

Semi-Modular Inference on Covid Dataset

Sitong Liu

June 3, 2022

Abstract

Model misspecification is a common issue in Bayesian statistical inference. Generalised Bayes methods, including Cut model and Semi-Modular Inference (SMI) have been designed to tackle with model misspecification problem in multi-modular setting. There are several classical examples that are commonly used to illustrate efficacy of model misspecification methods. In this paper, we propose a new dataset and analyze it under SMI scheme. The influence of misspecified module on the analysis is controlled by the influence parameter η . Expected log pointwise predictive density (elpd) is a criterion to select η , but it loses power when the number of data sample is small. In this paper, we use an alternative approach, based on prior-to-posterior Kullback-Leibler divergence to choose η .

1 Introduction

Bayesian analysis combines different sources of information or “modules” into a single analysis via Bayes theorem. Large-scale multi-modular analyses under conventional Bayesian inference are susceptible to problems of model misspecification, as any bad module may distort inference in the model as a whole (Liu et al., 2009).

Modular Bayesian Inference is one approach to deal with model misspecification in a multi-modular setting by controlling feedback from misspecified modules (Liu et al., 2009; Plummer, 2015; Jacob et al., 2017; Carmona and Nicholls, 2020; Nicholls et al., 2022). Modular inference is characterised by *Cut-model* inference (Spiegelhalter et al., 2014; Plummer, 2015). This completely eliminates the contribution from some modules to the posterior distribution of parameters in other modules. *Semi-Modular* inference (SMI) (Carmona and Nicholls, 2020) interpolates between Bayes and Cut, controlling feedback using an influence parameter $\eta \in [0, 1]$. At $\eta = 1$, η -SMI is standard Bayesian inference and at $\eta = 0$, η -SMI is the Cut model. η can be identified with the learning rate parameter in a power posterior (Walker and Hjort, 2001; Zhang, 2006; Grünwald and van Ommen, 2017). Carmona and Nicholls (2020) suggest choosing η maximising the expected log pointwise predictive density (elpd), which can be estimated using cross-validation and WAIC (Watanabe. 2009). Other criteria for choosing the influence parameter are partly summarized in Wu and Martin, 2020. Chakraborty et al., 2022 introduce a new way to choose η based on prior-to-posterior Kullback-Leibler divergence.

These approaches face computational challenges due to intractable normalising constant. Nested Monte Carlo samplers are given in Plummer (2015) and Liu and Goudie (2020) and used in Carmona and Nicholls (2020) to target SMI posteriors. New developments include variational approximation (Yu et al., 2021).

There are several classical examples that people use when illustrating the Cut model or SMI. Epidemiological dataset, studying the correlation between human papilloma virus (HPV) prevalence and cervical cancer incidence, is introduced by Maucort-Boulch et al. (2008), revisited by several

authors including Plummer (2015) and Jacob et al. (2017) in the context of Cut models and Carmona and Nicholls (2020) and Nicholls et al. (2022) for SMI posteriors. Biased data, a synthetic example, was used by Liu et al. (2009) and Jacob et al. (2017) under Cut model scheme and Carmona and Nicholls (2020) and Nicholls et al. (2022) within SMI. Pharmacokinetic/ pharmacodynamic (PK/ PD) modelling (Bennett and Wakefield, 2001; Lunn et al., 2009) is a typical example used in Cut model. Agricultural dataset was introduced by Styring et al. (2017), analysed using a Cut model, and revisited by Carmona and Nicholls (2020) for SMI analysis. Nicholson et al. (2021b) introduced Covid prevalence debiasing model and analysed it in Nicholson et al. (2021a) using a cut model.

In this work, we extend the analysis on Covid prevalence debiasing model in the context of semi-modular inference. The goal is to construct a new dataset by taking a subset on the Covid dataset and explore semi-modular inference on this new test dataset. In order to choose the influence parameter η , we use prior-to-posterior Kullback-Leibler divergence (Chakraborty et al., 2022) instead of elpd. For small datasets, estimating elpd is impractical as LOOCV and WAIC are not reliable.

In section 2, we introduce the debiasing model and how to choose priors. In section 3, we discuss this model in SMI setting. In section 4, we use both nested MCMC and inversion method to simulate η -smi posterior samples. In section 5, we discuss the method to choose η based on prior-to-posterior Kullback-Leibler divergence. In section 6, we extend the analysis to data from different weeks and analyse how the optimal η changes with data. In section 7, we analyse using data from consecutive, multiple weeks and choose η maximizing elpd estimated using leave-one-out cross-validation (LOOCV). All results are reproducible using the accompanying R code.¹

2 Data and Model

In this section, we describe the data used in the case study and the corresponding model that characterises the data.

2.1 The data

We first describe the data, following Nicholson et al. (2021a).

2.1.1 Randomized surveillance data

These record u positive tests out of U total subjects tested. The REal-time ASsessment of COMMunity Transmission (REACT) study is a nationally representative prevalence survey of SARS-CoV-2 based on repeated polymerase chain reaction (PCR) tests of cross-sectional samples from a representative subpopulation defined through stratified random sampling from England's National Health Service patient register.

2.1.2 Targeted surveillance data

These record n positive tests of N total subjects tested. Pillar 1 tests comprise “all swab tests performed in Public Health England (PHE) labs and National Health Service (NHS) hospitals for those with a clinical need, and health and care workers”, and Pillar 2 is defined as “swab testing for the wider population”. Pillar 1+2 testing has more capacity than REACT, but the protocol incurs ascertainment bias as those at higher risk of being infected are more likely to be tested, such

¹https://github.com/StoneLiu2001/Debiasing_SMI

as front-line workers, contacts traced to a COVID-19 case, or the sub population presenting with COVID-19 symptoms, such as loss of taste and smell. The ascertainment bias potentially varies over the course of the pandemic as the testing strategy and capacity changes. We exclude lateral flow tests and use exclusively test data from Pillar 1+2 PCR tests.

2.1.3 Exploratory Data Analysis

The dataset we use is `southeast_df`. This dataset includes data from 60 consecutive weeks in the southeast region and each week's data is composed of (M, U, N, u, n) . Among them, there are 20 weeks where $U = 0$, which is not valid data in this setting. We focus on the other 40 weeks data. The index for each week that we mention later in this paper is ordered index after deleting the 20 invalid weeks. In the first part of the study, i.e. from section 3 to 6, we analysis data from single week. In section 7, we use data from multiple weeks to analyze.

Fig. 1 is the data plot, i.e. $\frac{u}{U}, \frac{n}{N}$ as a time series.

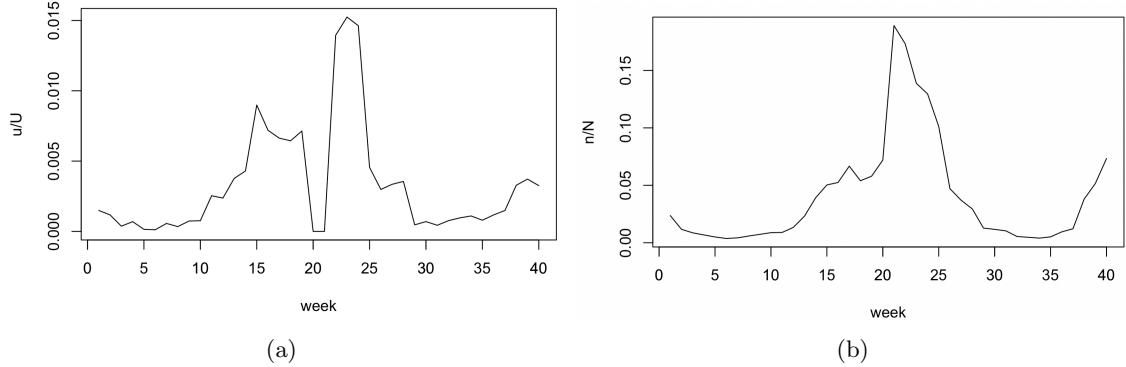


Figure 1: Plot of (a) $\frac{u}{U}$ and (b) $\frac{n}{N}$ as a time series.

2.2 Debiasing model

The model has two modules: a hyper-geometric model for the PCR-positive prevalence proportion π in the REACT data, and a causal model for the observation n of N positive targeted (e.g. Pillar 1+2) tests,

$$p(u|\pi) = \frac{\binom{\pi M}{u} \binom{M - \pi M}{U - u}}{\binom{\pi M}{U}}, \quad (1)$$

which is HyperGeometric($u|M, \pi M, U$), where M is the known population size, and πM is a latent integer number of infected.

$$p(n|\pi, q, p) = p_1(n|\pi, q)p_2(N - n|\pi), \quad (2)$$

where

$$p_1(n|\pi, q) = \text{Binomial}(n|\pi M, q) \quad (3)$$

$$p_2(N - n|\pi) = \text{Binomial}(N - n|(1 - \pi)M, p) \quad (4)$$

$$q = P(\text{Tested}|\text{Infected}) \quad (5)$$

$$p = P(\text{Tested}|\text{Not Infected}). \quad (6)$$

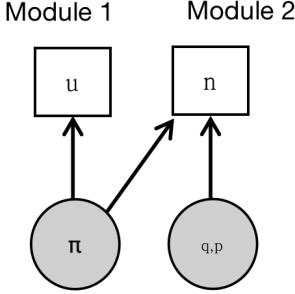


Figure 2: Graphical representation of the debiasing model. Grey circles denote unknown quantities to be inferred, and white boxes are fixed quantities.

Here is the intuition of the second model (Eq. 2). n and $N - n$ can be treated as draws from two different populations, i.e. n are the number of tested samples from infected population πM with tested probability q , $N - n$ are the number of tested samples from uninfected population $(1 - \pi)M$ with tested probability p . Of the ones who were infected, i.e. positive in the population, a binomial n number came forward to test. Of the ones who were not infected, i.e. negative in the population, a binomial $N - n$ number came forward to test.

Note that p is directly estimable from the targeted data, with $\hat{p} = \frac{N-n}{M}$ acting as a precise estimator with little bias when prevalence is low. We use p as a plug-in estimator in the second module likelihood. However, heuristically estimating p can contribute to model misspecification, we can use semi-modular inference to deal with model misspecification.

