Research Master’s programme: Methodology and Statistics for the Behavioral, Biomedical, and Social Sciences

Utrecht University, the Netherlands

Sensitivity of the Random Intercept Cross-lagged Panel Model to Unmodeled Measurement Error (Lina Kramer (0629863))

Supervisors:

Prof. Dr. Ellen Hamaker

Jeroen Mulder

Second grader:

Prof. Dr. Irene Klugkist

Preferred journal of publication: Structural Equation Modeling: A Multidisciplinary Journal

Word count: ~7500

Sensitivity of the Random Intercept Cross-lagged Panel Model to unmodeled measurement error

Lina Kramer

Methodology and Statistics

Faculty of Social and Behavioural Sciences

Utrecht University

The Netherlands

Author Note

Lina Kramer https://orcid.org/0000-0002-7282-555X Correspondence concerning this article should be addressed to Lina Kramer, Methodology and Statistics, Utrecht University, Utrecht 3584 CH, The Netherlands. E-mail: l.kramer@uu.nl

Abstract

The random-intercept cross-lagged panel model (RI-CLPM) lets researchers investigate within-relations between variables over time, while accounting for stable between-person differences. However, the RI-CLPM does not account for measurement error. This paper provides researchers with step-by-step instructions for a sensitivity analysis of RI-CLPM results to measurement error. This approach to sensitivity analysis is supported by the newly developed `sensRICLPM` `R` package. This tool has a minimum required input of a panel data set that includes measures from at least three time points for two variables. An example is provided.

In addition, a simulation study is used to illustrate how improperly accounting for measurement error can distort results. This is evaluated in terms of lagged parameter estimates and confidence interval coverage rates of the true effects. When there is measurement error in the data but errors are fixed to 0 (as they are in the RI-CLPM), the autoregressive and cross-lagged parameters are estimated to be closer to 0. Estimates usually move closer to the true value of the parameter when measurement error proportions are fixed to the true data generating proportion of error. When measurement errors are modeled with larger values than the true measurement errors, cross-lagged parameters can be both under- and overestimated.

Keywords: random intercept cross-lagged panel model, stable-trait autoregressive-state model, sensitivity, measurement error

Sensitivity of the Random Intercept Cross-lagged Panel Model to unmodeled measurement error

Longitudinal study designs let researchers investigate relations between variables over time. Cross-lagged panel models are an approach to data where two (or more) variables are measured at multiple time points; earlier measures of each variable may affect later measures of other variables, referred to as *cross-lagged* effects, and early measures of each variable will also have effects on later measures of the same variables, referred to as *autoregressive* effects. For example, in a psychological study a cross-lagged panel model could be used to model the relationships between happiness and sense of purpose over time - while accounting for the initial levels of happiness and sense of purpose. These types of research questions usually address *within*-person processes. For investigations of these processes, it is crucial to use models that distinguish the variability that occurs over time within persons from stable *between*-person variability (e.g., Curran & Bauer, 2011; de Haan-Rietdijk, Kuppens, & Hamaker, 2016; Schuurman, 2023).

One cross-lagged panel model that has gained popularity in investigations of these types of research questions is the random-intercept cross-lagged panel model (RI-CLPM; Hamaker, Kuiper, & Grasman, 2015). In the RI-CLPM, (stable trait-like) between-person differences are captured by random intercepts. This approach is inspired by multilevel thinking (for more explanation and applications of multilevel analysis see, e.g., Hox, Moerbeek, & Van de Schoot, 2017). This is how the RI-CLPM decomposes the within-person dynamics from the (stable trait-like) between-person variance. As a result of this decomposition, the cross-lagged and autoregressive paths in the RI-CLPM become interpretable as within-person dynamics. Extensions of the RI-CLPM exist for the inclusion of multiple indicators, multiple groups, and stable characteristics as predictors and/or outcomes (Mulder & Hamaker, 2021).

This paper addresses the issue that the RI-CLPM does not account for measurement error, and little is known about the sensitivity of its results to unmodeled measurement error. Measurement errors are the differences between the true value of a variable (a construct) and its measure (or indicator) (Miguel, Hernan, & James, 2023). For example, if we ask people to give their happiness a score, this will give us indicators of their happiness, but not the true values thereof. Because the RI-CLPM does not include measurement error, it poses the implicit assumption that all variables are measured perfectly. In reality, there is very likely some measurement error present in most data. When measurement error is present but not accounted for, it can negatively affect the validity of parameter estimates (e.g., Savalei, 2019). Previous research has already studied this in the context of some longitudinal data models. In $n = 1$ autoregressive (AR) models, not accounting for measurement error leads to underestimations of autoregressive parameters, especially when autoregressive parameters are weak (Schuurman, Houtveen, & Hamaker, 2015). For multilevel autoregressive (VAR) models, it had been shown that disregarding measurement error can result in both under- and over-estimations of autoregressive effects (Schuurman & Hamaker, 2019).

In the context of cross-lagged panel models, an approach to including measurement error is provided by the Stable Trait AutoRegressive Trait State (STARTS) model (Kenny & Zautra, 1995, 2001). Like the RI-CLPM, the STARTS model separates the stable between-person differences (the *Stable Trait*) from the within-person dynamics (the *AutoRegressive Trait*). Additionally, the STARTS model estimates *State* fluctuations in the observed variables, which are interpretable as measurement errors. The RI-CLPM can be understood as a special case of the bivariate STARTS model, in which *State* components (measurement errors) are fixed to 0 (Usami, Murayama, & Hamaker, 2019; Hamaker et al., 2015). Even though the STARTS model can be used to address the same research questions as the RI-CLPM, while additionally accounting for measurement error, researchers often use the RI-CLPM. One reason for this is that the RI-CLPM can be used for data with fewer measurement waves (time points). This distinction can be explained in terms of statistical and empirical identification of the two models. The RI-CLPM is empirically identified and generally does not result in many convergence issues when there are at least 3 waves (Orth, Clark, Donnellan, & Robins, 2021). To be statistically identified, a STARTS model requires 4 waves of data. Moreover, despite being statistically identified, the STARTS model is often not empirically identified when a study has less than 8 measurement waves, especially when sample sizes are small (Cole, Martin, & Steiger, 2005). A partial solution for a more stable estimation of the STARTS, using a Bayesian approach with weakly informative priors (that push potential parameter estimates into the admissible range) has been proposed by Lüdtke, Robitzsch, and Wagner (2017). Furthermore, Lüdtke, Robitzsch, and Ulitzsch (2022) showed that for small sample sizes and as few as 4 measurement waves, using Bayesian approaches or bounded ML-estimation can improve the accuracy of parameter estimates. However, these approaches are still limited to at least 4 waves of data. This is because the number of parameters we can estimate corresponds to the amount of information in the data. The arguably best solution would be to measure more time points, which would make the estimation of measurement errors possible. Naturally, this is often not feasible in practice and researchers will then prefer to use models that do not include measurement error - such as the RI-CLPM.

It is unclear how the RI-CLPM performs under conditions of measurement error, and, to date, there is no tool available to assess the sensitivity of results obtained with the RI-CLPM to potential measurement error. The main objective of this project is to provide researchers with a tool for sensitivity analysis of their own RI-CLPM results to potential measurement error. For this purpose, I first introduce an example that will be used throughout this paper. Second, I give a brief overview of the RI-CLPM, its link to the STARTS model, and their variance decomposition into distinct sources. Third, a strategy for sensitivity analysis is presented step by step. This includes instructions for the newly developed sensitivity analysis tool: the `sensRICLPM` package.[1] An exemplary sensitivity analysis is conducted for the example. Fourth, a simulation study is used to illustrate how improperly accounting for measurement error can distort results. This is evaluated in terms of lagged parameter estimates and confidence interval coverage rates

[1] This package is a preliminary version.

of the true effects. Lastly, results and limitations are discussed.

The current preliminary version of the package can be downloaded from GitHub here. It can be used to conduct a sensitivity analysis for RI-CLPM results. `R` code to reproduce all present data sets and analyses is available here. The current project was approved by the FETC.

1. Illustrative example

The following example is used throughout this paper to illustrate the application of the presented sensitivity analysis approach. It is a fictive study scenario introduced by Mulder (2022).

Imagine we want to study the relationships between self-alienation (the feeling that one does not know oneself) and academic amotivation (a lack of intrinsic or extrinsic motivation for pursuing academic goals) in college students over time. Self-alienation is going to be our variable x and amotivation is going to be our variable y. We have observational data for three time points of $N = 1000$ students and want to use the RI-CLPM to estimate the effects while controlling for stable trait-like differences. We assume we satisfy RI-CLPM assumptions, meaning the effects are linear, homogeneous, not modified by other factors, not affected by unobserved confounding, and errors are normally distributed (Gische & Voelkle, 2022).[2]

We obtain RI-CLPM estimates for our data set, and conclude that there is an effect of self-alientation on amotivation and an effect of amotivation of self-alienation. There are also spill-over (autoregressive) effects of both variables on themselves from time points $t - 1$ to t. However, we have reason to believe that there is some measurement error in the measurement of both of these variables. Now, we are unsure how this may affect our results. Would the results be different if we accounted for measurement error?

[2] A data set was generated for the illustrative example. Data-generating values: $\alpha = 0.4$, $\beta = 0.2$, $\delta = 0.3$, $\gamma = 0.2$; 30 % measurement error in x and y, $N = 2000$, RI_x and $RI_y = 1$, $cov(RI_{xy}) = 0.5$, W_x and $W_y = 1$ $W_{cov,1} = 0.3$, 3 waves.

