Journal of Biometrics & Biostatistics
ISSN: 2155-6180

Yue et al., J Biomet Biostat 2013, 4:5
DOI: 10.4172/2155-6180.1000177

Research Article Open Access

Building Cost-efficient Models using BLARS Method

Li Hua Yue*, Wenqing He, Duncan Murdoch and Hristo Sendov

Department of Statistical and Actuarial Sciences, Western University, London, ON N6A5B7, Canada

Abstract

Variable selection is a difficult problem in building statistical models. Identification of cost efficient diagnostic factors is very important to health researchers, but most variable selection methods do not take into account the cost of collecting data for the predictors. The trade-off between statistical significance and cost of collecting data for a statistical model is our focus. In this paper, we extend the LARS variable selection method to incorporate costs of factors in variable selection, which also works with other methods of variable selection, such as Lasso and adaptive Lasso. A branch and bound search method combined with LARS is employed to select cost-efficient factors. We apply the resulting branching LARS method to a dataset from an Assertive Community Treatment project conducted in Southwestern Ontario to demonstrate the cost-efficient variable selection process, and the results show that a "cheaper" model could be selected by sacrificing a user selected amount of model accuracy.

Keywords: BLARS; Branch and bound; Cost efficient; LARS; Lasso; Variable selection

Introduction

Several automatic variable selection and estimation techniques have emerged in the past two decades, including Lasso [1], LARS [2] and Adaptive Lasso [3]. The Lasso (which stands for "least absolute shrinkage and selection operator") is a popular technique for simultaneous variable selection and parameter estimation. It selects variables and estimates their coefficients by minimizing the residual sum of squares subject to a constraint to the sum of the absolute value of the coefficients. It shrinks some coefficients and sets the others to zero by the constraint, which adds a little bias but reduces the variance of the predicted values, thus improving the overall prediction accuracy [1]. Efron et al. [2] introduced Least Angle Regression, abbreviated LARS (the "S" suggesting "Lasso" and "Stagewise"). Both Lasso and Stagewise linear regressions are variants of LARS. A simple modification of the LARS algorithm implements the Lasso, but uses less computer time than the original Lasso algorithm. The key characteristic of LARS is its computational efficiency. Zou [3] proposed the adaptive Lasso, which adds weights in a data adaptive way to the Lasso penalty term. These weights provide less shrinkage to important predictors, thus leads to consistent variable selection results.

Although those methods have good performance in choosing statistically important factors, they do not take into account the cost of collecting data for the predictors. Identification of cost efficient diagnostic factors is of great interest to health researchers because of the heavy burden on the public health system. Due to the development and improvement of new technologies, such as nuclear medicine imaging and DNA microarray analysis, the costs of health care are escalating. In practice, inexpensive factors may have similar statistical significance as costly factors, thus could be used as diagnostic or prognostic variables by sacrificing minimal prediction accuracy, while reducing the health cost burden. This requires statisticians to search for new strategies in building statistical models to contain the effect of the cost of collecting data for diagnostic factors. The cost of collecting data for a variable may include the cost of material, equipment, time, human labor, etc. The costs may be different for collecting different variables. A model is more cost-efficient than another one if this model costs less, but with almost the same prediction accuracy, or this model costs much less but with only slightly less prediction power. A health researcher, as well as a decision maker, may prefer a more cost-efficient model in many situations. If there is a budget constraint on a research project or we are at the screening stage of diagnosing a disease, a more accurate but costly model may not be necessarily better than a less accurate but cheaper model.

There has been relatively little work on cost efficient variable selection. To incorporate cost in a predictive model, Lindley [4] suggested adding the cost of obtaining the covariates to the objective loss function in univariate multiple regression where a Bayesian approach was used. Brown et al. [5] worked on variable selection in multivariate linear regression using a non-conjugate Bayesian decision theory approach, where a terminal cost, a function of the cost of retaining the selected variables, was added to the loss function. Their approach balances prediction accuracy against costs and omits covariates when they cost too much relative to their predictive benefit.

Our goal is to develop a variable selection procedure that can simultaneously select the important predictors and estimate their effects to build a model that is not only good at prediction but also cost efficient. We concentrate on developing a method to select cost-efficient variables based on some existed variable selection algorithm. The cost effect is our focus and the developed algorithm can be adapted to a variety of variable selection methods. Since the LARS method is implemented in the R [6] package lars [7], and this package is publicly available, we can conveniently build our cost-efficient variable selection strategy, which extends the LARS method to incorporate variable costs penalized in the objective loss function. The total loss includes the error sum of squares, the Lasso type penalty, and the cost of collecting data for the predictors, where the first two parts compose the Lasso loss. It employs a branch and bound method to search for a model which minimizes total loss. The method is referred to as the Branching LARS (BLARS) search procedure in this paper.

***Corresponding author:** Li Hua Yue, Department of Statistical and Actuarial Sciences, Western University, London, ON N6A5B7, Canada, Tel: 1-519-661-2111; ext (86385); Fax: 1-519-661-3813; E-mail: lyue7@uwo.ca

Received October 15, 2013; **Accepted** November 07, 2013; **Published** November 13, 2013

Citation: Yue LH, He W, Murdoch D, Sendov H (2013) Building Cost-efficient Models using BLARS Method. J Biomet Biostat 4: 177. doi:10.4172/2155-6180.1000177

The rest of the paper is organized as follows. In the "Framework" section, we derive the theoretical basis of the BLARS method. We discuss details of the implementation in the section of "Issues in Implementation". In the "Numerical Studies" section, a simulation study is conducted to examine different ordering methods, and the proposed BLARS method is applied to a dataset from an Assertive Community Treatment (ACT) project conducted in Southwestern Ontario to demonstrate the cost-efficient variable selection process. Possible extension of the BLARS method is discussed in the "Discussion" section.

Framework

We want to select and simultaneously estimate the coefficients of covariates such that a loss function is minimized. The total loss in the loss function consists of 3 parts: the error sum of squares of the model, the l_1 penalty and the cost incurred by collecting those variables in the model. The proposed optimization problem P can be written as

$$min\ f(\boldsymbol{\beta},\boldsymbol{\alpha}) = \left\| \mathbf{y} - \sum_{j=1}^{p} \mathbf{x}_j \beta_j \right\|^2 + \lambda \sum_{j=1}^{p} |\beta_j| + n\gamma C(\alpha_1,\ldots,\alpha_p) \qquad (2.1)$$

with domain D:

$$\alpha_j \in \{0,1\}, \text{ for } j = 1,\ldots,p,$$

$$\boldsymbol{\beta} = (\beta_1,\ldots,\beta_p)' \in \Re^p,$$

and constraints S:

$$\alpha_j = 0 \Rightarrow \beta_j = 0, \text{ for } j = 1,\ldots,p,$$

where y is the n dimensional vector of observations, β is the regression coefficient vector to be estimated; p is the total number of covariates of interest; $\lambda \le 0$ is the regularization or tuning parameter; $\gamma \le 0$ is a user-defined weight imposed on costs, reflecting the level of reluctance to use high cost variables. The vector $\alpha=(\alpha_1,\ldots,\alpha_p)$ contains 0's and 1's, with $\alpha_j=1$ if the variable $\mathbf{X}_j$ is included in the model, as indicated in the constraints S. The cost function $C(\alpha_1,..,\alpha_p)$ is assumed to be non-decreasing in each α_j. For example, costs may accrue additively,

$$C(\alpha_1,\ldots,\alpha_p) = \sum_{j=1}^{p} \alpha_j c_j \qquad (2.2)$$

where $c_j \le 0$ is the cost of collecting the variable $\mathbf{X}_j$.

