Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN$_{Online}$: 2312-9883
QJAE, Volume 27, Issue 4 (2025) ISSN$_{Print}$: 1816-9171
DOI: https:// doi.org/ 10.33916/ qjae.2025 .04120129

Penalized Methods in Semiparametric Single Index Models Using MAVE-LASSO and MAVE-Elastic Net

Sanaa.J.Tuama
SANAA.J.TUAMA@qu.edu.id
University of Al-Qadisiyah

Received: 24/8/2025
Accepted: 6/10/2025
Available online: 15 /12 /2025

Corresponding Author : Sanaa.J.Tuama

Abstract : Semiparametric Single Index Models are important and essential tools for addressing the high-dimensional problem, as they play an important role in the model-building process and selecting marginal variables. In this research, some modern penal methods have been used, which assess the vector of parameters and simultaneously select the variable for the quasi-parameter single indicator models (MAVE-LASSO) and MAVE-Elastic net) to improve the accuracy and predictability of the model. In order to achieve this goal, simulation experiments were conducted to demonstrate the preferences of the methods used in estimating and selecting the variable for the model. Different models, different variations, different sample sizes, and real data of factors influencing patients with blood diabetes were used for comparison and verification of the performance of these methods in practice. The methods studied will be compared by relying on two benchmarks for comparison: the average mean square error (AMSE) and the absolute average mean square error (AMAE), and the results will be obtained based on the R-code.
Theoretically and simulation-wise, the MAVE-EN method for estimating and selecting important variables for a single-indicator semi-parametric model has been shown to be effective in dealing with cases of high correlation between explanatory variables in the presence of different variances.

Keywords: MAVE, Semiparametric Single Index Models, variable selection, LASSO, elastic net

INTRODUCTION: Previous studies have Semiparametric single index models have gained significant attention due to their flexibility and effectiveness in handling high-dimensional data[7]. These models extend traditional linear regression by incorporating a nonparametric link function, allowing for more complex relationships between variables[8]. Estimation and variable selection within these models have been enhanced through the development of modern penalized methods such as LASSO and Elastic Net, which enable simultaneous estimation of parameters and identification of important variables, [7].
Research has shown that penalty functions like LASSO, and Elastic Net [20], improve model interpretability and prediction accuracy, especially in datasets with multicollinearity or noise. Zou and Hastie demonstrated the advantages of Elastic Net in dealing with highly correlated variables by combining L1 and L2 penalties, leading to more stable variable selection[20], the oracle properties of adaptive LASSO, which can identify true significant variables with high probability. These techniques are particularly useful in high-dimensional settings where traditional methods may fail or suffer from over fitting[9].
In practical applications, such as medical studies, the use of penalized semi-parametric models has been instrumental in selecting relevant factors influencing health outcomes. For example, in the study of gestational diabetes, variable selection techniques have identified key predictors, which improved the understanding of the disease's risk factors and contributed to better clinical decision-making.
These methods simultaneously estimate model parameters, select relevant variables, and determine the appropriate link function.,
These methods have since become state-of-the-art tools, applied extensively in genomics for gene selection, in finance for constructing risk scores, and in econometrics for analyzing complex consumer behavior, demonstrating their power and flexibility across disciplines.

Semi-parametric single index model:

The single index model (SIM) is one of the most commonly used semi-parametric models in econometrics. It was introduced by [1] and its estimation methods were further developed by [2]. This model allows the conditional mean of the response variable to be a non-parametric function of a linear combination of the predictive variables [3], [4]. The semi-parametric single index model can be mathematically represented as follows:

$$Y = g(X^T\theta) + \varepsilon \quad (1)$$

And its conditional expectation function (link function):

(2)

Where:

- $Y=(y_1,y_2,\ldots, y_n)^T$ represents the response variable of order n×1,
- $X=(x_1,x_2,\ldots, x_p)^T$ represents the predictor variable vector of order p×1,
- $\theta=(\theta_1,\theta_2,\ldots, \theta_p)^T$ represents a p-dimensional vector of unknown parameters, representing the parametric part of the model, and satisfying the $\theta^T\theta=1$ or $||\theta||=1$ condition for model diagnosis.
- $g(X^T\theta)$ represents an unknown, measurable link function, constituting the nonparametric part of the model.
- ε represents a random error with a normal distribution, a mean of zero, and a variance specified by σ^2, satisfying $E(\varepsilon|X^T\theta)=0$

