APPLIED FINANCE LETTERS

VOLUME 11, 2022

BOAREN: IMPROVING REGULARIZATION IN LINEAR REGRESSION WITH AN APPLICATION TO INDEX TRACKING

JOHN E. ANGUS[1]*, YUJIA DING[1] AND QIDI PENG[1]

1. Claremont Graduate University, USA

* Corresponding Author: John E. Angus, Institute of Mathematical Sciences, Claremont Graduate University, Claremont, California, USA. * : john.angus@cgu.com

Abstract

In this paper we introduce the Arbitrary Rectangle-range Elastic Net (AREN): an elastic net with coefficients restricted to some rectangle in $\mathbb{R}^{\mathbf{p}}$, $\mathbf{p} \geq \mathbf{1}$. The AREN method is one of many regularization techniques intended to increase prediction accuracy in linear regression models by shrinking the magnitude (and possibly eliminating some) of the regression coefficients in an effort to control over-fitting and under-fitting. In this work we describe the AREN features and discuss its statistical consistency properties in estimation and in selecting the correct set of predictors. We also introduce bootstrapping as a way to improve the "small-sample" performance of AREN in selecting predictors. We then apply the AREN (with and without bootstrapping) to tracking the value of the S&P 500 index using a reduced set of stocks.

MSC2020 subject classifications: Primary 62P05, 62J07; secondary 62F12.

Keywords: Arbitrary rectangle-range elastic net, variable selection, asymptotic consistency, bootstrap.

1. Introduction

Variable selection and regularization are essential tools in high-dimensional data analysis. They aim at deriving the most valuable information from the data by finding the right balance of bias (under-fitting) and variance (over-fitting) to optimize the model's prediction capability. Perhaps the earliest example of this type of regularization is the so-called "Ridge Regression" which enforces a penalty proportional to the squared l_2-norm of the regression coefficients in the least squares estimation problem. The "lasso" (Least Absolute Shrinkage and Selection Operator) (Tibshirani, 1996) replaces the squared l_2-norm penalty in Ridge Regression with an l_1-norm penalty, which adds the benefit of actually assigning 0 to certain regression coefficients. Due to its computational efficiency (Efron et al., 2004), variable selection consistency (Zhao and Yu, 2006), and estimation consistency (Negahban et al., 2012), lasso has overtaken the popularity of Ridge Regression. Refer to (Bickel et al., 2009; Efron et al., 2007; Lounici, 2008; Wang et al., 2007; Yuan and Lin, 2006; Zhao et al., 2009; Zou, 2006) for more in-depth discussions of lasso. Recently, the elastic net was introduced to extend the lasso (Zou and Hastie, 2005). This method involves linearly combining the lasso and ridge regression-like penalties.

Recall that in the classical regression each regression coefficient can assume any value in the real numbers; they are not constrained in any way. However, there can be practical constraints on the regression coefficients; some may be bounded, some may be restricted to be positive or negative. For example, it is known that body height is positively correlated to age; allocations (as a fraction of the total) of assets in a fund should be in $[0,1]$. Based on the above concerns in practice, it is natural to consider regressions with coefficients restricted to some specific range. For instance, Wu et al. (2014) and Wu and Yang (2014) introduced the non-negative lasso and non-negative elastic net to solve the index tracking problem without short sales (with non-negative constraints on weights). More flexible methods are needed to address problems that require arbitrary constraints. To this end, in this paper

we examine a recently developed method that assumes the regression coefficients to be in some rectangular range. This model, the arbitrary rectangle-range elastic net method (abbreviated throughout AREN), is a regularization method that deals with high-dimensional problems, and most importantly, generalizes and outperforms the lasso, ridge, and non-negative elastic net. Compared with the non-negative elastic net, AREN allows adding arbitrary lower and upper constraints on the coefficients. This feature makes AREN more adaptable to practical problems. We summarize the contribution of our paper as follows:

1. We introduce AREN to increase the adaptability of the elastic net method when dealing with regressions with constrained-range coefficients.

2. Sufficient conditions for estimation consistency and variable selection consistency of the AREN are discussed.

3. We apply AREN to the problem of tracking the S&P 500 index and following show that AREN (and likely other similar regularization approaches) can be improved through the use of bootstrapping.

The paper is organized as follows. In Section 2, we introduce the mathematical model of AREN and survey its estimation consistency (Theorem 2.1) and variable selection consistency properties (Theorem 2.4). Section 3 is devoted to an application of two-step AREN and bootstrapped two-step AREN to the practical problem of S&P 500 index-tracking. The interested reader may refer to Ding et al. (2021) for simulations that compare the performance of a variety of similar methods of this type.

2. The AREN

2.1 Definition and Basic Setup

Throughout the paper, the transpose of a matrix A is denoted by A'. The i-th column of A is denoted by A_i, and the entry in the i-th row and j-th column of A is expressed as A_{ij}. The notation $\max(v)$ (resp. $\min(v)$) signifies the maximum (resp. minimum) element of the vector or the matrix v. When necessary to identify the elements of an $n \times n$ matrix X we write $X = (X_{ij})_{1\le i,j\le n}$. The element-wise absolute value of the matrix $X = (X_{ij})_{1\le i,j\le n}$ is $|X| = (|X_{ij}|)_{1\le i,j\le n}$ with obvious modification if X is a vector. From two vectors $\mathbf{x} = (x_1, \dots, x_p)$, $\mathbf{y} = (y_1, \dots, y_p)$, we define the corresponding rectangle in $\mathbb{R}^p$ as the cartesian product $[\mathbf{x}, \mathbf{y}] = [x_1, y_1] \times \cdots \times [x_p, y_p]$.

Let us consider the linear regression model

$$Y = X\beta^* + \epsilon, \tag{2.1}$$

where X is a deterministic $n \times p$ design matrix, $Y = (y_1 \dots y_n)'$ is an $n \times 1$ response vector and $\epsilon = (\epsilon_1 \dots \epsilon_n)'$ is a zero-mean Gaussian noise vector with $\mathbb{V}\mathrm{ar}(\epsilon_1) = \sigma^2$. Without loss of generality, we assume all the predictors are centered, so the intercept is not included. $\beta^* \in \mathbb{R}^p$ denotes the vector of regression coefficients.

When p is large, it is natural to assume that the linear model, Equation (2.1), is q-sparse; i.e., β^* has at most q ($q \ll p$) nonzero elements. For the AREN regularization, we assume there is a rectangular region $\mathcal{I} = [\mathbf{s}, \mathbf{t}]$ in $\mathbb{R}^p$ that contains β^*, with $\mathbf{s} = (s_1, \dots, s_p)$, $\mathbf{t} = (t_1, \dots, t_p)$, $s_i \in \mathbb{R} \cup \{-\infty\}$, $t_i \in \mathbb{R} \cup \{+\infty\}$, $s_i < t_i$ for all $i = 1, \dots, p$. For the linear model Equation (2.1), the AREN estimator of β^* is given by

$$\hat{\beta}\left(\lambda_n^{(1)},\lambda_n^{(2)}\right) = \underset{\beta\in\mathcal{I}}{\arg\min}\left(\frac{1}{2n}\|Y - X\beta\|_2^2 + \lambda_n^{(1)}\|\beta\|_1 + \lambda_n^{(2)}\|\beta\|_2^2\right). \quad (2.2)$$

Here $\lambda_n^{(1)},\lambda_n^{(2)} \geq 0$ are tuning parameters which control the importance of the l_1 and l_2 regularization terms, respectively. These are typically tuned to minimize the prediction mean-squared error by repeatedly performing AREN on training data for each pair in a lattice of values of $\lambda_n^{(1)},\lambda_n^{(2)} \geq 0$, using the resulting model to predict the responses for an out-of-sample testing set and computing the observed mean-squared error in those predictions. A search over that lattice can then select the pair with the smallest mean-squared error.