BLARS method

The sum of the first two terms in the objective function (2.1) is the Lasso objective function. The third term complicates the problem, but if we fix the value of α, then the third term becomes a constant and the problem reduces to Lasso variable selection and estimation, and *lars* may be used to solve it. A naive approach would be to try all 2^p different values of α, compare the results and select the best solution. In practice, this approach is not feasible when p is large. For example, we build a model to minimize the objective function (2.1) using the diabetes data used by Efron et al. [2], which contains 442 observations and 10 covariates: Age, Sex, BMI (body mass index), BP (average blood pressure), and S1 to S6 representing 6 serum measurements. For the purpose of illustration, we let the cost of Age and Sex be zero, let the cost of BMI and BP be 5 and 10, respectively, and let the 6 serum measurements have a group cost of 20 for the collection of blood sample and have additional individual cost of 30 for each blood test. Fixing λ=90 and γ=1, we use the naive approach to build the model with p=10, where 2^{10}=1,024 different results are compared to select the best solution. Using a computer with Intel Core6™ i7 CPU and 12GB memory and 64-bit R software environment, the computation time is 3.4 seconds. We then add 5 covariates into the design matrix: the squared term BMI^2 and the two-way interaction terms Age: Sex, Age: BP, BMI: BP and Age: S5. Fixing λ=90 and λ=1, with p=15, we need to compare 2^{15}=32,768 different results to select the best solution, and the computation time is increased to 145.2 seconds or 2.4 minutes. We further add in 5 covariates: $S3^2$, $S5^2$, SEX: BMI, SEX: BP, and AGE: S3. Still with λ=90 and γ=1, for p=20, we need to compare 2^{20}=1,048,576 different results to select the best one, and the computation time is dramatically increased to 6039 seconds or 1.7 hours. If we consider all squared terms and two-way interaction terms, we need to compare $2^{65}=3.69\times10^{19}$ different results, and the computation time cannot be imaginable, although p=65 is not a big number. The branch and bound search method can provide a solution to this problem, where relaxation is used to make the searching process easier and faster.

At each step in the BLARS process, we fix the value of one α_j to be 0 or 1. (The choice of j is discussed later; for simplicity in this discussion we will assume numerical order, fixing α_1 first, then α_2, etc.). At step 1, we branch on the problem P and create two subproblems: $P_{1\ (left)}$ with $\alpha_1=0$ and $P_{1\ (Right)}$ with $\alpha_1=1$. We continue to branch on the subproblems and create second-level subproblems by fixing $\alpha_2=0$ and $\alpha_2=1$, respectively. Suppose at some step k, we have fixed the value of $\alpha_1,\alpha_2,...,\alpha_k$, then the subproblem P_k of P has the objective function (2.1), the same domain D and constraints S, but with the given value of α_j, j=1,…,k. R_k is a relaxed problem of P_k with the same objective function (2.1), the same domain D and the same given value of of α_j, j=1,…,k, but constraints S_k:

$$\alpha_j = 0 \Rightarrow \beta_j = 0, \text{ for } j = 1,\ldots,k, \text{ with } 1 \le k \le p,$$

i.e. the only difference between R_k and P_k when $k<p$ is that we drop the constraints on β_j for $j>k$ for R_k. Therefore, the feasible region of R_k contains the feasible region of P_k, so the optimal objective value of the relaxed problem R_k will be a lower bound on the optimal objective value of the subproblem P_k. Without any constraints from $j=k+1$ to $j=p$, to minimize the total loss in R_k, we set all $\alpha_j=0$ for j=k+1,…,p. Then, the value of the vector α is known for the relaxed problem R_k, and R_k can be solved by calling the lars function. When $k=p$, R_k is the same as P_k which is the subproblem corresponding to a leaf node.

The branch and bound process makes use of the lower bound obtained from solving R_k to accelerate the search by avoiding solution of the generally harder problem P_k. Suppose some subproblems have been solved resulting in a best candidate solution found so far. If the optimal value of R_k, say v, is greater than or equal to the objective value of the best candidate solution found so far, then there is no need to solve P_k or branch on P_k since its optimal value cannot be better than v. Problem P_k is regarded as having been solved, even though it is not actually solved. In this case, the search tree is said to be "pruned" at P_k.

Note that for $l<k$ and the same fixed values of α_j, j=1,…,l, R_l is also a relaxation of R_k, so we may be able to prune certain relaxed problems to speed up the overall search even more. The detailed BLARS algorithm is shown in the Appendix.

For comparison to the example introduced at the beginning of this section, where the naive method is used to build the cost-efficient models on the diabetes data fixing λ=90 and γ=1, we apply the BLARS method developed in this paper to the 3 datesets with p=10, p=15, and p=20, respectively. The computation time is 0.04 seconds, 0.10 seconds and 0.13 seconds, respectively, and the results are exactly the same as the ones based on the native approach.

Pruning based on previous fits

The tuning parameter γ controls the importance of cost in the objective function. Often one wants to explore multiple values of γ to study the effect of cost. The following proposition allows the efficiency of the search.

Proposition 1: Given a fixed value of λ in the BLARS minimization procedure, the value of $C(\alpha_1,..,\alpha_p)$ in the optimal model is a non-increasing function of γ.

Proof: Suppose for a fixed value of λ, we selected an optimal BLARS model for γ_1 with the corresponding optimal values β_1 and α_1. The optimal total loss is

$$f(\boldsymbol{\beta}_1,\boldsymbol{\alpha}_1)=\left\|\mathbf{y}-\sum_{j=1}^{p}\mathbf{x}_j\beta_{1j}\right\|^2+\lambda\sum_{j=1}^{p}|\beta_{1j}|+n\gamma_1 C(\boldsymbol{\alpha}_1)\ .$$

Let $LassoLoss(\boldsymbol{\beta}_1)=\left\|\mathbf{y}-\sum_{j=1}^{p}\mathbf{x}_j\beta_{1j}\right\|^2+\lambda\sum_{j=1}^{p}|\beta_{1j}|$ and $C_1=nC(\boldsymbol{\alpha}_1)$, then

$$f(\boldsymbol{\beta}_1,\boldsymbol{\alpha}_1)=LassoLoss(\boldsymbol{\beta}_1)+\gamma_1 C_1.$$

Similarly, when we increased the γ value to γ_2, the optimal values have been changed to $\boldsymbol{\beta}_2$ and $\boldsymbol{\alpha}_2$. The optimal total loss is

$$f(\boldsymbol{\beta}_2,\boldsymbol{\alpha}_2)=LassoLoss(\boldsymbol{\beta}_2)+\gamma_2 C_2.$$

where $LassoLoss(\boldsymbol{\beta}_2)=\left\|\mathbf{y}-\sum_{j=1}^{p}\mathbf{x}_j\beta_{2j}\right\|+\lambda\sum_{j=1}^{2p}|\beta_{2j}|$ and $C_2=nC(\boldsymbol{\alpha}_2)$.