Fig. 2 illustrates the debiasing model with a graphical representation.

2.3 Prior elicitation

In this part, we show our choice for the prior based on common sense.

π is the PCR-positive prevalence proportion in the REACT (random) data, which is a quite small value in magnitude. Since π has value between 0 and 1, we assign a Beta distribution prior

$$\rho(\pi) = \text{Beta}(\pi; \alpha, \beta). \quad (7)$$

We require the mean to be 0.005, i.e. $\frac{\alpha}{\alpha+\beta} = 0.005$, and the probability that π is greater than 0.05 is 0.01, i.e 99th quantile of $\pi = 0.05$. Hence, we choose prior for π as Beta(1.01, 201).

q is the probability of tested conditional on being infected. It is not likely that everyone who is infected would get tested due to various reasons, such as mistake the symptoms as flu, or do not follow the government rules. Therefore, we assign a Beta distribution prior

$$\rho(q) = \text{Beta}(q; \alpha', \beta'). \quad (8)$$

We require the mean to be 0.75, i.e. $\frac{\alpha'}{\alpha'+\beta'} = 0.75$. Hence, we choose prior for q as Beta(4.5, 1.5). Consequently, \tilde{q} has the same prior as q . We will see later that this subjective prior is actually rather too weighted to large q .

3 Semi-Modular Inference

In this section, we analyze the debiasing model introduced in Section 2 under semi-modular inference scheme.

Let $p(u|\pi)$ and $p(n|\pi, q)$ denote the observation models for the two modules (refering to Eq. 1 and Eq. 2). Note that we use plug-in value for p , so we treat p as a constant in this setting and focus on π, q . We introduce an auxiliary parameter \tilde{q} which has the same distribution as q .

The conventional (full) Bayes posterior for this model is

$$\begin{aligned}\rho(\pi, q|u, n) &= \rho(\pi|u, n)\rho(q|n, \pi) \\ &\propto p(u|\pi)p(n|q, \pi)\rho(\pi, q),\end{aligned}\quad (9)$$

where $\pi \in [\frac{n}{M}, \frac{M-N+n}{M}]$, $q \in [0, 1]$.

The cut model posterior is

$$\begin{aligned}\rho_{cut}(\pi, q|u, n) &= \rho_{cut}(\pi|u)\rho(q|n, \pi) \\ &\propto p(u|\pi)\rho(\pi)\frac{p(n|q, \pi)\rho(\pi, q)}{\int p(n|q, \pi)\rho(\pi, q)dq}.\end{aligned}\quad (10)$$

The η -smi posterior is

$$\rho_{smi,\eta}(\pi, q, \tilde{q}|u, n) = \rho_{pow,\eta}(\pi, \tilde{q}|u, n)\rho(q|n, \pi), \quad (11)$$

where $\rho_{pow,\eta}(\pi, \tilde{q}|u, n)$ is the power posterior

$$\begin{aligned}\rho_{pow,\eta}(\pi, \tilde{q}|u, n) &\propto p(u|\pi)p(n|\pi, \tilde{q})^\eta\rho(\pi, \tilde{q}) \\ &= p(u|\pi)(p_1(n|\pi, \tilde{q})p_2(N-n|\pi))^\eta\rho(\pi, \tilde{q})\end{aligned}\quad (12)$$

$$\propto \rho_\eta(\tilde{q}|\pi, u, n)\rho_\eta(\pi|u, n). \quad (13)$$

We derive $\rho_\eta(\tilde{q}|\pi, u, n)$, $\rho_\eta(\pi|u, n)$ and $\rho(q|n, \pi)$ in Appendix A. Here we summarize the results.

$$\rho_\eta(\tilde{q}|\pi, u, n) = \text{Beta}(\tilde{q}; n\eta + \alpha', (\pi M - n)\eta + \beta'), \quad (14)$$

$$\rho_\eta(\pi|u, n) = \frac{g_\eta(\pi, u, n)}{c_\eta(u, n)}, \quad (15)$$

where

$$\begin{aligned}g_\eta(\pi, u, n) &= \frac{\Gamma(n\eta + \alpha')\Gamma((\pi M - n)\eta + \beta')}{\Gamma(\alpha' + \beta' + \pi M\eta)} \\ &\times \binom{\pi M}{u} \binom{M - \pi M}{U - u} \left(\binom{\pi M}{n} \binom{M - \pi M}{N - n} (1 - \hat{p})^{-\pi M} \right)^\eta \\ &\times \pi^{\alpha-1}(1 - \pi)^{\beta-1} \cdot \mathbb{1}\left\{ \frac{n}{M} \leq \pi \leq \frac{M - N + n}{M} \right\},\end{aligned}\quad (16)$$

$$c_\eta(u, n) = \int_{\frac{n}{M}}^{\frac{M-N+n}{M}} g_\eta(\pi, u, n) d\pi, \quad (17)$$

and

$$\begin{aligned}\rho(q|n, \pi) &\propto p(n|\pi, q)\rho(q|\pi) \\ &= \text{Beta}(q; n + \alpha', \pi M - n + \beta').\end{aligned}\quad (18)$$

The η -smi posterior of the original parameters is just the marginal

$$\rho_{smi,\eta}(\pi, q|u, n) = \int \rho_{smi,\eta}(\pi, q, \tilde{q}|u, n) d\tilde{q}. \quad (19)$$

4 Simulation

In this section, we use nested Markov Chain Monte Carlo (MCMC) to sample π, q from the η -smi posterior (Eq. 11). The data we use comes from the first week in `southeast_df` file mentioned in Nicholson et al. (2021a), where $M = 9180135, N = 77406, n = 1830, U = 2011, u = 3$.

The nested MCMC procedure is described in Algorithm 1: sample N_1 draws from $\rho_{pow,\eta}(\pi, \tilde{q}|u, n)$; for each sampled value of π , run a sub-chain targeting $\rho(q|n, \pi)$ for N_2 steps, where N_2 is large enough to avoid initialisation bias; keep only the last sampled value in this sub-chain. The resulting joint samples (π, \tilde{q}, q) are approximately distributed according to the η -smi posterior (Eq. 11). We typically ignore the output for \tilde{q} as we target the marginal in Eq. 19.

Algorithm 1 Nested MCMC for η -smi posterior Eq. 11

Input: influence $\eta \in [0,1]$, Data $(u, n) = \{(u_i, n_i)\}_{i=1}^n$; observational models $p(u|\pi)$ and $p(n|\pi, q)$; prior $\rho(\pi, q)$; run-lengths N_1 and N_2 .

Output: Samples $\{(\pi^{(s)}, q^{(s)})\}_{s=1}^{N_1}$ distributed approximately according to the η -smi posterior $\rho_{smi,\eta}(\pi, q|u, n)$ with influence parameter η .

```

for  $s = 1, \dots, N_1$  do
    Sample  $(\pi^{(s)}, \tilde{q}^{(s)}) \sim \rho_{pow,\eta}(\pi, \tilde{q}|u, n)$ , using any standard sampler.
end for
Let  $\{\pi^{(s)}\}_{s=1}^{N_1}$  be samples after burn-in and thinning.

for  $s = 1, \dots, N_1$  do
    for  $s = 1, \dots, N_2$  do
        Sample  $(q^{s,r}) \sim \rho(q|n, \pi^{(s)})$ , using any standard sampler.
    end for
    Let  $q^{(s)} = q^{(s,N_2)}$  (final state).
end for
return  $\{(\pi^{(s)}, q^{(s)})\}_{s=1}^{N_1}$ 

```

To update π, \tilde{q} in the first loop in Algorithm 1, we use Metropolis-within-Gibbs scheme. The procedure is displayed in Algorithm 2.

Algorithm 2 Update π, \tilde{q}

Starting from an arbitrary $(\pi^{(1)}, \tilde{q}^{(1)})$, iterate for $t = 2, \dots, N_1$

1. Sample $\pi \sim \rho(\pi)$.
 2. Compute $\alpha((\pi)|\pi^{(t-1)}) = \min\left(1, \frac{\rho_\eta(\pi|u,n)\rho(\pi^{t-1})}{\rho_\eta(\pi^{(t-1)}|u,n)\rho(\pi)}\right)$
 3. With probability $\alpha((\pi)|\pi^{(t-1)})$, set $\pi^{(t)} = \pi$, otherwise set $\pi^{(t)} = \pi^{(t-1)}$
 4. Sample $\tilde{q}^{(t)} \sim \rho_\eta(\tilde{q}|\pi^{(t)}, u, n)$ (refer to Eq. 14).
-

After obtaining $\{\pi^{(s)}\}_{s=1}^{N_1}$ after burn-in and thinning, we sample q directly from Eq. 18.

The trace plot and scatter plot of MCMC samples of π, q can be found in Appendix B.

In order to check MCMC result, we sample π, q from η -smi posterior using inversion method. The detail is included in Appendix E.

Fig. 3 shows the density plot of π and q individually where $\eta = 1$, including the prior (blue), posterior with subjective beta prior using MCMC sampling (red), posterior with subjective beta prior using inversion method (black), posterior with uniform prior using MCMC sampling (green), and posterior with uniform prior using inversion method (pink). The red line and the black line are nearly identical as expected, indicating that both nested MCMC and inversion method are valid approaches to sample η -smi posterior.