The AREN, Equation (2.2) method extends the elastic net method when $\mathcal{I} = \mathbb{R}^p$. It extends the non-negative elastic net when $\mathcal{I} = [0,+\infty)^p$.

Observe that, taking $\dot{X} = X/\sqrt{2n}$ and $\dot{Y} = Y/\sqrt{2n}$, the mean-squared error loss function in Equation *(2.2)* can be transformed to the residual sum of squares loss function, i.e., Equation *(2.2)* becomes

$$\hat{\beta}\left(\lambda_n^{(1)},\lambda_n^{(2)}\right) = \underset{\beta\in\mathcal{I}}{\arg\min}\left(\left\|\dot{Y} - \dot{X}\beta\right\|_2^2 + \lambda_n^{(1)}\|\beta\|_1 + \lambda_n^{(2)}\|\beta\|_2^2\right), \quad (2.3)$$

which is a particular case of the Arbitrary Rectangle-range Generalized Elastic Net (ARGEN) studied in Ding et al. (2021). As a result, the AREN problem, Equation (2.2), can be solved numerically using the so-called "multiplicative updates for solving quadratic programming with rectangle-range l_1 regularization" algorithm. We refer the reader to Ding et al. (2021, Algorithm 1) for more detail on this algorithm.

Also observe that the AREN problem can be transformed to a rectangle-range lasso problem. If we take

$$\tilde{X} = \frac{1}{\sqrt{1+\lambda_n^{(2)}}}\begin{pmatrix} X \\ \sqrt{\lambda_n^{(2)}}1_{p\times p}\end{pmatrix}_{(n+p)\times p}, \quad \tilde{Y} = \begin{pmatrix} Y \\ 0\end{pmatrix}_{(n+p)\times 1}, \quad \lambda_n = \frac{\lambda_n^{(1)}}{\sqrt{1+\lambda_n^{(2)}}},$$
$$\tilde{\beta}^* = \sqrt{1+\lambda_n^{(2)}}\beta^*, \quad \tilde{\mathcal{J}} = \prod_{i=1}^{p}\left[\sqrt{1+\lambda_n^{(2)}}s_i, \sqrt{1+\lambda_n^{(2)}}t_i\right],$$

then the problem, Equation (2.2), is equivalent to

$$\hat{\tilde{\beta}}(\lambda_n) = \underset{\beta\in\tilde{\mathcal{J}}}{\arg\min}\left(\frac{1}{2n}\left\|\tilde{Y} - \tilde{X}\beta\right\|_2^2 + \lambda_n\|\beta\|_1\right), \quad (2.4)$$

where $1_{p\times p}$ denotes the $p \times p$ identity matrix and $\hat{\tilde{\beta}}(\lambda_n)$ is the estimator of $\tilde{\beta}^*$. Both estimation and model consistencies of the lasso have been studied in the literature, and this work is easily adapted to apply to the estimation and model consistencies of the rectangle-range lasso, as well as the AREN. We therefore simply state these results for the AREN without proof in the next two sections.

2.2 Upper Bounds of Tail Probability and Estimation Consistency

We say the AREN has estimation consistency if the AREN estimator $\hat{\beta}$ satisfies

$$\left\|\hat{\beta}-\beta^*\right\|_1 \xrightarrow[n\to\infty]{\mathbb{P}} 0 \quad \text{or} \quad \left\|\hat{\beta}-\beta^*\right\|_2 \xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

where $\xrightarrow[n\to\infty]{\mathbb{P}}$ denotes the convergence in probability. As pointed out earlier, our AREN model is in fact equivalent to the rectangle-range lasso studied further in Ding et al. (2021, Corollary 2.5). The main difference between AREN and the model in Ding et al. (2021, Corollary 2.5) lies in whether there is the multiplier $1/(2n)$ in the loss function (see Equation (2.2)). This difference makes the conditions (ii) and (iii) below slightly different from those in Ding et al. (2021, Corollary 2.5). Nevertheless, those results prove the estimation consistency of the AREN, subject to the following conditions, adapted and modified from Ding et al. (2021):

(i) $\boldsymbol{\beta}^* \in \boldsymbol{\mathcal{I}}$.

(ii) The designed matrix $\mathbf{X}$ satisfies

$$\frac{X_j' X_j + \lambda_n^{(2)}}{\left(1+\lambda_n^{(2)}\right)n} \le 1, \text{ for all } j = 1, \dots, p.$$

(iii) There exists a constant $\boldsymbol{\kappa} > \mathbf{0}$, such that

$$\frac{\|X\beta\|_2^2 + \lambda_n^{(2)}\|\beta\|_2^2}{\left(1+\lambda_n^{(2)}\right)n} \ge \kappa\|\beta\|_2^2$$

for all $\beta \ge 0$ satisfying

$$\sum_{j\in\{1,\dots,p\}:\, \beta_j^*=0} |\beta_j| \le 3 \sum_{j\in\{1,\dots,p\}:\, \beta_j^*\neq 0} |\beta_j|.$$

(iv) $\boldsymbol{\lambda_n^{(1)}}$ and $\boldsymbol{\lambda_n^{(2)}}$ satisfy

$$\frac{\lambda_n^{(1)}}{1+\lambda_n^{(2)}} \xrightarrow[n\to\infty]{} 0 \quad \text{and} \quad \exp\left(-\frac{n}{8\sigma^2}\frac{(\lambda_n^{(1)})^2}{1+\lambda_n^{(2)}}\right) \xrightarrow[n\to\infty]{} 0,$$

where $\sigma > 0$ is the residual standard deviation of each component of the error term in the linear model, Equation (2.1).

The estimation consistency of the AREN is provided by this theorem:

Theorem 2.1: Consider a q-sparse instance of the AREN, Equation (2.2). Let X satisfy the conditions (i) - (iii) and let the regularization parameters satisfy $\lambda_n^{(1)} > 0, \lambda_n^{(2)} \ge 0$. Then the AREN solution $\hat{\beta} = \hat{\beta}(\lambda_n^{(1)}, \lambda_n^{(2)})$ satisfies:

$$\mathbb{P}\left(\left\|\hat{\beta}-\beta^*\right\|_2^2 > \frac{9q\left(\lambda_n^{(1)}\right)^2}{\kappa^2\left(1+\lambda_n^{(2)}\right)^2}\right) \le 2p\exp\left(-\frac{n\left(\lambda_n^{(1)}\right)^2}{8\sigma^2\left(1+\lambda_n^{(2)}\right)}\right),$$

$$\mathbb{P}\left(\left\|\hat{\beta}-\beta^*\right\|_1 > \frac{12q\lambda_n^{(1)}}{\kappa\left(1+\lambda_n^{(2)}\right)}\right) \le 2p\exp\left(-\frac{n\left(\lambda_n^{(1)}\right)^2}{8\sigma^2\left(1+\lambda_n^{(2)}\right)}\right).$$

In addition, if (iv) holds, the AREN, Equation (2.2), has the property of estimation consistency: $\left\|\hat{\beta}-\beta^*\right\|_2 \xrightarrow[n\to\infty]{\mathbb{P}} 0$.