2. The RI-CLPM, the STARTS model, and measurement error

Figure 1 provides a graphical representation of a bivariate RI-CLPM and a bivariate STARTS model for three waves of data.

The RI-CLPM seen in Figure 1a decomposes the variance of the observed variables X_{it} and Y_{it} of person i at time point t into stable between-person variance and dynamic within-person variance. The observed scores in an RI-CLPM can be expressed as

$$\begin{aligned} X_{it} &= \mu_{xt} + RI_{xi} + W_{xit} \\ Y_{it} &= \mu_{yt} + RI_{yi} + W_{yit}, \end{aligned} \tag{1}$$

where X_{it} and Y_{it} are the observed scores of person i at time point t in the variables x and y, μ_{xt} and μ_{yt} indicate the temporal group means in x and y at time point t; RI_{xi} and RI_{yi} denote the time-invariant stable trait factors (person i's trait like deviation from the temporal group means of x and y), and W_{xit} and W_{yit} are person-specific, temporal deviations from the expected scores of person i at time point t.[3]

The individual temporal deviations W_{xit} and W_{yit} are modeled as

$$\begin{aligned} W_{xit} &= \alpha_t W_{xi,t-1} + \beta_t W_{yi,t-1} + \mu_{xit} \\ W_{yit} &= \delta_t W_{yi,t-1} + \gamma_t W_{xi,t-1} + \mu_{yit}, \end{aligned} \tag{2}$$

where α_t and δ_t are the autoregressive parameters, which are interpreted as the within-person "carry-over" effect of a person's deviation from their expected score at an earlier time point $t-1$ on one variable on the person's current deviation from their expected score at time point t on the *same* variable. The cross-lagged parameters β_t and γ_t represent the within-person effect of a person's deviation from their expected score at an earlier time point $t-1$ on one variable on the person's current deviation from their expected score at time point t on the *other* variable, when controlling for the autoregressive effect.[4] μ_{xit} and μ_{yit} represent individuals' temporal deviations from the

[3] Note that i can also be used to denote any kind of cluster (e.g., families, groups, classes,...) instead of individuals.

[4] Note that the regression parameters are free to vary over waves here, but they can also be

time-varying group means.

The STARTS model (Figure 1b) additionally separates out measurement error, which is captured by the wave- and item-specific latent variables ϵ_{xit} and ϵ_{yit}.

Including measurement error in Equation (1) leads to the bivariate STARTS model, which can be expressed as

$$\begin{aligned} X_{it} &= \mu_{xt} + RI_{xi}{}^{*} + W_{xit}{}^{*} + \epsilon_{xit} \\ Y_{it} &= \mu_{yt} + RI_{yi}{}^{*} + W_{yit}{}^{*} + \epsilon_{yit}, \end{aligned} \tag{3}$$

where X_{it} and Y_{it} again represent the observed scores of person i at time point t on two variables x and y, and μ_{xt} and μ_{yt} denote the temporal group means, $RI_{xi}{}^{*}$ and $RI_{yi}{}^{*}$ denote individual trait like deviation from the temporal group means of x and y, and $W_{xit}{}^{*}$ and $W_{yit}{}^{*}$ are the individual temporal deviations. ϵ_{xit} and ϵ_{yit} represent the measurement errors on variables x and y for individual i at time point t. The individual temporal deviations can be expressed in terms of lagged parameters as

$$\begin{aligned} W_{xit}{}^{*} &= \alpha_t{}^{*} W_{xi,t-1}{}^{*} + \beta_t{}^{*} W_{yi,t-1}{}^{*} + \mu_{xit}{}^{*} \\ W_{yit}{}^{*} &= \delta_t{}^{*} W_{yi,t-1}{}^{*} + \gamma_t{}^{*} W_{xi,t-1}{}^{*} + \mu_{yit}{}^{*}. \end{aligned} \tag{4}$$

The parameters in the STARTS are different from the within-components in the RI-CLPM, as indicated by the *.

The RI-CLPM is a special case of the bivariate STARTS model with measurement error variances constrained to 0 at all waves. This constraint implies the assumption that there is no measurement error in the data that we model with the RI-CLPM. In practice, we cannot be certain that this assumption is true.

3. Sensitivity Analysis Strategy

A sensitivity analysis examines how various sources of uncertainty in a model's input can contribute to uncertainty in its output (Saltelli, 2002). In this case, the goal is to evaluate how measurement error as a source of uncertainty may affect RI-CLPM results. In a RI-CLPM, the measurement errors are omitted. For the sensitivity analysis, we can use the bivariate STARTS model and include measurement errors with

constrained to be equal.

fixed parameter values. The measurement error parameter values that can be fixed to specific values are the measurement error variances (indicated in Figure 1(b) by the double-headed arrows on the ϵs). By fixing these variances to different values, it is possible to see how estimates for the other model parameters (e.g., the lagged effects) change. This allows an assessment of whether the inferences drawn from results obtained with the RI-CLPM would change if there was measurement error and we accounted for it. This approach strongly resembles the phantom variable approach to sensitivity analysis for structural equation models introduced by Harring, McNeish, and Hancock (2017).

The approach to sensitivity analysis presented here can be summarized as follows: a STARTS model is fitted repeatedly, with different choices for the fixed amount of measurement error variance, and results are compared to the original RI-CLPM results.

I simulated one singular data set that corresponds to the self-alienation and amotivation example, which is used to explain the approach step by step. This includes instructions for the use of the `sensRICLPM` *R* function, which implements the more cumbersome steps for the user.

The sensitivity analysis approach presented here can be summarized in 5 steps:

1. Decide which measurement error proportions should be included in the sensitivity analysis.
2. Decide whether measurement error proportions or error variances should be stable over time.
3. Compute values for the error variances that correspond to these proportions.
4. Estimate the STARTS with error variances fixed to these values to obtain estimates for parameters of interest.
5. Assess results.

Step 0: required data format

A sensitivity analysis of RI-CLPM results to unmodeled measurement error can be conducted with any panel data suited for analysis with the RI-CLPM. There need to be at least three measurement waves of two variables.

The `sensRICLPM` function has a minimum required input of a data set. This data set needs to be in a format where all observations of x are in the first half of columns and all observations of y are in the second half of columns (e.g., the columns are $selfalienation_{t=1}$, $selfalienation_{t=2}$, $selfalienation_{t=3}$, $amotviation_{t=1}$, $amotviation_{t=2}$, $amotviation_{t=3}$). Each row then contains individual observations. A data set can be given to the function by typing:

```
sensRICLPM(data).
```

Step 1: Decide measurement error proportions

The first step is to decide in which increments measurement error will be included in the sensitivity analysis. This can be approached in terms of the *proportion* of measurement error variance in the total observed variance. The proportion of measurement error in a variable can be anything between 0 and 1, where 0 means there is no measurement error and 1 means all variation in the variable is due to measurement error.

The `sensRICLPM` function by default increases proportions in .1 steps from 0 to 0.9 for both observed variables x and y. It includes all possible combinations of these proportions. This is because the two variables do not necessarily contain the same amounts of error.

For our example, we think that both self-alienation and amotivation could contain anything between 0 and .8 proportions of measurement error. We therefore provide input for the measurement error proportion options `ME_prop_x` and `ME_prop_y` by typing:

```
sensRICLPM(data, ME_prop_x = c(0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6,
```

`0.7, 0.8), ME_prop_y = c(0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8)).`

Step 2: Decide whether measurement error proportions or error variances should be stable over time

Measurement error may not be of the same extent over waves. However, in this approach, we assume that at least either the proportions of measurement errors or the error variances are stable over waves. It may be that over waves, the observed variances are very similar at each time point. Then it will not make a difference if we keep the measurement error proportions or the measurement error variances equal over time. However, observed variances can also be grossly different over waves. It could be that the changes in total variance are due to true variation (e.g., less variance at time point t, because it is summer and many people feel less self-alienation when it is sunny). In this case, we could assume that the absolute measurement error variance at this time point is still the same (i.e., there is no association between people feeling more or less self-alienated and errors). On the other hand, there could be differences in observed variances that are more systematic and that could affect measurement error variances. An example of this could be issues with the administration of a test, which results in systematic differences in the observed scores variances over waves. In this case, it may be more sensible to keep measurement error proportions stable over waves.

The sensRICLPM by default keeps the proportion of measurement error variance (relative to observed variance) stable over time. This can be changed by including `equal_ME = TRUE`.

For our example, we decide to keep measurement error proportions stable over time, so that the absolute measurement error variances can be different for each wave.

