2023-08

Data Subset-Based Methods of Inference for Spatial Individual Level Epidemic Models

Nyein, Thet Htet Chan

UNIVERSITY OF CALGARY

Data Subset-Based Methods of Inference for Spatial Individual Level Epidemic Models

by

Thet Htet Chan Nyein

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN MATHEMATICS AND STATISTICS

CALGARY, ALBERTA

AUGUST, 2023

Abstract

Mathematical models are essential to understand infectious disease dynamics, enabling to control the spread of those diseases and preparing for public health measures. Since time and space are important factors affecting the transmission of infectious diseases, spatial individual-level models (ILM) with both temporal and spatial information are developed. Typically, Markov Chain Monte Carlo (MCMC) methods are utilized for the inference of ILM. Nonetheless, this approach can be computationally intensive for complex or large models, resulting in repeated likelihood calculations. This thesis explores various spatial and temporal subset methods to conduct statistical inference for spatial epidemic models, aiming to provide appropriate parameter estimates with minimum computational resources. In this thesis, we utilize the spatial ILM with the Euclidean distance between susceptible individuals and infectious individuals as a kernel function.

Preface

This thesis is an original work by Thet Htet Chan Nyein. No part of this thesis has been previously published.

Acknowledgements

I would like to extend my gratitude to my supervisor, Dr.Rob Deardon for his support, guidance, and mentorship throughout my Master's degree. I would also like to express my appreciation to my family for their financial assistance and emotional encouragement, throughout my undergraduate and graduate studies.

Table of Contents

Chapter 1

Introduction

Mathematical modelling is necessary to understand more about the dynamics of infectious diseases. With that information, appropriate public health measures can be implemented to control the spread of those diseases as well as prepare health systems with the necessary resources to treat the projected number of patients.

Mathematical models are useful before, during, and after a pandemic. Models are used for planning, identifying critical gaps, and creating plans to detect and respond to a pandemic before it occurs. Policy-makers are interested in learning the following information at the beginning of a pandemic: (i) where and how the pandemic originated; (ii) the likelihood that it will spread within the area; (iii) the likelihood that it will be imported into other parts of the world; and (iv) a basic understanding of the pathogen and its epidemiological traits. Researchers start looking into the following as the pandemic spreads: (i) different intervention and control strategies; non-pharmaceutical interventions are typically the most effective in the event of a pandemic, (ii) forecasting the epidemic incidence rate, hospitalization rate, and mortality rate; (iii) effectively allocating limited medical resources to treat the patients; and (iv) understanding the change in individual and collective behaviour. Modellers are interested in creating simulations of recovery and the pandemic's long-term effects after it begins to slow down [1].

Two key ideas that can affect the transmission of disease are time and space. As an epidemic's severity grows over time, it frequently spreads in space. The intensity of the disease may not be uniform throughout a community as a result of its distribution. Incorporating place and time's effects into a model is critical because they both hold crucial information about how a disease spreads[8]. Hence, in this research, we study individual-level models (ILM) which allows for modelling epidemics of infectious diseases through time and space at the individual level in the population [5].

In ILM, the probability of a susceptible individual being infected from the infectious pressure exerted upon it from the surrounding infectious population is quantified[5]. The ILM framework assumes that individuals in the population go through discrete time points in time and space, and those individuals are often positioned in a set of square or rectangular areas [19]. It allows us to express the probability of a susceptible individual becoming infected at a point in discrete time as a function of their interactions with the surrounding infectious population [11]. ILMs use the Bayesian MCMC framework which is known for having computational difficulties in computing the full likelihood [5]. Because of that, ILMs are computationally expensive in finding parameter estimates, especially for a large number of individuals.

There are several types of ILMs depending on different compartmental models. In compartmental models, individuals within a closed population are divided into compartments, or mutually exclusive groups according to their disease status. Each person is assumed to be in one compartment at a time, however, they are free to switch between compartments based on the model's characteristics [18]. In this thesis, we focus mainly on the SIR version of spatial ILM, which has susceptible (S), infectious (I) and removed (R) states, and uses a geometric kernel.

Typically, MCMC methods are used for inference in ILM models. However, for complex models or large populations, the MCMC approach can become computationally expensive due to repeated likelihood calculations [20]. To address this, we investigate data subset-based inference methods for spatial epidemic models that can provide appropriate parameter estimates with minimum computational resources. Given the spatial and temporal aspects of epidemic data in ILMs, we explore various spatial and temporal subset methods to conduct statistical inference for infectious disease models with minimal computational resources. This research focuses on the spatial ILM that uses the Euclidean distance between susceptible individuals and infectious individuals as a kernel function.

The structure of the thesis is as follows. In Chapter 2, we introduce the general form of ILM, explore the dynamics of the spatial epidemic model, and introduce concepts such as Bayesian inference, Markov Chain Monte Carlo and its properties. Additionally, it discusses general simulation methods and defines the absolute biases in the spatial ILM. Furthermore, it presents a measure of computational power and defines the spatial and temporal subset methods. In Chapter 3, we summarize the outputs from the simulations. In Chapter 4, we analyze the results from simulations, reached the conclusion of the "best" subset methods, and finally discuss how those subset methods can be refined for implementation in real-life epidemics.

Chapter 2

Methods

2.1 The ILM Epidemic Model

The general form of the ILM epidemic model is presented in this section. In this research, we use the SIR version of the general ILM. This assumes that at any given point in time t, either individual i is susceptible to disease (S), individual i is infected with the disease (I), or individual i is recovered from the disease (R).Individuals move through the states, $S \rightarrow I \rightarrow R$. For each time point, $t = 1, 2, ..., \infty$, a susceptible or infected is removed. individual i is said to be in the set $S(t)$, $I(t)$, or $R(t)$ respectively.[5]

Let $P(i,t)$ be the probability that a susceptible individual i is infected in the continuous interval $[t, t+1)$ under the ILM framework it is given by

$$P(i,t) = 1 - \exp[-\Omega_S(i) \sum_{j \in I(t)} \Omega_T(j)\kappa(i,j) + \epsilon(i,t)], \epsilon(i,t) < 0 \tag{2.1}$$

where, $I(t)$ is the set of infected and infectious individuals at time t, $\Omega_S(i)$ is a function of risk factors associated with a susceptible individual i acquiring the disease (susceptibility), $\Omega_T(j)$ is a function of risk factor associated with an infectious individual j (transmissibility), $\kappa(i,j)$ is an infection kernel that contains risk factors associated with pairs of susceptible and infectious individuals. For example, $\kappa(i,j)$ could be a function of the spatial proximity between a susceptible individual i and an infectious individual j. $\epsilon(i,t)$ represents some infection process otherwise unexplained by the model.It is often set to be zero, or assumed

to be constant over time and individuals. The likelihood over discrete time points $t = 1, ..., T$ is given by:

$$\begin{aligned} f(D \mid \theta) &= \prod_{i=1}^{T-1} f_t(S(t), I(t), R(t)) \mid \theta) \\ &= \prod_{i \in I(t+1) \setminus I(t)} P(i,t) \prod_{i \in S(t+1)} [1 - P(i,t)] \end{aligned} \tag{2.2}$$

where $i \in I(t+1) \setminus I(t)$ is the set of infectious individuals who are newly infected in time $t+1$, and $i \in S(t+1)$ is the set of susceptible individuals at time $t+1$.

2.2 Spatial ILM Epidemic Model

In this research, the following spatial ILM epidemic model is considered. Let $\Omega_T(j) = 1$, $\Omega_S(i) = \alpha$ and $\epsilon(i,t) = 0$. We then let $k(i,j) = d_{ij}^{-\beta}$, resulting in a power-law spatial ILM,

$$P(i,t) = 1 - \exp[-\alpha \sum_{j \in I(t)} d_{ij}^{-\beta}] \qquad \alpha, \beta > 0 \tag{2.3}$$

where d_{ij} is the Euclidean distance between a susceptible individual i and an infectious individual j, α is the susceptibility parameter and β is the spatial parameter.

2.3 Simulation of Epidemic Data

The SIR epidemic data is generated from the spatial model described in equation 2.3 by using the EpiILM package in R. In order to keep the population density consistent across different population sizes or n, individuals are uniformly positioned on the square area $\sqrt{n} \times \sqrt{n}$. Next, parameters $\alpha = 0.2, \beta = 2$ are assigned to the SIR epidemic model with the population sizes of $n = 100$, $n = 1000$ or $n = 5000$.

The SIR epidemic data contains information on not only the locations of the individuals but also infection time for each individual. The maximum infection time for this simulation is assigned as 15. In addition, we set the infection period for each individual as 3. Figure 2.1 represents that epidemic that contains 100 individuals along with their infection time.

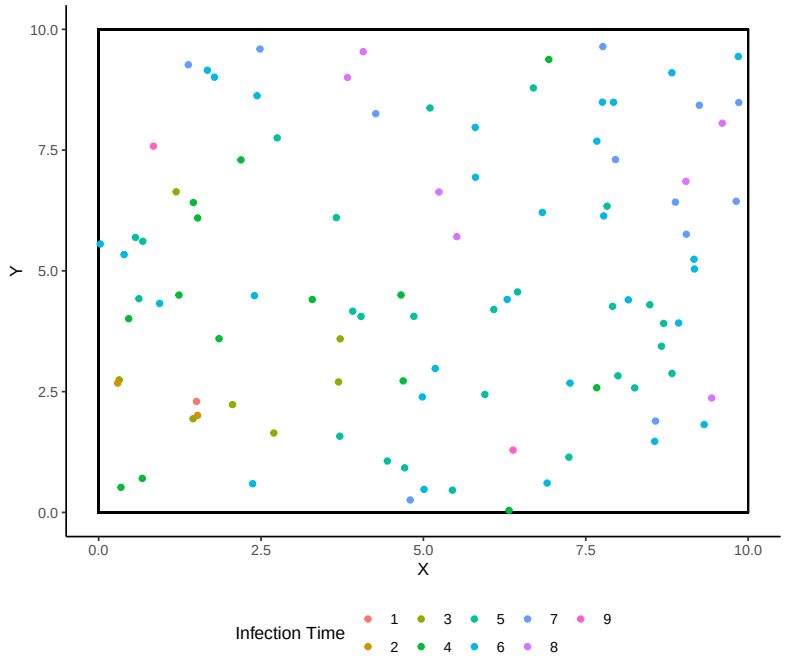


Figure 2.1: Example of SIR epidemic data with population size $n = 100$

2.4 Bayesian Inference

In this research, we use a Bayesian Markov Chain Monte Carlo (MCMC) framework. Specifically, Metropolis-Hastings Random Walk MCMC is used to sample from the posterior distribution to conduct inference for the parameters α and β.

Before moving on to the Bayesian MCMC framework, we will explore more about Bayesian inference. This inference is conducted by using the posterior distribution, which is composed of prior distribution and likelihood.

Let us assume an unidentified parameter θ is a random variable with proper prior distribution $f(\theta)$. It frequently happens that the researcher will be familiar with the likely value of θ before seeing the data [8]. Likelihood $l(\theta)$ or $f(D \mid \theta)$ represents the probability of observing the data D given the parameter θ. After observing the data (D), the posterior distribution, $f(\theta \mid D)$ is calculated by using Bayes's theorem. The posterior distribution can be written as

$$f(\theta \mid D) = \frac{f(D \mid \theta) f(\theta)}{\int f(D \mid \theta) f(\theta) d\theta} \tag{2.4}$$

The denominator in 2.4 represents the marginal probability of data or the normalization constant. The normalisation constant is used to guarantee that $f(\theta \mid D)$ is a probability density function. However, this integral is difficult to evaluate and obtain in the closed form. Hence, we rewrite the posterior distribution

up to proportionality:

$$f(\theta \mid D) \propto f(D \mid \theta) f(\theta) \tag{2.5}$$

Even in the proportionality format described in eq:2.5, we still need to find integrals for posterior mean and standard deviation for θ. When we are dealing with high-dimensional parameters, the integral can be challenging to execute. Because of this, Monte Carlo techniques are frequently employed as an effective means of obtaining an approximate result by sampling from the posterior distribution.

2.5 Markov Chain Monte Carlo

In this section, we introduce important ideas behind Markov Chain Monte Carlo (MCMC).

2.5.1 Markov Chain

In this thesis, we focus on discrete-time Markov chains. A collection of random variables, typically indexed by time, makes up a random process. Let the discrete-time random process X be a sequence of random variables $X = \{X_n, n \in N\}$ over the finite state space S. X is a discrete-time Markov chain if

- for every $n > 0$, $X_n \in S$,
- for every $n \in N$, for all $i_1, i_2, ..., i_t \in S$

$$P(X_t = i_t \mid X_{t-1} = i_{t-1}, ..., X_0 = i_0) = P(X_t = i_t \mid X_{t-1} = i_{t-1}) \tag{2.6}$$

when both conditional probabilities are defined.

The likelihood that the chain will move from location X_k to position X_{k+1} depends only on the transition probability matrix, $P_{i,j}$. For all $i, j \in S$, $P_{i,j}$ can be defined by

$$P_{i,j} = P\{X_n = j \mid X_{n-1} = i\}, \quad \forall n \geq 1 \tag{2.7}$$

A matrix having

$$P_{i,j} \geq 0, \quad \forall i, j \in S, \tag{2.8}$$

$$\sum_{j \in S} P_{i,j} = 1 \quad \forall i \in S \tag{2.9}$$

is called a stochastic matrix [17]. The discrete-time Markov chain described in this thesis is homogeneous since the probability of changing from one state to another depends only on the present state and not on any earlier states. Additionally, this Markov chain is memoryless in that its future evolution depends solely on its current state and not on any previous history. The transition probabilities are constant and do not change over time.

2.5.2 Stationarity

A probability distribution π which satisfies

$$\pi^T = \pi^T P \tag{2.10}$$

is referred to as a stationary distribution of the transition matrix P or that of the associated homogeneous Markov chain [2].

In order to reach stationarity, the Markov chain needs to be irreducible, aperiodic and positive recurrent.

Irreducibility

If and only if $x \to y$ and $y \to x$, or $x = y$, we say that x communicates with y, and we denote it as $x \leftrightarrow y$. The communication classes are the equivalence classes listed under $\leftrightarrow$. The communication class of x is indicated by [x] for $x \in X$. The Markov chain X is irreducible if there is only one communication class [2], which means that each state can be reached from any other state by utilising only transitions with a positive probability [10].

Aperiodicity

Let $T(X)$ be the set of times when it is possible for the chain to return to its initial position x. $T(X)$ can be denoted by $T(X) := \{t \geq 1 : P_t(x, x) > 0\}$.The greatest common divisor (GCD) of $T(X)$ is defined as the period of state x. A state x is said to be aperiodic if $gcd[T(X)] = 1$. A Markov chain is said to be aperiodic when all states have period 1 [10].

Positive Recurrence

A state j is said to be recurrent if and only if

$$\sum_{n=1}^{\infty} P_{jj}^n = \infty \tag{2.11}$$

Let μ_{jj} be the expected number of transitions needed to return to state j. Then μ_{jj} can be denoted as

$$\mu_{jj} = \sum_{n=1}^{\infty} f_{jj}^{n} \tag{2.12}$$

where f_{jj}^{n} is the probability that starting in state j the first transition into j occurs at time n.A state j is said to be positive recurrent if $\mu_{jj} < \infty$ [15].

2.5.3 Monte Carlo Markov Chain

The majority of problems involving Bayesian inference can be expressed as a function of interest, $g(\theta)$ evaluated over the posterior distribution [14], and hence, the expectation can be written as:

$$E[g(\theta) \mid D] = \int f(\theta \mid D) g(\theta) d\theta \tag{2.13}$$

where E represents the expectation.

Sampling from high-dimensional distributions can be challenging, which is why Markov chain Monte Carlo (MCMC) is often used to find $E[g(\theta) \mid D]$. The goal of MCMC is to create a random walk or Markov process that has $f(\theta \mid D)$ as its stationary distribution, and then run the process long enough so that the resulting sample closely resembles a sample from $f(\theta \mid D)$, ensuring the independence of the samples [3]. These samples can be used directly for parameter inference and prediction. For example, to determine $E[g(\theta) \mid D]$, Monte Carlo integration involves selecting n samples from $f(\theta \mid D)$, denoted θ^i $(i = 1, ..., n)$, and computing the mean as follows [14].

$$E[g(\theta) \mid D] \approx \frac{1}{n} \sum_{i=1}^{n} g(\theta^i) \tag{2.14}$$

The law of large numbers ensure that when the samples θ^i are independent, the approximation can be made as accurate as desired by increasing the sample size n.

Due to the memoryless property of the Markov chain, the chain will eventually forget its initial state and will converge to a unique stationary distribution. As time increases, θ^i will be similar to dependent samples from $f(\theta \mid D)$ after a sufficient long burn-in. When burn-in samples are discarded, an estimator $E[g(\theta) \mid D]$ is calculated by

$$\bar{g} = \frac{1}{n-m} \sum_{i=m+1}^{n} g(\theta^i) \tag{2.15}$$

where, m = number of burn-in iterations, and n = number of total iterations, and $\bar{g}$ is called an ergodic

average [7]. In this research, we use $n = 50000$ and $m = 10000$.

2.5.4 Metropolis-Hastings Algorithm

The Metropolis Algorithm is first created by Metropolis for producing random samples from a probability distribution by using a Markov chain. Later, Hastings Algorithm, which can produce samples from a larger range of probability distributions, was created to be more versatile than Metropolis' algorithm.

The steps for the MH algorithm are as follows.

1. Choose an arbiratiry θ^0 as a starting value.
2. Create a candidate value, θ^{new}, using proposal density $q(\theta^{new} \mid \theta_i)$ given the current location $\theta^{(i)} = \theta_i$.
3. Calculate acceptance probability, $\alpha(\theta^{new} \mid \theta_i)$.

$$\alpha(\theta^{new} \mid \theta_i) = min\left(1, \frac{\pi(\theta^{new})q(\theta_i \mid \theta^{new})}{\pi(\theta_i)q(\theta^{new} \mid \theta_i)}\right) \tag{2.16}$$

   where $\pi(.)$ is the target distribution.

4. Using a uniform distribution between 0 and 1, generate a random number u. If $u \leq \alpha(\theta^{new} \mid \theta_i)$, set $\theta_{i+1} = \theta^{new}$,otherwise, $\theta_{i+1} = \theta_i$.
5. Until a predetermined number of MCMC iterations have been completed or a stopping criterion has been met, go back to step 2.

2.5.5 Metropolis-Hastings Random Walk Algorithm

For the Metropolis-Hastings Random Walk Algorithm, we incorporate the random walk proposal into the proposal density, $q(\theta^{new} \mid \theta_i) = f(\mid \theta^{new} - \theta_i \mid)$. Hence, the proposed observation is obtained by adding the random displacement to the current state θ_i, and becomes,

$$\theta^{new} = \theta_i + z, \tag{2.17}$$

where $z \sim f$ and f is symmetric around zero.

There are many choices of distribution for f. Some of the popular distributions for f are the uniform distribution on the unit disk, a univaraiate or multivariate normal and student's t-distributions. In our research, we will use the multivariate normal distribution as f. Additionally, due to symmetry,

$$q(\theta_i \mid \theta^{new}) = q(\theta^{new} \mid \theta_i) \tag{2.18}$$

Therefore, the acceptance probability becomes

$$\alpha(\theta^{new} \mid \theta_i) = min\left(1, \frac{\pi(\theta^{new})}{\pi(\theta_i)}\right) \tag{2.19}$$

where θ_i is the current value.

2.5.6 Convergence Diagnostic

In order to ensure accurate results, it is crucial to determine whether the Markov chain has converged to the stationary distribution, which, in our case, is the posterior distribution. MCMC diagnostic tools are needed to check for convergence. While running the chain for a longer period generally leads to better Monte Carlo estimates, it is advisable to have stopping rules in place to use resources efficiently [16].

There are several methods to assess convergence, such as trace plots, Gelman-Rubin diagnostic, Honest MCMC, and so on. In our study, we used Geweke's diagnostic test, which evaluates the equality of the first and last parts of the Markov chain's means [6]. If the samples are drawn from the stationary distribution of the chain, the two means are equal and Geweke's statistic has an asymptotically standard normal distribution [13]. The null and alternative hypothesis of Geweke's diagnostic test is as follows:

H_0 : The means of the first and last parts are the same. Convergence has been reached.

H_a :The means of the first and last parts are different. Convergence has not been reached.

In our research, from Metropolis-Hastings random walk algorithm, we obtain the estimator of $g(\theta)$ or $E[g(\theta)]$ via

$$\bar{g}(\theta) = \frac{\sum_{i=1}^{n} g(\theta^i)}{n} \tag{2.20}$$

The z-score, which is calculated as follows, is the test statistic used in Geweke's diagnostic test.

$$z = \frac{\bar{g}(\theta)_n^A - \bar{g}(\theta)_n^B}{\sqrt{\frac{v(g(\theta)_n^A)}{n^A} + \frac{v(g(\theta)_n^B)}{n_B}}} \tag{2.21}$$

where, $\bar{g}(\theta)_n^A$ and $\bar{g}(\theta)_n^B$ are the means based on the first n_A and n_B iterations,$v(g(\theta)_n^A)$ and $v(g(\theta)_n^B)$ are the variances of those two disjoint segments [4]. In this thesis, we implement Geweke's diagnostic test by computing the first 10% after burn-in and the last part 25% of the MCMC chain. This test is evaluated in R using the CODA package.

2.6 Simulation Study

We implement the Metropolis-Hastings Random Walk Algorithm to sample from the posterior distribution for fitting our ILMs to simulated data. The procedure is as follows.

1. Establish the likelihood function $f(D \mid \theta)$. We use the likelihood function from eq:2.2.
2. Specify the prior distribution, which in this research, we use independent gamma distribution as prior distributions for α and β.
3. Define the proposal distribution. In this research, we use a multivariate normal distribution with correlation ϕ, and so have two variance parameters to tune.
4. Calculate the posterior distribution up to proportionality as follows.

$$f(\theta \mid D) \propto f(D \mid \theta) f(\theta) \tag{2.22}$$

5. Generate a candidate value, θ^{new},from the proposal distribution.
6. Calculate the acceptance probability $\alpha(\theta^{new} \mid \theta_i)$.

$$\alpha(\theta^{new} \mid \theta_i) = min\left(1, \frac{f(\theta^{new} \mid D)}{f(\theta^{i} \mid D)}\right) \tag{2.23}$$

7. Generate a random number u from a uniform distribution between 0 and 1. If the value of $u \leq \alpha(\theta^{new} \mid \theta_i)$, set the new value of the parameter as $\theta_{i+1} = \theta^{new}$. Otherwise, set the new value of the parameter as $\theta_{i+1} = \theta_i$.
8. Keep repeating step 5-7 until a stopping criterion has been met or the desired number of MCMC iterations has been completed.
9. Confirm the convergence of the MCMC chain by using Geweke's Convergence Diagnostic.
10. If the simulated epidemic dies out very quickly, i.e, all individuals in the population gets infected in the earlier infection times, and thus the resulting epidemic dataset is not informative enough for our ILM to be successfully fitted, the new SIR epidemic data is generated and steps 4-8 are repeated until we obtain the informative epidemic data.

2.7 Absolute Bias in ILMs

Absolute bias is used to measure the absolute difference between the parameter estimate and the true parameter. The absolute bias of a parameter θ can be defined as

$$abs[bias(\theta)] = abs[\hat{\theta} - \theta] \tag{2.24}$$

where, $\hat{\theta}$ is the parameter estimate and θ is the true parameter.

In our research, we explore the absolute biases of the posterior means ($E(.)$) and standard deviations ($SD(.)$) of the parameters α and β. The absolute biases of those parameters can be written as

$$abs[bias(\alpha)] = abs[E_{SS}(\alpha) - E(\alpha)] \tag{2.25}$$

$$abs[bias(\beta)] = abs[E_{SS}(\beta) - E(\beta)] \tag{2.26}$$

$$abs[bias(SD_\alpha)] = abs[SD_{SS}(\alpha) - SD(\alpha)] \tag{2.27}$$

$$abs[bias(SD_\beta)] = abs[SD_{SS}(\beta) - SD(\beta)] \tag{2.28}$$

where, $E_{SS}(.)$ and $SD_{SS}(.)$ are the posterior mean and standard deviation under the subsetting method, $E(.)$ and $SD(.)$ are the posterior mean and standard deviation under the full epidemic data.

2.8 Measure of Computational Power

The amount of time the processor uses to execute instructions for a particular task is known as CPU time. CPU time is composed of User and System time. User time is the amount of CPU time used by the process or task directly whereas system time is the amount of CPU time used by the operating system on behalf of the process or task.

We can estimate how much CPU power is being used by a project as a whole by calculating the total of the system and user CPU time of the job. In our research, we used the total system and user CPU per epidemic in CPU hours.

2.9 Model Prediction From Spatial Subsets

For this research, I intend to predict the posterior estimate for the full data set using models fitted to estimate from smaller through progressively larger subsets. In terms of spatial subset, it means creating smaller square(s) within the $\sqrt{n} \times \sqrt{n}$ square grid for the simulations described in the section 2.3.

To develop spatial subset methods that can produce appropriate parameters while remaining computationally efficient, it is interesting to assess the effect of changing the amount of epidemic data in a spatial subset. Consequently, we have opted to create subset squares centred within the $\sqrt{n} \times \sqrt{n}$ square grid that contains approximately $p\%$ of the original data $p = 10, 20, 30, 40, 50, 64$. Figure 2.2 demonstrates how a 50% spatial subset is created within an ILM epidemic data with 100 individuals. Next, we calculate and construct the posterior means and 95% credible intervals of α and β for each subset percentage as well as for the entire epidemic data. By doing so, we examine the relationship between the subset percentage and the accuracy of parameter estimation.

Figure 2.3 (a) and (b) show 95% credible intervals and posterior mean estimates for α and β respectively for a single typical epidemic and population. The red dotted line in Figure 2.3 (a) and (b) refer to parameters, $\alpha = 0.2$ and $\beta = 2$, that are used to generate the epidemic data. From this figure, it can be observed that as the size of the spatial subset increases, the credible intervals become narrower and their respective posterior mean estimates get closer to the true values.

Additionally, we observe exponential-like trends in the posterior mean estimates of α and β shown in Figures 2.3 (a) and (b). To better understand the trends of posterior mean estimates of α and β, we fitted a linear regression model with percent and percent-squared as predictors and a non-linear least square exponential model with ae^{percent}. We utilized the posterior mean estimates from subsets of the epidemic data as the training data for both models and predicted posterior mean estimates at 100% of the epidemic data for the test data.

The results of our models are shown in Figures 2.4 (a) and (b). The green line represents predictions from the non-linear least square exponential model, the blue line represents predictions from the ordinary least square model, and the red line represents the posterior mean estimates from different percentages of the original epidemic data. We can observe that the non-linear exponential model performs better compared to the ordinary least square model. However, the non-linear exponential model follows a similar trend as the spatial posterior mean estimates for both α and β, the difference is quite large at 100%. Therefore, we have decided not to explore exponential and ordinary least square models further.

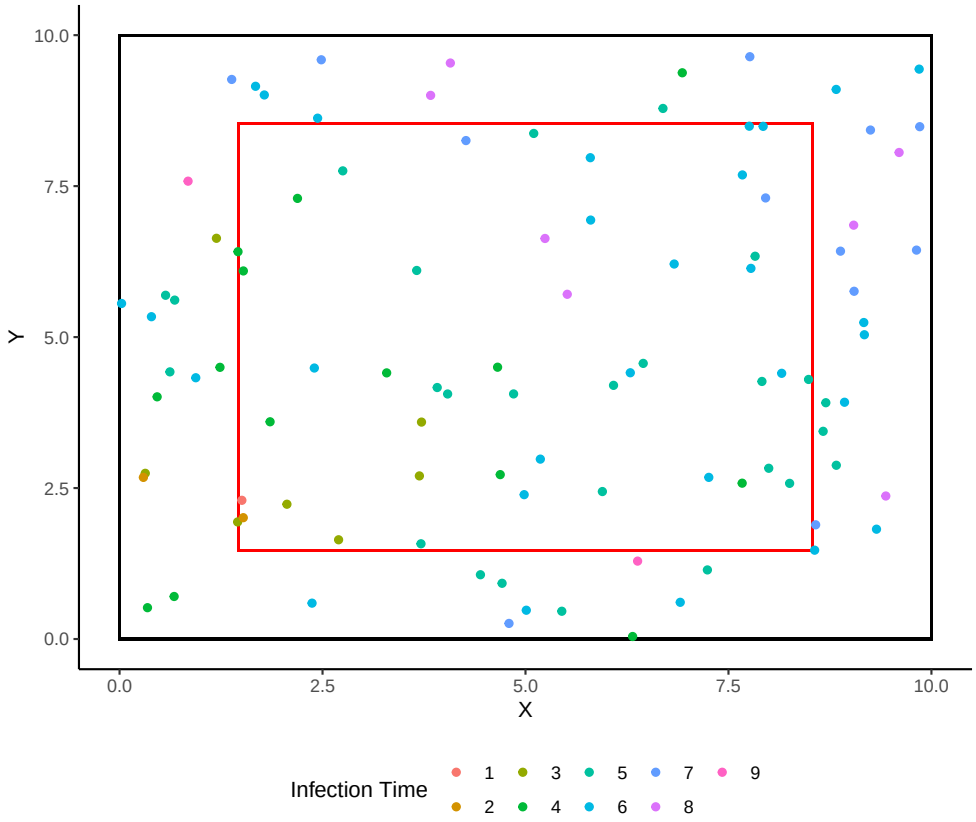
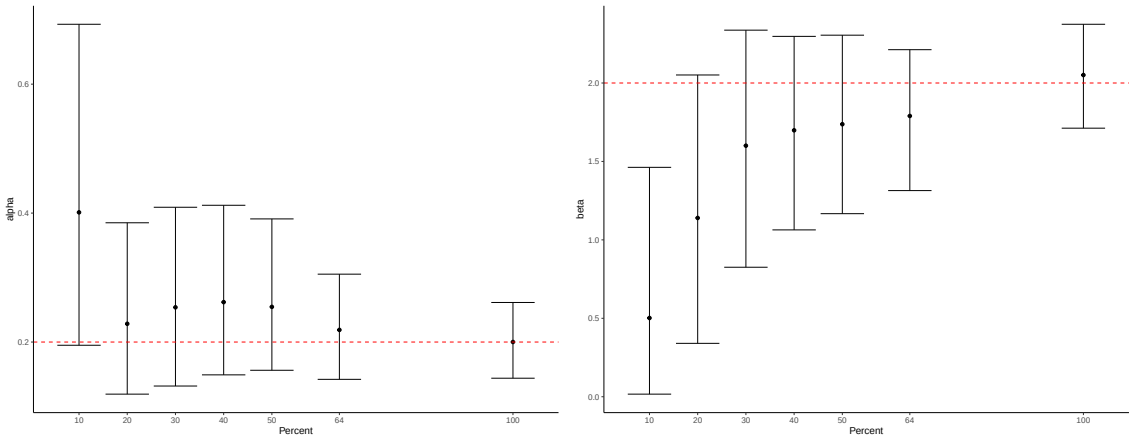
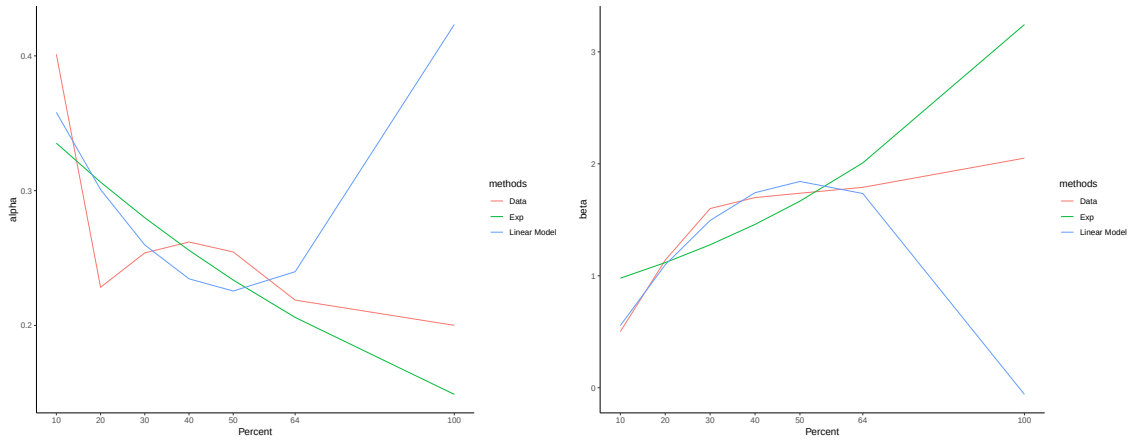


Figure 2.2: Example of spatial subset containing approximately 50% of data from SIR epidemic data with population size $n = 100$



(a) (b)

Figure 2.3: 95% Credible Intervals of (a) alpha (b) beta for spatial subsets for a single typical epidemic with $n = 100$



(a) (b)

Figure 2.4: Initial Modeling (a) alpha (b) beta parameters for a single typical epidemic with $n = 100$

2.10 Spatial Subset Methods

Figure 2.3 (a) and (b) indicates the positive association between spatial subset percentage and accuracy parameter estimation. To validate this finding for different population sizes ($n = 100, 1000, 5000$), we implemented the **k-percent center** method, which selects a square containing $k\%$ of the original epidemic data from the center of the $\sqrt{n} \times \sqrt{n}$ square where $k = 10\%, 20\%, 30\%, 40\%, 50\%, 64\%$. By computing $E_{SS}(.)$ and $SD_{SS}(.)$ using posterior means and standard deviations of the subset, we observed the relationship between population size and effective subset percentage.

To investigate the impact of the location of the spatial subset data on absolute biases, we modified the **k-percent center** method, and developed a new method herein, referred to as **k-percent ll** method. This method involves obtaining a subset from the lower left corner of the square from 0 to $\sqrt{nk}$ where $k = 10\%, 20\%, 30\%, 40\%, 50\%$ or 64%. Afterwards, we follow the same procedure as **k-percent center** method to compute $E_{SS}(.)$ and $SD_{SS}(.)$.

Next, we explore whether incorporating additional subset areas of different sizes would improve the results of inference. Hence, we expand two more methods named **k-percent corner** method,which obtains subsets from the four corners of the square, and **k-percent random** method, which takes four subsets randomly within the $\sqrt{n} \times \sqrt{n}$ square grid. We computed $E_{SS}(.)$ and $SD_{SS}(.)$ using posterior means and standard deviations of each subset.

To explore the effect of including different subset percentages on inference, we developed the **average center** method. This method selects four subsets, each containing approximately 25%, 36%, 49%, and 64% of the population from the center of the $\sqrt{n} \times \sqrt{n}$ square grid. The subsets are chosen to be perfect squares,

as the epidemics are generated in the square grid. We computed $E_{SS}(.)$ and $SD_{SS}(.)$ using posterior means and standard deviations of each subset and obtained the average of these estimates.

In addition, we modified the **average center** method and created a method, herein referred to as **average ll** method to examine the relationship between the location of the spatial subset to the absolute biases of the parameter estimates. In this method, we create smaller squares containing approximately 25%, 36%, 49%, and 64% of the population from the lower left of the $\sqrt{n} \times \sqrt{n}$ square grid, and we computed $E_{SS}(.)$ and $SD_{SS}(.)$ in the same manner as the **average center** method.

Finally, we examined whether incorporating considering different subset areas in the square grid would decrease the absolute biases of parameter estimates compared to **average center** and **average ll** methods. Therefore, we establish two methods named, **average corner** and **average random** methods. **Average corner** method obtains four subsets for each of the corners of the square, with each subset compromising approximately 25%, 36%, 49%, and 64% of the population. We then find the $E_{SS}(.)$ and $SD_{SS}(.)$ using posterior means and standard deviations of each subset, obtained the average of these estimates, and find the mean of those averages. **Average random** method is similar to **average corner** method except the subset areas are chosen randomly.

2.11 Temporal Subsetting

For temporal subsets, ILM epidemic data is placed in subsets based on infection time. In order to establish temporal subsets, we need to define the minimum and maximum infection times while implementing the Bayesian MCMC framework. For instance, in Figure 2.5, the area shaded in blue represents the temporal subset that contains a minimum infection time of 1 and a maximum infection time of 7.

To explore the effect of temporal subsetting, we first split the epidemic data into two temporal subsets: one containing infection time points from 1 to 7, and another containing infection time points from 8 to 15. We then calculate the posterior means and construct the 95% credible intervals of α and β for each of those subsets. We then proceed to follow the same procedure for the entire epidemic data that contains infection time points from 1 to 15.

Figures 2.6 (a) and 2.7 (a) display the resulting posterior mean estimates and 95% Credible Intervals of α for β for a single typical epidemic and population. The red dotted lines refer to the true susceptibility parameter $\alpha = 0.2$ in figure 2.6 (a) and the true spatial parameter $\beta = 2$ in figure 2.6 (b), which are used to generate the epidemic data. In both figures, the subset which contains the first half of the infection time points (from 1 to 7) performs almost the same as the entire epidemic data that contains all the infection time points (from 1 to 15). On the other hand, the subset data that contains the last half of the infection

time has the posterior mean and 95% Credible interval quite different from that of the entire epidemic data.

We then continue to break down more temporal subdivisions in order to precisely locate the infection times which have the most useful information for parameter prediction. Hence, we divide the epidemic data into three separate segments: the first segment contains epidemic data with infection times from 1 to 5, the second segment from 6 to 10, and the last segment from 1 to 15. The posterior means and 95% credible intervals for α and β are calculated for each segment, as well as for the entire dataset (infection times from 1 to 15).

The results for this analysis are provided in Figures 2.6 (b) and 2.7 (b). The second temporal segment consisting of the infection times from 6 to 10, produces similar posterior means and 95% credible intervals for α and β as that of the entire epidemic data. Even though the first segment (the infection time from 1 to 5) yields similar outcomes for the posterior mean and 95% credible interval for α as the entire epidemic data, the same scenario did not apply for β.

To further refine our analysis and identify the specific infection times that provide the most valuable information for parameter prediction, we proceeded to create four temporal subsets. The first subset include infection time from 1 to 3, the second from 4 to 7, the third from 8-11 and the last from 12 to 15. We then calculated the posterior means and generated their corresponding 95% credible intervals for α and β for each temporal subset. Subsequently, we repeated the same procedure using the complete epidemic data, which encompassed infection time points ranging from 1 to 15.

Figures 2.6 (c) and 2.7 (c) present the obtained posterior mean estimates and 95% credible intervals for the parameters α and β respectively. From those figures, it can be denoted that second temporal subset (infection time from 4 to 7) performs similarly as the entire epidemic data in terms of posterior means and 95% credible intervals for α and β. Aside from this subset, any other temporal subsets have far-off posterior means and 95% credible intervals for α and β than that of the entire epidemic data.

Based on the iteration, it is evident that the temporal subset comprising infection time from 4 to 7 yields the most accurate inference of α and β when compared to other temporal subsets. However, it is important to ascertain whether this observation is specific to this particular simulation or holds true for a range of different epidemics and population sizes.

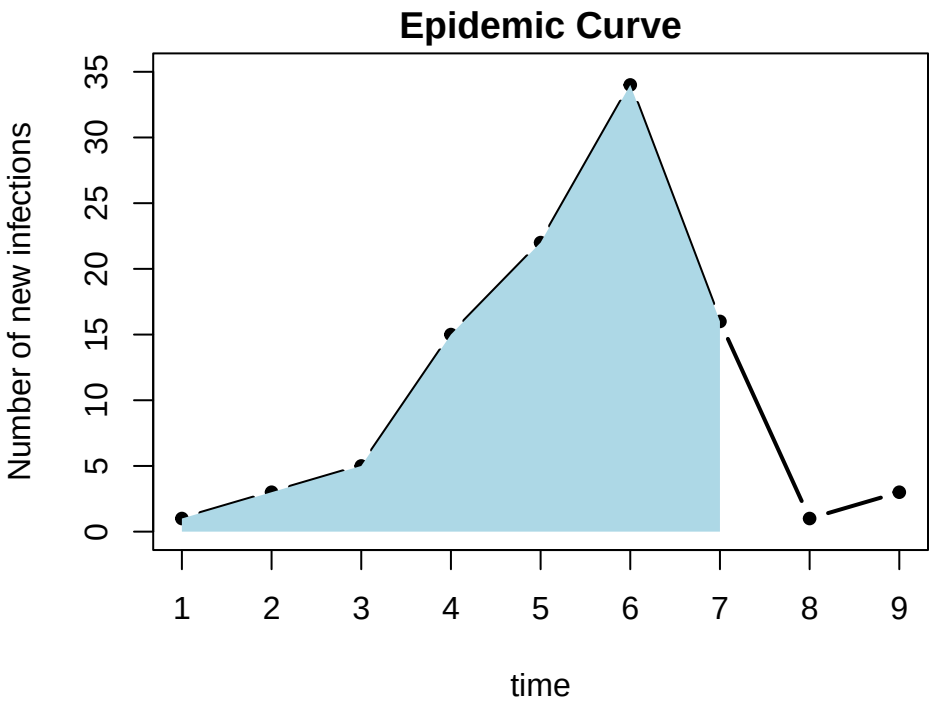
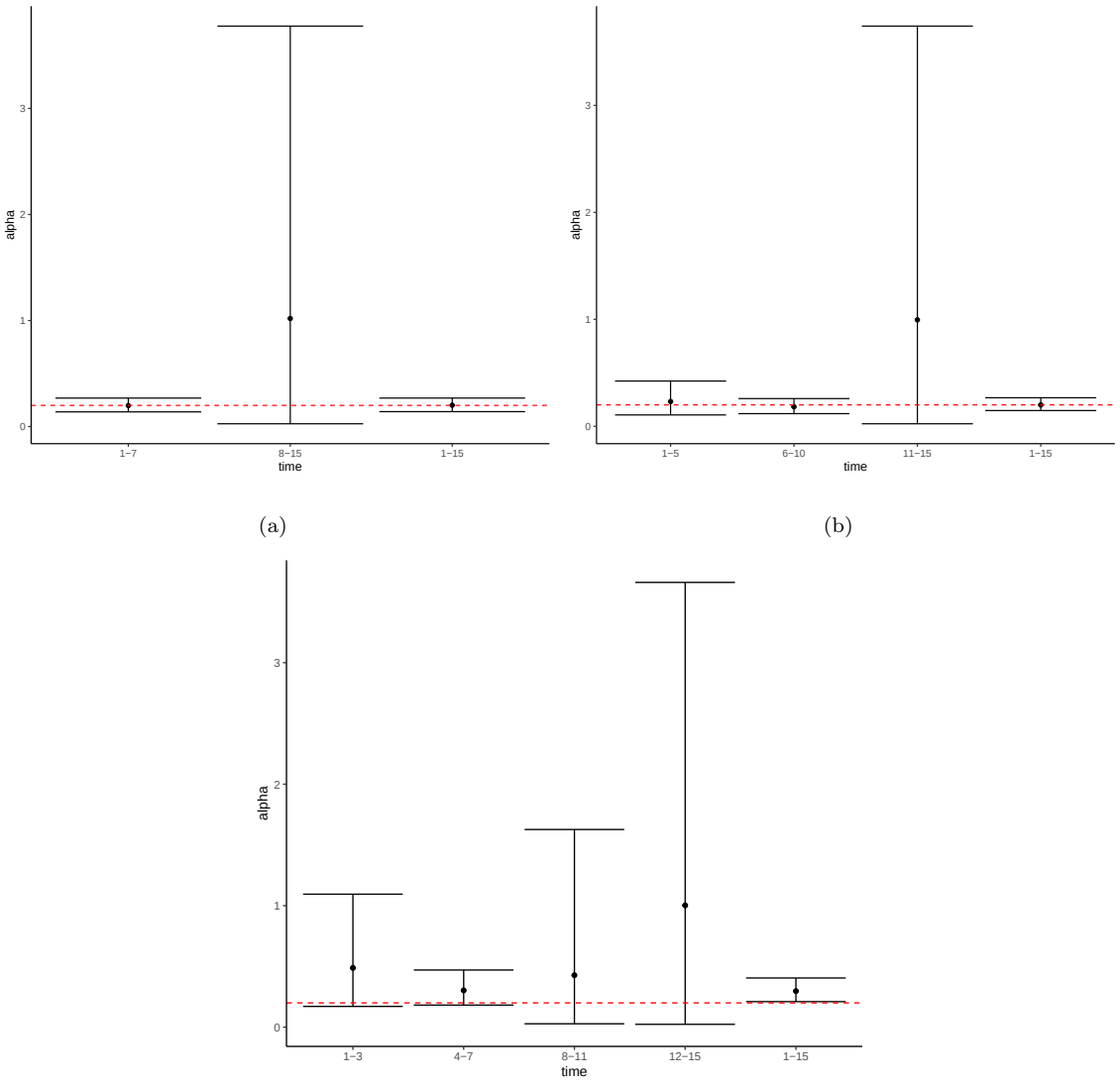


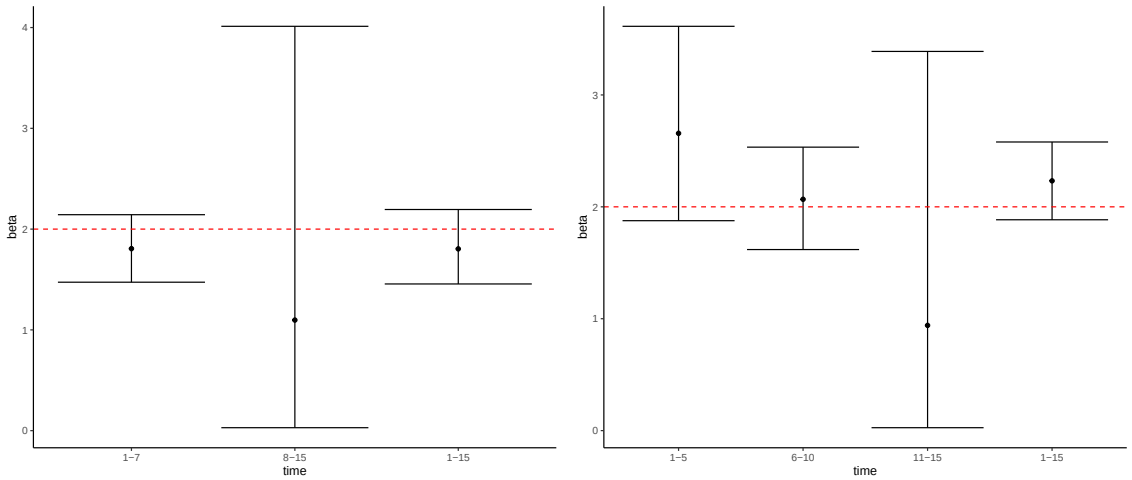
Figure 2.5: Example of temporal subset of individuals whose infected time is between 1 and 7 for population size $n = 100$



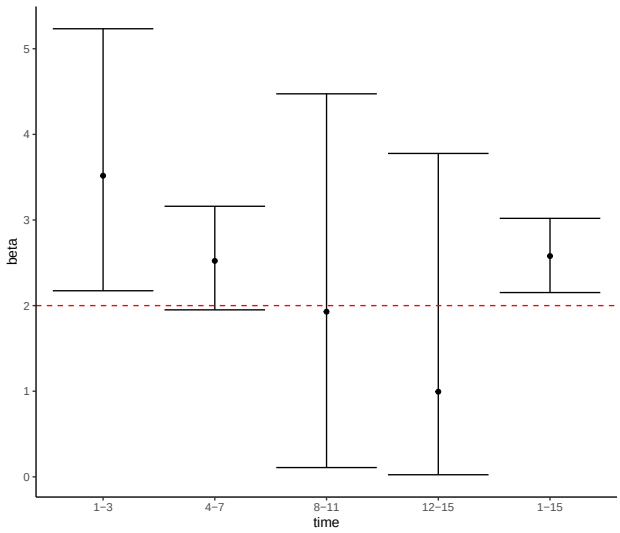
(a) (b)

(c)

Figure 2.6: 95% Credible Intervals of alpha for (a) 2 temporal subsets (b) 3 temporal subsets (c) 4 temporal subsets for $n = 100$



(a) (b)



(c)

Figure 2.7: 95% Credible Intervals of beta for (a) 2 temporal subsets (b) 3 temporal subsets (c) 4 temporal subsets for $n = 100$

2.12 Temporal Subsetting Methods

For this research, we explore two cases related to the temporal inference of posterior parameter estimates of α and β: which infection time frame is the most useful, and whether or not including infection time frames that contain few epidemic data improve the estimation process.