Theorem 2.1 provides fine upper bounds of the tail probabilities of the estimation errors in l_1 and l_2-norms. It then makes clear that the estimation consistency holds whenever λ_n goes to zero slower than $n^{-1/2}$. If we take $\mathcal{I} = [0, +\infty)^p$, $\lambda_n^{(2)} = 0$ and $\lambda_n^{(1)} = 4\sigma\sqrt{\log(p)/n}$ in Theorem 2.1, the tail probability bounds for the non-negative lasso follow as in Wu et al. (2014, Proposition 1). If we further assume $\mathcal{I} = \mathbb{R}^p$ in Theorem 2.1, the tail bounds for the unconstrained lasso follow (Negahban et al., 2012, Corollary 2). In the above two cases, if we assume $p = p_n$, $q = q_n$ with $p_n \longrightarrow +\infty$ and $p_n \log(q_n)/n \longrightarrow 0$, as $n \to \infty$, the estimation consistency holds. However, if $\lambda_n = \lambda_0 n^{-1/2}$ for some $\lambda_0 > 0$, whether the estimation consistency holds is an open problem. In this situation what can be derived is the sign pattern consistency: there is positive probability that all signs of $\hat{\beta}$ are consistent with those of β^*, i.e., Propositions 1 and 2 in Bach (2008) can be obtained for AREN. Such sign pattern consistency can be viewed as a weak form of variable selection model consistency. When λ_n goes to infinity, it is possible to establish the strong version of variable selection consistency for AREN. In the next section we present the variable selection consistency of AREN subject to the assumption that $\lambda_n^{(1)}$ goes to infinity faster than $\sqrt{n}$ and some other conditions.

2.3 Variable Selection Consistency

Recall that our AREN problem is equivalent to the problem, Equation (2.3), whose variable selection consistency can be derived as a special case of the generalized version in Ding et al. (2021, Theorem 2.3). In the interests of completeness, we state the variable selection consistency conditions for our AREN, which have been modified and adapted from those in Ding et al. (2021, Theorem 2.3). Denote by

$$G = \{i \in \{1, \dots, p\}: \beta_i^* = 0\} \text{ and } \widehat{G} = \{i \in \{1, \dots, p\}: \hat{\beta}_i = 0\},$$

and let $\#G$ be the cardinality of the group of indexes G. The variable selection consistency for the AREN is defined as follows.

Definition 2.2: We say that the AREN, Equation (2.2), satisfies variable selection consistency if there exist $\lambda_n^{(1)}$ and $\lambda_n^{(2)}$ such that $\mathbb{P}(\widehat{G} = G) \underset{n\to\infty}{\longrightarrow} 1$.

The above variable selection consistency is a stronger property than the sign pattern consistency discussed in Bach (2008). It says that, if $\beta_i^* = 0$, then with probability approaching 1, the i-th predictor will not be selected, as n becomes large. Note that the variable selection consistency of the non-negative elastic net and elastic net (Wu and Yang, 2014; Wu et al, 2014; Zhao and Yu, 2006) are implied by variable consistency of the AREN.

Let $X_{(1)} = (X_i)_{i\notin G}$ be the observed predictor values corresponding to the group of indexes G^c, the complementary of G. Let $\beta_{(1)}^* = (\beta_i^*)_{i\notin G}$, $s_{(1)} = (s_i)_{i\notin G}$ and $t_{(1)} = (t_i)_{i\notin G}$. Similarly let $X_{(2)} = (X_i)_{i\in G}$, $\beta_{(2)}^* = (\beta_i^*)_{i\in G}$, $s_{(2)} = (s_i)_{i\in G}$ and $t_{(2)} = (t_i)_{i\in G}$. Moreover, denote by

$$
\begin{aligned}
&C_{ij} = \frac{X'_{(i)}X_{(j)}}{2n^2}, \text{ for } i, j = 1,2;\\
&\rho_n^{(1)} = \max\left\{\left(C_{11} + \frac{\lambda_n^{(2)}}{n}1_{(p-\#G)\times(p-\#G)}\right)^{-1} C_{11}\beta^*_{(1)} - t_{(1)}\right\};\\
&\rho_n^{(2)} = \min\left\{\left(C_{11} + \frac{\lambda_n^{(2)}}{n}1_{(p-\#G)\times(p-\#G)}\right)^{-1} C_{11}\beta^*_{(1)} - s_{(1)}\right\};\\
&C_n = \left(C_{11} + \frac{\lambda_n^{(2)}}{n}1_{(p-\#G)\times(p-\#G)}\right)^{-1}\left(\frac{\lambda_n^{(1)}}{2}\text{sign}\left(\beta^*_{(1)}\right)\right);\\
&C_n^{\max} = \max C_n, \quad C_n^{\min} = \min C_n,
\end{aligned}
\tag{2.5}
$$

where for a vector $v = (v_1, \dots, v_n)$, $\text{sign}(v) = \left(\text{sign}(v_1) \dots \text{sign}(v_n)\right)'$ denotes the vector of signs of the elements in v. The sign equals 1 for positive entry, -1 for negative entry and 0 for zero entry. To show the AREN, Equation (2.2), admits the variable selection consistency, we assume that the following conditions hold:

$$
q > 1,\ p - q > 1,\ \frac{\lambda_n^{(1)}}{\sqrt{n}} \underset{n\to\infty}{\longrightarrow} +\infty,\ \frac{\max_{1\le i\le p} X_i'X_i}{n^2} \underset{n\to\infty}{\longrightarrow} 0, \tag{2.6}
$$

and

$$
\frac{1}{\rho_n^{(1)}}\left(\frac{8\sigma\sqrt{\#G_{(1)}\text{trace}(C_{11})\log\left(\#G_{(1)}\right)}}{n\Lambda_{\min}\left(C_{11} + \lambda_n^{(2)}1_{(p-\#G)\times(p-\#G)}/n\right)} + \frac{\left|C_n^{\min}\right|}{n}\right) \underset{n\to\infty}{\longrightarrow} 0, \tag{2.7}
$$

$$
\frac{1}{\rho_n^{(2)}}\left(\frac{8\sigma\sqrt{\#G_{(1)}\text{trace}(C_{11})\log\left(\#G_{(1)}\right)}}{n\Lambda_{\min}\left(C_{11} + \lambda_n^{(2)}1_{(p-\#G)\times(p-\#G)}/n\right)} + \frac{\left|C_n^{\max}\right|}{n}\right) \underset{n\to\infty}{\longrightarrow} 0, \tag{2.8}
$$

where $\text{trace}(C_{11})$ denotes the trace of the matrix C_{11} and $\Lambda_{\min}(M)$ denotes the minimal eigenvalue of the matrix M. In addition, we assume that the arbitrary rectangle-range elastic irrepresentable condition (AREIC), defined below, is satisfied.

Definition 2.3: If there exists a positive constant vector η, such that

$$
C_{21}\left(C_{11} + \frac{\lambda_n^{(2)}}{n}1_{(p-\#G)\times(p-\#G)}\right)^{-1}\left(\text{sign}\left(\beta^*_{(1)}\right) + \frac{2\lambda_n^{(2)}}{\lambda_n^{(1)}}\beta^*_{(1)}\right) - \frac{2\lambda_n^{(2)}}{\lambda_n^{(1)}}s_{(2)} \le \mathbf{1} - \eta,
$$

where $\mathbf{1} = (1 \dots 1)'$, we say that AREIC holds.

