A COMPUTATIONAL APPROACH TO NONPARAMETRIC REGRESSION: BOOTSTRAPPING CMARS METHOD

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
THE MIDDLE EAST TECHNICAL UNIVERSITY

BY

CEYDA YAZICI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
STATISTICS

SEPTEMBER 2011

Approval of the thesis:

A COMPUTATIONAL APPROACH TO NONPARAMETRIC REGRESSION: BOOTSTRAPPING CMARS METHOD

submitted by **Ceyda YAZICI** in partial fulfillment of the requirements for the degree of **Master of Science in Statistics Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen _____
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. H. Öztaş Ayhan _____
Head of Department, **Statistics**

Assoc. Prof. Dr. İnci Batmaz, _____
Supervisor, **Statistics Department, METU**

Examining Committee Members:

Prof. Dr. H. Öztaş Ayhan _____
Statistics Department, METU

Assoc. Prof. Dr. İnci Batmaz _____
Supervisor, Statistics Department, METU

Prof. Dr. Gülser Köksal _____
Industrial Engineering Department, METU

Assist. Prof. Dr. Özlem İlk _____
Statistics Department, METU

Assist. Prof. Dr. Ceylan Yozgatlıgil _____
Statistics Department, METU

Date: 15.09.2011

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Ceyda YAZICI

Signature:

ABSTRACT

A COMPUTATIONAL APPROACH TO NONPARAMETRIC REGRESSION: BOOTSTRAPPING THE CMARS METHOD

Yazıcı, Ceyda

M.Sc., Department of Statistics

Supervisor: Assoc. Prof. Dr. İnci Batmaz

September 2011, 114 pages

Bootstrapping is a resampling technique which treats the original data set as a population and draws samples from it with replacement. This technique is widely used, especially, in mathematically intractable problems. In this study, it is used to obtain the empirical distributions of the parameters to determine whether they are statistically significant or not in a special case of nonparametric regression, Conic Multivariate Adaptive Regression Splines (CMARS). Here, the CMARS method, which uses conic quadratic optimization, is a modified version of a well-known nonparametric regression model, Multivariate Adaptive Regression Splines (MARS). Although performing better with respect to several criteria, the CMARS model is more complex than that of MARS. To overcome this problem, and to improve the CMARS performance further, three different bootstrapping regression methods, namely, Random-X, Fixed-X and Wild Bootstrap are applied on four data sets with different size and scale. Then, the performances of the models are compared using various criteria including accuracy, precision, complexity, stability, robustness and efficiency. Random-X yields more precise, accurate and less complex models particularly for medium size and medium scale data even though it is the least efficient method.

Keywords: Bootstrapping Regression, Conic Multivariate Adaptive Regression Splines, Fixed-X Resampling, Random-X Resampling, Wild Bootstrap

ÖZ

PARAMETRİK OLMAYAN REGRESYON MODELİNE HESAPLAMALI BİR YAKLAŞIM: KONİK ÇOK DEĞİŞKENLİ UYARLANABİLİR REGRESYON EĞRİLERİNE (KÇURE) KORUYAN HALKA YÖNTEMİNİN UYGULANMASI

Yazıcı, Ceyda

Yüksek Lisans, İstatistik Bölümü

Tez Yöneticisi: Doç. Dr. İnci Batmaz

Eylül 2011, 114 sayfa

Koruyan Halka (Bootstrap) yöntemi esas veri kümesine kitle gibi davranarak ondan örneklem alan bir yeniden örnekleme yöntemidir. Bu teknik, özellikle matematiksel çözümü olmayan problemlerde yaygın olarak kullanılmaktadır. Bu çalışmada, özel bir parametrik olmayan regresyon yöntemi olan Konik Çok Değişkenli Uyarlanabilir Regresyon Eğrilerine (KÇURE) ilişkin parametrelerin istatistiksel olarak önemli olup olmadığına karar vermek amacı ile deneysel dağılımlarını elde etmek için kullanılmıştır. Burada KÇURE yöntemi, konik ikinci derece eniyileme yöntemini kullanan ve iyi bilinen bir parametrik olmayan regresyon yöntemi olan Çok Değişkenli Uyarlanabilir Regresyon Eğrilerinin (ÇURE) değiştirilmiş özel bir şeklidir. Birçok ölçüte göre daha iyi başarıma sahip olduğu halde, KÇURE modeli ÇURE modelinden daha karmaşıktır. Bu problemin üstesinden gelebilmek ve KÇURE modelinin başarımını daha da iyileştirmek amacı ile üç farklı Koruyan Halka yöntemi, Sabit-X, Rastgele-X ve de Aşırı (Wild) Koruyan Halka yöntemi, büyüklük ve ölçekleri bakımından farklı dört veri kümesine uygulanmıştır. Daha sonra geliştirilen modellerin başarımları doğruluk, kesinlik, karmaşıklık, durağanlık (stability), sağlamlık (robustness) ve etkinlik (efficiency) ölçütleri açısından karşılaştırılmıştır. Rastgele-X yöntemi, daha az etkin olmasına rağmen özellikle orta

büyüklükte ve ölçekte veri kümeleri için daha kesin, doğru ve daha az karmaşık modeller üretmiştir.

Anahtar Kelimeler: Koruyan Halka Regresyonu, Konik Çok Değişkenli Uyarlanabilir Regresyon Eğrileri, Sabit-X Yeniden Örnekleme, Rastgele-X Yeniden Örnekleme, Aşrı Koruyan Halka

To my family

for their unconditional love

and everlasting support

ACKNOWLEDGEMENTS

First of all, I would like to express my sincere gratitude to my thesis supervisor, Assoc. Prof. Dr. İnci Batmaz, who accepted me as a M.Sc. student. I am deeply indebted to her, for patience, help, suggestions and encouragements in all the process of research and writing of this thesis. Her guidance, support, and feedbacks have turned this study to an immeasurable learning experience for me.

I am also thankful to Prof. Dr.Öztaş Ayhan, Prof. Dr. Gülser Köksal, Assist. Prof. Dr. Özlem İlk and Assist. Prof. Dr. Ceylan Yozgatlıgil for their relevant discussions, suggestions and comments. I also would like to thank my examining committee members for their acceptance of reviewing my thesis study by spending their invaluable time.

I would also like to give special thanks to Fatma Yerlikaya-Özkurt and Elçin Kartal-Koç for their full support, everlasting patience, motivation and kindness. They always tried to answer my questions and contributed to this study. I also want to express my thanks to them for the amusing time we spent.

I owe my special thanks to my dear friend Derya Cidal for her perfect friendship, patience, kindness, confidence and the enjoyable time we spent from the first moment I have met her. I also want to express my thanks to Nazire Canver for her friendship and kindness.

I would like to give special thanks to Münire Tuğba Erdem, Könül Bayramoğlu and Gül İnan for their full support and help during this study.

But most of all, I am forever indebted to my parents, and my brother who have always stood by me, for their endless love, understanding, endless patience and encouragement when it was most required.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES

LIST OF TABLES

TABLES

LIST OF ABBREVIATIONS

ANN: Artificial Neural Network

ARIMA: Autoregressive Integrated Moving Average

B: Number of Bootstrap Samples

BCMARS: Bootstrap Conic Multivariate Adaptive Regression Splines

BF: Basis Function

BRT: Boosted Regression Trees

CART: Classification and Regression Trees

CI: Confidence Interval

CMARS: Conic Multivariate Adaptive Regression Splines

COX PH: Cox Proportional Hazards Model

CQP: Conic Quadratic Programming

CS: Concrete Slump Data Set

CV: Cross Validation

DGP: Data Generating Process

ECDF: Empirical Cumulative Distribution Function

FF: Forest Fires Data Set

GAM: Generalized Additive Models

GCV: Generalized Cross Validation

LAD: Least Absolute Deviation

LMS: Least Median Square

LP: Linear Programming

LR: Linear Regression

LS: Least Squares

M: Number of Maximum BFs

MAE: Mean Absolute Error

MAPE: Mean Absolute Percentage Error

MARS: Multivariate Adaptive Regression Splines

MLR: Multiple Linear Regression

MSE: Mean Square Error

PRESS: Prediction Error Residual Sum of Squares

PRSS: Penalized Regression Sum of Squares

PWI: Proportion of Residuals within Some User-Specified Range

$\mathbf{R^2}$: The Coefficient of Determination

RSS: Residual Sum of Squares

US: Uniform Sampling Data Set

CHAPTER 1

INTRODUCTION

Computational statistics is defined as "the method that heavily use computational techniques to create new statistical methodology" by Wegman (1988). In general, it includes computer-intensive methods of statistics and visualization. After the developments in the computer hardware and software, these methods have become feasible and popular, especially, since 1980s. Thus, producing and storing huge and high-dimensional data became easier and methods such as high-dimensional data representation became applicable.

Efron and Tibshirani (1991) give the following methods as examples of computational statistics which are called as *computer-intensive statistical methods*: bootstrap methods, generalized additive models (GAM), nonparametric regression, and classification and regression trees (CART). In these methods, modern computer algorithms are used instead of classical mathematical methods. The advantage of computational approach is that the analyst is free from choosing methods because of its mathematical tractability (Martinez and Martinez, 2002).

When computational statistics is compared with traditional statistics, it can be said that it is generally applicable to numerically tractable, computationally intensive, imprecise questions (Wegman, 1988). In these situations, the computational methods, including resampling, simulations and multiple views to make inferences for the parameters of the model are used (Gentle, 2009).

Multivariate Adaptive Regression Splines (MARS), a nonparametric regression method, is first introduced by Friedman (1991). It is widely used in every branch of science and engineering. Its main advantage lies in building models, which include main and interaction terms, for high-dimensional data. The MARS algorithm consists of two parts: *forward* and *backward*. In the forward part, a large model is obtained. The large model is reduced in the backward step. CMARS, which is introduced first by Yerlikaya (2008), and further improved by Batmaz et al. (2010) and evaluated by Weber et al. (2011), is a new approach to backward part of the algorithm by using conic quadratic optimization. In CMARS, the coefficients for the terms obtained from the forward step of MARS are calculated by conic quadratic optimization (CQP). As a result, there are more terms in a CMARS model than that of MARS, and hence, CMARS models are at least as complex as MARS models.

The mathematical intractability appears as the lack of distribution fitting to the parameters. If the distribution of the parameters were known, then the significance of the parameters could be determined by using the hypothesis testing or constructing confidence intervals (CIs) of model parameters by utilizing traditional statistical approaches. Unfortunately, in this special case of nonparametric regression, the distributions of the parameters are not known.

In this thesis, an empirical distribution is tried to be fitted to each parameter of CMARS models by using a computational method called *bootstrap*. Bootstrap is a resampling method that heavily depends on computer (Hjorth, 1994). The aim of this approach is to take samples with replacement from the original sample and calculate the parameter of interest. Using this approach, the statistically significant model parameters are determined, and thus, the large CMARS model (in other words the model complexity) is tried to be reduced. Three different bootstrapping regression methods, namely, Random-X, Fixed-X and Wild bootstrap methods are run on four different data sets with size and scale. These data sets are named as Concrete Slump (CS) (Yeh, 2007), Forest Fires (FF) (Cortez and Morais, 2007), PM10 (Aldrin, 2006) and Uniform Sampling (US) (Kartal, 2007). Then, performances of these methods

are compared with respect to various criteria including accuracy, precision, complexity, stability, robustness and efficiency.

In this study, a literature review for the applications of MARS and CMARS methods, are given in Chapter 2. This chapter also includes a literature survey of bootstrap applications for modeling. In Chapter 3, MARS, CMARS and bootstrap methods that are used in this study are explained in detail. Application results of the methods on the data sets are presented in Chapter 4. In Chapter 5, the findings are discussed. Conclusions and future studies, which can be developed depending on the findings of this thesis, are given in the last chapter.

CHAPTER 2

BACKGROUND

The analysis related with the social, physical or economic systems is based on a structure. This underlying, logical structure is defined as a *model* by Ramanathan (2002). A model carries the behavior of the members of the system, and it is the basic framework of the analysis. Hjorth (1994) defines the aim of models as a way to structure ideas and conclusions. According to him, the models are simple forms of the research phenomenon. Moreover, he classifies the models as "not true" but will have some mistakes except for some pure situations. In statistics, a model is an equation or a form of system which has several equations. To conduct an empirical study, formulating a model for the scientific question is the first step that should be conducted. After gathering the data, the model, including parameters should be estimated. If there are assumptions in the model, these must be checked by conducting hypothesis testing or with the help of visualization techniques. If the assumptions are satisfied, then the results can be interpreted statistically; otherwise, necessary attempts have to be made to validate them.

Modeling is widely used in various fields of study such as engineering, economics, finance, biology and genetics. Parametric and nonparametric approaches are the major two branches of statistical modeling. When assumptions are validated, parametric statistical models provide more trustable inferences. However, in certain situations, it may not be possible to satisfy some assumptions of a parametric approach. In these cases, nonparametric approaches are suggested to be used.

In 1991, famous statistician and physicist Jerome Friedman introduced the MARS model as a new method in nonparametric regression. The advantage of this model is

that it can handle the high-dimensional data easily and well approximate nonlinearity in case of high nonlinearity. MARS has a wide application area from biology to economy.

Lin et al. (2011) compare the MARS model of tourism demand with ARIMA and Artificial Neural Networks (ANN) according to Mean Absolute Percentage Error (MAPE). The results of this study indicate that the forecasting ability of ARIMA is the best among the others in the application of times series data. Zakeri et al. (2010) use MARS to model the prediction of energy expenditure for the first time in this research area. Moreover, Kriner (2007) uses MARS for survival analysis and show that the new model they obtained is a better fit than the classical method, called Cox PH approach.

MARS also has a wide application in biostatistics. For instance, York et al. (2006) compare the power of MARS with least squares (LS) curve fitting using polynomials. The results show that the power for MARS is higher than the LS method for detecting disease-risk relationship among different subgroups. Deconinck et al. (2008) compare the performances of the MARS and Boosted Regression Trees (BRT) by using a data set for blood-brain barriers passage. The authors conclude that MARS is performing superior to BRT in terms of fitting nonlinearities, being robust to small changes in the data and having easier interpretation.

The model is also applied to different research areas such as geology. Gutiérrez et al. (2011) use MARS to model soil properties of a region in Spain. In their study, the reason for choosing this predictive model is that it is faster, accurate and has easy interpretation compared with ANN and CART. Moreover, MARS is used for the first time to model the species distributions in freshwater environments in the study of Leathwick et al. (2005).

Denison et al. (1998) provide a Bayesian algorithm for MARS. In general, MARS has been applied to different data sets including time series, biostatistics, meteorology, geology and biology.

Yerlikaya (2008) proposes a contribution to MARS model, and call it as Conic MARS (CMARS). In CMARS, the backward step of MARS is replaced with *CQP*. Batmaz et al. (2010) improve CMARS to better model nonlinearities in data and compare it with MARS and Multiple Linear Regression (MLR). According to their results MARS and CMARS perform better than MLR. Moreover, CMARS has a higher performance in terms of R^2 measure. Weber et al. (2011) evaluate CMARS method rigorously, and state that it produces more accurate, robust and stable models than MARS under various data features.

In another study, Taylan et al. (2010), compare MARS and CMARS for classification and use a data set for diabetes. The analysis results show that the accuracy measures for both train and validation data sets are not different. Moreover, this study concludes that CMARS is also superior to MARS in terms of reducing the probability of committing Type-II Error. Özmen et al. (2011) propose a robustification on the CMARS method in order to reduce the estimation variance in case of modeling random (but not fixed) variables. Alp et al. (2011) compare GAM, CMARS, MARS and Logistic Regression (LR) to detect a financial crisis before it occurs. The authors conclude that CMARS has better results with higher correct classification rate and stability, and also, being robust.

In parametric modeling such as MLR, the significance of parameters can be tested by conducting hypothesis or with the help of CIs. For instance, in the MLR it is assumed that the parameters are distributed normally. However, if there is no information on the distribution of parameters and normality assumption is not plausible, methods in the computational statistics can be used.

Efron (1988) applies bootstrap to Least Absolute Deviation (LAD) method. Fox (2002) uses Random-X and Fixed-X Resampling methods for robust regression which uses *M-estimator* with the Huber weight function. Also, Salibian-Barrera and Zamar (2002) apply bootstrapping to robust regression. Austin (2008) replaces bootstrap with backward elimination which results a better coverage in percentile

CIs. Yetere-Kursun and Batmaz (2010) compare regression methods by employing different bootstrapping methods.

Bootstrapping for estimating regression parameters is also applicable in biostatistics and bioinformatics. Loughin and Koehler (1997) estimate parameters by bootstrapping in multivariate survival analysis. Kirk and Stumpf (2009) use bootstrap resampling for Gaussian Process Regression (GPR), which is a Bayesian Model. Flachaire (2003) compares the pairs bootstrap with wild bootstrap for heteroscedastic models. Gonçalves and White (2005) use moving blocks bootstrap approach for estimating standard errors of the parameters in the MLR.

Efron and Tibshirani (1993) apply resampling residuals to a model based on Least Median of Squares (LMS). The difference between this technique and LS approach depends on the fitting procedure. In LMS, the median of the residual sum of squares (RSS) is used. In the model, median is used since it is more resistant to influential observations.

Chernick (2008) uses vector resampling for a kind of nonlinear model that is used in aerospace engineering. Montgomery et al. (2001) conduct bootstrapping residuals method to Michaelis-Menten model, which is a nonlinear regression. Note however that the data set used in the study has a small sample size, whereas the theoretical approach for the model is valid only for large samples.

CHAPTER 3

METHODS

3.1. Multivariate Adaptive Regression Splines (MARS)

MARS, which is published by Jerome Friedman in 1991, is a nonparametric regression technique in which there is no assumption for the relationship between dependent and independent variables. According to Hastie et al. (2001) MARS is similar to stepwise MLR or another approach for CART model. MARS has an advantage for building models of high-dimensional data and approximating the nonlinearity in data. In addition to obtaining additive models, it is useful for constructing models including interaction terms. Due to recorded success of MARS in modeling real life data, it has a wide range of applications in various areas of research in recent years. As mentioned in Chapter 2, the application areas range from biology to meteorology.

MARS stands for "Multivariate Adaptive Regression Splines". Below each term is defined briefly:

- MARS has the capability of dealing with multi-dimensional data; that is why MARS is known as a multivariate procedure.
- MARS is able to reduce the model complexity (or terms) if any predictor does not contribute enough. Thus, it has special "adaptive" procedure.
- MARS is a regression technique which investigates the relationship between variables.
- "Splines" refer to the class of piecewise functions that are used in modeling. To obtain a spline, original space is divided into intervals separated by knot

values. Thus, any shape can be approximated with the help of sufficient number of knots.

MARS constructs a model after conducting forward and backward algorithms. In the forward part, a large model including many basis functions (BFs) is obtained. This large model may lead to overfitting. In this stage, some of the terms in the model may not contribute to the model. Thus, a backward algorithm is applied to reduce the number of terms and retain only the ones that contribute.

MARS is a nonparametric regression method. This regression is free from the assumption for the relationship that may exist between the dependent and independent variables. However, in parametric models such as MLR, the relationship between the response and predictor variables is predetermined. These parametric models are easy to implement but constrained by the assumptions of the model. The nonparametric models are defined to obtain a model where the assumptions of the parametric models are not satisfied.

The classical nonparametric regression is defined as

$$y_i = f\left(\beta, x_i^{'}\right) + \varepsilon_i, \quad i = 1, \ldots, n; \tag{1}$$

where $\boldsymbol{\beta}$ stands for the parameters, n represents the sample size of the data and x_i represents the independent variables. In this model, f is in an unknown functional form.

In the MARS model, the following forms of the independent variables are used as inputs to obtain a model. These are

$$\left(x-t\right)_+ = \begin{cases} x-t, \text{ if } x > t, \\ 0, \text{ otherwise} \end{cases} \qquad \left(t-x\right)_+ = \begin{cases} t-x, \text{ if } x < t, \\ 0, \text{ otherwise} \end{cases} \tag{2}$$

where $t \in \left\{x_{1j}, x_{2j}, \ldots, x_{nj}\right\}$

The first form of the BF takes the value of zero for the values less than *t*, and takes the value of the magnitude of the difference between *t* and *x*, otherwise. In the second form, the value of it is set to zero if the value of *x* is greater than *t*. Otherwise, it is set to the magnitude of the difference between *t* and the value of *x*. Thus, the value of the BF always takes a nonnegative value. Figure 1 represents the BFs (*x*-0.5) and (0.5-*x*). They are piecewise linear functions and the value of *t* is defined as the *knot value* for the BF. Moreover, these BFs are also called linear splines (Hastie et al., 2001). These two functions are reflected pairs of each other.

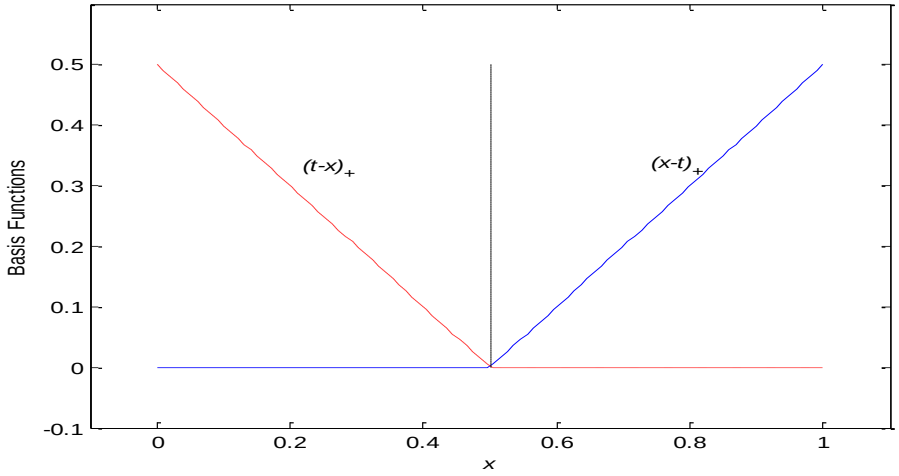

Figure 1. The BFs $(x-t)_+$ and $(t-x)_+$ used by MARS (Based on Hastie et al., 2001)

Here, the purpose is to obtain the reflected pairs for each x_j with knots at each observed value of x_{ij}. In this notation, p represents the number of independent variables. The set of BFs where $t \in \{x_{1j}, x_{2j}, \ldots, x_{nj}\}$ and $j = 1,2,\ldots, p$ is:

$$C = \{(x_j - t)_+, (t - x_j)_+\}. \tag{3}$$

There are $2np$ BFs, if all of the observed values of independent variables are distinct. The key point here is that each BF depends only on one independent variable, x_j.

The multivariate spline BFs take the following form to employ the BF that is tensor products of univariate spline functions:

$$B_m(x) = \prod_{k=1}^{K_m} \left[s_{km} \left(x_{(km)} - t_{km} \right) \right]_+, \tag{4}$$

where K_m represents the number of truncated functions in the m^{th} BF, x_{km} shows the input variable corresponding to the k^{th} truncated linear function in the m^{th} BF and t_{km} is the corresponding knot value and s_{km} takes the value of 1 or -1.

The method for obtaining the model is similar to forward stepwise MLR. However, in MARS, the BFs coming from the set C given in (3) are used instead of original independent variables. The MARS model is defined as

$$f(x) = \beta_0 + \sum_{m=1}^{M} \beta_m h_m(x) + \varepsilon, \tag{5}$$

where each h_m belongs to the set C and M represents the number of BFs in the current model.

Given a choice for the h_m, the coefficients for the parameters (β_m) are estimated by minimizing the RSS with the same method similar to the one used in the usual MLR. In this part of modelling, the key point is to determine the $h_m(x)$. The constant function $h_0(x) = 1$ is the first function that is used, and all functions in C are considered as candidate functions.

The functions below are possible alternatives for BFs (Kriner, 2007):

- 1
- x_j

- $x_l x_j$
- $(x_j - t_k)_+$
- $(x_j - t_k)_+ (x_l - t_h)_+$

In the BFs above which include multiplication of two BFs, the independent variables, x_l and x_j are different. This is due to the algorithm of the MARS model. According to the algorithm, BFs cannot include same independent variables.

The decision on adding a new BF to the current model is explained with the following algorithm. Let M represent the current model set. The BFs in the current model are multiplied by the BFs in the candidate set C (with their reflected pairs) as shown below:

$$\hat{\beta}_{M+1} h_l(x)(x_j - t)_+ + \hat{\beta}_{M+2} h_l(x)(t - x_j)_+, h_l \in M. \quad (6)$$

The BF which causes the most amount of reduction in the residual error is added to the model first. The estimates of coefficients (including $\hat{\beta}_{M+1}$ and $\hat{\beta}_{M+2}$) are determined by the LS approach. When the maximum number of terms, which is determined by the user, is reached, this process is finished.

According to (Kriner, 2007) the functions given below can be some candidate BFs:

- $x_j, j = 1,2,\ldots, p,$
- $x_l x_j$, if x_j and x_l are already in the model,
- $(x_j - t_k)_+$, if x_j is already in the model,
- $(x_j - t_k)_+ (x_l - t_h)_+$, if $x_j (x_l - t_h)_+$ and $(x_j - t_k)_+ x_l$, are already in the model.

The forward procedure of MARS is finished by yielding a large model. This model may causes overfitting in the data. In this situation, the model estimates the data well, however it is not safe to generalize it. So, a backward deletion procedure is

needed. In this step, a term in the model whose deletion causes the least amount of residual squared error is deleted first. This procedure estimates the best model, $\hat{f}_M$, of each size (number of terms) M. Cross validation (CV) is a possible solution for finding the optimal value of M. However, *generalized cross validation (GCV)* is used due to computational purposes. The GCV is defined as

$$GCV = \frac{\sum_{i=1}^{n}\left(y_i - \hat{f}_M(x_i)\right)^2}{\left(1 - C(M)/n\right)}, \tag{7}$$

where *n* represents the number of data samples. The numerator of the GCV is the usual RSS.

In general, $C(M)$ is calculated by using the following formula:

$$C(M) = trace\ (B(B^T B)^{-1} B^T) + 1. \tag{8}$$

$C(M)$ represents the cost penalty measure of a model in which there are M BFs. So, B is the matrix of BFs with dimensions $M \times n$.