This model has a single linear component for the explanatory variables, through which most of the information about the relationship between the response variable and the explanatory variables can be obtained, thus avoiding the curse of dimensionality. The semi-parametric single index model (SSIM) can be estimated using a two-step estimation procedure. In the first step, the parameter vector b is estimated to find the index value ($X_i^T\hat{\theta}$). In the second step, the index values for each observation are used to estimate the link function $g(X^T\theta)$ using univariate nonparametric regression as a preliminary function that depends on the data $(X_i^T\theta, Y_i)$ (Kong & Xia, 2007; Naik, 2001).

1. Advantages of SIM:

The SIM has numerous advantages, which have been studied by many researchers in various fields, including economics and sociology. These advantages can be explained as follows:

1- The SIM does not assume a known link function g(.), and thus is more flexible and less restrictive than parametric models for conditional mean functions such as linear models and binary choice models [4].

2- The use of the single index model (SIM) reduces the risk of reaching misleading results [4].

3- Although nonparametric estimation of the conditional mean function increases flexibility and reduces the risk of misspecification, it does not eliminate it, and the cost of this flexibility can be high for several reasons, which are [2]:

a. The accuracy of nonparametric estimates decreases rapidly as the dimension of X increases. To obtain a reasonable level of accuracy when X is high-dimensional, large sample sizes may be required.

b. Nonparametric estimates can be difficult to interpret when X is high-dimensional.

c. Nonparametric estimation does not allow for extrapolation and does not provide predictions for E[Y|X] at points X that are not in the data. This is a serious drawback for analysis and forecasting. In contrast, the single index model allows for boundary extrapolation and provides predictions for E[Y|X] at values of X that are not in the data but are within the range of $X^T\theta$.

4- The SSIM has the ability to overcome the curse of dimensionality [5] through the index structure, since $X^T\theta$ is an index of the dimensions of X. At the same time, the parameter vector θ can be estimated with the same $(-\sqrt{n})$ convergence rate achieved in parametric models for estimating the parameter vector θ, and it is as accurate as a one-dimensional nonparametric model in estimating the link function $g(X_i^T\theta)$ [6].

2. Variable Selection:

Variable selection technique is very important in many types of statistical modeling, including semiparametric modeling, as it seeks to simultaneously reduce the chances of including irrelevant data and minimize the effects of bias [7]. In many applications, we have large datasets where we need to identify the number of important (significant) variables. These applications include finance, business, medicine, engineering, environment, sociology, and others. The process of variable selection is an important part of most applications, and its importance becomes apparent when the true function is nonlinear and the dataset often contains multicollinearity or outliers. It is known that including a large number of variables in a regression equation for any phenomenon or study costs time, effort, and money, and the presence of some non-essential variables in their effect on the dependent variable or their effect is similar to the effect of other variables or these variables have a high internal correlation among them, which makes their effect insignificant, which calls for the exclusion of such variables. In this research, modern techniques were used in the process of selecting significant variables based on different penalty functions, namely the lasso [8] and elastic net [9] penalty functions. These techniques have emerged since 1990 and are called penalty methods, which is an alternative approach to variable selection in linear regression models. These methods have been described for semiparametric models, including the single index model, which has a behavior in estimating and selecting variables in the modeling process itself, i.e., it works on estimating the parameter vector θ and the link function g(.) and automatically selecting significant variables at the same time [10].

3. Penalized likelihood:

The role of the penalty function is to prevent overfitting, meaning that a model with many input variables is suboptimal (not significant). When the correct model is sparse, meaning that the dependent variable depends only on a small number of input variables (significant), the penalty function relies on the penalty parameter (λ). Through this parameter, a penalty limit is set, and this limit will push towards obtaining very large parameters, while parameters that are smaller than the penalty value will have a value of zero, thus the selection of the appropriate model is automatic [11], [12].

