Is Timing Everything? The Effects of Measurement Timing on the Performance of Nonlinear Longitudinal Models

by

Sebastian L.V. Sciarra

A Thesis
presented to
The University of Guelph

In partial fulfilment of requirements
for the degree of

Doctorate of Philosophy

in

Psychology

in the field of Industrial-Organizational Psychology

Guelph, Ontario, Canada

ABSTRACT

IS TIMING EVERYTHING? THE EFFECTS OF MEASUREMENT TIMING ON
THE PERFORMANCE OF NONLINEAR LONGITUDINAL MODELS

Sebastian L.V. Sciarra   Advisor(s):

University of Guelph, 2023   David Stanley

Despite the value that longitudinal research offers for understanding psychological processes, studies in organizational research rarely use longitudinal designs. One reason for the paucity of longitudinal designs may be the challenges they present for researchers. Three challenges of particular importance are that researchers have to determine 1) how many measurements to take, 2) how to space measurements, and 3) how to design studies when participants provide data with different response schedules (time unstructuredness). In systematically reviewing the simulation literature, I found that few studies comprehensively investigated the effects of measurement number, measurement spacing, and time structuredness (in addition to sample size) on model performance. As a consequence, researchers have little guidance when trying to conduct longitudinal research. To address these gaps in the literature, I conducted a series of simulation experiments. I found poor model performance across all measurement number/sample size pairings. That is, bias and precision were never concurrently optimized under any combination of manipulated variables. Bias was often low, however, with moderate measurement numbers and sample sizes. Although precision was frequently low, the greatest improvements in precision resulted from using either seven measurements with $N \geq 200$ or nine measurements with $N \leq 100$. With time-unstructured data, model performance systematically

decreased across all measurement number/sample size pairings when the model incorrectly assumed an identical response pattern across all participants (i.e., time-structured data). Fortunately, when models were equipped to handle heterogeneous response patterns using definition variables, the poor model performance observed across all measurement number/sample size pairings no longer appeared. Altogether, the results of the current simulation experiments provide guidelines for researchers interested in modelling nonlinear change.

DEDICATION

[To be completed after defence]

ACKNOWLEDGEMENTS

[To be completed after defence]

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABC   American Broadcasting Company
CBS   Colombia Broadcasting System
CUS   Computer User Services
PBS   Public Broadcasting System

LIST OF APPENDICES

1 Introduction

"Neither the behavior of human beings nor the activities of organizations can be defined without reference to time, and temporal aspects are critical for understanding them" (Navarro et al., 2015, p. 136).

The topic of time has received considerable attention in organizational psychology over the past 20 years. Examples of well-received articles published around the beginning of the 21st century have discussed how investigating time is important for understanding patterns of change and boundary conditions of theory (Zaheer et al., 1999), how longitudinal research is necessary for disentangling different types of causality (T. R. Mitchell & James, 2001), and explicated patterns of organizational change (or institutionalization; Lawrence et al., 2001). Since then, articles have emphasized the need to address time in specific areas such as performance (Dalal et al., 2014; C. D. Fisher, 2008), teams (Roe et al., 2012), and goal setting (Fried & Slowik, 2004) and, more generally, throughout organizational research (Aguinis & Bakker, 2021; George & Jones, 2000; Kunisch et al., 2017; Navarro et al., 2015; Ployhart & Vandenberg, 2010; Roe, 2008; Shipp & Cole, 2015; Sonnentag, 2012; Vantilborgh et al., 2018).

The importance of time has also been recognized in organizational theory. In defining a theoretical contribution, Whetten (1989) discussed that time must be discussed in setting boundary conditions (i.e., under what circumstances does the theory apply) and in specifying relations between variables over time (George & Jones, 2000; T. R. Mitchell & James, 2001). Even if a considerable number of organizational theories do not adhere to the definition of Whetten (1989), theoretical models in organizational psychology consist of path diagrams that delineate the causal events of processes. Given that temporal

1

precedence is a necessary condition for establishing causality (Mill, 2011), time has a role, whether implicitly or explicitly, in organizational theory.

Despite the considerable attention given towards investigating processes over time and the ubiquity of time in organizational theory, the prevalence of longitudinal research has historically remained low. One study examined the prevalence of longitudinal research from 1970–2006 across five organizational psychology journals and found that 4% of articles used longitudinal designs (Roe, 2014, September 22–26). Another survey of two applied psychology journals in 2005 found that approximately 10% (10 of 105 studies) of studies used longitudinal designs (Roe, 2008). Similarly, two surveys of studies employing longitudinal designs with mediation analysis found that, across five journals, only about 10% (7 of 72 studies) did so in 2005 (Maxwell & Cole, 2007) and approximately 16% (15 of 92 studies) did so in 2006 (M. A. Mitchell & Maxwell, 2013).[1] Thus, the prevalence of longitudinal research has remained low.

In the seven sections that follow, I will explain why longitudinal research is necessary and the factors that must be considered when conducting such research. In the first section, I will explain why conducting longitudinal research is essential for understanding the dynamics of psychological processes. In the second section, I will overview patterns of change that are likely to emerge over time. In the third section, I will overview design and analytical issues involved in conducting longitudinal studies. In the fourth section, I will explain how design and analytical issues encountered in conducting longitudinal research can be investigated. In the fifth section, I will provide a systematic review of the research

---

[1]Note that the definition of a longitudinal design in Maxwell and Cole (2007) and M. A. Mitchell and Maxwell (2013) required that measurements be taken over at least three time points so that measurements of the predictor, mediator, and outcome variables were separated over time.

that has investigated design and analytical issues involved in conducting longitudinal research. Finally, in the sixth and seventh sections, I will, respectively, discuss some methods for modelling nonlinear change and the frameworks in which they can be used. A summary of the three simulation experiments that I conducted in my dissertation will then be provided.

1.1 The Need to Conduct Longitudinal Research

Longitudinal designs provide several advantages over cross-sectional designs that allow them to more accurately investigate change (e.g., temporal precedence, testing reverse causality). Unfortunately, even though longitudinal studies often produce results that differ from those of cross-sectional studies, researchers commonly discuss the results of cross-sectional studies as if they have been obtained with a longitudinal design. One example of the assumption of equivalence between cross-sectional and longitudinal findings comes from the large number of studies employing mediation analysis. Given that mediation is used to understand chains of causality in psychological processes (Baron & Kenny, 1986), it would thus make sense to pair mediation analysis with a longitudinal design because understanding causality, after all, requires temporal precedence. Unfortunately, the majority of studies that have used mediation analysis have done so using cross-sectional designs—with estimates of approximately 90% (Maxwell & Cole, 2007) and 84% (M. A. Mitchell & Maxwell, 2013)—and often discuss the results as if they are longitudinal. Investigations into whether mediation results remain equivalent across cross-sectional and longitudinal designs have repeatedly concluded that using mediation analysis on cross-sectional data can return different, and sometimes completely opposite, results from using it on longitudinal data (Cole & Maxwell, 2003; Maxwell & Cole, 2007;

Maxwell et al., [2011](); M. A. Mitchell & Maxwell, [2013](); O'Laughlin et al., [2018()). Therefore, mediation analyses based on cross-sectional analyses may be misleading.

The non-equivalence of cross-sectional and longitudinal results that occurs with mediation analysis is, unfortunately, not due to a specific set of circumstances that only arise with mediation analysis, but a consequence of a broader systematic cause that affects the results of many analyses. The concept of ergodicity explains why cross-sectional and longitudinal analyses seldom yield similar results. To understand ergodicity, it is first important to realize that variance is central to many statistical analyses—correlation, regression, factor analysis, and mediation are some examples. Thus, if variance remains unchanged across cross-sectional and longitudinal data sets, then analyses of either data set would return the same results. Importantly, variance only remains equal across cross-sectional and longitudinal data sets if two conditions put forth by ergodic theory are satisfied (homogeneity and stationarity; Molenaar, [2004](); Molenaar & Campbell, [2009]()). If these two conditions are met, then a process is said to be ergodic. Unfortunately, the two conditions required for ergodicity are highly unlikely to be satisfied and so cross-sectional findings will frequently deviate from longitudinal findings (for a detailed discussion, see Appendix **??**).

Given that cross-sectional and longitudinal analyses are, in general, unlikely to return equivalent findings, it is unsurprising that several investigations in organizational research—and psychology as a whole—have found these analyses to return different results. Beginning with an example from Curran and Bauer ([2011]()), heart attacks are less likely to occur in people who exercise regularly (longitudinal finding), but more likely to happen when exercising (cross-sectional finding). Correlational studies find differences in

4

correlation magnitudes between cross-sectional and longitudinal data sets (for a meta-analytic review, see A. J. Fisher et al., 2018; Nixon et al., 2011).[2] Moving on to perhaps the most commonly employed analysis in organizational research of mediation, several articles have highlighted that cross-sectional data can return different, and sometimes completely opposite, results than those obtained from longitudinal data (Cole & Maxwell, 2003; Maxwell & Cole, 2007; Maxwell et al., 2011; O'Laughlin et al., 2018). Factor analysis is perhaps the most interesting example: The well-documented five-factor model of personality seldom arises when analyzing person-level data obtained by measuring personality on 90 consecutive days (Hamaker et al., 2005). Therefore, cross-sectional analyses are rarely equivalent to longitudinal analyses.

With longitudinal analyses often producing results that differ from those of cross-sectional analyses, it is paramount that longitudinal designs be used to more accurately understand change. Fortunately, technological advancements have allowed researchers to more easily conduct longitudinal research in two ways. First, the use of the experience sampling method (Beal, 2015) in conjunction with modern information transmission technologies—whether through phone applications or short message services—allows data to often be sampled over time with relative ease. Second, the development of longitudinal analyses (along with their integration in commonly used software) that enable person-level data to be modelled such as multilevel models (Raudenbush & Bryk, 2002), growth mixture models (M. Wang & Bodner, 2007), and dynamic factor analysis (Ram et al.,

---

[2]Note that A. J. Fisher et al. (2018) also found the variability of longitudinal correlations to be considerably larger than the variability of cross-sectional correlations.

[2013](#)) provide researchers with avenues to explore the temporal dynamics of psychological processes. With one recent survey estimating that 43.3% of mediation studies (26 of 60 studies) used a longitudinal design (O'Laughlin et al., [2018](#)), it appears that the prevalence of longitudinal research has increased from the 9.5% (Roe, [2008](#)) and 16.3% (M. A. Mitchell & Maxwell, [2013](#)) values estimated at the beginning of the 21st century. Although the frequency of longitudinal research appears to have increased over the past 20 years, several avenues exist where the quality of longitudinal research can be improved, and in my dissertation, I focus on investigating these avenues.

1.2 Understanding Patterns of Change That Emerge Over Time

Change can occur in many ways over time. One pattern of change commonly assumed to occur over time is that of linear change. When change follows a linear pattern, the rate of change over time remains constant. Unfortunately, a linear pattern places demanding restrictions on the possible trajectories of change. If change were to follow a linear pattern, then any pauses in change (or plateaus) or changes in direction could not occur: Change would simply grow over time. Unfortunately, effect sizes have been shown to diminish over time after peaking (for meta-analytic examples, see Cohen, [1993](#); Griffeth et al., [2000](#); Hom et al., [1992](#); Riketta, [2008](#); Steel & Ovalle, [1984](#); Steel et al., [1990](#)). Moreover, many variables display cyclic patterns of change over time, with mood (Larsen & Kasimatis, [1990](#)), daily stress (Bodenmann et al., [2010](#)), and daily drinking behaviour (Huh et al., [2015](#)) as some examples. Therefore, change over is unlikely to follow a linear pattern.

A more realistic pattern of change to occur over time is a nonlinear pattern (for a review, see Cudeck & Harring, [2007](#)). Nonlinear change allows the rate of change to

be nonconstant; that is, change may occur more rapidly during certain periods of time, stop altogether, or reverse direction. When looking at patterns of change observed across psychology, several examples of nonlinear change have been found in the declining rate of speech errors throughout child development (Burchinal & Appelbaum, 1991), rates of forgetting (Murre & Dros, 2015), development of habits (Fournier et al., 2017), and the formation of opinions (Xia et al., 2020). Given that nonlinear change appears more likely than linear change, my dissertation will assume change over time to be nonlinear.

1.3 Challenges Involved in Conducting Longitudinal Research

Conducting longitudinal research presents researchers with several challenges. Many challenges are those from cross-sectional research only amplified (for a review, see Bergman & Magnusson, 1990).[3] For example, greater efforts have to be made to to prevent missing data which can increase over time (Dillman et al., 2014; Newman, 2008). Likewise, the adverse effects of well-documented biases such as demand characteristics (Orne, 1962) and social desirability (Nederhof, 1985) have to be countered at each time point. Outside of challenges shared with cross-sectional research, conducting longitudinal research also presents new challenges. Analyses of longitudinal data have to consider complications such as how to model error structures (Grimm & Widaman, 2010), check for measurement non-invariance over time (the extent to which a construct is measured with the same measurement model over time; Mellenbergh, 1989), and how to center/process data to appropriately answer research questions (Enders & Tofighi, 2007; L. Wang & Maxwell, 2015).

---

[3]It should be noted that conducting a longitudinal study does alleviate some issues encountered in conducting cross-sectional research. For example, taking measurements over multiple time points likely reduces common method variance (Podsakoff et al., 2003; for an example, see Ostroff et al., 2002).

Although researchers must contend with several issues in conducting longitudinal research, three issues are of particular interest in my dissertation. The first issue concerns how many measurements to use in a longitudinal design. The second issue concerns how to space the measurements. The third issue focuses on how much error is incurred if the time structuredness of the data is overlooked. The sections that follow will review each of these issues.

1.3.1 Number of Measurements

Researchers have to decide on the number of measurements to include in a longitudinal study. Although using more measurements increases the accuracy of results—as noted in the results of several studies (e.g., Coulombe et al., 2016; Finch, 2017; Fine et al., 2019; Timmons & Preacher, 2015)—taking additional measurements often comes at a cost that a researcher may be unable to absorb given a limited budget. One important point to mention is that a researcher designing a longitudinal study must take at least three measurements to allow a reliable estimate of change and, perhaps more importantly, to allow a nonlinear pattern of change to be modelled (Ployhart & Vandenberg, 2010). In my dissertation, I hope to determine whether an optimal number of measurements exists when modelling a nonlinear pattern of change.

1.3.2 Spacing of Measurements

Additionally, a researcher must decide on the spacing of measurements in a longitudinal study. Although discussions of measurement spacing often recommend that researchers use theory and previous studies to determine measurement spacing (Cole & Maxwell, 2003; Collins, 2006; Dormann & Griffin, 2015; Dormann & van de Ven, 2014; T. R. Mitchell & James, 2001), organizational theories seldom delineate periods of time

8

over which a processes unfold, and so the majority of longitudinal research uses intervals of convention and/or convenience to space measurements (Dormann & van de Ven, 2014; T. R. Mitchell & James, 2001). Unfortunately, using measurement spacings that do not account for the temporal pattern of change of a psychological process can lead to inaccurate results (e.g., Chen et al., 2014). As an example, Cole and Maxwell (2009) show how correlation magnitudes are affected by the choice of measurement spacing intervals. In my dissertation, I hope to determine whether an optimal measurement spacing schedule exists when modelling a nonlinear pattern of change.