In order to delve deeper into the first topic, we generate a method which is referred to as **temporal 1 subset-m**. This method involves subsetting epidemic data from minimum infection time ($t = 1$) to desired infection time m, $m = 2, 3, 4, ..., t_{max}$. $E_S S(.)$ and $SD_{SS}(.)$ are derived via the posterior estimates of the

subset data. This method is computationally efficient as it only requires one Bayesian MCMC framework to obtain $E_S S(.)$ and $SD_{SS}(.)$.

Concerning the second scenario, we devise two methods, herein referred to as **temporal mean subset-h** and **temporal median subset-h** methods. Those two methods first require temporally subsetting the epidemic data into h-parts, $h = 2, 3, 4$. Finally, we then acquire $E_{SS}(.)$ and $SD_{SS}(.)$ by finding the average of posterior parameter estimates of α and β across all h groups for **temporal mean subset-h** method and median of those posterior estimates for **temporal mean subset-h**.

The following sets describe the temporal subsets of the epidemic data based on the value of h.When $h = 2$,

$$S_1 = \{1, 2, 3, 4, 5, 6, 7\}, S_2 = \{8, 9, 10, 11, 12, 13, 14, 15\} \tag{2.29}$$

When $h = 3$,

$$S_1 = \{1, 2, 3, 4, 5\}, S_2 = \{6, 7, 8, 9, 10\}, S_3 = \{11, 12, 13, 14, 15\} \tag{2.30}$$

When $h = 4$,

$$S_1 = \{1, 2, 3\}, S_2 = \{4, 5, 6, 7\}, S_3 = \{8, 9, 10, 11\}, S_4 = \{12, 13, 14, 15\} \tag{2.31}$$

where S_k refers to the k-th temporal subsets containing associated infection times, $k = 1, 2, 3, 4$

Chapter 3

Results

To understand more about the behaviour of absolute biases of α and β, we generate 20 epidemics under each subset method. Finally, we calculate the median and third quantile of absolute biases under those 20 epidemics to measure the performance of each method.

Additionally, to enhance the readability of the results, subset methods are grouped, and the minimum median and third-quantile absolute biases of the parameter estimate for each group are highlighted in green in Tables 3.1 to 3.6. The organization of the groups is as follows:

- **Average Center**, **Average ll**, **Average Corner**, and **Average Random** methods,
- **k-percent center** method, $k = 10, 20, 30, 40, 50, 64$,
- **k-percent ll** method,$k = 10, 20, 30, 40, 50, 64$,
- **k-percent corner** method,$k = 10, 20, 30, 40, 50, 64$,
- **k-percent random** method,$k = 10, 20, 30, 40, 50, 64$,
- **temporal 1 subset-m**, $m = 2, 3, 4, 5, 6, 7$,
- **temporal mean subset-h**, $h = 2, 3, 4$, and
- **temporal median subset-h**, $h = 2, 3, 4$

For instance, in Table 3.5, the **average center** method has the median absolute bias for α highlighted as green because it has the lowest median absolute bias value in the **average center**, the **average ll**, the **average corner**, and the **average random** method group. In this section, we mainly discuss methods that consistently demonstrate good performance within each group.

Information about the total system and user CPU per epidemic in CPU hours for each population size n are in Tables 3.7 to 3.9. The maximum total system and user CPU per epidemic in CPU hours for each group are highlighted red.

The simulations for all spatial and temporal subset methods are available in the GitHub repository "Data Subset Spatial Epidemic Models" by Nyein (May 2023). [1]

3.1 Spatial Subset Methods

3.1.1 Average Center, Average ll, Average Corner and Average Random Methods

When $n = 100$ and $n = 5000$, the **average center** method displays the lowest median absolute bias for α, measuring 0.044 and 0.005 respectively. In addition, it performs consistently well in predicting SD(α). For $n = 100$ and $n = 1000$, it exhibits the second-lowest median absolute values for SD(α), measuring 0.03 and 0.011. Additionally, for $n = 5000$, this method exhibits the lowest median absolute values for SD(α), measuring 0.005.

When $n = 1000$ and $n = 5000$, **average ll** method demonstrates the lowest median absolute bias for β, with values of 0.149 and 0.113 respectively. In addition, it displays the second-lowest third-quantile absolute bias for β, measuring 0.218 for $n = 1000$ and 0.141 for $n = 5000$. Furthermore, it portrays the lowest median and third-quantile absolute biases for SD(α) when $n = 100$ and $n = 5000$, measuring 0.028 and 0.005 for $n = 100$ and 0.037 and 0.006 for $n = 5000$. Furthermore, it demonstrates the lowest third-quantile absolute bias for SD(β) when $n = 100$ and $n = 1000$, yielding values of 0.176 and 0.042 respectively. Lastly, it demonstrates the second-lowest median absolute bias for SD(β) when $n = 5000$, measuring 0.019.

The **average corner** method yields the lowest third-quantile absolute bias values for α, β and SD(β) across all population sizes (n). Specifically, it records the following third-quantile absolute bias values for α: 0.062 for $n = 100$, 0.017 for $n = 1000$, and 0.008 for $n = 5000$. Similarly, for the third-quantile absolute biases for β, it displays the following values: 0.318 for $n = 100$, 0.181 for $n = 1000$, and 0.138 for $n = 5000$. Likewise, for SD(β), it records the following third-quantile absolute bias values: 0.176 for $n = 100$, 0.042 for $n = 1000$, and 0.018 for $n = 5000$. Additionally, when $n = 1000$ and $n = 5000$, this method exhibits the lowest median and third-quantile absolute bias for SD(α): 0.009 and 0.011 when $n = 1000$ and 0.005 and 0.006 for $n = 5000$ respectively. However,for $n = 100$, it displays the third-lowest median and third-quantile absolute bias for SD(α): yielding 0.036 and 0.041 respectively.

[1] Data Subset Spatial Epidemic Models. Available at: `https://github.com/tnyein99/Data-Subset-Spatial-Epidemic-Models`. Accessed on May 30, 2023.

The **average corner** method produces the lowest minimum third quantile absolute biases in the majority of the population sizes (n) except for $n = 100$. Even for $n = 100$, this method displays the lowest median and third-quantile absolute biases for the majority of parameter estimates, except for SD(α).

Additionally, **average ll** method performs similarly to the **average corner** method when $n = 5000$. However, the same scenario does not apply to the smaller population sizes $n = 100$ and $n = 1000$.

Furthermore, the **average center** and **average ll** methods exhibit comparable values of median and third-quantile absolute biases for all parameter estimates across all population sizes.

The **average random** method performs the worst in this group as it consistently yields the largest or the second-largest median and third-quantile absolute biases for the majority of parameter estimates across various population sizes (n). Hence, in this group, the **average corner** method would be the best method to use for estimation.

The **average center** and **average ll** methods consume a similar amount of computational resources per epidemic in terms of the total system and user CPU hours per epidemic. Likewise, the **average corner** and the **average random** methods also demonstrate a similar total system and user CPU per epidemic in CPU hours when $n = 100$ and $n = 1000$.However, as the population size increases to $n = 5000$, the difference in total CPU hours per epidemic between these two methods becomes more pronounced.

Hence, in this group, the **Average corner** method would be the best method to use for estimation since it has low median and third-quantile absolute biases.

3.1.2 k-percent Center Method

The **50-percent center** method displays the second-lowest median and third-quantile absolute bias values for β, SD(α), and SD(β) across all population sizes (n). Particularly, it yields the following values for median and third-quantile absolute bias for β: 0.266 and 0.425 for $n = 100$, 0.174 and 0.234 for $n = 1000$, and 0.092 and 0.114 for $n = 5000$. Similarly, for SD(α), it demonstrates the following values for median and third-quantile absolute biases: 0.018 and 0.024 for $n = 100$, 0.009 and 0.011 for $n = 1000$, and 0.004 and 0.005 for $n = 5000$. Likewise, for SD(β), it exhibits the following values for median and third-quantile absolute biases: 0.091 and 0.109 for $n = 100$, 0.032 and 0.035 for $n = 1000$, and 0.01 and 0.015 for $n = 5000$. Moreover, this method demonstrates the second-lowest median absolute bias for α when $n = 100$ and $n = 5000$, measuring 0.029 and 0.005 respectively. Additionally, this method displays the lowest median absolute bias for α when $n = 1000$ at a value of 0.009. Furthermore, it yields the second-lowest third-quantile absolute bias for α when $n = 1000$ with a value of 0.02. Lastly, it displays the lowest third-quantile absolute bias for α when $n = 100$ and $n = 5000$, measuring 0.049 and 0.007 respectively.

The **64-percent center** method exhibits the lowest median and third-quantile absolute bias values for β, SD(α), and SD(β) across all population sizes (n). Specifically, it produces the following values for median and third-quantile absolute bias for β: 0.201 and 0.305 for $n = 100$, 0.109 and 0.131 for $n = 1000$, and 0.069 and 0.094 for $n = 5000$. In particular, it demonstrates the following values for median and third-quantile absolute bias for SD(α): 0.016 and 0.021 for $n = 100$, 0.005 and 0.007 for $n = 1000$, and 0.002 and 0.003 for $n = 5000$. Likewise, it exhibits the following values for median and third-quantile absolute bias for SD(β): 0.069 and 0.081 for $n = 100$, 0.017 and 0.023 for $n = 1000$, and 0.006 and 0.009 for $n = 5000$. Additionally, when $n = 5000$, this method yields the lowest median and third-quantile absolute biases for α, measuring 0.004 and 0.007. Similarly, it yields the lowest median absolute bias for α when $n = 100$, measuring 0.026. Moreover, for $n = 1000$, this method displays the second-lowest median absolute bias for α at a value of 0.011. Likewise, this method demonstrates the lowest third-quantile absolute bias for α when $n = 1000$, measuring 0.017. Lastly, when $n = 100$, it demonstrates the second-lowest third-quantile absolute bias for α, measuring 0.053.

The **64-percent center** method performs the best compared to all other **k-percent center** methods. However, the **50-percent center** method has a similar performance as the **64-percent Center** in terms of median and third-quantile absolute biases of the parameter estimates for larger population sizes $n = 1000$ and $n = 5000$.

When taking in account for the computational resources, the **k-percent center** method tends to have higher total system and user CPU hours per epidemic as the subset size or the value of k increases. Hence, the **50-percent center** would be the best method out of **k-percent center** methods for larger population sizes (n) since **50-percent Center** and **64-percent center** methods perform similarly in the large value of n, and **k-percent center** method has the higher computational cost as the value of k increases.

3.1.3 k-percent ll Method

The **50-percent ll** method exhibits the second-lowest median and third-quantile absolute biases for SD(β) for all values of n. In particular, this method displays the following values for median and third-quantile absolute bias for SD(β): 0.102 and 0.12 when $n = 100$, 0.024 and 0.034 when $n = 1000$, and 0.009 and 0.016 when $n = 5000$. Similarly, it yields the second-lowest third-quantile absolute bias values for SD(α) for all population sizes (n): 0.033 when $n = 100$, 0.008 when $n = 1000$, and 0.005 when $n = 5000$. In addition, it demonstrates the second-lowest median SD(α) biases for $n = 100$ and $n = 1000$, resulting in values of 0.022 and 0.006 respectively. Moreover, it displays the lowest median absolute for SD(α) when $n = 5000$, yielding a value of 0.003. It also results in the second-lowest median and third-quantile absolute biases for

β when $n = 100$ and $n = 1000$. Specifically, it produces the following values for median and third-quantile absolute bias for β: 0.26 and 0.407 when $n = 100$ and 0.123 and 0.179 when $n = 1000$. Furthermore, when $n = 5000$, this method displays the lowest median and third-quantile absolute bias for β, measuring 0.057 and 0.065. Moreover, it also displays the lowest third-quantile absolute bias for α when $n = 1000$ and $n = 5000$, measuring 0.02 and 0.009 respectively. Finally, for $n = 100$, it produces the second-lowest third-quantile absolute bias for α with a value of 0.049.

The **64-percent ll** method demonstrates the lowest median and third-quantile absolute biases for SD(α) and SD(β) for all population sizes (n). Specifically, it yields the following values for the median and third-quantile absolute biases for SD(α): 0.01 and 0.015 when $n = 100$, 0.005 and 0.006 when $n = 1000$, and 0.003 and 0.004 when $n = 5000$. Similarly, it demonstrates the following values for the median and third-quantile absolute biases for SD(β): 0.058 and 0.068 when $n = 100$, 0.015 and 0.017 when $n = 1000$, and 0.007 and 0.008 when $n = 5000$. In addition, this method produces the minimum median and third-quantile absolute biases for α when $n = 100$, measuring 0.02 and 0.037. Similarly, when $n = 5000$, it exhibits the lowest median absolute bias for α at 0.005. Moreover, it displays the second-lowest median and third-quantile absolute bias values for α when $n = 1000$, resulting in values of 0.012 and 0.023 respectively. Furthermore, it exhibits the second-lowest third-quantile absolute bias for α when $n = 5000$: yielding a value of 0.01. In addition, it displays the lowest median and third-quantile absolute bias values for β when $n = 100$ and $n = 1000$: measuring 0.137 and 0.23 for $n = 100$ and 0.053 and 0.119 for $n = 1000$. Lastly, it exhibits the second-lowest median and third-quantile absolute biases for β when $n = 5000$, measuring 0.076 and 0.093 respectively.

The **64-percent ll** method performs the best in smaller population sizes $n = 100$ and $n = 1000$. For $n = 5000$, the **50-percent ll** method performs similar or even better in certain parameter estimates compared to **64-percent ll** method.

Similar to the **k-percent center** method, the total system and user CPU hours per epidemic increases as the subset size or the value of k increases. Therefore, the **50-percent ll** method can be considered the most efficient and effective method among **k-percent ll** methods for simulation with larger population due to low computational cost and small median and third-quantile absolute biases.

3.1.4 k-percent Corner Method

For all population sizes (n), the **64-percent corner** method exhibits the second-lowest third-quantile absolute bias of β: yielding 0.983 when $n = 100$, 0.563 when $n = 1000$, and 0.367 when $n = 5000$. Similarly, it produces the second-lowest median absolute bias for SD(β) when $n = 100$ and $n = 1000$, measuring 0.498 and 0.163 respectively. In addition, when $n = 1000$, this method displays the second-lowest median bias for

SD(α) at 0.04. Furthermore, when $n = 1000$, it exhibits the fourth-lowest third-quantile absolute bias for SD(α), yielding 0.048.

This method also displays the lowest median and third-quantile absolute bias values for α and SD(α) when $n = 100$ and $n = 5000$. Specifically, it yields the following values for median and third-quantile absolute bias values of α: 0.321 and 0.357 for $n = 100$, and 0.011 and 0.016 for $n = 5000$. Similarly, it produces the following values for median and third-quantile absolute bias values of SD(α):0.184 and 0.22 for $n = 100$ and 0.018 and 0.019 for $n = 5000$. Similarly, when $n = 100$ and $n = 5000$, it produces the lowest third-quantile absolute bias for SD(β), measuring 0.548 and 0.068 respectively.

However, it demonstrates the fifth-lowest median and third-quantile absolute bias values for α when $n = 1000$, yielding 0.051 and 0.071 respectively. Similarly, for $n = 5000$, it displays the fifth-lowest median absolute bias of SD(β), measuring 0.066. Likewise, for $n = 1000$, it demonstrates the fifth-lowest third-quantile absolute bias of SD(β), measuring 0.182.

Most of the **k-percent corner** method, $k = 10, 20, 30, 40, 50$, do not exhibit consistently good performance in terms of parameter estimation compared to **k-percent center**, **k-percent ll** and **k-percent random** methods. However, the **64-percent corner** method performs with much consistency in terms of low median and third-quantile absolute biases in the majority of population sizes $n = 100$ and $n = 5000$.

In a manner akin to the **k-percent ll** method, as the value of k increases, the total system and user CPU hours per epidemic tend to experience an upward trajectory.

3.1.5 k-percent Random Method

The **50-percent random** method displays the second-lowest median and third-quantile absolute bias values for SD(α) and SD(β) across all parameter estimates for all values of n. In particular, it results in the following values for the median and third-quantile absolute biases for SD(α):0.027 and 0.034 for $n = 100$,0.009 and 0.01 for $n = 1000$, and 0.004 and 0.005 for $n = 5000$. Likewise, it yields the following values for the median and third-quantile absolute biases for SD(β): 0.118 and 0.143 for $n = 100$, 0.029 and 0.033 for $n = 1000$, and 0.011 and 0.015 for $n = 5000$. Similarly, it displays the second-lowest median absolute bias for β across all population sizes: 0.243 for $n = 100$, 0.142 for $n = 1000$, and 0.119 for $n = 5000$. Likewise, this method demonstrates the second-lowest third-quantile absolute bias for β when $n = 1000$ and $n = 5000$: measuring 0.19 and 0.131 respectively. Furthermore, it yields the third-lowest third-quantile absolute bias for β, measuring 0.377. This method also produces the second-lowest third-quantile absolute bias for α when $n = 100$ and $n = 5000$, resulting in values of 0.069 and 0.01 respectively. Lastly, this method produces the fourth-lowest third-quantile absolute bias for α when $n = 1000$, measuring 0.021.

The **64-percent random** method yields the minimum median and third-quantile absolute bias values for all parameter estimates across all population sizes (n). In particular, it results in the following values for the median and third-quantile absolute biases for α: 0.022 and 0.04 for $n = 100$, 0.009 and 0.015 for $n = 1000$, and 0.005 and 0.007 for $n = 5000$. For β, it displays the following values for the median and third-quantile absolute biases: 0.136 and 0.215 for $n = 100$, 0.106 and 0.122 for $n = 1000$, and 0.062 and 0.087 for $n = 5000$. Likewise, for SD(α), it demonstrates the following values for the median and third-quantile absolute biases: 0.012 and 0.017 for $n = 100$, 0.005 and 0.006 for $n = 1000$, and 0.003 and 0.004 for $n = 5000$. Regarding SD(β), it exhibits the following values for the median and third-quantile absolute biases: 0.061 and 0.079 for $n = 100$, 0.015 and 0.017 for $n = 1000$, and 0.006 and 0.011 for $n = 5000$.

As the subset size or the value of k increases, the **k-percent random** methods typically exhibit higher total system and user CPU hours per epidemic.

The **64-percent Random** method outperforms any other **k-percent random** methods, $k = 10, 20, 30, 40, 50, 64$ for all values of n, $n = 100, 1000, 5000$. When we increase the value of k from 50% to 64% for $n = 5000$, the median absolute biases of β and SD(β) are reduced by approximately 50% (from 0.119 to 0.062 and from 0.011 to 0.006). Hence, **64-percent Random** method is the best method among **k-percent Random** method even though it takes approximately twice as long in terms of total CPU hours per epidemic as **50-percent Random** method.

3.2 Temporal Subset Methods

3.2.1 Temporal 1 Subset-m Method

The **temporal 1 subset-6** method displays the second-lowest median and third-quantile absolute bias values for α and β across all population sizes. Specifically, it yields the following values for the median and third-quantile absolute biases for α: 0.008 and 0.013 for $n = 100$, 0.006 and 0.01 for $n = 1000$, and 0.004 and 0.006 for $n = 5000$. Likewise, for β, it results in the following values for the median and third-quantile absolute biases: 0.041 and 0.052 for $n = 100$, 0.012 and 0.023 for $n = 1000$, and 0.012 and 0.017 for $n = 5000$. Similarly, this method exhibits the lowest median absolute bias values for SD(α) for all population sizes (n), measuring 0.003 for $n = 100$, 0.007 for $n = 1000$, and 0.008 for $n = 5000$ respectively. In a similar manner, it yields the lowest median absolute bias for SD(β) across all values of n, resulting in values of 0.011 for $n = 100$, 0.009 for $n = 1000$, and 0.013 for $n = 5000$ respectively. Furthermore, it exhibits the second-lowest third-quantile absolute bias values for SD(α) and SD(β) when $n = 100$ and $n = 1000$. For third-quantile absolute biases for SD(α), it results in values of 0.005 for $n = 100$ and 0.01 for $n = 1000$ respectively. For

third-quantile absolute biases for SD(β), it yields the values of 0.019 for $n = 100$ and 0.011 for $n = 1000$ respectively. Lastly, it demonstrates the lowest third-quantile absolute bias for SD(α) and SD(β) when $n = 5000$, measuring 0.009 for SD(α) and 0.014 for SD(β) respectively.

The **temporal 1 subset-7** method exhibits the lowest median and third-quantile absolute biases for α and β for all values of n. In particular, it results in the following values for the median and third-quantile absolute biases for α: 0.003 and 0.006 for $n = 100$, 0.001 and 0.002 for $n = 1000$, and 0.002 and 0.003 for $n = 5000$. Similarly, for β, it produces the following values for the median and third-quantile absolute biases: 0.015 and 0.031 when $n = 100$, 0.006 and 0.009 when $n = 1000$, and 0.004 and 0.006 when $n = 5000$. Likewise, it demonstrates the lowest median absolute bias for SD(α) for all population sizes (n), measuring 0.002 for $n = 100$, 0.003 for $n = 1000$, and 0.007 for $n = 5000$. In a similar manner, it displays the minimum median absolute bias for SD(β) for all values of n, yielding 0.006 for $n = 100$, 0.005 for $n = 1000$, and 0.012 for $n = 5000$. Analogously, when $n = 100$ and $n = 1000$, this method exhibits the lowest third-quantile absolute bias for SD(α) and SD(β). For third-quantile absolute biases of SD(α), it produces 0.004 when $n = 100$ and 0.008 when $n = 1000$ respectively. In addition, for third-quantile absolute biases of SD(β), it results in the values of 0.015 when $n = 100$ and 0.009 when $n = 1000$ respectively. Furthermore, for $n = 5000$, it yields the third-lowest third-quantile absolute bias for SD(α), measuring 0.014 as well as the second-lowest third-quantile absolute bias for SD(β), measuring 0.019.

The **Temporal 1 Subset-m Method** follows the same trend as the **k-percent center**, **k-percent ll**, **k-percent corner**, and **k-percent random** methods since the total CPU hours per epidemic tends to increase as higher values of m for $n = 5000$ and $n = 1000$. However, for smaller sample sizes like $n = 100$, the change in computational power is minimal.

The analysis of results in the Figures 3.1 to 3.4 for $n = 5000$ reveals a consistent pattern: as the value of m increases, both the median and third quantile absolute biases of all parameter estimates decrease. This trend is further supported by the findings for $n = 100$ and $n = 1000$, as shown in Figures A.62 to A.65 and Figures A.46 to A.49 in the Appendix.

Furthermore, it is worth noting that once the value of $m \geq 5$, the rate of change of the median and third quantile absolute biases diminishes to zero. Moreover, as depicted in the Figures 3.1 to 3.4, Figures A.46 to A.49 and A.62 to A.65 the interquartile range of absolute biases tends to converge towards zero, indicating a convergence of bias estimates as m increases.

The **temporal 1 subset-6** method is the most efficient. It produces similar absolute biases for parameter estimates across different population sizes compared to **temporal 1 subset-7** method. Moreover, this method consumes less total system and user CPU per epidemic than **temporal 1 subset-7** method.

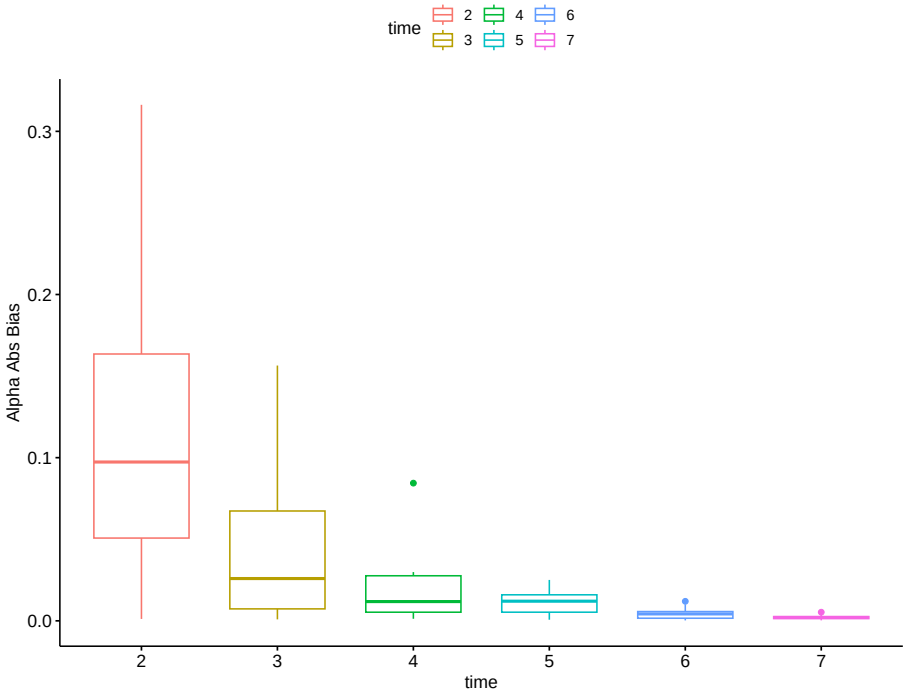


Figure 3.1: Alpha Absolute Bias of Temporal 1 Subset-m method for $n = 5000$

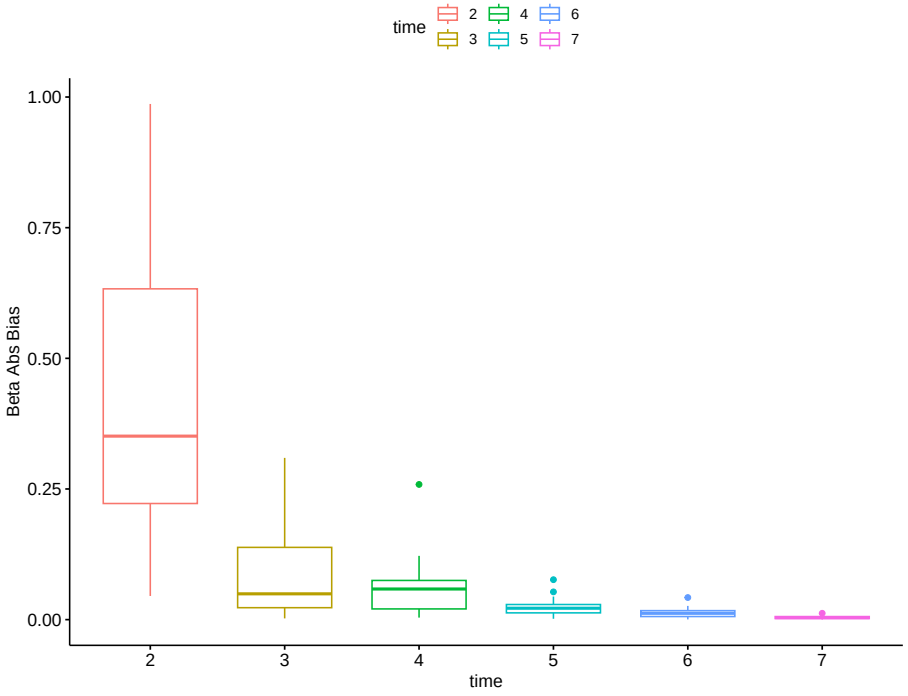


Figure 3.2: Beta Absolute Bias of Temporal 1 Subset-m method for $n = 5000$

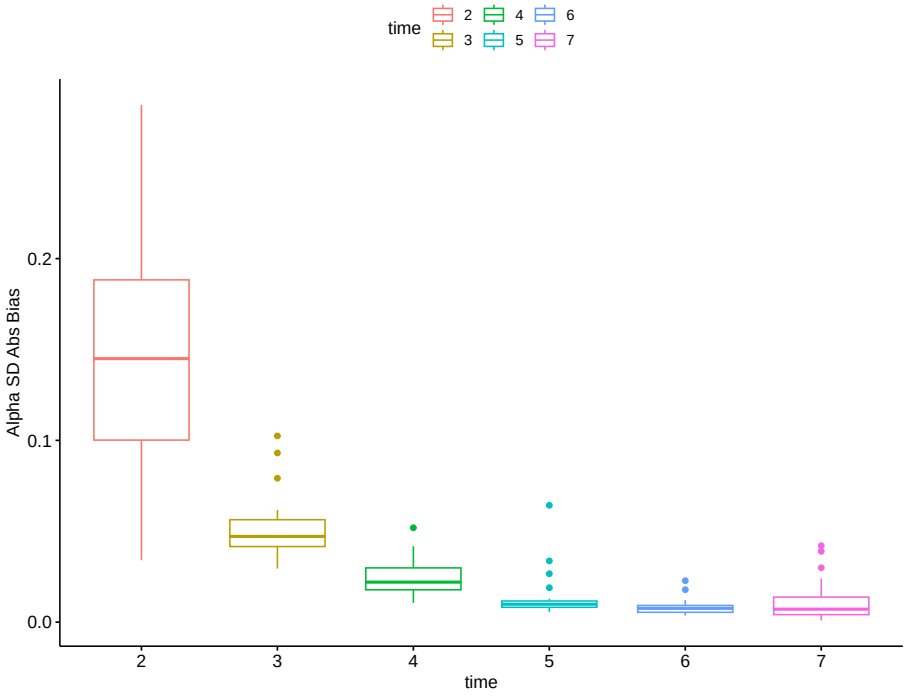


Figure 3.3: Alpha SD Absolute Bias Temporal 1 Subset-m method for $n = 5000$

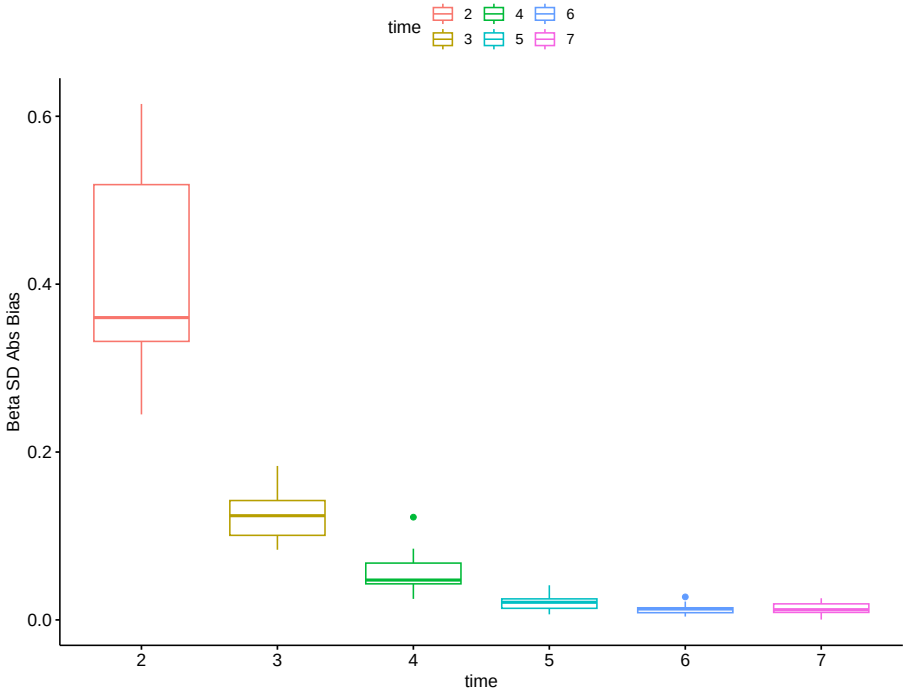


Figure 3.4: Beta SD Absolute Bias of Temporal 1 Subset-m method for $n = 5000$

3.2.2 Temporal Mean Subset-h Method

When $n = 100$ and $n = 1000$, the **temporal mean subset-3** method exhibits the lowest median and third-quantile absolute bias values for α and SD(α). In particular, it yields the following values for the median and third-quantile absolute biases for α: 0.269 and 0.327 for $n = 100$, and 0.263 and 0.275 for $n = 1000$. For SD(α), it exhibits the median and third-quantile absolute biases of 0.358 and 0.403 for $n = 100$ and 0.334 and 0.342 for $n = 1000$. Similarly, when $n = 5000$, it displays the lowest third-quantile absolute bias for α and SD(α) at values of 0.282 and 0.343 respectively. Moreover, when $n = 5000$, it demonstrates the second-lowest median absolute bias for α, and SD(α) measuring 0.273 and 0.339 respectively.

Moreover, for $n = 1000$, it displays the lowest median and third-quantile absolute bias for β: yielding 0.332 and 0.36 respectively. Likewise, this method demonstrates the lowest median absolute bias for β when $n = 100$: resulting in a value of 0.413. In a similar manner, when $n = 5000$, it exhibits the lowest third-quantile absolute bias for β at 0.337. When $n = 5000$, it displays the largest median absolute bias for β at 0.323. Additionally, when $n = 100$, it exhibits the largest third-quantile absolute bias for β at 0.545.

When $n = 1000$ and $n = 5000$, this method displays the lowest median and third-quantile absolute biases for SD(β). Specifically, it produces the following values for the median and third-quantile absolute biases for SD(β): 0.364 and 0.376 for $n = 1000$ and 0.346 and 0.362 for $n = 5000$. However, when $n = 100$, it yields the second-lowest median absolute bias for SD(β) at 0.433 as well as, the largest third-quantile absolute bias for SD(β) at 0.521.

The **temporal mean subset-h** method demonstrates comparable total CPU hours per epidemic for different values of h ($h = 2, 3, 4$) across all population sizes n ($n = 100, 1000, 5000$).

The **temporal mean subset-3** method tends to perform well consistently, in terms of having the lowest third quantile absolute bias for most parameter estimates across different population sizes (n). Even though **temporal mean subset-2** and **temporal mean subset-4** methods occasionally have the lowest median and/or third-quantile absolute biases for some parameters, they are not as consistent as **temporal mean subset-3** method. Therefore, it can be implied the **temporal mean subset-3** method has the best leverage against potentially high absolute bias for parameter estimation purposes. There is minimal disparity in the total system and user CPU per epidemic in CPU hours with **temporal mean subset-h** methods for all values of h, ($h = 2, 3, 4$).

3.2.3 Temporal Median Subset-h Method

When $n = 1000$ and $n = 5000$, the **temporal median subset-3** method displays the lowest median and third-quantile absolute bias for all parameter estimates. In particular, it produces the following values for

median and third-quantiles of α: 0.013 and 0.027 for $n = 1000$, and 0.01 and 0.018 for $n = 5000$. For β, it yields the following values for median and third-quantile absolute biases: 0.045 and 0.128 for $n = 1000$, and 0.017 and 0.036 for $n = 5000$. Regarding SD(α), it produces the following values for median and third-quantile absolute biases: 0.024 and 0.043 for $n = 1000$, and 0.012 and 0.013 for $n = 5000$. Likewise, for SD(β), this method results in the following values for median and third-quantile absolute biases: 0.11 and 0.191 for $n = 1000$, and 0.032 and 0.043 for $n = 5000$.

When $n = 100$, this method produces the lowest median absolute bias for α and SD(α), yielding 0.036 and 0.067 respectively. However, it displays the largest third-quantile absolute bias for α and SD(α), measuring 0.734 and 0.876 respectively. Furthermore, this method produces the largest median and third-quantile absolute bias for β and SD(β): measuring 0.673 and 0.953 for β and 0.549 and 0.626 for SD(β)

Analogous to the **temporal mean subset-h** method, the **temporal median subset-h** method demonstrates relatively consistent total CPU hours per epidemic across different values of h ($h = 2, 3, 4$) and population sizes n ($n = 100, 1000, 5000$).

The **temporal median subset-3** method produces the minimum median and third quantile absolute biases for all parameter estimates in larger populations $n = 1000$ and $n = 5000$. Moreover, across all values of h ($h = 2, 3, 4$), the total system and user CPU per epidemic in CPU hours show negligible variation when employing the **temporal median subset-h** methods. Due to having the best results in parameter predictions and also having the computaional cost as the **temporal median subset-2** and **temporal median subset-4**methods, **temporal median subset-3** method can be considered the best method in this group.

Table 3.1: Median and third-quantile for absolute biases of α and β for n=100

Methods	**Alpha**		**Beta**	
	Median	**3rd Quantile**	**Median**	**3rd Quantile**
Average Center	0.044	0.062	0.340	0.460
Average ll	0.050	0.077	0.341	0.502
Average Corner	0.057	0.062	0.263	0.318
Average Random	0.069	0.084	0.282	0.484
10 Percent Center	0.320	0.809	1.089	1.294
20 Percent Center	0.131	0.187	0.625	1.093
30 Percent Center	0.052	0.087	0.418	0.699
40 Percent Center	0.050	0.065	0.481	0.639
50 Percent Center	0.029	0.049	0.266	0.425
64 Percent Center	0.026	0.053	0.201	0.305
10 Percent ll	0.218	0.462	0.947	1.246
20 Percent ll	0.059	0.161	0.561	0.790
30 Percent ll	0.033	0.109	0.481	0.900
40 Percent ll	0.038	0.061	0.330	0.518
50 Percent ll	0.034	0.049	0.260	0.407
64 Percent ll	0.020	0.037	0.137	0.230
10 Percent Corner	0.400	0.493	0.739	1.098
20 Percent Corner	0.324	0.386	0.801	1.027
30 Percent Corner	0.414	0.487	0.915	1.061
40 Percent Corner	0.321	0.486	0.711	1.009
50 Percent Corner	0.334	0.444	0.737	0.917
64 Percent Corner	0.321	0.357	0.881	0.983
10 Percent Random	0.420	0.599	0.856	0.922
20 Percent Random	0.118	0.176	0.562	0.837
30 Percent Random	0.059	0.098	0.374	0.800
40 Percent Random	0.067	0.111	0.249	0.353
50 Percent Random	0.035	0.069	0.243	0.377
64 Percent Random	0.022	0.040	0.136	0.215
Temporal 1 Subset-2	0.101	0.138	0.382	0.619
Temporal 1 Subset-3	0.041	0.099	0.237	0.606
Temporal 1 Subset-4	0.025	0.051	0.118	0.225
Temporal 1 Subset-5	0.013	0.022	0.071	0.106
Temporal 1 Subset-6	0.008	0.013	0.041	0.052
Temporal 1 Subset-7	0.003	0.006	0.015	0.031
Temporal Mean Subset-2	0.406	0.458	0.456	0.540
Temporal Mean Subset-3	0.269	0.327	0.413	0.545
Temporal Mean Subset-4	0.395	0.421	0.467	0.525
Temporal Median Subset-2	0.400	0.431	0.511	0.547
Temporal Median Subset-3	0.036	0.734	0.673	0.953
Temporal Median Subset-4	0.392	0.415	0.493	0.580

Table 3.2: Median and third-quantile for absolute biases of SD(α) and SD(β) for n=100

Methods	**SD(Alpha)**		**SD(Beta)**	
	Median	**3rd Quantile**	**Median**	**3rd Quantile**
Average Center	0.030	0.039	0.161	0.178
Average ll	0.028	0.037	0.142	0.176
Average Corner	0.036	0.041	0.169	0.176
Average Random	0.043	0.056	0.182	0.199
10 Percent Center	0.176	0.504	0.410	0.666
20 Percent Center	0.067	0.116	0.283	0.366
30 Percent Center	0.037	0.059	0.238	0.296
40 Percent Center	0.029	0.039	0.156	0.169
50 Percent Center	0.018	0.024	0.091	0.109
64 Percent Center	0.016	0.021	0.069	0.081
10 Percent ll	0.137	0.219	0.462	0.551
20 Percent ll	0.046	0.096	0.315	0.366
30 Percent ll	0.033	0.070	0.249	0.270
40 Percent ll	0.024	0.046	0.167	0.206
50 Percent ll	0.022	0.033	0.102	0.120
64 Percent ll	0.010	0.015	0.058	0.068
10 Percent Corner	0.264	0.372	0.550	0.641
20 Percent Corner	0.200	0.252	0.456	0.619
30 Percent Corner	0.224	0.341	0.515	0.598
40 Percent Corner	0.229	0.280	0.503	0.602
50 Percent Corner	0.203	0.301	0.546	0.644
64 Percent Corner	0.184	0.220	0.498	0.548
10 Percent Random	0.272	0.405	0.590	0.634
20 Percent Random	0.081	0.117	0.353	0.413
30 Percent Random	0.045	0.056	0.235	0.251
40 Percent Random	0.042	0.057	0.175	0.196
50 Percent Random	0.027	0.034	0.118	0.143
64 Percent Random	0.012	0.017	0.061	0.079
Temporal 1 Subset-2	0.094	0.182	0.556	0.819
Temporal 1 Subset-3	0.054	0.090	0.285	0.317
Temporal 1 Subset-4	0.018	0.026	0.077	0.118
Temporal 1 Subset-5	0.007	0.015	0.030	0.040
Temporal 1 Subset-6	0.003	0.005	0.011	0.019
Temporal 1 Subset-7	0.002	0.004	0.006	0.015
Temporal Mean Subset-2	0.489	0.504	0.379	0.432
Temporal Mean Subset-3	0.358	0.403	0.433	0.521
Temporal Mean Subset-4	0.496	0.506	0.495	0.504
Temporal Median Subset-2	0.477	0.489	0.385	0.403
Temporal Median Subset-3	0.067	0.876	0.549	0.626
Temporal Median Subset-4	0.469	0.509	0.503	0.581

Table 3.3: Median and third-quantile for absolute biases of α and β for n=1000

Methods	**Alpha**		**Beta**	
	Median	**3rd Quantile**	**Median**	**3rd Quantile**
Average Center	0.020	0.026	0.153	0.282
Average ll	0.014	0.030	0.149	0.218
Average Corner	0.014	0.017	0.161	0.181
Average Random	0.015	0.020	0.180	0.222
10 Percent Center	0.046	0.127	0.494	0.817
20 Percent Center	0.035	0.045	0.344	0.502
30 Percent Center	0.019	0.036	0.237	0.333
40 Percent Center	0.014	0.027	0.215	0.275
50 Percent Center	0.009	0.020	0.174	0.234
64 Percent Center	0.011	0.017	0.109	0.131
10 Percent ll	0.048	0.082	0.419	0.565
20 Percent ll	0.027	0.038	0.373	0.472
30 Percent ll	0.012	0.024	0.178	0.235
40 Percent ll	0.023	0.029	0.151	0.248
50 Percent ll	0.011	0.020	0.123	0.179
64 Percent ll	0.012	0.023	0.053	0.119
10 Percent Corner	0.044	0.056	0.499	0.618
20 Percent Corner	0.048	0.059	0.477	0.599
30 Percent Corner	0.040	0.066	0.508	0.587
40 Percent Corner	0.072	0.098	0.441	0.527
50 Percent Corner	0.046	0.056	0.556	0.683
64 Percent Corner	0.051	0.071	0.504	0.563
10 Percent Random	0.050	0.090	0.514	0.588
20 Percent Random	0.037	0.047	0.287	0.361
30 Percent Random	0.014	0.017	0.244	0.356
40 Percent Random	0.009	0.020	0.207	0.291
50 Percent Random	0.010	0.021	0.142	0.190
64 Percent Random	0.009	0.015	0.106	0.122
Temporal 1 Subset-2	0.089	0.113	0.343	0.553
Temporal 1 Subset-3	0.054	0.079	0.169	0.248
Temporal 1 Subset-4	0.022	0.034	0.093	0.124
Temporal 1 Subset-5	0.008	0.017	0.020	0.053
Temporal 1 Subset-6	0.006	0.010	0.012	0.023
Temporal 1 Subset-7	0.001	0.002	0.006	0.009
Temporal Mean Subset-2	0.402	0.420	0.494	0.557
Temporal Mean Subset-3	0.263	0.275	0.332	0.360
Temporal Mean Subset-4	0.387	0.406	0.456	0.539
Temporal Median Subset-2	0.401	0.419	0.516	0.552
Temporal Median Subset-3	0.013	0.027	0.045	0.128
Temporal Median Subset-4	0.390	0.406	0.544	0.578

