# On Estimating the Causal Effects in Interrupted Time Series Design with Phase-in Period and Gradual Implementation: A Simulation Study

by

Tianyi Zheng

B.Sc., The University of British Columbia, 2017

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Measurement, Evaluation, and Research Methodology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2022

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

On Estimating the Causal Effects in Interrupted Time Series Design with Phase-in Period and Gradual Implementation: A Simulation Study

submitted by Tianyi Zheng in partial fulfillment of the requirements for

the degree of Master of Arts

in Measurement, Evaluation and Research Methodology

**Examining Committee:**

Dr. Amery Wu, Associate Professor, Measurement, Evaluation and Research Methodology, UBC
Supervisor

Dr. Ed Kroc, Assistant Professor, Measurement, Evaluation and Research Methodology, UBC
Supervisory Committee Member

# Abstract

Interrupted time series design has been widely applied to assess the causal effectiveness of an intervention. This thesis investigates potential estimation and modelling challenges that arise from interrupted time series (ITS) studies. Specially, for studies that may involve delayed effects and phase-in periods. A phase-in period is a special form of delayed intervention effect where the full effect occurs sometime after intervention. Through a simulation study, this thesis proposes various applicable analytical strategies for phase-in periods and highlights the different casual effects that they referred to. This thesis concludes that multiple counterfactual assumptions may exist, and different analytical strategies lead to different causal effects. Researchers should be cautious about forming their counterfactual assumptions and picking the appropriate analytical strategy accordingly.

## Lay Summary

Interrupted time series is a technique that often applied to assess the effectiveness of interventions, such as political policies. This technique is relatively straightforward and powerful. However, it relies on implicit assumptions. Phase-in period refers to a typical violation of the assumption that intervention may involves some time in order to be implemented or to take effect, although the classic interrupted time series framework assumes the intervention would be implemented instantly. This thesis aims to investigate any inferential issues that may arise with such delayed intervention effects. With these findings, researchers could enhance their understanding in how interrupted time series framework could be applied more robustly.

## Preface

This thesis is an original work by Tianyi Zheng. The research idea and study design were formulated by me and my research supervisor Dr. Amery Wu. I was solely responsible for coding, collecting data, reviewing literature and formulating experiments. Dr. Amery Wu and Dr. Ed Kroc have provided valuable inputs in revising and enhancing the research design and manuscript edition.

# Table of Contents

## List of Tables

## List of Figures

# Chapter 1: Introduction

This thesis investigates the features and modelling challenges that arise from the phase-in period in the interrupted time series (ITS) design. ITS has been a popular quasi-experimental method for determining the degree to which an intervention shifts the underlying trajectory of an outcome (Morgan & Winship, 2015). With the expectation of an intervention effect, the observations are hypothesized to have a different slope or level from the outcome trajectory than those before intervention (Shadish, et al., 2002). The term level and slope changes originates from the general practice of fitting a linear line to the outcome trajectory, and level change often represents a discontinued shift at the time of intervention, while slope change generally denotes a modification of long-term trend because of the intervention. Figure 1.1 includes a simple illustration on level and slope changes.

A phase-in period is a special form of delayed intervention effect where the full effect occurs sometime after intervention, instead of right after intervention. Despite the fact that this terminology is being used by a growing number of academics, to the best of my knowledge, there is still no clear definition, characteristics, and best practices for phase-in periods in ITS studies.

The purpose of this thesis is to highlight the fact that multiple counterfactual assumptions may exist, formally define the phase-in period, and conclude on how different analytical strategies could result under different circumstances.

## 1.1 Background

A quasi-experiment is an empirical and interventional research that uses non-random assignment to quantify the causal impact and test causal hypotheses of an intervention on the target population (Shadish, et al., 2002). Due to the lack of experiment control, the primary issue in this technique is the impossibility to observe the same individual or unit of observation in both intervention and control stages (Morgan & Winship, 2015). A time series is a set of observations made on the same variable over a period of time that is ideally evenly spaced. Time series quasi-experiments were initially proposed by Campbell and Stanley (1963). With this framework, it is feasible to monitor the same individual or unit at multiple periods in time (Brockwell & Davis, 2009), and the key is to know the precise time point at which an intervention occurred (Shadish, et al., 2002). The causal effect of an intervention is then expressed as the estimated difference between the pre-intervention and post-intervention time series; or, from another perspective, it is the difference between the counterfactual (unobserved potential outcomes) and the true post-intervention outcomes, with an arguably courageous assumption of potential outcomes being stable across time. If the intervention had an effect, the post-intervention observations may have a different level and/or slope than the pre-intervention data.

ITS is a reasonable alternative to randomized controlled trials (RCT) under certain circumstances, and RCT have been regarded as the gold standard design for evaluating the effectiveness of an intervention (Bernal et al., 2017). Despite its obvious advantages, a randomized controlled trial may not always be an option due to cost, feasibility, ethical considerations or the necessity to evaluate an intervention retrospectively (Bernal, et al., 2017).

## 1.2 Terminologies

In this sub-section, important terminologies in this thesis are introduced, including the idea of causal action, causal effect, counterfactual, delayed effect and phase-in period.

### 1.2.1 Causal Action and Causal Effect

In my opinion, a causal action is an action that directly created observable impact in some outcome measures, and this impact is the corresponding causal effect. Also, the magnitude of this impact is the causal effect size. For instance, if an investor holds some stocks and have decide to sell them, the action of selling would be the causal action, and the resulting profit or loss would be the causal effect. In addition, how much is the profit or loss relates to the causal effect size.

### 1.2.2 Counterfactual

Counterfactual refers to "what-if" type of potential outcomes that exist in theory but cannot be directly observed. With the advancement of philosophy on counterfactuals and causations, the counterfactual models are widely applied to formulate precise causal hypothesis with specific counterfactual contrasts (Collins, et al., 2004). A generic hypothesis from such framework could be formulated as: if individuals with independent variable $X = x_1$ instead of $X = x_2$, how much would their dependent variable $Y$ have changed.

In ITS, counterfactual refers to the hypothetical scenario under which the prior trend would be assumed to continue without the intervention (Bernal et al., 2017). The causal impact is measured based on differences from the observed post-intervention and the unobserved counterfactual scenarios. Interestingly, in academia, researchers often think prospectively.

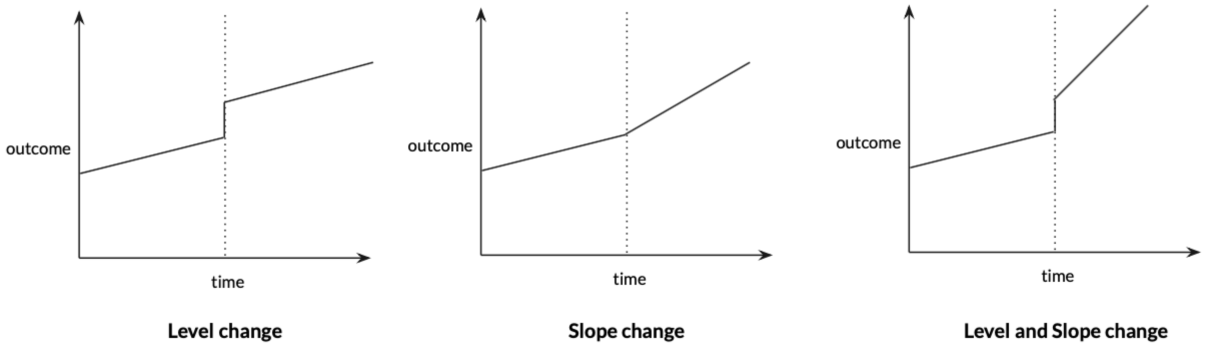
However, in real life, human being often thinks in both directions unconsciously. Back to the previous stock investor example, if this unlucky investor held a stock that was \$3 on the first trading day, \$2 on the second day, and \$1 on the last, and he/she decided to sell this stock on the second day. on the last day, he could optimistically say that “I am glad that I had sold on the second day”, or pessimistically say that “I wish I had sold on the first day”. In this case, the causal action is selling the stock on the second day, the effect is making a profit, and the effect size is \$2; With the optimistic view, the counterfactual (potential outcome) would be selling the stock on the last day, which would result in having \$1. In contrast, the pessimistically perspective compared to the counterfactual of selling on the first day, which could lead to having \$3. This discrepancy indicates that different counterfactual assumptions may result in different causal effects and causal effect sizes.

### 1.2.3 Level Change and Slope Change Effect

The post-intervention time series may differ from pre-intervention time series in several ways, and the most common two are level/intercept change and slope/trend change. In general, level changes correspond to the abrupt (often discontinued) shift in the observed outcome shortly after the intervention; slope changes represent how the rate of change differs when compared a relatively longer period right before the intervention, with a similar length period right after the intervention. Thus, a level change focuses on short-term lift or drop of the outcome from any interventions, while a slope change emphasizes more on long-term trend shifts. I believe that the terminologies of level and slope originated in the fact that researchers most often adopt segmented regression analysis, which will be introduced in-depth in the later chapter. Since the intercept (level) and slope coefficients are the most important components to describe a linear line, it makes sense to compare how much their intercept and slope coefficients differ in a segmented linear regression model. Level and slope changes are causal effects in ITS, and how much has changed refers to the causal effect sizes.

**Figure 1.1**

*A Simple Illustration on Level and Slope Changes*



As an illustration, consider the following hypothetical example series for pre-intervention: 1, 2, 3, 4, 5, 6, 7, 8 and post-intervention: 7, 8, 9, 10, 11, 12, 13, 14. In these two series, we would expect the next value of the series to be 9 without intervention, however, we observed a value of 7 instead, which resulted in a -2 point (i.e. =7-9) of level/intercept change. Also, as the rate of change stayed constant before and after intervention, there was no slope change. As another example where the pre-intervention series as: 1, 2, 3, 4, 5, 6, 7, 8 and post-intervention series as: 9, 11, 13, 15, 17, 19, 21, 23. In the above series, the post-intervention series increased by 2 points per unit, as supposed to only 1 in the pre-intervention series, which led to a double-sized slope/trend. In addition, the counterfactual assumed continuation of pre-intervention trend, and would expect the first element of the post-intervention series to be 9, same as observed; thus, there was no level change.

In the remaining chapters of this thesis, we will introduce two types of counterfactual assumptions as counterfactual from before (intervention) and counterfactual from after (intervention). The counterfactual from before assumes a continuation of the potential outcome prospectively; in our previous stock example, this would refer to assume the stock price would stay at $2 on the third trading day. In comparison, the counterfactual from after assumes the stock price would be $2 on the first trading day. The distinction of these two counterfactuals will be clear in later chapters through the simulation results.

Interestingly, the above series could also be analyzed in a more counter-intuitive way, like the previous stock investor example. In comparison with assuming the counterfactual from pre-intervention series, and calculating the causal effect from difference between the counterfactual outcome and observed outcome, it is also possible to assume the counterfactual from post-intervention series, and ask what would happen if the intervention happened a unit earlier. The counterfactual outcomes in the last point of pre-intervention series would be 6 for the first example, and 7 for the second example. It is compelling to notice the difference in causal effect sizes from these two counterfactual assumptions. Nevertheless, only the counterfactual from before (e.g. assume the stock price would stay the same on the next trading day) were applied in academic settings, because this approach was more intuitive. Moreover, since level change casual effect only compares the counterfactual to adjacent outcome observations before or after intervention, the time intervals were short for any long-term trend shifts (slope changes) to create an impact. Therefore, the counterfactual from before and counterfactual from after approaches are not likely to create a significant difference for casual effect estimation in abrupt intervention cases. However, things can be different when the intervention is not implemented abruptly, and this phenomenon will be further discussed in the later chapters.

### 1.2.4 Delayed Effect and Phase-in Period

Effects in levels and slopes may be immediate or delayed, continuous or abrupt (discontinuous). Delayed effect refers to cases where the intervention effect does not occur instantly. Such effect may involve an implementation process with an unclear end date. Ideally, ITS is intended for immediate and abrupt interventions, and any distortion would cause problems. Specifically, if the effect spanned over a longer time period rather than instantly, the more threat on internal validity (i.e. does the causal effect estimation reflect the actual impact from intervention) it might impose; therefore, it would be challenging to separate the true unknown intervention effect from the chaotic observed change in the outcome, which might be a composition of true intervention effect, co-intervention (confounding) effect, pre-existing trend, random noise and so on.

However, in some situations, a delayed effect may also have a pre-existing or definable end date, and such situations are often named as phase-in period (also referred as roll-in or roll-out period). Specially, phase-in period is a special case of delayed effect when there is a clearly defined start and end date within which the causal effect takes place gradually. In other words, phase-in period represents when we know that the full intervention effect does not occur immediately after the intervention (e.g. due to implementation process, due to shortage of resources, unable to reach all the target population at once), but we also know the specific time where we would expect the full intervention effect to take place, and the time period between the start of intervention and the known time for the full intervention effect is phase-in period. For example, if the research question is testing how Covid-19 vaccination has reduced hospital admission rate, it is possible to set the phase-in period to be between the beginning of first vaccination phase and the completion of the last vaccination phase. The end date could also be defined as when 80% of the population are vaccinated, if 80% is considered as a threshold of community immunity. Alternatively, the phase-in period can be specified by analysts based on the trend in observed time series. However, different analysts' subjective judgement about the phase-in period could lead to different conclusions about the casual effect. As briefly in Chapter 1.3, ITS studies often require strong assumptions to ensure internal validity.

## 1.3 Threats to Internal Validity

Internal validity is the credibility of inference that the observed change implies a causal link from the manipulated variables; to support this assertion, researchers must show that there is no plausible alternative explanation (Shadish, et al., 2002). However, in quasi-experiments, such a claim is not always straightforward, because the observed change may have happened without the intervention. Table 1.1 lists some of the most prevalent threats to internal validity. History (also known as co-intervention) is the most significant threat in ITS. This may be a concern if another intervention occurred at the same time, resulting in indistinguishable effects. Instrumentation and selection are also significant threats, with the former occurring if administrative process for record keeping or data definition changes; the latter implying that the composition of experimental group may change abruptly at time of intervention if the intervention causes or requires attrition from the measurement framework (Shadish, et al., 2002).

Internal validity for ITS may be improved by using a variety of techniques. The most popular one is multiple baseline design, which has played a pivotal role in the development of interventions in clinical psychology (Hayes, et al., 1999), education (Kratochwill, 2013), health promotion (Windsor, 1986) and reinforcement (Sidman, 1960). This technique is often implemented to include a no-intervention control group time series. As such, history is less of a concern because it is less likely that a co-intervention would occur and have a discretionary impact to make the two effects indistinguishable (Shadish, et al., 2002). Other methods for improving internal validity include introducing a second dependent variable that isn't anticipated to be impacted by the intervention, removing the intervention at a specific time, or adding multiple replications using an ABAB or ABBA design (Shadish, et al., 2002), where A and B typically denote treatment effect and placebo effect, respectively.

**Table 1.1**

*Common Threats to Internal Validity (Shadish, et al., 2002)*

| | |
|---|---|
| Ambiguous Temporal Precedence | Lack of clarity about which variable occurred first may yield confusion about which variable is the cause and which is the effect. |
| Selection | Systematic differences over conditions in respondent characteristics that could also cause the observed effect. |
| History | Events occurring concurrently with treatment could cause the observed effect. |
| Maturation | Naturally occurring changes over time could be confused with a treatment effect. |
| Regression | When units are selected for their extreme scores, they will often have fewer extreme scores on other variables, an occurrence that can be confused with a treatment effect. |
| Attrition | Loss of respondents to treatment or to measurement can produce artifactual effects when systematically correlated with conditions. |
| Testing | Exposure to a test can affect scores on subsequent exposures to that test, an occurrence that can be confused with a treatment effect. |

| | |
|---|---|
| Instrumentation | The nature of a measure may change over time or conditions in a way that could be confused with a treatment effect. |
| Additive and Interactive Effects | The impact of a threat can be added to that of another threat or may depend on the level of another threat. |

## 1.4 Purpose of the Thesis

This thesis advocates the investigation of characteristics and different analytical strategies for interrupted time series (ITS) design with a delayed effect from a phase-in period. To the best of my knowledge, there is currently no consensus regarding what a phase-in period is. As mentioned, the issues of validity are complicated and there is no universal guidance for evaluating validity. Thus, instead of dwelling on the issues of validity, this thesis assumes that all the validity requirements have been met and focus on providing some practical guidance for defining phase-in causal effect and the analytical strategies to evaluate it. To do so, I conducted a simulation study with the following foci:

1. Describing the characteristics of phase-in periods with gradual implementation.
2. Introducing various strategies of analyzing causal effect when there is a phase-in period.
3. Exploring the potentially misleading conclusions from inappropriate analytical strategies.

# Chapter 2: Literature Review

## 2.1 Applications

ITS has been widely used in various areas of research, including: political science (e.g. attorney advertising: Johnson, et al., 1993, community interventions: Biglan, et al., 2000, gun control: Carrington & Moyer, 1994; O'Carroll et al., 1991, human rights: Stanley, 1987, political participation: Seamon & Feiock, 1995), health science (e.g. Covid-19 related: Leske, et al., 2021; Pirkis, et al., 2021; Scortichini, et al., 2021; Hamadani, et al., 2020; Vokó, et al., 2020, epidemiology: Tesoriero, et al.,1995; medication use: Hawton, et al., 2013; Walley, et al., 2013; Derde, et al., 2014, Catalano & Serxner, 1987, surgery: Everitt, et al., 1990), economics (e.g. consumer behavior: Einav, et al., 2013, environmental risk analysis: Teague, et al., 1995, real estate values: Brunette, 1995; Murdoch, Singh, & Thayer; 1993, tax policy: Bloom & Ladd, 1982,), education (e.g. educational evaluation: Somers, et al., 2013, education policy: Hallberg, et al., 2018; Bloom, 2003), and psychology (e.g. emotions: Fan, et al., 2019, spouse abuse: Tilden & Shepherd, 1987, substance abuse: Velicer, 1994).

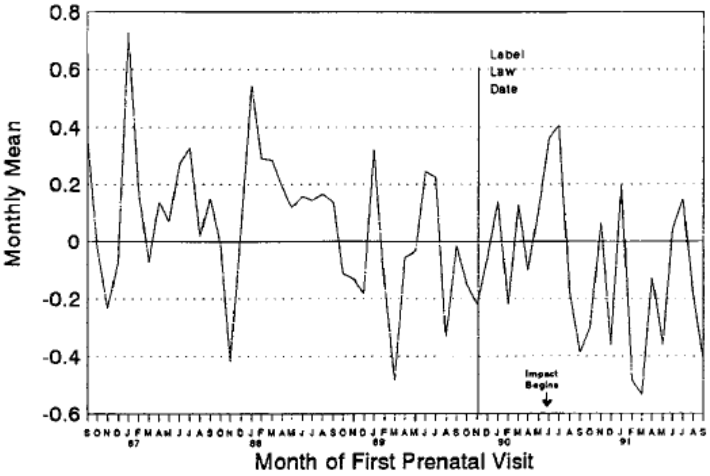
## 2.2 Studies with Clear and Unclear Intervention End Dates

In this sub-section, two example studies are included to highlight the difference between a clear intervention end date, with an ambiguous one.

