Published in final edited form as:

Comput Stat Data Anal. 2011 June 1; 55(6): 2026–2037. doi:10.1016/j.csda.2011.01.009.

FSR Methods for Second-Order Regression Models

Hugh B. Crews¹,

Center for Marine Science, University of North Carolina Wilmington

Dennis D. Boos, and

Department of Statistics, North Carolina State University

Leonard A. Stefanski

Department of Statistics, North Carolina State University

Summary

Most variable selection techniques focus on first-order linear regression models. Often, interaction and quadratic terms are also of interest, but the number of candidate predictors grows very fast with the number of original predictors, making variable selection more difficult. Forward selection algorithms are thus developed that enforce natural hierarchies in second-order models to control the entry rate of uninformative effects and to equalize the false selection rates from first-order and second-order terms. Method performance is compared through Monte Carlo simulation and illustrated with data from a Cox regression and from a response surface experiment.

Key words and phrases

Bagging; False selection rate; Model selection; Response optimization; Variable selection

1 Introduction

Variable selection techniques are used in a variety of settings with most attention focused on selecting a subset of the measured variables. In some applications such as response surface optimization, selecting interaction and quadratic terms is important. In such applications, second-order terms can increase the model's predictive accuracy and reveal patterns that would be missed when only considering the measured variables (see, for example, Hamada and Wu, 1992). However, the number of possible second-order terms grows exponentially with the number of predictors, making the problem of selecting the best subset difficult.

Several methods have been proposed to limit the number of terms under consideration. One natural approach is to restrict to models that are invariant to changes in measurement scale (see Peixoto, 1990, and Peixoto and Diaz, 1996). In a stepwise search this entails including both main effects before considering their interaction, called *strong heredity* or *strong hierarchy*. With *weak heredity* or *weak hierarchy* only one main effect is required to be in the model before considering the associated interaction. Chipman (1996) and Chipman, Hamada, and Wu (1997) use Bayesian methods to enforce these hierarchies in constructing

¹hugh.crews@gmail.com

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

second-order models. Yuan, Joseph, and Lin (2007) proposed LARS algorithms that enforce a hierarchy for analyzing experimental designs.

In the context of first-order regression models, Wu, Boos, and Stefanski (2007) developed a general simulation-based method for estimating the tuning parameter of variable selection techniques to control the False Selection Rate (FSR) of variables. More recently, Boos, Stefanski, and Wu (2009), henceforth BSW, proposed a “Fast” FSR approach that requires no simulation to estimate α -to-enter to use with stepwise forward regression. In this paper, we apply Fast FSR methodology to forward addition algorithms that enforce either the strong or weak hierarchies in second-order regression models. We also propose a new approach to forward addition sequence construction by using adjusted p -values for second-order terms. Then, by estimating the entry level of the associated forward selection method, we attain approximately equal contributions to the FSR from both first- and second-order effects.

Section 2 reviews the basic Fast FSR approach to variable selection and outlines the method to handle hierarchy-based algorithms. Section 3 presents the new forward addition algorithm and associated FSR selection methods. Section 5 compares methods via Monte Carlo simulation, and Sections 4 and 6 illustrate the methods with examples. Conclusions are summarized in Section 7.

2 Fast FSR Methods

2.1 Fast FSR, Forward Addition, and Forward Selection

When using a variable selection procedure on a data set with an $n \times 1$ response vector, $\mathbf{Y}$, and an $n \times k_T$ matrix of explanatory variables, $\mathbf{X}$, the False Selection Rate (FSR) is defined as

$$\gamma = E \left\{ \frac{U(\mathbf{Y}, \mathbf{X})}{1 + I(\mathbf{Y}, \mathbf{X}) + U(\mathbf{Y}, \mathbf{X})} \right\}, \quad (1)$$

where $I(\mathbf{Y}, \mathbf{X})$ and $U(\mathbf{Y}, \mathbf{X})$ are the number of informative and uninformative variables in the selected model. Informative variables are defined as those whose regression coefficients are nonzero. The goal of FSR variable selection is to tune the variable selection procedure such that the FSR is some desired level, γ_0 . Typically, $\gamma_0 = 0.05$, but other choices may be appropriate.

The FSR approach can be used with a variety of variable selection methods, but in this paper we use it with classical stepwise forward addition whereby variables are entered into the model in steps beginning with only an intercept in the model. The first variable to enter the model is the one whose p -value for the test of $H_0: \beta_j = 0$ is smallest, $j = 1, \dots, k_T$. On the next step, the variables not in the model are again tested, and the one with the smallest p -value is entered. Continuing in this fashion results in a sequence of models we call the *forward addition sequence*. It has $k_n = \min(k_T, n)$ steps and model sizes $\text{Size}(i) = i$, $i = 1, \dots, k_n$.

Given the forward addition sequence of models, a number of methods can be used to select a model such as choosing the one with the smallest AIC or BIC value. The most well-known method is to fix an α , say $\alpha_0 =$ the “ α -to-enter” value, and then the model of size k is selected when its associated minimum p -value from the k th step is less than or equal to α_0 but the

minimum p -value of the next step is greater than α_0 . We define this variable selection method as a function of α , Forward-Selection (α) = FS(α).

It is very important to distinguish between the forward addition sequence and Forward-Selection(α) that chooses a model from that sequence. The FSR selection method described below uses Forward-Selection ($\hat{\alpha}$) to select a model, where $\hat{\alpha}$ estimates the α that keeps the FSR at γ_0 . Often “forward selection” is used loosely to describe the forward addition sequence, Forward-Selection(α), or both, leading to confusion of (i) building a sequence of models and (ii) selecting a model from them.

A more precise definition of Forward-Selection (α) is as follows. Suppose that the minimum p -values in the steps of the forward addition sequence are labeled $p_1, p_2, \dots, p_{k_n}$, where recall that $k_n = \min(k_T, n)$. If these p -values are monotone increasing, $p_1 \leq p_2 \leq \dots \leq p_{k_n}$, then FS(α) chooses a model of size k , where $k = \max\{i: p_i \leq \alpha\}$. However, if the p -to-enter values are not monotone increasing, then they must be monotone increasing in order to use this set selection notation. We label the monotone increasing sequence of p -to-enter values as $\{\tilde{p}_1, \dots, \tilde{p}_{k_n}\}$, where $\tilde{p}_i = \max\{p_1, \dots, p_i\}$. Now, in general FS(α) chooses a model of size k , where $k = \max\{i: \tilde{p}_i \leq \alpha\}$. A curious feature of FS(α) is that when the p -values are not monotone, then certain models in the forward addition sequence are never selected.