We want to prove that $nC(\boldsymbol{\alpha}_1)\geq nC(\boldsymbol{\alpha}_2)$, i.e. $C_1\geq C_2$. Now we assume that $C_1<C_2$. Recall that the optimal BLARS solution can be regarded as the best one among the 2^p different results corresponding to the 2^p different α values. Thus, for $\gamma=\gamma_1$, we must have

$$LassoLoss(\boldsymbol{\beta}_1)+\gamma_1 C_1\leq LassoLoss(\boldsymbol{\beta}_2)+\gamma_1 C_2.$$

Equivalently,

$$LassoLoss(\boldsymbol{\beta}_1)-LassoLoss(\boldsymbol{\beta}_2)\leq\gamma_1(C_2-C_1) \qquad (2.3)$$

Similarly for $\gamma=\gamma_2$, we must have

$$LassoLoss(\boldsymbol{\beta}_2)+\gamma_2 C_2\leq LassoLoss(\boldsymbol{\beta}_1)+\gamma_2 C_1\ .$$

Equivalently,

$$LassoLoss(\boldsymbol{\beta}_1)-LassoLoss(\boldsymbol{\beta}_2)\geq\gamma_2(C_2-C_1) \qquad (2.4)$$

Since $C_1<C_2$ and $\gamma_2>\gamma_1$, we have $\gamma_2(C_2-C_1)>\gamma_1(C_2-C_1)$, and the inequalities (2.3) and (2.4) cannot hold simultaneously. Thus, the initial assumption of $C_1<C_2$ must be false, and we conclude that $C_1\geq C_2$, i.e. $nC(\boldsymbol{\alpha}_1)\geq nC(\boldsymbol{\alpha}_2)$.

Based on Proposition 1, we may prune a branch if the value of C of this branch is larger than the one in the optimal model for a smaller γ.

Issues in Implementation

Cost structure

The cost of collecting a variable may include the cost of material, equipment, time, human labor, etc. One way to assign a cost would be to use the dollar amount we have to pay to get that variable; a more sophisticated analysis might include both the monetary cost and the level of difficulty to collect the data.

The simplest cost structure is the additive cost (2.2), in which the total cost of obtaining data for a selected set of variables is the sum of the cost of getting data for each variable in the set. This cost structure applies to situations where the data for the variable are collected individually and independently. More generally, the cost structure can be non-additive, as there may be grouping effects. Grouping effects occur when selection of one variable causes other variables to decrease in cost. For example, the cost of collecting several blood test results for one patient may include a group cost of getting the blood sample and several additional costs for different blood tests. If one test result is selected into a statistical model, the other test results become cheaper if they are also selected, since we only need to count the group cost once. Suppose we can get two blood test results simultaneously from one test, then when one of them is selected into a statistical model, the other one becomes free. Another grouping cost may come from the situation where higher order or interaction terms are considered in a model. These terms become free once the variables involved in the terms have been selected.

We could treat additive cost as a special case of non-additive cost with all group costs being zero. In BLARS, we deal with non-additive cost by updating the cost of each of the undetermined variables (the variables that have not entered the search process) after each step based on which variables have been selected into the model.

The order of covariates in selection

The order of the variables entering the searching process is an important factor affecting the efficiency of the algorithm. Earlier pruning will avoid searching more paths, resulting less *lars* calls during the searching process.

Intuition suggests several possible orderings in which the variables should enter the search. We could use the order of the LARS entries. The covariate which is most highly correlated with the response is added first and less correlated covariates are added later. Alternatively, we could order the variables by their costs. If we let the most correlated covariate enter the BLARS searching process first, the Lasso loss (the first two terms in the objective function) may decrease dramatically, and the tree is more likely to be pruned at the node where we force this variable out of the model, i.e. the node where we let $\alpha_1=0$. Using this ordering method, the computing time may be reduced because the tree has more chance to be pruned at upper level left-path nodes. On the other hand, when the cost difference of the predictors is large (usually associated with a higher value of γ), the cost effect may dominate. Ordering variables by descending order of the costs could be a better approach in this case. If we let the most expensive covariate enter the BLARS searching process first, the gain by the decrease of the Lasso loss may be clearly surpassed by the increase of the cost, and the tree is more likely to be pruned at the node where we force this variable in the model, i.e. the node where we let $\alpha_1=1$. Using this ordering method, the computing time may be reduced because the tree has more chance to be pruned at upper level right-path nodes.

Our approach is to combine the *LARS* with the *COST* ordering method to make the search process more efficient. First, we divide the costs of potential predictors into bins. Each bin covers a range of costs defined as a multiple s of the observed variance of the responses:

$$B=s\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2, \qquad (3.1)$$

Through a simulation study described later, we found that the results are reasonably good when we set $s=10\gamma/[\log(1+\gamma)\log(1+\lambda)]$, where γ and λ are the tuning parameters in (2.1). The incremental cost

predictor $\mathbf{X}_j$ will be c_j (fixed for additive costs, varying depending on what is already in the model in the general case). This cost will fall into one bin $kB \le c_j < (k+1)B$, where k ≥ 0 is an integer and j=1,…,p, as shown in Figure 1. We order variables in different bins by the COST method, and order the variables in the same bin by the LARS method. Thus, for the case in Figure 1, x_5 and x_6 are the first two variables entering the BLARS search process since they have the highest costs, but which one enters first depends on the LARS entry order. The variables in the lowest cost bin, such as x_1, x_2, x_3 in Figure 1 are the last ones entering the BLARS search process. Note that the variables with zero cost require no search at all, so may always be placed last.

Note that with non-additive costs, each time after we update the costs for the undetermined variables, we may need to reorder them based on their new costs.

Tuning parameter and model selection criteria

A fast effective way of selecting the tuning parameter λ is another important issue in practice. The selection criteria in the literature include C_p, AIC, BIC, and Cross-validation [8]. Efron et al. [2] suggested selecting the tuning parameter and the optimal model based on C_p. Others claimed that AIC is asymptotically valid if no fixed-dimension correct model exists while BIC is preferred if there exist fixed-dimension correct models [9,10]. Zou et al. [11] proved without any special assumption on the predictors that the number of nonzero coefficients is an unbiased estimate for the degrees of freedom of the Lasso. The authors discussed C_p, AIC and BIC model selection criteria and suggested using BIC for the Lasso as the model selection criteria, when the sparsity of the model is the major concern. BIC for the Lasso can be written as

$$BIC(\hat{\boldsymbol{\mu}}) = \frac{\|\mathbf{y} - \hat{\boldsymbol{\mu}}\|^2}{n\sigma^2} + \frac{log(n)}{n}\widehat{df}(\hat{\boldsymbol{\mu}}) \tag{3.2}$$

where $\widehat{df}(\hat{\boldsymbol{\mu}})$ equals the number of nonzero coefficients.