In the first study, Hankin et al. (1993) examined the impact of an alcohol warning label law on reducing antenatal drinking. Starting November 18, 1989, all containers sold or distributed had to include a warning label saying that drinking alcohol during pregnancy might cause birth defects. Hankin et al. (1993) used a sample of 12026 African-American women visited at a prenatal clinic in Detroit between September 1986 and September 1991 to track monthly mean antenatal drinking scores, which represented how much alcohol was drunk in the two weeks leading up to their first prenatal appointment. The findings indicated that the label legislation had a small reduction of 0.28 on monthly mean drinking score, but there was a seven-month lag in the impact of the label, as shown in Figure 2.1. To back up their assumption about the transition phase, the authors collected information on label awareness, noticing a consistent trend until March 1990, four months after the label legislation went into effect. This delayed effect could be due to several factors. For example, the newly labelled containers would only appear on store shelves after the previous stock had been sold. For another example, it could take some time for people to become aware of the label and take actions accordingly. Although such a delayed impact is typical in ITS studies, there is still a lack of knowledge about its implication on inferences, as well as the best methods for dealing with it.

**Figure 2.1**

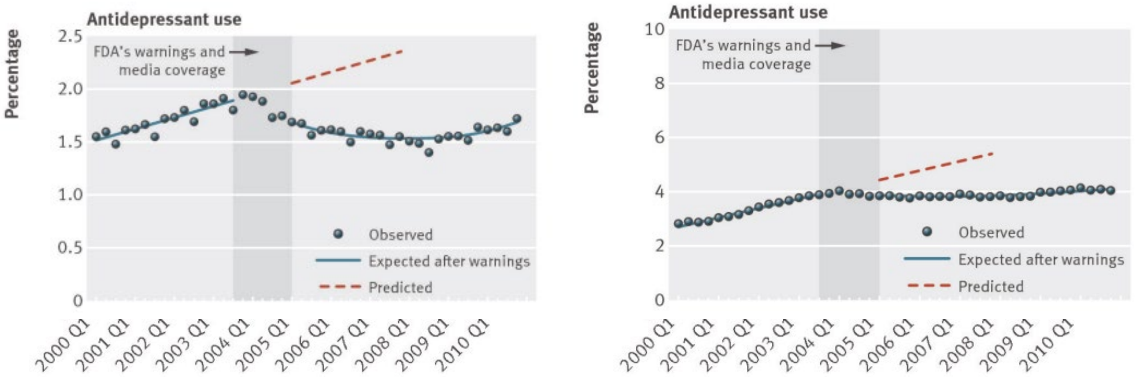
*Monthly Mean of Antenatal Drinking Scores (Hankin, et al., 1993)*

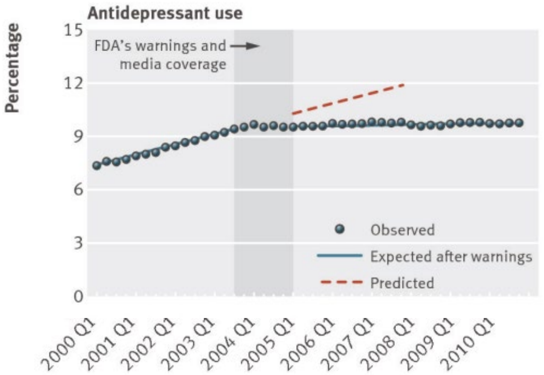


As the intervention end date from Hankin's study depended on their subjective observation from the outcome and interview information, in the following study, the end date was pre-defined from the implementation process. Specifically, Lu et al. (2014) investigated a possible association between changes in young people's antidepressant use with Food and Drug Administration (FDA) warnings and media coverages. The authors applied a quasi-experimental ITS design, and examined on abrupt rate changes of antidepressant dispensing, psychotropic drug poisonings and completed suicides. The time series of population rates were divided into three segments: pre-warning period (2000 Q1- 2003 Q3), phase-in period (2003 Q4 – 2004 Q4), and post-warning period (2005 Q1 – 2010 Q4). As a reference, FDA issued several health advisory warnings against the increased risk of suicidality from adolescent taking antidepressants. The phase-in period spanned the entire period of advisories, warning labels, and media coverages. Moreover, authors argued that excluding the phase-in period would accommodate the anticipatory response to the warnings, and thus resulted in a more reliable estimate on "full strength" effect from the policy (Lu, et al., 2014). As a result, the authors concluded on a 31% decrease of antidepressant use among adolescents, a 24.3% decline for young adults, and a 14.5% drop among adults, as shown in Figure 2.2.

**Figure 2.2**

*Rates of Antidepressant Use for Adolescents, Young Adults, and Adults*





*Note.* Top left graph is for children aged 10-17, top right is for young adults aged 18-29 and bottom is for adults aged 30-64 (Lu, et al., 2014). **The darker gray period is the phase-in time.**

In conclusion, these two studies both discussed issues of delayed effects due to gradual implementation with relatively short but significant spanned periods. However, authors from the studies handled these circumstances differently. Hankin et al. (1993) used Box-Jenkins intervention model (gradual-start-permanent-duration model), where they hypothesized the label effect started gradually and had a permanent effect, also included the potential lagged period in their analysis. Their strategy involved fitting a model that consisted of two parts: an autoregressive integrated moving average (ARIMA)  model, and an intervention component. The intervention component measured impact from the label law, and the ARIMA model provided a smooth fit to the observations by adjusting from seasonality and other noises. However, the appropriateness of this model might be questionable, because impact from label law might be diminishing over time instead of being permanent. Moreover, as this model measured a gradual and long-term effect, the observed intervention effect might be mixed up with any overall trend (e.g. if people tends to consume more and more alcohol over time). In comparison, Lu et al. (2014) applied a classic segmented regression model, measured the abrupt change in antidepressant use, with all the datapoints from phase-in period excluded in their model analysis. While the latter approach is more popular in ITS studies (e.g. Bou-Antoun, et al., 2018; Sruamsiri, et al., 2016; Leopold, et al., 2014; Bernal, et al., 2012; Garabedian, et al., 2012; Serumaga, et al., 2011), there is still a lack of explorations on examining its effectiveness. To solve this puzzle, a simulation study will be conducted in the later chapter to evaluate the appropriateness of this approach under different circumstances.

### 2.3 Simulation Study

Simulation studies are computer experiments that adapt pseudo-random sampling to generate data (Morris, et al., 2019). Because parameters are prespecified and known by the user, simulation studies provide a chance to evaluate the performance of statistical methods (Morris, et al., 2019; Burton, et al., 2006). For example, Hawley et al. (2019) investigated the statistical power to detect an intervention effect by varying the number of time points, average sample size per time point, average relative reduction post-intervention, intervention location in the time series, and reduction mediated via a slope change or level change. Turner et al. (2020) compared the performance of estimation methods for ITS, such as ordinary least square, Newey-West estimator, generalized least square, restricted maximum likelihood, and autoregressive integrated moving average model with varying level change, slope change, autocorrelation, noise, and number of datapoints. They also evaluated the Durbin-Watson (DW) test's ability in detecting autocorrelation.

# Chapter 3: Methods and Results

## 3.1 Study Design

In this section, I will report a simulation study that aims to evaluate the aptness of six strategies for analyzing the causal effect in a phase-in ITS. Simulation design allows me to draw sample data from a population with known intervention effect, both in terms of level and slope changes, varies in sample size and effect size. The six analytical strategies for analyzing the casual effect are labeled as: (1) as abrupt at start, (2) as abrupt at end, (3) excluding phase-in period - before, (4) excluding phase-in period - after, (5) counterfactual from before, and (6) counterfactual from after. These six analytical strategies are applied to the simulated data, respectively to see whether each of the estimate of intervention effect is biased. The performance of each study is determined by comparing estimates of level and slope changes to the true parameters.

## 3.2 Objective

The objectives of this study are as follows:

1. To simulate the trajectories of ITS with gradual implementation in a phase-in period under various circumstances.
2. To evaluate the estimates of causal effect by the six different analytical strategies at varying levels of effect sizes and sample sizes.
3. To compare the difference in performance among the six analytical strategies with a level change in the population, a slope change, or both.

## 3.3 Parameters and Scenarios

As described and justified in the following sections, this study investigates three population scenarios with varying parameters.

### 3.3.1 ITS Scenarios

In this study, three ITS population scenarios were created under the assumptions that the total number of time points was 200, intervention started at time point 100, and the length of phase-in period was 20:

**Scenario-1**: a time series with no prior trend and a considerable level shift but no slope change because of intervention.

**Scenario-2**: a time series with a prior trend and a considerable level shift but no slope change because of intervention.

**Scenario-3**: a time series with a prior trend, a slope shift because of intervention, and both with or without considerable level shifts.

Table 3.1 visualizes the difference between three population scenarios.

**Table 3.1**

*Type of Changes Comparison across the Scenarios*

| | Pre-existing trend | Level change | Slope change |
|---|---|---|---|
| Scenario-1 | ✗ | ✓ | ✗ |
| Scenario-2 | ✓ | ✓ | ✗ |
| Scenario-3 | ✓ | ✓ | ✓ |
| | ✓ | ✗ | ✓ |

### 3.3.2 Simulation Design

This section provides rationales for choosing the research settings for the simulations.

- Total number of time points = 200. The number of time point (denoted as N) was controlled (do not vary) in this simulation. The choice of 200 time points is reasonable because it represents approximately 50 years if they are quarterly observations, 16 years for monthly, and 4 years for weekly (weekly is often the lowest level of aggregation in ITS studies). Also, since modelling seasonality effects typically requires more than 2 years of data, N = 200 is relatively sufficient to provide any form of analysis (Wagner et al., 2002). In addition, Hawley et al. (2019) provided evidence that the number of time points may have the same effect on statistical power as the number of subjects (sample size, denoted as n) per time point; therefore, it is reasonable and more simplified to only include variation on n while having N as fixed.
- Length of phase-in period. This factor was controlled in the simulation and was set to be 20. In many ITS studies that involve phase-in periods, the periods typically represent 7.5% to 25% of the total time, with a median of 11.3% (Bou-Antoun, et al., 2018; Sruamsiri, et al., 2016; Leopold, et al., 2014; Garabedian, et al., 2012; Serumaga, et al., 2011). Accordingly, the phase-in period length was set to 20 (10%) out of the total 200 time points.
- Number of subjects per time point (n). This was a factor to be examined in the simulation study and it had three levels of 20, 100, 1000. The gradual impact model assumes that intervention effect is implemented at a constant rate and this will be explained in detail later; as a preview, since the length of phase-in period was set to 20, with at least one subject intervened per time point, the minimum possible sample size

for 20 time points would be 20. In addition, group size of 100 and 1000 were also included to compare how would increasing number of subjects at each time point impact the estimations for level and slope changes. Based on the simulation study from Hawley et al. (2019), 200 time points with 100 subjects at each point has led to sufficient statistical power (> 0.8) in detaching a reasonable level or slope change with their assumptions.

- Population variance = 10. The population variance parameter was only set to introduce some randomness in the sample groups (among the subjects at each time point). In fact, since the outcome of interest was the sample group mean, the expected variation would be tiny, as the variance of mean was inversely related to the sample group size. Therefore, a population variance of 10 would be able to create a small but noticeable variation and would not impact the result of estimations in this study. Since this research emphasized more on how each analytical strategy affect estimations, the population variance was set as a controlling variable rather than a parameter of investigation, although it would be interesting to explore how large variation could impact our model estimations.
- Level changes. The level change was a factor to be examined and there were four magnitudes: 0, -20, -50, and -100. In the simulation study from Hawley et al. (2019), the authors specified 15%, 34%, 50% and 75% reductions to represent reasonable levels of changes. Similarly, as we purposely fixed the value of the outcome variable to be 120 before intervention across all scenarios, level changes of -20, -50 and -100 would represent 16% (relatively small), 42% (relatively medium), and 83% (relatively

large) reductions. Level change of 0 was only applied in scenario-3 in order to investigate any estimation impact with only slope change but no level change.

- Slope change. In scenario-3, the post-intervention trend line's slope was doubled from a value of 1 to 2. The idea here was to introduce an additional slope change after intervention and investigate how would this contribute to the estimation on the level change. Moreover, the slope change and level change were at opposite directions (decreased level and increased slope), thus, it would be interesting to examine if this opposition could lead to misleading conclusions.

Because scenario-1 and -2 had the same outcome and the same amount of changes at time of intervention (i.e. in scenario-1 and -2 at time 100, the population outcomes were both at 120, and reduced to 100 at time 101 after receiving the intervention), any difference in estimations would be a result of the only distinction in the two scenarios, which was the existence of pre-existing trend. Similarly, scenario-2 and -3 had the same outcome trajectory patterns (shown in Figure 3.1) before the intervention, and had the same changes because of the intervention, the difference was whether a slope change existed in addition to level changes. Scenarios that represent no level or slope changes were not included in this study. This is because ITS is more often applied to justify and quantify an observed change. With no change observed, researchers typically choose not to proceed with further analysis.

### 3.3.3 Simulation Assumptions

Over the simulation study, all following assumptions were made for the sake of simplicity and feasibility:

1. No substantial threats to internal validity, which meant no other major factors such as confounders or seasonality that would affect the outcome.
2. Individual outcomes at each time point distributed normally with a constant variance of 10.
3. The phase-in period was predefined to start at time 100 and end at time 120, during which the intervention was gradually implemented among the subjects at a constant rate.
4. Variability in the outcome variable during the entire time series was solely explained by the trend lines as specified in Table 3.2.

## 3.4 Data Generating Process

The datasets for this study were created with R v4.1 according to the settings described in the previous section. The reproducible script was posted on GitHub (https://github.com/tianyica/SimulationITS). In order to simulate time series with phase-in periods, abrupt change scenarios were generated first, then the effect was gradually added during the phase-in period to reflect the process of gradual implementation. To provide a stronger evidence, one hundred iterations were generated, which corresponded to one hundred different datasets for each effect size and sample size combination; that was, nine hundred (3 effect sizes * 3 sample sizes * 100) different datasets generated for scenario-1 and -2, and one thousand and two hundred (4 effect sizes * 3 sample sizes * 100) different datasets generated for scenario-3. To obtain a better understanding with visual inspections and detailed estimation results, datasets from one iteration were randomly selected to create the trajectory plots (Figure 3.1-3.7), residual plots (Figure 3.8-3.13) and effect size point estimation tables (Table 3.4, 3.7-3.10). The data generating process for each dataset is described as follows.

### Abrupt Change

Recall that the outcome at time point 100, when the intervention started, was 120 for all three scenarios. The relationships for abrupt changes were shown in the Table 3.2 in terms of the slopes and intercepts for the trend lines. Taking scenario-3 with level change effect size of -50 as an example, the outcome would be 120 (=1*100+20) at time 100, and expected to be 121 (=1*101+20); however, because of the intervention, it was 71 (=2*101-131) instead. Thus, the estimated causal level change would be the difference from the observed and counterfactual, which was -50 (=71-121).

**Table 3.2**

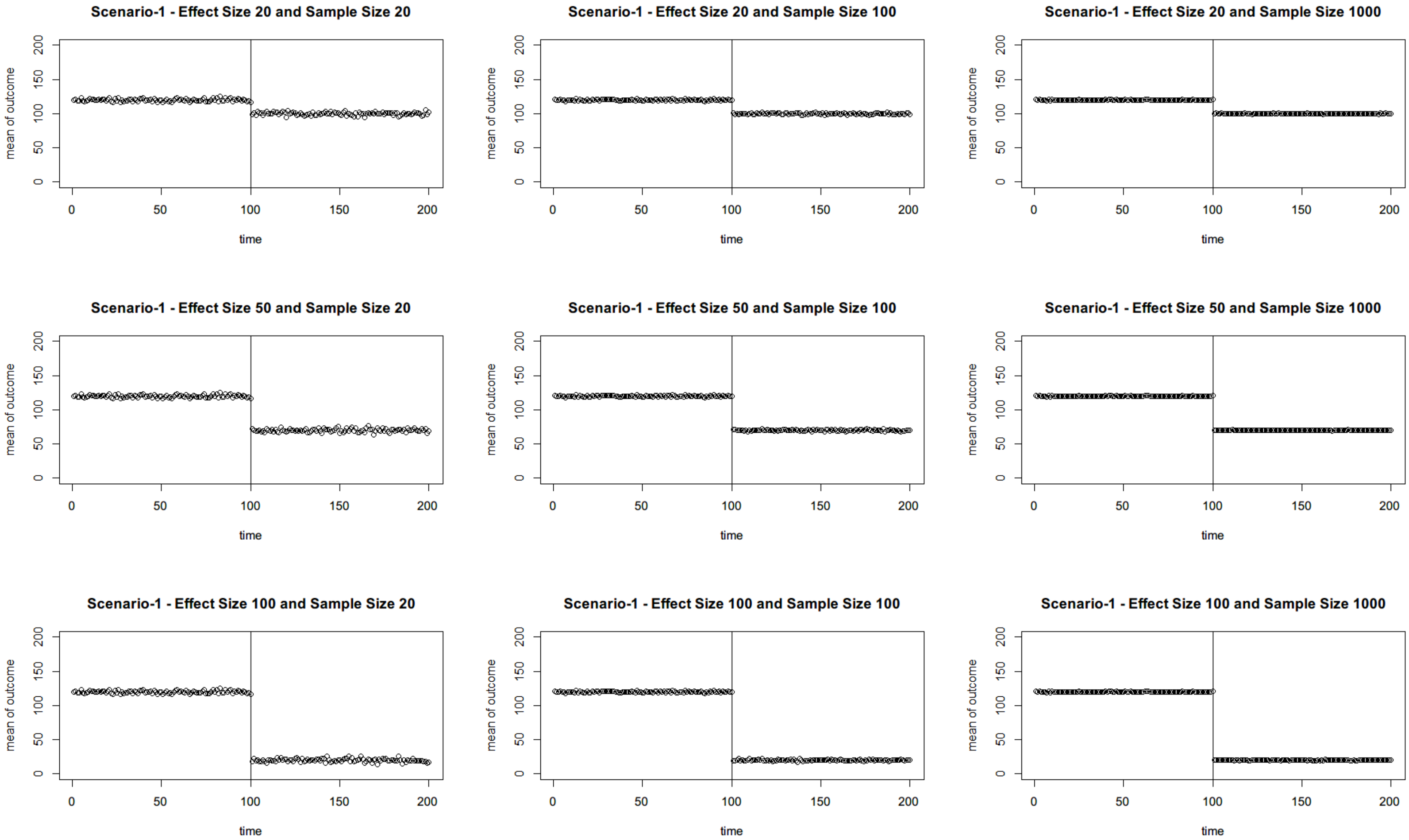
*True Relationships Before and After Intervention in Each Scenario*

| | Level Change Effect Size | Before Intervention (time = 1…100) | After Intervention (time = 101…200) |
|---|---|---|---|
| Scenario-1: Level change only, no pre-exiting trend | -20 | $Y_t = 0 * time_t + 120$ | $Y_t = 0 * time_t + 100$ |
| | -50 | $Y_t = 0 * time_t + 120$ | $Y_t = 0 * time_t + 70$ |
| | -100 | $Y_t = 0 * time_t + 120$ | $Y_t = 0 * time_t + 20$ |
| Scenario-2: Level change with pre-existing trend | -20 | $Y_t = 1 * time_t + 20$ | $Y_t = 1 * time_t + 0$ |
| | -50 | $Y_t = 1 * time_t + 20$ | $Y_t = 1 * time_t - 30$ |
| | -100 | $Y_t = 1 * time_t + 20$ | $Y_t = 1 * time_t - 80$ |
| Scenario-3: Level and slope change with pre-existing trend | 0 | $Y_t = 1 * time_t + 20$ | $Y_t = 2 * time_t - 81$ |
| | -20 | $Y_t = 1 * time_t + 20$ | $Y_t = 2 * time_t - 101$ |
| | -50 | $Y_t = 1 * time_t + 20$ | $Y_t = 2 * time_t - 131$ |
| | -100 | $Y_t = 1 * time_t + 20$ | $Y_t = 2 * time_t - 181$ |