When using Forward-Selection (α), let $U(\alpha) = U(\mathbf{Y}, \mathbf{X})$ be the number of uninformative variables selected, and let $S(\alpha)$ be the total number of variables selected. If $U(\alpha)$ were known, then a simple estimator for the FSR would be $U(\alpha)/\{1 + S(\alpha)\}$. Although $U(\alpha)$ is unknown, it can be estimated by $\hat{N}(\alpha)\theta(\alpha)$, where $\theta(\alpha)$ is the rate that uninformative variables enter the model, and $\hat{N}(\alpha)$ is an estimate of the total number of uninformative variables available for selection. Without hierarchy restrictions, $\hat{N}(\alpha)$ is taken to be $k_T - S(\alpha)$. In order to estimate $\theta(\alpha)$, Wu, Boos, and Stefanski (2007) generated phony explanatory variables and monitored their rate of entry over a grid of α values.

In the Fast FSR approach, BSW use $\theta(\alpha) = \alpha$, and therefore no phony variable simulation is required. This leads to the Fast FSR estimate,

$$\hat{\gamma}_F(\alpha) = \frac{\hat{N}(\alpha)\alpha}{1+S(\alpha)} = \frac{\{k_T - S(\alpha)\} \alpha}{1+S(\alpha)}. \quad (2)$$

The goal is to use the largest α such that $\hat{\gamma}_F(\alpha)$ is no greater than γ_0 . However, because $S(\alpha) \rightarrow k_n$ as $\alpha \rightarrow 1$, $\hat{\gamma}_F(\alpha)$ typically underestimates the FSR over the range $[\alpha_{\max}, 1]$, where $\alpha_{\max}$ is the α value such that $\hat{\gamma}_F(\alpha)$ is at its maximum. Therefore, α is estimated using

$$\hat{\alpha} = \sup_{\alpha \leq \alpha_{\max}} \{\alpha: \hat{\gamma}_F(\alpha) \leq \gamma_0\}. \quad (3)$$

Using these definitions leads to the Fast FSR rule for model size,

$$k(\gamma_0) = \max \left\{ i: \tilde{p}_i \leq \frac{\gamma_0[1+S(\tilde{p}_i)]}{k_T - S(\tilde{p}_i)} \text{ and } \tilde{p}_i \leq \alpha_{\max} \right\}. \quad (4)$$

Using $k(\gamma_0)$ from (4), the solution to (3) is

$$\widehat{\alpha} = \min \left\{ \frac{\gamma_0 \{1 + k(\gamma_0)\}}{k_T - k(\gamma_0)}, \alpha_{\max} \right\}.$$

BSW show that (4) can be viewed as a type of adaptive false discovery rate (FDR) method applied to the monotonized p -to-enter values (e.g., see Benjamini and Hochberg, 1995 and 2000).

2.2 Forward Addition Sequences for Second-Order Models

The observed data are n pairs $(Y_1, \mathbf{x}_1), \dots, (Y_n, \mathbf{x}_n)$, where $\mathbf{x}_i$ is a $p \times 1$ vector of design constants. We refer to these predictor variables as main effects. When estimating response surfaces, it is typical to also use the squares and products of the predictor variables. In other situations, it often makes sense to check for interactions and nonlinearities. Correlation among the predictor variables, however, generally makes variable selection more difficult. Thus, before adding quadratic terms, we first center the main effects to reduce correlation between second-order effects and parent main effects. For example, the sample correlation of X and X^2 when X consists of the integers 1 to 10 is .97, whereas the sample correlation of $X - 5.5$ and $(X - 5.5)^2$ is 0. One may also rescale the variables although this has no effect on

our forward selection approach. Then we relabel all $k_T = 2p + \binom{p}{2}$ variables $\mathbf{x}^* = (x_1^*, \dots, x_{k_T}^*)^T$ so that the first p of these are the centered and rescaled main effects, the second p are the squares of the first p , and the remaining $\binom{p}{2}$ are the main effect cross products. If some of the variables are binary, then the number of squared terms is less than p . The full $n \times k_T$ design matrix with rows $\mathbf{x}_i^{*T}$, $i = 1, \dots, n$, is $\mathbf{X}^*$.

No Hierarchy—A simple method of obtaining a forward addition sequence for a second-order model is to ignore the hierarchy between main effects and higher-order terms and treat each effect as a separate variable. If we create a forward addition sequence with this *No Hierarchy* approach, then each effect is a candidate for entry at the beginning of the forward addition process. Fast FSR with the forward addition sequence from using No Hierarchy

works exactly as described in Section 2.1 with $k_T = 2p + \binom{p}{2} - b$, where b is the number of binary variables. However, in the next section we propose an adjusted version of the No Hierarchy forward addition sequence that equalizes the entry of uninformative main effects and second-order terms. Otherwise second-order terms will tend to dominate the sparse models at the beginning of the forward addition sequence. Essentially the adjustment is to multiply the p -values of the second-order terms at each step by a constant that is the ratio of the number of second order terms available for entry divided by the number of main effects available. The Fast FSR formulas must be changed to accommodate this adjustment.

Strong Hierarchy—An approach with some philosophical appeal is to enforce a hierarchy throughout variable selection. When creating the forward addition sequence with *Strong Hierarchy* (or strong heredity), an interaction cannot enter the model until both of its parent main effects are in the model. Similarly, a squared term cannot enter the model until its parent main effect is in the model. Thus the Fast FSR formulas must be modified to adapt to the dynamically changing number of candidate terms available at each step of the forward addition sequence. However, p -value adjustment for over-selection of second-order terms is

not necessary because the number of second-order terms available at the beginning of the sequence is quite limited.

Weak Hierarchy—A less restrictive hierarchy, called *Weak Hierarchy* (or weak heredity), allows an interaction to enter the model provided at least one of its parent main effects is in the model. Similarly, a squared term cannot enter the model until its parent main effect is in the model. Although the number of second-order terms available at each step of the forward addition sequence starts out small, it grows quickly after a number of main effects enter the model. Thus, we also provide an adjusted forward addition sequence under Weak Hierarchy restrictions that takes into account the number of second order terms available. Two separate Fast FSR formulas are required for the two versions.

In summary, we consider three hierarchy principles to use in building second-order models via forward selection: No Hierarchy, Strong Hierarchy, and Weak Hierarchy. Associated with each of the No Hierarchy and Weak Hierarchy approaches are two versions of the forward addition sequence, the natural one and an adjusted one that takes into account the differing numbers of first-order and second-order terms available at each step.

Table 1 illustrates the first three steps of the forward addition sequences for the example in Section 4 where $p = 17$ and $k_T = 165$ (not 170 because there are $b = 5$ binary variables). Note that at Step 3 for the Strong Hierarchy, $x_8 * x_4$ is available to enter, but x_{17} has the smallest p -value (not displayed) and enters. For the No Hierarchy approach, the interaction $x_4 * x_7$ enters first, but the adjustments explained in the next section move x_8 ahead of $x_4 * x_7$ and thus reverse the order of the first two steps. The Weak Hierarchy has a different Step 3 from the others, but the p -value adjustments do not change the order.

