

## Confounder Selection via Penalized Credible Regions

Ander Wilson\* and Brian J. Reich\*\*

Department of Statistics, North Carolina State University, Raleigh, North Carolina, 27695, U.S.A.

\**email*: ander\_wilson@ncsu.edu

\*\**email*: brian\_reich@ncsu.edu

**SUMMARY:** When estimating the effect of an exposure or treatment on an outcome it is important to select the proper subset of confounding variables to include in the model. Including too many covariates increases mean square error on the effect of interest while not including confounding variables biases the exposure effect estimate. We propose a decision-theoretic approach to confounder selection and effect estimation. We first estimate the full standard Bayesian regression model and then post-process the posterior distribution with a loss function that penalizes models omitting important confounders. Our method can be fit easily with existing software and in many situations without the use of Markov chain Monte Carlo methods, resulting in computation on the order of the least squares solution. We prove that the proposed estimator has attractive asymptotic properties. In a simulation study we show that our method outperforms existing methods. We demonstrate our method by estimating the effect of fine particulate matter (PM<sub>2.5</sub>) exposure on birth weight in Mecklenburg County, North Carolina.

**KEY WORDS:** Adjustment uncertainty; Confounder selection; Consistency; Credible region; Variable selection.

## 1. Introduction

In many applications, the effect of an exposure or treatment  $X$  on an outcome  $Y$  is estimated while controlling for other explanatory or confounding variables. These additional covariates are often selected from a potentially large number of observed variables  $U$ , and the estimated exposure effect can be sensitive to the set of covariates included. In such cases, omitting confounding variables, those correlated with both  $X$  and  $Y$ , can bias the estimate of the exposure effect. On the other hand, including variables that are correlated with neither  $X$  nor  $Y$  increases the variance of the exposure effect. Estimating the exposure effect is particularly difficult when the number of potential confounders is large relative to the sample size. However, accounting for all confounding variables while minimizing the number of variables uncorrelated with the outcome that are included in the model is essential for inference on the exposure effect.

There are numerous variable selection methods that can be used to select a model that balances model fit and parsimony. Popular methods include a variety of penalized regression models: the least absolute shrinkage and selection operator (LASSO; Tibshirani, 1996), smoothly clipped absolute deviation (SCAD; Fan and Li, 2001), the elastic net (Zou and Hastie, 2005), the adaptive LASSO (Zou, 2006), and octagal shrinkage and clustering algorithm for regression (OSCAR; Bondell and Reich, 2008). Bayesian methods have also been developed for variable selection (George and McCulloch, 1993; George and Foster, 2000; Brown et al., 2002; Carvalho et al., 2010; Bondell and Reich, 2012). However, these variable selection methods emphasize prediction, not estimation of the effect of one exposure of interest while treating the other predictors as confounders. Applying these general variable selection methods without specifically treating covariates as confounders could be problematic. For example, if a confounder is highly correlated with the exposure and the exposure is forced to be included in the model, the confounder will likely be dropped, leading to bias.

Rather than selecting a single model, Bayesian model averaging (BMA) attempts to account

for effect uncertainty by averaging the effect estimate over the entire model space, giving higher weights to models that have greater support from the data. However, to insure an unbiased effect estimate the model should contain all confounding variables. Crainiceanu et al. (2008) noted that BMA averages over the subspace of models that includes all confounding variables and the subspace that does not, and demonstrated the bias in the effect estimate that can result. To address this, Crainiceanu et al. (2008) proposed a two-step approach. In the first step, the exposure is regressed on the other covariates to identify potential confounders. In the second step, the outcome is regressed on the exposure and the confounders identified in the first step. Additional explanatory variables not identified as confounders are selected in this second step.

Wang et al. (2012) combine the exposure and outcome models into a one-step Bayesian model. For both models, a set of indicator variables parameterize which covariates are included, and those included in the exposure model are included in the outcome model with probability one. Conceptually, this is BMA that only averages over the subspace of exposure models that include all confounders while simultaneously accounting for uncertainty in confounder selection. This method worked well for larger sample sizes, but is biased in simulations for smaller sample sizes and no optimality properties have been identified. In addition, it is computationally intensive.

In this paper we take a decision-theoretic approach to confounder selection and effect estimation. Our approach extends that of Bondell and Reich (2012) from the usual variable selection setting to the important problem of confounder adjustment. We first fit the standard Bayesian regression model and then post-process the posterior distribution in a decision-theoretic way using a loss function that penalizes models that do not include confounding variables and other important covariates. Thus, the proposed estimator is the Bayes rule associated with the proposed confounder-specific loss function. The distinction between the different losses associated with errors in the estimates of the variable of interest and confounders are made explicit in our loss function that leads to the final estimate. This approach allows the priors on the regression coefficients to represent the

state of knowledge before seeing the data as in the usual subjective Bayesian model, rather than assigning priors designed to avoid errors caused by failing to include important confounders, which is less natural in the Bayesian framework.

The proposed method has several appealing properties. Under general conditions it exhibits the oracle properties (Fan and Li, 2001) and simulation studies show good finite sample performance. Our method can be easily fit with existing software for a wide variety of outcome and exposure models, including all generalized linear models. In some cases Markov chain Monte Carlo methods are not required, for example, the normal linear model with normal or flat priors. We also establish a connection between our method and the adaptive LASSO (Zou, 2006), but with weights tailored to confounder adjustment rather than variable selection.

## 2. Methods

### 2.1 Modeling approach

The idea behind the proposed method is to find the simplest model among all feasible models. We define a feasible model as one with coefficients in the  $(1 - \alpha) \times 100\%$  posterior region of the full model. Within this credible region we select the model that maximizes parsimony. In this case, parsimony is defined as the model containing all confounders and other covariates associated with the outcome, but no other covariates.

Consider the normal linear outcome model

$$\mathbf{Y} = \beta_0 + \mathbf{X}\beta_x + \mathbf{U}\beta_u + \epsilon_y, \quad \epsilon_y \sim \mathbf{N}(0, \sigma_y^2 \mathbf{I}) \quad (1)$$

and exposure model

$$\mathbf{X} = \gamma_0 + \mathbf{U}\gamma_u + \epsilon_x, \quad \epsilon_x \sim \mathbf{N}(0, \sigma_x^2 \mathbf{I}). \quad (2)$$

For notational simplicity, let  $\boldsymbol{\beta} = (\beta_0, \beta_x, \boldsymbol{\beta}_u^T)^T$ ,  $\boldsymbol{\beta}_u = (\beta_1, \dots, \beta_p)^T$ ,  $\boldsymbol{\gamma} = (\gamma_0, \boldsymbol{\gamma}_u^T)^T$ , and  $\boldsymbol{\gamma}_u = (\gamma_1, \dots, \gamma_p)^T$  and assume  $\mathbf{W} = (\mathbf{1}, \mathbf{X}, \mathbf{U})$  is  $n \times (p + 2)$  with a column of ones for the intercept.

Let  $\mathcal{C}_\alpha^\beta$  and  $\mathcal{C}_\alpha^\gamma$  be  $(1 - \alpha) \times 100\%$  posterior credible regions of  $\beta$  and  $\gamma$ , respectively. In the case of flat or  $N(0, \sigma_y^2/\tau_y \mathbf{I})$  prior on  $\beta$  (with  $\tau_y$  fixed) and inverse-gamma prior on  $\sigma_y^2$ , the highest posterior density region is of the form  $\mathcal{C}_\alpha^\beta = \{\beta : (\beta - \hat{\beta})^T \Sigma_y^{-1} (\beta - \hat{\beta}) < C_\alpha\}$  where  $\hat{\beta}$  and  $\Sigma_y$  are the posterior mean and covariance. For regions of this form, there is a one-to-one relationship between the chosen  $\alpha$  level and the scalar  $C_\alpha$ . For the exposure model,  $\mathcal{C}_\alpha^\gamma$  has a similar form. In general, elliptical credible regions of this form exist for other priors on  $\beta$  and for other likelihoods such as generalized linear models, but may not be the highest posterior density regions.