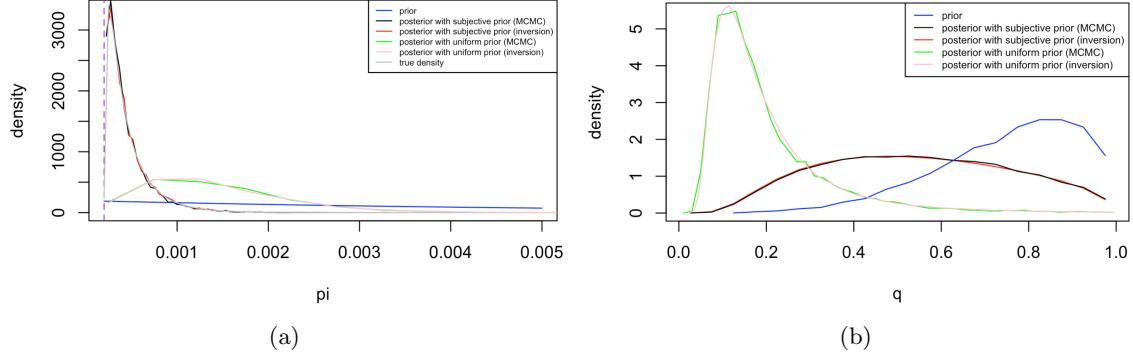


Figure 3: Density plot for π (a) and q with $\eta = 1$ (b) showing the prior (blue), posterior with subjective beta prior using MCMC sampling (red), posterior with subjective beta prior using inversion method (black), posterior with uniform prior using MCMC sampling (green), and posterior with uniform prior using inversion method (pink). In (a), the grey line is the true density of π and the purple dotted line is the lower bound $\frac{n}{M}$.

5 Choosing the Influence Parameter η

In Carmona and Nicholls (2020), they use expected log pointwise predictive density (elpd) to determine the optimal η . They suggest that when the true data-generating distribution p^* is unknown, WAIC or cross-validation can be used to estimate elpd. However, in the debiasing model example, since the number of data sample is one, we cannot estimate elpd via cross-validation. We can try WAIC since the samples are pseudo-replicate. We also consider prior-to-posterior divergence to choose the optimal influence parameter η .

The idea we use here is motivated by Chakraborty et al. (2022).

First, we note that the Kullback-Leibler divergence between the cut posterior distribution (Eq. 10) and the η -smi posterior distribution (Eq. 19) is the Kullback-Leibler divergence between their marginal posterior distribution for π . This follows from the fact that the conditional posterior distribution for q given π is the same in both distributions (see Proposition 1).

Proposition 1.

$$KL(\rho_{smi,\eta}(\pi, q|u, n) || \rho_{cut}(\pi, q|u, n)) = KL(\rho_{smi,\eta}(\pi|u, n) || \rho_{cut}(\pi|u)).$$

Proof.

$$\begin{aligned}
\text{KL}(\rho_{smi,\eta}(\pi, q|u, n) || \rho_{cut}(\pi, q|u, n)) &= \int \int \rho_{smi,\eta}(\pi, q|u, n) \log \frac{\rho_\eta(\pi|u, n)\rho(q|\pi, n)}{\rho(\pi|u)\rho(q|\pi, n)} dq d\pi \\
&= \int \log \frac{\rho_\eta(\pi|u, n)}{\rho(\pi|u)} \int \rho_{smi,\eta}(\pi, q|u, n) dq d\pi \\
&= \int \rho_{smi,\eta}(\pi|u, n) \log \frac{\rho_\eta(\pi|u, n)}{\rho(\pi|u)} d\pi \\
&= \text{KL}(\rho_{smi,\eta}(\pi|u, n) || \rho_{cut}(\pi|u)).
\end{aligned} \tag{20}$$

□

Since module 1 (Eq. 1) is reliable, inference for π from the cut model (Eq. 10) is accurate but has larger variance compared with semi-modular inference (Eq. 19).

We define

$$d(n; \eta) := \text{KL}(\rho_{smi,\eta}(\pi|u, n) || \rho_{cut}(\pi|u)), \tag{21}$$

which can be thought of as a prior-to-posterior divergence, where we treat u as known when forming the prior. $d(n; \eta)$ measures the divergence from the cut posterior after introducing information n .

Then we define a tail probability. Let $\varepsilon \in [0, 1]$ give tail probability

$$\varepsilon_\eta := P_r(d(n; \eta = 1) > d(n_{obs}; \eta)), \tag{22}$$

where $n \sim p(\cdot|q, \pi)$ (Eq. 2), $\pi \sim \rho_{cut}(\cdot|u)$, and $q \sim \rho(\cdot|\pi)$.

We then choose η as the largest value η' such that $\varepsilon_{\eta'} > \alpha$, where α is the significance level such as 0.05. The intuition of this criterion is: if there is no conflict at level α between the cut posterior and η -smi posterior, then we choose η -smi posterior ($\eta > 0$) as it has smaller variance; if there is a conflict, then we trust the cut posterior and choose smaller η .

We now consider how to compute the KL divergence in Eq. 21.

Define

$$f_\eta(\pi; n) = \log \left(\frac{\rho_{smi,\eta}(\pi|u, n)}{\rho_{cut}(\pi|u)} \right) \tag{23}$$

$$= \log \left(\frac{\rho_\eta(\pi|u, n)}{p(u|\pi)\rho(\pi)} \right) + \log \left(\int_{a_n}^{b_n} p(u|\pi)\rho(\pi) d\pi \right) \tag{24}$$

$$= \log \rho_\eta(\pi|u, n) - \log(p(u|\pi)\rho(\pi)) + \log \left(\int_{a_n}^{b_n} p(u|\pi)\rho(\pi) d\pi \right) \tag{25}$$

$$= \log g_\eta(\pi, u, n) - \log c_\eta(u, n) - \log(p(u|\pi)\rho(\pi)) + \log \left(\int_{a_n}^{b_n} p(u|\pi)\rho(\pi) d\pi \right), \tag{26}$$

where $p(u|\pi)$ is given in Eq. 1, $\rho(\pi)$ is given in Eq. 7, $g_\eta(\pi, u, n)$ is given in Eq. 16, $c_\eta(u, n)$ is given in Eq. 17, $a_n = \frac{n}{M}$, and $b_n = \frac{M-N+n}{M}$.

To calculate $d(n_{obs}; \eta)$, we evaluate $f_\eta(\pi; n)$ at $\{\pi^{(i)}\}_{i=1}^N$ samples from the η -smi posterior using the inversion method and $n = n_{obs}$, then take average of $\{f_\eta(\pi^{(i)}; n_{obs})\}_{i=1}^N$ to get $\bar{f}_\eta(\pi; n_{obs})$, which can be treated as an estimate of $d(n_{obs}; \eta)$. Here, we take $N = 1000$. We experiment with larger N value, i.e. 10000, the result does not change much, indicating that $N = 1000$ is a descent choice.

To estimate ε_η (Eq. 22), we first take $\{\pi^{(i)}\}_{i=1}^{N'}$ samples from the cut posterior using the inversion method. For each $\pi^{(i)}$, we sample $q^{(i)} \sim \rho(\cdot|\pi)$, and then generate $n^{(i)} \sim p(\cdot|\pi^{(i)}, q^{(i)})$ (Eq. 2) using

Algorithm 3 Estimate ε_η (Eq. 22)

```

for  $i = 1, \dots, N'$  do
    Sample  $\pi^{(i)} \sim \rho_{cut}(\cdot|u)$  using any standard sampler.
    Sample  $q^{(i)} \sim \rho(\cdot|\pi)$  using any standard sampler.
    Sample  $n^{(i)} \sim p(\cdot|q^{(i)}, \pi^{(i)})$  using any standard sampler.
    for  $j = 1, \dots, N$  do
        Sample  $\pi^{(i,j)} \sim \rho_{\eta=1}(\cdot|u, n^{(i)})$  using any standard sampler.
        Evaluate  $c_\eta(u, n)$  and  $\int_{a_n}^{b_n} p(u|\pi)\rho(\pi)d\pi$  by quadrature. Calculate  $f_{\eta=1}(\pi^{(i,j)}; n^{(i)})$ .
    end for
    Estimate  $\hat{d}(n^{(i)}; \eta = 1) = \frac{1}{N} \sum_{j=1}^N f_{\eta=1}(\pi^{(i,j)}; n^{(i)})$ 
end for
Estimate  $\widehat{\varepsilon}_\eta = \frac{1}{N'} \sum_{i=1}^{N'} \mathbb{1}\{\hat{d}(n^{(i)}; \eta = 1) \geq \hat{d}(n_{obs}; \eta)\}$ 

```

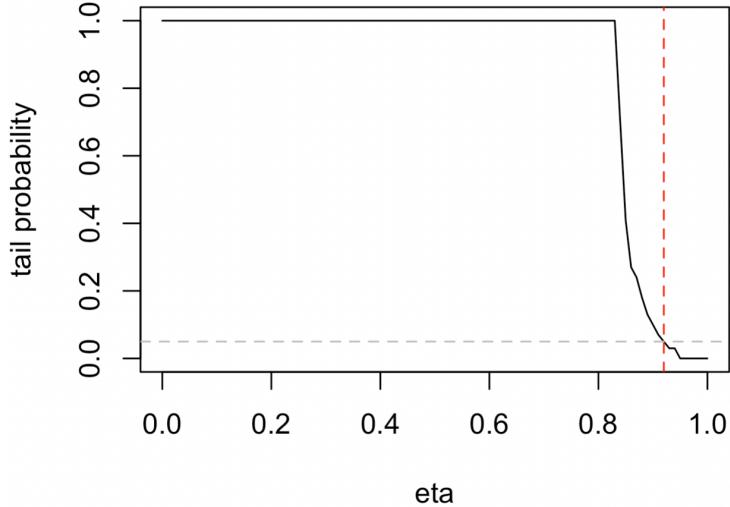


Figure 4: ε_η with η varying from 0 to 1. The grey dotted line is $\varepsilon = 0.05$. The red dotted line indicates the largest η satisfying $\varepsilon_\eta > 0.05$, which is 0.92 here.

random walk Metropolis-Hastings MCMC algorithm, calculate $d_{\pi^{(i)}}(n^{(i)}; \eta = 1)$, estimate $\varepsilon_{i,\eta}$ by the ratio of number of $\{d_{\pi^{(i)}}(n^{(i)}; \eta = 1)\}_{s=1}^{N'}$ exceeding $d(n_{obs}; \eta)$. Here, we take $N' = 100$. The procedure is shown in Algorithm 3.

