Sparse minimum average variance estimation via the adaptive elastic net when the predictors correlated

Esraa Rahman\(^1\) and Ali Alkenani\(^2\)

\(^1\)Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Al Diwaniyah, Iraq.
\(^2\)Correspondence: Ali Alkenani, E-mail: ali.alkenani@qu.edu.iq
\(^3\)http://orcid.org/0000-0001-5067-2321

Abstract

A new model-free variable selection method was proposed in this article, which is called SMAVE-AdEN. We combined the effective sufficient dimension reduction method MAVE with the variable selection method which is called adaptive Elastic Net (AdEN) to introduce SMAVE-AdEN. The SMAVE-AdEN produces a sparse and accurate estimate when the predictors are highly correlated. The advantage of SMAVE-AdEN is that SMAVE-AdEN extended Adaptive Elastic net (AdEN) to nonlinear and multi-dimensional regression. Also, the SMAVE-AdEN enables MAVE to work with problems were the predictors are highly correlated. In addition, SMAVE-AdEN can exhaustively estimate dimensions, while selecting informative covariates simultaneously. The performance of SMAVE-AdEN is evaluated by both simulation and real data analysis.

Keywords: Dimension reduction, Variable selection, Minimum average variance estimation, Adaptive Elastic Net.
1. Introduction

When the amount of predictors \( p \) is large, the regression analysis might be highly challenging. A beneficial mechanism to deal with this obstacle is decreasing \( p \)-dimensional predictors vector \( \mathbf{x} \) with no loss regarding regression information.

The sufficient dimension reduction (SDR) theory has been presented by Cook (1998) to perform the above aim. Assuming \( y \) is response variable, also \( \mathbf{x} = (x_1, \ldots, x_p)^T \) is a \( p \times 1 \) has been predictor vector. SDR explores \( p \times d \) matrix \( \mathbf{B} \), in a way that \( y \perp \mathbf{x}^T \mathbf{B} \), in which \( \perp \) indicates independence. Also, dimension reduction subspace (DRS) is the column space spanned by \( \mathbf{B} \). The intersection of all DRS has been referred to as the central subspace (\( S_{y|x} \)). The \( S_{y|x} \) has contained all regression information regarding \( y|x \) (Yu and Zhu, 2013). Several methods has been submitted for obtaining \( S_{y|x} \). For instance, SIR (Li, 1991), SAVE (Cook and Weisberg, 1991) as well as PHD (Li, 1992).

When the mean function is of interest, (Cook and Li, 2002) has presented the notion that is related to central mean subspace (\( S_{\mathbb{E}|y|x} \)). For the purpose of estimate \( S_{\mathbb{E}|y|x} \), a number of DR methods were presented, as the iterative Hessian transformation (Cook and Li, 2002) as well as MAVE (Xia et al., 2002).

SDR approaches supplying the researchers with useful tools for gaining adequate DR; nevertheless, these methods suffer from that each DR direction has been linear combination regarding original predictors. This makes the resulting measures not ease.

The process of electing the predictors is a very important in constructing the model of multiple regression. Besides, the selection of the significant predictors for being in the model improves the prediction accuracy related to the model. Furthermore, the small subset of predictors performs the interpretation of the results easy. The regularisation methods were applied for variable selection with regard to regression models from various researchers. See, for instance the Lasso (Tibshirani, 1996), SCAD (Fan and Li, 2001), Elastic Net (EN) (Zou and Hastie, 2005), adaptive Lasso (Zou, 2006) and MCP (Zhang, 2010).

Under SDR perspectives, the views of regularisation methods joined with several SDR methods from many researchers. For instance, Li et al. (2005), Ni et al. (2005), Li and Nachtsheim (2006), Li (2007), Li and Yin (2008). Wang and Yin (2008)
combined Lasso with MAVE to produce sparse MAVE (SMAVE) estimate. Wang et al. (2013) proposed penalised MAVE (P-MAVE) through combining bridge penalty with \( l_1 \)-norms of the rows of a basis matrix. Alkenani and Yu (2013) incorporate MAVE with SCAD, adaptive Lasso and the MCP to produce SCAD-MAVE, ALMAVE and MCP-MAVE, respectively. Wang et al. (2015) combined Lasso with the group-wise MAVE which suggested by Li et al. (2010). Alkenani and Rahman (2020) proposed SMAVE-EN method by combining MAVE with EN penalty to produce sparse and accurate estimates. The Lasso does not have oracle properties and it is not stable. Furthermore, adaptive Lasso achieves the oracle property, yet it inherits the instability related to Lasso with regard to the high-dimensional data. Elastic net handles collinearity, yet lacking oracle property. Zou and Zhan (2015) proposed Adaptive Elastic Net (AdEN) to tackle the limitations of EN. They proposed to employ adaptive Lasso instead of Lasso in EN penalty.