Table 3.4: Median and third-quantile for absolute biases of SD(α) and SD(β) for n=1000

Methods	**SD(Alpha)**		**SD(Beta)**	
	Median	**3rd Quantile**	**Median**	**3rd Quantile**
Average Center	0.011	0.013	0.042	0.046
Average ll	0.011	0.014	0.040	0.042
Average Corner	0.009	0.011	0.038	0.042
Average Random	0.011	0.012	0.039	0.045
10 Percent Center	0.050	0.069	0.177	0.241
20 Percent Center	0.022	0.025	0.087	0.095
30 Percent Center	0.014	0.019	0.058	0.063
40 Percent Center	0.010	0.011	0.038	0.042
50 Percent Center	0.009	0.011	0.032	0.035
64 Percent Center	0.005	0.007	0.017	0.023
10 Percent ll	0.034	0.038	0.142	0.149
20 Percent ll	0.021	0.025	0.087	0.095
30 Percent ll	0.013	0.015	0.052	0.062
40 Percent ll	0.011	0.014	0.040	0.042
50 Percent ll	0.006	0.008	0.024	0.034
64 Percent ll	0.005	0.006	0.015	0.017
10 Percent Corner	0.040	0.043	0.161	0.180
20 Percent Corner	0.040	0.046	0.170	0.180
30 Percent Corner	0.039	0.048	0.166	0.180
40 Percent Corner	0.047	0.050	0.172	0.176
50 Percent Corner	0.040	0.043	0.171	0.182
64 Percent Corner	0.040	0.048	0.163	0.182
10 Percent Random	0.043	0.050	0.170	0.186
20 Percent Random	0.023	0.028	0.087	0.097
30 Percent Random	0.014	0.015	0.055	0.061
40 Percent Random	0.010	0.011	0.039	0.041
50 Percent Random	0.009	0.010	0.029	0.033
64 Percent Random	0.005	0.006	0.015	0.017
Temporal 1 Subset-2	0.139	0.195	0.428	0.511
Temporal 1 Subset-3	0.049	0.065	0.141	0.266
Temporal 1 Subset-4	0.022	0.029	0.068	0.078
Temporal 1 Subset-5	0.010	0.015	0.024	0.031
Temporal 1 Subset-6	0.007	0.010	0.009	0.011
Temporal 1 Subset-7	0.003	0.008	0.005	0.009
Temporal Mean Subset-2	0.467	0.484	0.406	0.492
Temporal Mean Subset-3	0.334	0.342	0.364	0.376
Temporal Mean Subset-4	0.490	0.502	0.483	0.510
Temporal Median Subset-2	0.463	0.494	0.393	0.450
Temporal Median Subset-3	0.024	0.043	0.110	0.191
Temporal Median Subset-4	0.485	0.502	0.500	0.532

Table 3.5: Median and third-quantile for absolute biases of α and β for $n = 5000$

Methods	**Alpha**		**Beta**	
	Median	**3rd Quantile**	**Median**	**3rd Quantile**
Average Center	0.005	0.010	0.131	0.173
Average ll	0.007	0.010	0.113	0.141
Average Corner	0.006	0.008	0.122	0.138
Average Random	0.010	0.013	0.132	0.148
10 Percent Center	0.031	0.038	0.472	0.530
20 Percent Center	0.011	0.022	0.288	0.338
30 Percent Center	0.009	0.012	0.186	0.235
40 Percent Center	0.011	0.018	0.111	0.126
50 Percent Center	0.005	0.007	0.092	0.114
64 Percent Center	0.004	0.007	0.069	0.094
10 Percent ll	0.016	0.024	0.410	0.511
20 Percent ll	0.017	0.024	0.199	0.219
30 Percent ll	0.007	0.015	0.169	0.209
40 Percent ll	0.007	0.011	0.103	0.136
50 Percent ll	0.006	0.009	0.057	0.065
64 Percent ll	0.005	0.010	0.076	0.093
10 Percent Corner	0.014	0.024	0.330	0.364
20 Percent Corner	0.014	0.019	0.360	0.396
30 Percent Corner	0.019	0.024	0.338	0.389
40 Percent Corner	0.015	0.020	0.354	0.369
50 Percent Corner	0.011	0.019	0.350	0.384
64 Percent Corner	0.011	0.016	0.328	0.367
10 Percent Random	0.016	0.027	0.364	0.393
20 Percent Random	0.009	0.015	0.243	0.280
30 Percent Random	0.009	0.013	0.162	0.178
40 Percent Random	0.005	0.011	0.134	0.162
50 Percent Random	0.008	0.010	0.119	0.131
64 Percent Random	0.005	0.007	0.062	0.087
Temporal 1 Subset-2	0.097	0.164	0.351	0.633
Temporal 1 Subset-3	0.026	0.067	0.049	0.138
Temporal 1 Subset-4	0.012	0.028	0.059	0.075
Temporal 1 Subset-5	0.012	0.016	0.022	0.029
Temporal 1 Subset-6	0.004	0.006	0.012	0.017
Temporal 1 Subset-7	0.002	0.003	0.004	0.006
Temporal Mean Subset-2	0.054	0.363	0.230	0.440
Temporal Mean Subset-3	0.273	0.282	0.323	0.337
Temporal Mean Subset-4	0.324	0.388	0.242	0.533
Temporal Median Subset-2	0.085	0.396	0.205	0.525
Temporal Median Subset-3	0.010	0.018	0.017	0.036
Temporal Median Subset-4	0.279	0.407	0.215	0.512

Table 3.6: Median and third-quantile for absolute biases of SD(α) and SD(β) for $n = 5000$

Methods	**Alpha SD**		**Beta SD**	
	Median	**3rd Quantile**	**Median**	**3rd Quantile**
Average Center	0.005	0.007	0.016	0.021
Average ll	0.005	0.006	0.017	0.019
Average Corner	0.005	0.006	0.017	0.018
Average Random	0.006	0.008	0.016	0.021
10 Percent Center	0.022	0.027	0.072	0.076
20 Percent Center	0.011	0.012	0.036	0.039
30 Percent Center	0.007	0.009	0.023	0.026
40 Percent Center	0.005	0.007	0.016	0.019
50 Percent Center	0.004	0.005	0.010	0.015
64 Percent Center	0.002	0.003	0.006	0.009
10 Percent ll	0.021	0.023	0.070	0.079
20 Percent ll	0.011	0.013	0.039	0.042
30 Percent ll	0.007	0.008	0.022	0.025
40 Percent ll	0.004	0.006	0.014	0.019
50 Percent ll	0.003	0.005	0.009	0.016
64 Percent ll	0.003	0.004	0.007	0.008
10 Percent Corner	0.018	0.020	0.063	0.071
20 Percent Corner	0.019	0.019	0.064	0.068
30 Percent Corner	0.020	0.021	0.069	0.070
40 Percent Corner	0.018	0.019	0.064	0.068
50 Percent Corner	0.018	0.020	0.065	0.068
64 Percent Corner	0.018	0.019	0.066	0.068
10 Percent Random	0.020	0.022	0.070	0.074
20 Percent Random	0.011	0.012	0.035	0.041
30 Percent Random	0.008	0.009	0.026	0.028
40 Percent Random	0.005	0.006	0.016	0.018
50 Percent Random	0.004	0.005	0.011	0.015
64 Percent Random	0.003	0.004	0.006	0.011
Temporal 1 Subset-2	0.145	0.188	0.360	0.519
Temporal 1 Subset-3	0.047	0.056	0.124	0.142
Temporal 1 Subset-4	0.022	0.030	0.048	0.068
Temporal 1 Subset-5	0.010	0.012	0.021	0.025
Temporal 1 Subset-6	0.008	0.009	0.013	0.014
Temporal 1 Subset-7	0.007	0.014	0.012	0.019
Temporal Mean Subset-2	0.107	0.393	0.357	0.420
Temporal Mean Subset-3	0.339	0.343	0.346	0.362
Temporal Mean Subset-4	0.366	0.489	0.473	0.510
Temporal Median Subset-2	0.107	0.464	0.314	0.415
Temporal Median Subset-3	0.012	0.013	0.032	0.043
Temporal Median Subset-4	0.300	0.506	0.444	0.469

Table 3.7: Computational Power Results for $n = 100$

Methods	Total CPU in hours per epidemic
Average Center	0.02
Average ll	0.02
Average Corner	0.05
Average Random	0.05
10 Percent Center	0.01
20 Percent Center	0.01
30 Percent Center	0.01
40 Percent Center	0.01
50 Percent Center	0.01
64 Percent Center	0.01
10 Percent ll	0.01
20 Percent ll	0.01
30 Percent ll	0.01
40 Percent ll	0.01
50 Percent ll	0.01
64 Percent ll	0.01
10 Percent Corner	0.02
20 Percent Corner	0.02
30 Percent Corner	0.02
40 Percent Corner	0.02
50 Percent Corner	0.01
64 Percent Corner	0.01
10 Percent Random	0.01
20 Percent Random	0.02
30 Percent Random	0.02
40 Percent Random	0.02
50 Percent Random	0.02
64 Percent Random	0.02
Temporal 1 Subset-2	0.01
Temporal 1 Subset-3	0.01
Temporal 1 Subset-4	0.01
Temporal 1 Subset-5	0.02
Temporal 1 Subset-6	0.02
Temporal 1 Subset-7	0.02
Temporal Mean Subset-2	0.02
Temporal Mean Subset-3	0.02
Temporal Mean Subset-4	0.02
Temporal Median Subset-2	0.02
Temporal Median Subset-3	0.02
Temporal Median Subset-4	0.02

Table 3.8: Computational Power Results for $n = 1000$

Methods	Total CPU in hours per epidemic
Average Center	1.01
Average ll	1.11
Average Corner	2.06
Average Random	2.02
10 Percent Center	0.54
20 Percent Center	0.61
30 Percent Center	0.63
40 Percent Center	0.58
50 Percent Center	0.69
64 Percent Center	0.70
10 Percent ll	0.52
20 Percent ll	0.58
30 Percent ll	0.53
40 Percent ll	0.71
50 Percent ll	0.65
64 Percent ll	0.70
10 Percent Corner	0.56
20 Percent Corner	0.56
30 Percent Corner	0.54
40 Percent Corner	0.56
50 Percent Corner	0.66
64 Percent Corner	0.58
10 Percent Random	0.60
20 Percent Random	0.61
30 Percent Random	0.68
40 Percent Random	0.80
50 Percent Random	1.13
64 Percent Random	1.25
Temporal 1 Subset-2	0.51
Temporal 1 Subset-3	0.54
Temporal 1 Subset-4	0.64
Temporal 1 Subset-5	0.73
Temporal 1 Subset-6	0.78
Temporal 1 Subset-7	0.99
Temporal Mean Subset-2	1.02
Temporal Mean Subset-3	0.88
Temporal Mean Subset-4	1.03
Temporal Median Subset-2	1.07
Temporal Median Subset-3	0.88
Temporal Median Subset-4	0.96

Table 3.9: Computational Power Results for $n = 5000$

Methods	Total CPU in hours per epidemic
Average Center	22.14
Average ll	22.60
Average Corner	50.19
Average Random	47.71
10 Percent Center	12.31
20 Percent Center	12.17
30 Percent Center	13.75
40 Percent Center	13.98
50 Percent Center	15.10
64 Percent Center	16.27
10 Percent ll	14.32
20 Percent ll	12.90
30 Percent ll	13.46
40 Percent ll	13.42
50 Percent ll	13.92
64 Percent ll	15.62
10 Percent Corner	11.78
20 Percent Corner	15.29
30 Percent Corner	16.91
40 Percent Corner	18.69
50 Percent Corner	22.50
64 Percent Corner	29.40
10 Percent Random	14.00
20 Percent Random	19.50
30 Percent Random	19.07
40 Percent Random	17.95
50 Percent Random	23.95
64 Percent Random	45.49
Temporal 1 Subset-2	11.54
Temporal 1 Subset-3	11.68
Temporal 1 Subset-4	14.49
Temporal 1 Subset-5	13.07
Temporal 1 Subset-6	16.30
Temporal 1 Subset-7	23.27
Temporal Mean Subset-2	25.77
Temporal Mean Subset-3	24.93
Temporal Mean Subset-4	25.19
Temporal Median Subset-2	22.79
Temporal Median Subset-3	23.12
Temporal Median Subset-4	23.10

Chapter 4

Comparative Analysis of Spatial and Temporal Subset Methods, Conclusion and Future Work

4.1 Comparative Analysis of Spatial and Temporal Subset Methods

There are several spatial subset methods whose performance remains consistent across all population sizes (n) and parameter estimates. The **average corner** method consistently outperforms the **k-percent corner** method regardless of the values of k. In addition,the **k-percent Center** and **k-percent ll** methods exhibit similar median and third quantile absolute biases, as well as similar total CPU hours per epidemic, especially for larger populations such as $n = 1000$ and $n = 5000$.

Moreover, both the **k-percent center** and **k-percent ll** methods consume fewer total system and user CPU hours per epidemic in comparison to the **average Center** and **average ll** methods. Similar observations can be made for the **k-percent Random** method, which outperforms the **average random** method for larger values of k, starting from $k = 50\%$ and $k = 64\%$, and also exhibits lower total CPU hours per epidemic.

Overall, the **average**, **average corner**, and **average random** methods demonstrate similar performance across the majority of parameter estimates for all population sizes. Moreover, the **k-percent Center** and **k-percent ll** methods tend to exhibit similar median and third quantile absolute biases, as well as similar

total CPU hours per epidemic, especially for larger populations such as $n = 1000$ and $n = 5000$.

Comparing the **k-percent center**, **k-percent ll**, **k-percent corner**, and **k-percent random** methods for $n = 100$, the **k-percent center** method exhibits the largest absolute bias across all parameter estimates when $k = 10\%$. However, as the value of k increases, the **k-percent corner** method surpasses the **k-percent center** method as the method with the highest median and third quantile absolute bias across all parameter estimates. This same pattern persists when analyzing the results for $n = 1000$ and $n = 5000$. These findings are presented in the corresponding Figures 4.1 to 4.4, A.42 to A.45 and A.58 to A.61.

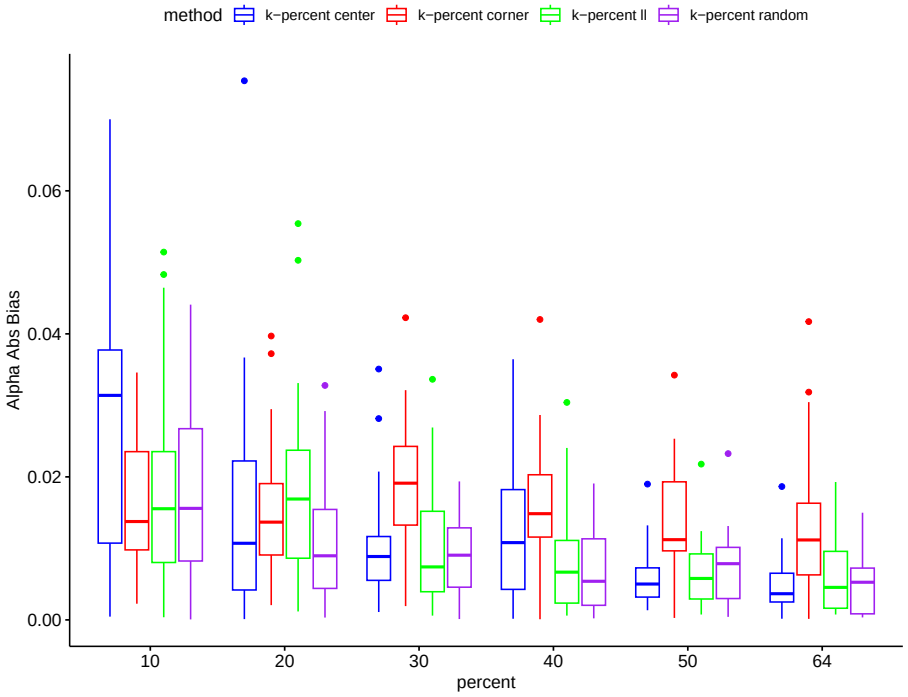


Figure 4.1: Alpha Absolute Bias of k-percent, k-percent Corner, and k-percent Random Methods for $n = 5000$

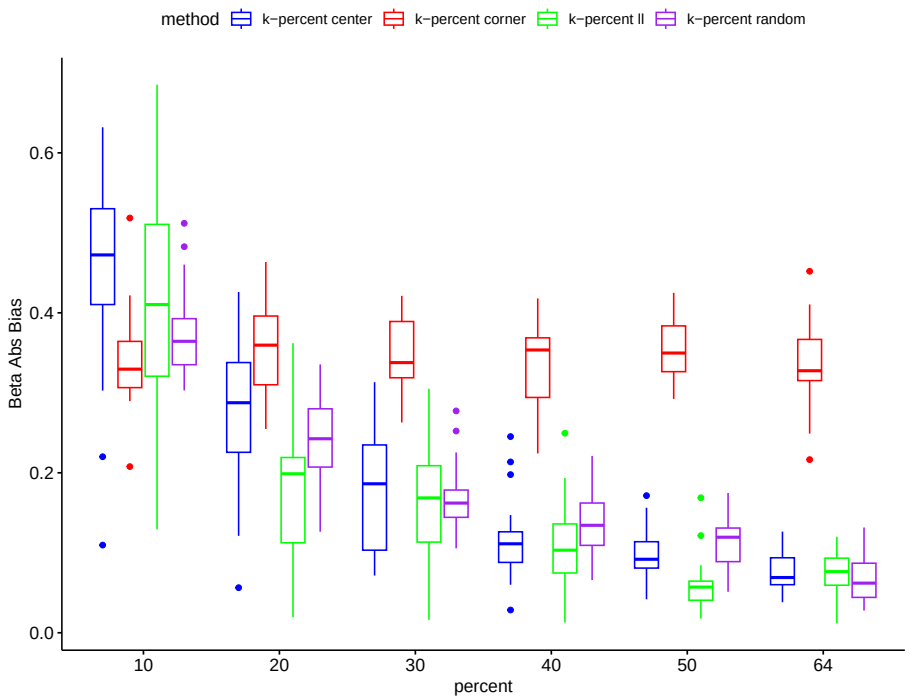


Figure 4.2: Beta Absolute Bias of k-percent, k-percent Corner, and k-percent Random Methods for $n = 5000$

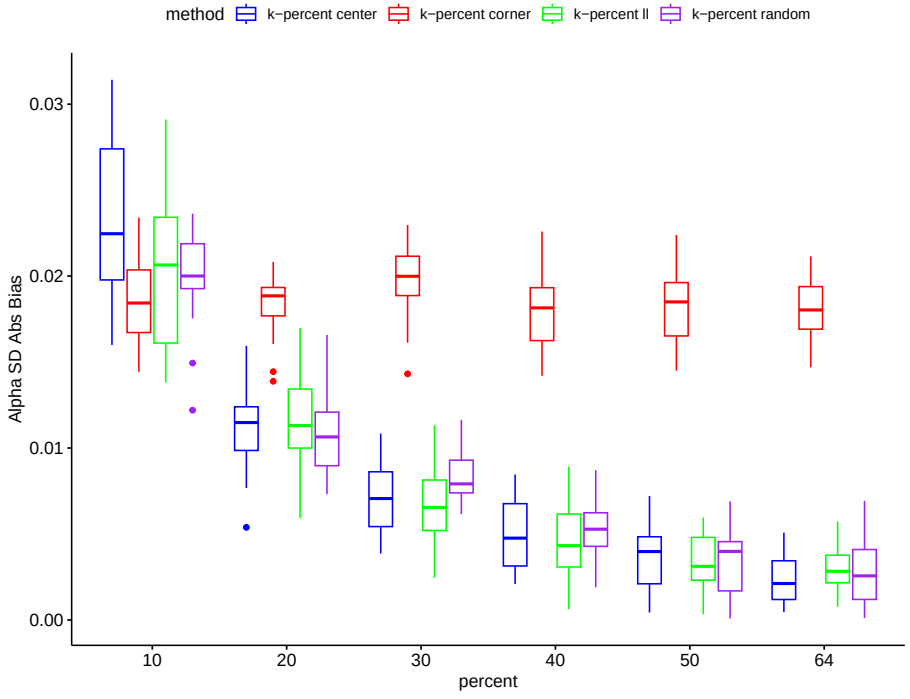


Figure 4.3: Alpha SD Absolute Bias of k-percent, k-percent Corner, and k-percent Random Methods for $n = 5000$

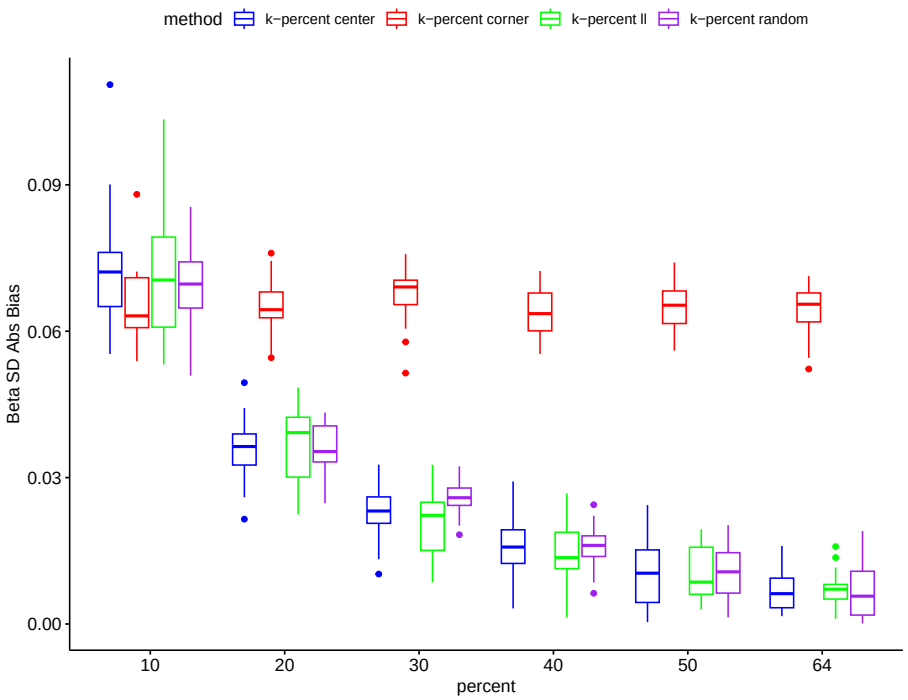


Figure 4.4: Beta SD Absolute Bias of k-percent, k-percent Corner, and k-percent Random Methods for $n = 5000$

For temporal subset methods, **temporal mean subset-h** and **temporal median subset-h** methods produce far-worse absolute biases compared to **temporal 1 subset-m** method for all values of h, m regardless of populations sizes n. Furthermore, the **temporal 1 subset-m** methods exhibit minimal absolute bias values that approach zero for $m \geq 5$ for all parameter estimates and population sizes (n). Finally, **temporal 1 subset-m** methods have comparatively lower absolute biases for all parameter estimates compared to their spatial counterparts starting from $m = 4$.

Regarding computational power, **average center**, **average ll**, **average corner** and **average random** methods require higher total system and user CPU times per epidemic compared to all other spatial and temporal methods for all values of n. **k-percent Center** and **k-percent ll** methods take approximately the same total CPU time. Relative to other spatial and temporal subset methods, **temporal 1 subset-m** is considerably faster, requiring very few computational resources per epidemic regardless of the value of n. However, **temporal mean subset-h** and **temporal median subset-h** methods consume significantly more average CPU time per epidemic compared to the **temporal 1 subset-m** methods for the majority of h, m, n.

4.2 Conclusion

Based on our findings given the values of $\alpha = 0.2$ and $\beta = 2$, it is apparent that, within our simulations, both the center and lower left corner of the square contain a substantial amount of relevant information, particularly for larger population sizes, given that the absolute biases for all parameter estimates are either relatively small or close to zero for **k-percent ll** and **k-percent center** methods. Hence, it can be concluded that for large sample sizes, the location where the subset areas occur do not play an influential role in the inference of posterior parameter estimates.

Moreover, we can deduce that the incorporation of subset areas lacking significant information regarding infected individuals can potentially worsen estimation accuracy, as evidenced by the results obtained from the **k-percent corner** approach. Furthermore, the addition of a randomization factor or random subset areas did not yield improvements in estimation accuracy, as concluded from our analyses.

Additionally, including multiple spatial subset areas in the analysis shown in the **average center**, **average ll**, **average corner** and **average random** methods do not give considerably smaller absolute biases compared to **k-percent center**,**k-percent ll**,**k-percent corner**, and **k-percent random** methods, especially for large value of k. Hence, considering additional subset squares of multiple areas increase the absolute bias, reducing the performance of the inference.

Based on the findings of **temporal 1 subset-m** method, it can be concluded that first-half of the infection time frame ($t = 1, 2, 3, 4, 5, 6, 7$) contains the most information about infected individuals. Moreover, results from **temporal mean subset-h** and **temporal median subset-h** methods suggest that taking in account for infection time frames that do not contain information about infected individuals can adversely affect the inference process. Therefore, it can be deduced that considering infection time frames that contain little to no epidemic data can potentially compromise the accuracy of the estimation process.

Based on the findings of the spatial and temporal subset method, we conclude that **temporal 1 subset-m** method, where $m \geq 5$ is the best subset method for the inference of the spatial and susceptibility parameters.

4.3 Future Works

This research involved simulations of epidemics with the same population density, despite differences in population sizes. In each population, individuals were uniformly distributed across a square grid. However, this scenario may not accurately reflect in models such as geographically dependent individual-level models (GD-ILMs) [12] that incorporate real-life geographical information. Therefore, spatial subset methods need to be refined to account for these factors, including adjusting for different population densities and non-square

geographic areas.

While temporal subset methods can be more accurate and versatile compared to spatial subset methods, they require further refinement to accommodate different values of α, β, and infection periods. Changes to these parameters may result in long computational times, which can limit the usefulness of temporal subset methods in assisting with public health measures during highly contagious epidemics such as Covid-19. However, the spatial subset methods can be applied as long as the spatial information is available regardless of changes in values of α, β, and infection periods. In addition, we need to investigate scenarios with time-varying covariates to check whether the temporal subset methods have the same effectiveness or not in predicting parameter estimates.

Furthermore, additional research is required to determine the effectiveness of both spatial and temporal subset methods for other compartmental models like SEIR, SIRS, SI and so on. Finally, we need to explore if subset methods derived in this thesis can be combined with other methods such as approximate Bayesian computation [20] and linear-based approximation [9], in order to reduce the computational burden.

Bibliography

[1] Aniruddha Adiga et al. "Mathematical models for covid-19 pandemic: a comparative analysis". In: *Journal of the Indian Institute of Science* 100.4 (2020), pp. 793–807.

[2] Pierre Brémaud. *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*. Vol. 31. Springer Science & Business Media, 2001.

[3] Stephen P Brooks and Andrew Gelman. "General methods for monitoring convergence of iterative simulations". In: *Journal of computational and graphical statistics* 7.4 (1998), pp. 434–455.

[4] Mary Kathryn Cowles and Bradley P Carlin. "Markov chain Monte Carlo convergence diagnostics: a comparative review". In: *Journal of the American Statistical Association* 91.434 (1996), pp. 883–904.

[5] Rob Deardon et al. "Inference for individual-level models of infectious diseases in large populations". In: *Statistica Sinica* 20.1 (2010), p. 239.

[6] John F Geweke et al. *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*. Tech. rep. Federal Reserve Bank of Minneapolis, 1991.

[7] Walter R Gilks, Sylvia Richardson, and David Spiegelhalter. *Markov chain Monte Carlo in practice*. CRC press, 1995.

[8] Babak Habibzadeh. "Misspecifying latent and infectious periods in space-time epidemic models". PhD thesis. University of Guelph, 2009.

[9] Grace PS Kwong and Rob Deardon. "Linearized forms of individual-level models for large-scale spatial infectious disease systems". In: *Bulletin of mathematical biology* 74 (2012), pp. 1912–1937.

[10] David A Levin and Yuval Peres. *Markov chains and mixing times*. Vol. 107. American Mathematical Soc., 2017.

[11] Zeyi Liu et al. "Estimating Parameters of Two-Level Individual-Level Models of the COVID-19 Epidemic Using Ensemble Learning Classifiers". In: *Frontiers in Physics* 8 (2021), p. 602722.

[12] MD Mahsin, Rob Deardon, and Patrick Brown. "Geographically dependent individual-level models for infectious diseases transmission". In: *Biostatistics* 23.1 (2022), pp. 1–17.

[13] Martyn Plummer et al. "CODA: convergence diagnosis and output analysis for MCMC". In: *R news* 6.1 (2006), pp. 7–11.

[14] Song S Qian, Craig A Stow, and Mark E Borsuk. "On monte carlo methods for Bayesian inference". In: *Ecological modelling* 159.2-3 (2003), pp. 269–277.

[15] Sheldon M Ross. *Stochastic processes*. John Wiley & Sons, 1995.

[16] Vivekananda Roy. "Convergence diagnostics for markov chain monte carlo". In: *Annual Review of Statistics and Its Application* 7 (2020), pp. 387–412.

[17] Gerardo Rubino and Bruno Sericola. "Discrete-time Markov chains". In: *Markov Chains and Dependability Theory*. Cambridge University Press, 2014, pp. 26–64. DOI: 10.1017/CBO9781139051705.002.

[18] Juliana Tolles and ThaiBinh Luong. "Modeling epidemics with compartmental models". In: *Jama* 323.24 (2020), pp. 2515–2516.

[19] Irene Vrbik et al. "Using Individual-Level Models for Infectious Disease Spread to Model Spatio-Temporal Combustion Dynamics". In: *Bayesian analysis* 7.3 (2012). ISSN: 1936-0975.

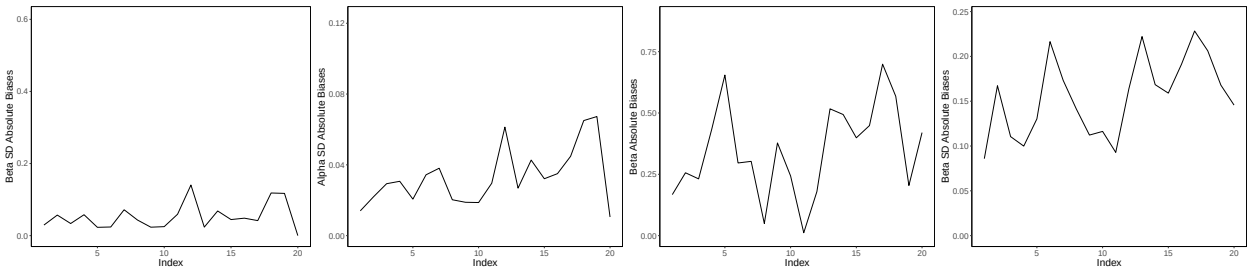
[20] Madeline A Ward, Lorna E Deeth, and Rob Deardon. "Computationally efficient parameter estimation for spatial individual-level models of infectious disease transmission". In: *Spatial and Spatio-temporal Epidemiology* 41 (2022), p. 100497.

Appendix A

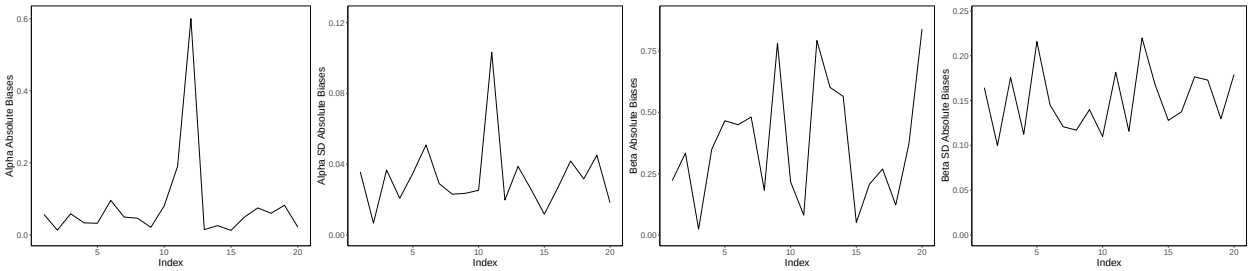
Name of appendix

List of Figures

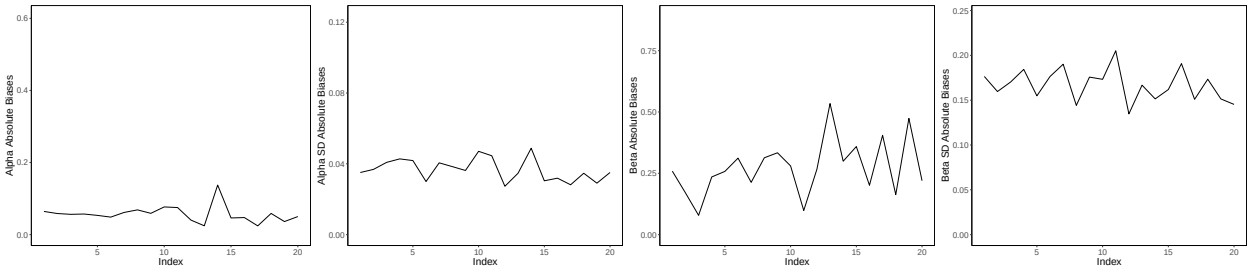
List of Tables



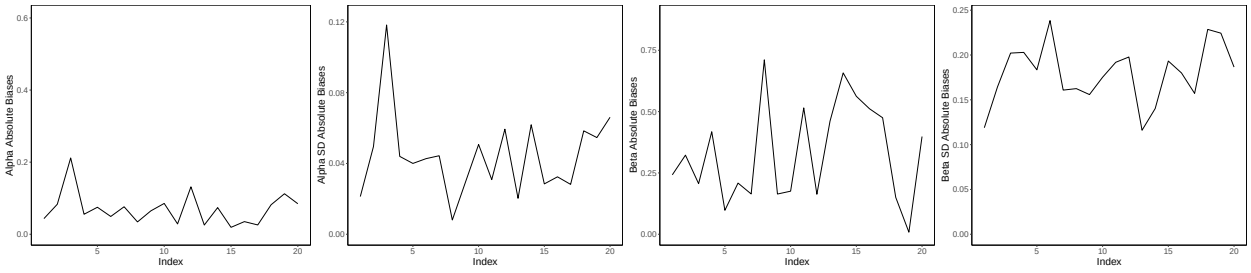
Average Center



Average ll

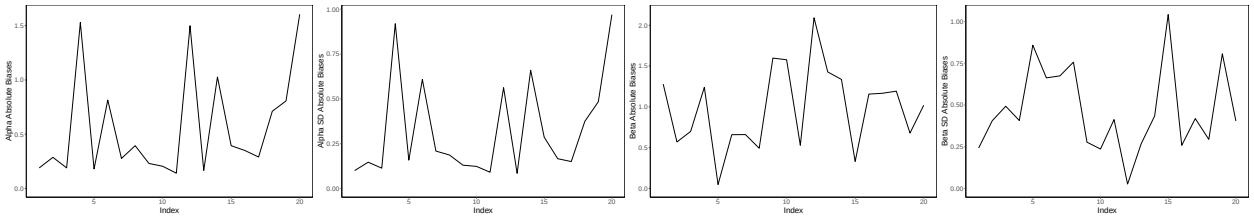


Average Corner

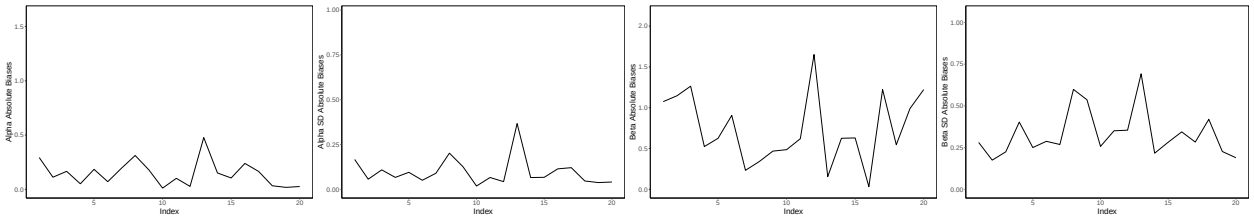


Average Random

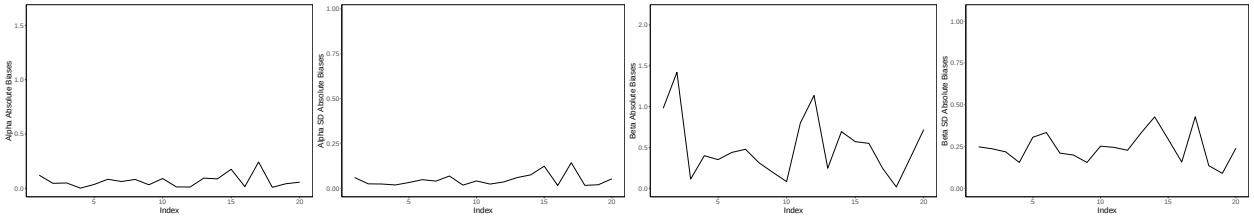
Figure A.1: Average Center, Average ll, Average Corner and Average Random Results of $n = 100$



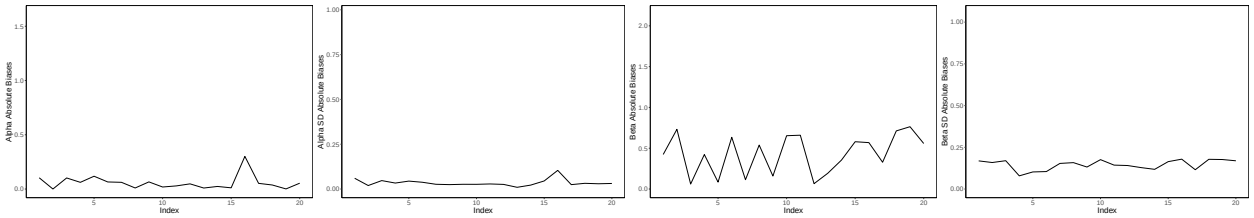
$k = 10\%$



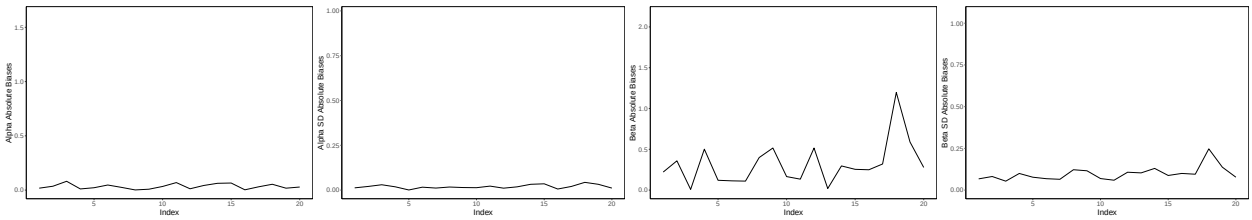
$k = 20\%$



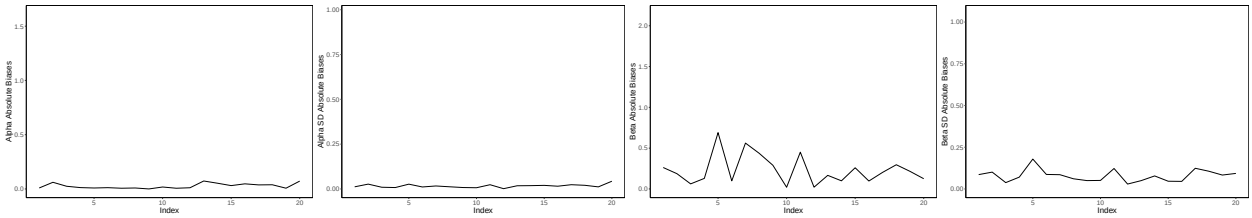
$k = 30\%$



$k = 40\%$

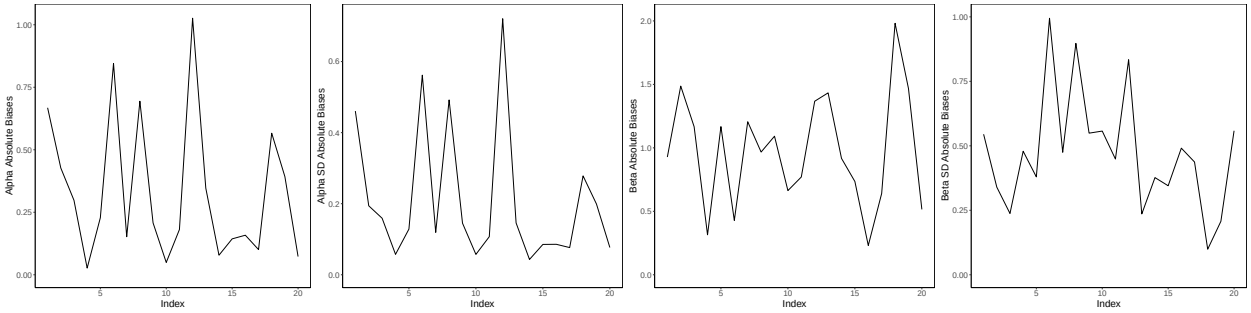


$k = 50\%$

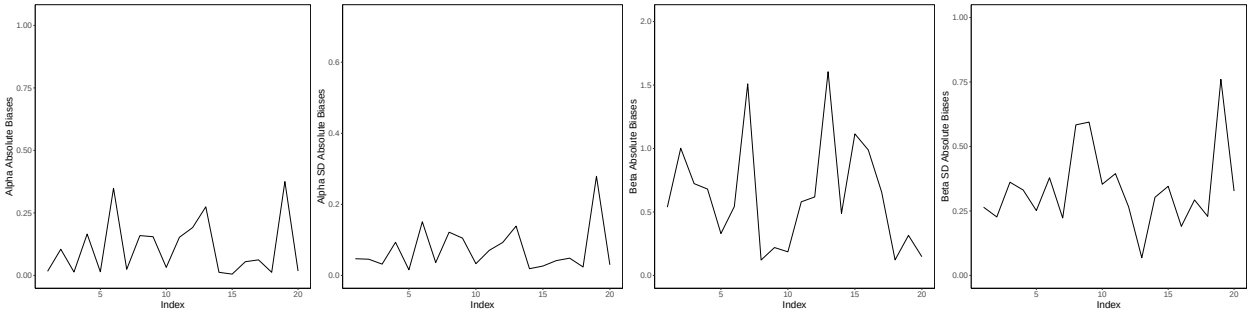


$k = 64\%$

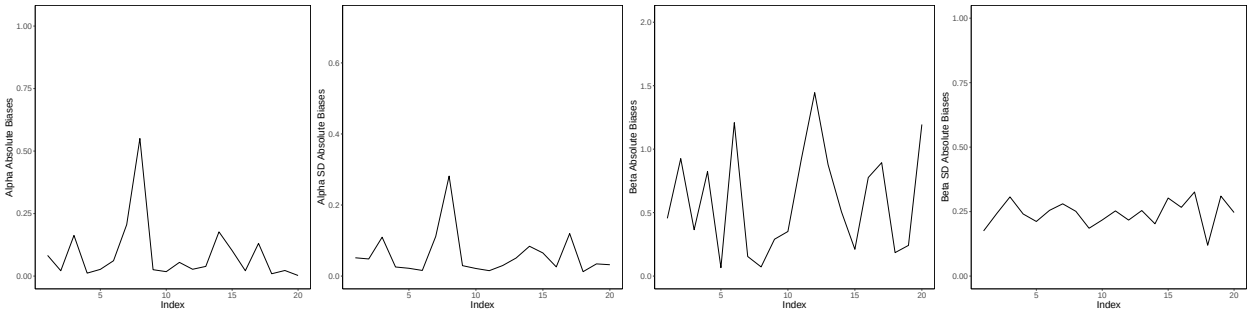
Figure A.2: $k-$Percent Center Results for n=100