The parameter γ is a user-defined weight imposed on costs, reflecting the level of reluctance to use high cost variables. When γ=0, we ignore the costs and selection becomes the standard Lasso variable selection. The higher the γ value, the more reluctant is the user to select high cost variables. Thus, when the user assigns a higher value to γ, the BLARS process will be less likely to select higher cost variables. The assignment of a γ value is thus based to a large extent on the opinions and judgments of the user or the decision maker. Sometimes, the user has to use a higher γ because of budget constraints. Once γ is fixed, the optimal value of λ and the corresponding optimal statistical model could be selected by a chosen model selection criterion. Note that LARS builds up estimates in successive steps, each step adding one covariate to the model, until all covariates are added [2]. The LARS result shows which variable enters the model at each step with the corresponding λ value, starting from the largest λ at the first step and ending to the smallest λ at the last step. Since our BLARS procedure calls *lars* function, the possible values of the λ from an initial *lars* call provide us a reasonable range of the tuning parameter λ of the BLARS procedure. A golden section search approach [12] can be implemented to choose the optimal λ value given a model selection criterion and a fixed γ, for example, the optimal λ could be selected as the one that gives a model with minimum BIC value when using BIC as the model selection criterion. In practice, we can start from a small value of γ, which usually gives the same result as a Lasso model where cost effect is ignored, and then we get a group of BLARS models when we gradually increase the value of γ and costly variables are gradually excluded. The percentage increase in Error Sum of Squares (SSE) is compared with the percentage decrease in cost of the group of BLARS models, and the user can select their preferred cost-efficient one that sacrificing minimal prediction accuracy, i.e. sacrificing a user selected amount of SSE increment that surpassed by the gain in cost reduction.

In the following ACT data analysis, we use both C_p and BIC for the Lasso as the tuning parameter and model selection criterion. We use C_p because it is the default selection criterion in the R package lars, and we use BIC for the Lasso as the selection criterion for its simplicity and effectiveness.

Numerical Studies

Simulation

The order of the variables entering the BLARS, searching process is an important factor affecting the efficiency of the algorithm, and we propose the *Bin* ordering method in Section 3.2. To compare this ordering method with other potential candidate methods, we conduct a simulation study. Another objective of the simulation study is to investigate a suitable scalar s in the Equation (3.1) for calculating the bin.

In the simulation study, we compare 7 ordering methods by assessing the number of calls to the *lars* function in the BLARS searching process. The 7 ordering methods are to order the potential covariates in descending order of the correlations with the updated response, i.e. the order of the LARS entries (*LARSd*), ascending order of the correlations (*LARSa*), descending order of the costs (*COSTd*), ascending order of the costs (*COSTa*), descending order of the absolute value of the OLS estimates (*OLSd*), ascending order of the absolute value of the OLS estimates (*OLSa*), and combined order of *LARSd* with *COSTd* (*Bin*). We change the order of the covariates at the beginning of the searching process, and once when using the order of *COSTd, COSTa, OLSd* or *OLSa*. For the order of *LARSd, LARSa* or *Bin*, we change the order of the covariates based on the *lars* calls during the searching process.

The data are simulated based on the diabetes data used by Efron et al. [2], where they have 10 covariates: Age, Sex, BMI, BP, and S1 to S6. For example, we simulate 1000 observations of BMI from the 442 observations of BMI in the diabetes data by random sampling with replacement. We choose 5 models in the simulation study. There are 10 potential predictors in each of the first 4 models as in the diabetes data, whereas there are 11 potential predictors in the last model.

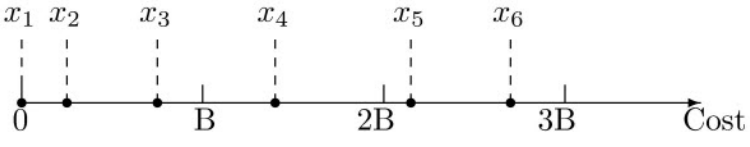

Figure 1: The bin of the costs.

The first model contains only one true predictor (BMI), which is the first one of the LARS entries when applying the *lars* function on the original diabetes data. Similarly based on the LARS entries, we choose 3, 5, and 7 true predictors in the second, third, and fourth model, respectively. The fifth model is the same as the third one with 5 true predictors, except that we add in a fake predictor called FS5 which has a correlation around 0.8 with S5 (the second one of the LARS entries), but costs much less than S5.

We simulate a dataset for each model and apply the BLARS algorithm with different ordering methods to the dataset for different combinations of λ and γ values. There are two non-additive cost structures used in the simulation study. The first one contains a small group cost and 6 large additional individual costs for the 6 blood test results (S1 to S6); and the second one contains a large group cost and 6 small additional individual costs for the 6 blood test results. Table 1 shows the details of the costs, where the cost of FS5 only applies to model 5.

When using the *Bin* ordering method, we also change the scalar s within a broad range in calculating the bin value, where s can be a fixed number, a function of γ, a function of λ or both. We found that s could relate to λ by a function $1/\log(1+\lambda)$, or relate to γ by a function $\gamma/\log(1+\gamma)$ through preliminary simulation studies. Then, a thorough comparison are made where 82 different s values are under consideration: fixed numbers (0.1,0.5,1,2,…,20), function of $\lambda\left(\frac{k}{\log(1+\lambda)}, k=1,2,\ldots,20\right)$ function of $\gamma\left(\frac{k\cdot\gamma}{\log(1+\gamma)}, k=1,2,\ldots,20\right)$ and function of both λ and γ $\left(\frac{k\cdot\gamma}{\log(1+\gamma)\log(1+\lambda)}, k=1,2,\ldots,20\right)$. We repeat this process for 100 times for model 1 and 50 times for each of the other 4 models, since more combinations of λ and γ are used there. The *Bin* ordering method shows promising result for a range of s values based on the simulation results, with different range of *s* for different situations. We emphasize that the ordering method only affects the efficiency of the BLARS algorithm; it does not affect the finally chosen model, which have been confirmed by the simulation results. Therefore, a relatively suitable s value is chosen as $\frac{10\gamma}{\log(1+\gamma)\log(1+\lambda)}$, which has the best overall result. Note that many other values of *s* give almost as good overall results as the chosen one, such as $\frac{j\cdot\gamma}{\log(1+\gamma)\log(1+\lambda)}$ for j=11,…,16 and $\frac{k}{\log(1+\lambda)}$ for k=15,…,18. Even other values of s do not give much worse results, and still make the *Bin* ordering method superior to other ordering methods.

We compare the times of *lars* function calls by the 7 ordering methods during the searching process. There are 100 replicate datasets simulated for model 1 with 8 combinations of cost, λ and γ, leading to 800 comparisons of the 7 ordering methods. The *Bin* ordering method is the fastest for 790 out of 800 simulations. Table 2 shows typical results for one simulated dataset. Similarly, with 18 combinations of cost, λ and γ, the *Bin* ordering method is the fastest for 900 out of 900 times for both model 2 and model 3. Table 3 presents typical results for one simulated dataset using model 3. With 24 combinations of cost, λ and γ in model 4, the *Bin* ordering method is the fastest for 1200 out of 1200 times. The *Bin* ordering method is the fastest for 889 out of 900 times using model 5, where a fake covariate is added.

Cost	Structure	Age	Sex	BMI	BP	S1	S2	S3	S4	S5	S6	FS5
	Group					80						
1	Additional	0	0	20	40	120	120	120	120	120	120	20
	Group					170						
2	Additional	0	0	20	40	30	30	30	30	30	30	20

Table 1: Non-additive cost structure used in simulation study.

			Times to Call *lars* Function						
Cost	λ	γ	LARSa	LARSd	COSTa	COSTd	OLSa	OLSd	Bin
1	1	0.3	147	48	52	88	147	49	32
	1	1.0	147	54	57	112	147	53	36
	3	0.3	24	17	36	12	26	18	9
	3	1.0	24	27	29	14	26	28	10
2	1	0.3	167	54	52	105	163	56	36
	1	1.0	159	60	57	117	156	60	40
	3	0.3	24	17	36	12	26	18	9
	3	1.0	24	27	29	14	26	28	10

Table 2: Comparison of 7 ordering methods using simulation model 1. Model 1 contains 10 potential predictors, and only one is the true predictor.