Associated with these five forward addition sequences are Fast FSR formulas that yield a selected model from each sequence. Extension of the Fast FSR formulas to handle the various hierarchy restrictions and p -value adjustments takes some care, and the next section explains these extensions.

3 Modified Fast FSR Formulas

3.1 Extended Definition of $\hat{N}$

At any given step of the FSR method based on $FS(\alpha)$, the number of uninformative variables available to enter the model is estimated by the total number of variables available to enter. In the case of No Hierarchy, this is simply $\hat{N}(\alpha) = k_T - S(\alpha)$. However, if a hierarchy is enforced, then the set of variables available to enter is not simply the complement of the set of variables already in the model, but depends dynamically on the form of the hierarchy and on the particular variables that have already entered. Thus under hierarchy restrictions $\hat{N}(\alpha) = k_T - S(\alpha)$ is no longer an appropriate estimate of the number of uninformative variables available for selection. The appropriate estimate is defined below at the values of the associated monotized p -value sequence (adjusted or otherwise) and requires information on variables available for selection used in forming the associated forward addition sequence. In the following $\tilde{p}_1 \leq \tilde{p}_2 \leq \dots \leq \tilde{p}_{k_n}$ are the monotized p -values of the variables that entered.

1. If $\alpha = \tilde{p}_i$ for i such that a single variable enters, $\hat{N}(\alpha)$ equals one less than the number of variables available to enter at $\alpha = \tilde{p}_i - \varepsilon$, for $\varepsilon > 0$ suitably small. The reduction by one is for the variable entering at $\alpha = \tilde{p}_i$.
2. Now consider the case where $\alpha = \tilde{p}_i$ and i is such that $\tilde{p}_i$ appears k times in the monotized sequence. Then all k variables enter Forward-Selection(α) at $\alpha = \tilde{p}_i$. Using any of the various forward addition sequences, we define $\hat{N}(\alpha)$ to be one less

than the number of candidate predictors available right before the last of the k steps. The reduction by one is for the last of the k variables that enter at $\alpha = \tilde{p}_i$.

For example, consider a Weak Hierarchy case with $p = 4$ main effects and the sequence $(p_1, p_2, p_3) = (.001, .0005, .0003)$ for entering terms (X_2, X_4, X_4^2) in the first three steps of forward addition. The corresponding number of available predictors before each step is $(4, 7, 9)$ because the corresponding sets of available predictors are $\{X_1, X_2, X_3, X_4\}$, $\{X_1, X_3, X_4, X_2^2, X_2 X_1, X_2 X_3, X_2 X_4\}$, and $\{X_1, X_3, X_2^2, X_2 X_1, X_2 X_3, X_2 X_4, X_4^2, X_4 X_1, X_4 X_3\}$. Then, monotonicizing gives $(\tilde{p}_1, \tilde{p}_2, \tilde{p}_3) = (.001, .001, .001)$ and $\hat{N}(.001) = 9 - 1 = 8$. Notice that in terms of monotonicized p -to-enter values, all three terms come in at $\alpha = .001$ so that $S(.001) = 3$, but we are using step information to compute $\hat{N}$. The value $\hat{N}(.001) = 8$ makes sense because there really were 9 terms available for entry before the 3rd step, but since X_4^2 entered on the 3rd step, we no longer consider it a possible uninformative variable. If the forward addition sequence had been based on the Strong Hierarchy principle, then the available predictors before each step would have been $\{X_1, X_2, X_3, X_4\}$, $\{X_1, X_2^2, X_3, X_4\}$, and $\{X_1, X_2^2, X_3, X_4^2, X_2 X_4\}$, and $\hat{N}(.001) = 5 - 1 = 4$.

This definition of $\hat{N}(\alpha)$ is consistent with $k_T - S(\alpha)$ when No Hierarchy is used. Replacing $k_T - S(\alpha)$ in (2) with $\hat{N}(\alpha)$, leads to the general Fast FSR formula

$$\hat{\gamma}_F(\alpha) = \frac{\hat{N}(\alpha)\alpha}{1 + S(\alpha)}. \quad (5)$$

Note that $\hat{N}(\alpha)$ differs by type of hierarchy constraint and p -value adjustment. The estimated α remains defined by (3), however, the rule for model size is now

$$k(\gamma_0) = \max \left\{ i: \tilde{p}_i \leq \frac{\gamma_0 \{1 + S(\tilde{p}_i)\}}{\hat{N}(\tilde{p}_i)} \text{ and } \tilde{p}_i \leq \alpha_{\max} \right\}. \quad (6)$$

The selection method defined by (6) is suitable for the forward addition sequences associated with the Strong, Weak, and No Hierarchy principles. In the next section, we adjust the Weak and No Hierarchy forward addition sequences to accommodate the possibly large number of second-order terms. The modified FSR formulas are then given in a later section.

3.2 Adjusted Forward Addition Sequences for No Hierarchy and Weak Hierarchy Approaches

With $p = 10$ main effects, there are $10 + 45 = 55$ second-order terms. With $p = 20$ there are $20 + 190 = 210$ second-order terms. Moreover, one might expect a larger percentage of second-order terms to be uninformative than first-order terms. If main effects and second-order terms are treated equally in the No- or Weak-Hierarchy approaches, then the number of uninformative second-order terms entering the model by chance will be much larger than the number of uninformative main effects entering the model. Thus in Algorithm 1 below we define adjusted forward addition sequences by (i) multiplying the p -values of the second-order terms by a constant that reflects the relative numbers of first-order and second-order terms available for entry at each step; and (ii) choosing the term at each step with the smallest adjusted p -value. Let $c^{(i)}$, $i = 1, \dots, k_n = \min(k_T, n)$, be the sequence of positive

constants used to adjust the second-order p -values. In the next section we motivate the choice of constants $c^{(i)}$ in terms of FSR.

Algorithm 1: Forward Addition Sequence with Adjusted P -Values

1. Starting with an intercept term in the model, calculate the p -to-enter values for adding any single effect from the candidate set of all main effects, interactions, and quadratic terms. Call these p -to-enter values $\{p_{1,1}, \dots, p_{1,k_T}\}$. For Step 1, define

$$\text{adjusted } p\text{-to-enter values} = \begin{cases} p_{1,j} & \text{if } j \in \mathcal{M}, \\ c^{(1)} p_{1,j} & \text{otherwise,} \end{cases} \quad (7)$$

where $\mathcal{M}$ is the set of main effect indices. Select the effect, $X_{(1)}$, corresponding to $p_1 = p_{1,(1)}$, the smallest adjusted p -to-enter value for the first step.

2. With $X_{(1)}$ and an intercept term in the model, calculate the p -to-enter values for adding any single effect remaining in the candidate set. Next, calculate the adjusted p -to-enter values using (7) with $c^{(2)}$ replacing $c^{(1)}$ and select the effect, $X_{(2)}$, corresponding to $p_2 = p_{2,(1)}$, the smallest adjusted p -to-enter value for the second step.
3. Repeat this process, updating $c^{(i)}$ at each step until $k_n = \min(k_T, n)$ terms have been entered.