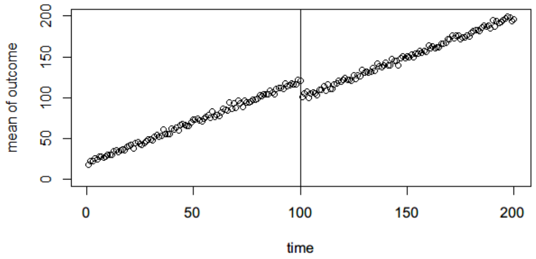
To generate datasets in each scenario, a random sample of the corresponding size was taken for each time point from a normal distribution with a mean specified by the trend line and a variance of 10, as $Y_t \sim N(a * time_t + b, 10)$, where a and b were slope and intercept terms shown in Table 3.2. Again, for this generation, we assumed that no residual autocorrelation existed. Theoretically, this would create random variations around the pre-specified trendlines. For instance, the simulated observations from time point 150 of scenario-3, sample size 100, and level change effect size -20 were created by randomly drawing 100 samples from a normal distribution with a mean of 200 (= 2*150-100) and variance of 10. The mean of each generated time point was recorded as the outcome variable of interest. Figure 3.1 depicted some of the situations, noticing that since sample mean variation was inversely linked to sample size, the observations tended to cluster closer to the trend line as the sample size grew at each time point.

**Figure 3.1**

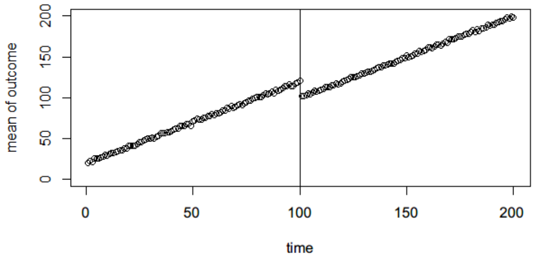
*Mean of Outcome Trajectories with No Phase-In Period (As Abrupt Change) from Scenario-1 to Scenario-3*



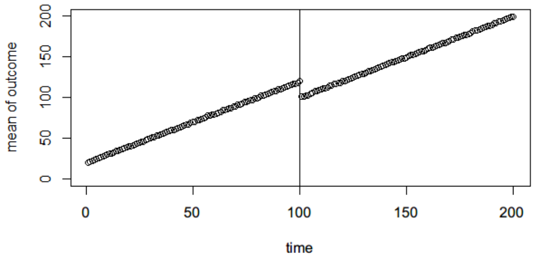
Scenario-2 - Effect Size 20 and Sample Size 20



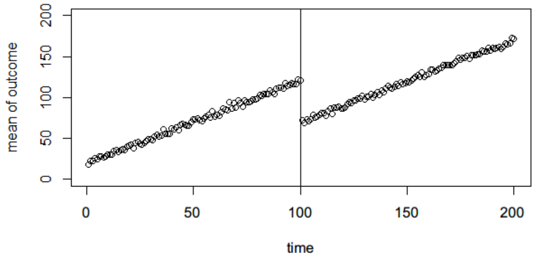
Scenario-2 - Effect Size 20 and Sample Size 100



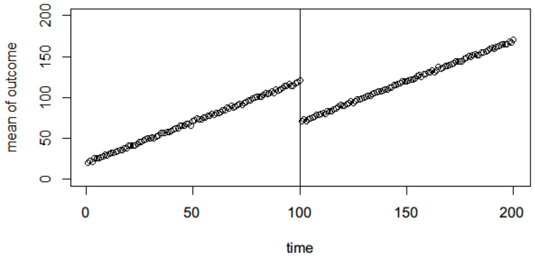
Scenario-2 - Effect Size 20 and Sample Size 1000



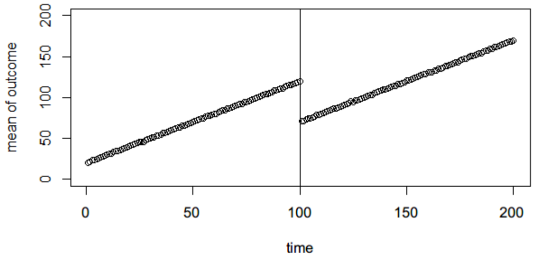
Scenario-2 - Effect Size 50 and Sample Size 20



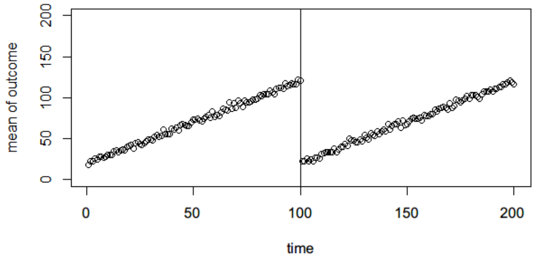
Scenario-2 - Effect Size 50 and Sample Size 100



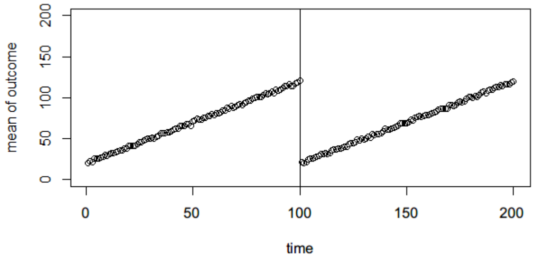
Scenario-2 - Effect Size 50 and Sample Size 1000



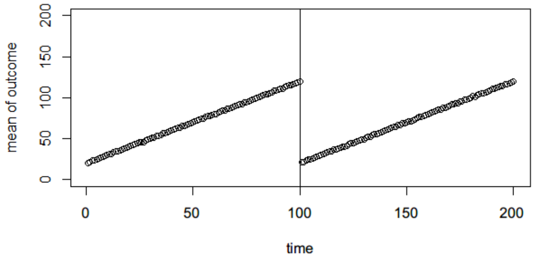
Scenario-2 - Effect Size 100 and Sample Size 20

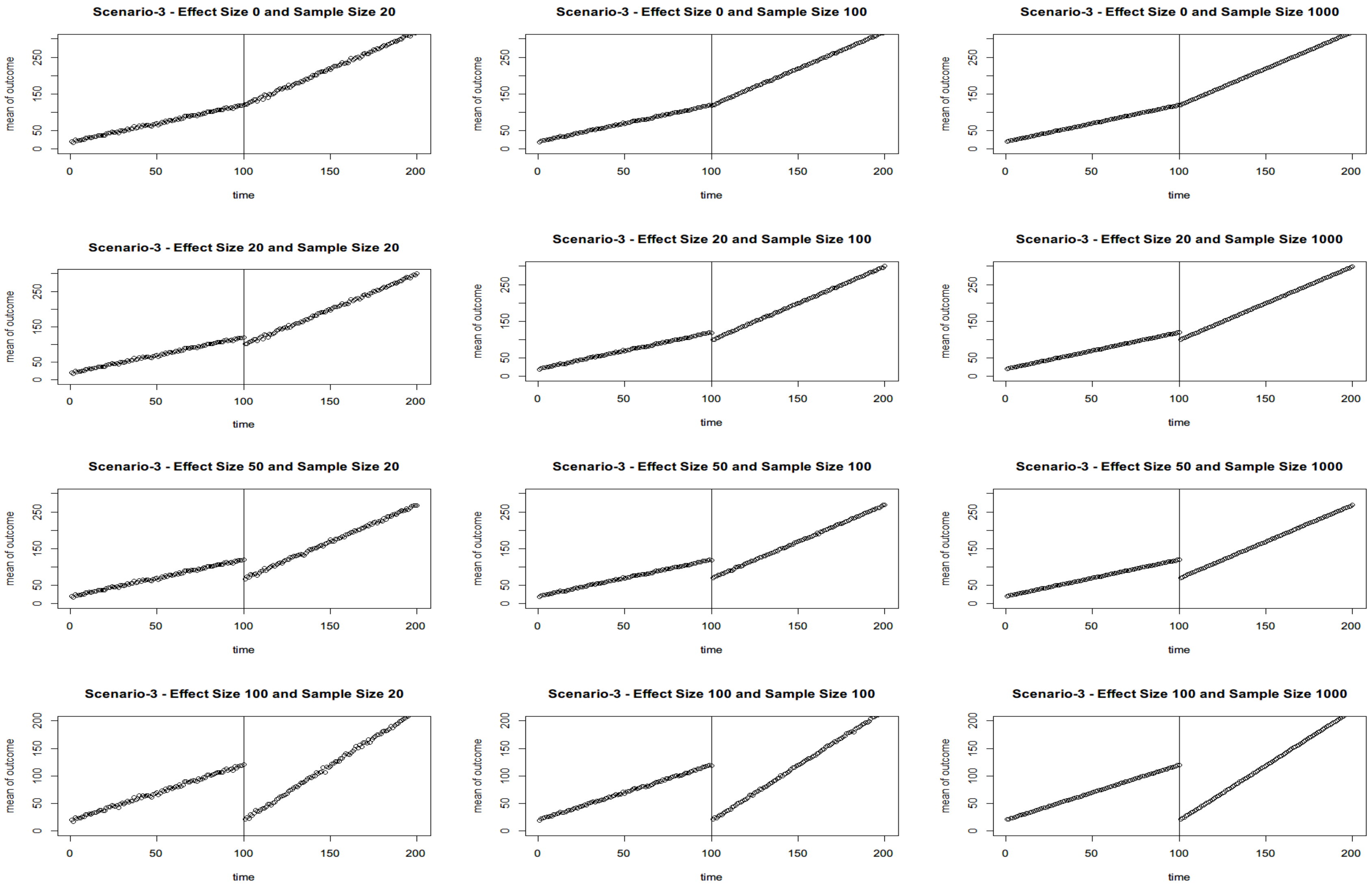


Scenario-2 - Effect Size 100 and Sample Size 100



Scenario-2 - Effect Size 100 and Sample Size 1000





*Note.* The vertical line at time 100 represents that the intervention occurred at time 100.

**Gradual Phase-in Period**

The data for the gradual phase-in period was created after obtaining an abrupt change time series. First, data from time 100 to 120 created by the abrupt change were specifically excluded and replaced to represent a gradually implemented phase-in period, while the observations outside of this time interval (i.e. time 1-99, 121-200) remained unchanged. In simple terms, it meant that we kept the same start and end points for the phase-in period, however, instead of believing the whole sample would switch suddenly because of the intervention effect, only 5% of the sample would receive intervention at each time point. This allowed for a constant gradual increase of intervened sample from 0% to 100%.

Taking sample size of 20 for example, at the start of phase-in period, when time = 100 and all 20 subjects have not received intervention, the expected mean outcome would follow the before intervention trend. At the next point when time = 101, one of the 20 subjects would receive intervention, and the mean outcome at time 101 would be slightly dragged towards the after-intervention trend. By the same token, at time 119, 19 out of the 20 subjects would receive intervention, thus, the mean outcome would be close to the after-intervention trend. At the end of phase-in period (when time = 120), all 20 subjects would have received intervention, and therefore the expected mean outcome would be equivalent to the after-intervention trend. Table 3.3 showed the implementation strategy using a sample size 20 at each time point as an example.

To illustrate the data values at subject level, Table 3.4 displayed a zoom-in example at time 105 for sample size 20 and level change of -50. In this case, five individual outcomes were generated at random from a normal distribution with a mean calculated from the counterfactual trend line of before intervention at time 105 and a variance of 10. Table 3.5 illustrated the summary of aggregated intervened vs. non-intervened group size comparison at every time point within the phase-in period. Means were computed at each time point after the aggregation for phase-in periods and attached to the original time series to construct the complete case.

Figure 3.1 demonstrated all the trajectories in each scenario when the intervention was abrupt. Subsequently, trajectories from Figure 3.2-3.4 indicated that the constant rate implementation of intervention resulted in a gradual shift within the phase-in period. Moreover, the shape was approximately linear without a slope change, and was approximately curvilinear with a slope change.

**Table 3.3**

*Implementation of Gradual Phase-in Period with Sample Size 20 at Each Time*

| Time / Subject | 100 | 101 | 102 | 103 | 104 | 105 | 106 | 107 | 108 | 109 | 110 | 111 | 112 | 113 | 114 | 115 | 116 | 117 | 118 | 119 | 120 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | N | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| 2 | N | N | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| 3 | N | N | N | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| 4 | N | N | N | N | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| 5 | N | N | N | N | N | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| 6 | N | N | N | N | N | N | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| 7 | N | N | N | N | N | N | N | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| 8 | N | N | N | N | N | N | N | N | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| 9 | N | N | N | N | N | N | N | N | N | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| 10 | N | N | N | N | N | N | N | N | N | N | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| 11 | N | N | N | N | N | N | N | N | N | N | N | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| 12 | N | N | N | N | N | N | N | N | N | N | N | N | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| 13 | N | N | N | N | N | N | N | N | N | N | N | N | N | Y | Y | Y | Y | Y | Y | Y | Y |
| 14 | N | N | N | N | N | N | N | N | N | N | N | N | N | N | Y | Y | Y | Y | Y | Y | Y |
| 15 | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | Y | Y | Y | Y | Y | Y |
| 16 | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | Y | Y | Y | Y | Y |
| 17 | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | Y | Y | Y | Y |
| 18 | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | Y | Y | Y |
| 19 | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | Y | Y |
| 20 | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | N | Y |

*Note*. Y denoted having received intervention and N as having not.

**Table 3.4**

*Individual Outcomes for Sample Size 20 and Level Change of -50 at Time 105*

| Individual | Outcome | Intervention | Group Mean | Grand Mean |
|---|---|---|---|---|
| 1 | 60.22 | Y | 74.70 | 109.76 |
| 2 | 73.97 | Y | 74.70 | 109.76 |
| 3 | 76.35 | Y | 74.70 | 109.76 |
| 4 | 83.96 | Y | 74.70 | 109.76 |
| 5 | 78.99 | Y | 74.70 | 109.76 |
| 6 | 116.68 | N | 121.45 | 109.76 |
| 7 | 117.49 | N | 121.45 | 109.76 |
| 8 | 96.79 | N | 121.45 | 109.76 |
| 9 | 126.03 | N | 121.45 | 109.76 |
| 10 | 138.61 | N | 121.45 | 109.76 |
| 11 | 118.78 | N | 121.45 | 109.76 |
| 12 | 103.99 | N | 121.45 | 109.76 |
| 13 | 113.56 | N | 121.45 | 109.76 |
| 14 | 119.33 | N | 121.45 | 109.76 |
| 15 | 128.54 | N | 121.45 | 109.76 |
| 16 | 127.51 | N | 121.45 | 109.76 |
| 17 | 122.39 | N | 121.45 | 109.76 |
| 18 | 132.47 | N | 121.45 | 109.76 |
| 19 | 130.43 | N | 121.45 | 109.76 |
| 20 | 129.08 | N | 121.45 | 109.76 |

*Note*. Individuals assumed to be randomly shuffled; Y as received intervention and N as not received.