1.3.3 Time Structuredness

Last, and perhaps most pernicious, latent variable analyses of longitudinal data are likely to incur error from an assumption they make about data collection conditions. Latent variable analyses assume that, across all collection points, participants provide their data at the same time. Unfortunately, such a high level of regularity in the response patterns of participants is unlikely: Participants are more likely to provide their data over some period of time after a data collection window has opened. As an example, consider a study that collects data from participants at the beginning of each month. If participants respond with perfect regularity, then they would all provide their data at the exact same time (e.g., noon on the second day of each month). If the participants respond with imperfect regularity, then they would provide their at different times after the beginning of each month. The regularity of response patterns observed across participants in a longitudinal study determines the time structuredness of the data and the sections that follow will provide overview of time structuredness.

1.3.3.1 Time-Structured Data

Many analyses assume that data are *time structured*: Participants provide data at the same time at each collection point. By assuming time-structured data, an analysis can incur error because it will map time intervals of inappropriate lengths onto the time intervals that occurred between participant's responses.[4] As an example of the consequences of incorrectly assuming data to be time structured, consider a study that assessed the effects of an intervention on the development of leadership by collecting leadership ratings at four time points each separated by four weeks (Day & Sin, 2011). The employed analysis assumed time-structured data; that is, each each participant provided ratings on the same day—more specifically, the exact same moment—each time these ratings were collected. Unfortunately, it is unlikely that the data collected from participants were time structured: At any given collection point, some participants may have provided leadership ratings at the beginning of the week, while others may only provide ratings two weeks after the survey opened. Importantly, ratings provided two weeks after the survey opened were likely influenced by changes in leadership that occurred over the two weeks. If an analysis incorrectly assumes time-structured data, then it assumes each participant has the same response pattern and, therefore, will incorrectly attribute the amount of time that elapses between most participants' responses. For instance, if a participant only provides a leadership rating two weeks after having received a survey (and six weeks after providing their previous rating), then using an analysis that assumes time-structured data would incorrectly assume that each collection point of this participant is separated

---

[4]It should be noted that, although seldom implemented, analyses can be accessorized to handle time-unstructured data by using definition variables (Mehta & West, 2000; Mehta & Neale, 2005).

by four weeks (the interval used in the experiment) and would, consequently, model the observed change as if it had occurred over four weeks. Therefore, incorrectly assuming data to be time structured leads an analysis to overlook the unique response rates of participants across the collection points and, as a consequence, incur error (Coulombe et al., 2016; Mehta & Neale, 2005; Mehta & West, 2000).

1.3.3.2 Time-Unstructured Data

Conversely, other analyses assume that data are *time unstructured*: Participants provide data at different times at each collection point. Given the unlikelihood of one response pattern describing the response rates of all participants in a given study, the data obtained in a study are unlikely to be time structured. Instead, and because participants are likely to exhibit unique response patterns in their response rates, data are likely to be time unstructured. One way to conceptualize the distinction between time-structured and time-unstructured data is on a continuum. On one end of the continuum, participants all provide data with identical response patterns, thus giving time-structured data. When participants exhibit unique response patterns, the resulting data are time unstructured, with the extent of time-unstructuredness depending on the average uniqueness of all response patterns. For example, if data are collected at the beginning of each month and participants only have one day to provide data at each time point, then the resulting data will have a low amount of time structuredness because response patterns can only differ from each other over the course of one day. Alternatively, if data are collected at the beginning of each month and participants have 30 days to provide data at each time point, then the resulting data will have a high amount of time structuredness because response patterns can differ from each other over the course of 30 days. Therefore, the continuum

11

of time struturedness has time-structured data on one end and time-unstructured data with long response windows on another end. In my dissertation, I hope to determine how much error is incurred when time-unstructured data of varying degrees are assumed to be time structured.

1.3.4 Summary

In summary, researchers must contend with several issues when conducting longitudinal research. In addition to contending with issues encountered in conducting cross-sectional research, researchers must contend with new issues that arise from conducting longitudinal research. Three issues of particular importance in my dissertation are the number of measurements, the spacing of measurements, and incorrectly assuming time-unstructured data to be time structured. These issues will be serve as a basis for a systematic review of the simulation literature.

1.4 Using Simulations To Assess Modelling Accuracy

In the next section, I will present the results of a systematic review of the literature that has investigated the issues of measurement number, measurement spacing, and time structuredness. Before presenting the results of the systematic review, I will provide an overview of the Monte Carlo method used to investigate the issues involved in conducting longitudinal research.

To understand how the effects of longitudinal issues on modelling accuracy can be investigated, the inferential method commonly employed in psychological research will first be reviewed with an emphasis on its shortcomings (see Figure 1.1). Consider an example where a researcher wants to understand how sampling error affects the accuracy with which a sample mean ($\bar{x}$) estimates a population mean (μ). Using the inferential

12

₂₆₇ method, the researcher samples data and then estimates the population mean (μ) by

₂₆₈ computing the mean of the sampled data ($\bar{x}_1$). Because collected samples are almost

₂₆₉ always contaminated by a variety of methodological and/or statistical deficiencies (such

₂₇₀ as sampling error, measurement error, assumption violations, etc.), the estimation of the

₂₇₁ population parameter is likely to be imperfect. Unfortunately, to estimate the effect of

₂₇₂ sampling error on the accuracy of the population mean estimate ($\bar{x}_1$), the researcher

₂₇₃ would need to know the value of the population mean; without knowing the value of the

₂₇₄ population mean, it is impossible to know how much error was incurred in estimating the

₂₇₅ population mean and, as as a result, impossible to know the extent to which sampling

₂₇₆ error contributed to this error. Therefore, a study following the inferential approach can

₂₇₇ only provide estimates of population parameters.

₂₇₈   The Monte Carlo method has a different goal. Whereas the inferential method

₂₇₉ focuses on estimating parameters from sample data, the Monte Carlo method is used to

₂₈₀ understand the factors that influence the accuracy of the inferential approach. Figure 1.1

₂₈₁ shows that the Monte Carlo method works in the opposite direction of the inferential

₂₈₂ approach: Instead of collecting a sample, the Monte Carlo method begins by assigning a

₂₈₃ value to at least one parameter to define a population. Many sample data sets are then

₂₈₄ generated from the defined population ($s_1, s_2, ..., s_n$) and the data from each sample are

₂₈₅ then modelled by computing a sample mean ($\bar{x}_1, \bar{x}_2, ..., \bar{x}_n$). Importantly, manipulations

₂₈₆ can be applied to the sampling and/or modelling of the data. In the current example,the

₂₈₇ population estimates of each statistical model are averaged ($\bar{\bar{x}}$) and compared to the

₂₈₈ pre-determined parameter value (μ). The difference between the average of the estimates

₂₈₉ and the known population value constitutes bias in parameter estimation (i.e., parameter

13

bias). In the current example, the manipulation causes a systematic underestimation, on average, of the population parameter. By randomly generating data, the Monte Carlo method can estimate how a variety of methodological and statistical factors affect the accuracy of a model (for a review, see Robert & Casella, 2010).

Monte Carlo simulations have been used to evaluate the effects of a variety of methodological and statistical deficiencies for several decades. Beginning with an early use of the Monte Carlo method, Boneau (1960) used it to evaluate the effects of assumption violations on the fidelity of t-value distributions. In more recent years, implementations of the the Monte Carlo method have shown that realistic values of sample size and measurement accuracy produce considerable variability in estimated correlation values (Stanley & Spence, 2014). Monte Carlo simulations have also provided valuable insights into more complicated statistical analyses. In investigating more complex statistical analyses, simulations have shown that mediation analyses are biased to produce results of complete mediation because the statistical power to detect direct effects falls well below the statistical power to detect indirect effects (Kenny & Judd, 2014). Given the ability of the Monte Carlo method to evaluate statistical methods, the experiments in my dissertation used it to evaluate the effects of measurement number, measurement spacing, and time structuredness on modelling accuracy.[5]

1.5 Systematic Review of Simulation Literature

To understand the extent to which issues involved in conducting longitudinal research had been investigated, I conducted a systematic review of the simulation literature.

---

[5]My simulation experiments also investigated the effects of sample size and nature of change on modelling accuracy.

14

Figure 1.1
Depiction of Monte Carlo Method
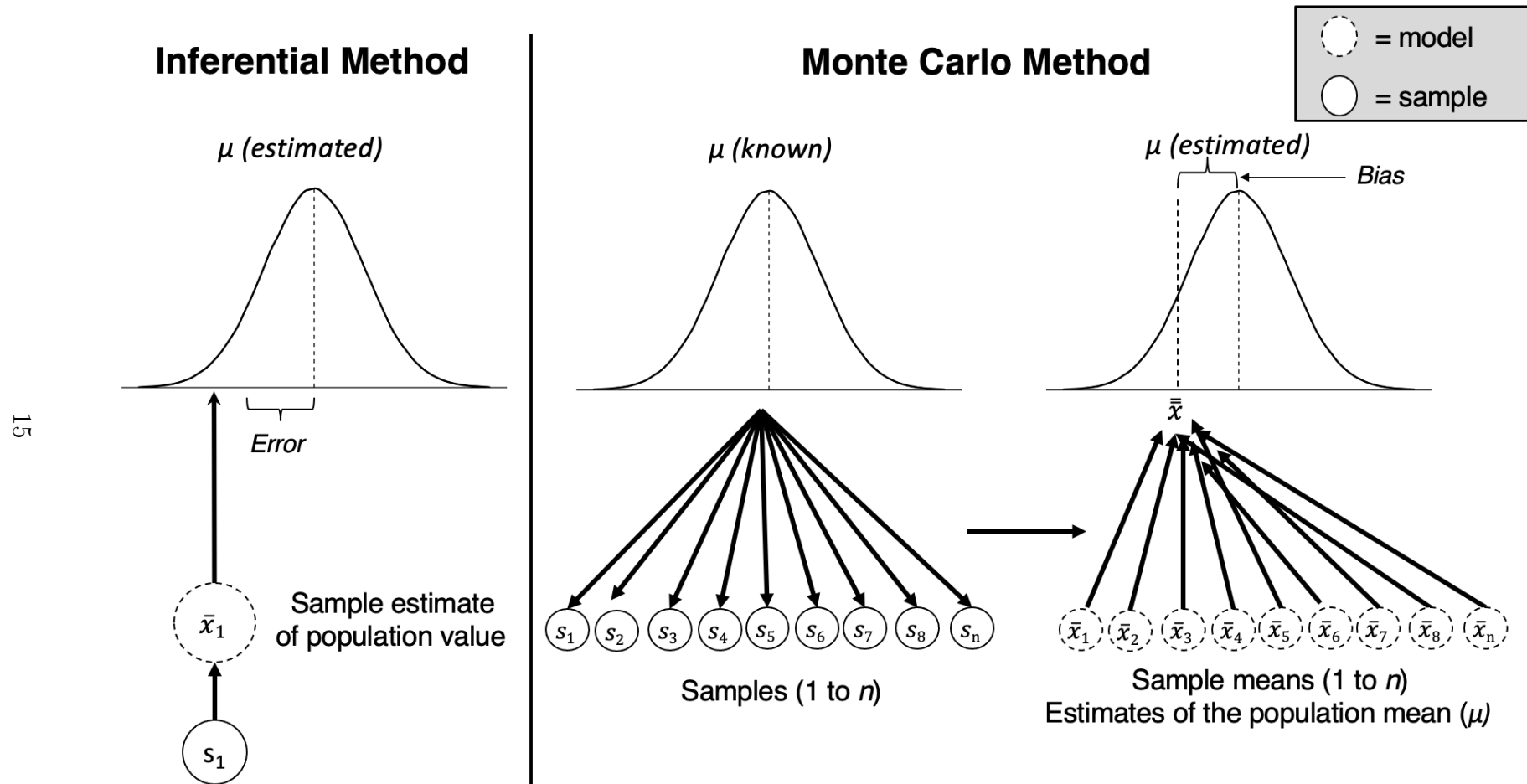
311   *Note.* Comparison of inferential approach with the Monte Carlo approach. The inferential approach begins with a collected sample and then estimates the

312   population parameter using an appropriate statistical model. The difference between the estimated and population value can be conceptualized as error.

Because the population value is generally unknown in the inferential approach, it cannot estimate how much error is introduced by any given methodological or statistical deficiency. To estimate how much error is introduced by any given methodological or statistical deficiency, the Monte Carlo method needs to be used, which constitutes four steps. The Monte Carlo method first defines a population by setting parameter values. Second, many samples are generated from the pre-defined population, with some methodological deficiency built in to each data set (in this case, each sample has a specific amount of missing data). Third, each generated sample is then analyzed and the population estimates of each statistical model are averaged and compared to the pre-determined parameter value. Fourth, the difference between the estimate average and the known population value defines the extent to which the missing data manipulation affected parameter estimation (the difference between the population and average estimated population value is the parameter bias).

³²⁰ The sections that follow will first present the method I followed in systematically
³²¹ reviewing the literature and then summarize the findings of the review.

1.5.1 Systematic Review Methodology

³²³ I identified the following keywords through citation searching and independent read-
³²⁴ ing: "growth curve", "time-structured analysis", "time structure", "temporal design", "in-
³²⁵ dividual measurement occasions", "measurement intervals", "methods of timing", "longi-
³²⁶ tudinal data analysis", "individually-varying time points", "measurement timing", "latent
³²⁷ difference score models", "parameter bias", and "measurement spacing". I entered these
³²⁸ keywords entered into the PsycINFO database (on July 23, 2021) along with the word
³²⁹ "simulation" in any field and considered any returned paper a viable ppaper (see Figure
³³⁰ 1.2 for a PRISMA diagram illustrating the filtering of the reports). The search returned
³³¹ 165 reports, which I screened by reading the abstracts. Initial screening led to the removal
³³² of 60 reports because they did not contain any simulation experiments. Of the remaining
³³³ 105 papers, I removed 2 more papers because they could not be accessed (Stockdale, 2007;
³³⁴ Tiberio, 2008). Of the remaining 103 identified simulation studies, I deemed a paper as
³³⁵ relevant if it investigated the effects of any design and/or analysis factor related to con-
³³⁶ ducting longitudinal research (i.e., number of measurements, spacing of measurements,
³³⁷ and/or time structuredness) and did so using the Monte Carlo simulation method. Of
³³⁸ the remaining 103 studies, I removed 89 studies because they did not meet the inclusion
³³⁹ criteria, leaving fourteen studies to be included in the review. I also found an additional
³⁴⁰ 3 studies through citation searching, giving a total of 17 studies.

