This is a peer-reviewed, accepted author manuscript of the following conference paper: Basu, T, Troffaes, MCM & Einbeck, J 2023, A robust Bayesian approach for causal inference problems. in Z Bouraoui & S Vesic (eds), Symbolic and Quantitative Approaches to Reasoning with Uncertainty: 17th European Conference, ECSQARU 2023, Arras, France, September 19–22, 2023, Proceedings. Lecture Notes in Artificial Intelligence, vol. 14294, Springer, Cham, pp. 359–371. https://doi.org/10.1007/978-3-031-45608-4_27

A Robust Bayesian Approach for Causal Inference Problems

Tathagata Basu[1][0000−0002−6851−154X], Matthias C. M. Troffaes[2][0000−0002−1294−600X], and Jochen Einbeck[2,3][0000−0002−9457−2020]

[1] Civil and Environmental Engineering, University of Strathclyde, UK
[2] Department of Mathematical Sciences, Durham University, UK
[3] Durham Research Methods Centre, UK

Abstract Causal inference concerns finding the treatment effect on subjects along with causal links between the variables and the outcome. However, the underlying heterogeneity between subjects makes the problem practically unsolvable. Additionally, we often need to find a subset of explanatory variables to understand the treatment effect. Currently, variable selection methods tend to maximise the predictive performance of the underlying model, and unfortunately, under limited data, the predictive performance is hard to assess, leading to harmful consequences. To address these issues, in this paper, we consider a robust Bayesian analysis which accounts for abstention in selecting explanatory variables in the high dimensional regression model. To achieve that, we consider a set of spike and slab priors through prior elicitation to obtain a set of posteriors for both the treatment and outcome model. We are specifically interested in the sensitivity of the treatment effect in high dimensional causal inference as well as identifying confounder variables. However, confounder selection can be deceptive in this setting, especially when a predictor is strongly associated with either the treatment or the outcome. To avoid that we apply a post-hoc selection scheme, attaining a smaller set of confounders as well as separate sets of variables which are only related to treatment or outcome model. Finally, we illustrate our method to show its applicability.

Keywords: high dimensional data · variable selection · Bayesian analysis · imprecise probability.

1 Introduction

In causal inference, we are interested in estimating the causal effect of independent variables on a dependent variable. Ideally, randomised trials are the most efficient way to perform this task. However, this is not always practical for several reasons; ethical concerns, design cost, population size, to name a few. This leaves us with observational studies which are usually obtained by means of collecting data though surveys or record keeping. But this can be problematic in the presence of confounders, which are variables associated with both the treatment and the outcome. In such cases, we need to be extra cautious as otherwise it will lead to unwanted bias in the treatment effect estimator [1]. Several works have been done in order to tackle the presence of confounder variables. One such work in the topic was by Robins [2] where the author used a graphical approach for the identification of the causal parameters. Rosenbaum and Robin [3] suggested the use of a link model to estimate the propensity scores for all individuals. Later on several other methods have been proposed based on propensity score matching. A brief review on such methods can be found in [4, 5].

The Bayesian approach in causal effect estimation is a popular strategy in the field and one of the earlier works on this can be found in [6]. Lately, with the rise of high dimensional data,

Bayesian methodologies have become more appealing. Crainiceanu et al. [7] proposed a bi-level Bayesian model averaging based method for estimating the causal effect. Wang et al. [8] suggested BAC (or, Bayesian adjustment for confounding) where they use an informative prior obtained from the treatment model and apply them on the outcome model for estimating causal effect. Several other methods were also proposed to tackle confounders from the point of view of Bayesian variable selection, see for instance [9, 10] among others.

In this paper we take inspiration from the approach of Koch et al. [11], who proposed a bi-level spike and slab prior for causal effect estimation. They considered a data-driven adaptive approach to propose their prior which reduces the variance of the causal estimate. In our approach, we perform a sensitivity analysis based approach where instead of using a single prior, we consider a set of priors [12]. This is particularly interesting as in many cases, causal effect estimation can be performed through a meta analysis and hence robust Bayesian analysis [13] can be beneficial under severe uncertainty. Moreover, for some problems we have to rely on very limited data to perform our Bayesian analysis and inference may not be reliable in presence of heteroscedasticity within the data. Instead, we use expert opinion and elicit a set of priors based on empirical evidence. This also allows us to construct the problem of confounder identification in a framework where abstention has a relatively positive gain i.e. when the cost of further tests/data collection is cheaper than mistreating a subject. To propose our framework, we consider a set of continuous spike and slab priors [14] for confounder identification and construct a Bayesian group LASSO [15] type problem. To perform the prior sensitivity analysis, we consider a set of beta priors on the covariate selection probability of the spike and slab priors. We use the posteriors of this covariate selection probability for identifying the confounders. Finally, we consider a post-hoc coefficient adjustment method [16] to recover sparse estimates associated with either the outcome or the treatment model.