**Table 3.5**

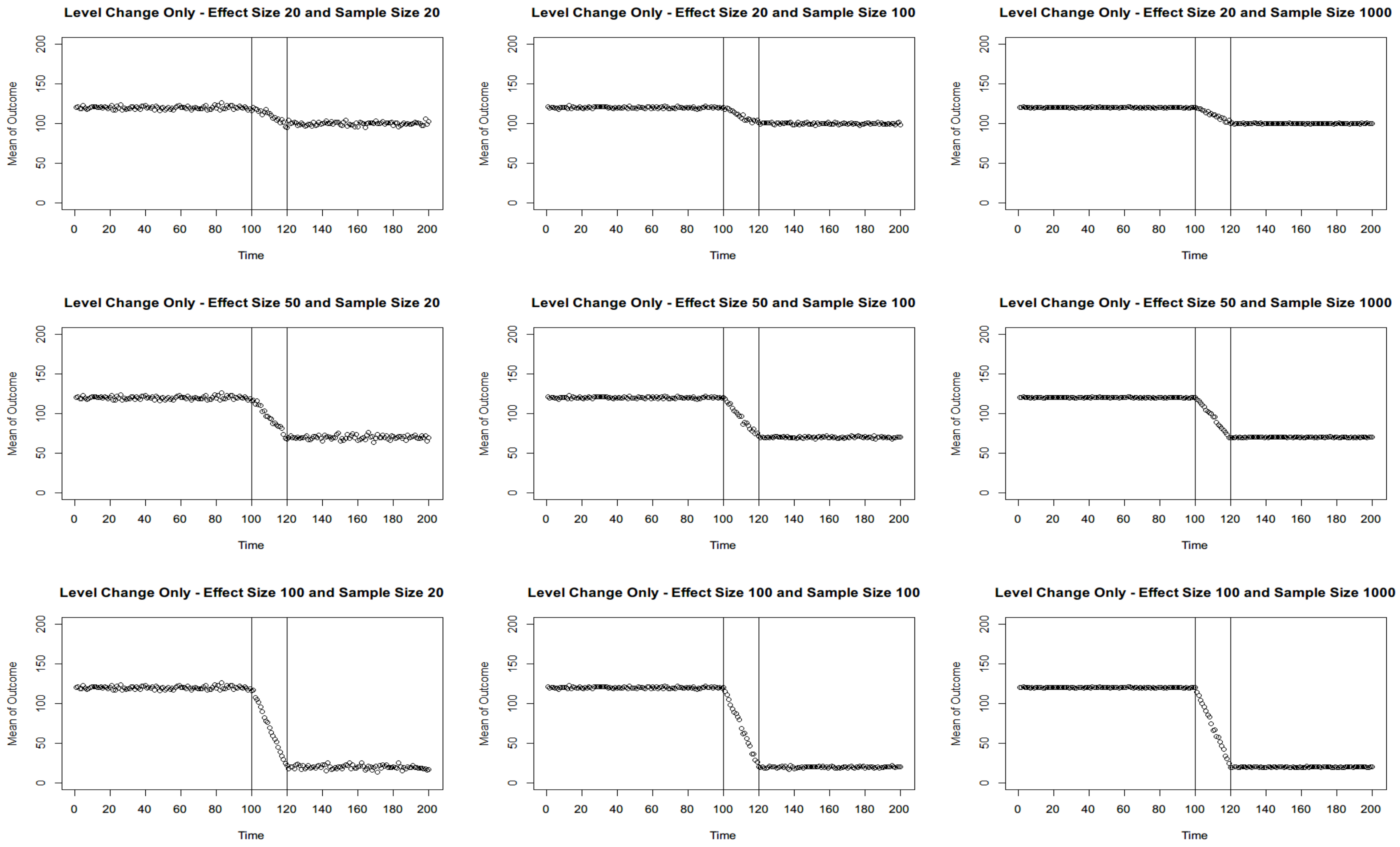
*Intervened vs. Not Intervened Group Sizes at Each Time within Phase-in Period*

| Sample Size→<br>Time↓ | 20 | 100 | 1000 |
|---|---|---|---|
| 100 | 0:20 | 0:100 | 0:1000 |
| 101 | 1:19 | 5:95 | 50:950 |
| 102 | 2:18 | 10:90 | 100:900 |
| 103 | 3:17 | 15:85 | 150:850 |
| 104 | 4:16 | 20:80 | 200:800 |
| 105 | 5:15 | 25:75 | 250:750 |
| 106 | 6:14 | 30:70 | 300:700 |
| 107 | 7:13 | 35:65 | 350:650 |
| 108 | 8:12 | 40:60 | 400:600 |
| 109 | 9:11 | 45:55 | 450:550 |
| 110 | 10:10 | 50:50 | 500:500 |
| 111 | 11:9 | 55:45 | 550:450 |
| 112 | 12:8 | 60:40 | 600:400 |
| 113 | 13:7 | 65:35 | 650:350 |
| 114 | 14:6 | 70:30 | 700:300 |
| 115 | 15:5 | 75:25 | 750:250 |
| 116 | 16:4 | 80:20 | 800:200 |
| 117 | 17:3 | 85:15 | 850:150 |
| 118 | 18:2 | 90:10 | 900:100 |
| 119 | 19:1 | 95:5 | 950:50 |
| 120 | 20:0 | 100:0 | 1000:0 |

*Note*. **Received intervention: not received** intervention.

**Figure 3.2**

*Mean of Outcome Trajectories with Phase-In Period for Scenario-1*



Level Change Only - Effect Size 20 and Sample Size 20

Level Change Only - Effect Size 20 and Sample Size 100

Level Change Only - Effect Size 20 and Sample Size 1000

Level Change Only - Effect Size 50 and Sample Size 20

Level Change Only - Effect Size 50 and Sample Size 100

Level Change Only - Effect Size 50 and Sample Size 1000

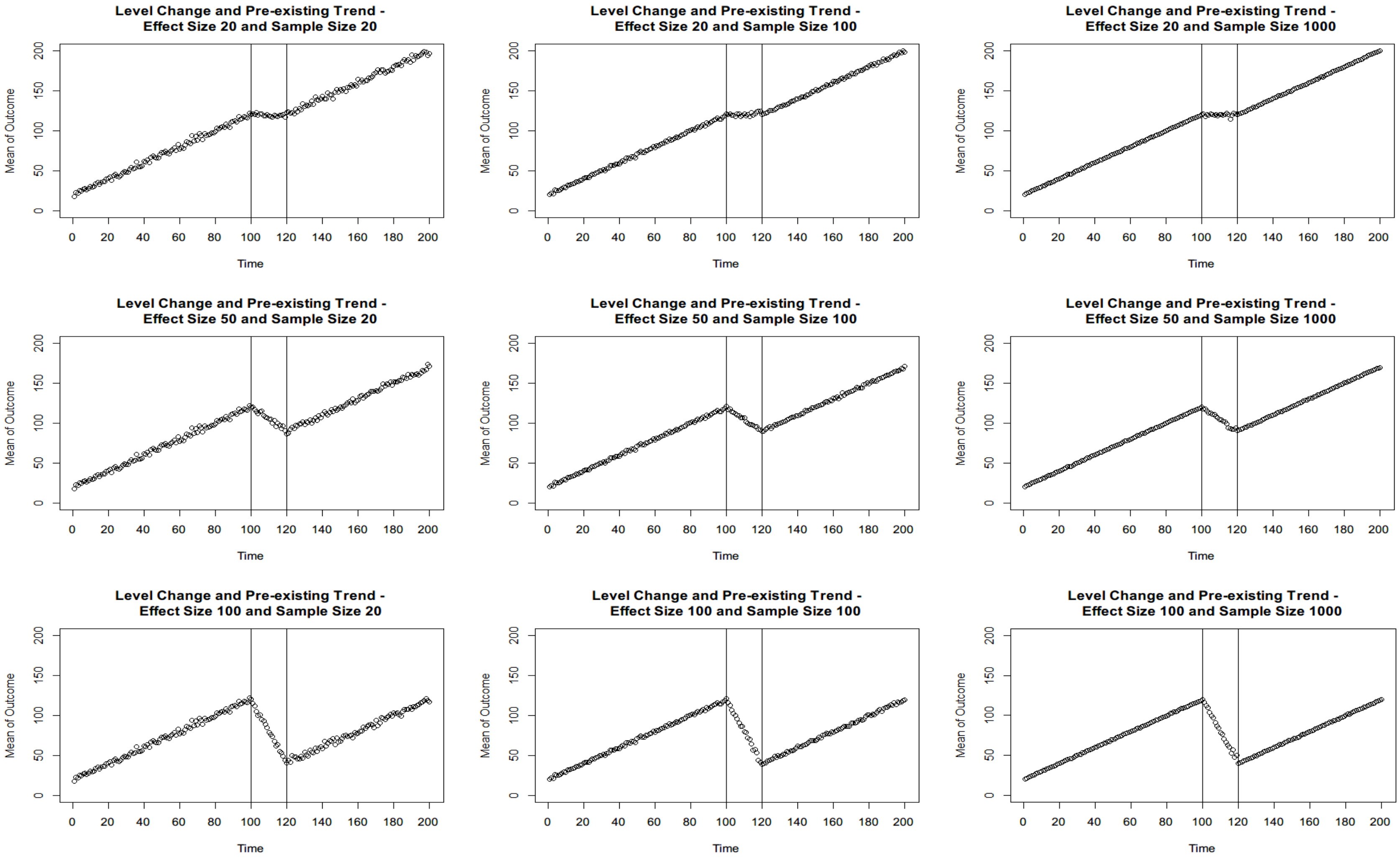
Level Change Only - Effect Size 100 and Sample Size 20

Level Change Only - Effect Size 100 and Sample Size 100

Level Change Only - Effect Size 100 and Sample Size 1000

**Figure 3.3**

*Mean of Outcome Trajectories with Phase-In Period for Scenario-2*



Level Change and Pre-existing Trend - Effect Size 20 and Sample Size 20

Level Change and Pre-existing Trend - Effect Size 20 and Sample Size 100

Level Change and Pre-existing Trend - Effect Size 20 and Sample Size 1000

Level Change and Pre-existing Trend - Effect Size 50 and Sample Size 20

Level Change and Pre-existing Trend - Effect Size 50 and Sample Size 100

Level Change and Pre-existing Trend - Effect Size 50 and Sample Size 1000

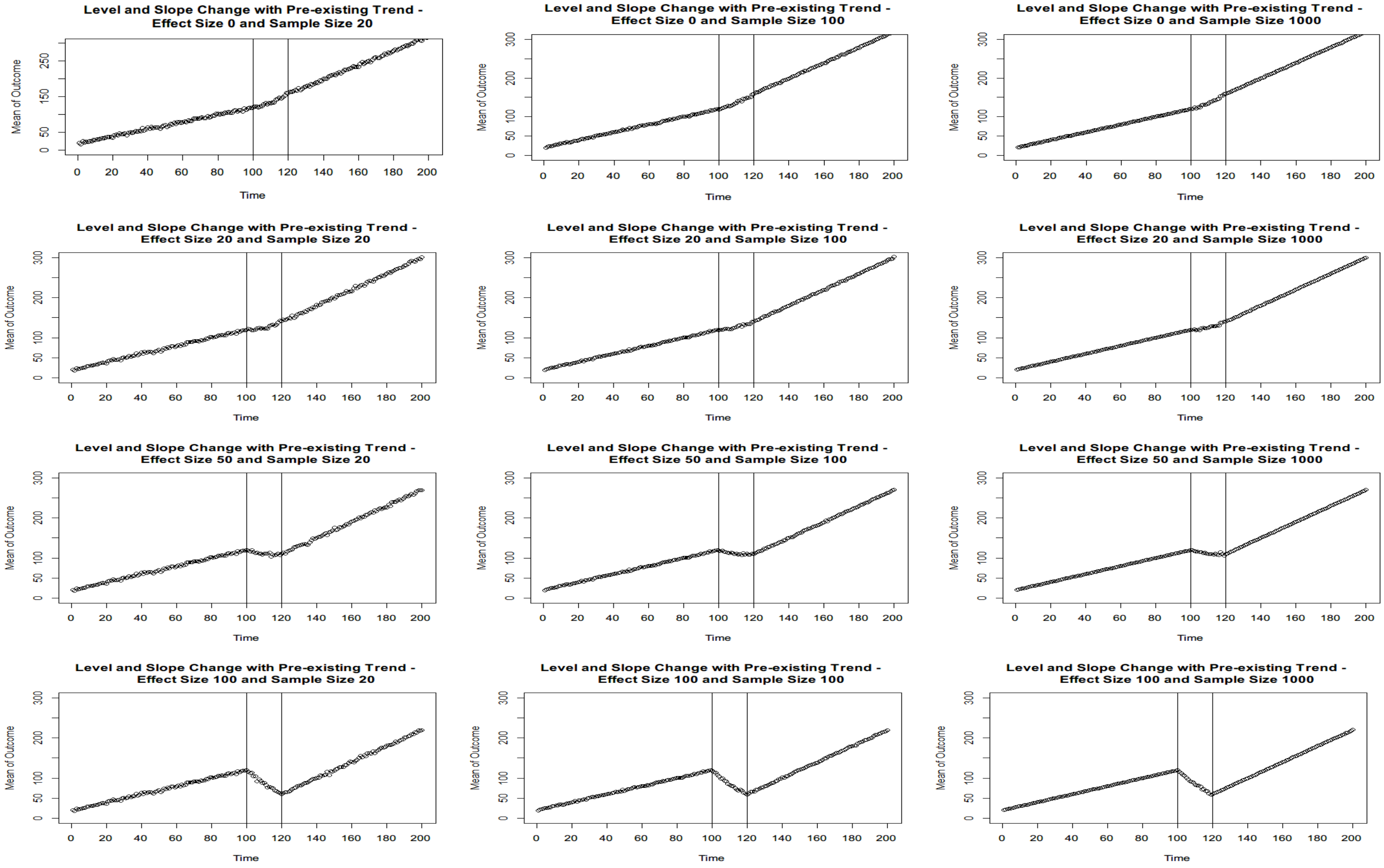
Level Change and Pre-existing Trend - Effect Size 100 and Sample Size 20

Level Change and Pre-existing Trend - Effect Size 100 and Sample Size 100

Level Change and Pre-existing Trend - Effect Size 100 and Sample Size 1000

**Figure 3.4**

*Mean of Outcome Trajectories with Phase-In Period for Scenario-3*



### Segmented Regression

When assessing intervention impact, a generic time series model could be formulated as (Morgan & Winship, 2015):

$$Y_t = f(t) + D_t * b + e_t \text{ ......................................} \quad (1)$$

Here, the outcome $Y_t$ is a certain function of $f(t)$ and an intervention variable $D_t$, where $D_t$ is a dummy variable indicating whether the intervention is in effect at time $t$. Also, a time varying noise $e_t$ is included. The function $f(t)$ describes how the outcome $Y_t$ behaves along the time (i.e. any pre-existing trend) while controlling for the intervention. The $b$ coefficient is a constant that measures the magnitude of intervention impact.

Alternatively, modern applied research uses the segmented regression approach (Bernal et al., 2017; Wagner et al., 2002) that is given as:

$$Y_t = \beta_0 + \beta_1 * time_t + \beta_2 * intervetion_t + \beta_3 * time\ after\ intervetion_t + e_t \text{ .......} \quad (2)$$

In this formula, $\beta_0 + \beta_1 * time_t$+$\beta_3 * time\ after\ intervetion_t$ corresponds to $f(t)$ as they are time-related terms, and $\beta_2 * intervetion_t$ relates to $D_t * b$, which represents intervention effect. As a detailed explanation for each term: $Y_t$ is an aggregated measure of interest at each time point (e.g. mean, rate, and proportion); $time_t$ $(t)$ is a continuous variable that indicates time since the inception of the study; $intervetion_t$ $(i_t)$ is an indicator that is 0 before intervention and 1 thereafter; $time\ after\ intervetion_t$ $(tai_t)$ is a continuous variable that counts time after intervention. In terms of parameters, $\beta_0$ estimates baseline level of the outcome; $\beta_1$ defines any pre-existing trend on how the outcome changes before intervention; $\beta_2$ measures level changes because of intervention; $\beta_3$ represents additional trend changes compared to before intervention. The error term $e_t$ denotes any random variability that not explained by the model, and it usually consists of a normally distributed random error with a possible correlated error term to account for autocorrelation issues (Nelson, 1998; Prais & Winsten, 1954).

Although the formulation of $time\ after\ intervetion_t\ (tai_t)$ is more widely accepted in ITS literatures (e.g. Bernal, et al., 2017; Wagner, et al., 2002), it suggests a fact that part of the intervention effect depends on time, which can be understood as an interaction effect between intervention and time. With this interpretation, equation (2) could also be expressed as:

$$Y_t = \beta_0 + \beta_1 * time_t + \beta_2 * intervetion_t + \beta_3 * time_t * intervetion_t + e_t \ldots\ldots (3)$$

Here, the $time\ after\ intervetion_t\ (tai_t)$ is re-expressed as an equivalent interaction effect between $time_t$ and $intervetion_t$. This expression should be preferred as being more precise from a mathematical perspective, and not introducing additional variable $time\ after\ intervetion_t\ (tai_t)$. On the other hand, it is understandable that many researchers might prefer the model expressed in formula (2), since interpreting and diagnosing an interaction term could be challenging and require more advanced knowledge of ordinary regression context.

In addition, a variety of statistical methods are available to estimate the model parameters, such as ordinary least square (OLS), generalized least squares (GLS), autoregressive integrated moving average (ARIMA), and restricted maximum likelihood (REML) (Turner, et al., 2021; Hudson, et al., 2019; Jandoc, et al., 2015; Wagner, et al., 2002). For simplicity, OLS was applied in our study, and a normally distributed random error $e_t$ was assumed (i.e. no autocorrelation of residuals).

### 3.5 Phase-in Period Analytical Strategies

As previously introduced in chapter one, a phase-in period is a delayed effect with pre-defined or rationalized end date. In this sub-section, six analytical strategies for ITS with phase-in periods were examined. Detailed descriptions and corresponding rationales were provided as follows.

1. As abrupt at start: modelling the full intervention effect as if it occurred abruptly at the start of phase-in period ($t = 100$). This is arguably the most naïve approach, because all the ITS studies should have a properly defined intervention start date, but not necessarily for an end date. In reality, researchers often encounter scenarios such as in the antenatal drinking study from Hankin et al. (1993), where although some forms of delayed effects were acknowledged, there would still be ambiguities for the end date of corresponding implementations, and thus a phase-in period would not be properly defined.
2. As abrupt at end: modelling the full intervention effect as if it occurred abruptly at the end of phase-in period ($t = 120$). Although this approach is rare, it is theoretically possible to model the intervention effect as if it happened abruptly at the end of phase-in period, given the end date can be clearly defined.
3. Excluding phase-in period - before: excluding observations in the phase-in period and modeling the remaining as an abrupt change (i.e. excluding $100 < t < 120$). This is currently the most popular method for dealing with well-defined phase-in periods, because the observations within the period may have ambiguous effects on estimating slope and level

changes. In addition, this strategy assumes that the full-strength intervention effect occurred at the end date of intervention (i.e. a prospective counterfactual).

4. Excluding phase-in period - after: similar as above, but with the assumption that the full-strength intervention effect occurred at the start date of intervention (i.e. a retrospective counterfactual).
5. Counterfactual from before: assuming the trend would stay the same from pre-intervention in the phase-in period, and creating projection observations accordingly. Technically, this means to extend the pre-intervention trend line to the end of phase-in period. This approach assumes that the full-strength intervention effect would happen abruptly at the end of phase-in period, which also corresponds to the prospective counterfactual introduced earlier.
6. Counterfactual from after: assuming that the trend would be the same with post-intervention in the phase-in period and creating projection observations accordingly. Technically, this means to extend the post-intervention trend line to the start of phase-in period. By extending the post-intervention trajectory, this assumes that the full-strength intervention effect would happen abruptly at the start of phase-in period, which also corresponds to the retrospective counterfactual introduced earlier.

Table 3.6 demonstrates the coding scheme in order to prepare the independent variables in equation (2) for the six analytical strategies, respectively. The important values are highlighted in bold face to reflect the setups of the six analytical strategies. To evaluate the appropriateness of the analytical strategies, each was applied to the simulated datasets with different sample sizes and effect sizes, generated under three different population scenarios.