When $\mathcal{J} = [0, +\infty)^p$, the AREIC becomes the non-negative elastic irrepresentable condition (NEIC) as follows:

$$C_{21}\left(C_{11}+\frac{\lambda_n^{(2)}}{n}1_{(p-\#G)\times(p-\#G)}\right)^{-1}\left(\mathbf{1}+\frac{2\lambda_n^{(2)}}{\lambda_n^{(1)}}\beta^*_{(1)}\right)\leq \mathbf{1}-\eta. \qquad (2.9)$$

The NEIC was crucial to get the variable selection consistency of the non-negative elastic net (Zhao et al., 2014). If further $\lambda_n^{(2)} = 0$ in Equation (2.9), the NEIC then becomes the non-negative irrepresentable condition (NIC): $C_{21}C_{11}^{-1}\mathbf{1} \leq \mathbf{1} - \eta$, which was needed to obtain the variable selection consistency of the non-negative lasso in Wu et al. (2014). Note that, NIC is a non-negative version of the following irrepresentable condition (IC): $\left|C_{21}C_{11}^{-1}\text{sign}\left(\beta^*_{(1)}\right)\right| \leq \mathbf{1} - \eta$, for the variable selection consistency of the lasso (Zhao and Yu, 2006). It was proved in Zhao and Yu (2006) that IC is a sufficient and necessary condition for the variable selection consistency of the lasso, while NIC is only a sufficient condition. However, since NIC is less restrictive than IC (it does not depend on the unknown parameters β^*), so easier verified than IC in practice. Nevertheless, AREIC is a natural extension of the previous conditions NEIC and NIC for the variable selection consistency. We state below the variable selection consistency theorem for the AREN.

Theorem 2.4: Assume that the conditions, Equations (2.6) - (2.8), and the AREIC hold. Then the AREN, Equation (2.2), possesses the variable selection consistency property given in Definition 2.2.

Estimation consistency and variable selection consistency are important statistical properties because they guarantee that as the sample size n increases (which is tantamount to a proportional increase in information), so does the accuracy of estimation and variable selection. Although the somewhat esoteric sufficient conditions outlined in this section are difficult to verify in practice, it is almost certain that less restrictive, more general conditions for these types of consistency hold and apply broadly. Still, this is no guaranty of accuracy in any specific moderate or large sample. Accordingly, it is possible that AREN could under-perform in practice if conditions are sufficiently extreme. To help mitigate this, we introduce BoAREN, or "Bootstrapped AREN", following Bach (2008). As long as n is not critically small, the estimation consistency and variable selection consistency would guarantee a high likelihood that most of the variables are correctly chosen and estimated accurately. By performing bootstrap replications of the data, each bootstrap replicate should also produce a result with most of the variables correctly chosen and estimated accurately. By definition of bootstrap replication, each bootstrap replicate of fitting the AREN is likely to be slightly different (i.e. contain a slightly different set of variables selected). By estimation consistency, each of these sets should be "close" to the correct set, though some may be lacking some important variables, while others may contain slightly too many. Intersecting (or alternatively, "soft" intersecting - see the next section) these sets can therefore improve the accuracy of variable selection. The two models, AREN and BoAREN, are applied to the index tracking problem in the next section.

3. BoAREN and S&P 500 Index Tracking[1]

An index fund is a passively managed mutual fund that is designed to track a given component of the market, for example, the S&P 500. Index tracking is a generic term for the various methods/algorithms used by portfolio managers to guarantee that the index fund remains in close agreement with the target market component. Here, the main objective for the index tracking problem is minimizing the tracking error, which we define as the standard deviation of the difference

[1] Code for the AREN/BoAREN computations in this section can be found in https://github.com/yujiading/bootstrapped-aren.

between the returns of the selected portfolio (R^P) and the benchmark (R^B). Assuming the total number of periods is n, the tracking error (TE) per perod is computed by

$$TE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\left(R_i^P - R_i^B\right) - \frac{1}{n}\sum_{l=1}^{n}\left(R_l^P - R_l^B\right)\right)^2}. \qquad (3.1)$$

Algorithm 1: BoAREN[2]

Input: $X \in \mathbb{R}^{n\times p}$
$Y \in \mathbb{R}^n$
Number of bootstrap replicates m
Soft index S
l_1 regularization parameter $\lambda_n^{(1)}$
l_2 regularization parameter $\lambda_n^{(2)}$
AREN coefficient lower constraints $\mathbf{s} \in (\mathbb{R} \cup \{-\infty\})^p$
AREN coefficient upper constraints $\mathbf{t} \in (\mathbb{R} \cup \{+\infty\})^p$

1 For $i \leftarrow 1$ to m do
2 Generate bootstrapped $X^{(i)} \in \mathbb{R}^{n\times p}$ and $Y^{(i)} \in \mathbb{R}^n$
3 Compute AREN estimate $\hat{\beta}^{(i)}$ from $X^{(i)}$ and $Y^{(i)}$ with $\lambda_n^{(1)}, \lambda_n^{(2)}, \mathcal{I} = [0,\infty)^p$
4 Compute support $J^{(i)} = \{j,\ \hat{\beta}_j^{(i)} \neq 0\}$
5 Compute $J = (\bigcap_S)_{i=1}^m J^{(i)}$
6 Compute AREN estimate $\hat{\beta}_J$ from X_J and Y with $\lambda_n^{(1)} = \lambda_n^{(2)} = 0, \mathcal{I} = [\mathbf{s}, \mathbf{t}]$

Algorithm 2: Two-step AREN

Input: $X \in \mathbb{R}^{n\times p}$
$Y \in \mathbb{R}^n$
l_1 regularization parameter $\lambda_n^{(1)}$
l_2 regularization parameter $\lambda_n^{(2)}$
AREN coefficient lower constraints $\mathbf{s} \in (\mathbb{R} \cup \{-\infty\})^p$
AREN coefficient upper constraints $\mathbf{t} \in (\mathbb{R} \cup \{+\infty\})^p$

1 Compute AREN estimate $\hat{\beta}$ from X and Y with $\lambda_n^{(1)}, \lambda_n^{(2)}, \mathcal{I} = [0,\infty)^p$
2 Compute support $J = \{j,\ \hat{\beta}_j \neq 0\}$
3 Compute AREN estimate $\hat{\beta}_J$ from X_J and Y with $\lambda_n^{(1)} = \lambda_n^{(2)} = 0, \mathcal{I} = [\mathbf{s}, \mathbf{t}]$

We apply the two-step AREN (Algorithm 2) and BoAREN (Algorithm 1) to creating an index fund to track the performance of S&P 500 index. The main idea is to follow a two-step method that applies a (bootstrapped) non-negative elastic net to select a subset of stocks, then apply constrained least squares on the selected stocks to estimate the unknown coefficients. In this section the two-step BoAREN is simply called BoAREN. The index tracking procedure is summarized in Algorithm 3. Wu and Yang (2014) find that the index tracking results can be greatly improved through this two-step method. For bootstrap approach, it has been observed in Bach (2008) that intersecting the supports for each

[2] "BoAREN" refers to the AREN algorithm enhanced by bootstrapping.

bootstrap replication might be too strict and a so-called "soft" version improves the performance. We include a soft index S (e.g. 90%, 80%, 70%) in the BoAREN algorithm so that $\cap_S$ (Algorithm 1, Line 5) selects the supports which are present in at least the percentage S of the bootstrap replications. The non-negative setting, i.e. $\mathcal{I} = [0, \infty)^p$ in Line 3 of Algorithm 1 and Line 1 of Algorithm 2, ensures that we focus on "long-only" strategies.