			Times to Call *lars* Function						
Cost	λ	γ	LARSa	LARSd	COSTa	COSTd	OLSa	OLSd	Bin
1	1	0.3	338	16	75	61	338	16	14
	1	1.0	431	23	100	76	431	25	19
	1	5.0	424	48	102	144	424	56	36
	3	0.3	247	14	67	41	247	14	12
	3	1.0	325	18	93	55	325	20	14
	3	5.0	357	42	94	106	357	50	30
	10	0.3	134	15	66	23	134	15	9
	10	1.0	202	14	81	38	202	16	10
	10	5.0	265	45	87	72	265	53	30
2	1	0.3	268	17	55	51	268	17	15
	1	1.0	396	17	89	59	396	17	15
	1	5.0	572	48	100	136	572	56	36
	3	0.3	229	14	49	36	229	14	12
	3	1.0	321	12	83	42	321	12	10
	3	5.0	513	45	94	118	513	53	33
	10	0.3	170	8	39	22	170	8	6
	10	1.0	230	8	73	24	230	8	6
	10	5.0	423	45	88	99	423	53	30

Table 3: Comparison of 7 ordering methods using simulation model 3. Model 3 contains 10 potential predictors, and there are three true predictors.

In model 5, S5 is one of the true predictors and FS5 is a fake covariate which is highly correlated with S5, but with much less cost (Table 1). S5 is selected into the BLARS model when we choose a small γ value, but FS5 is selected instead of S5 due to the cost effect when we increase the γ gradually. Choosing one simulated dataset, using the first cost structure, and fixing λ=10 and γ=1, we compare the *LARSd, OLSd, COSTd* and *Bin* ordering method by drawing the search trees in Figure 2. In Figure 2, the black path is the optimal path. The search trees show the difference in the order of covariate entering the searching process, resulting in different pruning of the trees and indicating the best result for the tree associated with the *Bin* ordering method.

ACT data analysis

A study was conducted in Southwestern Ontario to assess factors which would influence the outcomes of clients with severe mental illness (SMI) receiving care from the Assertive Community Treatment (ACT) [13] service. The patients recruited in the study were diagnosed

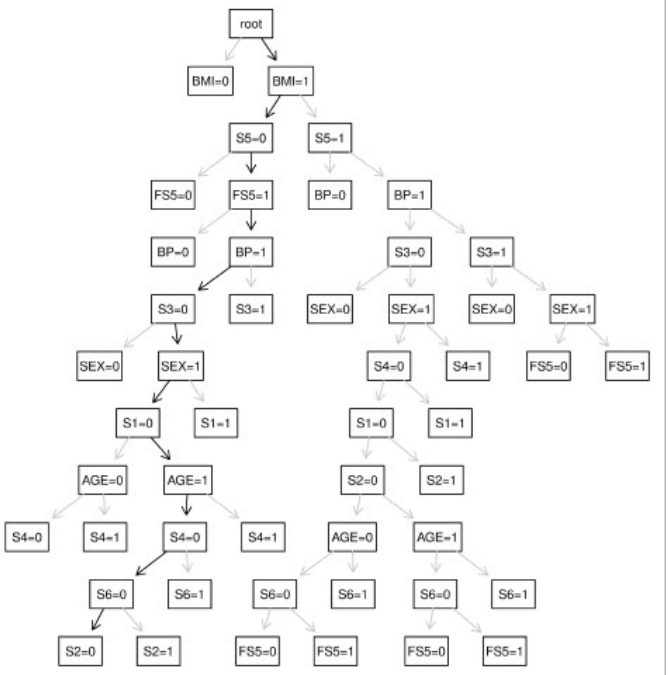

Figure 2: Search trees for simulation model 5 with λ=10 and γ=1 using different ordering methods.
Model 5 contains 11 potential predictors, and there are five true predictors including S5 which is highly correlated with a fake predictor FS5 but costs much more.
(a): Using *LARSd*.

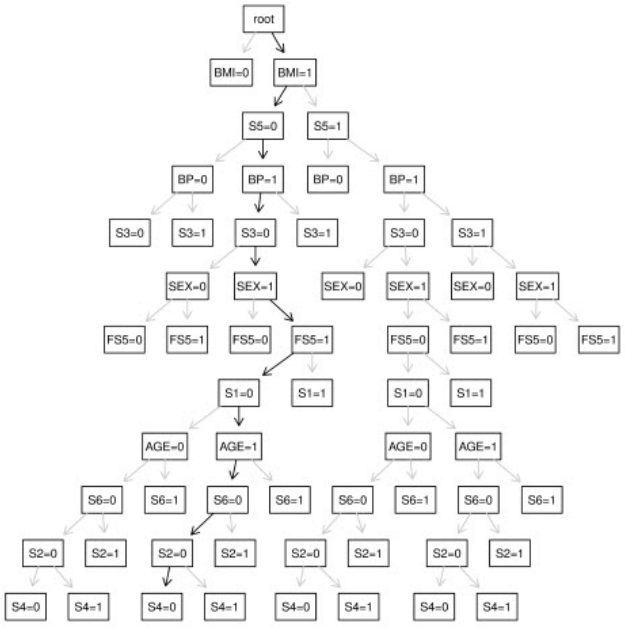

Figure 2: Search trees for simulation model 5 with λ=10 and γ=1 using different ordering methods.
Model 5 contains 11 potential predictors, and there are five true predictors including S5 which is highly correlated with a fake predictor FS5 but costs much more.
(b): Using *OLSd*.

as having psychosis or multiple co-morbid psychiatric and physical disorders, as well as a history of high hospital use, long-term illness, high needs and low functioning. There were about 19 potential predictive factors. Table 4 presents the names and descriptions of the variables used in the data analysis of the ACT project. Long term outcome was the overall Colorado Client Assessment Record (CCAR) score revised for use in Southwestern Ontario [14], which is the overall degree of problem severity (a larger score associates with a higher level of problem severity), and was measured at 12 and 24 months after enrollment in the project.

Our goal in this study was to assess what cost-efficient factors influence outcomes of clients with SMI receiving care from ACT. We wanted to find the risk factors not only with higher prediction accuracy, but also cheaper and easier to collect the data, so that we can reduce the burden of the ACT teams and the patients.

Cost structure

Since the sources of data collection were different, the costs of collecting data were different for the potential predictors. In the ACT project, data were collected from the following sources: client self-reports, ACT clinicians, client records, hospital archives, ACT team's staff activity records and ACT coordinators. The data that involved the professional work of clinicians cost more than the data from the work of research assistants, while the client self-reported data were harder to obtain than the data extracted from hospital archives due to the fact that the clients were having severe mental illness.

The cost of collecting the data had two components in the ACT project. The first was the monetary cost for human labor, time, material, equipment, compensation paid to the clients in some research activities, etc. The second was the level of difficulty to get an answer or a value for a potential predictor. For example, since the clients we dealt with were the patients with severe mental illness, they might refuse to provide some information and some results reported from the clients might need to be double checked or traced. This resulted in some variables being more "expensive" than others. We also needed to take into account the grouping effects of cost for both of the two components.

