{\tilde{x}_{i1} + \tilde{x}_{iT}}{T} &= \frac{1}{T}\sum_{t=1}^{T} \tilde{x}_{it} - \frac{1}{T}\sum_{t=2}^{T-1} \tilde{x}_{it} \\
&= \frac{1}{T}\sum_{t=1}^{T} \left(\mu_x + \upsilon_i + \tau_{it} - \underbrace{\frac{\sum_{t=1}^{T} x_{it}}{T}}_{\to \mu_x + \upsilon_i} \right) - \frac{1}{T}\sum_{t=2}^{T-1} \left(\mu_x + \upsilon_i + \tau_{it} - \underbrace{\frac{\sum_{t=1}^{T} x_{it}}{T}}_{\to \mu_x + \upsilon_i} \right) \\
&\longrightarrow \frac{1}{T}\sum_{t=1}^{T} \tau_{it} - \frac{1}{T}\sum_{t=2}^{T-1} \tau_{it} = \frac{\tau_{i1}}{T} + \frac{\tau_{iT}}{T} \longrightarrow 0.
\end{aligned}
$$

The top-right element almost surely converges to zero because

$$
\begin{aligned}
\frac{\tilde{x}_{i1}^2 + \tilde{x}_{iT}^2}{T} &= \frac{1}{T}\sum_{t=1}^{T} \tilde{x}_{it}^2 - \frac{1}{T}\sum_{t=2}^{T-1} \tilde{x}_{it}^2 \\
&= \underbrace{\frac{1}{T}\sum_{t=1}^{T} \left(\mu_x + \upsilon_i + \tau_{it} - \frac{\sum_{t=1}^{T} x_{it}}{T} \right)^2}_{\longrightarrow \mathbb{E}\left(\mu_x + \upsilon_i + \tau_{it} - \frac{\sum_{t=1}^{T} x_{it}}{T} \right)^2} - \underbrace{\left(\frac{T-2}{T} \right)}_{\longrightarrow 1} \underbrace{\frac{1}{T-2}\sum_{t=2}^{T-1} \left(\mu_x + \upsilon_i + \tau_{it} - \frac{\sum_{t=1}^{T} x_{it}}{T} \right)^2}_{\longrightarrow \mathbb{E}\left(\mu_x + \upsilon_i + \tau_{it} - \frac{\sum_{t=1}^{T} x_{it}}{T} \right)^2} = 0.
\end{aligned}
$$

since $\mathbb{E}(\mu_x+\upsilon_i+\tau_{it}-\frac{\sum_{t=1}^{T} x_{it}}{T})^2 = (\sigma_\upsilon^2+\sigma_\tau^2/(1-\rho_\tau^2)) < \infty$. For large T we can thus approximate $\mathbf{C}_i$ by a matrix $\mathbf{C}^*$ with the following elements:

$$
\begin{aligned}
c_{11}^* &= \frac{(T-2)(1-\rho_\epsilon)^2 + 2(1-\rho_\epsilon)}{\sigma_\epsilon^2(1-\rho_\epsilon^2)} \\
c_{12}^* &= c_{21}^* = 0 \\
c_{22}^* &= \frac{(1+\rho_\epsilon)^2(T-1)(\sigma_{\upsilon_k}^2 + \sigma_{\tau_k}^2/(1-\rho_{\tau_k}^2)) - 2\rho_\epsilon(T-1)(\sigma_{\upsilon_k}^2 + \rho_{\tau_k}\sigma_{\tau_k}^2/(1-\rho_{\tau_k}^2))}{\sigma_\epsilon^2(1-\rho_\epsilon^2)}
\end{aligned}
\tag{28}
$$

This approximation suffices for large T because the standard error of the effect of interest β_{10} is equal to $|\mathbf{C}_i|^{-1}c_{11} = |\mathbf{C}_i|^{-1}c_{11}^*$ and $\frac{|\mathbf{C}_i|}{|\mathbf{C}^*|} \longrightarrow 1$.