However, $C(M)$ also has a representation other than the above formula. For instance, it is also calculated as $r + cK$, where r represents the linearly independent BFs, and K shows the number of knot points used during the forward step and the value of c is generally taken as three. In case of additive models, the value is taken as two. If the value of $C(M)$ is small, it produces a model with many BFs. Otherwise, a smaller model with less BFs is obtained. This procedure continues for all BFs and then chooses the best model that has minimum GCV. .

MARS uses BFs for modeling instead of original variables and has a particular modeling procedure. Piecewise linear BFs allow model to operate locally instead of global modeling. Multiplication of two BFs produces a result which is nonzero only over the factor space where both components are nonzero (Figure 2). Thus, the regression surface is obtained by using only nonzero components locally- only when

they are needed. If polynomial BFs are used, then the multiplication of BFs would be nonzero everywhere and would not work as well.

The BF in Figure 2 is defined as the multiplication of two BFs such as $h(X_1, X_2) = (X_1 - x_{15})_+ (x_{27} - X_2)_+$.

The forward procedure of modeling is hierarchical. Multiway products of BFs are constructed from the terms already exist in the model. If a higher-order level BF is in the model, then its lower-order components must also exist in the model. For instance, a three-way product can only exist in the model if one of its two-way components is already in the model. This property avoids constructing a large number of alternatives and focuses on the BFs in the model.

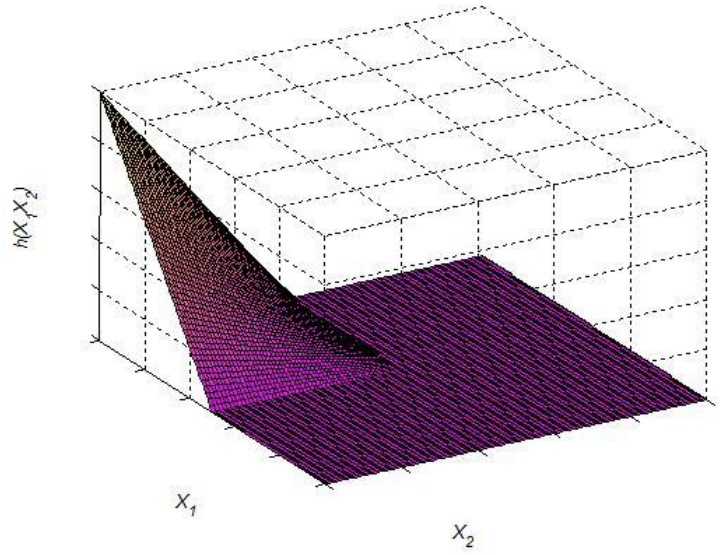

Figure 2. Two-way interactions BFs (Based on Hastie et al., 2001)

There is a limitation in the construction of the model: each independent variable can exist at most once in a product. This prevents the occurrence of higher-order degrees of a variable which increase or decrease too sharply near the boundaries of the

feature space. Piecewise linear function can approximate the higher-order powers in a more stable way.

There is an option to set the upper limit of the degree of interaction in the MARS model. By setting it to three, the powers of four and more interactions are not allowed. If it is set to one, an additive model is produced. This makes the interpretation of the model easier.

3.2. Conic Multivariate Adaptive Regression Splines (CMARS)

CMARS is a modified version of MARS developed by using the CQP. The letter "C" here stands for *"convex"* and *"continuous"*. Yerlikaya (2008) explains the modified model as the following. Note that the notations below will be used for the BFs:

$$c^{+}(x,\tau)=\left[+(x-\tau)\right]_{+},\ c^{-}(x,\tau)=\left[-(x-\tau)\right]_{+}. \tag{9}$$

Here, $[q]_{+}:=\max\{0,q\}$ and τ is a univariate knot. Each BF is a piecewise linear function which has a knot value at τ.

In the previous section, the nonparametric regression model is explained as

$$Y=f(\boldsymbol{\beta},\boldsymbol{X})+\varepsilon, \tag{10}$$

where Y represents the dependent variable whereas $\boldsymbol{X}=(X_1,X_2,\ldots,X_p)^T$ is a vector of independent variables. ε is the error term which is assumed to have zero mean and finite variance.

The objective here is to obtain reflected pairs for each independent variable, $X_j\,(j=1,2,\ldots,p)$ with knot values at $\tau_i=(\tau_{i,1},\tau_{i,2},\ldots\tau_{i,p})$ or at $\bar{x}_i=(\bar{x}_{i,1},\bar{x}_{i,2},\ldots,\bar{x}_{i,p})^T$ of that independent variable. These $\bar{x}_i$ values are very close but not equal to τ_i value. The aim of this modification is to take the derivatives during optimization process. The nonparametric regression model given in (9) can be explained by the following formula

$$Y = \theta_0 + \sum_{m=1}^{M} \theta_m(\boldsymbol{X}) + \varepsilon, \tag{11}$$

where θ_0 is the intercept and $\psi_m\,(m = 1,2,\ldots,M)$ are the BFs, where ψ_m comes from ζ which is defined by

$$\zeta = \left\{ \left(X_j - \tau\right)_+, \left(\tau - X_j\right)_+ \mid \tau \in \left\{x_{1,j}, x_{2,j}, \ldots, x_{N,j}\right\}, j \in \{1,2,\ldots,p\} \right\}. \tag{12}$$

The m^{th} BF is defined as

$$\psi_m(x) = \prod_{j=1}^{\kappa_m} \left[s_{\kappa_j^m} \left(x_{\kappa_j^m} - \tau_{\kappa_j^m} \right) \right]_+ , \tag{13}$$

where κ_m is the truncated linear functions multiplied in the m^{th} BF and $x_{\kappa_j^m}$ is the independent variable associated with m^{th} BF. $\tau_{\kappa_j^m}$ represents the knot value corresponding to the variable $x_{\kappa_j^m}$ and $s_{\kappa_j^m}$ takes the value of 1 or -1. There are *2np* BFs in total in case of distinct independent variables.

The BFs, which came from the set of ζ, are represented by ψ_m in the model. β_m is the coefficient for the m^{th} BF (m = 1,2,…,M).

In CMARS, the backward algorithm is eliminated. Instead of backward procedure, penalty terms after applying LS are used to control the lack of fit.

3.2.1. The Penalized Residual Sum of Squares Problem (PRSS)

For the MARS model, the Penalized Residual Sum of Squares **(**PRSS) has the following form:

$$PRSS = \sum_{i=1}^{n} \left(y_i - f(\bar{\boldsymbol{x}}_i) \right)^2 + \sum_{m=1}^{M_{\max}} \lambda_m \sum_{\substack{|\boldsymbol{\alpha}|=1 \\ \boldsymbol{\alpha}=(\alpha_1,\alpha_2)^T}}^{2} \sum_{\substack{r<s \\ r,s \in V(m)}} \int \theta_m^2 \left[D_{r,s}^{\alpha} \psi_m(\boldsymbol{t}^m) \right]^2 d\boldsymbol{t}^m, \tag{14}$$

where $M_{\max}$ is the number of BFs reached at the end of the forward algorithm; $V(m) = \left\{ \kappa_j^m \mid j = 1, 2, ..., K_m \right\}$ is the variable set associated with m^{th} BF, ψ_m. $\boldsymbol{t}^m = \left(t_{m_1}, t_{m_2}, \ldots, t_{m_{\kappa_m}} \right)^T$ represents the variables which contribute to the m^{th} BF, ψ_m.

The λ_m values are always nonnegative and used as the *penalty parameters* (m=1,2,…, M_{max}). Moreover, the α values in the following term (13) is taken as

$$D_{r,s}^{\boldsymbol{\alpha}} \psi_m(\boldsymbol{t}^m) = \frac{\partial^{|\boldsymbol{\alpha}|} \psi_m}{\partial^{\alpha_1} t_r^m \partial^{\alpha_2} t_s^m}(\boldsymbol{t}^m) \ , \tag{15}$$

$\boldsymbol{\alpha} = (\alpha_1, \alpha_2), \ |\boldsymbol{\alpha}| = \alpha_1 + \alpha_2, \ \text{where} \ \alpha_1, \alpha_2 \in \{0,1\}.$

If $\alpha_i = 2,$ the derivative $D_{r,s}^{\alpha} \psi_m(t^m)$ disappears, and by addressing indices $r < s,$ the Schwarz's Theorem can be applied.

The optimization approach to the problem takes both the accuracy and lower complexity into account. The term *accuracy* refers to the small sum of squares of errors. The tradeoff between these two terms are expressed by penalty parameters and solved by CQP.

The purpose of obtaining low complexity is explained by two ways. First, the areas where the base functions contribute to an explanation of the observations should be large. In the classification view, the classes should be large and this is obtained by the *flat* model. This flat model is defined as the linear combination of BFs which have small residual errors. This means the model is moved from the coordinate axes to the data points $(\bar{x}_i, \bar{y}_i)(i = 1,2,,n)$. The aim is to dampen the coefficients of the BFs, θ_m, while making no change in the goodness of data fitting. The second approach is to achieve the stability of the estimation by taking care that curvatures of the model function with its compartments.

By considering (11), (13) and (14), the objective function takes the following form:

$$PRSS = \sum_{i=1}^{n}\left(\bar{y}_i - \theta_0 - \sum_{m=1}^{M}\theta_m \psi_m\left(\bar{x}_i^m\right) - \sum_{m=M+1}^{M_{\max}}\theta_m \psi_m\left(\bar{\boldsymbol{x}}_i^m\right)\right)^2 + \sum_{m=1}^{M_{\max}}\lambda_m \sum_{\substack{|\alpha|=1 \\ \alpha=(\alpha_1,\alpha_2)}}^{2} \sum_{\substack{r<s \\ r,s\in V(m)}} \int \theta_m^2 \left[D_{r,s}^{\alpha}\psi_m(\boldsymbol{t}^m)\right]^2 d\boldsymbol{t}^m, \tag{16}$$

where $\bar{\boldsymbol{x}}_i = (\bar{x}_{i,1}, \bar{x}_{i,2}, ..., \bar{x}_{i,p})^T$ represents any of the independent variables and $\bar{\boldsymbol{x}}_i^m = (\bar{x}_{i,\kappa_1}, \bar{x}_{i,\kappa_2}, ..., \bar{x}_{i,\kappa_{K_m}})^T$ stands for the corresponding projection vectors of $\bar{\boldsymbol{x}}_i$ onto those coordinates which contribute to the m^{th} BF, and they are related with the i^{th} output $\bar{y}_i$. Those coordinates are collected from the set V_m.

To obtain discretized form, the following changes are made. The input data which is represented by $\bar{\boldsymbol{x}}_l = (\bar{x}_{l,1}, \bar{x}_{l,2}, ..., \bar{x}_{l,p})^T$ generates a subdivision of any sufficiently large parallelpipe Q of IR^{n}.

The parallelpipe, which is represented by Q, contains all the input data, and it is expressed as

$$Q = [a_1, b_1] \times [a_2, b_2] \times ... \times [a_p, b_p] = \prod_{j=1}^{p} Q_j \,, \tag{17}$$

where $Q_j = [a_j, b_j]$, $a_j \le \bar{x}_{l,j} \le b_j$ $(j = 1,2,..., p;\ l = 1,2,..., N)$.

It is assumed that $a_j < \bar{x}_{l,j} < b_j$. For all j reordering is done on the coordinates of the input data points as: $\bar{x}_{l_1^j, j} \le \bar{x}_{l_2^j, j} \le \ldots \le \bar{x}_{l_{n1}^j, j}$ where $\theta_\sigma^j = 1,2,\ldots,n$ $(\sigma = 1,2,\ldots,n)$ $j = 1,2,..., p)$, and $\bar{x}_{l_\sigma^j, j}$ is the j^{th} component of $\bar{\boldsymbol{x}}_{l_\sigma^j}$, the $l_\sigma^{j\,\text{th}}$ input vector after reordering. Without loosing generality, it can be assumed that $\bar{x}_{l_\sigma^j, j} \neq \bar{x}_{l_\varphi^j, j}$ for all $\sigma, \varphi = 1,2,\ldots,n$ with $\sigma \neq \varphi$; i.e. $\bar{x}_{l_1^j, j} \le \bar{x}_{l_2^j, j} \le \ldots \le \bar{x}_{l_{n1}^j, j}$.

By using the previous notations, the parallelpipe is expressed as

$$Q = \bigcup_{\sigma^j=0}^{n} \prod_{j=1}^{p} \left[\bar{x}_{l_{\sigma^j}, j}, \bar{x}_{l_{\sigma^{j+1}}, j}\right] \tag{18}$$

So the PRSS takes the following form;

$$PRSS \approx \sum_{i=1}^{n}\left(y_i - \boldsymbol{\theta}^T \boldsymbol{\psi}(\bar{\boldsymbol{d}}_i)\right)^2$$
$$+ \sum_{m=1}^{M_{\max}} \lambda_m \sum_{\substack{|\boldsymbol{\alpha}|=1 \\ \boldsymbol{\alpha}=(\alpha_1,\alpha_2)^T}}^{2} \sum_{r<s \; r,s\in V(m)} \sum_{(\sigma^{\kappa_j})} \theta_m^2 \left[D_{r,s}^{\boldsymbol{\alpha}} \psi_m(\bar{x}_{l_{\sigma^{\kappa_j^m},\kappa_j^m}}, \bar{x}_{l_{\sigma^{\kappa_j^m},\kappa_j^m}}, ..., \bar{x}_{l_{\sigma^{\kappa_j^m},\kappa_j^m}}) \right]^2 \prod_{j=1}^{K_m} \left(\bar{x}_{l_{\sigma^{\kappa_j^m+1},\kappa_j^m}} - \bar{x}_{l_{\sigma^{\kappa_j^m},\kappa_j^m}} \right),$$

where $\left(\sigma^{\kappa_j}\right)_{j\in\{1,2,\ldots,K_m\}} \in \{0,1,2,\ldots,n+1\}^{K_m}$. (19)

The following notations related with (σ^{κ_i}) are defined in order to use in the forward steps:

$$\hat{\boldsymbol{x}}_i^m = \left(\bar{x}_{l_{\sigma^{\kappa_j^m},\kappa_j^m}}, \bar{x}_{l_{\sigma^{\kappa_j^m},\kappa_j^m}}, \ldots, \bar{x}_{l_{\sigma^{\kappa_j^m},\kappa_j^m}} \right), \tag{20}$$

$$\Delta\hat{\boldsymbol{x}}_i^m = \prod_{j=1}^{K_m} \left(\bar{x}_{l_{\sigma^{\kappa_j^m+1},\kappa_j^m}} - \bar{x}_{l_{\sigma^{\kappa_j^m},\kappa_j^m}} \right). \tag{21}$$

The approximation to the PRSS is defined as

$$PRSS \approx \sum_{i=1}^{n}\left(y_i - \boldsymbol{\theta}^T \boldsymbol{\psi}(\bar{\boldsymbol{d}}_i)\right)^2 + \sum_{m=1}^{M_{\max}} \lambda_m \theta_m^2 \sum_{i=1}^{(N+1)^{K_m}} \left(\sum_{\substack{|\boldsymbol{\alpha}|=1 \\ \boldsymbol{\alpha}=(\alpha_1,\alpha_2)^T}}^{2} \sum_{r<s \; r,s\in V(m)} \left[D_{r,s}^{\boldsymbol{\alpha}} \psi_m(\hat{\boldsymbol{x}}_i^m) \right]^2 \right) \Delta\hat{\boldsymbol{x}}_i^m. \tag{22}$$

To simplify (22), the PRSS is redefined as follows:

$$PRSS \approx \left\| y - \psi(\bar{d})\theta \right\|_2^2 + \sum_{m=1}^{M_{\max}} \lambda_m \sum_{i=1}^{(n+1)^{K_m}} L_m^2 \theta_m^2, \tag{23}$$

where $\boldsymbol{\psi}(\bar{\boldsymbol{d}}) = \left(\boldsymbol{\psi}(\bar{\boldsymbol{d}}_1), ..., \boldsymbol{\psi}(\bar{\boldsymbol{d}}_N)\right)^T$ is a matrix with dimensions of $\left(n \times \left(M_{\max} + 1\right)\right)$, and $\|\cdot\|_2$ denotes the Euclidean norm and the numbers L_{im} are defined as

$$L_{im} = \left[\left(\sum_{\substack{|\boldsymbol{\alpha}|=1 \\ \boldsymbol{\alpha}=(\alpha_1,\alpha_2)^T}}^{2} \sum_{\substack{r<s \\ r,s \in V(m)}} \left[D_{r,s}^{\boldsymbol{\alpha}} \psi_m(\hat{\boldsymbol{x}}_i^m) \right]^2 \right) \Delta \hat{\boldsymbol{x}}_i^m \right]^{1/2}. \tag{24}$$

The first parts of the equations (22) and (23) are equal since it can be written as

$$\sum_{i=1}^{n} \left(y_i - \boldsymbol{\theta}^T \boldsymbol{\psi}\left(\bar{\boldsymbol{d}}_i\right)^2 \right) = \left\| \boldsymbol{y} - \boldsymbol{\psi}\left(\bar{\boldsymbol{d}}\right)\boldsymbol{\theta} \right\|_2^2. \tag{25}$$

3.2.2. Applying Tikhonov Regularization

The linear systems of equations $\boldsymbol{y} = \boldsymbol{\psi}(\bar{\boldsymbol{d}})\boldsymbol{\theta}$ can be solved approximately by using the PRSS. The problem is classified as ill-posed, which means irregular or unstable. Thus, *Tikhonov regularization problem* is considered for the PRSS problem since it is the most widely used method for converting the ill-posed problems to well-posed (regular or stable) ones.

When (23) is considered again, the PRSS can be written as

$$\begin{aligned} PRSS &\approx \left\| \boldsymbol{y} - \boldsymbol{\psi}(\bar{\boldsymbol{d}})\boldsymbol{\theta} \right\|_2^2 + \sum_{m=1}^{M_{\max}} \lambda_m \sum_{i=1}^{(n+1)^{K_m}} L_{im}^2 \, \theta_m^2 \\ &= \sum_{m=1}^{M_{\max}} \lambda_m \left[\left(L_{1m}\theta_m\right)^2 + \left(L_{2m}\theta_m\right)^2 + ... + \left(L_{(n+1)^{K_m} m}\theta_m\right)^2 \right] \text{ where} \\ &= \left\| \boldsymbol{y} - \boldsymbol{\psi}(\bar{\boldsymbol{d}})\boldsymbol{\theta} \right\|_2^2 + \sum_{m=1}^{M_{\max}} \lambda_m \left\| \boldsymbol{L}_m \theta_m \right\|_2^2, \end{aligned}$$

$$\boldsymbol{L}_m := \left(L_{1m}, L_{2m}, ..., L_{(n+1)^{K_m}, m}\right)^T \ (m = 1, 2, ..., M_{\max}). \tag{26}$$

Tikhonov regularization needs one λ parameter. However, in Equation 26, there is a vector of parameters, i.e. $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_{M_{\max}})^T$. To overcome this problem, same λ value for each derivative term is defined as $\lambda_1 = \lambda_2 = \ldots = \lambda_{M_{\max}} = \lambda$.

Then, the PRSS is expressed as

$$PRSS \approx \left\| \boldsymbol{y} - \boldsymbol{\psi}(\bar{\boldsymbol{d}})\boldsymbol{\theta} \right\|_2^2 + \lambda \left\| \boldsymbol{L}\boldsymbol{\theta} \right\|_2^2 \tag{27}$$

where $\boldsymbol{L}$ is an $(M_{\max}+1)\times(M_{\max}+1)$-diagonal matrix with first column $\boldsymbol{L}_0 = \boldsymbol{0}_{(n+1)^{K_m}}$ and the other columns being the vectors $\boldsymbol{L}_m$ introduced above. Here, θ, with the dimensions of $((M_{Max}+1)\times 1)$, is a vector contains parameters to be estimated.

After these modifications, the problem becomes a Tikhonov regularization problem with $\varphi > 0$, i.e. $\lambda = \varphi^2$ for some $\varphi \in R$.

Tikhonov regularization approach tries to find solutions to minimize the two objective functions: $\left\| \boldsymbol{y} - \boldsymbol{\psi}(\bar{\boldsymbol{d}})\boldsymbol{\theta} \right\|_2^2$ and $\left\| \boldsymbol{L}\boldsymbol{\theta} \right\|_2^2$. Thus, this is a multiobjective problem. Tikhonov regularization approach combines these two objective functions into a single functional form by using weighted linear sum of the functions with weights defined by λ.

3.2.3. An Alternative Approach: The Conic Quadratic Programming (CQP)

The PRSS can be taken from the view point of CQP, a technique used for continuous optimization. Thus, the Tikhonov regularization problem can be formulated again by using the CQP. The optimization problem below is considered by putting an appropriate bound, *M*.

$$\begin{aligned} &\min \quad \left\| \boldsymbol{\psi}(\bar{\boldsymbol{d}})\boldsymbol{\theta} - \boldsymbol{y} \right\|_2^2 \\ &\text{subject to} \quad \left\| \boldsymbol{L}\boldsymbol{\theta} \right\|_2^2 \le M. \end{aligned} \tag{28}$$

Here, the LS objective function is tried to be minimized subject to the inequality constraint function. To obtain feasible solution, this constraint function should be nonnegative.

By adding a new variable and taking the square root of the constraint function, the problem can be expressed as the following way:

$$\begin{aligned} &\min_{t,\boldsymbol{\theta}} \quad t, \\ &\text{subject to} \quad \left\|\boldsymbol{\psi}(\bar{\boldsymbol{d}})\boldsymbol{\theta}-\boldsymbol{y}\right\|_2 \le t, \\ &\qquad\qquad \left\|\boldsymbol{L}\boldsymbol{\theta}\right\|_2 \le \sqrt{M}. \end{aligned} \tag{29}$$

Thus, the problem can be expressed as a CQP problem with the following way

$$\min_{\boldsymbol{x}} \ \boldsymbol{c}^T\boldsymbol{x}, \quad \text{subject to} \ \ \left\|\boldsymbol{D}_i\boldsymbol{x}-\boldsymbol{d}_i\right\|_2^2 \le \boldsymbol{p}_i^T\boldsymbol{x}-q_i \ \ (i=1,2,...,k). \tag{30}$$

where

$$\boldsymbol{c}=(1,0_{M_{\max}+1}^T)^T,\ \boldsymbol{x}=(t,\boldsymbol{\theta}^T)^T,\ \boldsymbol{D}_1=(0_n,\boldsymbol{\psi}(\bar{\boldsymbol{d}})),\ \boldsymbol{d}_1=\boldsymbol{y},\ \boldsymbol{p}_1=(1,0,...,0)^T,\ q_1=0,$$

$$\boldsymbol{D}_2=(0_{M_{\mathrm{Max}+1}},\boldsymbol{L}),\ \boldsymbol{d}_2=\boldsymbol{0}_{\mathrm{M}_{\max}+1},\ \boldsymbol{p}_2=0_{M_{\max}+2} \ \text{ and } \ q_2=-\sqrt{M}\ .$$

The problem (Equation 29) should be reformulated to obtain the optimality condition as the following

$$\begin{aligned} &\min_{t,\boldsymbol{\theta}} \quad t, \\ &\text{such that} \quad \boldsymbol{\chi}=\begin{pmatrix}\boldsymbol{0}_n & \boldsymbol{\psi}(\bar{\boldsymbol{d}})\\ 1 & \boldsymbol{0}_{M_{\max}+1}^T\end{pmatrix}\begin{pmatrix}t\\ \boldsymbol{\theta}\end{pmatrix}+\begin{pmatrix}-\boldsymbol{y}\\ 0\end{pmatrix}, \\ &\qquad\qquad \boldsymbol{\eta}=\begin{pmatrix}\boldsymbol{0}_{M_{\max}+1} & \boldsymbol{L}\\ 0 & \boldsymbol{0}_{M_{\max}+1}^T\end{pmatrix}\begin{pmatrix}t\\ \boldsymbol{\theta}\end{pmatrix}+\begin{pmatrix}\boldsymbol{0}_{M_{\max}+1}\\ \sqrt{M}\end{pmatrix}, \\ &\qquad\qquad \boldsymbol{\chi}\in L^{n+1},\ \boldsymbol{\eta}\in L^{M_{\max}+2}, \end{aligned} \tag{31}$$

where L^{n+1}, $L^{M_{\max}+2}$ are the $(n+1)$ and $(M_{\max}+2)$ dimensional *ice-cream* (or *second-order*, or *Lorentz*) *cones*, defined by

$$L^{n+1}=\left\{\boldsymbol{x}=(x_1,x_2,...,x_{n+1})^T\in\mathbf{R}^{n+1}\mid x_{n+1}\ge\sqrt{x_1^2+x_2^2+...+x_n^2}\right\}\quad (n\ge 1). \tag{32}$$

The *dual problem* to the latter primal one is explained as below

$$
\begin{aligned}
&\max \quad (\boldsymbol{y}^T, 0)\,\boldsymbol{\omega}_1 + \left(\mathbf{0}^T_{M_{\max}+1}, -\sqrt{M}\right)\boldsymbol{\omega}_2 \\
&\text{such that} \quad \begin{pmatrix} \mathbf{0}^T_n & 1 \\ \boldsymbol{\psi}(\bar{\boldsymbol{d}})^T & \mathbf{0}_{M_{\max}+1} \end{pmatrix}\boldsymbol{\omega}_1 + \begin{pmatrix} \mathbf{0}^T_{M_{\max}+1} & 0 \\ \boldsymbol{L}^T & \mathbf{0}_{M_{\max}+1} \end{pmatrix}\boldsymbol{\omega}_2 = \begin{pmatrix} 1 \\ \mathbf{0}_{M_{\max}+1} \end{pmatrix}, \\
&\qquad \boldsymbol{\omega}_1 \in L^{n+1},\ \boldsymbol{\omega}_2 \in L^{M_{\max}+2}.
\end{aligned} \tag{33}
$$

Moreover, $(t, \boldsymbol{\theta}, \chi, \eta, \omega_1, \omega_2)$ is a *primal dual optimal solution* if and only if

$$
\begin{aligned}
&\chi = \begin{pmatrix} \boldsymbol{O}_n & \boldsymbol{\psi}(\bar{\boldsymbol{d}}) \\ 1 & \boldsymbol{O}^T_{M_{\max}+1} \end{pmatrix}\begin{pmatrix} t \\ \boldsymbol{\theta} \end{pmatrix} + \begin{pmatrix} -y \\ 0 \end{pmatrix}, \\
&\eta = \begin{pmatrix} \boldsymbol{O}_{M_{\max}+1} & 1 \\ 0 & \boldsymbol{O}^T_{M_{\max}+1} \end{pmatrix}\begin{pmatrix} t \\ \boldsymbol{\theta} \end{pmatrix} + \begin{pmatrix} \boldsymbol{O}_{M_{\max}+1} \\ \sqrt{M} \end{pmatrix}, \\
&\begin{pmatrix} \boldsymbol{O}^T_n & 1 \\ \psi(\bar{d}^T) & \boldsymbol{O}_{M_{\max}+1} \end{pmatrix}\boldsymbol{\omega}_1 + \begin{pmatrix} \boldsymbol{O}^T_{M_{\max}+1} & 0 \\ L^T & \boldsymbol{O}_{M_{\max}+1} \end{pmatrix}\boldsymbol{\omega}_2 = \begin{pmatrix} 1 \\ \boldsymbol{O}_{M_{\max}+1} \end{pmatrix}, \\
&\boldsymbol{\omega}_1^T \chi = 0,\ \boldsymbol{\omega}_2^T = 0, \\
&\boldsymbol{\omega}_1 \in L^{n+1},\ \boldsymbol{\omega}_2 \in L^{M_{\max}+2}, \chi \in L^{n+1},\ \eta \in L^{M_{\max}+2}.
\end{aligned} \tag{34}
$$

3.2.4. MOSEK: The Optimization Software Used in CMARS

In this thesis, MOSEK software is used for solving optimization problems. The package is used as an MATLAB add-on and can solve CQPs as well as linear convex quadratic, general convex and mixed integer problems.