Step 3: Compute values for the error variances

After deciding on the proportions of measurement error that will be included in the sensitivity analysis, the error variances that correspond to these proportions need to be computed. The proportion of measurement error variance are values between 0 and 1. We can make use of the expression of the proportion of measurement error as

$$p_{var(E_t)} = \frac{var(E_t)}{var(O_t)}, \tag{5}$$

where $p_{var(E_t)}$ denotes the proportion of measurement error variance $var(E_t)$ in the total observed variance $var(O_t)$ at time point t. The total observed variance can be decomposed into stable between (or trait) variance, within (or state) variance, and measurement error variance. For any proportion of measurement error, solving for $var(E_t)$ will provide the value that the measurement error variance needs to be fixed to to reach this proportion in the observed variance. Solving for $var(E_t)$ gives

$$var(E_t) = p_{var(E_t)} \times var(O_t). \tag{6}$$

For example, if the proportion of measurement error $p_{var(E_x)}$ in the variable self-alienation (x) is assumed to be .3 (for all time points), and the total observed variances for three measurement waves are $var(O_x) = \{2, 2.1, 1.9\}$, the measurement error variances for each time point t can be computed as

$$\begin{bmatrix} var(E_{x,1}) \\ var(E_{x,2}) \\ var(E_{x,3}) \end{bmatrix} = 0.3 \times \begin{bmatrix} 2.1 \\ 1.9 \\ 2.3 \end{bmatrix} = \begin{bmatrix} 0.63 \\ 0.57 \\ 0.69 \end{bmatrix}. \tag{7}$$

This computation needs to be done separately for the observed variances of both variables x and y at all measured time points, for all proportions of measurement error chosen in the previous steps. If measurement error variances are required to be equal over waves, they can be based on total observed variances averaged over waves. For the above example, this means the measurement error variances would be computed as

$$\begin{bmatrix} var(E_{x,1}) \\ var(E_{x,2}) \\ var(E_{x,3}) \end{bmatrix} = 0.3 \times \begin{bmatrix} \frac{2.1+1.9+2.3}{3} \\ \frac{2.1+1.9+2.3}{3} \\ \frac{2.1+1.9+2.3}{3} \end{bmatrix} = \begin{bmatrix} 0.63 \\ 0.63 \\ 0.63 \end{bmatrix}, \tag{8}$$

so that the measurement error variances for self-alienation would be fixed to the same value for each of the three waves.

The `sensRICLPM` function automatically computes the measurement error variances based on the observed variances in the input data set and the proportions that were chosen by the user. This step does not require any input from the user.

For our example, this means the values that measurement errors need to be fixed to (for proportions from 0 to .8) are computed from the observed variances in self-alienation and amotivation for each wave.

Step 4: Estimate the STARTS with error variances fixed

The next step is to run the STARTS model with error variances fixed to the values determined in the previous step, for each of the measurement error proportions. For this step, the `sensRICLPM` function creates `lavaan` model syntax (for each different measurement error proportion) and then fits the data set to the different models. This step does not require any input from the user.

Step 5: Assess results

The results of the sensitivity analysis can be visually assessed by plotting the parameter estimates of interest for each model that was run; and by comparing the point estimates of the parameters for each model to the RI-CLPM results. By comparing the RI-CLPM estimates (the case where measurement errors are fixed to 0) to models with increasing amounts of measurement error, it is possible to assess how (much) our conclusions would change - in terms of direction and magnitude - if there is measurement error and it is accounted for.

Sensitivity plots

The `sensRICLPM` function creates plots for the change in point estimates of the lagged parameters. Figure 2 shows a plot created for the amotivation and self-alienation data set. It depicts the estimates of the lagged parameters (autoregressive: $\hat{\alpha}$ and $\hat{\delta}$; cross-lagged: $\hat{\beta}$ and $\hat{\gamma}$)[5] against the (accounted for) proportions of measurement error in x on the x-axis and the proportion of measurement error in y as separate lines. The two separate rows are for the lagged estimates of $\hat{\alpha}_2$, $\hat{\beta}_2$, $\hat{\gamma}_2$, and $\hat{\delta}_2$ for the effects between

[5] Plot columns are "alpha", "delta", "beta", and "gamma".

waves one and two in the first row, and $\hat{\alpha_3}$, $\hat{\beta_3}$, $\hat{\gamma_3}$, and $\hat{\delta_3}$ for the effects between waves two and three in the second row. In this plot, the original RI-CLPM estimate is the one where we account for 0 proportions of measurement in both x and y. This is the leftmost red point in each sub-plot. In addition, we note that the x-axis does not extend beyond a measurement error proportion in x of .4, and some lines end before the other lines. This is because models did not converge to admissible solutions for those combinations of measurement errors.[6]

Zooming in to assess the sensitivity of one lagged parameter, the autoregressive parameter δ_2, we check the column named "delta" in the first row. This is the autoregressive effect of amotivation at time point one on amotivation at time point two. We see that the RI-CLPM estimate of this parameter is approximately $\hat{\delta_2} \approx 0.1$. If the proportion of measurement error is 0 in self-alienation (x) and .2 in amotivation (y), the parameter is estimated to be $\hat{\delta_2} \approx 0.18$. This can be seen by looking at the point on the green line (which includes all models in which the measurement error proportion in amotivation is .2) at which the x-axis is 0. If we follow this line to the right, we see that if the proportion of measurement error is .2 in both x and y, the estimate moves to $\hat{\delta_2} \approx 0.2$. We can see that with increasing (total) amounts of measurement error accounted for, the estimate of $\hat{\delta_2}$ moves further away from 0. This means that if there is measurement error in the data and we account for this, the estimate of the autoregressive effect of amotivation is larger than the RI-CLPM estimate.

By looking at the remaining plots in same way, we see that the RI-CLPM estimates all the autoregressive and cross-lagged parameters closer to 0 than the models that account for measurement error. In this example, the lagged estimates become stronger if there is measurement error and we account for this.

Sensitivity table

Beyond the lagged effects, Table 1 shows the sensitivity analysis results for all parameters. Included in the table are the parameter estimates and 95% confidence intervals (CIs) for those models in which the fixed proportions of measurement error in

[6] Admissibility issues are further addressed in the simulation study in section 4 of this paper.

the total observed variances of x and y were equal. Hence, it includes results for a model in which the fixed proportion of measurement error was 0 (i.e., the RI-CLPM), and for models with .1, .2, and .3 measurement error (respectively in x and y). These estimates can also be found the Figure 6, at those points in which the colored lines and the x-axis take on the same values. Table 1 does not include models with proportions higher than .3 in both x and y because those models did not converge to admissible solutions.

Included in this table is a column with all true parameter values, which are known in this example but would not be known in real empirical data. Analogous to the sensitivity plots, we see that lagged parameter estimates become stronger if there is measurement error and we account for it. Furthermore, we can see (for example) that the $95\% - CI$s for the estimates of the autoregressive effect of amotivation at time point two on amotivation at time point three ($\hat{\delta_3}$) include 0 for all models. This could imply that the conclusion we draw from the RI-CLPM estimate of $\hat{\delta_3}$ would not change if there is measurement error and we account for it. Regardless of if we account for measurement error, this autoregressive effect is estimated to be around 0. However, we also see that the CIs become wider the more measurement error we account for. This means the CIs should probably be interpreted with caution.

Additionally, in this table, we see that the variances of the random intercepts and the (residual) within variances are estimated to be smaller when more measurement error is accounted for. This makes sense because more of the observed variance is "taken up" by the measurement error variances.

We can now answer our initial question: Assuming there is measurement error in the data, would our inferences change if we accounted for it? From this sensitivity analysis, we conclude that the effects of self-alienation and amotivation would still be found if we were to account for measurement error. Moreover, we can conclude that effects should be at least as large as the RI-CLPM estimates them to be. This is because, in this example, the estimates from models that do include measurement error are of the same direction and of higher magnitude than the ones obtained with the RI-CLPM.

4. Simulation study: sensitivity of parameters

A simulation study was conducted to explore the sensitivity of lagged estimates obtained with the RI-CLPM to improperly modeled measurement errors. Using the strategy described in the previous section, the sensitivity was assessed across a wide range of conditions. For the simulation study, I generated data for several conditions and then conducted a sensitivity analysis with different choices for the proportions of measurement error on each data set. Results are evaluated in terms of bias in the lagged parameter estimates and confidence interval coverage rates of the true effects. Additionally, admissibility rates are discussed.

4.1 Data generation

Data were generated under the STARTS model in `lavaan`. Experimental conditions varied by sample size, number of time points, true cross-lagged parameter values, the proportion of between relative to within variance, and the true proportion of measurement error in the data (see Table 2). The true proportion of measurement error in the data varied from 0 to 0.6, meaning that measurement error was the source of up to 60 % of the total observed variance. All possible combinations of the simulation factors resulted in 144 conditions. Each condition represents a combination of values for each factor that influences estimates in the RI-CLPM. For each experimental condition, 100 data sets were generated from the STARTS model. This means that in one condition of this study, the population values are: 0.2 proportion of measurement error, a sample size of $N = 300$, 3 measurement waves, the true cross-lagged effects -.2 and -.2, and a proportion of 70:30 between to within variance. For this condition, 100 data sets were simulated.[7] Over the 144 conditions, this results in a total number of 14400 data sets.

For each of these data sets, a sensitivity analysis was conducted. In these sensitivity analyses, four fixed proportions of measurement errors were included: 0, .2, .4, and .6. The measurement error proportion accounted for in the included models

[7] The number of replications was capped at 100 due to computational limitations. The simulation should be repeated with a higher number of replications before publication.

were kept equal between the variables x and y (i.e., if the fixed proportion of error was .4 for x, it was fixed to the same value for y)

The results were evaluated in terms of point estimates of the lagged parameters, $95\% - CI$ coverage rates of the true value, and admissibility rates. Admissible results refer to only those results that do not contain any negative variance estimates.