The rest of the paper is organised as follows. In Section 2 we give a formal description of the causal estimation problem in the context of linear regression. Section 3 is focused on the Bayesian analysis of causal inference problems, followed by the motivation of a robust Bayesian analysis along with our proposed decision theoretic framework for confounder (variable) selection. In Section 4, we provide results of simulation studies under different scenarios and show the possible applications in real life problems. Finally, we discuss our findings and conclude this paper in Section 5.

2 Causal Estimation

Let an observational study give us the outcomes $Y = (Y_1, \ldots, Y_n)$ along with corresponding treatment indicators $T = (T_1, \ldots, T_n)$. Then the treatment effect in the population is given by the expectation of the difference in outcomes between the treatment and controls:

$$\delta = \mathbb{E}(Y \mid T = 1) - \mathbb{E}(Y \mid T = 0). \tag{1}$$

Similarly, the individual causal effect of treatment T_i on outcome Y_i is given by:

$$\delta_i \coloneqq \mathbb{E}(Y_i \mid T_i = 1) - \mathbb{E}(Y_i \mid T_i = 0). \tag{2}$$

That is, we are interested in the difference between the outcomes when the i-th subject receives the treatment and when it remains as a control.

In theory, both of these quantities exist. However, we cannot observe $\mathbb{E}(Y_i \mid T_i = 1)$ and $\mathbb{E}(Y_i \mid T_i = 0)$ the average causal effect of the treatment T by calculating the averaged outcome of

all the subjects that received the treatment and all the subjects that remained as control:

$$\hat{\delta} \coloneqq \frac{\sum_{i=1}^{n} Y_i \cdot \mathbb{I}(T_i = 1) - \sum_{i=1}^{n} Y_i \cdot \mathbb{I}(T_i = 0)}{n}. \tag{3}$$

However, this relies on an important assumption that the treatment effect on the i-th subject given that they received the treatment is the same as the (counterfactual) treatment effect when they remain as control [4].

2.1 Regression Model

Regression methods are widely used in causal effect estimation. The main idea behind these regression methods is to remove the correlation between the treatment indicator and the error term [4, 17]. To do so, we rely on p different observed quantities or predictors denoted by $X \coloneqq [X_1^T, \ldots, X_n^T]^T$ where each $X_i \in \mathbb{R}^p$. Each X_i is treated as a p-dimensional row vector, so X is a $n \times p$ matrix. Now, let $\beta \coloneqq (\beta_1, \ldots, \beta_p)^T$ denote the vector of regression coefficients related to the predictors, and β_T denote a regression coefficient related to the treatment. Then we can define a linear model for the outcome so that

$$Y_i = T_i\beta_T + X_i\beta + \epsilon_i \tag{4}$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$. Clearly, when the underlying true outcome model is linear with respect to the treatment,

$$\delta_i = \mathbb{E}(Y_i \mid T_i = 1) - \mathbb{E}(Y_i \mid T_i = 0) = \beta_T. \tag{5}$$

In the presence of confounders we also need to consider the association between the treatment indicators and the predictors. In literature, authors often suggest a probit link function to construct the regression model. This way, we can specify the conditional probability that subject i receives the treatment through a linear model. That is, for another vector of regression coefficients $\gamma \coloneqq (\gamma_1, \cdots, \gamma_p)^T$ we define

$$P(T_i = 1 \mid X_i) = \Phi(X_i\gamma) \tag{6}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of a standard normal distribution. To incorporate this probit link function, we assume that we can model the T_i through the following [18]:

$$T_i^* = X_i\gamma + u_i \tag{7}$$

$$T_i = \mathbb{I}(T_i^* > 0) = \begin{cases} 1 & \text{if } T_i^* > 0 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where $u_i \sim \mathcal{N}(0, 1)$.