The package can handle high-dimensional data and can be used with C/C++, JAVA, MATLAB, and Python. Moreover, MOSEK supplies an interior-point optimizer with basis identification. The package has an efficient presolver for reducing the problem size before optimization. For linear programming (LP), it provides primal and dual simplex optimizers.

MOSEK can be used with MPS, LP and XML formats for reading and writing. The package can also do sensitivity analysis for linear problems. It can solve a problem with different optimizers simultaneously with the help of concurrent optimizer.

3.3. The Bootstrap Method

The bootstrap is a resampling technique which takes samples from the original data set with replacement. It is a data-based simulation method useful for producing inferences. Even though the term "boot" is used in the computer science, the term bootstrap is different from it. The term *bootstrap* is used by Efron (1979), which is inspired from a story (Adventures of Baron Munchausen) written by Rudolph Erich Rapse. In the story, when the Baron fell to the bottom of a deep lake, he survived by "pulling himself up by his bootstraps," which hides the idea of being close to doing the impossible (Hjorth, 1994).

Two major problems in applied statistics are the determination of an estimator for a particular parameter of interest and the evaluation of its accuracy. The evaluation of accuracy is determined by the estimates of the standard error of the estimator and conducting CIs. Efron (1979) focused on these two problems when introducing the bootstrap methodology. The purpose of bootstrap is to conduct basic statistical concepts by using computer-based implementation. The application of this method is not difficult, but depends heavily on computers. Thus, they are called computer-intensive methods (Chernick, 2008).

The bootstrap is widely used for estimating bias, standard errors and parameters. The approach is generalized to solve problems in independent but not identically distributed data, dependent data, and discrimination and regression problems. The application of bootstrap includes the following:

- Estimation of standard errors and bias,
- Constructing CIs,
- Subset selection in regression,
- Classification,
- Kriging,

in the field of psychology, geology, econometrics, biology, engineering and accounting (Chernick, 2008).

The bootstrap procedure, which is graphically shown in Figure 3, can be explained with the following steps.

1. Generate a random sample (x^{*b}) of size *n* (the same sample size with the original data) from the empirical distribution with replacement.
2. Compute the value of the statistic of interest for this sample.
3. Repeat steps 1-2 *B* times (i.e. b = 1,…, B).

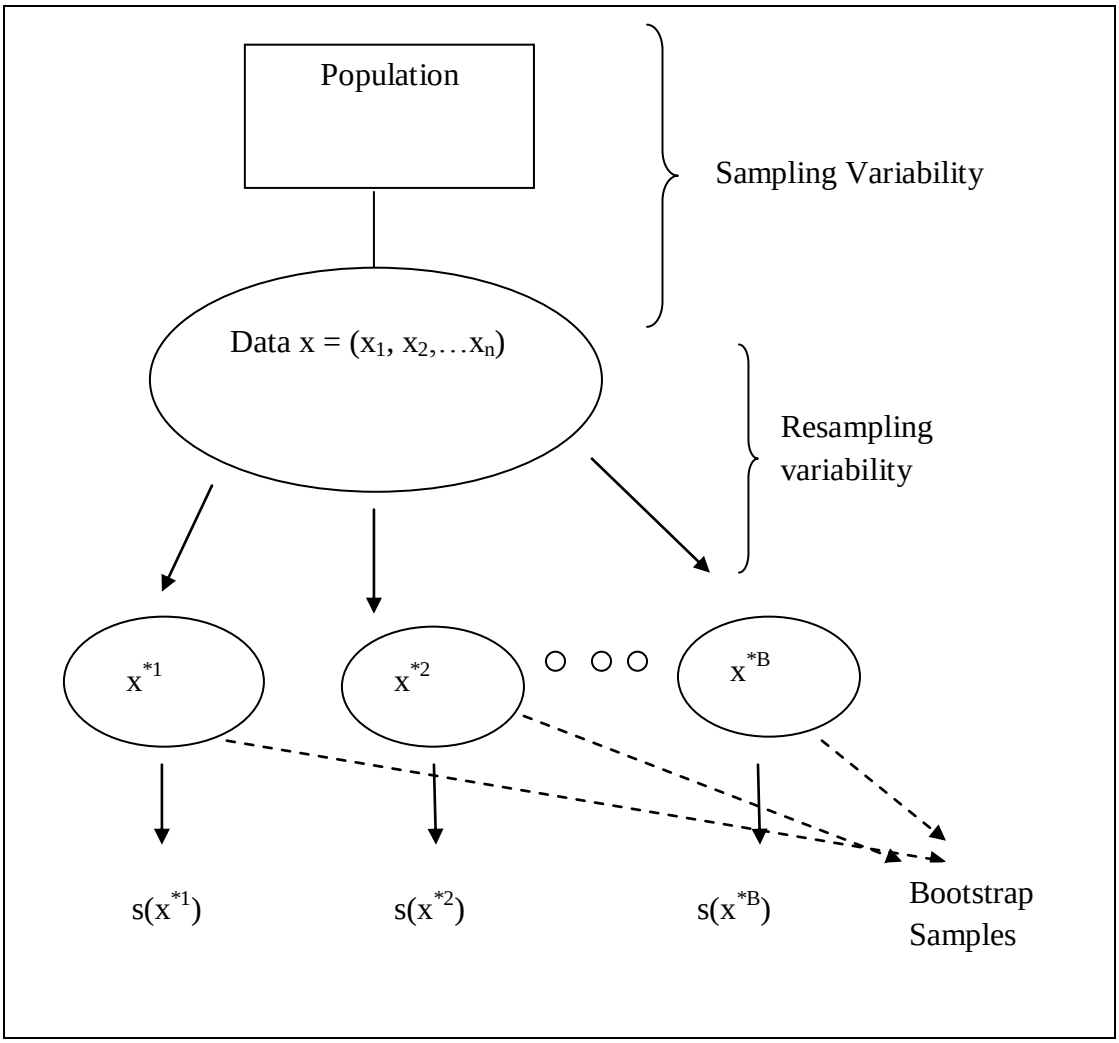

Figure 3. The Bootstrap Algorithm (Efron and Tibshirani, 1993)

Taking the observations with replacement from the sample to each bootstrap sample makes the observations independent from each other. Otherwise the observations become dependent. By treating the original sample as the population, the bootstrap algorithm is defined as "the population is to the sample as the sample is to the bootstrap samples" by Fox (2002).

Let $\theta = T(F)$ be a statistic of interest and $x_1, x_2, \cdots x_n$ be the data observed of the random variables $X_1, X_2, \cdots X_n$ are i.i.d. F and $\chi = \{X_1, X_2, \ldots X_n\}$ denote the entire data set. The key idea is to resample from the original data. Thus, the bootstrap methodology depends on treating the sample as a population. If $\hat{F}$ denotes the empirical distribution function (will be explained in the next subsection) of the observed data, then the estimator of the parameter is defined as $\hat{\theta} = T(\hat{F})$. Statistical inference deals with estimating the distribution of $T(\hat{F})$ or *R(χ,F),* which is defined as a function of the data and its unknown distribution function. Sometimes, the distribution of this random variable cannot be derived theoretically. In these cases, bootstrap method provides an approximation to the unknown distribution based on the empirical distribution function. This method can be applied to make the analyst free from assumptions, and thus, to obtain quick solutions (Davison and Hinkley, 1997).

3.3.1. Empirical Cumulative Distribution Function (ECDF)

It can be intractable to derive the distribution of the random variable R(χ,F) or it may be unknown. In these cases, bootstrap provides an approach to obtain the distribution and to make inferences. Use of the *Empirical Cumulative Distribution Function (ECDF)* is defined as one of the main approaches in statistical inference by Gentle (2009).

Suppose a researcher is interested in the statistic $\hat{\theta} = T(\hat{F})$. Moreover, assume that it is an estimate of the population parameter $\theta = T(F)$. To derive the sampling distribution of $\hat{\theta}$, some assumptions should be satisfied. Otherwise, the bootstrap method will help to obtain the distribution empirically.

The exact distribution of $\hat{\theta}$ cannot be obtained for some cases. For these cases, the most popular approach is to obtain the asymptotic distribution of *T*. However, this method has the following drawbacks (Fox, 2002):

- If the assumptions for the population are not satisfied, then the sampling distribution for the *T* will be probably incorrect. Even though the results can be relied on, the level of accuracy will not be obtained for small samples.
- Trying to obtain the sampling distribution of the statistic can be intractable.

The bootstrap provides an empirical approach for the solution of this problem. The nonparametric bootstrap method uses the data as the population and draws samples from it with replacement. Thus, the sampling distribution of the statistic is obtained from each sample by making the researcher free from the assumptions.

In mathematical statistics, the cumulative distribution function is defined as

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(x)dx \text{ for continuous cases}$$

and (35)

$$F(a) = \sum_{x_i \le a} f(x_i) \text{ for discrete cases.}$$

When the mathematical form of the distribution is not known, the ECDF is used as the estimate of the underlying distribution. This is also known as the nonparametric estimate. If the sample is estimated from a known form, then this is called as the parametric setting. So, in the parametric form, the CDF can easily be obtained with the help of parameters.

For the nonparametric setting, the CDF is obtained by using order statistics. Let

$X_{(1)} \le X_{(2)} \le \ldots \le X_{(n)}$, be the order statistics of a sample whose size is *n*, and $X_{(i)}$ be the i^{th} order statistic. Then, the ECDF, $F_n(x)$, given in (37) represents the number of observations less than or equal to *x* divided by the sample size:

$$\hat{F}_n(x) = \begin{cases} 0; & x < X_{(1)} \\ j/n; & X_{(j)} \le x < X_{(j+1)} \\ 1; & x \ge X_{(n)}. \end{cases} \tag{36}$$

In bootstrap inference, the sources of error are defined as follows:

1. The error originated from the sample by not representing the population, which is a common problem in every branch of statistics. The only assumption for the bootstrap is that the sample is a good representative for the population.

2. The sampling error originated by not covering all bootstrap samples. The solution for this kind of error is to make the number of bootstrap samples (B) large.

3.3.2 Bootstrapping Regression

Let $Y_i = x_i^T \beta + \varepsilon_i, \quad for\, i = 1, \ldots, n$ be the usual MLR model. In the model, x_i represents the independent variables and β shows the parameters. The error terms, ε_i, are normally distributed with zero mean and constant variance. The parameters, β, are distributed normally.

If all assumptions of the model are satisfied, then the model is appropriate for the data and the results will be reliable. However, in the following cases there are some problems (Hjorth, 1994). If

- the model is non-linear,
- the statistical analysis of estimation has no direct classical solution,
- errors are not normally distributed,
- there are parameters dependent on another function.

Efron and Tibshirani (1993) indicate that bootstrap is applicable to general models including non-linearity of parameters; fitting methods different from LS approach by

giving reasonable outputs. According to them, bootstrapping regression is applicable to models that have a mathematical form in addition to models that have no mathematical solution.

The principle of bootstrap for regression models is defined as "*To compute a test, the bootstrap principle is to construct a data-generating process (DGP), called the bootstrap DPG, based on estimates on the unknown parameters and probability distributions. The distribution of the test statistic under this artificial DGP is called bootstrap distribution"* (Flachaire, 2003).

Moreover, applying bootstrap to regression modeling, a computer-based result for accuracy of the parameters is yielded. Here, "accuracy" refers to how much $\hat{\beta}$ fluctuates, if independently generated replicates of $\hat{\beta}$ observed, which is not easy to obtain in real situations. Chernick (2008) expresses that in the non-Gaussian case of errors, the probability distribution of the parameters cannot be determined. Thus, constructing CIs, prediction intervals and obtaining standard errors are not possible. The bootstrap approach provides a method for approximating the distribution of parameters through bootstrap sample estimates.

The bootstrapping regression methods are classified into two parts: Random-X and Fixed-X Resampling. Each method has its own advantages and disadvantages.

3.3.2.1 Random-X Resampling (Pairs Bootstrap)

Freedman (1981) calls it as "correlation model" since it is applicable to heteroscedastic models. This method is also known as "vector resampling" (Hjorth, 1994). It is recommended to be used when there is heteroscedasticity in the residual variance or correlation structure in the residuals, or it is suspected that some important parameters are missing in the model (Chernick, 2008).

Response variable is represented by y_i, and predictors are denoted by x_i.

Step 1: Select *B* bootstrap samples of $z_i^{'} = (y_i, x_{i1}, x_{i2}, \ldots, x_{ik}),\ i = 1, \ldots, n$

Step 2: Fit a model to the vector $z_i^{'}$ and obtain the estimates of parameters *(β)* and save them.

Step 3: Repeat this procedure B times and obtain bootstrap estimates of parameters.

The design X is assumed to be deterministic as in the usual regression approach. But this approach makes the X matrix random so the estimates will lead to the variability.

This method can be more advantages to be used in the following cases:

- If the distribution of the error terms is different for the independent variables (i.e. heteroscedasticity, skewness),
- If the non-linearity part is not well-defined,
- For large sample sizes, if the data consists influential observations, in case of heteorscedasticity or skewness.

3.3.2.2 Fixed-X Resampling (Residual Resampling)

In this model, the response values are taken as random due to the error components. Its use is recommended in case of identically distributed errors (Fox, 2002).

Step 1: Fit a model to the data and obtain the fitted values, $\hat{y}_i$ and the residuals, $\hat{\varepsilon}_i$.

Step 2: Select a bootstrap sample of residuals and add them to the fitted values. These new fitted values are now new response variables, $y_{new} = \hat{y}_i + \hat{\varepsilon}_b$.

Step 3: Fit a model to the original independent variables and new response variables. Obtain the new parameters, $y_{new} = X\beta + \varepsilon$.

Step 4: Repeat this procedure B times and collect the parameters.

This method can be more advantages to be used in the following cases:

- If there is no doubt about the adequacy of the model,
- If the predictors are considered as fixed,

- For small data sets or data with influential observations.

3.3.2.3 Wild Bootstrap

The wild bootstrap is a new approach for heteroscedastic models. According to Liu (1988), the errors of the model have two-point distribution which is called *Rademacher* distribution, and defined as follows:

$$f(x)=\begin{cases}0.5, & x=1\\ 0.5, & x=-1\\ 0, & otherwise\end{cases}. \qquad (37)$$

Step 1: Fit a model to the data and obtain the fitted values, $\hat{y}_i$, and the residuals, $\hat{\varepsilon}_i$.

Step 2: These new fitted values are now new response variables, $y_{new}=\hat{y}_i+\hat{\varepsilon}_b$, where the error distribution is the f(x) given in (38).

Step 3: Repeat this procedure B times and collect the parameters.

In the wild bootstrap, the errors are randomly assigned as 1 or -1 and attached to the fitted values.

Flachaire (2005) suggests the use of wild bootstrap instead of pairs bootstrap in case of heteroscedasticity since the simulation studies give better results.

The choice of the bootstrapping regression model depends on how well the assumptions of the model are satisfied. For instance, if the model is MLR, then the errors must be independent from the covariates and must be i.i.d. Then, the Fixed-X resampling is reliable. However, Random-X resampling is not as conservative as the Fixed-X resampling. It performs better even when the assumptions are not satisfied.

3.3.3. Bootstrap Confidence Intervals

CIs are used to determine the reliability of the parameter estimates. Generally it is defined as the

$$\Pr(f_{(\alpha/2)} \le f(T,\theta) \le f_{(1-\alpha/2)}) = 1-\alpha. \quad (38)$$

In order to be able to use this method, the distribution, *f*, must be specified.

The types of CIs are described below (Martinez and Martinez, 2002).

i. Standard Normal Interval

For a parameter θ, the standard normal CI is defined as

$$\left(\hat{\theta} - z^{(1-\alpha/2)}\hat{se}, \hat{\theta} - z^{(\alpha/2)}\hat{se}\right), \quad (39)$$

where $\hat{se}$ represents the estimated standard error.

ii. Percentile Interval

Percentile interval uses the ECDF of the bootstrap sample to find the upper and lower endpoints. It is defined as

$$\left(\hat{\theta}_B^{*(\alpha/2)}, \hat{\theta}_B^{*(1-\alpha/2)}\right). \quad (40)$$

If the bootstrap distribution has a roughly normal shape, then the percentile and standard normal CIs give closer solutions (Efron and Tibshirani, 1993). According to the Central Limit Theorem the bootstrap histogram will approach to normal shape as the sample size gets larger. However, for small samples, two intervals may give different results. Efron and Tibshirani (1993) indicate that percentile method is a computational approach for the generalization of effectiveness of the standard normal interval. The percentile interval automatically transforms data; the user need not have to know the true type of transformation. However, in standard normal interval, the user must know the correct type of transformation to construct the CI.

iii. Bootstrap-t Interval

In this CI, there is no need to make normality assumption. The construction steps of the Bootstrap-t CIs are as follows:

Step 1: For each bootstrap sample, calculate the following value

$$z_b^* = \frac{\hat{\theta}_b^* - \hat{\theta}}{s\hat{e}_b^*}. \tag{41}$$

Step 2: Calculate the α^{th} percentile of the z_b^*, which is represented by $\hat{t}^{(\alpha)}$ with the following formula

$$\alpha = \frac{\#\left\{z_b^* \leq \hat{t}^{(\alpha)}\right\}}{B}. \tag{42}$$

Step 3: Construct the CI as

$$\left(\hat{\theta} - \hat{t}^{(1-\alpha)} s\hat{e}, \hat{\theta} - \hat{t}^{(\alpha)} s\hat{e}\right). \tag{43}$$

This type of CIs is applicable mostly to location statistics.

3.3.4. Bootstrap Estimate of Bias

Bias is used to investigate the performance of a measure (Martinez and Martinez, 2002). Actually, it measures the statistical accuracy of a measure. It is defined as

$$bias(T) = E[T] - \theta, \tag{44}$$

In general, it is the difference between the expected value of a statistic and the parameter value. For bootstrap estimate of bias, the empirical distribution of the parameter is used. It is defined as the following formula.

$$\hat{bias}_B = \bar{\hat{\theta}}^* - \hat{\theta}, \tag{45}$$

where $\bar{\hat{\theta}}^* = \frac{1}{B}\sum_{b=1}^{B}(\hat{\theta}^{*b})$.

where $\bar{\hat{\theta}}^*$ is the mean of the values obtained by bootstrapping. The bias corrected estimator of a parameter is defined as

$$\tilde{\theta} = \hat{\theta} - \hat{bias}_B. \tag{46}$$

When the Equation (45) is inserted to Equation (46) the bias corrected estimate is explained by the following formula.

$$\tilde{\theta} = 2\hat{\theta} - \bar{\hat{\theta}}^{*}. \quad (47)$$

3.3.4. Cross Validation (CV) Technique and the Performance Criteria

In the comparison of models, 3-fold CV technique is used (Martinez and Martinez, 2002; Gentle, 2009). In this technique, data sets are randomly divided into three parts (folds). At each attempt, two folds (66.6% of observations) are used to develop models while the other fold (33.3% of observations) is kept to test them.

The performances of the models developed are evaluated with respect to different criteria. These include accuracy, precision, complexity, stability, robustness and efficiency. The accuracy criterion is used to measure the predictive ability of the models while precision criterion is used to determine how variable the parameter estimates are; the less variability indicates more precision. The MAE, R^2, PWI and PRESS measures are used to evaluate the models according to accuracy. On the other hand, the precision of parameter estimates are determined by their empirical CIs. Other criterion used in the comparisons is the complexity; it is measured by the MSE. Besides, the stabilities of the accuracy and complexity measures obtained from the training and test data sets are also evaluated. The definitions of these measures are placed in Appendix A. Furthermore, robustness of the measures with respect to different data sets are evaluated by considering the standard deviations of the measures. Moreover, to assess the efficiency of the models build, computational run times are utilized.

CHAPTER 4

APPLICATIONS AND RESULTS

4.1. Data Sets

In order to evaluate and compare the performances of the models developed by using the MARS, CMARS and Bootstrapping CMARS (BCMARS) methods, they are run on four different data sets. These data sets are particularly selected to observe the effects of certain data characteristics such as size and complexity on the methods' performances. Here, the size and the complexity features are represented by the sample size (*n*) of the data set and the scale (*p*), the number of variables involved in the problems, respectively.

In this study, two different sample sizes (small and medium) and scales (small and medium) are considered. The data sets are gathered from variety of sources which meet the constraints of the study (Table 1). Before conducting any analysis, preprocessing is applied to all data sets including standardization of variables and handling missing values.

Table 1. Data Sets Used in the Comparisons

		Scale (p)	
Sample Size (n)	**(n, p)**	**Small**	**Medium**
	Small	Data Set 1: Concrete Slump (CS) (103,7)	Data Set 2: Uniform Sampling (US) (160,10)
	Medium	Data Set 3: PM10 (500,7)	Data Set 4: Forest Fires (FF) (517,11)

- *Small-size and small-scale:* CS, labeled as Data Set 1, includes seven independent variables with 103 observations (Yeh, 2007).
- *Small-size and medium-scale:* US, labeled as Data Set 2, consists of 10 independent variables and 160 records (Kartal, 2007).
- *Medium-size and small-scale:* PM10, labled as Data Set 3, has seven independent variables with 500 observations (Aldrin, 2006).
- *Medium-size and medium-scale:* FF, labeled as Data Set 4, contains 11 independent variables and 517 records (Cortez and Morais, 2007).

In this study, the CV approach is use as defined in Section 3.4. Therefore, instead of the original data sizes stated above, two-third of the observations are used for training and the rest is used for testing the model.

4.2. Application of the Methods

In this study, three different packages are used to obtain the CMARS and then BCMARS models. To develop a CMARS model, first, the R package (2.10.0, R Development Core Team, Austria) is used to obtain the BFs provided from the forward step of MARS. Then, the code written in MATLAB (2009a, The MathWorks, U.S.A.) by Yerlikaya (2008) and developed further by Batmaz et al. (2010) is used to obtain CMARS models. For optimization process in CMARS, the MOSEK optimization software (6, MOSEK ApS, Denmark) which is described in the next subsection (Section 4.2.1) is utilized. Then, all computations, including bootstrap, are run using the code written in MATLAB.

CMARS replaces the backward elimination of MARS with conic quadratic optimization. It utilizes all BFs yielded from the forward part of the MARS algorithm. To fit a MARS model to a data set, the R package *"Earth"* (Milborrow, 2009) is preferred due to the lack of MARS code in MATLAB. Nevertheless, there are some limitations of this package (Milborrow, 2009). These include the followings:

- R is not capable of obtaining piecewise cubic models which means that it can build at most two-way interactions.
- This package is not capable of handling missing values.

To carry out runs automatically, a link between MATLAB and R is conducted. With this link, the commands which belong to the R package can be run in MATLAB without opening its environment. After providing these BFs as inputs to the CMARS model, the program written by Batmaz et al. (2010) is run. Note here that, in this study, this program is improved further to make its interface more user-friendly. In its current version, the program can be run automatically by only supplying the data set. Hence, by developing codes for automatization, probable mistakes in the inputting procedure are prevented. After obtaining the CMARS model, three different bootstrap approaches are applied by using this improved program.

The following steps belong to the algorithm followed for obtaining three different BCMARS models, labeled as BCMARS-1 (uses Fixed-X Resampling), BCMARS-2 (uses Random-X Resampling) and BCMARS-3 (uses Wild Bootstrap).

Step 1: The set of BFs (from the first part of the MARS algorithm) are obtained. The BFs are considered fixed and they will be used for bootstrapping.

Step 2: A CMARS Model is constructed and the optimal value of $\sqrt{M}$ is found. To achieve this, the curve of $\sqrt{RSS}$ versus norm of $L\theta$ in the log-log scale is obtained (see Figure 4). The optimal value of this curve is the corner point which is demonstrated by a red point. The selected value gives the best solution for both accuracy and complexity.

Step 3: Since there is not a distributional assumption, nonparametric bootstrap is used for the analysis.

- *BCMARS-1:* the original data is used to obtain the residuals and fitted values. The bootstrap sample of residuals are selected with replacement and added to the fitted values, so the new dependent variable is obtained. A model is

constructed by using the fixed independent variables and this new dependent variable to obtain the parameters of BFs.

