Statistics and Applications {ISSN 2454-7395 (online)}
Special Issue in Memory of Prof. C R Rao
Volume 22, No. 3, 2024 (New Series), pp 361–399
http://www.ssca.org.in/journal

Bayesian Predictive Inference for Nonprobability Samples with Spatial Poststratification

Dhiman Bhadra[1] **and Balgobin Nandram**[2]
[1]*Operations and Decision Sciences Area, Indian Institute of Management Ahmedabad, Gujarat - 380015, India*
[2]*Department of Mathematical Sciences, Worcester Polyechnic Institute, Worcester, Massachusetts - 01609, USA*

Received: 27 April 2024; Revised: 07 September 2024; Accepted: 10 September 2024

Abstract

Non-probability sampling involves selecting samples from a population in which the probability of selection is unknown and some population units may have zero selection probabilities. This differentiates it from probability sampling where selection is governed by a probability model and every population unit has a non-zero chance of being selected. Non-probability samples usually suffer from selection bias and hence may not represent the target population accurately. An important problem that arises in this context is the prediction of responses corresponding to non-sampled units, which should ideally have been sampled. In this article, we propose three modeling frameworks to address this issue. We use propensity scores to balance the sampled and non-sampled units and a Bayesian estimation scheme for parameter inference and prediction. We incorporate a spatial poststratification scheme to assess the predictive ability of our models on a simulated dataset. In addition, we perform model selection routines to identify the optimal model having the best predictive ability.

Key words: Beta-Bernoulli; Metropolis Hastings sampler; Non-probability samples; Propensity scores; Spatial poststratification.

AMS Subject Classifications: 62K05, 05B05

1. Introduction

One of the most important aspects of any statistical investigation is the formulation of a realistic and objective plan for data collection. These data should ideally be derived from a sample that is a good representation of the target population in the sense that it reflects all the conspicuous categories of the population adequately. Traditionally, the selection of such samples is guided by an underlying probabilistic mechanism which ensures that each and every population unit has a positive probability of being selected. The most well known of these selection mechanisms is the so called simple random sampling, which has the property that every sample of size, say n, has the same chance of being selected. This implies that each

Corresponding Author: Dhiman Bhadra
Email: dhiman@iima.ac.in

population unit has the same chance of being selected in the final sample. As a result, this kind of sample is known as a probability sample and the corresponding plan is designated as a probability sampling plan. Commonly used sampling mechanisms such as stratified, cluster or systematic sampling (Neyman, 1934) and their combinations are all grounded in the principle of probability sampling as opposed to non-probability sampling.

However, obtaining a truly parsimonious and representative probability sample is often prohibitively difficult in a real setting due to various constraints. Even if such a sampling scheme is implemented, it is a formidable task to obtain the requisite responses from the selected sample units. In fact, response rates of major surveys have been declining rapidly, casting doubts on the validity of probability samples as a proper representation of the population. According to Pew Research Center, the response rates in typical telephone surveys dropped from 36% in 1997 to only 9% in 2012 (Kohut *et al.*, 2012). Such low response rates, coupled with the complexity of implementation of probability-based survey designs raise serious doubts as to the viability of such sampling frameworks in real-life settings.

The above considerations along with an explosion of data being generated through various channels have led to an upsurge in the usage of non-probability sampling schemes. These schemes, as the term suggests, does not involve any underlying probabilistic mechanism for implementation. As a result, such schemes are convenient to use and hence are also referred to as convenience sampling schemes. Inferences from such samples are generally model based. However, as population units "self-select" themselves, the samples so obtained, often suffer from selection bias. This often results in the sample being non-representative of the target population in the sense that the sample may fail to incorporate all the relevant segments of a target population in the correct proportion. For example, in an email survey, only those who are willing to participate respond, probably having particular demographic characteristics. As a result, the demographic characteristics of those who do not participate are under-represented in the sample. Having said that, there is a subtle difference between selection bias and undercoverage in which certain sections of the population have absolutely no representation in the sample. In other words, it can be said that undercoverage is an extreme form of selection bias where a certain section of the population have absolutely zero chance of being selected in the sample. In this context, we would like to state that the proposed modeling frameworks have been designed to account for selection bias, not necessarily undercoverage.

In order to explore the applicability of non-probability sampling schemes for sampling from finite populations, the American Association of Public Opinion Research (AAPOR) constituted two task forces, neither of which favoured their use (Baker *et al.*, 2013). It was also suggested that inferences about a population drawn from a non-probability sample is valid subject to the verification of the modeling assumptions underlying the sampling scheme, a rather difficult proposition. The report also outlined various forms of non-probability samples such as convenience, snowball, network, mall-intercept and volunteer samples. One common aspect of all these schemes is the non-probabilistic aspect of sample selection, which results in biases, as mentioned before. Techniques for controlling biases have also been proposed such as sampling match which involves selecting non-probability sample units such that their characteristics match those in the population. This leads to the reduction of selection bias specially when the distribution of covariates used for matching are similar for the non-probability sample and the target population. A modified matching principle can

be adopted for observational studies in which the non-probability sample units are matched with those in a probability sample. Each unit in the non-probability sample can then be assigned a weight as a form of quantification of its degree of matching with the probability sample (Rubin, 1979). Rosenbaum and Rubin (1983) illustrated the use of propensity scores in the context of observational studies when the distribution of covariates is different in the treatment and control groups. This technique can be adopted for non-probability sampling as well, since the covariate characteristics may differ between the sampled and non-sampled groups. An extensive overview of matching procedures for causal inference and their applicability in diverse fields have been provided by Stuart (2010).

Smith (1983) introduced the notion of non-probability sampling and discussed general approaches for making inference from such samples. The basic formulation outlined therein is to model the joint distribution of the response observations and the selection probabilities of the population units. This formulation resembles the works of Rubin (1976), Little (1982) and Little and Rubin (2002) on selection mechanisms and survey responses. Smith (1983) also introduced the concept of poststratification and discussed its application on quota sampling. In the context of the above framework, Elliott and Valliant (2017) proposed two specific approaches of inference from non-probability samples, namely quasi-randomization and super-population. The underlying idea for these two approaches is to decompose the aforementioned joint distribution into the product of a conditional distribution of the response vector given that of the selection probabilities and the distribution of the response vector given the covariates. Quasi-randomization involves modelling the first component and estimating the selection probabilities as a way of correcting for the selection bias. On the other hand, the superpopulation approach involves modeling the second component. Although both approaches involve modeling, those are fundamentally different in their character. However, both approaches are aimed at nullifying or correcting for the effect of selection bias so as to make the resulting non-probability sample a better representation of the population.

One approach is to use propensity scores to estimate the survey weights of the non-probability sample and then proceed as in a regular probability sample; see Elliott and Valliant (2017) for an informative review of quasi-randomization and the super-population approach for non-probability samples. Chen *et al.* (2020) supplemented a non-probability sample with a probability sample using only the observed covariates to estimate propensity scores via logistic regression. Another approach is to use a nonignorable selection model to remove the selection bias; see Smith (1983) for pioneering work in this direction. Xu and Nandram (2019) used this approach to obtain full Bayesian analyzes; the references therein provide a historical development of this area. It is difficult to make valid inference from a non-probability sample with considerable selection bias. After all, a probability sample is the gold standard (high quality), but a non-probability sample is likely to have low quality (large bias, large mean squared error but unrealistically small variance). The key problem of a non-probability sample is that it is very likely to lead to seriously biased estimates of finite population quantities. Therefore, the large well-documented literature on selection bias is pertinent in the study of non-probability samples; these articles are too numerous to mention here; see Xu *et al.* (2020) and Choi *et al.* (2021) for recent applications, and the references therein.

It is also possible to make inference about a finite population quantity using a sin-

gle non-probability sample only; see Rao (2021) for a discussion. Supplementing a non-probability sample with a small probability sample has recently received some attention. But it is not quite practical to run a small probability survey in parallel with a non-probability sample. Therefore, if one can make accurate finite population inference from a non-probability sample only, this can be useful and economical. After all it costs money and time to design and field even a small survey, and it is less practical that both a non-probability sample and a probability sample will be available at the same time.

Survey samplers have long been using probability samples from one or more sources in conjunction with census and administrative data to make valid and efficient inferences on finite population parameters. This topic has received a lot of attention more recently in the context of data from non-probability samples such as transaction data, web surveys and social media data. Rao (2021) reviewed various probability survey methods that are used to make valid inferences about finite population parameters. This allowed him to show how these models can be extended to non-probability samples that can lead to "valid inferences by themselves or when combined with probability samples". Beaumont (2020) also reviewed some approaches that can "reduce, or even eliminate the use of probability surveys, all while preserving a valid statistical inference framework". However, naive use of such data can lead to serious sample selection bias and without adjustment to reduce selection bias it can lead to the "big data paradox: the bigger the data, the surer we fool ourselves" (Meng, 2018). Inevitably, non-probability samples will be more widely used in the future, and we need to continue researching methods for obtaining valid (or at least acceptable) inferences from them, possibly in combination with probability samples as illustrated in several papers. Falling response rates and increasing respondent burden are often given as reasons for using non-probability samples, especially in socioeconomic surveys. Robustness to model misspecification is also important in non-probability samples; see, for example, Marella (2023) and Rafei *et al.* (2022).

It is possible to use post-stratification to make satisfactory inference from a non-probability sample only, and it is convenient to do so. It is not necessary to estimate directly the selection probabilities for the non-probability sample; see Little (1993), Wang *et al.* (2015), Wang *et al.* (2021), Nandram and Choi (2005, 2010). Propensity scores are used to stratify the population, and they are not used as survey weights. However, too many strata can lead to sparseness and some strata can be empty. Cochran and Chambers (1965) suggested an optimal number of five strata (using quintiles); while this is good for small samples, it may not be so good for large samples. For larger samples, we can use more thinning, and a larger number of strata might be more efficient, say ten strata (using deciles). We may not know the nonsampled covariates, but the minimal we can assume is that the population size and the average covariates are known, a practical scenario. It is then possible to generate surrogates of the the nonsampled covariates using a bootstrap procedure.

Here, we are not concerned with data integration nor small area estimation. But there is also an emerging area in this direction; see Nandram and Rao (2024), Nandram and Rao (2023), Nandram and Rao (2021) and Nandram *et al.* (2021) for a Bayesian approach using propensity scores to estimate the selection probabilities with assistance from a small probability sample; there are other Bayesian approaches such as Sakshaug *et al.* (2019), Wiśniowski *et al.* (2020), Salvatore *et al.* (2024) and Rafei *et al.* (2022), who used the non-probability sample to supplement the probability sample. There has been a non-Bayesian

literature also. One leading paper is Chen *et al.* (2020), who use the design approach with double robustness and asymptotic theory. However, models will be better, if in fact, inference about the finite population parameters is robust to the of assumptions of the models for both the study variable and the participation variable. Other non-Bayesian approaches are discussed by Elliot (2009), Elliott and Haviland (2007) and Robbins *et al.* (2021).

The primary objective of this article is to propose methodologies aimed towards reducing selection bias when there exists significant difference in the characteristics between sampled and non-sampled units. A related problem that will be addressed is the prediction of the response observations for the non-sampled units for given values of their covariates. In doing so, we will build on established ideas, for instance, post-stratification (Smith, 1983), superpopulation approach (Elliot and Valliant, 2017) and propensity scores matching (Rosenbaum and Rubin, 1983). The crux of our methodology will be to treat the non-sampled units vis-a-vis their response as missing data (Rubin, 1976; Little and Rubin, 2002). Propensity scores and post-stratification will be used to balance the covariate distributions between the sampled and non-sampled groups. Lastly, we will perform the analysis in a hierarchical Bayesian setup through Markov chain Monte Carlo methodology. Application of Bayesian methodology in the context of non-probability sampling is a relatively unexplored domain. A recent article in this space is by Sakshaug *et al.* (2019) which examines the exchangeability of probability and non-probability sampling schemes by supplementing small probability samples with non-probability ones in a Bayesian paradigm. Their proposed method is applied simultaneously on probability and non-probability surveys and is shown to reduce the variance and mean squared error of model based predictions corresponding to non-probability samples relative to probability-only samples. However, the novel aspect of the methodology proposed in this article is the integration of a spatial dimension in the model for binary responses which enables us to better predict the response for the non-sampled units. Having said so, we must emphasize that our target of inference and/or prediction is the finite population proportion of success in the non-sampled group. However, we believe that our framework can be effortlessly extended to estimate population means in general, arising from continuous or discrete response variables by broadening the distributional structure of the said variables.

This paper is organized as follows. In Section 2, we describe the simulation mechanism for generating test data. In Section 3, we outline a modeling framework based on the Beta-Bernoulli distribution for the purpose of prediction of responses for non-sampled units. In Section 4, we introduce a modified model that incorporates a spatial dimension to the existing modeling framework. In Section 5, we propose another spatial model that leads to more precise prediction of responses for the non-sampled units. For each of these frameworks, we discuss the mechanism of estimation and prediction in a hierarchical Bayesian setup. In Section 6, we discuss some diagnostic measures for comparing the relative predictive abilities of the aforementioned models, followed by concluding remarks and a discussion of future work in Section 7.

2. Data simulation

As mentioned before, one of the principal characteristics of non-probability sampling is that the distribution of covariates is different for the sampled and non-sampled groups. This will be the basis for our simulation exercise aimed at generating the dataset on which

our proposed methodologies will be tested later. For the purpose of simulation, let the population size be 10,000, denoted as N, while the size of the sampled group be 1000, denoted as n. We consider four (4) covariates, namely Age (X_1), Race (X_2), Gender (X_3) and Education level (X_4) and a response Y such that

$$X_{2i} = \begin{cases} 1 & \text{if } i^{th} \text{ subject is white} \\ 0 & \text{if } i^{th} \text{ subject is black;} \end{cases} \quad X_{3i} = \begin{cases} 1 & \text{if } i^{th} \text{ subject is male} \\ 0 & \text{if } i^{th} \text{ subject is female;} \end{cases}$$

$$X_{4i} = \begin{cases} 1 & \text{if } i^{th} \text{ subject's education is college or higher} \\ 0 & \text{if } i^{th} \text{ subject's education is highschool or lower;} \end{cases}$$

$$Y_i = \begin{cases} 1 & \text{if } i^{th} \text{ subject's response is Yes} \\ 0 & \text{if } i^{th} \text{ subject's response is No,} \end{cases}$$

where $i = 1, ..., N$. Finally, we assume the following distributions for the covariates

$$\begin{aligned} X_{1i} &\sim N(55, 5^2), & X_{1i} &\sim N(65, 5^2); \\ X_{2i} &\sim \text{Bernoulli}(0.3), & X_{2i} &\sim \text{Bernoulli}(0.5); \\ X_{3i} &\sim \text{Bernoulli}(0.5), & X_{3i} &\sim \text{Bernoulli}(0.4); \\ X_{4i} &\sim \text{Bernoulli}(0.5), & X_{4i} &\sim \text{Bernoulli}(0.6), \end{aligned}$$

where the distributions in the first (left) column correspond to the subjects in the sampled group ($i = 1, ..., n$) while those in the second (right) column correspond to those in the non-sampled group ($i = n+1, ..., N$). The above choice of parameters was guided by the fact that the distributions of each covariate for the sampled and non-sampled groups should not be too different. This is critical, because in the poststratification step to be implemented next, it is necessary for every stratum to have some sampled units. In other words, if the distributions of particular covariates in the sampled and non-sampled groups are very different, there may be stratum which will be devoid of any units from the sampled group. If so, it would not be possible for us to predict the response of the non-sampled units for that stratum.