In this article, the SMAVE-AdEN has been presented. The SMAVE-AdEN is a shrinkage estimation method under SDR settings, where there is a set of predictors among which the predictors are highly pairwise correlated. SMAVE-AdEN is a nice combination of MAVE and AdEN. SMAVE-AdEN has advantages over the SMAVE (Wang and Yin, 2008), SPMAVE (Alkenani and Yu, 2013), P-MAVE (Wang et al., 2013) and SMAVE-EN (Alkenani and Rahman, 2020). It benefits from the strength of AdEN. In AdEN, the variable selection and parameters estimation have been implemented in one process and AdEN has the oracle property to select groups of highly correlated variables. The mentioned ability does not hold for Lasso, adaptive lasso, SCAD, MCP, bridge penalties and EN which are employed in the existing methods.

The rest of this article is as follows. In Section 2, a summary of MAVE and SMAVE-EN method is presented. We proposed the SMAVE-AdEN in Section 3. Simulation studies are implemented in Section 4. In Section 5, the methods under consideration are applied to graduate student rate data. The conclusions are given in Section 6.
2. MAVE and SMAVE-EN

In this section, we propose a brief of MAVE and SMAVE-EN. Suppose the following model:

\[ y = f(x_1, x_2, \ldots, x_p) + \varepsilon, \]  

where \( y \), \( x \) and \( \varepsilon \) are the response variable, a \( p \times 1 \) predictor vector and the error term, respectively. In addition, \( E(y|x) = f(x_1, x_2, \ldots, x_p) \) and \( E(\varepsilon | x) = 0 \). For the mean function, the SDR aims to investigate a subspace \( S \) such that

\[ y \perp E(y|x) | P_{s}x, \]  

where \( P_{s} \) is a projection operator. The mean DR subspaces achieve (Cook and Li, 2002). If \( d = \text{dim}(S) \) and \( B = (B_1, B_2, \ldots, B_d) \) is a basis for \( S \), \( x \) can be replaced with \( LCx^T B_1, x^T B_2, \ldots, x^T B_d, d \leq p \) without a loss of information on \( E(y|x) \). Cook and Li, (2002) show that the central mean subspace \( S_{E(y|x)} \) is the intersection of all subspaces satisfying (2). Many methods were proposed to estimate \( S_{E(y|x)} \) and one of the more efficient methods is MAVE. 

Xia et al. (2002) proposed MAVE such that the matrix \( B \) is the solution of

\[ \min_{\mathbf{B}} \{ \mathbf{y} - E(y|x^T B) \}^2, \]  

where \( \mathbf{B}^T \mathbf{B} = \mathbf{I}_d \). The conditional variance given \( x^T \mathbf{B} \) is

\[ \sigma_B^2(x^T \mathbf{B}) = E[(y - E(y|x^T \mathbf{B}))^2 | x^T \mathbf{B}]. \]

Thus,

\[ \min_{\mathbf{B}} E[y - E(y|x^T \mathbf{B})]^2 = \min_{\mathbf{B}} E[\sigma_B^2(x^T \mathbf{B})]. \]

For any given \( x_0 \), \( \sigma_B^2(x_0^T \mathbf{B}) \) can be locally approximated as

\[ \sigma_B^2(x_0^T \mathbf{B}) \approx \sum_{i=1}^{n} \{ y_i - E(y_i|x_0^T \mathbf{B}) \}^2 \omega_{i0} \]

\[ \approx \sum_{i=1}^{n} \{ y_i - (a_0 + (x_i - x_0)^T \mathbf{B} b_0) \}^2 \omega_{i0}, \]

where \( \omega_{i0} \geq 0 \) are the kernel weights with \( \sum_{i=1}^{n} \omega_{i0} = 1 \). So, \( \mathbf{B} \) can be found by solving

\[ \min_{\mathbf{B}} \mathbf{B}^T \mathbf{B} = \mathbf{I}_m \left( \sum_{j=1}^{m} \sum_{i=1}^{n} \left[ y_i - \left( a_j + (x_i - x_j)^T \mathbf{B} b_j \right) \right]^2 \omega_{ij} \right). \]
Alkenani and Rahman (2020) incorporate EN penalty term in (6) to obtain a SMAVE. The SMAVE minimises:
\[
\sum_{j=1}^{n} \sum_{i=1}^{n} \left[ y_i - \left\{ a_j + (x_i - x_j)^T \mathbf{B} \mathbf{b}_j \right\} \right]^2 \omega_{ij} + \frac{\lambda_1}{n} \| \beta_m \|^2 + \lambda_2 \| \beta_m \|_1, \tag{7}
\]
for \( m = 1, \ldots, d \).

where, \( d \) has been known and it can be estimated by BIC. where, \( |\beta_m|^2 \) is \( l_2 \) norm related with ridge penalty and \( |\beta_m|_1 \) is \( l_1 \) norm related with Lasso penalty. The minimisation in (7) consists of three parts. The first part is the loss function of MAVE based on the least-squares formulation of MAVE. The second part is the ridge penalty function stabilizes the solution paths, handles the collinearity and, thus, improving the prediction. The third part is the Lasso penalty function which performs parameters estimation and variable selection simultaneously. The EN penalty consists of the second and the third parts. The EN inherit the advantages and disadvantages of ridge and Lasso. Also, \( \lambda_1 \) and \( \lambda_2 \) are the tuning parameters of Elastic Net. Such good properties making SMAVE-EN extremely significant variable selection approach. In spite of its significance, SMAVE-EN have two disadvantages: which are, lacking oracle property and bias for estimating.