³⁴¹ The findings of my systematic review are summarized in Tables 1.1–1.2. Tables
³⁴² 1.1–1.2 differ in one way: Table 1.1 indicates how many studies investigated each effect,

17

whereas Table 1.2 provides the reference of each study and detailed information about each study's method. Otherwise, all other details of Tables 1.1–1.2 are identical. The first column lists the longitudinal design factor (alongside with sample size) and the corresponding two- and three-way interactions. The second and third columns list whether each effect has been investigated with linear and nonlinear patterns of change, respectively. Shaded cells indicate effects that have not been investigated, with cells shaded in light grey indicating effects that have not been investigated with linear patterns of change and cells shaded in dark grey indicating effects that have not been investigated with nonlinear patterns of change.[6]

1.5.2 Systematic Review Results

Although previous research appeared to sufficiently fill some cells of Table 1.1, two patterns suggest that arguably the most important cells (or effects) have not been investigated. First, it appears that simulation research has invested more effort in investigating the effects of longitudinal design factors with linear patterns than with nonlinear patterns of change. In counting the number of effects that remain unaddressed with linear and nonlinear patterns of change, a total of five cells (or effects) have not been investigated, but a total of seven cells have not been investigated with nonlinear patterns of

---

[6]Table 1.2 lists the effects that each study (identified by my systematic review) investigated and notes the following methodological details (using superscript letters and symbols): the type of model used in each paper, assumption and/or manipulation of complex error structures (heterogeneous variances and/or correlated residuals), manipulation of missing data, and/or pseudo-time structuredness manipulation. Across all 17 simulation studies, 5 studies (29%) assumed complex error structures (Gasimova et al., 2014; Liu & Perera, 2022; Y. Liu et al., 2015; Miller & Ferrer, 2017; Murphy et al., 2011), 1 study (6%) manipulated missing data (Fine et al., 2019), and 2 studies (12%) contained a pseudo-time structuredness manipulation (Fine et al., 2019; Fine & Grimm, 2020). Importantly, the pseudo-time structuredness manipulation used in Fine et al. (2019) and Fine and Grimm (2020) differed from the manipulation of time structuredness used in the current experiments (and from previous simulation experiments of Coulombe et al., 2016; Miller & Ferrer, 2017) in that it randomly generated longitudinal data such that a given person could provide all their data before another person provided any data.

18

Figure 1.2

PRISMA Diagram Showing Study Filtering Strategy

360 *Note.* PRISMA diagram for systematic review of simulation research that investigates longitudinal design and analysis factors.

Table 1.1

Number of Simulation Studies That Have Investigated Longitudinal Issues with Linear and Nonlinear Change Patterns (n = 17)

Effect	Linear pattern	Nonlinear pattern
Main effects		
Number of measurements (NM)	11 studies	6 studies
Spacing of measurements (SM)	1 study	1 study
Time structuredness (TS)	2 studies	1 study
Sample size (S)	11 studies	7 studies
Two-way interactions		
NM x SM	1 study	1 study
NM x TS	1 study	**Cell 1 (Exp. 3)**
NM x S	9 studies	5 studies
SM x TS	**Cell 2**	**Cell 3**
SM x S	**Cell 4**	**Cell 5 (Exp. 2)**
TS x S	1 study	2 studies
Three-way interactions		
NM x SM x TS	**Cell 6**	**Cell 7**
NM x SM x S	**Cell 8**	**Cell 9 (Exp. 2)**
NM x TS x S	1 study	**Cell 10 (Exp. 3)**
SM x TS x S	**Cell 11**	**Cell 12**

Table 1.1

Number of Simulation Studies That Have Investigated Longitudinal Issues with Linear and Nonlinear Change Patterns (n = 17) (continued)

Effect	Linear pattern	Nonlinear pattern

Note. Cells are only numbered for effects that have not been investigated. Cells shaded in light grey indicate effects that have not been investigated with linear patterns of change and cells shaded in dark grey indicate effects that have not been investigated with nonlinear patterns of change.

Table 1.2

Summary of Simulation Studies That Have Investigated Longitudinal Issues with Linear and Nonlinear Change Patterns (n = 17)

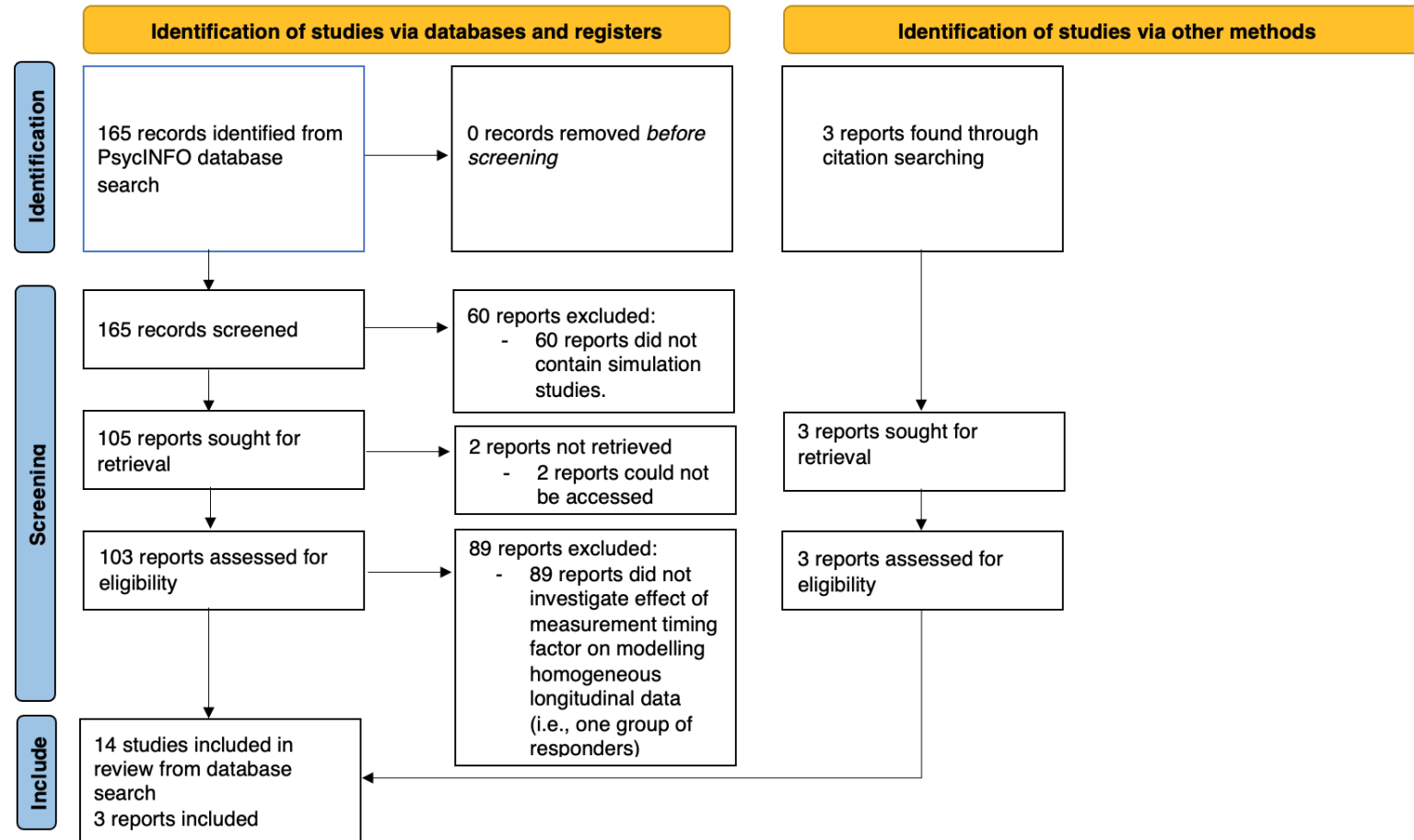| Effect	Linear pattern	Nonlinear pattern
Main effects		
Number of measurements (NM)	(Timmons & Preacher, 2015, [a]; Murphy et al., 2011, [b]⊍; Gasimova et al., 2014, [c]⊍; Wu et al., 2014, [a]; Coulombe, 2016, [a]; Ye, 2016, [a]; Finch, 2017, [a]; O'Rourke et al., 2022, [d]; Newsom & Smith, 2020, [a]; Coulombe et al., 2016, [a])	(Timmons & Preacher, 2015, [a]; Finch, 2017, [a]; Fine et al., 2019, [e○▽]; Fine & Grimm, 2020, [e,f▽]; J. Liu et al., 2021, [g]; Liu & Perera, 2022, [h]⊍; Y. Liu et al., 2015, [g]⊍)
Spacing of measurements (SM)	(Timmons & Preacher, 2015, [a])	(Timmons & Preacher, 2015, [a])
Time structuredness (TS)	(Aydin et al., 2014, [a]; Coulombe et al., 2016, [a])	(Miller & Ferrer, 2017, [a]⊍; Y. Liu et al., 2015, [g]⊍)
Sample size (S)	(Murphy et al., 2011, [b]⊍; Gasimova et al., 2014, [c]⊍; Wu et al., 2014, [a]; Coulombe, 2016, [a]; Ye, 2016, [a]; Finch, 2017, [a]; O'Rourke et al., 2022, [d]; Newsom & Smith, 2020, [a]; Coulombe et al., 2016, [a]; Aydin et al., 2014, [a])	(Finch, 2017, [a]; Fine et al., 2019, [e○▽]; Fine & Grimm, 2020, [e,f▽]; J. Liu et al., 2021, [g]; Liu & Perera, 2022, [h]⊍; Y. Liu et al., 2015, [g]⊍; Miller & Ferrer, 2017, [a]⊍)
Two-way interactions		
NM x SM	(Timmons & Preacher, 2015, [a])	(Timmons & Preacher, 2015, [a])
NM x TS	(Coulombe et al., 2016, [a])	**Cell 1 (Exp. 3)**

Table 1.2

Summary of Simulation Studies That Have Investigated Longitudinal Issues with Linear and Nonlinear Change Patterns (n = 17) (continued)

Effect	Linear pattern	Nonlinear pattern
NM x S	(Murphy et al., 2011, [b]℧; Gasimova et al., 2014, [c]℧; Wu et al., 2014, [a]; Coulombe, 2016, [a]; Ye, 2016, [a]; Finch, 2017, [a]; O'Rourke et al., 2022, [d]; Newsom & Smith, 2020, [a]; Coulombe et al., 2016, [a])	(Finch, 2017, [a]; Fine et al., 2019, [e○▽]; Fine & Grimm, 2020, [e,f▽]; J. Liu et al., 2021, [g]; Liu & Perera, 2022, [h℧])
SM x TS	**Cell 2**	**Cell 3**
SM x S	**Cell 4**	**Cell 5 (Exp. 2)**
TS x S	(Aydin et al., 2014, [a])	(Y. Liu et al., 2015, [g℧]; Miller & Ferrer, 2017, [a℧])
Three-way interactions		
NM x SM x TS	**Cell 6**	**Cell 7**
NM x SM x S	**Cell 8**	**Cell 9 (Exp. 2)**
NM x TS x S	(Coulombe et al., 2016, [a])	**Cell 10 (Exp. 3)**
SM x TS x S	**Cell 11**	**Cell 12**

Note. Cells are only numbered for effects that have not been investigated. Cells shaded in light and dark grey indicate effects that have not, respectively, been investigated with linear and nonlinear patterns of change.

[a] Latent growth curve model. [b] Second-order latent growth curve model. [c] Hierarchical Bayesian model. [d] Bivariate latent change score model. [e] Functional mixed-effects model. [f] Nonlinear mixed-effects model. [g] Bilinear spline model. [g] Parallel bilinear spline model.

[○] Manipulated missing data. [℧] Assumed complex error structure (heterogeneous variances and/or correlated residuals). [▽] Contained pseudo-time structuredness manipulation.

change. Given that change over time is more likely to follow a nonlinear than a linear pattern (for a review, see Cudeck & Harring, 2007), it could be argued that most simulation research has investigated the effect of longitudinal design factors under unrealistic conditions.

Second, all the cells corresponding to the three-way interactions with nonlinear patterns of change have not been investigated (Cells 7, 9, 10, and 12 in Table 1.1), meaning that almost no study has conducted a comprehensive investigation into measurement timing. Given that longitudinal research is needed to understand the temporal dynamics of psychological processes—as suggested by ergodic theory (Molenaar, 2004)—it is necessary to understand how longitudinal design and analysis factors interact with each other (and with sample size) in affecting the modelling accuracy of temporal dynamics. Given that no simulation study identified in my systematic review conducted a comprehensive investigation into the effects of longitudinal design and analysis factors on modelling nonlinear change, I designed simulation studies to address these gaps.

1.6 Methods of Modelling Nonlinear Patterns of Change Over Time

Because my simulation experiments assumed change over time to be nonlinear, it is important to provide an overview of how nonlinear change can be modelled. On this note, I will provide an overview of two commonly employed methods for modelling nonlinear change: 1) the polynomial approach and 2) the nonlinear function approach.[7,8]

---

[7]It should be noted that nonlinear change can be modelled in a variety of ways, with latent change score models (e.g., O'Rourke et al., 2022) and spline models (e.g., Fine & Grimm, 2020) offering some examples.

[8]The definition of a nonlinear function is mathematical in nature. Specifically, a nonlinear function contains at least one parameter that exists in its corresponding partial derivative (at any order). For example, in the logistic function $\theta + \frac{\alpha - \theta}{1 + exp(\frac{\beta - t}{\gamma}})$ is nonlinear because β exists in $\frac{\partial y}{\partial \beta}$ (in addition to γ

Importantly, the simulation experiments in my dissertation will use the nonlinear function approach to model nonlinear change.

Consider an example where an organization introduces a new incentive system with the goal of increasing the motivation of its employees. To assess the effectiveness of the incentive system, employees provide motivation ratings every month over a period of 360 days. Over the 360-day period, the motivation levels of the employees increase following an s-shaped pattern of change over time. One analyst decides to model the observed change using a *polynomial function* shown below in Equation 1.1:

$$y = a + bx + cx^2 + dx^3. \tag{1.1}$$

A second analyst decides to model the observed change using a *logistic function* shown below in Equation 1.2:

$$y = \theta + \frac{\alpha - \theta}{1 + e^{\frac{\beta - time}{\gamma}}} \tag{1.2}$$

Figure 1.3A shows the response pattern predicted by the polynomial function of Equation 1.1 with the estimated values of each parameter (a, b, c, and d) and Figure 1.3B shows the response pattern predicted by the logistic function (Equation 1.2) along with the values estimated for each parameter (θ, α, β, and γ). Although the logistic and polynomial

---

existing in its corresponding partial derivative). The nth order polynomial function of $y = a + bx + cx^2 + ... + nx^n$ is linear because the partial derivatives with respect to any of the parameters (i.e., $1, x^2, ..., x^n$) never contain the associated parameter.