Finally, we assume that $Y_i|p_i \overset{ind}{\sim} \text{Bernoulli}\ (p_i)$. Once the above covariate values are simulated, we generate the probability of success (*i.e..*, $Y_i = 1$) using the following logistic regression function

$$p_i = P(Y_i = 1) = \frac{e^{\alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{4i} + \epsilon_i}}{1 + e^{\alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{4i} + \epsilon_i}}, \quad i = 1, 2, ..., N,$$

where the ϵ follow a standard normal distribution *i.e..* $\epsilon_i \sim N(0, 1)$. We assume $(\alpha_0, \alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0.1, 0.01, 4, -5, -1)$. Once the N simulated values of p_i are obtained, the corresponding values of Y_i are drawn from Bernoulli (p_i). Table 1 depicts part of the simulated data.

Here R_i is such that

$$R_i = \begin{cases} 1 & \text{if unit } i \text{ belongs to the sampled group} \\ 0 & \text{if unit } i \text{ belongs to the non-sampled group,} \quad i = 1, 2, ..., N. \end{cases}$$

It is important to note that $Y_i, (i = 1001, ..., 10,000)$ will be assumed to be unobserved since they relate to the non-sampled units. However, the covariates, X_i's are always observed.

Table 1: Simulated data set of the population

i	R_i	X_{1i}	X_{2i}	X_{3i}	X_{4i}	Y_i
1	1	48	0	0	0	1
2	1	63	1	0	1	1
3	1	50	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
999	1	54	0	0	0	0
1000	1	56	1	1	1	0
1001	0	47	1	1	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
9998	0	64	1	0	0	1
9999	0	66	1	0	0	1
10000	0	59	1	0	0	1

Our sole purpose will be to predict these unobserved responses corresponding to the non-sampled units using data from the sampled units. Towards that end, we will formulate various modeling frameworks and will apply those on the above data. These will be illustrated in the next sections.

3. Non-spatial model

Here we describe the general approach to predict the finite population proportion using a non-spatial model.

3.1. Model specification

In order to specify the model framework, we need to define the propensity scores in the context of our setup. The propensity score for a subject/entity is the conditional probability of it being selected in a sample given its covariates. The foundational assumption in this regard is that all pertinent covariates related to the sample units are included in the study. Supposing $\boldsymbol{x}_i$ is the covariate vector corresponding to the i^{th} subject in the population, its propensity score, $\pi(\boldsymbol{x}_i)$ is given by

$$\pi(\boldsymbol{x}_i) = P(R_i = 1|\boldsymbol{x}_i, \boldsymbol{\phi}), \quad i = 1, 2, ..., N, \tag{1}$$

where R_i has been defined in Sec 2 and $\boldsymbol{\phi}$ is a vector of unknown parameters. We use a logistic regression model to model $\pi(\boldsymbol{x}_i)$ *i.e.*.

$$\pi(\boldsymbol{x}_i) = \frac{e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}{1 + e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}, \tag{2}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$ is our target of inference. We assume a non-informative prior on $\boldsymbol{\beta}$ *i.e.*. $\pi(\boldsymbol{\beta}) = 1$. Assuming independence, the conditional distribution of R_i is given by

$$R_i|\boldsymbol{\beta} \sim \text{Bernoulli}\{\pi(\boldsymbol{x}_i)\}.$$

Accordingly, the posterior density of $\boldsymbol{\beta}$ can be expressed as

$$\pi(\boldsymbol{\beta}|\boldsymbol{R}) \propto \prod_{i=1}^{N}\left(\frac{e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}\right)^{R_i}\left(1-\frac{e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}\right)^{1-R_i} = \prod_{i=1}^{N}\left(\frac{e^{R_i\boldsymbol{x}_i'\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_i'\boldsymbol{\beta}}}\right). \quad (3)$$

It is important to note that our target of inference in this case is the finite population proportion *i.e.*. $\frac{1}{N}\sum_{i=1}^{N} Y_i$, where the sample values $(Y_i, i = 1, 2, ..., n)$ are observed and the non-sample values $(Y_i, i = n+1, ..., N)$ are missing. In the context of non-probability sampling, the missing data mechanism can be assumed to be missing-at-random (MAR), given the covariates (Little and Rubin, 2002). However, this is not a binding condition since inference can be performed on non-probability samples accommodating for both nonignorable nonresponse and selection biases (Nandram and Choi, 2010; Nandram, 2022).

3.2. Bayesian computation

Since the above posterior is not in closed form, we will need to perform the Metropolis-Hastings algorithm (Hastings, 1970) in order to draw samples from it. For that purpose, we need to define a suitable proposal density. We use Laplace approximation for that purpose. The advantage of the Laplace approximation is that for small degrees of freedom, it has increased flexibility to accommodate skewness, thus enhancing its effectiveness as an approximation. It is worth noting that the Laplace approximation is simply used as a proposal density (first approximation) in the Metropolis sampler. Accordingly, we assume that $\boldsymbol{\beta}$ approximately follows a multivariate t distribution parametrized as

$$\boldsymbol{\beta}|\sigma^2 \sim N(\hat{\boldsymbol{\beta}}, \gamma^2\hat{\Sigma}); \ \frac{\nu}{\gamma^2} \sim \chi^2_\nu$$
$$\text{where} \quad \hat{\Sigma} = -(H(\hat{\boldsymbol{\beta}}))^{-1}, \ \hat{\boldsymbol{\beta}} \text{ being the mode of } \boldsymbol{\beta}.$$

Here ν is the degrees of freedom of the multivariate t distribution and acts as a tuning parameter. The values of $\hat{\boldsymbol{\beta}}$ and $-(H(\hat{\boldsymbol{\beta}}))^{-1}$ are obtained through numerical approximation. For posterior simulation, we use Metropolis-Hastings algorithm with the following candidate density for $\boldsymbol{\beta}$

$$p(\boldsymbol{\beta}) \propto \frac{1}{\left[1+\frac{(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})'\hat{\Sigma}^{-1}(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}})}{\nu}\right]^{\frac{5+\nu}{2}}}.$$

We first draw 10,000 sets of $\boldsymbol{\beta} = (\beta_0, ..., \beta_4)$. Then we drop the first $B = 5000$ iterates and take every 5^{th} of the remaining iterates *i.e.*. we take iterate number $B+1, B+1+k, B+1+2k, ..., B+1+m\times k$ where $k = 5$ and $m = 1000$ being the final sample size. Table 2 depicts the posterior summaries of $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$ obtained from the simulated samples, which are provided under Section 3.2 in the Annexure.

We use various diagnostics to assess the convergence of the chains, like trace and autocorrelation plots, Geweke test and effective sample sizes. The trace and auto-correlation plots are shown under Section 3.2 in the annexure. The plots and the diagnostics tests indicate satisfactory mixing and convergence of the chains. In the context of simulated data, the above estimates indicate that all the predictors have a significant effect on the response, R_i. Specifically, being younger, being black, being male or having a high school or lower degree significantly increase the odds of being sampled.

Table 2: Posterior summaries of β for Beta-Bernoulli model

Parameter	Mean	Standard Deviation	95% HPD Interval
β_0	9.957	0.448	(9.089, 10.775)
β_1	-0.196	0.006	(-0.219, -0.189)
β_2	-0.785	0.073	(-0.935, -0.652)
β_3	0.264	0.071	(0.122 0.398)
β_4	-0.416	0.074	(-0.551, -0.261)

3.3. Poststratification and prediction

As mentioned before, our principal aim is to predict the responses for the non-sampled subjects using data from the sampled subjects. Towards that end, it is imperative to balance (or adjust) the covariate distributions between the sampled and non-sampled groups. We will achieve that through a combination of propensity scores and poststratification (Baker *et al.*, 2013) as depicted by Rubin (1979) and in Nandram and Choi (2010) who applied it in the analysis of body mass index data in a small area context.

3.3.1. Poststratification

The poststratification procedure will be described in this section.

For the h^{th} set of simulated values of $\boldsymbol{\beta}$, the corresponding propensity score values are given by

$$\pi_i^{(h)} = \frac{e^{\boldsymbol{x}_i'\boldsymbol{\beta}^{(h)}}}{1 + e^{\boldsymbol{x}_i'\boldsymbol{\beta}^{(h)}}}, \quad h = 1, 2, ..., 1000; \;\; i = 1, 2, ..., 10,000.$$

Thus, we will have $m = 1000$ propensity score values for each of $N = 10,000$ simulated population units resulting in a $N \times m = 10,000 \times 1000$ matrix of propensity scores. Part of this matrix is shown under Sec 3.3.1 in the Annexure. Given the simulated values of the propensity scores, we create ten (10) strata by forming ten intervals from their deciles for implementing the poststratification procedure. The ten intervals are shown in Table 3, where I_j denotes the j^{th} interval. Now, for each simulated value of $\boldsymbol{\beta}$, we allocate the 10,000 population units into these strata/intervals based on their respective propensity score value. Table 4 depicts the number of subjects allocated to each of these strata corresponding to the sampled and non-sampled groups for four simulated values of $\boldsymbol{\beta}$. Note that the sample frequencies vary across the sub-strata because the deciles are based on $10^7(10,000 \times 1000)$ propensity score values.

Table 3: Propensity score intervals

I_1	I_2	I_3	I_4	I_5
(0, 0.0143]	(0.0143, 0.0237]	(0.0237, 0.0337]	(0.0337, 0.0459]	(0.0459, 0.0613]
I_6	I_7	I_8	I_9	I_{10}
(0.0613, 0.0815]	(0.0815, 0.1110]	(0.1110, 0.1550]	(0.1550, 0.2366]	(0.2366, 0.9302]

Table 4: Stratum allocation frequencies for sampled and non-sampled groups for different values of $\boldsymbol{\beta}$

		Stratum Frequency									
$\boldsymbol{\beta}$	Group	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}
	Sampled	8	20	28	36	36	70	105	133	225	339
$\boldsymbol{\beta}^{(1)}$	Non-sampled	985	916	955	1050	825	1042	911	874	820	622
	Sampled	8	30	23	38	38	77	92	115	240	339
$\boldsymbol{\beta}^{(2)}$	Non-sampled	1033	1208	852	967	825	1066	712	805	910	622
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	Sampled	9	20	32	35	48	59	101	134	205	357
$\boldsymbol{\beta}^{(999)}$	Non-sampled	1012	947	1057	967	960	926	846	823	783	679
	Sampled	13	25	27	37	40	74	83	133	206	362
$\boldsymbol{\beta}^{(1000)}$	Non-sampled	1203	1038	985	924	890	833	781	871	774	701

It can be easily verified that the cumulative frequencies of the sampled and non-sampled units across all the strata/intervals are 1000 and 9000 respectively for all simulated values of $\boldsymbol{\beta}$. Conditional on the above poststratification, the covariate distribution for the sampled and non-sampled groups can be assumed to be similar for each stratum. Hence, for each stratum/interval, we can predict the response for the non-sampled units using data for the sampled units. Prediction will be carried out using the superpopulation approach mentioned in Section 1 by modeling the conditional density of the response vector (say, $\boldsymbol{Y}_s$) given the covariate vector (say, $\boldsymbol{X}_s$) for the sampled group respectively.

3.3.2. Prediction

The prediction procedure is described in this section.

Let y_{ij} denote the response for the j^{th} unit in the i^{th} stratum for the sampled group, where $i = 1, 2, ..., 10$ and $j = 1, 2, ..., n_i$, n_i being the number of sampled subjects in stratum i. For example, for $\boldsymbol{\beta}^{(1)}$, the number of sampled subjects in the 1^{st} stratum is 8 *i.e.*. $n_i = 8$. Let p_i be the probability of success (*i.e.*. $y_{ij} = 1$) for the i^{th} stratum. We have the following model specification

$$\begin{aligned} Y_{ij}|p_i &\sim \text{Bernoulli}(p_i), \;\; i = 1, 2, ..., 10; j = 1, 2, ..., n_i, \\ p_i &\sim \text{Beta}(0, 0), \;\; i = 1, 2, ..., 10. \end{aligned} \tag{4}$$

The above prior is clearly improper and is also known as the Haldane prior. We have chosen this prior for p_i in order to make the inference as data driven as possible. The posterior of p_i is

$$\begin{aligned} \pi(p_i|\boldsymbol{y}_i) &\propto f(\boldsymbol{y}_i|p_i)\pi(p_i) = p_i^{\sum_{j=1}^{n_i} y_{ij}-1}(1-p_i)^{n_i-\sum_{j=1}^{n_i} y_{ij}-1} \\ i.e.. \; p_i|\boldsymbol{y}_i &\sim \text{Beta}\left(\sum_{j=1}^{n_i} y_{ij}, n_i - \sum_{j=1}^{n_i} y_{ij}\right). \end{aligned} \tag{5}$$

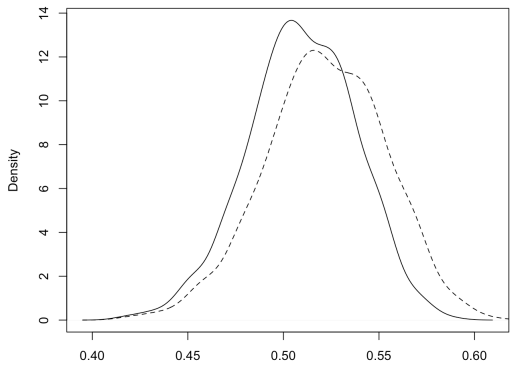
Here $\sum_{j=1}^{n_i} y_{ij}$ and $n_i - \sum_{j=1}^{n_i} y_{ij}$ are respectively the number of 1's and 0's of the response variable for the sampled data in the i^{th} stratum. As an illustration, for $\boldsymbol{\beta}^{(1)}$, there are 8

sampled units and 985 non-sampled units in the 1^{st} stratum. The response values for the 8 sampled units are $(1, 0, 0, 1, 1, 0, 1, 1)$ *i.e.*. $\sum_{j=1}^{n_i} y_{ij} = 5$ and $n_i - \sum_{j=1}^{n_i} y_{ij} = 3$. Hence, the posterior distribution of p_1 will be Beta(5, 3). Now, we can draw a random sample of p_1 from a Beta(5, 3) distribution (say $\tilde{p_1}$) and finally draw 985 values of y_{ij} from Bernoulli($\tilde{p_1}$). The simulated values of y so drawn will be the predicted values of Y corresponding to the 1^{st} stratum. Similarly, we can predict all the non-sampled response observations for all the strata.

Once the above prediction is complete, we compute the proportion of successes *i.e.*. $Y = 1$ for the sampled and non-sampled groups separately as well as for all the N subjects in the combined set corresponding to each $\boldsymbol{\beta}^{(h)}, h = 1, 2, ..., 1000$. These are given by

$$P_{all}^{(h)} = \frac{\sum_{k=1}^{N} Y_k^{(h)}}{N} \quad \text{and} \quad P_{ns}^{(h)} = \frac{\sum_{k=n}^{N} Y_k^{(h)}}{N-n}, \quad h = 1, 2, ..., 1000.$$

The true values of the above quantities for the sampled, non-sampled and all individuals taken together are 0.398, 0.509 and 0.498 respectively. The kernel density plots of the above quantities are given in Figure 1. Here the bold (dashed) curve correspond to $P_{all}^{(h)}(P_{ns}^{(h)})$ respectively.