4.2 Results

Point estimates and $95\% - CI$ *true-value coverage rates*

Figures 3 to 5 show simulation results for conditions in terms of the average point estimates and $95\% - CI$ true-value coverage rates for the autoregressive parameter α_2 and the cross-lagged parameter β_2 over 100 replications. Only admissible results are included here. Similarly to the previous sensitivity plots, results are plotted against the *accounted for* measurement error proportion. These plots have separate lines for the *true* proportions of measurement error in the data (with one line each for 0, .2, .4, .6). Each sub-plot includes information about the other simulation factor values. Sample size, the number of measurement waves, and true cross-lagged effects are represented in columns. The three factor levels for the ratio of between to within variance are represented in the grid rows. In these plots, the number of results that were used to compute the average point estimate are included for all those cases in which the fixed proportion of measurement error was larger than the true proportion in the data.

Autoregressive parameter. In Figure 3 we see the average point estimates (a) and $95\% - CI$ true value coverage rates (b) for the autoregressive parameter α_2. The true value for α_2 was 0.4. In the point estimate plots, the true value is represented by the grey horizontal lines.

The following is an example of how to read one of the sub-plots: Zooming in, the upper left plot in Figure 3(a) includes average estimates of α_2. For all levels of the simulation factor *true proportion of measurement error* (0, .2, .4, .6), it plots the average estimates of α_2. This is done for the conditions in which the other true simulation factor values were: measurement waves = 3, true cross-lagged effects = $[\beta = 0.2, \gamma = 0]$, ratio of between to within variance = 70:30, and sample size $N = 300$.

In this plot, the average point estimates that are obtained with an RI-CLPM are where we account for no measurement error (0 on the x-axis). The separate line that is solid represents the condition in which there was no (0) *true* measurement error variance in the data. We see that fixing the measurement error variance to 0 for 100 of these kinds of data sets results in an average estimate of α_2 that is very close to the true value of 0.4; these are the results obtained with an RI-CLPM. Accounting for higher measurement error proportions than the true proportions makes the line move away from the true autoregressive effect. The small number next to this line (28) denotes how many of the 100 replications with 0 measurement error were able to estimate a model while accounting for measurement error proportions of .2. The other 72 replications did not converge to admissible results, which seems to happen in most cases in which we attempt to account for more measurement error than there is in the data. From the solid line, we can conclude that, in this specific condition, when there is no measurement error in the sample data, the estimates of the autoregressive effects obtained with the RI-CLPM are close to the true value. If we over-correct for measurement error (i.e., include it even though it is not there), we obtain more biased estimates of the autoregressive effect.

Similarly, the dotted line represents the condition in which the *true* proportion of measurement error was .2. We see that the average estimate of α_2 is closer to 0 when measurement error is not accounted for (i.e., fixed to 0). The dotted line moves closer to the true autoregressive effect when the proportion of measurement error that we account for gets closer to the true measurement error proportion. Again, the small number next to this line (7) denotes the number of replications in the condition with .2 measurement error that were able to estimate the model where we account for measurement error proportions of .4. From the dotted line in this plot we can conclude that, for this condition, when there is some measurement error in the sample data and we analyze it with an RI-CLPM, the autoregressive effect is underestimated, (i.e., estimated closer to 0). If we over-correct for measurement error (i.e., account for proportions of measurement error larger than .2), obtain more biased estimates of the autoregressive effect; although most typically, in this condition, we would obtain inadmissible results.

For the same conditions from the above example, we can additionally check the $95\% - CI$ coverage rates of the true value of α_2 in the upper left sub-plot in Figure 3(b). On the y-axis is the rate out of all replications at which the true effect is covered by the confidence intervals. For instance, the condition with no measurement error is again represented by the solid line, and when the measurement error variance is fixed to 0 (i.e., the model is estimated with a RI-CLPM), the true-value coverage rate is .95. We can see that the true value is covered slightly less often when the measurement error proportion is fixed to a higher value than 0. For all lines in this plot, we see that coverage rates become very high (close to 1) when measurement error variances are fixed to high values. This primarily shows that confidence intervals become very wide (and standard errors large) when the fixed measurement error variances are large, which implies that confidence intervals obtained in three present sensitivity analysis approach (such as in Table 2 should not be overinterpreted.

Zooming out, the plots in Figure 3(a) are organized in a grid, with rows and columns representing levels of the simulation factors from Table 2. The rows include the three different ratios of between to within variance. The columns include the number of waves, the cross-lagged parameters, and the sample size. Each sub-plot contains four of the 144 simulation conditions.

For each of the conditions, these plots show that when there is measurement error in the data, but measurement error variances are fixed to 0 (as they are in the RI-CLPM), the autoregressive parameters are underestimated (estimated closer to 0). Estimates usually move closer to the true value of the parameter when the accounted-for measurement error proportions are equal to the true data-generating proportion of error. When we account for more measurement error than the true value, the autoregressive parameters can be both under- and over-estimated. For instance, in those cases where there is no measurement error present in the data (solid lines), including measurement error in the model results in estimates of the autoregressive paths that are farther from the true values. An example of such an overestimation of the autoregressive effect can be seen in the top leftmost plot in Figure 3(a), where the solid black line indicates a stronger effect (farther from 0) when the model accounts for a .2 proportion of measurement error, compared to accounting for no measurement error. An example of underestimation of the autoregressive effect can be seen in the second plot from the left in the second row of Figure 3. Here, the solid black line indicates a weaker effect (closer to 0) when the model accounts for a .2 proportion of measurement error, compared to accounting for no measurement error.

Cross-lagged parameter. For the cross-lagged parameter β_2, there are two separate figures. Figure 4 includes the results for the cross-lagged parameter β_2 for those conditions in which the true value of β_2 was 0.2. Figure 5 depicts the results for those conditions in which the true value of β_2 was -0.2. In both figures, the true values are again represented by horizontal lines in the (a) plots of the average point estimates. $95\% - CI$ coverage rates for the true values are again included in the (b) plots.

For the cross-lagged parameters, modeling the measurement errors to be 0 (as is done in the RI-CLPM) when they are not really 0 results in estimates of the parameters that are closer to 0 than the true values. This appears to be the case for cross-lagged parameters whose true values are negative and positive. When measurement errors are modeled with larger values than the true measurement errors, cross-lagged parameters can be both under- and overestimated. This corresponds to the conclusions drawn for the autoregressive effect.

In addition, we see that the $95\% - CI$ coverage rates, here for both of the true values of β_2, are again very high. This points towards wide CIs and large standard errors in the models with measurement error variances fixed to large values.

For researchers wanting to conduct a sensitivity analysis for RI-CLPM results to potential unmodeled measurement error, these simulation results may provide an insight into how parameters usually behave under different conditions of measurement error. Furthermore, the simulation study shows that the RI-CLPM will probably not falsely inflate autoregressive and/or cross-lagged effects by not accounting for measurement error. It may however estimate the effects to be closer to 0 than their true value if there is unmodeled measurement error. Moreover, the simulations study demonstrates that confidence intervals computed for parameters in models with large fixed error variances will be very wide, and should therefore be interpreted with caution in the sensitivity analysis.

Admissible results

In the sensitivity plots in Figure 2 and also in the simulation results in Figures 3-5, we see that the lines most often end before we account for the maximum amount of measurement error (here 0.8) that we decided to include in the sensitivity analysis. This is because no admissible results were obtained for these proportions. Software such as Mplus and lavaan mark results as inadmissible when negative variance estimates (also called Heywood cases) occur.

Figure 6 depicts the admissibility rates of results obtained for all conditions over 100 replications. The x-axis represents the accounted-for proportion of measurement error. The true proportions of measurement error are represented by different shades of grey. The simulation factor levels are represented in the rows and columns of the large plot. The ratio of between to within variance and the sample sizes of the conditions are indicated in the rows, and the number of measurement waves and the cross-lagged effects are included in the columns.

Zooming into the sub-plot in the leftmost lower corner, we can evaluate the results in terms of admissibility rates for four of the 144 conditions. For all levels of the simulation factor *true proportion of measurement error* (0, .2, .4, .6), the plot shows the number of inadmissible results out of 100 replications. The frour conditions included in this sub-plot are those where the other true simulation factor values were: measurement waves = 3, true cross-lagged effects = $[\beta = 0.2, \gamma = 0]$, ratio of between to within variance = 70:30, and sample size $N = 1000$. For these cases, none of the models that accounted for 0 or .2 proportions of measurement error resulted in inadmissible estimates. This means that for these two conditions, none of the models obtained negative variance estimates. Furthermore, we can see that for those data sets in which there was a true .6 proportion of measurement error (dark grey), the models start obtaining inadmissible estimates at higher accounted-for proportions of measurement error.

Overall, when models attempt to account for much more measurement error than there is in the data, this *usually* results in a heavy spike in the number of inadmissible results. This means that when the error variance is fixed to a high value, more Heywood cases occur: some of the within and/or random intercept variances are estimated to be negative. However, the spikes in inadmissible results appear to be less drastic in those conditions with smaller sample sizes and fewer measurement waves.

For researchers conducting sensitivity analyses of RI-CLPM results to unmodeled measurement error, this may offer some explanation for why some models in the sensitivity analysis do not converge to admissible solutions. This also illustrates why lines in the sensitivity plots break off before the modeled measurement error proportion is at the user-specified maximum (as seen in Figure 2.)