Fig. 4 shows the value of ε_η with η varying from 0 to 1. We set the significance level to be 0.05, then we choose the largest η that satisfies $\varepsilon_\eta > 0.05$, which is 0.92 approximately. We repeat the procedure ten times and each time we use independent samples. The ten plots are quite close to each other, i.e. the lines have little variation and all of them says that $\eta \approx 0.92$ is the optimal one, indicating that our choice of $N = 1000, N' = 100$ is suitable.

In Fig. 5(a), we plot $\rho_{smi,\eta}(\pi|u, n_{obs})$ with $\eta = 0, 0.1, 0.2, \dots, 1$. We can see that the distribution does not change much as η varies. We also plot $\rho_{\eta=1}(\pi|u, n)$ with 10 different n sampled according to Eq. 2. These plots differ quite a lot from $\rho_{cut}(\pi|u)$, indicating that $d(n; \eta = 1)$ in Eq. 22 is usually greater than $d(n_{obs}; \eta)$. This is in line with the fact in Fig. 4 that the tail probability remains 1 for a long time.

In this example, n is pseudo-replicate, so we also use WAIC to estimate elpd in terms of n . The result is in Fig. 6. The plot is very flat, i.e. elpd value varies in a small range. This is due to the

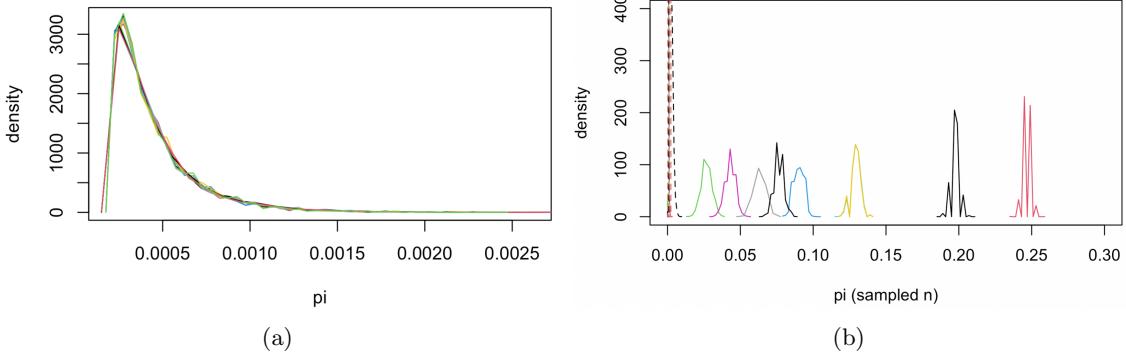


Figure 5: (a) $\rho_{smi,\eta}(\pi|u, n_{obs})$ with $\eta = 0, 0.1, 0.2, \dots, 1$; (b) $\rho_{\eta=1}(\pi|u, n)$ with 10 different n sampled according to Eq. 2. The dotted line is $\rho_{cut}(\pi|u)$ associated with each random sample.

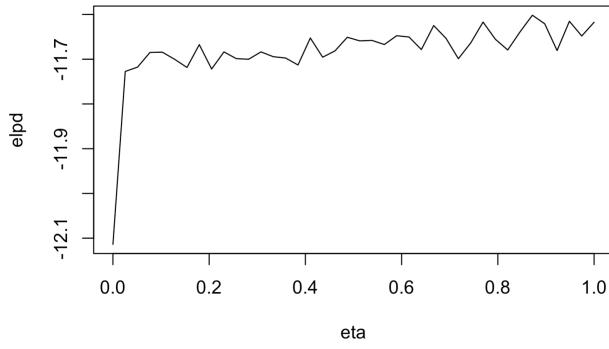


Figure 6: elpd estimated by WAIC.

fact that $\rho_{smi,\eta}(\pi|u, n)$ does not change much as η varies from 0 to 1.

6 Single Group of Data Analysis

In this section, we vary the value for M, U, N, u, n and evaluate the optimal η value. `southeast_df` contains data from 60 weeks, from 2020-05-31 to 2021-07-18. Among these 60 groups of data, there are 20 weeks with $U = 0$, i.e. no REACT data, which can be treated as invalid data. We use the remaining 40 groups of data, and choose the optimal η for each week, following procedures mentioned in previous sections.

The result is shown in Fig. 7(a).

To get an intuition about how the optimal η value is influenced by the data, we do a rough approximation. π can be approximated by $\frac{u}{U}$, and q can be approximated by $\frac{n}{\pi M} \approx \frac{nU}{Mu}$. From Fig. 7(b), we can observe that as estimated q increases, optimal η value increases generally. The reason is partly due to our poor prior distribution for q (Eq. 8). For small q value, since it is far from prior mean, which is 0.75 in our setting, model misspecification occurs. SMI inference tends to choose small η to mitigate the influence of misspecified module. We also experiment with q assigned a uniform[0,1] prior. It turns out that $\hat{\eta} = 1$ for all weeks.

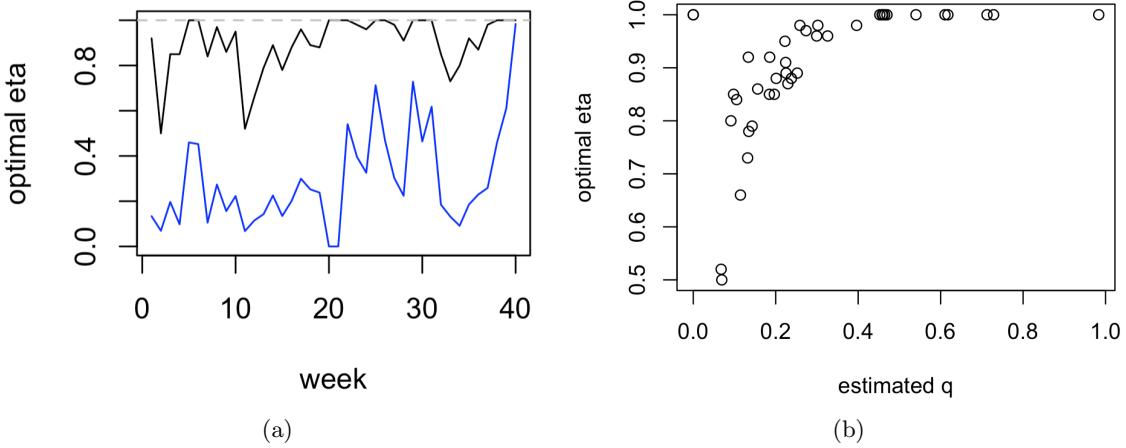


Figure 7: Different optimal η value: (a) Optimal η value in different weeks. The grey dotted line is the case where q has a uniform prior. The blue line is $\frac{nU}{Mu}$; (b) Relationship between optimal η value and $\frac{nU}{Mu}$.

7 Multiple Groups of Data Analysis

In this section, rather than analyzing with one-week data, we extend the sample size into multiple weeks. To do so, we have to ensure theta the data we use have approximately the same π, q values, such as data from consecutive weeks. We first calculate $\frac{u}{U}$, which can be treated as a rough estimate of π , and find that data from week 3, 4, 5, 6, 7, 8, 9, 10 in `southeast_df` is a reasonable choice (refer to Fig. 1(a), $\frac{u}{U}$ is quite flat during these weeks).

The analysis is essentially the same as what we have done in the previous sections, except that Eq. 14 becomes

$$\tilde{\rho}_\eta(\tilde{q}|\pi, u, n) = \text{Beta}(\tilde{q}; \eta \sum_{i=1}^k (n_i) + \alpha', \eta \sum_{i=1}^k (\pi M_i - n_i) + \beta'), \quad (27)$$

Eq. 15 becomes

$$\tilde{\rho}_\eta(\pi|u, n) = \frac{\tilde{g}_\eta(\pi, u, n)}{\tilde{c}_\eta(u, n)}, \quad (28)$$

where

$$\begin{aligned} \tilde{g}_\eta(\pi, u, n) &= \frac{\Gamma(\eta \sum_{i=1}^k (n_i) + \alpha') \Gamma(\eta \sum_{i=1}^k (\pi M_i - n_i) + \beta')}{\Gamma(\alpha' + \beta' + \eta \sum_{i=1}^k \pi M_i)} \\ &\times \prod_{i=1}^k \left\{ \frac{\binom{\pi M_i}{u_i} \binom{M_i - \pi M_i}{U_i - u_i}}{\binom{M_i}{U_i}} \left(\binom{\pi M_i}{n_i} \binom{M_i - \pi M_i}{N_i - n_i} (1 - \hat{p}_i)^{-\pi M_i} \right)^\eta \right. \\ &\times \left. \pi^{\alpha-1} (1 - \pi)^{\beta-1} \mathbb{1}\left\{ \frac{n_i}{M_i} \leq \pi \leq \frac{M_i - N_i + n_i}{M_i} \right\} \right\}, \end{aligned} \quad (29)$$

$$\tilde{c}_\eta(u, n) = \int \tilde{g}_\eta(\pi, u, n) d\pi, \quad (30)$$

and Eq. 18 becomes

$$\tilde{\rho}(q|n, \pi) = \text{Beta}(q; \sum_{i=1}^k (n_i) + \alpha', \sum_{i=1}^k (\pi M_i - n_i) + \beta'). \quad (31)$$

The trace plot of MCMC samples is displayed in Appendix B. The samples are more stable compared with single group of data analysis.

7.1 Choose the influence parameter η with elpd

In this setting, since we have multiple data, we can utilize *expected log pointwise predictive density* (elpd) as a criterion to select the optimal η ,

$$\text{elpd}(\eta) = \int \int p^*(u', n') \cdot \log \rho_{smi, \eta}(u', n' | u, n) du' dn', \quad (32)$$

where p^* is the distribution representing the true data-generating process and

$$p_{smi, \eta}(u', n' | u, n) = \int \int p(u', n' | \pi, q) \cdot \tilde{\rho}_{smi, \eta}(\pi, q | u, n) d\pi dq \quad (33)$$

is a candidate posterior predictive distribution, indexed by η . We take the value, $\eta = \eta^*$, which maximizes the estimated elpd function.