The two components of costs of the potential predictors were estimated between 0 and 100 by the ACT project researcher and coordinator and are listed in Table 5, where both monetary cost and level of difficulty consist of two parts: group cost and additional individual cost. We considered an overall cost for each predictive factor, which was a combination of the above two components. One predictor cost more than another if this predictor was more expensive overall. Since the scales of the two components were comparable (with minimum 0 and maximum 100), one simple way to combine them was to use summation. For convenience, we divided the combined costs by 200, which are also displayed in Table 5.

Cost-efficient variable selection

We applied the BLARS method to the ACT data to select cost-efficient variables and estimate their effects. First, we used BIC for the Lasso (Equation 3.2) as the tuning parameter and model selection criterion. When we assigned 0.1 to γ, there were 4 predictors selected into the BLARS model: number of months in ACT, average number of contacts per month, CCAR substance use subscale and CCAR functioning subscale. The same 4 variables were selected using the Lasso model (γ=0). When γ was increased to 0.2, 3 predictors remained in the BLARS model, where average number of contacts per month was dropped out. When γ was increased to 0.5, only number of months in

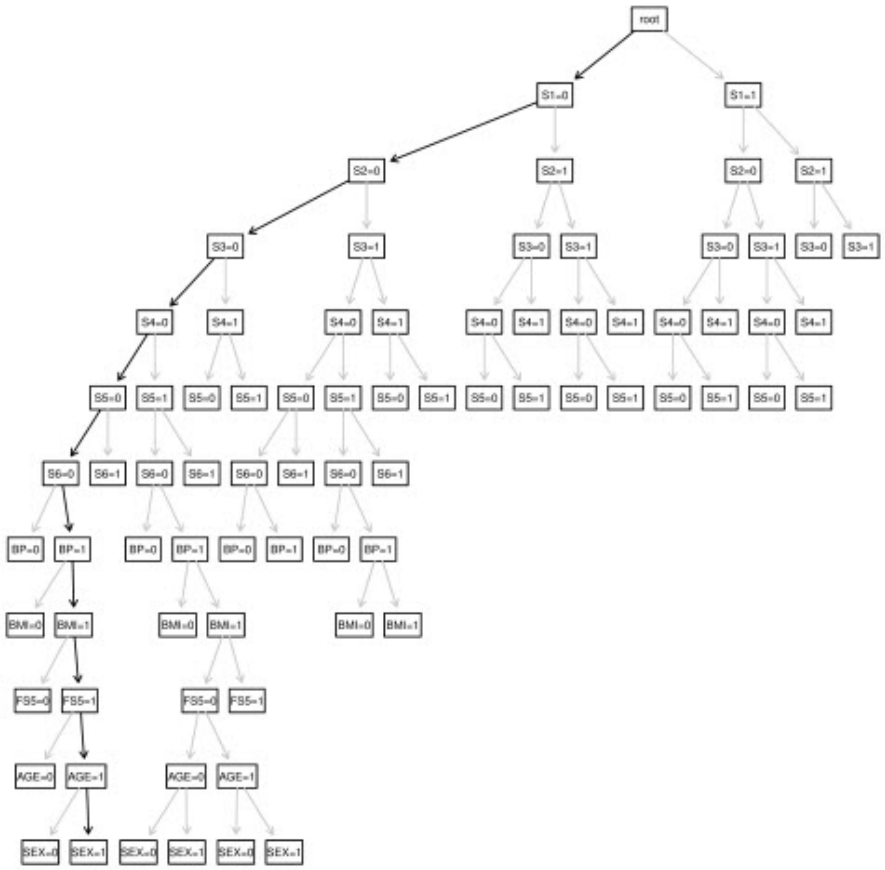

Figure 2: Search trees for simulation model 5 with λ=10 and γ=1 using different ordering methods. Model 5 contains 11 potential predictors, and there are five true predictors including S5 which is highly correlated with a fake predictor FS5 but costs much more.
(c): Using *COSTd*.

ACT remained in the model. For γ=1.0, no variable was selected in the BLARS model due to the cost effect and the best prediction in this case was the grand mean of the response. The Lasso model and BLARS models for different γ values are shown in Table 6, where some non-selected variables are not displayed. Table 7 gives the components in objective functions including SSE, l_1 penalty and cost penalty of the corresponding models, where the percentage increases or decreases are compared with the first BLARS model (γ=0.1). When we choose a small value of γ, as in the case of γ=0.1, the BLARS model select the same covariates as the Lasso model, although the estimated coefficients are slightly different; the SSE of the BLARS model is smaller than the SSE of the Lasso model. Second, we used C_p as the tuning parameter and model selection criterion. Table 8 presents Lasso model and BLARS models for different γ values and Table 9 displays the components in the objective functions of the corresponding models. When we choose a small value of γ, as in the case of γ=0.01, the BLARS result is exactly the same as the Lasso result, with the same estimated coefficients and the same SSE.

Compared with models using C_p as the model selection criterion, models selected by BIC were much more parsimonious for small values of γ. However, when γ was larger (γ>0.1), BLARS results were similar, regardless of which model selection criterion was used (Table 9).

The value of γ is user-defined and the selection criteria of tuning parameter and model selection are also user's choice. The health researchers or decision makers should make overall judgments based on the percentage increase of the error sum of squares and the percentage decrease of the cost to choose their preferred cost-efficient model from the BLARS results.

Discussion

We developed a cost-efficient variable selection method based on the LARS technique with focus on the cost effect. The proposed BLARS algorithm can be generalized by replacing the Lasso loss (the first two terms in Equation (2.1)) with other objective functions to incorporate the cost effect whenever we have a method to solve that minimization problem. For example, if we adjust the l_1 penalty (the second term in Equation (2.1)) by adaptive weights to penalize different coefficients, we obtain Adaptive Lasso type object function. The same efficient algorithm (LARS) for solving the Lasso can be employed to solve the problem by using a transformation to the design matrix [3]. Thus, our BLARS procedure can be easily adjusted to an Adaptive Lasso type cost-efficient variable selection method. Recently Friedman et al. [15] proposed new fast algorithms for regression estimation, which are based on cyclical coordinate descent methods. Their methods are a remarkably fast approach for solving convex problems with l_1 (the Lasso) penalty or l_2 (the ridge-regression) penalty, or mixtures

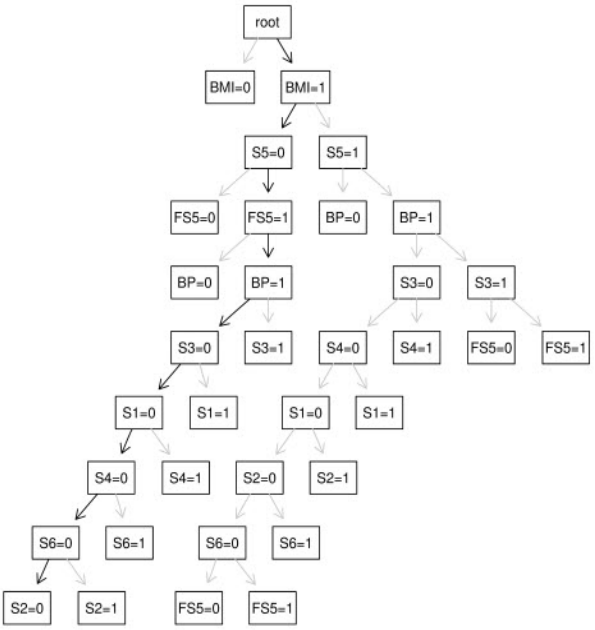

Figure 2: Search trees for simulation model 5 with λ=10 and γ=1 using different ordering methods.
Model 5 contains 11 potential predictors, and there are five true predictors including S5 which is highly correlated with a fake predictor FS5 but costs much more.
(d): Using *Bin*.