3.3 Fast FSR for Adjusted Forward Addition Sequences

Our goal here with the FSR approach is an estimate $\hat{\alpha}$ so that $\text{FS}(\hat{\alpha})$ selects a model from the adjusted forward addition sequences of Algorithm 1 such that the contribution of the FSR rate from first-order effects (FSR_m) is equal to the contribution to the false selection rate from second-order effects (FSR_q), in other words,

$$\text{E}[\text{FSR}_m] = \text{E}[\text{FSR}_q] = \gamma_0/2. \quad (8)$$

We proceed similar to the original Fast FSR approach and regard all effects in the model as informative and all effects in the candidate set as uninformative. Thus at the i th step, define $\text{Size}_m(i-1)$ and $\text{Size}_q(i-1)$ to be the numbers of main effects and second-order effects in the model after $i-1$ steps, respectively. For the No-Hierarchy approach, $\hat{N}_m(i-1) = p - \text{Size}_m(i-1)$ is the estimated number of uninformative main effects after $i-1$ steps, and similarly $\hat{N}_q(i-1) = k_T - p - \text{Size}_q(i-1)$ is the estimated number of uninformative second-order effects. For the Weak Hierarchy, $\hat{N}_m(i-1) = p - \text{Size}_m(i-1)$ is the same, but $\hat{N}_q(i-1)$ has to take into account what second-order effects are in the model. Then at Step i of Algorithm 1, we propose

$$c^{(i)} = \frac{1 + \hat{N}_q(i-1)}{1 + \hat{N}_m(i-1)}. \quad (9)$$

The justification for (9) is as follows. Recall that $\text{FS}(\alpha)$ only selects models when α is one of the monotonized minimum p -values $\tilde{p}_i$ (and here the $\tilde{p}_i$ are defined from the minimum adjusted p -values). Let α be one of these monotonized minimum p -values, and let i_α be the

largest i associated with all $\tilde{p}_i$ equal to α . Then running $\text{FS}(\alpha)$, we expect $\hat{N}_m(i_\alpha - 1)\alpha$ uninformative main effects to enter and

$$\hat{N}_q(i_\alpha - 1)\alpha / c^{(i_\alpha)} = \hat{N}_q(i_\alpha - 1)\alpha \left\{ \frac{1 + \hat{N}_m(i - 1)}{1 + \hat{N}_q(i - 1)} \right\} \approx \hat{N}_m(i_\alpha - 1)\alpha$$

uninformative second-order effects to enter. Adding 1 in the numerator and denominator of (9) ensures $0 < c^{(i)} < \infty$. Thus, use of (9) in Algorithm 1 attempts to equalize the number of uninformative first-order and second-order terms in the model at each of the possible models of $\text{FS}(\alpha)$.

After running $\text{FS}(\alpha)$ as described, we have a sequence of $c^{(i)}$ values, effects entered, and monotonized minimum adjusted p -values $\tilde{p}_1, \dots, \tilde{p}_{k_n}$. To define $\hat{\gamma}_F(\alpha)$ at all α , we define $S_m(\alpha) = \text{Size}_m(i_\alpha)$, $S_q(\alpha) = \text{Size}_q(i_\alpha)$, and $c(\alpha) = c^{(i_\alpha)}$ at $\alpha = \tilde{p}_i$, and let these functions be constant between values of $\tilde{p}_i$ so that the functions are right continuous step functions. Then

$$\hat{\gamma}_F(\alpha) = \frac{\hat{N}_m(\alpha)\alpha + \hat{N}_q(\alpha)\alpha / c(\alpha)}{1 + S_m(\alpha) + S_q(\alpha)}. \quad (10)$$

After calculating $\hat{\gamma}_F(\alpha)$ for each $\tilde{p}_i$, we choose the model of size

$$k(\gamma_0) = \max \left\{ i: \tilde{p}_i \leq \frac{\gamma_0 [1 + S_m(\tilde{p}_i) + S_q(\tilde{p}_i)]}{\hat{N}_m(\tilde{p}_i) + \hat{N}_q(\tilde{p}_i) / c(\tilde{p}_i)} \text{ and } \tilde{p}_i \leq \alpha_{\max} \right\}, \quad (11)$$

and let $\hat{\alpha} = \sup_{\alpha \leq \alpha_{\max}} \{ \alpha: \hat{\gamma}_F(\alpha) \leq \gamma_0 \}$.

4 Example with Cox Regression

Here we briefly illustrate the flexibility and simplicity of using the methods described in Sections 2 and 3. This analysis is made very easy due to a set of SAS macros found at <http://www4.stat.ncsu.edu/~boos/var.select>. We use the primary biliary cirrhosis data analyzed in Fleming and Harrington (1991, p. 153–162) and given in their Appendix D. This data set has been analyzed often and is available on the web and in the above website with full description of variables. There are 276 complete cases with $p = 17$ predictors and $k_T = 165$ total quadratic terms (five of the predictors are binary). The response variable is survival time and there are 187 (60%) censored cases.

We first used Fast FSR with the Cox proportional hazards model and just the 17 main effects. The first column of Table 2 shows the order in which the main effects entered the model. Fast FSR chose a model of size 9 with $\hat{\alpha} = .0625$. Zhang and Lu (2007, Table 5) show that the LASSO identifies exactly the same 9 variables, and the Adaptive LASSO chooses the same variables except for *ascites*. We now consider models with interactions. In column 2 of Table 2 we use Fast FSR with the Strong Hierarchy and obtain the same size model but with two interactions, *ascites***copper* and *ascites***edema*, replacing the main effects *albumin* and *protime*. Notice that $\hat{\alpha} = 0.012$ (lower than 0.0625 because more terms were considered) and that the BIC is somewhat lower.

Next, we ran Fast FSR with No Hierarchy because we wanted to see if there are important interactions that could not enter the model due to weak parent effects. In column 3 of Table 2 we see that four interactions are in this model, three that did not overlap with those chosen by the Strong Hierarchy. Note that $\hat{\alpha} = 0.0032$ is much smaller in order to control the FSR in the face of 165 possible predictors. To further analyze within the No Hierarchy setup, we used the adjusted p -values to limit the entry of second-order terms. The fourth column shows that this adjusted forward sequence gives a model not so different from the second column with the Strong Hierarchy. Either model might be of interest for clinicians to evaluate. The two Weak Hierarchy approaches resulted in models with different interactions and possibly harder interpretation.

5 Simulation Studies

In this section we summarize two simulation studies designed to assess the performance of the Fast FSR methods. In the first, performance criteria are related to prediction and interpretation. In the second, performance criteria are related to response surface optimization. Crews (2008) contains additional details and results for each study.

