

Abstract

CREWS, HUGH BATES. Fast FSR Methods for Second-Order Linear Regression Models. (Under the direction of Leonard Stefanski and Dennis Boos.)

Many variable selection techniques have been developed that focus on first-order linear regression models. In some applications, such as modeling response surfaces, fitting second-order terms can improve predictive accuracy. However, the number of spurious interactions can be large leading to poor results with many methods. We focus on forward selection, describing algorithms that use the natural hierarchy existing in second-order linear regression models to limit spurious interactions. We then develop stopping rules by extending False Selection Rate methodology to these algorithms. In addition, we describe alternative estimation methods for fitting regression models including the LASSO, CART, and MARS. We also propose a general method for controlling multiple-group false selection rates, which we apply to second-order linear regression models. By estimating a separate entry level for first-order and second-order terms, we obtain equal contributions to the false selection rate from each group. We compare the methods via Monte Carlo simulation and apply them to optimizing response surface experimental designs.

Fast FSR Methods for Second-Order Linear Regression Models

by
Hugh Crews

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina

April 28, 2008

APPROVED BY:

Dr. Leonard Stefanski (Co-chair)

Dr. Dennis Boos (Co-chair)

Dr. Jason Osborne

Dr. Howard Bondell

Dedication

To my family and friends,

Biography

Hugh Crews was born on March 1, 1982 in Wilmington, NC. After graduating from New Hanover High School in 2000, he began his undergraduate work at North Carolina State University in statistics. In May 2003, Hugh completed his Bachelor of Science degree in statistics. After completing this degree, he decided to continue his graduate work at NCSU. Hugh completed his Master of Statistics degree in May 2005 and began working toward a doctoral degree. After finishing coursework, he began his research under the direction of Dr. Leonard Stefanski and Dr. Dennis Boos. As a college student, Hugh had the opportunity to teach an introductory statistics course, advise undergraduates majoring in statistics, work on an air pollution problem with the Environmental Protection Agency, and spend numerous summers working at SAS.

Acknowledgements

First, I would like to thank my advisors Dr. Leonard Stefanski and Dr. Dennis Boos for their guidance and support throughout my graduate career. As professors, they inspired me to learn statistics, but as advisors, they challenged me to grow as a statistician. I am thankful for my committee members who have taken time to contribute to my research. I would also like to thank Dr. Bill Swallow who advised me through my years as an undergraduate and convinced me to pursue a doctoral degree. I would also like to thank the faculty and staff in the statistics department for making our department a great learning environment.

I would like to thank SAS for employing me many summers and supporting me as an undergraduate SAS Scholar. I would also like to thank GlaxoSmithKline for their generous award which I received in my first year of graduate school. I must also thank the National Science Foundation for supporting research in statistics.

I must thank Ben Ogorek for mentoring me in my first few years as a graduate student. Without his help, I would not have grasped many of the key concepts in statistics. I would also like to thank Clay Barker for mentoring me in the last few years. He answered my numerous questions promptly and offered encouragement. His support these last few years has gotten me through many tough situations. I would like to thank Staples, Jeff, West, Brett, Brent, Scotty, Steve, and Jonah for running with me throughout the years. I am grateful for Steve, Emily, David, Michael, McKay, Darryl, and Joe, all of whom I met in graduate school. Finally, I must thank Mike, Adam, Seth, Micah, Daniel, and Tirey for their friendship throughout my time in

college.

I must thank Marty Sugerik for introducing me to the field of statistics during my time in high school. He was the first teacher to show me that mathematics has many interesting applications.

Most importantly, I must thank my family for their support and encouragement throughout my life. No matter what I pursue, they always show interest and support. I am thankful for my brother and sister who always spend time with me when I visit Wilmington. My father has been a great inspiration to me. He has always trusted me to do what was best for myself and is always available for help, wisdom, and encouragement when I need it.

Table of Contents

List of Tables	x
List of Figures	xiv
1 Modeling Interactions in Regression Analysis	1
1.1 Introduction	1
1.2 Issues with Interactions in Regression	2
1.3 Subset Selection Methods	4
1.4 No Hierarchy	6
1.5 Strong Hierarchy	7
1.6 Weak Hierarchy	9
2 Other Modeling Techniques	11
2.1 Regression with Shrinkage	11
2.1.1 Nonnegative Garrote	12
2.1.2 Least Absolute Shrinkage and Selection Operator	14
2.2 Recursive Partitioning	14
2.2.1 Automatic Interaction Detection	14
2.2.2 Classification and Regression Trees	17
2.2.3 Chi Square Automatic Interaction Detection	19

2.2.4	Other Techniques	21
2.3	Multivariate Adaptive Regression Splines	23
3	Fast False Selection Rate	28
3.1	Introduction to Fast False Selection Rate	28
3.1.1	Calculating Alpha Grid	31
3.1.2	Choosing Alpha	33
3.2	False Selection Rate for Interactions	35
3.3	No Hierarchy with Fast FSR	36
3.4	Hierarchy Methods with Fast FSR	37
3.5	Fast FSR Summary	37
4	Controlling Group False Selection Rates	39
4.1	Forward Selection with Group Entry-Levels	39
4.1.1	Example	41
4.2	Adjustment Method 1	41
4.2.1	Summary of Algorithm for Adjustment Method 1	47
4.3	Adjustment Method 2	47
4.3.1	Summary of Algorithm for Adjustment Method 2	49
4.4	Weak Hierarchy with Adjustment Method 2	49
4.4.1	Summary of Algorithm for Weak Hierarchy with Adjustment Method 2	51
4.5	Method Extensions	52

5	Comparison of Methods	54
5.1	Power and Predictability	54
5.2	Interpretability	55
5.3	Simulation Design	55
5.3.1	Methods and Stopping Rules	56
5.3.2	Data Generation	58
5.3.3	Measures of Interest	60
5.4	Simulation Results	62
5.4.1	Overall Performance by Method	63
5.4.2	Overall Performance by Model	71
5.4.3	Comparison of Other Factors	72
5.4.4	Adjustment Methods	74
6	Applications	77
6.1	Response Surface Study	77
6.1.1	Introduction	77
6.1.2	Simulation Design	79
6.1.3	Measures of Interest	81
6.1.4	Simulation Results	81
6.2	Real Data Examples	85
6.2.1	Cutinase Study	85
6.2.2	Lipase Study	88
6.2.3	Diabetes Study	93

6.2.4	Pyrimidine Study	95
6.2.5	Triazine Study	96
7	Conclusions	104
	Appendix	111
	Appendix A Simulation Results	112
A.1	Results for Chapter 5 Simulation Study	112
A.2	Results for Chapter 6 Simulation Study	112
	Appendix B Response Surface Designs	144
B.1	Coded Response Surface Designs	144

List of Tables

Table 3.1	Example of Forward Sequence	32
Table 3.2	Example of Alpha Grid	35
Table 4.1	Forward Selection Step 1	42
Table 4.2	Forward Selection Step 2	42
Table 4.3	Forward Selection Step 3	43
Table 4.4	Forward Selection Step 4	43
Table 5.1	Methods and Stopping Rules	57
Table 5.2	Method Abbreviations	57
Table 6.1	Optimal Levels for Small Composite Design with 10 Factors	81
Table 6.2	Optimal Levels for Central Composite Design with 8 Factors . . .	81
Table 6.3	Variables in Cutinase Study	85
Table 6.4	Model Summaries for Cutinase Production	87
Table 6.5	Optimal Levels and Maximum Cutinase Production	87
Table 6.6	Variables in Lipase Study	88
Table 6.7	Model Summaries for Lipase Production	89
Table 6.8	Model Summaries for Specific Activity	90
Table 6.9	Type III p -values for Rathi et al. Models	92
Table 6.10	Optimal Levels for Maximum Lipase Production	93
Table 6.11	Optimal Levels for Maximum Specific Activity	93
Table 6.12	Variables in Diabetes Study	94

Table 6.13	Model Summaries for Disease Progression	95
Table 6.14	Variables in Pyrimidine Study	97
Table 6.15	Variables in Pyrimidine Study Continued	98
Table 6.16	Model Summaries for Pyrimidine Activity	98
Table 6.17	Effects Selected for Pyrimidine Activity	99
Table 6.18	Variables in Triazine Study	100
Table 6.19	Variables in Triazine Study Continued	101
Table 6.20	Variables in Triazine Study Continued	102
Table 6.21	Model Summaries for Triazine Activity	102
Table 6.22	Effects Selected for Triazine Activity	103
Table A.1	Comparison of Methods: Normal Uncorrelated, Model 1, $R^2 = 0.25$	113
Table A.2	Comparison of Methods: Normal Uncorrelated, Model 2, $R^2 = 0.25$	114
Table A.3	Comparison of Methods: Normal Uncorrelated, Model 3, $R^2 = 0.25$	115
Table A.4	Comparison of Methods: Normal Correlated, Model 1, $R^2 = 0.25$	116
Table A.5	Comparison of Methods: Normal Correlated, Model 2, $R^2 = 0.25$	117
Table A.6	Comparison of Methods: Normal Correlated, Model 3, $R^2 = 0.25$	118
Table A.7	Comparison of Methods: Chi Square Uncorrelated, Model 1, $R^2 = 0.25$	119
Table A.8	Comparison of Methods: Chi Square Uncorrelated, Model 2, $R^2 = 0.25$	120

Table A.9 Comparison of Methods: Chi Square Uncorrelated, Model 3, $R^2 = 0.25$	121
Table A.10 Comparison of Methods: Chi Square Correlated, Model 1, $R^2 = 0.25$	122
Table A.11 Comparison of Methods: Chi Square Correlated, Model 2, $R^2 = 0.25$	123
Table A.12 Comparison of Methods: Chi Square Correlated, Model 3, $R^2 = 0.25$	124
Table A.13 Comparison of Methods: Normal Uncorrelated, Model 1, $R^2 = 0.5$	125
Table A.14 Comparison of Methods: Normal Uncorrelated, Model 2, $R^2 = 0.5$	126
Table A.15 Comparison of Methods: Normal Uncorrelated, Model 3, $R^2 = 0.5$	127
Table A.16 Comparison of Methods: Normal Correlated, Model 1, $R^2 = 0.5$	128
Table A.17 Comparison of Methods: Normal Correlated, Model 2, $R^2 = 0.5$.	129
Table A.18 Comparison of Methods: Normal Correlated, Model 3, $R^2 = 0.5$.	130
Table A.19 Comparison of Methods: Chi Square Uncorrelated, Model 1, $R^2 = 0.5$	131
Table A.20 Comparison of Methods: Chi Square Uncorrelated, Model 2, $R^2 = 0.5$	132
Table A.21 Comparison of Methods: Chi Square Uncorrelated, Model 3, $R^2 = 0.5$	133

Table A.22 Comparison of Methods: Chi Square Correlated, Model 1, $R^2 = 0.5$	134
Table A.23 Comparison of Methods: Chi Square Correlated, Model 2, $R^2 = 0.5$	135
Table A.24 Comparison of Methods: Chi Square Correlated, Model 3, $R^2 = 0.5$	136
Table A.25 Comparison of Methods for Model 1a	140
Table A.26 Comparison of Methods for Model 1b	141
Table A.27 Comparison of Methods for Model 2a	142
Table A.28 Comparison of Methods for Model 2b	143
Table B.1 Small Composite Design with 10 Factors (Runs 1-25)	145
Table B.2 Small Composite Design with 10 Factors (Runs 26-50)	146
Table B.3 Small Composite Design with 10 Factors (Runs 51-73)	147
Table B.4 Orthogonal Central Composite Design with 8 Factors (Runs 1-25)	148
Table B.5 Orthogonal Central Composite Design with 8 Factors (Runs 26-50)	149
Table B.6 Orthogonal Central Composite Design with 8 Factors (Runs 51-75)	150
Table B.7 Orthogonal Central Composite Design with 8 Factors (Runs 76-100)	151

List of Figures

Figure 3.1	Example of step function for model size by α . \bullet indicates the closed lower bound for α corresponding to a specific model size. \circ indicates the open upper bound for α corresponding to specific model size.	33
Figure 3.2	Example plot of $\hat{\gamma}_{fast}(\alpha)$ by α	36
Figure 5.1	Correlation matrix for simulation study	59
Figure 5.2	AME ratio (larger is better) using Normal predictors for LASSO (L), FFSR-NH (N), FFSR-SH (S), FFSR-WH (W), FFSR-NHA2 (A), and FFSR-WHA2 (D). The first three values in each plot are uncorrelated predictors and the last three values are correlated predictors.	64
Figure 5.3	Average model size for LASSO (L), FFSR-NH (N), FFSR-SH (S), FFSR-WH (W), FFSR-NHA2 (A), and FFSR-WHA2 (D). The first three values in each plot are uncorrelated predictors and the last three values are correlated predictors.	65
Figure 5.4	Average FSR_m for LASSO (L), FFSR-NH (N), FFSR-SH (S), FFSR-WH (W), FFSR-NHA2 (A), and FFSR-WHA2 (D). The first three values in each plot are uncorrelated predictors and the last three values are correlated predictors.	66

Figure 5.5	Average FSR_q for LASSO (L), FFSR-NH (N), FFSR-SH (S), FFSR-WH (W), FFSR-NHA2 (A), and FFSR-WHA2 (D). The first three values in each plot are uncorrelated predictors and the last three values are correlated predictors.	67
Figure 5.6	Average FSR (left) and CSR (right) using Normal predictors for LASSO (L), FFSR-NH (N), FFSR-SH (S), FFSR-WH (W), FFSR-NHA2 (A), and FFSR-WHA2 (D). The first three values in each plot are uncorrelated predictors and the last three values are correlated predictors.	68
Figure 5.7	Average CF for LASSO (L), FFSR-NH (N), FFSR-SH (S), FFSR-WH (W), FFSR-NHA2 (A), and FFSR-WHA2 (D). The first three values in each plot are uncorrelated predictors and the last three values are correlated predictors.	69
Figure 5.8	Comparison of average FSR_m (left) and FSR_q (right) using Normal predictors for FFSR-NHA1 (1), FFSR-NHA2 (2), and FFSR-WHA2 (W). The first three values in each plot are uncorrelated predictors and the second three values are correlated predictors.	76
Figure 6.1	Scaled average difference in actual and optimal performance for Model 1. True Model (T), Standard Approach (U), LASSO (L), FFSR-NH (N), FFSR-SH (S), FFSR-WH (W), FFSR-NHA2 (A), and FFSR-WHA2 (D).	83

Figure 6.2	Scaled average difference in actual and optimal performance for Model 2. True Model (T), Standard Approach (U), LASSO (L), FFSR-NH (N), FFSR-SH (S), FFSR-WH (W), FFSR-NHA2 (A), and FFSR-WHA2 (D).	84
Figure 6.3	Boxplots of simulated R^2 values for lipase production and specific activity	91
Figure A.1	AME ratio (larger is better) using Chi Square predictors for LASSO (L), FFSR-NH (N), FFSR-SH (S), FFSR-WH (W), FFSR-NHA2 (A), and FFSR-WHA2 (D). The first three values in each plot are uncorrelated predictors and the last three values are correlated predictors.	137
Figure A.2	Average FSR (left) and CSR (right) using Chi Square predictors for LASSO (L), FFSR-NH (N), FFSR-SH (S), FFSR-WH (W), FFSR-NHA2 (A), and FFSR-WHA2 (D). The first three values in each plot are uncorrelated predictors and the last three values are correlated predictors.	138
Figure A.3	Comparison of average FSR_m (left) and FSR_q (right) using Chi Square predictors for FFSR-NHA1 (1), FFSR-NHA2 (2), and FFSR-WHA2 (W). The first three values in each plot are uncorrelated predictors and the last three values are correlated predictors.	139

CHAPTER 1

Modeling Interactions in Regression Analysis

1.1 Introduction

Suppose that a researcher is interested in developing a statistical model in order to explain the variation in a response, Y , for some population of interest. The researcher is able to collect data on Y and p predictors X_1, \dots, X_p for a set of n individuals. There are two main goals that could be of interest. The first is understanding the relationships between the response variable and the predictors, and the second is predicting future observations. Putting this in statistical terms, the model of interest is

$$Y = f(X_1, \dots, X_p) + \epsilon \tag{1.1}$$

where ϵ is random error and $f(\cdot)$ is a real-valued function, often a linear function of the predictors. In terms of this model, the first goal is obtaining an easily interpretable estimate of $f(\cdot)$ that can be used in explaining the science behind the variation in the response. The second goal is obtaining the best prediction for $f(\cdot)$ at fixed values of the predictor variables. Depending upon the study at hand, the researcher may focus more on one goal than the other.

A great amount of recent research has been done in the area of variable and model selection in this regression setting. However, the focus of this paper is less on variable and model selection, in general, and more on how to approach model selection in the presence of interactions. Thus, some background information is omitted, and it is assumed that the reader has basic knowledge of model selection techniques. However, a brief discussion of the problem and a summary of the model selection methods can be found in George (2000). For a more in-depth reference, refer to Miller (2002).

Much of the past research has been on simply choosing the best subset of variables in fitting the linear model $f(X_1, \dots, X_p) = \beta_0 + \sum_{j=1}^p \beta_j X_j$. However, in the real world more complex relationships between the response and the predictors exist. Thus the next logical step in the model fitting process is to include all pairwise interactions and quadratic effects, expanding the model of interest to $f(X_1, \dots, X_p) = \beta_0 + \sum_{j=1}^p \beta_j X_j + \sum_{j < k} \beta_{j,k} X_j X_k$. Of course the three-way interactions, four-way interactions, etc., could be added increasing the complexity of the problem. However, merely including the second-order effects greatly increases the space of response surfaces and provides sufficient flexibility for fitting a variety of models.

1.2 Issues with Interactions in Regression

Since linear models are the most widely used and best understood regression models, it is of interest to discuss how these methods handle interactions. We first describe certain problems encountered when fitting linear regression models when interactions are present. These problems reveal the two most common strategies currently employed in practice for handling interactions. The first is to treat each term as a separate

Chapter 1. Modeling Interactions in Regression Analysis

variable, whereas the second is to follow the natural ordering or “hierarchy” of the terms. An in-depth description of these approaches highlighting their advantages and disadvantages appears in later sections.

Interactions are often much harder to detect than main effects in regression modeling. Multicollinearity, measurement error, the numerous forms interactions can take, and lack of power for detection are all problems that make interactions troublesome. Jaccard et al. (1990) suggest centering predictors to reduce multicollinearity. They also argue that standardizing the predictors makes causal inference troublesome. The remainder of their suggestions for dealing with interactions are based on proper planning of experiments and using prior knowledge or scientific theory to aid in the model selection process. However, it is often the case that the researcher is doing exploratory analysis and has very little prior information on which variables affect the response in a simple manner, let alone as a complex interaction. Thus, automatic methods of variable selection such as stepwise regression, best subsets, shrinkage methods, etc., are appealing. They aid in identifying possible informative variables before conducting more controlled and costly studies or experiments.

Another issue that arises when fitting interactions is that significance tests are not invariant to linear transformations of the data. The problem occurs with the significance tests for main effects when corresponding second-order effects are also in the model. Uncentered main effects are often correlated with their second-order effects, and since this correlation changes depending upon the measurement units, the Type III significance tests for the main effects change as well. Therefore, it is possible for two researchers with the same data in different units of measurement using the same method to choose different models. For more explanation of this phenomenon refer to

Griepentrog et al. (1982). As a result, many statisticians recommend keeping main effects in the model if their interaction is significant regardless of whether the main effect is significant. This strategy is known as “maintaining the hierarchy” and is often associated with backward elimination of terms. Another solution to the linear transformation problem is to first standardize the predictors. If all researchers start by standardizing the set of predictors, then they will always choose the same model when applying the same method on the same dataset regardless of their choice of measurement units.

The idea of maintaining the hierarchy leads to the two most popular approaches in the model building process when interactions are of interest. The first is enforcing the hierarchy and the second is not enforcing the hierarchy. Enforcing the hierarchy means that when including interaction and quadratic terms in the model, we require their parent main effects to be included as well. The remainder of this chapter is organized as follows. The next section provides a short discussion on some commonly used searching algorithms: best subsets, forward selection, backward elimination, and stepwise regression. Then the approaches of enforcing the hierarchy and no hierarchy are discussed using forward selection, as well as some issues that arise with each. Finally, a “weak hierarchy” approach is discussed.

1.3 Subset Selection Methods

Many different strategies have been developed for searching through the possible models. We review the most basic and commonly used algorithms. “Best subsets” refers to generating all possible models and then choosing the best one based on some criterion

such as Mallows' C_p , Adjusted R^2 , Akaike's Information Criterion (AIC), Bayesian Information Criterion (BIC), etc. The number of possible models, 2^p , increases exponentially with p . Now consider including all the pairwise interactions and not enforcing the hierarchy. The number of total effects for a full quadratic is

$$p_q = 2p + \binom{p}{2}, \quad (1.2)$$

meaning that the number of subset models is 2^{p_q} . Even when enforcing the hierarchy, the number of total subsets is large. Therefore, it should be obvious that unless p is very small, best subsets is impractical for searching models with interactions.

Forward selection starts with an intercept term and builds a model by adding terms in a successive manner. At each step, the most informative variable is entered into the model. The process continues entering variables until either all have been selected or none meet the specified criterion for entry. Backward elimination starts with a full model and removes uninformative variables one at a time. At each step a variable is removed from the model until a stopping criterion is met or only an intercept term remains. Backward elimination cannot be used when the number of possible terms is greater than the sample size, which occurs more often when modeling interactions. However, a solution to this problem is to use forward selection to build a large model followed by backward elimination to "trim" it down. Another searching algorithm combines forward selection with backward elimination at each step. This method, called stepwise regression, is similar to forward selection in that it starts with an intercept term and then builds a model. However, unlike forward selection, variables in the model can be removed if deemed unnecessary. Thus, at any step, either the

most informative variable is entered or the most uninformative variable is removed depending upon the criterion specified. The process proceeds one step at a time until no variable meets the criterion for entry or removal. When including the pairwise interactions, the number of total effects often outnumbers the sample size. For this reason, forward selection and stepwise regression are obvious choices, since they can be used when $n < p_q$. The methods discussed and developed in this paper are applied to forward selection, but they could be developed using backward elimination (when $n > p_q$) or stepwise regression.

1.4 No Hierarchy

Since it is often impossible to use best subsets or backward elimination, forward selection is commonly used. The simplest approach for fitting interactions using forward selection is the *no hierarchy* approach where no restrictions are placed on the model. Forward selection runs freely and can select any effect whether linear, quadratic, or interaction regardless of the step in the model. That is, each term is a possible candidate for entry at the beginning of the forward selection process. Each term is treated as a separate variable and selected according to the entry criterion.

There are three major problems with not enforcing the hierarchy. The first is the tendency to choose larger and more complex models for a fixed α . As previously noted, each of the terms is treated as a separate variable, so that there are p_q candidate variables for entry on the first step. With so many more candidate variables at each step, there is a higher expected number of selected variables than when only choosing from the p linear terms. For example, assume that forward selection is employed with

α entry-level = α_0 . Then for a model where none of the effects are truly important, the probability for any single term to be entered is α_0 . When selecting from only p linear terms, the expected model size is $1 + p\alpha_0$, where 1 is added for the intercept. When selecting from all linear, quadratic, and pairwise interaction terms at the beginning of the selection process, the expected model size increases to $1 + p_q\alpha_0$. Also note that as p grows, the pairwise interactions outnumber the linear terms, so there is a tendency to choose them. Thus, not only are these models being overfitted, but they are often more complex. A second problem with not maintaining the hierarchy is that the method is not invariant to linear transformations of the predictors. However, henceforth we suggest that the predictors be centered or standardized in order to avoid the lack of invariance.

A third problem is that the quadratic or interaction effects and their corresponding main effects are often highly correlated. This problem is greatly alleviated by centering the predictors before creating the second-order terms. However, for highly skewed predictors, centering does not fully solve the problem. An alternative solution is sweeping out the linear effects from each of the second-order effects and replace them with the corresponding residuals. To sweep out the linear effects for quadratic terms, replace X_j^2 with $\hat{\epsilon}$ from the regression of X_j^2 onto X_j . For interactions, replace X_jX_k with $\hat{\epsilon}$ from the regression of X_jX_k onto X_j and X_k .

1.5 Strong Hierarchy

A standard approach for finding informative interactions is to maintain the hierarchy throughout variable selection. If we enforce the hierarchy with forward selection,

then an interaction cannot be selected until both of its parent main effects are in the model. Similarly, a quadratic term cannot be selected until its parent main effect is in the model. Therefore, given the effects in the model, a candidate set of effects can be formed. Only effects from this candidate set can be entered into the model. An algorithm for forward selection with *strong hierarchy* is given below. Note that the algorithm is limited to second-order terms.

1. Starting with an intercept term in the model, the candidate set includes all main effects X_1, \dots, X_p . Select the most informative effect from this set for model entry.
 2. Suppose X_k is chosen in Step 1. Then the candidate set includes all remaining main effects $X_j, j \neq k$ and the quadratic effect X_k^2 . Select the most informative effect from this set for model entry.
 - 3a. Suppose X_l is chosen at Step 2. Then the candidate set includes all remaining main effects $X_j, j \neq k$ and $j \neq l$, the quadratic terms X_k^2 and X_l^2 , and the interaction $X_k X_l$. Select the most informative effect from this set for model entry.
 - 3b. Suppose X_k^2 is chosen at Step 2. Then the candidate set includes all remaining main effects $X_j, j \neq k$. Select the most informative effect from this set for model entry.
- Proceed accordingly, adding new quadratic effects to the candidate set once their parent main effect is selected and pairwise interactions once both parent main effects are selected for the model.

Peixoto (1987) promotes using this practice of maintaining the hierarchy when searching for the best model by forward, backward, or stepwise regression. He refers to the models that maintain the hierarchy as “well-formulated” and suggests only search-

ing for such models. Stepwise regression would work just like the previous algorithm, except at any given step a main effect cannot be removed unless there are no second-order effects containing it currently in the model.

The advantage of the strong hierarchy is that it both enforces the hierarchy and tends to choose smaller, less complex models since it does not select pairwise interactions until both main effects have entered the model. The disadvantage of the strong hierarchy is that there is no chance for important interaction terms to be included unless both parent main effects are already in the model.

1.6 Weak Hierarchy

A less restrictive alternative to enforcing the hierarchy is to allow pairwise interactions to enter the model once at least one of their parent main effects is included in the model. Again, given the effects in the model, a candidate set of effects can be formed. This method is employed in MARS (Friedman, 1991) and Penalized Logistic Regression (Park and Hastie, 2006). An algorithm for forward selection with *weak hierarchy* is as follows. Note that the algorithm is limited to second-order terms.

1. Starting with an intercept term in the model, the candidate set includes all main effects X_1, \dots, X_p . Select the most informative effect from this set for model entry.
2. Suppose X_k is chosen in Step 1. Then the candidate set includes all remaining main effects $X_j, j \neq k$, the quadratic effect X_k^2 , and all pairwise interactions of the form $X_j X_k, j \neq k$. Select the most informative effect from this set for model entry.
- 3a. Suppose X_l is chosen at Step 2. Then the candidate set includes all remaining

Chapter 1. Modeling Interactions in Regression Analysis

main effects X_j , $j \neq k$ and $j \neq l$, the quadratic effects X_k^2 and X_l^2 , and all pairwise interactions containing either X_k or X_l . Select the most informative effect from this set for model entry.

3b. Suppose X_k^2 is chosen at Step 2. Then the candidate set includes all remaining main effects X_j , $j \neq k$ and all interactions of the form X_jX_k , $j \neq k$. Select the most informative effect from this set for model entry.

3c. Suppose X_lX_k is chosen at Step 2. Then the candidate set includes all remaining main effects X_j , $j \neq k$, the quadratic effect X_k^2 , and all interactions of the form X_jX_k , $j \neq k$ and $j \neq l$. Select the most informative effect from this set for model entry.

Proceed accordingly, considering new quadratic effects once the parent main effect is selected and pairwise interactions once either of the parent main effects is selected for the model.

In terms of selecting interactions, the weak hierarchy approach provides a good balance between enforcing the strong hierarchy and no hierarchy. The major weakness of the strong hierarchy method is that it can miss important interactions, especially in noisy models where both main effects are hard to detect. By making the selection process less restrictive, the weak hierarchy suffers less in these cases. However, there is no guarantee that all informative interactions will enter the candidate set early in the selection process. The cost of making it easier for interactions to enter is that the weak hierarchy still adds a large number of interactions to the candidate set. Therefore, it leads to more complex models than the strong hierarchy when using a fixed α in forward selection.

CHAPTER 2

Other Modeling Techniques

This chapter describes some alternative methods for estimating the regression function $f(X_1, \dots, X_p)$. The first section is a short discussion of regression with shrinkage focusing on the nonnegative garrote, the least absolute shrinkage and selection operator, and some related methods followed by a discussion of recursive partitioning methods leading to Classification and Regression Trees. Finally, Multivariate Adaptive Regression Splines that combines recursive partitioning with splines, is described.

2.1 Regression with Shrinkage

For some data sets, variable selection is not robust in the sense that small changes in the data can result in very different models. Such instability can adversely affect predictions. Therefore, alternative methods have been developed that are more stable. Instead of estimating coefficients for a chosen subset of variables, these methods shrink each variable's coefficient. If some of the coefficients are shrunk to zero, then variable selection is inherently conducted. In some sense, subset selection can be viewed as a shrinkage method that selectively shrinks some coefficients to zero, while estimating the others using ordinary least squares regression.

One of the shrinkage methods is ridge regression. It works by constraining the sum

of squared beta coefficients when solving the normal equations. That is, it minimizes

$$\sum_{i=1}^n (Y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 \quad (2.1)$$

subject to $\sum_{j=1}^p \beta_j^2 \leq s$ where $s \geq 0$. Although ridge regression improves prediction accuracy, its models are more difficult to interpret than subset selection methods since ridge regression models retain all p_q variables.

It is important to choose the right amount of shrinkage. For some variables the estimates change drastically for different values of s . In general, the coefficients are shrunk more for small values of s than large values. A standard way to estimate s is through cross validation or generalized cross validation.

2.1.1 Nonnegative Garrote

One limitation of ridge regression is that it does not perform variable selection, and therefore its models are often difficult to interpret. The nonnegative garrote (Breiman, 1995) was proposed in order to combine variable selection and shrinkage. It minimizes

$$\sum_{i=1}^n (Y_i - \beta_0 - \sum_{j=1}^p c_j \hat{\beta}_j^{ols} x_{ij})^2 \quad (2.2)$$

subject to $c_j \geq 0$ and $\sum_{j=1}^p c_j \leq s$. From (2.2) we see that the nonnegative garrote takes the ordinary least squares estimates, $\hat{\beta}_j^{ols}$, and shrinks them using a different scaling factor for each coefficient. Unlike ridge regression it shrinks some of the coefficients to zero which leads to simpler models. The nonnegative garrote has predictive ability that is competitive with ridge regression.

Chapter 2. Other Modeling Techniques

In its original form, the nonnegative garrote enforces no hierarchy. Therefore, it suffers from the same problems as forward selection with no hierarchy as described in Chapter 1. However, adaptations of the nonnegative garrote have been developed to handle interactions. Both the strong and weak hierarchy approaches have been applied to the nonnegative garrote (Yuan et al., 2007b). With the strong and weak hierarchy algorithms, main effects must be included in order for their interactions and quadratic terms to be included in the model.

For the strong hierarchy, an additional constraint requires the scaling factor for a quadratic term to be less than or equal to the scaling factor for its parent main effect. Similarly, the scaling factor for an interaction must be less than or equal to both scaling factors for its parent main effects. Let c_j be the scaling factor for any main effect X_j and $c_{j,k}$ be the scaling factor for a second-order term X_jX_k . Then the constraints imply that $c_{j,j} \leq c_j$ for quadratic terms, and both $c_{j,k} \leq c_j$ and $c_{j,k} \leq c_k$ for interactions. Any variable is included in the model if its scaling factor is nonzero. Therefore, with the strong hierarchy imposed, any main effect cannot have scaling factor equal zero if the scaling factor for any of its second-order terms is nonzero.

For the weak hierarchy, the constraint is changed so that for interactions the scaling factor must be less than or equal to the sum of the scaling factors for its parent main effects. That is, if X_jX_k is an interaction in the model, then $c_{j,k} \leq c_j + c_k$. As a result, any interaction in the model must have at least one of its parent main effects with a nonzero scaling factor.

2.1.2 Least Absolute Shrinkage and Selection Operator

The least absolute shrinkage and selection operator (LASSO) (Tibshirani, 1996) also combines variable selection and shrinkage. The LASSO works by constraining the sum of absolute beta coefficients when solving the normal equations. It imposes an L_1 penalty on the least squares estimate, whereas ridge regression imposes an L_2 penalty. That is, the LASSO minimizes (2.1) subject to $\sum_{j=1}^p |\beta_j| \leq s$ for some $s \geq 0$. Like the nonnegative garrote, the LASSO can shrink some of the coefficients to zero. An advantage of the LASSO is that it avoids the ordinary least squares estimates that can be adversely affected by multicollinearity.

The LASSO is a special case of least angle regression (LARS) (Efron et al., 2004). Under a simple modification, LARS can be used to obtain the LASSO solution. Like the nonnegative garrote, constraints can be added in order to achieve strong or weak hierarchy with LARS (Yuan et al., 2007a). Some other extensions of the LASSO include SCAD (Fan and Li, 2001) and the Adaptive LASSO (Zou, 2006) which both modify the penalization constraint.

2.2 Recursive Partitioning

2.2.1 Automatic Interaction Detection

Tree-based regression techniques trace their heritage to the method called Automatic Interaction Detection (AID) (Morgan and Sonquist, 1963). The basic idea of the method is very simple: partition or split the data set into groups of similar responses using the predictors to determine the splits. A binary split on the predictor X_i at a

Chapter 2. Other Modeling Techniques

real-value a creates the groups $X_i \leq a$ and $X_i > a$. These groups or nodes are also called t_L and t_R respectively. For each predictor, there exists at most $n - 1$ of these binary splits that result in different groupings of the actual observations. The goal is to choose these splits to best explain the variation in the response. The basic algorithm is given below.

1. Start with a single group t_s containing all observations. Define the unexplained variation for a group t as

$$R(t) = \sum_{\mathbf{X}_i \in t} (Y_i - \bar{Y}_t)^2 \quad (2.3)$$

For t_s , $R(t_s) = \sum_{i=1}^n (Y_i - \bar{Y}_{t_s})^2$ is the total sum of squares.

2. Split the data into two groups, t_L and t_R , that minimize the sum of squared errors. That is, search all predictors for the binary split such that

$$\Delta R = R(t) - R(t_L) - R(t_R) \quad (2.4)$$

is maximized.

3. Make sure that the reduction is greater than 1% of the total sum of squares. That is, $\Delta R > 0.01 \sum_{i=1}^n (Y_i - \bar{Y}_t)^2$. If not, then stop the splitting process.
4. Of the two groups, t_1 and t_2 , select the group t with the largest unexplained variation. Search all possible nonoverlapping splits and choose the one that maximizes (2.4). Split the group if the reduction, ΔR , is greater than 1% of the total sum of squares.
5. Continue searching and splitting groups until no group is more than 2% of the total sum of squares.
6. The estimator used for any final subgroup (or node) is the mean of the observations

in that group. That is,

$$\hat{Y}_t = n_t^{-1} \sum_{X_i \in t} Y_i. \quad (2.5)$$

Sonquist (1969) notes that AID is analogous to the sequential use of one-way Analysis of Variance (ANOVA) where the goal is to split the data into groups whose means explain as much of the variation within the group as possible. He suggests the use of AID to identify important interactions, but still recommends the use of scientific theory to avoid interactions that make no sense. In response to this suggestion, Staelin (1970) promotes the use of Tukey's test for additivity to test for the presence of interactions.

Although AID was met with much enthusiasm, it was by no means flawless. Doyle (1973) noted that:

1. By splitting the data, larger sample sizes are necessary to achieve high power to detect effects.
2. Multicollinearity leads to spurious results.
3. Selection bias is created when building a model using a forward searching algorithm.
4. Selection bias is highly affected by noise such as measurement error in the predictors.
5. Skewed predictors can lead to spurious results.