- *BCMARS-2:* the bootstrap sample of the data (including independent and dependent variables) is selected. This bootstrap sample and the BFs coming from Step 1 are used to obtain the parameters of the model (including the intercept).
- *BCMARS-3:* a model is fitted to the original data and the fitted values are obtained. The bootstrap sample of errors is obtained with Rademacher distribution. The fitted values and the bootstrap sample of errors are added to obtain new response variable. A model is constructed by using the fixed independent variables and this new dependent variable to obtain the parameters.

Step 4: Step 3 is repeated 1000 times and the ECDF of each parameter is obtained.

Step 5: For the significance level taken as $\alpha = 0.1$, the percentile CI of each parameter is constructed. If this interval includes zero, the corresponding BF is removed from the model.

Step 6: The Steps 2-5 are reapplied with the remaining BFs until all the CIs of the parameters do not include zero.

The percentile method is used for conducting the CIs, since there is no know form of the distribution of parameters. Efron and Tibshirani (1993) suggest the number of bootstrap samples to be as at least 1000 to construct percentile intervals. Then, the performance measures of each model obtained in three different ways are calculated. These are named as S_1, S_2 and S_3.

- S_1 is the solution obtained by taking the corner value of graph in Figure 4.
- S_2 is the solution selected as an alternative to S_1 and shown with the blue point in the Figure 4. This point may give better performance measures than S_1.
- S_3 is the solution obtained by bias-corrected method of parameters.

Moreover, the computational run time of the methods are recorded to be compared.

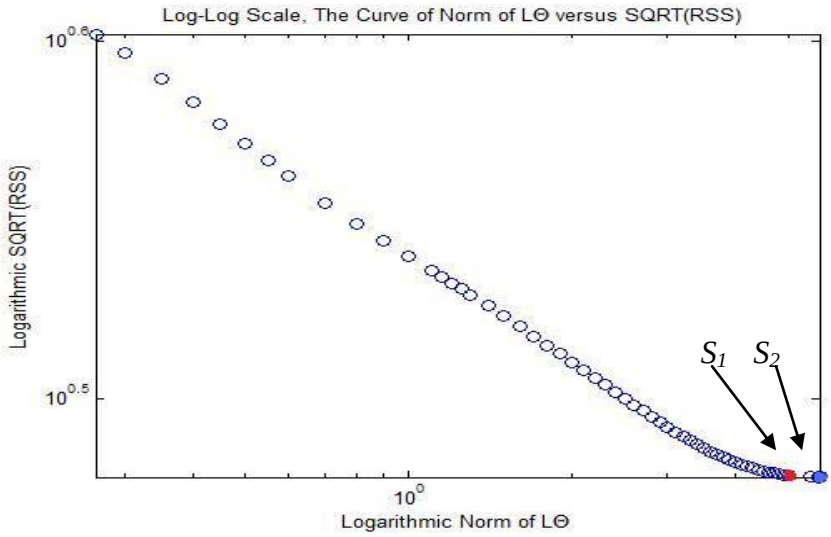

Figure 4. The plot of norm $L\theta$ versus $\sqrt{RSS}$

4.3. Results

The performance results of the MARS, CMARS, BCMARS models (BCMARS-1, BCMARS-2 and BCMARS-3, explained in the previous Chapter) with three solutions (S_1, S_2 and S_3) for training and test data sets and for the stabilities of measures are presented in Tables 2-4 for Fold 1; in Tables 5-7 for Fold 2; in Tables 8-10 for Fold 3. The performance criteria used to compare models are expressed in Appendix A. Small values of MAE, MSE and PRESS and higher values for R^2 and PWI measures indicate better performances. The comparisons are made by investigating the absolute differences. The results can be interpreted as follows:

For training data set of Fold 1 (Table 2):

- For all data sets, Random-X and Fixed-X Resampling methods produce the same results for all solutions (S_1, S_2 and S_3).
- In CS data set, Fixed-X Resampling gives the same results with the Random-X Resampling which has the best performance.
- In Forest data set, no significant BF is obtained for Random-X Resampling method.

Table 2. Performance Results of the Models Built for the Training Data Sets (Fold 1)

Data Set	Performance Measure	Training										
		MARS	CMARS	BCMARS-1			BCMARS-2			BCMARS-3		
				S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
CS	MAE	0.5746	1.4852	0.4210	0.4187*	0.4202	0.421	0.4187*	0.4202	0.5715	0.5700	0.5691
	MSE	0.4822	3.3023	0.2843	0.2775*	0.2806	0.2843	0.2775*	0.2806	0.5160	0.5134	0.5162
	R^2	0.5095	0.0139*	0.7132	0.7177	0.7157	0.7132	0.7177	0.7157	0.4771	0.4777	0.4772
	PWI	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
	PRESS	0.0023*	86.2012	0.0408	0.0706	0.2371	0.0408	0.0706	0.2371	0.1421	0.1643	0.4712
US	MAE	0.0050*	0.0652	0.0652	0.0221	0.0629	0.0652	0.0221	0.0629	0.0652	0.0221	0.0629
	MSE	0.0000*	0.0061	0.0061	0.0007	0.0056	0.0061	0.0007	0.0056	0.0061	0.0007	0.0056
	R^2	1.0000*	0.9995	0.9995	0.9999	0.9996	0.9995	0.9999	0.9996	0.9995	0.9999	0.9996
	PWI	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
	PRESS	0.0000*	0.0004	0.0004	0.0001	0.0000	0.0004	0.0001	0.0000*	0.0004	0.0001	0.0000*
PM10	MAE	0.5620	0.5510	0.5508*	0.5492	0.5498	0.5508*	0.5492	0.5498	0.5508*	0.5492	0.5498
	MSE	0.5441	0.5084	0.5095	0.5078*	0.5083	0.5095	0.5078*	0.5083	0.5095	0.5078*	0.5083
	R^2	0.4643	0.4999	0.4987	0.5000*	0.4996	0.4987	0.5000*	0.4996	0.4987	0.5000*	0.4996
	PWI	0.9913	0.9942*	0.9942*	0.9942*	0.9942*	0.9942*	0.9942*	0.9942*	0.9942*	0.9942*	0.9942*
	PRESS	0.0001*	0.0012	0.0040	0.0050	0.3791	0.0040	0.0050	0.3791	0.004	0.005	0.3791
FF	MAE	0.2258*	0.2469	0.245	0.2441	0.2453	-	-	-	0.2375	0.2375	0.4966
	MSE	0.3269	0.3025*	0.3087	0.3075	0.3082	-	-	-	0.3552	0.3552	44.398
	R^2	0.3847	0.4307*	0.4190	0.4212	0.4201	-	-	-	0.3314	0.3314	0.0001
	PWI	0.9742	0.9742	0.9742	0.9742	0.9742	-	-	-	0.9799	0.9799	0.9881*
	PRESS	0.0040	0.0008*	0.0086	0.0075	0.0312	-	-	-	0.0179	0.0179	13.184

* indicates a better performing model with respect to the corresponding performance measure

\- shows that there is not any statistically significant BF in the model

Table 3. Performance Results of the Models Built for the Testing Data Sets (Fold 1)

Data Set	Performance Measure	Testing										
		MARS	CMARS	BCMARS-1			BCMARS-2			BCMARS-3		
				S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
CS	MAE	0.5210*	1.6614	0.6055	0.6402	0.6181	0.6055	0.6402	0.6181	0.5759	0.5717	0.5717
	MSE	0.3628*	3.6901	0.5668	0.6334	0.5894	0.5668	0.6334	0.5894	0.4718	0.4812	0.4689
	R^2	0.6669*	0.0006	0.4924	0.4711	0.4844	0.4924	0.4711	0.4844	0.5421	0.5338	0.5421
	PWI	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
	PRESS	40.8108	517.2857	34.6381	33.9971	36.8989	34.6381	33.9971	36.8989	23.3453	26.2234	17.9402*
US	MAE	0.0044*	0.0660	0.0660	0.0227	0.0637	0.066	0.0227	0.0637	0.066	0.0227	0.0637
	MSE	0.0000*	0.0059	0.0059	0.0007	0.0055	0.0059	0.0007	0.0055	0.0059	0.0007	0.0055
	R^2	1.0000*	0.9996	0.9996	0.9999	0.9997	0.9996	0.9999	0.9997	0.9996	0.9999	0.9997
	PWI	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
	PRESS	0.0001	0.0001	0.0001	0.0002	0.0000*	0.0001	0.0002	0.0000*	0.0001	0.0002	0.0000*
PM10	MAE	0.6829	0.6857	0.6824	0.6839	0.6837	0.6824	0.6839	0.6837	0.6748	0.6829	0.6745*
	MSE	0.7171	0.7425	0.7375	0.7493	0.7442	0.7375	0.7493	0.7442	0.7008	0.7171	0.6995*
	R^2	0.2928	0.2772	0.2816	0.2824	0.2819	0.2816	0.2824	0.2819	0.2936	0.2928	0.2943*
	PWI	0.9936*	0.9936*	0.9936*	0.9936*	0.9936*	0.9936*	0.9936*	0.9936*	0.9936*	0.9936*	0.9936*
	PRESS	31.0120*	66.5730	61.7353	72.1223	73.0802	61.7353	72.1223	73.0802	31.0414	31.0119	24.3775
FF	MAE	0.545	0.5804	0.5582	0.5684	0.562	-	-	-	0.4947*	0.4947*	0.4966
	MSE	7.1841	7.4383	6.9455	7.5438	7.1681	-	-	-	4.4402	4.4404	4.4398*
	R^2	0.0001	0.0001	0.0002*	0.0002*	0.0002*	-	-	-	0.0001	0.0001	0.0001
	PWI	0.9881*	0.9881*	0.9881*	0.9881*	0.9881*	-	-	-	0.9881*	0.9881*	0.9881*
	PRESS	3575.863	3551.976	3506.1	3840.5	3606.5	-	-	-	2066.3	2066.4	2020.6*

* indicates a better performing model with respect to the corresponding performance measure

- shows that there is not any statistically significant BF in the model

Table 4. Stabilities of the Performance Results of the Models Built for the Data Sets (Fold 1)

Data Set	Performance Measure	Stability										
		MARS	CMARS	BCMARS-1			BCMARS-2			BCMARS-3		
				S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
CS	MAE	0.9067	0.8939	0.6953	0.6540	0.6798	0.6953	0.6540	0.6798	0.9924	0.9970*	0.9955
	MSE	0.7524	0.8949	0.5016	0.4381	0.4761	0.5016	0.4381	0.4761	0.9143	0.9373*	0.9084
	R^2	0.7640	0.0432	0.6904	0.6564	0.6768	0.6904	0.6564	0.6768	0.8801	0.8949*	0.8803
	PWI	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
	PRESS	0.0000	0.1666*	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
US	MAE	0.8800	0.9879*	0.9879*	0.9736	0.9874	0.9879*	0.9736	0.9874	0.9879*	0.9736	0.9874
	MSE	-	0.9672	0.9672	1.0000*	0.9821	0.9672	1.0000*	0.9821	0.9672	1.0000*	0.9821
	R^2	1.0000*	0.9999	0.9999	1.0000*	0.9999	0.9999	1.0000*	0.9999	0.9999	1.0000*	0.9999
	PWI	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
	PRESS	0.0009	0.2500	0.2500	0.5000*	0.0243	0.2500	0.5000*	0.0243	0.2500	0.5000*	0.0243
PM10	MAE	0.8230*	0.8036	0.8072	0.8030	0.8042	0.8072	0.8030	0.8042	0.8162	0.8042	0.8151
	MSE	0.7588*	0.6847	0.6908	0.6777	0.6830	0.6908	0.6777	0.6830	0.7270	0.7081	0.7267
	R^2	0.6306*	0.5545	0.5647	0.5648	0.5643	0.5647	0.5648	0.5643	0.5887	0.5856	0.5891
	PWI	0.9977	0.9994*	0.9994*	0.9994*	0.9994*	0.9994*	0.9994*	0.9994*	0.9994*	0.9994*	0.9994*
	PRESS	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*
FF	MAE	0.4143	0.4254	0.4389	0.4295	0.4365	-	-	-	0.4801	0.4801	1.0000*
	MSE	0.0000	0.0000	0.0000	0.0000	0.0000	-	-	-	0.0000	0.0000	1.0000*
	R^2	0.0003	0.0002	0.0005	0.0005	0.0005	-	-	-	0.0003	0.0003	1.0000*
	PWI	0.9859	0.9859	0.9859	0.9859	0.9859	-	-	-	0.9917	0.9917	1.0000*
	PRESS	0.0000	0.0000	0.0000	0.0000	0.0000	-	-	-	0.0000	0.0000	0.1533*

* indicates a better performing model with respect to the corresponding performance measure

\- shows that there is not any statistically significant BF in the model

Table 5. Performance Results of the Models Built for the Training Data Sets (Fold 2)

Data Set	Performance Measure	Training										
		MARS	CMARS	BCMARS-1			BCMARS-2			BCMARS-3		
				S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
CS	MAE	0.3662	0.3022*	0.3695	0.3661	0.3671	0.3673	0.3674	0.3666	0.5296	0.5293	0.5297
	MSE	0.2499	0.1692*	0.2402	0.2387	0.2389	0.2430	0.2400	0.2433	0.4285	0.4259	0.4292
	R^2	0.7387	0.8235*	0.7494	0.7504	0.7503	0.7472	0.7490	0.7465	0.5545	0.5547	0.5544
	PWI	1.0000*	0.9859	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
	PRESS	0.0178	0.0031*	0.0170	0.0184	0.0119	0.0087	0.0069	0.1253	0.0193	0.0222	0.8210
US	MAE	0.0045*	0.0045*	0.0045*	0.0045*	0.0045*	0.0045*	0.0045*	0.0045*	0.0953	0.0877	0.0953
	MSE	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0216	0.0173	0.0216
	R^2	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	0.9833	0.9833	0.9833
	PWI	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	0.9722	0.9722	0.9722
	PRESS	0.0004	0.0000*	0.0000*	0.0003	0.0274	0.0000*	0.0003	0.0274	0.0074	0.0077	0.0074
PM10	MAE	0.6076	0.5917*	0.6034	0.6051	0.6038	0.5989	0.5946	0.5995	0.6243	0.6235	0.6237
	MSE	0.5858	0.5585*	0.5845	0.5815	0.5826	0.5718	0.5637	0.5715	0.6183	0.6161	0.6168
	R^2	0.4382	0.4644*	0.4402	0.4424	0.4415	0.4519	0.4594	0.4524	0.4078	0.4091	0.4088
	PWI	1.0000*	1.0000*	0.9971	1.0000*	0.9971	0.9971	0.9971	0.9971	0.9912	0.9941	0.9912
	PRESS	0.0021	0.0023	2.5200	0.0016*	0.1414	9.8200	0.0038	0.4475	0.0688	0.0522	0.6029
FF	MAE	0.4144	0.4119	0.4171	0.4174	0.417	0.3990*	0.3990*	0.4026	0.4139	0.4144	0.4152
	MSE	0.8298	0.8143*	0.8229	0.8227	0.8228	11.5770	11.5770	11.5940	0.8300	0.8298	0.8298
	R^2	0.4004	0.4115*	0.4053	0.4055	0.4054	0.1634	0.1634	0.1622	0.4002	0.4004	0.4003
	PWI	0.9859	0.9859	0.9859	0.9859	0.9859	0.9915	0.9915	0.9887*	0.9887*	0.9859	0.9831
	PRESS	0.0054	0.0000*	0.0056	0.0048	0.283	0.1437	0.1437	0.3102	0.0017	0.0012	0.0011

* indicates a better performing model with respect to the corresponding performance measure

\- shows that there is not any statistically significant BF in the model

Table 6. Performance Results of the Models Built for the Testing Data Sets (Fold 2)

Data Set	Performance Measure	Testing										
		MARS	CMARS	BCMARS-1			BCMARS-2			BCMARS-3		
				S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
CS	MAE	0.8005	0.9002	0.7526	0.7640	0.7587	0.7779	0.7819	0.7683	0.7386	0.7331*	0.7381
	MSE	1.1056	1.4768	1.0608	1.0990	1.0780	1.0449	1.0802	0.9714	0.7124	0.7081*	0.7092
	R^2	0.2567	0.1820	0.2804	0.2821	0.2834	0.2639	0.2761	0.2862	0.3567	0.3654*	0.3543
	PWI	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	PRESS	23.9484	1.3163*	7.5776	7.6102	7.3092	28.3008	27.9363	19.5066	53.4781	52.9454	46.6092
US	MAE	0.0052*	0.0052*	0.0052*	0.0052*	0.0052*	0.0052*	0.0052*	0.0052	0.1002	0.0877	0.0897
	MSE	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0228	0.0173	0.0187
	R^2	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	0.9807	0.9833	0.9807
	PWI	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	0.9423	0.9722	0.9423
	PRESS	0.3000	0.3000	0.3000	0.3000	0.3000	0.3000	0.3000	0.3000	0.1318	0.0077*	0.0107
PM10	MAE	0.5283*	0.6547	0.6352	0.6518	0.6422	0.6424	0.6520	0.6451	0.6301	0.639	0.6332
	MSE	0.4927*	0.7402	0.6931	0.7337	0.7094	0.7112	0.7349	0.7183	0.6773	0.6998	0.6872
	R^2	0.4569*	0.3101	0.3109	0.3026	0.308	0.3116	0.3071	0.3056	0.2993	0.2941	0.2974
	PWI	0.9938*	0.9876	0.9876	0.9814	0.9876	0.9938*	0.9876	0.9876	0.9876	0.9876	0.9876
	PRESS	61.0080	8.6709	8.9899	7.3399	9.2137	7.1697	7.7663	8.2915	0.3726	0.3698*	1.4983
FF	MAE	0.7058	0.6844	0.6876	0.6857	0.6826	0.4134*	0.4134*	0.4239	0.7080	0.7058	0.7046
	MSE	2.4708	2.4061	2.3898	2.3835	2.3620	0.7112*	0.7112*	0.7369	24.7970	24.7080	24.5750
	R^2	0.0008	0.0006	0.0006	0.0007	0.0007	0.0006	0.0006	0.0005	0.0008	0.0008	0.0009*
	PWI	0.9632*	0.9632*	0.9632*	0.9632*	0.9632*	0.9632*	0.9632*	0.9693*	0.9632*	0.9632*	0.9632*
	PRESS	499.0300	452.680	482.9513	485.6970	497.3172	323.5600	323.559*	326.292	480.541	485.044	493.796

* indicates a better performing model with respect to the corresponding performance measure

- shows that there is not any statistically significant BF in the model

Table 7. Stabilities of the Performance Results of the Models Built for the Data Sets (Fold 2)

Data Set	Performance Measure	Stability										
		MARS	CMARS	BCMARS-1			BCMARS-2			BCMARS-3		
				S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
CS	MAE	0.4575	0.3357	0.4910	0.4792	0.4839	0.4722	0.4699	0.4772	0.7170	0.7220	0.7177*
	MSE	0.0000	0.0000	0.0000	0.0002	0.0002	0.0000	0.0000	0.2505	0.6015	0.6015	0.6052*
	R^2	0.3475	0.2210	0.3742	0.3759	0.3777	0.3532	0.3686	0.3834	0.6433	0.6587	0.6391
	PWI	1.0000*	0.9859	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
	PRESS	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*
US	MAE	0.8654	0.8654	0.8654	0.8654	0.8654	0.8654	0.8654	0.8654	0.9511	1.0000*	0.9412
	MSE	-	-	-	-	-	-	-	-	0.9474	1.0000*	0.8657
	R^2	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	0.9974	1.0000*	0.9974
	PWI	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	0.9692	1.0000	0.9692
	PRESS	0.0014	0.0000	0.0000	0.0013	0.0913	0.0000	0.0013	0.0913	0.0561	1.0000	0.6916
PM10	MAE	0.8695	0.9038	0.9499	0.9284	0.9402	0.9323	0.9120	0.9293	0.9908*	0.9757	0.9850
	MSE	0.8411	0.7545	0.8433	0.7926	0.8213	0.8040	0.7670	0.7956	0.9129*	0.8804	0.8976
	R^2	0.9591*	0.6677	0.7063	0.6840	0.6976	0.6895	0.6685	0.6755	0.7339	0.7189	0.7275
	PWI	0.9938	0.9876	0.9905	0.9814	0.9905	0.9967	0.9905	0.9905	0.9964*	0.9935	0.9964*
	PRESS	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0000	0.0000	0.1846*	0.1412	0.0000
FF	MAE	0.5871	0.6018	0.6066	0.6087	0.6109	0.9652*	0.9652*	0.9498	0.5846	0.5871	0.5893
	MSE	0.0000	0.0000	0.0000	0.0000	0.0003*	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000
	R^2	0.0020	0.0015	0.0015	0.0017	0.0017	0.0037	0.0037	0.0037	0.0020	0.0020	0.0022*
	PWI	0.9770	0.9770	0.9770	0.9770	0.9770	0.9715	0.9715	0.9742	0.9742	0.9770	0.9798*
	PRESS	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*

* indicates a better performing model with respect to the corresponding performance measure

\- shows that there is not any statistically significant BF in the model

Table 8. Performance Results of the Models Built for the Training Data Sets (Fold 3)

Data Set	Performance Measure	Training										
		MARS	CMARS	BCMARS-1			BCMARS-2			BCMARS-3		
				S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
CS	MAE	0.4507	0.2750*	0.3247	0.3234	0.3242	0.4385	0.4385	0.4387	0.7519	0.7519	0.7520
	MSE	0.3317	0.1341*	0.1787	0.1776	0.1778	0.3141	0.3141	0.3147	0.8229	0.8229	0.8229
	R^2	0.6763	0.8692*	0.8258	0.8267	0.8265	0.6935	0.6935	0.6932	0.1971	0.1971	0.1971
	PWI	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
	PRESS	0.0812	0.0017*	0.0244	0.0183	0.0143	0.0384	0.0384	0.8456	0.0324	0.0324	0.0323
US	MAE	0.0045*	0.0123	0.0123	0.0045*	0.0119	0.0123	0.0045*	0.0119	0.0916	0.0936	0.0868
	MSE	0.0000*	0.0003	0.0003	0.0000*	0.0003	0.0003	0.0000*	0.0003	0.0208	0.0182	0.0187
	R^2	1.0000*	0.9998	0.9998	1.0000*	0.9998	0.9998	1.0000*	0.9998	0.9810	0.9810	0.9810
	PWI	1.0000*	0.9519	0.9519	1.0000*	0.9519	0.9519	1.0000*	0.9519	0.9519	0.9519	0.9519
	PRESS	0.0000*	0.0001	0.0001	0.0000*	0.0001	0.0001	0.0000*	0.0001	0.0075	0.0094	0.0085
PM10	MAE	0.5764	0.5743	0.5741	0.5710*	0.5722	0.5761	0.5738	0.5752	0.9007	0.9135	0.9067
	MSE	0.5354	0.5159	0.5165	0.5149*	0.5156	0.5223	0.5211	0.5222	17.1420	17.7520	17.4110
	R^2	0.4241	0.4452	0.4446	0.4462*	0.4455	0.4384	0.4395	0.4385	0.1295	0.1299	0.1305
	PWI	0.9937	0.9969*	0.9937	0.9937	0.9937	0.9969*	0.9969*	0.9937	0.9748	0.9748	0.9748
	PRESS	0.0057	1.4500	0.0011*	0.0023	0.6961	3.7300	3.2400	0.2510	$1.62*10^8$	$1.75*10^8$	$1.67*10^8$
FF	MAE	0.3518	0.3231	-	-	-	0.3161*	0.3161*	0.3175	0.3161*	0.3161*	0.3175
	MSE	0.8932*	0.9756	-	-	-	10.6550	10.6550	10.6550	10.6550	10.6550	10.6550
	R^2	0.1692*	0.1350	-	-	-	0.0089	0.0089	0.0089	0.0089	0.0089	0.0089
	PWI	0.9879	0.9940*	-	-	-	0.9940*	0.9940*	0.9940*	0.9940*	0.9940*	0.9940*
	PRESS	0.0012	0.0069	-	-	-	0.0011*	0.0011*	0.3385	0.0011*	0.0011*	0.3385

* indicates a better performing model with respect to the corresponding performance measure

\- shows that there is not any statistically significant BF in the model

Table 9. Performance Results of the Models Built for the Testing Data Sets (Fold 3)

Data Set	Performance Measure	Testing										
		MARS	CMARS	BCMARS-1			BCMARS-2			BCMARS-3		
				S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
CS	MAE	0.6455*	0.6868	0.6603	0.6589	0.6602	0.6561	0.6561	0.6514	0.7644	0.7644	0.7643
	MSE	0.6440*	0.8351	0.7515	0.7611	0.7563	0.6643	0.6643	0.6607	0.9253	0.9253	0.9251
	R^2	0.3447*	0.2715	0.2865	0.2836	0.2854	0.3215	0.3215	0.323	0.0393	0.0393	0.0393
	PWI	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
	PRESS	73.882	217.44	168.11	179.66	173.43	58.085*	58.087	63.251	70.747	70.747	70.665
US	MAE	0.0052*	0.0137	0.0137	0.0052*	0.0132	0.0137	0.0052*	0.0132	0.098	0.0972	0.0923
	MSE	0.0000*	0.0004	0.0004	0.0000*	0.0004	0.0004	0.0000*	0.0004	0.0244	0.0203	0.0215
	R^2	1.0000*	0.9998	0.9998	1.0000*	0.9998	0.9998	1.0000*	0.9998	0.9814	0.9814	0.9814
	PWI	1.0000*	0.9821	0.9821	1.0000*	0.9821	0.9821	1.0000*	0.9821	0.9821	0.9821	0.9821
	PRESS	0.0070	0.0035*	0.0035*	0.0070	0.0035*	0.0035*	0.0070	0.0035*	0.8626	1.7572	1.3207
PM10	MAE	0.6910*	0.7039	0.7013	0.7011	0.7007	0.7025	0.7019	0.7004	0.9041	0.9159	0.9103
	MSE	0.7991	0.8153	0.8136	0.8153	0.8134	0.7974	0.7983*	0.7945*	16.023	16.563	16.281
	R^2	0.2975	0.2855	0.2872	0.2864	0.2883	0.2995*	0.2988	0.3025	0.1011	0.1007	0.101
	PWI	0.9945*	0.9945*	0.9945*	0.9945*	0.9945*	0.9945*	0.9945*	0.9945*	0.9780	0.9780	0.9780
	PRESS	594.95*	718.30	717.10	754.91	757.70	605.28	630.36	610.05	1730	1885	1790
FF	MAE	0.3518	0.3231	-	-	-	0.2929*	0.2929*	0.2947	0.2929*	0.2929*	0.2947
	MSE	0.8932	0.9756	-	-	-	0.8510*	0.8510*	0.8511	0.8510*	0.8510*	0.8511
	R^2	0.1692*	0.1350	-	-	-	0.0112	0.0112	0.0112	0.0112	0.0112	0.0112
	PWI	0.9879	0.9940*	-	-	-	0.9892	0.9892	0.9892	0.9892	0.9892	0.9892
	PRESS	0.0012*	0.0069	-	-	-	10.257	10.250	12.212	10.257	10.250	12.212