Since we do not know the true data-generating process p^* in this example, we can use leave-one-out cross validation (LOOCV) to approximate elpd (Eq. 33) as suggested by Carmona and Nicholls (2020). The procedure is as follows.

We can carry out LOOCV in three ways: LOOCV on n , on u , and on both n and u , depending on which parameter we focus on. Here we show how to conduct LOOCV on n . The other two follow similarly.

Let $M = (M_i)_{i=1}^8, U = (U_i)_{i=1}^8, N = (N_i)_{i=1}^8, u = (u_i)_{i=1}^8, n = (n_i)_{i=1}^8$ denote the groups of data we use. Let

$$p_\eta(n_i | u, n_{-i}) = \int p(n_i | \pi, q) \tilde{\rho}_{smi, \eta}(\pi, q | u, n_{-i}) d\pi dq, \quad (34)$$

where n_{-i} denote all n data except the i th one. Since analysis on q are essentially the same for η -smi posterior with different η values, we focus on π to choose the optimal η . We suspect that module 2 may contaminate π , so we do LOOCV on n .

To compute this predictive density (Eq. 34), we can evaluate the expectation using draws from $\rho_{smi, \eta}(\pi, q | u, n_{-i})$, the posterior simulations, which we label $\{(\pi^{(s)}, q^{(s)})\}_{s=1}^S$:

$$\hat{p}_\eta(n_i | u, n_{-i}) = \frac{1}{S} \sum_{s=1}^S p(n_i | \pi^{(s)}, q^{(s)}). \quad (35)$$

We assume that the number of simulation draws S is large enough to fully capture the posterior distribution.

Then, elpd is estimated by

$$\widehat{\text{elpd}} = \frac{1}{D} \sum_{i=1}^D \log(\hat{p}_\eta(n_i | u, n_{-i})), \quad (36)$$

where $D = 8$ is the number of groups of data we use here.

The result is shown in Fig. 8(a). We can see that elpd reaches its maximum at η close to 1. So we choose $\hat{\eta} = 1$ here, corresponding to the full Bayes model. This is in line with our result for choosing η using prior-to-posterior tail probability in section 6. In Fig. 8(a), optimal η in week 3 to 10 are close to 1 on average. We repeat estimating elpd ten times, each time we use samples from nested MCMC with $N = 1000000$ and thinning interval 100. Then ten plots vary to some extent, but all indicate that $\hat{\eta} \approx 1$, i.e. with the largest \widehat{elpd} .

We also do LOOCV on u and on both u and n . The results are shown in Fig. 8(b) and Fig. 8(c), respectively. The optimal η associated with the largest elpd is $\hat{\eta} \approx 0$ and $\hat{\eta} \approx 0.7$, respectively. The joint posterior of (π, q) associated with $\eta = 0, 1, 0.7$ is shown in Fig. 8(d).

The optimal η obtained in three cases meet our expectation. When we do LOOCV on n , we focus on parameter q . In this model, q depends on π (refer to Eq. 18), thus we tend to use as much information from module 1 as possible to make inference on q . When we do LOOCV on u , our focus is π . As module 2 might be misspecified, we tend to reduce its influence on π , so we choose small η value. Choosing which η value to use depends on our goal, i.e. which parameter we are interested in.

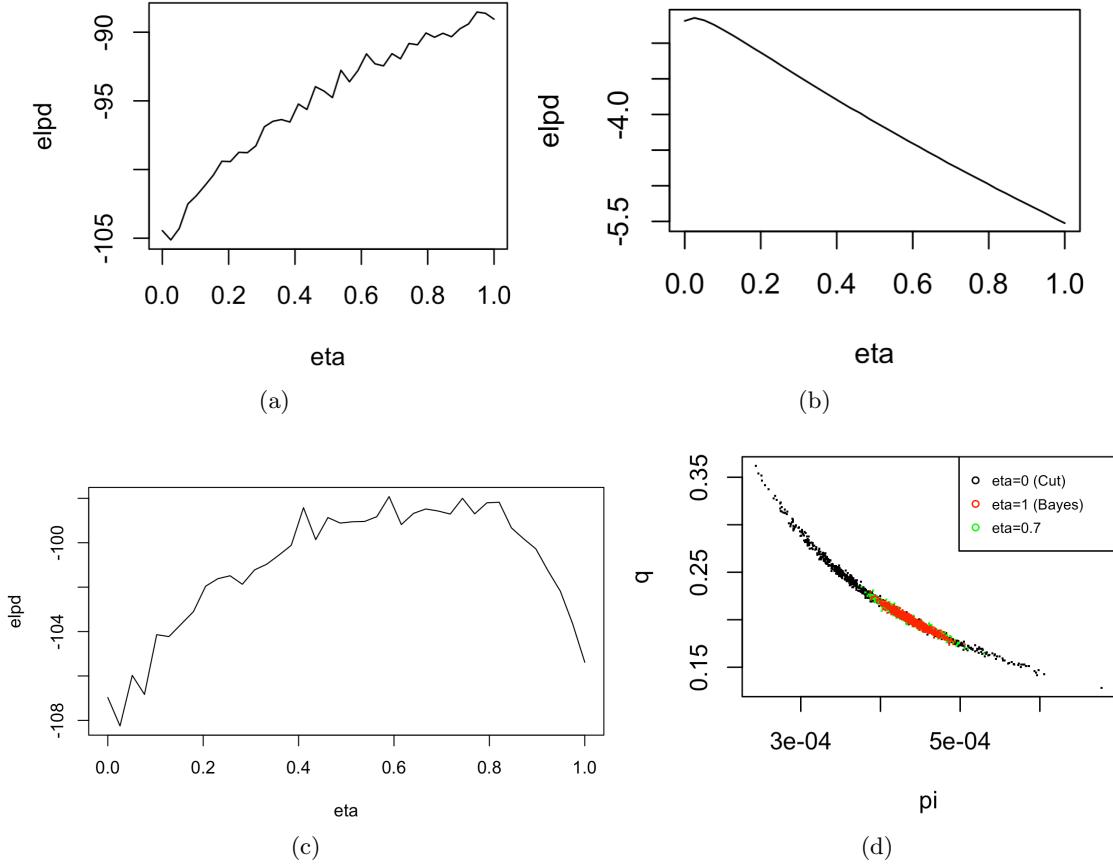


Figure 8: Estimated elpd with different η value: (a) LOOCV on n ; (b) LOOCV on u ; (c) LOOCV on both u and n . (d) joint posterior of (π, q)

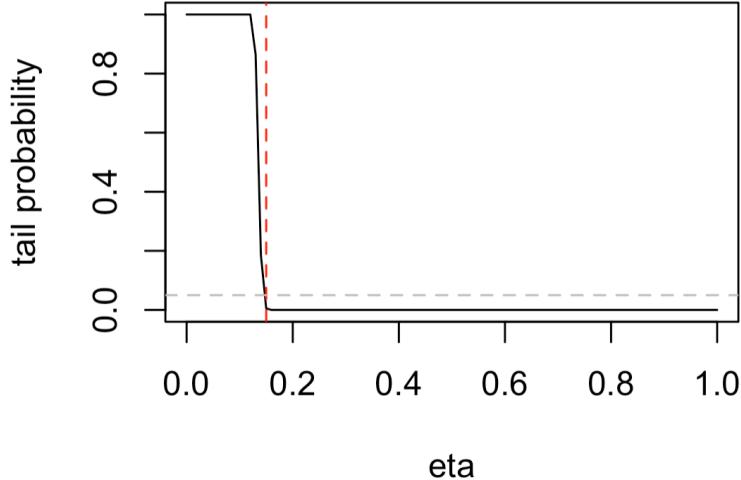


Figure 9: ε_η with η varying from 0 to 1 (multiple groups of data case). The grey dotted line is $\varepsilon = 0.05$. The red dotted line indicates the largest η satisfying $\varepsilon_\eta > 0.05$, which is 0.15 here.

7.2 Choose the influence parameter η with prior-to-posterior divergence

In addition to elpd, we also choose η based on prior-to-posterior KL divergence mentioned in section 5. The procedure is essentially the same as Algorithm 3, except that instead of sampling one n from $p(n|q, \pi)$ each time, we sample eight n samples.

Define

$$\tilde{f}_\eta(\pi; n) = \log \left(\frac{\tilde{\rho}_{smi,\eta}(\pi|u, n)}{\tilde{\rho}_{cut}(\pi|u)} \right) \quad (37)$$

$$= \log \tilde{g}_\eta(\pi, u, n) - \log \tilde{c}_\eta(u, n) - \log(p(u|\pi)\rho(\pi)) + \log \left(\int_{\tilde{a}_n}^{\tilde{b}_n} p(u|\pi)\rho(\pi)d\pi \right), \quad (38)$$

where $\tilde{g}_\eta(\pi, u, n)$ is given in Eq. 29, $\tilde{c}_\eta(u, n)$ is given in Eq. 30, $\tilde{a}_n = \max\{\frac{n_i}{M_i}, i = 1, \dots, 8\}$, $\tilde{b}_n = \min\{\frac{M_i - N_i + n_i}{M_i}, i = 1, \dots, 8\}$.

$d(n_{obs}; \eta)$ is calculated in the same way as that mentioned in section 5. To estimate ε_η (Eq. 22), we first take $\{\pi^{(i)}\}_{i=1}^{N'}$ samples from the cut posterior using the inversion method. For each $\pi^{(i)}$, we sample $q^{(i)} \sim \rho(\cdot|\pi)$, and then generate $n^{(i)} \sim p(\cdot|\pi^{(i)}, q^{(i)})$ (Eq. 2) using random walk Metropolis-Hastings MCMC algorithm, where $n^{(i)} = \{n^{(i,j)}\}_{j=1}^8$. The rest is the same as section 5.

The result is in Fig. 9. The optimal η is chosen as 0.15, which is close to the result using LOOCV on u (Fig. 8(b)).