Figure 1.3

Response Patterns Predicted by Polynomial (Equation 1.1) and Logistic (Equation 1.2) Functions

A: Response Pattern Predicted by Polynomial (Linear) Function

$$y = a + bx + cx^2 + dx^3$$

$a = 3.04$
$b = -0.00244$
$c = 2.58e{-}05$
$d = -4.77e{-}08$

B: Response Pattern Predicted by Logistic (Nonlinear) Function

$$y = \theta + \frac{\alpha - \theta}{1 + e^{\left(\frac{\beta - time}{\gamma}\right)}}$$

$\theta = 3.00$
$\alpha = 3.32$
$\beta = 180.00$
$\gamma = 20.00$
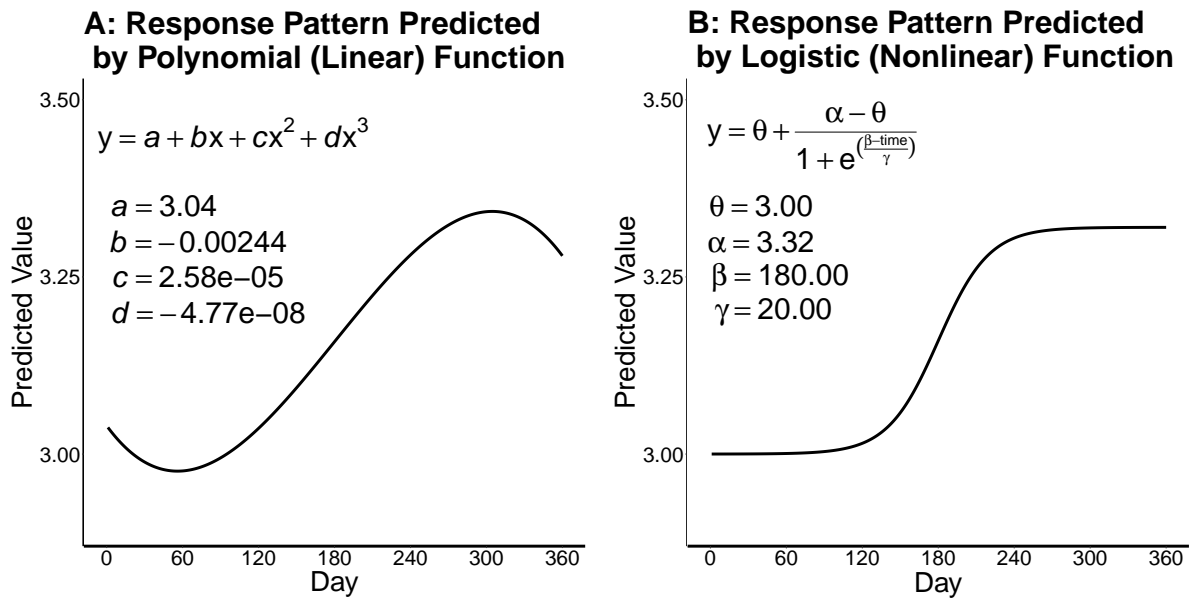
Note. Panel A: Response pattern predicted by the polynomial function of Equation (1.1). Panel B: Response pattern predicted by the logistic function of Equation (1.2).

functions predict nearly identical response patterns, the parameters of the logistic function have the following meaningful interpretations (see Figure 1.4):

- θ specifies the value at the first plateau (i.e., the starting value), and so is called the *baseline* parameter (see Figure 1.4A).

- α specifies the value at the second plateau (i.e., the ending value), and so is called the the *maximal elevation* parameter (see Figure 1.4B).

- β specifies the number of days required to reach half the difference between the first and second plateau (i.e., the midway point), and so is called the *days-to-halfway-elevation* parameter (see Figure 1.4C).

- γ specifies the number of days needed to move from the midway point to approximately 73% of the difference between the starting and ending values (i.e., satiation point), and so is called the *triquarter-halfway delta* parameter (see Figure 1.4D).

26

Figure 1.4

Description Each Parameters Logistic Function (Equation 1.2) Functions

A : **Baseline** (θ)

$$y = \theta + \frac{\alpha - \theta}{1 + e^{(\frac{\beta - time}{\gamma})}}$$

Baseline (y–axis)

$\theta = 3.00$

B : **Maximal Elevation** (α)

$$y = \theta + \frac{\alpha - \theta}{1 + e^{(\frac{\beta - time}{\gamma})}}$$

Maximal elevation (y–axis)

$\alpha = 3.32$

C : **Days to Halfway Elevation** (β)

$$y = \theta + \frac{\alpha - \theta}{1 + e^{(\frac{\beta - time}{\gamma})}}$$

Days to halfway elevation (x–axis)

$\beta = 180.00$

D : **Triquarter–Halfway Delta** (γ)

$$y = \theta + \frac{\alpha - \theta}{1 + e^{(\frac{\beta - time}{\gamma})}}$$

Triquarter–halfway delta (x–axis)

$\gamma = 20.00$

Curve Value

Day
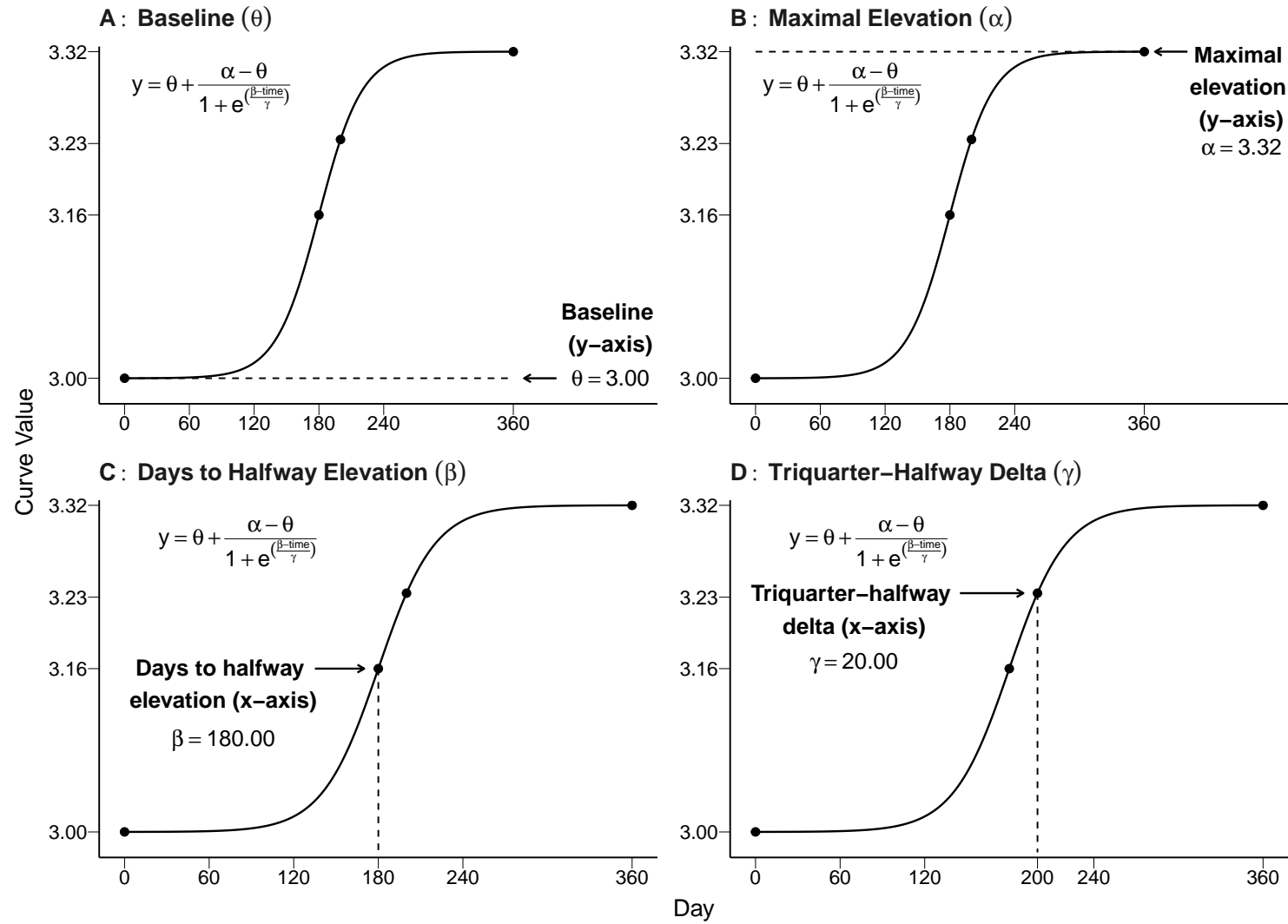
Note. Panel A: The baseline parameter (θ) sets the starting value of the of curve, which in the current example has a value of 3.00 $(\theta = 3.00)$. Panel B: The

maximal elevation parameter (α) sets the ending value of the curve, which in the current example has a value of 3.32 (α = 3.32). Panel C: The days-to-halfway elevation parameter (β) sets the number of days needed to reach 50% of the difference between the baseline and maximal elevation values. In the current example, the baseline-maximal elevation difference is 0.32 ($\alpha - \theta$ = 3.32 - 3.00 = 0.32), and so the days-to-halfway elevation parameter defines the number of days needed to reach a value of 3.16. Given that the days-to-halfway elevation parameter is set to 180 in the current example ($\beta = 180.00$), then 180 days are needed to go from a value of 3.00 to a value of 3.16. Panel D: The triquarter-halfway delta parameter (γ) sets the number of days needed to go from halfway elevation to approximately 73% of the baseline-maximal elevation difference of 0.32 ($\alpha - \theta$ = 3.32 - 3.00 = 0.32). Given that 73% of the baseline-maximal elevation difference is 0.23 and the triquarter-halfway delta is set to 20 days ($\gamma = 20.00$), then 20 days are needed to go from the halfway point of 3.16 to the triquarter point of approximately 3.23).

Applying the parameter meanings of the logistic function to the parameter values estimated by using the logistic function (Equation 1.2), the predicted response pattern begins at a value of 3.00 (baseline) and reaches a value of 3.32 (maximal elevation) by the end of the 360-day period. The midway point of the curve is reached after 180.00 days (days-to-halfway elevation) and the satiation point is reached 20.00 days later (triquarter-halfway delta; or 200.00 days after the beginning of the incentive system is introduced). When looking at the polynomial function, it is almost impossible to meaningfully interpret the values of any of the other parameter values (aside from the 'a' parameter, which indicates the starting value). Therefore, using a nonlinear function such as the logistic function provides a meaningful way to interpret nonlinear change.

1.7 Multilevel and Latent Variable Approach

In addition to using the logistic function to model nonlinear change, another modelling decision concerns whether to do so using the multilevel or latent growth curve framework. In my dissertation, I opted for the latent growth curve framework for two reasons. First, the latent growth curve framework allows data to be more realistically modelled than the multilevel framework. As some examples, the latent growth curve framework allows the modelling of measurement error, complex error structures, and time-varying covariates (for a review, see McNeish & Matta, 2017). Second, and perhaps more important, the likelihood of convergence with multilevel models decreases as the number of random-effect parameters increases due to nonpositive definitive covariance matrices (for a review, see McNeish & Bauer, 2020). With the model I used in my simulation experiments having four random-effect parameters, it is likely that my simulation experiments

would have considerable convergence issues if they use the multilevel framework. Therefore, given the convergence issues of multilevel models and the shortcoming realistically modelling data, I decided, on balance, that the strengths of the multilevel framework (e.g., more options for modelling small samples) were outweighed by its shortcomings, and decided to use a latent growth curve framework in my simulation experiments.

1.7.1 Next Steps

Given that longitudinal research is needed to understand the temporal dynamics of psychological processes, it is necessary to understand how longitudinal design and analysis factors interact with each other (and with sample size) in affecting the accuracy with which nonlinear patterns of change are modelled. With no study to my knowledge having conducted a comprehensive investigation into how longitudinal design and analysis factors affect the modelling of nonlinear change patterns, my simulation experiments are designed to address these gaps in the literature. Specifically, my simulation experiments investigate how measurement number, measurement spacing, and time structuredness affect the accuracy with which a nonlinear change pattern is modelled (see Cells 1, 5, 9, and 10 of Table 1.1/Table 1.2).

1.8 Overview of Simulation Experiments

To investigate the effects of longitudinal design and analysis factors on modelling accuracy, I conducted three Monte Carlo experiments. Before summarizing the simulation experiments, one point needs to be mentioned regarding the maximum number of independent variables used in each experiment. No simulation experiment manipulated more than three variables because of the difficulty associated with interpreting interactions

between four or more variables. Even among academics, the ability to correctly inter-pret interactions sharply declines when the number of independent variables increases from three to four (Halford et al., 2005). Therefore, none of my simulation experiments manipulated more than three variables so that results could be readily interpreted.

To summarize the three simulation experiments, the independent variables of each simulation experiment are listed below:

- Experiment 1: number of measurements, spacing of measurements, and nature of change.

- Experiment 2: number of measurements, spacing of measurements, and sample size.

- Experiment 3: number of measurements, sample size, and time structuredness.

The sections that follow will present each of the simulation experiments and their corresponding results.

2 Experiment 1

In Experiment 1, I investigated the number of measurements needed to obtain high model performance for the estimation of each logistic function parameter (i.e., unbiased and precise estimation) under different spacing schedules and natures of change. Before presenting the results of Experiment 1, I present my design and analysis goals. For my design goals, I conducted a 4 (measurement spacing:equal, time-interval increasing, time-interval decreasing, middle-and-extreme) x 4 (number of measurements: 5, 7, 9, 11) x 3 (nature of change: population value for the fixed-effect days-to-halfway elevation parameter [β_{fixed}] of 80, 180, or 280) study. For my analysis goals, I was interested in answering two questions. First, I was interested in whether placing measurements near periods of change increases model performance. To answer my first question, I determined whether

model performance under each spacing schedule increased when measurements were taken closer to periods of change.

Second, I was interested in how to space measurements when the nature of change is unknown. When the nature of change is unknown, this translates to a situation where a researcher has little to no knowledge of how change unfolds over time, and so any nature of change is a viable candidate for the true change. Therefore, to determine how to space measurements when the nature of change is unknown, I averaged the model performance of each spacing schedule across all possible nature-of-change curves and considered the spacing schedule with the highest model performance to be the best one.

2.1 Methods

2.1.1 Overview of Data Generation

2.1.1.1 Function Used to Generate Each Data Set

Data for each simulation experiment were generated using R (RStudio Team, 2020). To generate the data, the *multilevel logistic function* shown below in Equation (2.1) was used:

$$y_{ij} = \theta_j + \frac{\alpha_j - \theta_j}{1 + e^{\frac{\beta_j - time_i}{\gamma_j}}} + \epsilon_{ij}, \tag{2.1}$$

where θ represents the baseline parameter, α represents the maximal elevation parameter, β represents the days-to-halfway elevation parameter, and γ represents triquarter-halfway delta parameter. Note that, values for θ, α, β, and γ were generated for each j person across all i time points, with an error value being randomly generated at each i time point(ϵ_{ij}; see Figure 1.4 for a review of each parameter). In other words, unique response

32

patterns were generated for each person in each generated data set. Importantly, 1000 data sets were generated per cell.

The logistic growth function (Equation 2.1) was used because it is a common pattern of organizational change [or institutionalization; Lawrence et al. (2001)]. Institutionalization curves follow an s-shaped pattern (i.e., logistic growth), and so their rates of change can be represented by the days-to-halfway elevation and triquarter-halfway delta parameters (β, γ, respectively), and the success of the change can be defined by the magnitude of the difference between baseline and maximal elevation parameters (α - θ, respectively).

2.1.1.2 Population Values Used for Function Parameters

Table 2.1 lists the parameter values that were used for the population parameters. Given that the decisions for setting the values for the baseline, maximal elevation, and residual variance parameters were informed by past research, the discussion that follows highlights how these decisions were made. The difference between the baseline and maximal elevation parameters (θ and α, respectively) corresponded to the effect size most commonly observed in organizational research [i.e., the 50th percentile effect size value; Bosco et al. (2015)]. Because the meta-analysis of Bosco et al. (2015) computed effect sizes as correlations, the 50th percentile effect size value of $r = .16$ was computed to a standardized effect size using the following conversion function shown in Equation 2.2 (Borenstein et al., 2009, Chapter 7):

$$d = \frac{2r}{\sqrt{1 - r^2}}, \tag{2.2}$$

33

₅₂₄ where r is the correlation effect size. Using Equation 2.2, a correlation value of $r = .16$

₅₂₅ becomes a standardized effect size value of $d = 0.32$. For the value of the residual variance

₅₂₆ parameter (ϵ), Coulombe et al. (2016) set it to the value used for the intercept variance

₅₂₇ parameter. In the current context, the intercept of the logistic function (Equation 2.1)

₅₂₈ is the baseline parameter (θ).[9] Given that the value for the variability of the baseline

₅₂₉ parameter was 0.05 (albeit in standard deviation units), the value used for the residual

₅₃₀ variance parameter was 0.05 ($\epsilon = 0.05$). Importantly, because Coulombe et al. (2016)

₅₃₁ set covariances between parameters to zero, all the simulation experiments used zero-

₅₃₂ value covariances. Because justification for the other parameters could not be found in

₅₃₃ any of the simulation studies identified in my systematic review, values set for the other

₅₃₄ parameters were largely arbitrary.