[13] argues that a good penalty function should lead to an estimator with three essential properties: "unbiasedness", "sparsity", and "continuity". The penalty functions used in this research are:

5.1 LASSO Penalty function:

It was previously proposed by [8] and is also called the l_1 penalty function and has the following form:

$$P_\lambda(|\theta_j|)=\lambda|\theta_j| \quad (3)$$

5.2 Elastic Net Penalty function:

It was proposed by [9] and is a combination of the l_1 penalty (lasso) and l_2 penalty (ridge) [14]. It has advantages over the traditional l_1 penalty because the elastic net penalty function can handle highly correlated groups of variables. The elastic net penalty function is given by the following formula:

$$P_{\lambda 1}(|\theta_j|)+P_{\lambda 2}(\theta^2)=\lambda_1|\theta_j|+\lambda_2\theta^2 \quad (4)$$

4. Penalty parameter selection:

The penalty parameter is also known as the tuning parameter and is denoted by λ. The penalty parameter plays a crucial role in the process of selecting significant variables and also controls the degree of shrinkage of the estimator. Therefore, it is important to determine it accurately and appropriately. When analyzing data in the scientific field, the value of the penalty parameter is unknown, so the researcher Chakrabarti & Ghosh (2011) used the BIC criterion to determine the optimal penalty parameter lambda. This criterion was used because it requires less computational effort and is calculated using the following formula [15], [16]:

$$BIC(\lambda)=\ln\hat{\sigma}+df(\lambda)\frac{\ln(n)}{n} \quad (5)$$

Where $\hat{\sigma}$ represents the estimated standard error of the random error and is calculated according to the following formula:

$$\hat{\sigma}=\sqrt{\frac{1}{n-d}\sum_{i=1}^{n}\left[y_i-\hat{g}\left(X^T\theta\right)\right]^2} \quad (6)$$

Where:

- D: Represents the dimension of θ,

- $\hat{g}(X^T\hat{\theta})$: Represents the estimated value of the link function,
- $df(\lambda)$: Represents the model's degrees of freedom and is determined by the number of non-zero estimated parameters in $\hat{\theta}$,
- n: Sample size,
- λ: The penalty parameter, and the results of the optimal penalty parameter are known through $\hat{\lambda}_{BIC}$.

5. Kernel density function selection:

The choice of a kernel function does not matter much in terms of obtaining a good approximation to the correct density function, as its choice affects the shape of the estimate density. The researcher [16] points out an important point regarding the selection of kernel functions and their inclusion. Any reasonable choice of a kernel function produces reasonable results; thus, the choice itself is not of paramount importance. The kernel function is selected to fulfill the following conditions:

1- $\int_{-\infty}^{\infty} K(u)du=1$ is PDF

2- $K(u)=K(-u)$ is symmetric

3- $\int_{-\infty}^{\infty} uK(u)du=0$

4- $\int_{-\infty}^{\infty} u^2K(u)du=\mu_2(K)\neq 0$ moments of kernel

5- $K(u)\geq 0$

6. Bandwidth parameter selection:

The bandwidth parameter, denoted by h, controls the level of smoothing in the estimation [17]. Selecting an appropriate value for the bandwidth parameter is crucial for curve fitting, as this parameter plays a fundamental role in the performance of kernel estimation. The choice of the bandwidth parameter requires balancing bias and variance, which can be controlled by the mean squared error (MSE). If the bandwidth parameter is too small, the bias will be low but the variance will be high, resulting in an under smoothed density estimate. Conversely, if the bandwidth parameter is too large, the bias will be high but the variance will be low, leading to an over smoothed density estimate that distorts the true shape of the density function [12].

Since the basic idea of choosing the bandwidth is to minimize the MSE and in this research, the cross-validation method will be used for the ease of its calculation and its structure is applicable to any regression model and it is calculated by constructing a C.V function by excluding one observation at a time in the following way:

$$CV(h)=n^{-1}\sum_{i=1}^{n}\left[y_i-\hat{g}_{-i}(X^T\hat{\theta})\right]^2 \quad (7)$$

Then the estimation of the link function $\hat{g}_{-i}(X^T\hat{\theta})$ is calculated for all observations and in each time one observation is excluded and then the initial value (bandwidth) corresponding to the smallest C.V is chosen [18].