Now, to construct the joint likelihood function, we define an extended output $2n \times 1$ column vector $W \coloneqq \binom{Y}{T^*}$ and corresponding $2n \times (2p + 1)$ dimensional design matrix

$$Z \coloneqq \begin{bmatrix} T_1 & X_1 & 0 \\ \vdots & \vdots & 0 \\ T_n & X_n & 0 \\ 0 & 0 & X_1 \\ \vdots & \vdots & \vdots \\ 0 & 0 & X_n \end{bmatrix} = \begin{bmatrix} X_O & 0 \\ 0 & X_T \end{bmatrix} \tag{9}$$

where, $X_O = [T, X]$ and $X_T = X$. Then, considering the assumption of Gaussian error terms, we have the following likelihood distribution

$$W \mid Z, \beta_T, \beta, \gamma, \sigma^2 \sim \mathcal{N}\left(Z\nu, \Sigma\right), \tag{10}$$

where $\nu = (\beta_T, \beta^T, \gamma^T)^T$ and

$$\Sigma = \begin{bmatrix} \sigma^2 I_n & 0 \\ 0 & I_n \end{bmatrix}. \tag{11}$$

3 Bayesian Causal Estimation

The likelihood given by Eq. (10) gives us a foundation for a Bayesian group LASSO [15] type model. This way, we can look into the posterior selection probability associated with the j-th predictor. There are several ways to construct spike and slab priors which achieve variable selection. In our case, we consider a continuous type [14] prior for faster posterior computation.

3.1 Hierarchical model

Let π_j denote the prior probability that the j-th predictor is associated to the outcome or the treatment. That is,

$$\pi_j = P\left((\beta_j, \gamma_j) \neq (0, 0)\right). \tag{12}$$

Then we can define the following hierarchical model for spike and slab group LASSO so that, for $1 \leq j \leq p$,

$$(\beta_j, \gamma_j)^T \mid \pi_j, \sigma^2 \sim \pi_j \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tau_1^2 \begin{bmatrix} \sigma^2 & 0 \\ 0 & 1 \end{bmatrix}\right) + (1 - \pi_j)\mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tau_0^2 \begin{bmatrix} \sigma^2 & 0 \\ 0 & 1 \end{bmatrix}\right) \tag{13}$$

$$\beta_T \mid \sigma^2 \sim \mathcal{N}\left(0, \sigma^2\right) \tag{14}$$

$$\frac{1}{\sigma^2} \sim \text{Gamma}(a, b) \tag{15}$$

$$\pi_j \sim \text{Beta}\left(sq, s(1-q)\right). \tag{16}$$

In the hierarchical model, we fix sufficiently small τ_0 ($1 \gg \tau_0 > 0$) so that (β_j, γ_j) has its probability mass concentrated around zero. Therefore, this represents the spike component of our prior specification. For the slab component, we consider τ_1 to be large so that $\tau_1 \geq 1$. This allows the prior for (β_j, γ_j) to be flat, besides the spike component at the origin.. We illustrate the components of a bivariate spike and slab prior in Fig. 1 (with fixed $\sigma = 1$). We generate the spike component with $\tau_0 = 0.001$ and the slab component with $\tau_1 = 5$.

For the precision term $1/\sigma^2$, a natural choice of prior is the gamma distribution as it allows the control of both the location and the scale of the precision. To ensure that the prior is able to represent the data, we consider $b = 1$ and fix a so that it represents the prior mean of the precision. In cases where we have no prior information, we can simply consider a large value for a so that the interval $[0, 2a]$ contains the true value of the precision. As defined earlier, π_j is used as the selection probability of the j-th predictor in either of the models and we use a beta prior to specify these selection probabilities where q_j represents our prior expectation of the selection probability (π_j) and

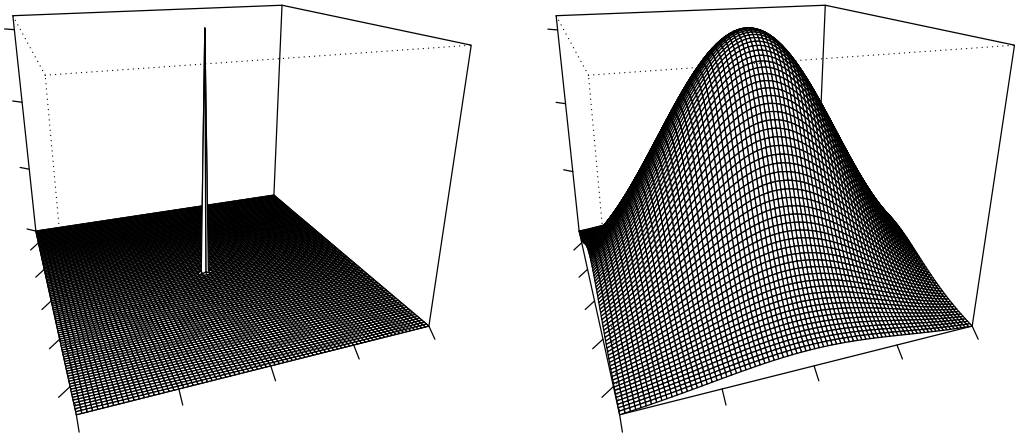

Figure 1: Spike (left) and slab (right) components of a bi-variate distribution for $\tau_0 = 0.001$, $\tau_1 = 5$ and $\sigma = 1$.

s acts as a concentration parameter. For the causal effect, we want to use a Gaussian distribution that matches the scale of the noise term. Therefore, we consider $\beta_T \sim \mathcal{N}(0, \sigma^2)$.