Algorithm 3: Index Tracking using Two-step AREN or BoAREN

Input: $X \in \mathbb{R}^{n \times p}$
$Y \in \mathbb{R}^n$
Number of bootstrap replicates m, if using BoAREN
Soft index S, if using BoAREN
$\lambda_n^{(1)}$ tuning grid $\Lambda^{(1)}$
$\lambda_n^{(2)}$ tuning grid $\Lambda^{(2)}$
AREN coefficient lower constraints $\mathbf{s} \in (\mathbb{R} \cup \{-\infty\})^p$
AREN coefficient upper constraints $\mathbf{t} \in (\mathbb{R} \cup \{+\infty\})^p$

1 $X_{train}, Y_{train}, X_{val}, Y_{val}, X_{test}, Y_{test} \leftarrow X, Y$
2 For each $\lambda_n^{(1)} \in \Lambda^{(1)}$ and $\lambda_n^{(2)} \in \Lambda^{(2)}$ do
3 If using BoAREN
4 Compute $\hat{\beta}_J$ from X_{train}, Y_{train}, m, S, $\lambda_n^{(1)}$, $\lambda_n^{(2)}$, $\mathbf{s}$, $\mathbf{t}$, using Algorithm 1
5 Else if using two-step AREN
6 Compute $\hat{\beta}_J$ from X_{train}, Y_{train}, $\lambda_n^{(1)}$, $\lambda_n^{(2)}$, $\mathbf{s}$, $\mathbf{t}$, using Algorithm 2
7 $R^B \leftarrow Y_{val}$
8 $R^P \leftarrow X_{val}\hat{\beta}_J$
9 Compute tracking error from R^B, R^P, using Equation (3.1)
10 Find the corresponding $\hat{\beta}_J$ with the smallest tracking error
11 $R^B \leftarrow Y_{test}$
12 $R^P \leftarrow X_{test}\hat{\beta}_J$
13 Compute tracking error from R^B, R^P, using Equation (3.1)

We use one year (from 2020-9-1 to 2021-9-1) of daily adjusted closing prices (253 observations) of the S&P 500 and 302[3] S&P 500 component stocks. In Algorithm 3, the input Y represents the daily percentage return of S&P 500, each column of the input X represents the daily percentage return of one of the 302 stocks. The total number of columns of X is $p = 302$. We use the first 70% of the 252 data points for training, the next 20% for validation, and the last 10% for testing. We use the mean value of the training set to center the whole data set so that the regression can be fit without intercept (Hastie et al., 2009, p. 64). To simplify the tuning process, we use the strategy in Friedman et al. (2010, Section 2.5) to rewrite the regularization as $\lambda(\alpha\|\beta\|_1 + 0.5(1-\alpha)\|\beta\|_2^2)$, where all coefficients will shrink to zero if $\lambda > \lambda_{max} = 2\max_l\{X_l Y\}/\alpha$. We use a grid of 10 equally spaced points on $[0,1]$ for α. We set $\lambda_{min} = 0.001\lambda_{max}$ and use a grid of 100 equally spaced points on $[\lambda_{min}, \lambda_{max}]$ for λ.

The possibility of adding constraints to coefficients in AREN allows us to select the values of $\mathbf{s}$ and $\mathbf{t}$ in Algorithm 3 to avoid concentrated stock positions, that is to avoid over investing in any single stock which can expose the investor to significant risk based on the fortunes of a few companies. To elaborate, suppose we invest money in $\#J$ (cardinality of J) stocks with returns $R_j^l = \left(P_j^l - P_j^{l-1}\right)/P_j^{l-1}, j = 1, \ldots, \#J$ to track S&P 500 with return $\widehat{R}_{SP}^l = \left(\widehat{P}_{SP}^l - P_{SP}^{l-1}\right)/P_{SP}^{l-1}$ using l as indexes of date. Let $\overline{R}_{SP}^{Train}, \overline{R}_j^{Train}, j = 1, \ldots, \#J$ represent the mean returns on the training set for the S&P 500 and the selected stocks. The regression gives:

[3] We only consider the daily prices from 302 stocks that have not been changed during the period of interest.

$$\hat{R}_{SP}^{l} - \overline{R}_{SP}^{Train} = \sum_{j=1}^{\#J} (\hat{\beta}_J)_j \, (R_j^l - \overline{R}_j^{Train});$$

$$\hat{P}_{SP}^{l} = \sum_{j=1}^{\#J} \frac{(\hat{\beta}_J)_j P_{SP}^{l-1} P_j^l}{P_j^{l-1}} + \left(1 + \overline{R}_{SP}^{Train} - \sum_{j=1}^{\#J} (1 + \overline{R}_l^{Train})(\hat{\beta}_J)_j \right) P_{SP}^{l-1},$$

which means to track P_{SP}^{l} dollar amount of S&P 500, we invest $(\hat{\beta}_J)_j P_{SP}^{l-1} P_j^l / P_j^{l-1}$ dollar amount on stock j for $j = 1, \dots, \#J$ and hold or borrow

$$\left(1 + \overline{R}_{SP}^{Train} - \sum_{j=1}^{\#J} (1 + \overline{R}_l^{Train})(\hat{\beta}_J)_j \right) P_{SP}^{l-1}$$

dollar amount. So, the percentage of money spent on each stock is

$$\frac{(\hat{\beta}_J)_i}{(\hat{\beta}_J)_i + \frac{P_i^{l-1}}{P_i^l} \sum_{j=1, j\neq i}^{\#J} \frac{(\hat{\beta}_J)_j P_j^l}{P_j^{l-1}}} = \frac{(\hat{\beta}_J)_i}{(\hat{\beta}_J)_i + \sum_{j=1, j\neq i}^{\#J} \frac{(\hat{\beta}_J)_j (1+R_j^l)}{1+R_i^l}}$$

for $i = 1, \dots, \#J$. To avoid concentrated stock positions, we want each percentage less than an amount M (e.g. 10%, 20%, 30%), i.e., for $i = 1, \dots, \#J$,

$$\frac{(\hat{\beta}_J)_i}{(\hat{\beta}_J)_i + \sum_{j=1, j\neq i}^{\#J} \frac{(\hat{\beta}_J)_j (1+R_j^l)}{1+R_i^l}} \leq \frac{t_i}{t_i + \frac{1+R^{min}}{1+R^{max}} \sum_{j=1, j\neq i}^{\#J} s_j} \leq M \leq 1,$$

where R^{min} and R^{max} are the smallest and largest prices for all stocks, respectively. Assume $s_i = s_0, t_i = t_0$ for all i, we guarantee that the percentage of money spent on a single stock is less that M through selecting s_0, t_0 such that

$$s_0 \leq t_0 \leq \frac{M}{1-M} \frac{1+R^{min}}{1+R^{max}} (\#J - 1) s_0.$$

A variety of approaches could be used to tune s_0, t_0 to improve performance, but here we use the following simple steps to select s_0, t_0. First, we find the maximum and minimum coefficients, $(\hat{\beta}_J)_{max}, (\hat{\beta}_J)_{min}$, when $M = 100\%$ (i.e., $[\boldsymbol{s}, \boldsymbol{t}] = [0, \infty)^p$). Then given M, we set $s_0 = (\hat{\beta}_J)_{min}$ and calculate the biggest t_0, and set $t_0 = (\hat{\beta}_J)_{max}$ and calculate the smallest s_0. The final s_0, t_0 are taken to be the case that has the largest distance between them. Note that the scale needs to be

$$\frac{M}{1-M} \frac{1+R^{min}}{1+R^{max}} (\#J - 1) \geq 1 \Leftrightarrow M \geq \frac{1}{1 + \frac{1+R^{min}}{1+R^{max}} (\#J - 1)}$$

to make this process work. Hence for this method there will be a bound below which M cannot be set, depending on the data.