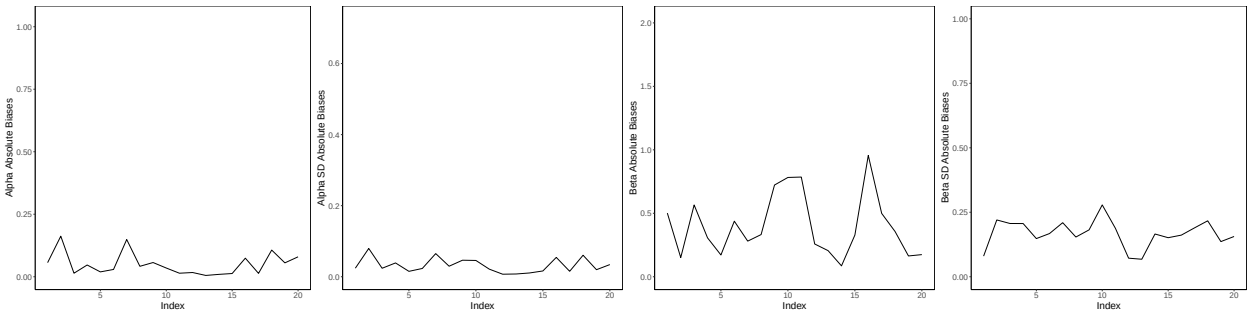
$k = 10\%$



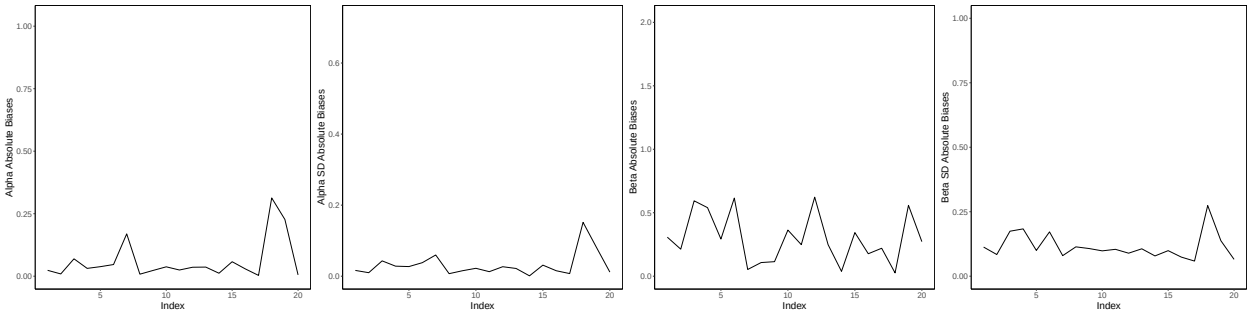
$k = 20\%$



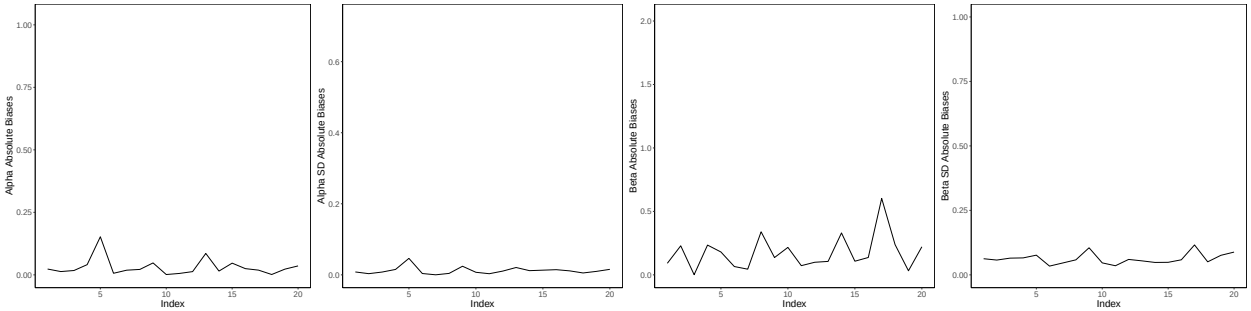
$k = 30\%$



$k = 40\%$

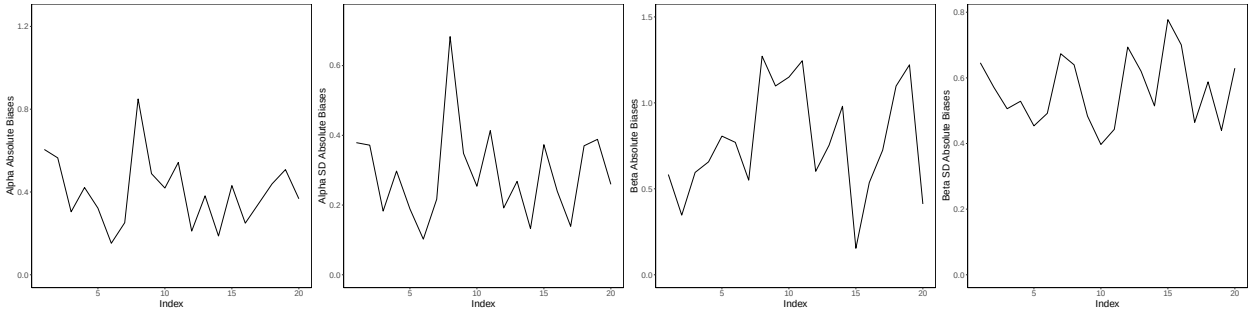


$k = 50\%$

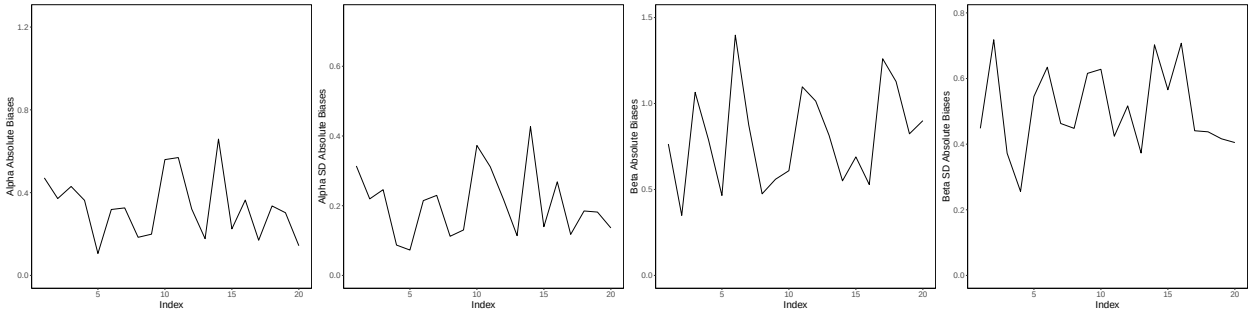


$k = 64\%$

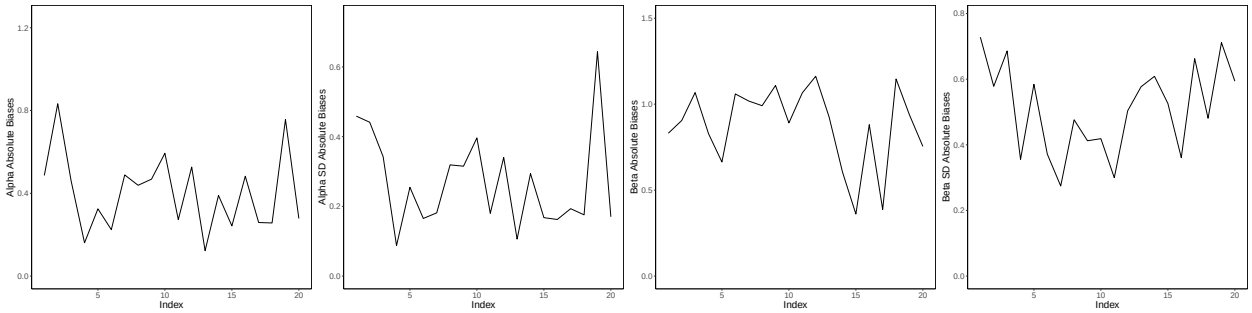
Figure A.3: $k-$Percent ll Results for n=100



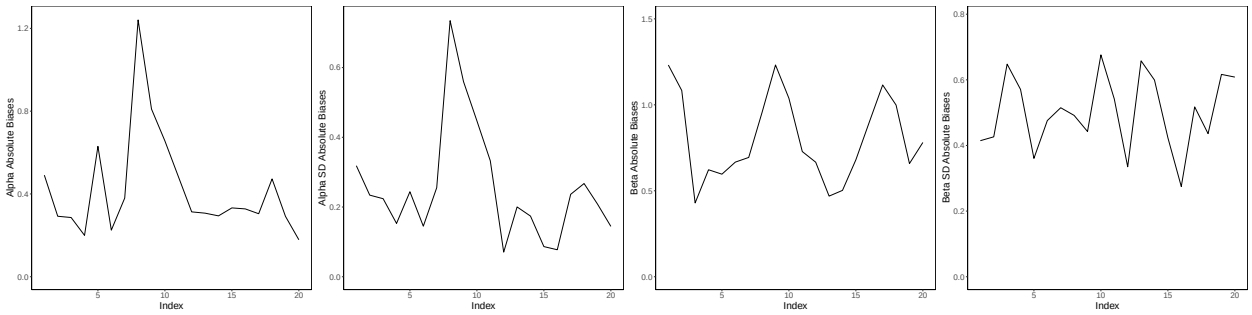
$k = 10\%$



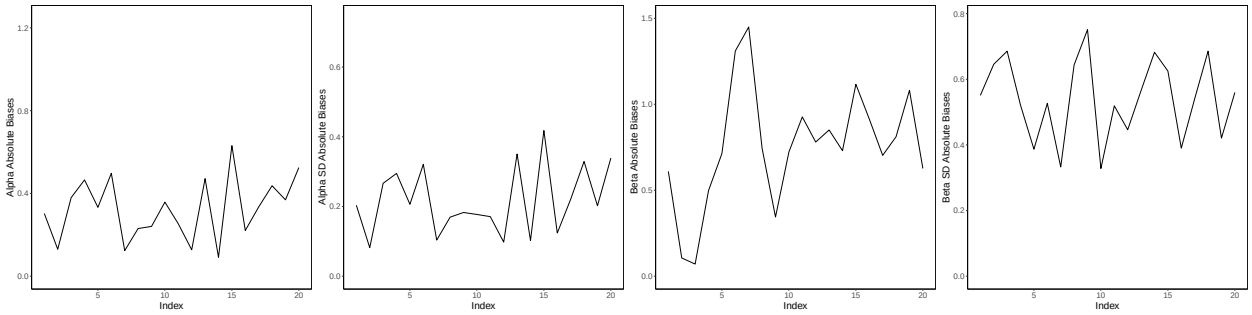
$k = 20\%$



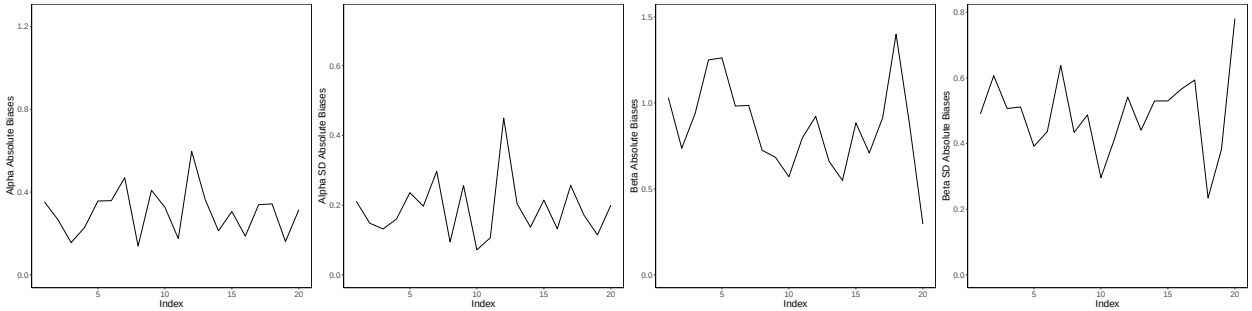
$k = 30\%$



$k = 40\%$

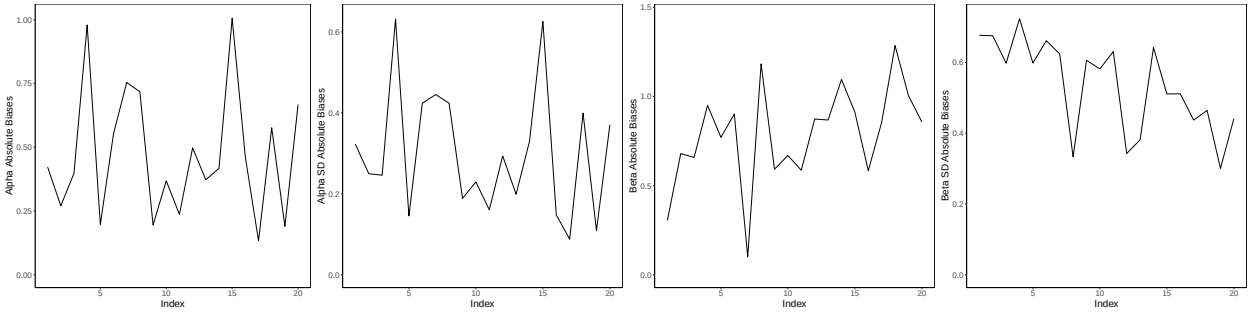


$k = 50\%$

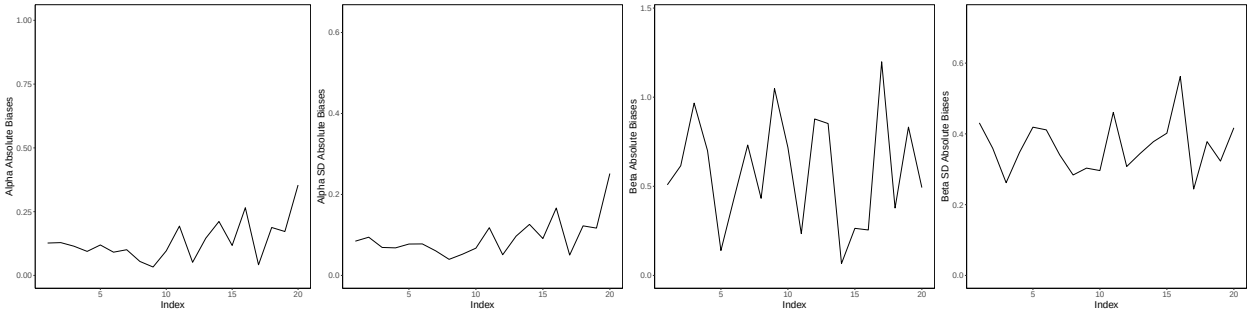


$k = 64\%$

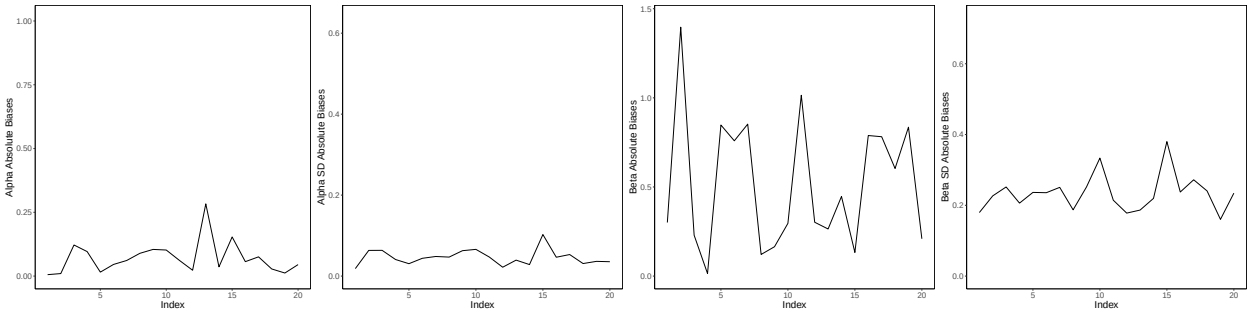
Figure A.4: $k-$Percent Corner Results for n=100



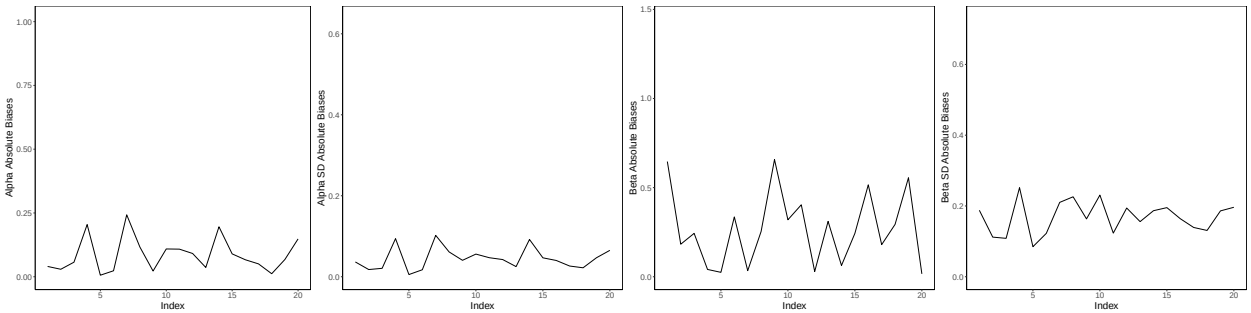
$k = 10\%$



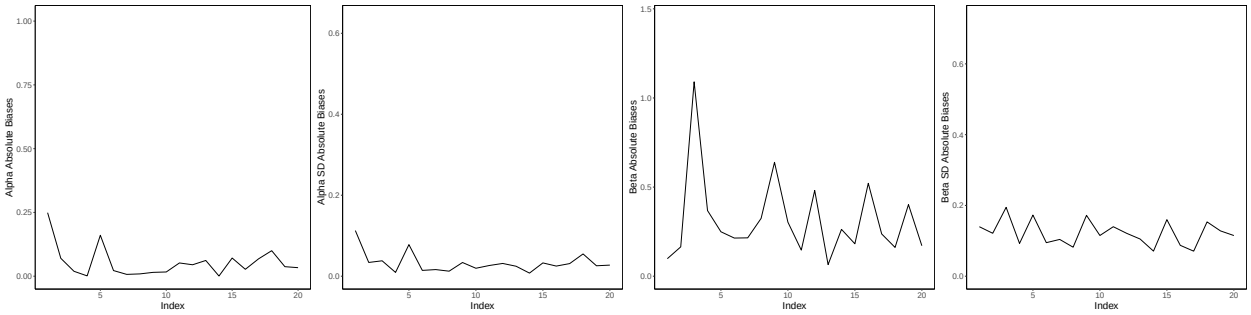
$k = 20\%$



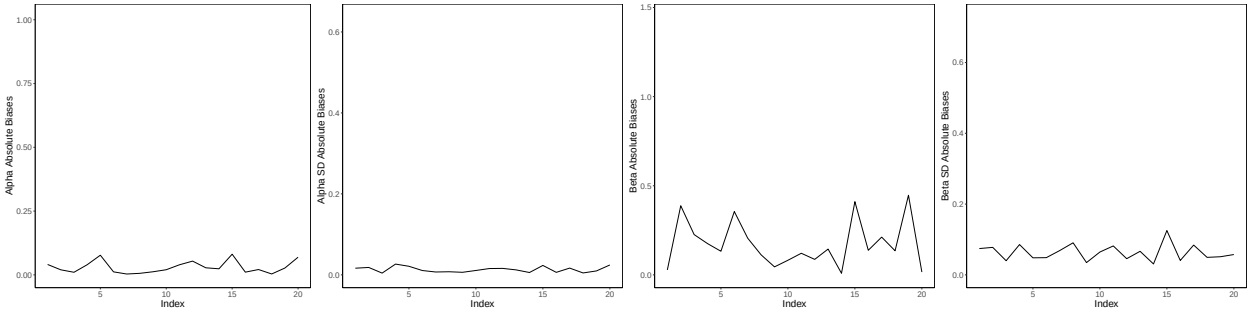
$k = 30\%$



$k = 40\%$

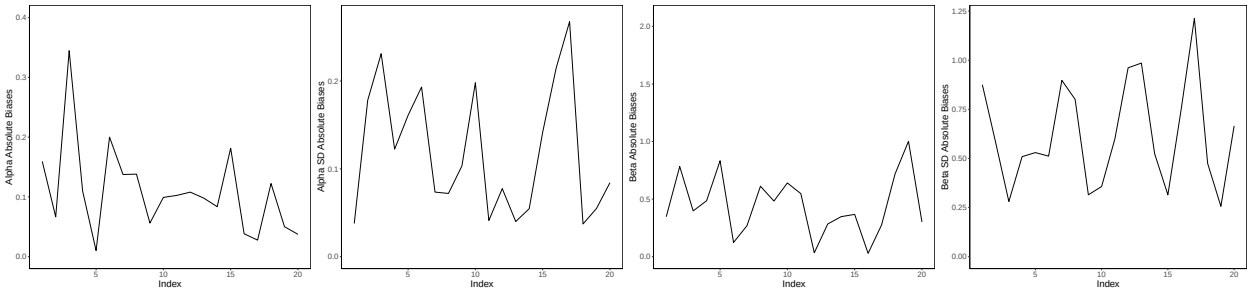


$k = 50\%$

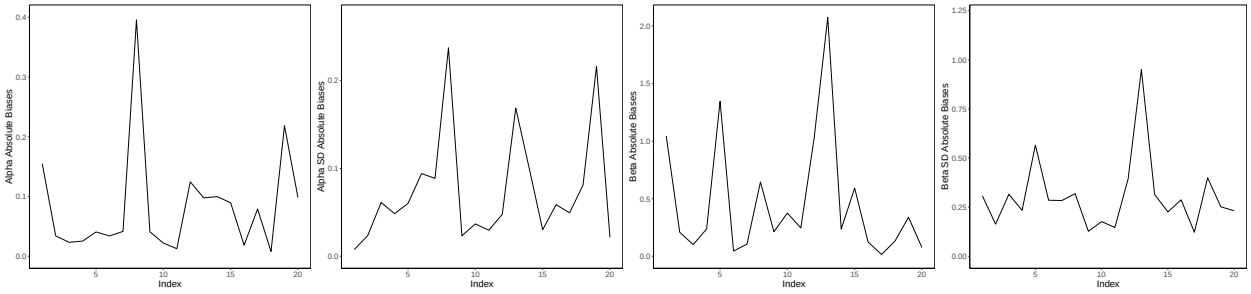


$k = 64\%$

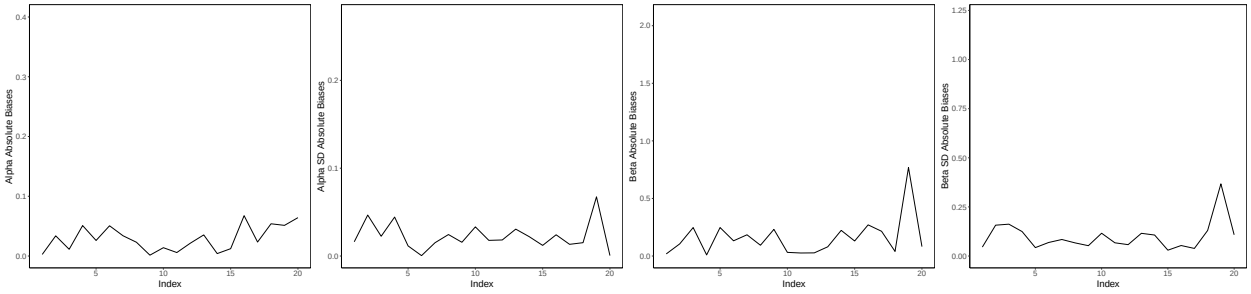
Figure A.5: $k-$Percent Random Results for n=100


Alpha Absolute Biases
Alpha SD Absolute Biases
Beta Absolute Biases
Beta SD Absolute Biases
Index

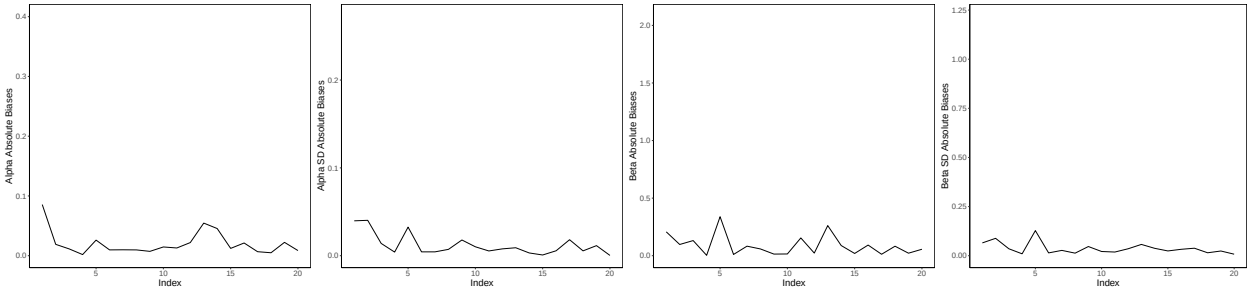
$m = 2$


Alpha Absolute Biases
Alpha SD Absolute Biases
Beta Absolute Biases
Beta SD Absolute Biases
Index

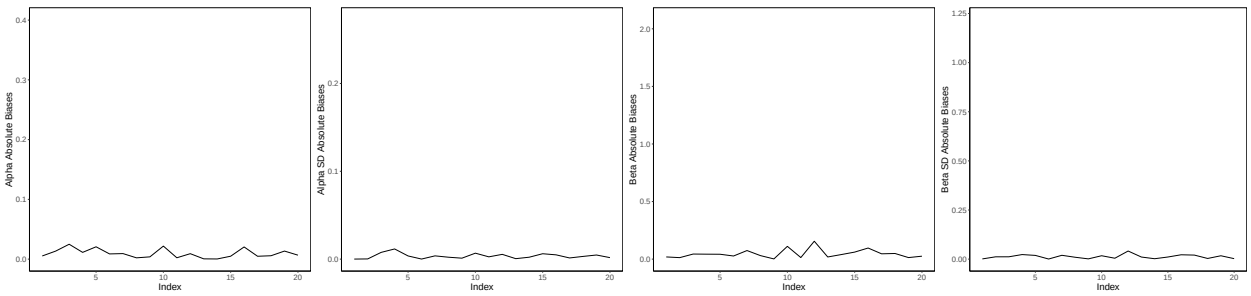
$m = 3$


Alpha Absolute Biases
Alpha SD Absolute Biases
Beta Absolute Biases
Beta SD Absolute Biases
Index

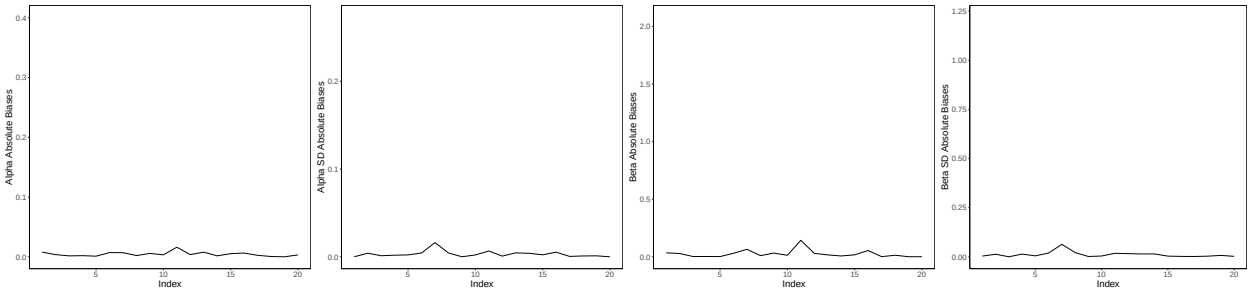
$m = 4$


Alpha Absolute Biases
Alpha SD Absolute Biases
Beta Absolute Biases
Beta SD Absolute Biases
Index

$m = 5$



$m = 6$



$m = 7$

Figure A.6: Temporal 1 Subset-m Results for n=100

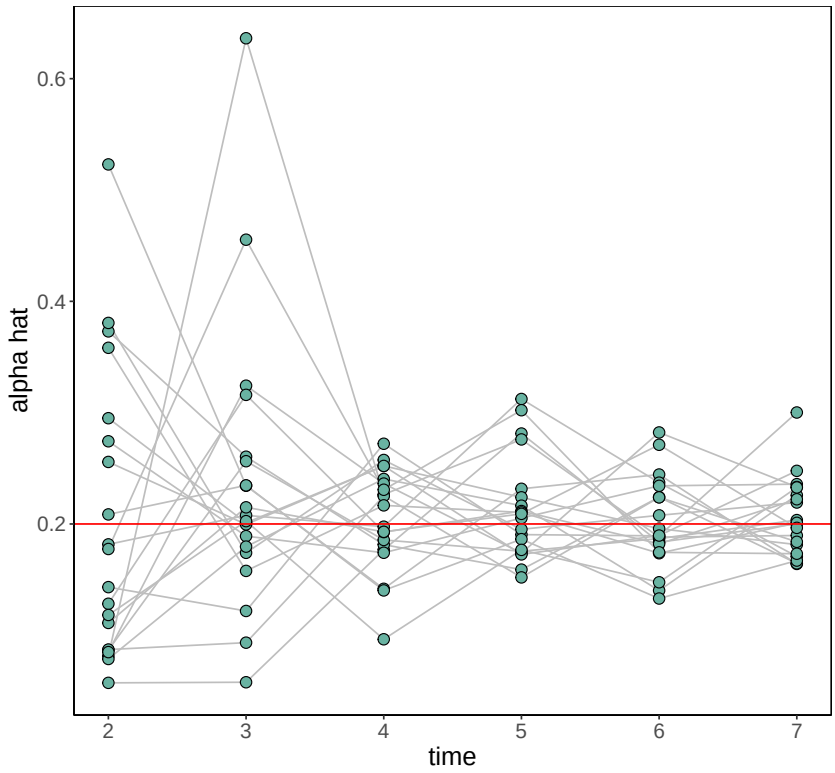


Figure A.7: Temporal 1 Subset-m $\hat{\alpha}$ vs α for n=100

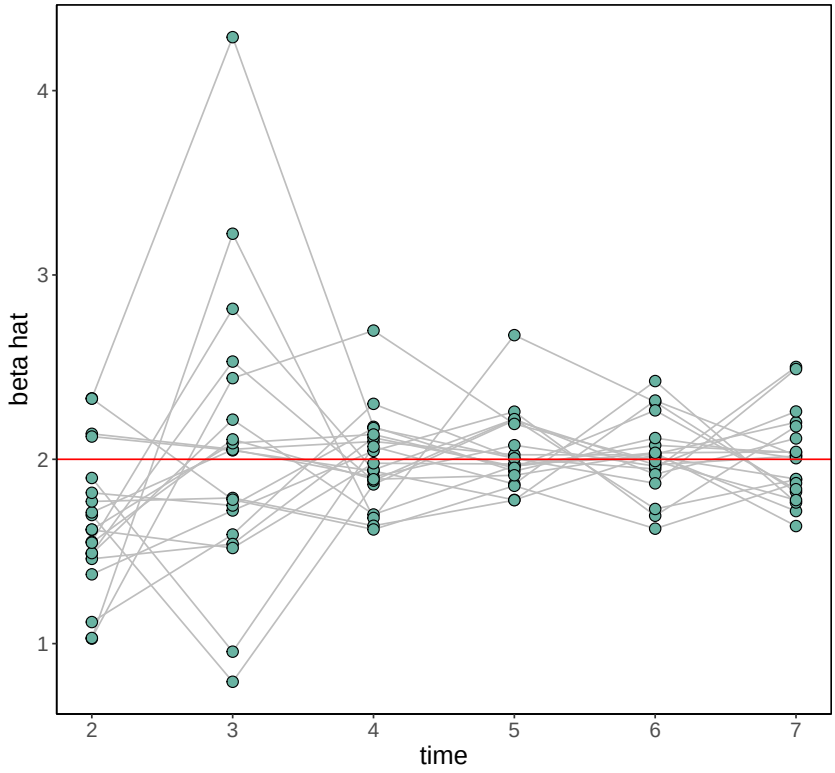
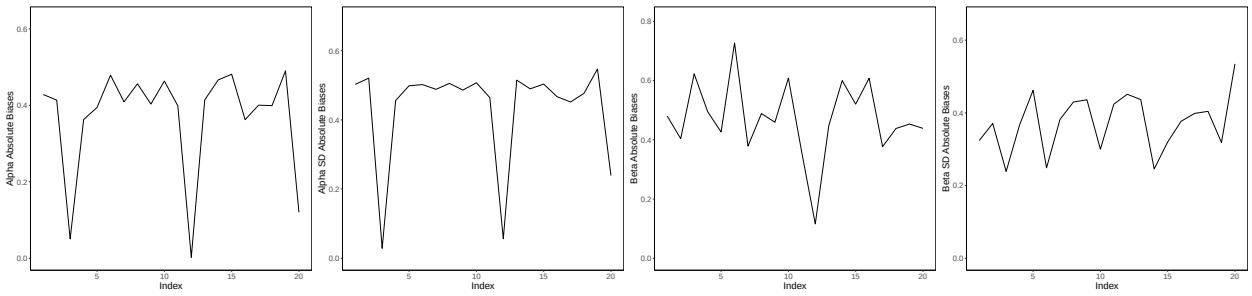
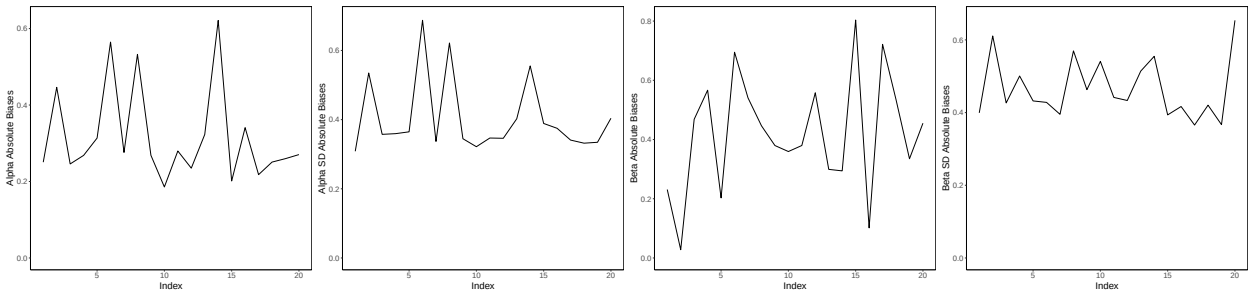


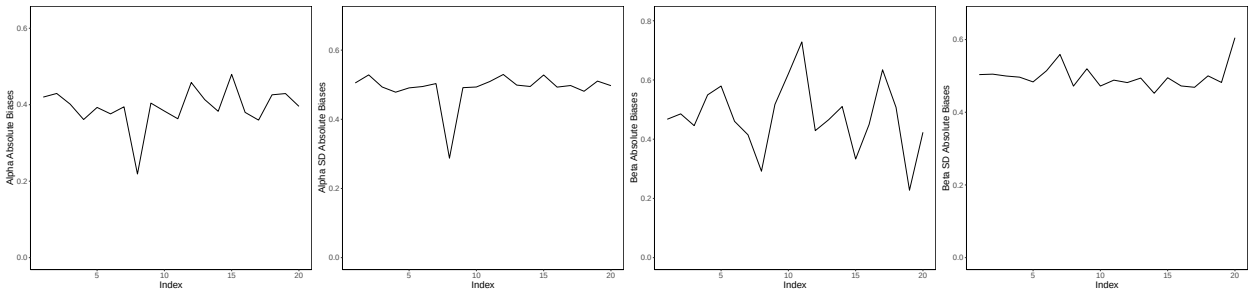
Figure A.8: Temporal 1 Subset-m $\hat{\beta}$ vs β for n=100



$h = 2$

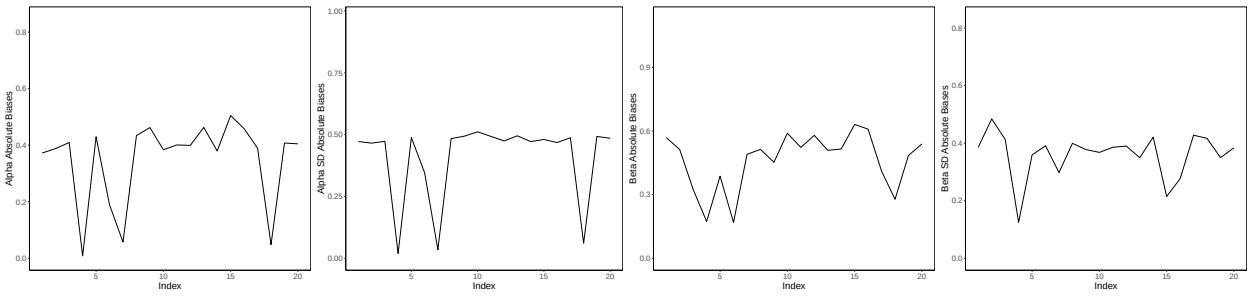


$h = 3$

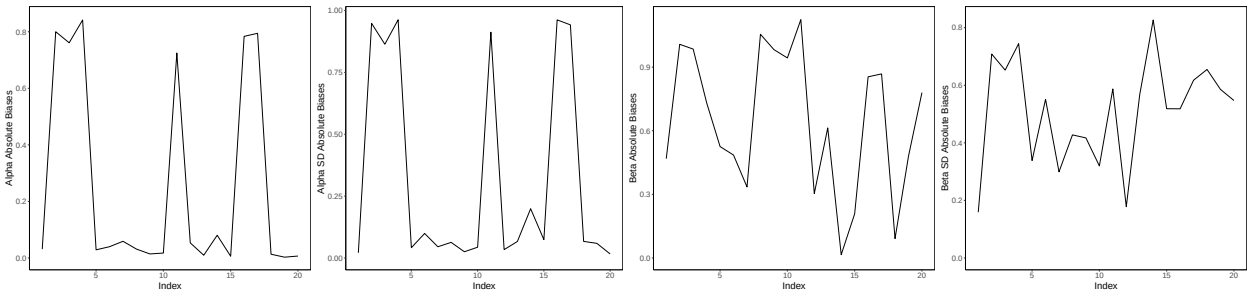


$h = 4$

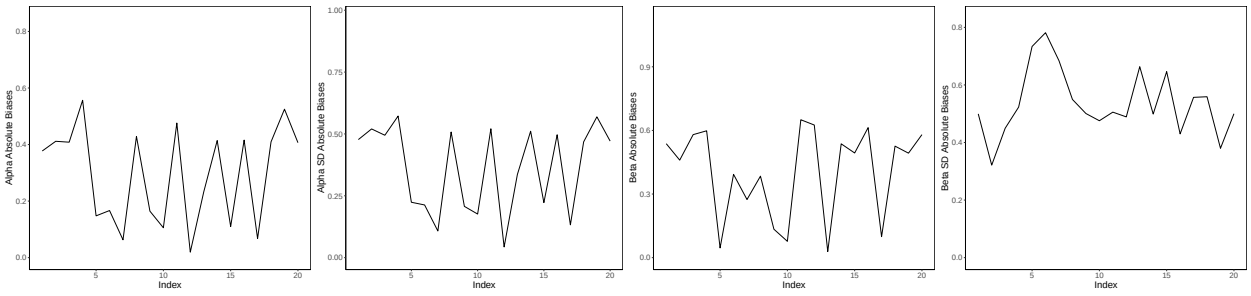
Figure A.9: Temporal Mean Subset-h results for n=100



$h = 2$

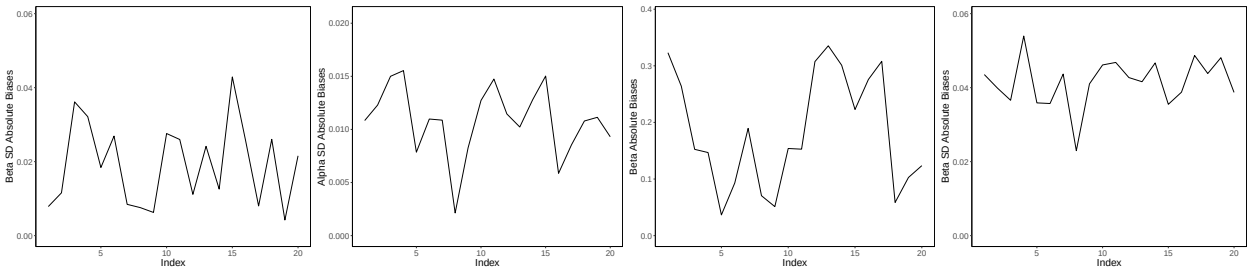


$h = 3$

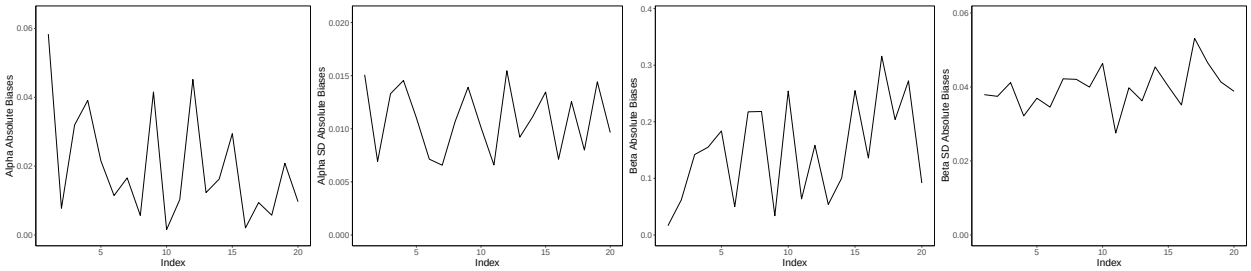


$h = 4$

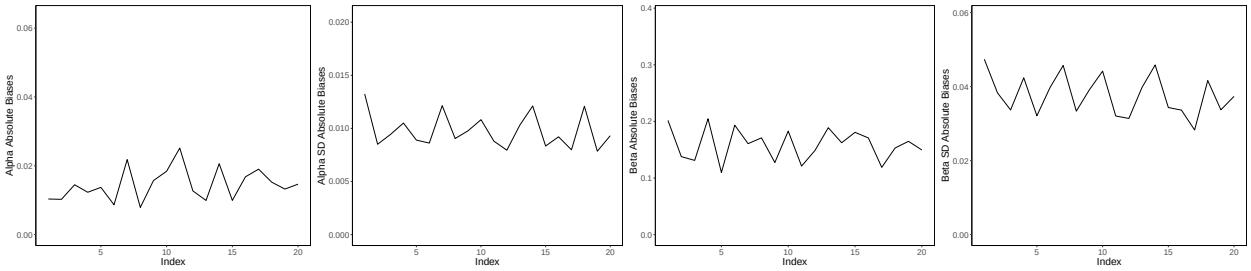
Figure A.10: Temporal Median Subset-h results for n=100



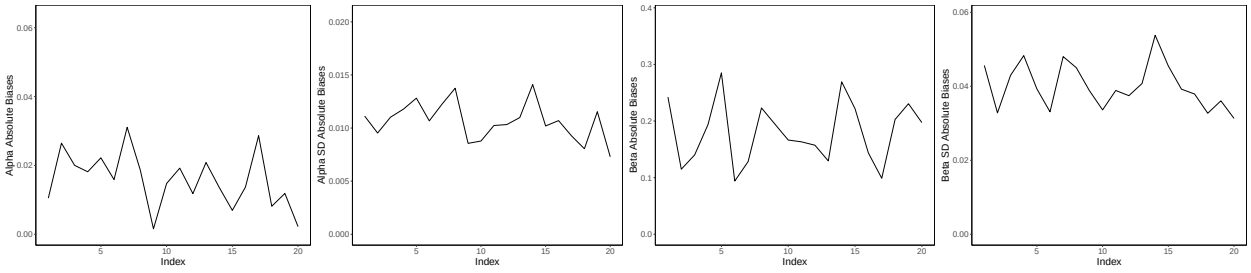
Average Center



Average ll

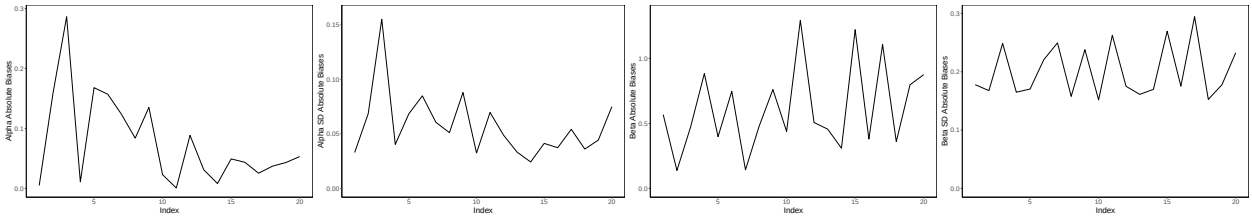


Average Corner

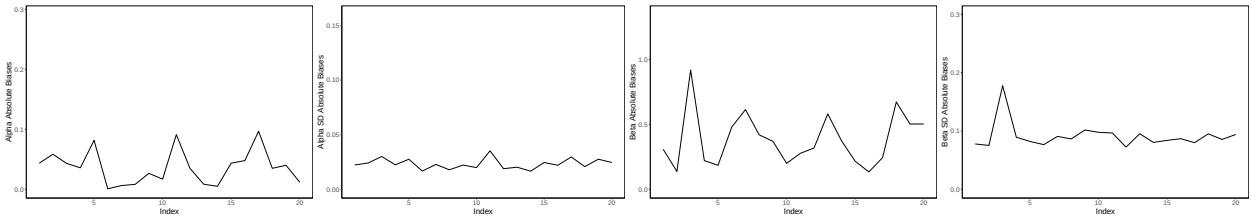


Average Random

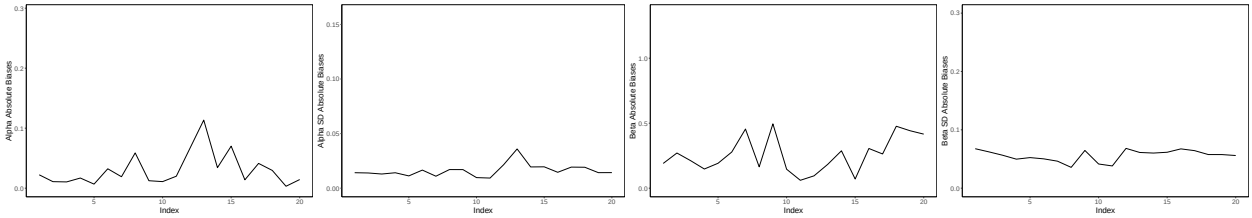
Figure A.11: Average Center, Average ll, Average Corner and Average Random Results of $n = 1000$