(a) Combined (bold: $P_{all}^{(h)}$; dashed: $P_{ns}^{(h)}$)

Figure 1: Kernel density plots of the proportion of positive responses predicted for all individuals and non-sampled individuals for Beta-Bernoulli model

3.4. Model accuracy

To evaluate the accuracy of our prediction, we compute the 95% highest posterior density (HPD) intervals of $P_{all}^{(h)}$ and $P_{ns}^{(h)}$. If the true proportion values, reported above, lies within and near the centre of the above intervals, it would indicate an accurate fit. However, if the true value lies outside the intervals or towards the edge, that would be indicative of a sub-optimal fit. The HPD interval for the true population proportion of positive responses for all the sampled units taken together ($P_{all}^{(h)}$) is found to be (.450, .559) while that for the proportion of non-sampled subjects is (0.456, 0.577). In both cases, the true values *i.e.*. 0.498 and 0.509 falls within and near the centre of the corresponding intervals. This indicates that our prediction is pretty accurate.

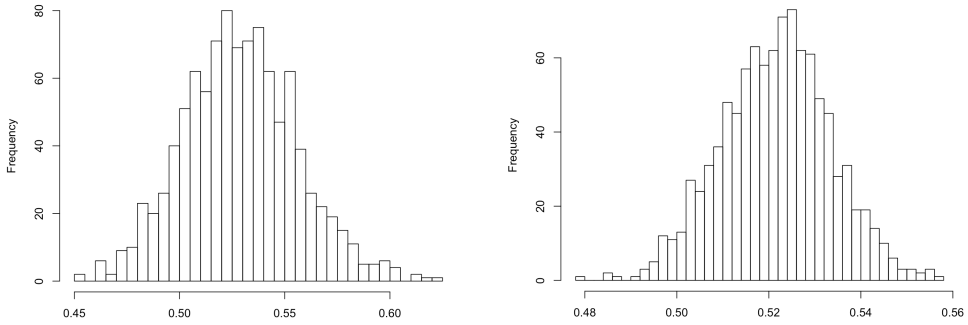
In addition to testing prediction accuracy, we also compare the predictive ability of our modeling framework, based on the superpopulation methodology, with that of quasi-randomization strategy, mentioned in Section 1, using the Horvitz-Thompson (Horvitz and Thompson, 1952) and the Hajek estimators of the population proportion (of positive responses for all the N units). These are given by

$$\hat{P}_{HT}^{(h)} = \frac{\hat{Y}_{HT}^{(h)}}{N} = \frac{1}{N}\sum_{i=1}^{n}\frac{Y_i}{Pr_i^{(h)}} \text{ and } \hat{P}_{H}^{(h)} = \frac{\sum_{i=1}^{n} Y_i/Pr_i^{(h)}}{\sum_{i=1}^{n} 1/Pr_i^{(h)}}, \quad h = 1, 2, ..., 1000,$$

where "HT" and "H" in the suffix denotes "Horvitz-Thompson" and "Hajek" respectively while

$$Pr_i^{(h)} = \frac{n\pi_i^{(h)}}{\sum_{i=1}^{N}\pi_i^{(h)}}, \quad h = 1, 2, ..., 1000; \quad i = 1, 2, ..., N.$$

Here $n = 1000$, $N = 10,000$ while $\pi_i^{(h)}$ is the propensity score for the i^{th} subject corresponding to the h^{th} case. The histograms of the 1000 simulated values of $\hat{P}_{HT}^{(h)}$ and $\hat{P}_{H}^{(h)}$ for the above estimators are shown in Figure 2. The corresponding 95% H.P.D interval of the true population proportion (for all subjects taken together) are (0.478, 0.584) and (0.496, 0.543) respectively. In both cases, the true value, 0.498, falls within the above intervals but more towards one of the edges. This is specially true for the Hajek estimator as the true population proportion is nearly equal to the lower bound of the H.P.D interval, 0.496. Thus, we can conclude that our proposed model has superior predictive properties compared to the Horvitz-Thompson and Hajek estimators. This also indicates that the superpopulation approach fares better than the quasi-optimization approach in predicting the response values for the non-sampled units.



(a) Horvitz-Thompson (b) Hajek

Figure 2: Histograms of the proportion of positive responses for Horvitz-Thompson and Hajek estimators

4. Standard spatial model

The Beta-Bernoulli framework developed in Section 3 enabled us to predict the proportion of positive responses in the non-sampled group for each stratum. In doing so, it was

assumed that the strata are independent of each other (*i.e.*. are uncorrelated). However, since the boundaries of strata are fuzzy, subjects close to the edges of two adjacent strata may have non-negligible correlation. Hence, it may be worthwhile to develop a modeling framework taking into account the spatial relationship between neighbouring strata. In this section, we will develop a Bayesian hierarchical model that incorporates this spatial association. Accordingly, we would like to test whether incorporating this spatial dimension in the modeling framework improves the ability of the model to predict the responses for the non-sampled individuals. This is a novel contribution in non-probability sampling.

4.1. Hierarchical model specification

For the proposed spatial modeling framework, the data and stratum-specific model specification remain the same as for the Beta-Bernoulli model, depicted in (3.4). As mentioned in He and Sun (2000), we specify the following logistic mixed model for p_i,

$$\log\left(\frac{p_i}{1-p_i}\right) = \theta + \nu_i,$$

where p_i is the i^{th} stratum-specific success probability for the sampled group, θ is the fixed effect and ν_i is the i^{th} stratum-specific random effect. Following He and Sun (2000), we use a simultaneous conditional autoregressive model (SCAR) to specify the prior of ν_i. Towards this end, we define the following 10×10 symmetric adjacency matrix (as we have 10 strata)

$$\mathbf{C} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix},$$

where $\mathbf{C}_{jk} = 0(1)$ means the j^{th} and k^{th} strata are non-adjacent(adjacent) *i.e..*, does not share (share) a boundary. According to He and Sun (2000), the eigenvalues of the adjacency matrix $\mathbf{C}$, given by $\lambda = (\lambda_1, \lambda_2, ..., \lambda_{10})$, can be defined so that the following inequality holds

$$\tfrac{1}{\lambda_{min}} \leq \rho \leq \tfrac{1}{\lambda_{max}}.$$

where λ_{min} and λ_{max} are the minimum and maximum eigenvalues of C. Then, based on the SCAR properties, discussed by Clayton and Kaldor (1987), the prior distribution of $\boldsymbol{\nu}$ can be shown to be

$$\boldsymbol{\nu} \sim MVN(\mathbf{0}, \delta^2(\mathbf{I} - \rho\mathbf{C})^{-1}), \tag{6}$$

where $\boldsymbol{\nu} = (\nu_1, \nu_2, \nu_3, \nu_4, \nu_5, \nu_6, \nu_7, \nu_8, \nu_9, \nu_{10})$ and $\mathbf{I}$ is a 10×10 identity matrix. In order to determine the prior distributions for θ, we employ the empirical logistic transform (Cox,

2018). Suppose we have g sets of binary observations and in the j^{th} set ($j = 1, 2, ..., g$), the success probability, p_j, is constant for that set. Let there be n_j trials and M_j successes in those trials. Then the empirical logistic transform is defined as

$$Z_j = \log\left(\frac{M_j + \frac{1}{2}}{n_j - M_j + \frac{1}{2}}\right) \text{ with mean } \phi_j = \log\left(\frac{p_j}{1-p_j}\right).$$

As per Gart and Zweifel (1967), an approximate unbiased estimator of the variance of Z_j is given by

$$V_j^2 = \frac{(n_j+1)(n_j+2)}{n_j(M_j+1)(n_j-M_j+1)}.$$

In fact, it can be shown that Z_j approximately follows a normal distribution with mean ϕ_j and variance V_j^2 *i.e.*. $Z_j \sim N(\phi_j, V_j^2)$ (McCullagh, 2019). Using the above transformation, the prior distribution of θ can be expressed as

$$\pi(\theta) = \frac{1}{V\pi\left(1 + \left(\frac{\theta - \hat{\theta}}{V}\right)^2\right)}, \quad -\infty < \theta < \infty, \tag{7}$$

which is a location-scale Cauchy distribution. Now, $y_k \sim \text{Bernoulli}\left(\frac{e^\theta}{1+e^\theta}\right)$ for $k = 1, 2, ..., n$ implying that $\hat{p} = \bar{y} = 0.398$. Thus, $\hat{\theta} = \log\left(\frac{\hat{p}}{1-\hat{p}}\right) = -0.4138$. On the other hand, V is obtained as

$$V = \sqrt{\frac{(n+1)(n+2)}{n(M+1)(n-M+1)}} = 0.0646,$$

where $M = \sum_{k=1}^{n} y_k = 398$ which is the total number of positive responses in the sampled group. Finally, the prior for (δ^2, ρ) is given by

$$\pi(\delta^2, \rho) = \frac{1}{(1+\delta^2)^2}, \ \ \delta^2 > 0, \ \ \frac{1}{\lambda_{min}} \le \rho \le \frac{1}{\lambda_{max}}. \tag{8}$$

Combining the likelihood and priors specified in (6-9), the joint posterior density of $(\theta, \delta^2, \boldsymbol{\nu}, \rho)$ is given by

$$\pi(\theta, \delta^2, \boldsymbol{\nu}, \rho | \mathbf{Y}) \propto f(\mathbf{Y}|\theta, \nu_i)\pi(\boldsymbol{\nu}|\delta^2, \rho)\pi(\delta^2, \rho)\pi(\theta),$$

where

$$\begin{aligned} f(\mathbf{Y}|\theta, \nu_i) &= \prod_{i=1}^{10}\prod_{j=1}^{n_i}\left\{\frac{e^{\theta+\nu_i}}{1+e^{\theta+\nu_i}}\right\}^{y_{ij}}\left\{1 - \frac{e^{\theta+\nu_i}}{1+e^{\theta+\nu_i}}\right\}^{1-y_{ij}} \\ &= \prod_{i=1}^{10}\prod_{j=1}^{n_i}\left\{\frac{e^{(\theta+\nu_i)y_{ij}}}{1+e^{\theta+\nu_i}}\right\} \end{aligned}$$

and

$$\pi(\boldsymbol{\nu}|\delta^2,\rho) \;=\; \frac{1}{\sqrt{|\delta^2(I-\rho C)^{-1}|}}\exp\left\{-\frac{1}{2}\boldsymbol{\nu}^T(\delta^2(I-\rho C)^{-1})^{-1}\boldsymbol{\nu}\right\}.$$

Combining these forms, the joint posterior density becomes

$$\begin{aligned}\pi(\theta,\delta^2,\boldsymbol{\nu},\rho|\mathbf{Y}) \propto \prod_{i=1}^{10}\prod_{j=1}^{n_i}\left\{\frac{e^{(\theta+\nu_i)y_{ij}}}{1+e^{\theta+\nu_i}}\right\} &\times \frac{1}{\sqrt{|\delta^2(I-\rho C)^{-1}|}}\exp\left\{-\frac{1}{2}\boldsymbol{\nu}^T(\delta^2(I-\rho C)^{-1})^{-1}\boldsymbol{\nu}\right\}\\ &\times \frac{1}{(1+\delta^2)^2}\times\frac{1}{V\pi((1+(\frac{\theta-\hat{\theta}}{V})^2)},\end{aligned} \tag{9}$$

where $\delta^2>0, \hat{\theta}-10\times V<\theta<\hat{\theta}+10\times V$ (the entire unimodal density lies in a narrower interval) and $\frac{1}{\lambda_{min}}<\rho<\frac{1}{\lambda_{max}}$.

4.2. Bayesian computation

Based on the full posterior density specified in (9) above, the full conditional posterior densities are given by

$$\boldsymbol{\nu}|\theta,\delta^2,\rho,\mathbf{Y}\propto\prod_{i=1}^{10}\left\{\frac{e^{(\theta+\nu_i)R_i}}{[1+e^{\theta+\nu_i}]^{n_i}}\right\}\exp\left\{-\frac{1}{2}\boldsymbol{\nu}^T(\delta^2(I-\rho C)^{-1})^{-1}\boldsymbol{\nu}\right\}; \tag{10}$$

$$\theta|\boldsymbol{\nu},\delta^2,\rho,\mathbf{Y}\propto\prod_{i=1}^{10}\left\{\frac{e^{(\theta+\nu_i)R_i}}{[1+e^{\theta+\nu_i}]^{n_i}}\right\}\times\frac{1}{V\pi((1+(\frac{\theta-\hat{\theta}}{V})^2)}; \tag{11}$$

$$\delta^2|\theta,\rho,\boldsymbol{\nu},\mathbf{Y}\propto\frac{1}{\sqrt{\delta^2}}\exp\left\{-\frac{1}{2}\boldsymbol{\nu}^T(\delta^2(I-\rho C)^{-1})^{-1}\boldsymbol{\nu}\right\}\times\frac{1}{(1+\delta^2)^2}; \tag{12}$$

$$\rho|\delta^2,\theta,\boldsymbol{\nu},\mathbf{Y}\propto\frac{1}{\sqrt{|(I-\rho C)^{-1}|}}\exp\left\{-\frac{1}{2}\boldsymbol{\nu}^T(\delta^2(I-\rho C)^{-1})^{-1}\boldsymbol{\nu}\right\}. \tag{13}$$

Since the full conditionals are not in closed form, we need to use a combination of specialized sampling schemes to draw sample from those. Specifically, we use

1. Metropolis-Hastings algorithm to sample from $\pi(\boldsymbol{\nu}|\theta,\delta^2,\rho,\mathbf{Y})$.
2. Grid method to sample from the remaining three full conditionals, namely $\pi(\theta|\boldsymbol{\nu},\delta^2,\rho,\mathbf{Y})$, $\pi(\delta^2|\theta,\rho,\boldsymbol{\nu},\mathbf{Y})$ and $\pi(\rho|\delta^2,\theta,\boldsymbol{\nu},\mathbf{Y})$.