In Fig. 2, we show a probabilistic graphical representation of our hierarchical model. In the figure, grey circular nodes represent the prior hyper-parameters which will be used for sensitivity analysis of the model. The transparent circular nodes are used to denote the modelling parameters which are our quantities of interest. The observed quantities are denoted with transparent rectangular nodes. We also use a grey rectangular node to denote the intermediate latent variable T^*. We use directed edges to denote the relationship between different nodes. However, we use a dashed edge between X and T as they are related through the latent variable T^*.

3.2 Robust Bayesian Analysis

The hierarchical model presented above is a standard spike and slab model for variable selection and performs well when we have sufficient data to begin with. However, especially in the case of causal inference having sufficient data may not be feasible. Moreover, we also need to be cautious about the side effects of a treatment. Therefore, we are particularly interested in constructing a robust Bayesian framework for variable selection. This way, when we are preparing a guideline for treatment, we can have the option to ask for more data before reaching any conclusion. To achieve this, we consider a utility based framework with three possible ways of determining a variable.

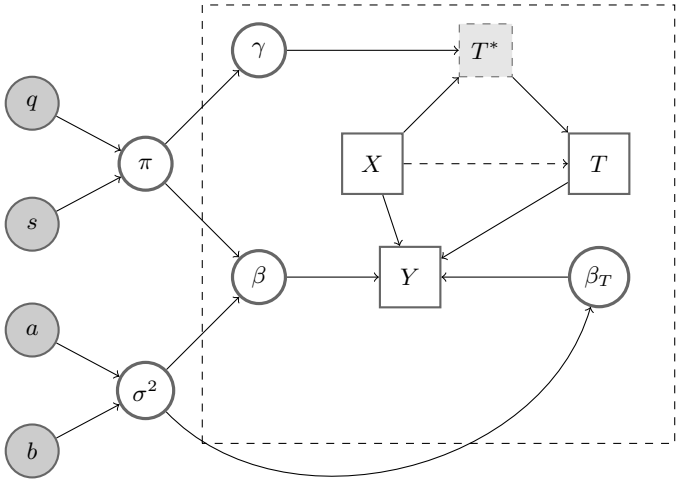

Figure 2: Probabilistic graphical representation for causal inference with Bayesian hierarchical model.

In general, an unsuccessful treatment of a subject can have severe consequences which cannot be associated with a suitable loss function. Instead, we assume that we can always revert any initial mistreatment by further treatments, and we can associate a loss function with the cost of further treatments. This way, in the simplest case, we can associate two constant loss values ℓ_1, ℓ_2 with false positives and false negatives respectively. Clearly, false positives will lead to unwanted side effects and false negatives will lead to mistreatment of the patient. Finally, we associate a loss value ℓ_3 for abstention which can be interpreted as the cost of further tests. Ideally, in most cases, $\ell_3 \ll \ell_1, \ell_2$. However, in certain scenarios, this might not be the case, especially when the condition of a subject deteriorates rapidly over time.

Now, based on this notion of abstaining from selecting a variable, we can perform a sensitivity analysis over a set of priors on the prior selection probability. That is, we can consider a set of possible values for q such that $q \in \mathcal{P}$, where $\mathcal{P} \subseteq (0,1)^p$. Here, the equality occurs for the near vacuous case. However, in real-life situations, performing a robust Bayesian analysis for the near vacuous case is not practical. Instead, we incorporate expert elicitation to define our model. For instance, we can consider $q \in \left[\underline{q}, \overline{q}\right]$ where $p\underline{q}$ and $p\overline{q}$ represent the bounds of the prior expectation on the total number of variables present in either of the models.

3.3 Variable selection and coefficient adjustment

For the co-variate selection, we look into the posterior expectation of π_j. We consider the j-th predictor to be removed from both the treatment and outcome model, if

$$\overline{\mathrm{E}}(\pi_j \mid W) \coloneqq \sup_{q \in \mathcal{P}} \mathbb{E}(\pi_j \mid W) < 1/2. \tag{17}$$

Similarly, we consider the j-th predictor to be present in at least one of the models, if

$$\underline{\mathrm{E}}(\pi_j \mid W) \coloneqq \inf_{q \in \mathcal{P}} \mathbb{E}(\pi_j \mid W) \geq 1/2. \tag{18}$$

Otherwise, we consider the variable to be indeterminate, in which case we abstain from putting it in any of the models but instead just report a lack of information.