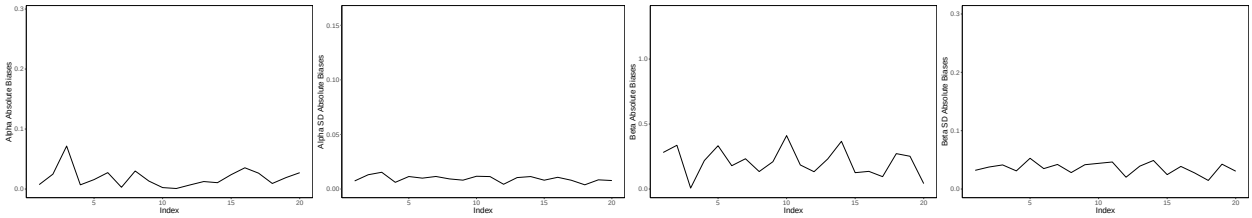
$k = 10\%$



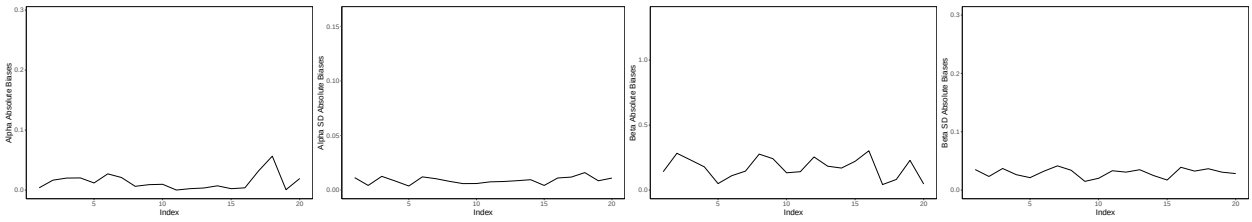
$k = 20\%$



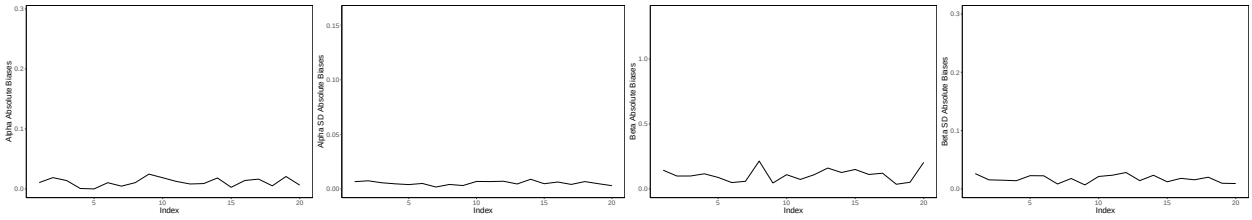
$k = 30\%$



$k = 40\%$

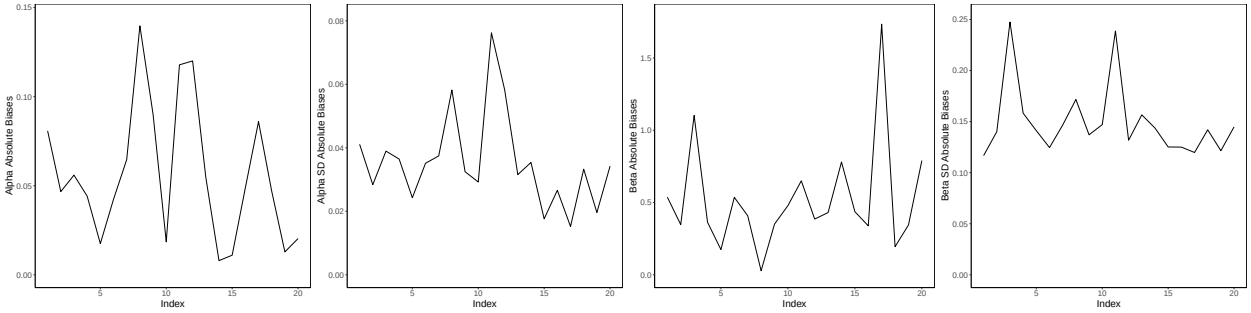


$k = 50\%$

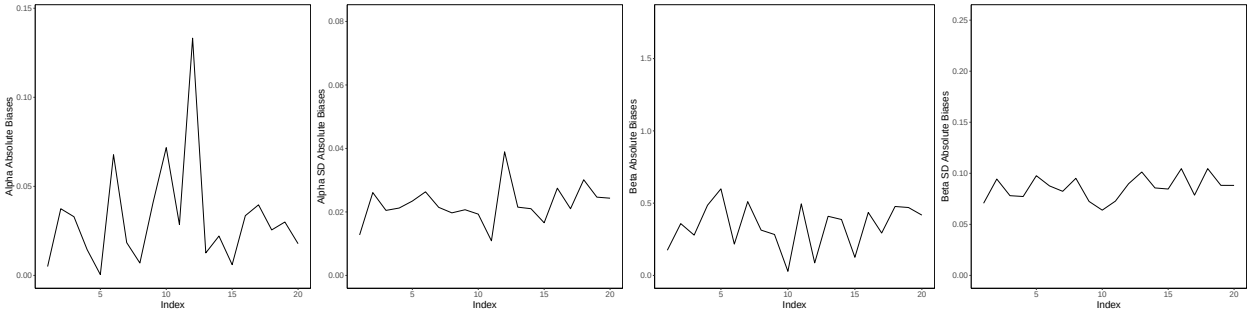


$k = 64\%$

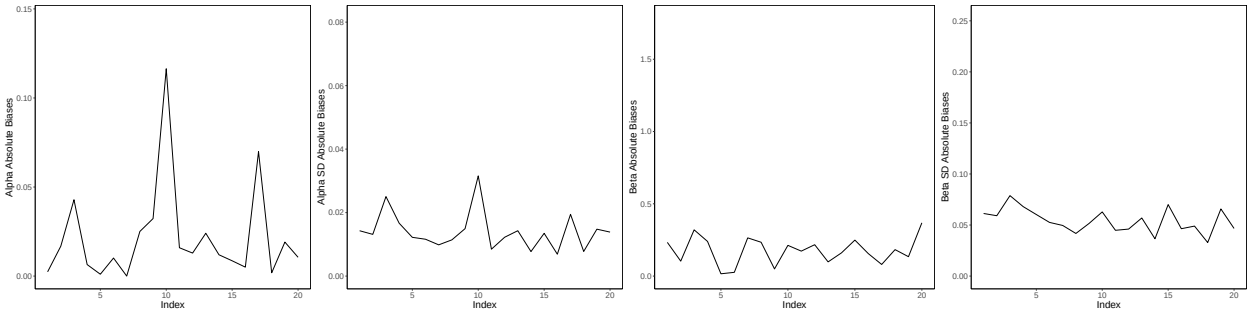
Figure A.12: $k-$Percent Center Results for n=1000



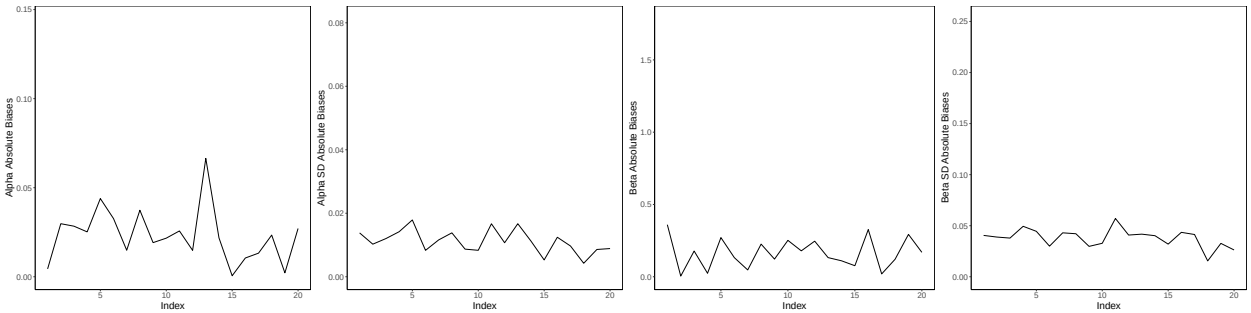
$k = 10\%$



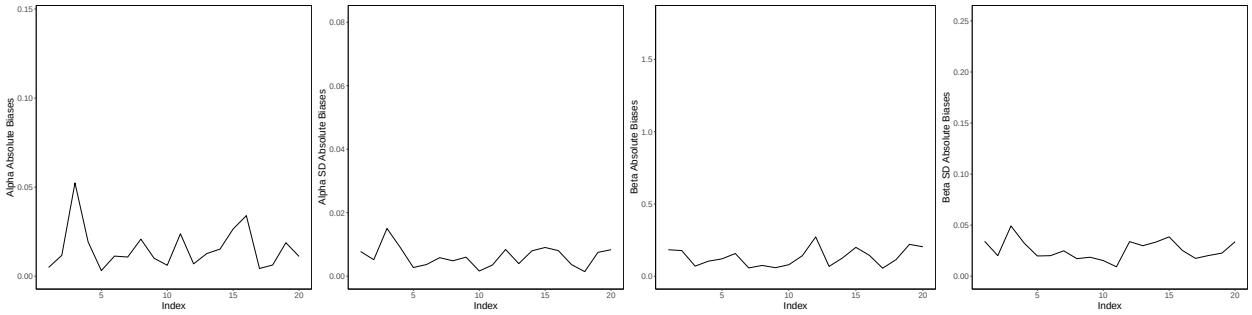
$k = 20\%$



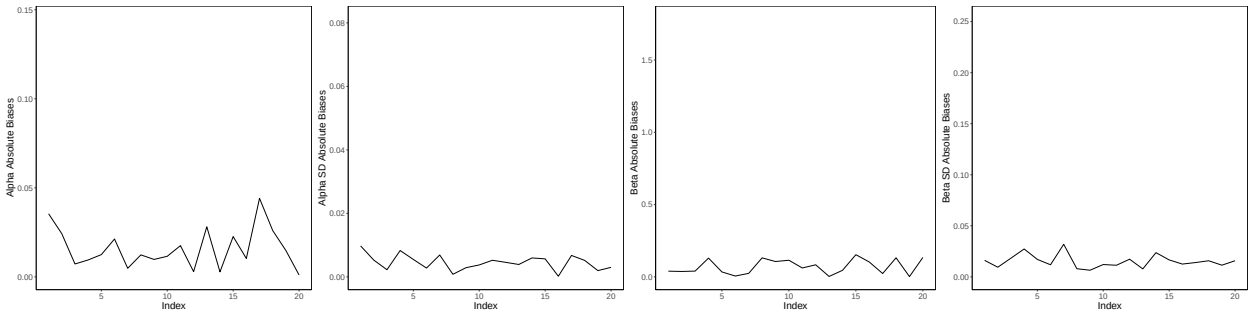
$k = 30\%$



$k = 40\%$

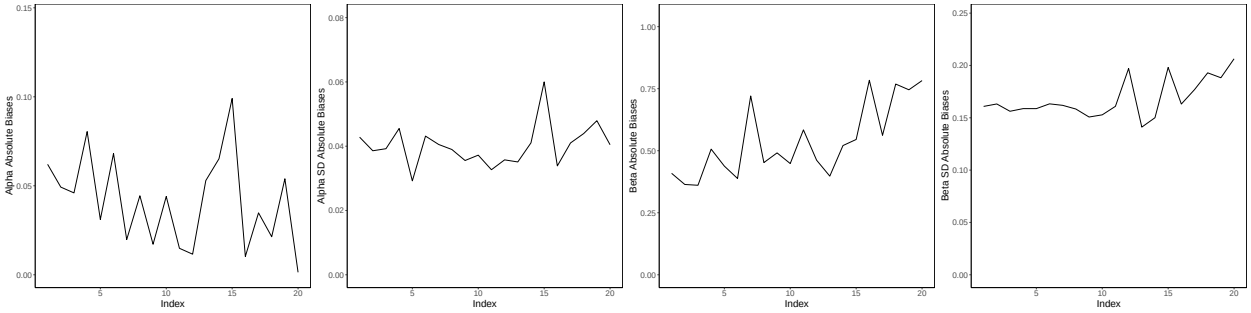


$k = 50\%$

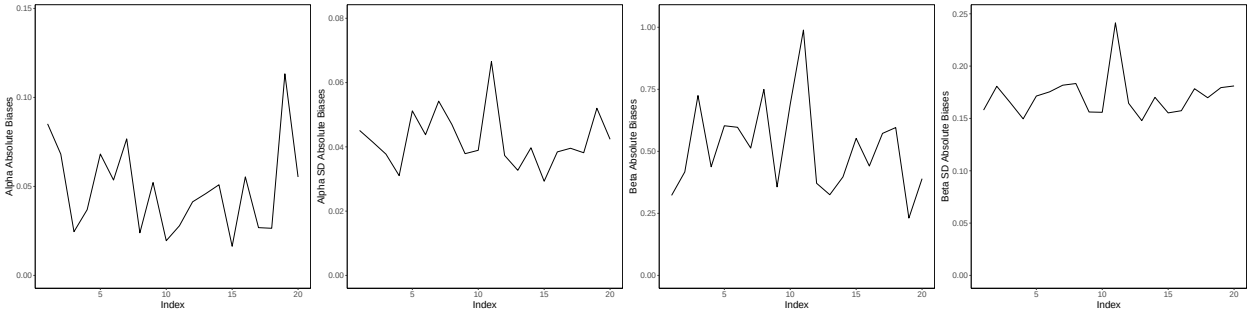


$k = 64\%$

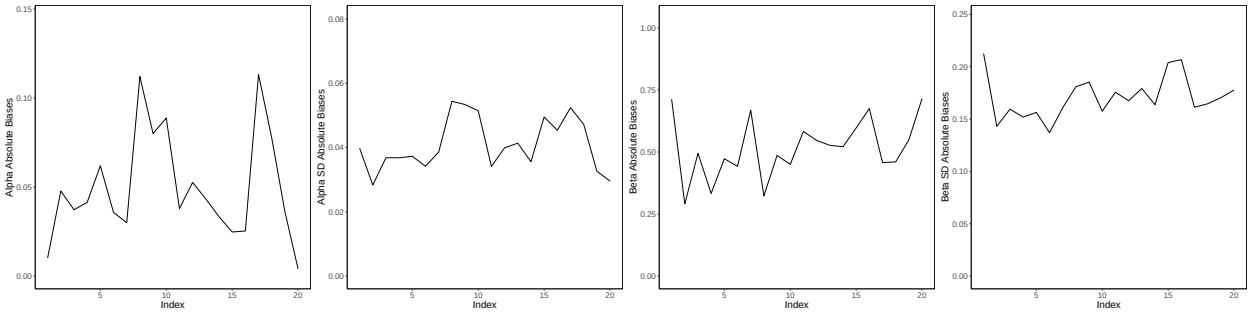
Figure A.13: $k-$Percent ll Results for n=1000



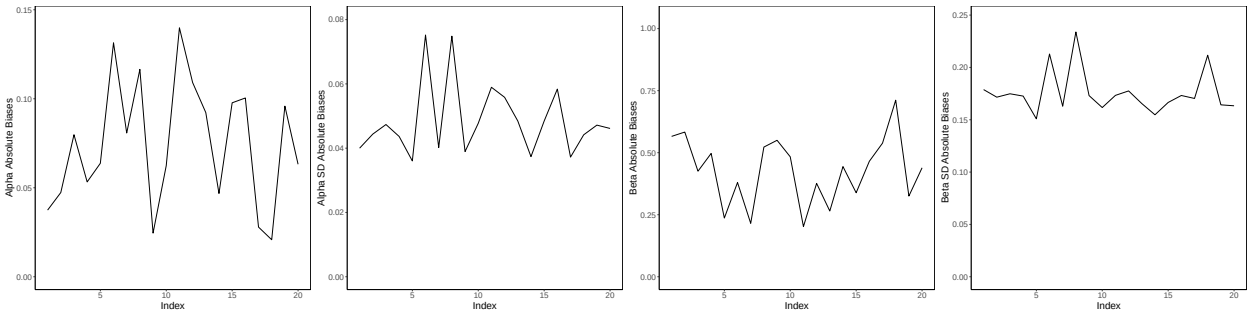
$k = 10\%$



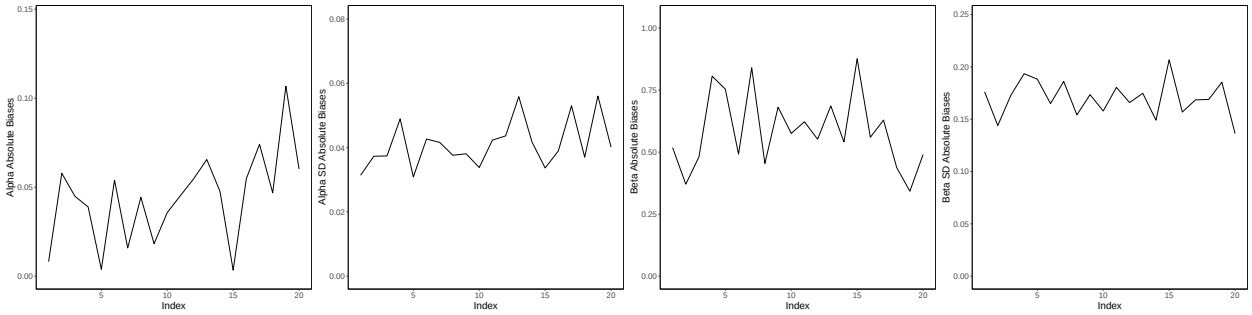
$k = 20\%$



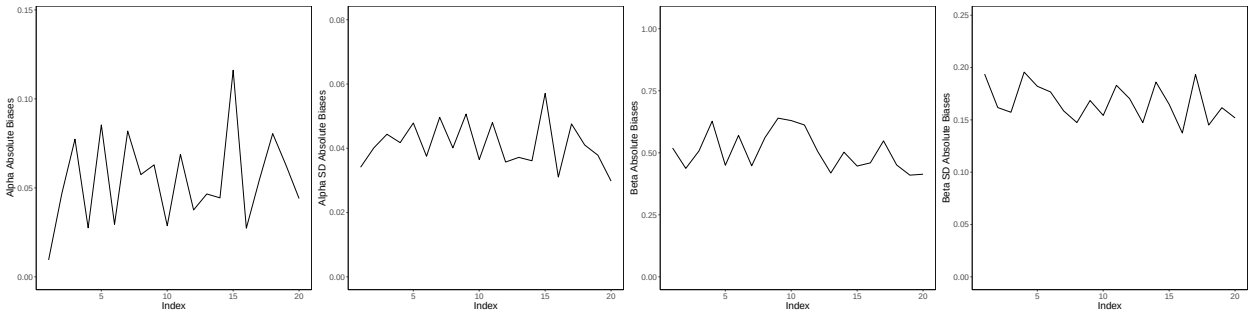
$k = 30\%$



$k = 40\%$

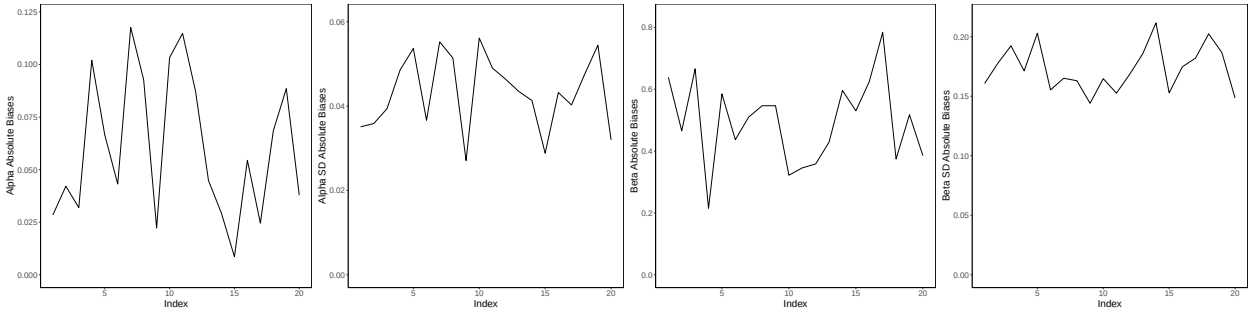


$k = 50\%$

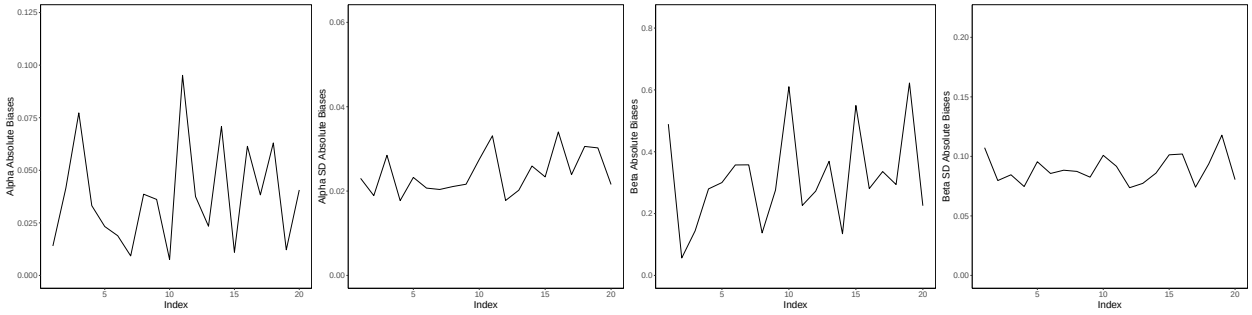


$k = 64\%$

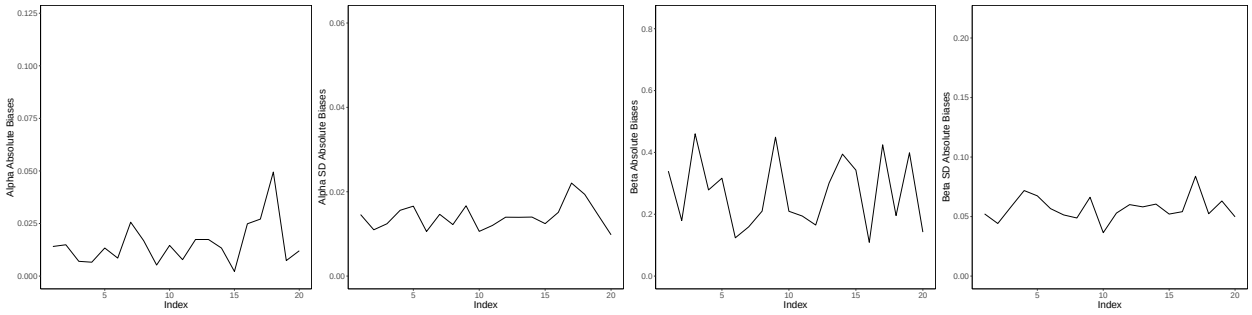
Figure A.14: $k-$Percent Corner Results for n=1000



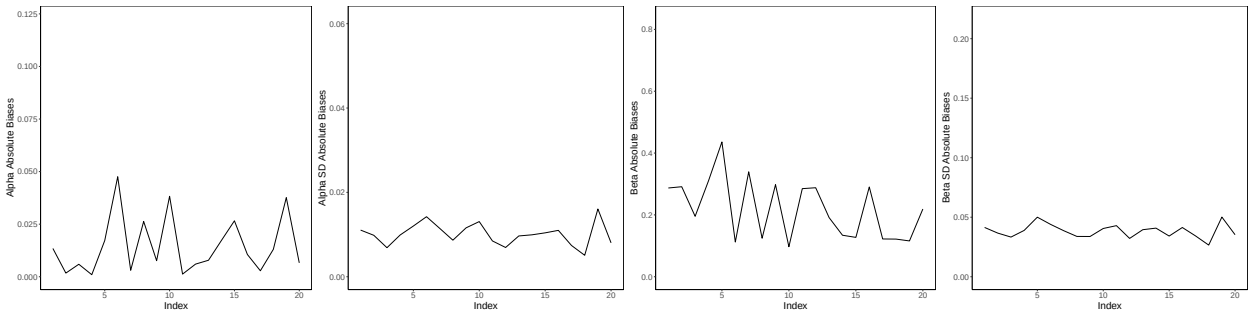
$k = 10\%$



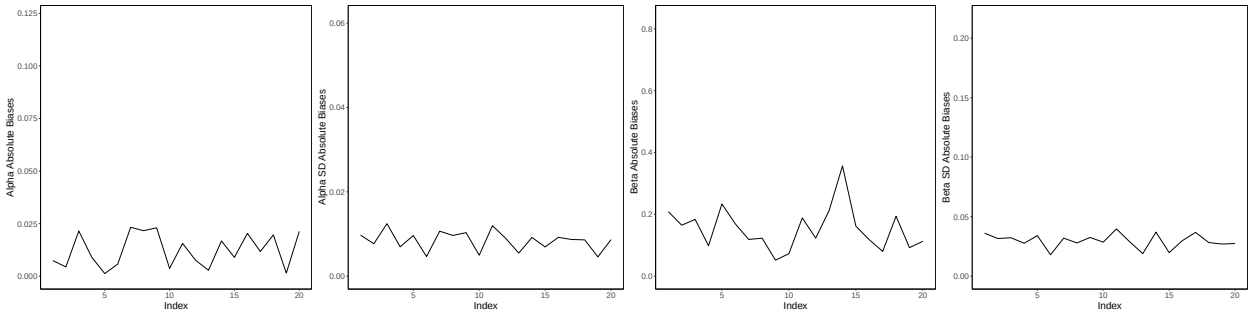
$k = 20\%$



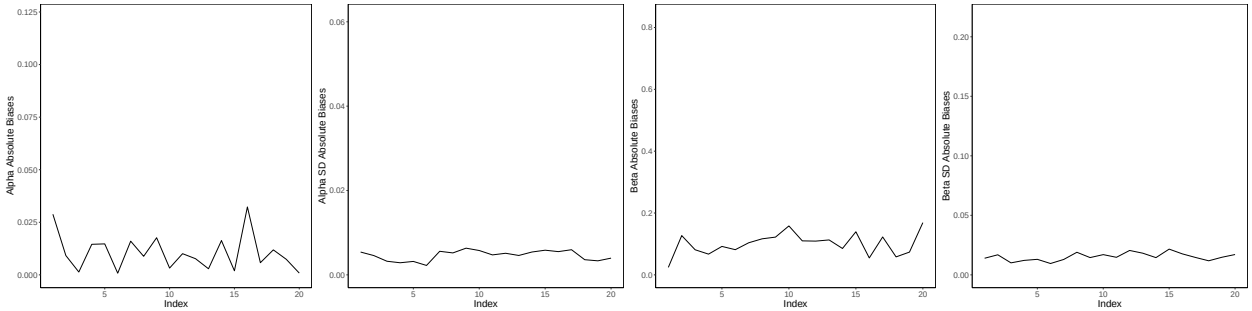
$k = 30\%$



$k = 40\%$



$k = 50\%$



$k = 64\%$

Figure A.15: $k-$Percent Random Results for n=1000

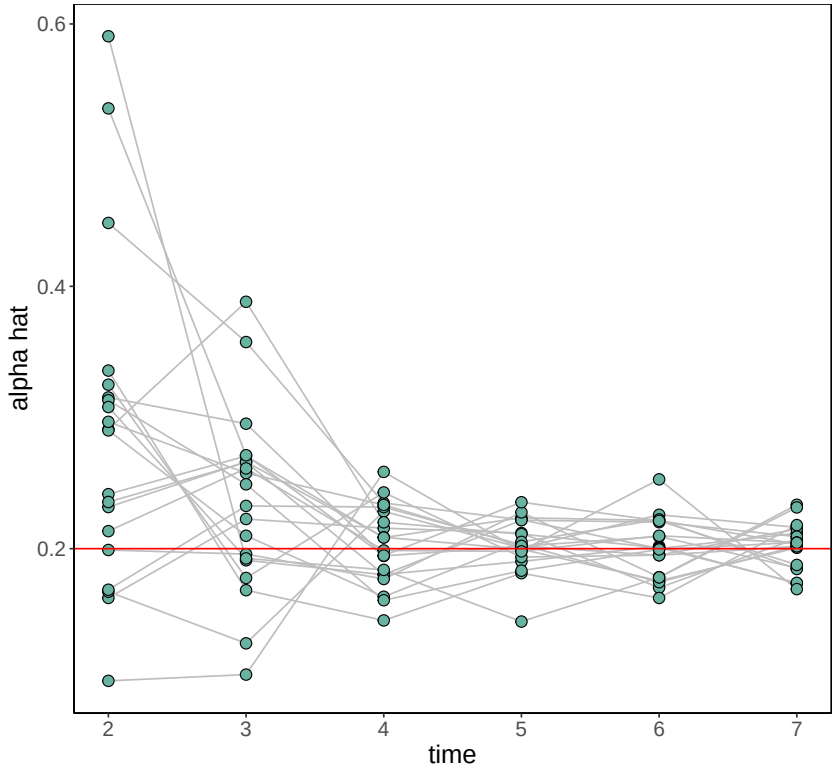


Figure A.16: Temporal 1 Subset-m $\hat{\alpha}$ vs α for n=1000

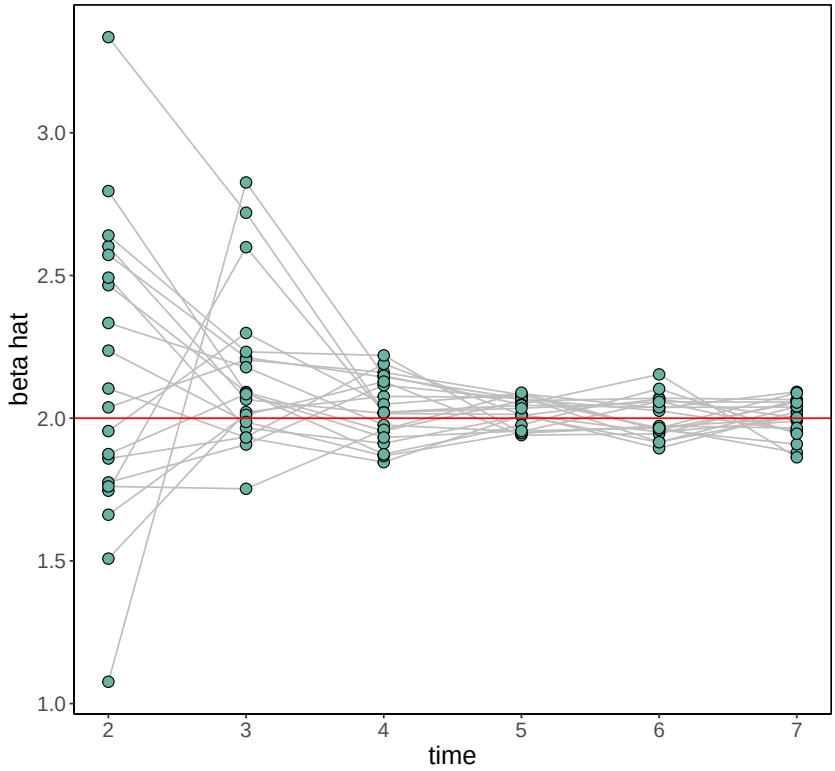
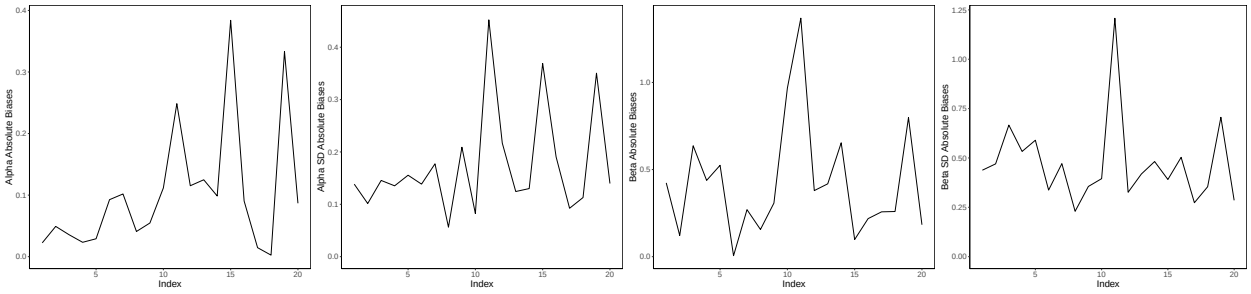
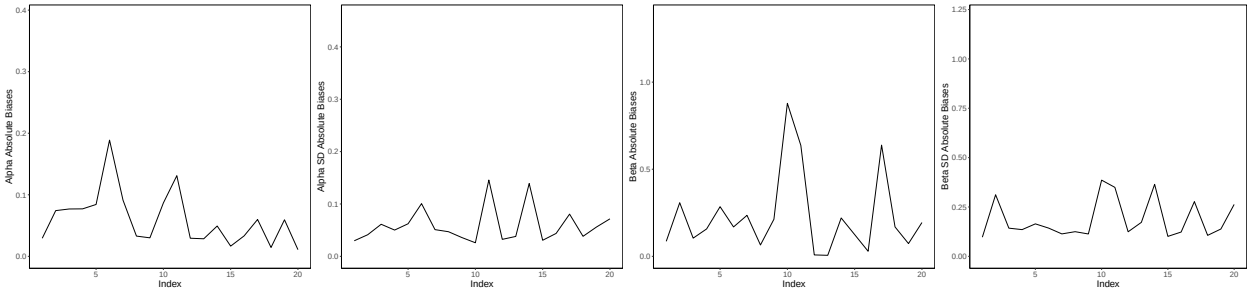


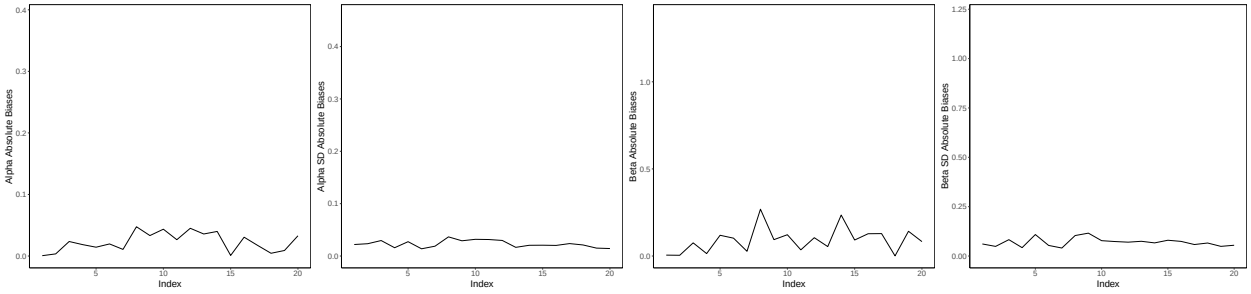
Figure A.17: Temporal 1 Subset-m $\hat{\beta}$ vs β for n=1000


Alpha Absolute Biases
Alpha SD Absolute Biases
Beta Absolute Biases
Beta SD Absolute Biases
Index

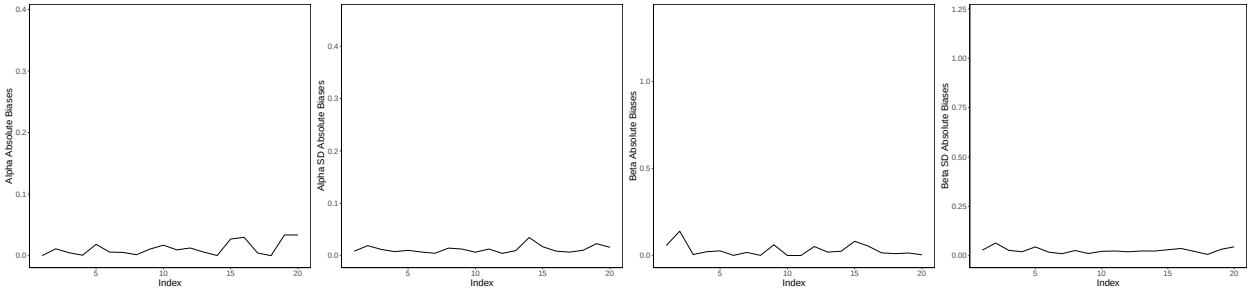
$m = 2$


Alpha Absolute Biases
Alpha SD Absolute Biases
Beta Absolute Biases
Beta SD Absolute Biases
Index

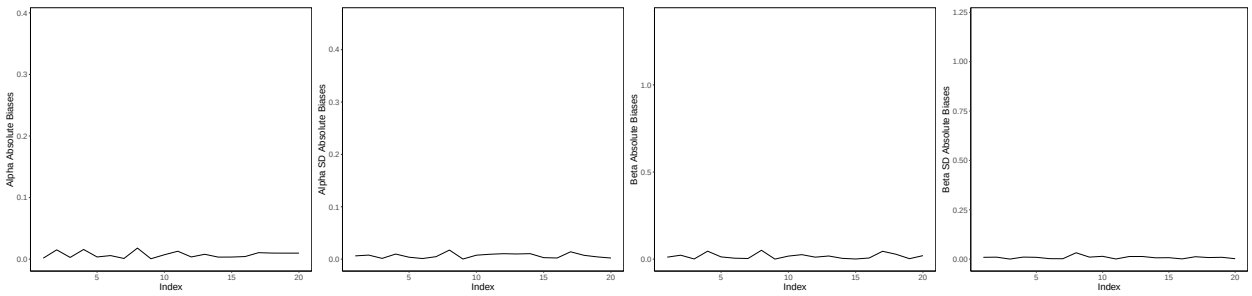
$m = 3$


Alpha Absolute Biases
Alpha SD Absolute Biases
Beta Absolute Biases
Beta SD Absolute Biases
Index

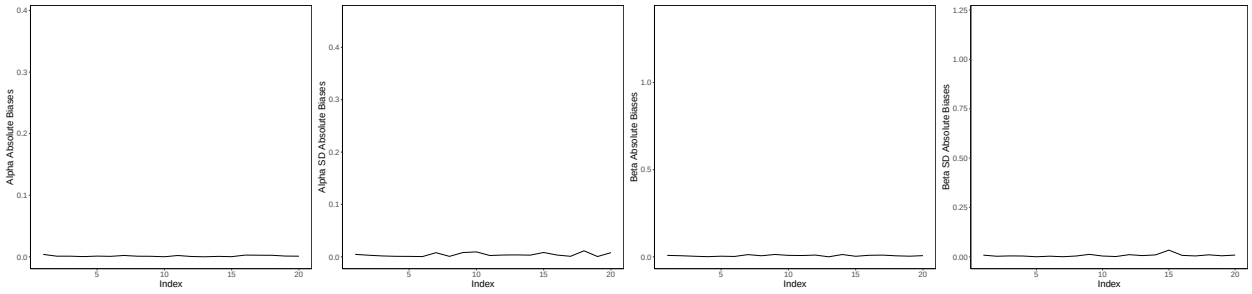
$m = 4$


Alpha Absolute Biases
Alpha SD Absolute Biases
Beta Absolute Biases
Beta SD Absolute Biases
Index

$m = 5$

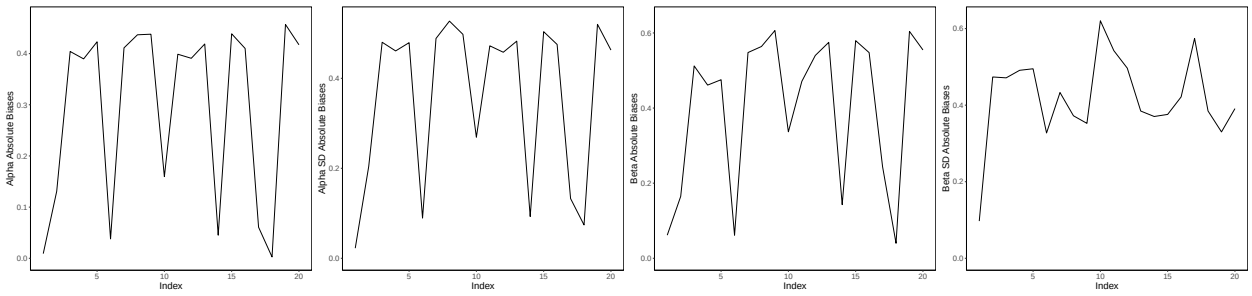


$m = 6$

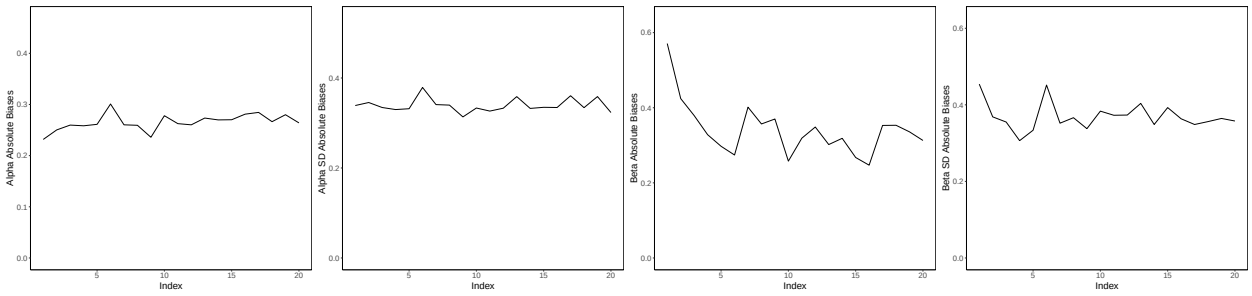


$m = 7$

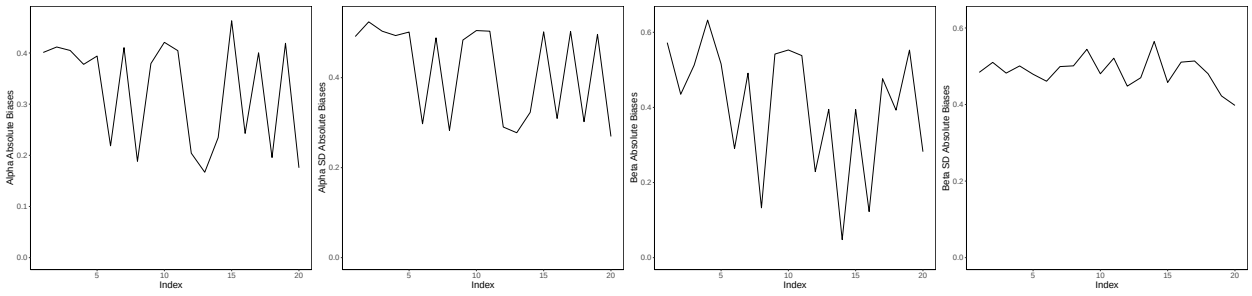
Figure A.18: Temporal 1 Subset-m Results for n=1000



$h = 2$

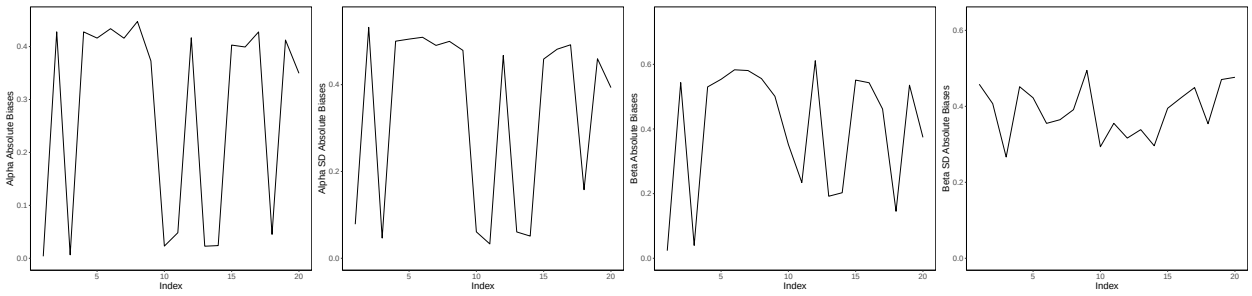


$h = 3$

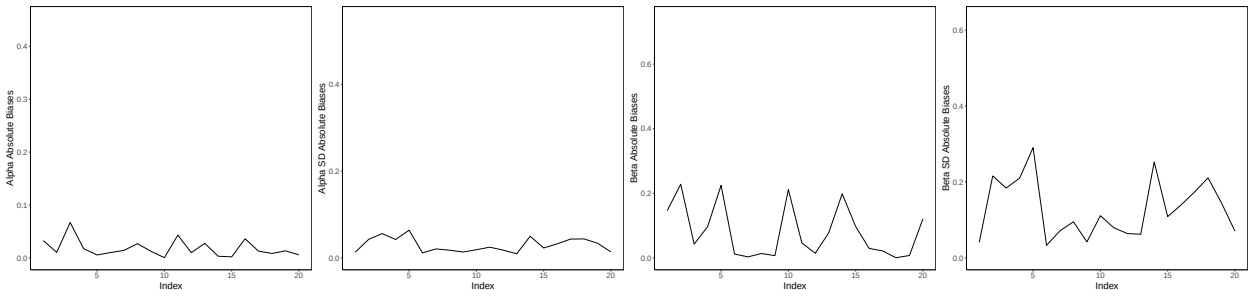


$h = 4$

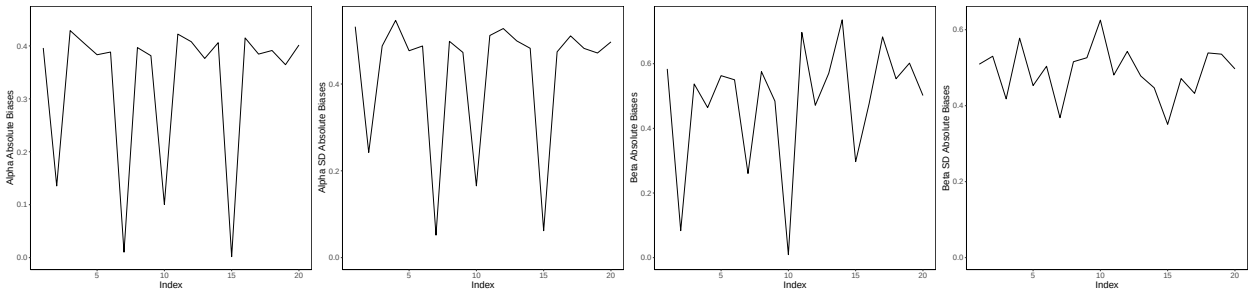
Figure A.19: Temporal Mean Subset-h results for n=1000