7. Semiparametric estimation methods for SIM:

9.1 MAVE-LASSO:

[19] introduced the "Minimum Average Variance Estimation (MAVE)" method for semiparametric models, particularly the SIM. This flexible and easy-to-implement method offers a simple computational approach through local linear approximation, transforming complex nonlinear optimization problems into linear minimization problems. [16] proposed a MAVE estimation method with a Lasso penalty for the SIM to simultaneously estimate and select variables.

All Lasso estimation solutions rely on the soft-thresholding rule when X is orthonormal, meaning: X'X = I (where I is the identity matrix)

$$\hat{\theta}_{LASSO}=sign\left(\hat{\theta}_{j_{ols}}\right)\left(\hat{\theta}_{j_{ols}}-\lambda/2\right)+ \quad j=1,2,\dots,p$$

$$\hat{\theta}=\begin{cases}\hat{\theta}_{j_{OLS}}-\lambda\backslash 2 & \text{if} \quad \hat{\theta}_{j_{OLS}}>\lambda\backslash 2\\ 0 & \text{if} \quad |\hat{\theta}_{j_{OLS}}|\leq\lambda\backslash 2\\ \hat{\theta}_{j_{OLS}}+\lambda\backslash 2 & \text{if} \quad \hat{\theta}_{j_{OLS}}<\lambda\backslash 2\end{cases} \tag{8}$$

Therefore, the lasso idea can be applied to SSIM to obtain both parameter vector estimates and the link function simultaneously through the following formula:

$$Q(\hat{g},\hat{\theta})=\arg\sum_{j=1}^{n}\sum_{i=1}^{n}\left(y_i-g\left(X_i^T\theta\right)\right)^2+\lambda\sum_{k=1}^{p}|\theta_{k_{LASSO}}| \tag{9}$$

When we have a large number of variables (P), this causes the problem of the curse of dimensionality, where the data becomes sparse. Consequently, the process of building the model becomes more complex. To overcome this problem, the MAVE method can be used in conjunction with the LASSO penalty function to estimate a semi-parametric single index model (SSIM). This method was proposed by [16], by estimating the parameter vector using the following approach.

$$\hat{\theta}_{MAVE\text{-}LASSO}=\arg\sum_{j=1}^{n}\sum_{i=1}^{n}\left(y_i-a_j-b_jX_{ij}^T\theta\right)^2.w_{ij}+\lambda\sum_{k=1}^{p}|\theta_{k_{LASSO}}| \tag{10}$$

$$\theta:||\theta||=1 a_j,b_j=1,2,\ldots,n$$

Where:

W_{ij}: weight (Kernel) and it is a function of the distance between Xi and Xj

$|\beta_k| = |\beta_1| + ... + |\beta_p|$: is the L_1 norm of β

And the calculation of the aforementioned minimization problem can be solved by minimizing two subproblems:

According to the (LASSO - MAVE) algorithm for SSIM. [4]

Step (1):

Suppose an initial estimate of the parameter θ is obtained using the OLS method or any arbitrary vector of dimension p.

Step (2):

We set $\hat{\theta}_0=\hat{\theta}$ and compute the solution vector (a_j,b_j) according to the following equation:

$$\begin{pmatrix}\hat{a}_j\\ \hat{b}_j\end{pmatrix}=\left\{\sum_{i,j}^{n}w_{ij}^{\hat{\theta}_0}\begin{pmatrix}1\\ X_{ij}^T\theta\end{pmatrix}\begin{pmatrix}1\\ X_{ij}^T\theta\end{pmatrix}\right\}^{-1}.\sum_{i,j}^{n}w_{ij}^{\hat{\theta}_0}\begin{pmatrix}1\\ X_{ij}^T\theta\end{pmatrix}y_i \tag{11}$$

$$\text{Where: } w_{ij}^{\hat{\theta}_0}=k_h\left(X_{ij}^T\hat{\theta}_0\right)=\frac{k((X_j^T\theta_0-x_j^T\theta_0)h)}{\sum_{i=1}^{n}k((X_i^T\theta_0-x_j^T\theta_0)h)} \tag{12}$$