Table 1: Tracking errors (TE in units of 10^{-3}) root mean-squared errors (RMSE in units of 10^{-3}), and number of selected stocks for two-step AREN and BoAREN with varying soft index S and number of bootstrap replicates m using Algorithm 3. M is the largest percentage of money spent on a single stock

M	Measure	Two-step AREN	BoAREN S=1 m=32	BoAREN S=1 m=64	BoAREN S=1 m=128	BoAREN S=1 m=256	BoAREN S=0.95 m=32	BoAREN S=0.95 m=64	BoAREN S=0.95 m=128	BoAREN S=0.95 m=256
	TE	1.11	1.25	1.41	1.50	1.40	1.13	**1.01***	1.13	1.24
1	RMSE	1.14	1.30	1.47	1.57	1.46	1.16	1.04	1.18	1.30
	Stocks	259	182	161	145	155	238	243	234	204
	TE	1.02	1.27	1.23	1.43	1.43	1.19	**1.01**	1.09	1.11
0.3	RMSE	1.11	1.36	1.31	1.52	1.52	1.27	1.08	1.18	1.19
	Stocks	259	184	192	154	155	216	243	234	236
	TE	1.13	1.24	1.29	1.40	1.98	1.25	1.13	1.19	1.19
0.2	RMSE	1.24	1.34	1.38	1.47	2.06	1.35	1.23	1.28	1.28
	Stocks	259	204	192	171	80	216	243	234	236
	TE	3.66	3.92	4.00	4.29	**3.59**	3.80	**3.53**	**3.24**	**3.34**
0.1	RMSE	3.82	3.98	4.01	4.30	3.72	3.90	3.66	3.34	3.47
	Stocks	190	67	31	26	33	95	152	121	121

M	Measure	Two-step AREN	BoAREN S=0.9 m=32	BoAREN S=0.9 m=64	BoAREN S=0.9 m=128	BoAREN S=0.9 m=256	BoAREN S=0.85 m=32	BoAREN S=0.85 m=64	BoAREN S=0.85 m=128	BoAREN S=0.85 m=256
	TE	1.11	1.15	1.15	**1.03**	**1.05**	1.20	1.11	1.23	1.13
1	RMSE	1.14	1.20	1.19	1.08	1.10	1.25	1.15	1.28	1.16
	Stocks	259	247	259	258	259	262	254	248	252
	TE	1.02	1.09	1.07	1.02	**1.01**	1.13	1.05	1.12	1.06
0.3	RMSE	1.11	1.17	1.15	1.10	1.09	1.22	1.12	1.20	1.13
	Stocks	259	247	259	258	259	251	254	248	248
	TE	1.13	1.24	1.24	1.24	1.24	1.20	1.17	1.22	1.19
0.2	RMSE	1.24	1.35	1.34	1.34	1.34	1.32	1.27	1.33	1.29
	Stocks	259	242	242	242	243	254	254	253	248
	TE	3.66	**2.97***	**3.23**	**3.11**	**3.31**	**3.41**	**3.34**	**3.13**	**3.57**
0.1	RMSE	3.82	3.10	3.34	3.21	3.44	3.53	3.47	3.24	3.68
	Stocks	190	133	153	150	147	144	121	118	154

M	Measure	Two-step AREN	BoAREN S=0.8 m=32	BoAREN S=0.8 m=64	BoAREN S=0.8 m=128	BoAREN S=0.8 m=256	BoAREN S=0.75 m=32	BoAREN S=0.75 m=64	BoAREN S=0.75 m=128	BoAREN S=0.75 m=256
	TE	1.11	1.35	1.14	1.13	1.12	1.34	1.38	1.39	1.42
1	RMSE	1.14	1.41	1.17	1.16	1.15	1.40	1.45	1.46	1.48
	Stocks	259	179	258	261	259	198	187	200	201
	TE	1.02	**0.95**	1.06	1.03	1.06	**0.94***	**0.95**	**0.95**	**0.95**
0.3	RMSE	1.11	1.08	1.14	1.11	1.13	1.07	1.07	1.07	1.08
	Stocks	259	295	258	261	259	299	297	295	295
	TE	1.13	**1.12**	1.18	1.14	1.17	**1.09**	**1.07***	1.13	1.11
0.2	RMSE	1.24	1.22	1.29	1.24	1.28	1.19	1.17	1.22	1.20
	Stocks	259	262	258	261	259	245	266	233	233
0.1	TE	3.66	**3.42**	3.94	**3.33**	**3.44**	**3.44**	**3.53**	**3.47**	**3.59**
	RMSE	3.82	3.54	4.06	3.45	3.56	3.57	3.65	3.59	3.70
	Stocks	190	114	116	145	147	166	154	154	156

Table 1 shows tracking errors (TE), root mean-squared errors (RMSE), and the number of selected stocks over two-step AREN and BoAREN using different bootstrap replicates m, different limits on the amount spent on each stock M, different BoAREN soft indexes S, and not limit on the number of stocks that can be selected. For all $M = 1,0.3,0.2,0.1$, two-step AREN performs well, and BoAREN performs even better. It appears that a soft index of $S = 1$ seems too strict, selecting only a few stocks, and as bootstrap replications increase, the tracking error does not converge and becomes increasingly large. In the case of no limit on the amount spent on a single stock (i.e., $M = 1$), BoAREN with a soft index of $S = 1$ also seems too strict, $S = 0.8,0.75$ seems too soft, and S in between works better with the lowest tracking error 1.01 obtained when $m = 64, S = 0.95$. By limiting the amount spent on each stock to $M = 0.3$ and $M = 0.2$, a softer index of $S = 0.75$ performs better, giving tracking errors of 0.94 and 1.07, respectively. In the case of no more than 10% of the total amount spent on each stock, BoAREN with $S = 1$ shows decreasing tracking errors as bootstrap replications increase and outperforms two-step AREN when $m = 256$. Moreover, most cases of BoAREN with $S = 0.95,0.9,0.85,0.8,0.75$ show improvement over the two-step AREN case in terms of tracking errors.

Table 1: Tracking errors (TE in units of 10^{-3}), root mean-squared errors (RMSE, in units of 10^{-3}), and number of selected stocks for two-step AREN and best BoAREN model using Algorithm 3. M is the largest percentage of money spent on a single stock. m is the number of bootstrap replicates. S is the soft index.

Stocks	M	Two-step AREN			Best BoAREN				
		TE	RMSE	Stocks	*TE*	RMSE	Stocks	m	*S*
No limit	1.0	1.11	1.14	259	1.01	1.04	243	64	0.95
	0.3	1.02	1.11	259	0.94	1.07	299	32	0.75
	0.2	1.13	1.24	259	1.07	1.17	266	64	0.75
	0.1	3.66	3.82	190	2.97	3.10	133	32	0.90
≤200	1.0	1.44	1.52	196	1.25	1.30	186	128	0.90
	0.3	1.33	1.42	196	1.20	1.29	192	32	0.95
	0.2	1.33	1.43	196	1.21	1.32	192	32	0.95
	0.1	3.66	3.82	190	2.97	3.10	133	32	0.9
≤150	1.0	1.47	1.55	146	1.37	1.43	147	64	1.00
	0.3	1.46	1.57	146	1.36	1.45	147	64	1.00
	0.2	1.49	1.58	149	1.39	1.47	144	32	0.85
	0.1	4.46	4.58	139	2.97	3.10	133	32	0.90
≤100	1.0	2.14	2.20	84	1.53	1.60	97	64	0.80
	0.3	2.24	2.32	84	1.61	1.70	97	64	0.80
	0.2	2.43	2.49	88	1.71	1.83	95	32	0.95
	0.1	6.40	6.52	88	3.59	3.72	33	256	1.00