* indicates a better performing model with respect to the corresponding performance measure

\- shows that there is not any statistically significant BF in the model

Table 10. Stabilities of the Performance Results of the Models Built for the Data Sets (Fold 3)

Data Set	Performance Measure	Stability									
		MARS	CMARS	BCMARS-1			BCMARS-2			BCMARS-3		
				S_1	S_2	S_3	S_1	S_2	S_3	S_1	S_2	S_3
CS	MAE	0.6982	0.4004	0.4917	0.4908	0.4911	0.6683	0.6683	0.6735	0.6735	0.9836*	0.9836*
	MSE	0.5151	0.1606	0.2378	0.2333	0.2351	0.4728	0.4728	0.4763	0.4763	0.8893*	0.8893*
	R^2	0.5097*	0.3124	0.3469	0.3431	0.3453	0.4636	0.4636	0.466	0.466	0.1994	0.1994
	PWI	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*
	PRESS	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*
US	MAE	0.8654	0.8978	0.8978	0.8654	0.9015	0.8978	0.8654	0.9015	0.9015	0.9347	0.9424*
	MSE	-	0.7500	0.7500	-	0.7500	0.7500	-	0.7500	0.7500	0.8525	0.9760*
	R^2	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	1.0000*	0.9996	0.9996
	PWI	1.0000*	0.9693	0.9693	1.0000*	0.9693	0.9693	1.0000*	0.9692	0.9693	0.9692	0.9693
	PRESS	0.0001	0.0286*	0.0286*	0.0011	0.0286*	0.0286*	0.0011	0.0286*	0.0286*	0.0087	0.0000
PM10	MAE	0.8342	0.8159	0.8186	0.8144	0.8166	0.8201	0.8175	0.8212	0.8212	0.9962*	0.9834
	MSE	0.6700	0.6328	0.6348	0.6315	0.6339	0.6550	0.6528	0.6573	0.6573	0.9347	0.9662*
	R^2	0.7015	0.6413	0.6460	0.6419	0.6471	0.6832	0.6799	0.6899	0.6899	0.7807*	0.7776
	PWI	0.9992*	0.9976	0.9992*	0.9992*	0.9992*	0.9976	0.9976	0.9992*	0.9992*	0.9967	0.9967
	PRSS	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*	0.0000*
FF	MAE	1.0000*	1.0000*	-	-	-	0.9266	0.9266	0.9282	0.9282	0.9266	0.9266
	MSE	1.0000*	1.0000*	-	-	-	0.0001	0.0001	0.0001	0.0000	0.0001	0.0001
	R^2	1.0000*	1.0000*	-	-	-	0.7946	0.7946	0.7946	0.0000	0.7946	0.7946
	PWI	1.0000*	1.0000*	-	-	-	0.9952	0.9952	0.9952	0.0000	0.9952	0.9952
	PRESS	1.0000*	1.0000*	-	-	-	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

* indicates a better performing model with respect to the corresponding performance measure

\- shows that there is not any statistically significant BF in the model

- CMARS and S_1 of the Fixed-X, Random-X Resampling and Wild bootstrap methods give the same results for US data set.
- In CS data, the best MSE values are obtained for the S_2 solution of Fixed-X and Random-X models; while in PM10 data, they are obtained for S_2 solution of all three BCMARS methods.
- In PM10 data, S_2 solutions of the three BCMARS methods provide the best method with respect to all measures (except MAE) considered.

For test data set of Fold 1 (Table 3):

- For CS and US data, MARS method gives the best results with respect to all measures except for PRESS. In US data set, S_1 solutions of Fixed-X and Random-X Resampling models yield the same results with that of CMARS.
- In PM10 and FF data sets, S_3 solution of Wild bootstrapping method is better than the other methods in terms of most of the measures.
- In PM10 and Forest data, the smallest MSE value occurred in the S_3 of the Wild bootstrap model.

In Fold 1, S_2 of Wild bootstrapping provides the most stable model for CS data (Table 4). Stability is explained in Appendix A in Equation (55). In US data, however, S_2 of the Fixed-X, Random-X Resampling and Wild bootstrap models indicate the highest stability for all measures. For PM10, each method has closer stability values, but MARS is more stable than the other models in terms of MAE, R^2 and MSE. For Forest data set, the most stable measures are obtained by S_3 of Wild bootstrapping.

For training data set of Fold 2 (Table 5):

- In general, Wild bootstrapping does not perform good at all.
- In US data, all methods, except for Wild bootstrap, produce the same performance results.
- CMARS produces the best model for PM10 data set for all performance measures.

- The smallest MAE and the largest PWI in FF data belongs to S_1 and S_2 solutions of Random-X Resampling model. However, CMARS gives the best performance with respect to the other measures.

For testing data of Fold 2 (Table 6):

- In CS data set, Wild bootstrap method yields the best MAE, MSE and R^2 values. Fixed-X is better than the other two bootstrap methods in terms of PRESS values.
- In US data, all methods, except for Wild bootstrap, give the same results.
- Wild bootstrap is better than CMARS in terms of MAE and MSE value for PM10 data. However, this method yields the worst result with respect to R^2. Moreover, MARS is the best method in this data set for all performance measures.
- In FF data, Random-X Resampling produces the best values for the MAE, MSE, PWI and PRESS measures.

The most stable performance measures for Fold 2, except for R^2, belong to Wild bootstrapping models (S_3) of CS data (Table 7). In the US data, S_2 solution of Wild bootstrapping gives the most stable measures among the others. The MSE, MAE and PRESS measures of the PM10 data set are the most stable ones for Wild bootstrapping. However, the best value of stability measure for R^2 belongs to MARS. In FF data, Random-X Resampling is the most stable method with respect to MAE, while Fixed-X Resampling is the most stable model for MSE. For the PWI and R^2 measures, Wild bootstrapping results in the best model in terms of stability. All methods are not stable at all with respect to PRESS measure.

For training data set of Fold 3 (Table 8):

- CMARS gives the best results for all performance measures in CS data. Following CMARS, the Fixed-X Resampling method yields the best results among all bootstrap methods.

- For US data, Fixed-X and Random-X models produce the same results with MARS.
- In PM10 data set, Wild bootstrapping does not yield good performance with respect to the MAE and MSE measures, whereas other methods give results similar to the best model, which is produced by the Fixed-X Resampling method.
- Fixed-X Resampling method does not produce any significant BFs for FF data. Random-X and Wild bootstrapping yield the same and the best results with respect to most of the measures, namely MAE, PWI and PRESS.
- When all bootstrapping models are compared, Fixed-X method has better performance measures.

For test data set of Fold 3 (Table 9):

- When MSE values are considered, Random-X and Fixed-X Resampling methods perform better than CMARS. Moreover, higher R^2 values are obtained by Random-X Resampling method. However, the best values of measures for CS data set, other than that of PRESS, belong to MARS.
- For US data, best performance measures are obtained by MARS, S_2 of Fixed-X and Random-X Resampling methods. And, the performance measures of CMARS, Fixed-X and Random-X are the same.
- In PM10 data, Random-X and Fixed-X methods give better results than CMARS in terms of MSE and R^2. And, S_3 of Random-X gives the best performance with respect to three measures, namely MSE, R^2 and PWI, while MARS produce the best performance in terms of MAE, PWI and PRESS.
- For FF data, Fixed-X Resampling gives no results at all. And, Random-X Resampling and Wild bootstrap performs better in terms of MAE and MSE while MARS performs better with respect to R^2 and PRESS measures.

For the stability of the measures obtained from the CS and US data sets, it is seen that Wild bootstrap is superior to other models in all measures other than R^2 (Table 10). For most data sets, Wild bootstrap is stable in terms of MSE values.

CHAPTER 5

DISCUSSION

In this section, it is aimed to compare the performances of the methods studied, namely MARS, CMARS, BCMARS (Fixed-X Resampling, Random-X Resampling and Wild Bootstrapping) in general (Section 5.1), and also, according to different features of data sets such as size (Section 5.2) and scale (Section 5.3). In these comparisons, various criteria including accuracy, precision, stability, efficiency (Section 5.4) and robustness are considered. Note here that while calculating the means (and standard deviations) for the BCMARS method, the best solution, S_1, S_2 or S_3, with respect to the measure considered is used.

5.1. Comparison with respect to Overall Performances

The mean and standard deviations of measures obtained from four data sets are given in Table 11. These values are calculated for training and testing data sets in addition to the stability of measures. Definitions of the measures are given in Appendix A. In this table, lower means for MAE, MSE and PRESS and higher means for R^2 and PWI measures indicate better performances. On the other hand, smaller standard deviations imply robustness for the corresponding measure. The following conclusions can be drawn from this table:

For training data sets:

- Fixed-X Resampling provides best performance with respect to MAE and R^2 accuracy measures. This method is the most robust among the others with respect to the same measures. These findings are also valid with respect to the complexity measure, MSE, as well.

- MARS, however, performs best with respect to the other accuracy measures PWI and PRESS. This method is the most robust among the others with respect to the same measures.
- When the bootstrapping models are compared among themselves, the Fixed-X Resampling method overperforms with respect to the means and spreads of all measures except the spread of PWI. Random-X Resampling is the most robust one with respect to the PWI measure.

For testing data sets:

- Random-X performs best with respect to most of the measures, namely MSE, R^2 and PRESS. It also produces more robust models for the same measures. Moreover, it gives the least complex models as well by providing the smallest MSE mean value.
- MARS has the best performance with respect to the only one accuracy measure, MAE. It is also the most robust for the same measure.
- CMARS, on the other hand, is the best performing and also the most robust method in terms of PWI.
- When only the bootstrapping methods are considered, Fixed-X Resampling is the best one with respect to the performance measure MAE, and Wild bootstrapping is the most robust one for the same performance measure. Moreover, Random-X Resampling has the highest PWI coverage, and Wild bootstrapping is the most robust with respect to PWI.

For stability;

- Random-X Resampling and Wild bootstrapping methods are more stable when compared to the other methods.
- Random-X is more stable with respect to R^2 and PWI; it has the most robust stability with respect to the same measures, and also has the most robust stability with respect to the MSE.

Table 11. Overall Performances (Mean±Std. Dev.) of the Methods

Performance Measures	**Training**				
	MARS	**CMARS**	**BCMARS-1**	**BCMARS-2**	**BCMARS-3**
MAE	0.3453 ±0.2336	0.4040 ±0.3980	0.3204*±0.2260**	0.3356 ±0.2263	0.4251 ±0.2797
MSE	0.4015 ±0.3064	0.6070 ±0.9080	0.3117*±0.2700**	0.4230 ±0.3990	0.5770 ±0.4950
R^2	0.6005 ±0.2797	0.5911 ±0.3407	0.6827*±0.2492**	0.6120 ±0.3350	0.5127 ±0.3398
PWI	0.9944*±0.0082**	0.9942 ±0.0082**	0.9909 ±0.0153	0.9932 ±0.0140	0.9855 ±0.0158
PRESS	0.0097*±0.0230**	72.0000 ±248.80	0.2390 ±0.7570	1.2090 ±3.0150	$13.5x10^6$±$4.7x10^6$
Performance Measures	**Testing**				
	MARS	**CMARS**	**BCMARS-1**	**BCMARS-2**	**BCMARS-3**
MAE	0.4576*±0.2956**	0.5800 ±0.4580	0.4838±0.3076	0.6460±0.6110	0.4977±0.2998
MSE	3.0700±7.0900	1.5780 ±2.1350	1.2670±1.9970	0.5480*±0.3660**	1.0720 ±1.2710
R^2	0.4480±0.3820	0.3630±0.4030	0.4500±0.3800	0.4530*±0.3770**	0.3840±0.4010
PWI	0.9930*±0.0108	0.9930*±0.0106**	0.9884±0.0177	0.989±0.0169	0.9878±0.0120
PRESS	470±996	491±287	459±1037	107.700*±189.10**	$1.4x10^6$±$0.5x10^6$
Performance Measures	**Stability**				
	MARS	**CMARS**	**BCMARS-1**	**BCMARS-2**	**BCMARS-3**
MAE	0.7657±0.1848	0.7440±0.2383	0.7252±0.1939	0.7375±0.2870	0.8690*±0.1783**
MSE	0.5500 ±0.3710	0.5690±0.3400	0.5550±0.3450	0.6374±0.2174**	0.7616*±0.2852
R^2	0.6070±0.3680	0.4690±0.3940	0.5750±0.3640	0.6577*±0.3063**	0.6300±0.3650
PWI	0.9950*±0.0070	0.9940±0.0070	0.9940±0.0070	0.9950*±0.0060**	0.9940±0.0080
PRESS	0.0003±0.0005	0.0±0.0**	0.0100±0.0270	0.0020±0.0050	0.1000*±0.2733

*indicates better performance with respect to means; **indicates better performance with respect to spread

- Besides, Wild bootstrapping is more stable in terms of MAE, MSE and PRESS; it has the most robust stability with respect to the MAE measure only.
- CMARS has the most robust stability with respect to PRESS.

5.2. Comparison with respect to Sample Sizes

Table 12 presents the performance measures of the studied methods with respect to two sample size categories: small and medium. Depending on the results given in the table, following conclusions can be reached:

- Small training and test data sets produce better models for all measures except PRESS compared to the medium training and testing data sets.
- All methods are more stable in small data sets with respect to R^2, PWI and PRESS.
- Wild bootstrapping is more stable in small data sets with respect to all measures except PWI.
- Fixed-X method produces the lowest MAE for training small size data sets, while MARS has the best value for this measure in testing samples.
- Fixed-X produces the lowest MSE value in both small and medium sized training samples. However, MARS is the best method for the MAE in small data while Random-X is the best one in medium size testing data.
- Fixed-X method is superior to other methods in terms of R^2 for small and medium size training data sets, while MARS is the best one for testing small and medium size data sets.
- MARS and CMARS are the best methods with respect to PWI measure in both types of testing data. Both methods also perform similar with respect to the same measure in training samples.
- Random-X Resampling is the best method for the PRESS measure in small testing samples while MARS is the best model for PRESS is medium training samples. On the other hand, Fixed-X gives the best result in small training samples with respect to the same measure.

Table 12. Averages of Performance Measures with Respect to Different Sample Sizes

Sample Size	Performance Measures	Training				
		MARS	**CMARS**	**BCMARS-1**	**BCMARS-2**	**BCMARS-3**
Small	**MAE**	0.2340	0.3570	0.1899*	0.2092	0.3410
	MSE	0.1773	0.6020	0.1158*	0.1387	0.3000
	R^2	0.8208	0.7840	0.8824*	0.8596	0.7350
	PWI	1.0000*	0.9970	0.9910	0.9910	0.9870
	PRESS	0.0170	144	0.0150*	0.0140	0.0340
Medium	**MAE**	0.4563	0.4498*	0.4769	0.4874	0.5090
	MSE	0.6257	0.6125	0.5469*	0.7630	0.8540
	R^2	0.3802	0.3978	0.4431*	0.3140	0.2908
	PWI	0.9888	0.9900*	0.9890	0.9940*	0.9830
	PRESS	0.0020*	0.2440	0.5080	2.6400	$27x10^6$
Sample Size	**Performance Measures**	**Testing**				
		MARS	**CMARS**	**BCMARS-1**	**BCMARS-2**	**BCMARS-3**
Small	**MAE**	0.3300*	0.5560	0.3440	0.7280	0.3790
	MSE	0.3520*	1.0010	0.3980	0.3670	0.3570
	R^2	0.7110*	0.5760	0.6770	0.6800	0.6500
	PWI	1.0000*	1.0000*	0.9910	0.9910	0.9920
	PRESS	23.200	122.70	35.000	18.650*	22.700
Medium	**MAE**	0.5849	0.6052	0.6518	0.5468*	0.6160
	MSE	5.7800	2.1500	2.3100	0.7658*	1.7880
	R^2	0.1853*	0.1497	0.1765	0.1817	0.1178
	PWI	0.9860*	0.9860*	0.9850	0.9860*	0.9830
	PRESS	918.00	860.00	968.00	215.00*	$2.9x10^6$
Sample Size	**Performance Measures**	**Stability**				
		MARS	**CMARS**	**BCMARS-1**	**BCMARS-2**	**BCMARS-3**
Small	**MAE**	0.2250	0.7300	0.7265	0.6110	0.9359*
	MSE	0.4980	0.5770	0.5750	0.5530	0.8835*
	R^2	0.7700	0.5960	0.7350	0.7510	0.7710*
	PWI	1.0000*	0.9970	0.9990	0.9990	0.9940
	PRESS	0.0007	0.0469	0.0189	0.0040	0.1660*
Medium	**MAE**	0.4431	0.7578	0.7236	0.8888*	0.8022
	MSE	0.5760	0.5620	0.4410	0.7049*	0.6150
	R^2	0.4450	0.3410	0.3830	0.5460*	0.4890
	PWI	0.9915	0.9900	0.9898	0.9920	0.9948*
	PRESS	0.0000	0.0003	0.0000	0.0012	0.0457*

*indicates better performance with respect to the corresponding measure and sample

- In terms of the complexity measure, MSE as well as the accuracy measures MAE and R^2, Wild bootstrapping and the Random-X are the most stable methods in small and medium size data sets.
- MARS and Wild bootstrapping methods are the most stable methods in small and medium size data sets with respect to PWI, respectively.
- Wild bootstrapping is the most stable method in both size data in terms of the PRESS measure.

5.3. Comparisons with respect to Scales

In Table 13, the performance measures of the studied methods with respect to two scale types; small and medium are presented. Depending on the results given in the table, following conclusions can be drawn

- Medium scale training data sets produce better models for all methods with respect to all measures except PWI.
- Small scale testing data sets produce better models for all methods with respect to MSE and PWI measures while medium scale testing data sets produce better models with respect to MAE and R^2.
- Medium scale data sets produce more stable models for all methods for MAE, R^2 and PRESS. On the other hand, small scale data yield more stable models for MSE and PWI.
- In small scale training samples, CMARS produces similar results with Fixed-X and Random-X Resampling for MAE and MSE measures. However, in small scale testing samples, MARS, Fixed-x and Random-x yield similar values for the same accuracy measures.
- Fixed-X Resampling is the best method with respect to the complexity measure, MSE, in small and medium scale training data sets. However, MARS and Random-X are the best methods for the same measure in small and medium scale test samples, respectively.

Table 13. Averages of Performance Measures with Respect to Different Scale

Scale	Performance Measures	Training				
		MARS	**CMARS**	**BCMARS-1**	**BCMARS-2**	**BCMARS-3**
Small	**MAE**	0.5229	0.4140*	0.4720	0.4910	0.6561
	MSE	0.4572	0.8992	0.3830*	0.4040	0.7728
	R^2	0.5483	0.4985	0.6139*	0.5928	0.4078
	PWI	0.9970	0.9924	0.9980*	0.9980*	0.9934
	PRESS	0.0214*	143.66	0.4320	2.1910	2.7000
Medium	**MAE**	0.1677	0.1773	0.1384*	0.1492	0.1940
	MSE	0.3417	0.3500	0.2260*	0.4450	0.3810
	R^2	0.6591	0.6630	0.7650*	0.6340	0.6170
	PWI	0.9913	0.9920*	0.9820	0.9870	0.9770
	PRESS	0.0017	0.0010*	0.0080	0.0300	0.0060
Scale	**Performance Measures**	**Testing**				
		MARS	**CMARS**	**BCMARS-1**	**BCMARS-2**	**BCMARS-3**
Small	**MAE**	0.6696	0.5445*	0.6747	0.6776	0.7130
	MSE	0.7327*	2.3959	0.7717	0.7443	0.8469
	R^2	0.3297	0.3377*	0.3240	0.3293	0.2721
	PWI	0.9964*	0.9901	0.9960	0.9960	0.9932
	PRESS	213.00	800.69	176.94	141.72*	$2.9x10^6$
Medium	**MAE**	0.2703	0.2790	0.2550*	0.6070	0.2820
	MSE	5.4630	1.7800	1.8600	0.3130*	1.2980
	R^2	0.5107	0.5040	0.6000	0.6020*	0.4960
	PWI	0.9892	0.9900*	0.9790	0.9810	0.9820
	PRESS	1012.8	714.00	714.00	66.800*	419.00
Scale	**Performance Measures**	**Stability**				
		MARS	**CMARS**	**BCMARS-1**	**BCMARS-2**	**BCMARS-3**
Small	**MAE**	0.5008	0.7801	0.7041	0.7300	0.9200*
	MSE	0.6515	0.5771	0.5183	0.5539	0.8378*
	R^2	0.6521*	0.3714	0.5540	0.5539	0.6277
	PWI	0.9984*	0.9930	0.9977	0.9979	0.9980
	PRESS	0.0003	0.0828*	0.0005	0.0013	0.0471
Medium	**MAE**	0.7666	0.7900	0.7505	0.7470	0.8100*
	MSE	0.3474	0.5920	0.3480	0.8040*	0.6850
	R^2	0.5628	0.5310	0.6000	0.7600*	0.6330
	PWI	0.9931*	0.9930*	0.9910	0.9930	0.9920
	PRESS	0.0004	0.0460*	0.0220	0.0040	0.1650

* indicates better performance with respect to the corresponding measure and scale

- The best model in terms of R^2 is yielded by Fixed-X in both scale training samples, while CMARS and Random-X produce the best for the same measure in small and medium scale testing data, respectively.
- Fixed-X and Random-X models are superior to others in terms of PWI in small scale training samples. In medium scaled training samples, however, CMARS produces the best value for the same measure. On the

other hand, in testing samples, MARS and CMARS give the best models with respect to PWI in both small and medium scale.

- Random-X Resampling is the best model for PRESS in all testing samples. But, MARS and CMARS result in better PRESS values all training data.
- Wild bootstrap is superior to other methods with respect to the stability of MAE in all type data sets.
- MARS seems more stable in terms of PWI in type data sets.
- Wild bootstrapping is superior to other methods with respect to the stability of MSE, the complexity measure, in small scaled while Random-X Resampling is the best method in medium scaled data with respect to the stability of the same measure.
- MARS and Random-X are the most stable in small scale data and medium scale data, respectively.
- CMARS is the most stable method with respect to the PRESS measure for both scales of data.
- The most stable model in small and medium scale data in terms of R^2 are MARS and Random-X Resampling, respectively.

5.4. Evaluation of the Efficiencies

The elapsed time of each method for each data set are recorded on Pentium (R) Dual-Core CPU 2.80 GHz processor and 32-bit operating system Windows ® computer during the runs (Table 14). Depending on the results, following conclusions can be stated:

- Run times increases as sample size and scale increases.
- As expected, it takes the bootstrap methods considerably longer times to run than MARS and CMARS.

Table 14. Runtimes (in seconds) of Methods with respect to Size and Scale of Data Sets

		Scale	
		Small	**Medium**
Sample Size	**Small**	MARS: < 0.0800 sec.*	MARS: < 0.0800 sec.*
		CMARS: < 4.4666 sec.	CMARS: < 19.5269 sec.
		BCMARS-1: < 1,595 sec.	BCMARS-1: < 13,262 sec.
		BCMARS-2: < 1,578 sec.	BCMARS-2: < 18,537 sec.
		BCMARS-3: < 1,599 sec.	BCMARS-3: < 15,617 sec.
	Medium	MARS: < 0.0840 sec.*	MARS: < 0.0900 sec.*
		CMARS: < 18.2008 sec.	CMARS: < 21.6737 sec.
		BCMARS-1: < 15,958 sec.	BCMARS-1: < 18,664 sec.
		BCMARS-2: < 7,076 sec.	BCMARS-2: < 31,590 sec.
		BCMARS-3: < 8,374 sec.	BCMARS-3: < 16,753 sec.

*indicates better performance with respect to run times

- Three bootstrap regression methods have almost the same efficiencies in small size and small scale data sets. Note that run times of these methods increases almost ten times as much as the scale increases from small to medium.
- Random-X and Wild bootstrapping have similar efficiencies in medium size small scale data sets; Fixed-X runs twice as much to those of Random-X and Wild bootstrapping, whose run times increase almost five times as much as the sample size increases.
- Fixed-X and Wild bootstrapping have similar run times for medium size medium scale data sets while Random-X runs twice as much to that of Fixed-X and Wild bootstrapping.

5.5. Evaluation of the Precisions of the Model Parameters

In addition to performance measures of the models, the CIs and standard deviations of the parameters are calculated after bootstrapping. These values are compared with those values obtained from bootstrapping CMARS. Table A1-A48 in Appendix B presents the length of CIs and standard deviations of the model parameters in addition to BFs of the models. The smaller the lengths of the CIs and the standard deviations, the more precise the parameter estimates are.

According to the results, following conclusions can be drawn:

In Forest data:

- The length of CIs is larger in Wild bootstrapping than the ones obtained by Fixed-X Resampling. Thus, Fixed-X gives more precise parameter estimates.
- The standard deviations obtained by bootstrapping (STD(BS)) are smaller for Wild bootstrapping method than for Fixed-X Resampling.
- In general, both types of standard deviations are smaller than the ones obtained from CMARS.

In US data set:

- In fold 2, standard deviations of Wild bootstrapping are smaller compared to those of CMARS, while the STD (BS) are not. However, the lengths of CIs become narrower after bootstrapping.

In PM10 data set:

- In general, the length of CIs of Random-X is smaller than CMARS. Thus, Random-X produces more precise parameter estimates.
- Random-X Resampling produces narrower CIs than Fixed-X. So, parameter estimates of Random-X are more precise.
- The standard deviations of parameters obtained by Random-X and Fixed-X are similar.