Step (3):

$\hat{a}$ and $\hat{b}$ are determined and the parameter vector θ is estimated according to the following formula:

$$\hat{\theta}_{MAVE\text{-}LASSO}=\arg\sum_{j=1}^{n}\sum_{i=1}^{n}\left(y_i-\hat{a}_j-\hat{b}_jX_{ij}^T\theta\right)^2.w_{ij}+\lambda\sum_{k=1}^{p}|\theta_k| \tag{13}$$

Step (4):

The iterative procedure of steps 1 and 2 is repeated using $\hat{\theta}_0=\text{sign}(\hat{\theta}_1)\frac{\hat{\theta}}{||\hat{\theta}||}$ until convergence to obtain the MAVE-LASSO estimator. The final link function g(.) is estimated as $\hat{a}_j=\hat{g}(u,\hat{\theta}_{MAVE\text{-}LASSO})$.

9.2 MAVE-Elastic Net:

[8] showed that the Lasso-penalty function lacks the Oracle property and these properties are represented by consistency in the choice of the variable, that is, the estimation of the parameters is exactly equal to zero,

and this means that the probability of excluding the non-significant variable tends to one when n→∞ and this property is called sparsity, and the second property is that non-zero parameters are efficiently estimated when the correct model is known and this property is approximated to optimization. The researchers also found that the lasso method has a bias in estimating non-zero coefficients and also showed that it does not have the Oracle property. In addition, Lasso does not deal with variables that are interconnected in groups, which prompted the researcher [20] to propose the elastic net method, where the Ridge penalty function was added to the Lasso penalty function to solve to deal with variables that are interconnected in groups.
According to the (EN - MAVE) algorithm for SSIM.

Step (1):

Suppose an initial estimate of the parameter θ is obtained using the OLS method or any arbitrary vector of dimension p.

Step (2):

We set $\hat{\theta}_0=\hat{\theta}$ and compute the solution vector \\\(a_j,b_j\\\) according to the following equation:

$$\begin{pmatrix}\hat{a}_j\\ \hat{b}_j\end{pmatrix}=\left\{\sum_{i,j}^{n} w_{ij}^{\hat{\theta}_0}\begin{pmatrix}1\\ X_{ij}^{T}\hat{\theta}\end{pmatrix}\begin{pmatrix}1\\ X_{ij}^{T}\hat{\theta}\end{pmatrix}\right\}^{-1}.\sum_{i,j}^{n} w_{ij}^{\hat{\theta}_0}\begin{pmatrix}1\\ X_{ij}^{T}\theta\end{pmatrix} y_i \quad (11)$$

$$\text{Where: } w_{ij}^{\hat{\theta}_0}=k_h\left(X_{ij}^{T}\hat{\theta}_0\right)=\frac{k((X_j^{T}\hat{\theta}_0-x_j^{T}\hat{\theta}_0)h)}{\sum_{i=1}^{n} k((X_i^{T}\hat{\theta}_0-x_j^{T}\hat{\theta}_0)h)} \quad (12)$$

Step (3):

$\hat{a}$ and $\hat{b}$ are determined and the parameter vector θ is estimated according to the following formula:

$$\hat{\theta}_{\text{MAVE-LASSO}}=\arg\sum_{j=1}^{n}\sum_{i=1}^{n}\left(y_i-\hat{a}_j-\hat{b}_j X_{ij}^{T}\theta\right)^2.w_{ij}+\lambda\sum_{k=1}^{p}|\theta_k|+\lambda\sum_{k=1}^{p}\theta_k^2 \quad (13)$$

Step (4):

The iterative procedure of steps 1 and 2 is repeated using $\hat{\theta}_0=\text{sign}\left(\hat{\theta}_1\right)\frac{\theta}{||\theta||}$ until convergence to obtain the EN-LASSO estimator. The final link function g(.) is estimated as $\hat{a}_j=\hat{g}(u,\hat{\theta}_{\text{MAVE-EN}})$.