In the first case, we need to determine the candidate generating density to be able to apply Metropolis-Hastings algorithm. We apply the empirical logistic transformation towards this end. As per this procedure,

$$Z_i|\nu_i \sim N(\theta+\nu_i, V_i^2) \text{ where } V_i^2 = \frac{(n_i+1)(n_i+2)}{n_i(M_i+1)(n_i-M_i+1)} \text{ and } Z_i = \log\left(\frac{M_i+0.5}{n_i-M_i+0.5}\right).$$

This implies that

$$\boldsymbol{\nu} \sim \text{MVN}((\mathbf{Z}-\theta), \boldsymbol{\Sigma}),$$

where MVN implies multivariate normal distribution and $\boldsymbol{\Sigma} = \text{diag}(V_1^2, V_2^2, ..., V_{10}^2)$. Assuming $H = \delta^2(I-\rho C)^{-1}$, the proposal density will be

$$\begin{aligned}
\boldsymbol{\nu}|\theta, \delta^2, \rho, \mathbf{Y} &\propto \exp\left\{-\frac{1}{2}\left[(\boldsymbol{\nu}-(\mathbf{Z}-\theta))^T\Sigma^{-1}(\boldsymbol{\nu}-(\mathbf{Z}-\theta))\right]\right\} \times \exp\left\{-\frac{1}{2}[\boldsymbol{\nu}^T H^{-1}\boldsymbol{\nu}]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\boldsymbol{\nu}^T\Sigma^{-1}\boldsymbol{\nu} - 2\boldsymbol{\nu}^T\Sigma^{-1}(\mathbf{Z}-\theta) + (\mathbf{Z}-\theta)^T\Sigma^{-1}(\mathbf{Z}-\theta) + \boldsymbol{\nu}^T H^{-1}\boldsymbol{\nu}\right]\right\} \\
&= \exp\left\{-\frac{1}{2}\left[\boldsymbol{\nu}^T(\Sigma^{-1}+H^{-1})\boldsymbol{\nu} - 2\boldsymbol{\nu}^T(\Sigma^{-1}+H^{-1})(\Sigma^{-1}+H^{-1})^{-1}\Sigma^{-1}(\mathbf{Z}-\theta)\right]\right\},
\end{aligned}$$

which implies that the proposal density of $\boldsymbol{\nu}|\theta, \delta^2, \rho, \mathbf{Y}$ is

$$\boldsymbol{\nu}|\theta, \delta^2, \rho, \mathbf{Y} \sim MVN\left\{(\Sigma^{-1}+H^{-1})^{-1}\Sigma^{-1}(\mathbf{Z}-\theta), (\Sigma^{-1}+H^{-1})^{-1}\right\}. \tag{14}$$

We use the grid method to draw samples from $\theta|\boldsymbol{\nu}, \delta^2, \rho, \mathbf{Y}$ and $\rho|\delta^2, \theta, \boldsymbol{\nu}, \mathbf{Y}$. This is particularly straightforward in the first case since θ and ρ are bounded, that is

$$\hat{\theta} - 10\times V < \theta < \hat{\theta} + 10\times V, \text{ and } \tfrac{1}{\lambda_{min}} < \rho < \tfrac{1}{\lambda_{max}}.$$

For the conditional posterior density $\rho|\delta^2, \theta, \boldsymbol{\nu}, \mathbf{Y}$, we apply the following transformation on δ^2, since δ^2, being positive, does not have an upper bound.

$$\phi = \frac{\delta^2}{1+\delta^2},\ 0 < \phi < 1,$$

which results in the transformed density

$$\phi|\theta, \rho, \boldsymbol{\nu}, \mathbf{Y} \propto \sqrt{\frac{1-\phi}{\phi}}\exp\left\{-\frac{1}{2}\boldsymbol{\nu}^T\left(\frac{\phi}{1-\phi}(I-\rho C)^{-1}\right)^{-1}\boldsymbol{\nu}\right\}.$$

Once we have simulated the values of ϕ, we can back-transform to obtain the corresponding values of δ^2 since $\delta^2 = \dfrac{\phi}{1-\phi}$.

Given the above discussion, it will now be straightforward to simulate observations from the respective full conditionals. In doing so, we randomly select 100 sets of propensity scores among the 1000 and the Gibbs sampler is run for each such set as follows:

1. Initial values for the parameters are selected as: $\rho^{(0)} = 0, \delta^{2(0)} = 1, \theta^{(0)} = 0$.

2. Given the intital values, a sample is drawn from $\boldsymbol{\nu}|\theta^{(0)}, \delta^{2(0)}, \rho^{(0)}, \mathbf{Y}$ using the Metropolis-Hastings sampler through the candidate density derived in (15). Let the sampled value be $\boldsymbol{\nu}^{(1)}$.

3. Given $\boldsymbol{\nu}^{(1)}$, a sample is drawn from $\theta|\boldsymbol{\nu}^{(1)}, \delta^{2(0)}, \rho^{(0)}, \mathbf{Y}$ through the Grid method. Let the sampled value be denoted as $\theta^{(1)}$.

4. Given $\boldsymbol{\nu}^{(1)}$ and $\theta^{(1)}$ obtained above, we sample from $\rho|\boldsymbol{\nu}^{(1)}, \theta^{(1)}, \delta^{2(0)}, \mathbf{Y}$ again using the Grid method. Let the sampled value be denoted as $\rho^{(1)}$.

5. Given the sampled values of $\boldsymbol{\nu}^{(1)}$, $\theta^{(1)}$ and $\rho^{(1)}$, we draw a sample from $\phi|\theta^{(1)}, \rho^{(1)}, \boldsymbol{\nu}^{(1)}, \mathbf{Y}$ by applying Grid method again and perform the transformation $\delta^2 = \dfrac{\phi}{1-\phi}$ to get the corresponding value of δ^2.

6. For implementing the grid samplers in steps (3 - 5), we use the upper and lower bounds of the respective parameters and come up with the grid points.

7. At the completion of the above iteration, we obtain the first set of simulated values of the parameters vis $(\boldsymbol{\nu}^{(1)}, \theta^{(1)}, \rho^{(1)}, \delta^{2(1)})$.

We repeat the above steps 2 to 5 step for 11,000 times and do a burn-in of the first 2000 iterates. Then we do some thinning and keep the following iterates,

$$(\boldsymbol{\nu}^{(2001+9m)}, \theta^{(2001+9m)}, \rho^{(2001+9m)}, \delta^{2(2001+9m)}),$$

where $m = 1, 2, ..., 1000$. In doing so, we are finally left with 1,000 sets of iterates.

As usual, we verify the convergence of the chains using trace and autocorrelation plots along with Gweke test and effective sample sizes. The associated plots and tables are shown under Section 4.2 in the Annexure. The plots are indicative of satisfactory convergence of the chains. Posterior summaries are shown in Table 5.

Some notable observations can be made from the above table. For instance, random effects corresponding to the second, fourth and tenth strata are significant which implies that observations/subjects within these sub-classes have significant dependence. In addition the fixed effect component, θ is also significantly negative. More importantly, the SCAR model of He and Sun (2000) cannot capture the spatial correlation as the 95% credible interval of ρ is (-0.342, 0.488) and hence contains zero. We will address this issue using an improved modeling framework discussed in the next section.

Table 5: Posterior summaries of parameters for first spatial model

Parameter	Mean	Standard deviation	95% Credible interval
ν_1	0.817	0.708	(-0.559, 2.165)
ν_2	0.886	0.458	(0.059, 1.846)
ν_3	0.703	0.396	(-0.061, 1.453)
ν_4	1.204	0.365	(0.420, 1.886)
ν_5	0.539	0.297	(-0.093, 1.131)
ν_6	0.438	0.260	(-0.046, 0.968)
ν_7	0.334	0.195	(-0.051, 0.697)
ν_8	0.007	0.174	(-0.371, 0.313)
ν_9	-0.223	0.141	(-0.505, 0.051)
ν_{10}	-0.431	0.119	(-0.665, -0.192)
θ	-0.413	0.016	(-0.440, -0.378)
ρ	0.0744	0.231	(-0.342, 0.488)
δ^2	3.925	3.554	(0.243, 11.423)

4.3. Prediction

Given the sampled values of the parameters obtained above, it is straightforward to predict the responses for the non-sampled units. For each $\boldsymbol{\beta}$, the number of non-sampled individuals for each stratum is known (see Table 4). Moreover, for the i^{th} stratum,

$$y_{ij}|p_i \sim \text{Bernoulli}(p_i), \;\; i = 1, 2, ..., 10; j = 1, 2, ..., n_i.$$

Hence, we can get a sample of the responses corresponding to the non-sampled group for the i^{th} stratum by drawing the requisite number of y_{ij}'s from Bernoulli(p_i). Based on the sampled values, we can evaluate the proportion of positive responses. This exercise should be repeated for other sets of propensity scores as well. Accordingly, we randomly selected 100 sets of propensity scores and obtained 100 proportion values (of positive responses in the non-sampled group). Based on those values, we form the highest posterior density (HPD) intervals of the true proportion of positive responses as was done for the Beta-Bernoulli model. The resulting interval is (0.449, 0.551) which is clearly narrower than those corresponding to the Beta-Bernoulli, Hajek and Horvitz-Thompson estimators. In addition, the true value of the proportion for all the subjects and for the non-sampled subjects lie near the centre of the interval corresponding to the spatial model. Both of these implies that the predictive ability of the spatial model is superior to the other models *i.e.*. the predicted values of the response in the non-sampled group and the corresponding proportions obtained from the spatial model is more accurate compared to those obtained from the other models, namely Beta-Bernoulli, Horvitz-Thompson and Hajek. Histograms and density plots of the proportions are shown under Section 5.2 in the Annexure.

5. Modified spatial model

In this section, we show how to improve the standard spatial model.

5.1. Model specification

The spatial regression model outlined in Section 4 is motivated by the work of He and Sun (2000). One shortcoming of their formulation is that it fails to account for positive and monotonically weakening spatial correlation. To account for that, we introduce a modified spatial model in this section for which we define the following 10×10 adjacency matrix:

$$\mathbf{A} = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 & \rho^7 & \rho^8 & \rho^9 \\ \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 & \rho^7 & \rho^8 \\ \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 & \rho^7 \\ \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 \\ \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^7 & \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^8 & \rho^7 & \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^9 & \rho^8 & \rho^7 & \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}, \quad 0 < \rho < 1.$$

The structure of the adjacency matrix distinguishes it from the spatial model discussed in Section 4. Specifically, the underlying assumption for the above structure is that subjects belonging to strata in close proximity have higher dependence than those belonging to strata which are further apart. The logistic mixed model specification for p_i remains the same as was done for the standard spatial model in Sec 4.1 *i.e.*.

$$\log\left(\frac{p_i}{1-p_i}\right) = \theta + \nu_i \quad i = 1, 2, ..., 10,$$

θ and ν_i having the same connotation as before. The conditional distribution of y_{ij} remains the same as for the standard spatial model *i.e.*.

$$y_{ij}|\theta, \nu_i \sim \text{Ber}\left(\frac{e^{\theta+\nu_i}}{1+e^{\theta+\nu_i}}\right), \quad i = 1, 2, ..., 10, j = 1, 2, ..., n_i.$$

The following priors are specified for the parameters $(\boldsymbol{\nu}, \theta, \delta^2, \rho)$

$$\begin{aligned} \boldsymbol{\nu}|\theta, \delta^2, \rho &\sim MVN(\theta\boldsymbol{j}, \delta^2\mathbf{A}), \\ \pi(\theta, \delta^2, \rho) &\propto \frac{1}{(1+\delta^2)^2}, \end{aligned}$$

where $\boldsymbol{j}$ is a 10×1 dimensional vector of 1's while $0 < \theta < 1$ and $0 < \rho < 1$. Combining the likelihood and priors, the joint posterior density of $(\boldsymbol{\nu}, \theta, \delta^2, \rho)$ is given by

$$\pi(\theta, \delta^2, \boldsymbol{\nu}, \rho|\mathbf{Y}) \propto f(\mathbf{Y}|\nu_i)\pi(\boldsymbol{\nu}|\theta, \delta^2, \rho)\pi(\theta, \delta^2, \rho),$$

where

$$\begin{aligned} f(\boldsymbol{Y}|\theta, \nu_i) &= \prod_{i=1}^{10}\prod_{j=1}^{n_i}\left\{\frac{e^{\theta+\nu_i}}{1+e^{\theta+\nu_i}}\right\}^{y_{ij}}\left\{1-\frac{e^{\theta+\nu_i}}{1+e^{\theta+\nu_i}}\right\}^{1-y_{ij}} \\ &= \prod_{i=1}^{10}\prod_{j=1}^{n_i}\left\{\frac{e^{(\theta+\nu_i)y_{ij}}}{1+e^{\theta+\nu_i}}\right\} \end{aligned}$$

and

$$\pi(\boldsymbol{\nu}|\theta,\delta^2,\rho) = \frac{1}{\sqrt{|\delta^2\mathbf{A}|}}\exp\left\{-\frac{1}{2}(\boldsymbol{\nu}-\theta\boldsymbol{j})^T(\delta^2\mathbf{A})^{-1}(\boldsymbol{\nu}-\theta\boldsymbol{j})\right\}.$$

Thus, the joint posterior density is

$$\begin{aligned}\pi(\theta,\delta^2,\boldsymbol{\nu},\rho|\mathbf{Y}) &\propto \prod_{i=1}^{10}\prod_{j=1}^{n_i}\left\{\frac{e^{(\theta+\nu_i)y_{ij}}}{1+e^{\theta+\nu_i}}\right\}\frac{1}{\sqrt{|\delta^2\mathbf{A}|}}\exp\left\{-\frac{1}{2}(\boldsymbol{\nu}-\theta\boldsymbol{j})^T(\delta^2\mathbf{A})^{-1}(\boldsymbol{\nu}-\theta\boldsymbol{j})\right\}\times\frac{1}{(1+\delta^2)^2}\\ &= \prod_{i=1}^{10}\left\{\frac{e^{(\theta+\nu_i)M_i}}{(1+e^{\theta+\nu_i})^{n_i}}\right\}\frac{1}{\sqrt{|\delta^2\mathbf{A}|}}\exp\left\{-\frac{1}{2}(\boldsymbol{\nu}-\theta\boldsymbol{j})^T(\delta^2\mathbf{A})^{-1}(\boldsymbol{\nu}-\theta\boldsymbol{j})\right\}\times\frac{1}{(1+\delta^2)^2},\end{aligned}$$

where $M_i=\sum_{j=1}^{n_i}y_{ij}$ is the total number of positive responses in the i^{th} subclass of the sampled group.

5.2. Bayesian computation

The following full conditional posterior densities can be derived from the full posterior shown above

$$\boldsymbol{\nu}|\theta,\delta,\rho,\mathbf{Y} \propto \prod_{i=1}^{10}\left\{\frac{e^{(\theta+\nu_i)M_i}}{(1+e^{\theta+\nu_i})^{n_i}}\right\}\frac{1}{\sqrt{|\delta^2\mathbf{A}|}}\exp\left\{-\frac{1}{2}(\boldsymbol{\nu}-\theta\boldsymbol{j})^T(\delta^2\mathbf{A})^{-1}(\boldsymbol{\nu}-\theta\boldsymbol{j})\right\}; \quad (15)$$

$$\theta|\boldsymbol{\nu},\delta,\rho,\mathbf{Y} \sim N\left(\frac{\boldsymbol{j}^T(\delta^2\mathbf{A})^{-1}\boldsymbol{\nu}}{\boldsymbol{j}^T(\delta^2\mathbf{A})^{-1}\boldsymbol{j}},\frac{1}{\boldsymbol{j}^T(\delta^2\mathbf{A})^{-1}\boldsymbol{j}}\right); \quad (16)$$

$$\delta^2|\theta,\rho,\boldsymbol{\nu},\mathbf{Y} \propto \frac{1}{(\delta^2)^5}\exp\left\{-\frac{1}{2}(\boldsymbol{\nu}-\theta\boldsymbol{j})^T(\delta^2\mathbf{A})^{-1}(\boldsymbol{\nu}-\theta\boldsymbol{j})\right\}\times\frac{1}{(1+\delta^2)^2}; \quad (17)$$

$$\rho|\delta^2,\theta,\boldsymbol{\nu},\mathbf{Y} \propto \frac{1}{\sqrt{|\mathbf{A}|}}\exp\left\{-\frac{1}{2}(\boldsymbol{\nu}-\theta\boldsymbol{j})^T(\delta^2\mathbf{A})^{-1}(\boldsymbol{\nu}-\theta\boldsymbol{j})\right\}. \quad (18)$$

Following a similar method that was detailed in Section 4.2, we obtain the following proposal density of $\boldsymbol{\nu}|\theta,\delta,\rho,\mathbf{Y}$

$$\boldsymbol{\nu}|\theta,\delta,\rho,\mathbf{Y} \sim MVN\left\{(\Sigma^{-1}+H^{-1})^{-1}(\Sigma^{-1}\mathbf{Z}+H^{-1}\theta\boldsymbol{j}),(\Sigma^{-1}+H^{-1})^{-1}\right\},$$

where $H=\delta^2\mathbf{A}$ and $\Sigma=\mathrm{diag}(V_1^2,V_2^2,...,V_{10}^2)$ and

$$Z_i=\log\frac{M_i+0.5}{n_i-M_i+0.5},\quad V_i^2=\frac{(n_i+1)(n_i+2)}{n_i(M_i+1)(n_i-M_i+1)},$$

The simulation steps will be similar to those mentioned in Section 4.2. As usual, convergence is verified using trace plots, autocorrelation plots, Geweke test and effective sample size procedures. All these tests indicate adequate convergence. Tables showing the p-values for the Geweke test and effective sample sizes are shown under Section 5.2 in the Annexure along with trace and kernel density plots of all the parameters. As shown in that table, the effective sample sizes of all but one parameter is 1000, the same length as the chain, thus indicating satisfactory convergence. Table 6 depicts the posterior summaries of all the parameters.