In Slump data set:

- The lengths of CIs become narrower and standard deviations of the parameters become smaller after bootstrapping, thus, resulting in more precise parameter estimates.
- STD(BS) values obtained for Fixed-X Resampling are smaller than ones obtained from Random-X.

CHAPTER 6

CONCLUSION AND FURTHER RESEARCH

In this study, three different bootstrap methods are applied to a nonparametric regression, called CMARS, which is an improved version of the backward step of the widely used method MARS. MARS has two-step algorithm to build a model: forward and backward. CMARS uses inputs obtained from the forward step of MARS, and then, by utilizing the CQP technique, it constructs the large model. Although CMARS overperforms MARS with respect to several criteria, it constructs models which are at least as complex as MARS (Weber et al., 2011).

In this thesis, it is aimed to reduce the complexity of CMARS models. To achieve this aim, bootstrapping regression methods, namely Fixed-X and Random-X Resampling, and Wild bootstrapping, are utilized by adopting an iterative approach to determine whether the parameters statistically contribute to the developed CMARS model or not. If there are any which do not contribute, they are removed from the model, and a new CMARS model is fitted to the data by only retaining the statistically significant parameters until none of them is found to be insignificant. The reason of using a computational method here is the lack of prior knowledge regarding the distributions of the model parameters.

The performances of the methods are empirically evaluated and compared with respect to several criteria by using four data sets which are selected in such a way that they can represent the small and medium sample size and scale categories. The criteria include accuracy (with MAE, R^2, PWI and PRESS measures), complexity (with the MSE measure), stability (by comparing the performances in training and test samples), robustness (by comparing the performances in different data sets),

efficiency (using run times) and precision (by evaluating the length of CIs of parameters). All performance criteria are explained in Appendix A. In order to validate all models developed; three-fold CV approach is used. For this purpose, these data sets are divided into three parts (folds) and two of them are used for building (training) and the remaining one is used for testing.

Depending on the comparisons presented in the previous chapter, Chapter 5, one may conclude the followings:

- In general, BCMARS methods perform better than MARS and CMARS with respect to most of the measures, and also lead to development of robust models with respect to the same measures.
- Either one of the BCMARS methods yields models which are less complex than that of MARS and CMARS.
- In overall, Random-X Resampling or Wild bootstrapping produce more stable models with respect to most of the measures considered.
- Fixed-X method performs the best in small size training data in terms of most measures.
- Fixed-X also performs the best in medium size training data sets with respect to MSE and R^2.
- MARS and Random-X Resampling overperform in small and medium size test data sets, respectively.
- Wild bootstrapping and Random-X methods are more stable in small and medium size test data sets, respectively.
- Fixed-X is performing equally well on both scale of training data sets.
- Random-X performs best in medium scale data while MARS and CMARS perform best in small scale data.
- Random-X Resampling is more stable in medium scale data set.
- It is apparent that by decreasing the number of terms in the model by bootstrapping, the CIs become narrower compared to those of CMARS. Moreover, the standard errors of the parameters which obtained empirically

decreases after bootstrapping. Thus, bootstrapping results in more precise parameter estimates.

- The main drawback of bootstrapping is its computational effort. Since it is heavily dependent on computers, it takes significantly more time than the other methods, MARS and CMARS.

In short, depending on the above conclusions, it may be suggested that Random-X Resampling method leads to more accurate and more precise and less complex models particularly for medium size and medium scale data. Nevertheless, it is the least efficient method among the others for this type of data set.

Future studies are planned in several directions. First, BCMARS methods are going to be applied on different data sets with small to large size and scale. Then, Repeated Analysis of Variance (RANOVA) will be applied to test whether there is statistically significant difference between the performances of methods. Besides, replicated CV is going to be used while validating the models. Then, after well-documented, the written MATLAB code will be issued as on open source to make it available for interested researchers.

REFERENCES

Aldrin, M. (2006). Improved Predictions Penalizing both Slope and Curvature in Additive Models. *Computational Statistics and Data Analysis*, 50 (2), 267–284.

Alp, O. S., Büyükbebeci, E., Iscanoglu Cekic, A., Yerlikaya-Özkurt, F., Taylan, P. and Weber, G.-W. (2011). CMARS and GAM & CQP - Modern Optimization Methods Applied to International Credit Default Prediction. *Journal of Computational and Applied Mathematics (JCAM)*, 235, 4639-4651.

Austin, P. (2008). Using the Bootstrap to Improve Estimation and Confidence Intervals for Regression Coefficients Selected using Backwards Variable Elimination. *Statistics in Medicine*, 27 (17), 3286–3300.

Batmaz, İ., Yerlikaya-Özkurt, F., Kartal-Koç, E., Köksal, G. and Weber, G. W. (2010). Evaluating the CMARS Performance for Modeling Nonlinearities. *Proceedings of the 3rd Global Conference on Power Control and Optimization, Gold Coast (Australia)*, 1239, 351-357.

Bootstrapping Regression Models. Retrieved from http://www.sagepub.com/upm-data/21122_Chapter_21.pdf (accessed on February 9, 2011).

Chernick, M. (2008). *Bootstrap Methods: A Guide for Practitioners and Researchers.* New York: Wiley.

Cortez, P., and Morais., A. (2007). Data Mining Approach to Predict Forest Fires using Meteorological Data. In J. Neves, M. F. Santos and J. Machado (Ed.), New Trends in Artificial Intelligence, *Proceedings of the 13th EPIA 2007 - Portuguese Conference on Artificial Intelligence*, December, Guimarães, Portugal, 512-523.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Application.* Cambridge: Cambridge Series in Statistical and Probabilistic Mathematics.

Deconinck, E., Zhang, M. H., Petitet, F., Dubus, E., Ijjaali, I., Coomans, D., and Vander Heyden, Y. (2008). Boosted Regression Trees, Multivariate Adaptive Regression Splines and Their Two-Step Combinations with Multiple Linear Regression or Partial Least Squares to Predict Blood-Brain Barrier Passage: A case study. *Analytica Chimica Acta*, 609 (1), 13–23.

Denison, D. G. T., Mallick, B. K., and Smith, F. M. (1998). Bayesian MARS. *Statistics and Computing,* 8 (4), 337-346.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7 (1), 1-26.

Efron, B. (1979). Computers and the Theory of Statistics: Thinking the Unthinkable. *SIAM Review*, 21 (4), 460-479.

Efron, B. (1988). Computer-Intensive Methods in Statistical Regression. *Society for Industrial and Applied Mathematics*, 30 (3), 421-449.

Efron, B. (1992). Jackknife-After-Bootstrap Standard Errors and Influence Functions. Journal *of the Royal Statistical Society*. Series B (Methodological), 54 (1), 83-127.

Efron, B., and Tibshirani, R.J. (1986). Bootstrap methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science,* 1 (1), 75-77.

Efron, B., and Tibshirani, R.J. (1991). Statistical Data Analysis in the Computer Age. *Science,* 253, 390-395.

Efron, B., and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* New York: Chapman & Hall.

Flachaire, E. (2003). *Bootstrapping Heteroskedastic Regression Models: Wild Bootstrap vs. Pairs Bootstrap*. Working paper.

Fox, J. (2002). Bootstrapping Regression Models. *An R and S-PLUS Companion to Applied Regression:Web Appendix to the Book.* Sage, CA: Thousand Oaks.

Freedman, D.A. (1981). Bootstrapping Regression Models. *The Annals of Statistics*, 9 (6), 1218-1228.

Friedman J. (1991). Multivariate Adaptive Regression Splines. *Annals of Statistics,* 19 (1), 1-67.

Gentle, J. E. (2009). *Computational Statistics*. New York: Springer.

Givens, G. H., and Hoeting, J. A. (2005). *Computational Statistics*. New York: John Wiley & Sons.

Godfrey, L. (2009). *Bootstrap Tests for Regression Models*. Palgrave Macmillian.

Gonçalves, S., White, H., (2005). Bootstrap Standard Error Estimates for Linear Regression. *American Statisticsl Association*, 100 (471), 970-979.

Gutièrrez, A. G., Contador, F. L., and Schnabel, S. (2011). Modeling Soil Properties at a Regional Scale Using GIS and Multivariate Adaptive Regression Splines. *Geomorphometry 2011 Conference Proceedings* In: *Geomorphometry 2011*, Edited by R. Purves, S. Gruber, R. Straumann and T. Hengl. California.

Hastie, T., Tibshirani, and R., Friedman, J., (2001). *The Elements of Statistical Learning, Data Mining, Inference and Prediction.* New York: Springer.

Hjorth, J. S. U., (1994). *Computer Intensive Statistical Methods: Validation Model Selection and Bootstrap.* New York: Chapman & Hall.

Kartal, E. (2007). *Metamodeling Complex Systems Using Linear and Nonlinear Regression Methods.* Master Thesis, Graduate School of Natural and Applied Sciences, Department of Statistics, METU, Ankara, Turkey.

Kirk, P.D.W., and Stumpf, M. P. H. (2009). Gaussian Process Regression Bootstrapping: Exploring the Effects of Uncertainty in Time Course Data. *Bioinformatics*, 25 (10), 1300-1306.

Kriner, M. (2007). *Survival Analysis with Multivariate Adaptive Regression Splines.* Dissertation, LMU Munchen: Faculty of Mathematics, Computer Science and Statistics, Munchen.

Leathwick, J. R., Rowe, D., Richardson, J., Elith J., and Hastie, T. (2005). Using Multivariate Adaptive Regression Splines to Predict the Distributions of New Zealand's Freshwater Diadromous Species. *Freshwater Biology*, 50, 2034–2052.

Lin, C. J., Chen, H. F., and Lee, T. S. (2011). Forecasting Tourism Demand Using Time Series, Artificial Neural Networks and Multivariate Adaptive Regression Splines: Evidence from Taiwan. *International Journal of Business Administration*, 2 (2), 14-24.

Liu, R. Y. (1988). Bootstrap Procedure under Some non-i.i.d. Models. *Annals of Statistics*, 16 (4), 1696-1708.

Loughin, T. M., and Koehler, K. J. (1997). Bootstrapping Regression Parameters in Multivariate Survival Analysis. *Lifetime Data Analysis*, 3(2), 157–177.

Martinez, W. L., and Martinez, A. R. (2002). *Computational Statistics Handbook with Matlab.* New York: Chapman & Hall.

MATLAB Version 7.8.0 (R2009a).

Matlab-R link Retrieved from http://www.mathworks.com/matlabcentral/fileexchange/5051 (accessed on February 24, 2011).

Milborrow, S. (2009). earth: Multivariate Adaptive Regression Spline Models. R Software Package. Retrieved from http://cran.r-project.org/web/packages/earth/index.html (accessed on February 24, 2011).

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2001). *Introduction to Linear Regression Analysis.* New York: John Wiley & Sons.

MOSEK, Version 6. A very powerful commercial software for CQP. Retrieved from http://www.mosek.com (accessed January 7, 2011).

Osei-Bryson, K. M. (2004). Evaluation of Decision Trees: A Multi-Criteria Approach. *Computers Operations Research*, 31 (11), 1933-1945.

R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org. (accessed on February 24, 2011).

Ramanathan, R. (2002). *Introductory Econometrics with Applications*. South Western College Publishing.

Salibian-Barrera, M., and Zamar, R. Z. (2002). Bootstrapping Robust Estimates of Regression. *The Annals of Statistics*, 30 (2), 556-582.

Taylan, P., Weber, G.-W. and Yerlikaya, F. (2010). A new approach to multivariate adaptive regression spline by using Tikhonov regularization and continuous optimization, *TOP (the Operational Research journal of SEIO (Spanish Statistics and Operations Research Society)*, 18 (2), 377-395.

Weber, G. W., Batmaz, I., Koksal, G., Taylan, P., and Yerlikaya-Ozkurt, F. (2011). CMARS: A New Contribution to Nonparametric Regression with

Multivariate Adaptive Regression Splines Supported by Continuous Optimisation, *Inverse Problems in Science and Engineering* (in print).

Wegman, E., (1988). Computational Statistics: A New Agenda for Statistical Theory and Practice. *Journal of the Washington Academy of Sciences*, 78, 310-322.

Wu, C. F. J. (1986). Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis. *The Annals of Statistics*, 14 (4), 1261-1295.

Yeh, I-Cheng, (2007). Modeling Slump Flow of Concrete using Second-order Regressions and Artificial Neural Networks. *Cement and Concrete Composites,* 29 (6), 474-480.

Yerlikaya, F. (2008). *A New Contribution to Nonlinear Robust Regression and Classification with MARS and its Applications to Data Mining for Quality Control in Manufacturing*. Master Thesis, Graduate School of Applied Mathematics, Department of Scientific Computing, METU, Ankara, Turkey.

Yetere-Kurşun, A., Batmaz, İ. (2010). Comparison of Regression Methods by Employing Bootstrapping Methods. *COMPSTAT2010: 19th International Conference on Computational Statistics.* Paris, France. August 22-27. Book of Abstracts, 92.

York, T. P., Eaves, L. J., and Van Den Oord, E., J., C., G. (2006). Multivariate Adaptive Regression Splines: A Powerful Method for Detecting Disease-Risk Relationship Differences among Subgroups. *Statistics in Medicine*, 25 (8), 1355–1367.

Zakeri, I. F., Adolph, A. L., Puyau, M., R., Vohra, F. A., Butte, N. F. (2010). Multivariate Adaptive Regression Splines Models for the Prediction of Energy Expenditure in Children and Adolescents. *Journal of Applied Physchology*, 108, 128–136.

APPENDIX A

DEFINITIONS OF COMPARISON MEASURES

Nomencleature:

y_i is the response value for the i^{th} observation,

$\hat{y}_i$ is the estimated response value for the i^{th} observation,

$\bar{y}$ is the value of the mean response,

n is the number of observations (sample size),

p is the number of terms (BFs) in the model,

$\bar{\hat{y}}$ is the value of the mean of the estimated responses,

$s(y)^2$ is the sample variance of the observed response values,

$s(\hat{y})^2$ is the sample variance of the estimated response values,

$e_i = y_i - \hat{y}_i$ is the residual for the i^{th} observation,

h_i is the leverage value of the i^{th} observation. It is obtained from the i^{th} diagonal element of the hat matrix; H. The hat matrix is defined with the following formula $H = X(X^T X)^{-1} X^T$. Here, X represents the design matrix and rank of it is p.

Accuracy Measures

Mean Absolute Error (MAE)

It is defined as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|e_i| = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|. \quad (48)$$

Small values are the better.

The Coefficient of Determination (R^2)

This value shows how much variation in the response variable is explained by the model. It is defined by the following formula:

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}. \tag{49}$$

Higher values indicate better fit.

Proportion of Residuals within Some User-Specified Range (PWI)

PWI is the proportion of residuals within some user-specified range such as two or three sigma. In this study, three sigma coverage is considered. The greater the percentage is the better the performance.

Prediction Error Residual Sum of Squares (PRESS)

PRESS measures the predictive capability of the model. The formula used to calculate this measure is defined as:

$$PRESS = \sum_{i=1}^{n}\left(\frac{e_i}{1-h_i}\right)^2. \tag{50}$$

Small values of PRESS, indicates a higher ability of prediction.

Precision Measure

Bootstrap Estimate of Standard Deviation

The bootstrap estimate of standard error is calculated with the following formula (Martinez and Martinez, 2002).

$$\hat{se}_B = \left\{ \frac{1}{B-1} \sum_{i=1}^{B} \left(\hat{\theta}^{*b} - \bar{\hat{\theta}}^{*} \right)^2 \right\}^{1/2}, \tag{51}$$

where

$$\bar{\hat{\theta}}^{*} = \frac{1}{B} \sum_{b=1}^{B} \hat{\theta}^{*b} \tag{52}$$

and $\hat{\theta}^{*b}$ is the bootstrap replication of $\hat{\theta}$.

It measures the variation around the mean. The standard deviations of the parameters from ECDF are obtained. Besides, standard deviations of the distributions of parameters are calculated in another way. First, an empirical distribution is obtained for each parameter. Then, by bootstrapping, the standard deviation of each parameter is calculated 1000 times. The standard deviations of these 1000 values are obtained and recorded in the tables presented in Appendix B under the label "STD (BS)." Thus, the spread of the standard deviations around the mean is obtained.

Efron and Tibshirani (1993) use the same method (second method) for correlation coefficient instead of standard error.

Complexity Measure

Mean Square Error (MSE)

In this study, the MSE is used to measure the model complexity. Larger values of the MSE indicate more complex models. The formula for the MSE is given below:

$$MSE = \frac{1}{n-p} \sum_{i=1}^{n} \left(y_i - \hat{y}_i \right)^2 \tag{53}$$

Stability Measure

The model is said to be stable if it performs well on both training and testing data sets. It is measured by the following formula (Osei-Bryson, 2004):

$$\min\left\{\frac{CR_{TR}}{CR_{TE}}, \frac{CR_{TE}}{CR_{TR}}\right\} \tag{54}$$

CR_{TR} and CR_{TE} represents the performance measures obtained from training and testing samples. If the stability measure is close to one, it indicates higher stability.

APPENDIX B

BASIS FUNCTIONS AND PERCENTILE INTERVALS

Table A1. BFs from Forward Step of MARS Obtained from the Main Effects Model with Interactions for Fold 1 FF Data Set

BF1 = max (0, 0, x5-1.74135)
BF2 = max (0, 0, 1.74135-x5)
BF3 = max (0, x5-1.5009)
BF4 = max (0, x5-1.5009)*max (0, x8-0.862951)
BF5 = max (0, x5-1.5009)*max (0, 0.862951-x8)
BF6 = max (0, x5-1.74135)*max(0, x8-0.759621)
BF7 = max (0, x5-1.74135)*max (0, 0.759621-x8)
BF8 = max (0, x4-0.75276)*max (0, x5-1.5009)
BF9 = max (0, 0.75276-x4)*max (0, x5-1.5009)
BF10 = max (0, x4-0.553489)*max (0, x5-1.74135)
BF11 = max (0, x2+1.05684)
BF12 = max(0, -1.05684-x2)
BF13 = max(0, x2+0.243765)*max(0, 1.74135-x5)
BF14 = max(0, -0.243765-x2)*max(0, 1.74135-x5)
BF15 = max(0, -1.05684-x2)*max(0, x8-0.518517)
BF16 = max(0, -1.05684-x2)*max(0, 0.518517-x8)
BF17 = max(0, -1.05684-x2)*max(0, x3-0.230308)
BF18 = max(0, -1.05684-x2)*max(0, 0.230308-x3)
BF19 = max(0, x2-1.38238)*max(0, x5-1.74135)
BF20 = max(0, 1.38238-x2)*max(0, x5-1.74135)

Table A2. Percentile Intervals of Parameters Obtained by Fixed-X Resampling of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 1 FF Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.4603	0.2559*	0.0033	0.0019*	0.1368	0.0784*
θ_1	3.7137	2.6732*	0.0578	0.0365*	1.2346	0.8627*
θ_2	0.1448	-	0.0011	-	0.0447	-
θ_3	2.5156	1.8946*	0.0325	0.03*	0.8194	0.646*
θ_4	7.8272	7.3439*	0.1265	0.1368	2.614	2.6197
θ_5	5.5938	4.6525*	0.0803	0.0595*	1.8364	1.5029*
θ_6	24.4399	22.4373*	0.4997	0.5158	8.5912	8.2625*
θ_7	6.8601	5.5717*	0.0991	0.0739*	2.2595	1.8446*
θ_8	98.5900	95.7020*	2.4622	2.8581	36.35	38.1427
θ_9	14.039	-	0.0288	-	0.6406	-
θ_{10}	72.3437	72.7325	1.4864	2.305	25.2854	29.9606
θ_{11}	0.2670	0.2349*	0.0024	0.0018*	0.0835	0.0722*
θ_{12}	0.982	-	0.0109	-	0.3031	-
θ_{13}	0.1223	0.1047*	0.001	0.001*	0.0381	0.0331*
θ_{14}	0.2173	0.1437*	0.0018	0.0012*	0.0666	0.0436*
θ_{15}	2.0120	1.3176*	0.0389	0.0331*	0.6712	0.4794*
θ_{16}	1.1293	-	0.0128	-	0.353	-
θ_{17}	2.3470	2.0170*	0.0267	0.031	0.7391	0.6868*
θ_{18}	0.9709	-	0.0164	-	0.3235	-
θ_{19}	11.2567	11.0153*	0.3327	0.2807	4.623	4.3802*
θ_{20}	1.1728	1.1216*	0.0111	0.0155	0.3458	0.3594

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A3. Percentile Intervals of Parameters Obtained by Random-X Resampling of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 1 FF Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.9803	-	0.0063	-	0.3063	-
θ_1	12.7135	-	0.0749	-	4.017	-
θ_2	0.1601	-	0.0012	-	0.0518	-
θ_3	11.927	-	0.0706	-	3.5503	-
θ_4	76.7845	-	7293.755	-	14291.84	-
θ_5	13.8178	-	0.0848	-	4.1067	-
θ_6	196.095	-	50992784	-	1.09E+08	-
θ_7	17.1029	-	0.1189	-	5.1721	-
θ_8	69,000,041	-	1.66E+11	-	2.68E+11	-
θ_9	5.4917	-	0.1246	-	2.08	-
θ_{10}	251.5979	-	15987320	-	26288490	-
θ_{11}	0.7684	-	0.0047	-	0.2404	-
θ_{12}	1.5667	-	0.0119	-	0.4804	-
θ_{13}	0.3131	-	0.002	-	0.0983	-
θ_{14}	0.4462	-	0.0116	-	0.1839	-
θ_{15}	5.3164	-	0.0967	-	1.9302	-
θ_{16}	1.6985	-	0.018	-	0.54	-
θ_{17}	7.5524	-	0.0939	-	2.3861	-
θ_{18}	1.3503	-	35.9922	-	57.0924	-
θ_{19}	1,984,276	-	1.24E+10	-	3.63E+10	-
θ_{20}	1.9581	-	0.0501	-	0.7559	-

- shows statistically insignificant model parameter

Table A4. Percentile Intervals of Parameters Obtained by Wild Bootstrapping of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 1 FF Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.821	0.037*	0.0054	0.0002*	0.2484	0.011*
θ_1	12.653	-	0.0839	-	3.8946	-
θ_2	0.290	-	0.002	-	0.0876	-
θ_3	7.737	-	0.0512	-	2.3136	-
θ_4	16.308	14.434*	0.0964	0.0874*	5.0666	4.281*
θ_5	11.264	-	0.072	-	3.3351	-
θ_6	50.145	48.527*	0.2932	0.2836*	14.9126	14.3465*
θ_7	14.011	-	0.091	-	4.1611	-
θ_8	234.212	221.133*	1.164	1.047*	69.0034	69.0773
θ_9	4.442	-	0.0245	-	1.3842	-
θ_{10}	191.733	-	1.0982	-	59.6608	-
θ_{11}	0.493	-	0.0032	-	0.1514	-
θ_{12}	1.788	-	0.0127	-	0.5471	-
θ_{13}	0.227	-	0.0015	-	0.069	-
θ_{14}	0.368	-	0.0024	-	0.1131	-
θ_{15}	3.667	2.791*	0.0236	0.0167*	1.1139	0.8465*
θ_{16}	2.114	-	0.0129	-	0.6326	-
θ_{17}	4.417	4.047*	0.0267	0.0257*	1.3321	1.182*
θ_{18}	1.848	-	0.0107	-	0.5752	-
θ_{19}	25.997	-	0.1522	-	7.997	-
θ_{20}	2.154	-	0.014	-	0.6703	-

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A5. BFs from Forward Step of MARS Obtained from the Main Effects Model with Interactions for Fold 2 FF Data Set

BF1 = max (0, x8-0.294634)
BF2 = max (0,0.294634-x8)
BF3 = max (0, x8-0.294634)*max (0, x10-0.492505)
BF4 = max (0, x8-0.294634)*max (0, 0.492505-x10)
BF5 = max (0, x8-0.294634)*max (0, x9+1.05949)
BF6 = max (0, x8-0.294634)*max (0, -1.05949-x9)
BF7 = max (0, x8-0.294634)*max (0, x9+0.936922)
BF8 = max (0, x8-1.44849)
BF9 = max (0, x6-0.611369)*max (0, x8-0.294634)
BF10 = max (0, 0.611369-x6)*max (0, x8-0.294634)
BF11 = max (0, x1-0.575144)*max (0, x8-0.294634)
BF12 = max (0, 0.575144-x1)*max (0, x8-0.294634)
BF13 = max (0, x4+0.0080942)*max (0, x8-0.294634)
BF14 = max (0, -0.0080942-x4)*max (0, x8-0.294634)
BF15 = max (0, x8-0.294634)*max (0, x10+0.735411)
BF16 = max (0, x1-0.14295)*max (0, x8-0.294634)
BF17 = max (0, x5-1.87719)*max (0, x8-0.294634)
BF18 = max (0, 1.87719-x5)*max (0, x8-0.294634)

Table A6. Percentile Intervals of Parameters Obtained by Fixed-X Resampling of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 2 FF Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.2615	0.1898*	0.0021	0.0013*	0.0808	0.0579
θ_1	2.4441	-	0.0216	-	0.7462	-
θ_2	0.2494	-	0.0021	-	0.0759	-
θ_3	3.1275	2.8766*	0.0298	0.0369	0.9813	0.9731*
θ_4	1.7759	1.0452*	0.0178	0.0128*	0.5643	0.3462*
θ_5	9.3435	9.4064	0.0787	0.1187	2.8484	2.9944
θ_6	5.4064	5.6300	0.0513	0.084	1.6834	1.8615
θ_7	9.6387	9.6081*	0.0809	0.1141	2.9521	3.097
θ_8	2.0566	1.3534*	0.0237	0.0287	0.6392	0.5379*
θ_9	7.5281	6.9701*	0.0755	0.0714*	2.3385	2.23*
θ_{10}	1.2823	1.2868	0.0182	0.0143*	0.4388	0.421*
θ_{11}	4.1136	3.3978*	0.0403	0.0552	1.2739	1.1904*
θ_{12}	0.7375	-	0.0083	-	0.2411	-
θ_{13}	1.6949	1.5016*	0.0156	0.0149*	0.5377	0.4706*
θ_{14}	1.2660	1.3234	0.0233	0.0181*	0.4527	0.4461*
θ_{15}	2.4694	1.7393*	0.0206	0.0261	0.7599	0.5809*
θ_{16}	3.2114	2.3321*	0.0326	0.04	0.9952	0.813*
θ_{17}	7.2183	7.2505	0.0947	0.1004	2.373	2.4001
θ_{18}	0.7454	0.7166*	0.0081	0.0081	0.2361	0.2426