In this article, for the reasons mentioned, SMAVE-AdEN is proposed to minimise
\[
\sum_{j=1}^{n} \sum_{i=1}^{n} \left[ y_i - \left\{ a_j + (x_i - x_j)^T \mathbf{B} \mathbf{b}_j \right\} \right]^2 \omega_{ij} + \frac{\lambda_2}{n} \| \beta_m \|^2 + \lambda_2 \| \beta_m \|_1 \tag{8}
\]
1. Let \( m = 1, \) and \( \mathbf{B} = \beta_0, \) any arbitrary \( p \times 1 \) vector.

2. For known \( \mathbf{B}, \) get \( (a_j, b_j) \) where \( j = 1, \ldots, n \), from
\[
\min_{a_j, b_j=1, \ldots, n} \left( \sum_{j=1}^{n} \sum_{i=1}^{n} \left[ y_i - \left\{ a_j + (x_i - x_j)^T \mathbf{B} b_j \right\} \right]^2 \omega_{ij} \right). \tag{10}
\]

3. For a given \( (\hat{a}_j, \hat{b}_j), j = 1, \ldots, n, \) solve \( \beta_{\text{mSMAVE-AdEN}} \) from
\[
\min_{\mathbf{B}: \mathbf{B}^T \mathbf{B} = \mathbf{I}_m} \left( \sum_{j=1}^{n} \sum_{i=1}^{n} \left[ y_i - \left\{ \hat{a}_j + (x_i - x_j)^T (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_{m-1}, \hat{\beta}_m) \mathbf{b}_j \right\} \right]^2 \omega_{ij} + \right.
\]
\[
\left. \left( \frac{\lambda_1}{n} \| \beta_m \|^2 + \frac{\lambda_2}{n} \sum_{k=1}^{p} w_k^* \| \beta_{m,k} \|_1 \right) \right\} \tag{11}
\]
4. Replace the \( m \)th column of \( \mathbf{B} \) by \( \hat{\beta}_{\text{mSMAVE-AdEN}} \) and repeat steps 2 and 3 until convergence.
5. Update $\mathbf{B}$ by $(\hat{\beta}_{1\text{SMAVE-AdEN}}, \hat{\beta}_{2\text{SMAVE-AdEN}}, \ldots, \hat{\beta}_{m\text{SMAVE-AdEN}}, \beta_0)$, and set $m$ to be $m + 1$.

6. If $m < d$, continue steps 2 to 5 until $m = d$.

Where $\omega_{ij}$ are the kernel weights and they are computed as follows

$$
\omega_{ij} = K_h \left\{ (x_i - x_j)^T \hat{\mathbf{B}} \right\} / \sum_{i=1}^{n} K_h \left\{ (x_i - x_j)^T \hat{\mathbf{B}} \right\},
$$

$K_h$ is the refined multidimensional Gaussian kernel and $h_{opt} = \mathcal{A}(d)n^{-1/(4+d)}$ is the bandwidth, where $\mathcal{A}(d) = \left\{ \frac{4}{(d+2)} \right\}^{1/(4+d)}$, see (Xia et al., 2002).

SMAVE-AdEN combines AdEN into the “OLS formulation” of MAVE. Thus, under the same conditions as those for MAVE and EN, the algorithm well be converging to global minimum. According to the simulations of this study, the algorithm of SMAVE-EN usually converges within seven to twelve iterations. The LARS algorithm of Efron et al. (2004) can be manipulated to get the efficient solution of EN with the order of computational efforts has been majorly comparable to that related to single OLS fit (Zou and T. Hastie, 2005). Compared to MAVE, the penalty term in SMAVE-AdEN appears in the “OLS formulation” of MAVE, so the algorithm is as efficient as MAVE. At an early stage of LARS-EN algorithm, the optimal results have been achieved. After $m$ steps, if the algorithm of LARS-EN is settled, then it demands $O(m^3 + pm^2)$ operations (Zou and T. Hastie, 2005). Due to (Xia et al., 2002), the rate of consistency for the MAVE estimator is $O(h_{opt}^3 \log(n))$. Because of that the rate of consistency of MAVE is a lower than that of AdEN, the rate of consistency of SMAVE-AdEN estimator is controlled by that of MAVE. In summary, under the same conditions of Zou and T. Hastie (2005) and Xia et al. (2002), one may show that the SMAVE-AdEN estimator has the same consistency rate as the MAVE estimator and it is also as efficient as MAVE asymptotically.