Like many others, he suggests the use of AID as an exploratory tool to help the researcher specify the form of the model. Similar to forward, backward, and stepwise regression techniques, AID also suffers from the problem that significance testing using the final model is invalid (Kass, 1975). Just like choosing α -to-enter for forward selec-

tion or the amount of smoothing for splines, the 2% stopping rule for AID is arbitrary. A later addition to the algorithm is limiting the minimum size of a final node, which is an arbitrary choice as well. Based on these issues, much of the research for extending tree-based methods involves the process of “growing an honest tree.” The basic elements of tree-based prediction include selecting the variables to use for splitting, how to split, deciding when to stop growing the tree, and determining a method for prediction once the tree is grown. For more history of AID and an updated version called SEARCH refer to Morgan (2005).

2.2.2 Classification and Regression Trees

There have been numerous developments to resolve the issue of how to grow regression trees. A popular method called Classification and Regression Trees (CART) (Breiman et al., 1998) makes use of the original AID algorithm but approaches the problem by growing overly large trees and then pruning them based on a cross-validation approach. CART chooses binary splits of the data like AID, but CART proceeds until every final node has a sample size of at most n_{min} . By choosing a small value for n_{min} , typically $n_{min} = 5$, $R(t)$ is often minimized. The largest tree grown using this stopping rule is referred to as T_{max} .

Despite having the “smallest” unexplained variation, T_{max} is an overgrown tree and needs to be pruned. In order to form subtrees, splits are deleted by remerging the subgroups into a single group in a backward sequential fashion. Since nodes are nonoverlapping, they cannot be removed alone or there will be holes in the prediction space. CART uses an error-complexity measure to find a sequence of subtrees for

choosing the final tree. This measure is defined as

$$R_\lambda(t) = R(T) + \lambda|\tilde{T}| \quad (2.6)$$

where λ is a complexity parameter and $|\tilde{T}|$ equals the number of final subgroups of the tree. $R_\lambda(t)$ can be thought of as unexplained error plus a penalty for complexity. Given a sequence of λ values, $0 = \lambda_1 < \lambda_2 < \dots$, define T_k as the smallest subtree that minimizes $R_\lambda(T)$ for $\lambda_k < \lambda < \lambda_{k+1}$. Note that as λ increases, smaller trees are favored.

In order to choose the best tree from the resulting sequence of subtrees, honest estimates of $R(T_k)$ are needed. This is done with V-fold cross validation. The process works by randomly dividing the data into V subsamples, L_1, \dots, L_V , such that each subsample has the same number of observations (or as close as possible). For $v = 1, \dots, V$, create the training sample that includes all observations not in L_v , and the validation sample, L_v . The training sample is used to grow and prune the particular tree, and the validation sample is used to estimate $R(t_k)$. After repeating this process for all v , the average of each $R(t_k)$ is taken, resulting in the V-fold cross validation estimate. This can also be defined as,

$$R^{CV}(T_k) = n^{-1} \sum_{v=1}^V \sum_{(X_i, Y_i) \in L_v} (Y_i - \bar{Y}^{(v)}(T_k))^2. \quad (2.7)$$

where $\bar{Y}^{(v)}(T_k)$ is the mean for the terminal node corresponding to observation i in the v^{th} validation sample. The overall goal of the pruning process is to choose the subtree

of T_{max} that minimizes $R^{CV}(T_k)$. Define this value as

$$R^{CV}(T_{k_0}) = \min_k R^{CV}(T_k). \quad (2.8)$$

However, T_{k_0} may not have a significantly lower value for R^{CV} than smaller trees. Since smaller trees are easier to interpret and therefore favored, the final tree chosen is the simplest tree whose estimate from (2.7) is within one standard error of $R^{CV}(T_{k_0})$. That is, the actual selected tree is the smallest tree such that

$$R^{CV}(T_k) \leq R^{CV}(T_{k_0}) + SE(R^{CV}(T_{k_0})) \quad (2.9)$$

where $SE(R^{CV}(T_{k_0}))$ is the standard error of the estimate from (2.8).

Understanding the importance of the variables chosen in the splitting process is an important issue. The measure of importance employed in CART is ΔR . That is, the variables that reduce the sum of squared errors the most are deemed the most important. CART uses the basic idea from AID, but implements an intelligent strategy for choosing the right-sized tree for interpretation and prediction. Breiman et al. (1998) emphasize that CART is one method for fitting regression models, but like any other method requires “intelligent and sensible application.”

2.2.3 Chi Square Automatic Interaction Detection

Another method for growing regression trees called Chi Square Automatic Interaction Detection or CHAID (Perreault and Barksdale, 1980) makes use of Bonferroni adjusted Chi-Square tests to determine if a group should be split. The test is an alternative

Chapter 2. Other Modeling Techniques

to the 1% rule in the original AID algorithm. Unlike AID and CART, which only permit binary splits, CHAID allows multiple splits of a single node. This provides more practical and interpretable results, especially for nominal predictors which can often be divided into more than just two distinct groups.

The CHAID algorithm works in the following way. The process begins by determining the number of splits that best describe each predictor. This is accomplished by a process of merging levels of the predictor based on similar response patterns and then splitting them if necessary. Note that if the predictor is continuous, then it must first be partitioned into levels. That is, the researcher must choose ranges of the predictor that make useful groups. The partitioning must be done in such a way that any value of the predictor will fall into only one range. The grouping process starts by a pairwise comparison of each level, merging those that are not significant based on a Chi-Square test. After merging is done, the final group(s) can also be split. The advantage of this additional process is that only predictors that are best described by more than one group are then considered for actually splitting the dataset. Once the number of splits is determined for each predictor, a Bonferroni-adjusted Chi-Square test is conducted for each eligible predictor. The adjustment helps deal with the issue of selection bias where predictors with more possible splits are favored. Assuming at least one of the tests is significant, the predictor with the smallest p -value is then chosen. The dataset is then divided based on the splits of that predictor from the grouping process. After splitting the dataset, the process is then repeated for each subgroup. Once all groups are too small to be split or no predictors are significant from the Chi-Square tests, then the process ends. Like AID, the predicted value for each final subgroup is the average of all its observations. For a more detailed look at the CHAID algorithm and a schematic

Chapter 2. Other Modeling Techniques

diagram of the process, refer to Perreault and Barksdale (1980). Extended Automatic Interaction Detection (XAID) (Heymann, 1981) is an extension of the CHAID algorithm to better handle continuous predictors. It employs two sample t -tests rather than Chi-Square tests in the merging process.

Using Chi-Square testing in a forward, semi-hierarchical manner is reminiscent of forward hierarchical regression presented in Chapter 1. One of the issues akin to such searching procedures is the problem of inference using the final model. However, a computational algorithm that calculates bootstrap standard errors for the predictions and their corresponding confidence intervals has been developed (van Diepen and Franses, 2006).

Computer applications for CHAID and XAID are provided via a program called Formal Inference-based Recursive Modeling (FIRM) (Hawkins, 1999). The documentation for the two algorithms, called CATFIRM and CONFIRM respectively, describes the main issues of recursive partitioning and the differences between FIRM and CART.

2.2.4 Other Techniques

Regression trees can be divided into two main groups: trees that use binary splits and trees that use multilevel splits. AID and CART are in the former group whereas CHAID and XAID are in the latter. There is a long list of similar methods that advance these two basic approaches. QUEST (Loh and Shih, 1997) is a multistage inference-based algorithm for fitting binary trees. By using multiple stages of testing in the variable selection process, selection bias is reduced. GUIDE (Loh, 2002) uses Chi-Square tests followed by bootstrap adjustments of the p -values to practically eliminate selection

Chapter 2. Other Modeling Techniques

bias. Another important aspect of GUIDE is the inclusion of a Chi-Square test for detecting pairwise interactions. CRUISE (Kim and Loh, 2001) uses multilevel splits, linear combination splits, and GUIDE's interaction test to produce less complex trees with low selection bias. Finally, LOTUS (Chan and Loh, 2004) is an algorithm for fitting logistic regression trees. Many more partitioning or tree-based algorithms exist, most of which deal with the issue of selection bias.

So far, the trees grown use the mean of each terminal node as the prediction for that group. However, other methods exist that make use of other estimators. Treed regression (Alexander and Grimshaw, 1996) grows less complex binary trees than the previously discussed methods. Instead of using the mean for prediction, the best simple linear regression of the group using one predictor is alternatively fit, and predictions are made using it. The goal of treed regression is to create less complex, more easily interpreted tree models.

A resampling method for increasing the accuracy of regression trees was developed by Brieman (1996). The method, called bagging, takes bootstrap samples from the original sample, and then grows regression trees using them. Finally it takes the average of the bootstrap predictors as the final predictor. Random Forests (Brieman, 2001) generalizes the idea of growing a large number of trees through resampling techniques. Brieman (2001) proves, via the Strong Law of Large Numbers, that as the number of trees grown increases, the final estimate converges to the right-sized tree and thus does not overfit. Although resampling methods can improve accuracy of predictions, they complicate the model interpretation.

A shrinkage method for growing regression trees is discussed in (LeBlanc and Tibshirani, 1998). The method uses a shrinkage constraint similar to that employed by

the LASSO in order to both prune branches of the tree and shrink the node estimates.

2.3 Multivariate Adaptive Regression Splines

As previously discussed, much of the need for automatic model selection techniques is for researchers who are unable or unwilling to specify the form of the regression model. Although recursive partitioning methods prove useful in detecting complex interactions, they do not produce continuous models and have trouble handling simple linear, additive, or interaction models of lower order. In fact, they cannot produce an additive model with more than one main effect. Unlike partitioning methods, splines produce continuous function estimates. Splines are a powerful nonparametric regression technique used to estimate smooth functions without specifying the form of the model. However, they suffer from the need for high-dimensional basis functions to estimate even simple functions. Multivariate Adaptive Regression Splines (MARS) was developed by Friedman (1991) in order to combine the adaptability of partitioning with the smoothness of splines. In this section, an overview of MARS and some enhancements to it are presented. For a review of spline estimation, refer to Eubank (1999).

The first step in understanding MARS is to see its relationship to the recursive partitioning methods previously discussed. The partitioning methods were simply described as splitting the data into subregions and then making predictions locally for each terminal node. This lead to the tree-based view referred to as regression trees. However, the partitioning methods can also be written as the sum of products of indicator functions and regression coefficients. The indicator functions identify the path to take along the tree to get to a specific node. By thinking of the indicators as basis

Chapter 2. Other Modeling Techniques

functions, the partitioning methods can be viewed as stepwise regression procedures, where the goal is to find a set of basis functions whose corresponding estimator best describes the data. Suppose there are $m = 1, \dots, M$ subregions. Then,

$$\hat{f}(\mathbf{X}) = \sum_{m=1}^M a_m B_m(\mathbf{X}) \quad (2.10)$$

where a_m is the corresponding estimator for the m^{th} subregion. $B_m(x)$ is the m^{th} basis function defined by

$$B_m(\mathbf{X}) = I[\mathbf{X} \in R_m] \quad (2.11)$$

with R_m representing the terminal region of interest. The right hand side of (2.11) can be broken further into the product of the K_m indicator functions which give the direct path to the specific region of interest. In order to see this, suppose that X_j is the predictor used in the splitting process, and τ is the splitting point or knot. Also define,

$$S_{k,m} = \begin{cases} 1 & \text{if right path,} \\ -1 & \text{if left path} \end{cases} \quad (2.12)$$

Then,

$$B_m(\mathbf{X}) = \prod_{k=1}^{K_m} I\{S_{k,m}(X_{j(k,m)} - \tau_{k,m}) \geq 0\} \quad (2.13)$$

where $S_{k,m}$ tells which path, left or right, to take to get to a particular node, and $(X_j - \tau_{k,m})$ tells the path that the observation takes. If these paths match, then the indicator function evaluates to 1. Otherwise, it evaluates to 0. Thus, by taking the product of these indicators, the resulting basis is itself an indicator as specified in (2.11).

Chapter 2. Other Modeling Techniques

In order to generalize the procedure, MARS substitutes the q^{th} order spline basis function $[S_{k,m}(X_{j(k,m)} - \tau_{k,m})]_+^q$ for the indicator function $I\{S_{k,m}(X_{j(k,m)} - \tau_{k,m}) \geq 0\}$ in (2.13). This leads to multivariate spline basis functions which take the form

$$B_m^q(\mathbf{X}) = \prod_{k=1}^{K_m} [S_{k,m}(x_{p(k,m)} - \tau_{k,m})]_+^q. \quad (2.14)$$

The advantage of using such basis functions is that if $q > 0$, they lead to continuous estimates of the regression surface. Taking $q = 0$ in equation (2.14) yields (2.13), i.e. the usual regression tree.

One of the major drawbacks of partitioning techniques is that simple functions are not handled well. This is a direct consequence of limiting splitting to only currently terminal nodes. As the level of splitting increases, the degree of the interaction increases, yielding models with higher order interactions. In order to overcome this problem, MARS allows all basis functions, whether parent or daughter, to be considered throughout the selection process. For example, to fit a purely additive function, the parent basis can be chosen multiple times using different variables. As a result, MARS can handle linear, additive, and other simple functions as well as more complex interactions and nonlinear models.

Except for $q = 0$, the product of two basis functions involving X_j does not yield a spline basis of power q . In order to solve this problem, a final addition to the MARS procedure is that a basis cannot be split on the same variable more than once. Instead the same basis and variable combination can be chosen multiple times to introduce additional terms without increasing the level of splitting.

Employing these major changes to the usual partitioning methodology, MARS pro-

ceeds in a forward stepwise fashion building a large set of basis functions to approximate the overall model. Like CART, MARS employs this forward step followed by a backward deletion. However, instead of deleting whole splits, MARS is able to delete specific basis functions if they are deemed unnecessary. This is due to the fact that the parent basis is not deleted to create daughter bases. Instead, the parent is maintained and considered again in the selection process. A consequence is that the basis functions overlap and can be removed individually without leaving holes in the predictor space.

Once MARS chooses a set of basis functions the model, can be rewritten from the form in (2.10) to a more interpretable form. Using an ANOVA-like decomposition, the estimator can be broken up into a constant basis, plus basis functions involving only one variable, plus basis functions involving two variables, and so forth. By rearranging the terms in such a manner the estimator is defined as

$$\hat{f}(\mathbf{X}) = a_0 + \sum_{K_m=1} f_j(X_j) + \sum_{K_m=2} f_{j,k}(X_j, X_k) + \sum_{K_m=3} f_{j,k,l}(X_j, X_k, X_l) + \dots \quad (2.15)$$

Using (2.15), it is easy to see which variables interact and to what degree. This makes for much easier interpretation than was present with regression trees.

Although there are many more specific details on the MARS procedure, such as its cost complexity function, knot optimization, etc., they are excluded, and the reader is referred to Friedman (1991). Some advancements to MARS include POLYMARS (Koopberg et al., 1997), which is a customized version used for fitting logistic regression models. Another advancement makes use of linear discriminant analysis to help prespecify linear combinations of the predictors to include in the model selection process (Zhang et al., 2003). By identifying important linear combinations beforehand,

Chapter 2. Other Modeling Techniques

MARS becomes a more powerful tool for fitting nonlinear functions.

CHAPTER 3

Fast False Selection Rate

With forward selection, the probability of any variable being included in the final model can be broken into two parts: the probability of the variable being a candidate for entry and the probability of the variable entering the model given it is a candidate for entry. That is, for an variable X_j ,

$$P(X_j \text{ in final model}) = P(X_j \text{ in candidate set})P(X_j \text{ is selected} \mid X_j \text{ in candidate set}). \quad (3.1)$$

Equation (3.1) holds true for both informative and uninformative variables. In this chapter, we shift our focus to controlling the entry of uninformative variables into the final model.

3.1 Introduction to Fast False Selection Rate

Wu et al. (2007) developed methods for controlling the proportion of uninformative variables in the final model using forward selection. This proportion is referred to as the false selection rate (FSR). By adding phony variables into the variable selection process, they keep track of the proportion in the final model for different levels of the tuning parameter. This information is used to estimate the correct amount of tuning to achieve a desired false selection rate. In this section, we develop “Fast” FSR methods

Chapter 3. Fast False Selection Rate

which require no phony variable generation.

The empirical false selection rate is defined as the proportion of total variables in a fitted regression model that are uninformative. In general, a variable X_j is deemed informative if in the true model $\beta_j \neq 0$ and uninformative if $\beta_j = 0$. For any dataset with response vector, \mathbf{Y} , and matrix of predictors, \mathbf{X} , we calculate the empirical false selection rate as

$$\gamma(\mathbf{Y}, \mathbf{X}) = \frac{U(\mathbf{Y}, \mathbf{X})}{1 + U(\mathbf{Y}, \mathbf{X}) + I(\mathbf{Y}, \mathbf{X})}, \quad (3.2)$$

where $U(\mathbf{Y}, \mathbf{X})$ is the number of uninformative variables and $I(\mathbf{Y}, \mathbf{X})$ is the number of informative variables. In order to avoid division by 0 and account for the intercept term, 1 is added to the number of total variables. By adding 1 to the denominator, the empirical false selection rate can also be viewed as the proportion of total regression degrees of freedom that are unnecessarily spent, or the proportion of estimated beta coefficients that are truly zero. That is,

$$\gamma(\mathbf{Y}, \mathbf{X}) = \frac{\text{false regression df}}{\text{total regression df}}. \quad (3.3)$$

Suppose we desire the empirical false selection rate to equal some value, say γ_0 . Since we are unlikely to hit this target for every dataset, we would instead like to achieve this level in the long run. Therefore, given γ_0 , a properly tuned model selection process would satisfy $\gamma = \gamma_0$, where

$$\gamma = E\{\gamma(\mathbf{Y}, \mathbf{X})\}. \quad (3.4)$$

Our goal is to choose the proper tuning to achieve the desired false selection rate. In or-

Chapter 3. Fast False Selection Rate

der to do so, we use information from the particular dataset of interest to best estimate (3.4) for different values of the tuning parameter. In this paper we focus on forward selection with α -to-enter as the tuning parameter. Before providing more information on estimating γ , some definitions are necessary. For a specific data set denote:

K_t = total number of variables.

$U(\alpha)$ = number of uninformative variables given α .

$I(\alpha)$ = number of informative variables given α .

$S(\alpha) = U(\alpha) + I(\alpha)$.

From the definitions provided, we see that given any α , the expected false selection rate is

$$\gamma(\alpha) = E \left\{ \frac{U(\alpha)}{1 + S(\alpha)} \right\}. \quad (3.5)$$

In order to estimate γ we can focus on each part of equation (3.5). First, we know there are $S(\alpha)$ variables in the model for any α . The difficult part is deciding what proportion are uninformative. We turn to the definition of expectation to estimate this quantity. Assuming the uninformative variables are independent, the expected number in the final model is equal to the probability of one entering times the number being considered for entry. The probability of any uninformative variable, X_j , entering the model is some unknown function of α , say $\theta(\alpha)$. That is,

$$P(X_j \text{ is selected} \mid X_j \text{ in candidate set}) = \theta(\alpha).$$

If we assume that only the $S(\alpha)$ variables in the current model are informative then

we can estimate $U(\alpha)$ as $K_t - S(\alpha)$. This leads to the FSR estimate,

$$\hat{\gamma}(\alpha) = \frac{\{K_t - S(\alpha)\}\theta(\alpha)}{1 + S(\alpha)}. \quad (3.6)$$

When using an α entry-level, a variable is only selected if its corresponding p -value is less than α . If we assume that the p -values for uninformative variables follow independent Uniform(0,1) distributions, then the probability that any p -value is less than α is in fact α . Therefore, we set $\theta(\alpha) = \alpha$ leading to the Fast FSR estimate,

$$\hat{\gamma}_{fast}(\alpha) = \frac{\{K_t - S(\alpha)\}\alpha}{1 + S(\alpha)}. \quad (3.7)$$

3.1.1 Calculating Alpha Grid

Now that we have a formula for estimating γ , our next goal is to determine a grid of possible choices for α . For any particular dataset, the model size, and therefore false selection rate, only changes at certain values of α . These values correspond to the p -values of the variable being included at a given step of the selection process. While obtaining the forward sequence, we can keep track of the p -values to enter at each step. In particular, the maximum p -value of the current and previous steps is of interest. This list of maximum p -values can be used as the possible estimates for α .

For example, suppose that Table 3.1 gives the 15 steps of a forward selection sequence. Clearly $S(\alpha) = 0$ if $\alpha = 0$, since we allow no variables to enter the model. Based on the p -value in step 1, we can see that $S(\alpha)$ does not change until $\alpha \geq 0.003$. Skipping to step 3, we see that the p -value to enter was 0.002. However, in order to get to step 3, α needs to be at least 0.005. Thus the maximum p -value at step 3 is

Table 3.1: Example of Forward Sequence

Step	p -value	Max p -value	Model Size
0	0.000	0.000	1
1	0.003	0.003	2
2	0.005	0.005	3
3	0.002	0.005	4
4	0.013	0.013	5
5	0.025	0.025	6
6	0.023	0.025	7
7	0.049	0.049	8
8	0.075	0.075	9
9	0.063	0.075	10
10	0.189	0.189	11
11	0.222	0.222	12
12	0.307	0.307	13
13	0.346	0.346	14
14	0.564	0.564	15
15	0.701	0.701	16

0.005. Once α gets to 0.005, the model size jumps from 2 to 4. An easy way to see the grid and resulting model sizes is by plotting $1 + S(\alpha)$ by α . Figure 3.1 gives this plot for the example data in Table 3.1. From the plot we can see that in order to obtain the model sizes, $\{1, 2, 4, 5, 7, 8, 10, 11, 12, 13, 14, 15, 16\}$, the minimum values of α necessary are $\{0, 0.003, 0.005, 0.013, 0.025, 0.049, 0.075, 0.189, 0.222, 0.307, 0.346, 0.564, 0.701\}$.

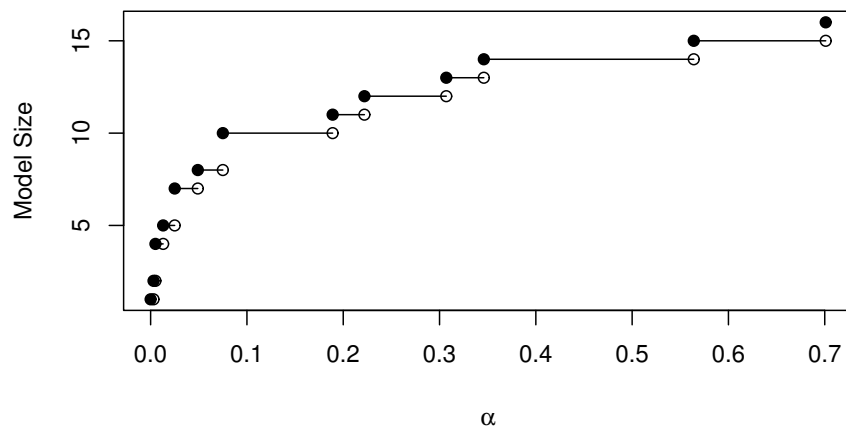


Figure 3.1: Example of step function for model size by α . ● indicates the closed lower bound for α corresponding to a specific model size. ○ indicates the open upper bound for α corresponding to specific model size.

3.1.2 Choosing Alpha

Once we have our grid of choices for α , we can plug these values into (3.7). Rather than choosing the value which yields $\hat{\gamma}_{fast}(\alpha)$ closest to γ_0 , we choose the largest α such that $\hat{\gamma}_{fast}(\alpha) \leq \gamma_0$. The reasoning is that by choosing the largest model meeting our false selection goal, a larger proportion of the informative variables are included. When $\alpha = 0$, no variables are entered into the model leading to $\hat{\gamma}_{fast} = 0$. But also when $S(\alpha) = K_t$, $\hat{\gamma}_{fast} = 0$ as well. These are the cases of fitting a null model and a full model respectively. Of course we do not believe that the false selection rate of the full model is always 0. Otherwise there would be no need for the variable selection

Chapter 3. Fast False Selection Rate

process. Therefore, constraints must be added when choosing α to avoid fitting the full model every time. Since $\hat{\gamma}_{fast}$ starts and ends at 0 and is nonnegative, there must be a maximum. Call the α entry-level corresponding to this maximum α_{max} . Beyond it, the false selection rates tend to be underestimated with the extreme case being $\hat{\gamma}_{fast} = 0$ for the full model. Therefore, the necessary constraint is to limit our choice of α to those which do not exceed this point. This leads to the estimate,

$$\hat{\alpha} = \sup_{\alpha \leq \alpha_{max}} \{\alpha : \hat{\gamma}_{fast}(\alpha) \leq \gamma_0\}. \quad (3.8)$$

Although the continuous interval $[0,1]$ is used in (3.8), it is sufficient to search over α from the discrete grid. Choosing α from the grid yields an expected false selection rate less than γ_0 , but results in the same model as choosing over the interval $[0,1]$. In general, if α_d is the value from the discrete grid that solves (3.8), then the solution over $[0, 1]$ is

$$\hat{\alpha} = \frac{\gamma_0\{1 + S(\alpha_d)\}}{K_t - S(\alpha_d)}.$$

Using $K_t = 15$ for the data in Table 3.1, the grid and corresponding estimates of false selection rate can be found in Table 3.2. Also, Figure 3.2 gives the corresponding plot of $\hat{\gamma}_{fast}(\alpha)$ by α . Notice that $\hat{\gamma}_{fast}$ peaks at 0.0859 when $\alpha = 0.189$. Using the discrete grid to solve (3.8), we get $\hat{\alpha} = 0.075$, whereas solving (3.8) on $[0,1]$ yields $\hat{\alpha} \approx 0.083$. In either case, the corresponding fitted regression model would contain the first 9 variables and an intercept term.

Table 3.2: Example of Alpha Grid

α	$\hat{\gamma}_{fast}(\alpha)$	$S(\alpha)$
0.000	0.0000	0
0.003	0.0210	1
0.005	0.0150	3
0.013	0.0286	4
0.025	0.0321	6
0.049	0.0490	7
0.075	0.0450	9
0.189	0.0859	10
0.222	0.0740	11
0.307	0.0708	12
0.346	0.0494	13
0.564	0.0376	14
0.701	0.0000	15

3.2 False Selection Rate for Interactions

When considering higher-order effects, the way we approach uninformative variables changes somewhat. In Chapter 1, a discussion was given on how different units for predictors can lead to different results when using a model selection process. It was then suggested that the original predictors be standardized before creating the derived variables. This would lead an experimenter to the same \mathbf{X} matrix no matter what units he used for measurement. Given that the original predictors are standardized, we assume that the hierarchy is maintained. Therefore, if an interaction or quadratic effect is informative, then its parent main effects are informative as well. Using this

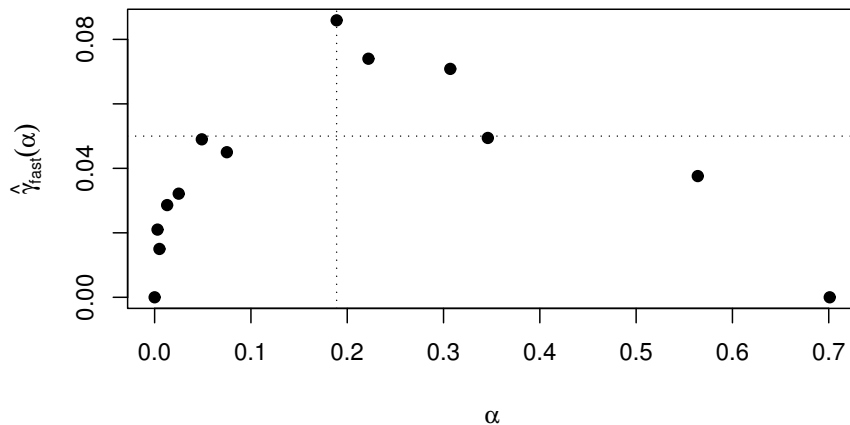


Figure 3.2: Example plot of $\hat{\gamma}_{fast}(\alpha)$ by α

information, we now develop the Fast FSR formulas for the algorithms from Chapter 1.

3.3 No Hierarchy with Fast FSR

In the no hierarchy approach, the Fast FSR methodology works exactly as previously described. Each effect is a candidate for entry at the beginning of the forward selection process. Therefore, the probability of any effect being in the candidate set is 1, and the probability of any uninformative effect being included in the final model is α . Therefore, equation (3.7) can be used to estimate the false selection rate with $K_t = p_q$.

3.4 Hierarchy Methods with Fast FSR

In (3.7), we use $K_t - S(\alpha)$ to estimate the number of uninformative variables. Since we start with K_t variables in the candidate set under no hierarchy, the estimated number of uninformative variables is also the number of variables in the candidate set after $S(\alpha)$ variables have entered the model. For the hierarchy methods, all variables are not in the candidate set when the selection process begins. Instead, the number of candidate variables depends on α . At any step of the forward selection process, it is possible to count the number of effects in the candidate set. If we let $C(\alpha)$ equal the total number of effects ever in the candidate set for a given α and assume that all effects currently in the candidate set are uninformative, then we can replace $K_t - S(\alpha)$ with $C(\alpha) - S(\alpha)$. This leads to the Fast FSR formula for the hierarchy methods,

$$\hat{\gamma}_{fast}(\alpha) = \frac{\{C(\alpha) - S(\alpha)\}\alpha}{1 + S(\alpha)}. \quad (3.9)$$

Notice that $C(\alpha)$ is monotonically increasing to K_t , implying that (3.7) is a specific case of (3.9) where $C(\alpha) = K_t$ for all α .

3.5 Fast FSR Summary

The Fast FSR process can be summarized in the following steps:

1. Pick a selection method and a target false selection rate, i.e. $\gamma_0 = 0.05$.
2. Obtain the forward sequence using the particular method, keeping track of the p -value at each step.
3. Using the maximum p -values, figure out the proper grid of α levels.

Chapter 3. Fast False Selection Rate

4. Using this grid, calculate $\hat{\gamma}_{fast}(\alpha)$ for each α using the Fast FSR formula corresponding to the particular method.
5. Find α where $\hat{\gamma}_{fast}$ peaks to obtain $\hat{\alpha}_{max}$.
6. Calculate $\hat{\alpha}$ using (3.8).
7. Fit the model corresponding to this entry-level.

CHAPTER 4

Controlling Group False Selection Rates

Recall from equation (3.1), $P(\text{effect in final model}) = P(\text{effect in candidate set})P(\text{effect is selected} \mid \text{effect in candidate set})$. The hierarchy methods discussed in Chapters 1 and 3 focused on the first part of the equation by making the probability of being a candidate for entry dependent upon the type of effect. Requiring main effects to enter the model before second-order effects can enter greatly reduces the number of quadratic and interaction effects included in the model. A disadvantage of enforcing the hierarchy is that informative effects are never in the candidate set for many α entry-levels. In order to avoid this problem, we turn our focus to the second part of (3.1). By using separate entry-levels for main effects and second-order effects, we can control the selection of each type of effect, while still allowing each effect to enter at Step 1.

4.1 Forward Selection with Group Entry-Levels

Suppose we want to run forward selection using separate α entry-levels for main effects and second-order effects. That is, we choose α_1 for main effects and α_2 for second-order effects and want a corresponding model. In general, we choose $\alpha_2 < \alpha_1$ to limit the number of second-order effects in our model.

In ordinary forward selection, we use one entry-level and select the effect from the

Chapter 4. Controlling Group False Selection Rates

candidate set with the smallest p -value to enter. When using forward selection with group entry-levels, we propose to select the effect with the smallest p -value after adjusting the p -values of the second-order terms. Given α_1 and α_2 , we can find c such that $\alpha_1 = c\alpha_2$. Therefore, we can multiply the p -values for second-order effects by c to put them on the same scale as the p -values for main effects. Assuming $\alpha_2 < \alpha_1$, c measures how much more we want main effects to be selected than second-order effects. Once we adjust all p -values to the same scale, we can simply choose the smallest one. An algorithm for forward selection using α_1 for main effects and α_2 for second-order effects is given below.

1. Starting with an intercept term in the model, calculate the Type III p -values for adding any single effect from the candidate set. The candidate set consists of all main effects, interactions, and quadratic terms. Call these p -values $\{p_1, \dots, p_{K_t}\}$. Define the adjusted p -values as

$$\tilde{p}_i = \begin{cases} p_i & \text{if } i \in \mathcal{M}, \\ cp_i & \text{otherwise} \end{cases} \quad (4.1)$$

where $c = \alpha_1/\alpha_2$ and \mathcal{M} corresponds to the main effects. Select the effect, $X_{(1)}$, corresponding to $\tilde{p}_{(1)}$, the smallest adjusted p -value.

2. With $X_{(1)}$ and an intercept term in the model, calculate the Type III p -values for adding any single effect remaining in the candidate set. Next, calculate the adjusted p -values using (4.1) and select the effect, $X_{(2)}$, corresponding to the smallest adjusted p -value.

3. Repeat this process until no adjusted p -value is less than or equal to α_1 .

4.1.1 Example

Suppose we have three predictor variables and want to run forward selection using $\alpha_1 = 0.05$ and $\alpha_2 = 0.01$. Using these entry-levels, we get $c = 5$. Table 4.1 contains the p -values and adjusted p -values for adding any one of the nine total effects into the model. We see that X_1^2 has the smallest adjusted p -value at 0.020. Since this value is smaller than 0.05, we enter X_1^2 into the model.

Table 4.2 contains the p -values and adjusted p -values for adding any one of the eight remaining effects into the model. We see that X_1 has the smallest adjusted p -value at 0.005. Since this value is smaller than 0.05, we enter X_1 into the model.

Table 4.3 contains the p -values and adjusted p -values for adding any one of the seven remaining effects into the model. We see that X_2 has the smallest adjusted p -value at 0.022. Since this value is smaller than 0.05, we enter X_2 into the model.

Table 4.4 contains the p -values and adjusted p -values for adding any one of the six remaining effects into the model. We see that X_1X_2 has the smallest adjusted p -value at 0.125. However, since this value is larger than 0.05, we stop forward selection without entering X_1X_2 into the model. Therefore, our final model using $\alpha_1 = 0.05$ and $\alpha_2 = 0.01$ contains an intercept, X_1 , X_1^2 , and X_2 .

4.2 Adjustment Method 1

Our goal in Chapter 3 was to choose α_0 such that $E[FSR] = \gamma_0$. Suppose we want to choose α_1 and α_2 such that the contribution to the false selection rate from main effects, FSR_m , is equal to the contribution to the false selection rate from second-order

Chapter 4. Controlling Group False Selection Rates

Table 4.1: Forward Selection Step 1

Variable	p -value	Adjusted p -value
X_1	0.031	0.031
X_2	0.064	0.064
X_3	0.279	0.279
X_1^2	0.004	0.020
X_2^2	0.732	3.660
X_3^2	0.581	2.905
X_1X_2	0.008	0.040
X_1X_3	0.232	1.160
X_2X_3	0.109	0.545

Table 4.2: Forward Selection Step 2

Variable	p -value	Adjusted p -value
X_1	0.005	0.005
X_2	0.032	0.032
X_3	0.393	0.393
X_2^2	0.549	2.745
X_3^2	0.612	3.060
X_1X_2	0.011	0.055
X_1X_3	0.472	2.360
X_2X_3	0.184	0.920

Table 4.3: Forward Selection Step 3

Variable	p -value	Adjusted p -value
X_2	0.022	0.022
X_3	0.410	0.410
X_2^2	0.592	2.960
X_3^2	0.391	1.955
X_1X_2	0.010	0.050
X_1X_3	0.317	1.585
X_2X_3	0.166	0.830

Table 4.4: Forward Selection Step 4

Variable	p -value	Adjusted p -value
X_3	0.360	0.360
X_2^2	0.491	2.455
X_3^2	0.502	2.510
X_1X_2	0.025	0.125
X_1X_3	0.334	1.670
X_2X_3	0.105	0.525

effects, FSR_q . That is, we want

$$E[FSR_m] = E[FSR_q] \tag{4.2}$$

and

$$E[FSR_m] + E[FSR_q] = \gamma_0. \tag{4.3}$$

Chapter 4. Controlling Group False Selection Rates

We know that for any α_1 and α_2 , there exists c such that $\alpha_1 = c\alpha_2$. In order to get a forward sequence of effects using the algorithm above, we need to know c . One approach is to find c that satisfies (4.2), run forward selection using the algorithm above, and then find α_1 to satisfy (4.3).