Although excluding phase-in period - before and counterfactual from before, excluding phase-in period - after and counterfactual from after were expressed as different strategies, in fact, they

**Table 3.6**

*Coding Independent Variables in Equation (2) for the Six Analytical Strategies*

| | As abrupt at start | | | As abrupt at end | | | Excluding phase-in period – before | | | Excluding phase-in period – after | | | Counterfactual from before | | | Counterfactual from after | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Time Period** | $t$ | $i_t$ | $tai_t$ | $t$ | $i_t$ | $tai_t$ | $t$ | $i_t$ | $tai_t$ | $t$ | $i_t$ | $tai_t$ | $t$ | $i_t$ | $tai_t$ | $t$ | $i_t$ | $tai_t$ |
| 1-100 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| | 2 | 0 | 0 | 2 | 0 | 0 | 2 | 0 | 0 | 2 | 0 | 0 | 2 | 0 | 0 | 2 | 0 | 0 |
| (Before | 3 | 0 | 0 | 3 | 0 | 0 | 3 | 0 | 0 | 3 | 0 | 0 | 3 | 0 | 0 | 3 | 0 | 0 |
| | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| intervention) | 98 | 0 | 0 | 98 | 0 | 0 | 98 | 0 | 0 | 98 | 0 | 0 | 98 | 0 | 0 | 98 | 0 | 0 |
| | 99 | 0 | 0 | 99 | 0 | 0 | 99 | 0 | 0 | 99 | 0 | 0 | 99 | 0 | 0 | 99 | 0 | 0 |
| | **100** | **0** | **0** | 100 | 0 | 0 | **100** | **0** | **0** | **100** | **0** | **0** | 100 | 0 | 0 | **100** | **0** | **0** |
| 101-120 | **101** | **1** | **1** | 101 | 0 | 0 | 101 | | | 101 | | | 101 | 0 | 0 | **101** | **1** | **1** |
| | 102 | 1 | 2 | 102 | 0 | 0 | 102 | | | 102 | | | 102 | 0 | 0 | 102 | 1 | 2 |
| (Phase-in) | 103 | 1 | 3 | 103 | 0 | 0 | 103 | | | 103 | | | 103 | 0 | 0 | 103 | 1 | 3 |
| | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | | | ⋮ | | | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| | 118 | 1 | 18 | 118 | 0 | 0 | 118 | | | 118 | | | 118 | 0 | 0 | 118 | 1 | 18 |
| | 119 | 1 | 19 | 119 | 0 | 0 | 119 | | | 119 | | | 119 | 0 | 0 | 119 | 1 | 19 |
| | 120 | 1 | 20 | **120** | **0** | **0** | 120 | | | 120 | | | 120 | **0** | **0** | 120 | 1 | 20 |
| 121-200 | 121 | 1 | 21 | **121** | **1** | **1** | **121** | **1** | **1** | **121** | **1** | **21** | **121** | **1** | **1** | 121 | 1 | 21 |
| | 122 | 1 | 22 | 122 | 1 | 2 | 122 | 1 | 2 | 122 | 1 | 22 | 122 | 1 | 2 | 122 | 1 | 22 |
| (After | 123 | 1 | 23 | 123 | 1 | 3 | 123 | 1 | 3 | 123 | 1 | 23 | 123 | 1 | 3 | 123 | 1 | 23 |
| | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| intervention) | 198 | 1 | 98 | 198 | 1 | 78 | 198 | 1 | 78 | 198 | 1 | 98 | 198 | 1 | 78 | 198 | 1 | 98 |
| | 199 | 1 | 99 | 199 | 1 | 79 | 199 | 1 | 79 | 199 | 1 | 99 | 199 | 1 | 79 | 199 | 1 | 99 |
| | 200 | 1 | 100 | 200 | 1 | 80 | 200 | 1 | 80 | 200 | 1 | 100 | 200 | 1 | 80 | 200 | 1 | 100 |
| Other Notes | | | | | | | Phase-in period observations were excluded | | | Phase-in period observations were excluded | | | Phase-in period projections based on counterfactual from before intervention | | | Phase-in period projections based on counterfactual from after intervention | | |

*Note.* $t$ is a continuous variable that indicating time since the inception of the study; ($i_t$) is an indicator that is 0 before the intervention and 1 thereafter; $\text{time after intervetion}_t$ ($tai_t$) is a continuous variable that counts time after the intervention.

were actually different expressions on the same counterfactual assumptions. The link between these shared assumptions will be more obvious in the simulation results section.

## 3.6 Result

In this sub-section, figures and tables based on one randomly selected iteration with the simulated trajectories were included, with summarized key findings. Also, Table 3.7 included aggregated accuracy performance results from the one hundred iterations of simulations, which suggested that abrupt modelling constantly led to poor accuracy in estimations, while excluding phase-in period - after and counterfactual from after provided satisfactory results. It was interesting to observe the accuracy discrepancies for excluding phase-in period - before and counterfactual from before, where they were only not able to obtain accurate level change estimations in scenario-3. Since there was no significant difference in accuracy across different effect sizes and sample sizes, Table 3.7 was aggregated at the strategy level to highlight the different performance from different analytical strategies. However, this does not suggest that effect sizes and sample sizes have no effect on estimation accuracy. In our study, the models from different strategies were either perfectly specified, or extremely mis-specified (from the abrupt change); in both cases, effect sizes and sample sizes may not contribute significantly to the estimation performance.

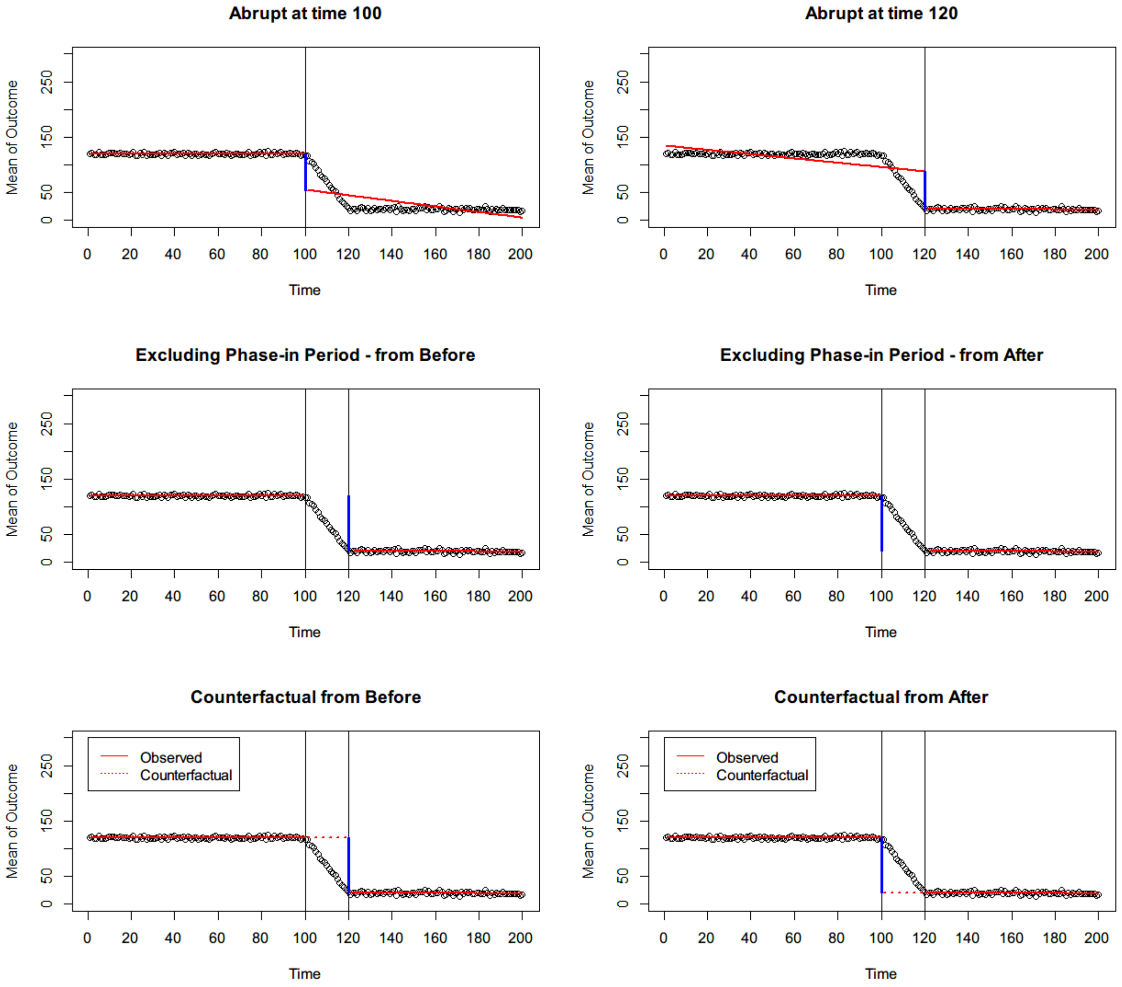
To better observe and understand the performance differences from each analytical strategy, the remaining of this sub-section focused on the one randomly selected iteration from the one hundred simulations. For scenario-1 to scenario-3 respectively, Figure 3.5-3.7 displayed model fitting results, shown as red line segments, from the six analytical strategies. In addition, different level change causal effects were labelled with blue vertical bars in each figure. They demonstrated that, for all three scenarios, abrupt change modelling (as abrupt at start or as abrupt at end) constantly yield to inaccurate approximation in pre- or post-intervention sections. Specifically, the pre-intervention series was unfit when assuming the full intervention effect happened abruptly at the end of phase-in period, and the post-intervention series was unfit when assuming the full intervention effect happened abruptly at the start of phase-in period. The remaining techniques provided reasonable fittings to the simulated observations. This could be observed from Figure 3.5-3.7, where all the fitted lines from these techniques aligned closely with the observations.

One thing that worth noting was the discrepancy between the level change causal effects for scenario-3 in Figure 3.7; the differences were resulted from the change of slope. Table 3.8-3.10 compared the performance of six analytical strategies by reporting causal effect point estimates regarding level and slope changes with 95% confidence intervals. Overall, the abrupt methods led to inaccurate estimates, and the inaccuracy in level change exacerbated as effect sizes increased. Counterfactual from before and excluding phase-in period - before produced correct estimations for scenario-1 and scenario-2 when there was no slope change; however, they led to imprecise level change estimations for scenario-3 when there was a slope change. With the additional slope change, the prospective counterfactual would introduce a bias with size of the difference in slope between before and after intervention trend lines, similar to what the simple series demonstrated in Chapter 1.2.3. In contrast, excluding phase-in period - after and counterfactual from after strategies constantly provided accurate estimations. It was also interesting to notice that the causal effect points estimations and 95% confidence intervals between excluding phase-in period - before and counterfactual from before, excluding phase-in period - after and counterfactual from after matched almost perfectly, since they shared the same counterfactual assumptions.

In addition, Figure 3.8-3.13 contained residual plots and normal Q-Q plots from different analytical strategies in each scenario, as a reference for possible modelling issues from the corresponding ordinary linear regression fits.

**Figure 3.5**

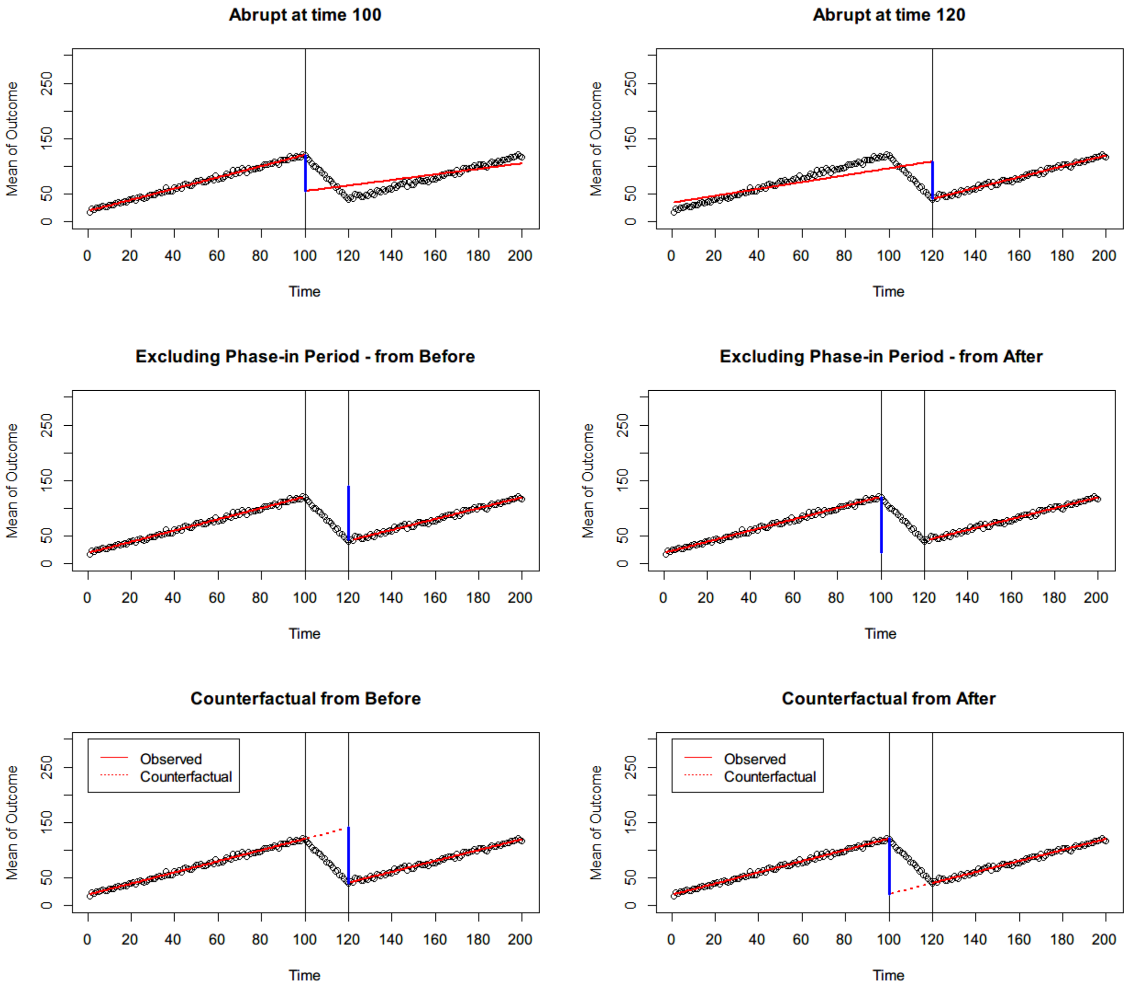
*Mean of Outcome Trajectories with Phase-In Period and Fitted Models for Scenario-1*



*Note.* Red line: fitted linear regression, blue bars: level change casual effect.

**Figure 3.6**

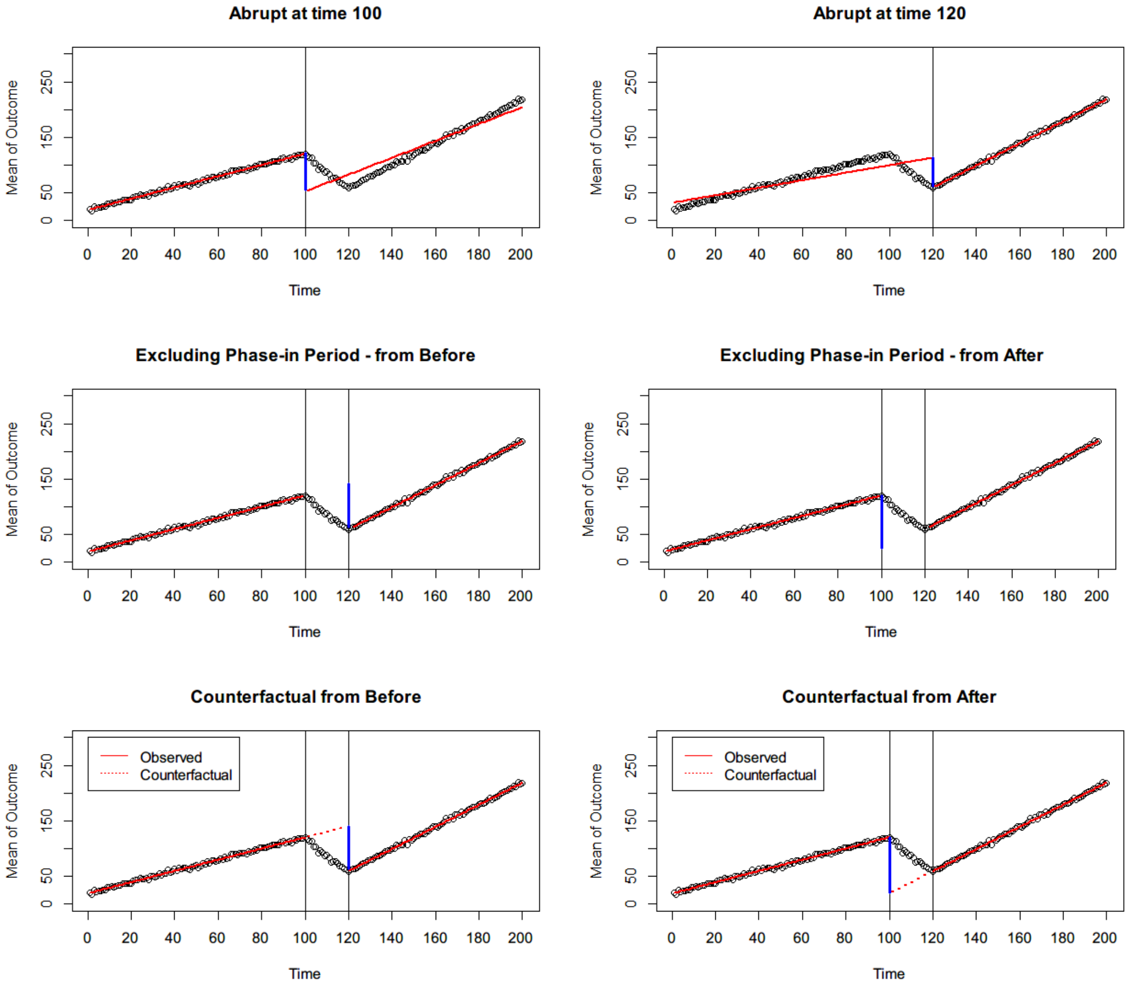
*Mean of Outcome Trajectories with Phase-In Period and Fitted Models for Scenario-2*



*Note.* Red line: fitted linear regression, blue bars: level change casual effect.