5.1 Simulation for Prediction and Interpretation

We compare the Fast FSR methods with the Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani, 1996) and with Bayesian Additive Regression Trees (BART) (Chipman et al., 2006). For LASSO we used 5-fold cross validation to determine a model, whereas for BART we used the default settings, i.e., BART-default as described in Chipman et al. (2006). Recall that all predictor variables are first centered by subtracting the mean and then are divided by the sample standard deviation. Forward selection with p -values determined from the usual least squares F -tests are used in all the Fast FSR methods. Terms above second-order are not considered. The following Fast FSR methods are studied.

Fast FSR with No Hierarchy (FFSR-NH)—All $k_r = 2p + \binom{p}{2}$ terms are available at all steps, and model size is chosen by (4).

Fast FSR with Strong Hierarchy (FFSR-SH)—Interactions $X_i X_j$ are available only after both X_i and X_j are in the model, whereas X_i^2 is available after X_i is in. Model size is chosen by (6).

Fast FSR with Weak Hierarchy (FFSR-WH)—Interactions $X_i X_j$ are available only after X_i or X_j are in the model, whereas X_i^2 is available after X_i is in. Model size is chosen by (6).

Fast FSR with P -Value Adjustment for No Hierarchy (FFSR-NH_{adj})—Same as FFSR-NH except that the second-order p -to-enter values are multiplied by $c^{(i)}$ of (9) at step i . Model size is chosen by (11).

Fast FSR with P -Value Adjustment for Weak Hierarchy (FFSR-WH_{adj})—Same as FFSR-WH except that the second-order p -to-enter values are multiplied by $c^{(i)}$ of (9) at step i . Model size is chosen by (11).

We studied models with $p = 20$ original predictors, and so there were $p_q = 230$ total predictors. The original predictors were generated as either $N(10, 20)$ or χ_{10}^2 random variables with both correlated and uncorrelated cases and sample sizes $n = 100$ and $n = 500$.

$N = 100$ independent data sets were generated for each situation. Correlated predictors had the following correlation structure:

$$\text{Corr}(X_i, X_j) = \begin{cases} 0.7 - 0.1(|i - j| - 1) & \text{if } 1 \leq |i - j| < 8, \\ 0 & \text{if } 8 \leq |i - j| < 13, \\ 0.7 - 0.1(19 - |i - j|) & \text{if } 13 \leq |i - j| \leq 19. \end{cases}$$

Note that this correlation initially decays linearly, is zero for lags 8 through 12, and then rises linearly again. However, because we randomly permute the columns of each data set, the only important fact is that there are 20 pairs of X columns with correlations 0.1 to 0.7, respectively, and 50 pairs of columns with no correlation. The models used are:

1. $Y = -100 + 25X_1 + 15X_{13} - 20X_{17} + X_1^2 - 3X_1X_9 + \varepsilon;$
2. $Y = -3 + X_1 - X_4 + 2X_9 - 1.2X_{13} + 1.6X_{17} + \varepsilon;$
3. $Y = 50 + 15X_1 - 25X_9 + 1.2X_1^2 - 1.6X_9^2 + 3X_1X_9 + \varepsilon.$

For each model, $\varepsilon \sim N(0, \sigma^2)$, where σ was chosen to achieve *theoretical* R^2 values 0.25 and 0.50, where

$$\text{theoretical } R^2 = \text{Var} \left(\sum_{j=1}^p \beta_j X_j \right) / \text{Var} \left(\sum_{j=1}^p \beta_j X_j + \varepsilon \right).$$

The key measure of performance used was average model error (AME),

$$\text{AME} = (nN)^{-1} \sum_{i=1}^N \sum_{j=1}^n (\widehat{Y}_{ij} - \mu_{ij})^2.$$

Although this definition of model error corresponds to a fixed design, we use it here because of the random permutation of the design matrices after generation. To maintain similar scales, results are given in terms of the ratio of the AME of the true model to the AME of a particular method. We call this measure the AME Ratio and note that methods with a high AME Ratio are preferred.

Treating the simulation results as repeated measures ANOVA with 7 methods and a $2^4 \times 3$ factorial treatment structure, we fit a linear model in SAS `proc mixed` with AME Ratio as our response and with the following factors: X distribution ($N(10, 20)$ or χ_{10}^2), predictor correlation (presence or absence), theoretical R^2 (0.25 or 0.50), sample size (100 or 500), and model (1–3). Tables 3 and 4 present part of the ANOVA results.

Table 3 shows that FFSR-SH was the best overall performer. Statistical significance at the experiment .05 level was assessed using Tukey's range test. Among the FSR methods, No Hierarchy fared the worst, and FFSR-WH and FFSR-NH_{adj} were roughly equivalent, with FFSR-WH_{adj} slightly better than those two. Clearly, p -value adjustment made a major improvement in the No Hierarchy method and a minor improvement in the Weak Hierarchy approach. Overall, BART and the LASSO were not competitive except when $n = 100$ and $R^2 = 0.25$. The LASSO generally captured a large proportion of the informative effects, but

because it tends to include a large number of effects, it also had large AFSR. Neither the LASSO nor BART used any hierarchy structure, and therefore suffered from overfitting interactions, in addition to lessening interpretability. Yuan, Joseph, and Lin (2007) show how to enforce hierarchy restrictions with LARS, a close relative of the LASSO.

Analysis among only the FSR methods reveals that Method does not interact strongly with the other factors. Among the five factors, model and sample size had large main effects as well as interactions with each other and with sample size. Theoretical R^2 did not have a strong main effect, but it did have a strong interaction with model. Distribution type for the X matrix (means 0.39 for χ^2 and 0.38 for normal) had no significant difference, and correlation within the X matrix (means 0.40 for no correlation and 0.37 for correlation) had only small effects. Table 4 shows the means for the important effects.

Model 1 was a combination of main effects and second-order terms, whereas Model 2 contained all main effects. Model 3 was a full quadratic in two variables and the toughest model to fit. For Models 1 and 2, as the sample size and R^2 increased, the methods performed better relative to the true model. However, for Model 3 this was not the case.

The average false selection rate (AFSR) is defined as the average over the Monte Carlo data sets of the number of uninformative effects selected divided by 1 plus the total number of effects selected. AFSR can be partitioned into contributions from main effects, AFSR_m , and second-order effects, AFSR_q .

Figure 1 illustrates that for uncorrelated predictors, the Fast FSR methods performed as expected choosing models whose AFSR were close to $\gamma_0 = 0.05$. For correlated predictors and $n = 100$, however, AFSR rates were on average around .10 although they generally improved for $n = 500$. The adjusted Fast FSR methods were designed to ensure that $E(\text{FSR}_m) = E(\text{FSR}_q) \approx E(\text{FSR})/2$. The simulation showed that if the original p predictors were uncorrelated, then AFSR_m and AFSR_q were approximately equal when using the adjustment methods. However, when the predictors were correlated, the AFSR contributions from the two groups were often unequal. For example, at $n = 100$, $R^2 = 0.25$, and Model 2, FFSR-NH_{adj} had $\text{AFSR}_m = 0.104$ and $\text{AFSR}_q = 0.012$. For $n = 500$, those rates improved to 0.035 and 0.018, respectively.