The set of feasible exposure and confounder coefficients are  $\{\beta : \beta \in \mathcal{C}_\alpha^\beta\}$  and  $\{\gamma : \gamma \in \mathcal{C}_\alpha^\gamma\}$  for a given probability level  $\alpha \in (0, 1)$ . This feasible space will potentially include coefficient vectors with some parameters equal to zero, and thus a reduced model.

The model of interest is  $\mathbf{Y} = \beta_0 + \mathbf{X}\beta_x + \mathbf{U}_{\mathcal{A}_u}\beta_{\mathcal{A}_u} + \epsilon_y$ , where  $\mathcal{A}_u = \{j | \beta_j \neq 0\} \cup \{j | \gamma_j \neq 0\}$  and  $\mathbf{U}_{\mathcal{A}_u}$  and  $\beta_{\mathcal{A}_u}$  are the corresponding subsets of  $\mathbf{U}$  and  $\beta$ . This includes all variables correlated with the outcome or exposure. It is possible that this may include some variables correlated with  $\mathbf{X}$  but not  $\mathbf{Y}$ ; however, the errors resulting from including these variables are less severe than errors from omitting variables that are correlated with both the outcome and exposure.

To find a sparse estimate we are interested in the feasible coefficient vector that minimizes the cardinality of  $\mathcal{A}_u$ . The proposed estimator is

$$\tilde{\beta} = \underset{\beta}{\operatorname{argmin}} \epsilon (\beta_0^2 + \beta_x^2) + \|\beta_u\|_0 + \|\gamma_u\|_0, \quad \text{subject to } \beta \in \mathcal{C}_\alpha^\beta \text{ and } \gamma \in \mathcal{C}_\alpha^\gamma \quad (3)$$

where  $\|\cdot\|_0$  denotes the  $L_0$  norm of a vector, that is, the number of nonzero elements, and  $\epsilon$  is fixed and small so the intercept and exposure effect are essentially unpenalized.

## 2.2 Penalized regression reformulation

The solution to (3) can be hard to find in practice as it requires a search over potentially high-dimensional posterior regions. To ease computation, we use the smooth homotopy between  $L_0$  and  $L_1$  proposed by Lv and Fan (2009) and used by Bondell and Reich (2012) for the usual linear

regression setting,  $\rho_a(t) = \{(a+1)t\}/(a+t)$ . The proposed criterion is  $\sum_{j=1}^p \rho_a(|\beta_j|+|\gamma_j|)$  and  $\lim_{a \rightarrow 0^+} \sum_{j=1}^p \rho_a(|\beta_j|+|\gamma_j|) = \|(|\beta_u|+|\gamma_u|)\|_0$ . The penalty can be further simplified with a local linear approximation (Zou and Li, 2008). The local linear approximation of  $\rho_a(|\beta_j|+|\gamma_j|)$  around the posterior mean  $(\hat{\beta}_j, \hat{\gamma}_j)$  is

$$\rho_a(|\beta_j|+|\gamma_j|) \approx \frac{a(a+1)(|\beta_j|+|\gamma_j|)}{(a+|\hat{\beta}_j|+|\hat{\gamma}_j|)^2} + \frac{a(a+1)(|\hat{\beta}_j|+|\hat{\gamma}_j|)}{(a+|\hat{\beta}_j|+|\hat{\gamma}_j|)^2}. \quad (4)$$

Since both credible regions are convex sets of the form  $\{\beta : \beta \in \mathcal{C}_\alpha^\beta\}$  and  $\{\gamma : \gamma \in \mathcal{C}_\alpha^\gamma\}$ , the proposed estimator in (3) with the local linearized penalty (4) is equivalent to the Lagrangian optimization problem

$$\begin{aligned} \tilde{\beta} = \underset{\beta}{\operatorname{argmin}} & (\beta - \hat{\beta})^T \Sigma_y^{-1} (\beta - \hat{\beta}) + (\gamma - \hat{\gamma})^T \Sigma_x^{-1} (\gamma - \hat{\gamma}) \\ & + \lambda \epsilon (\beta_0^2 + \beta_x^2) + \lambda \sum_{j=1}^p \left[ \frac{(|\beta_j|+|\gamma_j|)}{(a+|\hat{\beta}_j|+|\hat{\gamma}_j|)^2} + \frac{(|\hat{\beta}_j|+|\hat{\gamma}_j|)}{(a+|\hat{\beta}_j|+|\hat{\gamma}_j|)^2} \right]. \end{aligned} \quad (5)$$

In (5), the Lagrangian multiplier  $\lambda$  absorbs the  $a(a+1)$  in the numerator of (4).

Because our primary interest is in the effect of  $X$ , we let  $\epsilon = 0$  to completely remove the penalty from the intercept and exposure. The second and fifth terms on the right hand side of (5) and the  $\gamma_j$  from the numerator of the fourth term do not effect the minimization with respect to  $\beta$  and can be removed. Finally, by letting  $a \rightarrow 0$  we get the objective function

$$\tilde{\beta} = \underset{\beta}{\operatorname{argmin}} (\beta - \hat{\beta})^T \Sigma_y^{-1} (\beta - \hat{\beta}) + \lambda \sum_{j=1}^p \frac{|\beta_j|}{(|\hat{\beta}_j|+|\hat{\gamma}_j|)^2}. \quad (6)$$

For any given dataset, there is a one-to-one and decreasing relationship between  $\alpha$  and  $\lambda$  because of the one-to-one relationships between  $\alpha$  and  $\mathcal{C}_\alpha$  and between  $\mathcal{C}_\alpha$  and  $\lambda$ . The path of estimates obtained by varying  $\alpha$  is identical to the path obtained by varying  $\lambda$ . This converts (3) into a convex optimization problem with a single tuning parameter that can be easily solved with existing software.

### 2.3 Simplification under flat prior

Assuming the linear model in (1) and a flat prior for  $\beta$ ,  $\text{pr}(\beta) \propto 1$ , the posterior mean is the least squares estimate  $\hat{\beta} = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathbf{Y}$  and the posterior covariance is proportional to  $\Sigma_y \propto (\mathbf{W}^T \mathbf{W})^{-1}$ . In this case,  $(\beta - \hat{\beta})^T \Sigma_y^{-1} (\beta - \hat{\beta})$  is a linear function of the sum of squared errors and the estimate can be written

$$\tilde{\beta} = \underset{\beta}{\text{argmin}} (\mathbf{Y} - \mathbf{W}\beta)^T (\mathbf{Y} - \mathbf{W}\beta) + \lambda \sum_{j=1}^p \hat{w}_j |\beta_j| \quad (7)$$

where  $\hat{w}_j = 1/(|\hat{\gamma}_j| + |\hat{\beta}_j|)^2$  for confounders  $j = 1, \dots, p$ . This is a special case of the adaptive LASSO solution (Zou, 2006) with data-driven weights  $w_j$  tailored to confounder adjustment so that variables associated with either the exposure (large  $|\hat{\gamma}_j|$ ) or the response (large  $|\hat{\beta}_j|$ ) have small penalty and are thus encouraged to be included in the outcome model.