In general, this framework is self sufficient for variable selection. However, for model fitting and prediction, we need to evaluate the values of the regression coefficients. For that we first need to find the set of active predictors with respect to our prior expectation of the selection probability q. For any fixed q, we define the set $S(q)$ as the set of all variables which are active in the treatment model or in the outcome model:

$$S(q) \coloneqq \{j \colon E(\pi_j \mid W) \geq 1/2\}\,. \tag{19}$$

For sensitivity analysis, the intersection of $S(q)$ over all q gives us the set of active variables obtained through Eq. (18). Similarly, the union gives us the set of variables that are not removed through Eq. (17). That is:

$$\mathcal{S}_* \coloneqq \{j : \underline{\mathrm{E}}(\pi_j \mid W) \geq 1/2\} = \bigcap_{q \in \mathcal{P}} S(q), \qquad \mathcal{S}^* \coloneqq \left\{j : \overline{\mathrm{E}}(\pi_j \mid W) \geq 1/2\right\} = \bigcup_{q \in \mathcal{P}} S(q). \tag{20}$$

Clearly, $\mathcal{S}_* \subseteq \mathcal{S}^*$. $\mathcal{S}_*$ represents the set of variables that are sure to be selected, $\{1, \ldots, p\} \setminus \mathcal{S}^*$ represents the set of variables that are sure to be removed, and $\mathcal{S}^* \setminus \mathcal{S}_*$ represents the set of variables about which we are undecided. In this way, through sensitivity analysis, our approach incorporates robustness.

Now, for each fixed value of q, let $\hat{\beta}_{S(q)}$ be the posterior means of the regression coefficients of the outcome model with respect to the predictors that belong to $S(q)$. Similarly, $\hat{\gamma}_{S(q)}$ be the posterior means of the regression coefficients for the treatment effects. Since we use continuous spike and slab priors, these regression coefficients are not sparse. Moreover, with our variable selection we only determine whether the variable is included in at least one of the models. But, we cannot determine a specific association. Therefore, to adjust the sparsity of the estimates and understand the specific association with the treatment/outcome/both, we apply the "decoupled shrinkage and selection" method proposed by [16]. For that, we solve the following adaptive LASSO-type [19] problems

$$\hat{\beta}^D_{S(q)} = \arg\min_{\beta_{S(q)}} \frac{1}{n} \|X_{S(q)}\hat{\beta}_{S(q)} - X_{S(q)}\beta_{S(q)}\|_2^2 + \lambda \sum_{j \in S(q)} \frac{|\beta_{j,S(q)}|}{|\hat{\beta}_{j,S(q)}|} \tag{21}$$

and

$$\hat{\gamma}^D_{S(q)} = \arg\min_{\gamma_{S(q)}} \frac{1}{n} \|X_{S(q)}\hat{\gamma}_{S(q)} - X_{S(q)}\gamma_{S(q)}\|_2^2 + \lambda \sum_{j \in S(q)} \frac{|\gamma_{j,S(q)}|}{|\hat{\gamma}_{j,S(q)}|} \tag{22}$$

where $q \in \mathcal{P}$.

4 Simulation Studies

For the simulation studies, we consider 2 different settings. In each case, we generate the design matrix X such that $X_i \sim \mathcal{N}(0, \Sigma)$ for $1 \leq i \leq n$ where $[\Sigma]_{ij} = 0.3^{|i-j|}$. This way, we generate 50

predictors for our model with mild correlations among them. We then use the following generation schemes to generate the outcome and treatment indicator:

$$T_i \sim \text{Bernoulli}\left(1/(1+\exp(-X_i\gamma))\right) \quad \text{and} \quad Y_i = 4T_i + X_i\beta. \tag{23}$$

Scenario 1 — $|\gamma_j|, |\beta_j| > 0$ for $j \leq 10$
Scenario 2 — $|\gamma_j| > 0$ for $j \leq 10$ and $|\beta_j| > 0$ for $j \leq 15$

For both the cases, we consider different numbers of observations n where $n = 25 + 5k$ for $k = 0, 1, 2, \cdots, 10$.

We present our analyses in Table 1 and Table 2. For the sake of clarity we use the following accronyms: RBCE for robust Bayesian causal estimation (our method); SSCE for spike and slab causal estimation [11]; BSSCE for bi-level spike and slab causal estimation [11]; and BSSL for Bayesian spike and slab lasso [15]. As it can be seen from both the tables, SSCE and BSSCE are formulated for problems where $p \leq n$ and therefore we do not have any results for $n < 50$.