8 Conclusion

In this report, we analyze the debiasing model under semi-modular inference scheme. We choose beta prior for both parameters π, q . We sample π, q from η -smi posterior using both Metropolis-within-Gibbs sampling and inversion method. Due to the limited number of data in this example, we choose the influence parameter η based on prior-to-posterior Kullback-Leibler divergence, and the result shows that $\eta = 0.92$ is approximately the optimal choice. We then extend the analysis to data from different weeks and see whether there is some connection between the optimal η and

data. We also do multiple week analysis and choose the influence parameter via elpd. The optimal influence parameter given by the two methods match well.

9 References

- Bennett, J., Wakefield, J. (2001). Errors-in-variables in joint population phar- macokinetic/pharmacodynamic. *Biometrics*, 57:803–812.
- Carmona, C. U., & Nicholls, G. K. (2020). Semi-Modular Inference: enhanced learning in multi- modular models by tempering the influence of components. 108, 4226-4235.
- Chakraborty, A., Nott, D. J., Drovandi, C., Frazier, D. T., & Sisson, S. A. (2022). Modular- ized Bayesian analyses and cutting feedback in likelihood-free inference. *arXiv preprint arXiv: 2203.09782*.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin. D. B. (2013). *Bayesian Data Analysis, Third Edition*. Chapman and Hall/CRC.
- Grünwald, P. and van Ommen, T. (2017). Inconsistency of Bayesian Inference for Misspecified Linear Models, and a Proposal for Repairing It. *Bayesian Analysis*, 12(4):1069–1103.
- Jacob, P. E., Murray, L. M., Holmes, C. C., and Robert, C. P. (2017). Better together? Statistical learning in models made of modules.
- Liu, F., Bayarri, M. J., and Berger, J. O. (2009). Modularization in Bayesian analysis, with emphasis on analysis of computer models. *Bayesian Analysis*, 4(1):119–150.
- Liu, Y. and Goudie, R. J. B. (2020). Stochastic Approximation Cut Algorithm for Inference in Modularized Bayesian Models. *Statistics and Computing*, 32.
- Lunn, D., Best, N., Spiegelhalter, D., Graham, G., and Neuenschwander, B. (2009). Combining MCMC with ‘sequential’ PKPD modelling. *Journal of Pharmacokinetics and Pharmacodynamics*, 36(1):19–38.
- Maucort-Boulch, D., Franceschi, S., and Plummer, M. (2008). International Correlation between Human Papillomavirus Prevalence and Cervical Cancer Incidence. *Cancer Epidemiology Biomarkers & Prevention*, 17(3):717–720.
- Nicholls, G. K., Lee, J. E., Wu, C.-H., & Carmona, C. U. (2022). Valid belief updates for pre- quentially additive loss functions arising in Semi-Modular Inference. *arXiv preprint arXiv: 2201.09706*.
- Nicholson, G., Lehmann, B., Padellini, T., Pouwels, K. B., Jersakova, R., Lomax, J., King, R. E., Mallon, A. E., Diggle, P. J., Richardson, S., Blangiardo, M., & Holmes, C. (2021b). Local prevalence of transmissible SARS-CoV-2 infection: an integrative causal model for debiasing fine-scale targeted testing data. *medRxiv*
- Plummer, M. (2015). Cuts in Bayesian graphical models. *Statistics and Computing*, 25(1):37–43.
- Spiegelhalter, D. J., Thomas, A., Best, N., and Lunn, D. (2014). OpenBUGS User Manual.
- Styring, A. K., Charles, M., Fantone, F., Hald, M. M., McMahon, A., Meadow, R. H., Nicholls, G. K., Patel, A. K., Pitre, M. C., Smith, A., Sołtysiak, A., Stein, G., Weber, J. A., Weiss, H., and Bogaard, A. (2017). Isotope evidence for agricultural intensification reveals how the world’s first cities were fed. *Nature Plants*, 3(6).
- Walker, S. and Hjort, N. L. (2001). On Bayesian consistency. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(4):811–821.
- Watanabe, S. (2009). *Algebraic Geometry and Statistical Learning Theory*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press.
- Wu, P.-S. and Martin, R. (2020). A comparison of learning rate selection methods in generalized Bayesian inference. *arXiv preprint arXiv: 2012.11349*

- Yu, X., Nott, D. J., and Smith, M. S. (2021). Variational inference for cutting feedback in misspecified models. *arXiv preprint arXiv: 2108.11066*.
- Zhang, T. (2006). Information-theoretic upper and lower bounds for statistical estimation. *IEEE Transactions on Information Theory*, 52(4): 1307–1321.

Appendix A: Explicit formulae for η -smi posterior

In this appendix, we present the details of how to derive η -smi posterior, corresponding to the results in section 3.

$$\begin{aligned}\rho_{pow,\eta}(\pi, \tilde{q}|u, n) &\propto p(u|\pi)p(n|\pi, \tilde{q})^\eta \rho(\pi, \tilde{q}) \\ &= p(u|\pi)(p_1(n|\pi, \tilde{q})p_2(N-n|\pi))^\eta \rho(\pi, \tilde{q}) \\ &\propto \binom{\pi M}{u} \binom{M-\pi M}{U-u} \left(\binom{\pi M}{n} \tilde{q}^n (1-\tilde{q})^{\pi M-n} \binom{M-\pi M}{N-n} (1-\hat{p})^{-\pi M} \right)^\eta \\ &\quad \times \tilde{q}^{\alpha'-1} (1-\tilde{q})^{\beta'-1} \pi^{\alpha-1} (1-\pi)^{\beta-1}\end{aligned}$$

plug in Eq. 1, 2, 7, 8

$$\propto \left(\frac{\Gamma(\alpha' + \beta' + \pi M \eta)}{\Gamma(n\eta + \alpha')\Gamma((\pi M - n)\eta + \beta')} \tilde{q}^{n\eta + \alpha' - 1} (1 - \tilde{q})^{(\pi M - n)\eta + \beta' - 1} \right)$$

\tilde{q} follows a Beta distribution

$$\begin{aligned}&\times \frac{\Gamma(n\eta + \alpha')\Gamma((\pi M - n)\eta + \beta')}{\Gamma(\alpha' + \beta' + \pi M \eta)} \binom{\pi M}{u} \binom{M - \pi M}{U - u} \\ &\times \left(\binom{\pi M}{n} \binom{M - \pi M}{N - n} (1 - \hat{p})^{-\pi M} \right)^\eta \\ &\times \pi^{\alpha-1} (1 - \pi)^{\beta-1} \cdot \mathbb{1}\left\{ \frac{n}{M} \leq \pi \leq \frac{M - N + n}{M} \right\} \\ &\propto \rho_\eta(\tilde{q}|\pi, u, n) \rho_\eta(\pi|u, n),\end{aligned}$$

where $\rho_\eta(\tilde{q}|\pi, u, n)$ is given in Eq. 14 and $\rho_\eta(\pi|u, n)$ is given in Eq. 15.

The condition distribution of q given u, n is

$$\begin{aligned}\rho(q|n, \pi) &\propto p(n|\pi, q)\rho(q|\pi) \\ &= p_1(n|\pi, q)p_2(N-n|\pi)\rho(q)\end{aligned}$$

plug in Eq. 2, 8

$$\propto q^{n+\alpha'-1} (1 - q)^{\pi M - n + \beta' - 1}.$$

Therefore, $\rho(q|n, \pi)$ is the same as that in Eq. 16.

Appendix B: Trace plot, autocorrelation plot, scatter plot and effective sample size of MCMC samples

In the appendix, we provide the trace plot, scatter plot and effective sample size (ESS) of MCMC samples (π, q) under subjective prior and uniform prior respectively.

Fig. 10 is the trace plot of MCMC samples of π, q , with $N_1 = 100000, \eta = 1$ with thinning interval of 100 under the subjective prior, i.e. $\pi \sim \text{Beta}(\alpha = 1.01, \beta = 201), q \sim \text{Beta}(\alpha = 4.5, \beta = 1.5)$. Fig. 11 is the corresponding autocorrelation plot.

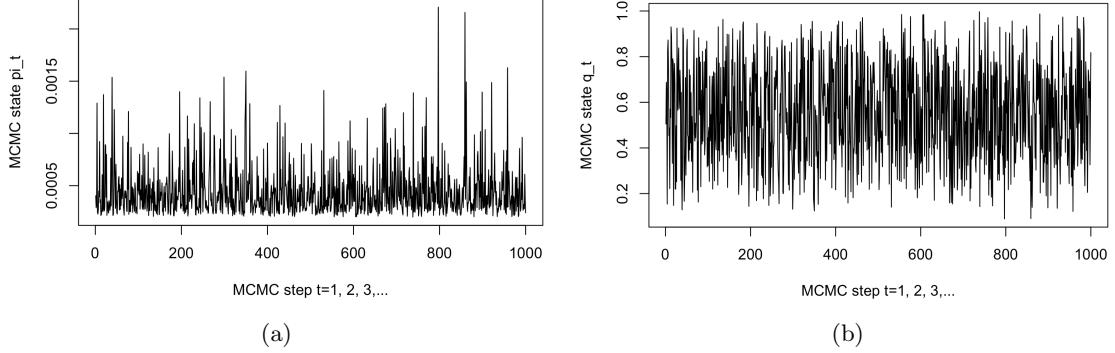


Figure 10: Nested MCMC targeting the η -smi posterior with $\eta = 1$ given subjective prior with thinning interval of 100 (Eq. 5): (a) MCMC trace of π^t ; (b) MCMC trace of q^t .