**Figure 3.7**

*Mean of Outcome Trajectories with Phase-In Period and Fitted Models for Scenario-3*



*Note.* Red line: fitted linear regression, blue bars: level change casual effect.

**Table 3.7**

*Aggregated Performance from One Hundred Simulations on Different Strategies*

| Scenario/ Analytical Strategies | As abrupt at time 100 | As abrupt at time 120 | Excluding phase-in period - before | Excluding phase-in period - after | Counterfactual from before | Counterfactual from after |
|---|---|---|---|---|---|---|
| Scenario-1 | Level: 0.11% | 0.00% | 100.00% | 100.00% | 99.78% | 100.00% |
| | Slope: 0.56% | 0.89% | 95.44% | 97.11% | 93.89% | 95.89% |
| Scenario-2 | 0.11% | 0.22% | 97.56% | 99.33% | 97.89% | 99.00% |
| | 1.56% | 2.89% | 92.22% | 95.22% | 91.67% | 96.33% |
| Scenario-3 | 0.58% | 0.42% | 0.50% | 97.75% | 0.92% | 98.00% |
| | 4.50% | 6.00% | 97.00% | 95.25% | 92.58% | 94.83% |

*Note.* Red color represents low accuracy (i.e. accuracy< 10%), while green color denotes high accuracy (i.e. accuracy> 90%).

**Table 3.8**

*Performance from One Simulation on Different Strategies for Scenario-1*

| Modelling methods/ Effect size & Sample size | | As abrupt at time 100 | As abrupt at time 120 | Excluding phase-in period - before | Excluding phase-in period - after | Counterfactual from before | Counterfactual from after |
|---|---|---|---|---|---|---|---|
| True level change: -20<br>True slope change: 0 | Sample size 20 | Level: $-12.97^{***}$ [-14.93, -11.01] | $-13.72^{***}$ [-15.84, -11.60] | $-20.23^{***}$ [-21.60, -18.85] | $-20.29^{***}$ [-21.75, -18.82] | $-20.43^{***}$ [-21.52, -19.33] | $-20.04^{***}$ [-21.12, -18.96] |
| | | Slope: $-0.10^{***}$ [-0.13, -0.06] | $0.08^{***}$ [0.04, 0.12] | 0 [-0.02, 0.03] | 0 [-0.02, 0.03] | 0.01 [-0.01, 0.03] | 0 [-0.02, 0.02] |
| | 100 | $-13.08^{***}$ [-14.60, -11.57] | $-13.42^{***}$ [-15.18, -11.66] | $-19.84^{***}$ [-20.53, -19.16] | $-19.76^{***}$ [-20.49, -19.03] | $-19.86^{***}$ [-20.40, -19.33] | $-19.81^{***}$ [-20.33, -19.28] |
| | | $-0.10^{***}$ [-0.13, -0.07] | $0.07^{***}$ [0.04, 0.11] | 0 [-0.02, 0.01] | 0 [-0.02, 0.01] | -0.01 [-0.02, 0] | 0 [-0.01, 0.01] |
| | 1000 | $-12.54^{***}$ [-14.06, -11.03] | $-14.11^{***}$ [-15.70, -12.53] | $-19.97^{***}$ [-20.20, -19.73] | $-19.98^{***}$ [-20.23, -19.73] | $-20.02^{***}$ [-20.20, -19.83] | $-19.98^{***}$ [-20.16, 19.80] |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | -0.11$^{***}$<br>[-0.13, -0.08] | 0.07$^{***}$<br>[0.04, 0.10] | 0<br>[0, 0] | 0<br>[0, 0] | 0<br>[0, 0] | 0<br>[0, 0] |
| True level change:<br>-50<br>True slope change:<br>0 | 20 | -32.75$^{***}$<br>[-36.52, -28.98] | -33.89$^{***}$<br>[-38.21, -29.57] | -50.01$^{***}$<br>[-51.51, -48.51] | -50.04$^{***}$<br>[-51.64, -48.43] | -50.07$^{***}$<br>[-51.24, -48.89] | -49.86$^{***}$<br>[-51.02, -48.70] |
| | | -0.24$^{***}$<br>[-0.31, -0.18] | 0.20$^{***}$<br>[0.12, 0.28] | 0<br>[-0.02, 0.03] | 0<br>[-0.02, 0.03] | 0<br>[-0.02, 0.03] | 0<br>[-0.02, 0.02] |
| | 100 | -32.82$^{***}$<br>[-36.40, -29.24] | -33.83$^{***}$<br>[-38.01, -29.65] | -49.84$^{***}$<br>[-50.52, -49.17] | -49.83$^{***}$<br>[-50.55, -49.10] | -49.85$^{***}$<br>[-50.38, -49.32] | -49.84$^{***}$<br>[-50.36, -49.32] |
| | | -0.25$^{***}$<br>[-0.31, -0.18] | 0.19$^{***}$<br>[0.11, 0.27] | 0<br>[-0.01, 0.01] | 0<br>[-0.01, 0.01] | 0<br>[-0.01, 0.01] | 0<br>[-0.01, 0.01] |
| | 1000 | -32.78$^{***}$<br>[-36.44, -29.11] | -33.95$^{***}$<br>[-38.18, -29.72] | -49.94 $^{***}$<br>[-50.16, -49.72] | -49.98 $^{***}$<br>[-50.22, -49.74] | -50.01 $^{***}$<br>[-50.19, -49.83] | -50.00 $^{***}$<br>[-50.18, 49.83] |
| | | -0.25$^{***}$<br>[-0.31, -0.18] | 0.20$^{***}$<br>[0.12, 0.28] | 0<br>[0, 0.01] | 0<br>[0, 0.01] | 0<br>[-0.02, 0.06] | 0<br>[0, 0] |
| True level change:<br>-100<br>True slope change:<br>0 | 20 | -64.82$^{***}$<br>[-71.95, -57.69] | -67.18$^{***}$<br>[-75.50, -58.87] | -99.12$^{***}$<br>[-100.59, -97.65] | -98.72$^{***}$<br>[-100.29, -97.15] | -99.04$^{***}$<br>[-100.19, -97.89] | -99.19$^{***}$<br>[-100.33, -98.05] |
| | | -0.50$^{***}$<br>[-0.63, -0.38] | 0.37$^{***}$<br>[0.21, 0.53] | -0.02<br>[-0.05, 0] | -0.02<br>[-0.05, 0] | -0.02<br>[-0.06, 0.10] | -0.01<br>[-0.03, 0.01] |
| | 100 | -65.34$^{***}$<br>[-72.56, -58.13] | -68.47$^{***}$<br>[-76.71, -60.22] | -100.10$^{***}$<br>[-100.76, -99.44] | -100.20$^{***}$<br>[-100.91, -99.50] | -100.14$^{***}$<br>[-100.65, -99.62] | -100.03$^{***}$<br>[-100.53, -99.52] |
| | 100 | -0.50$^{***}$<br>[-0.62, -0.37] | 0.39$^{***}$<br>[0.23, 0.55] | 0.01<br>[-0.01, 0.02] | 0.01<br>[-0.01, 0.02] | 0<br>[-0.01, 0.01] | 0<br>[-0.01, 0.01] |
| | 1000 | -65.29$^{***}$<br>[-72.46, -58.11] | -76.12$^{***}$<br>[-85.84, -66.40] | -99.95$^{***}$<br>[-100.18, -99.72] | -100.02$^{***}$<br>[-100.27, -99.77] | -100.03$^{***}$<br>[-100.21, -99.84] | -100.03$^{***}$<br>[-100.21, -99.85] |
| | | -0.50$^{***}$<br>[-0.62, -0.37] | 0.39$^{***}$<br>[0.23, 0.54] | 0<br>[0, 0.01] | 0<br>[0, 0.01] | 0<br>[0, 0.01] | 0<br>[0, 0.01] |

*Note.* Values in the table were point estimates, and the 95% confidence intervals are shown in squared brackets. Red color represents true value is not in 95% CI, while green color denotes correct conclusion. * p-value< 0.05, ** < 0.01, *** < 0.001.

**Table 3.9**

*Performance from One Simulation on Different Strategies for Scenario-2*

| Modelling methods/ Effect size & Sample size | | As abrupt at time 100 | As abrupt at time 120 | Excluding phase-in period - before | Excluding phase-in period - after | Counterfactual from before | Counterfactual from after |
|---|---|---|---|---|---|---|---|
| True level change: -20<br>True slope change: 0 | Sample size 20 | Level: -13.66$^{***}$<br>[-15.53, -11.79] | -13.09$^{***}$<br>[-15.35, -10.84] | -20.27$^{***}$<br>[-21.87, -18.67] | -19.68$^{***}$<br>[-21.39, -17.97] | -19.73$^{***}$<br>[-20.98, -18.48] | -20.18$^{***}$<br>[-21.41, -18.95] |
| | | Slope: -0.12$^{***}$<br>[-0.15, -0.08] | 0.06$^{***}$<br>[0.01, 0.10] | -0.03$^{*}$<br>[-0.06, 0] | -0.03$^{*}$<br>[-0.06, 0] | -0.03$^{*}$<br>[-0.04, 0] | -0.02$^{*}$<br>[-0.04, 0] |
| | 100 | -12.61$^{***}$<br>[-14.11, -11.11] | -13.88$^{***}$<br>[-15.48, -12.28] | -19.71$^{***}$<br>[-20.36, -19.05] | -19.76$^{***}$<br>[-20.46, -19.06] | -19.94$^{***}$<br>[-20.45, -19.43] | -19.76$^{***}$<br>[-20.27, -19.26] |
| | | -0.10$^{***}$<br>[-0.13, -0.08] | 0.07$^{***}$<br>[0.04, 0.10] | 0<br>[-0.01, 0.01] | 0<br>[-0.01, 0.01] | 0<br>[-0.01, 0.01] | 0<br>[-0.01, 0.01] |
| | 1000 | -13.18$^{***}$<br>[-14.63, -11.74] | -13.44$^{***}$<br>[-15.13, -11.74] | -19.89$^{***}$<br>[-20.10, -19.68] | -19.91$^{***}$<br>[-20.14, -19.69] | -19.96$^{***}$<br>[-20.12, -19.80] | -19.90$^{***}$<br>[-20.06, -19.74] |
| | | -0.10$^{***}$<br>[-0.12, -0.07] | 0.08$^{***}$<br>[0.05, 0.11] | 0<br>[0, 0] | 0<br>[0, 0] | 0<br>[0, 0] | 0<br>[0, 0] |
| True level change: -50<br>True slope change: 0 | 20 | -33.35$^{***}$<br>[-37.17, -29.53] | -34.68$^{***}$<br>[-39.08, -30.29] | -50.98$^{***}$<br>[-52.44, -49.52] | -50.83$^{***}$<br>[-52.39, -49.26] | -50.67$^{***}$<br>[-51.82 -49.52] | -50.85$^{***}$<br>[-51.98, -49.72] |
| | | -0.26$^{***}$<br>[-0.32, -0.19] | 0.19$^{***}$<br>[0.11, 0.27] | -0.01<br>[-0.03, 0.02] | -0.01<br>[-0.03, 0.02] | 0<br>[-0.02, 0.02] | -0.01<br>[-0.03, 0.01] |
| | 100 | -32.79$^{***}$<br>[-36.34, -29.25] | -33.55$^{***}$<br>[-37.72, -29.38] | -49.53$^{***}$<br>[-50.20, -48.86] | -49.55$^{***}$<br>[-50.27, -48.83] | -49.76$^{***}$<br>[-50.29, -49.23] | -49.68$^{***}$<br>[-50.20, -49.17] |
| | | -0.24$^{***}$<br>[-0.30, -0.18] | 0.20$^{***}$<br>[0.12, 0.28] | 0<br>[-0.01, 0.01] | 0<br>[-0.01, 0.01] | 0<br>[-0.01, 0.01] | 0<br>[-0.01, 0.01] |
| | 1000 | -32.72$^{***}$<br>[-36.30, -29.13] | -33.74$^{***}$<br>[-37.92, -29.56] | -49.75$^{***}$<br>[-49.96, -49.54] | -49.73$^{***}$<br>[-49.95, -49.50] | -49.80$^{***}$<br>[-49.97, -49.64] | -49.81 $^{***}$<br>[-49.97, -49.65] |
| | | -0.25$^{***}$<br>[-0.31, -0.18] | 0.19$^{***}$<br>[0.11, 0.27] | 0<br>[0, 0] | 0<br>[0, 0] | 0<br>[-0.01, 0] | 0<br>[0, 0] |
| True level change: -100 | 20 | -65.47$^{***}$<br>[-72.51, -58.42] | -67.16$^{***}$<br>[-75.54, -58.77] | -98.77$^{***}$<br>[-100.53, -97.02] | -98.77$^{***}$<br>[-100.53, -97.02] | -99.05$^{***}$<br>[-100.33, -97.77] | -99.91$^{***}$<br>[-101.18, -98.64] |
| | | -0.51$^{***}$<br>[-0.63, -0.39] | 0.36$^{***}$<br>[0.20, 0.52] | -0.03$^{*}$<br>[-0.06, -0.01] | -0.03$^{*}$<br>[-0.06, 0] | -0.03$^{*}$<br>[-0.05, 0] | -0.02<br>[-0.04, 0] |
| | 100 | -65.38$^{***}$<br>[-72.57, -58.18] | -68.01$^{***}$<br>[-76.30, -59.72] | -99.73$^{***}$<br>[-100.44, -99.03] | -99.82$^{***}$<br>[-100.57, -99.06] | -99.99$^{***}$<br>[-100.54, -99.44] | -99.81$^{***}$<br>[-100.35, -99.26] |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| True slope change:<br>0 | 100 | -0.49$^{***}$<br>[-0.62, -0.37] | 0.39$^{***}$<br>[0.23, 0.55] | 0<br>[-0.01, 0.02] | 0<br>[-0.01, 0.02] | 0<br>[-0.01, 0.01] | 0<br>[-0.01, 0.01] |
| | 1000 | -65.36$^{***}$<br>[-72.50, -58.21] | -76.07$^{***}$<br>[-85.78, -66.36] | -99.95$^{***}$<br>[-100.16, -99.73] | -100.02$^{***}$<br>[-100.25, -99.79] | -100.02$^{***}$<br>[-100.19, -99.86] | -99.96$^{***}$<br>[-100.12, -99.80] |
| | | -0.49$^{***}$<br>[-0.62, -0.37] | 0.39$^{***}$<br>[0.23, 0.55] | 0$^{*}$<br>[0, 0.01] | 0$^{*}$<br>[0, 0.01] | 0$^{*}$<br>[0, 0.01] | 0<br>[0, 0.01] |

*Note.* Values in the table were point estimates, and the 95% confidence intervals are shown in squared brackets. Red color represents true value is not in 95% CI, while green color denotes correct conclusion. * p-value< 0.05, ** < 0.01, *** < 0.001.