Elicitation For the elicitation of $\mathcal{P}$, we use marginal correlation between Y and X to determine the bounds on number of active variables. We set the thresholds to be 0.15 and 0.35 for the correlations. We compute the number of variables with marginal correlation greater than 0.15 (say p_1) and number of variables with marginal correlation greater than 0.35 (say p_2). We use these numbers to obtain the bounds on the number of active variables so that $\mathcal{P} = [p_2/p, p_1/p]$.

Initialisation To implement our method, we use `rjags` and for the other three methods we use the code provided in the appendix of [11]. For our method, we set $\tau_0 = 10^{-6}$ and $\tau_1 = 1$ to construct the spike and slab prior. For the noise term, we set $a = 10$ and $b = 1$. To perform our Bayesian analysis with `rjags`, we first consider an adaptive stage with 2000 iterations followed by discarding of 2000 burn in samples to refine the posteriors. We consider 5000 MCMC samples to compute the posterior estimates. For the other methods we use the in-built settings to initiate the analyses.

Results We provide our result for causal estimate in Table 1. As we perform a sensitivity analysis, our method gives an interval estimate for the causal effect and we show that in two different rows where the first row gives the lower bound and the second row gives the upper bound. We notice that our method is somewhat in agreement with the other methods but much more consistent in terms of estimating the treatment effect. However, this is not the case for other methods and sometimes those methods produce extreme values. This can be observed in Fig. 3 as well. Here, the true value is represented by the straight line for $\beta_T = 4$.

From the figure, we can notice that our method tends to underestimate the causal effect. This suggests that we may want to have a different value of a for these sets of observations instead of a fixed value of $a = 10$ for all of our analyses. We can also see that the lower bound tends to improve with increasing number of observations which validates the assumption that as we accumulate more information, the interval becomes smaller and converges towards the true value.

For the variable identification, we use the notion of different losses as described earlier. We consider $\ell_1 = \ell_2 = 1$ and $\ell_3 = 0.2$. This is a simplified way of choosing the loss function, we can choose more sophisticated loss functions based on [20]. We use this associated loss to obtain the total loss, which we present in Table 2. In the table we denote the misspecification by counting the number of false positives (FP) and false negatives (FN). For RBCE, we have an additional column 'ID' which denotes the number of variables which remain as indeterminate. From the table it can

Table 1: Causal estimates obtained from different methods for 6 different numbers of observations.

First scenario: $|\gamma_j|, |\beta_j| > 0$ for $j \leq 10$

	25	30	35	40	45	50	55	60	65	70	75
RBCE (low)	3.22	3.54	3.30	3.66	3.77	3.80	3.85	3.89	3.90	3.91	3.89
RBCE (up)	4.03	3.96	3.50	3.77	3.82	3.83	3.90	3.93	3.92	3.92	3.91
SSCE	–	–	–	–	–	4.24	4.11	3.99	4.00	4.00	3.99
BSSCE	–	–	–	–	–	4.02	4.01	4.01	4.01	4.01	4.01
BSSL	-0.23	4.07	6.80	4.05	4.00	4.00	4.01	3.98	3.99	3.99	3.99

Second scenario: $|\gamma_j| > 0$ for $j \leq 10$ and $|\beta_j| > 0$ for $j \leq 15$

	25	30	35	40	45	50	55	60	65	70	75
RBCE (low)	2.79	3.70	3.77	3.56	3.69	3.70	3.81	3.78	3.81	3.82	3.85
RBCE (up)	3.65	4.01	3.96	3.82	3.90	3.86	3.92	3.89	3.91	3.88	3.91
SSCE	–	–	–	–	–	4.80	4.05	4.06	6.02	4.04	4.04
BSSCE	–	–	–	–	–	10.34	8.12	4.17	4.04	4.06	4.05
BSSL	-6.68	3.62	4.06	4.07	4.06	4.02	4.05	4.07	4.03	4.05	4.04

be seen that for the first scenario, our method abstains from identifying some variables for $n < 50$. Especially for $n = 25$, our method identifies 26 and 23 variables as indeterminate for the first setting and second setting respectively. However, later on our method gives more precise results in terms of variable selection. We also notice that BSSL tends to perform poorly in terms of variable selection for $n = 25$, this can be seen from the treatment effect estimation as well. Moreover, we observe that for the second setting both SSCE and BSSCE underperform in identifying the active variables, which can be explained from Table 1 as well.