A.5 Asymptotic approximation of the precision matrix for Model 4

Note that Model 4 is a special case of the general multilevel model with P level 1 predictors, where the number of time-varying predictors $P = 1$. In this case, the precision matrix is given by Equation (27). Assuming that both N and T are sufficiently large, $\mathbf{C}_i$ converges in probability to $\mathbf{C}^*$ as defined in Eq. 28, $\bar{w}$ converges in probability to μ_w, and S_w^2 converges in probability to σ_w^2. Thus, the asymptotic approximation of the precision matrix for Model 4 is:

$$
\boldsymbol{I}(\boldsymbol{\beta}) = \begin{bmatrix} N\mathbf{E}^* & N\mu_w\mathbf{E}^* \\ N\mu_w\mathbf{E}^* & ((N-1)\sigma_w^2 + N\mu_w^2)\mathbf{E}^* \end{bmatrix}
$$

where $\mathbf{E}^* = (\mathbf{C}^* - \mathbf{C}^*(\mathbf{I}_2 + \boldsymbol{\Sigma}_\upsilon\mathbf{C}^*)^{-1}\boldsymbol{\Sigma}_\upsilon\mathbf{C}^*)$, with $\mathbf{C}^*$ being the matrix in Eq. 28.

A.6 Asymptotic approximation of the precision matrix for Model 5

Model 5 is also a special case of the general model with P level 1 predictors, where the total number of participants N is equal to $N_0 + N_1$ and w_i is a dummy variable indicating the group membership. Furthermore, we assume that the distribution of the level 1 predictor for participants in the reference group (i.e., group 0) is:

$$x_{0it} = \mu_{0x} + \upsilon_{0i} + \tau_{0it}$$

where μ_{0x} is the mean of the level 1 predictor across all persons of group 0, υ_{0i} is the deviation in the mean of each person in the reference group from μ_{0x} with variance denoted by $\sigma_{\upsilon_0}^2$, and τ_{0it} denotes the within-person error of x_{0it}, which is assumed to follow an AR(1) process with autocorrelation denoted by ρ_{τ_0} and variance $\sigma_{\tau_0}^2$. In this model, the mean of x_{0it} for each person in the reference group is $\mu_{x_{0i}} = \mu_{0x} + \upsilon_{0i}$, the variance of x_{0it} is $\sigma_{\upsilon_0}^2 + \sigma_{\tau_0}^2/(1-\rho_{\tau_0}^2)$, and the correlation between two consecutive measures of the level 1 predictor is $\sigma_{\upsilon_0}^2 + \rho_{\tau_0}\sigma_{\tau_0}^2/(1-\rho_{\tau_0}^2)$. Next, for persons in Group 0, if N_1 and T are sufficiently large, the matrix $\mathbf{C}_i$ for participants in Group 0 converges in probability to $\mathbf{C}_0^*$, with $\mathbf{C}_0^*$ being the matrix in Eq. 28 for persons in the reference group:

$$
\begin{aligned}
c_{0_{11}}^* &= \frac{(T-2)(1-\rho_\epsilon)^2 + 2(1-\rho_\epsilon)}{\sigma_\epsilon^2(1-\rho_\epsilon^2)} \\
c_{0_{12}}^* &= c_{0_{21}}^* = 0 \\
c_{0_{22}}^* &= \frac{(1+\rho_\epsilon)^2(T-1)(\sigma_{\upsilon_0}^2 + \sigma_{\tau_0}^2/(1-\rho_{\tau_0}^2)) - 2\rho_\epsilon(T-1)(\sigma_{\upsilon_0}^2 + \rho_{\tau_0}\sigma_{\tau_0}^2/(1-\rho_{\tau_0}^2))}{\sigma_\epsilon^2(1-\rho_\epsilon^2)}
\end{aligned}
$$

Therefore, $\mathbf{E}_i$ for participants in the reference group converges in probability to $\mathbf{E}_0^* = (\mathbf{C}_0^* - \mathbf{C}_0^*(\mathbf{I}_2 + \boldsymbol{\Sigma}_\upsilon\mathbf{C}_0^*)^{-1}\boldsymbol{\Sigma}_\upsilon\mathbf{C}_0^*)$.

The distribution of the level 1 predictor for participants in the non-reference group (i.e., group 1) is:

$$x_{1it} = \mu_{1x} + \upsilon_{1i} + \tau_{1it}$$

where the mean of x_{1it} for each person in the non-reference group is $\mu_{x_{1i}} = \mu_{1x} + \upsilon_{1i}$, the variance of x_{1it} is $\sigma_{\upsilon_1}^2 + \sigma_{\tau_1}^2/(1-\rho_{\tau_1}^2)$, and the correlation between two consecutive measures of the level 1 predictor is $\sigma_{\upsilon_1}^2 + \rho_{\tau_1}\sigma_{\tau_1}^2/(1-\rho_{\tau_1}^2)$. For persons in Group 1, if N_1 and T are sufficiently large, the matrix $\mathbf{C}_i$ for participants in Group 1 converges in probability to $\mathbf{C}_1^*$, with $\mathbf{C}_1^*$ being the matrix in Eq. 28 for persons in the