In order to estimate c , we assume that there are U_m uninformative main effects and U_q uninformative second-order effects in the candidate set. If we use forward selection with fixed α_1 and α_2 , then

$$E[FSR_m] = \frac{U_m\alpha_1}{1 + S(\alpha_1, \alpha_2)} \quad (4.4)$$

and

$$E[FSR_q] = \frac{U_q\alpha_2}{1 + S(\alpha_1, \alpha_2)}, \quad (4.5)$$

where $S(\alpha_1, \alpha_2)$ is the number of effects selected. Using (4.4) and (4.5) to solve (4.2) yields

$$\frac{U_m\alpha_1}{1 + S(\alpha_1, \alpha_2)} = \frac{U_q\alpha_2}{1 + S(\alpha_1, \alpha_2)}. \quad (4.6)$$

Substituting $\alpha_1 = c\alpha_2$ into (4.6) leads to

$$\frac{U_m c\alpha_2}{1 + S(\alpha_1, \alpha_2)} = \frac{U_q\alpha_2}{1 + S(\alpha_1, \alpha_2)}. \quad (4.7)$$

Finally, solving for c in (4.7) leads to

$$c = \frac{U_q}{U_m}. \quad (4.8)$$

Chapter 4. Controlling Group False Selection Rates

Therefore, c is a ratio of uninformative second-order effects to uninformative main effects. In order to avoid division by zero, we add 1 to the numerator and denominator, leading to

$$c = \frac{1 + U_q}{1 + U_m}. \quad (4.9)$$

In the Fast FSR methodology, we assume all effects in the model are informative and all effects in the candidate set are uninformative. Therefore, with only an intercept in the model we have p uninformative main effects and $K_t - p$ uninformative second-order effects. Using these values, an initial estimate for c is

$$c^{(1)} = \frac{1 + K_t - p}{1 + p}. \quad (4.10)$$

After initially estimating c , we use forward selection with group entry-levels to get a sequence of effects, $\{X_{(1)}, \dots, X_{(K_t)}\}$, and a corresponding sequence of adjusted p -values, $\{\tilde{p}_{(1)}, \dots, \tilde{p}_{(K_t)}\}$. Using the sequence of adjusted p -values, we obtain a sequence of maximum adjusted p -values, $\{\tilde{\tilde{p}}_{(1)}, \dots, \tilde{\tilde{p}}_{(K_t)}\}$, where

$$\tilde{\tilde{p}}_{(i)} = \max\{\tilde{p}_{(1)}, \dots, \tilde{p}_{(i)}\}. \quad (4.11)$$

As described in Chapter 3, we use the sequence of maximum adjusted p -values as our grid of α levels. The next step is to calculate the estimated false selection rates using the α levels in the grid. The Fast FSR formula when using group entry-levels is

$$\hat{\gamma}_{fast}(\alpha, c) = \frac{\{p - S_m(\alpha, c)\}\alpha + \{K_t - p - S_q(\alpha, c)\}\{\alpha/c\}}{1 + S(\alpha, c)}, \quad (4.12)$$

Chapter 4. Controlling Group False Selection Rates

where $S_m(\alpha, c)$ is the number of main effects selected and $S_q(\alpha, c)$ is the number of interactions and quadratic terms selected given α and c . That is,

$$S_m(\alpha, c) = \sum_{i=1}^{K_t} I(\tilde{p}_{(i)} \leq \alpha) I(i \in \mathcal{M}) \quad (4.13)$$

and

$$S_q(\alpha, c) = \sum_{i=1}^{K_t} I(\tilde{p}_{(i)} \leq \alpha) I(i \in \mathcal{M}^c). \quad (4.14)$$

In order to select a model, we need to choose α_1 . We want to choose the largest value of α such that $\hat{\gamma}_{fast}(\alpha, c) \leq \gamma_0$. As discussed in Chapter 3, we limit our choice to $\alpha \leq \alpha_{max}$, where α_{max} is the entry-level from the grid where $\hat{\gamma}_{fast}(\alpha, c)$ is largest. Therefore, estimate α_1 using

$$\hat{\alpha}_1 = \sup_{\alpha \leq \alpha_{max}} \{\alpha : \hat{\gamma}_{fast}(\alpha, c) \leq \gamma_0\}. \quad (4.15)$$

Corresponding to $\hat{\alpha}_1$ and $c^{(1)}$, we have a model containing $S_m(\hat{\alpha}_1, c^{(1)})$ main effects and $S_q(\hat{\alpha}_1, c^{(1)})$ second-order effects. If we assume this model is the true model, then we have $p - S_m(\hat{\alpha}_1, c^{(1)})$ uninformative main effects and $K_t - p - S_q(\hat{\alpha}_1, c^{(1)})$ uninformative second-order effects. By substituting these values into (4.9), we get an updated estimate for c ,

$$c^{(j)} = \frac{1 + K_t - p - S_q(\hat{\alpha}_1, c^{(j-1)})}{1 + p - S_m(\hat{\alpha}_1, c^{(j-1)})}, \quad (4.16)$$

where $j = 2$. Using $c^{(2)}$, we repeat forward selection to get a second sequence of effects, adjusted p -values, and eventually calculate $c^{(3)}$. We continue the process of estimating c and $\hat{\alpha}_1$ until $c^{(j)} - c^{(j-1)} = 0$. However, for computational purposes we often limit

the number of iterations to three or four. Finally, we fit the model corresponding to $\hat{\alpha}_1$ and $c^{(j)}$ from the final iteration.

4.2.1 Summary of Algorithm for Adjustment Method 1

1. Calculate $c^{(1)}$ using (4.10).
2. Using $c^{(1)}$, run forward selection with group entry-levels to get a full sequence of effects, adjusted p -values, and corresponding maximum adjusted p -values.
3. Using the maximum adjusted p -values, figure out the proper grid of α levels.
4. Using this grid, calculate $\hat{\gamma}_{fast}(\alpha, c^{(1)})$ using (4.12) for each α .
5. Calculate $\hat{\alpha}_1$ using (4.15).
6. Calculate $S_m(\hat{\alpha}_1, c^{(1)})$ and $S_q(\hat{\alpha}_1, c^{(1)})$.
7. Calculate $c^{(2)}$ using (4.16).
8. If $c^{(2)} - c^{(1)} = 0$, then choose the model corresponding to $\hat{\alpha}_1$ and $c^{(2)}$. Otherwise, repeat Steps 2 - 7 and continue the process until $c^{(j)} - c^{(j-1)} = 0$.

4.3 Adjustment Method 2

In Method 1 for controlling group false selection rates, our goal was to find α_1 and α_2 that satisfied both (4.2) and (4.3). Using the relation, $\alpha_1 = c\alpha_2$, we iteratively solved for c and α_1 to get a final model. In our second approach, we only run forward selection once. However, we update c after each selection step based on the type of effect entered.

As in Method 1, we start by calculating $c^{(1)}$ using (4.10). Using $c^{(1)}$, we run the first step of forward selection with group entry-levels to obtain a variable, $X_{(1)}$, that

Chapter 4. Controlling Group False Selection Rates

has the smallest adjusted p -value, \tilde{p}_1 . Up to this point, Methods 1 and 2 are exactly alike. However, instead of continuing to use $c^{(1)}$, we update c by assuming that $X_{(1)}$ is an informative variable. Thus, if $X_{(1)}$ is a main effect, then $c^{(2)} = (1 + K_t - p)(p)^{-1}$. Alternatively, if $X_{(1)}$ is a second-order effect, then $c^{(2)} = (K_t - p)(1 + p)^{-1}$. After updating c , we run another step of forward selection with group entry-levels to get $X_{(2)}$ and \tilde{p}_2 . That is, assuming $X_{(1)}$ is in the model, we calculate the p -values for any single effect to enter and adjust the p -values using $c^{(2)}$. We continue updating c after each step and running forward selection to obtain a full sequence of effects and adjusted p -values. In general, at the beginning Step i of forward selection with group entry-levels, we update c using

$$c^{(i)} = \frac{1 + K_t - p - S_q^{(i-1)}}{1 + p - S_m^{(i-1)}}, \quad (4.17)$$

where $S_m^{(i-1)}$ is the number of main effects and $S_q^{(i-1)}$ is the number of second-order effects in the model after $i - 1$ steps.

After running forward selection as described, we have a sequence of estimates for c , effects, adjusted p -values, and maximum adjusted p -values. Again, we can use the sequence of maximum adjusted p -values as our grid of α levels. The next step is to calculate the estimated false selection rates using (4.12) for each α in the grid. For a given α , use $c = c(\alpha)$, where

$$c(\alpha) = \frac{1 + K_t - p - S_q(\alpha, c)}{1 + p - S_m(\alpha, c)}. \quad (4.18)$$

After calculating $\hat{\gamma}_{fast}(\alpha, c(\alpha))$ for each α , we choose $\hat{\alpha}_1$ using (4.15). Finally, we fit

the model corresponding to $\hat{\alpha}_1$ and $c(\hat{\alpha}_1)$.

4.3.1 Summary of Algorithm for Adjustment Method 2

1. Calculate $c^{(1)}$ using (4.10).
2. Using $c^{(1)}$, run one step of forward selection with group entry-levels to get $X_{(1)}$ and $\tilde{p}_{(1)}$.
3. Update $c^{(2)}$ using (4.17).
4. Using $c^{(2)}$, run another step of forward selection with group entry-levels to get $X_{(2)}$ and $\tilde{p}_{(2)}$.
5. Repeat Steps 3 - 4 to obtain a full sequence of estimates for c , effects, adjusted p -values, and corresponding maximum adjusted p -values.
6. Using the maximum adjusted p -values, figure out the proper grid of α levels.
7. Using this grid, calculate $\hat{\gamma}_{fast}(\alpha, c(\alpha))$ using (4.12) for each α .
8. Calculate $\hat{\alpha}_1$ using (4.15).
9. Choose the model corresponding to $\hat{\alpha}_1$ and $c(\hat{\alpha}_1)$.

4.4 Weak Hierarchy with Adjustment Method 2

Recall that in Method 2, c changes at each step of the forward selection process. In the weak hierarchy approach, the number of terms in the candidate set is also changing at each step. Since we use these counts to estimate c , it is natural to combine the two approaches. Using this hybrid approach, we can control the rate at which uninformative second-order effects enter the model.

Under the weak hierarchy, we start with only main effects in the candidate set, \mathcal{C}_1 .

Chapter 4. Controlling Group False Selection Rates

That is, $\mathcal{C}_1 = \{X_1, \dots, X_p\}$. Therefore, we do not need to specify $c^{(1)}$ for this step. We select the main effect with the smallest p -value to enter, $\tilde{p}_{(1)}$. After $X_{(1)}$ enters the model, we update the candidate set by adding $X_{(1)}^2$ and all pairwise interactions containing $X_{(1)}$. That is, $\mathcal{C}_2 = \{\{X_1, \dots, X_p\} - \{X_{(1)}\}\} \cup \{X_{(1)}X_1, \dots, X_{(1)}X_p\}$. In order to estimate c , we need to keep track of the number of variables of each type in the candidate set. At the beginning of Step 2, we have $p - 1$ main effects and p second-order effects in the candidate set, and since we assume these variables are uninformative, $c^{(2)} = (p)(p - 1)^{-1}$. Using $c^{(2)}$, run one step of forward selection with group-entry levels to get $X_{(2)}$ and $\tilde{p}_{(2)}$. If $X_{(2)}$ is a main effect add $X_{(2)}^2$ and the $p - 2$ interactions containing $X_{(2)}$ not already in the the candidate set. Otherwise, if $X_{(2)}$ is a second-order effect, then no new variables are added to the candidate set. In general, at the beginning Step i of forward selection with group entry-levels, we update c using

$$c^{(i)} = \frac{1 + C(i, c^{(i-1)}) - p - S_q^{(i)}}{1 + p - S_m^{(i)}}, \quad (4.19)$$

where $C(i, c^{(i-1)})$ is the number of variables ever in the candidate set at the beginning of Step i . We continue updating the candidate set and c at the beginning of each step and running forward selection to obtain a full sequence of effects and adjusted p -values.

After running forward selection as described, we have a sequence of estimates for c , effects, adjusted p -values, and maximum adjusted p -values. Again, we can use the sequence of maximum adjusted p -values as our grid of α levels. The next step is to calculate the estimated false selection rates using

$$\hat{\gamma}_{fast}(\alpha, c) = \frac{\{p - S_m(\alpha, c)\}\alpha + \{C(\alpha, c) - p - S_q(\alpha, c)\}\{\alpha/c\}}{1 + S(\alpha, c)}, \quad (4.20)$$

for each α in the grid. For a given α , use $c = c(\alpha)$, where

$$c(\alpha) = \frac{1 + C(\alpha, c) - p - S_q(\alpha, c)}{1 + p - S_m(\alpha, c)}. \quad (4.21)$$

After calculating $\hat{\gamma}_{fast}(\alpha, c(\alpha))$ for each α , choose $\hat{\alpha}_1$ using (4.15). Finally, fit the model corresponding to $\hat{\alpha}_1$ and $c(\hat{\alpha}_1)$.

4.4.1 Summary of Algorithm for Weak Hierarchy with Adjustment Method 2

1. Run one step of forward selection with only main effects in the candidate set, i.e. $\mathcal{C}_1 = \{X_1, \dots, X_p\}$, to get $X_{(1)}$ and $\tilde{p}_{(1)}$.
2. Update the candidate set by adding $X_{(1)}^2$ and adding all pairwise interactions containing $X_{(1)}$. There are $p - 1$ main effects and p second-order effects in the candidate set. That is, $\mathcal{C}_2 = \{\{X_1, \dots, X_p\} - \{X_{(1)}\}\} \cup \{X_{(1)}X_1, \dots, X_{(1)}X_p\}$.
3. Calculate $c^{(2)} = (p)(p - 1)^{-1}$.
4. Using $c^{(2)}$, run another step of forward selection with group entry-levels to get $X_{(2)}$ and $\tilde{p}_{(2)}$, where $X_{(2)}$ can be any effect in \mathcal{C}_2 .
5. Update the candidate set by adding the necessary quadratic and interaction effects following the weak hierarchy. Keep track of the number of effects in the candidate set for each group.
6. Estimate $c^{(3)}$ using (4.19).
7. Using $c^{(3)}$, run another step of forward selection with group entry-levels to get $X_{(3)}$ and $\tilde{p}_{(3)}$.
8. Repeat Steps 5 - 7 to obtain a full sequence of estimates for c , effects, adjusted

p -values, and corresponding maximum adjusted p -values.

9. Using the maximum adjusted p -values, figure out the proper grid of α levels.

10. Using this grid, calculate $\hat{\gamma}_{fast}(\alpha, c(\alpha))$ using (4.20) for each α .

11. Calculate $\hat{\alpha}_1$ using (4.15).

12. Choose the model corresponding to $\hat{\alpha}_1$ and $c(\hat{\alpha}_1)$.

4.5 Method Extensions

The methods discussed focus on yielding equal contributions to the false selection rate. However, they are not limited to equal contributions. A more general approach is to yield an $a : b$ ratio in terms of the contribution. Then the goal is to estimate c , such that

$$\frac{E[FSR_m]}{E[FSR_q]} = \frac{a}{b} \quad (4.22)$$

and

$$E[FSR_m] + E[FSR_q] = \gamma_0. \quad (4.23)$$

With a little bit of math, we find that (4.9) can be written in the more general form,

$$c = \frac{a\{1 + U_q\}}{b\{1 + U_m\}}. \quad (4.24)$$

In the case of equal contributions, $a = b = 1$. For example, suppose we want to focus our attention on including main effects. In doing so, we might prefer a falsely selected main effect to be included twice as often as a falsely included interaction. That is, we want $a = 2$ and $b = 1$. Notice that if we substitute these values into (4.24) then c is

Chapter 4. Controlling Group False Selection Rates

twice as large than when $a = b = 1$. In order to control the entry of the second-order effects, the probability of their entry must be penalized twice as much as before.

As described in the methods, the quadratic and interaction effects are combined into one group. However, it may be of interest to group the variables differently. For example, we might combine main and quadratic effects into one group and pairwise interactions into another. We can then proceed with the methods exactly as described, except now α_1 corresponds to main and quadratic effects and α_2 corresponds to interactions. One change we might make to the weak hierarchy approach under this new grouping is to include all main and quadratic effects in the candidate set at Step 1. This makes all the uninformative effects in group 1 (main and quadratic effects) have the same probability of entry into the model. To further extend this idea, variables can essentially be grouped any way we would like, and then the methods can be applied using no hierarchy. For example, in many studies the variables can be grouped as environmental factors or genetic factors. We may want an equal contribution to the false selection rate from both groups or we might favor one group over the other.

One final extension is that the methods are not limited to two groups. Multiple groups can be created and resulting false selection rates controlled. However, as the number of group FSRs being controlled increases, the power for detecting an informative effect decreases.

CHAPTER 5

Comparison of Methods

In this chapter we compare the different methods for fitting models with interactions using a simulation study. We evaluate each method's performance using measures for power, predictability, and interpretability.

5.1 Power and Predictability

First and foremost, a useful method should have high power in approximating the true function. Even if the method can be easily interpreted, it is of no use if it yields poorly fitted models. The most important difference between linear regression and recursive partitioning is that the former fits a continuous surface, whereas the latter fits a discrete surface. In recursive partitioning the data are used less efficiently, since information is lost when a predictor is used in a split. However, the gain in efficiency for linear regression is at the cost of strong modeling assumptions. By enforcing strong modeling assumptions, linear regression methods often have trouble in the presence of interaction and nonlinear effects. Addressing marketing studies in particular, Armstrong and Andress (1970) conclude that regression trees should be preferred over linear regression for exploratory analyses in the presence of complex effects, assuming large sample sizes are available. However, their basis of comparison is strictly on predictive ability which is less often the focus for exploratory analyses.

5.2 Interpretability

Understanding a method and how to properly interpret the results are often more important to scientists than prediction. Of course, linear regression models are widely used and well understood. They provide simple interpretation by enforcing a linear relationship between the response and predictors. Regression trees can vary in their interpretability. On one hand, regression trees are able to divide the observations into groups that can be classified and understood through the use of tree-based diagrams. However, with each level of splitting, the order of interaction between the splitting variables increases. This makes for models with great complexity. Also, although each individual variable is measured for its importance, no insight on the importance of specific interactions is given. This makes it harder for researchers to understand which of the complex interactions are most interesting. Through the use of the ANOVA-like decomposition in (2.15), MARS overcomes the problem of interpretation seen in regression trees. Although this makes it easier to see which effects contribute to the model, understanding the spline basis functions and how they are selected is not obvious.

5.3 Simulation Design

A simulation study was conducted to compare the predictive ability, power in capturing important effects, and ability to exclude uninformative effects. The following factors were used in the study:

$p = 20$ original predictors ($p_q = 230$ total predictors)

Chapter 5. Comparison of Methods

Replicates: $N = 100$

Sample Sizes: $n = 100, 500$

Models:

1. $Y = -100 + 25X_1 + 15X_{13} - 20X_{17} + X_1^2 - 3X_1X_9 + \epsilon$

2. $Y = -3 + X_1 - X_4 + 2X_9 - 1.2X_{13} + 1.6X_{17} + \epsilon$

3. $Y = 50 + 15X_1 - 25X_9 + 1.2X_1^2 - 1.6X_9^2 + 3X_1X_9 + \epsilon$

Theoretical R-Square: $R^2 = 0.25, 0.5$

Distributions for predictors (Correlation structure is in Figure 5.1):

1. $N(10, 20)$ uncorrelated
2. $N(10, 20)$ correlated
3. χ_{10}^2 uncorrelated
4. χ_{10}^2 correlated

For the error terms, $\epsilon \sim N(0, \sigma^2)$ where σ^2 is chosen to achieve the desired theoretical R^2 in the particular model, where the theoretical $R^2 = Var\{\beta_0 + \sum_{j=1}^p \beta_j X_j\} / Var\{Y\}$.

5.3.1 Methods and Stopping Rules

Table 5.1 is a list of the methods being compared along with their corresponding stopping rules. The “best” methods are not feasible statistical methods, but are included in order to compare the strategy of maintaining the hierarchy versus not. Table 5.2 lists the method abbreviations used.

Table 5.1: Methods and Stopping Rules

Method	Stopping Rule
CART	Cross Validation
MARS	Cross Validation
LASSO	Cross Validation
No Hierarchy	Fast FSR with $\gamma_0 = 0.5$
No Hierarchy with Sweeping	Fast FSR with $\gamma_0 = 0.5$
Strong Hierarchy	Fast FSR with $\gamma_0 = 0.5$
Weak Hierarchy	Fast FSR with $\gamma_0 = 0.5$
No Hierarchy with Adjustment Method 1	Fast FSR with $\gamma_0 = 0.5$
No Hierarchy with Adjustment Method 2	Fast FSR with $\gamma_0 = 0.5$
Weak Hierarchy with Adjustment Method 2	Fast FSR with $\gamma_0 = 0.5$
Best No Hierarchy	Minimum Model Error
Best Strong Hierarchy	Minimum Model Error

Table 5.2: Method Abbreviations

Method Name	Abbreviated Name
Fast FSR with No Hierarchy	FFSR-NH
Fast FSR with No Hierarchy and Sweeping	FFSR-NHS
Fast FSR with Strong Hierarchy	FFSR-SH
Fast FSR with Weak Hierarchy	FFSR-WH
Fast FSR with No Hierarchy and Adjustment Method 1	FFSR-NHA1
Fast FSR with No Hierarchy and Adjustment Method 2	FFSR-NHA2
Fast FSR with Weak Hierarchy and Adjustment Method 2	FFSR-WHA2
Best No Hierarchy	Best No
Best Strong Hierarchy	Best Strong

5.3.2 Data Generation

In an attempt to cover a variety of modeling situations, the predictors were generated from both a symmetric and a skewed distribution. For each distribution, the predictors are generated as both correlated and uncorrelated. For uncorrelated data, the predictors are simply generated from independent $N(10, 20)$ and χ_{10}^2 accordingly. For the cases where the predictors are correlated, the structure is such that each predictor is correlated with 14 other predictors. The correlation matrix is in Figure 5.1.

Data with correlation structure as in Figure 5.1 were generated as follows. For normal predictors, let C_1, \dots, C_{20} be independently distributed $N(1, 2)$ and D_1, \dots, D_{20} be independently distributed $N(2, 4)$. Then define,

$$X_1 = D_1 + C_1 + C_2 + C_3 + C_4 + C_5 + C_6 + C_7 + C_8$$

$$X_2 = D_2 + C_2 + C_3 + C_4 + C_5 + C_6 + C_7 + C_8 + C_9$$

⋮

$$X_{19} = D_{19} + C_{19} + C_{20} + C_1 + C_2 + C_3 + C_4 + C_5 + C_6$$

$$X_{20} = D_{20} + C_{20} + C_1 + C_2 + C_3 + C_4 + C_5 + C_6 + C_7$$

Similarly for chi-squared predictors, let C_1, \dots, C_{20} be independently distributed χ_1^2 , and D_1, \dots, D_{20} be independently distributed χ_2^2 . Then define,

$$X_1 = D_1 + C_1 + C_2 + C_3 + C_4 + C_5 + C_6 + C_7 + C_8$$

$$X_2 = D_2 + C_2 + C_3 + C_4 + C_5 + C_6 + C_7 + C_8 + C_9$$

⋮

$$X_{19} = D_{19} + C_{19} + C_{20} + C_1 + C_2 + C_3 + C_4 + C_5 + C_6$$

$$X_{20} = D_{20} + C_{20} + C_1 + C_2 + C_3 + C_4 + C_5 + C_6 + C_7$$

	X_1	X_2	X_3	X_4	X_5	...	X_{16}	X_{17}	X_{18}	X_{19}	X_{20}
X_1	1	0.7	0.6	0.5	0.4	...	0.3	0.4	0.5	0.6	0.7
X_2	0.7	1	0.7	0.6	0.5	...	0.2	0.3	0.4	0.5	0.6
X_3	0.6	0.7	1	0.7	0.6	...	0.1	0.2	0.3	0.4	0.5
X_4	0.5	0.6	0.7	1	0.7	...	0	0.1	0.2	0.3	0.4
X_5	0.4	0.5	0.6	0.7	1	...	0	0	0.1	0.2	0.3
X_6	0.3	0.4	0.5	0.6	0.7	...	0	0	0	0.1	0.2
X_7	0.2	0.3	0.4	0.5	0.6	...	0	0	0	0	0.1
X_8	0.1	0.2	0.3	0.4	0.5	...	0	0	0	0	0
X_9	0	0.1	0.2	0.3	0.4	...	0	0	0	0	0
\vdots	\vdots					\ddots					\vdots
X_{13}	0	0	0	0	0	...	0.5	0.4	0.3	0.2	0.1
X_{14}	0.1	0	0	0	0	...	0.6	0.5	0.4	0.3	0.2
X_{15}	0.2	0.1	0	0	0	...	0.7	0.6	0.5	0.4	0.3
X_{16}	0.3	0.2	0.1	0	0	...	1	0.7	0.6	0.5	0.4
X_{17}	0.4	0.3	0.2	0.1	0	...	0.7	1	0.7	0.6	0.5
X_{18}	0.5	0.4	0.3	0.2	0.1	...	0.6	0.7	1	0.7	0.6
X_{19}	0.6	0.5	0.4	0.3	0.2	...	0.5	0.6	0.7	1	0.7
X_{20}	0.7	0.6	0.5	0.4	0.3	...	0.4	0.5	0.6	0.7	1

Figure 5.1: Correlation matrix for simulation study

Once each set of predictors are generated as discussed above, the three functions are used to generate corresponding responses.

After the predictors and corresponding response values have been generated, the predictors are then standardized. Finally, the quadratic and interaction effects are derived using the standardized predictors. As previously discussed, standardization lessens multicollinearity between the derived variables and their corresponding main effects.

This process is repeated for each replicate of the study, generating a new set of predictors and responses each time. The input variables for the linear regression methods on a particular replicate are the original response variables, the standardized predictors, and the second-order effects derived from them. For CART and MARS, only the original response variables and standardized predictors are input.

5.3.3 Measures of Interest

In this section we define each measure of interest and provide the defining equation where necessary. The goals of the different measures are to provide a means of comparing the methods in terms of prediction and interpretation.

Average model error (AME),

$$\text{AME} = (nN)^{-1} \sum_{i=1}^N \sum_{j=1}^n (\hat{Y}_{i,j} - \mu_{i,j})^2 \quad (5.1)$$

is a common measure for the predictability of a model fitting method. AME Ratio is the ratio of the average model error of the true model to the average model error of the method.

$$\text{AME Ratio} = \frac{\text{AME}_{True}}{\text{AME}_{Method}} \quad (5.2)$$

Methods with a high AME Ratio are preferred.

Model Size is the number of estimated regression parameters. For standard regression techniques, model size equals the number of effects in the model plus the intercept. For CART, model size is the number of terminal nodes in the final model. For MARS,

the model size is the number of basis functions in the final model.

The False Selection Rate (FSR) is defined as the fraction of total regression degrees of freedom deemed uninformative. FSR can be partitioned into contributions from main effects and second-order effects. That is, $FSR = FSR_m + FSR_q$, where

$$FSR = \frac{\text{Number of uninformative effects selected}}{\text{Size of selected model}}, \quad (5.3)$$

$$FSR_m = \frac{\text{Number of uninformative main effects selected}}{\text{Size of selected model}}, \quad (5.4)$$

$$FSR_q = \frac{\text{Number of uninformative second-order effects selected}}{\text{Size of selected model}}. \quad (5.5)$$

For CART, FSR was calculated as the weighted fraction of uninformative variables involved in a split. For MARS, FSR was calculated as the fraction of total basis functions associated with uninformative predictors. As a result of the differing approach to false selection, FSR was not broken into the two components for CART or MARS.

The Correct Selection Rate (CSR) is defined as the fraction of possible informative effects chosen in the model. It is used as a measure of the power to detect informative effects. This value was not calculated for CART and MARS, as neither can fit all of the informative effects.

$$CSR = \frac{\text{Number of informative effects selected}}{\text{Total number of informative effects}} \quad (5.6)$$

The Coefficient of Fiction (CF) is a measure of how much a method is fooling its

user in terms of R^2 .

$$CF = \frac{R_{Method}^2}{R_{True}^2} \quad (5.7)$$

For a data set, R_{Method}^2 is the observed coefficient of determination for the model selected by a particular method, whereas R_{True}^2 is the observed coefficient of determination for the true model.

5.4 Simulation Results

Figures 5.2 - 5.7 illustrate the results for LASSO (L), FFSR-NH (N), FFSR-SH (S), FFSR-WH (W), FFSR-NHA2 (A), and FFSR-WHA2 (D). In Figure 5.2, the left column of plots illustrate AME Ratio for the methods that enforce no hierarchy, whereas the right column of plots illustrate AME Ratio for the methods that enforce strong or weak hierarchy. Similar plots for Chi Square predictors can be found in Appendix A. Some of the methods were left out, so the plots could be more easily read. However, the full simulation results can be found in Tables A.1 - A.24. They are organized by R^2 , distribution of predictors, function, and sample size. The methods follow the same order as those in Table 5.1. The values in italics represent the average standard error of the corresponding column.

Before discussing the results in the various areas of interest, some general impressions are presented. Overall, FFSR-SH and FFSR-WHA2 performed the best in terms of average model error and correct selection. The LASSO generally captured a large proportion of the informative effects but was only competitive, in terms of model error, for smaller values of R^2 . The partition-based methods generally did not fit the models very well. Finally, adjusting the p -values for the quadratic and interaction effects

improved upon FFSR-NH and FFSR-WH.

5.4.1 Overall Performance by Method

In this section we summarize the simulation results for each method. As previously noted, CART and MARS were left out of Figures 5.2 - 5.7, but their results can be found in Tables A.1 - A.24.

CART had a tendency to underfit. The model size was generally smaller than many of the other methods. It was never competitive in terms of model error, although it did pretty well in terms of false selection. Another downside of the CART method is that it cannot fit the exact models. For example, it cannot fit more than one linear effect as previously discussed.

MARS had a tendency to overfit. The model size was generally larger than many of the other methods. It was neither competitive in terms of model error nor false selection.

LASSO was very competitive in terms of both model error and correct selection. Tables A.1 - A.6 indicate that the LASSO performs relatively best in models where the effects are harder to identify ($n = 100$ and $R^2 = 0.25$). However, since it tends to include a large number of effects, it did not do as well in terms of false selection. The ordinary LASSO suffers from searching over all the interaction effects and has a tendency to include a large number of them. This makes the final model tough to interpret. Using LASSO with strong hierarchy or weak hierarchy might improve false selection rates.

Fast FSR with No Hierarchy (FFSR-NH) was not very competitive in terms

Chapter 5. Comparison of Methods

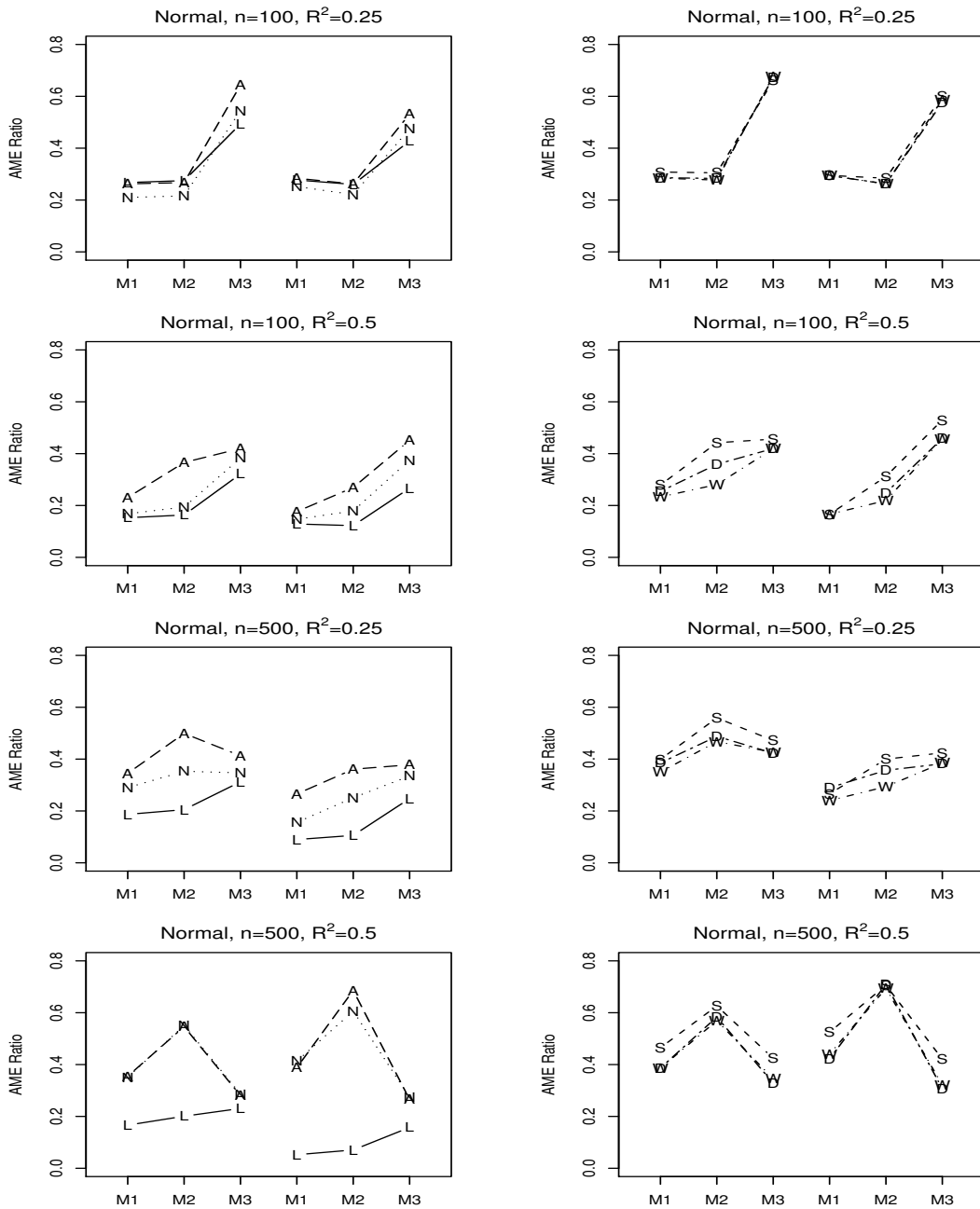


Figure 5.2: AME ratio (larger is better) using Normal predictors for LASSO (L), FFSR-NH (N), FFSR-SH (S), FFSR-WH (W), FFSR-NHA2 (A), and FFSR-WHA2 (D). The first three values in each plot are uncorrelated predictors and the last three values are correlated predictors.