**Table 3.10**

*Performance from One Simulation on Different Strategies for Scenario-3*

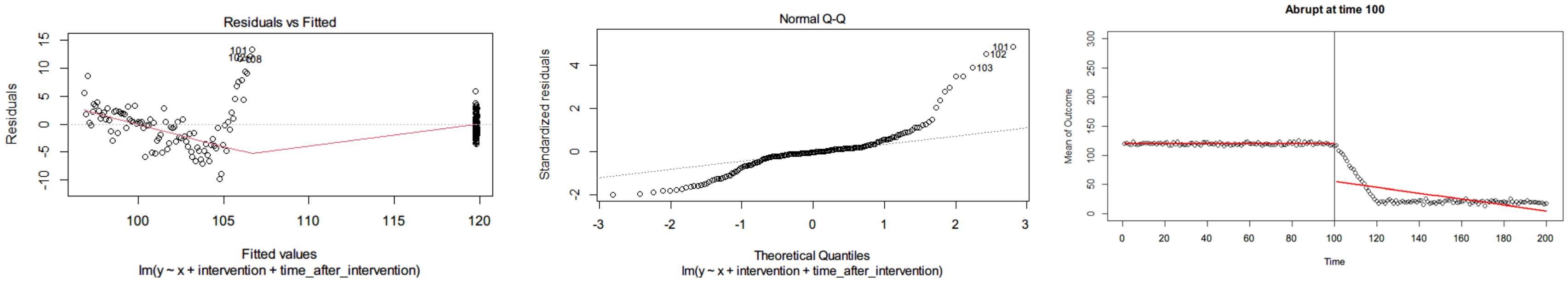
| Modelling methods/<br>Effect size & Sample size | | As abrupt at time 100 | As abrupt at time 120 | Excluding phase-in period - before | Excluding phase-in period - after | Counterfactual from before | Counterfactual from after |
|---|---|---|---|---|---|---|---|
| True level change:<br>0<br><br>True slope change:<br>1 | Sample size<br>20 | Level: -3.44$^{***}$<br>[-4.70, -2.18] | 14.38$^{***}$<br>[12.56, 16.20] | 18.36$^{***}$<br>[16.97, 19.76] | -1.41<br>[-2.89, 0.08] | 18.55$^{***}$<br>[17.45, 19.65] | -1.36$^{*}$<br>[-2.46, 0.25] |
| | | Slope: 1.02$^{***}$<br>[0.99, 1.04] | 0.94$^{***}$<br>[0.91, 0.98] | 0.99$^{***}$<br>[0.96, 1.01] | 0.99$^{***}$<br>[0.96, 1.01] | 0.99$^{***}$<br>[0.97, 1.01] | 0.99$^{***}$<br>[0.97, 1.01] |
| | 100 | -2.12$^{***}$<br>[-2.78, -1.46] | 14.88$^{***}$<br>[13.37, 16.39] | 19.45$^{***}$<br>[18.81, 20.10] | -0.43<br>[-1.13, 0.26] | 19.45$^{***}$<br>[18.94, 19.95] | -0.75$^{**}$<br>[-1.25, 0.24] |
| | | 1.02$^{***}$<br>[1.01, 1.03] | 0.94$^{***}$<br>[0.91, 0.97] | 0.99$^{***}$<br>[0.98, 1.01] | 0.99$^{***}$<br>[0.98, 1.01] | 0.99$^{***}$<br>[0.99, 1.00] | 1$^{***}$<br>[0.99, 1.01] |
| | 1000 | -2.74$^{***}$<br>[-3.23, -2.24] | 14.69$^{***}$<br>[13.24, 16.14] | 19.04$^{***}$<br>[18.82, 19.25] | -0.23$^{***}$<br>[-1.23, 0.76] | 19.03$^{***}$<br>[18.86, 19.20] | -0.36$^{***}$<br>[-1.16, 0.83] |
| | | 1.03`$^{***}$<br>[1.02, 1.03] | 0.95$^{***}$<br>[0.92, 0.98] | 1$^{***}$<br>[1, 1.01] | 1$^{***}$<br>[1, 1.01] | 1$^{***}$<br>[1, 1] | 1$^{***}$<br>[1, 1] |
| True level change:<br>-20 | 20 | -16.53$^{***}$<br>[-18.23, -14.83] | 0.75$^{***}$<br>[-0.63, 2.12] | -0.74$^{*}$<br>[-2.16, -1.33] | -20.56$^{***}$<br>[-22.07, -19.05] | -0.45<br>[-1.55, 0.65] | -20.31$^{***}$<br>[-21.39, -19.22] |
| | | 0.92$^{***}$<br>[0.89, 0.95] | 1.02$^{***}$<br>[0.99, 1.05] | 0.99$^{***}$<br>[0.97, 1.01] | 0.99$^{***}$<br>[0.97, 1.01] | 0.99$^{***}$<br>[0.97, 1.01] | 0.99$^{***}$<br>[0.97, 1.01] |
| | 100 | -14.56$^{***}$<br>[-15.87, -13.25] | 1.93$^{***}$<br>[1.24, 2.61] | 0.54<br>[-0.19, 1.10] | -19.43$^{***}$<br>[-20.13, -18.74] | 0.45<br>[-0.06, 0.95] | -19.70$^{***}$<br>[-20.20, -19.20] |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| True slope change:<br>1 | 100 | 0.93$^{***}$<br>[0.90, 0.95] | 1.01$^{***}$<br>[1, 1.03] | 0.99$^{***}$<br>[0.98, 1.01] | 0.99$^{***}$<br>[0.98, 1.01] | 0.99$^{***}$<br>[0.99, 1.00] | 1$^{***}$<br>[0.99, 1.01] |
| | 1000 | -15.17$^{***}$<br>[-16.37, -13.98] | 1.73$^{***}$<br>[1.16, 2.31] | 0.06<br>[-0.18, 0.25] | -20.00$^{***}$<br>[-20.23, -19.76] | 0.03<br>[-0.14, 0.20] | -19.99$^{***}$<br>[-20.16, -19.83] |
| | | 0.93$^{***}$<br>[0.91, 0.95] | 1.02$^{***}$<br>[1.01, 1.03] | 1$^{***}$<br>[1, 1.01] | 1$^{***}$<br>[1, 1.01] | 1$^{***}$<br>[1, 1] | 1$^{***}$<br>[1, 1] |
| True level change:<br>-50<br>True slope change:<br>1 | 20 | -35.42$^{***}$<br>[-38.97, -31.87] | -19.00$^{***}$<br>[-22.10, -15.89] | -31.67$^{***}$<br>[-33.10, -30.24] | -50.60$^{***}$<br>[-52.13, -49.07] | -31.41$^{***}$<br>[-32.53, -30.29] | -50.39$^{***}$<br>[-51.49, -49.29] |
| | | 0.78$^{***}$<br>[0.72,0.84] | 1.14$^{***}$<br>[1.08, 1.20] | 1$^{***}$<br>[0.97, 1.02] | 1$^{***}$<br>[0.97, 1.02] | 1$^{***}$<br>[0.98, 1.02] | 0.99$^{***}$<br>[0.97, 1.01] |
| | 100 | -34.68$^{***}$<br>[-38.05, -31.32] | -18.85$^{***}$<br>[-21.64, -16.07] | -31.05$^{***}$<br>[-31.71, -30.39] | -50.09$^{***}$<br>[-50.80, -49.39] | -30.02$^{***}$<br>[-30.54, -29.51] | -49.95$^{***}$<br>[-50.46, -49.44] |
| | | 0.78$^{***}$<br>[0.72,0.84] | 1.14$^{***}$<br>[1.09, 1.19] | 1$^{***}$<br>[0.99, 1.01] | 1$^{***}$<br>[0.99, 1.01] | 1$^{***}$<br>[0.99, 1.01] | 1$^{***}$<br>[0.99, 1.01] |
| | 1000 | -34.48$^{***}$<br>[-37.86, -31.10] | -18.80$^{***}$<br>[-21.57, -16.03] | -30.88$^{***}$<br>[-31.10, -30.67] | -49.85$^{***}$<br>[-50.08, -49.61] | -29.87$^{***}$<br>[-30.04, -29.71] | -49.91$^{***}$<br>[-50.08, -49.75] |
| | | 0.78$^{***}$<br>[0.72, 0.84] | 1.13$^{***}$<br>[1.08, 1.19] | 1$^{***}$<br>[0.99, 1] | 1$^{***}$<br>[0.99, 1] | 1$^{***}$<br>[0.99, 1] | 1$^{***}$<br>[1, 1] |
| True level change:<br>-100<br>True slope change:<br>1 | 20 | -67.05$^{***}$<br>[-74.01, -60.09] | -52.80$^{***}$<br>[-59.75, -45.84] | -80.83$^{***}$<br>[-82.20, -79.46] | -99.42$^{***}$<br>[-100.88, -97.95] | -80.65$^{***}$<br>[-81.72, -79.58] | -99.92$^{***}$<br>[-100.98, -98.87] |
| | | 0.51$^{***}$<br>[0.39,0.63] | 1.31$^{***}$<br>[1.18, 1.44] | 0.98$^{***}$<br>[0.96, 1] | 0.98$^{***}$<br>[0.96, 1] | 0.98$^{***}$<br>[0.96, 1] | 0.99$^{***}$<br>[0.97, 1.00] |
| | 100 | -67.87$^{***}$<br>[-74.71, -61.03] | -52.08$^{***}$<br>[-59.06, -45.09] | -80.74$^{***}$<br>[-81.44, -80.04] | -99.68$^{***}$<br>[-100.43, -98.93] | -79.78$^{***}$<br>[-80.33, -79.24] | -99.82$^{***}$<br>[-100.36, -99.28] |
| | | 0.54$^{***}$<br>[0.42,0.66] | 1.33$^{***}$<br>[1.20, 1.47] | 1$^{***}$<br>[0.98, 1.01] | 1$^{***}$<br>[0.98, 1.01] | 1$^{***}$<br>[0.99, 1.01] | 1$^{***}$<br>[0.99, 1.01] |
| | 1000 | -67.87$^{***}$<br>[-74.74, -61.01] | -79.00$^{***}$<br>[-87.25, -70.75] | -80.87$^{***}$<br>[-81.09, -80.66] | -99.85$^{***}$<br>[-100.08, -99.62] | -79.88$^{***}$<br>[-80.05, -79.71] | -99.92$^{***}$<br>[-100.09, -99.76] |
| | | 0.54$^{***}$<br>[0.42, 0.66] | 1.33$^{***}$<br>[1.20, 1.47] | 1$^{***}$<br>[0.99, 1] | 1$^{***}$<br>[0.99, 1] | 1$^{***}$<br>[1, 1] | 1$^{***}$<br>[1, 1] |

*Note.* Values in the table were point estimates, and the 95% confidence intervals are shown in squared brackets. Red color represents true value is not in 95% CI, while green color denotes correct conclusion. * p-value< 0.05, ** < 0.01, *** < 0.001.

## Figure 3.8

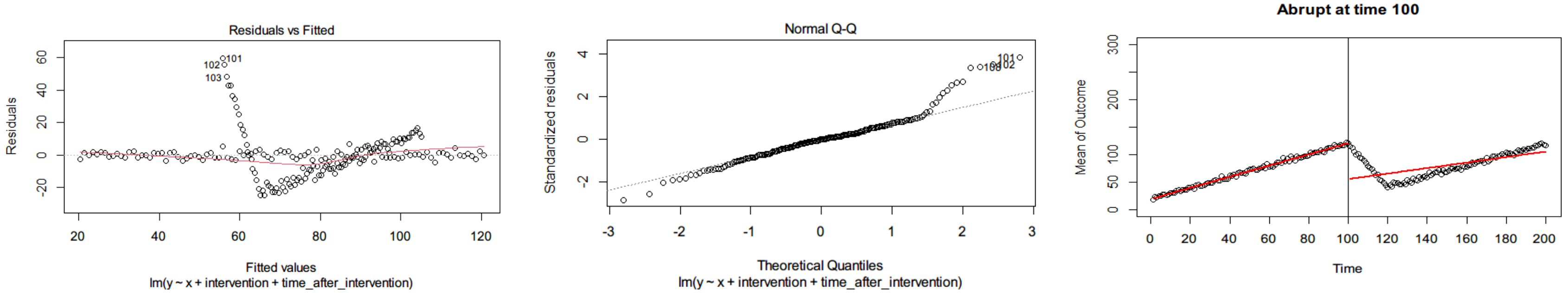
*Fitted Trajectory, Residual Plot and Normal Q-Q Plot from Scenario-1*



*Note.* Level change of -100, sample size 20, as abrupt at start.

## Figure 3.9

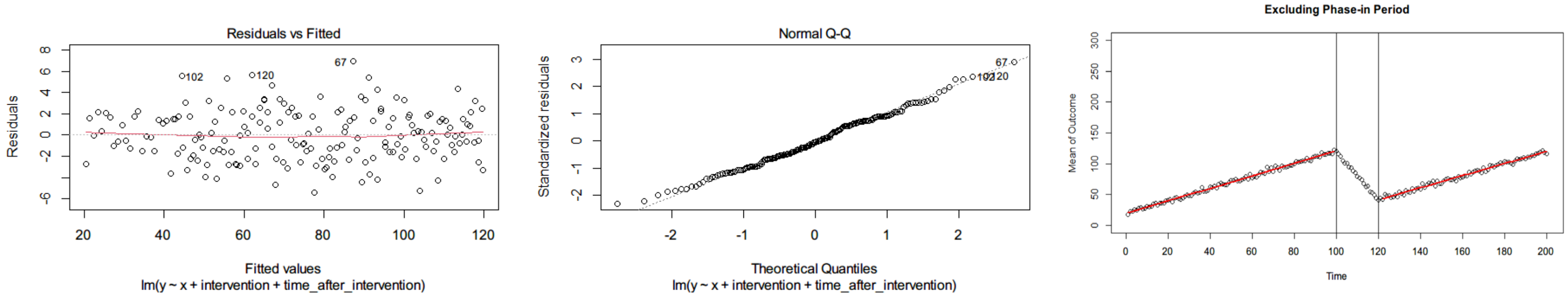
*Fitted Trajectory, Residual Plot and Normal Q-Q Plot from Scenario-2*



*Note.* Level change of -100, sample size 20, as abrupt at start.

**Figure 3.10**

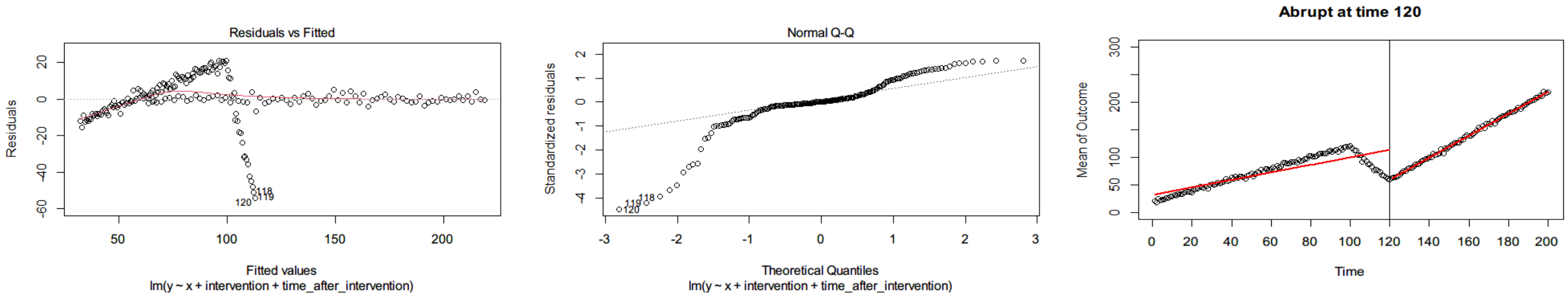
*Fitted Trajectory, Residual Plot and Normal Q-Q Plot from Scenario-2*



*Note.* Level change of -100, sample size 20, excluding phase-in period - before.

**Figure 3.11**

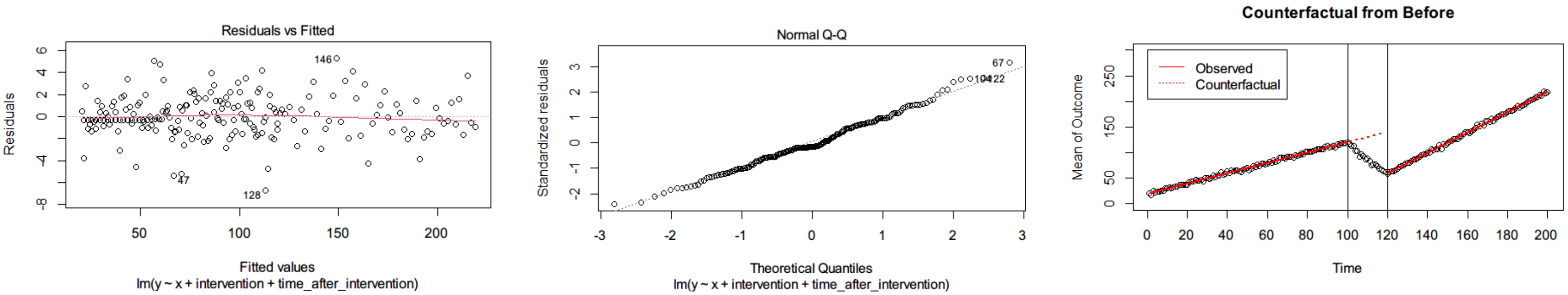
*Fitted Trajectory, Residual Plot and Normal Q-Q Plot from Scenario-3*



*Note.* Level change of -100, sample size 20, as abrupt at end.

**Figure 3.12**

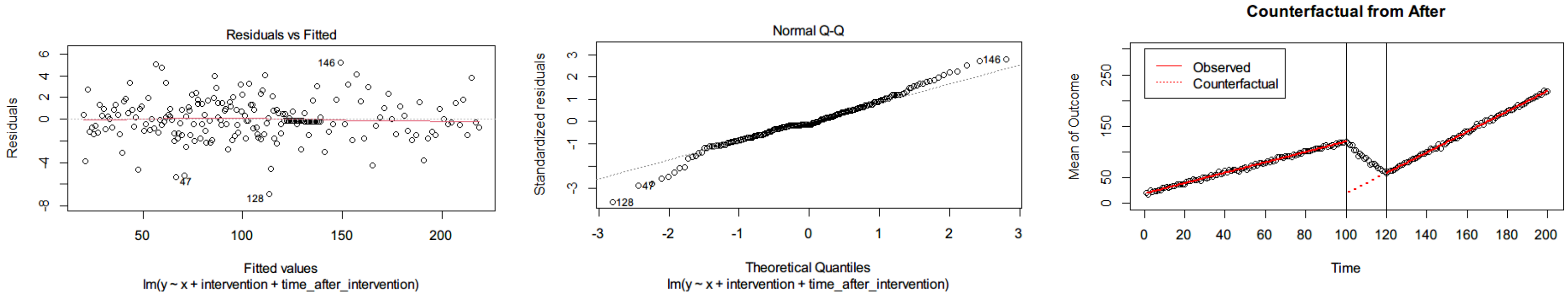
*Fitted Trajectory, Residual Plot and Normal Q-Q Plot from Scenario-3*



*Note.* Level change of -100, sample size 20, counterfactual from before.

**Figure 3.13**

*Fitted Trajectory, Residual Plot and Normal Q-Q Plot from Scenario-3*



*Note.* Level change of -100, sample size 20, counterfactual from after.

# Chapter 4: Conclusions and Discussions

This thesis showcases the characteristics of ITS trajectories under three different scenarios and clarifies the formal definition of phase-in periods. Six analytical strategies for estimating level and slope changes are proposed and examined by segmented linear regression using ordinary least square estimation method.

This chapter reflects on the findings from the simulation study, evaluates the significance and limitation of this thesis, and discusses potential future studies.

## Conclusions

### 4.1 Phase-in Period Definition

Although earlier we argued that phase-in period was a special form of delayed effect, not all delayed effects were guaranteed as phase-in periods. Both delayed effect and phase-in period in ITS studies would have a clearly defined start date (when the intervention begins), and some forms of detained full-strength effect; however, the key difference between these two is whether a clearly pre-defined or rationalized end date (when the intervention completes) exists.

In the previous antenatal drinking study from Hankin et al. (1993), although a clear delayed effect was identified, which was both proven by a seven-month lag of observed impact from outcome score, and a four-month delay in awareness trend collected from interviews, there was no pre-defined end date for this delayed effect. Actually, the delayed duration was identified by analysis instead of based on rationales, and such duration might vary upon modifications of analysis strategies or criteria (e.g. how was a meaningful impact of outcome score defined).

Therefore, although the antenatal drinking study involved with a delayed effect, the implementation should not be modelled as a phase-in period.

In comparison, the antidepressant usage study from Lu et al. (2014) examined a delayed effect with pre-defined start and end date. Generally speaking, the warnings and media coverages were unlikely to cover all the target population at once, thus resulted in a delayed full effect. Moreover, the start and end dates were defined as the first and last issues of warnings (2003 Q4 – 2004 Q4), which were specific dates defined prior to any analysis phases; also, setting first and last issues of warnings was a reasonable choice in terms of measuring the impact from corresponding policy. In conclusion, this delayed effect could be modelled as a phase-in period.

## 4.2 Phase-in Period Characteristics

In the current simulation study, the phase-in periods' trajectories demonstrated gradual transition over a short period of time from the pre-intervention trends to the post-intervention trends. The magnitude of this transition was determined by the size of corresponding level changes. Also, the shapes were approximately linear without slope changes and curvilinear with a slope change. This is reasonable because when there was no slope change, and the outcome was assumed to change at a constant rate within the phase-in period, as imagining the former scenario to shift the trend line gradually from pre-intervention to post-intervention, and the resulted trajectory would be linear. When there was slope change, because the aggregated outcome would change at a varying rate, the line would follow a curvilinear trajectory. Moreover, as expected, a bigger sample group size (n) turned out to have less fluctuation in the series, and vice versa.

However, it is important to realize that all the simulations were performed with assumptions that both pre- and post-intervention trends were strictly linear, and the intervention was implemented at a constant rate. In contrast, the intervention might be implemented at changing rates, and the pre- and post-intervention trends could be in any forms. For example, the rates of antidepressant use for adolescents gradually shifted from pre-intervention to post-intervention trend, as indicated in Figure 2.2, which aligned with scenarios from our simulations. However, rates for young adults and adults shifted after two to three data points and remained consistent with the post-intervention trends.