5. Discussion

So far, there was no cohesive guide for applied researchers who wanted to know how sensitive the results they obtain with the RI-CLPM are to unmodeled measurement error. In this paper, a step-wise strategy for sensitivity analysis of RI-CLPM results to measurement error was introduced. This strategy is supported by the `sensRICLM` package. Applied researchers can use this function to conduct their own sensitivity analysis of RI-CLPM results to unmodeled measurement error, without the requirement of vast statistical programming skills. This allows users to simply input a panel data set (with at least three waves of two variables), so that they can obtain sensitivity plots for lagged parameters, as well as tables with estimates obtained by fixing measurement error variances to increasingly large values. This allows an assessment of how inferences drawn from an RI-CLPM may be sensitive to unmodelled measurement error, in terms of direction and magnitude of estimates.

A simulation study showed how improperly accounting for measurement error can distort results. When there is measurement error in the data but errors are omitted from the model (as they are in the RI-CLPM), autoregressive and cross-lagged parameters are estimated closer to 0. When we over-account for measurement error, the lagged parameters can be both under- and overestimated. These results are consistent with previous research on the effects of disregarding measurement error in the context of autoregressive models (Schuurman et al., 2015; Schuurman & Hamaker, 2019). In addition, the conclusions from the simulation study in this paper extend these findings to cross-lagged parameters. For applied researchers, these results should facilitate the interpretation of their own sensitivity analysis results.

Based on the results of the simulation study for the true value coverage rates of the $95\% - CI$s, researchers are advised to interpret the confidence intervals for parameters that are obtained in a sensitivity analysis with caution. The confidence intervals may be too wide to allow much inference.

To further assist researchers in the interpretation of sensitivity analysis results, the simulation study was used to show where and when inadmissible solutions may be found in the analysis. Especially when sample sizes are small and there are only a few measurement time points, there may be fewer (combinations of) proportions of measurement error for which researchers will be able to obtain and compare results.

Limitations

The current version of the `sensRICLPM` package cannot yet handle missing data. Most (psychological) studies are affected by missingness, and handling it inadequately can be detrimental (e.g., Little, Jorgensen, Lang, & Moore, 2014). This issue still needs to be addressed in future advancements of the presented approach.

A further limitation of the presented package is that it does not allow different proportions of measurement error at different time points. In steps 2 and 3 of the sensitivity analysis strategy, researchers chose different (combinations of) proportions of measurement error that they want to include for all measured time points of two variables x and y. This approach was chosen because it seems theoretically sound that one variable x would have about the same amount of error each time it is measured. Plus, this approach confines the complexity of the sensitivity analysis. However, the proportion of measurement error in x could also, for example, be .1 at time point one,

.4 at time point two, and .2 at time point three. Beyond this, so far we have implicitly assumed that measurement error proportions are equal across individuals. However, measurement error variances may not even be the same for all persons (or groups) (Schuurman et al., 2015).

Moreover, the conceptualization of the State component in the STARTS model as "measurement error" in this study is a simplification of its meaning. The State components as modeled in the STARTS model capture more than just missmeasurements, they capture any moment-specific fluctuations that are unrelated to other variables in the model (e.g., Usami et al., 2019). One such moment-specific fluctuation for the academic amotivation example is this: a student receives negative feedback and experiences a sudden and short-lived dip in motivation - right before the assessment of amotivation. The then measured amotivation would not necessarily be indicative of that student's amotivation in that week, if we are interested in the relations between amotivation and self-alienation over several weeks. That one week's measure then contains more of a moment-specific fluctuation. It can be argued that this may not be indicative of the student's amotivation, but it is also not exactly measurement error.

In addition, the current approach is yet to be tested for cases in which measurement errors may be correlated. It may be that the measurement errors in two variables x and y are correlated at some or all time points. This seems to make sense theoretically, for instance, if something disrupts the administration of the measurements in one variable x at any measurement time point, it may also disrupt the measurements in another measured variable y.

References

Cole, D. A., Martin, N. C., & Steiger, J. H. (2005). Empirical and conceptual problems with longitudinal trait-state models: Introducing a trait-state-occasion model. *Psychological Methods*, *10*, 3–20. doi: 10.1037/1082-989X.10.1.3

Curran, P. J., & Bauer, D. J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual review of psychology*, *62*, 583–619.

de Haan-Rietdijk, S., Kuppens, P., & Hamaker, E. L. (2016). What's in a day? a guide to decomposing the variance in intensive longitudinal data. *Frontiers in Psychology*, *7*, 891.

Gische, C., & Voelkle, M. C. (2022). Beyond the mean: A flexible framework for studying causal effects using linear models. *psychometrika*, *87*(3), 868–901.

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*. doi: 10.1037/a0038889

Harring, J. R., McNeish, D. M., & Hancock, G. R. (2017). Using phantom variables in structural equation modeling to assess model sensitivity to external misspecification. *Psychological Methods*, *22*(4), 616.

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

Kenny, D., & Zautra, A. (1995). The trait-state-error model for multiwave data. *Journal of consulting and clinical psychology*, *63 1*, 52–9.

Kenny, D., & Zautra, A. (2001). Trait-State Models for Longitudinal Data. In *New methods for the analysis of change* (pp. 1–21). doi: 10.1037/10409-008

Little, T. D., Jorgensen, T. D., Lang, K. M., & Moore, E. W. G. (2014). On the joys of missing data. *Journal of pediatric psychology*, *39*(2), 151–162.

Lüdtke, O., Robitzsch, A., & Ulitzsch, E. (2022). A bayesian approach to estimating reciprocal effects with the bivariate starts model. *Multivariate Behavioral Research*, 1–20.

Lüdtke, O., Robitzsch, A., & Wagner, J. (2017). More stable estimation of the starts model: A bayesian approach using markov chain monte carlo techniques. *Psychological Methods*, *23*. doi: 10.1037/met0000155

Miguel, A., Hernan, R., & James, M. (2023). *Causal inference: What if.* CRC PRESS.

Mulder, J. D. (2022). Power analysis for the random intercept cross-lagged panel model using the powriclpm r-package. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–14.

Mulder, J. D., & Hamaker, E. L. (2021). Three Extensions of the Random Intercept Cross-Lagged Panel Model. *Structural Equation Modeling: A Multidisciplinary Journal*, *28*(4), 638–648. doi: 10.1080/10705511.2020.1784738

Orth, U., Clark, D. A., Donnellan, M., & Robins, R. (2021). Testing Prospective Effects in Longitudinal Research: Comparing Seven Competing Cross-Lagged Models. *Journal of Personality and Social Psychology*, *120*, 1013–1034. doi: 10.1037/pspp0000358

Saltelli, A. (2002). Sensitivity analysis for importance assessment. *Risk Analysis*, *22*(3), 579-590. doi: https://doi.org/10.1111/0272-4332.00040

Savalei, V. (2019). A comparison of several approaches for controlling measurement error in small samples. *Psychological methods*, *24*(3), 352.

Schuurman, N. K. (2023). A "within/between problem" primer: About (not) separating within-person variance and between-person variance in psychology. doi: 10.31234/osf.io/7zgkx

Schuurman, N. K., & Hamaker, E. L. (2019). Measurement error and person-specific reliability in multilevel autoregressive modeling. *Psychological methods*, *24*(1), 70.

Schuurman, N. K., Houtveen, J. H., & Hamaker, E. L. (2015). Incorporating measurement error in n= 1 psychological autoregressive modeling. *Frontiers in psychology*, *6*, 1038.

Usami, S., Murayama, K., & Hamaker, E. L. (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods*, *24*, 637–657. doi: 10.1037/met0000210