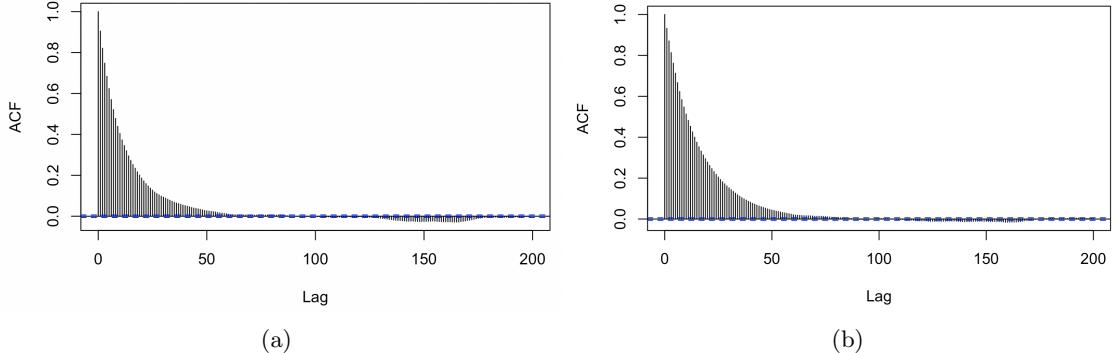


Figure 11: Autocorrelation plots for the MCMC output sequences targeting the η -smi posterior with $\eta = 1$ given subjective prior (Eq. 5): (a) ACF plot of $\{\pi^{(t)}\}_{t=1}^N$; (b) ACF plot of $\{q^{(t)}\}_{t=1}^N$.

Fig. 12 is the trace plot of MCMC samples of π, q , with $N = 1000000, \eta = 1$ and thinning interval of 100 under the uniform prior.

Fig. 13 is the scatter plot of (π, a) from η -smi posterior with $\eta = 1$ using MCMC and inversion method.

Table. 1 shows the effective sample size (ESS) for the related MCMC samples.

$\{\pi^{(i)}\}_{i=1}^{N=10000}$ subjective	$\{q^{(i)}\}_{i=1}^{N=10000}$ subjective	$\{\pi^{(i)}\}_{i=1}^{N=10000}$ uniform	$\{q^{(i)}\}_{i=1}^{N=10000}$ uniform
503.8	359.0	1193.2	1211.2

Table 1: ESS for the related MCMC samples

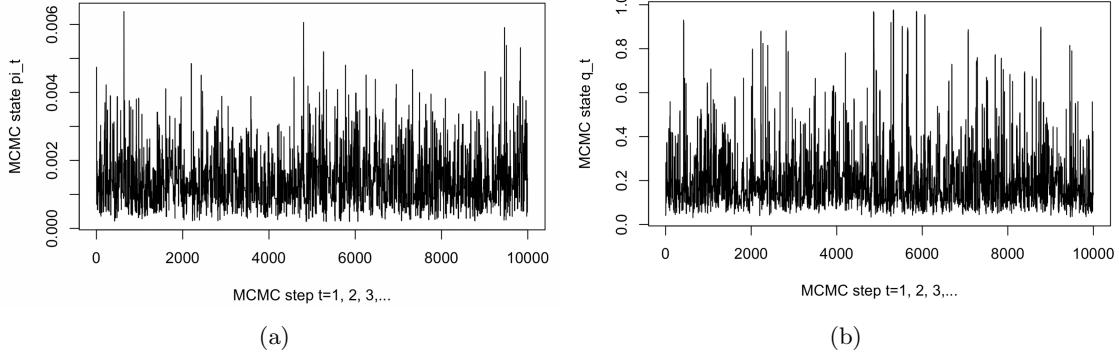


Figure 12: Nested MCMC targeting the η -smi posterior with $\eta = 1$ given uniform prior: (a) MCMC trace of π^t ; (b) MCMC trace of q^t .

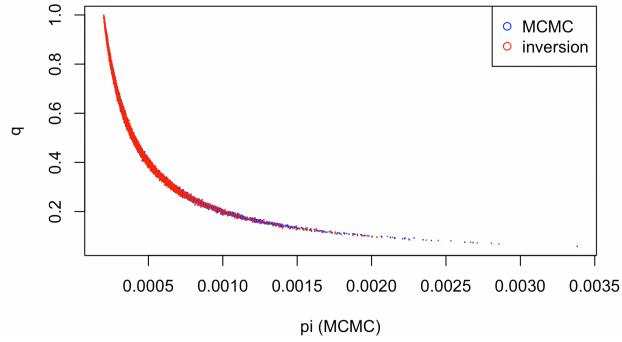


Figure 13: Scatter plot of (π, q) from η -smi posterior with $\eta = 1$ MCMC (blue) and inversion (red) superposed.

Fig. 14 is trace plot of MCMC samples of π, q , with $N_1 = 100000, \eta = 0.92$ (optimal case) with thinning interval of 100 under the subjective prior, i.e. $\pi \sim \text{Beta}(\alpha = 1.01, \beta = 201), q \sim \text{Beta}(\alpha = 4.5, \beta = 1.5)$.

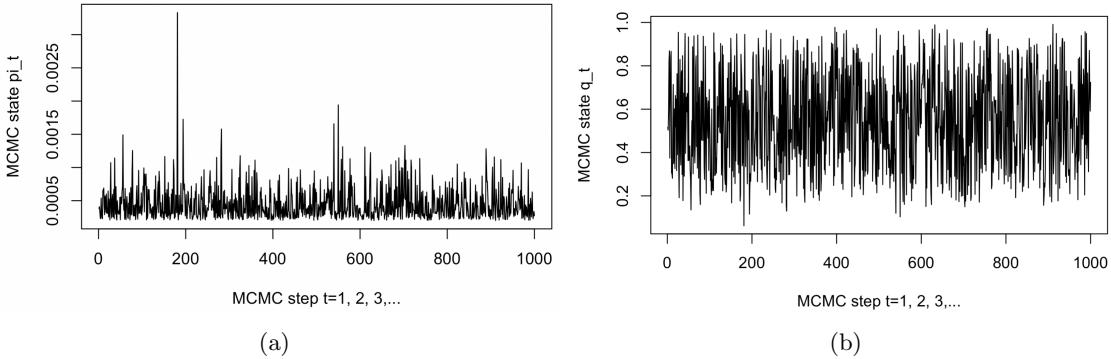


Figure 14: Nested MCMC targeting the η -smi posterior with $\eta = 0.92$ (optimal case) given subjective prior with thinning interval of 100 (Eq. 5): (a) MCMC trace of π^t ; (b) MCMC trace of q^t .

Fig. 15 is the trace plot of MCMC samples of π, q , with $N = 1000000, \eta = 0.926$ (optimal case) and thinning interval of 1000 under the uniform prior.

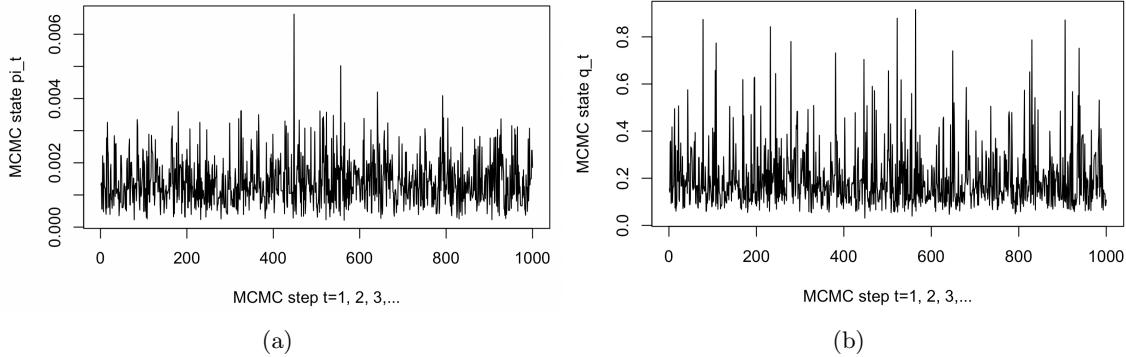


Figure 15: Nested MCMC targeting the η -smi posterior with $\eta = 0.92$ (optimal case) given uniform prior: (a) MCMC trace of π^t ; (b) MCMC trace of q^t .

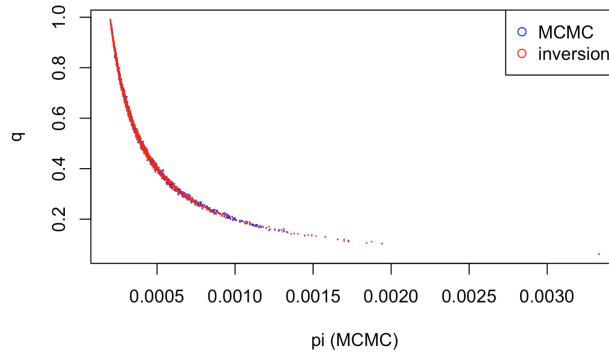


Figure 16: Scatter plot of (π, q) from η -smi posterior with $\hat{\eta} = 0.92$ MCMC (blue) and inversion (red) superposed.

Fig. 16 is the scatter plot of (π, q) from η -smi posterior with $\hat{\eta} = 0.92$ using MCMC and inversion method.

Fig. 17 is the trace plot of MCMC samples π, q given multiple weeks data. We first generate 100000 MCMC samples, and use a thinning interval of 10 after removing the burn-in samples. Fig. 18 is the corresponding autocorrelation plot

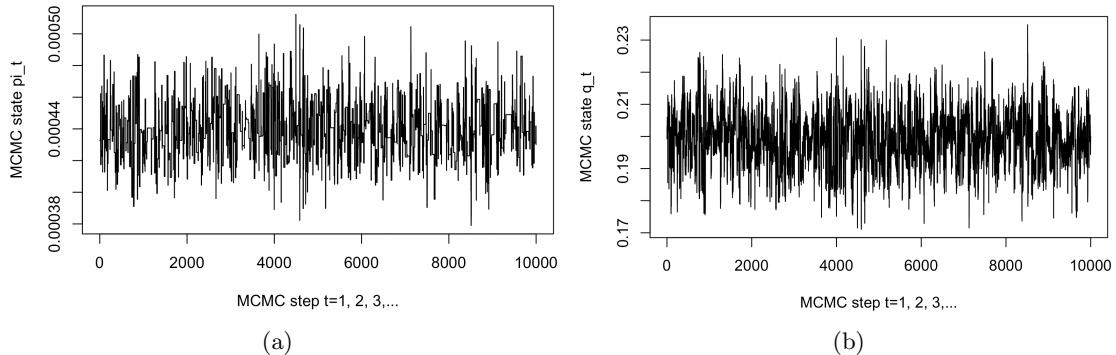


Figure 17: Nested MCMC targeting the η -smi posterior with $\eta = 1$ given multiple data under subjective prior: (a) MCMC trace of π^t ; (b) MCMC trace of q^t .