$h = 2$

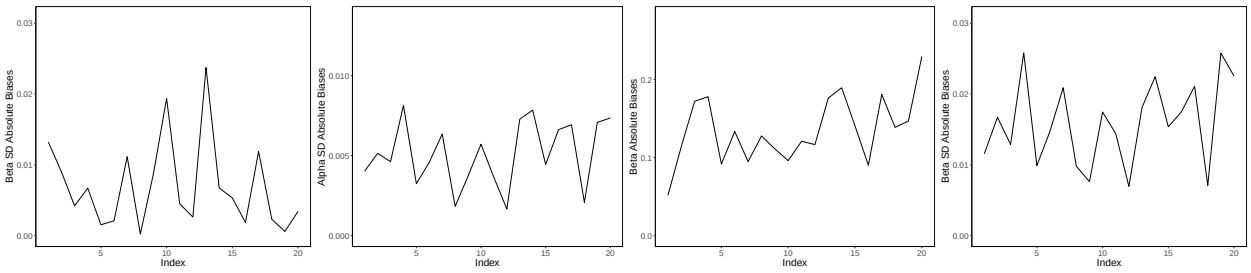


$h = 3$

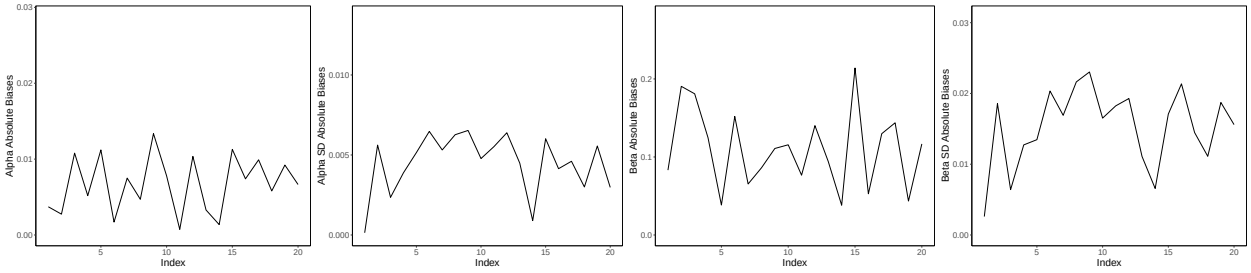


$h = 4$

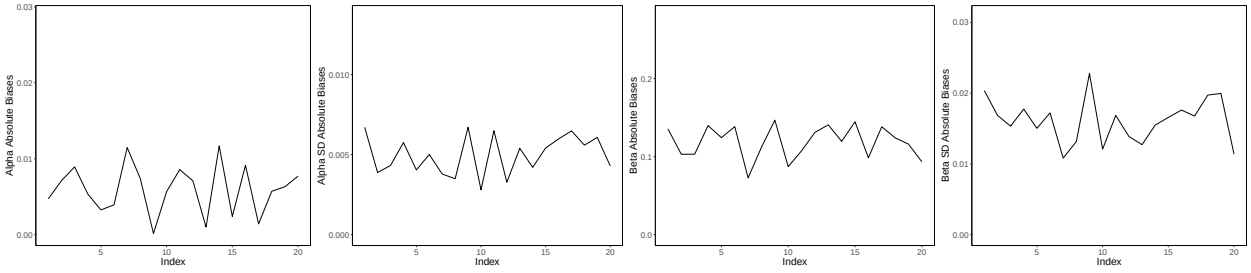
Figure A.20: Temporal Median Subset-h results for n=1000



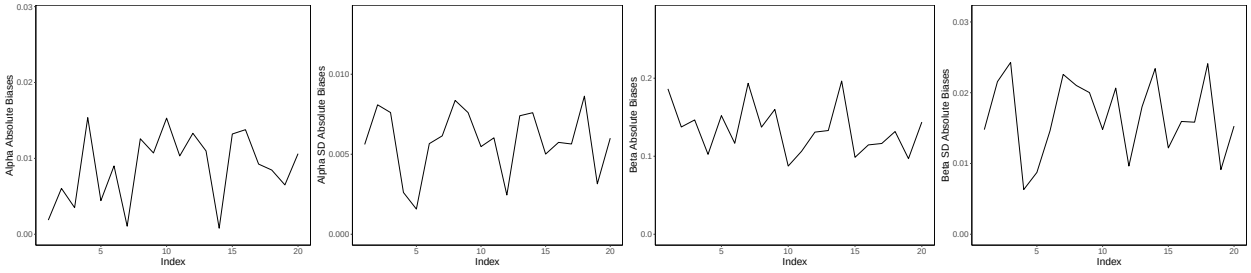
Average Center



Average ll

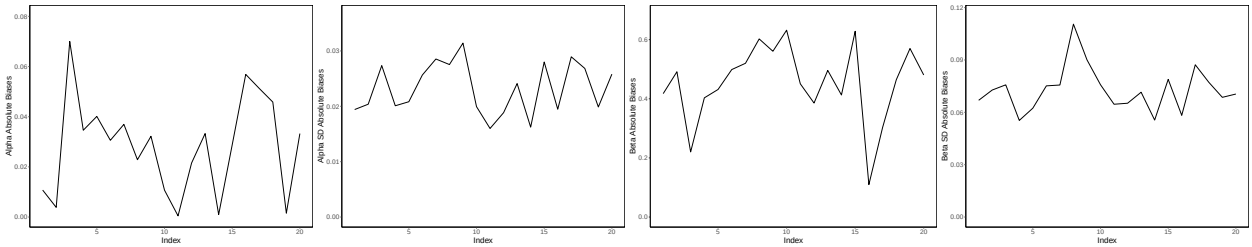


Average Corner

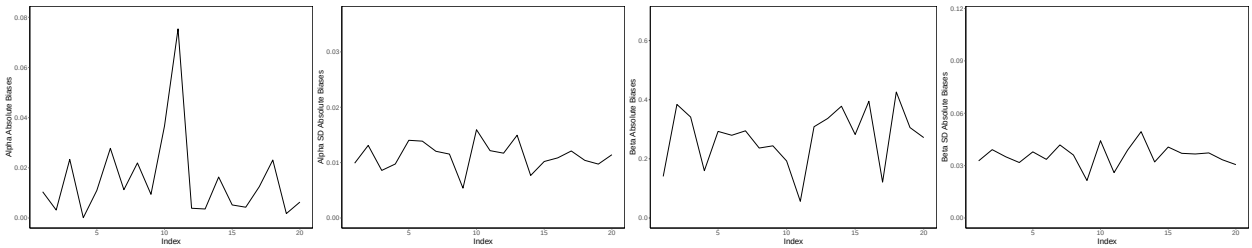


Average Random

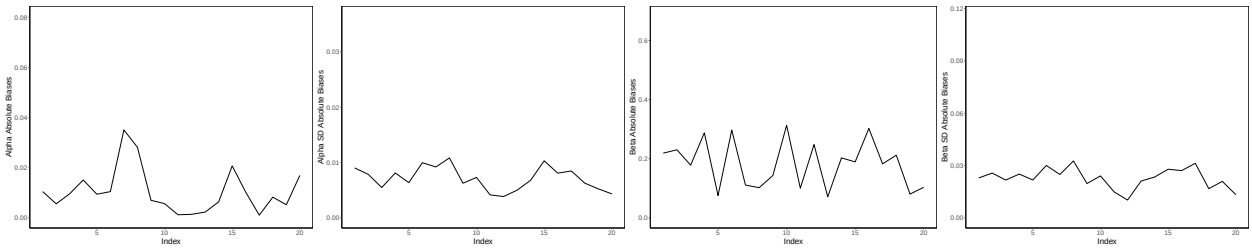
Figure A.21: Average Center, Average ll, Average Corner and Average Random Results of $n = 5000$



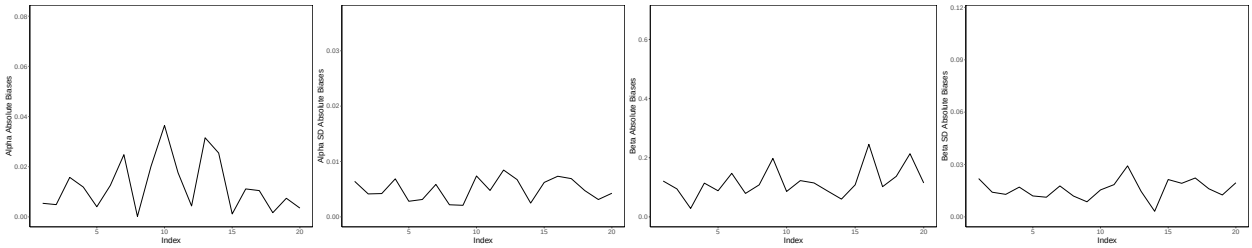
$k = 10\%$



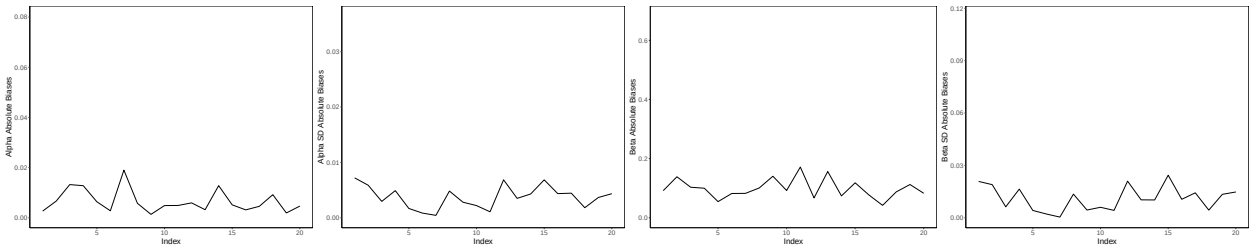
$k = 20\%$



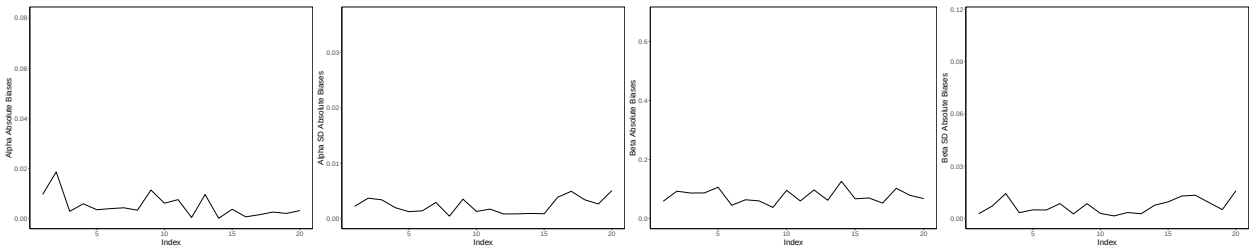
$k = 30\%$



$k = 40\%$

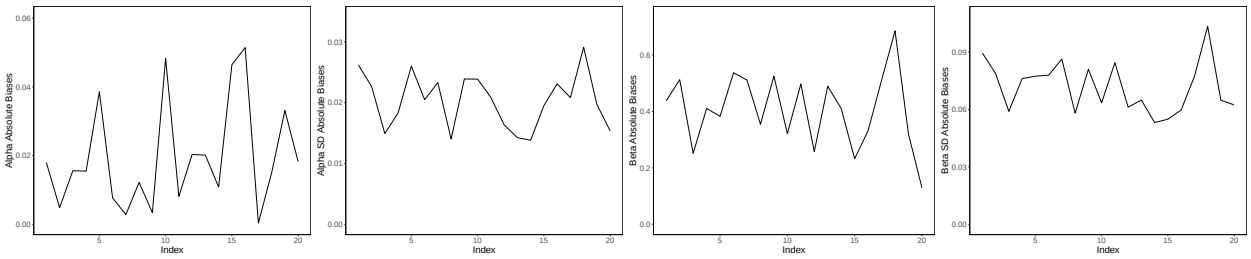


$k = 50\%$

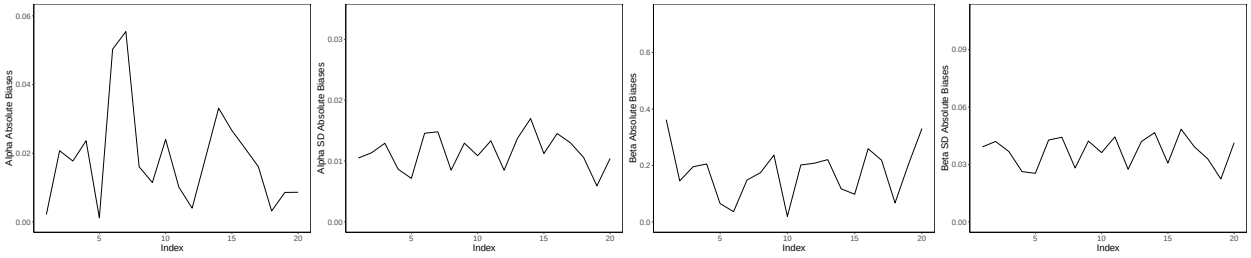


$k = 64\%$

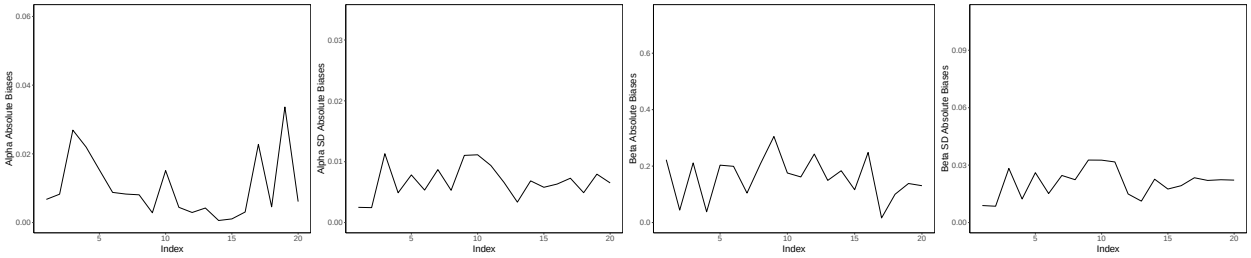
Figure A.22: $k-$Percent Center Results for n=5000



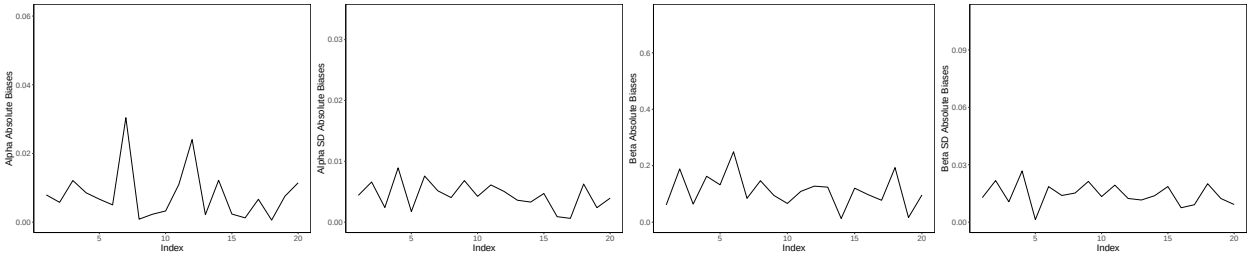
$k = 10\%$



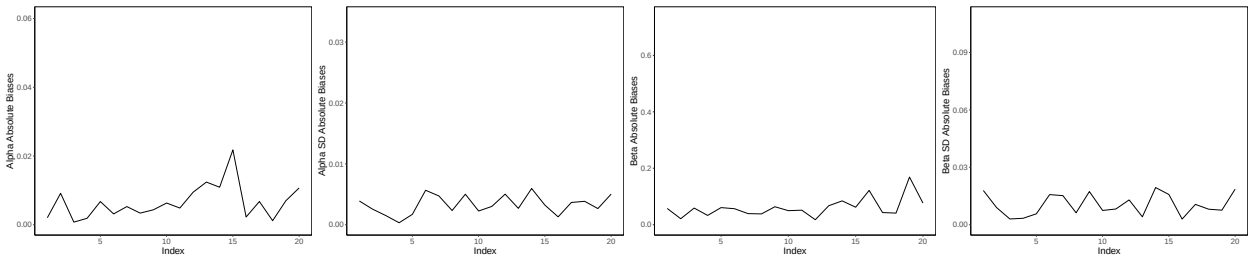
$k = 20\%$



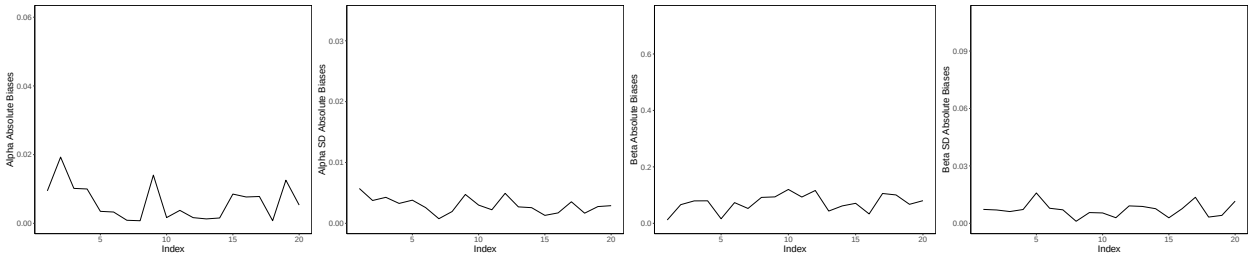
$k = 30\%$



$k = 40\%$

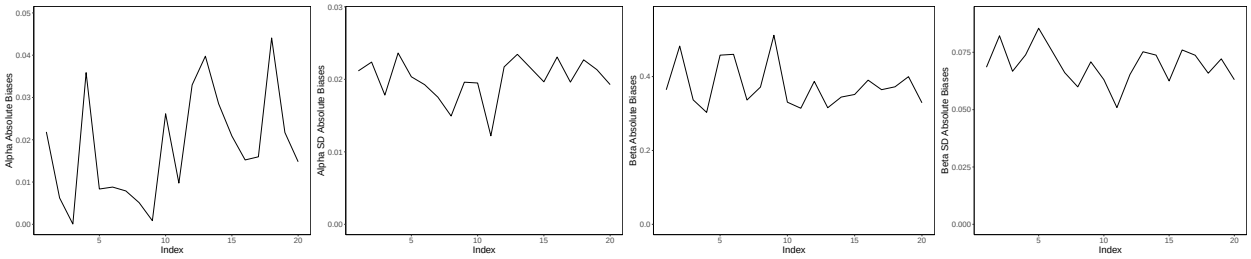


$k = 50\%$

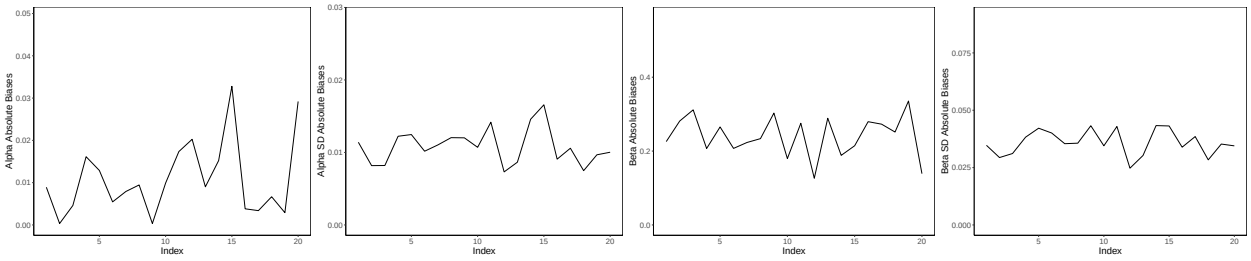


$k = 64\%$

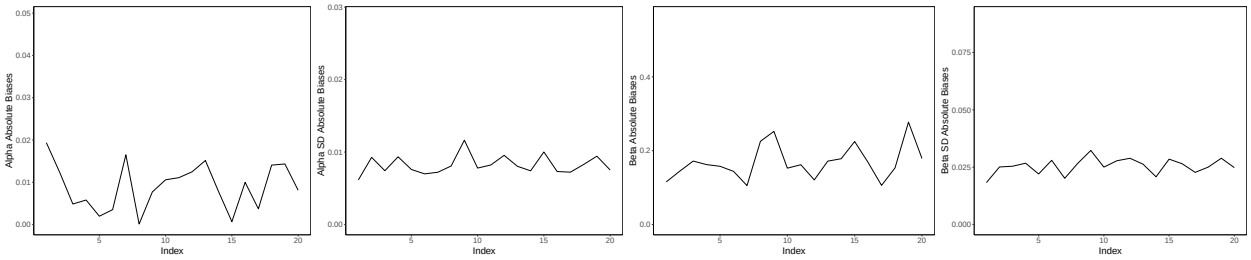
Figure A.23: $k-$Percent ll Results for n=5000



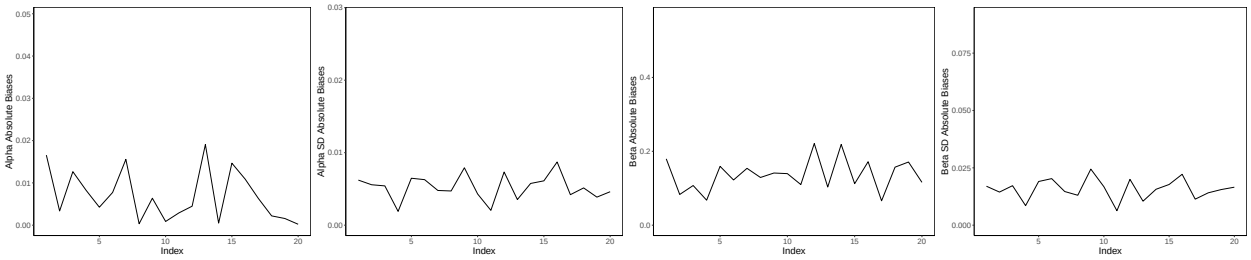
$k = 10\%$



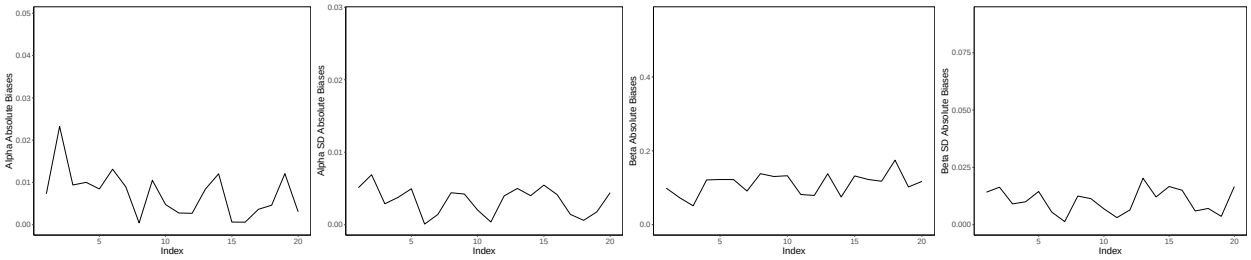
$k = 20\%$



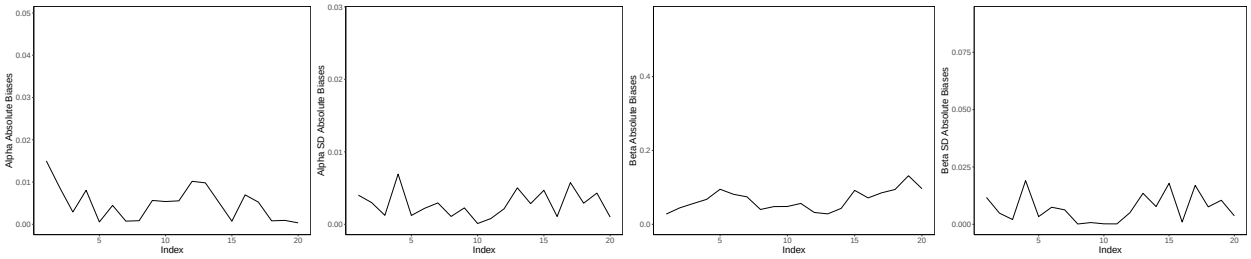
$k = 30\%$



$k = 40\%$

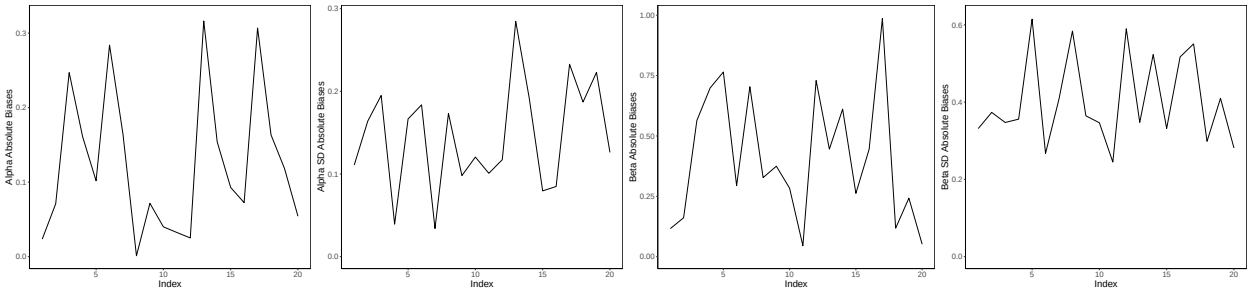


$k = 50\%$

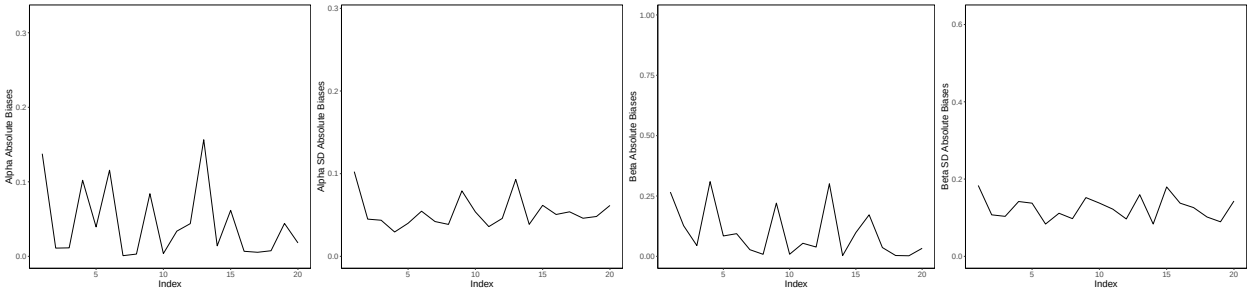


$k = 64\%$

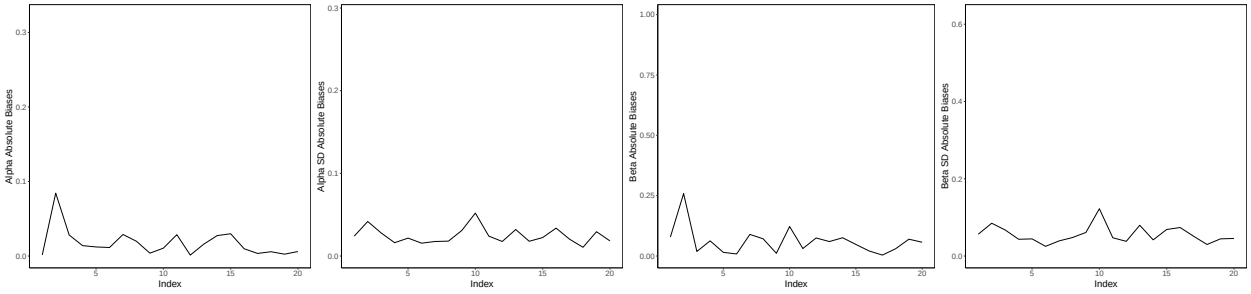
Figure A.24: $k-$Percent Random Results for n=5000


Alpha Absolute Biases
Alpha SD Absolute Biases
Beta Absolute Biases
Beta SD Absolute Biases
Index

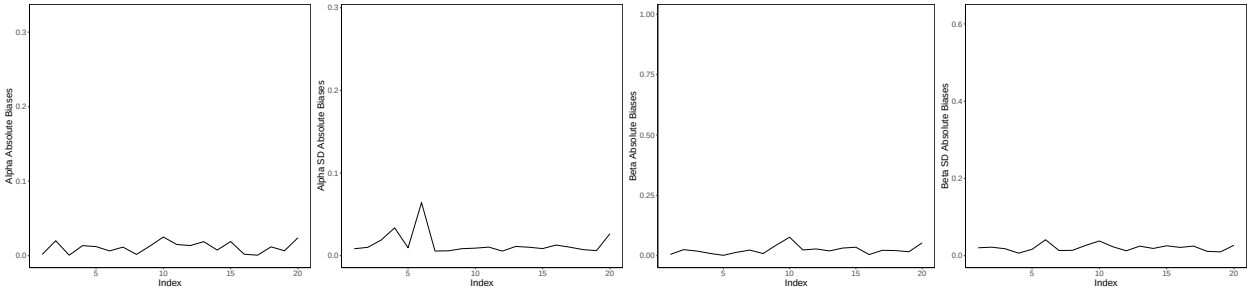
$m = 2$


Alpha Absolute Biases
Alpha SD Absolute Biases
Beta Absolute Biases
Beta SD Absolute Biases
Index

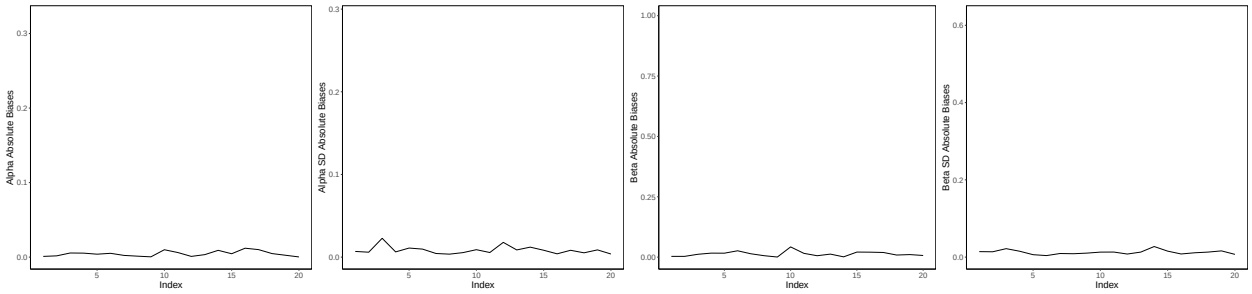
$m = 3$


Alpha Absolute Biases
Alpha SD Absolute Biases
Beta Absolute Biases
Beta SD Absolute Biases
Index

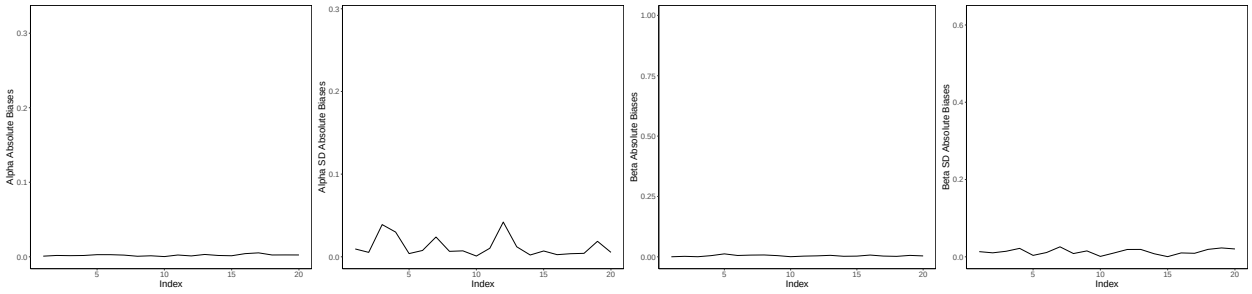
$m = 4$


Alpha Absolute Biases
Alpha SD Absolute Biases
Beta Absolute Biases
Beta SD Absolute Biases
Index

$m = 5$



$m = 6$



$m = 7$

Figure A.25: Temporal 1 Subset-m Results for n=5000

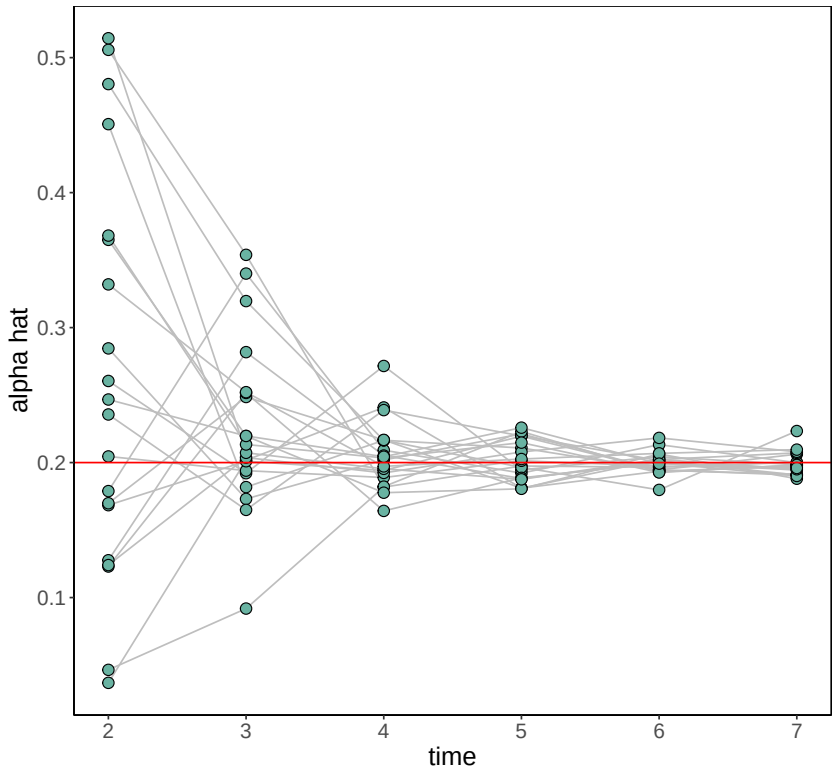


Figure A.26: Temporal 1 Subset-m $\hat{\alpha}$ vs α for n=5000

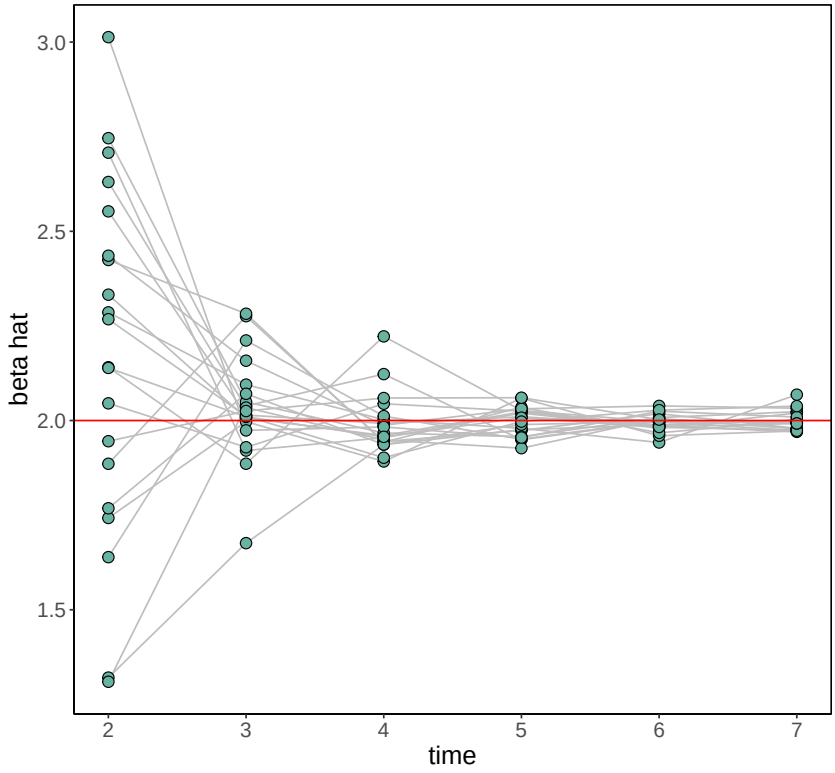
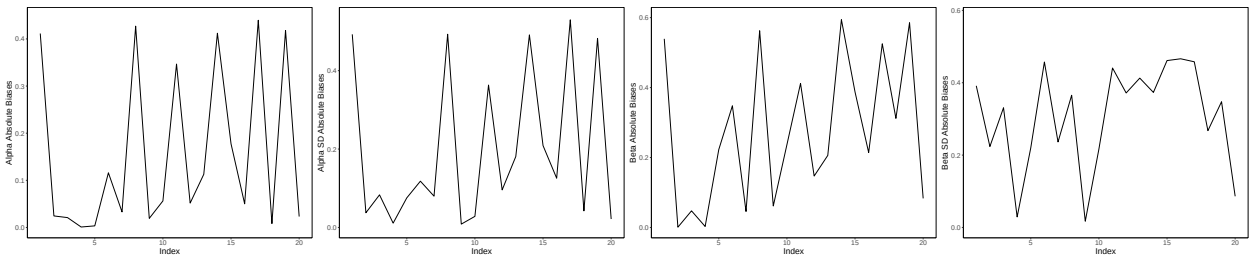
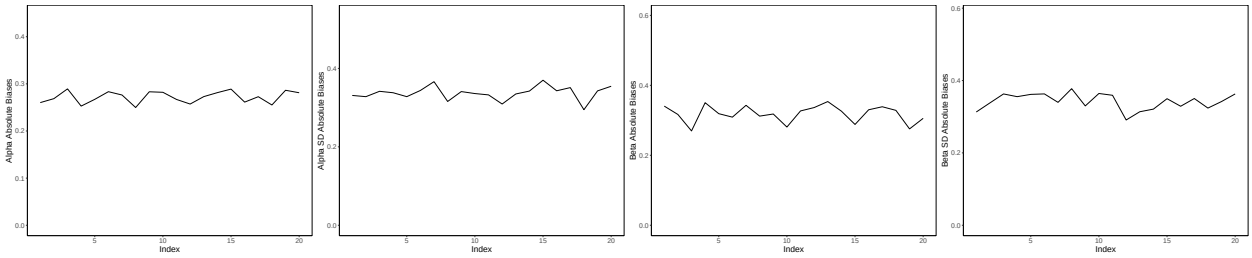


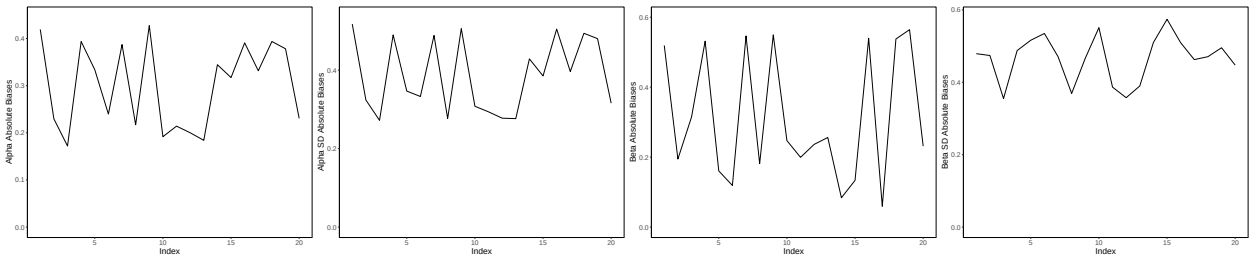
Figure A.27: Temporal 1 Subset-m $\hat{\beta}$ vs β for n=5000



$h = 2$

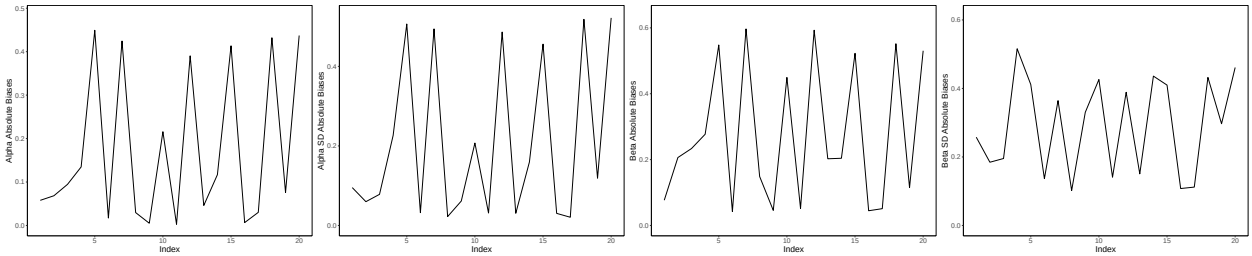


$h = 3$

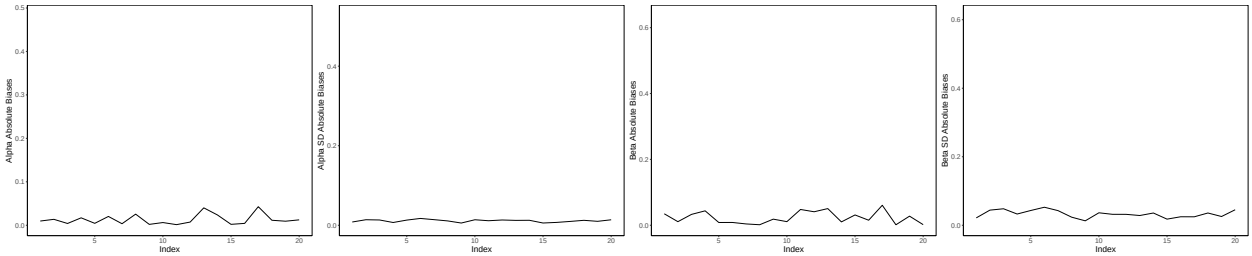


$h = 4$

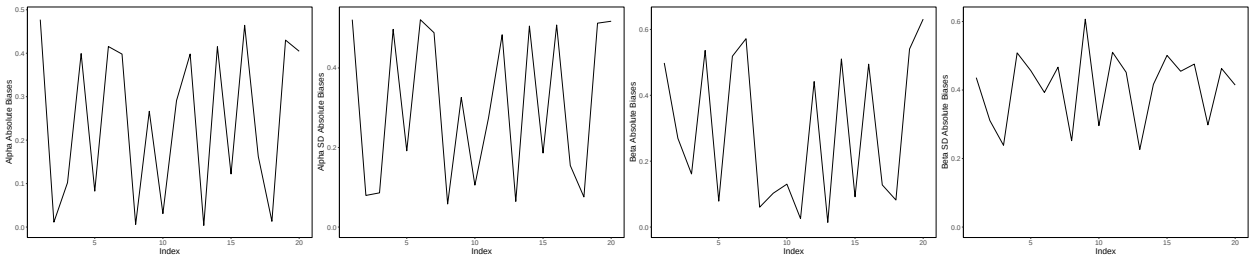
Figure A.28: Temporal Mean Subset-h results for n=5000



$h = 2$



$h = 3$



$h = 4$

Figure A.29: Temporal Median Subset-h results for n=5000

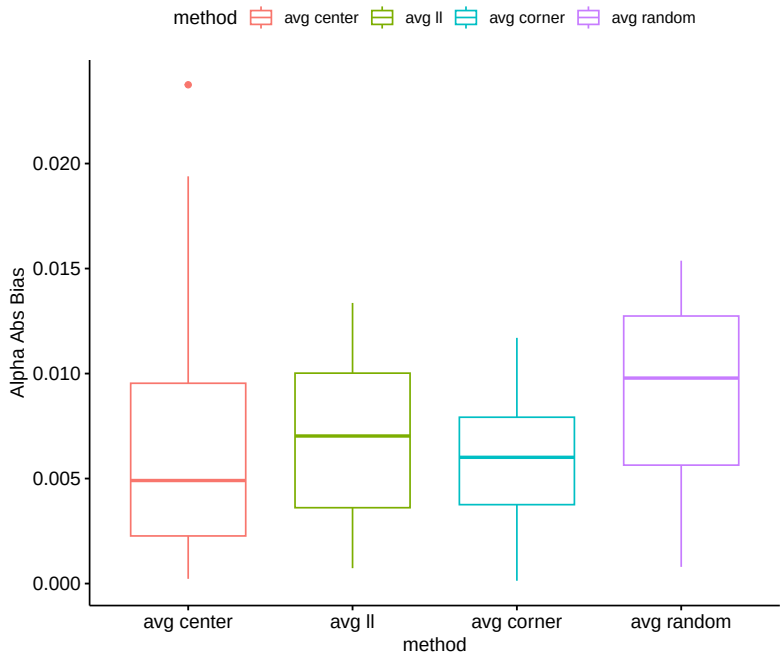


Figure A.30: Alpha Absolute Bias of Average Center, Average ll, Average Corner, and Average Random Methods for $n = 5000$