₅₃₅ Two last brief points need to be mentioned about how data were generated to

₅₃₆ facilitate the interpretation of the results. First, data were generated to take on units

₅₃₇ similar to that of a Likert scale (range of 1–5) by assuming a standard deviation of

₅₃₈ 1.00. Thus, previously established effect size of $d = 0.32$ standard deviations implies

₅₃₉ an effect size of 0.32 units. Second, change was assumed to occur over a period of 360

₅₄₀ days because many organizational processes are often governed by annual events (e.g.,

₅₄₁ performance reviews, annual returns, regulations, etc.).

₅₄₂ **2.1.2 Modelling of Each Generated Data Set**

₅₄₃ Previously, I described how data were generated. Here, I describe how the generated

₅₄₄ data were modelled.

---

[9]The definition of an intercept parameter is the value of a curve when no time has elapsed, and this is precisely the definition of the baseline parameter (θ). Therefore, the variance of the intercept parameter carries the same meaning as the variance of the baseline parameter (θ_{random}).

545   Each data set generated by the multilevel logistic function (Equation 2.1) was an-

546 alyzed using a modified latent growth curve model known as a structure latent growth

547 curve model (Preacher & Hancock, 2015).

Table 2.1

Values Used for Multilevel Logistic Function Parameters

Parameter Means	Value
Baseline, θ	3.00
Maximal elevation, α	3.32
Days-to-halfway elevation, β	180.00
Triquarter-halfway delta, γ	20.00

Variability and Covariability Parameters (in Standard Deviations)	
Baseline standard deviation, ψ_θ	0.05
Maximal elevation standard deviation, ψ_α	0.05
Days-to-halfway elevation standard deviation, ψ_β	10.00
Triquarter-halfway delta standard deviation, ψ_γ	4.00
Baseline-maximal elevation covariability, $\psi_{\theta\alpha}$	0.00
Baseline-days-to-halfway elevation covariability, $\psi_{\theta\beta}$	0.00
Baseline-triquarter-halfway delta covariability, $\psi_{\theta\gamma}$	0.00
Maximal elevation-days-to-halfway elevation covariability, $\psi_{\alpha\beta}$	0.00
Maximal elevation-triquarter-halfway delta covariability, $\psi_{\alpha\gamma}$	0.00
Days-to-halfway elevation-triquarter-halfway delta covariability, $\psi_{\beta\gamma}$	0.00
Residual standard deviation, ψ_ϵ	0.05

Note. The difference between α and θ corresponds to the 50th percentile Cohen's d value of 0.32 in organizational psychology (Bosco et al., 2015).

548 Importantly, the model fit to each generated data set estimated nine parameters: A fixed-

549 effect parameter for each of the four logistic function parameters, a random-effect param-

550 eter for each of the four logistic function parameters, and an error parameter. As with

551 a multilevel model, a fixed-effect parameter has a constant value across all individuals,

552 whereas a random-effect parameter represents the variability of values across all modelled

35

people.[10] To fit the logistic function to a given data set (Equation 2.1), a linear approximation of the logistic function was needed so that it could fit within the linear nature of structural equation modelling framework.[11] To construct a linear approximation of the logistic function, a first-order Taylor series was constructed for the logistic function. For a detailed explanation of how the logistic function was fit into the structural equation modelling framework, see Appendix **??** for an explanation of the model and Appendix **??** for the code used to create the model.

2.1.3 Variables Used in Simulation Experiment

2.1.3.1 Independent Variables

To build on current research, Experiment 1 used independent variable manipulations from a select number of previous studies. In looking at the summary of the simulation literature in Table 1.2, the study by Coulombe et al. (2016) was the only one to investigate three longitudinal issues of interest to my dissertation, and so represented the most comprehensive investigation. Because I was also interested in investigating measurement spacing, manipulations were inspired from the only other simulation study identified by my systematic review to manipulate measurement spacing (the study by Timmons & Preacher, 2015). The sections that follow will discuss each of the variables manipulated in Experiment 1.

---

[10]Estimating a random-effect for a parameter allows person- or data-point-specific values to be computed for the parameter.

[11]The logistic function (Equation 2.1) is a nonlinear function and so cannot be directly inserted into the structural equation modelling framework because this framework only allows linear computations of matrix-matrix, matrix-vector, and vector-vector operations. Unfortunately, the algebraic operations permitted in a linear framework cannot directly reproduce the operations in the logistic function (Equation 2.1) and so a linear approximation of the logistic function must be constructed so that the logistic function can be inserted into the structural equation modelling framework.

36

2.1.3.1.1 Spacing of Measurements

The only simulation study identified by my systematic review that manipulated measurement spacing was Timmons and Preacher (2015). Measurement spacing in Timmons and Preacher (2015) was manipulated in the following four ways:

1) *Equal spacing*: measurements are divided by intervals of equivalent lengths.

2) *Time-interval increasing spacing*: intervals that divide measurements increase in length over time.

3) *Time-interval decreasing spacing*: intervals that divide measurements decrease in length over time.

4) *Middle-and-extreme spacing*: measurements are clustered near the beginning, middle, and end of the data collection period.

To maintain consistency with the established literature, I manipulated measurement spacing in the same way as Timmons and Preacher (2015) presented above. Importantly, because Timmons and Preacher (2015) did not create their measurement spacing schedules with any systematicity, I developed a novel and replicable procedure for generating measurement schedules for each of the four measurement spacing conditions, which is described in Appendix **??**. I also automated the generation of measurement schedules by creating a set of functions in R (RStudio Team, 2020).

Table 2.2 lists the measurement days that were used for all measurement spacing-measurement number cells. The first column lists the type of measurement spacing (i.e., equal, time-interval increasing, time-interval decreasing, or middle-and-extreme); the second column lists the number of measurements (5, 7, 9, or 11); the third column lists the measurement days that correspond to each measurement number-measurement spacing

37

condition; and the fourth column lists the interval lengths between the measurements. Note that the interval lengths are equal for equal spacing, increase over time for time-interval increasing spacing, and decrease over time for time-interval decreasing spacing. For cells with middle-and-extreme spacing, the measurement days and interval lengths in the middle of the measurement window have been emboldened.

2.1.3.1.2 Number of Measurements

The smallest measurement number value in Coulombe et al. (2016) of three measurements could not be used in Experiment 1 (or any other simulation experiment that manipulated measurement number in my dissertation) because doing so would have created non-identified models The model used in my simulations estimated 9 parameters (p = 9; 4 fixed-effects + 4 random-effects + 1 error)[12] and so the minimum number of measurements (or observed variables) required for model identification (and to allow model comparison) was 4. Although a measurement number of three could not be used in my manipulation of measurement number, the next highest measurement number values in Coulombe et al. (2016) of 5, 7, and 9 were used. Importantly, a larger value of 11 was added to test for a possible effect of a high measurement number. Therefore, my simulation experiments used the following values in manipulating the number of measurements: 5, 7, 9, and 11.

---

[12]Degrees of freedom is calculated by multiplying the number of observed variables (p) by $p + 1$ and dividing it by 2 ($\frac{p[p+1]}{2}$; Loehlin & Beaujean, 2017).

Table 2.2

Measurement Days Used for All Measurement Number-Measurement Spacing Conditions

Spacing Schedule	Number of Measurements	Measurement Days	Interval Lengths
Equal	5	0, 90, 180, 270, 360	90, 90, 90, 90
	7	0, 60, 120, 180, 240, 300, 360	60, 60, 60, 60, 60, 60
	9	0, 45, 90, 135, 180, 225, 270, 315, 360	45, 45, 45, 45, 45, 45, 45, 45
	11	0, 36, 72, 108, 144, 180, 216, 252, 288, 324, 360	36, 36, 36, 36, 36, 36, 36, 36, 36, 36
Time-interval increasing	5	0, 30, 100, 210, 360	30, 70, 110, 150
	7	0, 30, 72, 126, 192, 270, 360	30, 42, 54, 66, 78, 90
	9	0, 30, 64.29, 102.86, 145.71, 192.86, 244.29, 300, 360	30, 34.29, 38.57, 42.86, 47.14, 51.43, 55.71, 60
	11	0, 30, 61.33, 94, 128, 163.33, 200, 238, 277.33, 318, 360	30, 31.33, 32.67, 34, 35.33, 36.67, 38, 39.33, 40.67, 42
Time-interval decreasing	5	0, 150, 260, 330, 360	150, 110, 70, 30
	7	0, 90, 168, 234, 288, 330, 360	90, 78, 66, 54, 42, 30
	9	0, 60, 115.71, 167.14, 214.29, 257.14, 295.71, 330, 360	60, 55.71, 51.43, 47.14, 42.86, 38.57, 34.29, 30
	11	0, 42, 82.67, 122, 160, 196.67, 232, 266, 298.67, 330, 360	42, 40.67, 39.33, 38, 36.67, 35.33, 34, 32.67, 31.33, 30
Middle-and-extreme	5	1, **150, 180, 210**, 360	150, **30, 30**, 150

Table 2.2

Measurement Days Used for All Measurement Number-Measurement Spacing Conditions (continued)

Spacing Schedule	Number of Measurements	Measurement Days	Interval Lengths
	7	1, 30, **150, 180, 210**, 330, 360	30, 120, **30, 30**, 120, 30
	9	1, 30, 60, **150, 180, 210**, 300, 330, 360	30, 30, 90, **30, 30**, 90, 30, 30
	11	1, 30, 60, **120, 150, 180, 210, 240,** 300, 330, 360	30, 30, 60, **30, 30, 30, 30**, 60, 30, 30

Note. For middle-and-extreme spacing levels, the measurement days and and interval lengths corresponding to the middle of measurement windows have been emboldened.

2.1.3.1.3 Population Values Set for The Fixed-Effect Days-to-Halfway Elevation Parameter β_{fixed} (Nature of Change)

The nature of change was manipulated by setting the days-to-halfway elevation parameter (β_{fixed}) to a value of either 80, 180, or 280 days (see Figure 1.4A). Note that no other study in my systematic review manipulated nature of change using logistic curves and so its manipulation in Experiment 1 is, to the best of my knowledge, unique. Importantly, nature of change was manipulated to simulate situations where uncertainty exists in how change unfolds over time.

2.1.3.2 Constants

Given that each simulation experiment manipulated no more than three independent variables so that results could be readily interpreted (Halford et al., 2005), other variables had to be set to constant values. In Experiment 1, two important variables were set to constant values: sample size and time structuredness. For sample size, I set the value across all cells to the average sample size used in organizational research ($n = 225$; Bosco et al., 2015). For time structuredness, data across all cells were generated to be time structured (i.e., all participants provide data according to one response pattern; that is, at each time point, participants provide their data at the exact same moment).

2.1.3.3 Dependent Variables

2.1.3.3.1 Convergence Success Rate

The proportion of iterations in a cell where models converged defined the *convergence success rate.*[13] Equation (2.3) below shows the calculation used to compute the

---

[13]Specifically, convergence was obtained if the convergence code returned by OpenMx was 0.

convergence success rate:

$$\text{Convergence success rate} = \frac{\text{Number of models that successfully converged in a cell}}{n},$$

(2.3)

where n represents the total number of models run in a cell.

2.1.3.3.2 Model Performance

Model performance was the combination of two metrics: bias and precision. More specifically, two questions were of importance in the estimation of a given logistic function parameter: 1) How well was the parameter estimated on average (bias) and 2) what was a range of values that could be expected for an estimate from the output of a single model (precision). In the two sections that follow, I will discuss each metric of model performance and the cutoffs used to determine whether estimation was unbiased and precise.

2.1.3.3.2.1 Bias

Bias was calculated to evaluate the accuracy with which each logistic function parameter was estimated in each experimental cell. As shown below in Equation (2.4), *bias* was obtained by summing the differences between the population value set for a parameter and the value estimated for the parameter by each i converged model and then dividing the sum by the number of N converged models.

$$\text{Bias} = \frac{\sum_i^N \left(\text{Population value for parameter} - \text{Average estimated value}_i\right)}{N}$$

(2.4)

42

Bias was calculated for the fixed- and random-effect parameters of the baseline (θ_{fixed}, θ_{random}), maximal elevation (α_{fixed}, α_{random}), days-to-halfway elevation (β_{fixed}, β_{random}), and the triquarter-halfway delta parameters (γ_{fixed}, γ_{random}) and the error parameter (ϵ).

2.1.3.3.2.2 Precision

In addition to computing bias, precision was calculated to evaluate the variability with which each parameter was estimated. Importantly, metrics used to evaluate precision in previous studies aoften ssume estimates are normally distributed (e.g., mean-squared error and empirical standard error). Because some parameters in my simulations had skewed distributions, using a metric that assumed a normal distribution would likely yield inaccurate results. Correspondingly, I used a distribution-independent definition of precision. In my simulations, *precision* was defined as the range of values covered by the middle 95% of values estimated for a logistic parameter.

2.1.4 Analysis of Data Modelling Output and Accompanying Visualizations

To analyse and visualize modelling performance, I calculated values for convergence success rate, bias, and precision in each experimental cell (see dependent variables). The sections that follow provide details on how I analysed each dependent variable and constructed plots to visualize bias and precision.

2.1.4.1 Analysis of Convergence Success Rate

For the analysis of convergence success rate, the mean convergence success rate was computed for each cell in each experiment (see section on convergence success rate). Because convergence rates exhibited little variability across cells due to the nearly unanimous high rates (almost all cells across all experiments had convergence success rates

above 90%), examining the effects of any independent variable on these values would have provided little information. Therefore, I only reported the average convergence success rate for each cell (see Appendix **??**).

**2.1.4.2 Analysis and Visualization of Bias**

In accordance with several simulation studies, an estimate with a bias value within a $\pm 10\%$ margin of error of the parameter's population value was deemed unbiased (Muthén et al., 1997). To visualize bias, I constructed bias/precision plots. Figure 2.1 shows a bias/precision plot for the fixed-effect triquarter-halfway parameter (γ_{fixed}) for each measurement number and nature of change. The dots (squares, circles, triangles, diamonds) indicate the average estimated value (see bias). The horizontal blue line indicates the population value ($\gamma_{fixed} = 4.00$) and the gray band indicates the acceptable margin of error of $\pm 10\%$ of the parameter's population value. Dots that lie within the gray margin of error are filled and dots that lie outside of the margin remain unfilled. In the current example, the average value estimated for the fixed-effect triquarter-halfway delta parameter (γ_{fixed}) is only biased (i.e., lies outside the margin of error) with five measurements with a nature-of-change value of 80 ($\beta_{fixed} = 80$). Therefore, estimates are unbiased in almost all cells.