5.2 Simulation for Response Optimization

For the response optimization study, two response surface designs were used to generate data. The first design was a 73-run, small composite, design with $p = 10$ factors, where main effects are orthogonal but interactions are correlated with main effects and/or other interactions. The second design was a 100-run, orthogonal, central composite, design with $p = 8$ factors, where all $k_T = p_q = (2)(8) + 28 = 44$ variables are orthogonal.

For each design, responses were generated from two models: Models 1a and 1b used the small composite design, whereas Models 2a and 2b used the central composite design. The models are defined as follows.

$$\begin{aligned} 1a: Y &= 15 - 3X_1 + 1.5X_1^2 + 3X_1X_9 + X_2 - 2X_3 + 1.5X_4 + X_5 - 2X_6 + X_7 + X_8 - X_9 - 5X_9^2 + \varepsilon \\ 1b: Y &= 15 - 5X_1 + 1.5X_1^2 + 3X_1X_9 + X_2 - 2X_3 + 1.5X_4 + X_5 - 2X_6 + X_7 + X_8 - 7X_9 - 5X_9^2 + \varepsilon \end{aligned}$$

$$2a: Y = 20 + 2X_1 - 4X_1^2 + 5X_1X_2 + 3X_1X_3 - 3X_2 - 3X_2^2 + 1.5X_2X_4 + 2X_3 + 4X_3^2 - 3X_4 - 2X_4^2 + 2X_4X_5 + 2.5X_5 + 2X_6 + 1.5X_7 + \varepsilon$$

$$2b: Y = 20 + 5X_1 - 3.5X_1^2 - X_1X_2 + 3X_1X_3 - 3X_2 - 3X_2^2 + 1.5X_2X_4 + 2X_3 + X_3^2 - 4X_4 - 2X_4^2 + 2X_4X_5 + 3.5X_5 + 2X_6 + 1.5X_7 + \varepsilon.$$

For each model, $\varepsilon \sim N(0, \sigma^2)$, where σ was chosen to achieve theoretical R^2 values .050, 0.75, or 0.90. As in the first study, $N = 100$ independent data sets were generated from each model.

Because we are mimicking the situation where screening is conducted prior to the response surface design, we created models with most main effects present. In all the models, only one variable has no effect on the response (X_{10} in Models 1a and 1b, X_8 in Models 2a and 2b). For Model 1a, the variable X_9 has a small main effect but a large interaction with X_1 and a large quadratic effect. The purpose of this model is to illustrate the lack of power of the hierarchy-based approaches to select second-order effects when their parent main effects are small. In Model 1b, the main effects of X_1 and X_9 are larger. Therefore, we expect the hierarchy methods to perform better. For Model 2a, the variable X_1 has a small main effect but a large interaction with X_3 and a large quadratic effect. As for Model 1a, the hierarchy-based approaches are at a disadvantage for Model 2a because X_1 must first enter before the large second-order effects have a chance to enter. In Model 2b, the effects of X_1 , X_4 , and X_5 are larger to give the hierarchy methods an advantage.

The goal of response surface modeling is usually to estimate the levels of a process that yield an optimal response. After fitting by Fast FSR, LASSO, or a standard approach where a full model was fit and terms removed if not significant at the $\alpha = 0.05$ level, each fitted model was optimized to get a set of optimal factor levels. The optimization was carried out subject to the constraint that each X lies in $(-2, 2)$; we call this constrained factor space the *region of interest*.

Tables 5 and 6 give the optimal factor levels for the four models as well as the maximum response. For most factors, the optimal level lies on the boundary of the region of interest. In Models 1a and 1b, only variable X_9 has an optimal level in the interior the region. For Models 2a and 2b, variables X_1 , X_2 , and X_4 all have optimal levels in the interior of the region.

To compare the methods we need a measure of how well a method identifies the optimum levels. The true mean optimal response is $\mu(\mathbf{X}_{\text{opt}})$, and the true mean response using $\hat{\mathbf{X}}_{\text{opt}}$ is $\mu(\hat{\mathbf{X}}_{\text{opt}})$. For any factor not selected we set the optimal level at the center point, 0. We refer to $\mu(\hat{\mathbf{X}}_{\text{opt}})$ as the actual performance, whereas $\mu(\mathbf{X}_{\text{opt}})$ is the optimal performance. The standardized difference, $\{\mu(\hat{\mathbf{X}}_{\text{opt}}) - \mu(\mathbf{X}_{\text{opt}})\} / \mu(\mathbf{X}_{\text{opt}})$, is a measure of how close a method performs relative to the true optimal performance. When analyzing a real data set, the estimate of the optimal response is $\hat{\mu}(\hat{\mathbf{X}}_{\text{opt}})$, but that is not used in this simulation.

Figure 2 illustrates the mean performance for each method. Because Fast FSR with P -Value Adjustment for Weak Hierarchy (FFSR-WH_{adj}) was always better than Fast FSR with Weak Hierarchy (FFSR-WH), and Fast FSR with P -Value Adjustment for No Hierarchy (FFSR-NH_{adj}) was always better than Fast FSR with No Hierarchy (FFSR-NH), FFSR-WH and FFSR-NH were left off the figure. For Model 1a, Fast FSR with Strong Hierarchy (FFSR-SH) and FFSR-WH_{adj} performed poorly. The reason is that X_9^2 and X_1X_9 are both large effects, but the main effect X_9 is relatively small, thus making it hard for these second-order terms to enter. The LASSO and FFSR-NH_{adj} performed the best in Model 1a, with the LASSO better for $R^2 = 0.5$ and FFSR-NH_{adj} better for $R^2 = 0.9$. For Model 1b, the methods performed fairly equally with FFSR-SH, FFSR-NH_{adj}, and LASSO among the best. The LASSO does best with smaller R^2 , and FFSR-NH_{adj} is better for larger R^2 . For Model 2a,

FFSR-SH and FFSR-WH_{adj} performed poorly. In this model, X_1^2 is very important, but its main effect is relatively small. Therefore, FFSR-NH_{adj} performed best. For Model 2b, FFSR-NH_{adj} again performed the best overall.

BSW use bagging (Breiman, 1996) to improve predictions. The basic idea of bagging is to take a random sample with replacement of size n from the full data set and use this bootstrap sample to obtain β^* . After repeating the process B times, average the β^* to obtain $\bar{\beta}^*$. BSW note that β^* typically has no zeros, so there is no variable selection in the averaged model. However, the model can be used for prediction or for determining optimal factor levels. Bagged versions of the Fast FSR methods were also used in the response surface simulations. Generally, bagging improved performance when using FFSR-SH, but the improvement was not as large for FFSR-NH_{adj}. In general, when $R^2 = 0.5$, the bagged Fast FSR methods were superior to their regular versions in estimating the optimal factor levels. However, as R^2 increased to 0.9, bagging yielded little or no improvement. The only exception was for the hierarchy-based approaches on Models 1a and 2a. In these models, bagging overcomes the problems of fitting large second-order terms with weak parent main effects.