### 2.4 Extension to multiple exposures

There is often interest in the effect of multiple exposures on an outcome. The penalized credible region confounder selection method naturally extends to the multiple exposure case. For a second exposure  $\mathbf{X}_2$  there is a second exposure model, for example

$$\mathbf{X}_2 = \delta_0 + \mathbf{U}\delta_u + \epsilon_{x_2}, \quad \epsilon_{x_2} \sim \mathbf{N}(0, \sigma_{x_2}^2 \mathbf{I}). \quad (8)$$

Using a similar approach to the single exposures setting the penalized regression estimator for multiple exposures is

$$\tilde{\beta} = \underset{\beta}{\text{argmin}} (\beta - \hat{\beta})^T \Sigma_y^{-1} (\beta - \hat{\beta}) + \lambda \sum_{j=1}^p \frac{|\beta_j|}{\left(|\hat{\beta}_j| + |\hat{\gamma}_j| + |\hat{\delta}_j|\right)^2}. \quad (9)$$

All the computation remains the same with (9). Additional exposure models can be used when there are more than two exposures of interests. In addition, more complex relationships can be estimated when it is appropriate. For example,  $\mathbf{X}_2$  could be a covariate in (2) if it is believed that confounders are correlated with the first exposure conditionally on  $\mathbf{X}_2$ , i.e. Simpson's paradox.

### 3. Theoretical Results

Although motivated by Bayesian decision theory, the proposed estimator can take the form of a standard penalized regression objective function. Therefore, it is of interest to evaluate this new estimate using the techniques of penalized regression. Theorem 1 demonstrates that the credible region confounder method has the oracle property for a properly chosen  $\lambda_n$ . Without loss of generality let  $\mathcal{A} = \{0, x, 1, \dots, p_0\}$  and make the following assumptions:

ASSUMPTION 1: The posterior covariance  $\Sigma_y^{-1}/n \xrightarrow{d} \mathbf{C}$ , where  $\mathbf{C}$  is a  $(p+2) \times (p+2)$  positive definite matrix that can be partitioned as

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix}. \quad (10)$$

ASSUMPTION 2: The posterior means are asymptotically normal,  $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(0, \mathbf{C}^{-1})$ , and analogously for  $\boldsymbol{\gamma}$ .

THEOREM 1: Under conditions 1 and 2, if  $\lambda_n/\sqrt{n} \rightarrow 0$  and  $\lambda_n\sqrt{n} \rightarrow \infty$  the penalized credible region confounder method is consistent in variable selection,  $\lim_{n \rightarrow \infty} pr(\mathcal{A}_n = \mathcal{A}) = 1$ , and asymptotically normal,  $\sqrt{n}(\tilde{\boldsymbol{\beta}}_{\mathcal{A}} - \boldsymbol{\beta}_{\mathcal{A}}) \xrightarrow{d} N(0, \mathbf{C}_{11}^{-1})$ .

REMARK 1: Theorem 1 requires that the posterior mean is consistent for  $\boldsymbol{\beta}$ , which holds if  $\boldsymbol{\beta}$ 's prior does not change with the sample size and has positive mass in the neighborhood of the truth, and that the asymptotic covariance matrix of  $\mathbf{W}$  is full rank.

REMARK 2: Assumption 2 is satisfied in many settings. In the case of flat or  $N(0, \sigma_y^2/\tau_y \mathbf{I})$  prior on  $\boldsymbol{\beta}$  (with  $\tau_y$  fixed) and inverse-gamma prior on  $\sigma_y^2$ , as used in this paper, the posterior of  $\boldsymbol{\beta}$  is multivariate- $t$  with degrees of freedom on the order of  $n$ . Hence, the posterior is asymptotically normal. More generally, the Bayesian central limit theorem implies asymptotic normality under loose regularity conditions.



Theorem 1 assumes a rate on  $\lambda_n$ . However, for the Bayesian credible regions approach it would be more natural to study asymptotics as a function of the confidence level  $\alpha_n$ . For a given dataset, there is a one-to-one correspondence between  $\lambda$  and  $\alpha$ . However, the limiting behavior as a function of  $\alpha_n$  can be quite different than those for  $\lambda_n$  (Gunes and Bondell, 2012; Bondell and Reich, 2012). Theorem 1 implies path consistency as a function of  $\alpha_n$ . Because the solution paths obtained by varying  $\alpha$  and  $\lambda$  are identical for each dataset, the selection consistency result in Theorem 1 implies that the correct model will be included in the solution path with probability tending to one for the Bayesian credible regions approach. A proof of Theorem 1 is in the Web Appendix.

#### 4. Computation and tuning

We first fit the full exposure and outcome models separately. This can be done with MCMC or with a closed form solution when it is known to exist. Given the posterior means  $\hat{\beta}$  and  $\hat{\gamma}$  and posterior covariances  $\Sigma_x$  and  $\Sigma_y$ , we obtain the solution to (6) using least angle regression (LARS; Efron et al., 2004). We then refit the outcome model using only confounders with nonzero coefficients for our final estimates.

Let  $X^* = \Sigma_y^{-1/2}D$  where  $D = \text{diag}\{\delta^{-1}, \delta^{-1}, (|\hat{\beta}_1| + |\hat{\gamma}_1|)^2, \dots, (|\hat{\beta}_p| + |\hat{\gamma}_p|)^2\}$  and  $\delta$  is a very small number that effectively removes the penalty on  $\beta_0$  and  $\beta_x$ . If  $\delta$  is 0 then the model fit is exactly (6); however, approximation with a very small number results in the same solution and a nonzero value is required for LARS. Then let  $Y^* = \Sigma_y^{-1/2}\hat{\beta}$ . Solving the  $L_1$  penalized regression function with  $Y^*$  and  $X^*$  gives the solution  $\beta^*$ . The solution to (3) is  $\tilde{\beta} = D\beta^*$ . The full solution can be estimated with the `BayesPen` package in R which is available on CRAN (Wilson et al., 2014).

The method must be tuned by selecting either the confidence level  $\alpha \in (0, 1)$  or the penalty parameter  $\lambda > 0$ . For any given dataset, there is a one-to-one relationship between  $\alpha$  and  $\lambda$  and the path of estimates obtained by varying  $\alpha$  is identical to the path obtained by varying  $\lambda$ . Gunes and Bondell (2012) discuss this relationship in detail and note that most tuning methods (AIC,

BIC, and cross-validation) can result in similar models but very different  $\alpha$  levels for different data sets. Further, the solution corresponding to a certain  $\alpha$  level is not related to controlling the false positive rate at that level. As such, tuning based on a pre-specified  $\alpha$  level can result in poor selection properties, while using the solution path obtained from varying  $\alpha$  or  $\lambda$  performs well. For convenience, we tune based on  $\lambda$  since this is the approach taken in the LARS package in R. This gives the entire solution path.

There are several approaches to selecting a model from that path. We found that forward selection performed best in our simulation study. We performed forward selection on the exposure and outcome model simultaneously using the same path. Forward selection stops when the minimum of the  $\chi^2$  tests for adding an additional covariate to the outcome and exposure model fails to reject at a particular  $\alpha$  level denoted by  $\alpha^{fs}$ . Larger  $\alpha^{fs}$  values will include more covariates and can be considered more conservative. The LARS path may have a variable appear in a model and then be removed before reappearing (Efron et al., 2004). For forward selection, we modified the LARS path to include a covariate in every model after it first appears in order to ensure nested models for testing. Alternative selection methods include cross-validation, Mallows's  $C_p$  (Mallows, 1973), Akaike information criterion (Akaike, 1973), and Bayesian information criterion (Schwarz, 1978). The Web Appendix contains a simulation comparing these methods.