2.1.4.3 Analysis and Visualization of Precision

As discussed previously, precision was defined as the range of values covered by the middle 95% of estimated values for a given parameter (see precision). The cutoff value used to estimate precision directly followed from the cutoff value used for bias. Given that bias values within a $\pm 10\%$ of a parameter's population value were deemed acceptable, an acceptable value for precision should not allow any bias value above the $\pm 10\%$ cutoff.

44

⁶⁹⁵ That is, the range covered by the middle 95% of estimated values should not contain a

⁶⁹⁶ bias value outside the ±10% cutoff. If the range of values covered by the middle 95% of

⁶⁹⁷ estimate values is conceptualized as an error bar centered on the population value, then

⁶⁹⁸ an acceptable value for precision implies that neither

Figure 2.1

Bias/Precision Plot for the Fixed-Effect Days-to-Halfway Elevation Parameter (γ_{fixed})
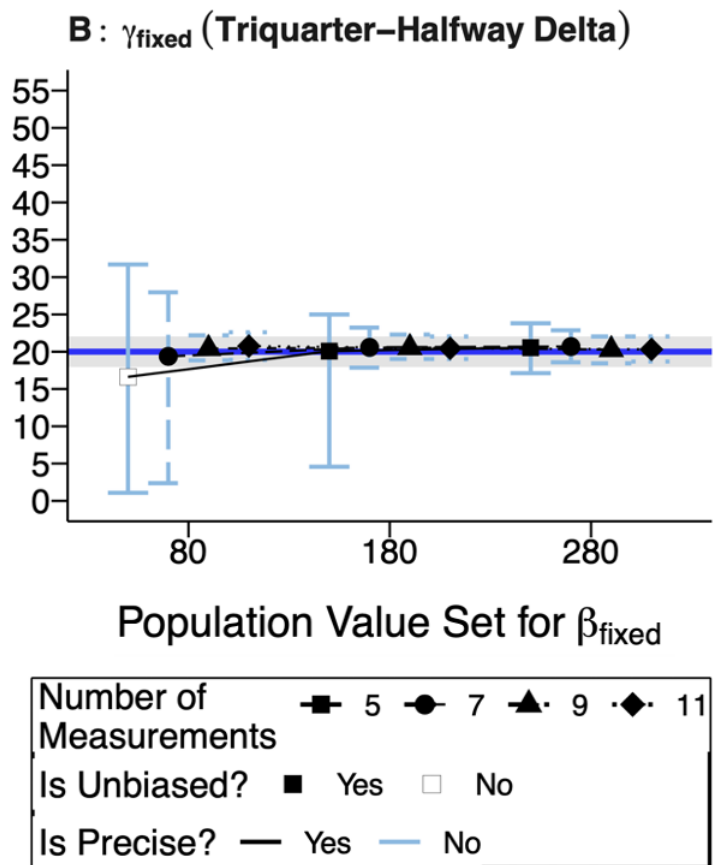
⁶⁹⁹ *Note.* Dots (squares, circles, triangles, diamonds) indicate the average estimated value and error bars show

⁷⁰⁰ the range of values covered by the middle 95% of the estimated values (see Precision). The horizontal blue

⁷⁰¹ line indicates the population value (γ_{fixed} = 4.00) and the gray band indicates the acceptable margin of error

⁷⁰² (i.e., ±10% of the population value) for bias. Dots that lie outside of the margin of error are unfilled and are

⁷⁰³ considered biased estimates. Dots that lie inside the margin of error are filled and considered unbiased

⁷⁰⁴ estimates. Error bars whose upper and/or lower whisker lengths exceed 10% of the parameter's population

⁷⁰⁵ value are light blue and indicate parameter estimation that is not precise. Error bars whose upper and/or

⁷⁰⁶ lower whisker lengths do not exceed 10% of the parameter's population value are black and indicate

⁷⁰⁷ parameter estimation that is precise.

45

the lower nor upper whiskers have a length greater than 10% of the parameter's population value. In summary, I deemed precision acceptable if no estimate within the range of values covered by the middle 95% of estimated values had a bias value greater than 10% of the population value, which also means that neither the lower nor upper whiskers of the error bar have a length greater than 10% of the population value.

Like bias, I also depicted precision in bias/precision plots using error bars. Each error bar in the bias/precision plot of Figure 2.1 indicates the range of values covered by the middle 95% of estimated values in the given cell for the fixed-effect triquarter-halfway delta parameter (γ_{fixed}). Importantly, if estimation is not precise, then at least one of the lower and/or upper whisker lengths exceeds 10% of the parameter's population value. When estimation is not precise, the error bar is light blue. When estimation is precise (i.e., neither of the lower or upper whisker lengths exceed 10% of the parameter's population value), the corresponding error bar is black. In the current example, all error bars are light blue and so precision is low in all cells.

2.1.4.3.1 Effect Size Computation for Precision

One last statistic I calculated was an effect size value to estimate the variance in parameter estimates accounted for by each effect. Among the several effect size metrics—at a broad level, effect size metrics can represent standardized differences or variance-accounted-for measures that are corrected or uncorrected for sampling error—the corrected variance-accounted-for effect size metric of partial ω^2 was chosen because of three desirable properties. First, partial ω^2 provides a less biased estimate of effect size than other variance-accounted-for measures (Okada, 2013). Second, partial ω^2 is more robust to assumption violations of normality and homogeneity of variance (Yigit & Mendes,

46

$_{731}$ 2018). Given that parameter estimates were often non-normally distributed across cells,

$_{732}$ effect size values computed with partial ω^2 should be relatively less biased than other

$_{733}$ variance-accounted-for effect size metrics (e.g., η^2). Third, partial ω^2 provides an effect

$_{734}$ size estimate that is not diluted by the inclusion of unaccountable variance in the denom-

$_{735}$ inator. To compute partial ω^2 value for each experimental effect, Equation 2.5 shown

$_{736}$ below was used:

$$\text{partial}\omega^2 = \frac{\sigma^2_{effect}}{\sigma^2_{effect} + MSE} \tag{2.5}$$

$_{737}$ where σ^2_{effect} represents the variance accounted by an effect and MSE is the mean squared

$_{738}$ error. Importantly, σ^2_{effect} values were corrected values obtained by using the following

$_{739}$ formula in Equation 2.6 for a two-way factorial design with fixed variables (Howell, 2009):

$$\sigma^2_{effect} = \frac{(a-1)(MS_{effect} - MS_{error})}{nab}, \tag{2.6}$$

$_{740}$ where a is the number of levels in the effect, b is the number of levels in the second effect,

$_{741}$ and n is the cell size. The variance accounted by the interaction was computed using the

$_{742}$ following formula in Equation 2.7:

$$\sigma^2_{AxB} = \frac{(a-1)(b-1)(MS_{AxB} - MS_{error})}{nab}. \tag{2.7}$$

$_{743}$ To compute partial ω^2 values for effects, a Brown-Forsythe test was computed and the

$_{744}$ appropriate sum-of-squares terms were used to compute partial ω^2 values. A Brown-

$_{745}$ Forsythe test was used to protect against the biasing effects of skewed distributions

(Brown & Forsythe, 1974), which were observed in the parameter estimate distributions in the current simulation experiments. To compute the Brown-Forsythe test, median absolute deviations in each cell were computed by calculating the absolute difference between each i estimate and the median estimated value in the given experimental cell as shown in Equation 2.8 below:

$$\text{Median absolute deviation}_i = |\text{Parameter estimate}_i - \text{Median parameter estimate}_{cell}|.$$

$$(2.8)$$

An ANOVA was then computed on the median absolute deviation values (using the independent variables of the experiment and the associated interactions as predictors), with the terms in Equation 2.5 extracted from the ANOVA output to compute partial ω^2 values.

2.2 Results and Discussion

In the sections that follow, I organize the results by presenting them for each spacing schedule (equal, time-interval increasing, time-interval decreasing, middle-and-extreme). The results are presented for each spacing schedule because answering my research questions first requires knowledge of these results. To answer my first question of whether model performance increases from placing measurements during periods of change, I need to determine whether model performance under each spacing schedule increases when measurements are placed near periods of change. To answer my second question of how to space measurements when the nature of change is unknown, model performance across all manipulated nature-of-change values must first be calculated for each spacing schedule. The spacing schedule that obtains the highest model performance across all

48

nature-of-change values can then be determined as the best schedule to use.

For each spacing schedule, I will first present a concise summary table of the results and then provide a detailed report for each column of the summary table. Because the detailed reports are of considerable length, I provide concise summaries before the detailed reports to establish a framework to help interpret the detailed reports. The detailed report of each spacing schedule presents the results of each day-unit's bias/precision plot, model performance under each nature-of-change value, and then provides a qualitative summary. After providing the results for each spacing schedule, I then use the results to answer my research questions.

2.2.1 Framework for Interpreting Results

To conduct Experiment 1, the three variables of number of measurements (4 levels), measurement spacing (4 levels), and nature of change (3 levels) were manipulated, which yielded a total of 48 cells. Importantly, within each cell, bias and precision values were also computed for each of the nine parameters estimated by the structured latent growth curve model (for a review, see modelling of each generated data set). Thus, because the analysis of Experiment 1 computed values for many dependent variables, interpreting the results can become overwhelming. Therefore, I will provide a framework to help the reader efficiently navigate the results section.

Because I will present the results of Experiment 1 by each level of measurement spacing, the framework I will describe in Figure 2.2 shows a template for the bias/precision plots that I will present for each spacing schedule. The results of each spacing schedule contain a bias/precision plot for each of the nine estimated parameters. Each bias/precision

49

plot shows the bias and precision for the estimation of one parameter across all measurement number and nature-of change levels. Within each bias/precision plot, dots indicate the average estimated value (which indicates bias bias) and error bars represent the middle 95% range of estimated values (which indicates precision). Bias/precision plots with black outlines show the results for day-unit parameters and plots with gray outlines show the results for Likert-unit parameters. Importantly, only the results for the day-unit parameters will be presented (i.e., fixed- and random-effect days-to-halfway elevation and triquarter-halfway delta parameters [β_{fixed}, β_{random}, γ_{fixed}, γ_{random}, respectively]). The results for the Likert-unit parameters (i.e., fixed- and random-effect baseline and maximal elevation parameters [θ_{fixed}, θ_{random}, α_{fixed}, α_{random}, respectively]) were largely trivial and so are presented in Appendix **??**. Therefore, the results of each spacing schedule will only present the bias/precision plots for four parameters (i.e., the day-unit parameters).

### 2.2.2 Pre-Processing of Data and Model Convergence

After collecting the output from the simulations, non-converged models (and their corresponding parameter estimates) were removed from subsequent analyses. Table **??** in Appendix **??** provides the convergence success rates for each cell in Experiment 1. Model convergence was almost always above 90% and convergence rates, with rates only going below 90% in two cells (or instances) with five measurements.

2.2.3 Equal Spacing

For equal spacing, Table 2.3 provides a concise summary of the results for the day-unit parameters (see Figure 2.5 for the corresponding bias/precision plots). The sections that follow will present the results for each column of Table 2.3 and provide elaboration when necessary.

50

Figure 2.2

Set of Bias/Precision Plots Constructed for Each Spacing Schedule in Experiment 1
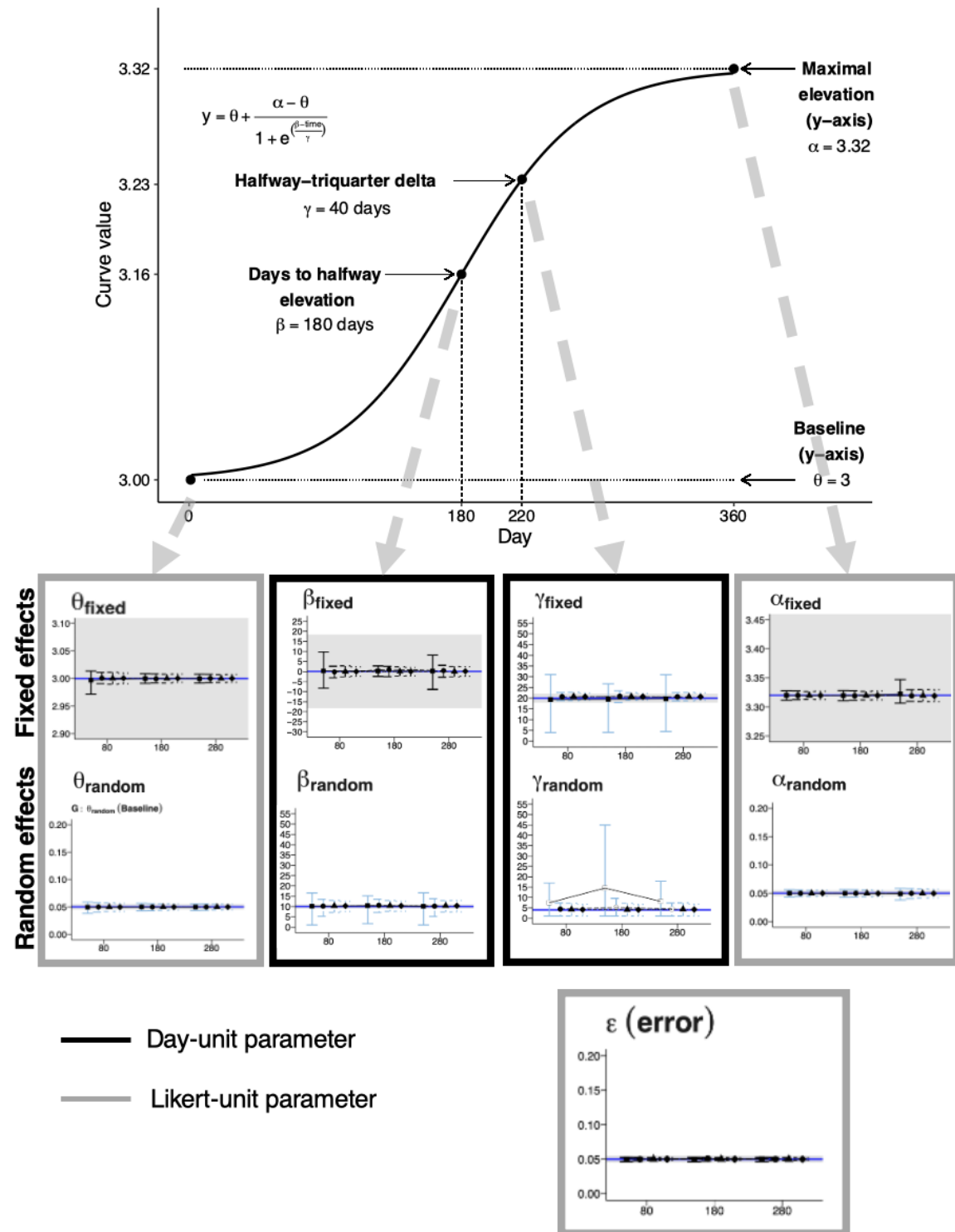
Note. A bias/precision plot is constructed for each parameter of the logistic function (see Equation 2.1).

Bias/precision plots with black borders show the results for day-unit parameters and plots with gray border

show the results for Likert-unit parameters. For each parameter, bias and precision are shown across each

combination of measurement number and nature of change.