Chapter 5. Comparison of Methods

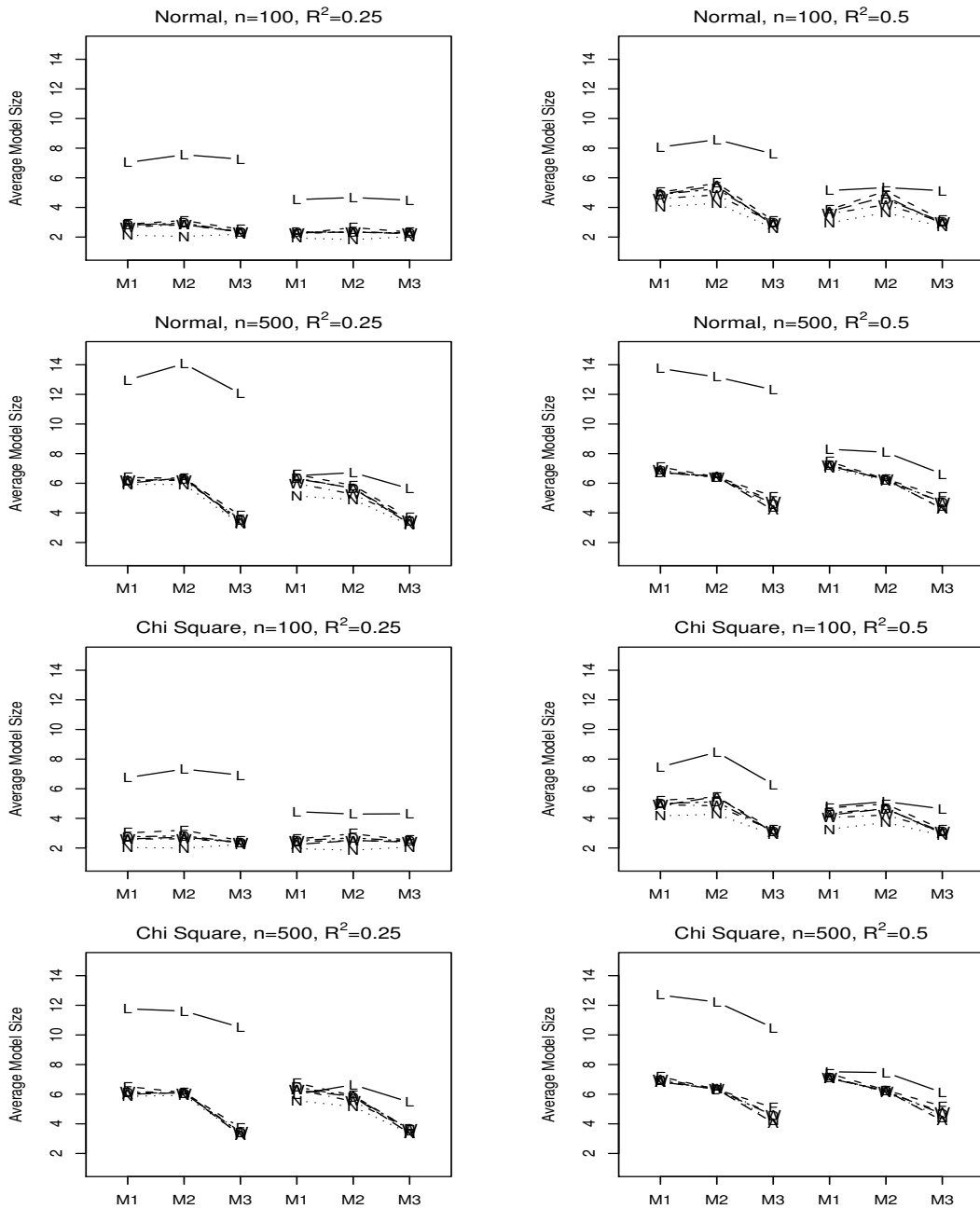


Figure 5.3: Average model size for LASSO (L), FFSR-NH (N), FFSR-SH (S), FFSR-WH (W), FFSR-NHA2 (A), and FFSR-WHA2 (D). The first three values in each plot are uncorrelated predictors and the last three values are correlated predictors.

Chapter 5. Comparison of Methods

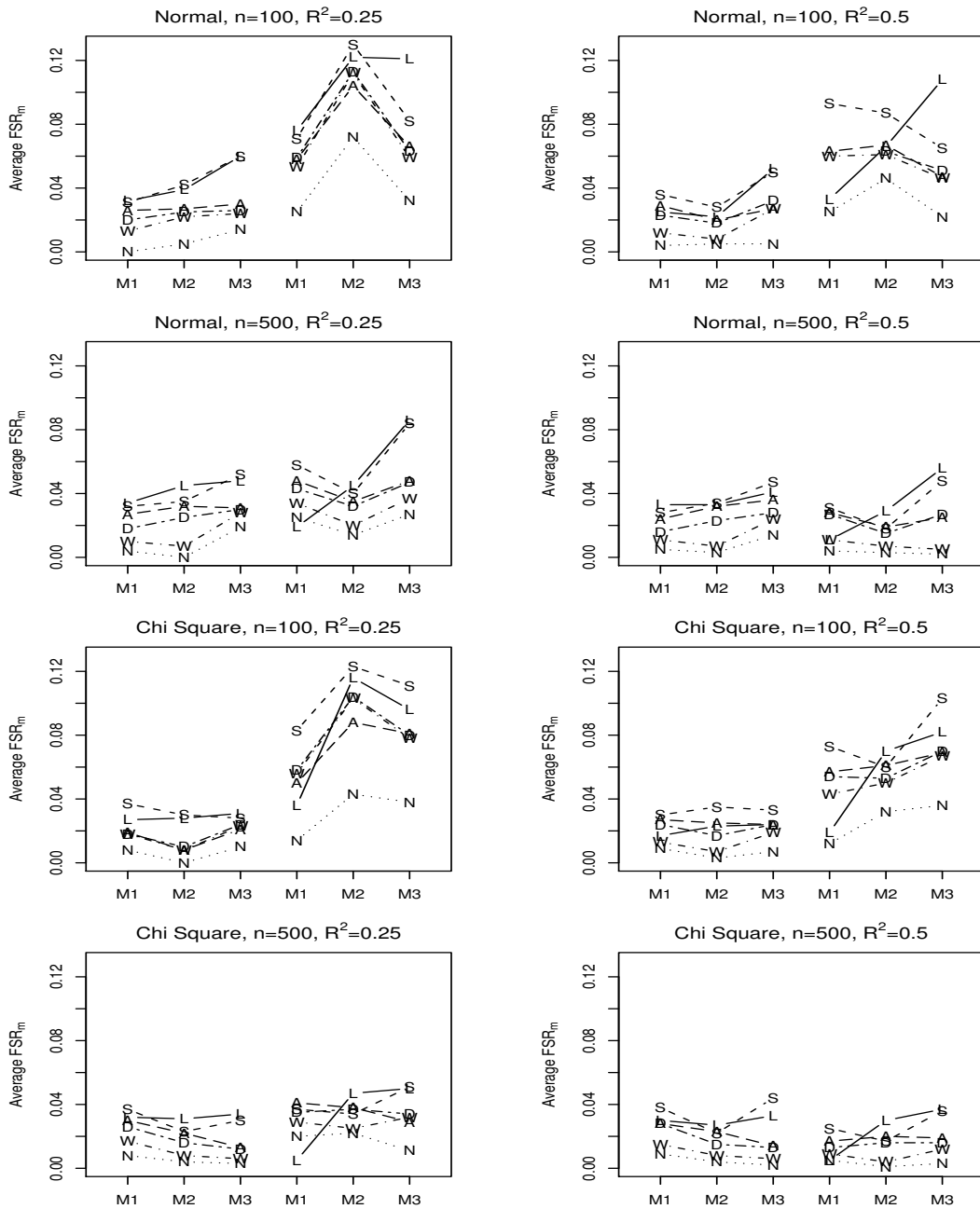


Figure 5.4: Average FSR_m for LASSO (L), FFSR-NH (N), FFSR-SH (S), FFSR-WH (W), FFSR-NHA2 (A), and FFSR-WHA2 (D). The first three values in each plot are uncorrelated predictors and the last three values are correlated predictors.

Chapter 5. Comparison of Methods

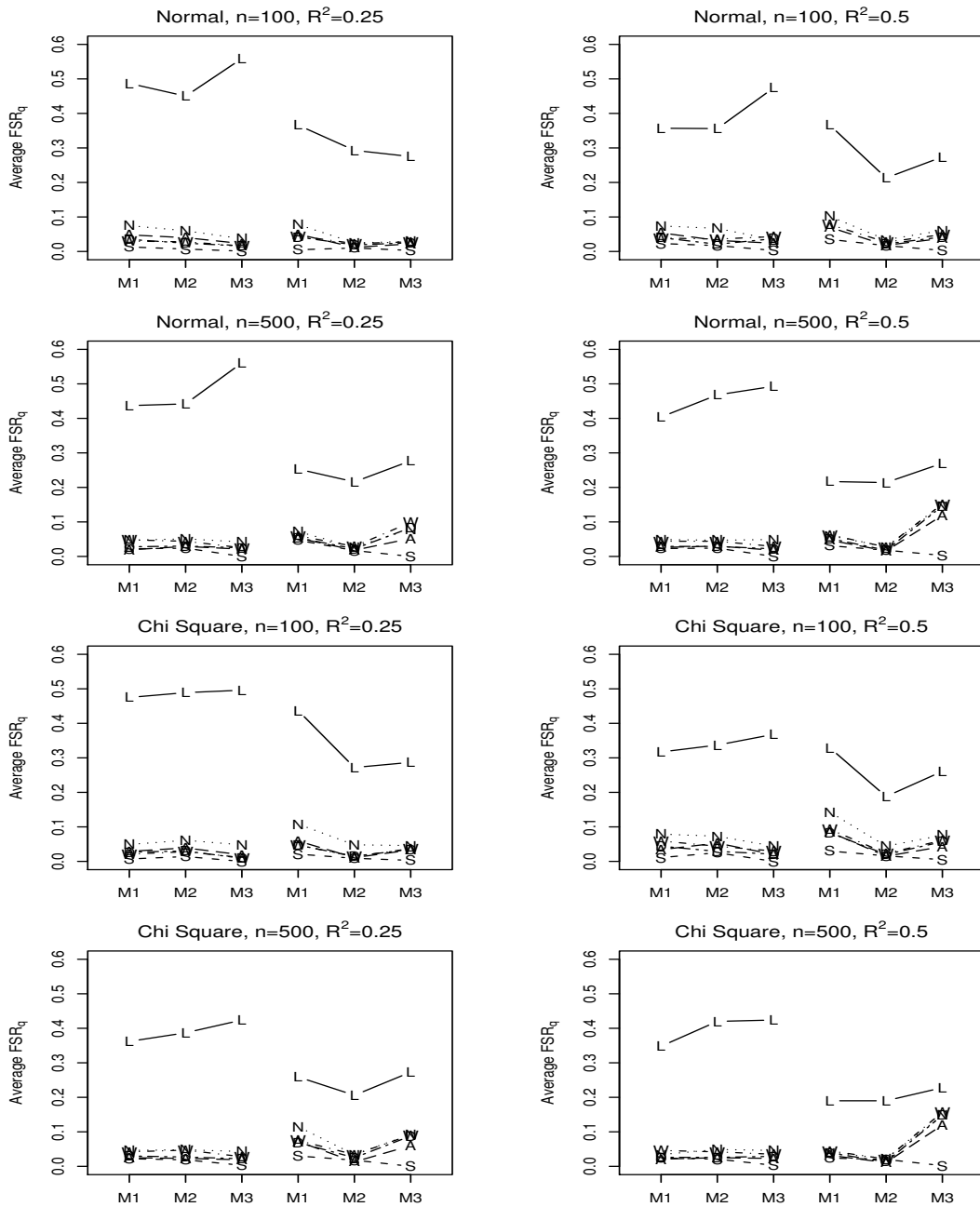


Figure 5.5: Average FSR_q for LASSO (L), FFSR-NH (N), FFSR-SH (S), FFSR-WH (W), FFSR-NHA2 (A), and FFSR-WHA2 (D). The first three values in each plot are uncorrelated predictors and the last three values are correlated predictors.

Chapter 5. Comparison of Methods

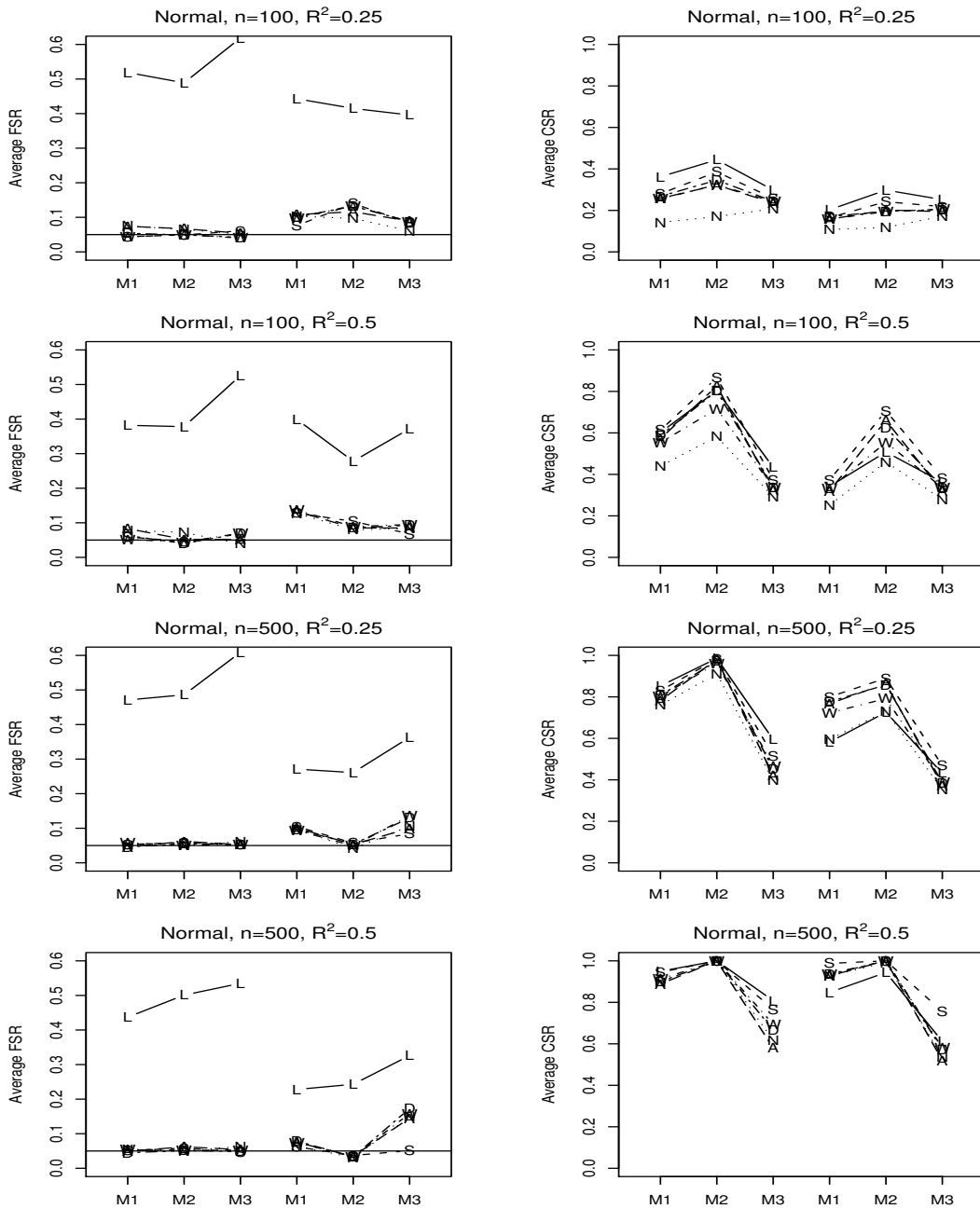


Figure 5.6: Average FSR (left) and CSR (right) using Normal predictors for LASSO (L), FFSR-NH (N), FFSR-SH (S), FFSR-WH (W), FFSR-NHA2 (A), and FFSR-WHA2 (D). The first three values in each plot are uncorrelated predictors and the last three values are correlated predictors.

Chapter 5. Comparison of Methods

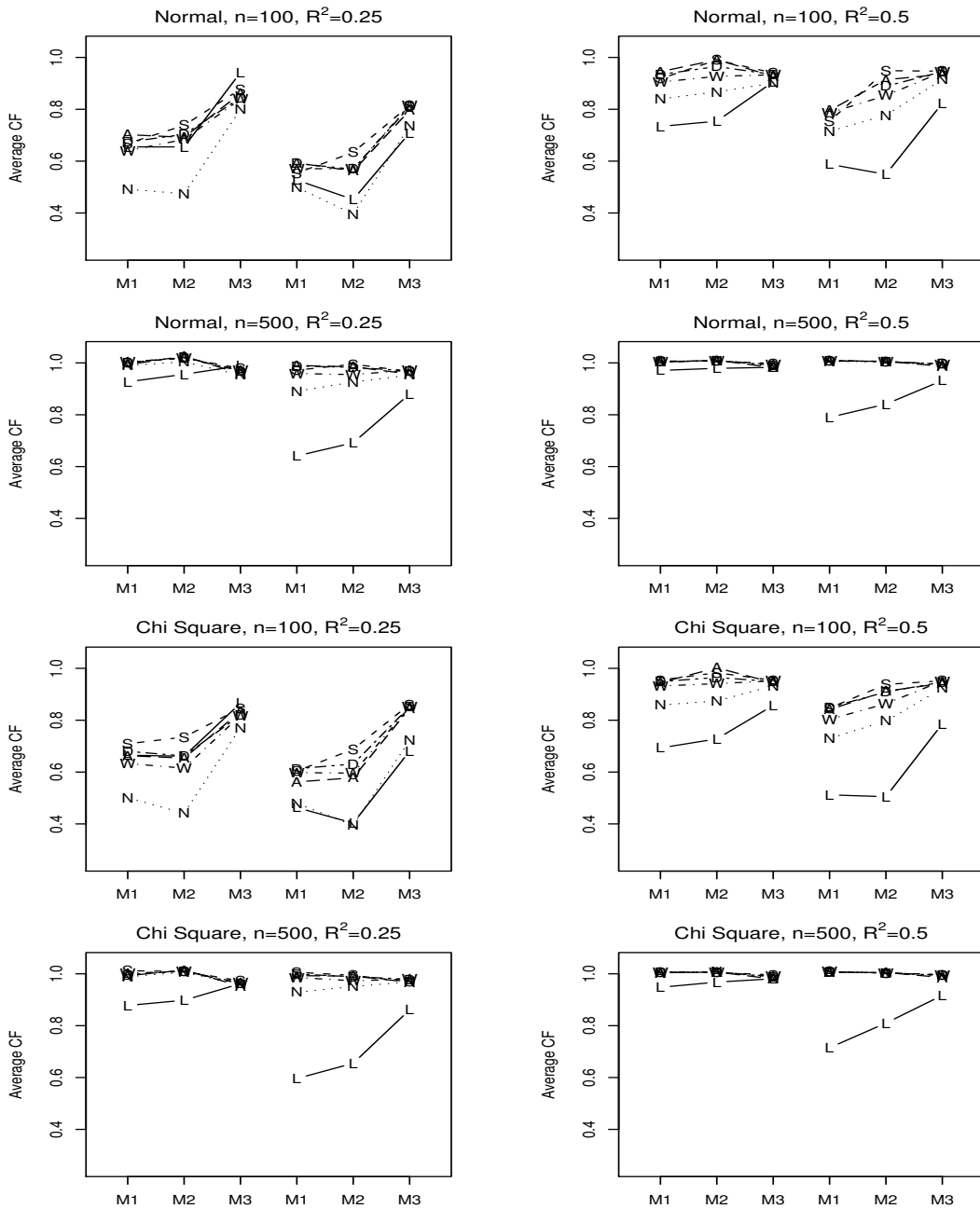


Figure 5.7: Average CF for LASSO (L), FFSR-NH (N), FFSR-SH (S), FFSR-WH (W), FFSR-NHA2 (A), and FFSR-WHA2 (D). The first three values in each plot are uncorrelated predictors and the last three values are correlated predictors.

of model error and correct selection although it typically had false selection rates less than the specified level. Like the LASSO, it suffers from searching over all interaction effects. Note that even the best that the no hierarchy can do in terms of model error is still outperformed by some of the other methods.

Fast FSR with No Hierarchy and Sweeping (FFSR-NHS) had performance very similar to FFSR-NH, and therefore was not very competitive.

Fast FSR with Strong Hierarchy (FFSR-SH) was the overall best method in terms of model error, false selection, and correct selection. The advantage of the strong hierarchy method is that fewer interaction effects are in the candidate set early in the selection process. This is important when the data are correlated because the quadratic and interaction effects can be highly correlated leading to higher false selection rates. In fact, the false selection rate for FFSR-SH is closer to the specified level in some cases where the other methods had much larger false selection rates. For example, in Table A.18 when $n = 500$, the false selection rate is 0.051 whereas it is nearly three times as large for many of the other methods.

Fast FSR with Weak Hierarchy (FFSR-WH) often improved on the no hierarchy method, but not the strong hierarchy. Main effects enter the model more easily than under no hierarchy, but as Table A.1 illustrates, it still had a tendency to choose uninformative second-order effects in the model over uninformative main effects when the data are uncorrelated.

Fast FSR with No Hierarchy and Adjustment Method 1 (FFSR-NHA1) performs very well in all models, although it had cases where the false selection rates were slightly higher than the specified level even when the data were uncorrelated. The method accomplished its goal of having approximately equal contributions to the

false selection rate from main effects and interactions, at least when the data are uncorrelated. Controlling the separate false selection rates is discussed in more detail later.

Fast FSR with No Hierarchy and Adjustment Method 2 (FFSR-NHA2) performs almost equivalently to FFSR-NHA1. In Tables A.1 - A.24, we see the results for all measures of interest are the same in many cases and very close in others.

Fast FSR with Weak Hierarchy and Adjustment Method 2 (FFSR-WHA2) was very competitive in terms of model error, false selection, and correct selection. It also has the good interpretation of a balanced contribution to the false selection rate between main effects and second-order effects. Adjusting the p -values in combination with a weak hierarchy limits the entry of second-order effects, allowing more main effects to enter the model.

5.4.2 Overall Performance by Model

In the first model, four main effects, one interaction and one quadratic effect are informative. With more main effects than quadratic and interaction effects, the model is easier to fit than Model 3. Based on correct selection rates, this model was the second hardest to fit. This can be seen in Figure 5.6. FFSR-SH and FFSR-WHA2 performed the best in terms of model error and correct selection.

In the second model, five main effects are informative. Of the three models, Model 2 was the easiest to fit since it had no informative interactions. As with Model 1, the FFSR-SH performed the best in terms of model error and correct selection. However, the LASSO performed best in terms of correct selection when $R^2 = 0.25$ and $n = 100$.

In the third model, two main effects, two interactions and one quadratic effect are informative. This model was the toughest for the methods to fit. The FFSR-WH and FFSR-WHA2 performed very well in terms of model error. However, when the data are correlated, they tend to have false selection rates higher than desired. This problem is not exclusive to Model 3 but is most prevalent with this model. The LASSO generally had the largest correct selection rates for this model, but its large model error is likely due to the large number of uninformative variables included as well. FFSR-SH does well in all measures. It had false selection rates which were less often inflated when the data were correlated, while still approximating the function well.

5.4.3 Comparison of Other Factors

Predictor Distribution: Normal vs. Chi Square Distribution Overall there was relatively no difference between the performance of the methods for normally distributed predictors versus Chi Square distributed predictors. One of the intentions of comparing a symmetric versus a skewed distribution is to see if there is any gain from sweeping out the main effects. Recall that centering a variable and then deriving its quadratic term greatly reduces the correlation between the two. The same holds true for pairwise interactions. However, when the data are skewed, some of this correlation remains. Sweeping, on the other hand, creates quadratic and interaction effects that are orthogonal to their parent main effects. However, Tables A.1 - A.24 show that there were no major differences in the results for sweeping versus not sweeping with either distribution. The skewness of the distributions used in the simulation was fairly small, but for more highly skewed distributions there may be an advantage to sweeping.

Predictor Correlation: Uncorrelated vs. Correlated

When the predictors are generated independently of one another, the Fast FSR methods performed as expected. That is, they chose models whose average false selection rates were close to 0.05. However, problems occur when the data are correlated. There were numerous cases where the average false selection rates were inflated for the Fast FSR methods. A simple explanation is that when the predictors are correlated and the signal weak, an uninformative variable could explain more of the variation in the response than the informative counterpart just by chance. For example, suppose X_1 and X_2 are highly correlated, and the signal for X_1 in the model is fairly weak. Just by chance we might expect X_2 to be selected instead in some of the replicates. Partitioning the false selection rate into the contribution from main effects and second-order effects is informative. In Figure 5.4, notice that for Model 2, where all of the informative variables are main effects, the values of FSR_m are often larger than expected. Models 1 and 3 have both informative main effects and interactions. In these models, notice in Figures 5.4 and 5.5 that both false selection rates can be larger than expected. Unfortunately, the theory does not take into account correlation between the predictors, and the problems caused by correlated predictors are not easily resolved.

Sample Size: n=100 vs. n=500

With larger sample sizes, the linear regression methods performed better. Their AME ratio and CSR increased as the sample size increased. However, relative to the true model, CART and MARS performed worse. That is, the AME ratio decreased as the sample size increased. In general, the model sizes increased as the sample size increased. Also, the problems with inflated false selection rates caused by correlated data occurred more at the smaller sample size.

Coefficient of Determination: $R^2 = 0.25$ Vs. $R^2 = 0.5$

As expected, the methods perform much better when R^2 is larger. The problem with inflated false selection rates caused by correlated data occurred more at the smaller values of R^2 than larger. When the signal is smaller it is more likely for an informative variable to be replaced by an uninformative variable.

5.4.4 Adjustment Methods

As briefly mentioned, the adjustment methods had very similar performance in all the cases. This suggests that the two approaches for estimating c selected the same model for the most part. Method 1 might be preferred due to its simple interpretation of estimating α_1 for main effects and $\alpha_2 = \alpha_1/c$ for the quadratic and interaction effects. However, Method 1 requires multiple forward sequences, whereas Method 2 requires only one. Therefore, we prefer to apply Method 2 due to its faster computational speed.

The overall goal of the adjustment methods in the study was for $E[FSR_m] = E[FSR_q] = 0.025$. Figure 5.8 plots the contributions to the false selection rate for each modeling situation. The left column of plots is FSR_m and the right column is FSR_q . The first three values in each plot correspond to Models 1 - 3 with uncorrelated predictors and the final three values correspond to Models 1 - 3 with correlated predictors. Similar plots for Chi Square predictors can be found in Appendix A.

When the data are uncorrelated, the goals were somewhat accomplished. In Chapter 4, it was shown that a good estimate for c is a ratio of number of uninformative quadratic and interaction effects to the number of uninformative main effects. If we

miss some of the informative effects, then this could lead to c which is too large or too small. When $n = 500$ and $R^2 = 0.5$, the methods captured nearly all of the informative effects. As a result, we see in Figure 5.8 that FSR_m and FSR_q were close to 0.025. In the cases where the informative effects are hard to detect, we see that the false selection rate is larger than our desired level. For example, in Model 2, if we classify some of the informative main effects as uninformative then c will be too small. This leads to an inflation in FSR_q as seen most prevalently in Figure 5.10 when $n = 100$.

When the data are correlated, the false selection rates are often larger than expected. As already discussed, uninformative effects that are correlated with informative effects are being selected. In Figure 5.8, we see that for Model 2 when $n = 100$ and $R^2 = 0.25$, FSR_m is four times our desired level for many of the methods. However, as the sample size and R^2 increase, FSR_m gets closer to our desired level. The opposite effect occurs for FSR_q , except in Model 2. In general, we see that as the sample size and R^2 increase, FSR_q increases. A possible explanation is that when the main effects are correlated, interactions and quadratic terms are also correlated. Since each main effect is correlated with over half the main effects, each second-order effect is correlated with over half the second-order effects. When all of these effects are in the candidate set, an uninformative interaction might be chosen over an informative interaction by chance. For example, since X_1 and X_2 are highly correlated, the variables X_1^2 and X_1X_2 are highly correlated. In Model 1, X_1^2 is an informative effect. Since X_1^2 is highly correlated with X_1X_2 , it is possible that X_1X_2 might be chosen instead by chance. As the sample size and R^2 increase, more effects are selected and this problem is observed more often.

Chapter 5. Comparison of Methods

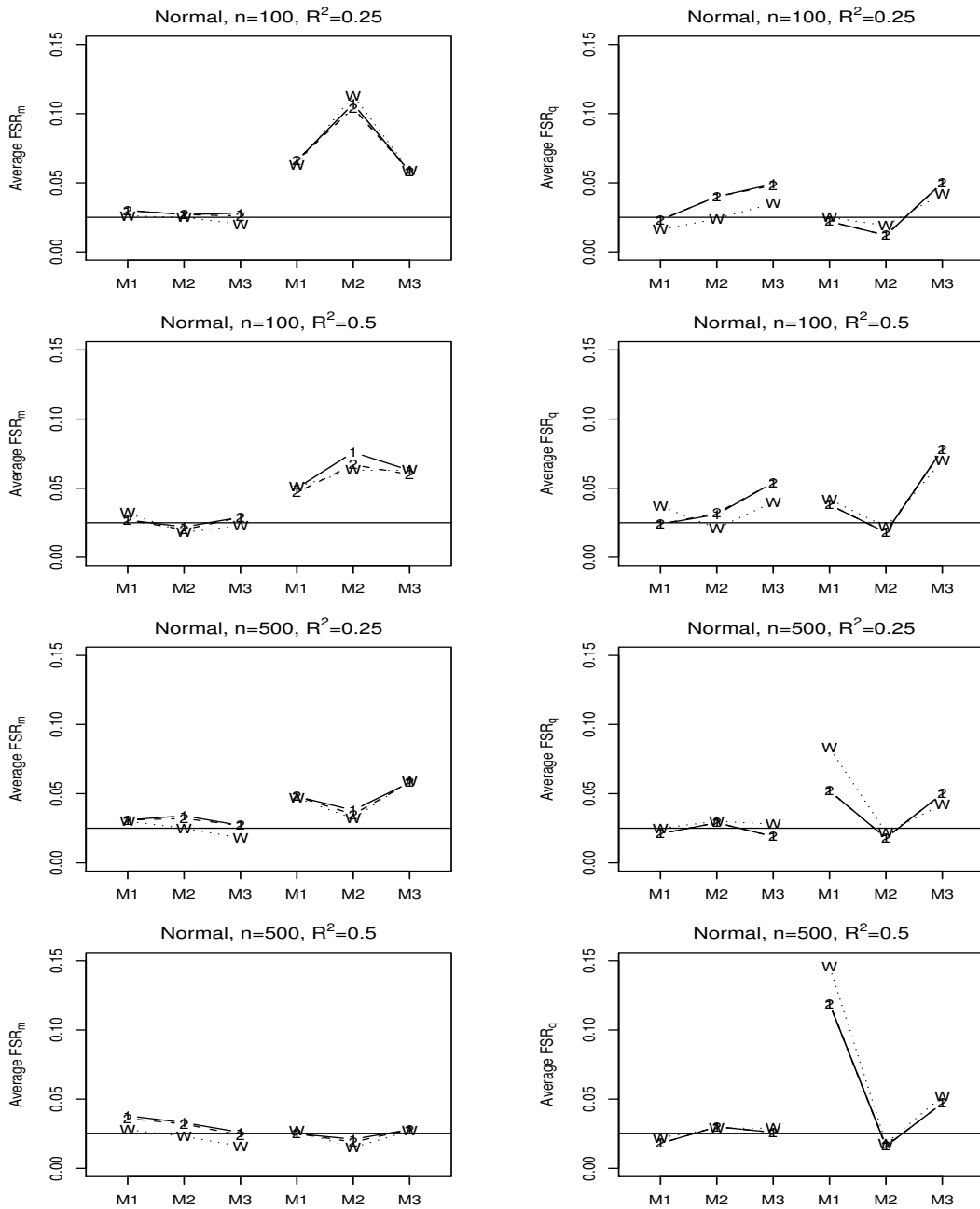


Figure 5.8: Comparison of average FSR_m (left) and FSR_q (right) using Normal predictors for FFSR-NHA1 (1), FFSR-NHA2 (2), and FFSR-WHA2 (W). The first three values in each plot are uncorrelated predictors and the second three values are correlated predictors.

CHAPTER 6

Applications

In this chapter we apply our methods to several response surface experiments to see how our methods compare to the standard approach for optimizing a response. First, we give a brief discussion of the basic approach for response surfaces, and then we conduct a simulation study to compare our methods on their ability to optimize a response. Following the study, we illustrate our methods using real data examples.

6.1 Response Surface Study

6.1.1 Introduction

Response Surface Methodology (RSM) is the process of designing and analyzing experiments in order to optimize some response(s). RSM has many applications including economics, chemical processes, industrial processes, food science, etc. A brief description of the basic approach follows. For more details, refer to Myers and Montgomery (2002).

RSM can be broken into two stages: variable screening and response surface exploration. The first stage of response optimization is determining which variables are most influential in the variability of the response. This is often referred to as variable or factor screening. Some screening designs include two-level factorial, fractional

Chapter 6. Applications

factorial, and Plackett-Burmann designs. Typically only main effects are examined in screening designs, since many researchers believe that the most influential variables also have strong main effects. By ignoring interactions, they are able to reduce the number of runs in the experiment. However, screening designs can miss important variables that have strong interactions but weak main effects. Therefore, some researchers have proposed expanded screening approaches to test for interactions (Lewis and Dean, 2001).

Screening designs are followed by response surface designs, such as central composite, small composite, hybrid, and Box-Behnken designs. These designs are more focused on improving the predictive accuracy of the model by allowing for the estimation of interaction and quadratic terms. This stage often consists of many sequential experiments where researchers try to find the region of optimality. But the ultimate goal of these experiments is finding the factor levels that optimize the response given a set of constraints. Therefore, at some point in this stage an experiment is run to determine these levels and estimate the optimal response. The standard approach for optimization is to fit a full response surface and eliminate effects that are not significant at the $\alpha = 0.05$ level, except main effects with significant quadratic or interaction terms. Then, the resulting model is used to optimize the response.

Fast FSR methods are useful for analyzing experimental designs in both stages of RSM. For screening designs focused on main effects the basic Fast FSR method can be applied. In some cases screening designs allow interaction effects to be estimated. In these situations the methods discussed in Chapters 3 and 4 are appropriate. They are also useful in the second stage of RSM for tweaking the process to optimize the response. In the following section, we study the usefulness of the Fast FSR methods

when optimizing a response in the second stage of RSM.

6.1.2 Simulation Design

In this section, we conduct a simulation study for response optimization. We assume that a screening experiment has been conducted and analyzed, yielding a set of variables. Using these variables, a second experiment is conducted in order to optimize the response. We use our methods to analyze the experiment and then optimize the response to get $\widehat{\mathbf{X}}_{opt}$. More specifically, we maximize the response subject to the constraint, $-2 \leq X_j \leq 2$, for each factor. In practice, these constraints are determined by the process, cost, resources, etc. We call this constrained factor space the region of interest.

Two different response surface designs are used to generate data. The first experimental design is a 73-run small composite design with $p = 10$ factors using the Draper-Lin method (Draper and Lin, 1990). The second experimental design is a 100-run orthogonal central composite design with $p = 8$ factors. The coded designs can be found in the Appendix B.

For each experimental design, responses are generated from two models using $N = 100$ replicates and three theoretical values of R^2 : 0.5, 0.75, and 0.9. Models 1a and 1b are generated using the small composite design, whereas Models 2a and 2b use the central composite design. The models are defined below.

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

Chapter 6. Applications

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

$$\begin{aligned} \text{2a: } Y = & 20 + 2X_1 - 3X_2 + 2X_3 - 3X_4 + 2.5X_5 + 2X_6 + 1.5X_7 - 4X_1^2 - 3X_2^2 + \\ & 4X_3^2 - 2X_4^2 + 5X_1X_2 + 3X_1X_3 + 1.5X_2X_4 + 2X_4X_5 + \epsilon \end{aligned}$$

$$\begin{aligned} \text{2b: } Y = & 20 + 5X_1 - 3X_2 + 2X_3 - 4X_4 + 3.5X_5 + 2X_6 + 1.5X_7 - 3.5X_1^2 - 3X_2^2 + \\ & X_3^2 - 2X_4^2 - X_1X_2 + 3X_1X_3 + 1.5X_2X_4 + 2X_4X_5 + \epsilon \end{aligned}$$

For each model being generated, $\epsilon \sim N(0, \sigma^2)$, where σ is chosen to achieve the desired theoretical R^2 .

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 in Model 1a, the hierarchy-based approaches are at a disadvantage. In Model 2b, the effects of X_1 , X_4 , and X_5 are larger to give the hierarchy-based methods an advantage.

Since we assume that screening is conducted prior to the response surface design, we created models where most main effects are present. In all the models, only one variable has no effect on the response. Tables 6.1 and 6.2 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 our region of interest. In Models 1a and 1b, only variable, X_9 , has an optimal level within the region. For Models 2a and 2b, variables X_1 , X_2 , and X_4 all have optimal levels within the region.