Note that in the majority of cases in Table 1, a large number of stocks were selected. Due to transaction fees and management effort for retail or individual fund managers, it is of interest to examine cases with the limit on stocks to be no greater than $50,100,150,200$ during the tuning process (Lines 2-10) of Algorithm 3. We examine models with $S = 1,0.95,0.9,0.85,0.8,0.75$, $m = 32,64,128,256$ and summarize the best model for each M and limit on stocks in Table 2. As compared with two-step AREN which generally performs well, BoAREN performs better in terms of tracking errors, mean-squared errors, and picking about the same number or even fewer stocks. To see the sensitivity of the BoAREN parameters, in Figure 1, we plot the predicted S&P 500 index using the best BoAREN models for each stocks number constraint and each M (see Table 2) and compare them with the actual S&P 500 index values. We use predicted returns for next time step R_{pred}^{next} and the actual price from last time step P_{real}^{last} to calculate each fitted or predicted S&P 500 index $(R_{pred}^{next} + 1)P_{real}^{last}$. To make the difference between the actual and predicted values more visible, in Figure 2, we plot the ratio of actual to predicted S&P 500 index using the best BoAREN models in Table 2. From Figure 1 and Figure 2 we see that BoAREN

tends to show better performance as the limits on stocks and M get larger. However, this difference is not significant among the cases $M = 1, 0.3, 0.2$ and among the various conditions imposed on stock count, i.e. no limit, ≤ 200, ≤ 150. Since there is a trade-off between the model accuracy and the expense in applying the model, these results imply that portfolio managers may consider using a relatively small number of stocks and a suitable constraint on the amount spent on each stock to track the S&P 500 index while retaining high tracking accuracy.

Figure 1: **Predicted S&P 500 index (from 2021-2-24 to 2021-9-1) using the best BoAREN models in Table 2. Green dot lines correspond to actual values of S&P 500 index; red solid lines correspond to predicted values by BoAREN; blue dash lines correspond to predicted values by two-step AREN.**

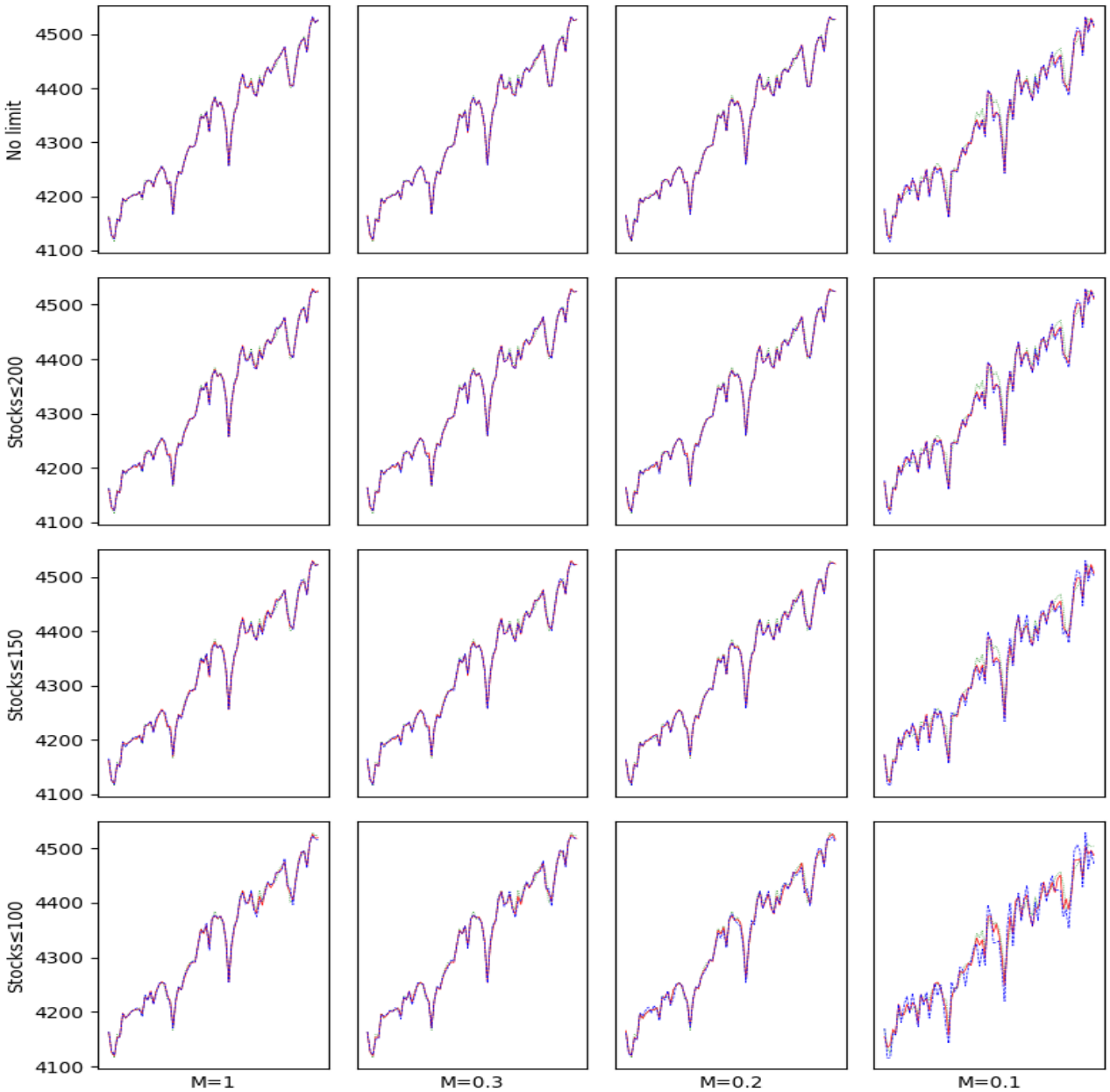
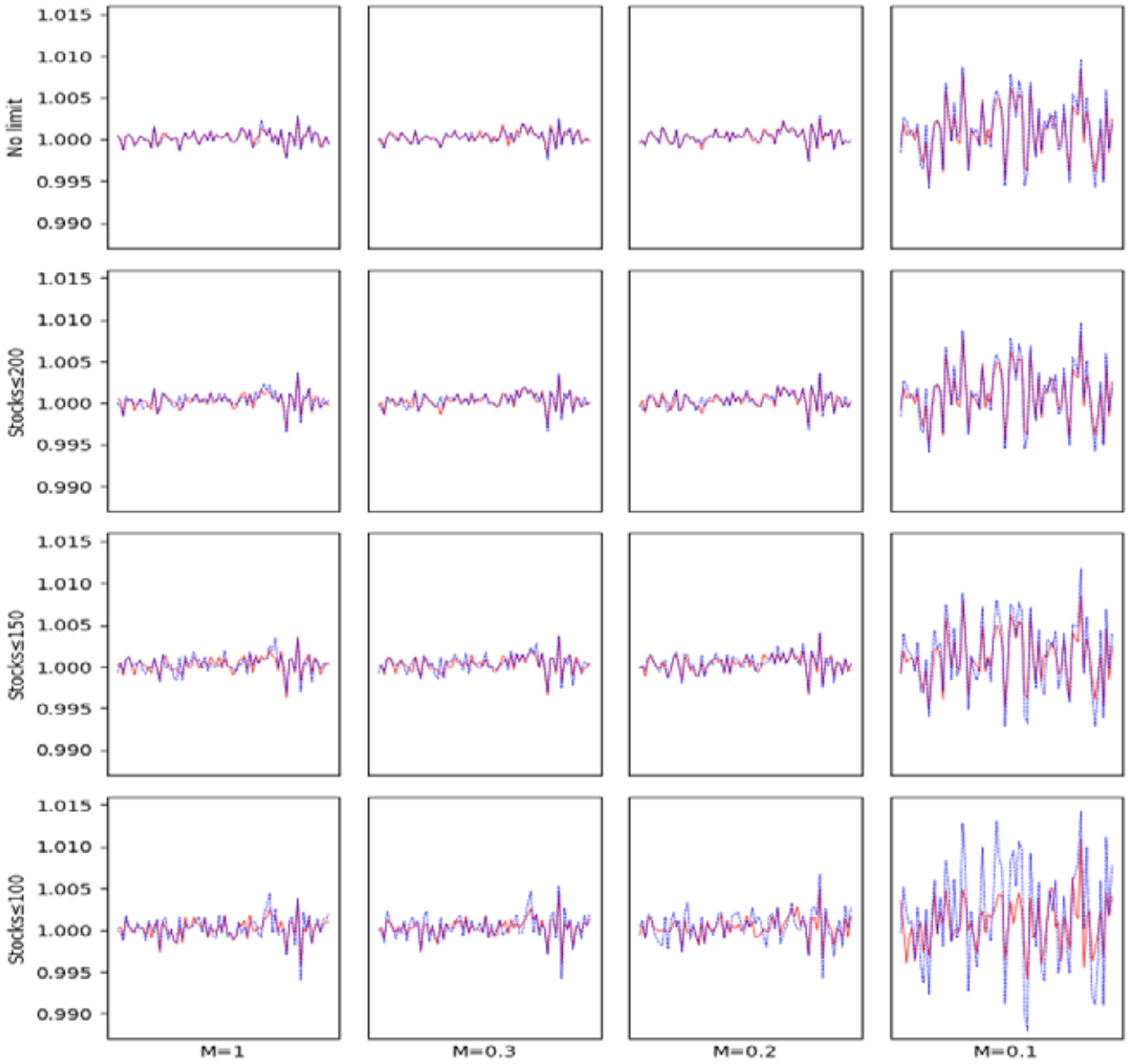