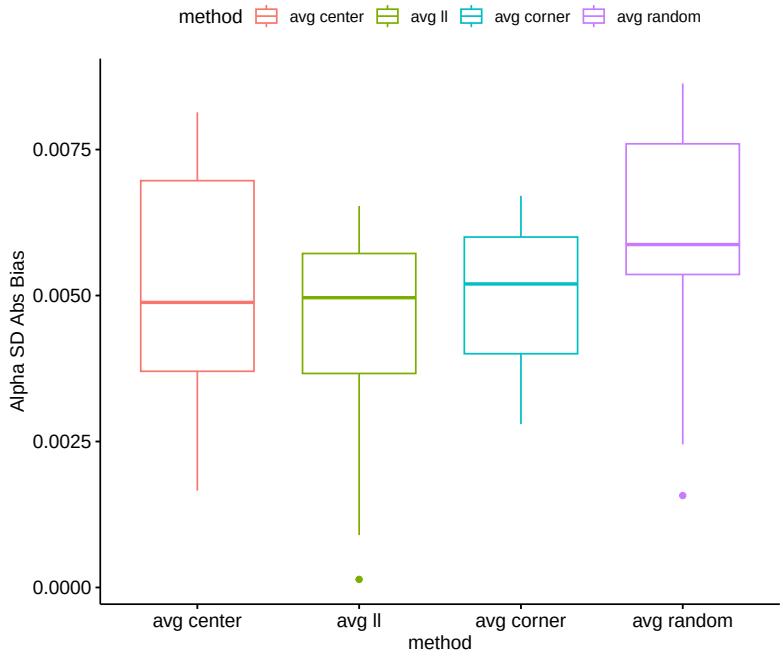


Figure A.31: Alpha SD Absolute Bias of Average Center, Average ll, Average Corner, and Average Random Methods for $n = 5000$

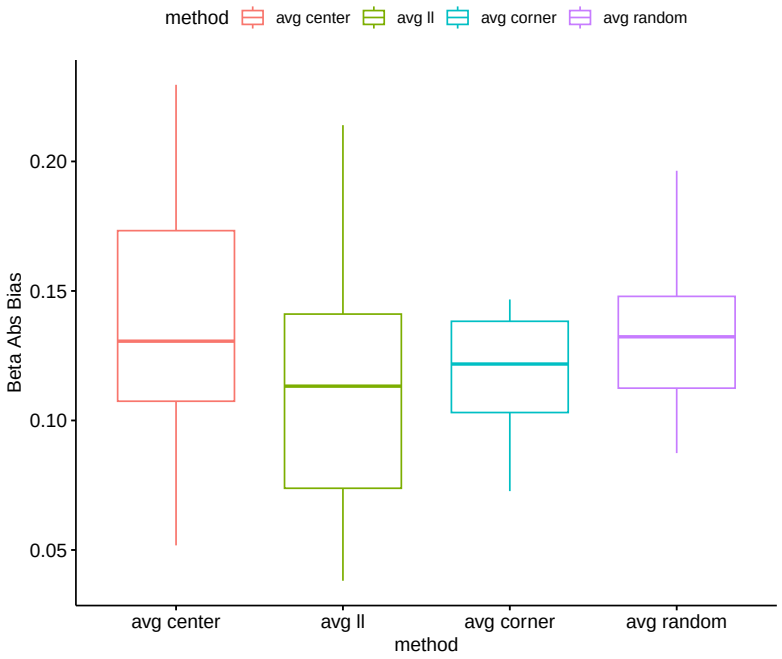


Figure A.32: Beta Absolute Bias of Average Center, Average ll, Average Corner, and Average Random Methods for $n = 5000$

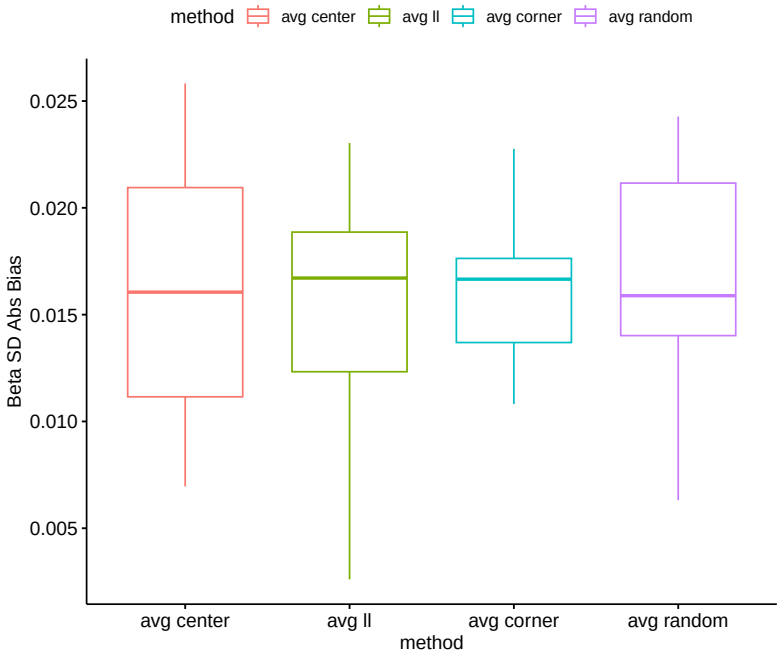


Figure A.33: Beta SD Absolute Bias of Average Center, Average ll, Average Corner, and Average Random Methods for $n = 5000$

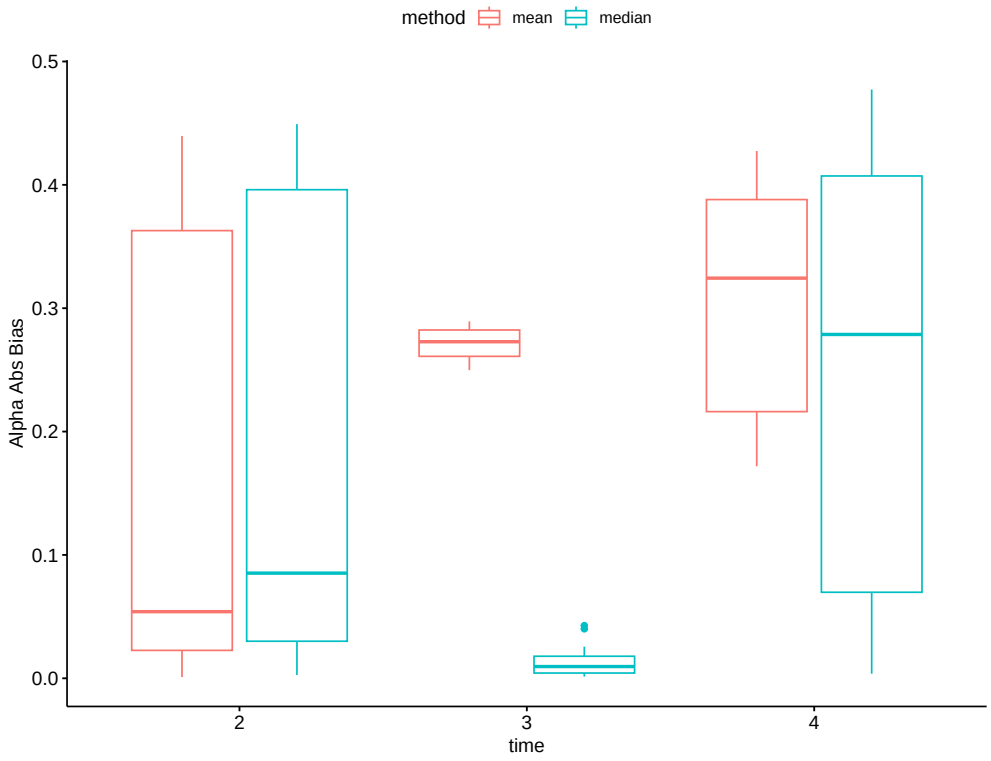


Figure A.34: Alpha Absolute Bias of Temporal Mean Subset-h and Temporal Median Subset-h methods for $n = 5000$

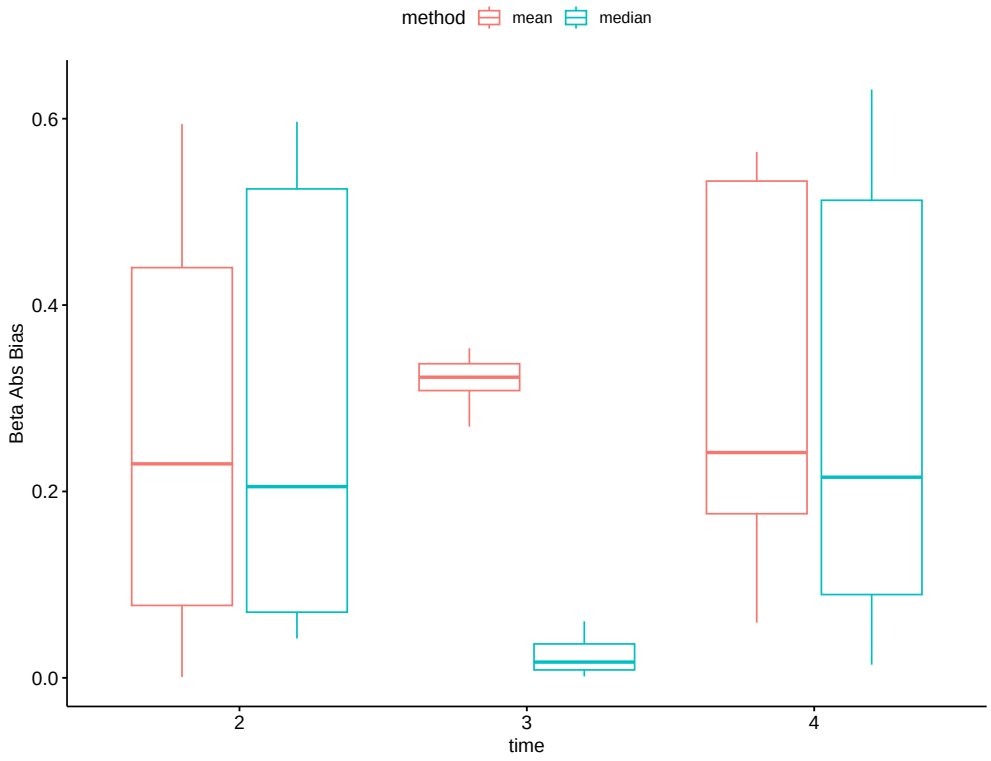


Figure A.35: Beta Absolute Bias of Temporal Mean Subset-h and Temporal Median Subset-h methods for $n = 5000$

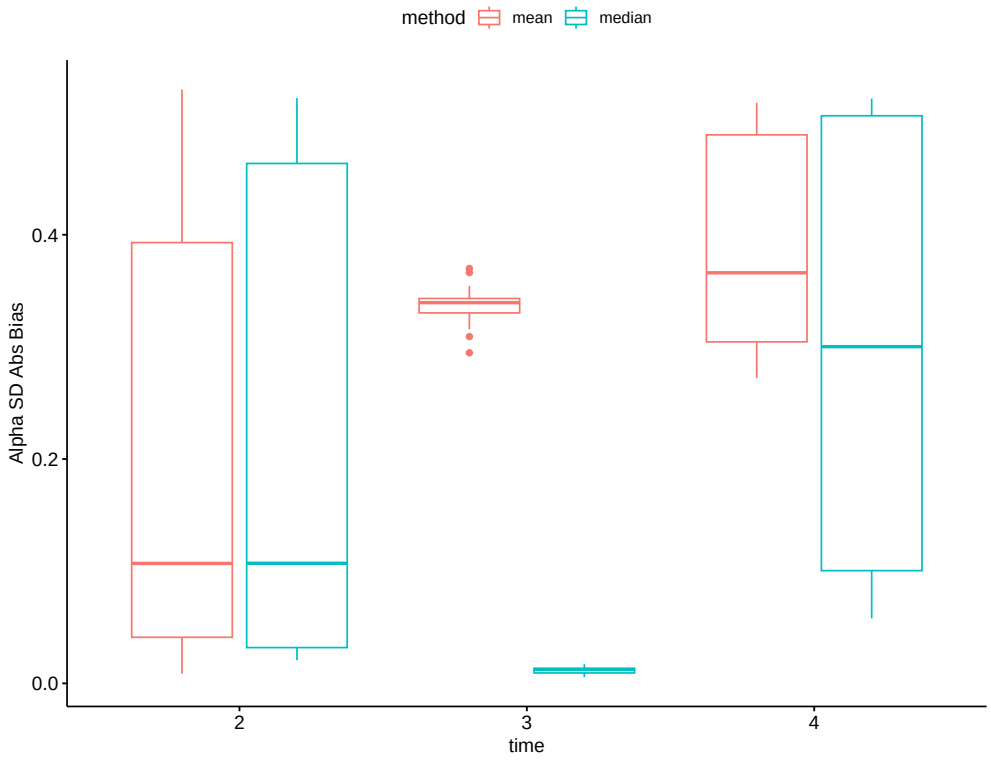


Figure A.36: Alpha SD Absolute Bias Temporal Mean Subset-h and Temporal Median Subset-h methods for $n = 5000$

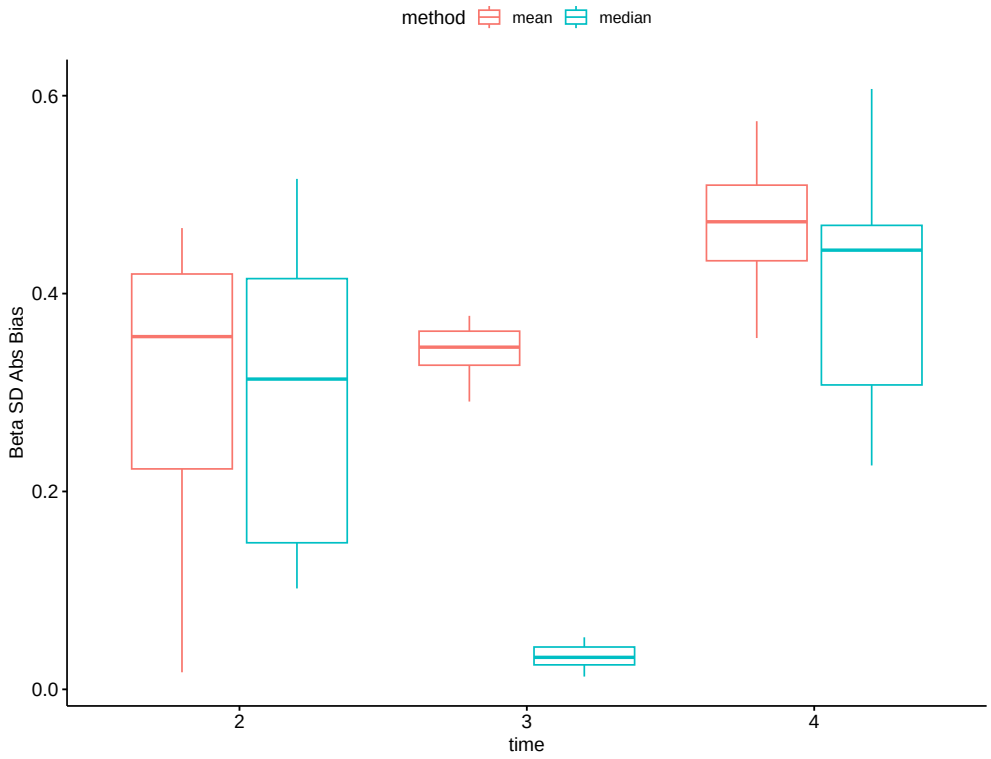


Figure A.37: Beta SD Absolute Bias of Temporal Mean Subset-h and Temporal Median Subset-h methods for $n = 5000$

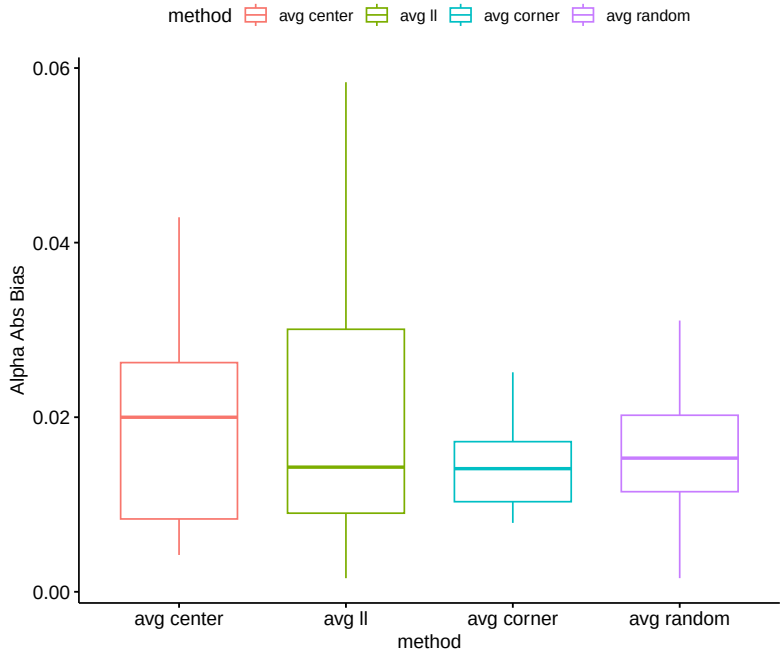


Figure A.38: Alpha Absolute Bias of Average Center, Average ll, Average Corner, and Average Random Methods for $n = 1000$

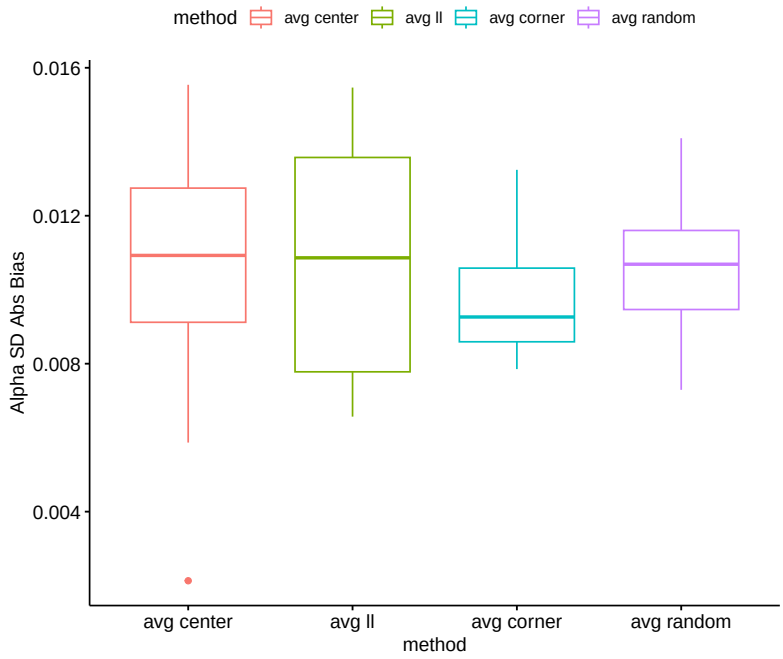


Figure A.39: Alpha SD Absolute Bias of Average Center, Average ll, Average Corner, and Average Random Methods for $n = 1000$

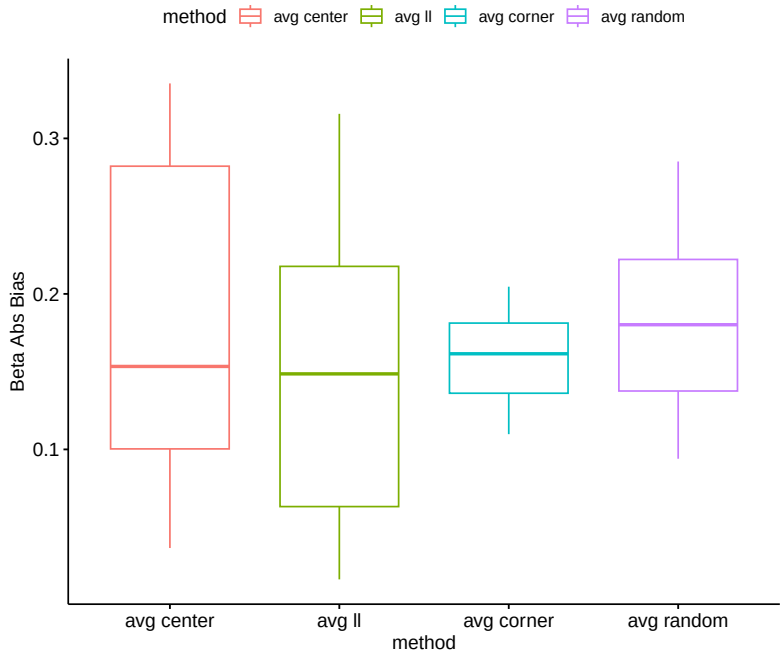


Figure A.40: Beta Absolute Bias of Average Center, Average ll, Average Corner, and Average Random Methods for $n = 1000$

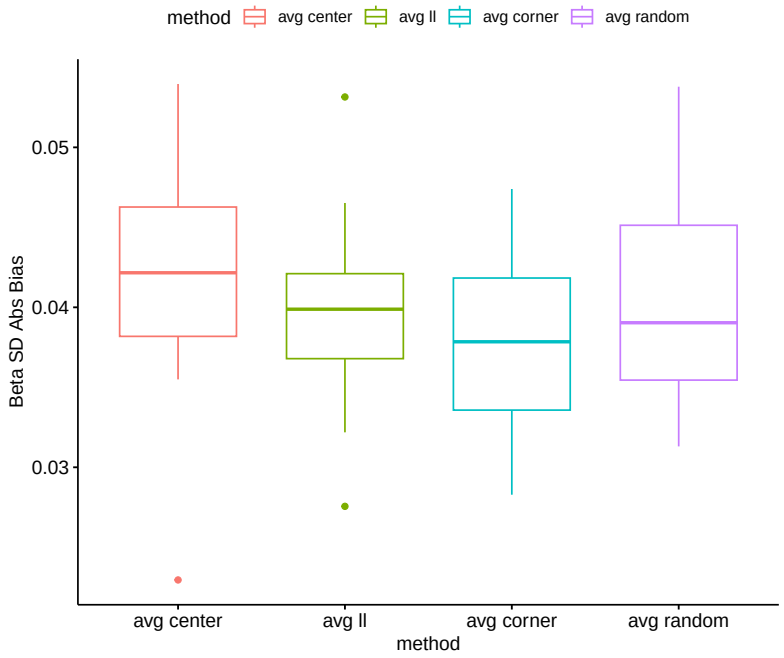


Figure A.41: Beta SD Absolute Bias of Average Center, Average ll, Average Corner, and Average Random Methods for $n = 1000$

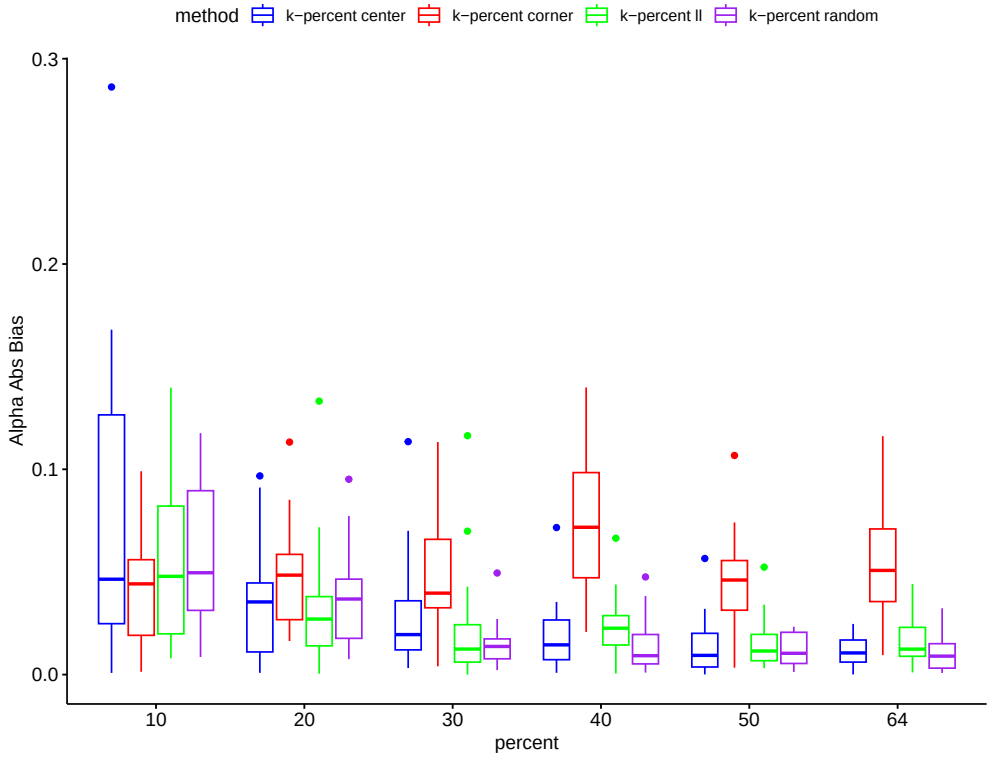


Figure A.42: Alpha Absolute Bias of k-percent Center,k-percent ll, k-percent Corner, and k-percent Random Methods for $n = 1000$

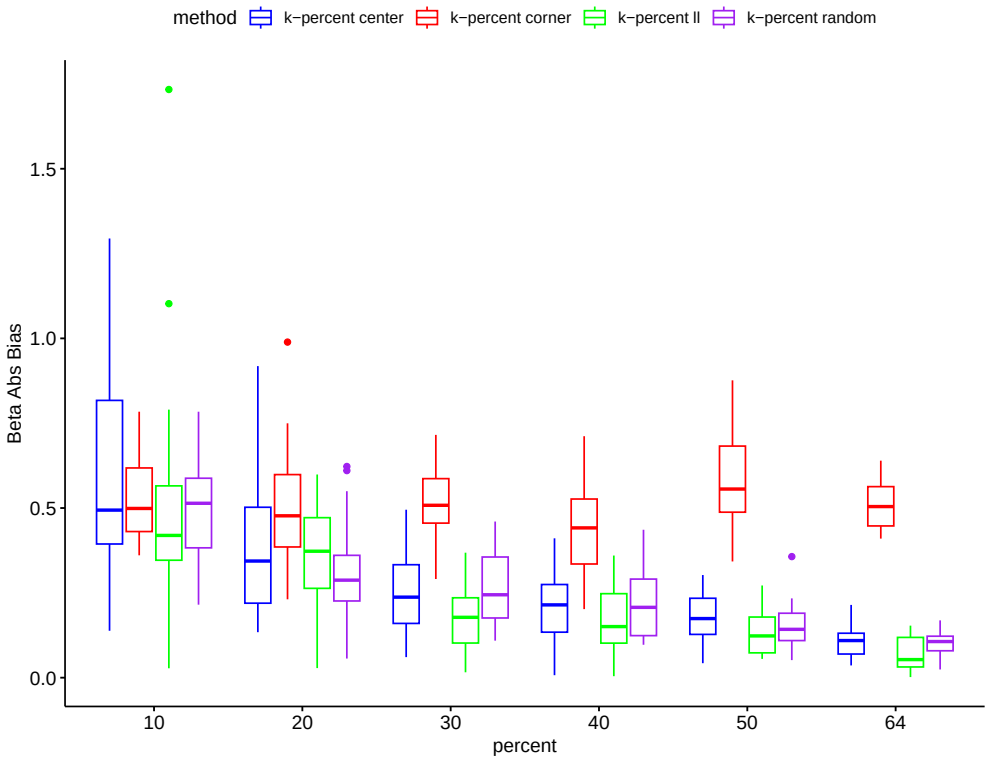


Figure A.43: Beta Absolute Bias of k-percent Center,k-percent ll, k-percent Corner, and k-percent Random Methods for $n = 1000$

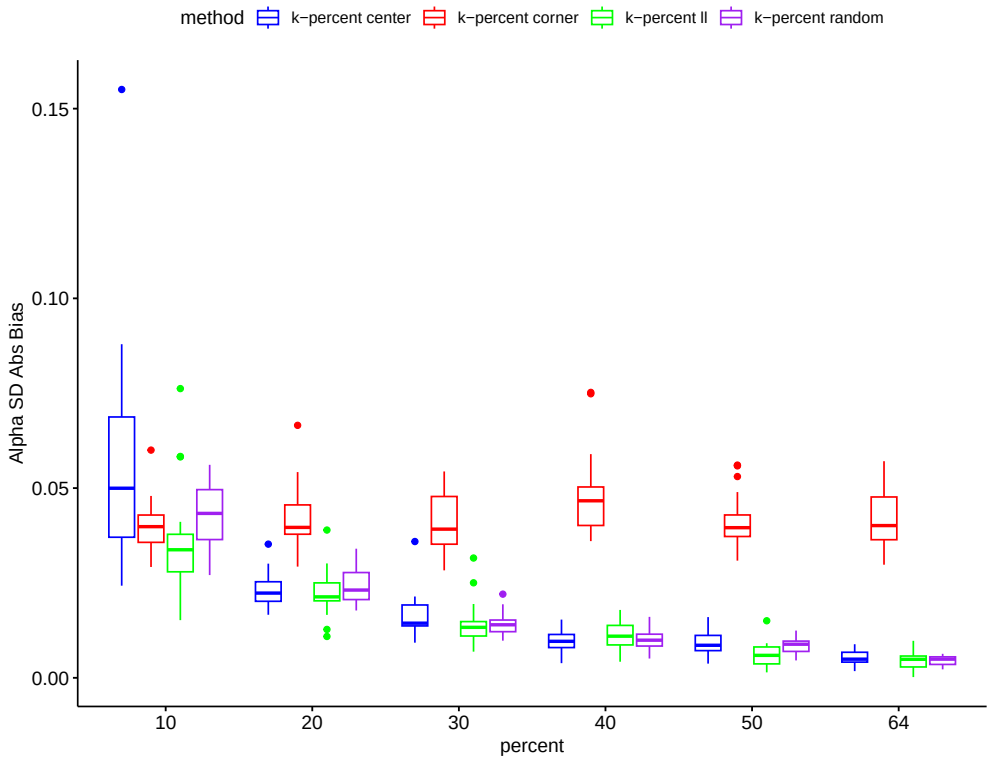


Figure A.44: Alpha SD Absolute Bias of k-percent Center,k-percent ll, k-percent Corner, and k-percent Random Methods for $n = 1000$

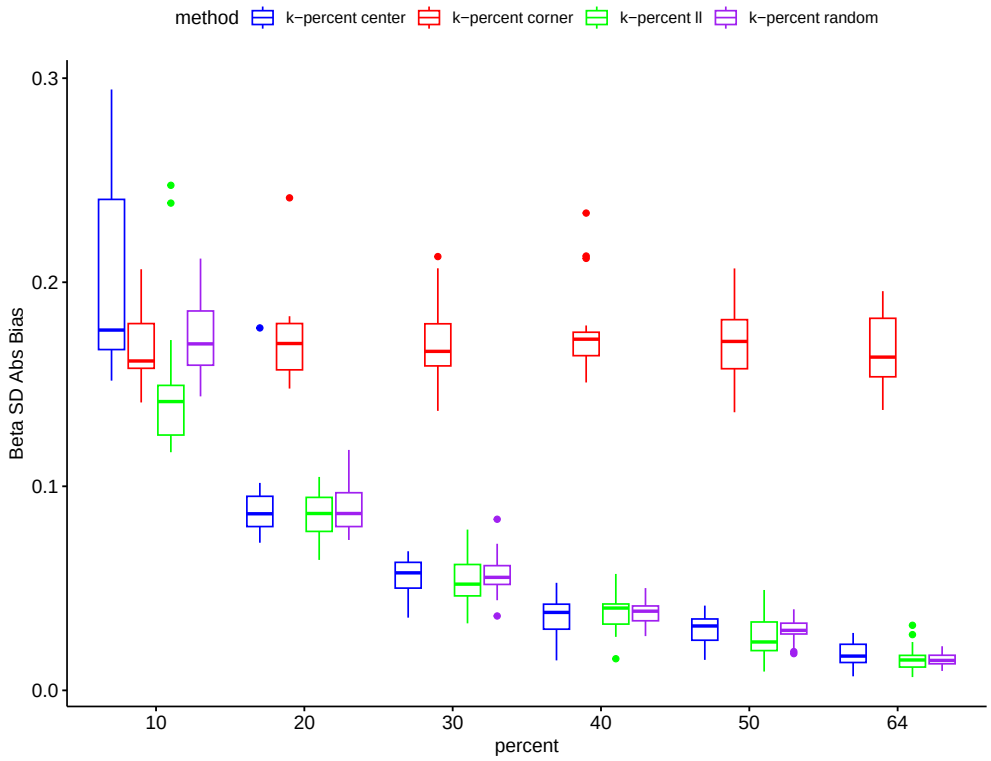


Figure A.45: Beta SD Absolute Bias of k-percent Center,k-percent ll, k-percent Corner, and k-percent Random Methods for $n = 1000$

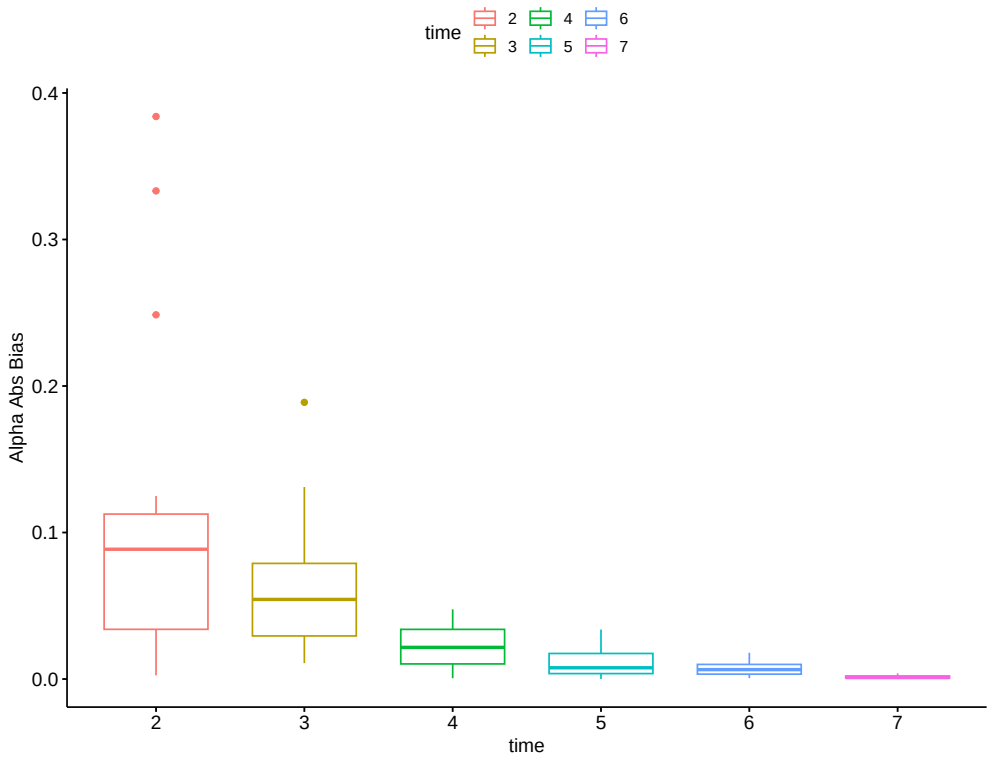


Figure A.46: Alpha Absolute Bias of Temporal 1 Subset-m method for $n = 1000$

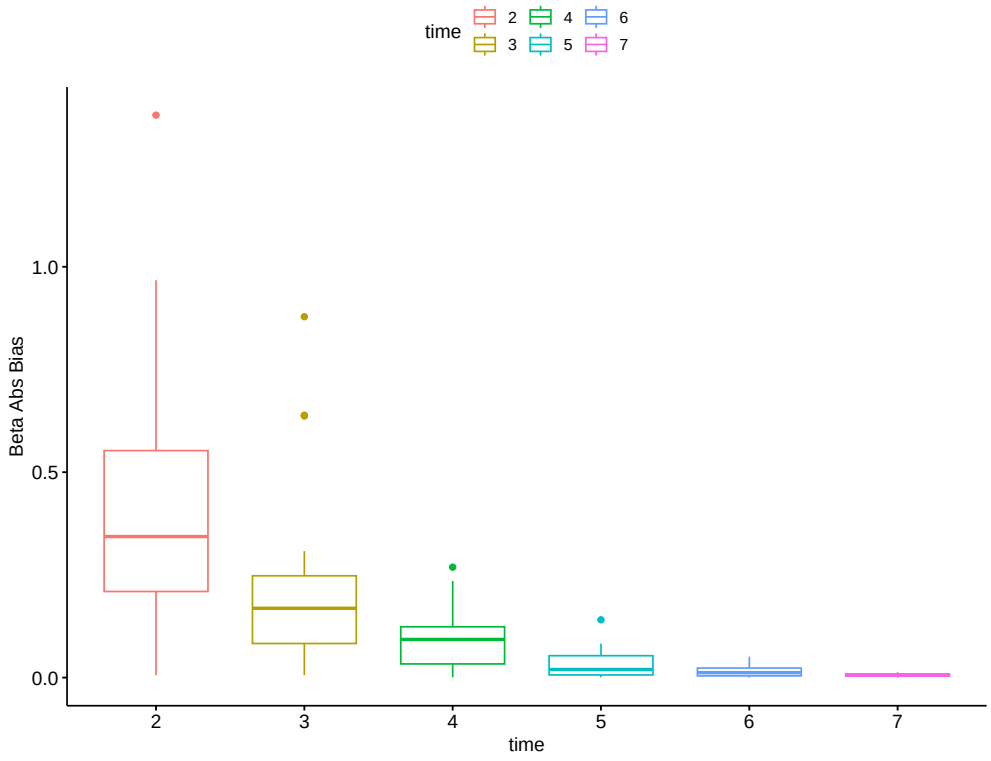


Figure A.47: Beta Absolute Bias of Temporal 1 Subset-m method for $n = 1000$

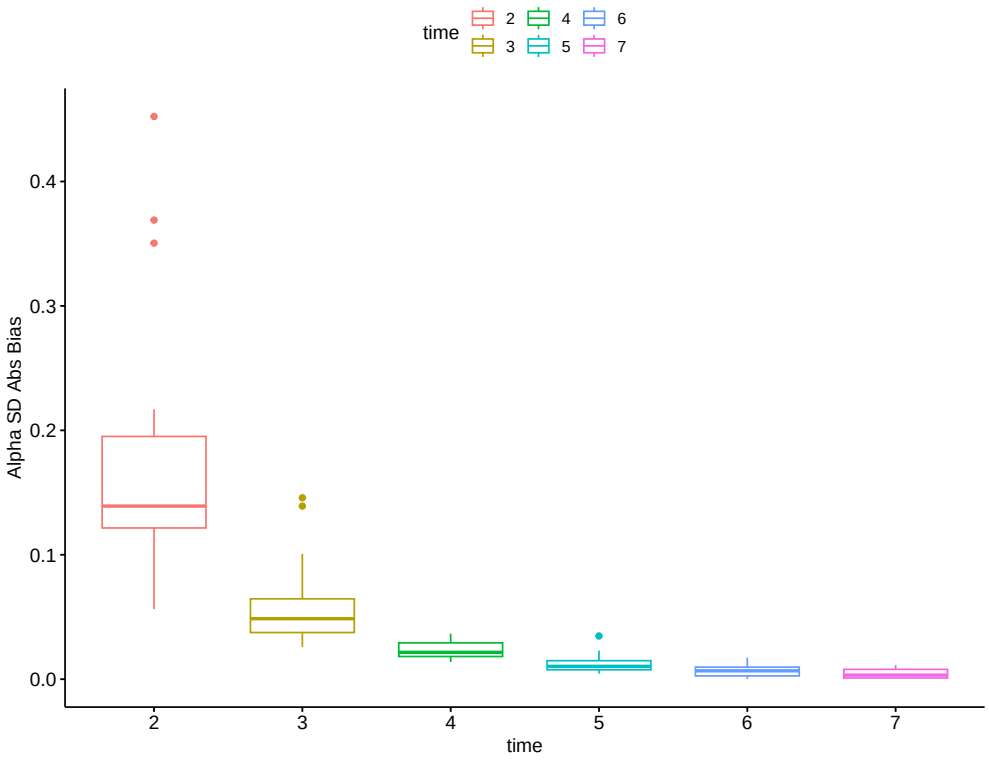


Figure A.48: Alpha SD Absolute Bias Temporal 1 Subset-m method for $n = 1000$

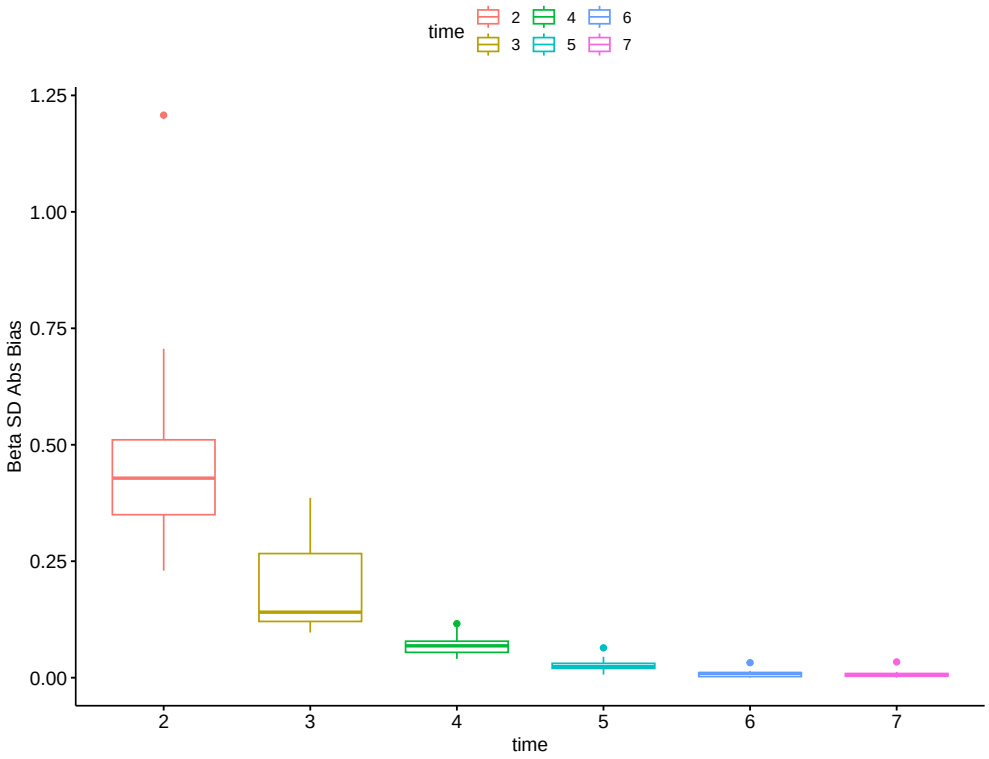


Figure A.49: Beta SD Absolute Bias of Temporal 1 Subset-m method for $n = 1000$

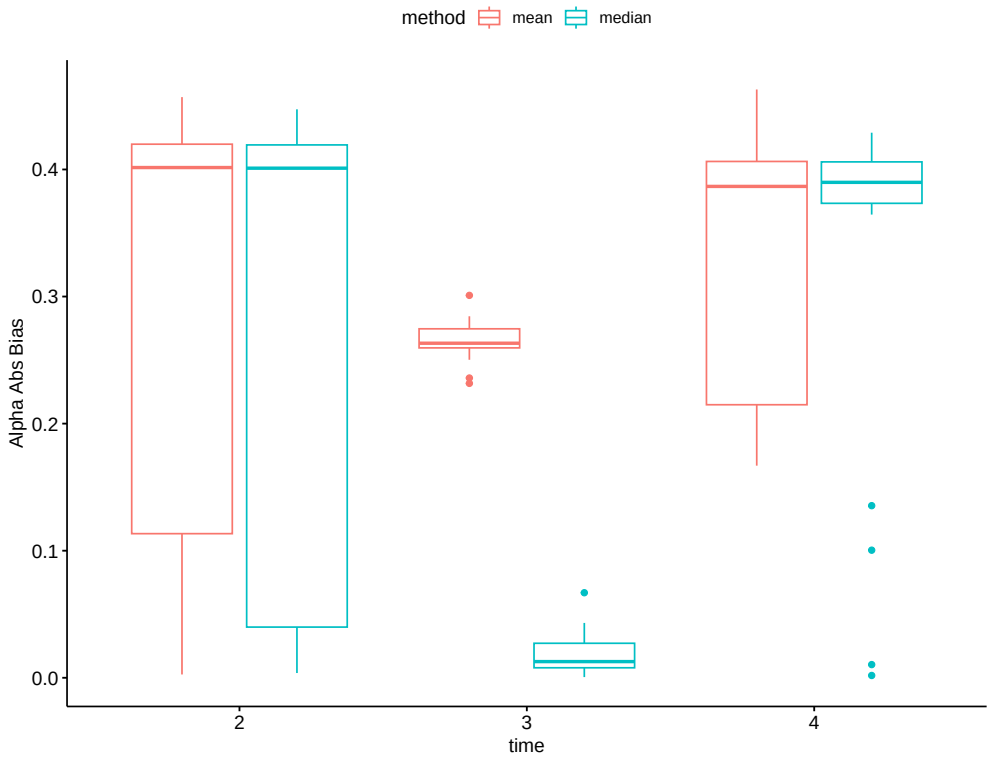


Figure A.50: Alpha Absolute Bias of Temporal Mean Subset-h and Temporal Median Subset-h methods for $n = 1000$