51

Before presenting the results for equal spacing, I provide a brief description of the concise summary table created for each spacing schedule and shown for equal spacing below in Table 2.3. Text within the 'Highest Model Performance' column indicates the nature-of-change value that resulted in the highest model performance for each day-unit parameter. Text within the 'Unbiased' and 'Precise' columns indicates the number of measurements that were needed to, respectively, obtain unbiased and precise parameter estimation across all manipulated nature-of-change values. Emboldened text in the 'Unbiased' and 'Qualitative Description' columns indicates the measurement number that, respectively, resulted in unbiased estimation and the greatest improvements in bias and precision across all day-unit parameters and manipulated nature-of-change values. The 'Error Bar Length' column indicates the average error bar length across all manipulated nature-of-change values that resulted from using the measurement number listed in the 'Qualitative Description' column.

2.2.3.1 Nature of Change That Leads to Highest Model Performance

For equal spacing, Table 2.4 lists the precision values (i.e., error bar lengths) for each day-unit parameter across each nature-of-change value. The 'Total' column indicates the total error bar length, which is a sum of the lower ('Lower') and upper ('Upper') whisker lengths. Given that the lower and upper whisker lengths were largely equivalent for each parameter, they were largely redundant and so were not reported for equal spacing. Although model performance was determined by bias and precision, results for bias were not reported because the differences in bias across the nature-of-change values were negligible.

Table 2.3

Concise Summary of Results for Equal Spacing in Experiment 1

Parameter	Highest Model Performance	Unbiased	Precise	Summary	
				Qualitative Description	Error Bar Length
β_{fixed} (Figure 2.5A)	β_{fixed} = 180	All cells	All cells	Largest improvements in precision with **NM = 7**	5.64
γ_{fixed} (Figure 2.5B)	β_{fixed} = 180	All cells	No cells	Largest improvements in precision with **NM = 7**	4.37
β_{random} (Figure 2.5C)	β_{fixed} = 180	All cells	No cells	Largest improvements in precision with **NM = 7**	7.74
γ_{random} (Figure 2.5D)	β_{fixed} = 180	**NM ≥ 9**	No cells	Largest improvements in bias and precision with **NM = 7**	7.02

Note. 'Highest Model Performance' indicates the curve that resulted in the highest model performance (largely determined by precision; see nature of change). Emboldened text in the 'Unbiased' and 'Qualitative Description' columns indicates the number of measurements that, respectively, resulted in unbiased estimation and the greatest improvements in bias and precision across all day-unit parameters (note that acceptable precision was not obtained in the estimation of all day-unit parameters with equal spacing). 'Error Bar Length' indicates the average error bar length value across all nature-of-change values that resulted from using the measurement number in the 'Qualitative Description' column. Parameter names and population values are as follows: β_{fixed} = fixed-effect days-to-halfway elevation parameter = 80, 180, 280; γ_{fixed} = fixed-effect triquarter-halfway delta parameter = 20; β_{random} = random-effect days-to-halfway elevation parameter = 10; γ_{random} = random-effect triquarter-halfway delta parameter = 4. NM = number of measurements.

Note that error bar lengths were obtained by computing the average length across all

manipulated number of measurements. The columns shaded in gray indicate the nature-

of-change value that resulted in the highest precision (i.e., shortest error bar lengths) for

equal spacing. For equal spacing, precision was highest (i.e., shortest error bar lengths)

with a nature-of-change value of 180 for all day-unit parameters with one exception (i.e.,

see the 'Highest Model Performance' column in Table 2.3). Importantly, with a nature-

of-change value of 180, measurements were taken closer to periods of change under equal

spacing than with other nature-of-change values (see Figure 2.3). Therefore, it appears

that placing measurements closer to periods of change increased model performance with

equal spacing.

Table 2.4

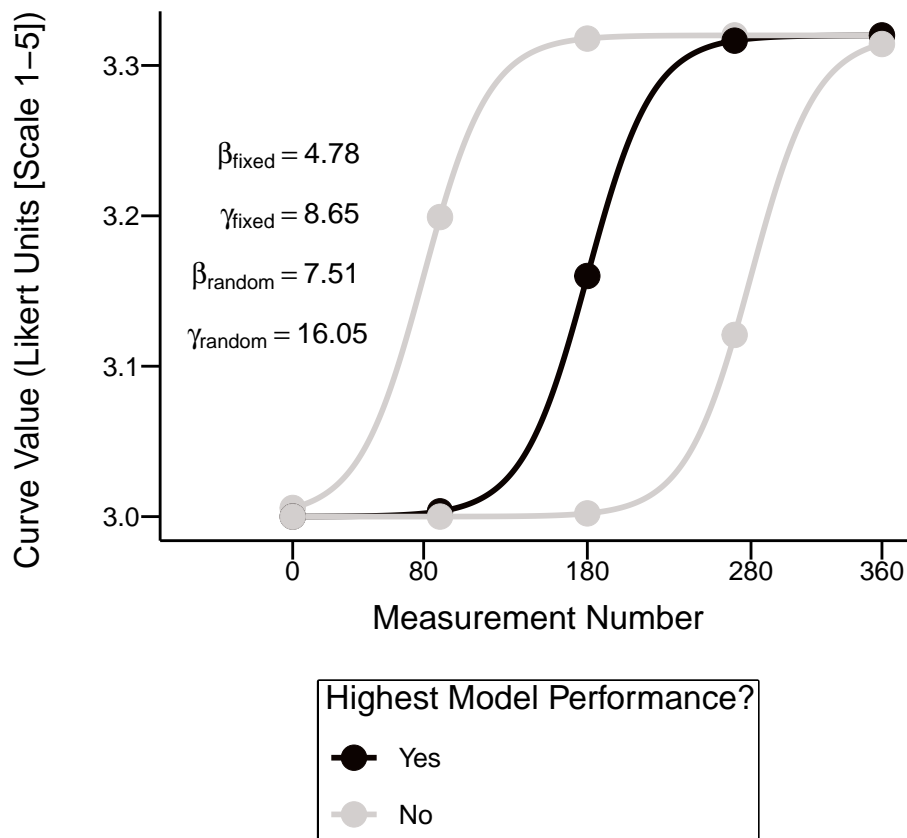*Error Bar Lengths Across Nature-of-Change Values Under Equal Spacing in Experiment 1*

	Population Value of β_{fixed}								
	80			180			280		
Parameter	Lower	Upper	Total	Lower	Upper	Total	Lower	Upper	Total
β_{fixed} (Figure 2.5A)	4.42	4.12	8.54	2.46	2.32	4.78	4.09	4.16	8.25
γ_{fixed} (Figure 2.5B)	4.84	4.69	9.53	4.95	3.7	8.65	4.79	4.65	9.44
β_{random} (Figure 2.5C)	4.74	3.88	8.62	3.96	3.55	7.51	4.77	4.05	8.82
γ_{random} (Figure 2.5D)	3.00	5.52	8.52	3.00	13.05[a]	16.05	3.00	5.78	8.78

Note. 'Total' indicates the total error bar length, which is a sum of the lower ('Lower') and upper ('Upper') whisker lengths. Parameter names and population values are as follows: β_{fixed} = fixed-effect days-to-halfway elevation parameter = 80, 180, 280; γ_{fixed} = fixed-effect triquarter-halfway delta parameter = 20; β_{random} = random-effect days-to-halfway elevation parameter = 10; γ_{random} = random-effect triquarter-halfway delta parameter = 4. Note that error bar lengths were calculated by computing the average error bar length value across all number-of-measurement (NM) values (NM $\in$ {5, 7, 9, 11}). Columns shaded in gray indicate the nature-of-change value that results in the shortest error bar and whisker lengths.

[a] Error bar length is longest in this case because of the existence of high-value outliers (see Figure 2.4).

Figure 2.3
Model Performance Status Across Nature-of-Change Values With Equal Spacing

Note. Model performance was highest when measurements were taken closer to periods of greater change, which resulted with a nature-of-change value of 180 with equal spacing. Text prints error bar lengths that resulted when model performance was highest (see Table 2.4).

55

To understand why precision for the random-effect triquarter-halfway elevation parameter (γ_{random}) was lower with a nature-of-change value of 180, I looked at the distribution of estimated values. Figure 2.4 shows the distribution of values (i.e., density plots) estimated for the random-effect triquarter-halfway elevation parameter (γ_{random}) for each nature-of-change level with five measurements. Importantly, the error bars in the bias/precision plot of Figure 2.5D with five measurements are created from the density plots shown in Figure 2.4. Panel A shows the density plot with a nature-of-change value of 80 ($\beta_{fixed} = 80$). Panel B shows the density plot with a with a nature-of-change value of 180 ($\beta_{fixed} = 180$). Panel C shows the density plot with a with a nature-of-change value of 280 ($\beta_{fixed} = 280$). Regions shaded in gray represent the middle 95% of estimated values and the width of the shaded regions is indicated by the length of the horizontal error bars. As originally confirmed by Table 2.4, Figure 2.4B shows that precision was indeed lowest (i.e., longer error bars) with a nature of change of 180. In looking across the density plots in Figure 2.4, precision was lowest (i.e., longest error bars) for the random-effect triquarter-halfway parameter (γ_{random}) with a nature-of-change value of 180 because of the existence of high-value outliers.
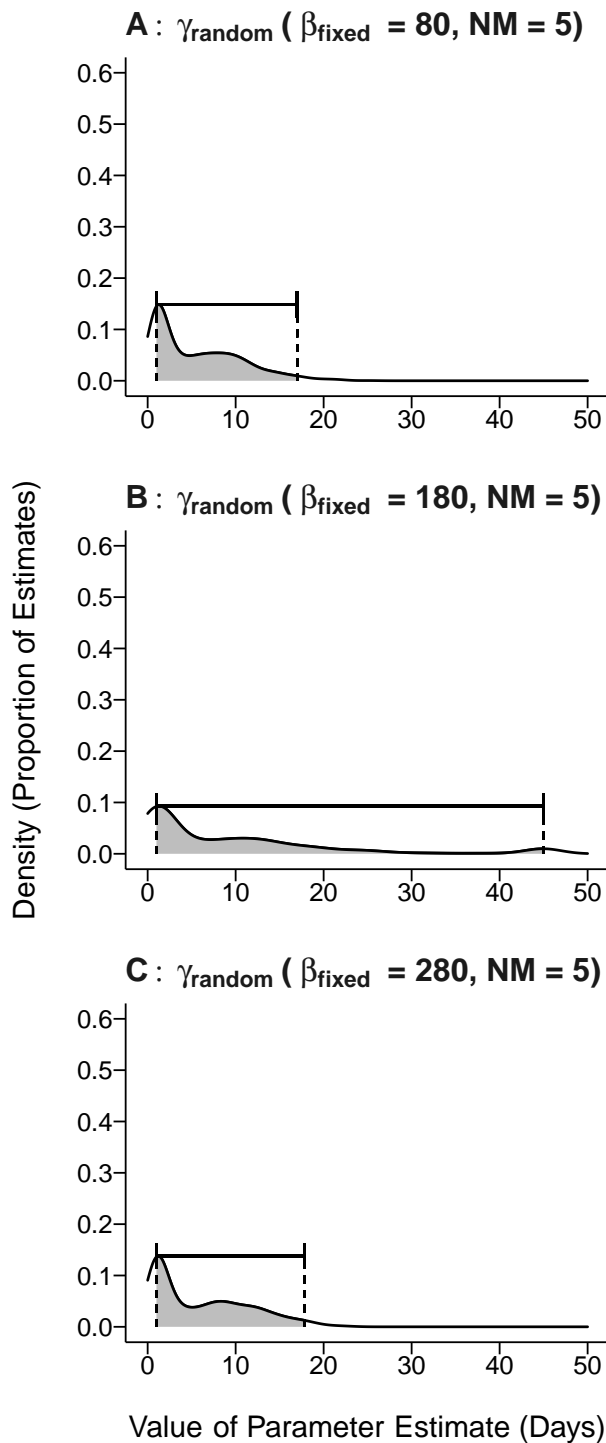
In summary, under equal spacing, model performance for all the day-unit parameters was greatest when the nature-of-change value set by the fixed-effect days-to-halfway elevation parameter (β_{fixed}) had a value of 180. The one exception to this result was that model performance (as indicated by precision) was lower for the random-effect triquarter-halfway elevation parameter (γ_{random}) with a nature-of-change value of 180 because of high-value outliers.

2.2.3.2 Bias

Before presenting the results for bias, I provide a description of the set of bias/precision plots shown in Figure 2.5 and in the results sections for the other spacing schedules in Experiment 1. Figure 2.5 shows the bias/precision plots for each day-unit parameter and Table 2.5 provides the partial ω^2 values for each independent variable of each day-unit parameter. In Figure 2.5, blue horizontal lines indicate the population values for each parameter (with population values of $\beta_{fixed} \in \{80, 180, 280\}$, $\beta_{random} = 10.00$, $\gamma_{fixed} = 20.00$, and $\gamma_{random} = 4.00$). Gray bands indicate the $\pm 10\%$ margin of error for each parameter and unfilled dots indicate cells with average parameter estimates outside of the margin. Error bars represent the middle 95% of estimated values, with light blue error bars indicating imprecise estimation. I considered dots that fell outside the gray bands as biased and error bar lengths with at least one whisker length exceeding the 10% cutoff (i.e., or longer than the portion of the gray band underlying the whisker) as imprecise. Panels A–B show the bias/precision plots for the fixed- and random-effect days-to-halfway elevation parameters (β_{fixed} and β_{random}, respectively). Panels C–D show the bias/precision plots for the fixed- and random-effect triquarter-halfway delta parameters (γ_{fixed} and γ_{random}, respectively). Note that random-effect parameter units are in standard deviation units. Importantly, across all population values used for the fixed-effect days-to-halfway elevation parameter (β_{fixed}), the acceptable amount of bias and precision was based on a population value of 180.

57

Figure 2.4

Density Plots of the Random-Effect Triquarter-Halfway Delta (γ_{random}; Figure 2.5D) With Equal Spacing in Experiment 1 (95% Error Bars)

A : $\gamma_{\textbf{random}}$ ($\beta_{\textbf{fixed}}$ = 80, NM = 5)

B : $\gamma_{\textbf{random}}$ ($\beta_{\textbf{fixed}}$ = 180, NM = 5)

C : $\gamma_{\textbf{random}}$ ($\beta_{\textbf{fixed}}$ = 280, NM = 5)

Density (Proportion of Estimates)

Value of Parameter Estimate (Days)

892 *Note.* Regions shaded in in gray represent the middle 95% of estimated values and the width of the shaded

893 regions is indicated by the length of the horizontal error bars. The error bar length if longest when the

894 nature-of-change value is 180. γ_{random} = random-effect triquarter-halfway delta parameter, with population

895 value of 4.00, NM = number of measurements.

58

With respect to bias for equal spacing, estimates were biased (i.e., above the acceptable 10% cutoff) for each day-unit parameter in the following cells:

- fixed-effect days-to-halfway elevation parameter (β_{fixed}; Figure 2.5A): no cells.

- fixed-effect triquarter-halfway delta parameter (γ_{fixed}; Figure 2.5B): no cells.

- random-effect days-to-halfway elevation parameter (β_{random}; Figure 2.5C): no cells.

- random-effect triquarter-halfway elevation parameter (γ_{random}; Figure 2.5D): five measurements with all manipulated nature-of-change values and seven measurements with nature-of-change values of 180 and 280.