Table 6: Posterior summaries for modified spatial model

Parameter	Mean	Standard deviation	95% Credible interval
ν_1	0.429	0.442	(-0.491, 1.229)
ν_2	0.722	0.340	(0.061, 1.382)
ν_3	0.691	0.299	(0.125, 1.264)
ν_4	0.449	0.271	(-0.131, 0.955)
ν_5	0.171	0.240	(-0.316, 0.614)
ν_6	-0.029	0.212	(-0.426, 0.418)
ν_7	-0.210	0.170	(-0.547, 0.137)
ν_8	-0.377	0.151	(-0.684, -0.100)
ν_9	-0.417	0.141	(-0.671, -0.127)
ν_{10}	-0.914	0.113	(-1.116, -0.687)
θ	-0.024	0.463	(-0.963, 0.897)
ρ	0.664	0.203	(0.274, 0.972)
δ^2	0.405	0.254	(0.077, 0.864)

A comparison of Tables 5 and 6 results in some important observations. Firstly, in the modified spatial model, the random effects corresponding to five sub-classes are significant, namely those for second, third, eighth, ninth and tenth subclasses. For the first spatial model, this was true for only three subclasses. This indicates that the modified spatial model has better discriminatory ability in capturing intra-subclass-specific spatial dependence compared to the first spatial model. Secondly, the credible intervals for the modified spatial model are in general narrower than those corresponding to the previous spatial model. This implies that the modified spatial model generates more precise estimates of the parameters relative to the original spatial model. Moreover, the correlation parameter (ρ) is significant for the modified spatial model but was insignificant in the previous model. This is a major finding since it implies that the modified model is more capable of capturing the underlying spatial dependence between the sub-strata compared to the previous model. Thirdly, the estimate for the variance component δ^2 is much smaller for the modified model as compared to the previous model. This indicates that the modified model has superior ability to control for variance inflation of the strata specific random effects which indicate a better predictive ability of the responses for the non-sampled units.

5.3. Prediction

Since the chains have converged, we can use the parameter estimates to predict the responses corresponding to the non-sampled units as was done for the Beta-Bernoulli and standard spatial models. The 95% highest posterior density intervals corresponding to the modified spatial model along with those for the Beta-Bernoulli model, standard spatial model and those of Horvitz-Thompson and Hajek estimators are shown in Table 7. All the intervals relate to the prediction of the proportion of positive responses for all the subjects (sampled + non-sampled). It is clear from Table 7 that the modified spatial model has superior predictive ability compared to all the other models since it results in the narrowest HPD interval among the model-based intervals; the width under the Hajek model is much too small. Moreover, the true value of the proportion of positive responses for all the units *viz.* 0.4976, lies near the centre of the above interval as well. So, we conclude that the modified spatial model is

the optimal model for prediction. It is important to note here that the Hajek estimator is usually more precise in a design-based situation Särndal *et al.* (1992). However, it is difficult to evaluate the standard errors because it involves second-order inclusion probabilities. So, we have used an output analysis from the Metropolis sampler to get repeated values of the Hajek estimator like a bootstrap sample. These may not be the best estimates of the standard errors as they might be small.

Table 7: 95% credible intervals for the different models

Model	95% HPD interval	Width
Horvitz-Thompson	(0.478, 0.584)	0.106
Hajek	(0.497, 0.543)	0.046
Beta-Bernoulli	(0.455, 0.559)	0.104
Spatial	(0.433, 0.517)	0.084
Modified spatial	(0.456, 0.537)	0.081

6. Model comparison

Based on the discussion in the previous section, specifically with regard to the parameter estimates and credible intervals depicted in Tables 5 and 6, it is evident that the modified spatial model is more robust and has better predictive ability than the Beta-Bernoulli and standard spatial models. In this section, we will use two more diagnostic tools, namely conditional predictive ordinate (CPO) and log-pseudo marginal likelihood (LPML) to validate this fact.

The conditional predictive ordinate (CPO), introduced by Geisser (1980), is used to detect observations which are fitted poorly by a given parametric model. The CPO values can be calculated based on the output of the Markov chain Monte Carlo simulation procedure. The Monte Carlo approximation of CPO for the i^{th} stratum is given by

$$\hat{CPO}_i = \left[\frac{1}{M}\sum_{h=1}^{M}\frac{1}{f(y_i|p_i^{(h)})}\right]^{-1}, \quad i = 1, 2, ..., 10; \quad h = 1, 2, ..., 1000,$$

where $\hat{CPO}_i$ is the harmonic mean of $f(y_i|p_i^{(h)})$. For the Beta-Bernoulli model, $M = 1000$, $p_i^{(h)}$ is the h^{th} sample drawn from the posterior density of $p_i|\mathbf{y}_i$ while $\mathbf{y}_i \sim \text{Binomial}(n_i, p_i)$ for i^{th} stratum ($i = 1, 2, ..., 10,\ h = 1, 2, ..., 1000$). For the spatial model, $M = 100$ and p_i is obtained from the following expression

$$p_i = \frac{e^{\theta+\nu_i}}{1 + e^{\theta+\nu_i}},$$

where (θ, ν_i) are drawn from their respective posterior densities through the Monte Carlo simulation. Here also, $\mathbf{y}_i|p_i \sim \text{Binomial}(n_i, p_i)$ for the i^{th} stratum ($i = 1, 2, ..., 10$). For our proposed frameworks, each CPO value will correspond to a particular stratum and will indicate which, if any, stratum is an outlier in terms of model fit. Table 8 depicts the CPO values for each strata corresponding to the Beta-Bernoulli, basic spatial and modified spatial

Table 8: CPO values for the proposed models

Stratum	Beta-Bernoulli	Spatial I	Spatial II
1	0.061	0.119	0.206
2	0.029	0.041	0.071
3	0.046	0.021	0.074
4	0.041	0.077	0.082
5	0.037	0.037	0.063
6	0.039	0.047	0.047
7	0.028	0.039	0.041
8	0.023	0.035	0.034
9	0.021	0.027	0.027
10	0.0004	0.009	0.017

models. The spatial models are denoted as Spatial I (basic spatial) and Spatial II (modified spatial) respectively.

In terms of assessing model fit, observations with CPO values less than 0.025 are deemed as possible outliers while those with values less than .014 are regarded as extreme observations (Ntzoufras, 2011). From the CPO values depicted in Table 8, it can be concluded that for Beta-Bernoulli model, there are three outlying strata, namely strata 8, 9 and 10. Of this, stratum 10 seems to be an influential point since the CPO value is lesser than 0.014. For Spatial model I, there are two outlying strata (strata 3 and 10). Again, stratum 10 seems to be an influential point. Finally, for Spatial model II, only the last stratum is identified as an outlier but not an influential point. Hence, it is apparent that the modified spatial model (Spatial II) performs better than the other models as per this diagnostic measure since it has the lowest number of outlier strata and no influential strata.

In order to have a confirmatory assessment of model fit, we next calculate the log-pseudo marginal likelihood (LMPL), which is a function of CPO, given by

$$\text{LPML} = \sum_{i=1}^{N} \log(\hat{CPO}_i).$$

Larger values of LMPL indicate a better fit. The following table depicts the values of LMPL for all the three models. Since the modified spatial model has the highest value of LMPL, we conclude that it has superior predictive ability compared to the Beta-Bernoulli and the standard spatial models. This validates the findings derived in Section 5.

Model	LMPL
Beta-Bernoulli	-38.11
Spatial	-32.97
Modified spatial	-29.39

7. Discussion

The standard method of obtaining a representative sample from a target population is through a probability sampling scheme which involves the selection of population units

according to a certain specified probability distribution. The most common of these methods is simple random sampling in which each and every population unit is assigned the same probability of selection. Having said that, implementation of an ideal probability sampling scheme in a real life setting is prohibitively difficult due to restrictions on costs, manpower and time among other things. This has led to the popularization of alternate sampling schemes which are easier to implement on the field as well in the online space. Some examples are convenience sample, volunteer sample, online polls etc.

However, one major disadvantage of these schemes is that selection of units are heavily dependent on the choice and preference of the survey designer and is often guided by convenience rather than an underlying probabilistic framework. Hence, these kinds of samples are known as non-probability samples and the generating scheme, a non-probability sampling scheme. Consequently, the sample, so chosen, often comes with various biases which may lead to a unreliable estimate of the parameter of interest. Selection bias is one such bias which results in a sample that may lack representation of certain segments of the target population. This results in a sample that is not a proper representation of the target population.

Regardless of the above shortcomings, non-probability sampling schemes are becoming increasingly popular due to the ease with which they can be implemented, both on the ground and in the virtual space. However, it is equally critical not to sacrifice on the "representativeness" of the final sample and the unbiasedness of the final estimate as it reflects the true population parameter. Hence, it is utmost importance to come up with a general framework that would enable us to predict the responses of sample units which should have been sampled but were left out in a non-probability sampling scheme.

In this article, we have proposed three modeling frameworks that will enable us to predict the non-sampled individuals responses from information obtained from the sampled units. The underlying idea behind each of these frameworks was to first balance the covariate distributions of the sampled and non-sampled groups/units. This was implemented using the propensity scores for those units. The propensity scores quantified the probability that a particular unit is incorporated in a sample given the values of its covariates and were obtained using a Bayesian hierarchical model. Ten strata were constructed based on the quantiles of the propensity scores so obtained. Finally, prediction of the unknown responses of the non-sampled units were carried out using three models - a Beta-Bernoulli model and two spatial models which accounted for possible spatial autocorrelation between the strata. Between the two spatial models we proposed, one incorporated postive and gradually weakening correlation structure while the other did not. We tested our models on a simulated dataset. Estimation was carried out through Markov chain Monte Carlo simulation and Bayesian bootstrap.

A comparison of the predictive abilities of the aforementioned models unambiguously indicated the superiority of the spatial modeling framework over the non-spatial ones, namely the Beta-Bernoulli, Horvitz-Thompson and Hajek estimators. Moreover, the spatial model incorporating the gradually weakening spatial correlation structured performed considerably better than the one which did not incorporate this feature and had the best predictive ability of all the models. This points to the veracity of our assumption about the presence of long range but diminishing spatial autocorrelation between strata, which was an interesting

finding in its own right. Overall, we believe that our proposed methodology will contribute to ongoing research in this important field of research. Our proposed methodology was built on the superpopulation modeling framework. An interesting extension of our work would be the formulation of predictive approaches combining the superpopulation and quasi-randomisation frameworks.

Acknowledgements

It is an honour and privilege to be invited to submit this paper for the upcoming special issue in honour of Prof. C. R. Rao. The authors gratefully acknowledge the assistance received from Hanqi Cao and Zhiqing Xu on an earlier version of this paper. Balgobin Nandram was supported by a grant from the Simons Foundation (#353953, Balgobin Nandram). Last but not the least, we would like to thank the editor, Prof. Vinod Gupta and the reviewers for their careful reading of the manuscript and for their encouraging and insightful comments which led to significant improvement of the article.

References

Beaumont, J. F. (2020). Are probability surveys bound to disappear for the production of official statistics. *Survey Methodology*, **46**, 1–29.

Chen, Y., Li, P., and Wu, C. (2020). Doubly robust inference with nonprobability survey samples. *Journal of the American Statistical Association*, **115**, 2011–2021.

Choi, S., Nandram, B., and Kim, D. (2021). Bayesian predictive inference of small area proportions under selection bias. *Survey Methodology*, **47**, 91–123.

Clayton, D. and Kaldor, J. (1987). Empirical bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, **1**, 671–681.

Cochran, W. G. and Chambers, S. P. (1965). The planning of observational studies of human populations. *Journal of the Royal Statistical Society. Series A*, **128**, 234–266.

Cox, D. R. (2018). *Analysis of Binary Data*. Routledge.

Elliot, M. R. (2009). Combining data from probability and non-probability samples using pseudo-weights. *Survey Practice*, **2**, 813–845.

Elliott, M. N. and Haviland, A. (2007). Use of a web-based convenience sample to supplement a probability sample. *Survey Methodology*, **33**, 211–215.

Elliott, M. R. and Valliant, R. (2017). Inference for nonprobability samples. *Statistical Science*, **32**, 249–264.

Gart, J. J. and Zweifel, J. R. (1967). On the bias of various estimators of the logit and its variance with application to quantal bioassay. *Biometrika*, **1**, 181–187.

Geisser, S. (1980). Discussion on sampling and Bayes' inference in scientific modeling and robustness (by gep box). *Journal of the Royal Statistical Society A*, **143**, 416–417.

Hastings, W. K. (1970). *Monte Carlo Sampling Methods using Markov Chains and their Applications*. Oxford University Press.

He, Z. and Sun, D. (2000). Hierarchical Bayes estimation of hunting success rates with spatial correlations. *Biometrics*, **56**, 360–367.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, **47**, 663–685.

Kohut, A., Keeter, S., Doherty, C., Dimock, M., and Christian, L. (2012). Assessing the representativeness of public opinion surveys. Washington, DC: Pew Research Center.

Little, R. J. (1982). Models for nonresponse in sample surveys. *Journal of the American statistical Association*, **77**, 237–250.

Little, R. J. (1993). Post-stratification: a modeler's perspective. *Journal of the American Statistical Association*, **88**, 1001–1012.

Little, R. J. and Rubin, D. B. (2002). Bayes and multiple imputation. *Statistical Analysis with Missing Data*, , 200–220.

Marella, D. (2023). Adjusting for selection bias in nonprobability samples by empirical likelihood approach. *Journal of Official Statistics*, **39**, 151–172.

McCullagh, P. (2019). *Generalized Linear Models*. Routledge.

Meng, X. L. (2018). Statistical paradises and paradoxes in big data (i) law of large populations, big data paradox, and the 2016 us presidential election. *The Annals of Applied Statistics*, **12**, 685–726.

Nandram, B. (2022). A Bayesian assessment of non-ignorable selection of a non-probability sample. *Indian Bayesians News Letter*, **14**, 7–20.

Nandram, B. and Choi, J. W. (2005). Hierarchical Bayesian nonignorable nonresponse regression models for small areas: An application to the nhanes data. *Survey Methodology*, **31**, 73–84.