## 5. Simulation Study

### 5.1 Simulation with linear model

Our first simulation design follows from Wang et al. (2012). We assume  $\mathbf{Y} \sim \mathbf{N}(\mathbf{W}\boldsymbol{\beta}, \mathbf{I})$  where  $\mathbf{W}$  is an  $n \times (p+1)$  design matrix with the variable of interest in the first column and  $p = 57$  potential confounders. We let the covariates for the  $i$ th observation,  $\mathbf{W}_i$ , be independent  $\mathbf{N}(0, \boldsymbol{\Sigma})$ . The covariance matrix  $\boldsymbol{\Sigma} = \{\sigma_{jk}\}$  has diagonal elements  $\sigma_{jj} = 1$  and nonzero off-diagonal elements  $\sigma_{jk} = 0.7^{j+k-2}$  for  $j \neq k = 0, \dots, 7$ . The remaining off-diagonal elements are zero; hence,  $w_{ik}$

for  $k > 7$  are independent standard normal. The coefficients are  $\beta_x = \beta_1 = \dots = \beta_{14} = 0.1$  and  $\beta_{15} = \dots = \beta_{57} = 0$ . Hence, we have seven confounding variables correlated with  $\mathbf{Y}$  and  $\mathbf{X}$ , seven additional explanatory variables correlated with  $\mathbf{Y}$  but not  $\mathbf{X}$ , and 43 variables uncorrelated with both  $\mathbf{X}$  and  $\mathbf{Y}$ .

We present the results for the penalized credible regions using three different priors on  $\beta$  and  $\gamma$ : flat as assumed in Section 2.3;  $\beta \sim N(0, \sigma_y^2/\tau_y)$  and  $\gamma \sim N(0, \sigma_x^2/\tau_x)$  with  $\tau_y$  and  $\tau_x$  estimated from a linear model and fixed (empirical Bayes); and with independent Gamma(0.001,0.001) priors on  $\tau_y$  and  $\tau_x$  for a fully Bayesian approach with proper priors. For the empirical Bayes approach we let  $\hat{\tau}_y = \hat{\sigma}_y^2/\{(p+1)^{-1} \sum_{j=0}^p \hat{\beta}_j^2\}$  and  $\hat{\tau}_x = \hat{\sigma}_x^2/(p^{-1} \sum_{j=1}^p \hat{\gamma}_j^2)$  where  $\hat{\sigma}$ ,  $\hat{\beta}$ , and  $\hat{\gamma}$  are estimated from the full linear model. In all cases we put independent Gamma(0.001,0.001) priors on  $\sigma_y^{-2}$  and  $\sigma_x^{-2}$ .

Table 1 compares model performance for the credible region method with the three priors. For comparison we evaluate the simulated data with the true frequentist linear model that includes only the first 14 covariates; the full frequentist linear model that includes all 57 covariates; BMA with  $\beta_j = \eta_j \alpha_j$ ,  $\eta_j \sim \text{Bern}(0.5)$ , and  $\alpha_j \sim N(0, 10^2)$ ; Bayesian adjustment for confounding (BAC; Wang et al., 2012) with  $\omega = \infty$  using the R package BEAU; and adaptive LASSO with  $\gamma = 2$ .

[Table 1 about here.]

In general, the fully Bayesian credible region approach performed similarly to the true model in terms of bias, MSE, and interval coverage for  $\hat{\beta}_x$ , with the exception of the smallest sample size where it was biased and had lower coverage. At the larger sample sizes, the empirical Bayes and flat prior versions perform well and provide faster computation. These methods (and the adaptive LASSO) require a large enough sample size to get good initial least squares estimates  $\hat{\beta}$  and  $\hat{\gamma}$ , as well as  $\hat{\tau}_y$  and  $\hat{\tau}_x$  for the empirical Bayes priors, and thus perform poorly for small  $n$ . BMA, BAC, and adaptive LASSO had higher MSE, were more biased, and had lower interval coverage compared to the fully Bayesian credible region method, although the MSE for BAC and adaptive

LASSO approaches that of the credible region at higher sample sizes. At the larger sample sizes, BAC was notably slower than the other methods. For the moderate and larger sample sizes the interval coverage for  $\hat{\beta}_x$  achieved the nominal 95% level indicating that proper inference is being made with the credible region method.

We computed the AUC for the credible region approach and adaptive LASSO by calculating the sensitivity and specificity as additional variables are added along the path formed by varying  $\lambda$ . The credible region approach had larger AUC showing the benefit of weights specifically tailored to confounder adjustment. For BAC and BMA the AUC is estimated by adding variables to the model according to their inclusion probability. BAC had slightly higher AUC than the credible region method at the smallest sample size but smaller AUC at the largest sample size.

Despite the higher AUC at small sample sizes, BAC tended to give more weight to smaller models. For  $n = 100$ , Figure 1 compares the inclusion probability for each of the important covariates and the average inclusion probability of the other covariates, using the empirical Bayes priors for the confounder method. This shows that the credible region approach and BMA had a higher inclusion rate for the confounding variables than BAC and adaptive LASSO.

[Figure 1 about here.]

While variables 1 through 14 have the same effect size, their correlation with the exposure varies. The credible region approach includes confounding variables with greater frequency. The same is true for BAC but the pattern is not as strong. Adaptive LASSO and BMA do not account for the correlation with the outcome resulting in similar inclusion probabilities for all true predictors, not an emphasis on those correlated with the exposure that can impact the exposure effect estimation. The inclusion rate for the true confounders is below one because confounders that do not reduce the sum of squared errors for either model are excluded. By excluding true confounders that do not reduce the sum of squared errors the credible region approach can have lower MSE than the true model due to the larger error degrees of freedom.

These simulation results highlight the advantage of the penalized credible region approach rather than BMA or BAC which use zero-inflated priors for selection. Specifically, at smaller sample sizes the credible regions approach has smaller bias and MSE and higher interval coverage. The credible regions approach with flat priors has a similar objective function to the adaptive LASSO, except the weights in the penalty are augmented for confounder selection. Hence, the comparison of the results for the adaptive LASSO and the penalized credible region approach with flat priors highlights the advantage gained by including the confounder specific weights. These advantages include smaller bias and MSE as well as better interval coverage and AUC. Using the approach of Bondell and Reich (2012) with flat priors results in an estimator identical to the adaptive LASSO; hence, the use of the exposure model and confounder-specific loss function proposed in this paper improve performance when the primary interest is in estimating an exposure effect. Compared to a purely penalized regression approach, by using the Bayesian framework that motivated this method a practitioner can select priors for the regression coefficients. Even a vague normal prior improves finite sample performance as demonstrated by the empirical Bayes and fully Bayesian penalized credible region methods used here, and there is potential for additional gains when more prior information is available.

### 5.2 *Simulation in the ultra high-dimensional setting.*

In many cases there is interest in inference for  $p > n$ . To test the confounder selection performance in this setting we add additional covariates to the simulation that are independent standard normal uncorrelated with both the outcome and the exposures. Hence, the simulation design remains unchanged with the exception of additional noise variables added to the design. We use the fully Bayesian prior with gamma prior on  $\tau_x$  and  $\tau_y$  and let  $n$  grow at three rates:  $n = 100$ ,  $n = 15 + \sqrt{p}$ , and  $n = p/2$ . Figure 2 shows the resulting AUC. When  $n = p/2$  the credible regions approach performs well for variable selection at large sample sizes. However, for slower rates of growth the

credible regions approach does not adequately select confounding variables. Hence, for  $p > n$  if  $n$  grows at a sufficiently rate the penalized credible region method performs well.

[Figure 2 about here.]

### 5.3 Simulation with binary treatment and logistic confounder model

Our second simulation design assumes two treatment groups with no treatment effect. We assume that there is a set of five confounders that are predictors of  $\mathbf{Y}$  and have different distribution for the two treatment groups. We generate  $X_i$  as independent Bernoulli with probability 0.5 and let  $U_{ij}$  be normal with mean  $j^{-1}X_i$  and standard deviation  $j^{-1}$  for  $j = 1, \dots, 5$ . The remaining variables  $(U_{i6}, \dots, U_{i50})$  are independent standard normal. The regression coefficients are  $\beta_1, \dots, \beta_5 = 0.1$  and  $\beta_x = \beta_6 = \dots = \beta_{50} = 0$  and  $\mathbf{Y}$  is  $\mathbf{N}(\mathbf{W}\boldsymbol{\beta}, \mathbf{I})$ .