3. Simulation study

The goal in this section is to compare the performance of SMAVE-AdEN with the ALMAVE, SCAD-MAVE, MCP-MAVE, SMAVE, P-MAVE and SMAVE-EN methods in terms of prediction accuracy and variable selection. Also, to show the preference of SMAVE-AdEN when the grouped selection is required.

An amount of examples is reported to show the performance of the SMAVE-AdEN. The simulated data consist of a training set, an independent validation set and
an independent test set within each example. The training data were employed to fit the models, and the tuning parameters were selected by using the validation data. The test data were employed to compute the mean-squared error (MSE). The notation \( \ldots / \ldots \) refers to the number of observations in the mentioned three sets, respectively.

The ALMAVE, SCAD-MAVE and MCP-MAVE methods were computed using R codes made by Alkenani and Yu (2013). The SMAVE, P-MAVE and SMAVE-EN methods were computed using R codes made by Wang and Yin (2008), Wang et al. (2013) and Alkenani and Rahman (2020), respectively. The R code for SMAVE-AdEN is available from the authors. For each competitor, the tuning parameters was chosen via tenfold cross-validation (C.V).

**Example 1.** We generated data from the linear regression model.

\[
y = X^T \beta^* + \epsilon,
\]

where \( \beta^* \) is a \( p \) -dimensional vector and \( \epsilon \sim N(0, \sigma^2) \), \( \sigma = 6 \), and \( X \) is from multivariate normal distribution with zero mean and covariance \( \Sigma \) whose \((j,k)\) entry is \( \Sigma_{jk} = \rho^{\left|j-k\right|}, 1 \leq k, j \leq p \).

We considered \( \rho = 0.5 \) and \( \rho = 0.75 \). Let \( p = p_n = \left[4 n^{1/2}\right] - 5 \) for \( n = 50, 100 \). Let \( 1_{m}/0_{m} \) denote a \( m \) -vector of \( 1 \)'s/0's. The vector of true coefficients is \( \beta^* = (3.1q, 3.1q, 3.1q, 0_{p-3q})^T \) and \( |\mathcal{A}| = 3q \) and \( q = \left[ p_n/9 \right] \). In this example \( v = \frac{1}{2} \) hence, we used \( y = 3 \) for computing the adaptive weights in the adaptive elastic-net. Let \( \mathcal{A} = \{ j : \beta_j^* \neq 0, j = 1, 2, ..., p \} \), \( \gamma > \frac{2v}{1-v}, 0 \leq v < 1 \)

\( \mathcal{A} \) is the intrinsic dimension of the underlying model.

With regard to each estimator \( \hat{\beta} \), the estimation accuracy will be evaluated via mean squared error (MSE) that is specified as \( E[(\hat{\beta} - \beta^*)^T \Sigma (\hat{\beta} - \beta^*)] \). Also, variable selection performance has been gauged via \((C, IC)\), in which \( C \) has been the number of zero coefficients which have been estimated correctly via zero, also \( IC \) representing the number of the nonzero coefficients which have been estimated incorrectly via zero.

**Example 2.** We generated data from the linear regression model
\begin{align*}
y = \frac{x^T \beta_1}{0.5 + (15 + x^T \beta_2)} + \sigma \epsilon,
\end{align*}

\text{where } \beta^* \text{ is a } p \text{ -dimensional vector and } \epsilon \sim N(0, \sigma^2), \sigma = 6, \text{ and } X \text{ is from multivariate normal distribution with zero mean and covariance } \Sigma \text{ whose } (j, k) \text{ entry is } \Sigma_{j,k} = \rho^{|j-k|}, 1 \leq k, j \leq p. \text{ We considered } \rho = 0.5 \text{ and } \rho = 0.75. \text{ Let } p = p_n = \left[4 n^{1/2}\right] - 5 \text{ for } n = 100, 200, 400. \text{ Let } 1_m/0_m \text{ denote a } m \text{ -vector of } 1' \text{s/0' s.} \text{ The vector of true coefficients are } \\
\beta_1 = (3.1_q, 3.1_q, 3.1_q, 0_{p-3q})^T \text{ and } \beta_2 = (0_{p-3q}, 3.1_q, 3.1_q, 3.1_q)^T \text{ and } |\mathcal{A}_0| = 3q \text{ and } q = \left[p_n/9\right]. \text{ In this example } v = \frac{1}{2} \text{ hence, we used } y = 3 \text{ for computing the adaptive weights in the adaptive elastic-net.} \n
\gamma > \frac{2v}{1-v}, \quad 0 \leq v < 1, \quad \mathcal{A}_0 = \{j : \beta_j^* \neq 0, j = 1, 2, ..., p\}
Table 1: Model selection (ME) and fitting results based on 100 replications for $\rho = 0.5$ in example 1