Nandram, B. and Choi, J. W. (2010). A Bayesian analysis of body mass index data from small domains under nonignorable nonresponse and selection. *Journal of the American Statistical Association*, **105**, 120–135.

Nandram, B., Choi, J. W., and Liu, Y. (2021). Integration of nonprobability and probability samples via survey weights. *Journal of Statistics and Probability*, **10**, 1–5.

Nandram, B. and Rao, J. (2021). A Bayesian approach for integrating a small probability sample with a non-probability sample. In *Proceedings of the American Statistical Association (JSM2021-Virtual Conference)*, pages 1568–1603.

Nandram, B. and Rao, J. (2023). Bayesian predictive inference when integrating a non-probability sample and a probability sample. *arXiv preprint arXiv:2305.08997*, .

Nandram, B. and Rao, J. N. K. (2024). Bayesian integration for small areas by supplementing a probability sample with a non-probability sample. *Statistics and Applications*, **22**, 345–376.

Neyman, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, **97**, 558–625.

Ntzoufras, I. (2011). *Bayesian Modeling using WinBUGS*, volume 698. John Wiley & Sons.

Rafei, A., Flannagan, C. A., West, B. T., and Elliott, M. R. (2022). Robust Bayesian inference for big data: Combining sensor-based records with traditional survey data. *The Annals of Applied Statistics*, **16**, 1038–1070.

Rao, J. (2021). On making valid inferences by integrating data from surveys and other sources. *Sankhya B*, **83**, 242–272.

Robbins, M. W., Ghosh-Dastidar, B., and Ramchand, R. (2021). Blending probability and nonprobability samples with applications to a survey of military caregivers. *Journal of Survey Statistics and Methodology*, **9**, 1114–1145.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, **70**, 41–55.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, **63**, 581–592.

Rubin, D. B. (1979). Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association*, **74**, 318–328.

Sakshaug, J. W., Wiśniowski, A., Ruiz, D. A. P., and Blom, A. G. (2019). Supplementing small probability samples with nonprobability samples: A Bayesian approach. *Journal of Official Statistics*, **35**, 653–681.

Salvatore, C., Biffignandi, S., Sakshaug, J. W., Wiśniowski, A., and Struminskaya, B. (2024). Bayesian integration of probability and nonprobability samples for logistic regression. *Journal of Survey Statistics and Methodology*, **12**, 458–492.

Smith, T. (1983). On the validity of inferences from non-random samples. *Journal of the Royal Statistical Society: Series A (General)*, **146**, 394–403.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, **25**, 1.

Särndal, C. E., Swensson, B., Wretman, J., and et al. (1992). *Model Assisted Survey Sampling*. NewYork: Springer-Verlag.

Wang, L., Valliant, R., and Li, Y. (2021). Adjusted logistic propensity weighting methods for population inference using nonprobability volunteer-based epidemiologic cohorts. *Statistics in Medicine*, **40**, 5237–5250.

Wang, W., Rothschild, D., Goel, S., and Gelman, A. (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting*, **31**, 980–991.

Wiśniowski, A., Sakshaug, J. W., Perez Ruiz, D. A., and Blom, A. G. (2020). Integrating probability and nonprobability samples for survey inference. *Journal of Survey Statistics and Methodology*, **8**, 120–147.

Xu, Z. and Nandram, B. (2019). Bayesian inference of non-probability samples. In *Proceedings of the American Statistical Association, Bayesian Statistics Section)*, pages 2585–2593.

Xu, Z., Nandram, B., and Manandhar, B. (2020). Bayesian inference of a finite population mean under length-biased sampling. *Statistical Methods and Applications in Forestry and Environmental Sciences*, 79–103.

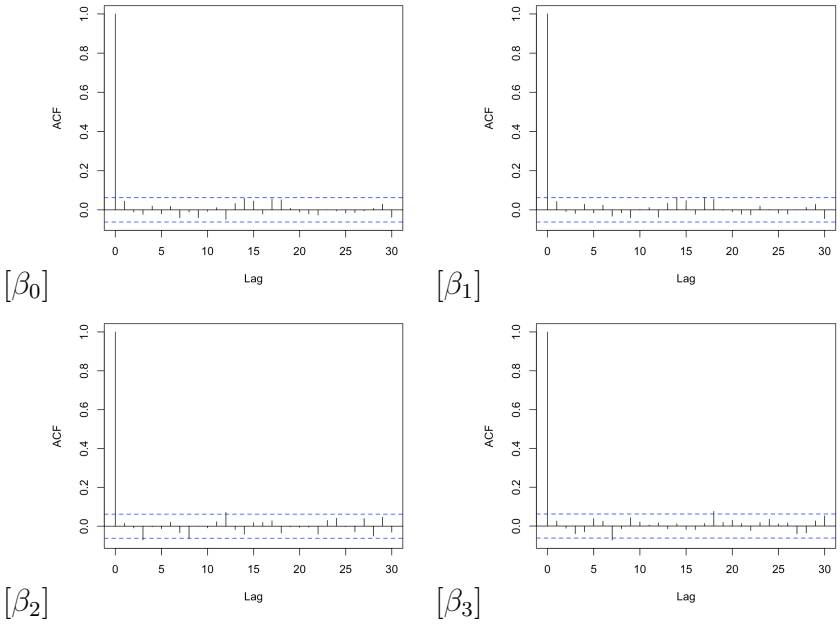
ANNEXURE BY SECTIONS

Section 3.2

The following table depicts part of the 1000 simulated values of $\boldsymbol{\beta}$ obtained using the Metropolis-Hastings sampler on $\pi(\boldsymbol{\beta}|\boldsymbol{R})$.

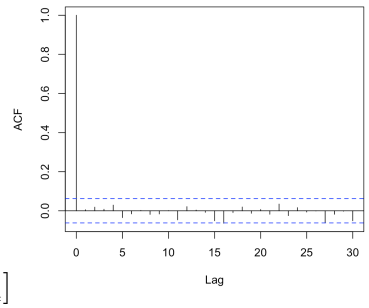
m	β_0	β_1	β_2	β_3	β_4
1	9.7473	-0.2002	-0.8234	0.2845	-0.4375
2	10.3579	-0.2096	-0.8713	0.2202	-0.4657
3	9.5018	-0.1960	-0.6962	0.1437	-0.5264
⋮	⋮	⋮	⋮	⋮	
998	9.1859	-0.1921	-0.6716	-0.2068	-0.3391
999	10.1561	-0.2094	-0.6304	0.3223	-0.4608
1000	10.3528	-0.2090	-0.9777	0.1989	-0.5132

The following figures depict the autocorrelation plots, trace plots and the kernel density plots for the simulated values of $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)$ obtained from the Markov Chain Monte Carlo run of the Bayesian Bootstrap model.



$[\beta_0]$ $[\beta_1]$

$[\beta_2]$ $[\beta_3]$



$[\beta_4]$

Section 3.3.1

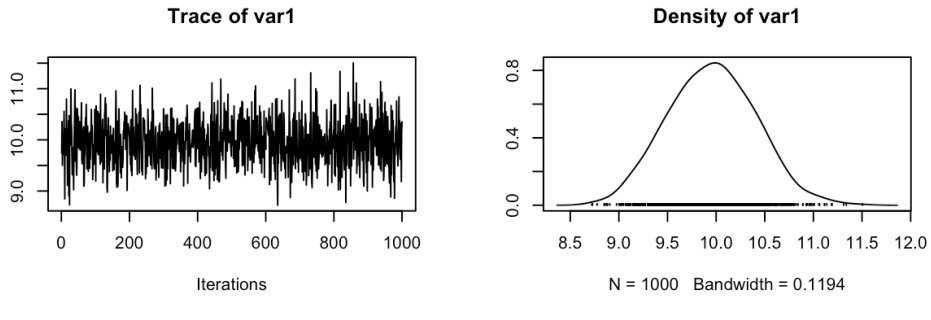
Following is part of the propensity score matrix where the rows correspond to the subjects ($N = 10000$) and columns correspond to 1000 simulated values of $\boldsymbol{\beta}$.

i	$\boldsymbol{\beta}^{(1)}$	$\boldsymbol{\beta}^{(2)}$	$\boldsymbol{\beta}^{(3)}$	⋯	$\boldsymbol{\beta}^{(1000)}$
1	0.5331	0.5614	0.5552		0.5836
2	0.0158	0.0146	0.0178		0.0134
3	0.4336	0.4593	0.4518		0.4790
⋮	⋮	⋮	⋮	⋮	⋮
9998	0.0199	0.0188	0.0246		0.0181
9999	0.0135	0.0125	0.0168		0.0119
10000	0.0526	0.0519	0.0632		0.0497

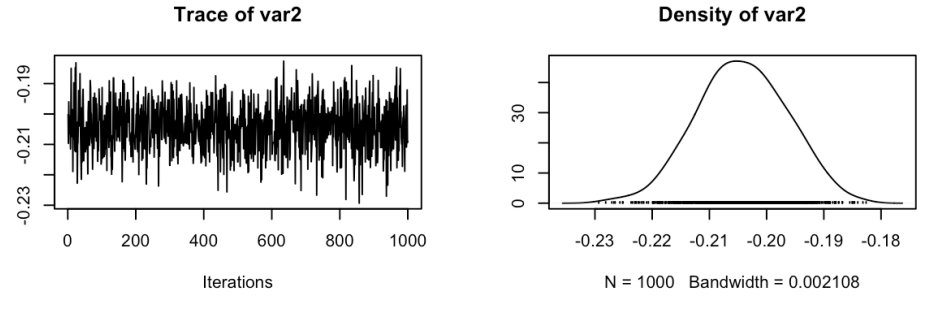
Section 4.2

The following table shows the p-values corresponding to the Gweke test and the effective sample sizes for $(\boldsymbol{\nu}, \theta, \rho, \delta^2)$ of the spatial model. All of the effective sample sizes are close to the size of chain $1,000$, which is desirable.

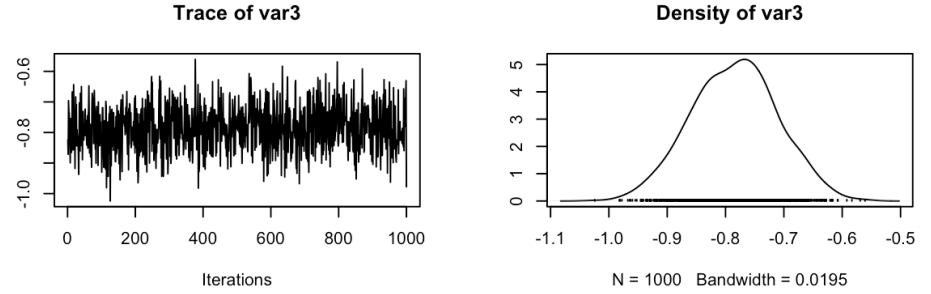
Parameter	P-value	Effective sample size
ν_1	0.098	1000
ν_2	0.186	1000
ν_3	0.459	874
ν_4	0.357	1000
ν_5	0.881	1000
ν_6	0.752	1000
ν_7	0.978	1000
ν_8	0.049	899
ν_9	0.285	1000
ν_{10}	0.768	1000
θ	0.667	1000
ρ	0.796	890
δ^2	0.721	926



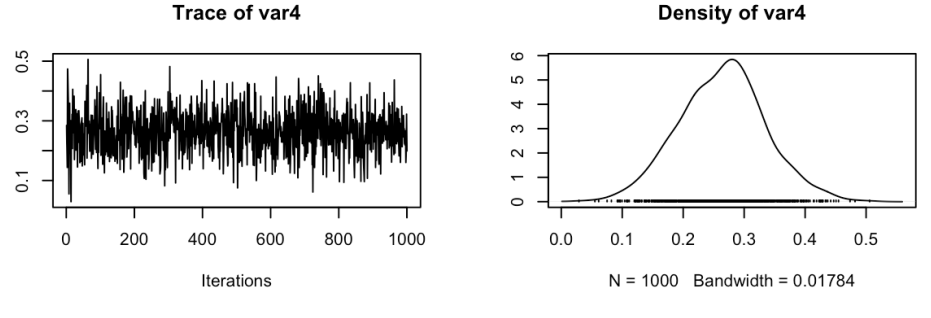
$[\beta_0]$



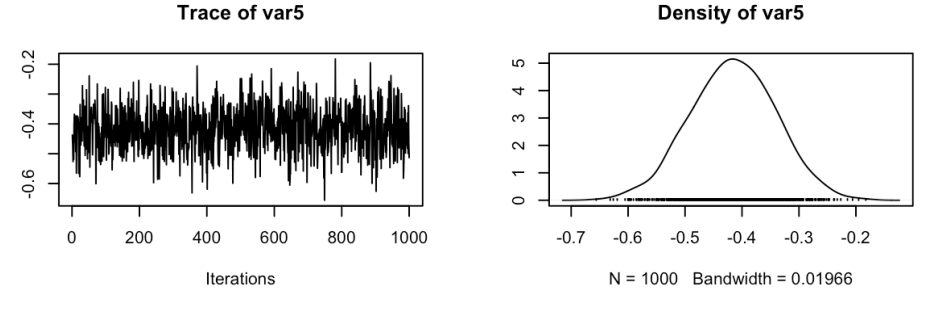
$[\beta_1]$



$[\beta_2]$

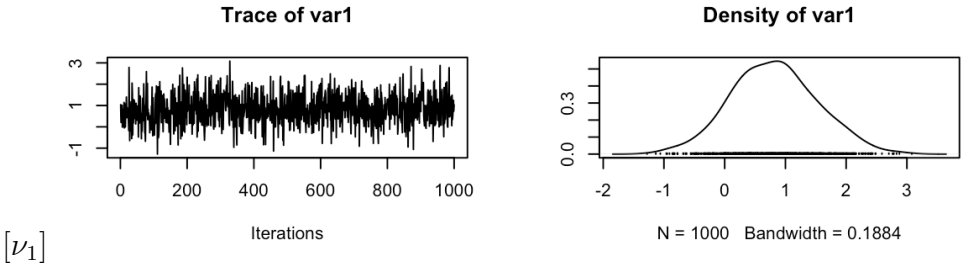


$[\beta_3]$

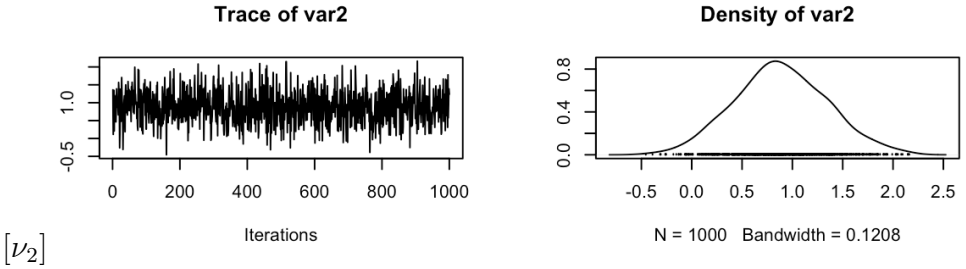


$[\beta_4]$

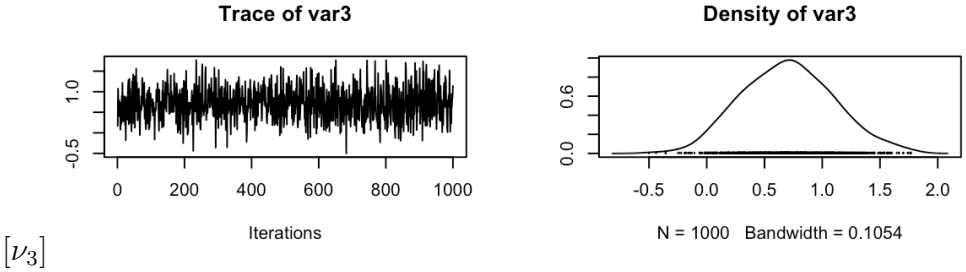
The following figures depict the trace plots, autocorrelation plots and the kernel density plots for the simulated values of $(\boldsymbol{\nu}, \theta, \rho, \delta^2)$ obtained from the Markov Chain Monte Carlo run of the Spatial model.