A standard approach is to fit the full response surface and eliminate terms not significant at the $\alpha = 0.05$ level. This approach performed poorly for Models 1a and 1b, but performed very well for Models 2a and 2b. Possible reasons for the poor performance in Models 1a and 1b are the sparsity of the true models, the large number of factors, and the correlation between interactions in the design matrix. Even when the standard approach performed well, it still had large false selection rates. Therefore, we recommend FFSR-NH_{adj}, especially in studies with a large number of factors; or a bagged version of FFSR-SH. From this study, it is clear that the power of a method to select informative quadratic and interaction terms is important when optimizing a response.

6 Example

Dual Response Optimization in the Lipase Study

Lipase is an enzyme used in industrial and food processes for its ability to break down lipids. Rath et al. (2002) used response surface modeling to maximize both the production of lipase and its ability to break down fatty acids or specific activity. In order to produce lipase, the bacteria *Burkholderia cepacia* was cultivated with concentrations of glucose and palm oil added as nutrients. In addition to the nutrient factors, Rath et al. (2002) were interested in the effect of incubation time, inoculum density, and agitation on the two response variables. Table 7 lists the variables in their study.

Rath et al. (2002) used a 32-run face-centered, central composite design and fit a second-order linear model in all five factors excluding their interactions. Using their models, the estimated maximum lipase production is 31 U/mL and maximum activity is 110 units/mg.

We used Fast FSR methods with $\gamma_0 = 0.05$ to analyze the data. The selected variables for each method and response are listed in Tables 8 and 9. The Fast FSR methods did not choose the same effects in their final models. FFSR-SH and FFSR-WH appear to underfit the data. Because the main effects were not selected, these approaches were unable to fit the significant quadratic terms. Conversely, FFSR-NH_{adj} fit larger, more reasonable models. For lipase production no effects were common to all models, although it is likely that palm oil, incubation time, and inoculum density all influence lipase production in some manner. For specific activity only incubation time is common to all the models, whereas glucose was the only factor not selected by any Fast FSR method.

The standard approach models in Tables 8 and 9 show which variables had Type III p -values less than 0.05. Thus, of the full ten-variable model used by Rathi et al. (2002), only X_2^2 , X_3 , X_4^2 , and X_5^2 are statistically significant at 0.05 level for both responses. Additional simulations in Crews (2008) suggest that their ten-variable models are too large.

Following Rathi et al. (2002), we maximized lipase production and specific activity subject to $-1 \leq X_j \leq 1$. Tables 10 and 11 give the coded optimal factor levels and estimated maximum for each method and response. Because the hierarchy-based approaches fit very small models, it is likely that they underestimate maximum lipase production and activity. The models fit using FFSR-NH_{adj} are reasonable but differ somewhat for the two responses.

The quadratic term for inoculum density, X_4^2 , is not chosen in the model for lipase activity. Without this term the model estimates maximum activity to be approximately 73 U/mg, whereas including this term increases the estimate to approximately 90 U/mg. Further investigation shows that the next possible model in the forward sequence, which adds X_4^2 and X_5^2 , has $\hat{\gamma}_F$ just over the 0.05 limit. Using this model for maximization leads to 100.63 U/mg. Based on these results, the optimal factor levels for glucose, palm oil and agitation are close to the center point. The optimal factor level for incubation time is the low level, and the optimal level for inoculum density is the high level. Using these optimal levels, the estimated maximum lipase production is approximately 23 U/mL, and the estimated maximum activity is approximately 90–100 U/mg. Therefore, it is likely that the maximum lipase production and specific activity observed in practice would be smaller than the estimates 31 U/mL and 110 units/mg, provided by Rathi et al.

7 Conclusion

Fast FSR methods provide a simple approach for variable selection in applications where quadratic terms are of interest. Although the Strong Hierarchy restriction (FFSR-SH) has intuitive appeal and performed best in our first simulation study that had $p = 20$ original variables, it can perform poorly in models where there are important second-order effects but weak parent main effects. In particular it did not perform very well in our second simulation study involving response optimization. Using no hierarchy restrictions (FFSR-NH) can prevent strong second-order effects from being missed. However, second-order terms often dominate the forward sequence, so adjusting the p -values with FFSR-NH_{adj} is recommended. In general, bagging both FFSR-SH and FFSR-NH_{adj} show improvement when estimating optimal factor levels.

Acknowledgments

This work was supported by NSF grants DMS-0504283 and DMS-0906421 and NIH grant P01 CA142538-01. SAS macros for computing Fast FSR variable selection are available at <http://www4.stat.ncsu.edu/~boos/var.select>.

References

- Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society B*. 1995; 57:289–300.
- Benjamini Y, Hochberg Y. On the Adaptive Control of the False Discovery Rate in Multiple Testing with Independent Statistics. *Journal of Educational and Behavioral Statistics*. 2000; 25:60–83.
- Boos DD, Stefanski LA, Wu Y. Fast FSR variable selection with applications to clinical trials. *Biometrics*. 2009; 65:692–700. [PubMed: 18945266]
- Breiman L. Bagging Predictors. *Machine Learning*. 1996; 24:123–140.
- Chipman HA. Bayesian Variable Selection with Related Predictors. *The Canadian Journal of Statistics*. 1996; 24:17–36.

- Chipman HA. A Bayesian Variable-Selection Approach for Analyzing Designed Experiments with Complex Aliasing. *Technometrics*. 1997; 39:372–381.
- Chipman, HA.; George, EL.; McCulloch, RE. BART: Bayesian Additive Regression Trees. 2006. preprint
- Crews, HB. unpublished PhD dissertation. North Carolina State University, Dept. of Statistics; 2008. Fast FSR Methods for Second-Order Linear Regression Models.
- Fleming, TR.; Harrington, DP. Counting Processes and Survival Analysis. Chichester: John Wiley and Sons; 1991.
- Hamada M, Wu CFJ. Analysis of Designed Experiments with Complex Aliasing. *Journal of Quality Technology*. 1992; 24:130–137.
- Peixoto JL. A Property of Well-Formulated Polynomial Regression Models. *The American Statistician*. 1990; 44:26–30.
- Peixoto JL, Diaz J. On the Construction Hierarchical Polynomial Regression Models. *Estadística*. 1990; 48:175–210.
- Rathi P, Goswami VK, Sahai V, Gupta R. Statistical Medium Optimization and Production of a Hyperthermostable Lipase from *Burkholderia cepacia* in a Bioreactor. *Journal of Applied Microbiology*. 2002; 35:59–67.
- Tibshirani RJ. Regression Shrinkage and Selection via the LASSO. *Journal of the Royal Statistic Society, Series B*. 1996; 58:267–288.
- Wu Y, Boos DD, Stefanski LA. Controlling Variable Selection by the Addition of Pseudovariables. *Journal of the American Statistical Association*. 2007; 102:235–243.
- Yuan M, Joseph VR, Lin Y. An Efficient Variable Selection Approach for Analyzing Designed Experiments. *Technometrics*. 2007; 49:430–439.
- Zhang HH, Lu W. Adaptive-LASSO for Cox's Proportional Hazard Model. *Biometrika*. 2007; 94:691–703.