We fit the penalized credible regions using a logistic exposure model:  $X_i$  is normal Bernoulli( $p_i$ ),  $p_i = \text{logit}(\gamma_0 + \mathbf{u}_i^T \boldsymbol{\gamma}_u)$ , and the elements of  $\boldsymbol{\gamma}$  have independent standard normal priors. We present the same three priors for the outcome model as used in simulation design one: flat, empirical Bayes, and Gamma. For comparison we show the true and full regression model as well as Bayesian model averaging and adaptive LASSO, but not BAC because the BEAU package does not model a binary treatment effect.

The credible regions methods and BMA had smaller MSE than the true model but were biased at the smaller sample sizes. For  $n = 200$  and  $500$  the credible region methods were unbiased and had MSE similar to the true model, whereas BMA and adaptive LASSO were still biased and had larger MSE. In addition, the credible region method had interval coverage near the nominal level throughout, whereas BMA and adaptive LASSO were lower. The AUC with the credible region method was larger than with the adaptive LASSO at all sample sizes and larger than with BMA for all except the smallest sample size.

[Table 2 about here.]

## **6. Data Analysis**

To illustrate the credible region method we estimate the effect of mean fine particulate matter ( $PM_{2.5}$ ) in the first trimester of pregnancy on birth weight in Mecklenburg County, North Carolina, while accounting for several potential confounders related to the mother's socioeconomic status, medical history, seasonality, and other weather variables.  $PM_{2.5}$  air pollution has been associated with various adverse pregnancy outcomes (Bosetti et al., 2010; Šrám et al., 2005). We study the effect of  $PM_{2.5}$  on birth weight using data similar to Chang et al. (2012).  $PM_{2.5}$  levels were obtained from the U.S. Environmental Protection Agency's Fused Air Quality Predictions Using Downscaling ([http://www.epa.gov/esd/land-sci/lcb/lcb\\_faqs.html](http://www.epa.gov/esd/land-sci/lcb/lcb_faqs.html)). From these data we computed average  $PM_{2.5}$  over the first trimester for each birth. We used birth and covariate data from the North Carolina Vital Statistics – Births 2003 through 2007 (State Center for Health Statistics, 2008) from the State Center for Health Statistics (SCHS) and the Howard W. Odum Institute for Research in Social Science at University of North Carolina at Chapel Hill. Temperature and dew point data were obtained from National Oceanic and Atmospheric Administration Climate Data Online (NOAA-CDO; <http://www.ncdc.noaa.gov/cdo-web/>).

We looked at the group of at risk women age 40 and over and limited the sample to single births that reached at least 37 weeks of gestation who were self-reported non-hispanic white, non-hispanic black, or hispanic. The exposure variable of interest is first trimester mean  $PM_{2.5}$ . In addition, following Warren et al. (2012), we include as potential confounders other variables observed during pregnancy, including: principal components (PC) of mean daily temperature throughout the pregnancy, PCs of mean daily dew point throughout the pregnancy, PCs for interaction of temperature and dew point, indicators for a birth in spring, summer, or fall. The PCs that explained 0.99 percent of the week-to-week variation in co-exposures were included. These covariates are correlated with  $PM_{2.5}$  and potentially impact birth weight. In addition we included variables relating to the birth and mother's medical history. These variables are listed in the Web Appendix. All

observations with missing data were excluded. The final sample contains 1399 women of which 902 are non-hispanic white, 308 are non-hispanic black, and 171 are hispanic. There are  $p = 65$  covariates in total. Within each subgroup the race/ethnicity covariates were removed as well as any other covariates that were not observed in the subgroup, resulting in 58, 61, and 63 covariates in the hispanic, non-hispanic black, and non-hispanic white models, respectively.

We used the credible region method with fully Bayesian priors. The resulting analysis shows that the credible regions method results in a similar point estimate for the  $PM_{2.5}$  exposure effect as the full model but has smaller variance. This is appealing because the full model includes all observed confounding variables; hence, a point estimate that substantially differs from the full model implies that important confounding variables in the data are omitted from the model.

Figure 3 shows the solution paths for each subgroup analysis. Figure 3a shows that for the hispanic subgroup the model with only  $PM_{2.5}$  included (step 0) indicates that  $PM_{2.5}$  is associated with increased birth weight. However, as additional confounding variables correlated with both  $PM_{2.5}$  and birth weight are added to the model the point estimate becomes negative. Starting at step 20, the exposure effect is unchanged by including additional covariates, indicating that all important confounding variables have been included.

Figure 4a shows that the credible region approach results in similar estimates to the full model overall and for each subgroup, but the standard errors of the effect estimate are 10% to 20% smaller with the credible regions approach (see Figure 4b). On the other hand, BMA and adaptive LASSO have smaller variances, but very different point estimates compared to the full model. This stems from selecting smaller models and omitting important confounders, particularly in the smaller subgroups. As a result, adaptive LASSO and the exposure only model find a statistically significant positive association between  $PM_{2.5}$  and birth weight among hispanics, a result that contradicts previous findings (Savitz et al., 2014; Pearce et al., 2012). This result mirrors that of the simulation study which showed these methods can be biased at small sample sizes.



[Figure 3 about here.]

[Figure 4 about here.]

## **7. Discussion**

This paper presents a new method for confounder selection and effect estimation using a decision-theoretic approach. The proposed estimator is the Bayes rule estimate associated with the confounder-specific loss function and allows the practitioner freedom to select an appropriate prior. Given the posterior mean and covariance for the coefficients of any generalized linear model the credible region confounder method can easily be applied and, in most cases, existing software to be used.

The proposed method outperformed alternative Bayesian confounder or variable selection approaches that utilize zero-inflated priors (BMA and BAC) in a simulation study. In addition to good finite sample performance, the credible region confounder method is consistent in variable selection and asymptotically normal under general conditions.

With flat priors on the regression coefficients, the proposed method is an adaptive LASSO-type estimator with data-driven weights tailored to confounder selection. The simulation study demonstrated that the confounder-specific weights in the penalty improved performance at small sample sizes (i.e. comparing adaptive LASSO to the credible region method with flat priors). While an alternative approach might be to skip the Bayesian motivation for the estimators, adding even a vague normal prior made a noticeable improvement in confounder selection at small sample sizes. In addition, the adaptive LASSO solution is identical to that of Bondell and Reich (2012) with flat priors, thus, demonstrating the advantage of the confounder-specific loss function in this similar Bayesian framework.

## 8. Supplementary Material

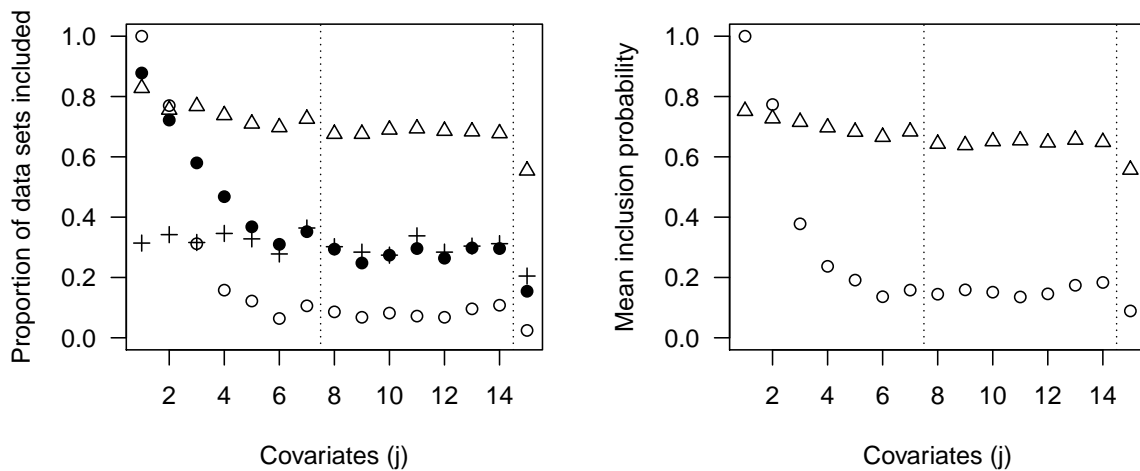
The Web Appendix referenced in Section 3, 4, and 6 is available at Biometrics website on Wiley Online Library. The R package `BayesPen` is available on CRAN.