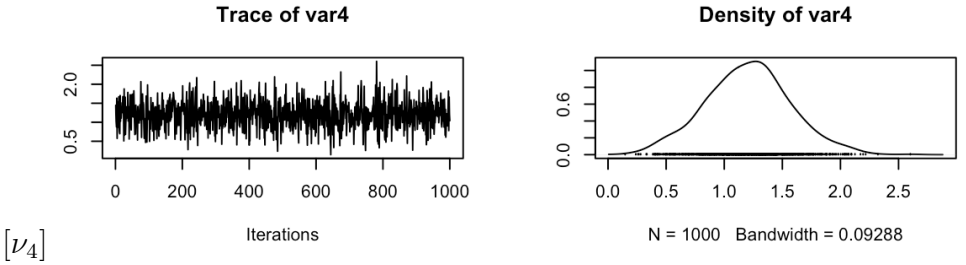
$[\nu_1]$



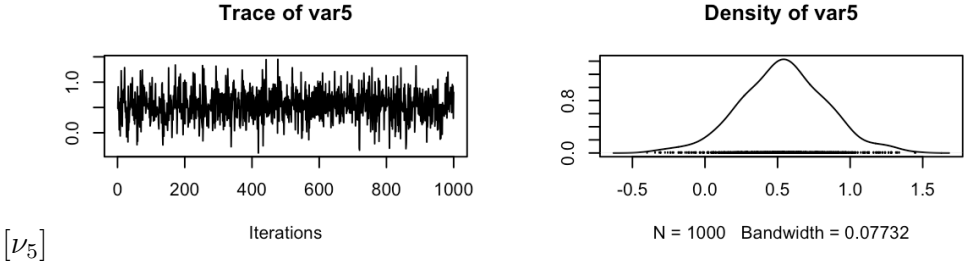
$[\nu_2]$



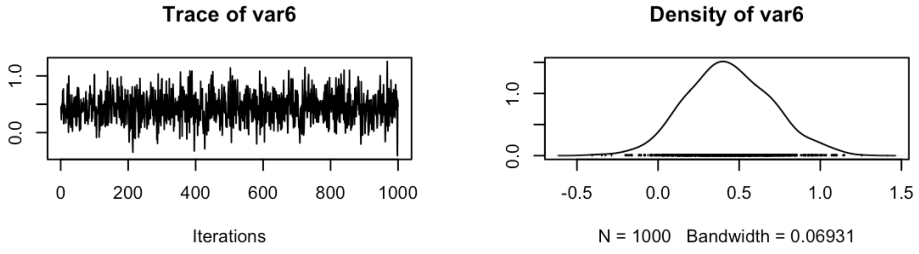
$[\nu_3]$



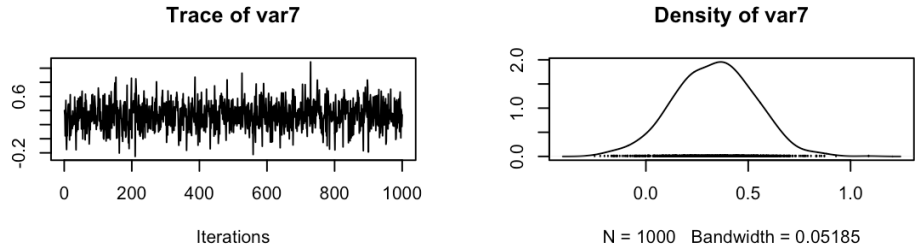
$[\nu_4]$



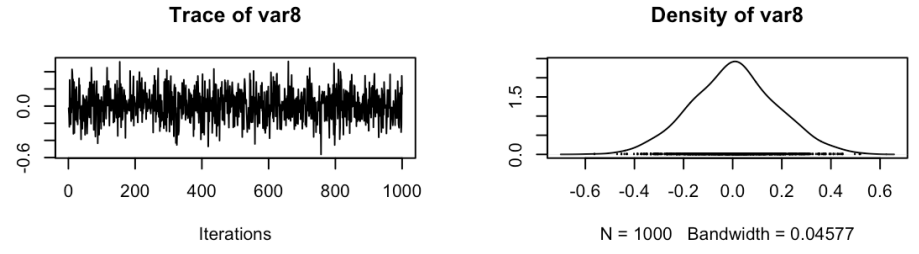
$[\nu_5]$



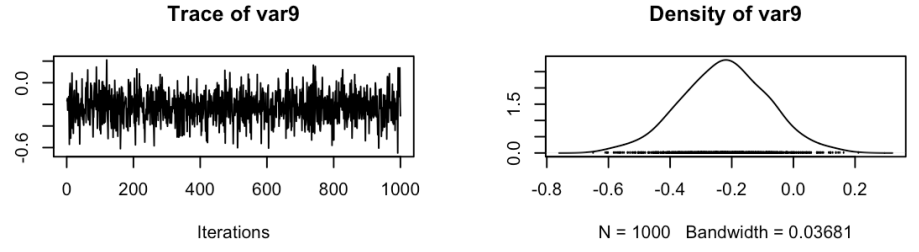
$[\nu_6]$



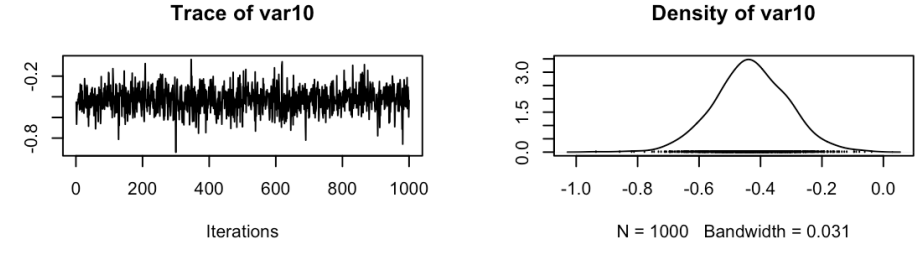
$[\nu_7]$



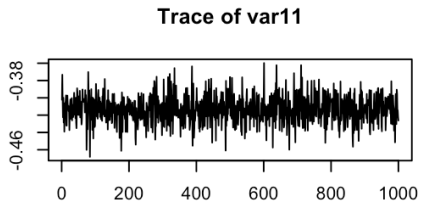
$[\nu_8]$

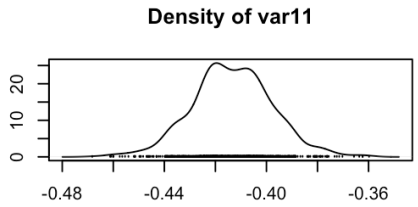


$[\nu_9]$

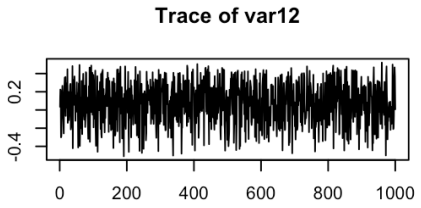


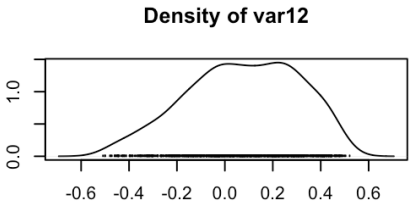
$[\nu_{10}]$



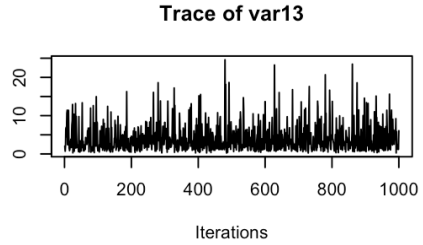


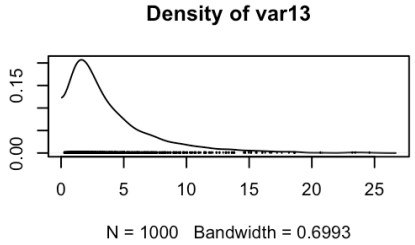
$[\nu_{11}]$



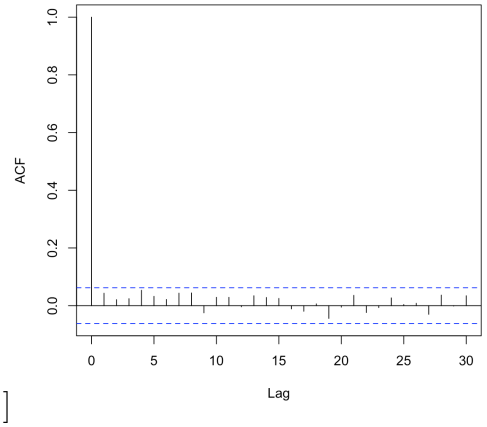


$[\nu_{12}]$

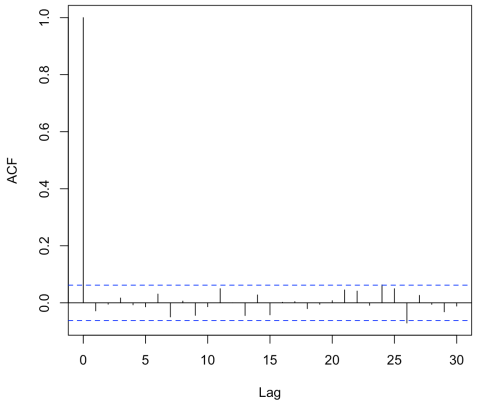




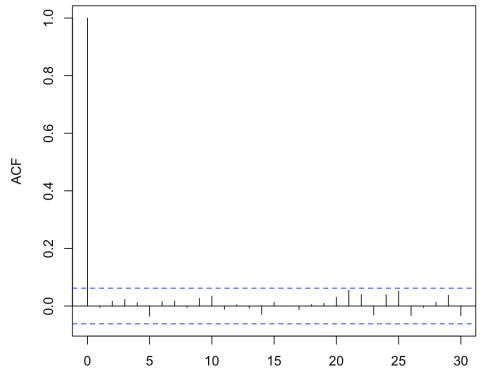
$[\nu_{13}]$



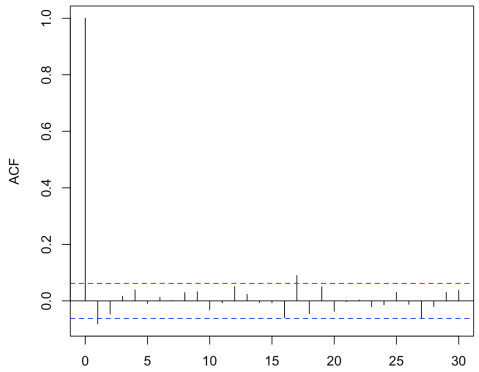
$[\nu_1]$



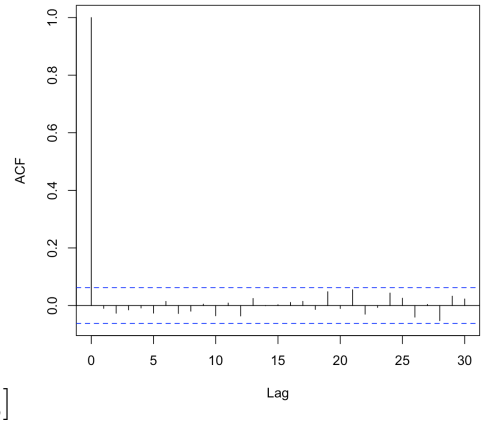
$[\nu_2]$



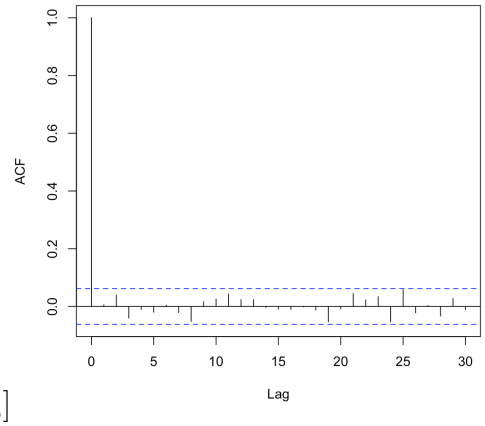
$[\nu_3]$



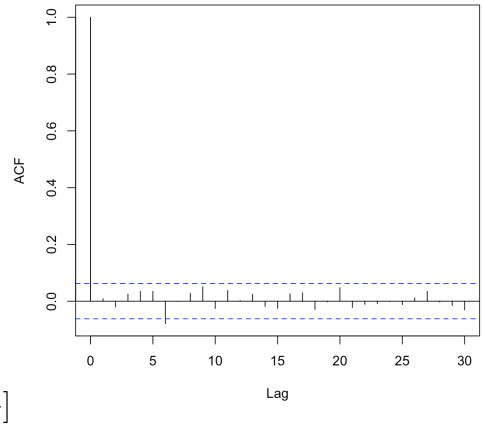
$[\nu_4]$



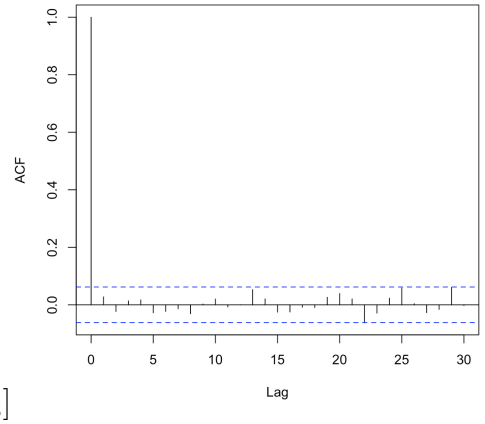
$[\nu_5]$



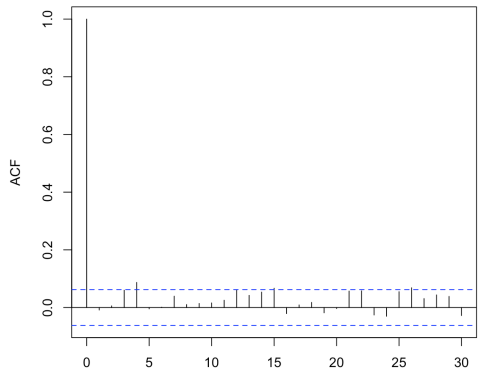
$[\nu_6]$



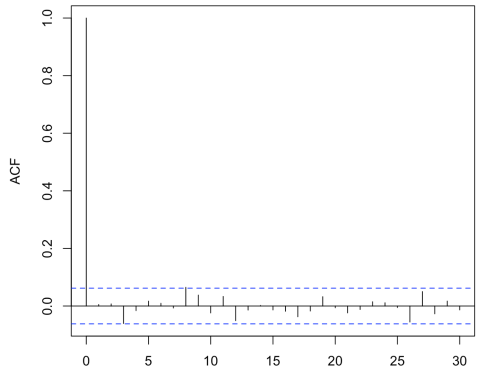
$[\nu_7]$



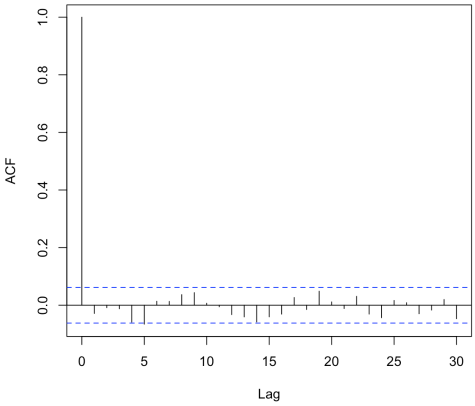
$[\nu_8]$



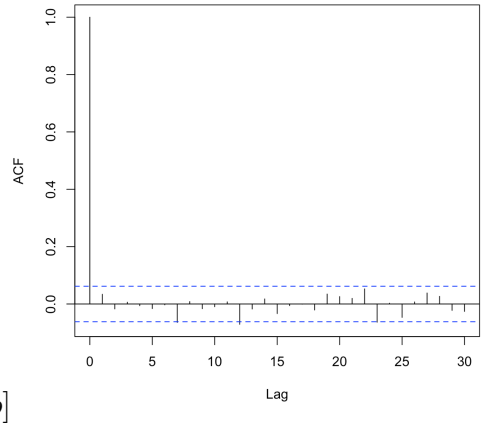
$[\nu_9]$



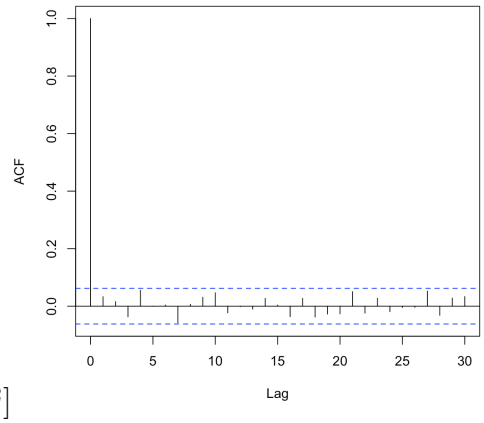
$[\nu_{10}]$



$[\theta]$



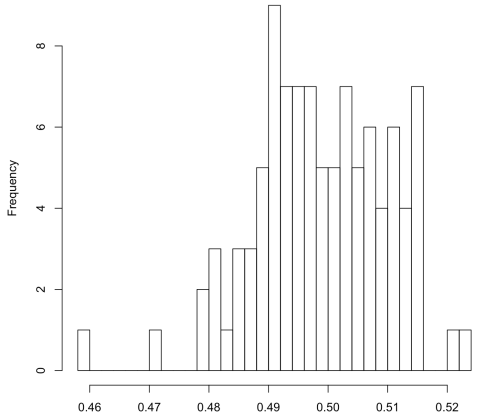
$[\rho]$



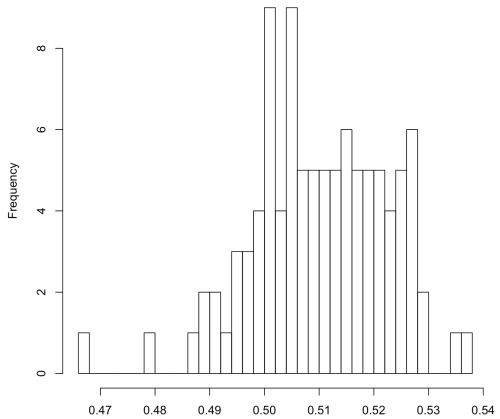
$[\delta^2]$

Section 4.3

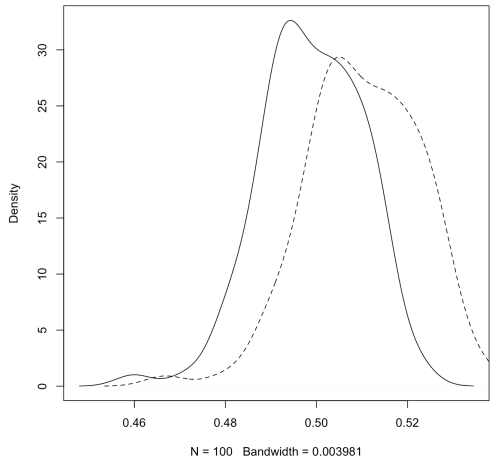
The following figures depict the histogram and kernel density plots of the proportions of positive responses predicted for i) all individuals $P_{all}^{(h)}$ and ii) non-sampled individuals ($P_{ns}^{(h)}$) based on the Spatial model. In the kernel density plot, the bold (dashed) curve corresponds to $P_{all}^{(h)}(P_{ns}^{(h)})$.