Table 6.1: Optimal Levels for Small Composite Design with 10 Factors

Model	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	Y_{max}
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 6.2: Optimal Levels for Central Composite Design with 8 Factors

Model	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	Y_{max}
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

6.1.3 Measures of Interest

The goal of RSM is to estimate the levels of a process that yield an optimal response. Therefore, we need a measure of how well a method identifies this optimum. First, we define the true mean optimal response as $\mu(\mathbf{X}_{opt})$. Then given a set of optimal factor levels, the mean response using these levels is defined as $\mu(\widehat{\mathbf{X}}_{opt})$. For any factor not selected we set the optimal level at the center point, 0. We might also refer to $\mu(\widehat{\mathbf{X}}_{opt})$ as our actual performance, whereas $\mu(\mathbf{X}_{opt})$ is the optimal performance. Their difference, $\mu(\widehat{\mathbf{X}}_{opt}) - \mu(\mathbf{X}_{opt})$, will be our measure of how close a method performs relative to the true optimal performance.

6.1.4 Simulation Results

The tables of results can be found in Appendix A. They are organized by model and R^2 . The values in parenthesis represent the standard error of the corresponding mean.

Chapter 6. Applications

Figures 6.1 - 6.2 illustrate the mean performance for True Model (T), Standard Approach (U), LASSO (L), FFSR-NH (N), FFSR-SH (S), FFSR-WH (W), FFSR-NHA2 (A), and FFSR-WHA2 (D).

For Model 1a, Fast FSR with strong hierarchy (FFSR-SH) and weak hierarchy (FFSR-WH) performed poorly. The reason is that the quadratic term, X_9^2 , and interaction, X_1X_9 , are both large effects, but the main effect X_9 is relatively small. The LASSO and no hierarchy with adjustment (FFSR-NHA2) performed the best. The LASSO was better for $R^2 = 0.5$ whereas FFSR-NHA2 was better for $R^2 = 0.9$. For Model 1b, the methods performed fairly equally with FFSR-SH, FFSR-NHA2, and LASSO among the best. As usual, the LASSO does best with smaller R^2 , and FFSR-NHA2 performs better for larger R^2 . For Model 2a, FFSR-SH and FFSR-WH performed poorly. In this model, X_1 is a very important variable, but its main effect is relatively small. Therefore, Fast FSR with no hierarchy (FFSR-NH) and FFSR-NHA2 performed best. For Model 2b, FFSR-NHA2 performed the best overall.

Recall that the standard approach is to fit the full response surface and eliminate effects not significant at the 0.05 level. This standard approach performed poorly for Models 1a and 1b, but performed very well for Models 2a and 2b. A few possible reasons for the poor performance in Models 1a and 1b are the sparsity of the true models, the larger 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-NHA2 especially in studies with a large number of factors. From this study, it is clear that the power for a method to select informative quadratic and interaction terms is important when optimizing a response.

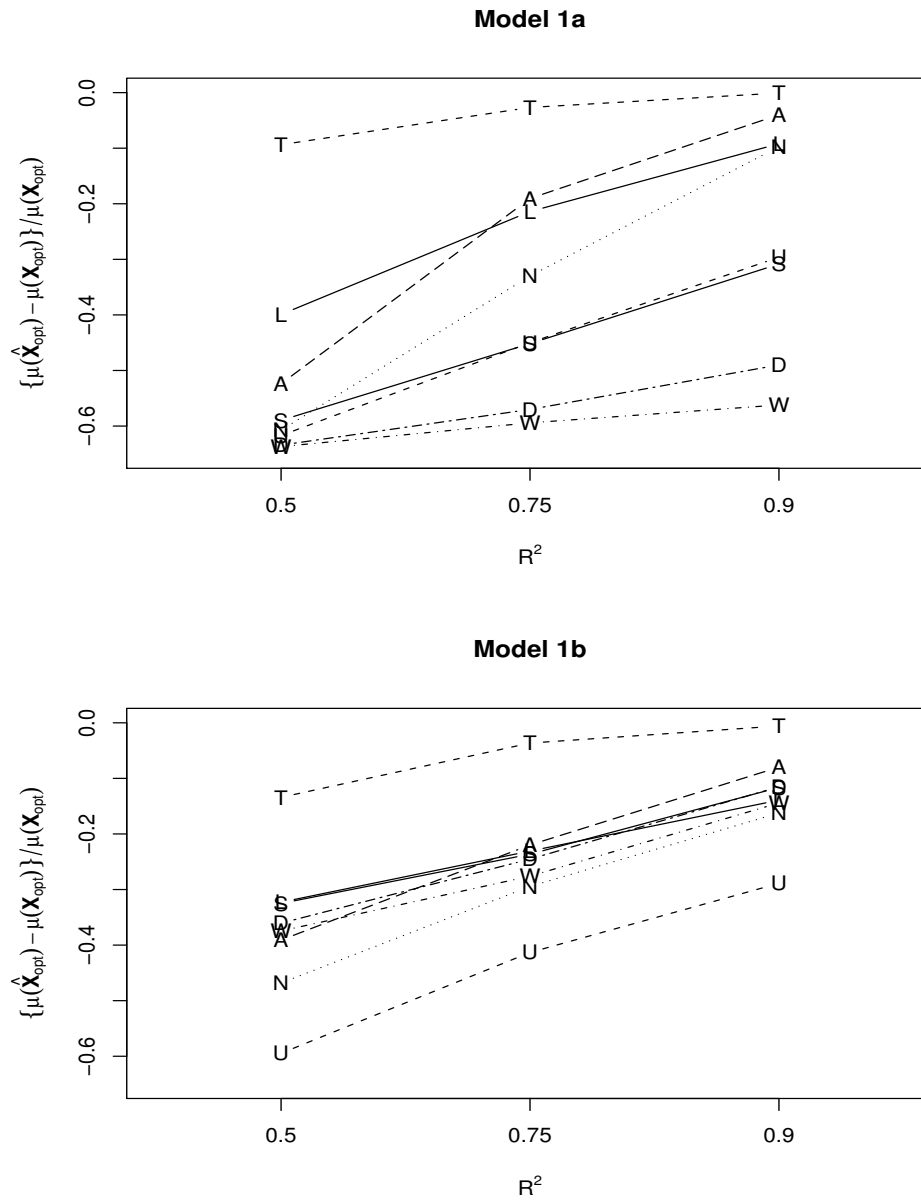


Figure 6.1: Scaled average difference in actual and optimal performance for Model 1. True Model (T), Standard Approach (U), LASSO (L), FFSR-NH (N), FFSR-SH (S), FFSR-WH (W), FFSR-NHA2 (A), and FFSR-WHA2 (D).

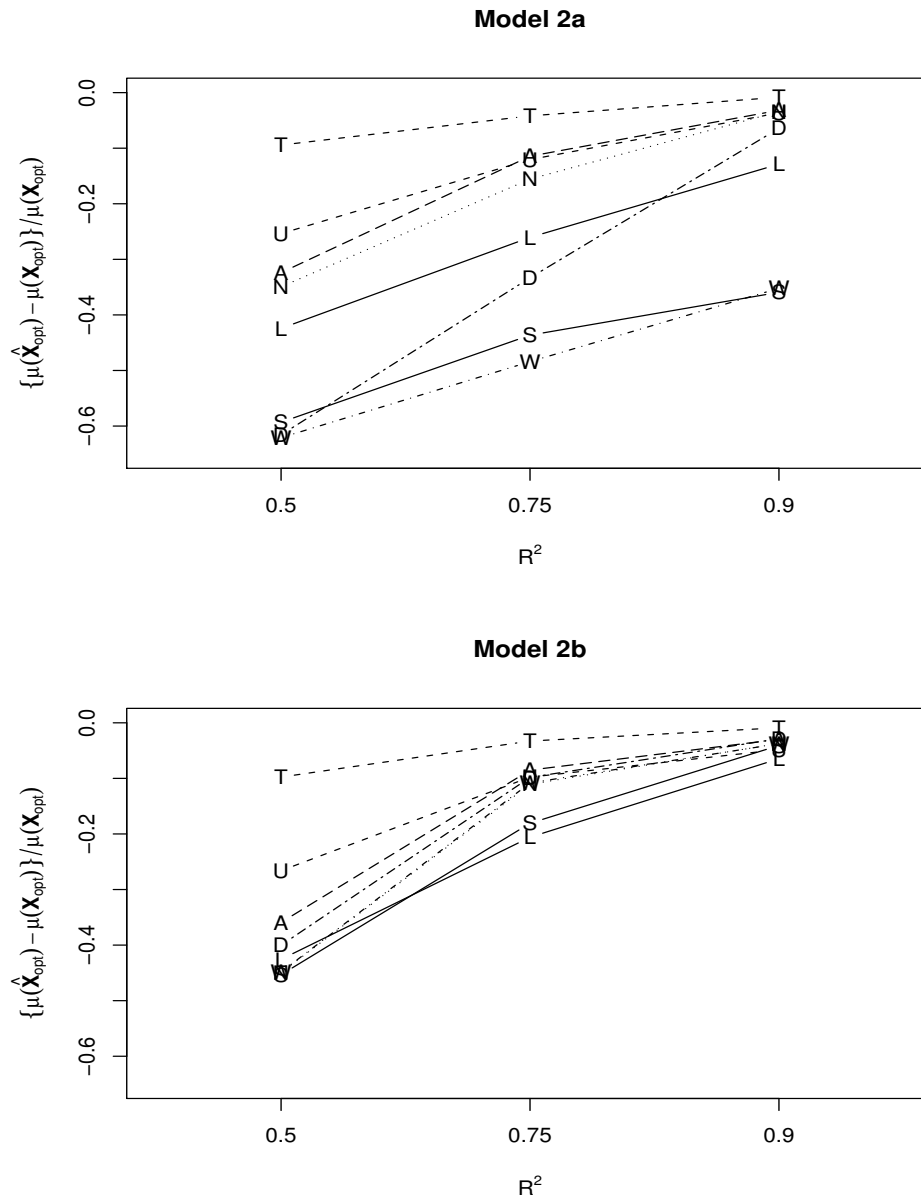


Figure 6.2: Scaled average difference in actual and optimal performance for Model 2. True Model (T), Standard Approach (U), LASSO (L), FFSR-NH (N), FFSR-SH (S), FFSR-WH (W), FFSR-NHA2 (A), and FFSR-WHA2 (D).

Table 6.3: Variables in Cutinase Study

Variable	Name	Units
X_1	flaxseed oil	%
X_2	K_2HPO_4	%
X_3	$MgSO_4$	%
X_4	$NaNO_3$	%
X_5	KCl	%
X_6	$FeSO_4 \cdot 7H_2O$	%
Y	cutinase	U/mL

6.2 Real Data Examples

In this section we apply our methods to real data examples. The first two examples focus on optimizing response surface experiments. The remainder of the examples focus on fitting a simple, but accurate predictive model.

6.2.1 Cutinase Study

Cutinase is an enzyme excreted by fungi that shows potential for use in many industrial products and processes, including laundry and dish washing detergents. Many current procedures to produce cutinase are time-consuming, expensive, and have low yield. Pio and Macedo (2008) used RSM to study the key factors for maximizing the production of cutinase. Table 6.3 describes the variables from the study.

In order to produce cutinase, they prepared a medium of potato dextrose agar and added nutrients before cultivation. They were interested in determining the optimal concentration levels of the nutrients: flaxseed oil; K_2HPO_4 ; $MgSO_4$; $NaNO_3$; KCl; and

Chapter 6. Applications

$\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$. Using a 20-run fractional factorial design, they determined that flaxseed oil, NaNO_3 , KCl , and $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ were all significant factors. Despite this find, they proceeded with an 81-run central composite design using all six factors to optimize cutinase production. After eliminating insignificant terms they report the final model:

$$Y = 10.73 - 0.66X_1 - 0.79X_1^2 - 0.70X_2 - 0.94X_2^2 - 0.84X_3 - 1.08X_3^2 - 1.56X_4 - 0.74X_4^2 - 0.61X_5 - 0.67X_5^2 - 1.17X_6 - 0.96X_6^2.$$

Using this model, they conclude that the optimal conditions for maximizing cutinase production can be met by setting the factors at their low level, -1 . Setting the nutrients at this level yields an estimated maximum production of $\hat{Y}_{max} = 11.09$ U/mL.

We used our methods from Chapters 3 and 4 to analyze the experiment. The selected models are listed in Table 6.4. The Fast FSR model suggests that all six nutrients are important for maximizing cutinase production. K_2HPO_4 , MgSO_4 , and NaNO_3 all interacted positively with $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$. Similarly, K_2HPO_4 and MgSO_4 interacted positively with NaNO_3 . The only nutrients that seemed to counteract each other were MgSO_4 and KCl .

Using the models, we maximized cutinase production subject to the constraint, $-1 \leq X_j \leq 1$, for each factor. The constraint is chosen based on the results from the original paper. Before discussing our results, a few notes on the data are necessary. In the paper, for run 65 the coded value for flaxseed oil is given as $+\alpha$, but should be $-\alpha$. For run 77, the uncoded value for KCl should be 0.03 instead of 0.00. Also the value for α is omitted in the original paper, so the results are given in terms of standardized factor levels.

Table 6.4: Model Summaries for Cutinase Production

Method	Effects in Model	R^2	Adjusted R^2
Pio and Macedo	$X_1, X_1^2, X_2, X_2^2, X_3, X_3^2,$ $X_4, X_4^2, X_5, X_5^2, X_6, X_6^2$	0.548	0.469
Fast FSR Methods	$X_1, X_1^2, X_2, X_2^2, X_2X_4, X_2X_6,$ $X_3, X_3^2, X_3X_4, X_3X_5, X_3X_6, X_4,$ $X_4^2, X_4X_6, X_5, X_5^2, X_6, X_6^2$	0.781	0.717
Standard Approach	$X_1, X_1^2, X_2, X_2^2, X_2X_6, X_3,$ $X_3^2, X_3X_4, X_3X_5, X_3X_6, X_4,$ $X_4^2, X_4X_6, X_5, X_5^2, X_6, X_6^2$	0.770	0.708

Table 6.5: Optimal Levels and Maximum Cutinase Production

Method	X_1	X_2	X_3	X_4	X_5	X_6	\hat{Y}_{max}
Pio and Macedo	-0.42	-0.38	-0.38	-1	-0.32	-0.59	12.33
Fast FSR Methods	-0.42	-1	-1	-1	-0.14	-1	15.56
Standard Approach	-0.42	-0.80	-1	-1	-0.15	-1	15.11

Maximizing the Pio and Macedo model results in $\hat{Y}_{max} = 12.33$ U/mL. Using $\gamma_0 = 0.05$, all the Fast FSR methods chose the 18 variable model listed in Table 6.4. Maximizing this model results in $\hat{Y}_{max} = 15.56$ U/mL. Fitting the full response surface, eliminating terms not significant at the 0.05 level, and maximizing leads to $\hat{Y}_{max} = 15.11$ U/mL. Table 6.5 gives the coded optimal factor levels and resulting maximum for these three models. Based on these results, response optimization using Pio and Macedo's model leads to underestimates of maximum cutinase production.

Table 6.6: Variables in Lipase Study

Variable	Name	Units
X_1	glucose	mg/mL
X_2	palm oil	% v/v
X_3	incubation time	hrs
X_4	inoculum density	%
X_5	agitation	rev/min
Y_1	lipase	U/mL
Y_2	specific activity	U/mg

6.2.2 Lipase Study

Lipase is another enzyme gaining use in industrial and food processes for its ability to break down various lipids. Rathi et al. (2002) used RSM to maximize both the production of lipase and its ability to break down fatty acids or specific activity. In order to produce lipase, they cultivated the bacteria, *Burkholderia cepacia*, adding concentrations of glucose and palm oil as nutrients. In addition to the nutrient factors they were interested in the effect of incubation time, inoculum density, and agitation on the two response variables. Table 6.6 describes the variables from the study.

In order to maximize lipase production and activity, a 32-run face-centered central composite design was used. From the experiment Rathi et al. fit a second-order linear model in all five factors excluding their interactions. Using their models, they conclude that maximum lipase production is 31 U/mL and activity is 110 U/mg.

We used our methods with $\gamma_0 = 0.05$ to analyze this experiment. The selected variables for each method and response are listed in Tables 6.7 - 6.8. Unlike the

Table 6.7: Model Summaries for Lipase Production

Method	Effects in Model	R^2	Adjusted R^2
Rathi et al.	$X_1, X_1^2, X_2, X_2^2, X_3, X_3^2, X_4, X_4^2, X_5, X_5^2$	0.742	0.619
FFSR-SH	No effects	0.000	0.000
FFSR-WH	No effects	0.000	0.000
FFSR-WHA2	No effects	0.000	0.000
FFSR-NH	X_2^2	0.260	0.235
FFSR-NHA2	X_2^2, X_3, X_4^2	0.534	0.484
Standard Approach	$X_2, X_2^2, X_3, X_4, X_4^2$	0.584	0.504

cutinase example, the Fast FSR methods did not choose the same effects in their final models. FFSR-SH, FFSR-WH, and FFSR-WHA2 fit small models in terms of R^2 . Since the main effects were not selected, these approaches were unable to fit the significant quadratic terms. Conversely, FFSR-NHA2 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.

Based on the Type III p -values for each effect in the two Rathi et al. models, there is evidence of overfitting. These p -values are listed in Table 6.9. Only X_2^2, X_3, X_4^2 , and X_5^2 are statistically significant for both responses. We ran a simulation analysis to see if these model R^2 are reasonable under the assumption that the FFSR-NHA2 models are true. Assuming that the FFSR-NHA2 models were the true models, we simulated

Table 6.8: Model Summaries for Specific Activity

Method	Effects in Model	R^2	Adjusted R^2
Rathi et al.	$X_1, X_1^2, X_2, X_2^2, X_3, X_3^2, X_4, X_4^2, X_5, X_5^2$	0.806	0.714
FFSR-SH	X_3	0.171	0.144
FFSR-WH	X_3	0.171	0.144
FFSR-WHA2	X_3	0.171	0.144
FFSR-NH	$X_2, X_2^2, X_3, X_4^2, X_5^2$	0.753	0.706
FFSR-NHA2	X_2, X_2^2, X_3	0.568	0.522
Standard Approach	$X_2, X_2^2, X_3, X_4, X_4^2, X_5, X_5^2$	0.778	0.713

1000 experiments using the 32-run design. From each experiment, we fit the “true” and Rathi models to obtain R^2 values. Boxplots of these values are shown in Figure 6.3. For lipase production, the average R^2 of the Rathi model is 0.664 with standard deviation of 0.097. For specific activity, the average R^2 of the Rathi model is 0.690 with standard deviation of 0.084. Based on these results, it is not unusual to observe R^2 values as large as in the actual experiment, since they fall within approximate 95% prediction intervals.

Using each model, we maximized lipase production and specific activity subject to the constraint, $-1 \leq X_j \leq 1$, for each factor. Tables 6.10 - 6.11 give the coded optimal factor levels and resulting maximum for the each method and response. Since the hierarchy-based approaches fit very small models, it seems likely that these methods underestimate maximum lipase production and activity. The models fit using FFSR-NHA2 are reasonable but differ somewhat. The quadratic term for inoculum density,

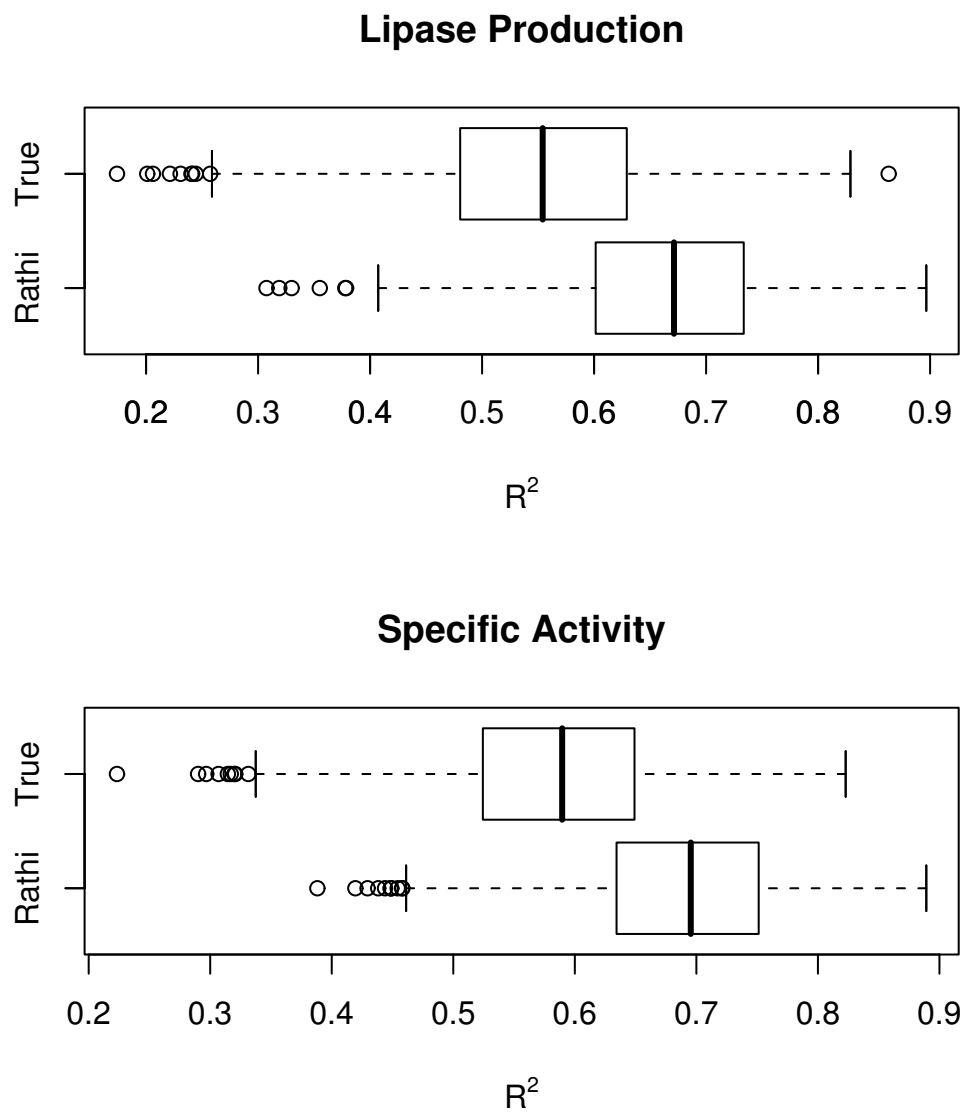


Figure 6.3: Boxplots of simulated R^2 values for lipase production and specific activity

Table 6.9: Type III p -values for Rathi et al. Models

Effect	Lipase Production	Specific Activity
X_1	0.953	0.902
X_2	0.068	0.003
X_3	0.004	<0.001
X_4	0.544	0.951
X_5	0.231	0.119
X_1^2	0.122	0.200
X_2^2	0.001	<0.001
X_3^2	0.092	0.172
X_4^2	<0.001	<0.001
X_5^2	0.024	0.007

X_4^2 , is not chosen in the model for lipase activity. Without this term the model estimates maximum activity around 73 U/mg, whereas including this term increases the estimate to around 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 $\hat{\gamma}_{fast} = 0.0507$. Using this model for maximization leads to $\hat{Y}_{2,max} = 100.63$ U/mg. Based on these results, the optimal factor levels for glucose, palm oil and agitation are basically at 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, maximum lipase production is around 23 U/mL and maximum activity around 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 provided by Rathi et al.

Table 6.10: Optimal Levels for Maximum Lipase Production

Method	X_1	X_2	X_3	X_4	X_5	$\hat{Y}_{1,max}$
Rathi et al.	0.01	0.09	-1	1	0.09	31.11
FFSR-SH	-	-	-	-	-	10.01
FFSR-WH	-	-	-	-	-	10.01
FFSR-WHA2	-	-	-	-	-	10.01
FFSR-NH	-	0	-	-	-	13.67
FFSR-NHA2	-	0	-1	1	-	22.98
Standard Approach	-	0.09	-1	1	-	26.59

Table 6.11: Optimal Levels for Maximum Specific Activity

Method	X_1	X_2	X_3	X_4	X_5	$\hat{Y}_{2,max}$
Rathi et al.	0.02	0.16	-1	1	0.10	110.99
FFSR-SH	-	-	-1	-	-	59.94
FFSR-WH	-	-	-1	-	-	59.94
FFSR-WHA2	-	-	-1	-	-	59.94
FFSR-NH	-	0.16	-1	1	0	100.63
FFSR-NHA2	-	0.19	-1	-	-	73.61
Standard Approach	-	0.16	-1	-1	0.10	100.45

6.2.3 Diabetes Study

Efron et al. (2004) use the following example to illustrate the LARS method. The study contains ten baseline measurements and a response for 442 diabetes patients. The response variable of interest is a numerical measure of diabetes progression one year after baseline. The predictors are age, gender, body mass index, average blood

Table 6.12: Variables in Diabetes Study

Variable	Name	Description
X_1	age	Patient's age
X_2	gender	Patient's gender
X_3	BMI	Body mass index
X_4	BP	Average blood pressure
X_5	S1	Blood serum measurement 1
X_6	S2	Blood serum measurement 2
X_7	S3	Blood serum measurement 3
X_8	S4	Blood serum measurement 4
X_9	S5	Blood serum measurement 5
X_{10}	S6	Blood serum measurement 6
Y	disease	Disease progression one year after baseline

pressure, and six serum measurements. The goals of the study were to develop a model that accurately predicts baseline responses for future patients and to understand which effects are most important for disease progression. Table 6.12 describes the variables from the study.

We used the Fast FSR methods with $\gamma_0 = 0.05$ and the LASSO with 5-fold Cross Validation to fit quadratic models using the data set. Note that there are only nine candidate quadratic terms, because gender is a binary variable. The selected effects for each method are listed in Table 6.13. In terms of R^2 , all the methods chose similar models, but the LASSO chose larger sized models. The main effects of gender and serum measurement 5 are common to all methods as well as the interaction between BMI and blood pressure. The interaction between a patient's age and gender was

Table 6.13: Model Summaries for Disease Progression

Method	Effects in Model	R^2	Adjusted R^2
LASSO with CV	$X_1X_2, X_1X_4, X_1X_{10}, X_2, X_3, X_3^2,$ $X_3X_4, X_4, X_7, X_9, X_{10}^2$	0.519	0.507
FFSR-SH	$X_2, X_3, X_3X_4, X_4, X_5, X_6, X_9$	0.526	0.518
FFSR-WH	$X_1X_2, X_2, X_3, X_3X_4, X_4, X_5, X_6, X_9$	0.539	0.530
FFSR-WHA2	$X_1X_2, X_2, X_3, X_3X_4, X_4, X_5, X_6, X_9$	0.539	0.530
FFSR-NH	$X_1X_2, X_2, X_3, X_3X_4, X_4, X_7, X_9$	0.534	0.527
FFSR-NHA2	$X_1X_2, X_2, X_3, X_3X_4, X_4, X_5, X_6, X_9$	0.539	0.530

selected by all the methods except the strong hierarchy. Since the main effect for age is not selected in the model, its interaction with gender cannot be selected using the strong hierarchy. Of the remaining blood serum measurements, some of the methods chose serum measurement 3, whereas the others chose serum measurements 1 and 2.

6.2.4 Pyrimidine Study

The pyrimidine data set was studied by Hirst et al. (1994a). They used neural networks and inductive logic programming to model the quantitative structure-activity relationships (QSAR) of the inhibition of dihydrofolate reductase (DHFR) by pyrimidines. QSAR is the process of relating physicochemical and/or structural properties with some known biological or chemical process. That is, activity = f(physicochemical and structural properties). The actual data set contains structural information on 74 2,4-diamino-5-(substituted benzyl) pyrimidines used as inhibitors of DHFR in *E. coli*. There are 3 positions where chemical activity occurs and 9 attributes per position lead-

Chapter 6. Applications

ing to 27 total predictors. Tables 6.14 - 6.15 describe the variables used in the study. The size of a substituent is the number of carbon, nitrogen, and oxygen atoms that it contains.

In this example, we again use the Fast FSR methods with $\gamma_0 = 0.05$ and the LASSO with 5-fold Cross Validation to build a predictive model for understanding the drug-receptor interactions. X_{25} was removed from the study due to lack of variability. Table 6.16 gives the number of effects selected for each method along with the R^2 and Adjusted R^2 , and Table 6.17 lists the actual effects selected. None of the effects were chosen by all methods. However, the polarizability at position 1 and number of hydrogen-bond donors at position 3 are selected as main effects or an interaction in most models.

6.2.5 Triazine Study

The triazine data set was studied by Hirst et al. (1994b). They used neural networks and inductive logic programming to model the QSAR of the inhibition of DHFR by triazines. The actual data set contains structural information on 186 2,4-diamino-6,6-dimethyl-5-phenyl-dihydrotriazines used as inhibitors of DHFR in rodents. There are 6 regions of chemical variability and 10 attributes per region leading to 60 total predictors. Tables 6.18 - 6.20 describe the variables used in the study.

In this example, we again use the Fast FSR methods with $\gamma_0 = 0.05$ and the LASSO with 5-fold Cross Validation to build a predictive model for understanding the drug-receptor interactions. X_{43} and X_{44} were removed from the study due to lack of variability. Also, X_{13} and X_{14} had identical values for all observations. Similarly, X_{23} ,

Table 6.14: Variables in Pyrimidine Study

Variable	Name	Description
X_1	p1_polar	Position 1 polarity
X_2	p1_size	Position 1 number of carbon, nitrogen, and oxygen atoms
X_3	p1_flex	Position 1 flexibility
X_4	p1_h_donor	Position 1 number of hydrogen-bond donors
X_5	p1_h_acceptor	Position 1 number of hydrogen-bond acceptors
X_6	p1_pi_donor	Position 1 presence and strength of π -donors
X_7	p1_pi_acceptor	Position 1 presence and strength of π -acceptors
X_8	p1_polarizable	Position 1 polarizability
X_9	p1_sigma	Position 1 σ -effect
X_{10}	p2_polar	Position 2 polarity
X_{11}	p2_size	Position 2 number of carbon, nitrogen, and oxygen atoms
X_{12}	p2_flex	Position 2 flexibility
X_{13}	p2_h_donor	Position 2 number of hydrogen-bond donors
X_{14}	p2_h_acceptor	Position 2 number of hydrogen-bond acceptors
X_{15}	p2_pi_donor	Position 2 presence and strength of π -donors
X_{16}	p2_pi_acceptor	Position 2 presence and strength of π -acceptors
X_{17}	p2_polarizable	Position 2 polarizability
X_{18}	p2_sigma	Position 2 σ -effect
X_{19}	p3_polar	Position 3 polarity
X_{20}	p3_size	Position 3 number of carbon, nitrogen, and oxygen atoms
X_{21}	p3_flex	Position 3 flexibility
X_{22}	p3_h_donor	Position 3 number of hydrogen-bond donors
X_{23}	p3_h_acceptor	Position 3 number of hydrogen-bond acceptors
X_{24}	p3_pi_donor	Position 3 presence and strength of π -donors

Table 6.15: Variables in Pyrimidine Study Continued

Variable	Name	Description
X_{25}	p3_pi_acceptor	Position 3 presence and strength of π -acceptors
X_{26}	p3_polarizable	Position 3 polarizability
X_{27}	p3_sigma	Position 3 σ -effect
Y	activity	$\log 1/K_i$, where K_i is the inhibition constant as experimentally assayed

Table 6.16: Model Summaries for Pyrimidine Activity

Method	Number of Effects Selected	R^2	Adjusted R^2
LASSO with CV	21	0.852	0.792
FFSR-SH	5	0.687	0.664
FFSR-WH	4	0.746	0.732
FFSR-WHA2	4	0.746	0.732
FFSR-NH	7	0.871	0.858
FFSR-NHA2	7	0.881	0.869

X_{24} , and X_{25} were identical, as well as X_{53} and X_{54} . Therefore, X_{14} , X_{24} , X_{25} , and X_{54} removed from the study. From this, we can see that flexibility and hydrogen-bond donors are often equivalent within a region. Table 6.21 gives the number of effects selected for each method along with the R^2 and Adjusted R^2 , and Table 6.22 lists the actual effects selected. Hirst et al. report $R^2 = 0.252$ for their neural network. They noted that R^2 values in this example were relatively small when compared to other QSAR studies. None of the effects were chosen by all methods, but region 1

Table 6.17: Effects Selected for Pyrimidine Activity

Method	Effects Selected in Model
LASSO with CV	$X_1X_5, X_1X_{22}, X_3X_8, X_5^2, X_5X_9, X_5X_{11}, X_5X_{19}, X_6^2,$ $X_8X_{10}, X_8X_{11}, X_8X_{22}, X_9X_{15}, X_{11}X_{15}, X_{12}X_{15},$ $X_{12}X_{17}, X_{13}X_{17}, X_{14}, X_{15}X_{17}, X_{17}, X_{20}, X_{26}$
FFSR-SH	$X_8, X_{17}, X_{22}, X_{26}, X_{26}^2$
FFSR-WH	$X_1X_{22}, X_8X_{26}, X_{22}, X_{26}$
FFSR-WH	$X_1X_{22}, X_8X_{26}, X_{22}, X_{26}$
FFSR-NH	$X_1X_5, X_8X_{10}, X_8X_{22}, X_{11}X_{15}, X_{13}X_{21}, X_{20}, X_{23}X_{27}$
FFSR-NHA2	$X_5X_9, X_8X_{22}, X_{11}, X_{11}X_{12}, X_{20}, X_{21}, X_{23}X_{27}$

polarizability and branching are selected by most.