## References

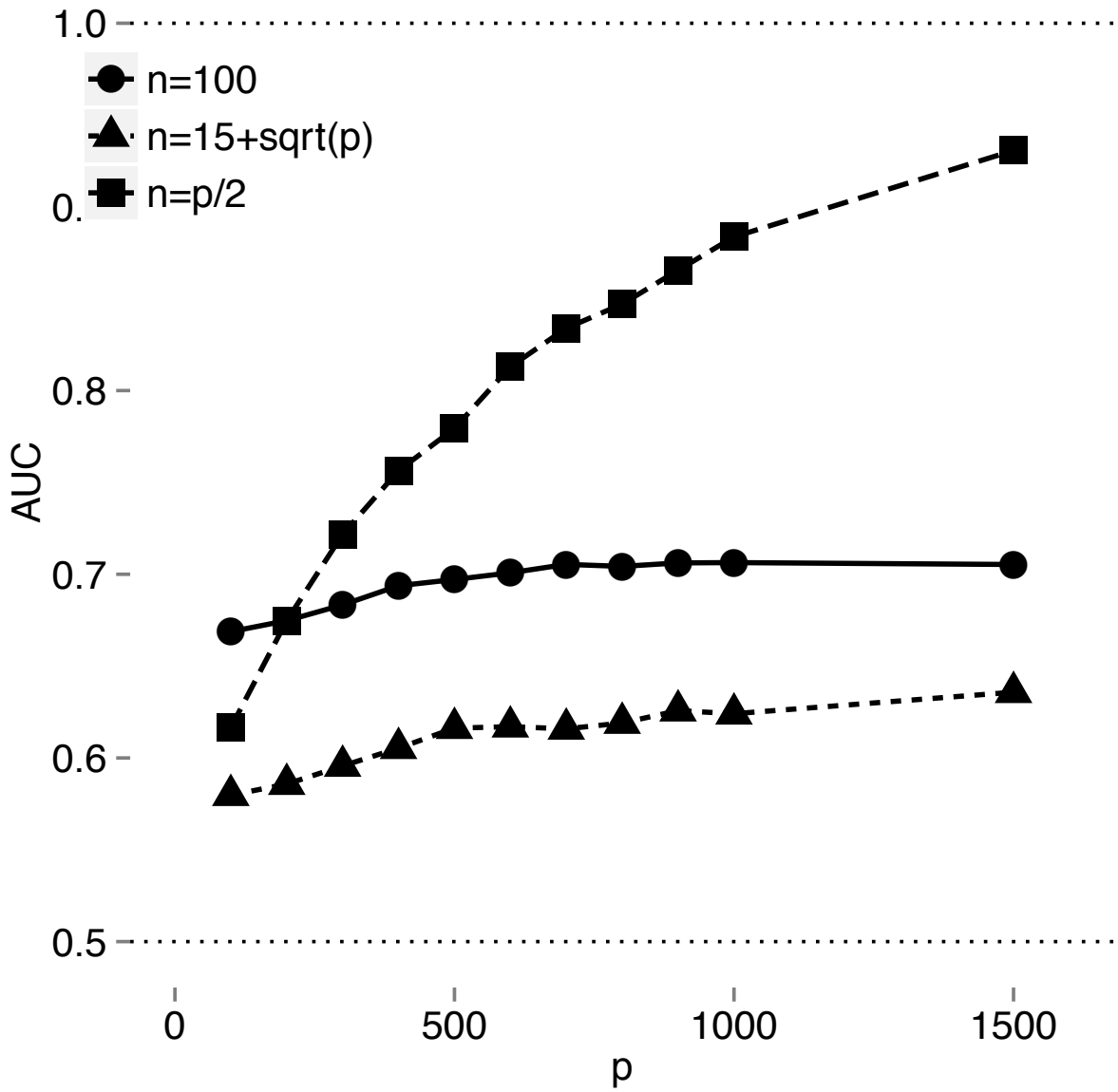
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Petrov, B. N. and Csàki, F., editors, *Proceedings of the 2nd International Symposium on Information Theory*, Springer Series in Statistics, page 267281. Akademiai Kiado, Budapest.
- Bondell, H. D. and Reich, B. J. (2008). Simultaneous regression shrinkage, variable selection, and supervised clustering of predictors with OSCAR. *Biometrics* **64**, 115–123, 322–323.
- Bondell, H. D. and Reich, B. J. (2012). Consistent high-dimensional bayesian variable selection via penalized credible regions. *J. Am. Statist. Assoc.* **107**, 1610–1624.
- Bosetti, C., Nieuwenhuijsen, M. J., Gallus, S., Cipriani, S., La Vecchia, C., and Parazzini, F. (2010). Ambient particulate matter and preterm birth or birth weight: a review of the literature. *Arch. Toxicol.* **84**, 447–460.
- Brown, P. J., Vannucci, M., and Fearn, T. (2002). Bayes model averaging with selection of regressors. *J. R. Stat. Soc. B* **64**, 519–536.
- Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* **97**, 465–480.
- Chang, H. H., Reich, B. J., and Miranda, M. L. (2012). Time-to-event analysis of fine particle air pollution and preterm birth: Results from North Carolina, 20012005. *Am. J. Epidemiol.* **175**, 91–98.
- Crainiceanu, C. M., Dominici, F., and Parmigiani, G. (2008). Adjustment uncertainty in effect estimation. *Biometrika* **95**, 635–651.
- Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *Ann. Statist.*

- 32**, 407–499. With discussion, and a rejoinder by the authors.
- Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Statist. Assoc.* **96**, 1348–1360.
- George, E. I. and Foster, D. P. (2000). Calibration and empirical Bayes variable selection. *Biometrika* **87**, 731–747.
- George, E. I. and McCulloch, R. E. (1993). Variable selection via gibbs sampling. *J. Am. Statist. Assoc.* **88**, 881–889.
- Gunes, F. and Bondell, H. D. (2012). A confidence region approach to tuning for variable selection. *J. Comput. Graph. Statist.* **21**, 295–314.
- Lv, J. and Fan, Y. (2009). A unified approach to model selection and sparse recovery using regularized least squares. *Ann. Statist.* **37**, 3498–3528.
- Mallows, C. L. (1973). Some comments on  $C_p$ . *Technometrics* **15**, 661–675.
- Pearce, M. S., Glinianaia, S. V., Ghosh, R., Rankin, J., Rushton, S., Charlton, M., Parker, L., Pless-Mulloli, T., et al. (2012). Particulate matter exposure during pregnancy is associated with birth weight, but not gestational age, 1962-1992: a cohort study. *Environmental Health* **11**, 13.
- Savitz, D. A., Bobb, J. F., Carr, J. L., Clougherty, J. E., Dominici, F., Elston, B., Ito, K., Ross, Z., Yee, M., and Matte, T. D. (2014). Ambient fine particulate matter, nitrogen dioxide, and term birth weight in New York, New York. *American Journal of Epidemiology* **179**, 457–466.
- Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6**, pp. 461–464.
- State Center for Health Statistics (2008). North Carolina vital statistics – births 2003 through 2007. <http://hdl.handle.net/1902.29/10131>.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B* **58**, 267–288.
- Šrám, R. J., Binková, B., Dejmek, J., and Bobak, M. (2005). Ambient air pollution and pregnancy outcomes: A review of the literature. *Environ. Health. Perspect.* **113**, pp. 375–382.