(a) All individuals



(b) Non-sampled individuals



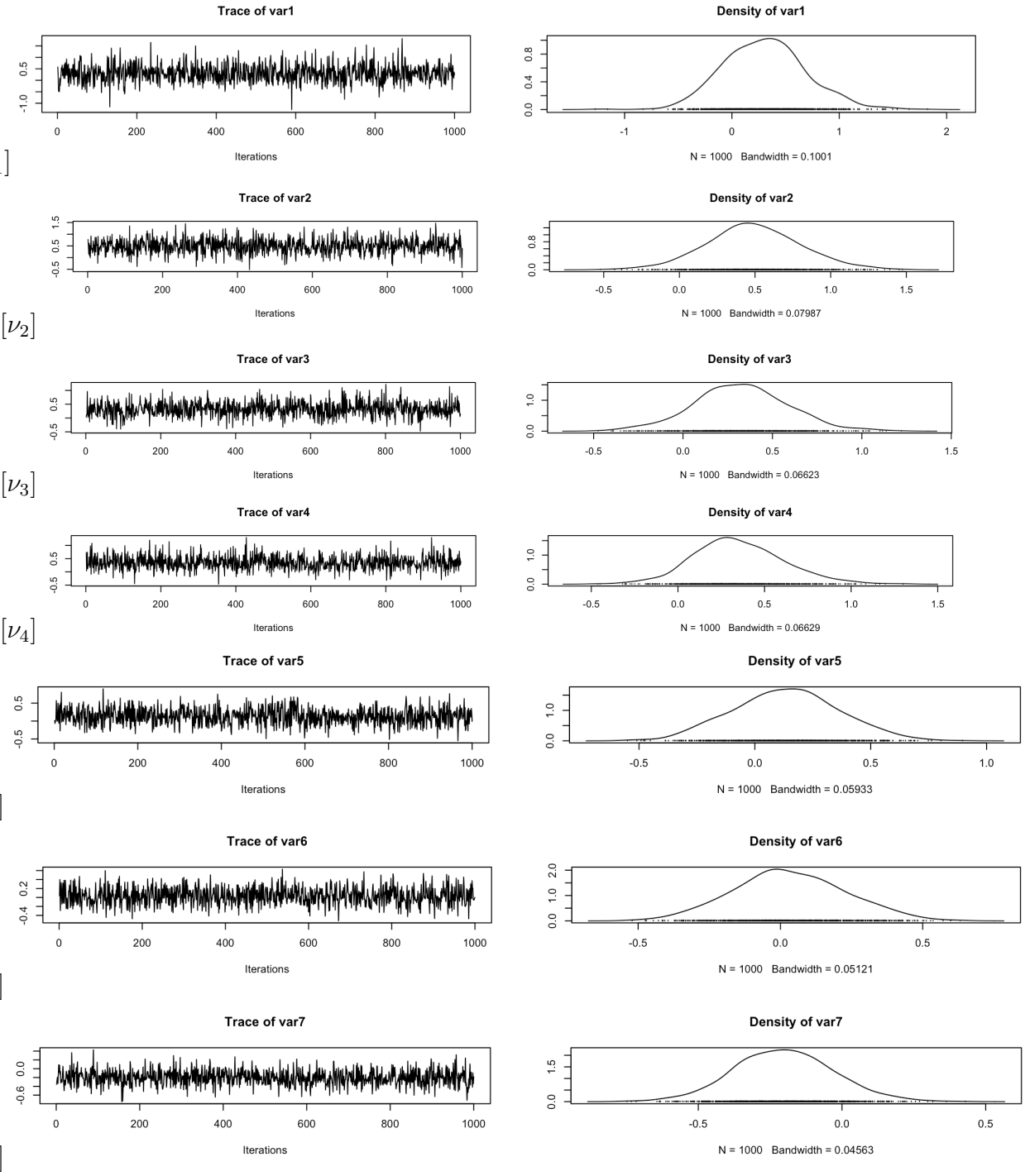
(c) Combined (bold: all subjects; dashed: non-sampled subjects)

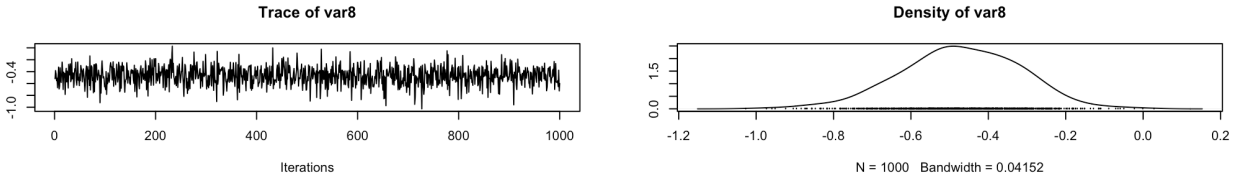
Section 5.2

The following table shows the p-values corresponding to the Gweke test and the effective sample sizes for $(\boldsymbol{\nu}, \theta, \rho, \delta^2)$ corresponding to the modified spatial model. All but one of the effective sample sizes are equal to the size of chain *i.e.*. $1,000$, which is desirable.

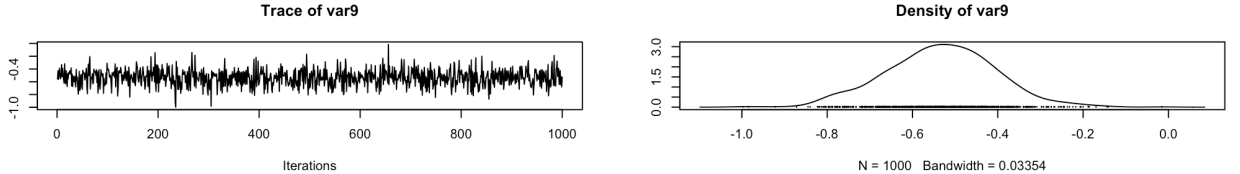
Parameter	P-value	Effective sample size
ν_1	0.10080200	1000
ν_2	0.95000090	1000
ν_3	0.45674993	1000
ν_4	0.28452094	1000
ν_5	0.91671578	1000
ν_6	0.22139337	1000
ν_7	0.47038949	1000
ν_8	0.06734535	1000
ν_9	0.11521862	1000
ν_{10}	0.34214527	1000
θ	0.38160805	1000
ρ	0.74996683	905
δ^2	0.94700833	1000

The following figures depict the trace and kernel density plots for the simulated values of $(\boldsymbol{\nu}, \theta, \rho, \delta^2)$ obtained from the Markov Chain Monte Carlo run of the modified Spatial model.

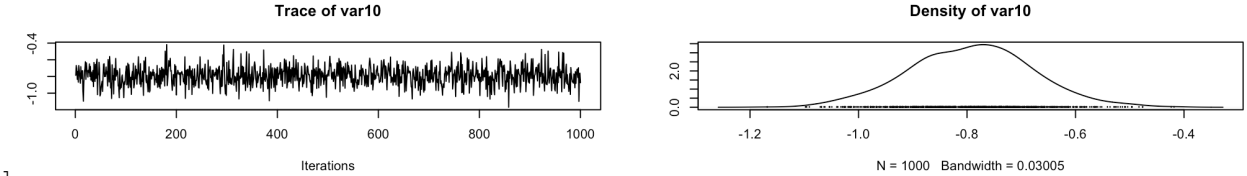

Trace of var1
Density of var1
Iterations
N = 1000 Bandwidth = 0.1001
$[\nu_1]$
Trace of var2
Density of var2
Iterations
N = 1000 Bandwidth = 0.07987
$[\nu_2]$
Trace of var3
Density of var3
Iterations
N = 1000 Bandwidth = 0.06623
$[\nu_3]$
Trace of var4
Density of var4
Iterations
N = 1000 Bandwidth = 0.06629
$[\nu_4]$
Trace of var5
Density of var5
Iterations
N = 1000 Bandwidth = 0.05933
$[\nu_5]$
Trace of var6
Density of var6
Iterations
N = 1000 Bandwidth = 0.05121
$[\nu_6]$
Trace of var7
Density of var7
Iterations
N = 1000 Bandwidth = 0.04563
$[\nu_7]$



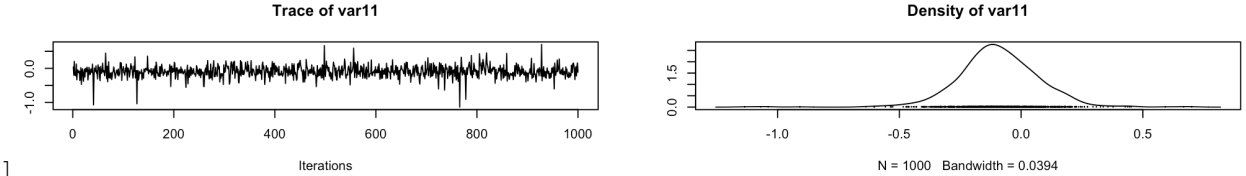
$[\nu_8]$



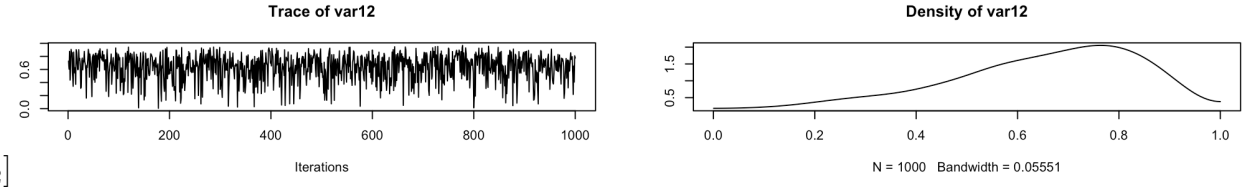
$[\nu_9]$



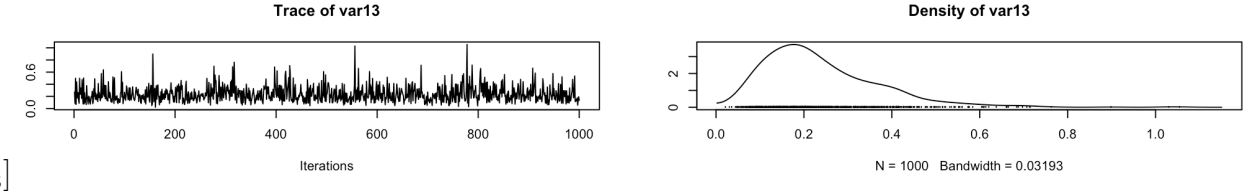
$[\nu_{10}]$



$[\nu_{11}]$



$[\nu_{12}]$



$[\nu_{13}]$