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A7. Percentile Intervals of Parameters Obtained by Random-X Resampling of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 2 FF Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.1738	0.1826	0.0013	0.0013*	0.0524	0.0568
θ_1	7.2361	-	0.0719	-	2.3023	-
θ_2	0.1092	-	0.0007	-	0.0331	-
θ_3	12.1772	5.8499*	0.0787	0.0384*	3.4944	1.7307*
θ_4	4.9731	-	0.033	-	1.529	-
θ_5	48.8433	22.0268*	0.3168	0.1512*	14.1354	6.4682*
θ_6	30.9479	12.9475*	0.2195	0.0919*	9.5091	3.9949*
θ_7	48.0485	22.2773*	0.2981	0.1578*	13.8268	6.6065*
θ_8	7.8384	-	0.0547	-	2.4153	-
θ_9	18.1939	-	0.1402	-	5.8189	-
θ_{10}	2.4921	-	0.0185	-	0.7795	-
θ_{11}	18.8273	-	0.1205	-	6.079	-
θ_{12}	1.2867	-	0.013	-	0.3945	-
θ_{13}	5.294	-	0.0381	-	1.7073	-
θ_{14}	2.8865	-	0.0755	-	1.1333	-
θ_{15}	9.2723	3.4522*	0.0584	0.0211*	2.7038	1.0296*
θ_{16}	12.5715	1.4830*	0.089	0.012*	4.0295	0.474*
θ_{17}	20.3324	-	10026.41	-	51681.67	-
θ_{18}	1.6136	-	0.0101	-	0.5262	-

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A8. Percentile Intervals of Parameters Obtained by Wild Bootstrapping of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 2 FF Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.2416	0.1112*	0.0016	0.0007*	0.0739	0.0333*
θ_1	2.7944	-	0.0195	-	0.8573	-
θ_2	0.2693	-	0.0018	-	0.0826	-
θ_3	3.9781	3.0994*	0.0277	0.0213*	1.1851	0.9786*
θ_4	2.3040	1.2159*	0.0153	0.0076*	0.6696	0.363*
θ_5	10.3385	10.1477*	0.0676	0.0627*	3.1805	3.1021*
θ_6	6.5232	6.5581	0.0386	0.0405	1.969	1.9497*
θ_7	10.6166	10.4488*	0.0679	0.0688	3.2873	3.208*
θ_8	2.3615	1.6832*	0.0168	0.0168*	0.7419	0.579*
θ_9	8.1361	-	0.0539	-	2.5172	-
θ_{10}	1.5512	1.4582*	0.0105	0.0095*	0.478	0.4475*
θ_{11}	4.8457	4.1868*	0.0314	0.0272*	1.5174	1.267*
θ_{12}	0.8703	-	0.0058	-	0.2683	-
θ_{13}	1.9068	1.7376*	0.0125	0.0105*	0.5708	0.5053*
θ_{14}	1.6044	1.4854*	0.0108	0.0098*	0.4902	0.455*
θ_{15}	3.0240	2.0019*	0.0205	0.0125*	0.9096	0.5871*
θ_{16}	3.8541	2.7509*	0.0242	0.0175*	1.1816	0.8386*
θ_{17}	8.6764	5.7258*	0.0573	0.0318*	2.5329	1.7714*
θ_{18}	0.8584	0.8137*	0.0061	0.0055*	0.2629	0.2463*

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A9. BFs from Forward Step of MARS Obtained from the Main Effects Model with Interactions for Fold 3 FF Data Set

BF1 = max (0, x8-1.12128)
BF2= max (0, 1.12128-x8)
BF3 = max (0, x8-0.845729)
BF4 = max (0, x8-1.29349)
BF5 = max (0, x8-0.845729)*max (0, x9+1.05949)
BF6 = max (0, x8-0.845729)*max (0, -1.05949-x9)
BF7 = max (0, x8-1.12128)*max (0, x9+1.05949)
BF8 = max (0, x8-1.12128)*max (0, -1.05949-x9)
BF9 = max (0, x8-0.845729)*max (0, x10+0.0098242)
BF10 = max (0, x8-0.845729)*max (0, -0.0098242-x10)
BF11 = max (0, x8-1.12128)*max (0, x10+2.01914)
BF12 = max (0, x8-1.12128)*max (0, x10+0.0098242)
BF13 = max (0, x8-1.29349)*max (0, x9+1.18206)
BF14 = max (0, x8-1.29349)*max (0, -1.18206-x9)
BF15 = max (0, x8-1.29349)*max (0, x9+0.998206)
BF16 = max (0, x8-0.845729)*max (0, x9+0.936922)

Table A10. Percentile Intervals of Parameters Obtained by Fixed-X Resampling of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 3 FF Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.3524	-	0.0031	-	0.1108	-
θ_1	5.6518	-	0.0376	-	1.6553	-
θ_2	0.2221	-	0.0021	-	0.0656	-
θ_3	2.8373	-	0.0555	-	1.0483	-
θ_4	5.3950	-	0.0466	-	1.6727	-
θ_5	22.7249	-	0.4397	-	8.4137	-
θ_6	30.1681	-	0.7874	-	10.8025	-
θ_7	22.1346	-	0.5879	-	9.2691	-
θ_8	79.9583	-	1.9622	-	28.226	-
θ_9	7.5260	-	0.2921	-	3.3645	-
θ_{10}	2.9897	-	0.0619	-	1.0936	-
θ_{11}	4.8080	-	0.0901	-	1.7008	-
θ_{12}	14.1384	-	0.3529	-	5.8088	-
θ_{13}	45.4657	-	0.9245	-	15.7589	-
θ_{14}	58.4440	-	1.2106	-	21.5196	-
θ_{15}	63.0700	-	1.657	-	22.2767	-
θ_{16}	22.3437	-	0.521	-	8.5695	-

- shows statistically insignificant model parameter

Table A11. Percentile Intervals of Parameters Obtained by Random-X Resampling of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 3 FF Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.5943	-	0.0056	-	0.1886	-
θ_1	4.1307	-	0.0248	-	1.5353	-
θ_2	0.3145	-	0.003	-	0.0995	-
θ_3	6.0062	-	0.0279	-	2.5694	-
θ_4	3.9385	-	0.0314	-	1.3953	-
θ_5	25.3066	-	0.1337	-	8.5121	-
θ_6	39.1122	-	0.1906	-	14.2317	-
θ_7	29.5122	-	0.1749	-	9.5268	-
θ_8	91.9701	-	0.4401	-	32.9339	-
θ_9	8.3818	-	0.0898	-	2.6828	-
θ_{10}	3.9866	-	0.02	-	1.4636	-
θ_{11}	6.8726	-	0.0374	-	2.4567	-
θ_{12}	16.9952	-	0.1444	-	5.6448	-
θ_{13}	47.8843	-	0.2759	-	15.5051	-
θ_{14}	62.1864	-	25077.87	-	107107.7	-
θ_{15}	67.5898	-	0.4042	-	21.7303	-
θ_{16}	19.6854	-	0.1213	-	6.3671	-

- shows statistically insignificant model parameter

Table A12. Percentile Intervals of Parameters Obtained by Wild Bootstrapping of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 3 FF Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.3411	0.2198*	0.0023	0.0014*	0.1056	0.0684*
θ_1	20.5739	-	0.13	-	6.2377	-
θ_2	0.2462	0.1950*	0.0016	0.0013*	0.0748	0.0607*
θ_3	7.8834	-	0.0494	-	2.4079	-
θ_4	15.3645	-	0.1	-	4.6484	-
θ_5	34.7498	-	0.2111	-	10.2602	-
θ_6	56.4426	-	0.3494	-	16.8484	-
θ_7	38.4642	-	0.2425	-	11.3621	-
θ_8	132.6501	-	0.8702	-	40.3004	-
θ_9	10.9232	-	0.0701	-	3.3922	-
θ_{10}	6.0988	-	0.036	-	1.8604	-
θ_{11}	9.995	-	0.0587	-	3.0387	-
θ_{12}	22.7681	-	0.1551	-	6.9473	-
θ_{13}	72.6011	-	0.4712	-	21.854	-
θ_{14}	85.6487	-	0.585	-	26.1419	-
θ_{15}	90.8052	-	0.6072	-	28.1857	-
θ_{16}	32.953	-	0.2088	-	10.0022	-

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A13. BFs from Forward Step of MARS Obtained from the Main Effects Model with Interactions for Fold 1 PM10 Data Set

BF1 = max (0, x1+0.533237)
BF2 = max (0, -0.533237-x1)
BF3 = max (0, x3-0.692675)
BF4 = max (0, 0.692675-x3)
BF5 = max (0, x7+0.610358)
BF6 = max (0, -0.610358-x7)
BF7 = max (0, x2-0.0959798)
BF8 = max (0, 0.0959798-x2)
BF9 = max (0, x2-0.0959798)*max (0, x7-1.16449)
BF10 = max (0, x2-0.0959798)*max (0, 1.16449-x7)
BF11 = max (0, x5+0.70273)
BF12 = max (0, -0.70273-x5)
BF13 = max (0, x3-0.692675)*max (0, x7-0.431617)
BF14 = max (0, x3-0.692675)*max (0, 0.431617-x7)
BF15 = max (0, x4+0.0554302)
BF16 = max (0, -0.0554302-x4)
BF17 = max (0, x1-0.386576)*max (0, -0.610358-x7)
BF18 = max (0, 0.386576-x1)*max (0, -0.610358-x7)
BF19 = max (0, x2-0.0959798)*max (0, x4+1.26923)
BF20 = max (0, x2-0.0959798)*max (0, -1.26923-x4)

Table A14. Percentile Intervals of Parameters Obtained by Fixed-X of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 1 PM10 Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.5354	0.4495*	0.0038	0.003*	0.1656	0.1357*
θ_1	0.3363	0.2746*	0.0023	0.0018*	0.1022	0.0821*
θ_2	0.3597	-	0.0025	-	0.1107	-
θ_3	0.7599	0.7424*	0.0049	0.005	0.2262	0.2273
θ_4	0.2209	0.2236	0.0014	0.0014*	0.0671	0.0671*
θ_5	0.2682	0.2377*	0.0018	0.0016*	0.0841	0.0709*
θ_6	1.1244	-	0.0079	-	0.3429	-
θ_7	0.5882	0.5893	0.0045	0.004*	0.1874	0.1771*
θ_8	0.2247	0.2327	0.0016	0.0016	0.0691	0.0704
θ_9	3.1358	2.9515*	0.0213	0.0204*	0.9489	0.9062*
θ_{10}	0.3752	0.3667*	0.0026	0.0025*	0.1154	0.1121*
θ_{11}	0.1722	0.1777	0.0011	0.0011*	0.0516	0.0528
θ_{12}	0.8167	0.7840*	0.0054	0.0056	0.2426	0.2364*
θ_{13}	1.2167	1.1739*	0.008	0.0079*	0.3676	0.3593*
θ_{14}	0.5859	0.5634*	0.0039	0.0039*	0.1771	0.1766*
θ_{15}	0.2508	0.2684	0.0018	0.0018*	0.0774	0.0794
θ_{16}	0.4012	0.3665*	0.0027	0.0025*	0.1189	0.1101*
θ_{17}	2.6372	2.1034*	0.0179	0.0146*	0.7876	0.6372*
θ_{18}	0.7346	0.5137*	0.0051	0.0037*	0.2175	0.1599*
θ_{19}	0.4529	-	0.0031	-	0.1363	-
θ_{20}	0.3221	0.3228	0.0023	0.0021*	0.0958	0.0961

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A15. Percentile Intervals of Parameters Obtained by Random-X of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 1 PM10 Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.5615	0.4343*	0.0037	0.003*	0.1651	0.1374*
θ_1	0.3492	0.2929*	0.0026	0.002*	0.1041	0.0871*
θ_2	0.3567	-	0.0027	-	0.1107	-
θ_3	0.5636	0.5450*	0.0052	0.0041*	0.1758	0.1698*
θ_4	0.2154	0.2141*	0.0014	0.0014*	0.0666	0.0646*
θ_5	0.2757	0.2517	0.0018	0.0018*	0.0842	0.0766*
θ_6	0.9608	-	0.0066	-	0.2983	-
θ_7	0.5003	0.4397*	0.0034	0.0031*	0.1524	0.1345*
θ_8	0.1963	0.1979	0.0014	0.0014*	0.061	0.061*
θ_9	2.3160	2.0397*	0.0162	0.0162*	0.7057	0.6428*
θ_{10}	0.3703	0.3370*	0.0025	0.0027	0.1108	0.103*
θ_{11}	0.1404	0.1359*	0.001	0.0009*	0.043	0.0411*
θ_{12}	0.7446	0.7156*	0.0053	0.0048*	0.2251	0.2124*
θ_{13}	0.9753	0.9691*	0.0076	0.0065*	0.2999	0.2906*
θ_{14}	0.6143	0.6149	0.0056	0.0047*	0.1909	0.1917
θ_{15}	0.2763	0.2678*	0.0017	0.0016*	0.0827	0.08*
θ_{16}	0.2832	0.2630*	0.0021	0.0017*	0.0847	0.08*
θ_{17}	2.4831	2.1737*	0.0192	0.0172*	0.7719	0.6891*
θ_{18}	0.7831	0.6699*	0.0051	0.0044*	0.2323	0.2013*
θ_{19}	0.4619	-	0.0032	-	0.1409	-
θ_{20}	0.2880	0.2780*	0.0027	0.0027*	0.0895	0.0886*

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A16. Percentile Intervals of Parameters Obtained by Wild Bootstrapping of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 1 PM10 Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.8329	0.4999*	0.0055	0.0034*	0.2558	0.1531*
θ_1	0.5040	0.3467*	0.0034	0.0024*	0.1507	0.1074*
θ_2	0.5321	-	0.0035	-	0.1632	-
θ_3	1.3528	-	0.0083	-	0.4173	-
θ_4	0.3139	0.2764*	0.002	0.002*	0.0958	0.0884*
θ_5	0.4175	-	0.0028	-	0.1275	-
θ_6	1.6759	-	0.0112	-	0.5094	-
θ_7	1.0458	0.5419*	0.0082	0.0036*	0.3254	0.1645*
θ_8	0.3279	0.3060*	0.0021	0.0021*	0.0969	0.0943*
θ_9	4.9698	3.1303*	0.0363	0.0202*	1.4948	0.9399*
θ_{10}	0.6321	-	0.0041	-	0.1942	-
θ_{11}	0.2435	0.2457	0.0016	0.0017	0.0752	0.0754
θ_{12}	1.3032	1.1804*	0.0092	0.0079*	0.3917	0.3577*
θ_{13}	2.0937	0.7942*	0.013	0.0052*	0.6352	0.2355*
θ_{14}	1.0284	-	0.0068	-	0.3114	-
θ_{15}	0.3740	0.3926	0.0025	0.0025*	0.1129	0.1169
θ_{16}	0.5692	0.5167*	0.0038	0.0034*	0.1764	0.159*
θ_{17}	3.7852	-	0.0249	-	1.1492	-
θ_{18}	1.0094	0.6595*	0.007	0.0046*	0.306	0.2023*
θ_{19}	0.6435	-	0.0042	-	0.1981	-
θ_{20}	0.4430	0.4490	0.0032	0.0034	0.137	0.1364*

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A17. BFs from Forward Step of MARS Obtained from the Main Effects Model with Interactions for Fold 2 PM10 Data Set

BF1 = max (0, x1-0.913661)
BF2 = max (0, 0.913661-x1)
BF3 = max (0, x3-0.423946)
BF4 = max (0, 0.423946-x3)
BF5 = max (0, x4+0.0554302)
BF6 = max (0, -0.0554302-x4)
BF7 = max (0, 0.913661-x1)*max (0, x5+0.737813)
BF8 = max (0, 0.913661-x1)*max (0, -0.737813-x5)
BF9 = max (0, 0.913661-x1)*max (0, x7+0.535575)
BF10 = max (0, 0.913661-x1)*max (0, -0.535575-x7)
BF11 = max (0, -0.0554302-x4)*max (0, x7-1.22431)
BF12 = max (0, -0.0554302-x4)*max (0, 1.22431-x7)
BF13 = max (0, 0.423946-x3)*max (0, x4-0.652619)
BF14 = max (0, 0.423946-x3)*max (0, 0.652619-x4)
BF15 = max (0, x2-0.127295)
BF16 = max (0, 0.127295-x2)
BF17 = max (0, -0.0554302-x4)*max (0, x7-0.496428)
BF18 = max (0, 0.913661-x1)*max (0, x3-2.03632)
BF19 = max (0, 0.913661-x1)*max (0, 2.03632-x3)

Table A18. Percentile Intervals of Parameters Obtained by Fixed-X of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 2 PM 10 Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.3611	0.2641	0.0024	0.0019*	0.1122	0.079*
θ_1	2.1895	-	0.0164	-	0.6777	-
θ_2	0.4845	-	0.0033	-	0.1476	-
θ_3	0.3981	0.3185	0.0026	0.0022*	0.1195	0.0963*
θ_4	0.4406	0.2763	0.0028	0.002*	0.1355	0.0847*
θ_5	0.5199	0.2461	0.0035	0.0016*	0.155	0.0748*
θ_6	1.2036	-	0.0119	-	0.3849	-
θ_7	0.1107	0.1034	0.0008	0.0007*	0.0342	0.031*
θ_8	0.7916	0.7216	0.0054	0.0051*	0.2367	0.2233*
θ_9	0.1813	0.1321	0.0012	0.0009*	0.0556	0.0403*
θ_{10}	0.4804	0.3388	0.0032	0.0022*	0.1488	0.1035*
θ_{11}	4.1837	3.8085	0.028	0.0267*	1.2868	1.1424*
θ_{12}	0.6572	0.2614	0.0062	0.0019*	0.2051	0.0775*
θ_{13}	0.4449	-	0.003	-	0.1347	-
θ_{14}	0.3715	0.3054	0.0026	0.0022*	0.1143	0.0959*
θ_{15}	0.3359	0.3159	0.0023	0.0021*	0.1058	0.0938*
θ_{16}	0.2561	0.2299	0.0017	0.0015*	0.0787	0.07*
θ_{17}	1.8645	0.7622	0.0171	0.0053*	0.5953	0.2299*
θ_{18}	3.7680	3.3986	0.0237	0.0223*	1.1172	1.0337*
θ_{19}	0.1851	-	0.0012	-	0.0546	-

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A19. Percentile Intervals of Parameters Obtained by Random-X of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 2 PM 10 Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.4008	0.2950*	0.0029	0.002*	0.1229	0.0883*
θ_1	2.1105	1.2882*	0.0149	0.0089*	0.6542	0.3942*
θ_2	0.5339	-	0.0037	-	0.1568	-
θ_3	0.4798	0.3713*	0.0032	0.0026*	0.1446	0.1155*
θ_4	0.4533	0.3113*	0.0031	0.0021*	0.1352	0.0953*
θ_5	0.5188	0.2444*	0.0034	0.0018*	0.1534	0.075*
θ_6	0.5193	0.3490*	0.0081	0.002*	0.1719	0.1075*
θ_7	0.1306	0.1234*	0.001	0.0008*	0.041	0.0375*
θ_8	1.2387	0.9917*	0.0125	0.0096*	0.3673	0.3041*
θ_9	0.1853	0.1173*	0.0013	0.0008*	0.0567	0.0366*
θ_{10}	0.5032	0.3695*	0.0036	0.0025*	0.152	0.1109*
θ_{11}	3.4493	3.1655*	0.0355	0.0424*	1.111	1.0531*
θ_{12}	0.3386	0.3145*	0.0028	0.0021*	0.1041	0.095*
θ_{13}	0.4286	-	0.0032	-	0.13	-
θ_{14}	0.3417	0.3037*	0.0025	0.0021*	0.1034	0.0925*
θ_{15}	0.3397	0.3013*	0.0024	0.0021*	0.1012	0.0895*
θ_{16}	0.2455	0.2177*	0.0017	0.0014*	0.0734	0.0652*
θ_{17}	1.0092	0.8058*	0.0093	0.0052*	0.3058	0.2461*
θ_{18}	3.8624	3.8507*	0.0283	0.0337	1.1823	1.1975
θ_{19}	0.1957	-	0.0013	-	0.0601	-

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A20. Percentile Intervals of Parameters Obtained by Wild Bootstrapping of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 2 PM10 Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.4689	0.2449*	0.0032	0.0017*	0.1421	0.0754*
θ_1	3.1026	-	0.0206	-	0.9653	-
θ_2	0.6381	-	0.0045	-	0.2034	-
θ_3	0.5241	-	0.0038	-	0.1661	-
θ_4	0.5978	0.2165*	0.0039	0.0021*	0.1787	0.0673*
θ_5	0.6711	-	0.0043	-	0.202	-
θ_6	3.0574	-	0.0188	-	0.9354	-
θ_7	0.1542	0.1298*	0.0011	0.0009*	0.0473	0.0403*
θ_8	1.0190	1.0139*	0.0066	0.0063*	0.3151	0.2984*
θ_9	0.2440	0.1742*	0.0017	0.0012*	0.0747	0.0543*
θ_{10}	0.6449	0.4331*	0.004	0.0029*	0.1902	0.1323*
θ_{11}	7.6795	4.8378*	0.0503	0.0287*	2.2867	1.4354*
θ_{12}	1.5390	0.3038*	0.0095	0.002*	0.4701	0.093*
θ_{13}	0.5922	-	0.0041	-	0.1812	-
θ_{14}	0.4926	0.3266*	0.0033	0.0025*	0.1509	0.1016*
θ_{15}	0.4728	-	0.0033	-	0.1439	-
θ_{16}	0.3440	0.2966*	0.0023	0.0019*	0.1069	0.0887*
θ_{17}	4.4368	0.9381*	0.0284	0.0061*	1.3589	0.2896*
θ_{18}	4.9904	3.6743*	0.0327	0.0221*	1.5194	1.1068*
θ_{19}	0.2453	-	0.0016	-	0.0756	-

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A21. BFs from Forward Step of MARS Obtained from the Main Effects Model with Interactions for Fold 3 PM10 Data Set

BF1 = max (0, x1+0.401651)
BF2 = max (0, -0.401651-x1)
BF3 = max (0, x3+0.221003)
BF4 = max (0, -0.221003-x3)
BF5 = max (0, x2+0.138881)*max (0, x3+0.221003)
BF6 = max (0, -0.138881-x2)*max (0, x3+0.221003)
BF7 = max (0, x7-1.25921)
BF8 = max (0, 1.25921-x7)
BF9 = max (0, x3+0.221003)*max (0, x5+0.74483)
BF10 = max (0, x3+0.221003)*max (0, -0.74483-x5)
BF11 = max (0, x4-0.0457198)*max (0, 1.25921-x7)
BF12 = max (0, 0.0457198-x4)*max (0, 1.25921-x7)
BF13 = max (0, x5-0.852629)*max (0, 1.25921-x7)
BF14 = max (0, 0.852629-x5)*max (0, 1.25921-x7)
BF15 = max (0, -0.221003-x3)*max (0, x6+0.794168)
BF16 = max (0, -0.221003-x3)*max (0, -0.794168-x6)
BF17 = max (0, x7-0.960082)
BF18 = max (0, x3+0.221003)*max (0, x4-0.34917)
BF19 = max (0, x3+0.221003)*max (0, 0.34917-x4)

Table A22. Percentile Intervals of Parameters Obtained by Fixed-X of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 3 PM10 Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.4940	0.3860*	0.0032	0.0027*	0.1524	0.1191*
θ_1	0.3181	0.2604*	0.0023	0.0018*	0.0987	0.0809*
θ_2	0.3821	-	0.0026	-	0.1175	-
θ_3	0.4669	-	0.0029	-	0.1451	-
θ_4	0.6524	-	0.0043	-	0.1964	-
θ_5	0.3859	0.4029	0.0026	0.0026*	0.1157	0.1194
θ_6	0.4207	0.3914*	0.0028	0.0026*	0.1255	0.1197*
θ_7	3.4964	3.0799*	0.0432	0.0321*	1.2001	1.0628*
θ_8	0.2630	0.2390*	0.0019	0.0015*	0.0806	0.072*
θ_9	0.2654	0.2109*	0.0018	0.0015*	0.0803	0.0644*
θ_{10}	1.0060	0.7516*	0.0067	0.0049*	0.3121	0.2258*
θ_{11}	0.1470	0.1475	0.001	0.001*	0.0451	0.0451*
θ_{12}	0.1719	0.1744	0.0011	0.0012	0.0512	0.0518
θ_{13}	0.3044	-	0.0021	-	0.0941	-
θ_{14}	0.1444	0.1238*	0.001	0.0008*	0.0425	0.0373*
θ_{15}	0.3842	0.2551*	0.0026	0.0017*	0.1153	0.0784*
θ_{16}	1.5930	1.2644*	0.0105	0.0084*	0.4778	0.3833*
θ_{17}	1.6896	1.6429*	0.0138	0.0116*	0.5229	0.4919*
θ_{18}	1.7358	1.6004*	0.0119	0.011*	0.5357	0.4757*
θ_{19}	0.3639	0.3569*	0.0026	0.0026*	0.1125	0.1115*