non-reference group:

$$c^*_{1,11} = \frac{(T-2)(1-\rho_\epsilon)^2 + 2(1-\rho_\epsilon)}{\sigma^2_\epsilon(1-\rho^2_\epsilon)}$$
$$c^*_{1,12} = a^*_{21} = 0$$
$$c^*_{1,22} = \frac{(1+\rho_\epsilon)^2(T-1)(\sigma^2_{\upsilon_1} + \sigma^2_{\tau_1}/(1-\rho^2_{\tau_1})) - 2\rho_\epsilon(T-1)(\sigma^2_{\upsilon_1} + \rho_{\tau_1}\sigma^2_{\tau_1}/(1-\rho^2_{\tau_1}))}{\sigma^2_\epsilon(1-\rho^2_\epsilon)}$$

Thus, $\mathbf{E}_i$ for participants in the non-reference group converges in probability to $\mathbf{E}^*_1 = (\mathbf{C}^*_1 - \mathbf{C}^*_1(\mathbf{I}_2 + \mathbf{\Sigma}_\nu \mathbf{C}^*_1)^{-1}\mathbf{\Sigma}_\nu \mathbf{C}^*_1)$.

As a result, the asymptotic approximation of the precision matrix for Model 5 can be written as:

$$\boldsymbol{I}(\boldsymbol{\beta}) = \sum_{i=1}^{N} \begin{bmatrix} \mathbf{E}_i & w_i\mathbf{E}_i \\ w_i\mathbf{E}_i & w_i^2\mathbf{E}_i \end{bmatrix}$$
$$\boldsymbol{I}(\boldsymbol{\beta}) = \sum_{i=1}^{N_0} \begin{bmatrix} \mathbf{E}^*_0 & 0 * \mathbf{E}^*_0 \\ 0 * \mathbf{E}^*_0 & 0 * \mathbf{E}^*_0 \end{bmatrix} + \sum_{i=1}^{N_1} \begin{bmatrix} \mathbf{E}^*_1 & \mathbf{E}^*_1 \\ \mathbf{E}^*_1 & \mathbf{E}^*_1 \end{bmatrix}$$
$$\boldsymbol{I}(\boldsymbol{\beta}) = \begin{bmatrix} N_0\mathbf{E}^*_0 + N_1\mathbf{E}^*_1 & N_1\mathbf{E}^*_1 \\ N_1\mathbf{E}^*_1 & N_1\mathbf{E}^*_1 \end{bmatrix}.$$

Acknowledgements We thank Kristof Meers for the assistance in using the HPC cluster. The resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish government.

Author Contributions The authors made the following contributions. Ginette Lafit: Conceptualization, Formal Analysis, Methodology, Software, Visualization, Writing - Original Draft Preparation, Writing -Review & Editing; Richard Artner: Conceptualization, Formal Analysis, Methodology, Validation, Writing - Original Draft Preparation, Writing - Review & Editing; Eva Ceulemans: Conceptualization, Methodology, Funding acquisition, Supervision, Writing - Review & Editing.

Funding The research presented in this article was supported by research grants from the Fund for Scientific Research-Flanders (FWO; Project No. G0C9821N) and from the Research Council of KU Leuven (C14/19/054; iBOF/21/090) awarded to E. Ceulemans.

Data Availability The clinical data are available upon request on the EMOTE database https://emotedatabase.com/datasets/3. The popularity data set is available at https://multilevel-analysis.sites.uu.nl/datasets/

Declarations

Conflict of interest The authors declare that there are no conflicts of interest concerning the authorship or the publication of this article.

Supplemental Material The supplemental material for this paper is available at the OSF repository https://osf.io/8maqf/ and the GitLab repository https://gitlab.kuleuven.be/ppw-okpiv/researchers/u0119584/ApproxPowerIL.

References

Adolf, J.K, Loossens, T., Tuerlinckx, F., & Ceulemans, E (2021). Optimal sampling rates for reliable continuous-time first-order autoregressive and vector autoregressive modeling. *Psychological Methods*

Arend, M. G., & Schäfer, T. (2019). Statistical power in two-level models: A tutorial based on Monte Carlo simulation. *Psychological Methods, 24*(1), 1–19.

Ariens, S., Adolf, J. K., & Ceulemans, E. (2023). One does not simply correct for serial dependence. *Structural Equation Modeling: A Multidisciplinary Journal, 30*(5), 807–821.