| $n$ | $p_n$ | $|\mathcal{A}|$ | Model       | MSE     | C    | IC  |
|-----|-------|---------------|-------------|---------|------|-----|
| 100 | 35    | 11            | Truth       | 26.00   | 0    |     |
|     |       |               | SMAVE       | 8.87 (0.65) | 22.74 | 0.41 |
|     |       |               | ALMAVE      | 5.22 (0.31) | 24.25 | 0.13 |
|     |       |               | SCADMAVE    | 9.05 (0.53) | 22.00 | 0.52 |
|     |       |               | MCPMAVE     | 7.53 (0.62) | 22.70 | 0.34 |
|     |       |               | SMAVE-EN    | 4.92 (0.24) | 24.75 | 0.09 |
|     |       |               | SMAVE-AdEN  | 4.02 (0.29) | 24.90 | 0.08 |

| 200 | 51    | 17            | Truth       | 36.00   | 0    |     |
|     |       |               | SMAVE       | 6.12 (0.57) | 33.24 | 0.12 |
|     |       |               | ALMAVE      | 4.01 (0.24) | 34.25 | 0    |
|     |       |               | SCADMAVE    | 3.88 (0.33) | 33.94 | 0.08 |
|     |       |               | MCPMAVE     | 5.50 (0.50) | 33.83 | 0.09 |
|     |       |               | SMAVE-EN    | 3.22 (0.17) | 35.02 | 0    |
|     |       |               | SMAVE-AdEN  | 2.94 (0.19) | 35.32 | 0    |

| 400 | 75    | 25            | Truth       | 51.00   | 0    |     |
|     |       |               | SMAVE       | 4.11 (0.17) | 48.34 | 0    |
|     |       |               | ALMAVE      | 2.61 (0.18) | 50.01 | 0    |
|     |       |               | SCADMAVE    | 3.02 (0.12) | 49.44 | 0    |
|     |       |               | MCPMAVE     | 3.78 (0.22) | 48.73 | 0    |
Table 2: ME and fitting results based on 100 replications for $\rho = 0.75$ in example 1

| $n$  | $p_n$ | $|\mathcal{A}|$ | Model     | MSE    | $C$     | IC     |
|------|-------|-----------------|-----------|--------|---------|--------|
| 100  | 35    | 11              | Truth     | 26     | 0       |        |
|      |       |                 | SMAVE     | 6.77(0.28) | 22.95  | 0.49 |
|      |       |                 | ALMAVE    | 5.12(0.17) | 24.75  | 0.17 |
|      |       |                 | SCADMAVE  | 10.05(0.22) | 22.13  | 1.11 |
|      |       |                 | MCPMAVE   | 7.75(0.24) | 22.82  | 1.24 |
|      |       |                 | SMAVE-EN  | 4.95(0.15) | 24.79  | 0.14 |
|      |       |                 | SMAVE-AdEN| 4.21(0.18) | 24.98  | 0.10 |
| 200  | 51    | 17              | Truth     | 36     | 0       |        |
|      |       |                 | SMAVE     | 6.44(0.51) | 34.11  | 0.13 |
|      |       |                 | ALMAVE    | 4.33(0.23) | 35.01  | 0.03 |
|      |       |                 | SCADMAVE  | 4.02(0.31) | 34.99  | 0.55 |
|      |       |                 | MCPMAVE   | 5.61(0.42) | 34.32  | 0.89 |
|      |       |                 | SMAVE-EN  | 3.35(0.14) | 35.41  | 0     |
|      |       |                 | SMAVE-AdEN| 3.12(0.15) | 35.63  | 0     |
| 400  | 75    | 25              | Truth     | 51     | 0       |        |
|      |       |                 | SMAVE     | 4.33(0.13) | 49.34  | 0.11 |
|      |       |                 | ALMAVE    | 2.98(0.18) | 50.33  | 0     |
|      |       |                 | SCADMAVE  | 3.45(0.12) | 49.95  | 0.07 |
|      |       |                 | MCPMAVE   | 3.90(0.12) | 49.11  | 0.09 |
From Table 1 and 2 the prediction results can be summarized as follows. First, it is clear that the SMAVE has the worst performance. SMAVE-AdEN is considerably more accurate than all the considered methods. In general, the ALMAVE and SMAVE-EN was competitor for SMAVE-EN and its performance was better than the rest methods for all the examples. The results of simulation indicate that the SMAVE-AdEN dominates the SMAVE-EN, ALMAVE, SCAD-MAVE, P-MAVE, MCP-MAVE and SMAVE methods under collinearity.

Example 1.
Figure 1: MSE for the considered methods based on example 1.