Table 1

Illustrative example: sensitivity analysis results

		Proportion of measurement error accounted for			
		0 (RI-CLPM)	0.1	0.2	0.3
Parameter	True value[a]	Estimates [95 % CI]			
$\hat{\alpha}_2$	0.4	0.197 [0.112, 0.283]	0.245 [0.140, 0.350]	0.323 [0.187, 0.458]	0.465 [0.281, 0.649]
$\hat{\alpha}_3$	0.4	0.185 [0.110, 0.261]	0.229 [0.137, 0.321]	0.300 [0.181, 0.419]	0.434 [0.261, 0.607]
$\hat{\beta}_2$	0.2	0.082 [0.010, 0.154]	0.102 [0.016, 0.189]	0.135 [0.028, 0.242]	0.195 [0.060, 0.331]
$\hat{\beta}_3$	0.2	0.090 [0.019, 0.161]	0.113 [0.027, 0.199]	0.152 [0.040, 0.263]	0.230 [0.072, 0.387]
$\hat{\gamma}_2$	0.2	0.109 [0.035, 0.182]	0.134 [0.044, 0.224]	0.173 [0.058, 0.288]	0.240 [0.084, 0.395]
$\hat{\gamma}_3$	0.2	0.070 [0.002, 0.137]	0.085 [0.004, 0.167]	0.111 [0.008, 0.213]	0.156 [0.014, 0.297]
$\hat{\delta}_2$	0.3	0.134 [0.055, 0.214]	0.164 [0.068, 0.260]	0.210 [0.088, 0.332]	0.291 [0.124, 0.458]
$\hat{\delta}_3$	0.3	0.041 [-0.039, 0.121]	0.052 [-0.047, 0.150]	0.069 [-0.059, 0.197]	0.105 [-0.078, 0.289]
$var(\hat{RI}_x)$	1	1.044 [0.880, 1.209]	1.020 [0.840, 1.200]	0.972 [0.757, 1.187]	0.834 [0.478, 1.191]
$var(\hat{RI}_y)$	1	1.167 [1.018, 1.316]	1.158 [1.003, 1.313]	1.140 [0.972, 1.308]	1.088 [0.867, 1.308]
$var(\hat{W}_{x,t=1})$	1	1.691 [1.515, 1.867]	1.441 [1.251, 1.630]	1.214 [0.992, 1.437]	1.079 [0.718, 1.440]
$var(\hat{W}_{x,t=2})$	1	1.745 [1.575, 1.916]	1.452 [1.286, 1.618]	1.154 [0.995, 1.312]	0.844 [0.697, 0.990]
$var(\hat{W}_{x,t=3})$	1	1.713 [1.571, 1.856]	1.430 [1.289, 1.571]	1.147 [1.008, 1.286]	0.863 [0.725, 1.001]
$var(\hat{W}_{y,t=1})$	1	1.714 [1.553, 1.876]	1.436 [1.269, 1.602]	1.165 [0.988, 1.343]	0.928 [0.702, 1.154]
$var(\hat{W}_{y,t=2})$	1	1.700 [1.523, 1.877]	1.405 [1.231, 1.579]	1.107 [0.938, 1.276]	0.804 [0.643, 0.965]
$var(\hat{W}_{y,t=3})$	1	1.590 [1.440, 1.740]	1.314 [1.164, 1.464]	1.040 [0.888, 1.191]	0.770 [0.616, 0.923]
$cov(\hat{RI}_{xy})$	0.5	0.630 [0.513, 0.748]	0.616 [0.490, 0.742]	0.587 [0.441, 0.733]	0.504 [0.276, 0.732]
$cov(\hat{W}_{xy,t=1})$	0.3	0.105 [-0.016, 0.227]	0.120 [-0.009, 0.248]	0.148 [0.001, 0.294]	0.229 [0.004, 0.455]
$cov(\hat{W}_{xy,t=2})$	0.112	0.121 [-0.004, 0.245]	0.117 [-0.005, 0.239]	0.111 [-0.006, 0.229]	0.099 [-0.011, 0.209]
$cov(\hat{W}_{xy,t=3})$	0.112	0.079 [-0.025, 0.183]	0.081 [-0.023, 0.184]	0.084 [-0.019, 0.186]	0.088 [-0.014, 0.191]

Note. Parameter estimates for different proportions of measurement error. For clarity only contains estimates of models in which measurement error proportions accounted for in x and y are equal.

[a]This column cannot be produced for non-simulated data since true values are unknown.

Table 2

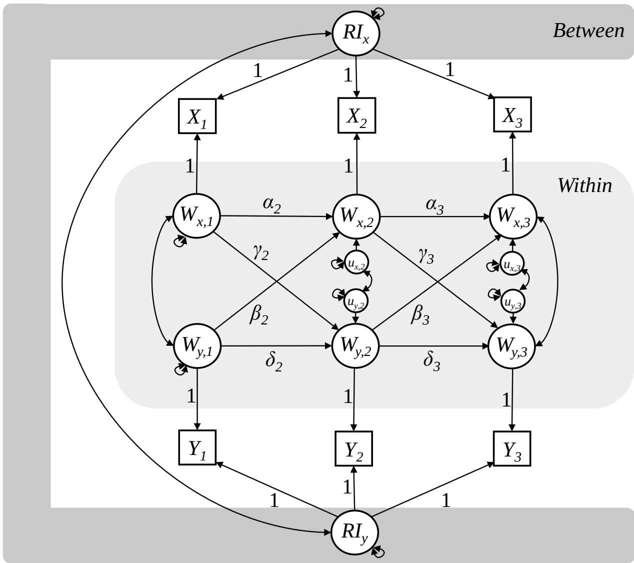
Simulation conditions.

Simulation factors	Values
True proportion of measurement errors	0, 0.2, 0.4, 0.6
Sample size	300, 1000
Number of time points	3, 8
Cross-lagged effects	[.2, 0], [-.2, .2], [-.2, -.2]
Proportion of between to within variance	70:30, 50:50, 30:70

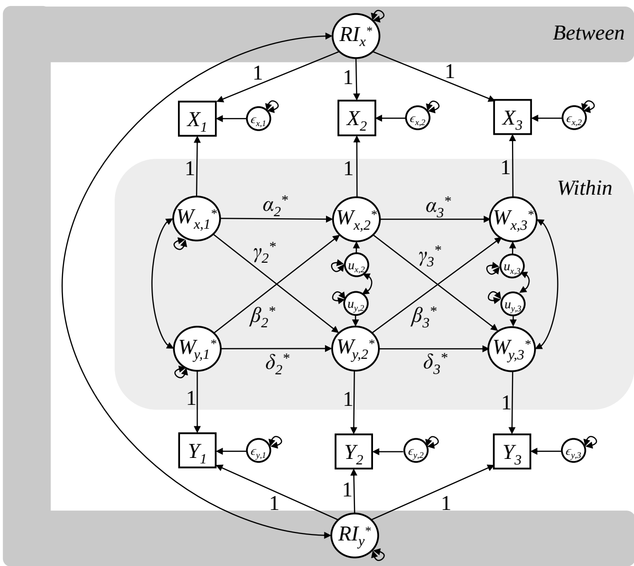
Figure 1

Bivariate RI-CLPM and STARTS model for three waves of data.

(a) *RI-CLPM.*



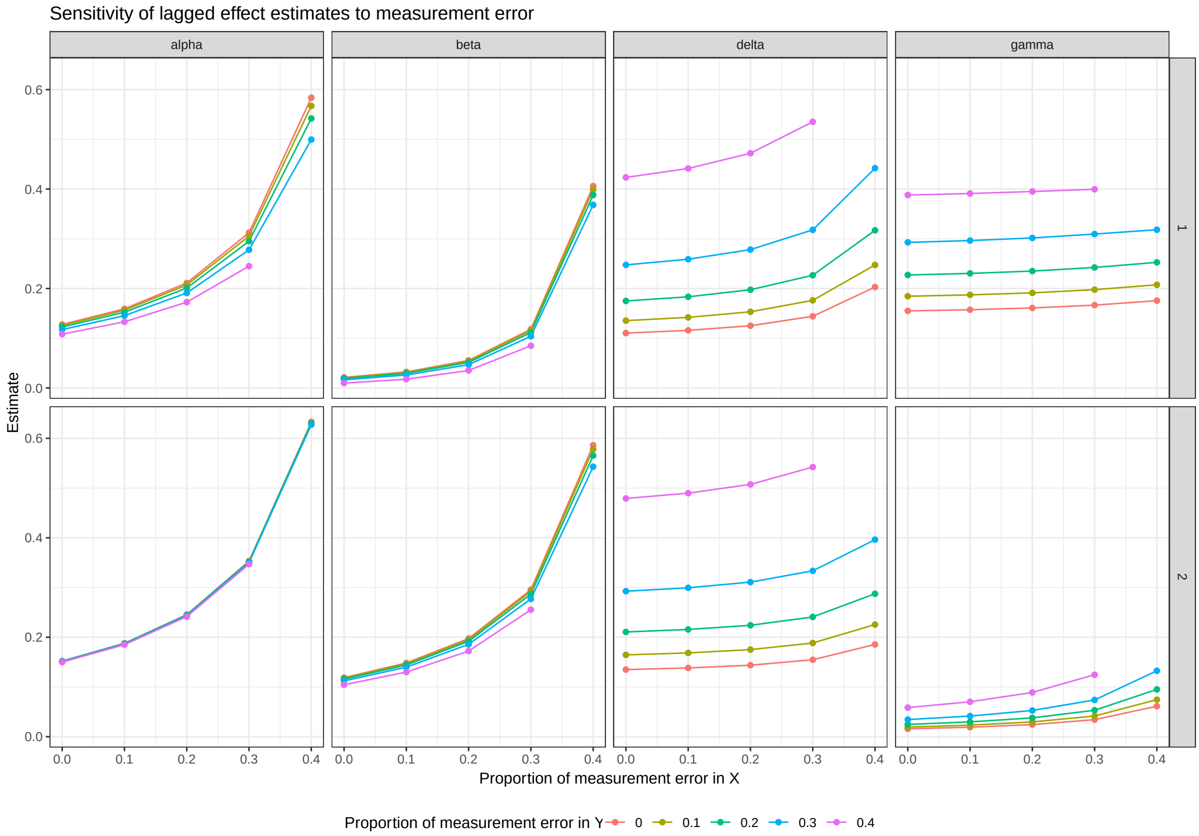
(b) *STARTS.*



Note. Mean structures are not modeled here.

Figure 2

Illustrative example: sensitivity analysis plots of lagged parameters.