Predictors	Description
Age	Age in years
Sex	1: Female ; 0: Male
Mstatus	Marital Status, 1: Married or Common-law; 0: Otherwise
CoMorbid	Number of co-morbid diagnoses
Duration	Number of years since first diagnosis
Lifetime	Lifetime days in hospital
Jail	Ever in jail, 1: No; 0: Yes
EmpSC	CCAR employment subscale. 1: Employed (full-time or part-time); 0: Otherwise
SubSC	CCAR substance use subscale. A larger score associates with a higher level of substance abuse.
FunSC	CCAR functioning subscale. A larger score associates with a lower level of functioning.
ACTmo	Total service use: number of months in ACT
Medtype	Medications prescribed: number of medication categories
Contacts	Intensity of contacts: average number of contacts per month by ACT staff
DACTS	Fidelity of team to ACT model: Dartmouth ACT Scale
	A larger score associates with a higher level of fidelity to ACT model
WAI	Therapeutic alliance: Working Alliance Inventory
	A larger score associates with a higher level of alliance between patient and therapist
PSE	Insight into psychosis: Present State Exam-insight score. 1: Limited insight; 0: Full insight
EMP	Empowerment scale
	A higher score associates with a higher level of client's participation in their recovery
DAI	Satisfaction with medications: Drug Attitude Inventory
	A higher score associates with a higher level of client's satisfaction with medication
MEDC	Medication compliance: Adherence to medication scale
	A higher score associates with a lower level of client's adherence to medication

Table 4: Potential predictive factors in ACT project.

	Monetary Cost		Level of Difficulty		Overall Cost	
Predictors	**Group**	**Additional**	**Group**	**Additional**	**Group**	**Additional**
Age	15	0	10	0	0.125	0
Sex		0		0		0
Mstatus		0		0		0
CoMorbid		0		0		0
Duration		10		0		0.1
ACTmo		0		0		0
Medtype		0		0		0
Contacts	25	0	20	0	0.225	0
Jail	20	0	30	0	0.25	0
MEDC	20	0	30	0	0.25	0
WAI	30	0	30	0	0.3	0
PSE		0		0		0
EMP		0		0		0
DAI		0		0		0
Lifetime	30	0	70	0	0.5	0
EmpSC	30	20	50	20	0.4	0.2
SubSC		20		20		0.2
FunSC		20		20		0.2
DACTS	60	0	100	0	0.8	0

Table 5: Two cost components and the overall costs used in ACT data analysis.

BLARS(k, BestLoss, PreSolution, PreBound, $\boldsymbol{\alpha}^+$) :
  Solve $R_{k(right)}$:
    Solution$_{(right)} \leftarrow$ PreSolution
    Bound$_{(right)} \leftarrow$ PreBound $+ \gamma c_k$.
    If Bound$_{(right)} >$ BestLoss, then
      Solution$_{(right)} \leftarrow$ "pruned"
      Loss$_{(right)} \leftarrow \infty$.
      $\boldsymbol{\alpha} \leftarrow \boldsymbol{\alpha}^+$
    Else:
      Loss$_{(right)} \leftarrow$ Bound$_{(right)} + \gamma \sum_{j>k;\beta_j \neq 0} c_j$.
      If Loss$_{(right)} <$ BestLoss, then
        BestLoss $\leftarrow$ Loss$_{(right)}$
      If $k < p$, then
        (Solution$_{(right)}$, Loss$_{(right)}$)
          $\leftarrow$ BLARS($k+1$, BestLoss, Solution$_{(right)}$, Bound$_{(right)}$, $\boldsymbol{\alpha}^+$)
      Compute $\boldsymbol{\alpha}$ from Solution$_{(right)}$.
  Solve $R_{k(left)}$:
    If $\alpha_k = 0$, then
      Solution$_{(left)} \leftarrow$ Solution$_{(right)}$
      Loss$_{(left)} \leftarrow$ Loss$_{(right)}$
    Else:
      $\alpha_k^+ \leftarrow 0$ and select design matrix based on $\boldsymbol{\alpha}^+$.
      Call *lars* to get Solution$_{(left)}$, Bound$_{(left)}$ and $\boldsymbol{\alpha}$.
      If Bound$_{(left)} >$ BestLoss, then
        Solution$_{(left)} \leftarrow$ "pruned"
        Loss$_{(left)} \leftarrow \infty$.
      Else:
        Loss$_{(left)} \leftarrow$ Bound$_{(left)} + \gamma \sum_{j>k;\beta_j \neq 0} c_j$.
        If Loss$_{(left)} <$ BestLoss, then
          BestLoss $\leftarrow$ Loss$_{(left)}$
        If $k < p$, then
          (Solution$_{(left)}$, Loss$_{(left)}$)
            $\leftarrow$ BLARS($k+1$, BestLoss, Solution$_{(left)}$, Bound$_{(left)}$, $\boldsymbol{\alpha}^+$)
  If Loss$_{(right)} <$ Loss$_{(left)}$, then
    Return (Solution$_{(right)}$, Loss$_{(right)}$)
  Else:
    Return (Solution$_{(left)}$, Loss$_{(left)}$)

Figure 3: The recursive step of the BLARS algorithm.

Solve R_0 to find Solution$_0$ and Bound$_0$.
Loss$_0 \leftarrow$ Bound$_0 + \gamma \sum_{j;\beta_j \neq 0} c_j$.
Call BLARS($k = 1$, BestLoss $=$ Loss$_0$, PreSolution $=$ Solution$_0$,
  PreBound $=$ Bound$_0$, $\boldsymbol{\alpha} = (1, \ldots, 1)$) and return the result.

Figure 4: The initialization step of the BLARS algorithm.

of the two (the elastic-net penalty). Since these alternatives are well developed, they can be adapted to the node-level in our cost efficient variable searching approach, but unfortunately they are not directly applicable to minimizing the full problem (2.1), which is not convex.

We illustrated the cost-efficient variable selection procedure in

		Estimated Coefficients										
Method	γ	**ACTmo**	**Contacts**	**Jail**	**MEDC**	**WAI**	**EMP**	**DAI**	**EmpSC**	**SubSC**	**FunSC**	**DACTS**
Lasso	-	-0.006	0.0028	0	0	0	0	0	0	0.07	0.16	0
BLARS	0.10	-0.011	0.0074	0	0	0	0	0	0	0.11	0.20	0
	0.20	-0.010	0	0	0	0	0	0	0	0.10	0.21	0
	0.50	-0.015	0	0	0	0	0	0	0	0	0	0
	1.00	0	0	0	0	0	0	0	0	0	0	0

Table 6: Optimal Lasso and BLARS models for different γ values using BIC as the model selection criterion.