Table 3: ME and fitting results based on 100 replications for $\rho = 0.5$ in example 2

| $n$ | $p$ | $n$ | $|\mathbf{A}|$ | Method | $\beta_1$ | $\beta_2$ |
|-----|-----|-----|----------------|--------|-----------|-----------|
|     |     |     |                |        | MSE       | C         | IC        |
| 100 | 81  | 27  | Truth          | 54     | 0         | 54        | 0         |
|     |     |     | SMAVE          | 41.17(1.60) | 47.04    | 2.31      | 43.88(2.50) | 46.80    | 2.41 |
|     |     |     | ALMAVE         | 21.31(0.31) | 53.25    | 0.25      | 23.78(1.11) | 53.00    | 0.28 |
|     |     |     | SCADMAVE       | 43.15(1.42) | 50.20    | 1.32      | 44.25(1.33) | 49.91    | 1.44 |
|     |     |     | MCPMAVE        | 34.11(1.55) | 48.82    | 2.22      | 36.51(1.22) | 47.14    | 2.32 |
|     |     |     | SMAVE- EN      | 19.02(0.32) | 54.24    | 0.19      | 19.82(0.45) | 53.17    | 0.20 |
|     |     |     | SMAVE- AdEN    | 18.42(0.65) | 54.97    | 0.17      | 19.10(0.51) | 54.74    | 0.17 |
| 200 | 131 | 43  | Truth          | 89     | 0         | 89        | 0         |
|     |     |     | SMAVE          | 22.00(0.67) | 81.32    | 0.22      | 22.50(0.37) | 80.40    | 0.24 |
|     |     |     | ALMAVE         | 16.74(0.54) | 86.33    | 0         | 17.00(0.43) | 86.21    | 0     |
|     |     |     | SCADMAVE       | 20.45(0.42) | 82.90    | 0.07      | 21.22(0.42) | 81.55    | 0.10 |
|     |     |     | MCPMAVE        | 20.81(0.66) | 33.83    | 0.14      | 22.40(0.66) | 35.20    | 0.17 |
|     |     |     | SMAVE- EN      | 11.05(0.30) | 87.52    | 0         | 11.47(0.30) | 87.11    | 0     |
| $n$ | $\rho$ | $|A|$ | Method | $\beta_1$ | $\beta_2$ |
|-----|-------|------|--------|----------|----------|
|     |       |      |        | MSE      | C        | IC |
|     |       |      |        | MSE      | C        | IC |
| 100 | 81    | 27   | Truth  |          | 54       | 0  |
|     |       |      | SMAVE  | 33.18(1.53) | 50.00    | 4.21 |
|     |       |      | ALMAVE | 16.22(0.97) | 53.21    | 1.15 |
|     |       |      | SCADMAVE | 26.15(1.22) | 52.97    | 2.24 |
|     |       |      | MCPMAVE | 31.52(1.37) | 51.80    | 2.92 |
|     |       |      | SMAVE-EN | 13.53(0.46) | 54.88    | 0.56 |
|     |       |      | SMAVE-AdEN | 13.11(0.28) | 54.90    | 0.49 |
| 200 | 131   | 43   | Truth  |          | 89       | 0  |
|     |       |      | SMAVE  | 21.44(0.56) | 83.17    | 3.01 |
|     |       |      | ALMAVE | 12.88(0.48) | 86.93    | 0.52 |
|     |       |      | SCADMAVE | 17.30(0.32) | 85.90    | 1.22 |
|     |       |      | MCPMAVE | 18.75(0.57) | 84.20    | 2.14 |
|     |       |      | SMAVE-EN | 9.05(0.29)  | 88.13    | 0.12 |

Table 4: ME and fitting results based on 100 replications for $\rho = 0.75$ in example 2
From Table 3 and 4, it is obvious that the SMAVE-AdEN produces sparse models. The SMAVE-AdEN tends to select the true important predictors more than the all competitors. The performance of SMAVE-AdEN was very well when grouped selection is required. The ability of ‘grouped selection’ of AdEN makes the SMAVE-AdEN a better than the SMAVE-EN, ALMAVE, SCAD-MAVE, P-MAVE, MCP-MAVE and SMAVE methods in term of variable selection. In all the examples, the performance of SMAVE-EN and ALMAVE for variable selection was better than the the performance of SCAD-MAVE, P-MAVE, MCP-MAVE and SMAVE methods. In terms of selection accuracy, the performance of SMAVE was the worst for all the examples.
4. Graduate student rate (GSR) data

In this part, the GSR data has been analysed by SMAVE-AdEN, SMAVE-EN, ALMAVE, SCAD-MAVE, P-MAVE, MCP-MAVE and SMAVE methods. The performance of the studied methods was compared via computing their prediction MSE on the test data. We apply the most important factors that affect performance of postgraduate students in Iraq data. The data had been collected by the students from the Al-Qadisiyah University in Al-Diwaniya to achieve this aim.