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A23. Percentile Intervals of Parameters Obtained by Random-X of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 3 PM10 Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.4856	0.3747*	0.0034	0.0025*	0.1465	0.1162*
θ_1	0.3262	0.2936*	0.0022	0.0018*	0.0999	0.0878*
θ_2	0.4045	-	0.0028	-	0.1217	-
θ_3	0.6624	-	0.0043	-	0.1984	-
θ_4	0.6449	-	0.0045	-	0.1991	-
θ_5	0.4013	0.3064*	0.0029	0.0023*	0.1238	0.0946*
θ_6	0.5206	0.5004*	0.0048	0.0044*	0.164	0.1562*
θ_7	1.2349	1.8221	0.0317	0.0344	0.4341	0.5991
θ_8	0.2997	0.2528*	0.002	0.0016*	0.0911	0.0772*
θ_9	0.3532	0.2557*	0.0022	0.0017*	0.107	0.0788*
θ_{10}	1.3608	0.8236*	0.0097	0.0069*	0.4162	0.2487*
θ_{11}	0.1789	0.1738*	0.0014	0.0016	0.0545	0.055
θ_{12}	0.1893	0.1768*	0.0016	0.0013*	0.0597	0.053*
θ_{13}	0.299	-	0.002	-	0.0926	-
θ_{14}	0.1649	0.1376*	0.0011	0.0009*	0.0502	0.0421*
θ_{15}	0.3752	0.2654*	0.0025	0.0018*	0.1118	0.0798*
θ_{16}	1.5711	1.1490*	0.012	0.0075*	0.4907	0.3487*
θ_{17}	1.2086	1.2142	0.008	0.0114	0.3665	0.3844
θ_{18}	1.6782	1.6414*	0.0178	0.0163*	0.5348	0.5313*
θ_{19}	0.3525	-	0.0028	-	0.1078	-

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A24. Percentile Intervals of Parameters Obtained by Wild Bootstrapping of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 3 PM10 Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.6632	0.5067*	0.0047	0.0035*	0.2	0.1524*
θ_1	0.4780	0.3929*	0.003	0.0026*	0.1416	0.1192*
θ_2	0.5606	-	0.0038	-	0.1661	-
θ_3	0.7077	-	0.0046	-	0.2136	-
θ_4	0.8693	-	0.0064	-	0.2641	-
θ_5	0.5330	0.3745*	0.0036	0.0026*	0.163	0.1159*
θ_6	0.5864	-	0.0038	-	0.1788	-
θ_7	4.7908	4.5729*	0.0536	0.0483*	1.6091	1.5516*
θ_8	0.3532	0.3185*	0.0024	0.002*	0.1078	0.0989*
θ_9	0.3701	0.2947*	0.0023	0.0017*	0.1124	0.0865*
θ_{10}	1.5266	1.0100*	0.0105	0.0065*	0.4459	0.2936*
θ_{11}	0.2162	0.2116*	0.0013	0.0013*	0.0657	0.064*
θ_{12}	0.2406	0.2324*	0.0015	0.0015*	0.0733	0.0702*
θ_{13}	0.4462	-	0.0029	-	0.1346	-
θ_{14}	0.1946	0.1764*	0.0013	0.0012*	0.0588	0.0515
θ_{15}	0.5267	0.3383*	0.0035	0.0023*	0.1576	0.1051
θ_{16}	2.2290	1.6441*	0.0151	0.0113*	0.6772	0.5134
θ_{17}	2.2199	2.1806*	0.0168	0.0174	0.6998	0.6802
θ_{18}	2.4348	-	0.0154	-	0.729	-
θ_{19}	0.527	-	0.0037	-	0.1603	-

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A25. BFs from Forward Step of MARS Obtained from the Main Effects Model with Interactions for Fold 1 CS Data Set

BF1 = max (0, x4+0.305221)
BF2 = max (0, -0.305221-x4)
BF3 = max (0, x2-0.132749)
BF4 = max (0, 0.132749-x2)
BF5 = max (0, x4+0.305221)*max (0, x6-0.000241644)
BF6 = max (0, x4+0.305221)*max (0, 0.000241644-x6)
BF7 = max (0, x2-0.132749)*max (0, x7+0.383708)
BF8 = max (0, x2-0.132749)*max (0, -0.383708-x7)
BF9 = max (0, x1-0.936973)*max (0, x2-0.132749)
BF10 = max (0, 0.936973-x1)*max (0, x2-0.132749)
BF11 = max (0, x4+0.305221)*max (0, x7-0.274622)
BF12 = max (0, x4+0.305221)*max (0, 0.274622-x7)
BF13 = max (0, x5+0.0854152)
BF14 = max (0, -0.0854152-x5)
BF15 = max (0, x2-0.228678)*max (0, x5+0.0854152)
BF16 = max (0, 0.228678-x2)*max (0, x5+0.0854152)
BF17 = max (0, x3-0.175436)*max (0, x4+0.305221)
BF18 = max (0, 0.175436-x3)*max (0, x4+0.305221)
BF19 = max (0, x3-1.05581)
BF20 = max (0, 1.05581-x3)

Table A26. Percentile Intervals of Parameters Obtained by Fixed-X of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 1 CS Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.8349	0.4618*	0.0056	0.0033*	0.2512	0.1439*
θ_1	1.1581	-	0.0078	-	0.3587	-
θ_2	0.6468	0.5672*	0.0046	0.0037*	0.1969	0.1719*
θ_3	1.6632	1.3112*	0.0101	0.008*	0.5048	0.3979*
θ_4	0.4771	-	0.0032	-	0.1443	-
θ_5	1.2562	1.2432*	0.009	0.0084*	0.3744	0.3599*
θ_6	0.6149	-	0.0042	-	0.1934	-
θ_7	0.6993	0.7179	0.0048	0.0048*	0.2115	0.2214
θ_8	1.2971	-	0.0086	-	0.3949	-
θ_9	4.6015	-	0.0329	-	1.4312	-
θ_{10}	0.6470	0.6700	0.0044	0.0043*	0.1983	0.2039
θ_{11}	0.4756	-	0.0032	-	0.1415	-
θ_{12}	0.693	-	0.0046	-	0.2054	-
θ_{13}	0.6083	0.5739*	0.0042	0.0041*	0.1849	0.1715*
θ_{14}	0.6488	-	0.0043	-	0.1984	-
θ_{15}	0.8789	0.6978*	0.0054	0.0056	0.2577	0.2172*
θ_{16}	0.7496	0.6847*	0.0051	0.0044*	0.2312	0.2023*
θ_{17}	0.7511	-	0.005	-	0.2311	-
θ_{18}	0.5839	0.4484*	0.004	0.003*	0.1815	0.1342*
θ_{19}	10.1595	-	0.0603	-	3.0448	-
θ_{20}	0.3558	0.3215*	0.0024	0.0021*	0.1109	0.0973*

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A27. Percentile Intervals of Parameters Obtained by Random-X of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 1 CS Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	1.1536	0.5221*	0.0082	0.0034*	0.352	0.1564*
θ_1	1.3211	-	0.0098	-	0.3964	-
θ_2	0.7964	0.6475*	0.0065	0.0043*	0.2441	0.1936*
θ_3	1.7350	1.6127*	0.02	0.0130*	0.5507	0.4801*
θ_4	0.5835	-	0.0041	-	0.1766	-
θ_5	2.0669	1.6124*	0.022	0.0163*	0.6481	0.5043*
θ_6	0.7511	-	0.006	-	0.2321	-
θ_7	1.0749	0.7490*	0.011	0.0054*	0.3335	0.2315*
θ_8	1.873	-	0.0246	-	0.6126	-
θ_9	5.8572	-	0.0431	-	1.7984	-
θ_{10}	0.8894	0.8761*	0.0085	0.0066*	0.2681	0.2554*
θ_{11}	0.7806	-	0.008	-	0.2363	-
θ_{12}	0.8853	-	0.0075	-	0.2774	-
θ_{13}	0.6347	0.3962*	0.0092	0.0082*	0.2117	0.1386*
θ_{14}	0.9289	-	0.0063	-	0.2866	-
θ_{15}	0.8715	0.6244*	0.0122	0.0092*	0.2773	0.2001*
θ_{16}	0.7506	0.6612*	0.0082	0.0405*	0.2462	0.2554*
θ_{17}	1.2385	-	0.01	-	0.3803	-
θ_{18}	0.8918	0.5349*	0.0065	0.0040*	0.2577	0.1591*
θ_{19}	5.3385	-	0.0524	-	1.7238	-
θ_{20}	0.4319	0.2995*	0.0034	0.0023*	0.1326	0.0911*

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A28. Percentile Intervals of Parameters Obtained by Wild Bootstrapping of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 1 CS Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	1.8352	0.3401*	0.012	0.0022 *	0.5486	0.1039*
θ_1	2.9865	-	0.0178	-	0.8939	-
θ_2	1.5587	0.9879*	0.0101	0.0069*	0.4805	0.3094*
θ_3	5.1004	-	0.0332	-	1.533	-
θ_4	1.0801	-	0.0073	-	0.3291	-
θ_5	2.8467	-	0.019	-	0.8729	-
θ_6	1.598	-	0.01	-	0.4841	-
θ_7	1.6127	-	0.0102	-	0.4951	-
θ_8	2.9106	-	0.0188	-	0.8953	-
θ_9	12.5849	-	0.0778	-	3.816	-
θ_{10}	2.1023	-	0.0136	-	0.6324	-
θ_{11}	1.1046	-	0.0068	-	0.3286	-
θ_{12}	1.6252	-	0.0108	-	0.4983	-
θ_{13}	1.7630	1.1073*	0.0104	0.0071*	0.5351	0.3327*
θ_{14}	1.4583	-	0.01	-	0.4464	-
θ_{15}	2.3503	1.1835*	0.0144	0.0083 *	0.7114	0.3609*
θ_{16}	1.905	-	0.0121	-	0.5626	-
θ_{17}	1.6945	-	0.0112	-	0.5173	-
θ_{18}	1.4136	-	0.0098	-	0.4283	-
θ_{19}	32.4422	-	0.1873	-	9.5966	-
θ_{20}	0.8394	-	0.0052	-	0.2519	-

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A29. BFs from Forward Step of MARS Obtained from the Main Effects Model with Interactions for Fold 2 CS Data Set

BF1 = max (0, x4+0.161715)
BF2 = max (0, -0.161715-x4)
BF3 = max (0, x2+0.548677)
BF4 = max (0, -0.548677-x2)
BF5 = max (0, x2-0.860487)*max (0, x4+0.161715)
BF6 = max (0, 0.860487-x2)*max (0, x4+0.161715)
BF7 = max (0, x1-0.852031)*max (0, x2+0.548677)
BF8 = max (0, 0.852031-x1)*max (0, x2+0.548677)
BF9 = max (0, -0.548677-x2)*max (0, x5+0.192271)
BF10 = max (0, -0.548677-x2)*max (0, -0.192271-x5)
BF11 = max (0, x4+0.161715)*max (0, x7-0.274622)
BF12 = max (0, x4+0.161715)*max (0, 0.274622-x7)
BF13 = max (0, x2+0.548677)*max (0, x4+0.899041)
BF14 = max (0, x2+0.548677)*max (0, -0.899041-x4)
BF15 = max (0, x4+0.161715)*max (0, x6-1.31485)
BF16 = max (0, x4+0.161715)*max (0, 1.31485-x6)
BF17 = max (0, -0.548677-x2)*max (0, x3-0.902449)
BF18 = max (0, -0.548677-x2)*max (0, 0.902449-x3)
BF19 = max (0, x1-1.14236)*max (0, -0.161715-x4)
BF20 = max (0, 1.14236-x1)*max (0, -0.161715-x4)

Table A30. Percentile Intervals of Parameters Obtained by Fixed-X of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 2 CS Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.6194	0.4656*	0.0042	0.003*	0.1823	0.1378*
θ_1	1.4771	-	0.0104	-	0.4384	-
θ_2	0.9296	0.5569*	0.0067	0.0035*	0.2792	0.1669*
θ_3	0.6639	0.4111*	0.0051	0.0035*	0.2064	0.1265*
θ_4	1.4081	-	0.0097	-	0.4252	-
θ_5	3.1569	3.5235	0.0249	0.0245*	0.9732	1.0368
θ_6	0.6804	-	0.0048	-	0.212	-
θ_7	1.2528	-	0.0087	-	0.3829	-
θ_8	0.3053	0.2139*	0.002	0.0017*	0.0937	0.0665*
θ_9	1.726	-	0.0122	-	0.5221	-
θ_{10}	1.2674	0.9742*	0.0088	0.0071*	0.3853	0.3014*
θ_{11}	0.6364	-	0.0045	-	0.1915	-
θ_{12}	0.7621	-	0.0058	-	0.2326	-
θ_{13}	0.8308	0.3796*	0.0057	0.0026*	0.2525	0.1168*
θ_{14}	1.0763	-	0.0077	-	0.3215	-
θ_{15}	10.1908	-	0.0764	-	3.0414	-
θ_{16}	0.4930	0.1710*	0.0036	0.0012*	0.1534	0.0527*
θ_{17}	7.5678	-	0.0516	-	2.3094	-
θ_{18}	1.039	-	0.0078	-	0.32	-
θ_{19}	14.0316	-	0.1124	-	4.1751	-
θ_{20}	0.4856	-	0.0031	-	0.146	-

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A31. Percentile Intervals of Parameters Obtained by Random-X of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 2 CS Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.541	0.4054*	0.005	0.0029*	0.1722	0.1274*
θ_1	1.625	-	0.0097	-	0.4923	-
θ_2	1.037	0.5729*	0.0163	0.0046*	0.3315	0.1749*
θ_3	0.862	0.3410*	0.0062	0.0044*	0.2617	0.1126*
θ_4	1.377	-	0.0103	-	0.4175	-
θ_5	14.925	5.8688*	0.3832	0.2498*	5.1734	3.5565*
θ_6	0.951	-	0.0074	-	0.3018	-
θ_7	3.882	-	0.0394	-	1.1874	-
θ_8	0.487	0.2579*	0.0033	0.002*	0.1459	0.0786*
θ_9	2.227	-	0.0275	-	0.7411	-
θ_{10}	1.276	0.9635*	0.0409	0.0071*	0.4683	0.2856*
θ_{11}	1.059	-	0.0096	-	0.3097	-
θ_{12}	1.062	-	0.0117	-	0.34	-
θ_{13}	0.994	0.3209*	0.0075	0.0031*	0.3058	0.0993*
θ_{14}	1.062	-	0.0164	-	0.3667	-
θ_{15}	200.365	-	37.4586	-	204.2981	-
θ_{16}	0.717	-	0.0077	-	0.2209	-
θ_{17}	9.743	-	0.183	-	3.4413	-
θ_{18}	1.271	-	0.0213	-	0.4129	-
θ_{19}	17.690	-	3.7782	-	32.0593	-
θ_{20}	0.546	0.3031*	0.0072	0.0022*	0.1751	0.0922*

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A32. Percentile Intervals of Parameters Obtained by Wild Bootstrapping of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 2 CS Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	1.8182	0.5300*	0.0124	0.0036 *	0.5626	0.1592*
θ_1	4.9591	-	0.031	-	1.4927	-
θ_2	2.5576	0.8149*	0.0167	0.0059*	0.7677	0.2543*
θ_3	2.3307	0.5567*	0.0147	0.0040*	0.701	0.1750*
θ_4	4.2330	-	0.0286	-	1.2871	-
θ_5	8.6484	-	0.0485	-	2.6596	-
θ_6	2.0279	-	0.0131	-	0.5945	-
θ_7	3.3377	-	0.0195	-	1.017	-
θ_8	0.9646	-	0.0059	-	0.2919	-
θ_9	4.2401	-	0.0267	-	1.2775	-
θ_{10}	3.4505	-	0.0218	-	1.0069	-
θ_{11}	1.5732	-	0.0098	-	0.4709	-
θ_{12}	1.9820	-	0.0128	-	0.6068	-
θ_{13}	2.2874	-	0.0158	-	0.6967	-
θ_{14}	2.6431	-	0.0178	-	0.8293	-
θ_{15}	24.4069	-	0.1399	-	7.8168	-
θ_{16}	1.4307	-	0.0095	-	0.4446	-
θ_{17}	18.3490	-	0.1171	-	5.6617	-
θ_{18}	2.7393	-	0.0173	-	0.8269	-
θ_{19}	34.3476	-	0.1895	-	10.3674	-
θ_{20}	1.2522	-	0.0084	-	0.3806	-

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A33. BFs from Forward Step of MARS Obtained from the Main Effects Model with Interactions for Fold 3 CS Data Set

BF1 = max (0, (x4+0.701101)
BF2 = max (0, (-0.701101-x4)
BF3 = max (0, (x2+0.0326454)
BF4 = max (0, (-0.0326454-x2)
BF5 = max (0, (x1+0.860758)*max (0, (x2+0.0326454)
BF6 = max (0, (-0.860758-x1)*max (0, (x2+0.0326454)
BF7 = max (0, (x3-0.491529)*max (0, (-0.701101-x4)
BF8 = max (0, (0.491529-x3)*max (0, (-0.701101-x4)
BF9 = max (0, (x1-0.229544)*max (0, (-0.701101-x4)
BF10 = max (0, (0.229544-x1)*max (0, (-0.701101-x4)
BF11 = max (0, (-0.701101-x4)*max (0, (x6-0.776335)
BF12 = max (0, (-0.701101-x4)*max (0, (0.776335-x6)
BF13 = max (0, (-0.0326454-x2)*max (0, (x3-1.04762)
BF14 = max (0, (-0.0326454-x2)*max (0, (1.04762-x3)
BF15 = max (0, (x3-1.00547)
BF16 = max (0, (1.00547-x3)
BF17 = max (0, (x1-0.394357)*max (0, (1.00547-x3)
BF18 = max (0, (0.394357-x1)*max (0, (1.00547-x3)
BF19 = max (0, (x2+0.0326454)*max (0, (x3+0.32797)
BF20 = max (0, (x2+0.0326454)*max (0, (-0.32797-x3)

Table A34. Percentile Intervals of Parameters Obtained by Fixed-X of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 3 CS Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.8402	0.4044*	0.0056	0.0027*	0.2585	0.1212*
θ_1	0.2615	0.2375*	0.0017	0.0016*	0.0789	0.0732*
θ_2	1.8106	0.5181*	0.0142	0.005*	0.5657	0.1607*
θ_3	1.651	-	0.0114	-	0.5046	-
θ_4	0.8568	-	0.0057	-	0.2604	-
θ_5	0.7944	0.6826*	0.0056	0.0043*	0.2399	0.2075*
θ_6	2.3505	2.6173	0.017	0.0171	0.6977	0.787
θ_7	2.7446	-	0.0193	-	0.8563	-
θ_8	4.3385	4.1189*	0.028	0.0269*	1.3232	1.2308*
θ_9	2.5137	2.0898*	0.0178	0.0133*	0.7559	0.6254*
θ_{10}	1.552	-	0.0103	-	0.479	-
θ_{11}	1.9196	-	0.0133	-	0.5777	-
θ_{12}	3.1979	2.8287*	0.0227	0.0189*	0.9513	0.8528*
θ_{13}	10.1284	-	0.0679	-	3.058	-
θ_{14}	0.6883	0.5021*	0.0047	0.0033*	0.2121	0.1549*
θ_{15}	9.7712	-	0.0604	-	2.9353	-
θ_{16}	0.563	-	0.0036	-	0.1691	-
θ_{17}	0.5640	0.5211*	0.0038	0.0034*	0.1675	0.1576*
θ_{18}	0.5008	0.2856*	0.0032	0.0021*	0.1505	0.0854*
θ_{19}	1.3428	0.4459*	0.009	0.0031*	0.4081	0.1362*
θ_{20}	0.9351	-	0.0061	-	0.2819	-

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A35. Percentile Intervals of Parameters Obtained by Random-X of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 3 CS Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	1.2749	0.4408*	0.0114	0.0032*	0.3872	0.1368*
θ_1	0.3411	0.2921*	0.0026	0.0021*	0.106	0.0883*
θ_2	6.6228	-	0.0564	-	1.9939	-
θ_3	2.4685	-	0.0172	-	0.737	-
θ_4	1.2987	-	0.0099	-	0.3962	-
θ_5	1.8741	0.5438*	0.013	0.0051*	0.5977	0.1699*
θ_6	4.7389	-	0.041	-	1.5475	-
θ_7	8.1824	-	0.126	-	2.7145	-
θ_8	18.3398	-	1.1407	-	10.3677	-
θ_9	35.9401	-	367736.2	-	1668763	-
θ_{10}	6.1838	-	0.3062	-	2.5232	-
θ_{11}	6.7286	-	0.8044	-	3.9658	-
θ_{12}	6.4461	2.8157*	0.5899	0.1193*	3.9535	0.997*
θ_{13}	19.0219	-	0.3998	-	6.8181	-
θ_{14}	1.1699	0.4074*	0.0091	0.0044*	0.3614	0.1302*
θ_{15}	10.8229	-	0.0682	-	3.2157	-
θ_{16}	0.9299	-	0.0065	-	0.2749	-
θ_{17}	1.2579	-	0.0127	-	0.3787	-
θ_{18}	0.6791	-	0.0058	-	0.208	-
θ_{19}	2.5027	0.6750*	0.0233	0.0067*	0.7733	0.2156*
θ_{20}	1.5654	-	0.0208	-	0.5121	-

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A36. Percentile Intervals of Parameters Obtained by Wild Bootstrapping of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 3 CS Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	3.0397	0.1216*	0.0187	0.0007*	0.9236	0.0360*
θ_1	0.7821	-	0.0049	-	0.2306	-
θ_2	7.6769	-	0.0482	-	2.3349	-
θ_3	5.0627	-	0.0333	-	1.5455	-
θ_4	2.6152	-	0.0171	-	0.795	-
θ_5	2.2105	-	0.0146	-	0.658	-
θ_6	6.1088	-	0.0391	-	1.9091	-
θ_7	11.1186	-	0.0738	-	3.3338	-
θ_8	13.0314	-	0.0756	-	3.9102	-
θ_9	7.5366	-	0.0485	-	2.254	-
θ_{10}	4.8937	-	0.0322	-	1.4809	-
θ_{11}	5.9532	-	0.0388	-	1.7927	-
θ_{12}	9.0160	3.7621*	0.0525	0.0205*	2.7377	1.1144*
θ_{13}	46.2720	-	0.2842	-	13.8551	-
θ_{14}	2.0107	-	0.0123	-	0.5995	-
θ_{15}	48.6094	-	0.2929	-	14.73	-
θ_{16}	1.8723	-	0.0132	-	0.5794	-
θ_{17}	1.5336	-	0.0097	-	0.4765	-
θ_{18}	1.4380	-	0.0096	-	0.4351	-
θ_{19}	3.9585	-	0.0257	-	1.2165	-
θ_{20}	3.0357	-	0.0194	-	0.9174	-

* indicates more precise parameter estimate

- shows statistically insignificant model parameter

Table A37. BFs from Forward Step of MARS Obtained from the Main Effects Model with Interactions for Fold 1 US Data Set

BF1 = max (0, x1-0.384375)
BF2 = max (0, 0.384375-x1)

Table A38. Percentile Intervals of Parameters Obtained by Fixed-X of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 1 US Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.0481	0.0481	0.0003	0.0003	0.0147	0.0147
θ_1	0.1377	0.1377	0.001	0.001	0.0422	0.0422
θ_2	0.2185	0.2185	0.0015	0.0015	0.0675	0.0675

Table A39. Percentile Intervals of Parameters Obtained by Random-X of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 1 US Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.0203	0.0203	0.0001	0.0001	0.0063	0.0063
θ_1	0.1162	0.1162	0.0008	0.0008	0.0353	0.0353
θ_2	0.1769	0.1769	0.0012	0.0012	0.0537	0.0537

Table A40. Percentile Intervals of Parameters Obtained by Wild Bootstrapping of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 1 US Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.5498	0.5498	0.0037	0.0037	0.0037	0.0037
θ_1	2.0171	2.0171	0.0133	0.0133	0.0133	0.0133
θ_2	3.2669	3.2669	0.0219	0.0219	0.0219	0.0219

Table A41. BFs from Forward Step of MARS Obtained from the Main Effects Model with Interactions for Fold 2 US Data Set

BF1 = max (0,x1-0.190625)
BF2 = max (0, 0.190625-x1)

Table A42. Percentile Intervals of Parameters Obtained by Fixed-X of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 2 US Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.0034	0.0034	0.0001	0.0001	0.0011	0.0011
θ_1	0.0075	0.0075	0.0001	0.0001	0.0023	0.0023
θ_2	0.0485	0.0485	0.0003	0.0003	0.0148	0.0148

Table A43. Percentile Intervals of Parameters Obtained by Random-X of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 2 US Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.0037	0.0037	0.0001	0.0001	0.0011	0.0011
θ_1	0.0073	0.0073	0.0001	0.0001	0.0023	0.0023
θ_2	0.0515	0.0515	0.0003	0.0003	0.0158	0.0158

Table A44. Percentile Intervals of Parameters Obtained by Wild Bootstrapping of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 2 US Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.4365	0.2056*	0.0027	0.0025*	0.1357	0.0754*
θ_1	1.0947	0.6084*	0.007	0.0076	0.3417	0.2231*
θ_2	5.2515	-	0.038	-	1.6408	-

Table A45. The BFs from Forward Step of MARS Obtained from the Main Effects Model with Interactions for Fold 3 US Data Set

BF1 = max (0, x1-0.196875)
BF2 = max (0, 0.196875-x1)

Table A46. Percentile Intervals of Parameters Obtained by Fixed-X of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 2 US Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.0108	0.0108	0.0001	0.0001	0.0033	0.0033
θ_1	0.0244	0.0244	0.0002	0.0002	0.0075	0.0075
θ_2	0.0992	0.0992	0.0008	0.0008	0.0304	0.0304

Table A47. Percentile Intervals of Parameters Obtained by Random-X of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 2 US Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.0039	0.0039	~0	~0	0.0012	0.0012
θ_1	0.0077	0.0077	0.0001	0.0001	0.0023	0.0023
θ_2	0.0444	0.0444	0.0003	0.0003	0.0137	0.0137

Table A48. Percentile Intervals of Parameters Obtained by Wild Bootstrapping of the Main Effects Model with Interactions at $\alpha = 0.1$ for Fold 2 US Data Set

Parameter	Length of CIs for CMARS	Length of CIs for BCMARS	STD (BS)		STD	
			CMARS	BCMARS	CMARS	BCMARS
θ_0	0.5445	0.3914*	0.0037	0.0028*	0.1661	0.1187*
θ_1	1.4967	1.2621*	0.0099	0.0090*	0.4423	0.3829*
θ_2	8.5869	-	0.0531	-	2.582	-