Figure 2: **Actual over predicted S&P 500 index (from 2021-2-24 to 2021-9-1) using the best BoAREN models in Table 2. Red solid lines correspond to actual over predicted values by BoAREN; blue dash lines correspond to actual over predicted values by two-step AREN.**



4. Summary and Conclusions

The objective of this study has been to illuminate a growing body of research that aims at improving the generality and prediction capability of linear statistical models applied to large quantities of data. The prototypical example we have chosen here is that of index tracking in financial modeling, but the potential applications to large-scale data analysis in finance extend well beyond this. To accomplish this, we have chosen to present an exposition of the Arbitrary Rectangle-range Elastic Net (AREN), one of many algorithmic approaches to regularization of linear statistical models having a large number of unknown parameters. The challenge for these models is to find the most influential predictors and to estimate their coefficients in a way that minimizes the model prediction error. The

AREN is a special case of the more general ARGEN model studied in Ding et al. (2021) and is ideal in this context because it is broadly applicable, and its important properties (tractability, estimation consistency, and variable selection consistency) follow from the more general ARGEN, allowing these results to be described without lengthy proofs. Progress in this field of research has been accelerating along with the influence of data science and the availability of extensive and inexpensive computing resources. Accordingly, following the work of Bach (2008) our main contribution here has been to demonstrate that the prediction capability of the AREN method can be further improved through the use of bootstrapping. This has been shown here to be the case for the index-tracking problem applied to the S&P 500. The literature in mathematical finance has been for some time mostly dominated by stochastic calculus and derivations of derivative-pricing formulae. Less well represented are methods for carefully analyzing financial data to design portfolios or reliably estimate the many unknown parameters that the aforementioned pricing formulae require. It is our hope that this work has reinforced the importance of methods essential to the empirical side of finance.

References

Bach, F. R. (2008). Bolasso: model consistent Lasso estimation through the bootstrap. In Proceedings of the 25th international conference on Machine learning (ICML '08) (pp. 33–40.). New York, NY, USA: Association for Computing Machinery. DOI:10.1145/1390156.1390161.

Bickel, P. J., Ritov, Y., & Tsybakov, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. Ann. Statist., 37(4), 1705-1732. MR2533469. DOI: 10.1214/08-AOS620.

Ding, Y., Peng, Q., Song, Z., & Chen, H. (2021). Variable selection and regularization via arbitrary rectangle-range generalized elastic net. arXiv:2112.07785.

Efron, B., Hastie, T., & Tibshirani, R. (2007). Discussion: "The Dantzig selector: statistical estimation when p is much larger than n". Ann. Statist., 35(6), 2358-2364. MR2382646. DOI: 10.1214/009053607000000433.

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. Ann. Statist., 32(2), 407-499. MR2060166. DOI: 10.1214/009053604000000067.

Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw., 33(1), 1-22. DOI:10.18637/jss.v033.i01.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Data mining, Inference, and Prediction. Springer, New York: Springer Series in Statistics. MR2722294. DOI:10.1007/978-0-387-84858-7.

Lounici, K. (2008). Sup-norm convergence rate and sign concentration property of Lasso and Dantzig estimators. Electron. J. Stat., 2, 90-102. MR2386087. DOI: 10.1214/08-EJS177.

Negahban, S. N., Ravikumar, P., Wainwright, M. J., & Yu, B. (2012). A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers. Statist. Sci., 27(4), 538-557. MR3025133. DOI: 10.1214/12-STS400.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. J. Roy. Statist. Soc. Ser. B, 58(1), 267-288. MR1379242. DOI:10.1111/j.2517-6161.1996.tb02080.x.

Wang, H., Li, G., & Tsai, C.-L. (2007). Regression coefficient and autoregressive order shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B Stat. Methodol., 69(1), 63-78. MR2301500. DOI:10.1111/j.1467-9868.2007.00577.x.

Wu, L., & Yang, Y. (2014). Nonnegative Elastic Net and application in index tracking. *Appl. Math. Comput., 227*, 541-552. MR3146340. DOI:10.1016/j.amc.2013.11.049.

Wu, L., Yang, Y., & Liu, H. (2014). Nonnegative-lasso and application in index tracking. *Comput. Statist. Data Anal., 70*, 116-126. MR3125482. DOI:10.1016/j.csda.2013.08.012.

Yuan, M., & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B Stat. Methodol., 68*(1), 49-67. MR2212574. DOI:10.1111/j.1467-9868.2005.00532.x.

Zhao, P., & Yu, B. (2006). On model selection consistency of Lasso. *J. Mach. Learn. Res., 7*, 2541-2563. MR2274449.

Zhao, P., Rocha, G., & Yu, B. (2009). The composite absolute penalties family for grouped and hierarchical variable selection. *Ann. Statist., 37*(6A), 3468-3497. MR2549566. DOI: 10.1214/07-AOS584.

Zhao, W., Zou, W., & Chen, J. J. (2014). Topic modeling for cluster analysis of large biological and medical datasets. *BMC Bioinformatics, 15*(S11), DOI:10.1186/1471-2105-15-S11-S11.

Zou, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc., 101*(476), 1418-1429. MR2279469. DOI:10.1198/016214506000000735.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol., 67*(2), 301-320. MR2137327. DOI:10.1111/j.1467-9868.2005.00503.x.