The researcher prepared a questionnaire to obtain the data required to study the most important factors that affect performance of postgraduate students in Iraq (University of Al-Qadisiyah as an example). To approve its validity, the text of the questionnaire was given to a jury to evaluate its appropriateness to investigate the subject under study. The jury consists of tutors of high experience and direct contact with the postgraduate students. The form of the questionnaire was first distributed to a population of postgraduate students in the college of Economy and Administration to explore their opinion about it. After getting their views, the final version of the form became ready to distribute as a procedure to collect the data required. The questionnaire is made up of 57 variables that are viewed to affect performance of the postgraduate students in Iraq. Fifty forms of the questionnaire were distributed to the
postgraduate students in University of Al-Qadisiyah. After responding to the questionnaires, the data were analysed by R code which was written by the author. The suggested method was compared with the methods already available. After analysing the data, we obtained some results which are mentioned in the tables 1 and 2.

GSR data contains $n = 50$ observations. The response $Y$ is GSR. The covariates are $X_1$ (Sex), $X_2$ (age), $X_3$ (marital state), $X_4$ (number of children of the students who are married), $X_5$ (number of brothers), $X_6$ (number of sisters), $X_7$ (students order in the family), $X_8$ (education level of the older brother whose level is higher than that of the student), $X_9$ (mother's occupation), $X_{10}$ (father's occupation), $X_{11}$ (education level of the mother), $X_{12}$ (education level of the father), (family income) $X_{13}$, $X_{14}$ (student income), $X_{15}$ (how far the student is materially and morally responsible in his family), $X_{16}$ (stable and positive family environment), $X_{17}$ (family support to the student), $X_{18}$ (the very ambitions fathers who practice pressure on their children to get higher average in their study), $X_{19}$ (death of one of the parents), $X_{20}$ (preference of one of the two sexes on the other), $X_{21}$ (computer and internet proficiency of the student), $X_{22}$ (English proficiency of the student), $X_{23}$ (teaching training and experience of the student), $X_{24}$ (psychology of the student), $X_{25}$ (health of the student (chronic)), $X_{26}$ (treatment of the student inside the classroom), $X_{27}$ (student's participation inside the classroom), $X_{28}$ (availability of appropriate environment in the surroundings of the student), $X_{29}$ (number of hours assigned for reading per a day), $X_{30}$ (worry and confusion during tests), $X_{31}$ (support of the society when the student spends some leisure time with the others), $X_{32}$ (full cooperation and spirit of team work among the students), $X_{33}$ (difficulty of the study material), $X_{34}$ (relation of teachers and students and the way teachers treat students), $X_{35}$ (availability of references and well-equipped labs and classrooms), $X_{36}$ (scientific level of the teachers), $X_{37}$ (presenting lectures in modern and developed methods), $X_{38}$ (student's average in the Baccalaureate), $X_{39}$ (student's score in the competition test), $X_{40}$ (ability to concentrate and paying attention), $X_{41}$ (modernity of the topic of student's thesis), $X_{42}$ (ability of the student to defend his thesis), $X_{43}$ (degree of agreement between the supervisor and the student), $X_{44}$ (the specific specialization of members of the examining committee is like that of the student's thesis), $X_{45}$ (how much objective the members of the examining committee are when granting the student's score in his thesis debate), $X_{46}$
(how far the house of the student from the place where he is studying), \(X_{47}\) (social class of the student), \(X_{48}\) (security in the environment surrounding the student), \(X_{49}\) (degree of availability of the electrical power), \(X_{50}\) (the duration between the last degree the student got and his current study), \(X_{51}\) (emotional relation (stability) with the other sex), \(X_{52}\) (number of years of no-pass during the undergraduate study), \(X_{53}\) (how much the student likes his specialization), \(X_{54}\) (student’s satisfaction of belonging to the university), \(X_{55}\) (number of student's in the graduation group), \(X_{56}\) (future opportunities available after getting the high certificate), \(X_{57}\) (attitude of society towards the postgraduate student).

Table 5: the adjusted R-square values for the model fit based on the real data

| Model Fit    | SMAVE-EN | SMAVE | ALMAVE | SCADMAVE | MCPMAVE | SMAVE-AdEN |
|--------------|----------|-------|--------|----------|---------|------------|
| Linear       | 0.93     | 0.77  | 0.93   | 0.77     | 0.77    | 0.93       |
| Quadratic    | 0.94     | 0.91  | 0.94   | 0.91     | 0.91    | 0.94       |
| Cubic        | 0.94     | 0.92  | 0.94   | 0.92     | 0.92    | 0.94       |
| Quartic      | 0.94     | 0.92  | 0.94   | 0.92     | 0.92    | 0.94       |

Table 5, reports the adjusted R-squared values for the model fit, based on the GSR data. The studied methods have discovered a nonlinear structure, which can be approximated by a cubic fit. Also, it can be observed that the adjusted R-squared values for the SMAVE-EN and SMAVE-AdEN methods are bigger than the values of adjusted R-squared for the SMAVE, SCADMAVE and MCPMAVE. The adjusted R-squared values for the SMAVE, SCADMAVE and MCPMAVE are similar.