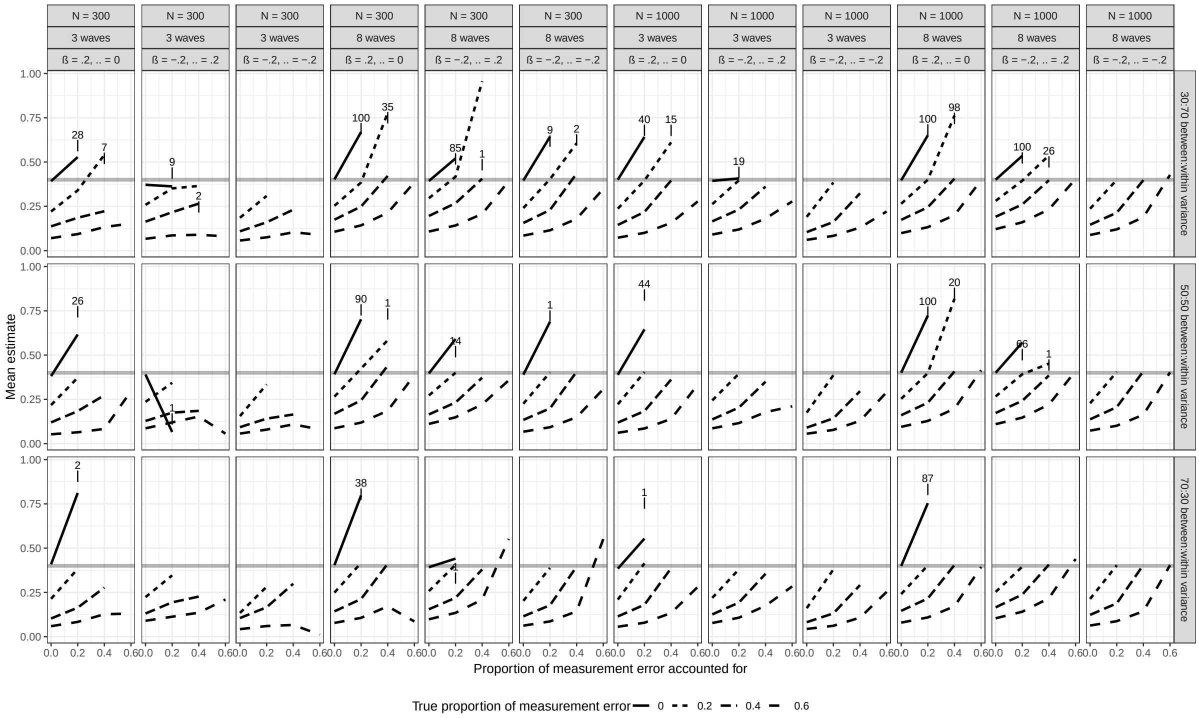


Note. Separate rows for parameter estimates at wave 1 and 2. True (data-generating) values were $\alpha = 0.4$, $\beta = 0.2$, $\delta = 0.3$, $\gamma = 0.2$; 30 % measurement error in x and y, $N = 2000$, RI_x and $RI_y = 1$, $cov(RI_{xy}) = 0.5$, $W_{x,t}$ and $W_{y,t} = 1$, $cov(W_{xy,1}) = 0.3$.

Figure 3

Simulation results: autoregressive parameter

(a) α_2: *point estimates*



(b) α_2: 95% *confidence interval true value coverage rates*

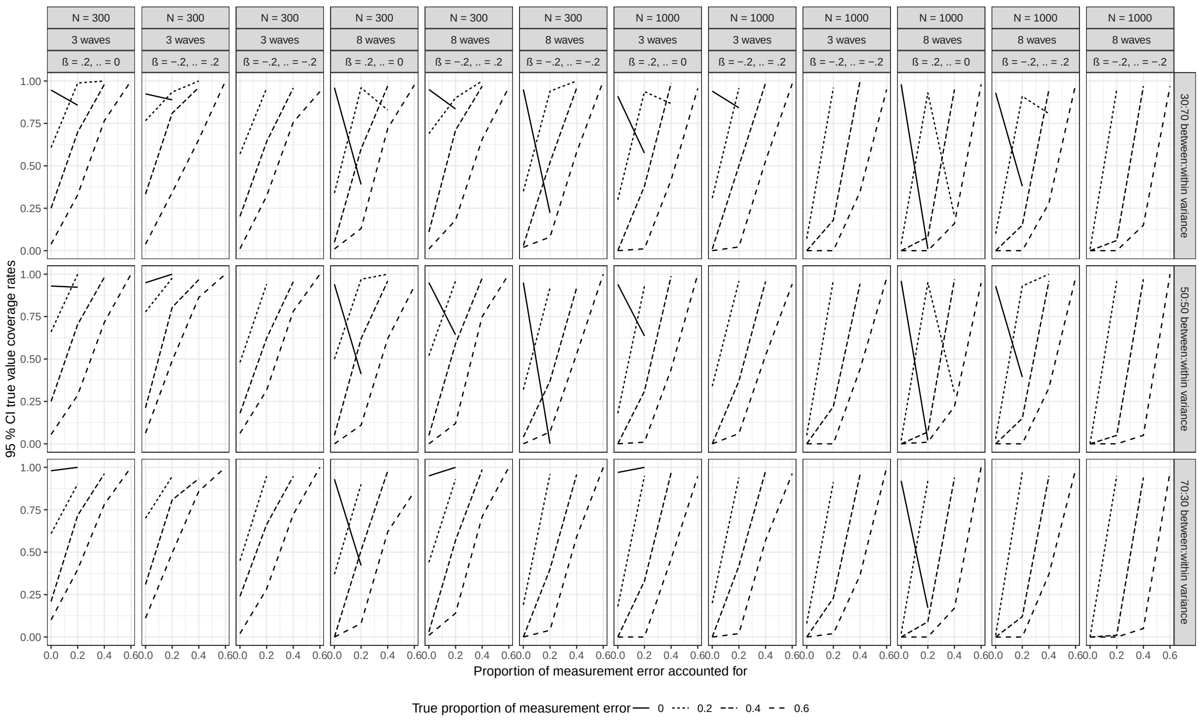
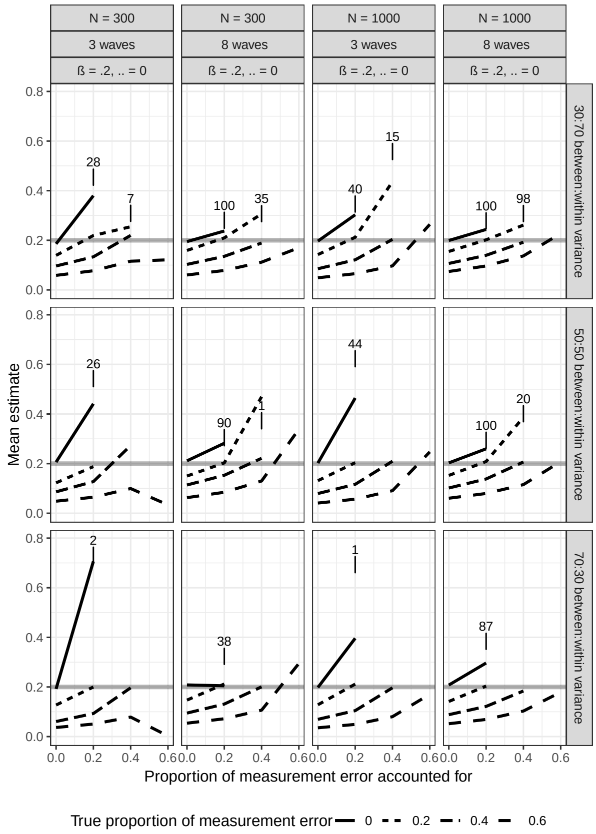