Method	γ	**Total Loss**	l_1 **Penalty**	**SSE**	**SSE Increase**	**Cost Penalty**	**Cost (per patient)**	**Cost Decrease**
Lasso	-	348	17	331	-	-	-	-
BLARS	0.1	348	9	317	-	22	1.150	-
	0.2	370	10	325	2.5%	35	0.925	19.6%
	0.5	384	4	368	15.8%	12	0.125	89.1%
	1.0	388	0	388	22.2%	0	0.000	100.0%

Table 7: Components in objective functions for different γ values using BIC as the model selection criterion.
The percentage increase or decrease are compared with the first BLARS model (γ=0.1). Lasso model was fitted without considering cost effect, and the total loss has only two components.

		Estimated Coefficients										
Method	γ	**ACTmo**	**Contacts**	**Jail**	**MEDC**	**WAI**	**EMP**	**DAI**	**EmpSC**	**SubSC**	**FunSC**	**DACTS**
Lasso	-	-0.014	0.0070	-0.11	0.086	-0.057	-0.32	-0.026	0.17	0.096	0.18	0.33
BLARS	0.01	-0.014	0.0070	-0.11	0.086	-0.057	-0.32	-0.026	0.17	0.096	0.18	0.33
	0.02	-0.012	0.0069	-0.15	0.091	-0.060	-0.35	-0.046	0.17	0.090	0.18	0
	0.04	-0.012	0.0067	0	0	-0.055	-0.29	-0.070	0.17	0.100	0.19	0
	0.10	-0.011	0.0074	0	0	0	0	0	0	0.112	0.20	0
	0.20	-0.010	0	0	0	0	0	0	0	0.101	0.21	0
	0.50	-0.015	0	0	0	0	0	0	0	0	0	0
	1.00	0	0	0	0	0	0	0	0	0	0	0

Table 8: Optimal Lasso and BLARS models for different γ values using Cp as the model selection criterion.

Method	γ	**Total Loss**	l_1 **Penalty**	**SSE**	**SSE Increase**	**Cost Penalty**	**Cost (per patient)**	**Cost Decrease**
Lasso	-	316	15	301	-	-	-	-
BLARS	0.01	322	15	301	-	7	2.950	-
	0.02	325	14	303	0.9%	8	2.150	27.1%
	0.04	335	14	308	2.5%	13	1.650	44.1%
	0.10	348	9	317	5.5%	22	1.150	61.0%
	0.20	370	10	325	8.2%	35	0.925	68.6%
	0.50	384	4	368	22.2%	12	0.125	95.8%
	1.00	388	0	388	28.9%	0	0.000	100.0%

Table 9: Components in objective functions for different γ values using *Cp* as the model selection criterion. The percentage increase or decrease are compared with the first BLARS model (γ=0.01). Lasso model was fitted without considering cost effect, and the total loss has only two components.

this paper with either BIC or C_p as the turning parameter and model selection criteria. There is a lot of controversy on which criterion is the best, and it seems that no one surpasses others in all situations. Researchers may have their preferred selection criteria other than BIC or C_p, and they have to make the judgment based on their own experience. But the BLARS algorithm is the same, regardless which model selection criterion is used.

We considered two cost components, monetary cost and level of difficulty, in the ACT data analysis. Because the two components were estimated in the same scale, we used the combined overall costs in the data analysis. In general cases, the two cost components may not be in the same scale, therefore, it may be better to consider them separately by using two cost terms in Equation (2.1) with two user-defined weights γ_1 and γ_2, and it will give researchers more flexibility to balance between the two kind of costs.

Acknowledgements

The first author was supported by a scholarship from the Natural Sciences and Engineering Research Council of Canada (NSERC); the other authors are all supported by grants from NSERC.

Appendix

The detailed BLARS algorithm can be described as follows.

At step *k*, we let α^+ be equal to $(\alpha_1,\ldots,\alpha_k,1,\ldots,1)$. This vector indicates which variables are passed to lars for optimization. Once we have the lars result in hand, we set each of $\alpha_k+1,\ldots,\alpha_p$ to 0 if the corresponding β_k is zero and 1 otherwise. This gives $\boldsymbol{\alpha}$ to use in the cost calculations. We also calculate $\boldsymbol{\alpha}^-$ as $(\alpha_1,\ldots,\alpha_k,0,\ldots,0)$ to use in the cost calculations for the bound.

In the algorithm below, we use the following notation. *lars* refers to the R package or the *lars* function in that package. Our own variables and functions will be written in small caps, e.g. SOLUTION below. For $0 \le k \le p$, the solution of a relaxation R_k is denoted by $SOLUTION_k$; the corresponding objective value uses $\boldsymbol{\alpha}^-$ to give the lower bound for P_k, and is referred to as $BOUND_k$. (We suppress the dependence on $\boldsymbol{\alpha}^-$, but in fact there are potentially 2^k different relaxations called

R_k, and a corresponding number of other entities subscripted with *k*.) The real total loss of the model selected by R_k computed using α is denoted by $LOSS_k$. The *lars* solution from the previous step is denoted by PRESOLUTION with corresponding objective value PREBOUND. Note that P_0=P, and plain *lars* is sufficient to solve R_0, since there are no restrictions on it. The best total loss seen so far is BESTLOSS.

The recursive step of the BLARS algorithm is shown in Figure 3. This is invoked as shown in Figure 4.

References

1. Tibshirani R (1996) Regression shrinkage and selection *via* the lasso. J Royal Stat Soc Ser B (Method) 58: 267-288.
2. Efron B, Hastie T, Johnston I, Tibshirai R (2004) Least angle regression. Ann Stat 32: 407-451.
3. Zou H (2006) The adaptive lasso and its oracle properties. J Am Stat Assoc 101: 1418-1429.
4. Lindley DV (1968) The choice of variables in multiple regression. J Royal Stat Soc Ser B (Method) 30: 31-66.
5. Brown PJ, Fearn T, Vannucci M (1999) The choice of variables in multivariate regression: A non-conjugate Bayesian decision theory approach. Biometrika 86: 635-648.
6. Dean CB, Nielsen JD (2007) Generalized linear mixed models: A review and some extensions. Lifetime Data Anal 13: 497-512.
7. http://CRAN.R-project.org/package=lars.
8. Hesterberg T, Choi NH, Meier L, Fraley C (2008) Least angle and l1 penalized regression: A review. Stat Surv 2: 61-93.
9. Shao J (1997) An asymptotic theory for linear model selection. Stat Sinica 7: 221-264.
10. Yang Y (2005) Can the strengths of AIC and BIC be shared? A conflict between model identification and regression estimation. Biometrika 92: 937-950.
11. Zou H, Hastie T, Tibshirani R (2007) On the “degrees of freedom” of the lasso. Ann Stat 35: 2173-2192.
12. Press WH, Teukolsky SA, Vetterling WT, Flannery BP (2007) Numerical recipes: The art of scientific computing. (3rd Edn), Cambridge University Press, New York, USA.
13. Lehman AF, Steinwachs DM (1998) Translating research into practice: The Schizophrenia Patient Outcomes Research Team (PORT) treatment recommendations. Schizophr Bull 24: 1-10.
14. Swabey T, Suskin N, Arthur HM, Ross J (2004) The Ontario Cardiac Rehabilitation Pilot Project. Can J Cardiol 20: 957-961.
15. Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models *via* coordinate descent. J Stat Softw 33: 1-22.