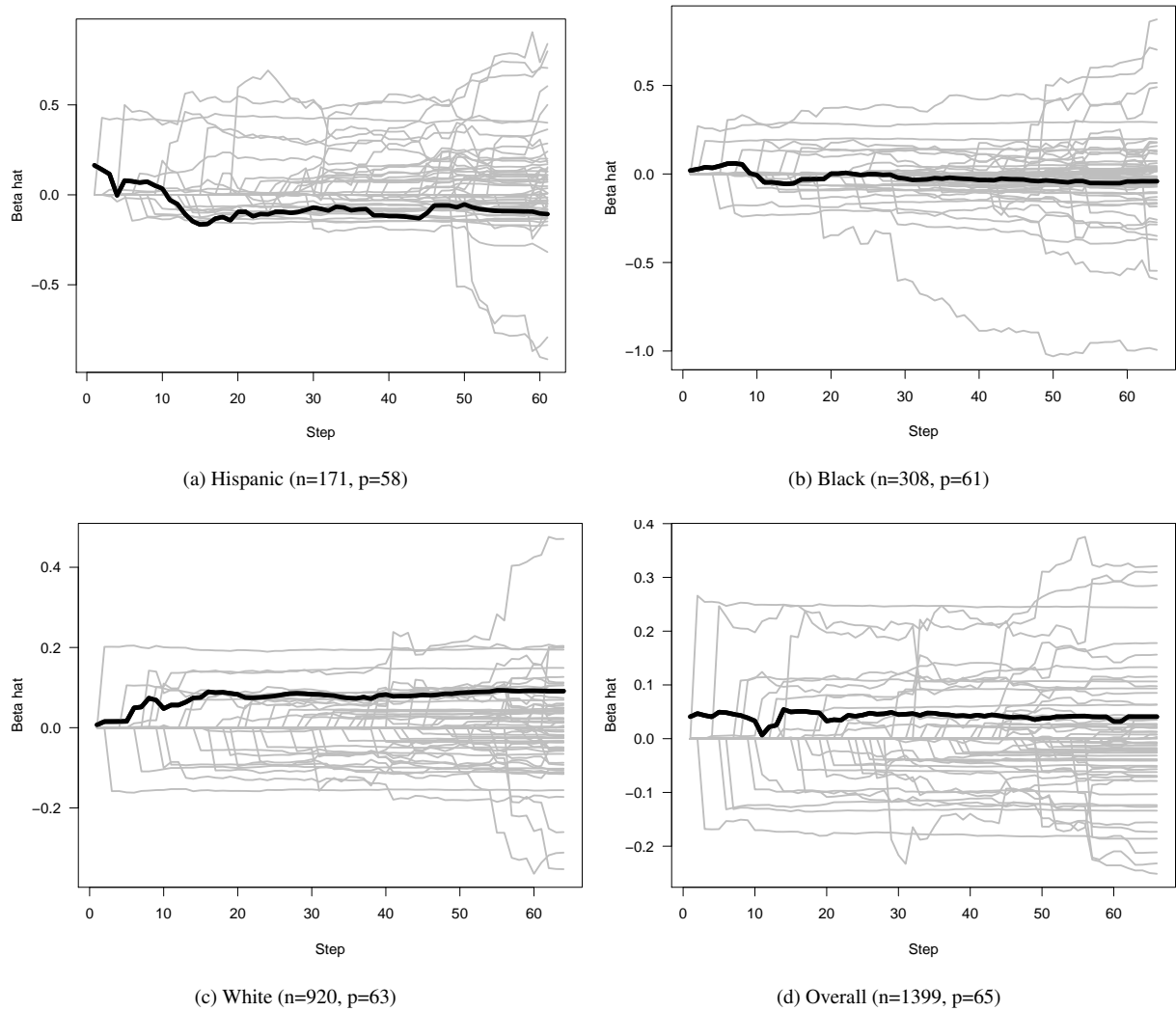
- Wang, C., Parmigiani, G., and Dominici, F. (2012). Bayesian effect estimation accounting for adjustment uncertainty. *Biometrics* **68**, 661–671.
- Warren, J., Fuentes, M., Herring, A., and Langlois, P. (2012). Spatial-temporal modeling of the association between air pollution exposure and preterm birth: Identifying critical windows of exposure. *Biometrics* **68**, 1157–1167.
- Wilson, A., Bondell, H. D., and Reich, B. J. (2014). *BayesPen: Bayesian Penalized Credible Regions*. R package version 1.0.
- Zou, H. (2006). The adaptive lasso and its oracle properties. *J. Am. Statist. Assoc.* **101**, 1418–1429.
- Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B* **67**, 301–320.
- Zou, H. and Li, R. (2008). One-step sparse estimates in nonconcave penalized likelihood models. *Ann. Statist.* **36**, 1509–1533.



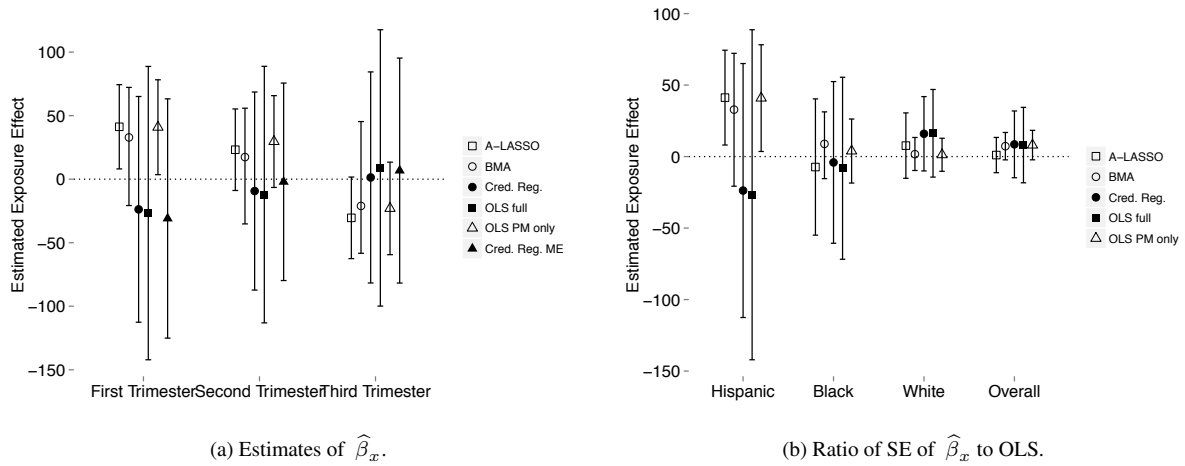
**Figure 1:** Probability of including covariates for simulation design one with  $n = 100$  for the credible region method (●), BAC (○), BMA (△), and adaptive LASSO (+). Left: the proportion of simulated data sets for which each variable is selected. For BMA and BAC a variable is counted as included if its inclusion probability is greater than 0.5. Right: the average inclusion probability for BMA and BAC. Covariates in the first section (1 to 7) are confounders, in the second section (8 to 14) are other explanatory variables, and the far right dot is the average of the non-important covariates (15 to 57).



**Figure 2:** AUC for simulation design one with  $p > n$  for three different rates of growth for  $n$ .



**Figure 3:** Solution path for three subgroups and overall. The thick black line is the estimated PM effect,  $\hat{\beta}_x$ . The other lines are the estimates for the other covariates in the model. The regression coefficients correspond to centered and scaled variables.



**Figure 4:** Estimates of the PM effect ( $\hat{\beta}_x$ ). Left: The point estimates with each model and the 95% interval. Right: The ratio of the SE or SD of  $\hat{\beta}_x$  with each model compared to the full OLS model.