As a result, a gradual shift from a stable pre-intervention trend to a stable post-intervention trend could be an indication of potential phase-in period. Still, such conclusion would need to be backed up by evidence from the underlying processes. Some examples of these evidence could be the implementation process were completed via multiple stages or throughout a continuous time period (e.g. multiple campaigns for a policy, different roll-out phases).

## 4.3 Analytical Strategies on Phase-in Period

The model estimation results from simulations suggested that excluding phase-in period observations - after and the counterfactual from after strategies yielded unbiased estimations. Also, ignoring potential delayed effects and modelling abruptly has constantly led to biased estimations, both in level and slope changes; this was resulted as fitting either pre- or post-intervention series jointly with the phase-in period series. The counterfactual from before and excluding phase-in period observations – before strategies performed well in the scenarios without slope changes (only level changes); however, with slope changes, these strategies produced biased estimations in level changes. Results from Figure 3.7 and Table 3.10 revealed such discrepancies visually and numerically.

In addition, same counterfactual assumption (retrospective or prospective) led to aligned estimation results, which was shown in Table 3.8-3.10. Although modeling ITS is challenging, and there is no gold standard approach, our simulation results indicated that excluding phase-in period observations might be preferred as best practices for ITS analysis involving phase-in periods. After all, if a simple exclusion guarantees unbiased estimation, why bother to create complicated projections? In fact, the counterfactual from before and counterfactual from after strategies were primarily designed to visually help researchers to understand what counterfactual assumptions would be from excluding phase-in period, as indicated through projection lines and vertical bars in Figure 3.5-3.7. Researchers are also encouraged to inspect the phase-in period trajectory in a preliminary analysis. In the case of possible autocorrelation (particularly seasonality), it may be helpful to first control or remove the autocorrelation by including variables that are highly correlated with seasonality, or by using an autoregressive integrated moving average (ARIMA) model.

**Discussions**

### 4.4 Limitations of the Current Study

This study is subject to various limitations. First, autocorrelation issues are persistent in time series analyses, and the assumption of strictly linear trend with normally distributed noises in our simulation may rarely hold in practice. Second, subjects within each sample may have influences with each other, especially when the sample size is small; thus, these correlated observations would violate the independence assumption from ordinary linear regression. Third, the effect size (level and slope changes) and some parameters (e.g. sample size) were chosen discretely rather than performing numerous simulations and having the corresponding effect size on a continuous scale. Finally, only three scenarios were included in the simulation study, the impacts from some interesting factors were not fully explored, such as relative length of phase-in period, and any potential floor or celling effects within the phase-in periods.

## 4.5 Additional Concerns for Segmented Regression Analysis

Results in Table 3.8-3.10 suggested that modelling phase-in period as abrupt changes led to biased effect estimations. Moreover, from ordinary linear regression prospective, there might be additional concerns. As indicated in Figure 3.5-3.7, the abrupt models fitted linear lines to V-shaped trajectories. Consequently, model residuals were not randomly distributed as assumed. In Figure 3.8, 3.9, and 3.11, clear V-shaped residual patterns can be observed in residual plots, also with significant deviations from tails in normal Q-Q plots. This suggested that both parameter and standard error estimations from abrupt modelling should not be trusted. In contrast, Figure 3.10, 3.12 and 3.13 included residual plots from non-abrupt strategies, where the plots were less problematic.

## 4.6 Suggestions for Future Studies

The interrupted time series method could be better understood with explorations in the following problems. To begin with, how does the estimation performance vary with the relative length of phase-in period compared to total time series? For simplicity, this thesis has fixed the length of phase-in period and the overall time series. However, it would be interesting to investigate if the phase-in period last longer (e.g. more than 50% of the total collected time series), could researchers still obtain reliable estimations? Second, what are the possible techniques to handle potential floor or celling effects within the phase-in periods? This might be important practically; as an example, if researchers are measuring how vaccines reduced hospital admission rate, it is impossible to vaccinate the entire population because of resources constraints or personal preferences. In this case, how should researchers measure the full causal effect if only a fraction of population could be covered. Last but not least, how to effectively communicate the assumptions and create visualizations for ITS studies? Although one of the strengths from ITS studies is the possibility to visualize the series and observe the hypothesized causal effect, it could also be misleading. The majority of researchers used before intervention trend to extend the counterfactual projections (e.g. Bou-Antoun, et al., 2018; Sruamsiri, et al., 2016); however, there is a conceptual and estimation difference between the counterfactual from before intervention and the counterfactual from after intervention in the case of phase-in periods; the counterfactual from before assumes that the intervention was not fully implemented until the end of phase-in period, whereas the counterfactual from after projects the hypothetical case as if the intervention was implemented abruptly at the beginning of phase-in period. These calculations differ when there are slope changes, as shown in scenario-3. Therefore, although the true effect size of changes remains unknown, it is important for researchers to choose their own ways of visualizations and communicate the underlying assumptions transparently.

To summarize, multiple causal hypothesizes may exist within an interrupted time series framework, and each may correspond to different casual effect in terms of level and slope changes. With the possible existence of a delayed effect and phase-in period, researchers may apply a number of different analytical strategies to obtain effect estimations, including as abrupt at start, as abrupt at end, excluding phase-in period –before, excluding phase-in period –after, counterfactual from before and counterfactual from after. Researchers should also be aware of any counterfactual assumptions each analytical strategy assumes, and which causal effect it corresponds to.

## Bibliography

Biglan, A., Metzler, C. W., & Ary, D. V. (1994). Increasing the prevalence of successful children: The case for community intervention research. *The Behavior Analyst, 17*(2), 335-351.

Biglan, A., Ary, D., & Wagenaar, A. C. (2000). The value of interrupted time-series experiments for community intervention research. *Prevention Science, 1*(1), 31-49.

Bloom, H. S., & Ladd, H. F. (1982). Property tax revaluation and tax levy growth. *Journal of Urban Economics, 11*(1), 73-84.

Bloom, H. S. (2003). Using "short" interrupted time-series analysis to measure the impacts of whole-school reforms: With applications to a study of accelerated schools. *Evaluation Review, 27*(1), 3-49.

Bou-Antoun, S., Costelloe, C., Honeyford, K., Mazidi, M., Hayhoe, B. W. J., Holmes, A., Johnson, A. P., & Aylin, P. (2018). Age-related decline in antibiotic prescribing for uncomplicated respiratory tract infections in primary care in England following the introduction of a national financial incentive (the quality premium) for health commissioners to reduce use of antibiotics in the community: An interrupted time series analysis. *Journal of Antimicrobial Chemotherapy, 73*(10), 2883-2892.

Brockwell, P. J., & Davis, R. A. (2009). *Time Series: Theory and Methods*. Springer.

Brunette, D. (1995). Natural disasters and commercial real estate returns. *Real Estate Finance, 11*(4), 67-72.

Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine, 25*(24), 4279-4292.

Campbell, D. T., & Stanley, J. (1963). *Experimental and quasi-experimental designs for research.* Chicago, IL: Rand McNally.

Carrington, P. J., & Moyer, S. (1994). Gun availability and suicide in Canada: Testing the displacement hypothesis. *Studies on Crime and Crime Prevention, 3*, 168-178.

Catalano, R., & Senmer, S. (1987). Time series designs of potential interest to epidemiologists. *American Journal of Epidemiology, 126*(4), 724-731.

Collins, J., Hall, R. I., & Paul, L. (2004). *Causation and counterfactuals: History, problems, and prospects.* MIT Press.

Derde, L., Cooper, B. S., Goossens, H., Malhotra-Kumar, S., Willems, R., Gniadkowski, M., Hryniewicz, W., Empel, J., Dautzenberg, M, Annane, D., Aragão, I., Chalfine, A., Dumpis, U., Esteves, F., Giamarellou, H., Muzlovic, I., Nardi, G., Petrikkos, G. L., Tomic, V., . . . MOSAR WP3 Study Team. (2014). Interventions to reduce colonisation and transmission of antimicrobial-resistant bacteria in intensive care units: An interrupted time series study and cluster randomised trial. *The Lancet Infectious Diseases, 14*(1), 31-39.

Einav, L., Jenkins, M., & Levin, J. (2013). The impact of credit scoring on consumer lending. *The Rand Journal of Economics, 44*(2), 249-274

Everitt, D. E., Soumerai, S. B., Avorn, J., Klapholz, H., & Wessels, M. (1990). Changing surgical antimicrobial prophylaxis practices through education targeted at senior department leaders. *Infectious Control and Hospital Epidemiology, 11*(11), 578-583.

Fan, R., Varol, O., Varamesh, A., Barron, A., van de Leemput, Ingrid A, Scheffer, M., & Bollen, J. (2019). The minute-scale dynamics of online emotions reveal the effects of affect labeling. *Nature Human Behaviour, 3*(1), 92.

Garabedian, L. F., Ross-Degnan, D., Ratanawijitrasin, S., Stephens, P., & Wagner, A. K. (2012). Impact of universal health insurance coverage in Thailand on sales and market share of medicines for non-communicable diseases: An interrupted time series study. *BMJ Open, 2*(6), e001686.

Hallberg, K., Williams, R., Swanlund, A., & Eno, J. (2018). Short comparative interrupted time series using aggregate school-level data in education research. *Educational Researcher, 47*(5), 295-306.

Hamadani, J. D., Hasan, M. I., Baldi, A. J., Hossain, S. J., Shiraji, S., Bhuiyan, M. S. A., Mehrin, S. F., Fisher, J., Tofail, F., Tipu, S M Mulk Uddin, Grantham-McGregor, S., Biggs, B., Braat, S., & Pasricha, S. (2020). Immediate impact of stay-at-home orders to control COVID-19 transmission on socioeconomic conditions, food insecurity, mental health, and intimate partner violence in Bangladeshi women and their families: An interrupted time series. *The Lancet Global Health, 8*(11), e1380-e1389.

Hankin, J. R., Sloan, J. J., Firestone, I. J., Ager, J. W., Sokol, R. J., & Martier, S. S. (1993). A time series analysis of the impact of the alcohol warning label on antenatal drinking. *Alcoholism, Clinical and Experimental Research, 17*(2), 284-289.

Hawley, S., Ali, M. S., Berencsi, K., Judge, A., & Prieto-Alhambra, D. (2019). Sample size and power considerations for ordinary least squares interrupted time series analysis: A simulation study. *Clinical Epidemiology, 11*, 197-205.

Hawton, K., Bergen, H., Simkin, S., Dodd, S., Pocock, P., Bernal, W., Gunnell, D., & Kapur, N. (2013). Long term effect of reduced pack sizes of paracetamol on poisoning deaths and liver transplant activity in England and Wales: Interrupted time series analyses. *BMJ (Online), 346*(feb07 1), f403.

Bernal, L., J. A., Lu, C. Y., Gasparrini, A., Cummins, S., Wharam, J. F., & Soumerai, S. B. (2017). Association between the 2012 health and social care act and specialist visits and hospitalisations in England: A controlled interrupted time series analysis. *PLoS Medicine, 14*(11), e1002427-e1002427.

Lu, C. Y., Zhang, F., Lakoma, M. D., Madden, J. M., Rusinak, D., Penfold, R. B., Simon, G., Ahmedani, B. K., Clarke, G., Hunkeler, E. M., Waitzfelder, B., Owen-Smith, A., Raebel, M. A., Rossom, R., Coleman, K. J., Copeland, L. A., & Soumerai, S. B. (2014). Changes in antidepressant use by young people and suicidal behavior after FDA warnings and media coverage: Quasi-experimental study. *BMJ: British Medical Journal*, *348*(Jun18 24), g3596-g3596.

Hayes, S. C., Barlow, D. H., & Nelson-Gray, R. O. (1999). *The scientist practitioner: Research and accountability in the age of managed care.* Allyn and Bacon.

Hudson, J., Fielding, S., & Ramsay, C. R. (2019). Methodology and reporting characteristics of studies using interrupted time series design in healthcare. *BMC Medical Research Methodology, 19*(1), 137-137.

Jandoc, R., Burden, A. M., Mamdani, M., Lévesque, L. E., & Cadarette, S. M. (2015). Interrupted time series analysis in drug utilization research is increasing: Systematic review and recommendations. *Journal of Clinical Epidemiology, 68*(8), 950–956.

Johnson, M., Yazdi, K., & Gelb, B. D. (1993). Attorney advertising and changes in the demand for wills. *Journal of Advertising, 22*(1), 35-45.

Kratochwill, T. R. (2013). *Single subject research: Strategies for evaluating change*. Academic Press.

Leopold, C., Zhang, F., Mantel-Teeuwisse, A. K., Vogler, S., Valkova, S., Ross-Degnan, D., & Wagner, A. K. (2014). Impact of pharmaceutical policy interventions on utilization of antipsychotic medicines in Finland and Portugal in times of economic recession: Interrupted time series analyses. *International Journal for Equity in Health, 13*(1), 53-53.

Leske, S., Kõlves, K., Crompton, D., Arensman, E., & de Leo, D. (2021). Real-time suicide mortality data from police reports in Queensland, Australia, during the COVID-19 pandemic: An interrupted time-series analysis. *The Lancet. Psychiatry, 8*(1), 58-63.

Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference: Methods and principles for social research.* Cambridge University Press.

Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine, 38*(11), 2074-2102.

Murdoch, J. C., Singh, H., & Thayer, M. (1993). The impact of natural hazards on housing values: The Loma Prieta earthquake. *Real Estate Economics, 21*(2), 167-184.

Nelson, B. K. (1998). Time series analysis using autoregressive integrated moving average (ARIMA) models. *Academic Emergency Medicine, 5*(7), 739-744.

O'Carroll, P. W., Loftin, C., Waller, J. B., McDowall, D., Bukoff, A., Scott, R. O., Mercy, J. A., & Wiersema, B. (1991). Preventing homicide: An evaluation of the efficacy of a Detroit gun ordinance. *American Journal of Public Health, 81*(5), 576-581.

Pirkis, J., John, A., Shin, S., DelPozo-Banos, M., Arya, V., Analuisa-Aguilar, P., Appleby, L., Arensman, E., Bantjes, J., Baran, A., Bertolote, J. M., Borges, G., Brečić, P., Caine, E., Castelpietra, G., Chang, S., Colchester, D., Crompton, D., Curkovic, M., . . . Spittal, M. J. (2021). Suicide trends in the early months of the COVID-19 pandemic: An interrupted time-series analysis of preliminary data from 21 countries. *The Lancet. Psychiatry, 8*(7), 579-588.

Prais, G.J., & Winsten, C.B. (1954). Trend estimators and serial correlation. *Cowles Commission discussion paper Stat No. 383*, Chicago.

Scortichini, M., Schneider dos Santos, R., De' Donato, F., De Sario, M., Michelozzi, P., Davoli, M., Masselot, P., Sera, F., & Gasparrini, A. (2021). Excess mortality during the COVID-19 outbreak in Italy: A two-stage interrupted time-series analysis. *International Journal of Epidemiology, 49*(6), 1909-1917.

Seamon, F., & Feiock, R. C. (1995). Political participation and city county consolidation: Jacksonville-Duval County. *International Journal of Public Administration, 18*(11), 1741-1752.

Serumaga, B., Ross-Degnan, D., Avery, A. J., Elliott, R. A., Majumdar, S. R., Zhang, F., & Soumerai, S. B. (2011). Effect of pay for performance on the management and outcomes of hypertension in the United Kingdom: Interrupted time series study. *BMJ, 342*(7792)

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference.* Houghton, Mifflin and Company.

Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology.* Basic Books.

Somers, M., Zhu, P., Jacob, R. & Bloom, H. (2013). *The validity and precision of the comparative interrupted time series design and the difference-in-difference design in educational evaluation*. MDRC.

Sruamsiri, R., Wagner, A. K., Ross-Degnan, D., Lu, C. Y., Dhippayom, T., Ngorsuraches, S., & Chaiyakunapruk, N. (2016). Expanding access to high-cost medicines through the E2 access program in Thailand: Effects on utilisation, health outcomes and cost using an interrupted time-series analysis. *BMJ Open, 6*(3), e008671-e008671.

Stanley, W. D. (1987). Economic migrants or refugees from violence? A time series analysis of Salvadoran migration to the United States. *Latin American Law Review, 22*(1), 132-154.

Teague, M. L., Bernardo, D. J., & Mapp, H. P. (1995). Farm-level economic analysis incorporating stochastic environmental risk assessment. *American Journal of Agricultural Economics, 77*(1), 8-19.

Tesoriero, J. M., Sorin, M. D., Burrows, K. A., & LaChance-McCullough, M. L. (1995). Harnessing the heightened public awareness of celebrity HIV disclosures: "Magic" and "Cookie " Johnson and HIV testing. *AIDS Education and Prevention, 7*(3), 232-250.

Tilden, V. P., & Shepherd, P. (1987). Increasing the rate of identification of battered women in an emergency department: Use of a nursing protocol. *Research in Nursing and Health, 10*(4), 209-215.

Turner, S. L., Karahalios, A., Forbes, A. B., Taljaard, M., Grimshaw, J. M., Cheng, A. C., Bero, L., & McKenzie, J. E. (2020). Design characteristics and statistical methods used in interrupted time series studies evaluating public health interventions: A review. *Journal of Clinical Epidemiology, 122*, 1–11.

Turner, S. L., Forbes, A. B., Karahalios, A., Taljaard, M., & McKenzie, J. E. (2021). Evaluation of statistical methods used in the analysis of interrupted time series studies: A simulation study. *BMC Medical Research Methodology, 21*(1), 1-181.

Velicer, W. F. (1994). Time series models of individual substance abusers. *NIDA Research Monograph, 142*, 264.

Vokó, Z., & Pitter, J. G. (2020). The effect of social distance measures on COVID-19 epidemics in Europe: An interrupted time series analysis. *Geroscience, 42*(4), 1075-1082.

Wagner, A. K., Soumerai, S. B., Zhang, F., & Ross-Degnan, D. (2002). Segmented regression analysis of interrupted time series studies in medication use research. *Journal of Clinical Pharmacy and Therapeutics, 27*(4), 299–309.

Walley, A. Y., Xuan, Z., Hackman, H. H., Quinn, E., Doe-Simkins, M., Sorensen-Alawad, A., Ruiz, S., & Ozonoff, A. (2013). Opioid overdose rates and implementation of overdose education and nasal naloxone distribution in Massachusetts: Interrupted time series analysis. *BMJ (Online), 346*(jan30 5), f174.

Windsor, R. A. (1986). The utility of time series designs and analysis in evaluating health promotion and education programs. *Advances in Health Education and Promotion*, *1*, 435-465.