KU Leuven

Faculty of Psychology and Educational Sciences

THE EFFECT OF MISSING DATA ON THE ESTIMATION BIAS, PRECISION, AND STATISTICAL POWER IN MULTILEVEL AUTOREGRESSIVE(1) MODELS

Benjamin Simsa

2023

Contents

Introduction

Methods

<div align="right">

Data structure

First-order autoregressive model

Simulation procedure

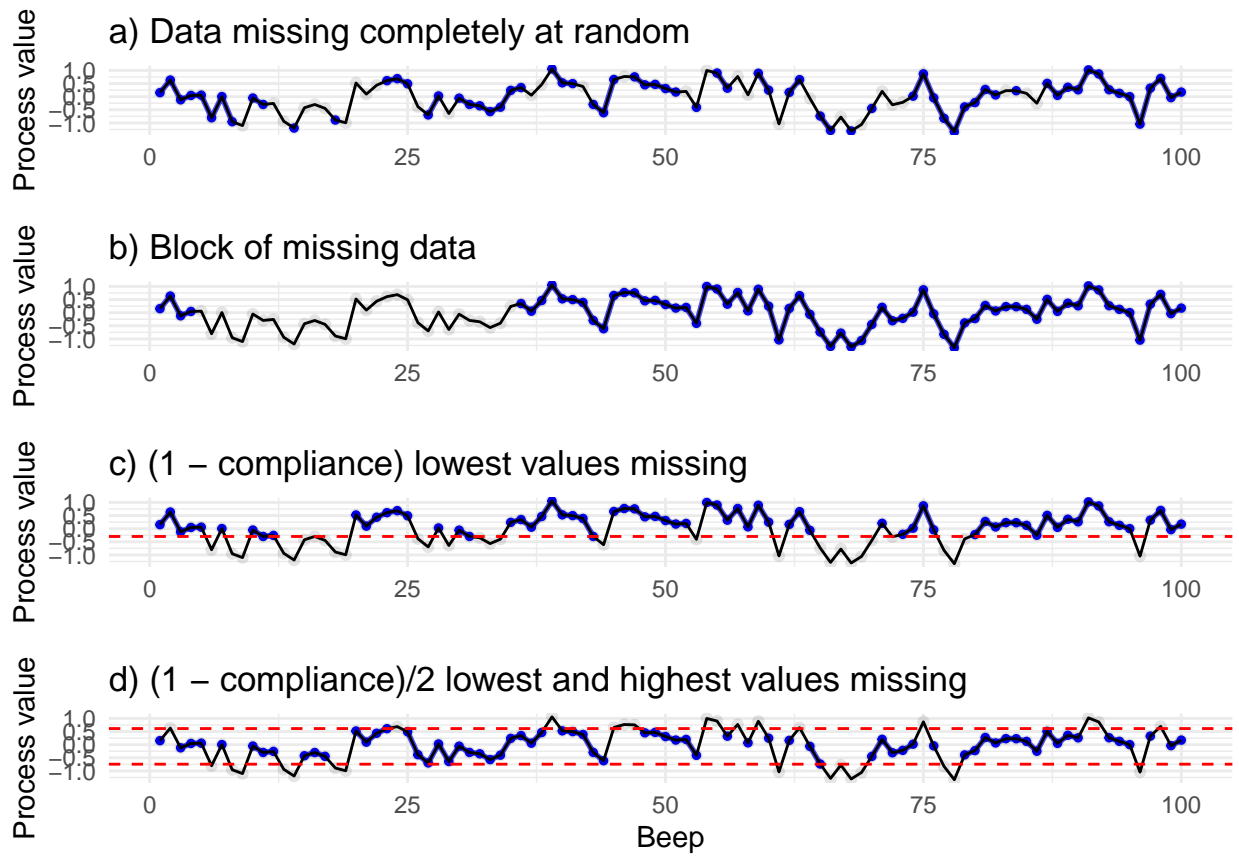## Simulation design

Manipulated parameters

</div>

Missingness pattern   Four different missingness patterns were introduced to the data: a) data missing completely at random (MCAR); b) data missing in blocks of consecutive observations; c) lowest (100%-compliance) observations set as missing, and d) highest and lowest (100%-compliance)/2 observations set as missing.

Each of these missingness patterns correspond to a hypothetical scenario in an ESM study.

The MCAR pattern assumes that the participants miss responding to beeps randomly, and each beep has the same probability of being missed, regardless on any other factors (e.g., whether the previous beep was missed, or the intensity of the emotion measured by ESM). When there is a block of missing data present, all missing observations follow each other, there is a block of non-missing values and missing values. The start and the endpoint of the missing block do not depend on the intensity of the emotion. This can correspond to a situation where a participant misses a series of beeps because they are attending a social event. For patterns c) and d), the missingness is dependent on the value of the process itself. Pattern c) represent a situation in which a participant does not respond to an ESM measure of a positive mood because they are not feeling well, while pattern d) represents a situation where a participant misses an ESM beep when they either do not feel well enough, or they feel too good to respond to their phone beeping.

It can be expected that the different missingness patterns will differ in their effects on the simulation outcomes (estimation bias, standard error, power etc.). This is because with identical proportion in missing data, datasets with different missingness patterns will have different proportions of effective observation-pairs (i.e., proportion of timepoints for which both the observation at t and the observation at $t\text{-}1$ are not missing) used to estimate the autoregressive effect.

Figure VF illustrates the four different missingness patterns in a single ESM time-series.

a) Data missing completely at random

b) Block of missing data

c) (1 – compliance) lowest values missing

d) (1 – compliance)/2 lowest and highest values missing

Beep

Other simulation parameters

Simulation outcomes

Code and reproducibility

We proved stuff

yes.

Or did we

References