Table 1: Simulation results for design 1. Bias, MSE, and coverage are for the effect of interest  $\hat{\beta}_x$ . Coverage is 95% confidence or credible interval coverage. CPU Time is reported in seconds on a MacBook Pro with OS X, 8 GB RAM, and 2 GHz Intel Core i7. SEs for the AUC range from 0.002 to 0.004, and for CPU time from less than 0.001 to 0.498.

$n = 60$	Bias		MSE		Coverage	AUC	CPU Time
True	0.005	(0.011)	0.060	(0.004)	0.93	NA	0.003
Full	0.007	(0.073)	2.624	(0.392)	0.95	NA	0.003
Cred. Reg. (Flat Priors)	0.181	(0.008)	0.065	(0.003)	0.69	0.530	0.084
Cred. Reg. (Empirical Bayes)	0.145	(0.009)	0.062	(0.003)	0.73	0.547	0.089
Cred. Reg. (Gamma Priors)	0.060	(0.010)	0.050	(0.003)	0.82	0.593	2.971
BMA	0.090	(0.009)	0.045	(0.003)	0.86	0.577	1.615
BAC ( $\omega = \infty$ )	0.080	(0.014)	0.110	(0.008)	0.76	0.633	6.166
Adaptive LASSO	0.076	(0.049)	1.214	(0.200)	0.58	0.521	0.049
$n = 100$	Bias		MSE		Coverage	AUC	CPU Time
True	-0.005	(0.008)	0.032	(0.002)	0.95	NA	0.003
Full	-0.013	(0.011)	0.065	(0.004)	0.95	NA	0.004
Cred. Reg. (Flat Priors)	0.051	(0.009)	0.044	(0.003)	0.76	0.624	0.108
Cred. Reg. (Empirical Bayes)	0.001	(0.007)	0.028	(0.002)	0.93	0.680	0.127
Cred. Reg. (Gamma Priors)	0.006	(0.007)	0.028	(0.002)	0.91	0.670	3.001
BMA	0.100	(0.006)	0.030	(0.002)	0.80	0.627	1.671
BAC ( $\omega = \infty$ )	0.050	(0.008)	0.033	(0.002)	0.91	0.701	7.874
Adaptive LASSO	0.071	(0.009)	0.043	(0.003)	0.71	0.571	0.041
$n = 200$	Bias		MSE		Coverage	AUC	CPU Time
True	0.004	(0.005)	0.015	(0.001)	0.95	NA	0.003
Full	0.004	(0.006)	0.018	(0.001)	0.95	NA	0.005
Cred. Reg. (Flat Priors)	0.012	(0.006)	0.017	(0.001)	0.90	0.748	0.165
Cred. Reg. (Empirical Bayes)	0.003	(0.005)	0.014	(0.001)	0.95	0.787	0.184
Cred. Reg. (Gamma Priors)	0.008	(0.005)	0.014	(0.001)	0.94	0.776	3.093
BMA	0.114	(0.005)	0.024	(0.001)	0.73	0.718	1.832
BAC ( $\omega = \infty$ )	0.044	(0.006)	0.017	(0.001)	0.93	0.779	40.779
Adaptive LASSO	0.048	(0.006)	0.020	(0.001)	0.78	0.687	0.051
$n = 500$	Bias		MSE		Coverage	AUC	CPU Time
True	0.000	(0.003)	0.005	(0.000)	0.95	NA	0.004
Full	0.000	(0.003)	0.006	(0.000)	0.95	NA	0.010
Cred. Reg. (Flat Priors)	0.001	(0.003)	0.006	(0.000)	0.94	0.906	0.268
Cred. Reg. (Empirical Bayes)	0.000	(0.003)	0.006	(0.000)	0.95	0.917	0.367
Cred. Reg. (Gamma Priors)	0.001	(0.003)	0.006	(0.000)	0.95	0.914	3.313
BMA	0.101	(0.003)	0.016	(0.001)	0.64	0.868	2.198
BAC ( $\omega = \infty$ )	0.024	(0.003)	0.006	(0.000)	0.95	0.896	499.235
Adaptive LASSO	0.022	(0.004)	0.008	(0.000)	0.80	0.861	0.098

Table 2: Simulation results for design 2. Bias, MSE, and coverage are for the effect of interest  $\hat{\beta}_x$ . Coverage is 95% confidence or credible interval coverage. CPU Time is reported in seconds on a MacBook Pro with OS X, 8 GB RAM, and 2 GHz Intel Core i7. SEs for the AUC range from less than 0.001 to 0.006, and for CPU time from less than 0.001 to 0.003.

$n = 60$	Bias		MSE		Coverage	AUC	CPU Time
True	-0.029	(0.015)	0.119	(0.008)	0.96	NA	0.003
Full	0.020	(0.040)	0.815	(0.062)	0.97	NA	0.003
Cred. Reg. (Flat Priors)	0.113	(0.013)	0.095	(0.006)	0.86	0.713	0.842
Cred. Reg. (Empirical Bayes)	0.114	(0.013)	0.093	(0.005)	0.87	0.718	0.836
Cred. Reg. (Gamma Priors)	0.132	(0.012)	0.095	(0.006)	0.88	0.718	1.855
BMA	0.115	(0.012)	0.080	(0.005)	0.91	0.736	1.143
Adaptive LASSO	0.151	(0.022)	0.260	(0.022)	0.61	0.532	0.035
$n = 100$	Bias		MSE		Coverage	AUC	CPU Time
True	-0.008	(0.012)	0.069	(0.004)	0.95	NA	0.003
Full	-0.005	(0.016)	0.124	(0.008)	0.97	NA	0.004
Cred. Reg. (Flat Priors)	0.029	(0.011)	0.059	(0.004)	0.93	0.883	1.342
Cred. Reg. (Empirical Bayes)	0.025	(0.011)	0.060	(0.004)	0.94	0.896	1.346
Cred. Reg. (Gamma Priors)	0.034	(0.011)	0.059	(0.004)	0.94	0.887	2.363
BMA	0.155	(0.009)	0.062	(0.003)	0.87	0.760	1.186
Adaptive LASSO	0.133	(0.012)	0.091	(0.005)	0.65	0.545	0.036
$n = 200$	Bias		MSE		Coverage	AUC	CPU Time
True	0.003	(0.009)	0.038	(0.003)	0.94	NA	0.003
Full	0.007	(0.010)	0.049	(0.003)	0.93	NA	0.005
Cred. Reg. (Flat Priors)	0.003	(0.009)	0.037	(0.002)	0.92	0.965	2.599
Cred. Reg. (Empirical Bayes)	0.001	(0.009)	0.037	(0.002)	0.94	0.975	2.603
Cred. Reg. (Gamma Priors)	0.002	(0.008)	0.036	(0.002)	0.94	0.971	3.637
BMA	0.174	(0.007)	0.055	(0.003)	0.75	0.812	1.267
Adaptive LASSO	0.115	(0.010)	0.058	(0.003)	0.66	0.621	0.043
$n = 500$	Bias		MSE		Coverage	AUC	CPU Time
True	-0.002	(0.005)	0.013	(0.001)	0.95	NA	0.003
Full	-0.001	(0.005)	0.014	(0.001)	0.96	NA	0.011
Cred. Reg. (Flat Priors)	-0.001	(0.005)	0.013	(0.001)	0.95	0.995	6.379
Cred. Reg. (Empirical Bayes)	0.000	(0.005)	0.013	(0.001)	0.96	0.998	6.433
Cred. Reg. (Gamma Priors)	0.000	(0.005)	0.013	(0.001)	0.96	0.996	7.460
BMA	0.164	(0.005)	0.039	(0.002)	0.64	0.878	1.635
Adaptive LASSO	0.065	(0.006)	0.023	(0.001)	0.74	0.731	0.077