Figure 4

Simulation results: cross-lagged parameter

(a) β_2: *point estimates (true value = 0.2)*



(b) β_2: $95\% - CI$ *true value coverage rates*

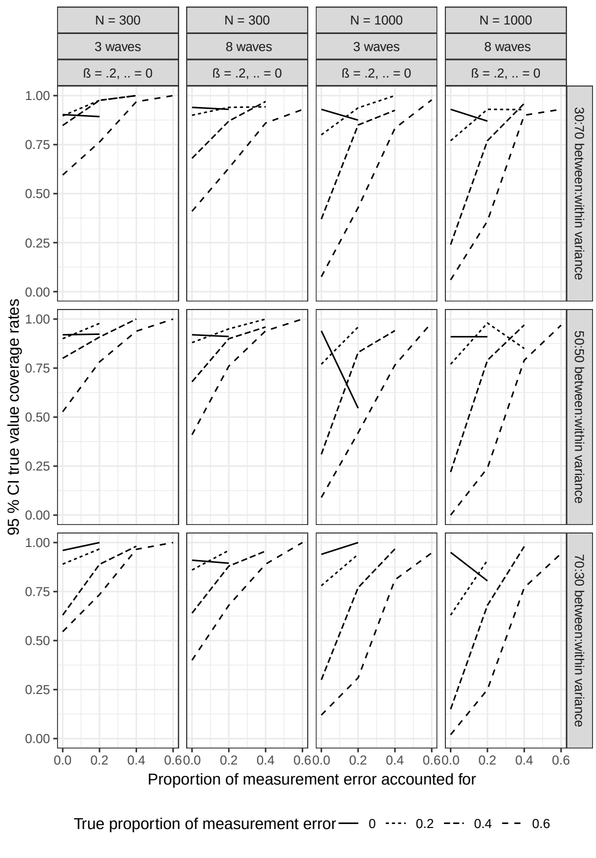
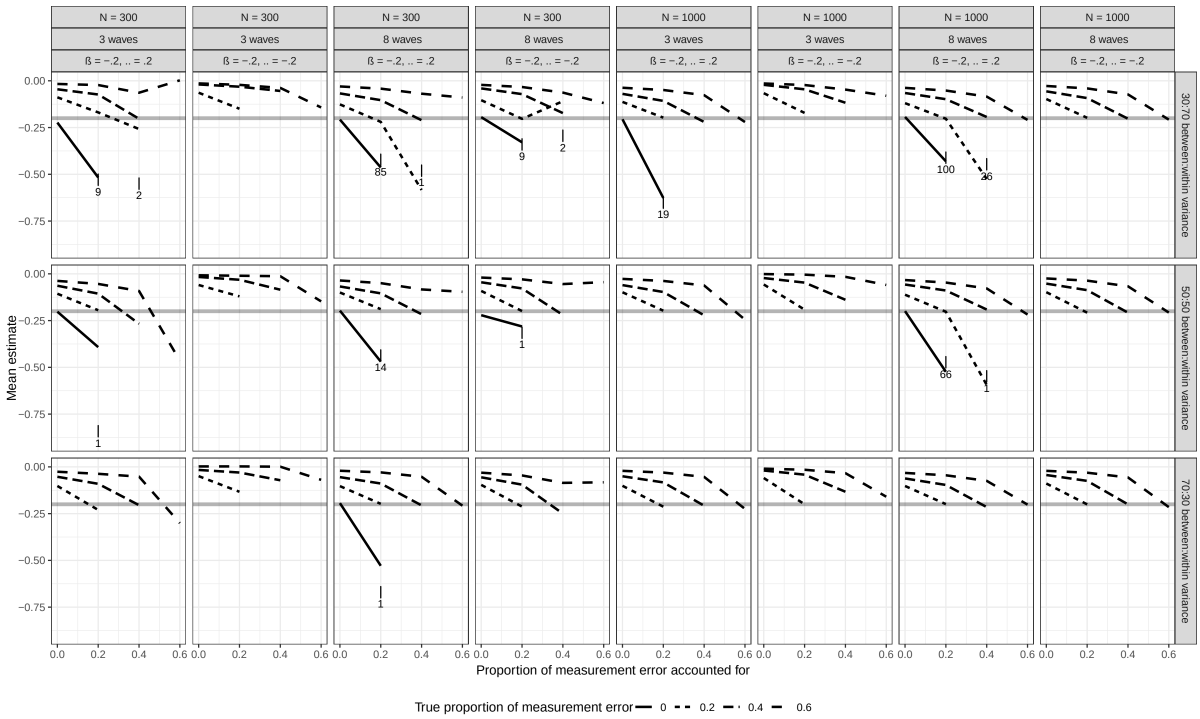


Figure 5

Simulation results: cross-lagged parameter.

(a) β_2*: point estimates (true value = -0.2)*



(b) β_2: $95\% - CI$ *true value coverage rates*

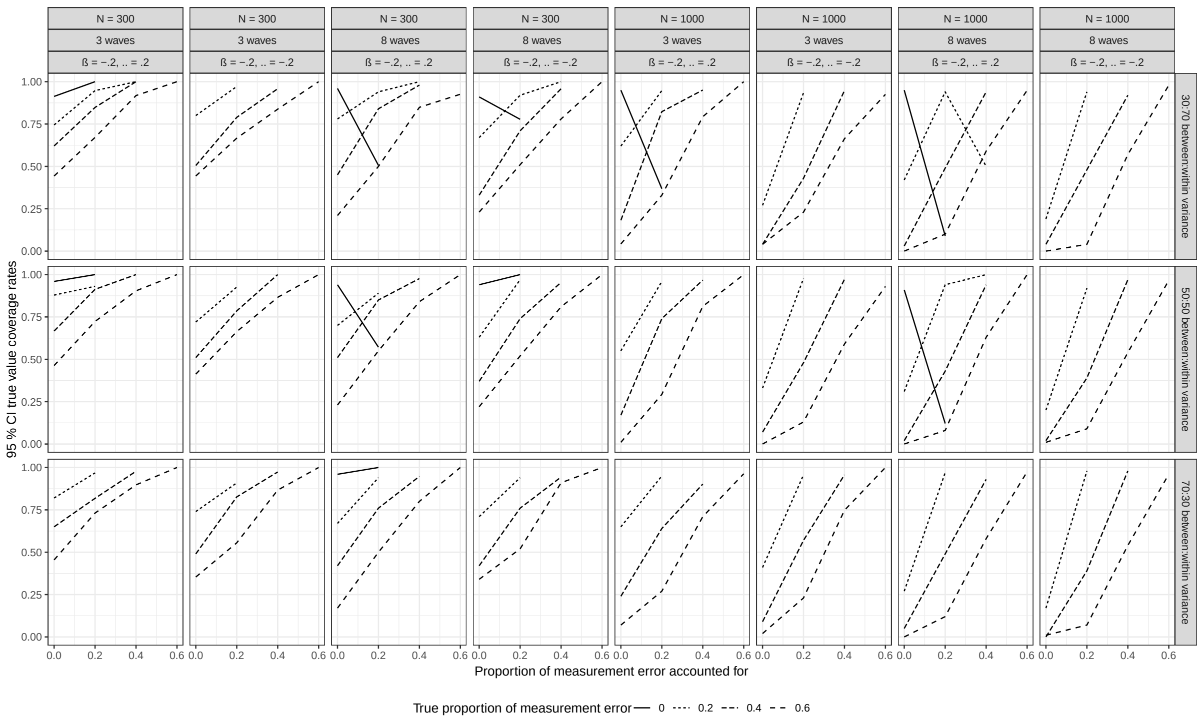
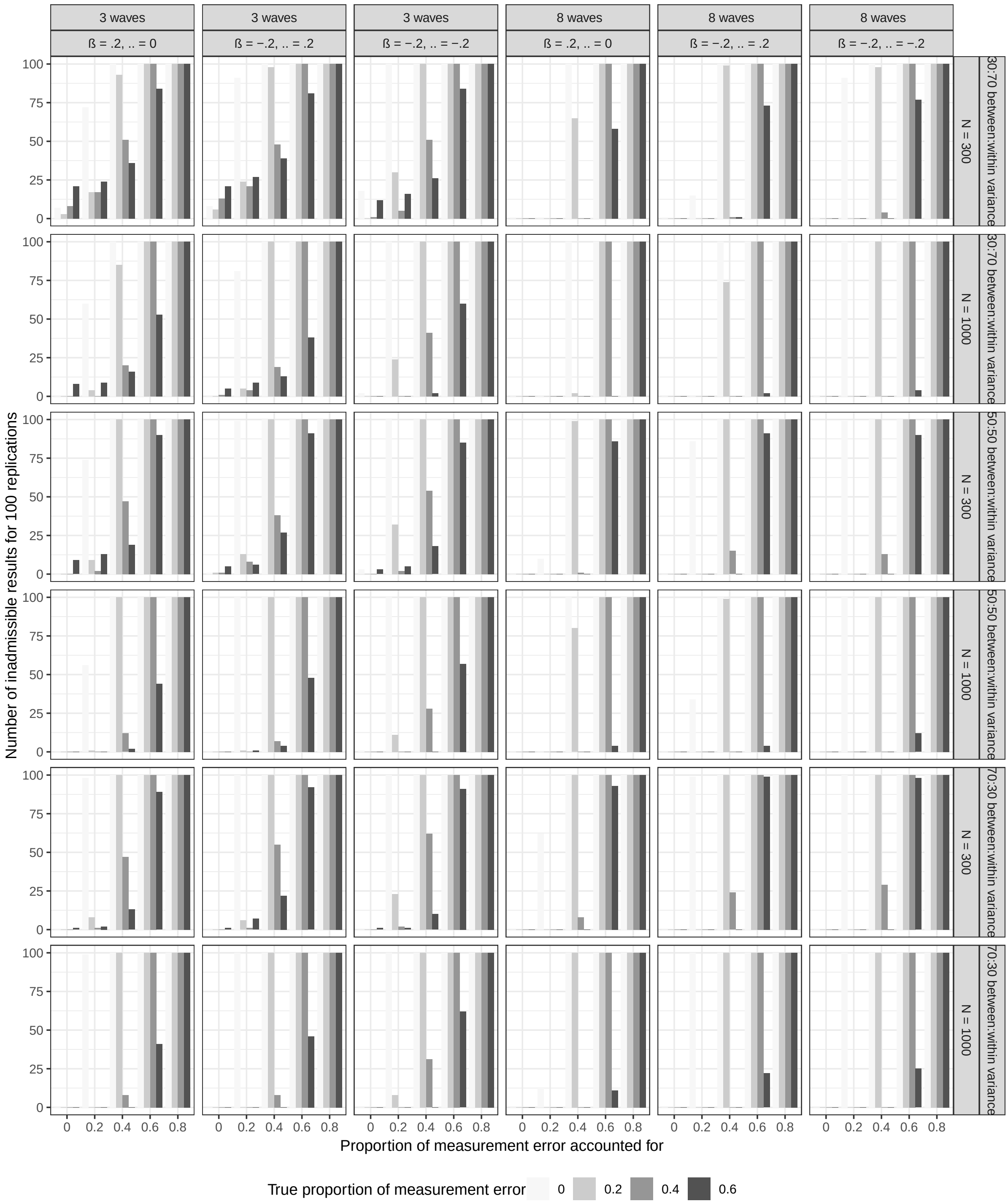


Figure 6

Simulation results: frequencies of inadmissible results (negative variance estimates) for 100 replications.



Note. Grid includes cross-lagged parameters and number of measurement waves in columns. Rows contain ratios of between to within variance and sample size.