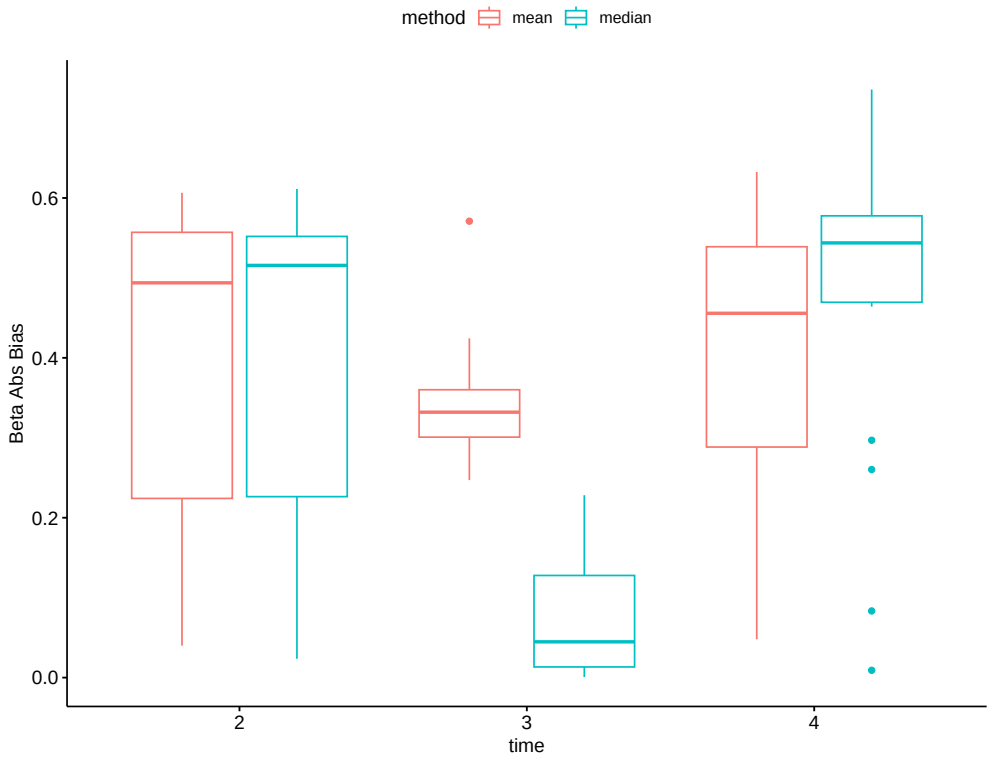


Figure A.51: Beta Absolute Bias of Temporal Mean Subset-h and Temporal Median Subset-h methods for $n = 1000$

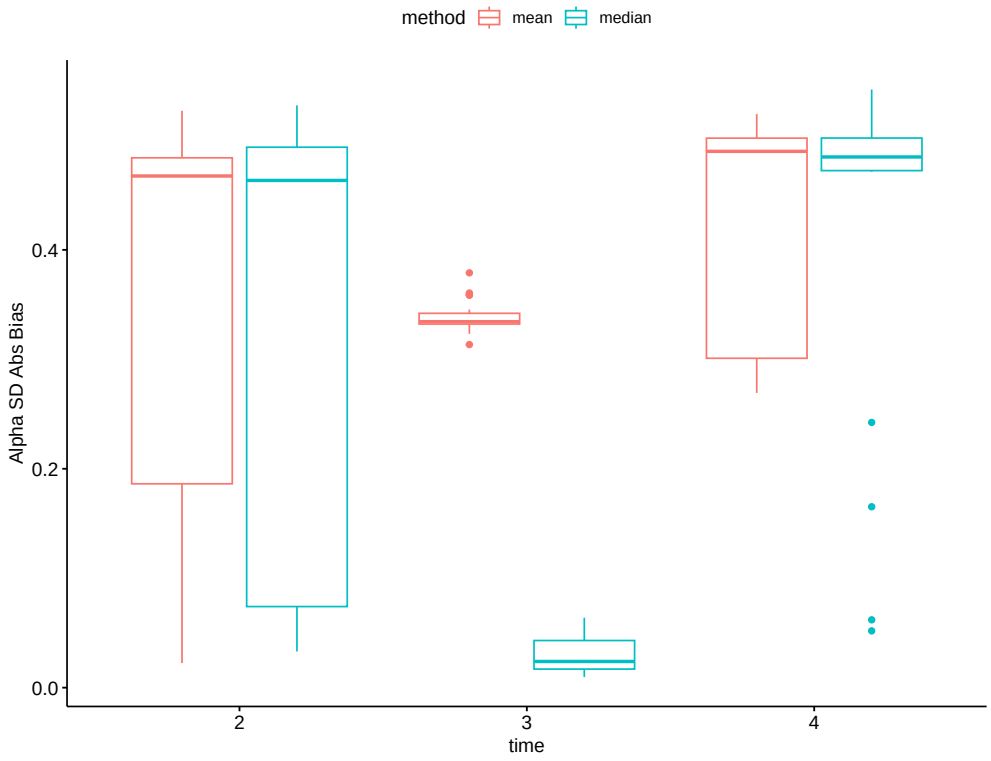


Figure A.52: Alpha SD Absolute Bias of Temporal Mean Subset-h and Temporal Median Subset-h methods for $n = 1000$

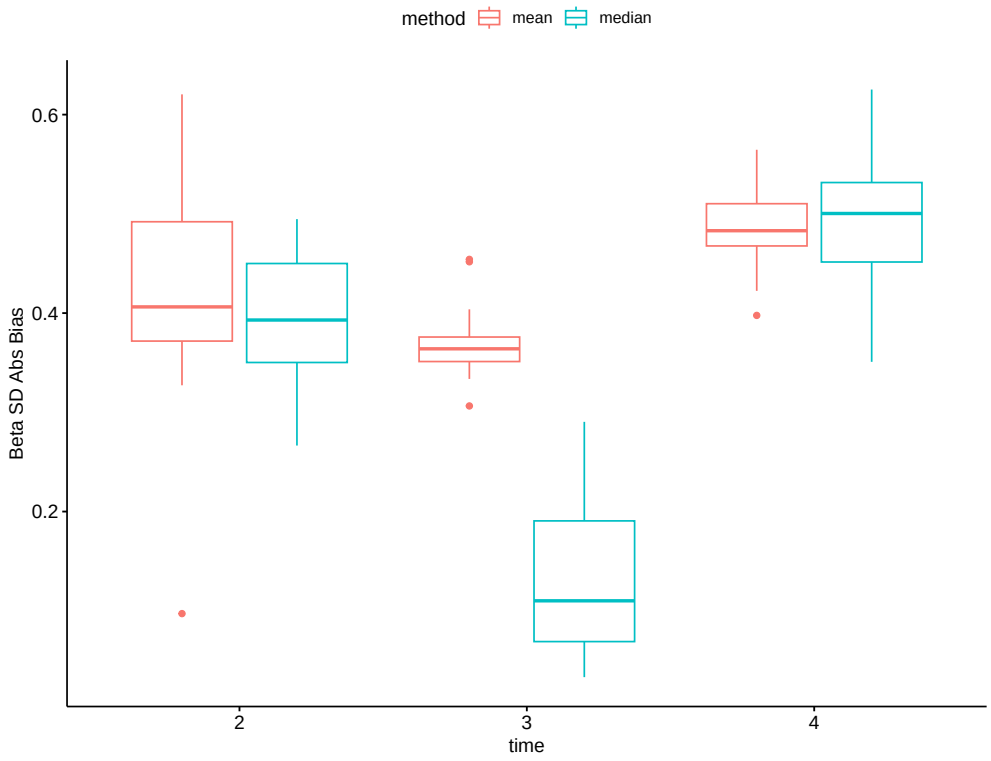


Figure A.53: Beta SD Absolute Bias of Temporal Mean Subset-h and Temporal Median Subset-h methods for $n = 1000$

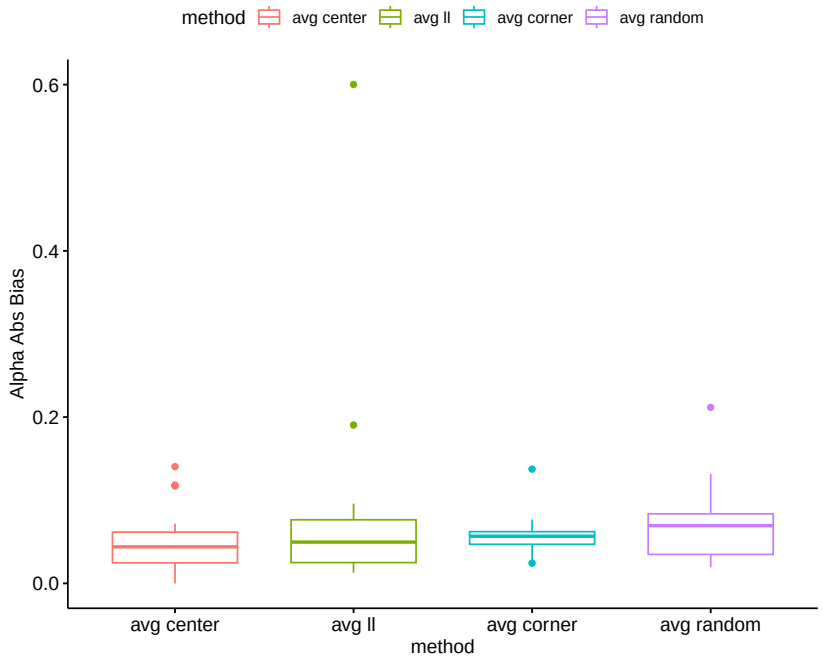


Figure A.54: Alpha Absolute Bias of Average, Average Corner, and Average Random Methods for $n = 100$

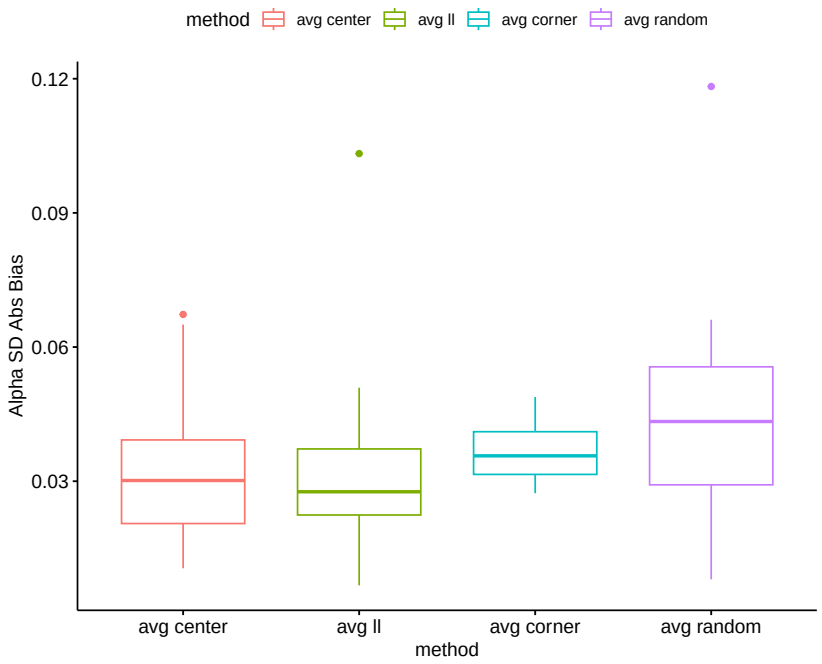


Figure A.55: Alpha SD Absolute Bias of Average, Average Corner, and Average Random Methods for $n = 100$

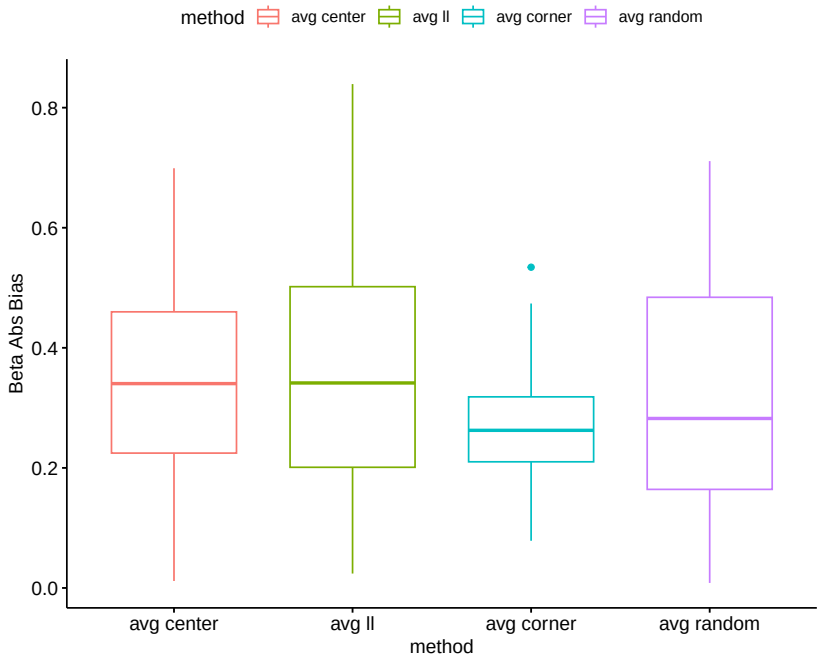


Figure A.56: Beta Absolute Bias of Average, Average Corner, and Average Random Methods for $n = 100$

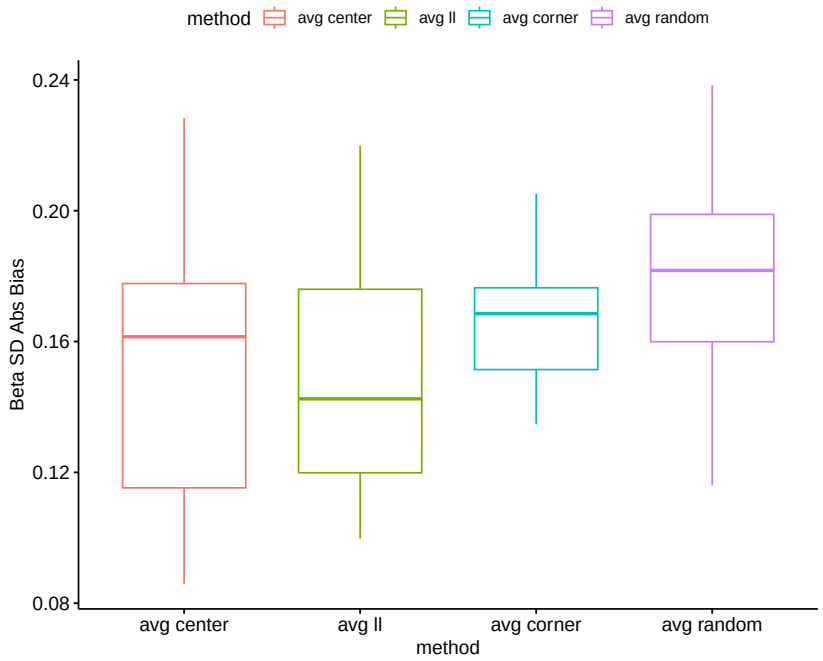


Figure A.57: Beta SD Absolute Bias of Average, Average Corner, and Average Random Methods for $n = 100$

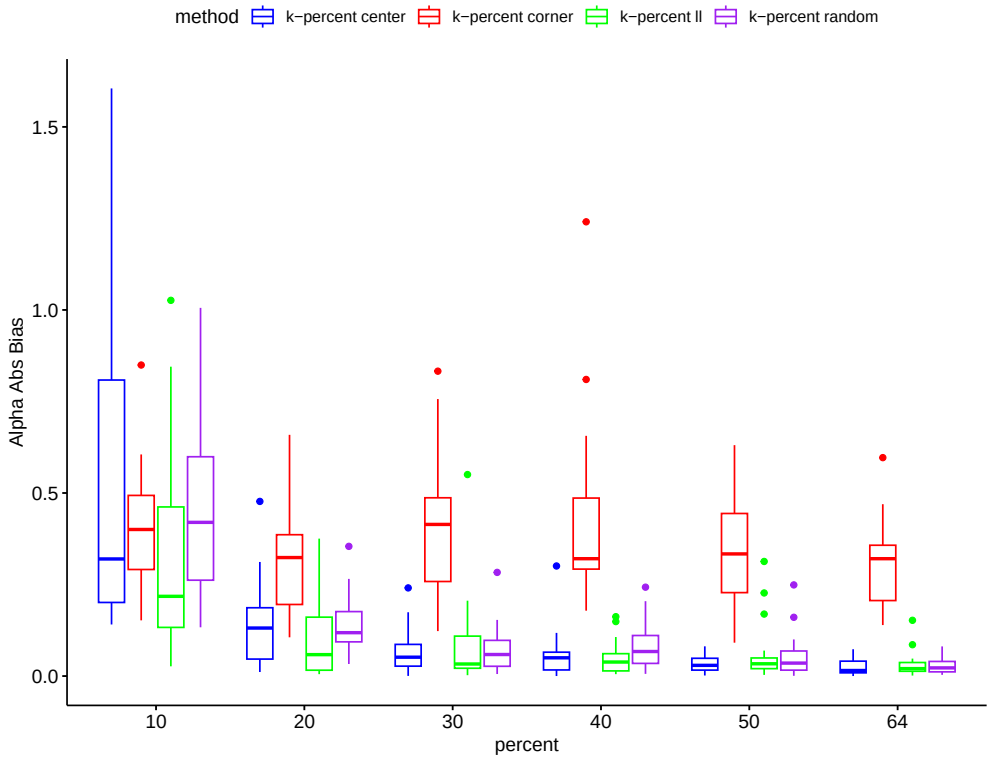


Figure A.58: Alpha Absolute Bias of k-percent Center,k-percent ll, k-percent Corner, and k-percent Random Methods for $n = 100$

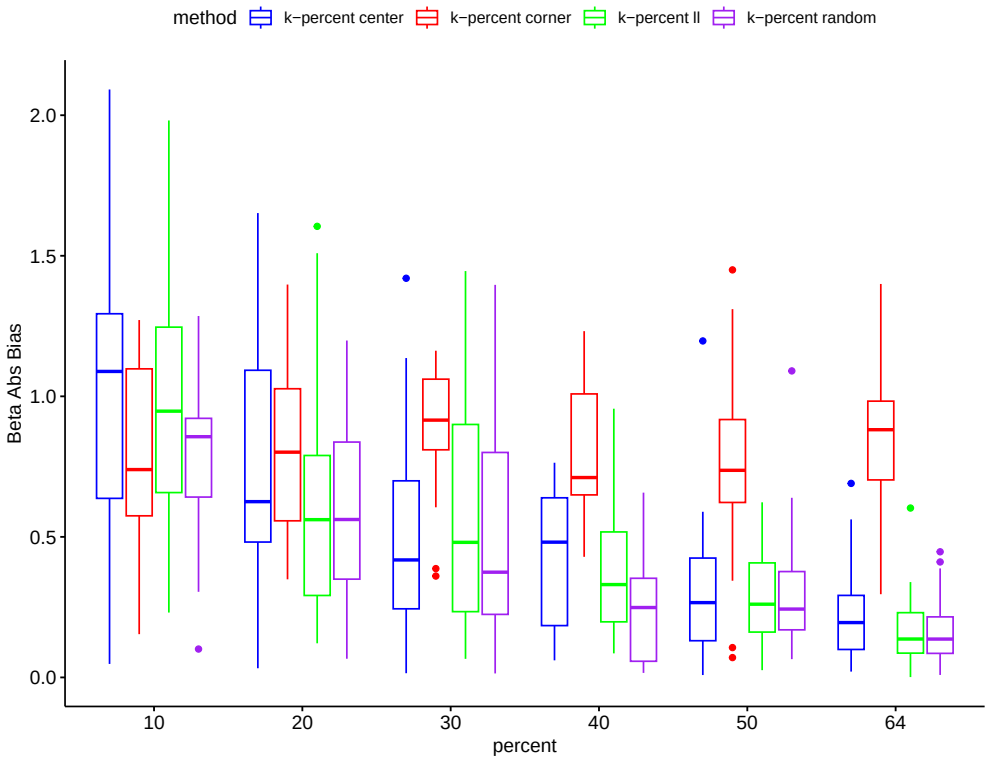


Figure A.59: Beta Absolute Bias of k-percent Center,k-percent ll, k-percent Corner, and k-percent Random Methods for $n = 100$

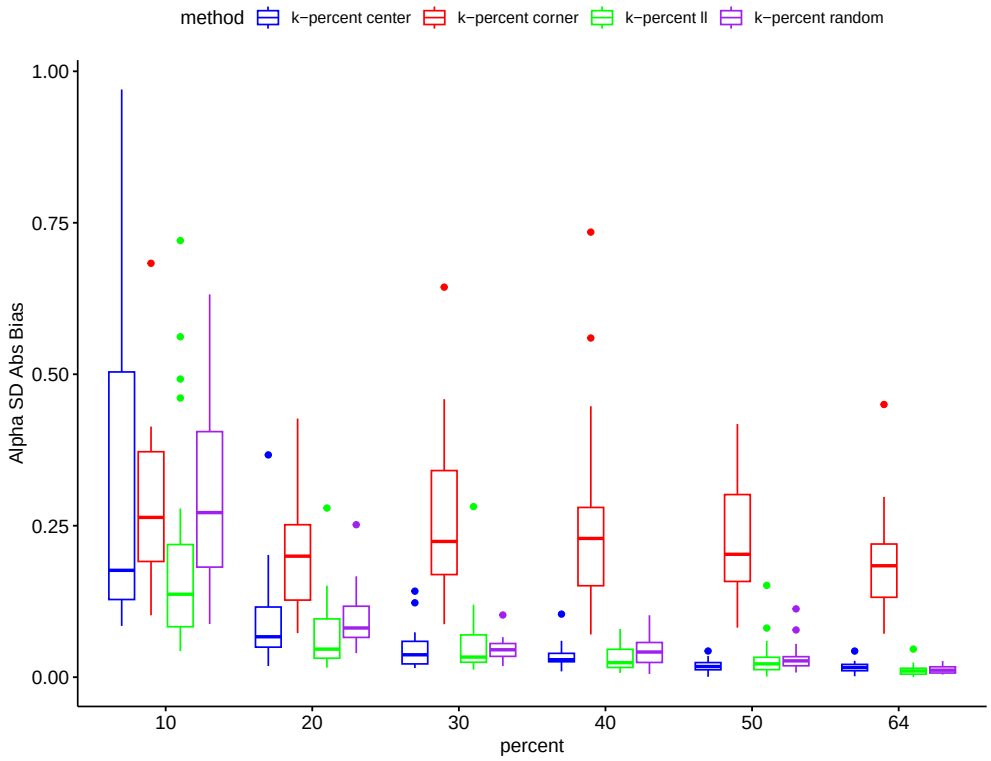


Figure A.60: Alpha SD Absolute Bias of k-percent Center,k-percent ll, k-percent Corner, and k-percent Random Methods for $n = 100$

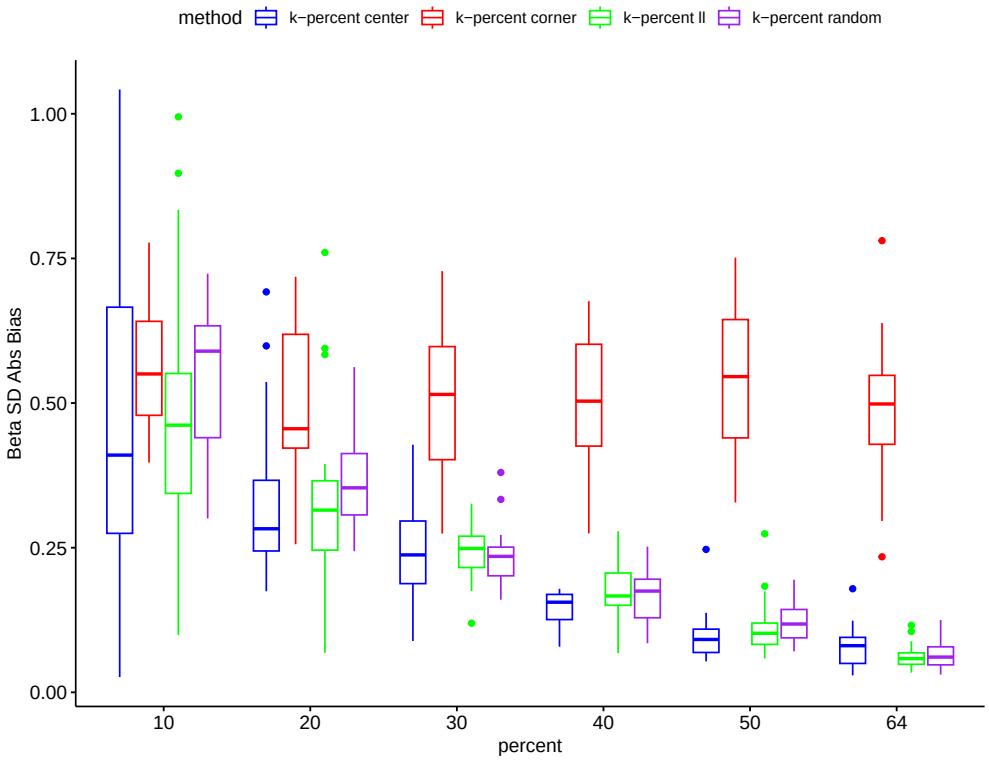


Figure A.61: Beta SD Absolute Bias of k-percent Center,k-percent ll, k-percent Corner, and k-percent Random Methods for $n = 100$

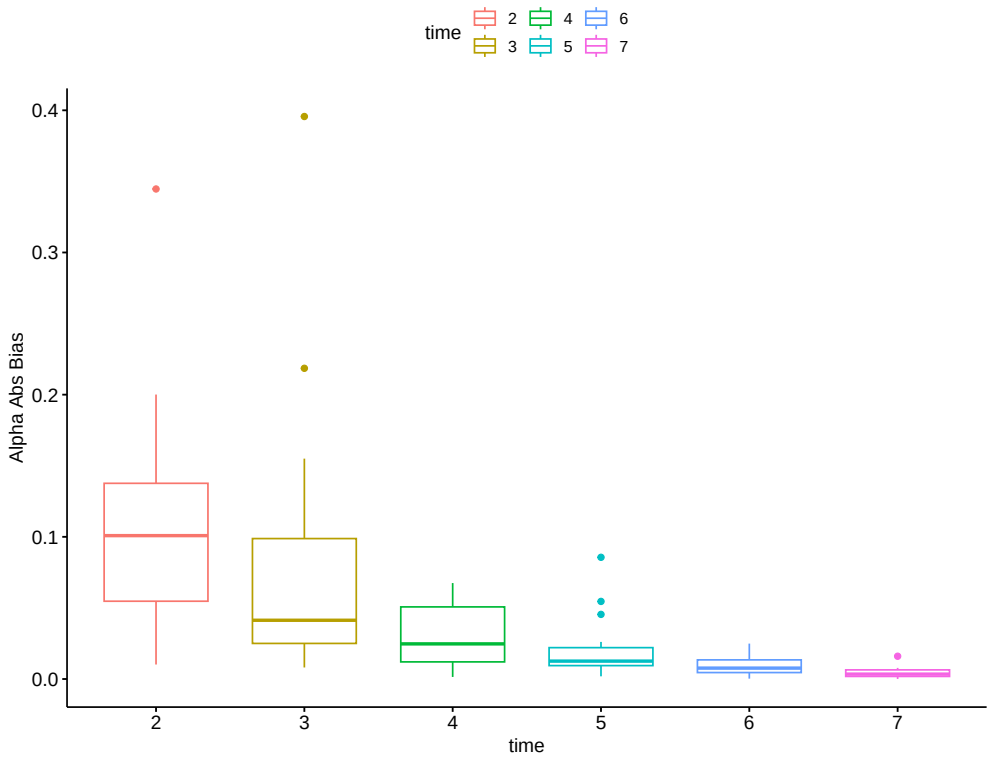


Figure A.62: Alpha Absolute Bias of Temporal 1 Subset-m method for $n = 100$

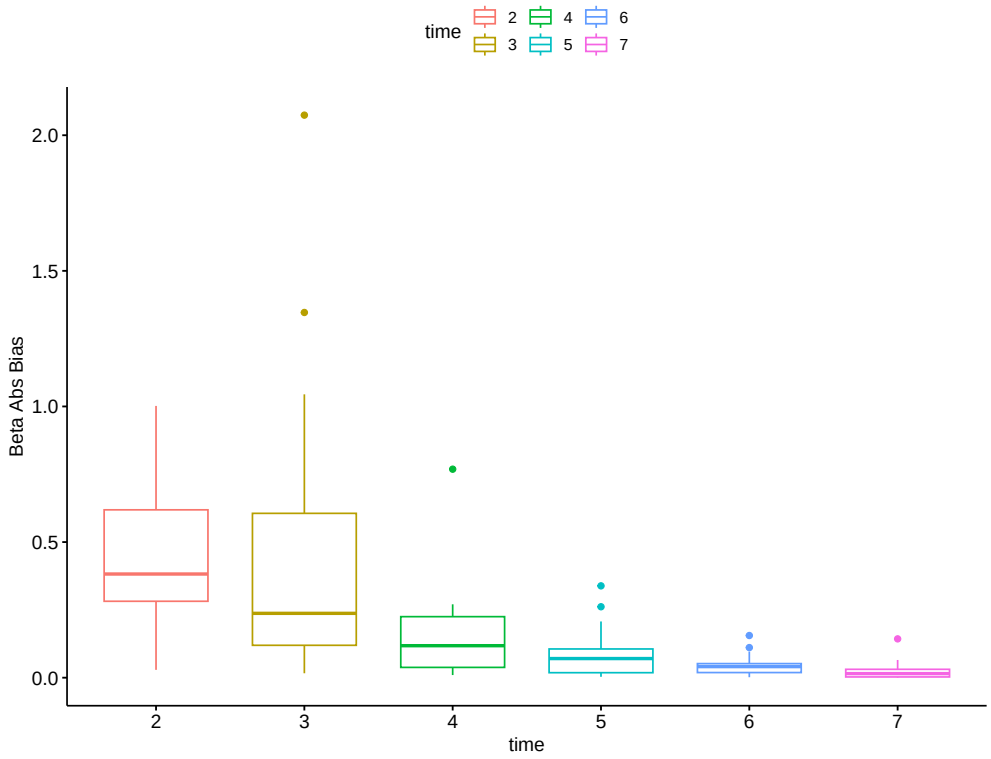


Figure A.63: Beta Absolute Bias of Temporal 1 Subset-m method for $n = 100$

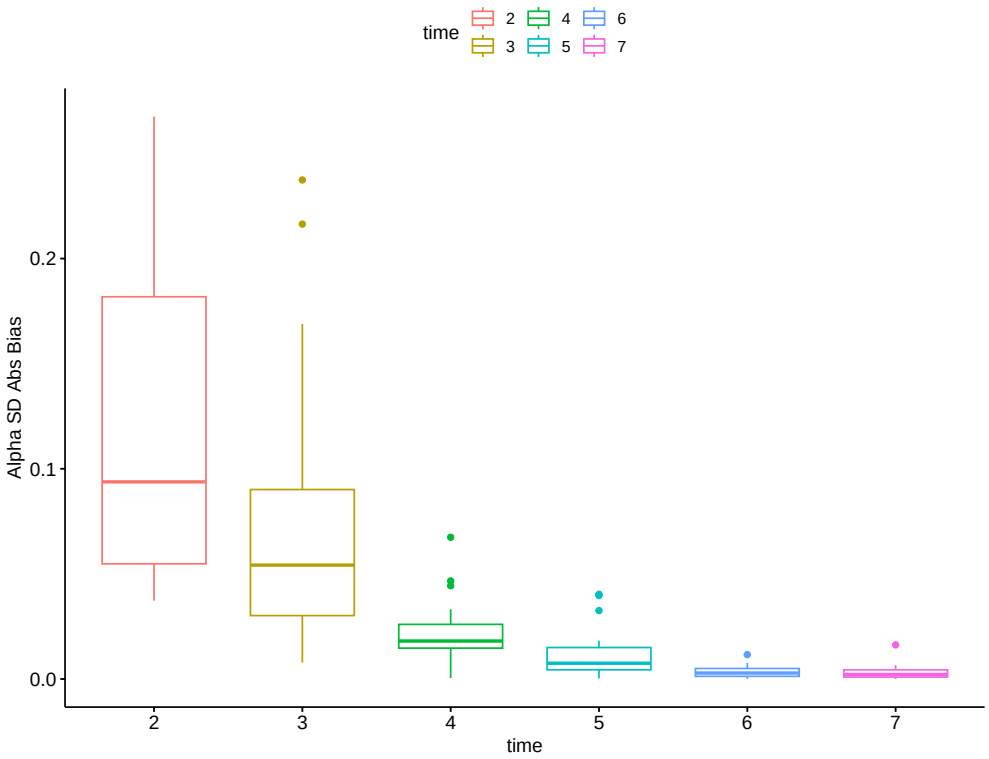


Figure A.64: Alpha SD Absolute Bias Temporal 1 Subset-m method for $n = 100$

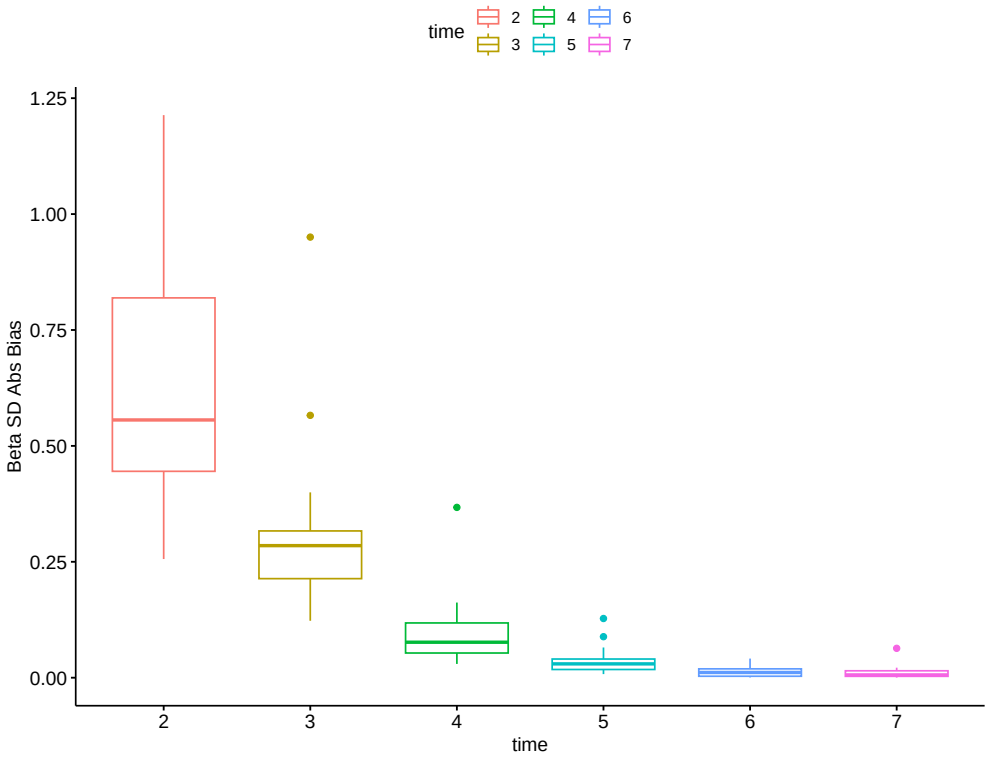


Figure A.65: Beta SD Absolute Bias of Temporal 1 Subset-m method for $n = 100$

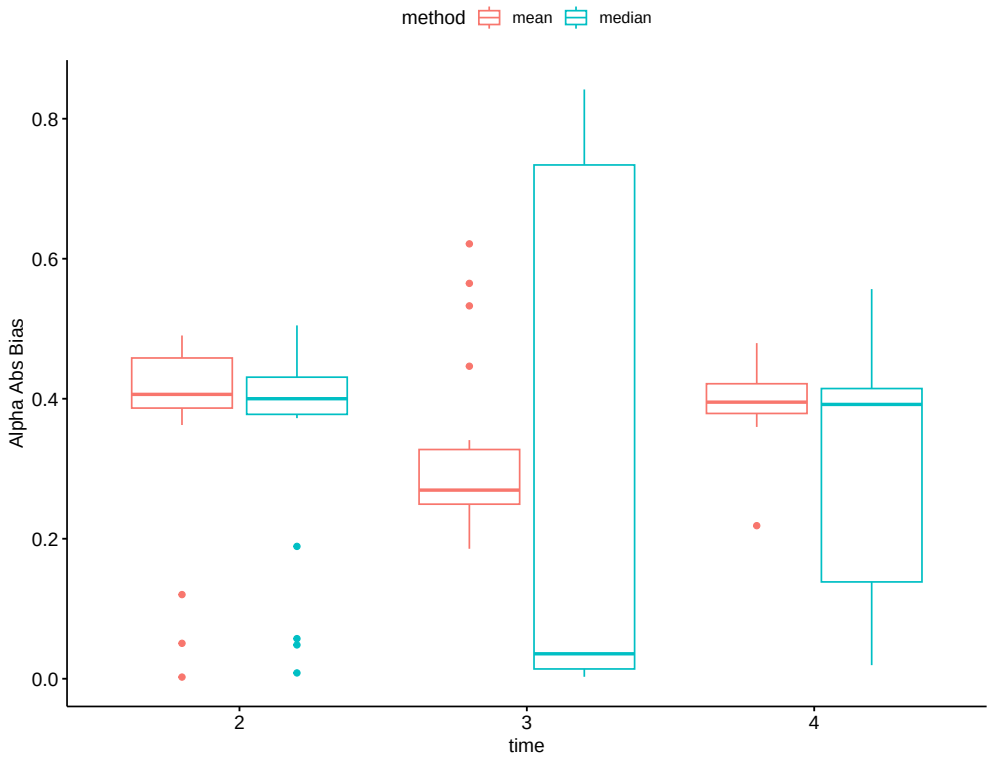


Figure A.66: Alpha Absolute Bias of Temporal Mean Subset-h and Temporal Median Subset-h methods for $n = 100$

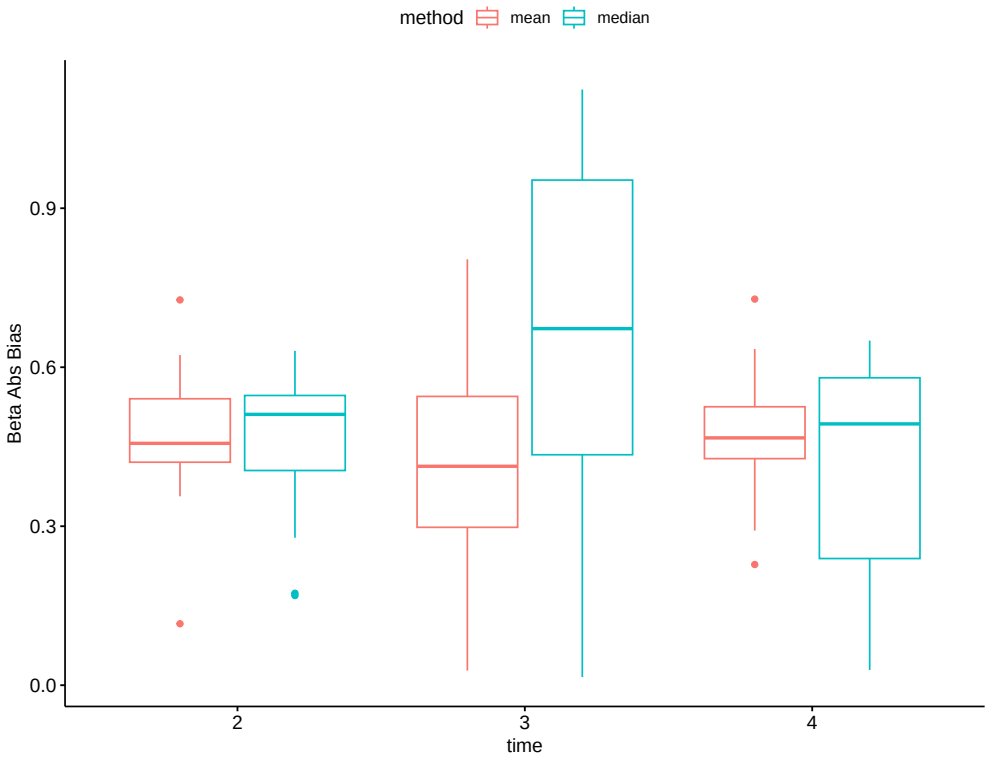


Figure A.67: Beta Absolute Bias of Temporal Mean Subset-h and Temporal Median Subset-h methods for $n = 100$

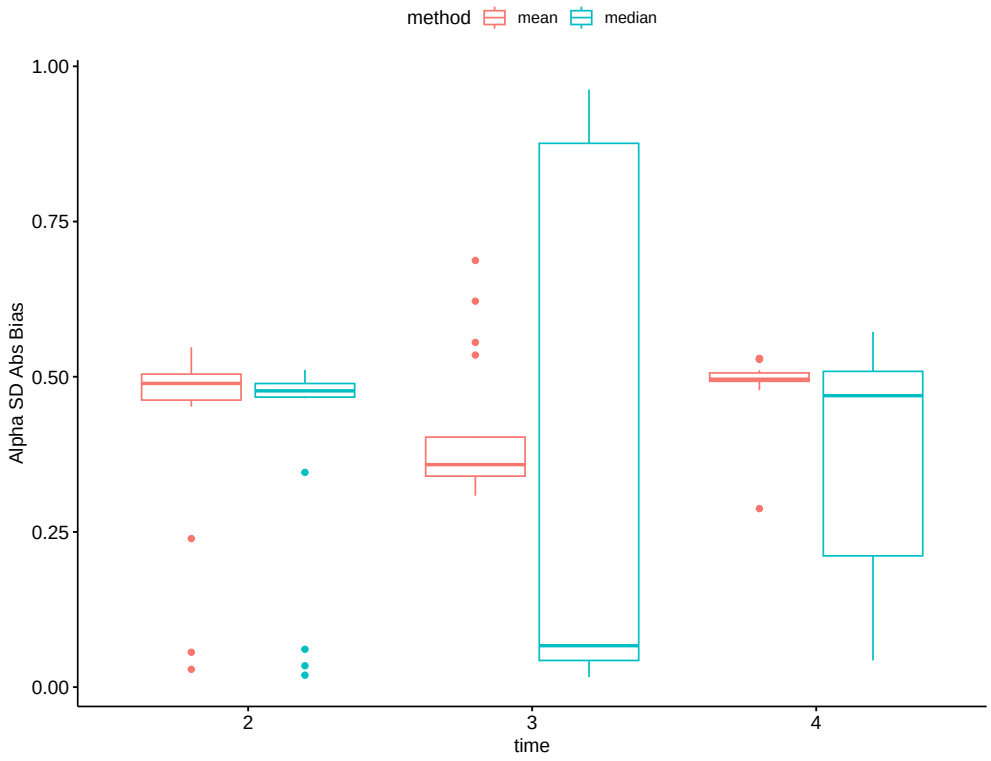


Figure A.68: Alpha SD Absolute Bias of Temporal Mean Subset-h and Temporal Median Subset-h methods for $n = 100$

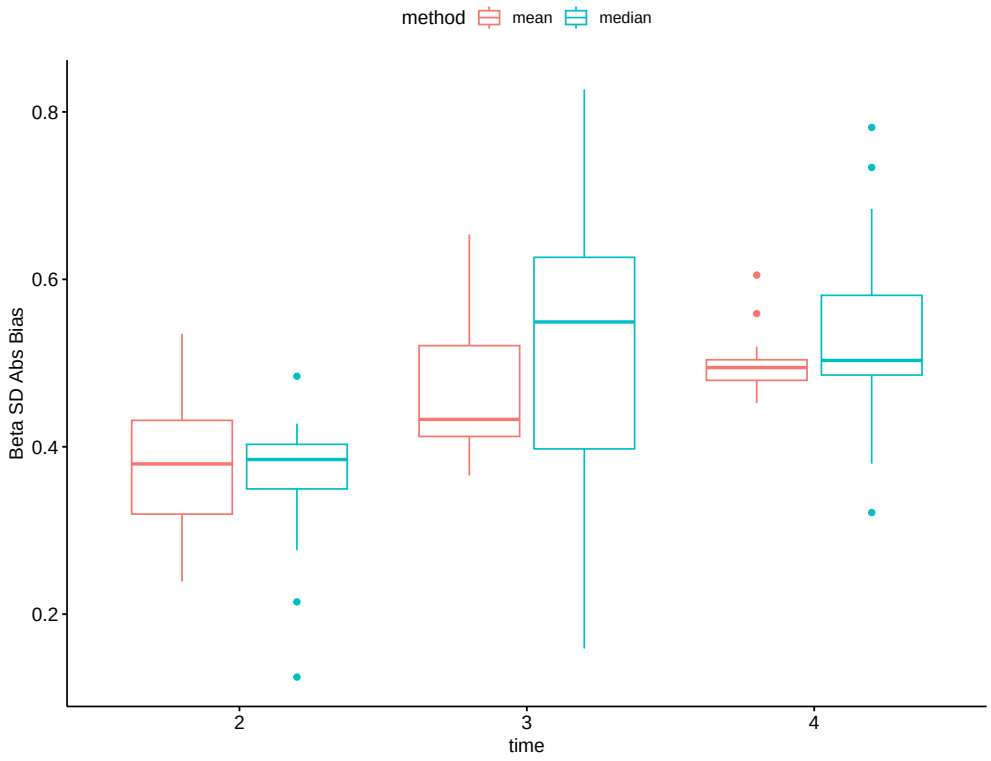


Figure A.69: Beta SD Absolute Bias of Temporal Mean Subset-h and Temporal Median Subset-h methods for $n = 100$