Figure 1.

Average false selection rates (AFSR) for each method using normal predictors. First three points are uncorrelated and second three points are correlated predictors. LASSO (L), FFSR-NH (N), FFSR-WH (W), FFSR-SH (S), FFSR-NH_{adj} (A), and FFSR-WH_{adj} (D). The standard errors of all plotted points are bounded by 0.02.

Figure 2.

Scaled average difference in actual and optimal performance. Values close to zero are better. True Model (T), Standard Approach (U), LASSO (L), FFSR-SH (S), FFSR-NH_{adj} (A), and FFSR-WH_{adj} (D). The standard errors of all plotted points are bounded by 0.03.

Table 1

Three Steps of Forward Addition for Example of Section 4

Forward Addition Method	Step 1	Step 2	Step 3
Strong Hierarchy	x_8	x_4	x_{17}
No Hierarchy	$x_4 * x_7$	x_8	x_{17}
No Hierarchy-adj	x_8	$x_4 * x_7$	x_{17}
Weak Hierarchy	x_8	x_4	$x_4 * x_{17}$
Weak Hierarchy-adj	x_8	x_4	$x_4 * x_{17}$

Table 2

Models Selected for the Primary Biliary Cirrhosis Data

Variable	Main Effects Only	Strong Hierarchy	No Hierarchy	Adjusted <i>p</i> -val No Hierarchy
bili	1	1	2	1
ascites	2	2		
stage	3	3	3	3
copper	4	4	5	5
albumin	5		7	9
protime	6			
age	7	6		6
sgot	8	8		8
edema	9	7		
gender			8	
ascites*copper		5		7
ascites*edema		9	1	2
gender*edema			4	4
ascites*hepato			6	
spiders*protime			9	
$\hat{a}$	0.063	0.012	0.0032	.026
BIC	979.1	961.3	952.3	946.0

Entries are the order than terms entered the model.

Table 3

Comparison of AME Ratio Means Using Tukey Range Test

Method	Mean	Grouping
FFSR-SH	0.44	A
FFSR-WH _{adj}	0.40	B
FFSR-WH	0.38	C
FFSR-NH _{adj}	0.38	C
FFSR-NH	0.32	D
LASSO	0.21	E
BART	0.18	E

Methods with the same letter are not significantly different.

Standard errors for entries and differences are 0.01 – 0.02.

Table 4

AME Ratio Means for Assessing Factor Interactions

	$R^2 = .25$	$R^2 = .50$	Model 1	Model 2	Model 3
$n = 100$	0.38	0.32	0.24	0.28	0.52
$n = 500$	0.37	0.46	0.36	0.53	0.36
$R^2 = .25$			0.28	0.35	0.49
$R^2 = .50$			0.32	0.47	0.39

Standard errors of entries are at most .02.

Table 5

Optimal Levels for 10-Factor Small Composite Design

Model	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	$\mu(X_{opt})$
1a	-2	2	-2	2	2	-2	2	2	-0.7	-	48.45
1b	-2	2	-2	2	2	-2	2	2	-1.3	-	58.45

Table 6

Optimal Levels for Central 8-Factor Composite Design

Model	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	$\mu(X_{opt})$
2a	1.64	1.03	2	0.64	2	2	2	-	57.35
2b	1.70	-0.86	2	-0.32	2	2	2	-	52.62

Table 7

Variables in Lipase Study

Variable	Name	Measurement Units
X_1	glucose	mg/mL
X_2	palm oil	% by volume (% v/v)
X_3	incubation time	hours
X_4	inoculum density	%
X_5	agitation	rev/min
Y_1	lipase	units/mL (U/mL)
Y_2	specific activity	units/mg (U/mg)

Table 8

Model Summaries for Lipase Production

Method	Effects in Model	R^2	BIC
Rathi et al.	$X_1, X_1^2, X_2, X_2^2, X_3, X_3^2, X_4, X_4^2, X_5, X_5^2$	0.74	112.9
FFSR-SH	No effects	0.00	121.6
FFSR-WH	No effects	0.00	121.6
FFSR-WH _{adj}	No effects	0.00	121.6
FFSR-NH	X_2^2	0.26	115.4
FFSR-NH _{adj}	X_2^2, X_3, X_4^2	0.53	107.6
Standard Approach	$X_2, X_2^2, X_3, X_4, X_4^2$	0.58	110.8

Table 9

Model Summaries for Specific Activity

Method	Effects in Model	R^2	BIC
Rathi et al.	$X_1, X_1^2, X_2, X_2^2, X_3, X_3^2, X_4, X_4^2, X_5, X_5^2$	0.81	179.9
FFSR-SH	X_3	0.17	195.2
FFSR-WH	X_3	0.17	195.2
FFSR-WH _{adj}	X_3	0.17	195.2
FFSR-NH	$X_2, X_2^2, X_3, X_4, X_5^2$	0.75	170.3
FFSR-NH _{adj}	X_2, X_2^2, X_3	0.57	181.2
Standard Approach	$X_2, X_2^2, X_3, X_4, X_4^2, X_5^2$	0.78	173.9

Table 10

Optimal Levels for Maximum Lipase Production

Method	X_1	X_2	X_3	X_4	X_5	$\hat{\mu}_1(\mathbf{X}_{opt})$
Rathi et al.	0.01	0.09	-1	1	0.09	31.11
FFSR-SH	-	-	-	-	-	10.01
FFSR-WH	-	-	-	-	-	10.01
FFSR-WH _{adj}	-	-	-	-	-	10.01
FFSR-NH	-	0	-	-	-	13.67
FFSR-NH _{adj}	-	0	-1	1	-	22.98
Standard Approach	-	0.09	-1	1	-	26.59

Table 11

Optimal Levels for Maximum Specific Activity

Method	X_1	X_2	X_3	X_4	X_5	$\hat{\mu}_2(\mathbf{X}_{\text{opt}})$
Rathi et al.	0.02	0.16	-1	1	0.10	110.99
FFSR-SH	-	-	-1	-	-	59.94
FFSR-WH	-	-	-1	-	-	59.94
FFSR-WH _{adj}	-	-	-1	-	-	59.94
FFSR-NH	-	0.16	-1	1	0	100.63
FFSR-NH _{adj}	-	0.19	-1	-	-	73.61
Standard Approach	-	0.16	-1	-1	0.10	100.45