In summary, with equal spacing, estimation of all the day-unit parameters across all manipulated nature-of-change values was unbiased using nine or more measurements, which is indicated by the emboldened text in the 'Unbiased' column of Table 2.3.
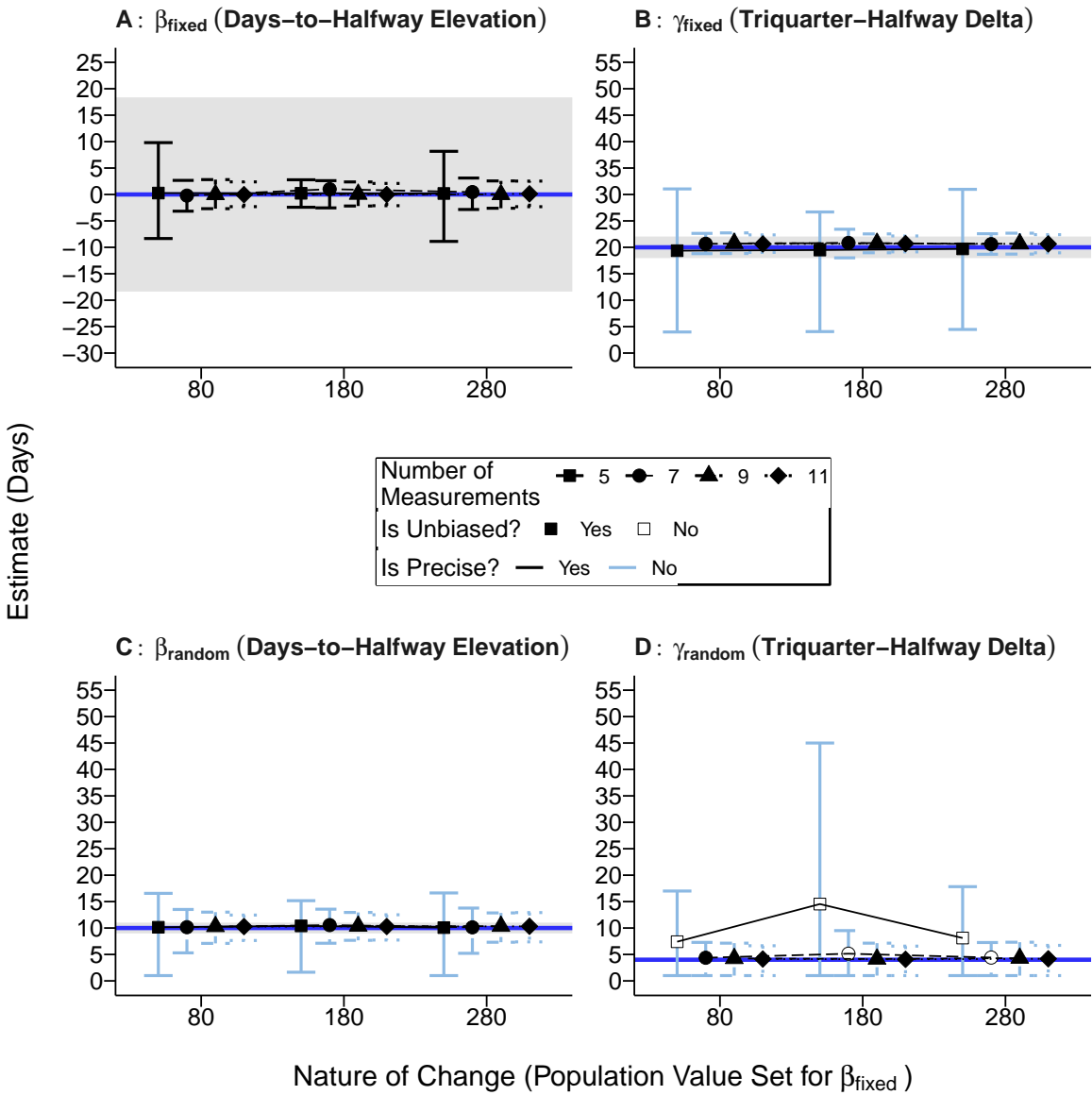
2.2.3.3 Precision

With respect to precision for equal spacing, estimates were imprecise (i.e., error bar length with at least one whisker length exceeding 10% of a parameter's population value) in the following cells for each day-unit parameter:

- fixed-effect days-to-halfway elevation parameter (β_{fixed}; Figure 2.5A): no cells.

- fixed-effect triquarter-halfway delta parameter (γ_{fixed}; Figure 2.5B): all cells.

- random-effect days-to-halfway elevation parameter (β_{random}; Figure 2.5C): all cells.

- random-effect triquarter-halfway delta parameter [γ_{random}] in Figure 2.5D): all cells.

In summary, with equal spacing, estimation across all manipulated nature-of-change values was only precise for the fixed-effect days-to-halfway elevation parameter (β_{fixed}) with five or more measurements. No manipulated measurement number resulted in precise

59

918 estimation of the fixed-effect triquarter-halfway delta parameter (γ_{fixed}) or the random-

919 effect day-unit parameters (see the 'Precise' column of Table 2.3).

Figure 2.5
Bias/Precision Plots for Day-Unit Parameters With Equal Spacing in Experiment 1

920 *Note.* Panel A: Bias/precision plot for the fixed-effect days-to-halfway elevation parameter (β_{fixed}). Panel B:

921 Bias/precision plot for the fixed-effect triquarter-halfway elevation parameter (γ_{fixed}). Panel C: Bias/precision

922 plot for the random-effect days-to-halfway elevation parameter (β_{random}). Panel D: Bias/precision plot for the

923 random-effect triquarter-halfway elevation parameter (γ_{random}). Blue horizontal lines in each panel

924 represent the population value for each parameter. Population values for each day-unit parameter are as

925 follows: $\beta_{fixed} \in$ 80, 180, 280, β_{random} = 10.00, γ_{fixed} = 20.00, γ_{random} = 4.00. Gray bands indicate the

926 $\pm 10\%$ margin of error for each parameter and unfilled dots indicate cells with average parameter estimates

outside of the margin or biased estimates. Error bars represent the middle 95% of estimated values, with

light blue error bars indicating imprecise estimation. I considered dots that fell outside the gray bands as

biased and error bar lengths with at least one whisker length exceeding the 10% cutoff (i.e., or longer than

the portion of the gray band underlying the whisker) as imprecise. Note that random-effect parameter units

are in standard deviation units. Importantly, across all nature-of-change values (i.e., population values used

for β_{fixed}), the acceptable amount of bias and precision was based on a population value of 180. See Table

?? for specific values estimated for each parameter and Table 2.5 for ω^2 effect size values.

Table 2.5

Partial ω^2 Values for Manipulated Variables With Equal Spacing in Experiment 1

Parameter	Effect		
	NM	NC	NM x NC
β_{fixed} (Figure 2.5A)	0.02	0.00	0.01
β_{random} (Figure 2.5B)	0.29	0.02	0.02
γ_{fixed} (Figure 2.5C)	0.36	0.01	0.03
γ_{random} (Figure 2.5D)	0.21	0.03	0.04

Note. NM = number of measurements $\in \{5, 7, 9, 11\}$, NC = nature of change (population value set for $\beta_{fixed} \in \{80, 180, 280\}$), NM x NC = interaction between number of measurements and population value set for β_{fixed}. β_{fixed} = fixed-effect days-to-halfway elevation parameter, γ_{fixed} = fixed-effect triquarter-halfway delta parameter, β_{random} = random-effect days-to-halfway elevation parameter, and γ_{random} = random-effect triquarter-halfway delta parameter.

2.2.3.4 Qualitative Description

Although no manipulated measurement number resulted in precise estimation of all

the day-unit parameters, the largest improvements in precision (and bias) resulted from

using moderate measurements numbers. With respect to bias under equal spacing, the

largest improvements in bias across all manipulated nature-of-change values resulted from

using the following measurement numbers for the following day-unit parameters (note

that only the random-effect triquarter halfway delta parameter [γ_{random}] had instances of

high bias):

- random-effect triquarter-halfway delta parameters (γ_{random}): seven measurements.

With respect to precision under equal spacing, the largest improvements precision in the estimation of all day-unit parameters (except the fixed-effect days-to-halfway elevation parameter [β_{fixed}]) were obtained with following measurement numbers:

- fixed-effect triquarter-halfway delta parameter (γ_{fixed}): seven measurements, which resulted in a maximum error bar length of 4.37 days.

- random-effect days-to-halfway elevation parameter (β_{random}): seven measurements, which resulted in a maximum error bar length of 7.74 days.

- random-effect triquarter-halfway delta parameter (γ_{random}): seven measurements, which resulted in a maximum error bar length of 7.02 days.

Therefore, for equal spacing, seven measurements led to the greatest improvements in bias and precision in the estimation of all day-unit parameters across all manipulated nature-of-change values (see the emboldened text in the 'Qualitative Description' column of Table 2.3).

2.2.3.5 Summary of Results With Equal Spacing

In summarizing the results for equal spacing, model performance was highest across all day-unit parameters (with the random-effect days-to-halfway elevation parameter (γ_{random}) being an exception) when measurements were placed closer to periods of change, which occurred with a nature-of-change value of 180 (see highest model performance). Unbiased estimation of all the day-unit parameters across all manipulated nature-of-change values resulted from using nine or more measurements (see bias). Precise estimation of all

62

the day-unit parameters was never obtained with any manipulated measurement number (see precision). Although it may be discouraging that no manipulated measurement number under equal spacing resulted in precise estimation of all the day-unit parameters, the largest improvements in precision (and bias) across all day-unit parameters were obtained with moderate measurement numbers. With equal spacing, the largest improvements in bias and precision in the estimation of all day-unit parameters across all manipulated nature-of-change values were obtained using seven measurements (see Qualitative Description).

2.2.4 Time-Interval Increasing Spacing

For time-interval increasing spacing, Table **??** provides a concise summary of the results for the day-unit parameters (see Figure **??** for the corresponding bias/precision plots). The sections that follow will present the results for each column of Table **??** and provide elaboration when necessary (for a description of Table **??**, see concise summary table).

2.2.4.1 Nature of Change That Leads to Highest Model Performance

For time-interval increasing spacing, Table **??** lists the precision values (i.e., error bar lengths) for each day-unit parameter across each nature-of-change value. The 'Total' column indicates the total error bar length, which is a sum of the the lower ('Lower') and upper ('Upper') whisker lengths. Given that the lower and upper whisker lengths were largely equivalent for each parameter, they were largely redundant and so were not reported for the remainder of the results for time-interval increasing spacing. Although model performance was determined by bias and precision, results for bias were not reported because the differences in bias across the nature-of-change values were negligible.

Note that error bar lengths were obtained by computing the average length across all

manipulated number of measurements. The columns shaded in gray indicate the nature

of change where precision is highest (i.e., shortest error bar lengths). For time-interval

increasing spacing, precision was lowest (i.e., longest error bars) with a nature-of-change

value of 80 for all day-unit parameters (see the 'Highest Model Performance' in Table

??). Importantly, with a nature-of-change value of 80, measurements were taken closer

to periods of change under time-interval increasing spacing than with other nature-of-

change values (see Figure 2.6). Therefore, it appears that placing measurements closer to

periods of change increased model performance with time-interval increasing spacing.

Figure 2.6
*Model Performance Status Across Nature-of-Change Values With Time-Interval Increasing
Spacing*
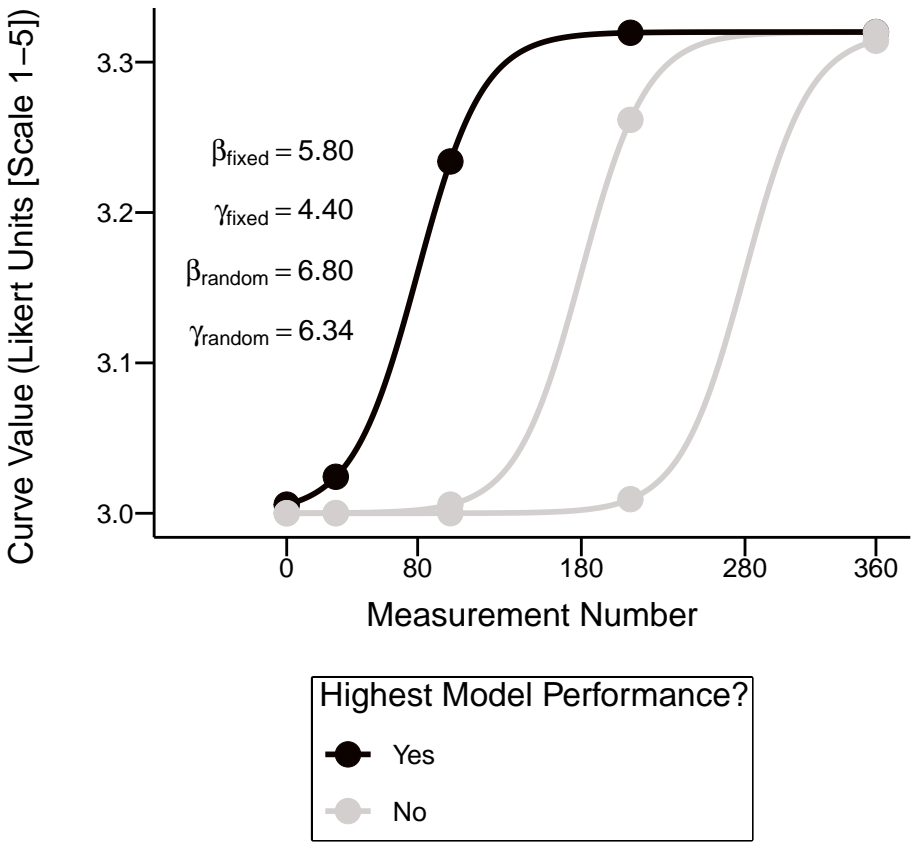
$\beta_{fixed} = 5.80$

$\gamma_{fixed} = 4.40$

$\beta_{random} = 6.80$

$\gamma_{random} = 6.34$

Note. Model performance was highest when measurements were taken closer to periods of greater change,

which resulted with a nature-of-change value of 80 with equal spacing. Text prints error bar lengths that resulted when model performance was highest (see Table ??).

2.2.4.2 Bias

With respect to bias for time-interval increasing spacing, estimates were biased (i.e., above the acceptable 10% cutoff) for each day-unit parameter in the following cells:

- fixed-effect days-to-halfway elevation parameter (β_{fixed}; Figure ??A): no cells.

- fixed-effect triquarter-halfway delta parameter (γ_{fixed}; Figure ??B): no cells

- random-effect days-to-halfway elevation parameter (β_{random}; Figure ??C): five measurements with a nature-of-change value of 280.

- random-effect days-to-halfway elevation parameter (β_{random}; Figure ??C): five measurements with all nature-of-change values and seven measurements with nature-of-change values of 180 and 280.

In summary, with time-interval increasing spacing, estimation of all the day-unit parameters across all manipulated nature-of-change values was unbiased using nine or more measurements, which is indicated by the emboldened text in the 'Unbiased' column of Table ??.

2.2.4.3 Precision

With respect to precision for time-interval increasing spacing, estimates were imprecise (i.e., error bar length with at least one whisker length exceeding 10% of a parameter's population value) in the following cells for each day-unit parameter:

65