5 Conclusion

Causal effect estimation is an important tool in statistical learning and needs to be performed with utmost care as in many cases we may have severe consequence of poor estimation. In this paper, we tackle this issue by proposing a robust Bayesian analysis of causal effect estimation problem for high dimensional data. Our framework is focused on the effect of prior elicitation on confounder selection as well as causal effect estimation. We consider a spike and slab type prior for confounder selection and discuss the possible sources of uncertainty that need to be tackled carefully. We were particularly focused on the uncertainty associated with prior selection probabilities for which we consider a set of beta priors to perform sensitivity analysis. We showed that the sensitivity analysis on the prior selection probability gives us a robust confounder selection scheme. In this way, we can abstain from selecting a confounder when the available data is not sufficient. We also propose a generalised utility based framework, where we associate a loss for abstaining which can be interpreted as the cost of further data collection. Finally, we illustrate our method with synthetic dataset and compare with other state of the art Bayesian methods.

Currently, the paper proposes a robust Bayesian approach for causal effect estimation where we rely on sampling strategies to obtain the posterior bounds as well as performing variable selection. In future, it will be interesting to derive inner approximation bounds for the posterior estimates to

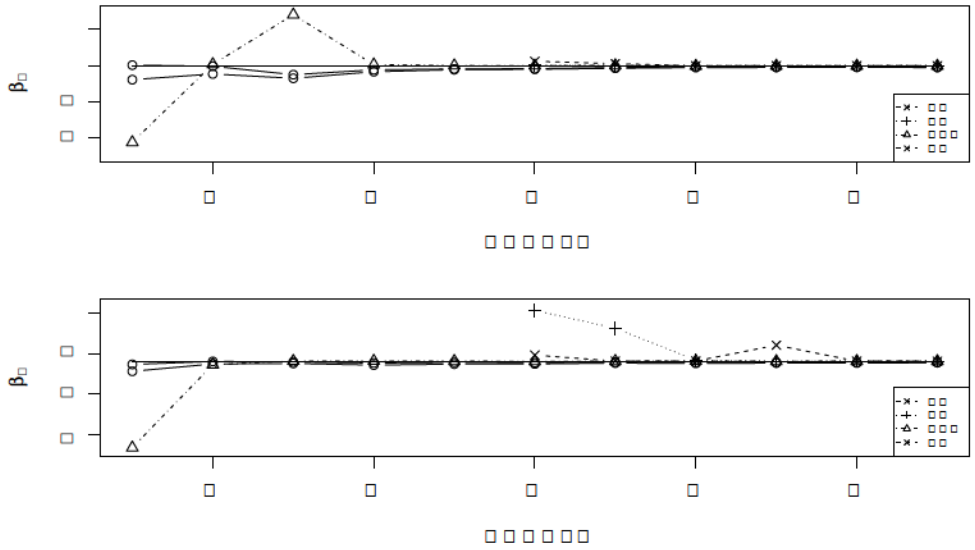

Figure 3: Comparison of different methods in estimating the treatment effect.

reduce the computational cost. Moreover, for the sake of illustration, we rely on simple loss functions and elicitation strategy. In future, we would like to investigate different elicitation strategies for the method and explore alternative loss functions for formulating a decision theoretic framework. Last but not the least, we noticed that our method is in good agreement with other methods with an added level of robustness. This confirms that our method has good potential for real-life problems, and we intend to apply it on a real dataset in future work.

References

[1] Rosenbaum, P.R., Rubin, D.B.: The central role of the propensity score in observational studies for causal effects. Biometrika **70**(1) (04 1983) 41–55

[2] Robins, J.M.: A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. Mathematical Modelling **7** (1986) 1393–1512

[3] Rosenbaum, P.R., Rubin, D.B.: Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. The American Statistician **39**(1) (1985) 33–38

[4] Winship, C., Morgan, S.L.: The estimation of causal effects from observational data. Annual Review of Sociology **25**(1) (1999) 659–706

[5] Stuart, E.A.: Matching Methods for Causal Inference: A Review and a Look Forward. Statistical Science **25**(1) (2010) 1 – 21

[6] Rubin, D.B.: Bayesian Inference for Causal Effects: The Role of Randomization. The Annals of Statistics **6**(1) (1978) 34 – 58

[7] Crainiceanu, C.M., Dominici, F., Parmigiani, G.: Adjustment Uncertainty in Effect Estimation. Biometrika **95**(3) (2008) 635–651

Table 2: Loss based on misspecification of active variables in different models.