Ariens, S., Ceulemans, E., & Adolf, J.K. (2020). Time series analysis of intensive longitudinal data in psychosomatic research: A methodological overview. *Journal of Psychosomatic Research,* 110191

Astivia, O. L. O., Gadermann, A., & Guhn, M. (2019). The relationship between statistical power and predictor distribution in multilevel logistic regression: A simulation-based approach. *BMC Medical Research Methodology, 19*(1), 97–117.

Berkhof, J., & Snijders, T. A. (2001). Variance component testing in multilevel models. *Journal of Educational and Behavioral Statistics, 26*(2), 133–152.

Bolger, N. (2011). Power analysis for intensive longitudinal studies. In N. Bolger, G. Stadler, & J.-P. Laurenceau (Eds.), *Handbook of research methods for studying daily life* (pp. 285–301). New York: Guilford.

Browne, W. J., Lahi, M. G., & Parker, R. M. (2009). *A guide to sample size calculations for random effect models via simulation and the MLPowSim software package*. Bristol, United Kingdom: University of Bristol.

Chang, W., Cheng, J., Allaire, J., Xie, Y., McPherson, J., et al. (2019). Shiny: Web Application Framework for R. *R package version, 1*(3), 2.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

Dejonckheere, E., Mestdagh, M., Houben, M., Erbas, Y., Pe, M., Koval, P., ..., & Kuppens, P. (2018). The bipolarity of affect and depressive symptoms. *Journal of Personality and Social Psychology, 114*(2), 323–341.

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2022). The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *Assessment, 29*(2), 136–151.

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.

Ferron, J., Dailey, R., & Yi, Q. (2002). Effects of misspecifying the first-level error structure in two-level models of change. *Multivariate Behavioral Research, 37*(3), 379–403.

Goldstein, H., Healy, M. J., & Rasbash, J. (1994). Multilevel time series models with applications to repeated measures data. *Statistics in Medicine, 13*(16), 1643–1655.

Green, P., & MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution, 7*(4), 493–498.

Hager, W. W. (1989). Updating the inverse of a matrix. *SIAM Review, 31*(2), 221–239.

Hamilton, J.D. (1994). Time series analysis, volume 2. Princeton New Jersey

Hedeker, D., Gibbons, R. D., & Waternaux, C. (1999). Sample size estimation for longitudinal designs with attrition: Comparing time-related contrasts between two groups. *Journal of Educational and Behavioral Statistics, 24*(1), 70–93.

Heininga, V. E., Dejonckheere, E., Houben, M., Obbels, J., Sienaert, P., Leroy, B., ..., & Kuppens, P. (2019). The dynamical signature of anhedonia in major depressive disorder: Positive emotion dynamics, reactivity, and recovery. *BMC Psychiatry, 19*(1), 59.

Hox, J.J., Moerbeek, M., & Van de Schoot, R. (2017). *Multilevel analysis: Techniques and applications*. Routledge

Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine, 2*(8), e124.

Kalokerinos, E., Russo-Batterham, D., Koval, P., Moeck, E.K.., Shrestha, K.M., Garrett, P., ..., & Kuppens, P. (2022) The EMOTE Database: An open, searchable database of experience sampling data mapping everyday life. unpublished

Kwok, O. M., West, S. G., & Green, S. B. (2007). The impact of misspecifying the within-subject covariance structure in multiwave longitudinal multilevel models: A Monte Carlo study. *Multivariate Behavioral Research, 42*(3), 557–592.

Lafit, G., Adolf, J. K., Dejonckheere, E., Myin-Germeys, I., Viechtbauer, W., & Ceulemans, E. (2021). Selection of the number of participants in intensive longitudinal studies: A user-friendly shiny app and tutorial for performing power analysis in multilevel regression models that account for temporal dependencies. *Advances in Methods and Practices in Psychological Science, 4*(1), 2515245920978738.

Lafit, G., Sels, L., Adolf, J. K., Loeys, T., & Ceulemans, E. (2022). PowerLAPIM: An application to conduct power analysis for linear and quadratic longitudinal actor-partner interdependence models in intensive longitudinal dyadic designs. *Journal of Social and Personal Relationships, 39*(10), 3085–3115.

Laird, N.M. & Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics,* pp. 963–974

Lakens, D. (2022). Sample size justification. Collabra. *Psychology, 8*(1), 33267.

Lane, S. P., & Hennes, E. P. (2018). Power struggles: Estimating sample size for multilevel relationships research. *Journal of Social and Personal Relationships, 35*(1), 7–31.