Table 6.18: Variables in Triazine Study

Variable	Name	Description
X_1	p1_polar	Region 1 polarity
X_2	p1_size	Region 1 number of carbon, nitrogen, and oxygen atoms
X_3	p1_flex	Region 1 flexibility
X_4	p1_h.donor	Region 1 number of hydrogen-bond donors
X_5	p1_h.acceptor	Region 1 number of hydrogen-bond acceptors
X_6	p1_pi.donor	Region 1 presence and strength of π -donors
X_7	p1_pi.acceptor	Region 1 presence and strength of π -acceptors
X_8	p1_polarizable	Region 1 polarizability
X_9	p1_sigma	Region 1 σ -effect
X_{10}	p1_branch	Region 1 branching
X_{11}	p2_polar	Region 2 polarity
X_{12}	p2_size	Region 2 number of carbon, nitrogen, and oxygen atoms
X_{13}	p2_flex	Region 2 flexibility
X_{14}	p2_h.donor	Region 2 number of hydrogen-bond donors
X_{15}	p2_h.acceptor	Region 2 number of hydrogen-bond acceptors
X_{16}	p2_pi.donor	Region 2 presence and strength of π -donors
X_{17}	p2_pi.acceptor	Region 2 presence and strength of π -acceptors
X_{18}	p2_polarizable	Region 2 polarizability
X_{19}	p2_sigma	Region 2 σ -effect
X_{20}	p2_branch	Region 2 branching
X_{21}	p3_polar	Region 3 polarity
X_{22}	p3_size	Region 3 number of carbon, nitrogen, and oxygen atoms
X_{23}	p3_flex	Region 3 flexibility
X_{24}	p3_h.donor	Region 3 number of hydrogen-bond donors
X_{25}	p3_h.acceptor	Region 3 number of hydrogen-bond acceptors

Table 6.19: Variables in Triazine Study Continued

Variable	Name	Description
X_{26}	p3_pi_donor	Region 3 presence and strength of π -donors
X_{27}	p3_pi_acceptor	Region 3 presence and strength of π -acceptors
X_{28}	p3_polarizable	Region 3 polarizability
X_{29}	p3_sigma	Region 3 σ -effect
X_{30}	p3_branch	Region 3 branching
X_{31}	p4_polar	Region 4 polarity
X_{32}	p4_size	Region 4 number of carbon, nitrogen, and oxygen atoms
X_{33}	p4_flex	Region 4 flexibility
X_{34}	p4_h_donor	Region 4 number of hydrogen-bond donors
X_{35}	p4_h_acceptor	Region 4 number of hydrogen-bond acceptors
X_{36}	p4_pi_donor	Region 4 presence and strength of π -donors
X_{37}	p4_pi_acceptor	Region 4 presence and strength of π -acceptors
X_{38}	p4_polarizable	Region 4 polarizability
X_{39}	p4_sigma	Region 4 σ -effect
X_{40}	p4_branch	Region 4 branching
X_{41}	p5_polar	Region 5 polarity
X_{42}	p5_size	Region 5 number of carbon, nitrogen, and oxygen atoms
X_{43}	p5_flex	Region 5 flexibility
X_{44}	p5_h_donor	Region 5 number of hydrogen-bond donors
X_{45}	p5_h_acceptor	Region 5 number of hydrogen-bond acceptors
X_{46}	p5_pi_donor	Region 5 presence and strength of π -donors
X_{47}	p5_pi_acceptor	Region 5 presence and strength of π -acceptors
X_{48}	p5_polarizable	Region 5 polarizability
X_{49}	p5_sigma	Region 5 σ -effect
X_{50}	p5_branch	Region 5 branching

Table 6.20: Variables in Triazine Study Continued

Variable	Name	Description
X_{51}	p6_polar	Region 6 polarity
X_{52}	p6_size	Region 6 number of carbon, nitrogen, and oxygen atoms
X_{53}	p6_flex	Region 6 flexibility
X_{54}	p6_h_donor	Region 6 number of hydrogen-bond donors
X_{55}	p6_h_acceptor	Region 6 number of hydrogen-bond acceptors
X_{56}	p6_pi_donor	Region 6 presence and strength of π -donors
X_{57}	p6_pi_acceptor	Region 6 presence and strength of π -acceptors
X_{58}	p6_polarizable	Region 6 polarizability
X_{59}	p6_sigma	Region 6 σ -effect
X_{60}	p6_branch	Region 6 branching
Y	activity	$\log 1/C$, where C is the molar concentration that produces 50 % reversible inhibition of dihydrofolate reductase

Table 6.21: Model Summaries for Triazine Activity

Method	Number of Effects Selected	R^2	Adjusted R^2
LASSO with CV	28	0.309	0.243
FFSR-SH	2	0.193	0.184
FFSR-WH	3	0.256	0.244
FFSR-WHA2	3	0.256	0.244
FFSR-NH	2	0.242	0.234
FFSR-NHA2	3	0.295	0.284

Table 6.22: Effects Selected for Triazine Activity

Method	Effects Selected in Model
LASSO with CV	$X_1X_2, X_1X_5, X_1X_{13}, X_5X_{31}, X_6X_{10}, X_8X_{33}, X_9, X_9X_{23},$ $X_{10}X_{16}, X_{10}X_{36}, X_{10}X_{40}, X_{11}X_{33}, X_{11}X_{39}, X_{11}X_{42}, X_{11}X_{48},$ $X_{21}X_{38}, X_{26}X_{56}, X_{31}^2, X_{31}X_{38}, X_{32}^2, X_{32}X_{37}, X_{33}, X_{33}X_{36},$ $X_{33}X_{41}, X_{34}X_{35}, X_{36}^2, X_{36}X_{50}, X_{37}X_{51}$
FFSR-SH	X_8, X_{10}
FFSR-WH	X_8, X_8X_{33}, X_{10}
FFSR-WHA2	X_8, X_8X_{33}, X_{10}
FFSR-NH	$X_{10}X_{40}, X_{11}X_{39}$
FFSR-NHA2	$X_8, X_{10}X_{40}, X_{11}X_{39}$

CHAPTER 7

Conclusions

This research has focused on studies where fitting second-order linear regression models is appropriate. When the number of predictors or factors is large, the full second-order model is often unestimable. Therefore, variable and model selection are necessary. A simple approach is to treat each term as a separate variable and apply standard variable and model selection techniques. However, many of the standard approaches have a tendency to select a large number of uninformative interactions. Part of the problem is the correlation between interactions and their parent main effects. In order to alleviate this multicollinearity, the predictors should be centered and possibly standardized before creating the interactions and quadratic terms. After resolving the issue of multicollinearity, handling the large number of interactions is a problem in itself.

Enforcing some level of hierarchy is one approach to limit the number of interactions in the model. The strong and weak hierarchy require parent main effects in the model before interactions and quadratic terms can enter. These constraints work well in some cases, as indicated by the simulation study in Chapter 5. However, some situations arise where an interaction or quadratic term has a large effect on the response, but its parent main effect is small. In these situations, the strong hierarchy is often unable to fit informative interactions or quadratic terms. The weak hierarchy handles interactions better, but the quadratic terms are still problematic. The lipase example is a good illustration of this limitation in a real study. One suggestion for improving these

Chapter 7. Conclusions

methods is to allow quadratic terms to enter the model without requiring their parent main effects.

A second approach for limiting the number of interactions is by lowering their probability of entry into the model. We accomplish this task by choosing separate entry criteria for main effects and interactions. We choose the entry-level to control both group false selection rates as well as the overall rate. By decreasing the selection of uninformative interactions, we also decrease the power to select informative interactions. Despite this drawback, the approach performed well in our simulations, especially for optimizing response surfaces. Also, the approach can be applied in more general situations where the researcher can identify two groups of variables to control.

In this paper, we have presented many approaches for fitting second-order response surfaces. Our methods avoid overfitting by controlling the selection of uninformative effects. This is especially appealing in studies where researchers are focused on data exploration and model interpretation. The hierarchy approaches with forward selection performed well in the Chapter 5 simulations, but often performed poorly for the response surface experiments. Therefore, these approaches are recommended for most studies. However, in response surface experiments, no hierarchy with adjustment (FFSR-NHA1 or FFSR-NHA2) is recommended for use. In either case, if $p < 5$, Fast FSR with no hierarchy should suffice. The weak hierarchy and adjustment methods are not available in statistical packages, but the strong hierarchy and no hierarchy algorithms are available for use. If limited to these two approaches, then Fast FSR with strong hierarchy is preferred for large p .

It is crucial to center/standardize the predictors before using methods that do not maintain the hierarchy. Also, sweeping out the linear effects from interactions and

Chapter 7. Conclusions

quadratic terms could be a useful alternative for highly skewed predictors but needs further examination. In either case, it is important to transform the predictors back to their original form when interpreting the final model. Variable and model selection methods are very useful for data modeling but cannot take the place of scientific reasoning. Therefore, researchers should use their scientific knowledge in combination with these methods to assure that the goals of their study are met.

Bibliography

- W.P. Alexander and S.D. Grimshaw. Treed regression. *Journal of Computational and Graphical Statistics*, 5:156–175, 1996.
- J.S. Armstrong and J.G. Andress. Exploratory analysis of marketing data: Trees vs. regression. *Journal of Marketing Research*, 7:487–492, 1970.
- L. Breiman. Better subset regression using the nonnegative garrote. *Technometrics*, 37:373–384, 1995.
- L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone. *Classification and Regression Trees*. Chapman & Hall/CRC, 1998.
- L. Brieman. Bagging predictors. *Machine Learning*, 24:123–140, 1996.
- L. Brieman. Random forests. *Machine Learning*, 45:5–32, 2001.
- K.-Y. Chan and W.-Y. Loh. Lotus: An algorithm for building accurate and comprehensible logistic regression trees. *Journal of Computational and Graphical Statistics*, 13:826–852, 2004.
- P. Doyle. The use of automatic interaction detector and similar search procedures. *Operational Research Quarterly*, 24:465–467, 1973.
- N. Draper and D. Lin. Small response-surface designs. *Technometrics*, 32:187–194, 1990.
- B. Efron, I. Johnstone, T. Hastie, and R. Tibshirani. Least angle regression. *The Annals of Statistics*, 32:407–499, 2004.
- R.L. Eubank. *Nonparametric Regression and Spline Smoothing*. Marcel Dekker, second edition, 1999.
- J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96:1348–1360, 2001.
- J.H. Friedman. Multivariate adaptive regression splines. *The Annals of Statistics*, 19:1–67, 1991.
- E.I. George. The variable selection problem. *Journal of the American Statistical Association*, 95:1304–1308, 2000.

BIBLIOGRAPHY

- G.L. Griepentrog, J.M. Ryan, and L.D. Smith. Linear transformations of polynomial regression models. *The American Statistician*, 36:171–174, 1982.
- D.M. Hawkins. FIRM: Formal inference-based recursive modeling. Technical Report 546, School of Statistics, University of Minnesota, 1999.
- C. Heymann. XAID: an extended automatic interaction detector. Internal Report, SWISK 28, Council for Scientific and Industrial Research, Pretoria, South Africa, 1981.
- J.D. Hirst, R.D. King, and M.J.E. Sternberg. Quantitative structure-activity relationships by neural networks and inductive logic programming. I: The inhibition of dihydrofolate reductase by pyrimidines. *Journal of Computer-Aided Molecular Design*, 8:405–420, 1994a.
- J.D. Hirst, R.D. King, and M.J.E. Sternberg. Quantitative structure-activity relationships by neural networks and inductive logic programming. II: The inhibition of dihydrofolate reductase by triazines. *Journal of Computer-Aided Molecular Design*, 8:421–432, 1994b.
- J. Jaccard, C.K. Wan, and R. Turrisi. The detection and interpretation of interaction effects between continuous variables in multiple regression. *Multivariate Behavioral Research*, 25:467–478, 1990.
- G.V. Kass. Significance testing in automatic interaction detection (A.I.D.). *Applied Statistics*, 24:178–189, 1975.
- H. Kim and W.-Y. Loh. Classification trees with unbiased multiway splits. *Journal of the American Statistical Association*, 96:598–604, 2001.
- C. Kooperberg, S. Bose, and C.J. Stone. Polychotomous regression. *Journal of the American Statistical Association*, 92:117–127, 1997.
- M. LeBlanc and R. Tibshirani. Monotone shrinkage of trees. *Journal of Computational and Graphical Statistics*, 7:417–433, 1998.
- S.M. Lewis and A.M. Dean. Detection of interactions in experiments on large numbers of factors. *Journal of the Royal Statistical Society*, 63:633–673, 2001.
- W.-Y. Loh. Regression trees with unbiased variable selection and interaction detection. *Statistica Sinica*, 12:361–386, 2002.

BIBLIOGRAPHY

- W.-Y. Loh and Y.-S. Shih. Split selection methods for classification trees. *Statistica Sinica*, 7:815–840, 1997.
- A. Miller. *Subset Selection in Regression*. Chapman & Hall/CRC, second edition, 2002.
- J.N. Morgan. History and potential of binary segmentation for exploratory data analysis. *Journal of Data Science*, 3:123–136, 2005.
- J.N. Morgan and J.A. Sonquist. Problems in the analysis of survey data, and a proposal. *Journal of the American Statistical Association*, 58:415–434, 1963.
- R.H. Myers and D.C. Montgomery. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. Wiley, second edition, 2002.
- M.Y. Park and Trevor Hastie. Penalized logistic regression for detecting gene interactions. Technical Report, Stanford University, 2006.
- J.L. Peixoto. Hierarchical variable selection in polynomial regression models. *The American Statistician*, 41:311–313, 1987.
- W.D. Perreault and H.C. Barksdale. A model-free approach for analysis of complex contingency data in survey research. *Journal of Marketing Research*, 17:503–515, 1980.
- T.F. Pio and G.A. Macedo. Cutinase production by *Fusarium oxysporum* in liquid medium using central composite design. *Journal of Industrial Microbiology and Biotechnology*, 35:59–67, 2008.
- P. Rathi, V.K. Goswami, V. Sahai, and R. Gupta. Statistical medium optimization and production of a hyperthermostable lipase from *Burkholderia cepacia* in a bioreactor. *Journal of Applied Microbiology*, 93:930–936, 2002.
- J.A. Sonquist. Finding variables that work. *The Public Opinion Quarterly*, 33:83–95, 1969.
- R. Staelin. A note on detection of interaction. *The Public Opinion Quarterly*, 34:408–411, 1970.
- R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, 58:267–288, 1996.
- M. van Diepen and P.H. Franses. Evaluating chi-squared automatic interaction detection. *Information Systems*, 31:814–831, 2006.

BIBLIOGRAPHY

- Y. Wu, D.D. Boos, and L.A. Stefanski. Controlling variable selection by the addition of pseudovariables. *Journal of the American Statistical Association*, 102:235–243, 2007.
- M. Yuan, V.R. Joseph, and Y. Lin. An efficient variable selection approach for analyzing designed experiments. *Technometrics*, 49:430–439, 2007a.
- M. Yuan, V.R. Joseph, and H. Zou. Structured variable selection and estimation. Technical Report, 2007b.
- H. Zhang, C.-Y. Yu, H. Zhu, and J. Shi. Identification of linear directions in multivariate adaptive spline models. *Journal of the American Statistical Association*, 98:369–376, 2003.
- H. Zou. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101:1418–1429, 2006.

Appendix

APPENDIX A

Simulation Results

A.1 Results for Chapter 5 Simulation Study

1. Simulation results for $R^2 = 0.25$: Tables A.1 - A.12
2. Simulation results for $R^2 = 0.5$: Tables A.13 - A.24
3. Plots of AME ratio using Chi Square predictors: Figure A.1
4. Plots of average FSR and CSR using Chi Square predictors: Figure A.2
5. Plots of average FSR_m and FSR_q using Chi Square predictors: Figure A.3

A.2 Results for Chapter 6 Simulation Study

1. Simulation results for Model 1a: Table A.25
2. Simulation results for Model 1b: Table A.26
3. Simulation results for Model 2a: Table A.27
4. Simulation results for Model 2b: Table A.28

Appendix A. Simulation Results

Table A.1: Comparison of Methods for Normal Uncorrelated, Model 1, $R^2 = 0.25$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.171	1.980	NA	NA	0.031	NA	0.493
	MARS	0.105	10.310	NA	NA	0.478	NA	1.999
	LASSO	0.267	7.030	0.032	0.487	0.519	0.362	0.655
	FFSR-NH	0.210	2.120	0.000	0.075	0.075	0.142	0.492
	FFSR-NHS	0.214	2.080	0.000	0.063	0.063	0.145	0.486
	FFSR-SH	0.308	2.870	0.031	0.013	0.044	0.280	0.668
	FFSR-WH	0.284	2.680	0.013	0.030	0.043	0.255	0.640
	FFSR-NHA1	0.261	2.890	0.028	0.049	0.077	0.263	0.712
	FFSR-NHA2	0.262	2.840	0.026	0.048	0.074	0.258	0.702
	FFSR-WHA2	0.286	2.820	0.020	0.035	0.054	0.268	0.674
	Best No	0.288	2.650	0.003	0.024	0.026	0.257	0.624
	Best Strong	0.445	4.320	0.049	0.017	0.066	0.483	0.868
	<i>Average SE</i>	<i>0.014</i>	<i>0.160</i>	<i>0.006</i>	<i>0.010</i>	<i>0.013</i>	<i>0.016</i>	<i>0.040</i>
500	CART	0.054	4.540	NA	NA	0.026	NA	0.731
	MARS	0.062	17.030	NA	NA	0.402	NA	1.443
	LASSO	0.187	12.970	0.034	0.437	0.471	0.853	0.927
	FFSR-NH	0.290	5.920	0.004	0.047	0.051	0.760	0.990
	FFSR-NHS	0.296	5.900	0.004	0.044	0.048	0.762	0.989
	FFSR-SH	0.399	6.390	0.032	0.022	0.053	0.830	1.001
	FFSR-WH	0.350	6.210	0.010	0.048	0.058	0.802	1.005
	FFSR-NHA1	0.345	6.010	0.027	0.019	0.046	0.782	0.993
	FFSR-NHA2	0.345	6.010	0.027	0.019	0.046	0.782	0.993
	FFSR-WHA2	0.386	6.180	0.018	0.028	0.046	0.808	1.002
	Best No	0.453	5.990	0.001	0.015	0.016	0.813	0.993
	Best Strong	0.614	6.390	0.012	0.007	0.019	0.873	1.003
	<i>Average SE</i>	<i>0.021</i>	<i>0.159</i>	<i>0.004</i>	<i>0.007</i>	<i>0.009</i>	<i>0.012</i>	<i>0.011</i>

Appendix A. Simulation Results

Table A.2: Comparison of Methods for Normal Uncorrelated, Model 2, $R^2 = 0.25$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.178	2.030	NA	NA	0.046	NA	0.477
	MARS	0.113	10.670	NA	NA	0.481	NA	1.991
	LASSO	0.275	7.560	0.039	0.450	0.489	0.444	0.655
	FFSR-NH	0.216	2.050	0.005	0.060	0.065	0.170	0.474
	FFSR-NHS	0.216	2.020	0.005	0.051	0.056	0.170	0.459
	FFSR-SH	0.305	3.120	0.042	0.007	0.049	0.386	0.738
	FFSR-WH	0.277	2.830	0.022	0.028	0.050	0.324	0.684
	FFSR-NHA1	0.268	2.850	0.027	0.040	0.067	0.324	0.694
	FFSR-NHA2	0.267	2.840	0.027	0.040	0.067	0.322	0.692
	FFSR-WHA2	0.283	2.940	0.025	0.024	0.049	0.346	0.703
	Best No	0.294	2.780	0.000	0.036	0.036	0.328	0.669
	Best Strong	0.470	4.580	0.038	0.021	0.059	0.648	0.951
	<i>Average SE</i>	<i>0.014</i>	<i>0.156</i>	<i>0.007</i>	<i>0.010</i>	<i>0.013</i>	<i>0.019</i>	<i>0.033</i>
500	CART	0.049	3.740	NA	NA	0.017	NA	0.597
	MARS	0.060	17.510	NA	NA	0.368	NA	1.481
	LASSO	0.205	14.070	0.045	0.442	0.486	0.986	0.956
	FFSR-NH	0.353	5.920	0.000	0.051	0.051	0.912	1.004
	FFSR-NHS	0.352	5.930	0.000	0.052	0.052	0.912	1.005
	FFSR-SH	0.559	6.340	0.035	0.024	0.059	0.982	1.019
	FFSR-WH	0.468	6.190	0.007	0.044	0.051	0.962	1.018
	FFSR-NHA1	0.496	6.380	0.034	0.029	0.063	0.982	1.027
	FFSR-NHA2	0.499	6.370	0.032	0.029	0.062	0.982	1.026
	FFSR-WHA2	0.488	6.280	0.025	0.030	0.055	0.972	1.021
	Best No	0.568	5.990	0.000	0.023	0.023	0.962	1.003
	Best Strong	0.844	6.010	0.006	0.003	0.008	0.990	1.001
	<i>Average SE</i>	<i>0.029</i>	<i>0.156</i>	<i>0.004</i>	<i>0.008</i>	<i>0.009</i>	<i>0.009</i>	<i>0.011</i>

Appendix A. Simulation Results

Table A.3: Comparison of Methods for Normal Uncorrelated, Model 3, $R^2 = 0.25$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.237	2.140	NA	NA	0.077	NA	0.674
	MARS	0.122	10.390	NA	NA	0.612	NA	2.162
	LASSO	0.493	7.260	0.060	0.559	0.619	0.298	0.942
	FFSR-NH	0.545	2.190	0.014	0.037	0.051	0.208	0.800
	FFSR-NHS	0.558	2.190	0.012	0.045	0.057	0.206	0.796
	FFSR-SH	0.661	2.520	0.060	0.001	0.061	0.258	0.878
	FFSR-WH	0.678	2.340	0.024	0.016	0.040	0.242	0.839
	FFSR-NHA1	0.639	2.360	0.030	0.023	0.054	0.238	0.863
	FFSR-NHA2	0.643	2.350	0.030	0.023	0.054	0.236	0.861
	FFSR-WHA2	0.673	2.360	0.026	0.016	0.042	0.244	0.844
	Best No	0.747	2.170	0.006	0.006	0.012	0.222	0.780
	Best Strong	0.911	2.520	0.025	0.000	0.025	0.286	0.846
	<i>Average SE</i>	<i>0.040</i>	<i>0.131</i>	<i>0.008</i>	<i>0.008</i>	<i>0.013</i>	<i>0.012</i>	<i>0.037</i>
500	CART	0.107	4.030	NA	NA	0.043	NA	0.917
	MARS	0.074	15.560	NA	NA	0.623	NA	1.478
	LASSO	0.312	12.050	0.048	0.561	0.609	0.594	0.989
	FFSR-NH	0.347	3.240	0.019	0.042	0.061	0.396	0.953
	FFSR-NHS	0.354	3.250	0.019	0.042	0.061	0.398	0.954
	FFSR-SH	0.472	3.820	0.052	0.000	0.052	0.516	0.979
	FFSR-WH	0.426	3.570	0.028	0.025	0.052	0.466	0.972
	FFSR-NHA1	0.413	3.320	0.031	0.021	0.053	0.420	0.958
	FFSR-NHA2	0.413	3.320	0.031	0.021	0.053	0.420	0.958
	FFSR-WHA2	0.423	3.510	0.030	0.025	0.054	0.452	0.969
	Best No	0.499	3.240	0.005	0.008	0.013	0.438	0.946
	Best Strong	0.656	4.070	0.016	0.000	0.016	0.598	0.974
	<i>Average SE</i>	<i>0.024</i>	<i>0.159</i>	<i>0.007</i>	<i>0.007</i>	<i>0.011</i>	<i>0.015</i>	<i>0.015</i>

Appendix A. Simulation Results

Table A.4: Comparison of Methods for Normal Correlated, Model 1, $R^2 = 0.25$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.190	2.050	NA	NA	0.154	NA	0.577
	MARS	0.124	9.880	NA	NA	0.558	NA	2.000
	LASSO	0.277	4.530	0.076	0.366	0.443	0.203	0.527
	FFSR-NH	0.252	1.930	0.025	0.078	0.103	0.108	0.499
	FFSR-NHS	0.254	1.910	0.022	0.073	0.094	0.108	0.486
	FFSR-SH	0.295	2.230	0.071	0.005	0.076	0.165	0.554
	FFSR-WH	0.295	2.250	0.053	0.045	0.098	0.158	0.572
	FFSR-NHA1	0.284	2.290	0.058	0.050	0.108	0.157	0.593
	FFSR-NHA2	0.284	2.290	0.058	0.050	0.108	0.157	0.593
	FFSR-WHA2	0.294	2.350	0.059	0.042	0.101	0.170	0.591
	Best No	0.315	2.550	0.043	0.090	0.133	0.185	0.678
	Best Strong	0.381	4.120	0.153	0.048	0.201	0.338	0.817
	<i>Average SE</i>	<i>0.014</i>	<i>0.146</i>	<i>0.011</i>	<i>0.011</i>	<i>0.018</i>	<i>0.013</i>	<i>0.039</i>
500	CART	0.057	3.530	NA	NA	0.065	NA	0.640
	MARS	0.067	16.330	NA	NA	0.417	NA	1.429
	LASSO	0.090	6.510	0.019	0.253	0.271	0.582	0.641
	FFSR-NH	0.156	5.140	0.025	0.072	0.097	0.595	0.891
	FFSR-NHS	0.162	5.210	0.026	0.076	0.101	0.605	0.896
	FFSR-SH	0.266	6.590	0.058	0.047	0.105	0.800	0.972
	FFSR-WH	0.240	5.980	0.034	0.060	0.094	0.723	0.959
	FFSR-NHA1	0.273	6.360	0.050	0.053	0.103	0.773	0.994
	FFSR-NHA2	0.266	6.300	0.048	0.053	0.101	0.767	0.990
	FFSR-WHA2	0.290	6.350	0.043	0.054	0.097	0.777	0.988
	Best No	0.342	6.940	0.032	0.081	0.113	0.838	1.027
	Best Strong	0.490	7.260	0.043	0.035	0.078	0.920	1.015
	<i>Average SE</i>	<i>0.016</i>	<i>0.181</i>	<i>0.007</i>	<i>0.010</i>	<i>0.013</i>	<i>0.020</i>	<i>0.019</i>

Appendix A. Simulation Results

Table A.5: Comparison of Methods for Normal Correlated, Model 2, $R^2 = 0.25$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.182	1.870	NA	NA	0.136	NA	0.435
	MARS	0.120	9.920	NA	NA	0.492	NA	1.940
	LASSO	0.260	4.670	0.122	0.293	0.415	0.298	0.452
	FFSR-NH	0.221	1.830	0.072	0.025	0.097	0.118	0.396
	FFSR-NHS	0.223	1.840	0.072	0.019	0.090	0.122	0.397
	FFSR-SH	0.284	2.640	0.130	0.010	0.140	0.242	0.634
	FFSR-WH	0.262	2.340	0.112	0.023	0.134	0.194	0.569
	FFSR-NHA1	0.261	2.350	0.107	0.012	0.119	0.202	0.564
	FFSR-NHA2	0.262	2.340	0.104	0.012	0.116	0.202	0.562
	FFSR-WHA2	0.263	2.330	0.113	0.019	0.132	0.196	0.569
	Best No	0.308	3.070	0.103	0.045	0.147	0.298	0.697
	Best Strong	0.439	4.500	0.158	0.013	0.171	0.530	0.917
	<i>Average SE</i>	<i>0.013</i>	<i>0.144</i>	<i>0.017</i>	<i>0.008</i>	<i>0.019</i>	<i>0.017</i>	<i>0.037</i>
500	CART	0.052	3.850	NA	NA	0.109	NA	0.606
	MARS	0.064	17.600	NA	NA	0.383	NA	1.429
	LASSO	0.106	6.720	0.045	0.216	0.261	0.726	0.692
	FFSR-NH	0.251	4.900	0.014	0.027	0.041	0.732	0.926
	FFSR-NHS	0.252	4.890	0.014	0.026	0.040	0.732	0.925
	FFSR-SH	0.400	5.850	0.040	0.017	0.057	0.890	0.995
	FFSR-WH	0.294	5.260	0.020	0.029	0.049	0.792	0.955
	FFSR-NHA1	0.374	5.740	0.038	0.018	0.056	0.874	0.990
	FFSR-NHA2	0.361	5.670	0.035	0.018	0.053	0.864	0.985
	FFSR-WHA2	0.357	5.650	0.032	0.022	0.054	0.858	0.983
	Best No	0.455	5.940	0.016	0.030	0.046	0.916	0.993
	Best Strong	0.682	6.260	0.039	0.008	0.047	0.980	1.010
	<i>Average SE</i>	<i>0.019</i>	<i>0.154</i>	<i>0.007</i>	<i>0.006</i>	<i>0.010</i>	<i>0.018</i>	<i>0.015</i>

Appendix A. Simulation Results

Table A.6: Comparison of Methods for Normal Correlated, Model 3, $R^2 = 0.25$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.217	2.140	NA	NA	0.149	NA	0.720
	MARS	0.117	10.280	NA	NA	0.643	NA	2.183
	LASSO	0.428	4.490	0.121	0.275	0.396	0.252	0.709
	FFSR-NH	0.475	2.030	0.032	0.029	0.061	0.172	0.736
	FFSR-NHS	0.486	2.010	0.032	0.024	0.056	0.172	0.731
	FFSR-SH	0.601	2.340	0.082	0.003	0.084	0.218	0.815
	FFSR-WH	0.588	2.280	0.059	0.028	0.087	0.206	0.816
	FFSR-NHA1	0.532	2.230	0.066	0.022	0.089	0.194	0.796
	FFSR-NHA2	0.532	2.230	0.066	0.022	0.089	0.194	0.796
	FFSR-WHA2	0.577	2.270	0.063	0.025	0.088	0.204	0.811
	Best No	0.702	2.330	0.051	0.023	0.074	0.220	0.846
	Best Strong	0.868	2.690	0.090	0.007	0.097	0.270	0.869
	<i>Average SE</i>	<i>0.031</i>	<i>0.111</i>	<i>0.013</i>	<i>0.008</i>	<i>0.017</i>	<i>0.010</i>	<i>0.036</i>
500	CART	0.099	3.960	NA	NA	0.081	NA	0.930
	MARS	0.082	14.110	NA	NA	0.589	NA	1.437
	LASSO	0.248	5.630	0.086	0.277	0.363	0.432	0.880
	FFSR-NH	0.335	3.180	0.027	0.079	0.106	0.352	0.953
	FFSR-NHS	0.340	3.160	0.027	0.074	0.101	0.352	0.952
	FFSR-SH	0.424	3.670	0.084	0.000	0.084	0.468	0.967
	FFSR-WH	0.385	3.510	0.037	0.099	0.136	0.388	0.972
	FFSR-NHA1	0.378	3.260	0.048	0.052	0.100	0.372	0.957
	FFSR-NHA2	0.378	3.260	0.048	0.052	0.100	0.372	0.957
	FFSR-WHA2	0.382	3.470	0.047	0.083	0.130	0.386	0.969
	Best No	0.420	3.310	0.041	0.048	0.089	0.392	0.954
	Best Strong	0.581	4.200	0.065	0.013	0.077	0.560	0.978
	<i>Average SE</i>	<i>0.022</i>	<i>0.134</i>	<i>0.010</i>	<i>0.010</i>	<i>0.015</i>	<i>0.014</i>	<i>0.017</i>

Appendix A. Simulation Results

Table A.7: Comparison of Methods for Chi Square Uncorrelated, Model 1, $R^2 = 0.25$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.179	1.920	NA	NA	0.066	NA	0.428
	MARS	0.118	10.400	NA	NA	0.497	NA	2.037
	LASSO	0.278	6.750	0.027	0.475	0.502	0.325	0.665
	FFSR-NH	0.236	2.040	0.008	0.049	0.056	0.142	0.501
	FFSR-NHS	0.232	1.920	0.008	0.035	0.043	0.132	0.431
	FFSR-SH	0.316	3.020	0.037	0.007	0.044	0.302	0.709
	FFSR-WH	0.286	2.600	0.018	0.021	0.039	0.242	0.633
	FFSR-NHA1	0.283	2.610	0.019	0.029	0.047	0.240	0.663
	FFSR-NHA2	0.283	2.610	0.019	0.029	0.047	0.240	0.663
	FFSR-WHA2	0.294	2.760	0.018	0.027	0.045	0.263	0.680
	Best No	0.300	2.720	0.003	0.057	0.060	0.248	0.671
	Best Strong	0.464	4.580	0.060	0.019	0.078	0.515	0.938
	<i>Average SE</i>	<i>0.017</i>	<i>0.162</i>	<i>0.008</i>	<i>0.010</i>	<i>0.014</i>	<i>0.016</i>	<i>0.042</i>
500	CART	0.055	4.340	NA	NA	0.013	NA	0.714
	MARS	0.058	17.460	NA	NA	0.437	NA	1.454
	LASSO	0.152	11.760	0.032	0.362	0.394	0.868	0.878
	FFSR-NH	0.268	5.860	0.008	0.046	0.054	0.750	0.989
	FFSR-NHS	0.280	5.900	0.008	0.043	0.051	0.758	0.989
	FFSR-SH	0.405	6.520	0.037	0.022	0.059	0.845	1.012
	FFSR-WH	0.318	6.190	0.017	0.042	0.059	0.793	1.002
	FFSR-NHA1	0.302	6.010	0.033	0.025	0.058	0.768	0.994
	FFSR-NHA2	0.304	5.990	0.030	0.025	0.055	0.768	0.993
	FFSR-WHA2	0.321	6.110	0.026	0.030	0.056	0.785	0.997
	Best No	0.372	6.190	0.003	0.039	0.041	0.812	1.000
	Best Strong	0.582	6.350	0.012	0.001	0.013	0.875	1.000
	<i>Average SE</i>	<i>0.018</i>	<i>0.167</i>	<i>0.005</i>	<i>0.007</i>	<i>0.009</i>	<i>0.013</i>	<i>0.012</i>

Appendix A. Simulation Results

Table A.8: Comparison of Methods for Chi Square Uncorrelated, Model 2, $R^2 = 0.25$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.181	2.170	NA	NA	0.079	NA	0.480
	MARS	0.117	10.830	NA	NA	0.461	NA	2.094
	LASSO	0.293	7.330	0.028	0.489	0.515	0.402	0.661
	FFSR-NH	0.236	1.990	0.000	0.061	0.061	0.162	0.441
	FFSR-NHS	0.231	2.010	0.000	0.065	0.065	0.162	0.446
	FFSR-SH	0.334	3.200	0.030	0.014	0.044	0.398	0.735
	FFSR-WH	0.291	2.620	0.008	0.028	0.036	0.294	0.615
	FFSR-NHA1	0.291	2.790	0.008	0.041	0.049	0.324	0.665
	FFSR-NHA2	0.291	2.730	0.008	0.039	0.047	0.314	0.654
	FFSR-WHA2	0.294	2.820	0.010	0.030	0.040	0.326	0.662
	Best No	0.300	2.710	0.002	0.067	0.069	0.286	0.638
	Best Strong	0.493	4.240	0.030	0.014	0.044	0.596	0.895
	<i>Average SE</i>	<i>0.018</i>	<i>0.164</i>	<i>0.004</i>	<i>0.010</i>	<i>0.013</i>	<i>0.021</i>	<i>0.037</i>
500	CART	0.050	3.920	NA	NA	0.017	NA	0.610
	MARS	0.062	17.380	NA	NA	0.371	NA	1.478
	LASSO	0.180	11.610	0.031	0.387	0.418	0.970	0.898
	FFSR-NH	0.367	5.920	0.004	0.048	0.052	0.910	1.006
	FFSR-NHS	0.356	5.980	0.004	0.053	0.057	0.910	1.010
	FFSR-SH	0.559	6.110	0.023	0.019	0.042	0.962	1.009
	FFSR-WH	0.436	6.080	0.008	0.047	0.056	0.938	1.011
	FFSR-NHA1	0.508	6.150	0.026	0.024	0.049	0.960	1.015
	FFSR-NHA2	0.510	6.110	0.022	0.023	0.045	0.958	1.014
	FFSR-WHA2	0.502	6.060	0.016	0.027	0.043	0.952	1.010
	Best No	0.550	5.890	0.005	0.017	0.022	0.948	1.002
	Best Strong	0.843	6.050	0.008	0.004	0.013	0.992	1.004
	<i>Average SE</i>	<i>0.024</i>	<i>0.150</i>	<i>0.004</i>	<i>0.007</i>	<i>0.008</i>	<i>0.010</i>	<i>0.011</i>

Appendix A. Simulation Results

Table A.9: Comparison of Methods for Chi Square Uncorrelated, Model 3, $R^2 = 0.25$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.219	2.470	NA	NA	0.088	NA	0.800
	MARS	0.132	9.340	NA	NA	0.583	NA	1.975
	LASSO	0.470	6.920	0.031	0.496	0.527	0.334	0.866
	FFSR-SH	0.756	2.500	0.028	0.000	0.028	0.282	0.846
	FFSR-NH	0.456	2.240	0.010	0.048	0.058	0.206	0.771
	FFSR-NHS	0.452	2.190	0.013	0.047	0.060	0.196	0.743
	FFSR-WH	0.666	2.380	0.023	0.011	0.033	0.254	0.818
	FFSR-NHA1	0.636	2.380	0.021	0.019	0.040	0.250	0.833
	FFSR-NHA2	0.636	2.380	0.021	0.019	0.040	0.250	0.833
	FFSR-WHA2	0.651	2.380	0.024	0.009	0.033	0.254	0.820
	Best No	0.693	2.330	0.002	0.012	0.014	0.254	0.809
	Best Strong	1.058	2.820	0.017	0.003	0.020	0.344	0.870
	<i>Average SE</i>	<i>0.039</i>	<i>0.126</i>	<i>0.007</i>	<i>0.007</i>	<i>0.012</i>	<i>0.012</i>	<i>0.034</i>
500	CART	0.089	4.250	NA	NA	0.038	NA	0.932
	MARS	0.062	15.920	NA	NA	0.614	NA	1.470
	LASSO	0.259	10.510	0.034	0.424	0.457	0.646	0.962
	FFSR-NH	0.374	3.280	0.003	0.043	0.045	0.416	0.958
	FFSR-NHS	0.395	3.260	0.003	0.033	0.036	0.420	0.955
	FFSR-SH	0.465	3.740	0.030	0.003	0.033	0.510	0.973
	FFSR-WH	0.440	3.460	0.006	0.028	0.034	0.460	0.965
	FFSR-NHA1	0.423	3.230	0.013	0.019	0.032	0.418	0.953
	FFSR-NHA2	0.423	3.230	0.013	0.019	0.032	0.418	0.953
	FFSR-WHA2	0.434	3.420	0.012	0.022	0.034	0.448	0.962
	Best No	0.467	3.310	0.000	0.011	0.011	0.452	0.957
	Best Strong	0.631	4.060	0.009	0.000	0.009	0.600	0.975
	<i>Average SE</i>	<i>0.021</i>	<i>0.147</i>	<i>0.005</i>	<i>0.007</i>	<i>0.010</i>	<i>0.013</i>	<i>0.013</i>