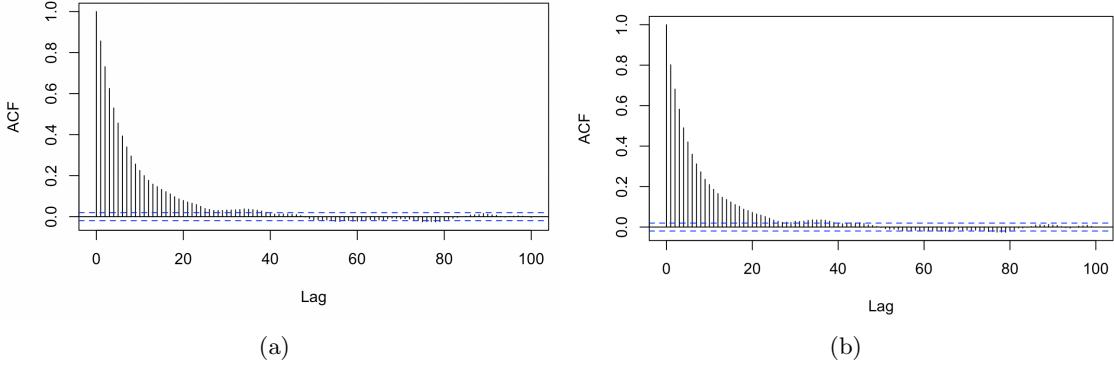


Figure 18: Autocorrelation plots for the MCMC output sequences targeting the η -smi posterior with $\eta = 1$ given multiple data under subjective prior : (a) ACF plot of $\{\pi^{(t)}\}_{t=1}^N$; (b) ACF plot of $\{q^{(t)}\}_{t=1}^N$.

Appendix C: Choosing the influence parameter η with elpd using synthetic data

In this appendix, we choose the influence parameter η with elpd using synthetic data following section 7.1, to test reliability of elpd method.

Here we introduce how we generate the synthetic data. M, U, N are identical to data from week 1 to 10 in `southeast_df`. We sample π, q from corresponding prior (Eq. 7, 8). Then, sample u according to Eq. 1, sample n using MCMC according to Eq. 2.

We estimate elpd using LOOCV. The result is shown in Fig. 19. η associated with the biggest elpd value is the optimal influence parameter we want to choose, which is approximately 1 in the figure. We repeatedly calculate elpd five times and they all give similar results. This is expected because there is no model misspecification, the full Bayes should give a good result.

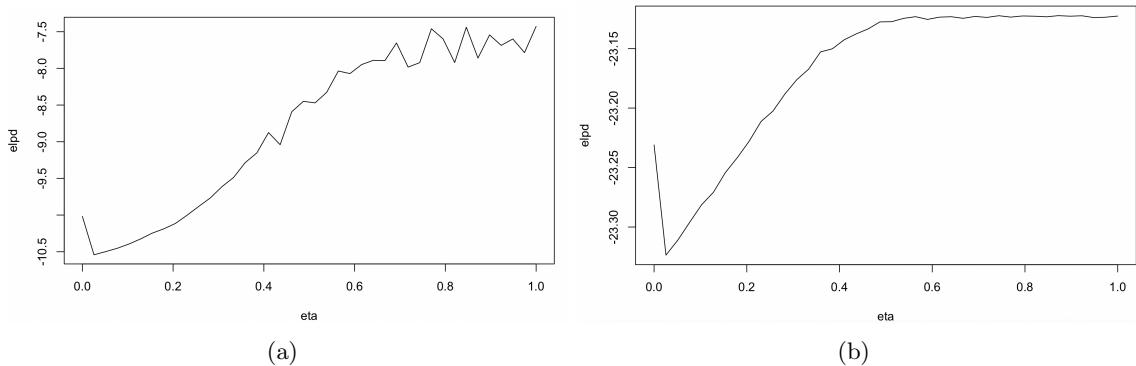


Figure 19: Estimated elpd with different η value using synthetic data: (a) LOOCV on n ; (b) LOOCV on u .

Appendix D: Data source

In Appendix D, we give `southeast_df` data we use in this paper. The data was first introduced by Nicholson et al. (2021a). `southeast_df` contains weekly Pillar 1+2 data and REACT study data for the whole of the South East, from the end of May 2020 to the beginning of August 2021. The columns **nt** and **Nt** contain the number of positive and total Pillar 1+2 tests respectively, while the columns **nr** and **Nr** contain the corresponding number of REACT tests. The column **M** contains the population of the relevant region. The data is displayed in Fig. 20.

Appendix E: Inversion method

In Appendix E, we sample from the η -smi posterior (Eq. 11) using inversion method.

Let F be the cdf of π from η -smi posterior

$$F(x) = \int_0^x \rho_\eta(\pi|u, n)d\pi, \quad (39)$$

and $U \sim \text{Unif}[0, 1]$. Then $X = F^{-}(U)$ has cdf F . After sampling π , we can sample q from $\rho(q|n, \pi)$ (refer to Eq. 18) to get marginal of q .

The implementation details are as follows. Let $L := \frac{n}{M}$ be the lower bound of π , $K = 1000$ be the number of points we evaluate to estimated cdf F . We evaluate $\rho_\eta(\pi|u, n)$ at

$$\pi_i = L + \delta^{-i}, \quad i = 1, \dots, K,$$

and the estimate cdf becomes

$$\hat{F}(\pi) = \sum_{j=1}^{i(\pi)} \rho_\eta(\pi_j|u, n)(\pi_{j+1} - \pi_j),$$

where $i(\pi) := \sum_{j=1}^K \mathbb{1}\{\pi \geq \pi_j\}$. If $\hat{F}(\pi_i) \leq u \leq \hat{F}(\pi_{i+1})$, then $\pi = \pi_i$ is the value we sample by inversion method.

The scatter plot of (π, q) inversion samples can be found in Appendix B.

	phe_region	mid_week	Nt	nt	Nr	nr	M
1	South East	2020-05-31	77406	1830	2011	3	9180135
2	South East	2020-06-07	81878	1311	0	0	9180135
3	South East	2020-06-14	68227	932	0	0	9180135
4	South East	2020-06-21	63759	746	8519	10	9180135
5	South East	2020-06-28	77542	671	21516	8	9180135
6	South East	2020-07-05	91351	622	4329	3	9180135
7	South East	2020-07-12	107026	641	0	0	9180135
8	South East	2020-07-19	116526	663	0	0	9180135
9	South East	2020-07-26	123565	622	13572	2	9180135
10	South East	2020-08-02	129301	483	17201	2	9180135
11	South East	2020-08-09	128856	547	5300	3	9180135
12	South East	2020-08-16	140184	680	0	0	9180135
13	South East	2020-08-23	146590	857	14647	5	9180135
14	South East	2020-08-30	145196	1058	14914	11	9180135
15	South East	2020-09-06	177048	1555	5253	4	9180135
16	South East	2020-09-13	177351	1122	0	0	9180135
17	South East	2020-09-20	175738	1576	14973	38	9180135
18	South East	2020-09-27	186278	2490	18996	45	9180135
19	South East	2020-10-04	213022	4940	5308	20	9180135
20	South East	2020-10-11	212521	6266	0	0	9180135
21	South East	2020-10-18	226363	8880	14891	64	9180135
22	South East	2020-10-25	220856	11113	15688	141	9180135
23	South East	2020-11-01	253012	13276	5425	39	9180135
24	South East	2020-11-08	266181	17285	0	0	9180135
25	South East	2020-11-15	273320	18228	11760	78	9180135
26	South East	2020-11-22	276566	14928	18774	121	9180135
27	South East	2020-11-29	269108	15598	7706	55	9180135
28	South East	2020-12-06	277473	19967	4	0	9180135
29	South East	2020-12-13	348807	35217	0	0	9180135
30	South East	2020-12-20	449015	50038	0	0	9180135
31	South East	2020-12-27	332988	62945	5	0	9180135
32	South East	2021-01-03	399472	69204	5160	72	9180135
33	South East	2021-01-10	399672	55450	26108	398	9180135
34	South East	2021-01-17	337712	43779	6834	100	9180135
35	South East	2021-01-24	294263	29775	659	3	9180135
36	South East	2021-01-31	290072	20070	0	0	9180135
37	South East	2021-02-07	273282	12865	13077	39	9180135
38	South East	2021-02-14	251086	9293	17910	60	9180135
39	South East	2021-02-21	248334	7315	5626	20	9180135
40	South East	2021-02-28	233811	4566	0	0	9180135
41	South East	2021-03-07	229659	3530	0	0	9180135
42	South East	2021-03-14	243857	3114	10737	5	9180135
43	South East	2021-03-21	254796	2969	15789	11	9180135
44	South East	2021-03-28	238442	2473	4585	2	9180135
45	South East	2021-04-04	210322	1795	0	0	9180135
46	South East	2021-04-11	232726	1623	0	0	9180135
47	South East	2021-04-18	237610	1296	9156	7	9180135
48	South East	2021-04-25	245216	1166	14543	14	9180135
49	South East	2021-05-02	228371	921	4553	5	9180135
50	South East	2021-05-09	239647	876	0	0	9180135
51	South East	2021-05-16	246471	918	0	0	9180135
52	South East	2021-05-23	264095	1353	6290	5	9180135
53	South East	2021-05-30	260898	2472	8524	10	9180135
54	South East	2021-06-06	290452	3534	4032	6	9180135
55	South East	2021-06-13	304583	4839	0	0	9180135
56	South East	2021-06-20	345273	6512	0	0	9180135
57	South East	2021-06-27	362866	13797	4573	15	9180135
58	South East	2021-07-04	404021	20842	9143	34	9180135
59	South East	2021-07-11	400513	29384	3687	12	9180135
60	South East	2021-07-18	413838	39568	0	0	9180135

Figure 20: southeast_df data