Lane, S. P., & Hennes, E. P. (2019). Conducting sensitivity analyses to identify and buffer power vulnerabilities in studies examining substance use over time. *Addictive Behaviors, 94*, 117–123.

Ly, A., Marsman, M., Verhagen, J., Grasman, R. P., & Wagenmakers, E. J. (2017). A tutorial on fisher information. *Journal of Mathematical Psychology, 80*, 40–55.

McNeish, D. (2017). Small sample methods for multilevel modeling: A colloquial elucidation of REML and the Kenward-Roger correction. *Multivariate Behavioral Research, 52*(5), 661–670.

Mestdagh, M., & Dejonckheere, E. (2021). Ambulatory assessment in psychopathology research: Current achievements and future ambitions. *Current Opinion in Psychology*

Miller, K. S. (1981). On the Inverse of the Sum of Matrices. *Mathematics Magazine, 54*(2), 67–72.

Moerbeek, M. (2008). Powerful and cost-efficient designs for longitudinal intervention studies with two treatment groups. *Journal of Educational and Behavioral Statistics, 33*(1), 41–61.

Moerbeek, M., & Maas, C. J. (2005). Optimal experimental designs for multilevel logistic models with two binary predictors. *Communications in Statistics-Theory and Methods, 34*(5), 1151–1167.

Moerbeek, M., van Breukelen, G. J., & Berger, M. P. (2000). Design issues for experiments in multilevel populations. *Journal of Educational and Behavioral Statistics, 25*(3), 271–284.

Moerbeek, M., Van Breukelen, G. J., & Berger, M. P. (2001). Optimal experimental designs for multilevel logistic models. *Journal of the Royal Statistical Society: Series D (The Statistician), 50*(1), 17–30.

Murayama, K., Usami, S., & Sakaki, M. (2022). Summary-statistics-based power analysis: A new and practical method to determine sample size for mixed-effects modeling. *Psychological Methods*

Myin-Germeys, I., Kasanova, Z., Vaessen, T., Vachon, H., Kirtley, O., Viechtbauer, W., & Reininghaus, U. (2018). Experience sampling methodology in mental health research: New insights and technical developments. *World Psychiatry, 17*(2), 123–132.

Myin-Germeys, I., Oorschot, M., Collip, D., Lataster, J., Delespaul, P., & van Os, J. (2009). Experience sampling research in psychopathology: Opening the black box of daily life. *Psychological Medicine, 39*(9), 1533–1547.

Myin-Germeys, I., Peeters, F., Havermans, R., Nicolson, N., DeVries, M., Delespaul, P., & Van Os, J. (2003). Emotional reactivity to daily life stress in psychosis and affective disorder: An experience sampling study. *Acta Psychiatrica Scandinavica, 107*(2), 124–131.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., et al. (2019). *Linear and Nonlinear Mixed Effects Models.* R package version 3.1-141

R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna: R Core Team.

Raudenbush, S. W. (1997). Statistical analysis and optimal design for cluster randomized trials. *Psychological Methods, 2*(2), 173–185.

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Sage, Thousand Oaks, CA, 2nd edition

Raudenbush, S. W., & Liu, X. F. (2001). Effects of study duration, frequency of observation, and sample size on power in studies of group differences in polynomial change. *Psychological Methods, 6*(4), 387–401.

Rights, J. D., & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods, 24*(3), 309.

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., Markowitz, J. C., Ninan, P. T., Kornstein, S., Manber, R., et al. (2003). The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry, 54*(5), 573–583.

Snijders, T. A., & Bosker, R. J. (1993). Standard errors and sample sizes for two-level research. *Journal of Educational Statistics, 18*(3), 237–259.

Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. arXiv:1906.02243

Sutradhar, B., & Kumar, P. (2003). The inversion of correlation matrix for MA(1) process. *Applied Mathematics Letters, 16*(3), 317–321.

Szucs, D., & Ioannidis, J. P. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS Biology, 15*(3), e2000797.

Trull, T. J., & Ebner-Priemer, U. W. (2020). Ambulatory assessment in psychopathology research: A review of recommended reporting guidelines and current practices. *Journal of Abnormal Psychology, 129*(1), 56–63.

Wang, C., Hall, C. B., & Kim, M. (2015). A comparison of power analysis methods for evaluating effects of a predictor on slopes in longitudinal designs with missing data. *Statistical Methods in Medical Research, 24*(6), 1009–1029.

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