Appendix A. Simulation Results

Table A.10: Comparison of Methods for Chi Square Correlated, Model 1, $R^2 = 0.25$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.174	2.000	NA	NA	0.110	NA	0.478
	MARS	0.121	9.570	NA	NA	0.521	NA	1.891
	LASSO	0.238	4.450	0.036	0.437	0.473	0.205	0.463
	FFSR-NH	0.219	1.950	0.014	0.108	0.123	0.100	0.479
	FFSR-NHS	0.218	1.910	0.005	0.075	0.080	0.112	0.439
	FFSR-SH	0.266	2.610	0.083	0.021	0.104	0.200	0.604
	FFSR-WH	0.257	2.460	0.056	0.047	0.103	0.183	0.598
	FFSR-NHA1	0.240	2.290	0.061	0.056	0.118	0.150	0.574
	FFSR-NHA2	0.238	2.240	0.050	0.059	0.109	0.148	0.562
	FFSR-WHA2	0.257	2.510	0.058	0.047	0.105	0.187	0.613
	Best No	0.282	2.810	0.025	0.149	0.174	0.200	0.704
	Best Strong	0.368	5.040	0.158	0.054	0.212	0.450	0.928
	<i>Average SE</i>	<i>0.012</i>	<i>0.159</i>	<i>0.012</i>	<i>0.013</i>	<i>0.019</i>	<i>0.013</i>	<i>0.040</i>
500	CART	0.059	3.770	NA	NA	0.054	NA	0.653
	MARS	0.070	16.550	NA	NA	0.405	NA	1.406
	LASSO	0.084	5.990	0.005	0.259	0.264	0.570	0.596
	FFSR-NH	0.166	5.560	0.020	0.115	0.134	0.630	0.929
	FFSR-NHS	0.175	5.730	0.018	0.124	0.142	0.647	0.940
	FFSR-SH	0.388	6.750	0.037	0.030	0.067	0.872	1.006
	FFSR-WH	0.264	6.310	0.029	0.075	0.104	0.767	0.986
	FFSR-NHA1	0.247	6.270	0.042	0.072	0.114	0.753	0.991
	FFSR-NHA2	0.249	6.290	0.041	0.072	0.113	0.757	0.993
	FFSR-WHA2	0.289	6.480	0.035	0.069	0.104	0.790	0.999
	Best No	0.322	7.320	0.028	0.122	0.151	0.845	1.040
	Best Strong	0.558	6.970	0.033	0.017	0.050	0.928	1.013
	<i>Average SE</i>	<i>0.015</i>	<i>0.173</i>	<i>0.006</i>	<i>0.010</i>	<i>0.012</i>	<i>0.020</i>	<i>0.015</i>

Appendix A. Simulation Results

Table A.11: Comparison of Methods for Chi Square Correlated, Model 2, $R^2 = 0.25$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.189	1.810	NA	NA	0.172	NA	0.415
	MARS	0.129	9.720	NA	NA	0.513	NA	1.902
	LASSO	0.270	4.290	0.116	0.272	0.388	0.302	0.403
	FFSR-NH	0.237	1.850	0.043	0.048	0.091	0.128	0.395
	FFSR-NHS	0.235	1.840	0.043	0.035	0.078	0.126	0.380
	FFSR-SH	0.328	2.960	0.123	0.009	0.132	0.306	0.688
	FFSR-WH	0.303	2.490	0.103	0.014	0.117	0.238	0.596
	FFSR-NHA1	0.288	2.520	0.088	0.011	0.099	0.248	0.579
	FFSR-NHA2	0.288	2.510	0.088	0.011	0.099	0.246	0.578
	FFSR-WHA2	0.305	2.680	0.104	0.015	0.120	0.266	0.631
	Best No	0.336	3.160	0.059	0.074	0.133	0.324	0.732
	Best Strong	0.489	4.420	0.107	0.017	0.124	0.554	0.898
	<i>Average SE</i>	<i>0.018</i>	<i>0.146</i>	<i>0.016</i>	<i>0.009</i>	<i>0.019</i>	<i>0.019</i>	<i>0.035</i>
500	CART	0.052	4.050	NA	NA	0.111	NA	0.646
	MARS	0.062	17.990	NA	NA	0.384	NA	1.451
	LASSO	0.106	6.640	0.047	0.206	0.253	0.658	0.655
	FFSR-NH	0.270	5.160	0.022	0.034	0.055	0.770	0.951
	FFSR-NHS	0.273	5.070	0.020	0.020	0.040	0.77	0.943
	FFSR-SH	0.462	5.910	0.034	0.018	0.052	0.912	0.994
	FFSR-WH	0.327	5.520	0.025	0.034	0.059	0.832	0.973
	FFSR-NHA1	0.444	5.900	0.038	0.013	0.051	0.910	0.995
	FFSR-NHA2	0.429	5.840	0.038	0.013	0.051	0.902	0.991
	FFSR-WHA2	0.395	5.830	0.037	0.025	0.061	0.886	0.989
	Best No	0.494	6.350	0.020	0.058	0.078	0.954	1.017
	Best Strong	0.790	6.160	0.024	0.006	0.031	0.986	1.006
	<i>Average SE</i>	<i>0.023</i>	<i>0.152</i>	<i>0.007</i>	<i>0.006</i>	<i>0.010</i>	<i>0.019</i>	<i>0.015</i>

Appendix A. Simulation Results

Table A.12: Comparison of Methods for Chi Square Correlated, Model 3, $R^2 = 0.25$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.206	2.330	NA	NA	0.203	NA	0.760
	MARS	0.131	9.230	NA	NA	0.624	NA	1.935
	LASSO	0.385	4.310	0.096	0.287	0.383	0.258	0.681
	FFSR-NH	0.426	2.070	0.038	0.046	0.083	0.168	0.722
	FFSR-NHS	0.448	2.090	0.041	0.030	0.071	0.178	0.730
	FFSR-SH	0.573	2.520	0.111	0.003	0.114	0.234	0.861
	FFSR-WH	0.556	2.450	0.078	0.037	0.114	0.222	0.851
	FFSR-NHA1	0.530	2.410	0.079	0.037	0.116	0.216	0.849
	FFSR-NHA2	0.530	2.400	0.081	0.033	0.114	0.216	0.846
	FFSR-WHA2	0.560	2.460	0.080	0.035	0.115	0.216	0.852
	Best No	0.609	2.520	0.043	0.055	0.088	0.234	0.855
	Best Strong	0.753	2.810	0.129	0.007	0.136	0.272	0.875
	<i>Average SE</i>	<i>0.027</i>	<i>0.118</i>	<i>0.014</i>	<i>0.009</i>	<i>0.019</i>	<i>0.012</i>	<i>0.033</i>
500	CART	0.080	4.160	NA	NA	0.100	NA	0.907
	MARS	0.070	14.050	NA	NA	0.600	NA	1.397
	LASSO	0.184	5.460	0.050	0.273	0.323	0.482	0.861
	FFSR-NH	0.327	3.360	0.011	0.093	0.104	0.390	0.969
	FFSR-NHS	0.347	3.430	0.018	0.082	0.100	0.400	0.970
	FFSR-SH	0.435	3.620	0.051	0.000	0.051	0.482	0.969
	FFSR-WH	0.375	3.650	0.032	0.091	0.123	0.420	0.980
	FFSR-NHA1	0.359	3.350	0.029	0.060	0.089	0.400	0.968
	FFSR-NHA2	0.359	3.350	0.029	0.060	0.089	0.400	0.968
	FFSR-WHA2	0.373	3.640	0.034	0.088	0.122	0.420	0.979
	Best No	0.417	3.380	0.011	0.060	0.071	0.412	0.959
	Best Strong	0.590	4.360	0.028	0.011	0.040	0.614	0.981
	<i>Average SE</i>	<i>0.035</i>	<i>0.133</i>	<i>0.007</i>	<i>0.011</i>	<i>0.013</i>	<i>0.012</i>	<i>0.012</i>

Appendix A. Simulation Results

Table A.13: Comparison of Methods for Normal Uncorrelated, Model 1, $R^2 = 0.5$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.082	3.410	NA	NA	0.049	NA	0.649
	MARS	0.095	11.930	NA	NA	0.380	NA	1.373
	LASSO	0.153	8.070	0.025	0.357	0.382	0.602	0.734
	FFSR-NH	0.168	4.050	0.004	0.074	0.078	0.440	0.840
	FFSR-NHS	0.177	4.040	0.000	0.072	0.072	0.443	0.842
	FFSR-SH	0.279	5.020	0.036	0.023	0.059	0.613	0.918
	FFSR-WH	0.233	4.600	0.012	0.039	0.052	0.553	0.905
	FFSR-NHA1	0.232	4.900	0.029	0.054	0.083	0.573	0.946
	FFSR-NHA2	0.231	4.890	0.029	0.054	0.083	0.572	0.945
	FFSR-WHA2	0.254	4.880	0.023	0.040	0.063	0.588	0.934
	Best No	0.255	5.120	0.001	0.066	0.068	0.608	0.955
	Best Strong	0.446	5.890	0.024	0.016	0.039	0.762	0.981
	<i>Average SE</i>	<i>0.020</i>	<i>0.177</i>	<i>0.006</i>	<i>0.009</i>	<i>0.012</i>	<i>0.018</i>	<i>0.020</i>
500	CART	0.027	9.130	NA	NA	0.026	NA	0.843
	MARS	0.060	18.850	NA	NA	0.378	NA	1.156
	LASSO	0.166	13.750	0.033	0.404	0.437	0.950	0.971
	FFSR-NH	0.351	6.800	0.005	0.049	0.054	0.897	1.005
	FFSR-NHS	0.360	6.780	0.005	0.046	0.051	0.898	1.005
	FFSR-SH	0.465	7.100	0.028	0.023	0.051	0.945	1.007
	FFSR-WH	0.386	6.890	0.011	0.044	0.055	0.912	1.005
	FFSR-NHA1	0.352	6.740	0.026	0.026	0.052	0.888	1.003
	FFSR-NHA2	0.355	6.720	0.024	0.026	0.049	0.888	1.003
	FFSR-WHA2	0.387	6.740	0.016	0.029	0.045	0.900	1.003
	Best No	0.512	6.570	0.001	0.008	0.009	0.917	0.999
	Best Strong	0.642	6.930	0.014	0.003	0.017	0.965	1.003
	<i>Average SE</i>	<i>0.020</i>	<i>0.174</i>	<i>0.004</i>	<i>0.006</i>	<i>0.008</i>	<i>0.010</i>	<i>0.005</i>

Appendix A. Simulation Results

Table A.14: Comparison of Methods for Normal Uncorrelated, Model 2, $R^2 = 0.5$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.081	3.340	NA	NA	0.048	NA	0.603
	MARS	0.094	13.050	NA	NA	0.371	NA	1.405
	LASSO	0.163	8.580	0.022	0.356	0.378	0.804	0.754
	FFSR-NH	0.194	4.280	0.005	0.067	0.072	0.582	0.867
	FFSR-NHS	0.204	4.210	0.005	0.040	0.045	0.594	0.857
	FFSR-SH	0.442	5.640	0.028	0.017	0.046	0.866	0.990
	FFSR-WH	0.281	4.850	0.008	0.036	0.044	0.716	0.927
	FFSR-NHA1	0.371	5.500	0.022	0.031	0.053	0.832	0.994
	FFSR-NHA2	0.366	5.460	0.020	0.032	0.052	0.826	0.991
	FFSR-WHA2	0.359	5.260	0.018	0.021	0.040	0.802	0.965
	Best No	0.320	5.170	0.001	0.049	0.050	0.768	0.961
	Best Strong	0.687	6.060	0.022	0.013	0.036	0.960	1.009
	<i>Average SE</i>	<i>0.018</i>	<i>0.166</i>	<i>0.004</i>	<i>0.008</i>	<i>0.010</i>	<i>0.018</i>	<i>0.016</i>
500	CART	0.024	8.510	NA	NA	0.021	NA	0.755
	MARS	0.058	19.990	NA	NA	0.353	NA	1.174
	LASSO	0.201	13.170	0.033	0.468	0.501	1.000	0.979
	FFSR-NH	0.551	6.400	0.003	0.047	0.050	1.000	1.010
	FFSR-NHS	0.539	6.430	0.002	0.052	0.054	1.000	1.010
	FFSR-SH	0.626	6.430	0.034	0.024	0.058	1.000	1.007
	FFSR-WH	0.569	6.410	0.007	0.044	0.052	1.000	1.009
	FFSR-NHA1	0.543	6.480	0.033	0.030	0.063	1.000	1.010
	FFSR-NHA2	0.546	6.470	0.032	0.030	0.062	1.000	1.010
	FFSR-WHA2	0.583	6.400	0.023	0.029	0.051	1.000	1.009
	Best No	1.000	6.000	0.000	0.000	0.000	1.000	1.000
	Best Strong	1.000	6.000	0.000	0.000	0.000	1.000	1.000
	<i>Average SE</i>	<i>0.025</i>	<i>0.167</i>	<i>0.004</i>	<i>0.007</i>	<i>0.008</i>	<i>0.005</i>	<i>0.005</i>

Appendix A. Simulation Results

Table A.15: Comparison of Methods for Normal Uncorrelated, Model 3, $R^2 = 0.5$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.150	3.330	NA	NA	0.049	NA	0.883
	MARS	0.123	10.040	NA	NA	0.556	NA	1.371
	LASSO	0.322	7.610	0.052	0.474	0.526	0.434	0.906
	FFSR-NH	0.383	2.600	0.005	0.034	0.039	0.292	0.901
	FFSR-NHS	0.379	2.650	0.005	0.041	0.046	0.294	0.904
	FFSR-SH	0.456	3.120	0.050	0.003	0.053	0.374	0.940
	FFSR-WH	0.419	2.960	0.027	0.043	0.070	0.336	0.933
	FFSR-NHA1	0.423	2.820	0.027	0.024	0.052	0.326	0.928
	FFSR-NHA2	0.420	2.810	0.027	0.024	0.052	0.324	0.928
	FFSR-WHA2	0.420	2.980	0.032	0.037	0.069	0.338	0.934
	Best No	0.497	2.690	0.000	0.010	0.010	0.330	0.901
	Best Strong	0.685	3.360	0.017	0.000	0.017	0.456	0.939
	<i>Average SE</i>	<i>0.026</i>	<i>0.145</i>	<i>0.007</i>	<i>0.008</i>	<i>0.012</i>	<i>0.014</i>	<i>0.020</i>
500	CART	0.064	7.180	NA	NA	0.038	NA	0.984
	MARS	0.072	16.300	NA	NA	0.574	NA	1.164
	LASSO	0.231	12.310	0.041	0.493	0.535	0.808	0.983
	FFSR-NH	0.286	4.420	0.014	0.049	0.063	0.616	0.988
	FFSR-NHS	0.290	4.390	0.013	0.044	0.057	0.616	0.987
	FFSR-SH	0.424	5.110	0.047	0.000	0.047	0.766	0.998
	FFSR-WH	0.345	4.790	0.024	0.029	0.053	0.694	0.994
	FFSR-NHA1	0.280	4.190	0.038	0.018	0.056	0.580	0.982
	FFSR-NHA2	0.280	4.180	0.036	0.018	0.054	0.580	0.982
	FFSR-WHA2	0.328	4.610	0.028	0.022	0.049	0.666	0.991
	Best No	0.395	4.640	0.003	0.023	0.026	0.696	0.990
	Best Strong	0.634	5.550	0.030	0.002	0.031	0.868	0.999
	<i>Average SE</i>	<i>0.019</i>	<i>0.181</i>	<i>0.007</i>	<i>0.007</i>	<i>0.010</i>	<i>0.019</i>	<i>0.004</i>

Appendix A. Simulation Results

Table A.16: Comparison of Methods for Normal Correlated, Model 1, $R^2 = 0.5$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.090	2.880	NA	NA	0.094	NA	0.615
	MARS	0.109	11.190	NA	NA	0.421	NA	1.330
	LASSO	0.129	5.140	0.033	0.366	0.399	0.342	0.588
	FFSR-NH	0.147	2.960	0.025	0.102	0.127	0.252	0.715
	FFSR-NHS	0.148	2.970	0.022	0.101	0.123	0.257	0.717
	FFSR-SH	0.166	3.850	0.093	0.035	0.128	0.373	0.754
	FFSR-WH	0.169	3.410	0.056	0.068	0.124	0.320	0.762
	FFSR-NHA1	0.166	3.580	0.063	0.078	0.141	0.332	0.787
	FFSR-NHA2	0.166	3.570	0.060	0.078	0.137	0.332	0.786
	FFSR-WHA2	0.176	3.730	0.063	0.070	0.133	0.320	0.796
	Best No	0.214	6.480	0.059	0.219	0.279	0.552	1.014
	Best Strong	0.357	7.420	0.141	0.092	0.232	0.742	1.037
	<i>Average SE</i>	<i>0.009</i>	<i>0.179</i>	<i>0.011</i>	<i>0.012</i>	<i>0.017</i>	<i>0.019</i>	<i>0.024</i>
500	CART	0.027	7.570	NA	NA	0.070	NA	0.781
	MARS	0.065	18.430	NA	NA	0.352	NA	1.148
	LASSO	0.052	8.300	0.011	0.217	0.228	0.848	0.790
	FFSR-NH	0.413	7.040	0.004	0.060	0.064	0.927	1.006
	FFSR-NHS	0.406	7.090	0.005	0.060	0.065	0.930	1.007
	FFSR-SH	0.525	7.450	0.031	0.031	0.062	0.988	1.009
	FFSR-WH	0.440	7.220	0.011	0.063	0.074	0.938	1.009
	FFSR-NHA1	0.388	7.160	0.028	0.047	0.075	0.928	1.007
	FFSR-NHA2	0.388	7.160	0.028	0.047	0.075	0.928	1.007
	FFSR-WHA2	0.422	7.240	0.027	0.052	0.079	0.935	1.009
	Best No	0.547	6.960	0.003	0.045	0.048	0.937	1.003
	Best Strong	0.780	7.030	0.004	0.005	0.009	0.993	1.003
	<i>Average SE</i>	<i>0.021</i>	<i>0.158</i>	<i>0.004</i>	<i>0.007</i>	<i>0.009</i>	<i>0.013</i>	<i>0.007</i>

Appendix A. Simulation Results

Table A.17: Comparison of Methods for Normal Correlated, Model 2, $R^2 = 0.5$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.083	3.140	NA	NA	0.165	NA	0.585
	MARS	0.106	12.050	NA	NA	0.364	NA	1.361
	LASSO	0.122	5.340	0.066	0.213	0.278	0.508	0.549
	FFSR-NH	0.179	3.650	0.046	0.034	0.080	0.460	0.777
	FFSR-NHS	0.178	3.620	0.049	0.029	0.077	0.460	0.773
	FFSR-SH	0.312	5.110	0.087	0.017	0.104	0.706	0.949
	FFSR-WH	0.219	4.190	0.061	0.027	0.089	0.554	0.853
	FFSR-NHA1	0.274	4.830	0.076	0.018	0.094	0.668	0.924
	FFSR-NHA2	0.269	4.730	0.067	0.018	0.085	0.660	0.914
	FFSR-WHA2	0.249	4.550	0.063	0.022	0.085	0.626	0.889
	Best No	0.332	5.590	0.039	0.061	0.101	0.752	0.980
	Best Strong	0.496	6.200	0.094	0.023	0.116	0.880	1.011
	<i>Average SE</i>	<i>0.014</i>	<i>0.167</i>	<i>0.011</i>	<i>0.007</i>	<i>0.015</i>	<i>0.021</i>	<i>0.021</i>
500	CART	0.024	6.730	NA	NA	0.068	NA	0.680
	MARS	0.066	18.620	NA	NA	0.339	NA	1.158
	LASSO	0.071	8.090	0.029	0.214	0.243	0.944	0.841
	FFSR-NH	0.606	6.180	0.003	0.028	0.031	0.992	1.004
	FFSR-NHS	0.613	6.170	0.003	0.026	0.029	0.992	1.003
	FFSR-SH	0.706	6.280	0.018	0.018	0.037	1.000	1.005
	FFSR-WH	0.694	6.250	0.007	0.027	0.034	1.000	1.006
	FFSR-NHA1	0.686	6.270	0.021	0.016	0.037	1.000	1.006
	FFSR-NHA2	0.686	6.260	0.019	0.016	0.035	1.000	1.006
	FFSR-WHA2	0.710	6.240	0.015	0.018	0.033	1.000	1.005
	Best No	0.968	6.020	0.001	0.001	0.003	1.000	1.000
	Best Strong	0.984	6.010	0.000	0.001	0.001	1.000	1.000
	<i>Average SE</i>	<i>0.025</i>	<i>0.126</i>	<i>0.004</i>	<i>0.005</i>	<i>0.007</i>	<i>0.008</i>	<i>0.006</i>

Appendix A. Simulation Results

Table A.18: Comparison of Methods for Normal Correlated, Model 3, $R^2 = 0.5$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.132	3.710	NA	NA	0.155	NA	0.947
	MARS	0.118	10.150	NA	NA	0.541	NA	1.354
	LASSO	0.267	5.130	0.108	0.274	0.372	0.362	0.823
	FFSR-NH	0.375	2.690	0.022	0.061	0.083	0.280	0.915
	FFSR-NHS	0.374	2.660	0.022	0.049	0.071	0.280	0.910
	FFSR-SH	0.530	3.140	0.065	0.003	0.068	0.382	0.947
	FFSR-WH	0.457	3.030	0.046	0.049	0.095	0.336	0.944
	FFSR-NHA1	0.455	2.900	0.050	0.038	0.088	0.320	0.937
	FFSR-NHA2	0.451	2.880	0.047	0.038	0.085	0.318	0.935
	FFSR-WHA2	0.458	3.030	0.051	0.042	0.093	0.338	0.945
	Best No	0.531	3.050	0.042	0.060	0.102	0.360	0.937
	Best Strong	0.685	3.680	0.068	0.011	0.079	0.468	0.961
	<i>Average SE</i>	<i>0.026</i>	<i>0.128</i>	<i>0.010</i>	<i>0.010</i>	<i>0.015</i>	<i>0.014</i>	<i>0.019</i>
500	CART	0.061	6.710	NA	NA	0.056	NA	0.970
	MARS	0.075	15.640	NA	NA	0.569	NA	1.152
	LASSO	0.157	6.620	0.056	0.270	0.326	0.612	0.932
	FFSR-NH	0.274	4.360	0.002	0.146	0.148	0.534	0.990
	FFSR-NHS	0.271	4.370	0.002	0.146	0.148	0.534	0.990
	FFSR-SH	0.421	5.090	0.048	0.003	0.051	0.756	0.996
	FFSR-WH	0.322	4.680	0.005	0.152	0.158	0.580	0.995
	FFSR-NHA1	0.266	4.280	0.025	0.119	0.145	0.520	0.987
	FFSR-NHA2	0.266	4.280	0.025	0.119	0.145	0.520	0.987
	FFSR-WHA2	0.307	4.720	0.027	0.146	0.173	0.572	0.996
	Best No	0.343	4.700	0.000	0.137	0.137	0.596	0.991
	Best Strong	0.628	5.720	0.037	0.007	0.044	0.882	1.001
	<i>Average SE</i>	<i>0.016</i>	<i>0.141</i>	<i>0.006</i>	<i>0.011</i>	<i>0.012</i>	<i>0.015</i>	<i>0.004</i>

Appendix A. Simulation Results

Table A.19: Comparison of Methods for Chi Square Uncorrelated, Model 1, $R^2 = 0.5$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.084	2.900	NA	NA	0.047	NA	0.530
	MARS	0.111	12.110	NA	NA	0.370	NA	1.364
	LASSO	0.152	7.470	0.017	0.317	0.334	0.593	0.693
	FFSR-NH	0.189	4.170	0.009	0.079	0.088	0.453	0.859
	FFSR-NHS	0.194	4.300	0.010	0.075	0.085	0.472	0.861
	FFSR-SH	0.357	5.220	0.030	0.011	0.041	0.662	0.955
	FFSR-WH	0.255	4.920	0.013	0.059	0.072	0.578	0.932
	FFSR-NHA1	0.285	4.910	0.026	0.038	0.064	0.590	0.951
	FFSR-NHA2	0.281	4.870	0.027	0.034	0.061	0.588	0.945
	FFSR-WHA2	0.281	5.000	0.024	0.044	0.068	0.598	0.949
	Best No	0.283	5.080	0.009	0.067	0.077	0.598	0.964
	Best Strong	0.521	6.000	0.023	0.017	0.039	0.782	0.993
	<i>Average SE</i>	<i>0.017</i>	<i>0.177</i>	<i>0.006</i>	<i>0.009</i>	<i>0.012</i>	<i>0.018</i>	<i>0.020</i>
500	CART	0.027	8.720	NA	NA	0.023	NA	0.833
	MARS	0.057	19.070	NA	NA	0.389	NA	1.158
	LASSO	0.128	12.690	0.030	0.348	0.378	0.978	0.948
	FFSR-NH	0.381	6.800	0.009	0.034	0.043	0.912	1.004
	FFSR-NHS	0.372	6.830	0.009	0.038	0.046	0.912	1.005
	FFSR-SH	0.433	7.190	0.038	0.024	0.062	0.947	1.007
	FFSR-WH	0.367	7.020	0.015	0.046	0.061	0.922	1.007
	FFSR-NHA1	0.371	6.840	0.029	0.021	0.050	0.908	1.004
	FFSR-NHA2	0.372	6.830	0.028	0.021	0.050	0.908	1.004
	FFSR-WHA2	0.378	6.920	0.028	0.026	0.054	0.917	1.005
	Best No	0.526	6.600	0.000	0.006	0.006	0.925	0.998
	Best Strong	0.668	6.870	0.005	0.002	0.007	0.968	1.001
	<i>Average SE</i>	<i>0.022</i>	<i>0.162</i>	<i>0.004</i>	<i>0.006</i>	<i>0.008</i>	<i>0.009</i>	<i>0.005</i>

Appendix A. Simulation Results

Table A.20: Comparison of Methods for Chi Square Uncorrelated, Model 2, $R^2 = 0.5$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.087	3.390	NA	NA	0.047	NA	0.619
	MARS	0.109	12.170	NA	NA	0.358	NA	1.386
	LASSO	0.167	8.470	0.023	0.337	0.359	0.708	0.728
	FFSR-NH	0.214	4.280	0.003	0.073	0.076	0.578	0.874
	FFSR-NHS	0.221	4.240	0.003	0.042	0.045	0.596	0.862
	FFSR-SH	0.382	5.440	0.035	0.025	0.060	0.810	0.983
	FFSR-WH	0.292	4.850	0.007	0.044	0.051	0.712	0.941
	FFSR-NHA1	0.332	5.460	0.024	0.052	0.076	0.794	1.001
	FFSR-NHA2	0.328	5.470	0.025	0.053	0.078	0.792	1.002
	FFSR-WHA2	0.339	5.120	0.017	0.029	0.046	0.768	0.965
	Best No	0.302	5.230	0.004	0.087	0.090	0.732	0.973
	Best Strong	0.622	5.830	0.024	0.009	0.033	0.918	1.000
	<i>Average SE</i>	<i>0.018</i>	<i>0.173</i>	<i>0.005</i>	<i>0.009</i>	<i>0.011</i>	<i>0.021</i>	<i>0.019</i>
500	CART	0.025	8.940	NA	NA	0.021	NA	0.782
	MARS	0.061	19.140	NA	NA	0.348	NA	1.166
	LASSO	0.174	12.220	0.027	0.420	0.446	1.000	0.968
	FFSR-NH	0.577	6.400	0.004	0.049	0.053	1.000	1.009
	FFSR-NHS	0.554	6.450	0.004	0.053	0.058	1.000	1.010
	FFSR-SH	0.701	6.320	0.022	0.021	0.043	1.000	1.005
	FFSR-WH	0.625	6.370	0.008	0.042	0.050	1.000	1.007
	FFSR-NHA1	0.624	6.380	0.028	0.025	0.053	1.000	1.008
	FFSR-NHA2	0.640	6.340	0.023	0.025	0.047	1.000	1.007
	FFSR-WHA2	0.670	6.300	0.015	0.026	0.041	1.000	1.006
	Best No	1.000	6.000	0.000	0.000	0.000	1.000	1.000
	Best Strong	1.000	6.000	0.000	0.000	0.000	1.000	1.000
	<i>Average SE</i>	<i>0.023</i>	<i>0.148</i>	<i>0.003</i>	<i>0.007</i>	<i>0.007</i>	<i>0.004</i>	<i>0.005</i>

Appendix A. Simulation Results

Table A.21: Comparison of Methods for Chi Square Uncorrelated, Model 3, $R^2 = 0.5$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.133	3.520	NA	NA	0.055	NA	0.882
	MARS	0.124	10.600	NA	NA	0.509	NA	1.360
	LASSO	0.295	6.290	0.024	0.368	0.392	0.472	0.857
	FFSR-NH	0.420	2.900	0.007	0.044	0.051	0.338	0.933
	FFSR-NHS	0.412	2.950	0.008	0.054	0.062	0.340	0.928
	FFSR-SH	0.604	3.250	0.033	0.000	0.033	0.422	0.954
	FFSR-WH	0.545	3.170	0.019	0.030	0.048	0.394	0.952
	FFSR-NHA1	0.542	3.050	0.024	0.020	0.044	0.374	0.946
	FFSR-NHA2	0.538	3.040	0.024	0.020	0.044	0.372	0.945
	FFSR-WHA2	0.550	3.130	0.024	0.021	0.044	0.388	0.949
	Best No	0.568	2.950	0.003	0.010	0.013	0.380	0.932
	Best Strong	0.827	3.480	0.016	0.000	0.016	0.480	0.955
	<i>Average SE</i>	<i>0.031</i>	<i>0.137</i>	<i>0.007</i>	<i>0.008</i>	<i>0.012</i>	<i>0.013</i>	<i>0.018</i>
500	CART	0.055	6.880	NA	NA	0.022	NA	0.967
	MARS	0.063	16.060	NA	NA	0.567	NA	1.152
	LASSO	0.209	10.480	0.033	0.424	0.456	0.860	0.981
	FFSR-NH	0.258	4.210	0.002	0.046	0.047	0.592	0.984
	FFSR-NHS	0.262	4.210	0.002	0.044	0.046	0.594	0.984
	FFSR-SH	0.395	5.090	0.044	0.003	0.047	0.758	0.996
	FFSR-WH	0.328	4.620	0.006	0.035	0.041	0.678	0.990
	FFSR-NHA1	0.255	4.080	0.016	0.023	0.039	0.572	0.980
	FFSR-NHA2	0.255	4.070	0.014	0.023	0.037	0.572	0.980
	FFSR-WHA2	0.314	4.560	0.013	0.029	0.042	0.662	0.989
	Best No	0.351	4.480	0.000	0.016	0.016	0.676	0.986
	Best Strong	0.611	5.530	0.025	0.001	0.026	0.868	0.998
	<i>Average SE</i>	<i>0.017</i>	<i>0.174</i>	<i>0.005</i>	<i>0.007</i>	<i>0.009</i>	<i>0.017</i>	<i>0.004</i>

Appendix A. Simulation Results

Table A.22: Comparison of Methods for Chi Square Correlated, Model 1, $R^2 = 0.5$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.082	3.210	NA	NA	0.128	NA	0.637
	MARS	0.104	10.910	NA	NA	0.420	NA	1.305
	LASSO	0.103	4.820	0.019	0.329	0.348	0.322	0.512
	FFSR-NH	0.136	3.270	0.012	0.143	0.156	0.282	0.730
	FFSR-NHS	0.145	3.390	0.019	0.109	0.128	0.315	0.735
	FFSR-SH	0.201	4.710	0.073	0.030	0.103	0.520	0.847
	FFSR-WH	0.168	4.050	0.043	0.093	0.136	0.400	0.803
	FFSR-NHA1	0.175	4.280	0.057	0.085	0.142	0.433	0.848
	FFSR-NHA2	0.172	4.220	0.057	0.085	0.142	0.423	0.841
	FFSR-WHA2	0.186	4.380	0.054	0.083	0.137	0.450	0.849
	Best No	0.215	6.310	0.059	0.219	0.279	0.550	1.005
	Best Strong	0.383	7.070	0.118	0.047	0.165	0.792	1.035
	<i>Average SE</i>	<i>0.011</i>	<i>0.187</i>	<i>0.009</i>	<i>0.013</i>	<i>0.017</i>	<i>0.021</i>	<i>0.024</i>
500	CART	0.029	7.420	NA	NA	0.071	NA	0.770
	MARS	0.069	18.470	NA	NA	0.349	NA	1.140
	LASSO	0.042	7.510	0.005	0.190	0.195	0.797	0.715
	FFSR-NH	0.459	7.080	0.005	0.039	0.044	0.955	1.006
	FFSR-NHS	0.395	7.120	0.005	0.063	0.068	0.932	1.007
	FFSR-SH	0.570	7.390	0.025	0.024	0.050	0.993	1.008
	FFSR-WH	0.478	7.140	0.009	0.045	0.054	0.953	1.008
	FFSR-NHA1	0.433	7.060	0.017	0.032	0.049	0.947	1.006
	FFSR-NHA2	0.433	7.060	0.017	0.032	0.049	0.947	1.006
	FFSR-WHA2	0.465	7.110	0.013	0.039	0.052	0.950	1.007
	Best No	0.640	7.060	0.000	0.029	0.029	0.973	1.004
	Best Strong	0.777	7.030	0.004	0.002	0.006	0.997	1.003
	<i>Average SE</i>	<i>0.022</i>	<i>0.148</i>	<i>0.003</i>	<i>0.007</i>	<i>0.008</i>	<i>0.013</i>	<i>0.007</i>

Appendix A. Simulation Results

Table A.23: Comparison of Methods for Chi Square Correlated, Model 2, $R^2 = 0.5$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.089	2.880	NA	NA	0.211	NA	0.567
	MARS	0.110	11.050	NA	NA	0.389	NA	1.304
	LASSO	0.128	5.130	0.070	0.188	0.258	0.468	0.505
	FFSR-NH	0.202	3.720	0.032	0.044	0.076	0.484	0.797
	FFSR-NHS	0.203	3.700	0.040	0.030	0.070	0.486	0.794
	FFSR-SH	0.333	4.970	0.060	0.016	0.076	0.706	0.938
	FFSR-WH	0.253	4.230	0.050	0.023	0.073	0.582	0.864
	FFSR-NHA1	0.286	4.670	0.063	0.015	0.078	0.654	0.912
	FFSR-NHA2	0.287	4.660	0.061	0.015	0.076	0.654	0.912
	FFSR-WHA2	0.292	4.630	0.053	0.020	0.072	0.652	0.909
	Best No	0.337	5.140	0.039	0.061	0.101	0.706	0.953
	Best Strong	0.535	5.970	0.069	0.017	0.086	0.872	1.003
	<i>Average SE</i>	<i>0.016</i>	<i>0.159</i>	<i>0.011</i>	<i>0.007</i>	<i>0.015</i>	<i>0.022</i>	<i>0.021</i>
500	CART	0.025	7.120	NA	NA	0.078	NA	0.711
	MARS	0.064	19.050	NA	NA	0.348	NA	1.154
	LASSO	0.063	7.460	0.030	0.190	0.220	0.906	0.810
	FFSR-NH	0.651	6.130	0.001	0.023	0.024	0.992	1.003
	FFSR-NHS	0.659	6.120	0.001	0.021	0.023	0.992	1.003
	FFSR-SH	0.743	6.270	0.017	0.020	0.037	1.000	1.004
	FFSR-WH	0.713	6.160	0.004	0.021	0.025	0.996	1.003
	FFSR-NHA1	0.714	6.250	0.021	0.013	0.034	1.000	1.005
	FFSR-NHA2	0.728	6.230	0.020	0.011	0.031	1.000	1.005
	FFSR-WHA2	0.744	6.220	0.016	0.013	0.029	1.000	1.004
	Best No	0.985	6.010	0.000	0.001	0.001	1.000	1.000
	Best Strong	0.982	6.010	0.000	0.001	0.001	1.000	1.000
	<i>Average SE</i>	<i>0.026</i>	<i>0.127</i>	<i>0.003</i>	<i>0.005</i>	<i>0.007</i>	<i>0.009</i>	<i>0.006</i>