Table 6: the prediction error (P.E) of the cubic fit for the studied methods based on the real data

| Methods     | Prediction error |
|-------------|------------------|
| SMAVE       | 0.8212           |
| ALMAVE      | 0.6737           |
| SCADMAVE    | 0.8010           |
| MCPMAVE     | 0.8011           |
SMAVE-EN  0.6481
SMAVE- AdEN  0.6443

From Table 6, it is obvious that the SMAVE-AdEN method has a lower P.E than the SMAVE-EN, SMAVE, ALMAVE, SCADMAVE and MCPMAVE methods. This means that SMAVE-AdEN has better performance than the SMAVE-EN, SMAVE, ALMAVE, SCADMAVE and MCPMAVE methods.

6. Conclusion

In this article, SMAVE-AdEN has been proposed. SMAVE-AdEN combined the AdEN with MAVE method. MAVE can estimate $S_{E(y|x)}$ while AdEN performs a shrinkage estimation and variable selection simultaneously and it encourages groups selection of correlated predictor. The SMAVE-AdEN benefits from the advantages of MAVE and Adaptive Elastic Net. The SMAVE-AdEN enable AdEN to work with nonlinear and multi-dimensional regression. Computationally, the SMAVE-AdEN is proved to be ease implemented with an effective algorithm. From the results of simulation and real data, it is clear that SMAVE-AdEN can have good predictive accuracy, as well as encourages groups variable selection for the strongly correlated predictors under SDR settings.

The proposed approach can be extended to SIR (Li, 1991), SAVE (Cook and Weisberg, 1991) and PHD (Li, 1992). Also, the SMAVE-AdEN can be extended to binary response models. Moreover, robust SMAVE-AdEN is another possible extension of the proposed method.

7. References

1. Alkenani, A. and Yu, K. (2013). Sparse MAVE with oracle penalties. Advances and Applications in Statistics 34, 85–105.
2. Cook, R. (1998). Regression graphics: ideas for studying the regression through graphics. New York, Wily.
3. Cook, R. D. and Li, B. (2002). Dimension reduction for the conditional mean in regression. The Annals of Statistics 30, 455–474.
3. Cook, R. D. and Weisberg, S. (1991). Discussion of Li (1991). Journal of the American Statistical Association 86, 328–332.

4. Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression. The Annals of Statistics 32, 407–499.

5. Fan, J. and Li, R. Z. (2001). Variable selection via non-concave penalized likelihood and its oracle properties. Journal of the American Statistical Association 96, 1348–1360.

6. Li, L. (2007). Sparse sufficient dimension reduction. Biometrika 94, 603–613.

7. Li, K. (1991). Sliced inverse regression for dimension reduction (with discussion). Journal of the American Statistical Association 86, 316–342.

8. Li, K. C. (1992). On principal Hessian directions for data visualization and dimension reduction: Another application of Stein’s lemma. Journal of the American Statistical Association 87, 1025–1039.

9. Li, L., Cook, R. D. and Nachtsheim, C. J. (2005). Model-free variable selection. Journal of the Royal Statistical Society Series B, 67, 285–299.

10. Li, L., Li, B. and Zhu, L.-X. (2010). Groupwise dimension reduction. Journal of the American Statistical Association. 105, 1188–1201.

11. Li, L. and Nachtsheim, C. J. (2006). Sparse sliced inverse regression. Technometrics 48, 503–510.

12. Li, L. and Yin, X. (2008). Sliced Inverse Regression with regularizations. Biometrics 64, 124–131.

13. Ni, L., Cook, R. D. and Tsai, C. L. (2005). A note on shrinkage sliced inverse regression. Biometrika 92, 242–247.

14. Stamey, T., Kabalin, J., McNeal, J., Johnstone, I., Freiha, F., Redwine, E. and Yang, N. (1989) Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate ii: radical prostatectomy treated patients. J. Urol., 16, 1076–1083.

15. Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. Journal of the Royal Statistical Society, Series B 58, 267–288.

16. Wang, Q. and Yin, X. (2008). A Nonlinear Multi-Dimensional Variable Selection Method for High Dimensional Data: Sparse MAVE. Computational Statistics and Data Analysis 52, 4512–4520.

17. Wang, T., Xu, P. and Zhu, L. (2013). Penalized minimum average variance estimation. Statist. Sinica 23 543–569.
18. Wang, T., Xu, P., Zhu, L. (2015). Variable selection and estimation for semiparametric multiple-index models. Bernoulli 21 (1), 242–275.
19. Xia, Y., Tong, H., Li, W. and Zhu, L. (2002). An adaptive estimation of dimension reduction space. Journal of the Royal Statistical Society Series B 64, 363–410.
20. Yu, Z. and Zhu, L. (2013). Dimension reduction and predictor selection in semiparametric models. Biometrika, 100, 641-654.
21. Zhang, C. H. (2010). Nearly unbiased variable selection under minimax concave penalty. Annals of Statistics 38, 894–942.
22. Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net, Journal of the Royal Statistical Society, Series B 67, 301–320.
23. Zou, H. (2006). The adaptive Lasso and its oracle properties. Journal of the American Statistical Association 101, 1418–142.