First scenario: $|\gamma_j|, |\beta_j| > 0$ for $j \leq 10$

	RBCE				SSCE			BSSCE			BSSL		
Samples	FP	FN	ID	Tot	FP	FN	Tot	FP	FN	Tot	FP	FN	Tot
25	0	0	26	5.2	–	–	–	–	–	–	12	2	14
30	0	1	1	1.2	–	–	–	–	–	–	0	0	0
35	0	0	1	0.2	–	–	–	–	–	–	0	0	0
40	0	1	0	1.0	–	–	–	–	–	–	0	0	0
45	0	0	1	0.2	–	–	–	–	–	–	0	0	0
50	0	0	0	0.0	0	0	0	0	0	0	0	0	0
55	0	0	0	0.0	0	0	0	0	0	0	0	0	0
60	0	0	0	0.0	0	0	0	0	0	0	0	0	0
65	0	0	0	0.0	0	0	0	0	0	0	0	0	0
70	0	0	0	0.0	0	0	0	0	0	0	0	0	0
75	0	0	0	0.0	0	0	0	0	0	0	0	0	0

Second scenario: $|\gamma_j| > 0$ for $j \leq 10$ and $|\beta_j| > 0$ for $j \leq 15$

	RBCE				SSCE			BSSCE			BSSL		
Samples	FP	FN	ID	Tot	FP	FN	Tot	FP	FN	Tot	FP	FN	Tot
25	0	2	23	6.6	–	–	–	–	–	–	9	4	13
30	0	2	9	3.8	–	–	–	–	–	–	0	0	0
35	0	0	18	3.6	–	–	–	–	–	–	0	0	0
40	0	0	5	1.0	–	–	–	–	–	–	0	0	0
45	0	0	3	0.6	–	–	–	–	–	–	0	0	0
50	0	0	1	0.2	0	7	7	0	14	14	0	0	0
55	0	0	1	0.2	0	0	0	0	12	12	0	0	0
60	0	0	1	0.2	1	0	1	0	0	0	0	0	0
65	0	0	0	0.0	0	12	12	0	0	0	0	0	0
70	0	0	0	0.0	0	0	0	0	0	0	0	0	0
75	0	0	0	0.0	0	0	0	0	0	0	0	0	0

[8] Wang, C., Dominici, F., Parmigiani, G., Zigler, C.M.: Accounting for uncertainty in confounder and effect modifier selection when estimating average causal effects in generalized linear models. Biometrics **71**(3) (2015) 654–665

[9] Zigler, C.M., Dominici, F.: Uncertainty in Propensity Score Estimation: Bayesian Methods for Variable Selection and Model-Averaged Causal Effects. Journal of the American Statistical Association **109**(505) (2014) 95–107

[10] Hahn, P.R., Carvalho, C.M., Puelz, D., He, J.: Regularization and Confounding in Linear Regression for Treatment Effect Estimation. Bayesian Analysis **13**(1) (2018) 163 – 182

[11] Koch, B., Vock, D.M., Wolfson, J., Vock, L.B.: Variable selection and estimation in causal inference using Bayesian spike and slab priors. Statistical Methods in Medical Research **29**(9) (2020) 2445–2469

[12] Berger, J.O.: Robust Bayesian analysis: sensitivity to the prior. Journal of Statistical Planning and Inference **25**(3) (1990) 303 – 328

[13] Raices Cruz, I., Troffaes, M.C.M., Lindström, J., Sahlin, U.: A robust Bayesian bias-adjusted random effects model for consideration of uncertainty about bias terms in evidence synthesis. Statistics in Medicine **41**(17) (2022) 3365–3379

[14] Ishwaran, H., Rao, J.S.: Spike and slab variable selection: Frequentist and Bayesian strategies. Ann. Statist. **33**(2) (04 2005) 730–773

[15] Xu, X., Ghosh, M.: Bayesian Variable Selection and Estimation for Group Lasso. Bayesian Analysis **10**(4) (2015) 909 – 936

[16] Hahn, P.R., Carvalho, C.M.: Decoupling Shrinkage and Selection in Bayesian Linear Models: A Posterior Summary Perspective. Journal of the American Statistical Association **110**(509) (2015) 435–448

[17] Heckman, J.J., Robb, R.: Alternative methods for evaluating the impact of interventions: An overview. Journal of Econometrics **30**(1) (1985) 239–267

[18] Albert, J.H., Chib, S.: Bayesian Analysis of Binary and Polychotomous Response Data. Journal of the American Statistical Association **88**(422) (1993) 669–679

[19] Zou, H.: The Adaptive Lasso and Its Oracle Properties. Journal of the American Statistical Association **101**(476) (2006) 1418–1429

[20] Zaffalon, M., Corani, G., Mauá, D.: Evaluating credal classifiers by utility-discounted predictive accuracy. International Journal of Approximate Reasoning **53**(8) (2012) 1282–1301 Imprecise Probability: Theories and Applications (ISIPTA'11).