Appendix A. Simulation Results

Table A.24: Comparison of Methods for Chi Square Correlated, Model 3, $R^2 = 0.5$

N	Method	AME Ratio	Size	FSR_m	FSR_q	FSR	CSR	CF
100	CART	0.121	3.820	NA	NA	0.170	NA	0.907
	MARS	0.127	9.360	NA	NA	0.518	NA	1.322
	LASSO	0.224	4.640	0.082	0.261	0.343	0.376	0.786
	FFSR-NH	0.360	2.840	0.036	0.076	0.112	0.292	0.923
	FFSR-NHS	0.375	2.860	0.043	0.054	0.097	0.304	0.922
	FFSR-SH	0.483	3.240	0.103	0.005	0.108	0.374	0.952
	FFSR-WH	0.443	3.130	0.067	0.061	0.128	0.336	0.951
	FFSR-NHA1	0.440	3.010	0.072	0.043	0.114	0.324	0.943
	FFSR-NHA2	0.438	2.990	0.069	0.043	0.112	0.322	0.942
	FFSR-WHA2	0.439	3.090	0.070	0.057	0.127	0.330	0.947
	Best No	0.495	3.130	0.042	0.060	0.102	0.346	0.943
	Best Strong	0.637	3.600	0.085	0.011	0.096	0.434	0.957
	<i>Average SE</i>	<i>0.024</i>	<i>0.130</i>	<i>0.012</i>	<i>0.010</i>	<i>0.016</i>	<i>0.013</i>	<i>0.017</i>
500	CART	0.051	7.320	NA	NA	0.064	NA	0.977
	MARS	0.067	15.170	NA	NA	0.571	NA	1.138
	LASSO	0.112	6.110	0.037	0.228	0.265	0.654	0.917
	FFSR-NH	0.236	4.420	0.003	0.148	0.151	0.536	0.991
	FFSR-NHS	0.239	4.480	0.003	0.158	0.161	0.536	0.992
	FFSR-SH	0.428	5.190	0.036	0.002	0.038	0.790	0.997
	FFSR-WH	0.284	4.780	0.012	0.157	0.169	0.580	0.996
	FFSR-NHA1	0.219	4.260	0.022	0.120	0.141	0.510	0.987
	FFSR-NHA2	0.220	4.220	0.019	0.120	0.139	0.508	0.986
	FFSR-WHA2	0.279	4.660	0.016	0.148	0.164	0.566	0.994
	Best No	0.310	4.910	0.002	0.146	0.148	0.620	0.995
	Best Strong	0.621	5.770	0.033	0.012	0.044	0.888	1.000
	<i>Average SE</i>	<i>0.016</i>	<i>0.146</i>	<i>0.005</i>	<i>0.011</i>	<i>0.013</i>	<i>0.016</i>	<i>0.004</i>

Appendix A. Simulation Results

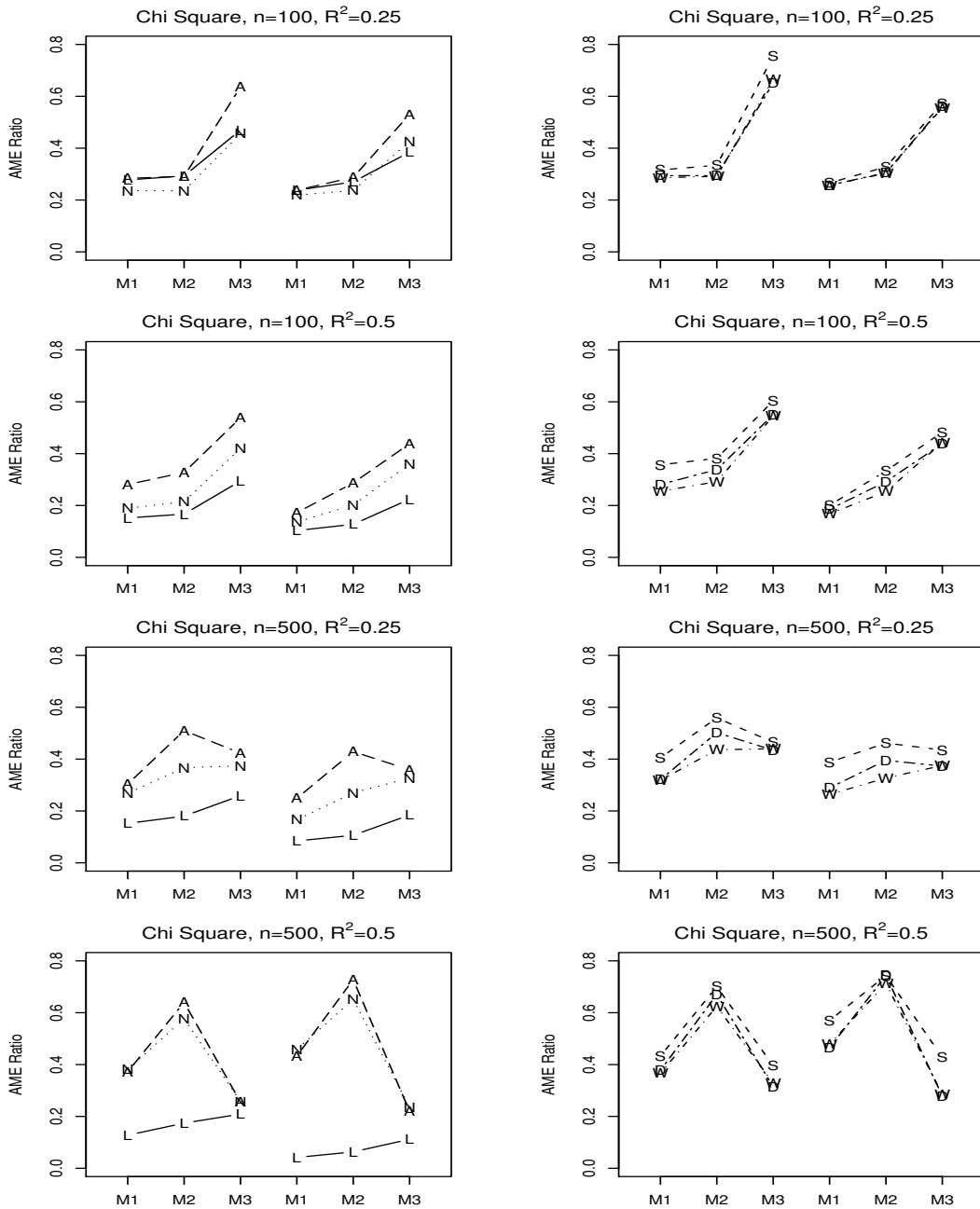


Figure A.1: AME ratio (larger is better) using Chi Square predictors for LASSO (L), FFSR-NH (N), FFSR-SH (S), FFSR-WH (W), FFSR-NHA2 (A), and FFSR-WHA2 (D). The first three values in each plot are uncorrelated predictors and the last three values are correlated predictors.

Appendix A. Simulation Results

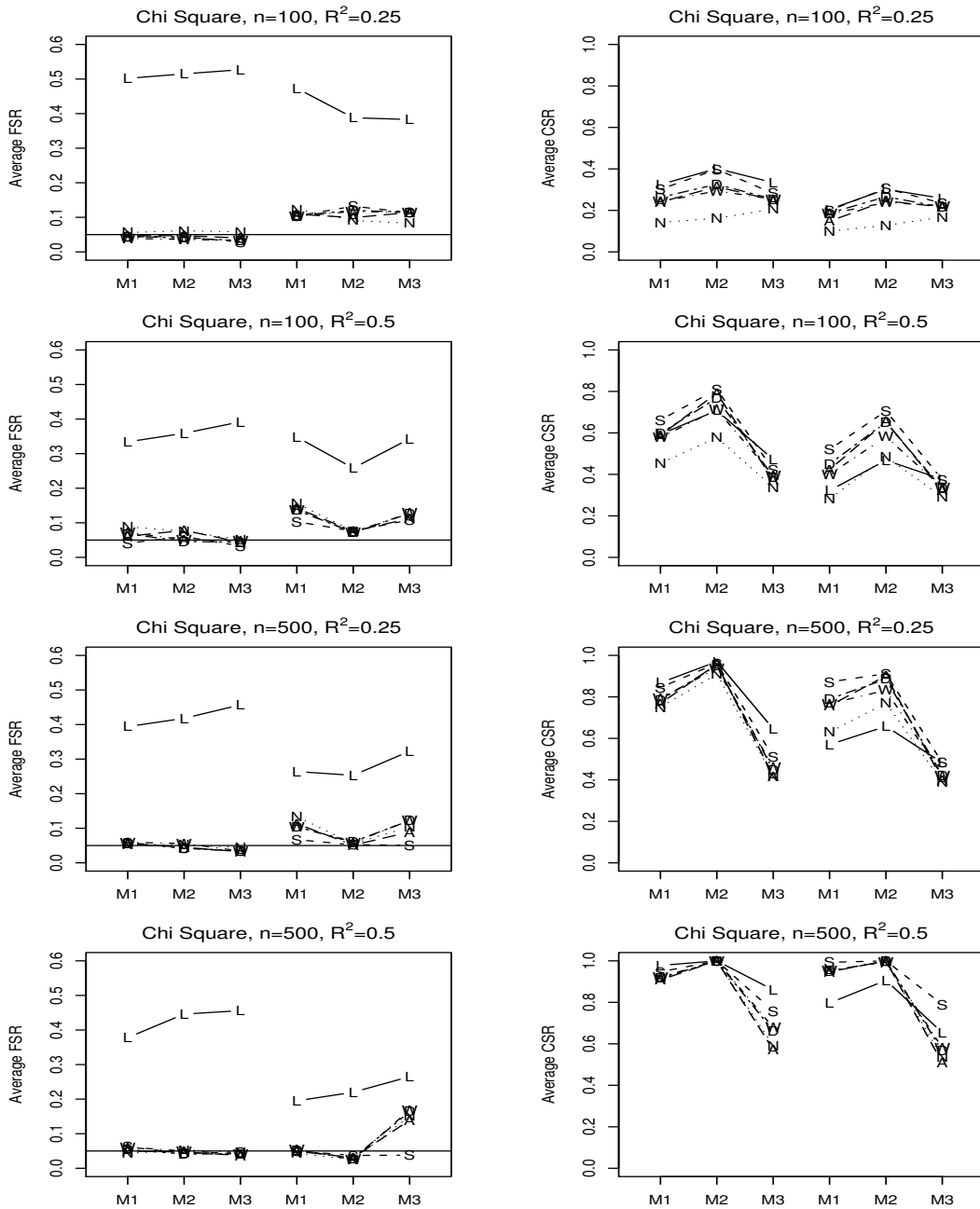


Figure A.2: Average FSR (left) and CSR (right) using Chi Square predictors for LASSO (L), FFSR-NH (N), FFSR-SH (S), FFSR-WH (W), FFSR-NHA2 (A), and FFSR-WHA2 (D). The first three values in each plot are uncorrelated predictors and the last three values are correlated predictors.

Appendix A. Simulation Results

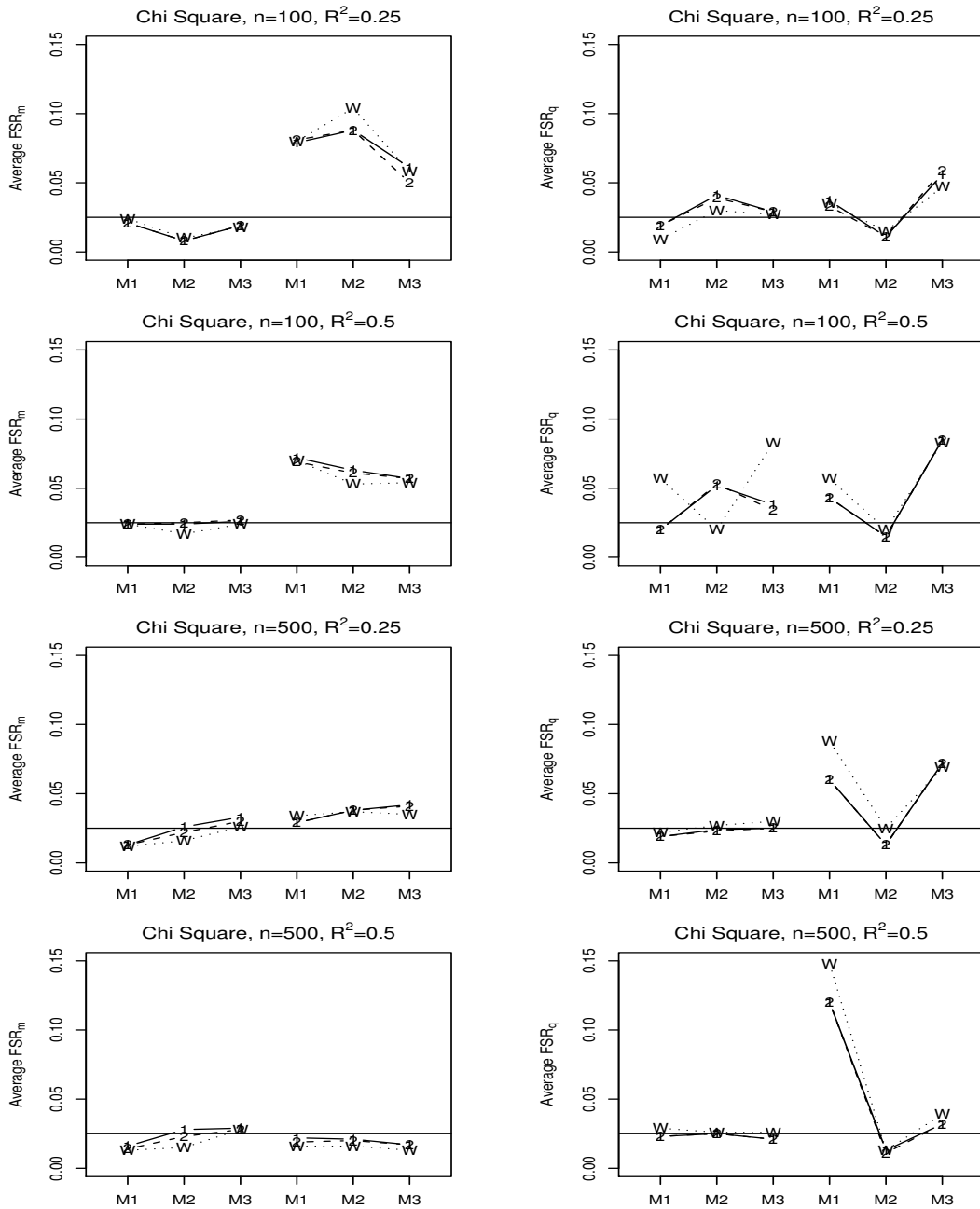


Figure A.3: Comparison of average FSR_m (left) and FSR_q (right) using Chi Square predictors for FFSR-NHA1 (1), FFSR-NHA2 (2), and FFSR-WHA2 (W). The first three values in each plot are uncorrelated predictors and the last three values are correlated predictors.

Appendix A. Simulation Results

Table A.25: Comparison of Methods for Model 1a

R^2	Method	$\mu(\widehat{\mathbf{X}}_{opt}) - \mu(\mathbf{X}_{opt})$	FSR	CSR
0.5	True Model	-4.540 (0.451)	0.000 (0.000)	1.000 (0.000)
	Standard Approach	-29.895 (0.948)	0.131 (0.018)	0.364 (0.023)
	LASSO	-19.343 (0.812)	0.252 (0.016)	0.356 (0.014)
	FFSR-SH	-28.573 (0.647)	0.004 (0.003)	0.068 (0.009)
	FFSR-WH	-30.860 (0.367)	0.003 (0.003)	0.053 (0.007)
	FFSR-NH	-29.427 (0.673)	0.031 (0.009)	0.172 (0.009)
	FFSR-NHA2	-25.379 (0.821)	0.021 (0.007)	0.233 (0.013)
	FFSR-WHA2	-30.730 (0.384)	0.003 (0.003)	0.055 (0.008)
0.75	True Model	-1.292 (0.191)	0.000 (0.000)	1.000 (0.000)
	Standard Approach	-21.800 (0.933)	0.138 (0.019)	0.497 (0.022)
	LASSO	-10.370 (0.636)	0.171 (0.012)	0.555 (0.016)
	FFSR-SH	-21.881 (0.777)	0.006 (0.003)	0.161 (0.016)
	FFSR-WH	-28.790 (0.498)	0.005 (0.005)	0.111 (0.013)
	FFSR-NH	-15.981 (0.631)	0.062 (0.010)	0.398 (0.015)
	FFSR-NHA2	-9.227 (0.502)	0.045 (0.007)	0.596 (0.019)
	FFSR-WHA2	-27.620 (0.645)	0.007 (0.005)	0.134 (0.016)
0.9	True Model	-0.035 (0.004)	0.000 (0.000)	1.000 (0.000)
	Standard Approach	-14.345 (0.655)	0.108 (0.018)	0.621 (0.020)
	LASSO	-4.453 (0.330)	0.115 (0.010)	0.787 (0.015)
	FFSR-SH	-14.940 (0.297)	0.000 (0.000)	0.252 (0.008)
	FFSR-WH	-27.230 (0.374)	0.000 (0.000)	0.168 (0.010)
	FFSR-NH	-4.715 (0.322)	0.082 (0.009)	0.805 (0.012)
	FFSR-NHA2	-1.936 (0.278)	0.075 (0.009)	0.959 (0.007)
	FFSR-WHA2	-23.742 (0.866)	0.007 (0.003)	0.264 (0.026)

Appendix A. Simulation Results

Table A.26: Comparison of Methods for Model 1b

R^2	Method	$\mu(\widehat{\mathbf{X}}_{opt}) - \mu(\mathbf{X}_{opt})$	FSR	CSR
0.5	True Model	-7.840 (0.533)	0.000 (0.000)	1.000 (0.000)
	Standard Approach	-34.759 (1.713)	0.212 (0.023)	0.446 (0.029)
	LASSO	-18.869 (1.020)	0.217 (0.015)	0.357 (0.013)
	FFSR-SH	-18.997 (0.608)	0.024 (0.007)	0.368 (0.009)
	FFSR-WH	-21.892 (0.796)	0.028 (0.007)	0.296 (0.009)
	FFSR-NH	-27.368 (1.285)	0.031 (0.009)	0.245 (0.011)
	FFSR-NHA2	-22.850 (1.097)	0.021 (0.007)	0.290 (0.011)
	FFSR-WHA2	-21.071 (0.831)	0.030 (0.008)	0.312 (0.010)
0.75	True Model	-2.129 (0.255)	0.000 (0.000)	1.000 (0.000)
	Standard Approach	-24.155 (1.393)	0.193 (0.020)	0.507 (0.025)
	LASSO	-13.492 (0.544)	0.192 (0.014)	0.512 (0.014)
	FFSR-SH	-13.819 (0.413)	0.043 (0.008)	0.548 (0.012)
	FFSR-WH	-16.130 (0.416)	0.047 (0.008)	0.447 (0.010)
	FFSR-NH	-17.209 (0.362)	0.033 (0.007)	0.401 (0.009)
	FFSR-NHA2	-12.848 (0.548)	0.041 (0.006)	0.523 (0.016)
	FFSR-WHA2	-14.335 (0.449)	0.037 (0.006)	0.492 (0.013)
0.9	True Model	-0.332 (0.100)	0.000 (0.000)	1.000 (0.000)
	Standard Approach	-16.884 (0.560)	0.109 (0.016)	0.546 (0.021)
	LASSO	-8.133 (0.469)	0.108 (0.011)	0.683 (0.015)
	FFSR-SH	-6.798 (0.341)	0.045 (0.007)	0.778 (0.010)
	FFSR-WH	-8.426 (0.420)	0.069 (0.008)	0.711 (0.014)
	FFSR-NH	-9.485 (0.479)	0.072 (0.009)	0.668 (0.016)
	FFSR-NHA2	-4.661 (0.458)	0.069 (0.008)	0.866 (0.014)
	FFSR-WHA2	-6.726 (0.386)	0.060 (0.008)	0.753 (0.013)

Appendix A. Simulation Results

Table A.27: Comparison of Methods for Model 2a

R^2	Method	$\mu(\widehat{\mathbf{X}}_{opt}) - \mu(\mathbf{X}_{opt})$	FSR	CSR
0.5	True Model	-5.412 (0.479)	0.000 (0.000)	1.000 (0.000)
	Standard Approach	-14.551 (0.793)	0.123 (0.009)	0.761 (0.011)
	LASSO	-24.331 (1.320)	0.092 (0.012)	0.417 (0.026)
	FFSR-SH	-34.017 (0.860)	0.005 (0.003)	0.135 (0.019)
	FFSR-WH	-35.611 (0.727)	0.007 (0.004)	0.091 (0.016)
	FFSR-NH	-20.039 (0.834)	0.028 (0.006)	0.371 (0.020)
	FFSR-NHA2	-18.580 (0.823)	0.022 (0.005)	0.450 (0.023)
	FFSR-WHA2	-35.312 (0.799)	0.006 (0.004)	0.123 (0.019)
0.75	True Model	-2.397 (0.273)	0.000 (0.000)	1.000 (0.000)
	Standard Approach	-6.907 (0.468)	0.106 (0.008)	0.909 (0.006)
	LASSO	-14.939 (1.125)	0.078 (0.009)	0.673 (0.032)
	FFSR-SH	-25.029 (1.357)	0.012 (0.004)	0.448 (0.032)
	FFSR-WH	-27.764 (1.472)	0.013 (0.004)	0.347 (0.034)
	FFSR-NH	-8.923 (0.550)	0.053 (0.007)	0.813 (0.013)
	FFSR-NHA2	-6.565 (0.485)	0.060 (0.008)	0.885 (0.010)
	FFSR-WHA2	-19.112 (1.565)	0.018 (0.003)	0.599 (0.032)
0.9	True Model	-0.475 (0.044)	0.000 (0.000)	1.000 (0.000)
	Standard Approach	-2.092 (0.267)	0.088 (0.007)	0.994 (0.002)
	LASSO	-7.381 (1.203)	0.089 (0.009)	0.818 (0.034)
	FFSR-SH	-20.606 (1.517)	0.032 (0.006)	0.582 (0.034)
	FFSR-WH	-20.189 (1.811)	0.037 (0.006)	0.573 (0.043)
	FFSR-NH	-1.952 (0.274)	0.063 (0.006)	0.990 (0.003)
	FFSR-NHA2	-1.819 (0.268)	0.069 (0.007)	0.991 (0.002)
	FFSR-WHA2	-3.681 (0.897)	0.042 (0.006)	0.943 (0.016)

Appendix A. Simulation Results

Table A.28: Comparison of Methods for Model 2b

R^2	Method	$\mu(\widehat{\mathbf{X}}_{opt}) - \mu(\mathbf{X}_{opt})$	FSR	CSR
0.5	True Model	-5.137 (0.395)	0.000 (0.000)	1.000 (0.000)
	Standard Approach	-14.000 (0.881)	0.121 (0.010)	0.651 (0.011)
	LASSO	-22.399 (1.040)	0.070 (0.010)	0.403 (0.020)
	FFSR-SH	-23.850 (0.831)	0.017 (0.004)	0.415 (0.016)
	FFSR-WH	-23.595 (0.925)	0.019 (0.005)	0.341 (0.016)
	FFSR-NH	-23.654 (0.915)	0.018 (0.005)	0.358 (0.017)
	FFSR-NHA2	-18.903 (0.982)	0.020 (0.005)	0.463 (0.017)
	FFSR-WHA2	-21.061 (0.934)	0.015 (0.004)	0.399 (0.017)
0.75	True Model	-1.726 (0.170)	0.000 (0.000)	1.000 (0.000)
	Standard Approach	-5.144 (0.404)	0.107 (0.009)	0.857 (0.008)
	LASSO	-10.833 (0.951)	0.077 (0.009)	0.684 (0.022)
	FFSR-SH	-9.514 (0.753)	0.046 (0.006)	0.776 (0.014)
	FFSR-WH	-5.693 (0.405)	0.052 (0.006)	0.769 (0.012)
	FFSR-NH	-5.821 (0.398)	0.048 (0.006)	0.767 (0.010)
	FFSR-NHA2	-4.479 (0.371)	0.060 (0.007)	0.804 (0.010)
	FFSR-WHA2	-5.168 (0.362)	0.031 (0.005)	0.753 (0.010)
0.9	True Model	-0.472 (0.063)	0.000 (0.000)	1.000 (0.000)
	Standard Approach	-2.539 (0.325)	0.103 (0.009)	0.967 (0.004)
	LASSO	-3.401 (0.638)	0.092 (0.009)	0.882 (0.018)
	FFSR-SH	-2.143 (0.261)	0.075 (0.008)	0.966 (0.004)
	FFSR-WH	-1.980 (0.234)	0.081 (0.009)	0.961 (0.005)
	FFSR-NH	-1.956 (0.234)	0.081 (0.009)	0.961 (0.005)
	FFSR-NHA2	-1.560 (0.227)	0.079 (0.008)	0.961 (0.005)
	FFSR-WHA2	-1.506 (0.230)	0.053 (0.007)	0.956 (0.005)

APPENDIX B

Response Surface Designs

B.1 Coded Response Surface Designs

1. Small Composite Design with 10 Factors: Tables B.1 - B.3
2. Orthogonal Central Composite Design with 8 Factors: Tables B.4 - B.7

Appendix B. Response Surface Designs

Table B.1: Small Composite Design with 10 Factors (Runs 1-25)

Run	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
1	1	-1	-1	-1	-1	1	-1	-1	-1	-1
2	1	1	-1	-1	-1	-1	-1	-1	1	-1
3	1	1	1	-1	-1	-1	1	-1	-1	1
4	1	1	1	1	-1	-1	-1	1	1	-1
5	1	1	1	1	1	-1	-1	-1	-1	1
6	-1	1	1	1	1	1	-1	-1	1	-1
7	1	-1	1	1	1	1	-1	-1	1	1
8	1	1	-1	1	1	1	1	-1	-1	1
9	1	1	1	-1	1	1	1	1	-1	-1
10	1	1	1	1	-1	1	1	1	-1	-1
11	-1	1	1	1	1	-1	1	1	-1	-1
12	-1	-1	1	1	1	1	1	1	1	-1
13	1	-1	-1	1	1	1	-1	1	-1	1
14	-1	1	-1	-1	1	1	1	-1	-1	-1
15	1	-1	1	-1	-1	1	1	1	-1	-1
16	-1	1	-1	1	-1	-1	1	1	-1	-1
17	1	-1	1	-1	1	-1	1	1	1	-1
18	1	1	-1	1	-1	1	-1	1	1	1
19	1	1	1	-1	1	-1	-1	-1	1	1
20	-1	1	1	1	-1	1	1	-1	1	1
21	-1	-1	1	1	1	-1	-1	1	1	1
22	1	-1	-1	1	1	1	1	-1	-1	1
23	-1	1	-1	-1	1	1	-1	1	1	-1
24	-1	-1	1	-1	-1	1	1	-1	1	1
25	1	-1	-1	1	-1	-1	1	1	1	1

Appendix B. Response Surface Designs

Table B.2: Small Composite Design with 10 Factors (Runs 26-50)

Run	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
26	1	1	-1	-1	1	-1	1	1	1	1
27	-1	1	1	-1	-1	1	-1	1	-1	1
28	1	-1	1	1	-1	-1	-1	-1	-1	-1
29	1	1	-1	1	1	-1	1	-1	1	-1
30	-1	1	1	-1	1	1	-1	1	-1	1
31	-1	-1	1	1	-1	1	-1	-1	1	-1
32	-1	-1	-1	1	1	-1	1	-1	-1	1
33	1	-1	-1	-1	1	1	1	1	1	-1
34	-1	1	-1	-1	-1	1	-1	1	1	1
35	1	-1	1	-1	-1	-1	1	-1	1	1
36	-1	1	-1	1	-1	-1	1	1	-1	1
37	1	-1	1	-1	1	-1	-1	1	-1	-1
38	1	1	-1	1	-1	1	-1	-1	1	-1
39	-1	1	1	-1	1	-1	-1	-1	-1	1
40	-1	-1	1	1	-1	1	1	-1	-1	-1
41	-1	-1	-1	1	1	-1	-1	1	1	-1
42	-1	-1	-1	-1	1	1	1	-1	1	1
43	1	-1	-1	-1	-1	1	-1	1	-1	1
44	-1	1	-1	-1	-1	-1	1	-1	1	-1
45	-1	-1	1	-1	-1	-1	1	1	1	1
46	-1	-1	-1	1	-1	-1	-1	1	-1	1
47	-1	-1	-1	-1	1	-1	-1	-1	-1	-1
48	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
49	-2.63	0	0	0	0	0	0	0	0	0
50	2.63	0	0	0	0	0	0	0	0	0

Appendix B. Response Surface Designs

Table B.3: Small Composite Design with 10 Factors (Runs 51-73)

Run	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
51	0	-2.63	0	0	0	0	0	0	0	0
52	0	2.63	0	0	0	0	0	0	0	0
53	0	0	-2.63	0	0	0	0	0	0	0
54	0	0	2.63	0	0	0	0	0	0	0
55	0	0	0	-2.63	0	0	0	0	0	0
56	0	0	0	2.63	0	0	0	0	0	0
57	0	0	0	0	-2.63	0	0	0	0	0
58	0	0	0	0	2.63	0	0	0	0	0
59	0	0	0	0	0	-2.63	0	0	0	0
60	0	0	0	0	0	2.63	0	0	0	0
61	0	0	0	0	0	0	-2.63	0	0	0
62	0	0	0	0	0	0	2.63	0	0	0
63	0	0	0	0	0	0	0	-2.63	0	0
64	0	0	0	0	0	0	0	2.63	0	0
65	0	0	0	0	0	0	0	0	-2.63	0
66	0	0	0	0	0	0	0	0	2.63	0
67	0	0	0	0	0	0	0	0	0	-2.63
68	0	0	0	0	0	0	0	0	0	2.63
69	0	0	0	0	0	0	0	0	0	0
70	0	0	0	0	0	0	0	0	0	0
71	0	0	0	0	0	0	0	0	0	0
72	0	0	0	0	0	0	0	0	0	0
73	0	0	0	0	0	0	0	0	0	0

Appendix B. Response Surface Designs

Table B.4: Orthogonal Central Composite Design with 8 Factors (Runs 1-25)

Run	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	-1	-1	-1	-1	-1	-1	1	-1
2	-1	-1	-1	-1	-1	1	-1	1
3	-1	-1	-1	-1	1	-1	-1	1
4	-1	-1	-1	-1	1	1	1	-1
5	-1	-1	-1	1	-1	-1	-1	1
6	-1	-1	-1	1	-1	1	1	-1
7	-1	-1	-1	1	1	-1	1	-1
8	-1	-1	-1	1	1	1	-1	1
9	-1	-1	1	-1	-1	-1	-1	-1
10	-1	-1	1	-1	-1	1	1	1
11	-1	-1	1	-1	1	-1	1	1
12	-1	-1	1	-1	1	1	-1	-1
13	-1	-1	1	1	-1	-1	1	1
14	-1	-1	1	1	-1	1	-1	-1
15	-1	-1	1	1	1	-1	-1	-1
16	-1	-1	1	1	1	1	1	1
17	-1	1	-1	-1	-1	-1	-1	-1
18	-1	1	-1	-1	-1	1	1	1
19	-1	1	-1	-1	1	-1	1	1
20	-1	1	-1	-1	1	1	-1	-1
21	-1	1	-1	1	-1	-1	1	1
22	-1	1	-1	1	-1	1	-1	-1
23	-1	1	-1	1	1	-1	-1	-1
24	-1	1	-1	1	1	1	1	1
25	-1	1	1	-1	-1	-1	1	-1

Appendix B. Response Surface Designs

Table B.5: Orthogonal Central Composite Design with 8 Factors (Runs 26-50)

Run	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
26	-1	1	1	-1	-1	1	-1	1
27	-1	1	1	-1	1	-1	-1	1
28	-1	1	1	-1	1	1	1	-1
29	-1	1	1	1	-1	-1	-1	1
30	-1	1	1	1	-1	1	1	-1
31	-1	1	1	1	1	-1	1	-1
32	-1	1	1	1	1	1	-1	1
33	1	-1	-1	-1	-1	-1	1	1
34	1	-1	-1	-1	-1	1	-1	-1
35	1	-1	-1	-1	1	-1	-1	-1
36	1	-1	-1	-1	1	1	1	1
37	1	-1	-1	1	-1	-1	-1	-1
38	1	-1	-1	1	-1	1	1	1
39	1	-1	-1	1	1	-1	1	1
40	1	-1	-1	1	1	1	-1	-1
41	1	-1	1	-1	-1	-1	-1	1
42	1	-1	1	-1	-1	1	1	-1
43	1	-1	1	-1	1	-1	1	-1
44	1	-1	1	-1	1	1	-1	1
45	1	-1	1	1	-1	-1	1	-1
46	1	-1	1	1	-1	1	-1	1
47	1	-1	1	1	1	-1	-1	1
48	1	-1	1	1	1	1	1	-1
49	1	1	-1	-1	-1	-1	-1	1
50	1	1	-1	-1	-1	1	1	-1

Appendix B. Response Surface Designs

Table B.6: Orthogonal Central Composite Design with 8 Factors (Runs 51-75)

Run	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
51	1	1	-1	-1	1	-1	1	-1
52	1	1	-1	-1	1	1	-1	1
53	1	1	-1	1	-1	-1	1	-1
54	1	1	-1	1	-1	1	-1	1
55	1	1	-1	1	1	-1	-1	1
56	1	1	-1	1	1	1	1	-1
57	1	1	1	-1	-1	-1	1	1
58	1	1	1	-1	-1	1	-1	-1
59	1	1	1	-1	1	-1	-1	-1
60	1	1	1	-1	1	1	1	1
61	1	1	1	1	-1	-1	-1	-1
62	1	1	1	1	-1	1	1	1
63	1	1	1	1	1	-1	1	1
64	1	1	1	1	1	1	-1	-1
65	-2.83	0	0	0	0	0	0	0
66	2.83	0	0	0	0	0	0	0
67	0	-2.83	0	0	0	0	0	0
68	0	2.83	0	0	0	0	0	0
69	0	0	-2.83	0	0	0	0	0
70	0	0	2.83	0	0	0	0	0
71	0	0	0	-2.83	0	0	0	0
72	0	0	0	2.83	0	0	0	0
73	0	0	0	0	-2.83	0	0	0
74	0	0	0	0	2.83	0	0	0
75	0	0	0	0	0	-2.83	0	0

Appendix B. Response Surface Designs

Table B.7: Orthogonal Central Composite Design with 8 Factors (Runs 76-100)

Run	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
76	0	0	0	0	0	2.83	0	0
77	0	0	0	0	0	0	-2.83	0
78	0	0	0	0	0	0	2.83	0
79	0	0	0	0	0	0	0	-2.83
80	0	0	0	0	0	0	0	2.83
81	0	0	0	0	0	0	0	0
82	0	0	0	0	0	0	0	0
83	0	0	0	0	0	0	0	0
84	0	0	0	0	0	0	0	0
85	0	0	0	0	0	0	0	0
86	0	0	0	0	0	0	0	0
87	0	0	0	0	0	0	0	0
88	0	0	0	0	0	0	0	0
89	0	0	0	0	0	0	0	0
90	0	0	0	0	0	0	0	0
91	0	0	0	0	0	0	0	0
92	0	0	0	0	0	0	0	0
93	0	0	0	0	0	0	0	0
94	0	0	0	0	0	0	0	0
95	0	0	0	0	0	0	0	0
96	0	0	0	0	0	0	0	0
97	0	0	0	0	0	0	0	0
98	0	0	0	0	0	0	0	0
99	0	0	0	0	0	0	0	0
100	0	0	0	0	0	0	0	0