10- A simulation study

In this section the simulation was used for the purpose of comparing the methods used in this study. We used the criteria (AMSE) and (AMAE) for the purpose of comparing the method (MAVE-LASSO) and (MAVE-EN) using the R code.

$$AMSE=\frac{1}{n}E\ \sum_{i=1}^{n}\left(Y_i-\hat{g}\left(X_i^{T}\hat{\beta}\right)\right)^2$$

$$AMAE=\frac{1}{n}E\ \sum_{i=1}^{n}|Y_i-\hat{g}\left(X_i^{T}\beta\right)|$$

Example 10.1: R = 200 datasets with size n = 25, 50, and 100 were generated from $y=1+2\left(X^{T\beta}+1\right)\log(3|X^{T}\beta|+1)+\varepsilon$ and ε are independent and are identically distributed $(i.i.d)$ from an $N(0,\sigma^2)$, and $\sigma^2=(1,5,9)$, $\beta=(0.4,\ -0.4,\ 0.8,\ -0.2,\ 0,\ 0,\ 0,\ 0)^{T}$, $X\in\mathbb{R}^8$, The multivariate normal distribution N(0, Σ) is used to simulate the design matrix X, where $\Sigma_{ij}=\rho^{|i-j|}$ for all $1\le i\le j\le p$, and $\rho=(0.1,\ 0.9)$.

Example 10.2: R = 200 datasets with size n = 25, 50, and 100 were generated from $y=5\cos(X^{T}\beta)+\exp(-X^{T}\beta)+\varepsilon$ and ε are independent and are identically distributed $(i.i.d)$ from an $N(0,\sigma^2)$, and $\sigma^2=(1,5,9)$, $\beta=(0.4,\ -0.4,\ 0.8,-0.2,0,0,0,0)^{T}$, $X\in\mathbb{R}^8$, The multivariate normal distribution N(0, Σ) is used to simulate the design matrix X, where $\Sigma_{ij}=\rho^{|i-j|}$ for all $1\le i\le j\le p$, and $\rho=(0.1,0.9)$.

Example 10.3: R = 200 datasets with size n = 25, 50, and 100 were generated from $y=\sin(X^{T}\beta)+\varepsilon$ and ε are independent and are identically distributed $(i.i.d)$ from an $N(0,\sigma^2)$, and $\sigma^2=(1,5,9)$, $\beta=$

$(0.4,\ -0.4,\ 0.8, -0.2, 0, 0, 0, 0)^T$, $X \in \mathbb{R}^8$, The multivariate normal distribution N(0, Σ) is used to simulate the design matrix X, where $\Sigma_{ij} = \rho^{|i-j|}$ for all $1 \leq i \leq j \leq p$, and $\rho = (0.1, 0.9)$.

Table (1): Results of example 10.1

ρ	n	Methods	AMSE			AMAE		
			$\sigma^2 = 1$	$\sigma^2 = 5$	$\sigma^2 = 9$	$\sigma^2 = 1$	$\sigma^2 = 5$	$\sigma^2 = 9$
0.1	25	MAVE-LASSO	0.0237	0.2013	0.2262	0.1116	0.2886	0.3192
		MAVE-EN	0.0243	0.2011	0.2244	0.1118	0.2881	0.3192
	50	MAVE-LASSO	0.0169	0.1001	0.2215	0.0842	0.1825	0.2628
		MAVE-EN	0.0171	0.1002	0.1957	0.0844	0.1826	0.2488
	100	MAVE-LASSO	0.0097	0.0559	0.0928	0.0582	0.1326	0.1729
		MAVE-EN	0.0097	0.0559	0.0928	0.0582	0.1326	0.1729
0.9	25	MAVE-LASSO	0.0199	0.1315	0.2477	0.0979	0.2526	0.3508
		MAVE-EN	0.0195	0.1306	0.2584	0.0960	0.2499	0.3729
	50	MAVE-LASSO	0.0160	0.0788	0.1515	0.0893	0.1648	0.2554
		MAVE-EN	0.0159	0.0783	0.1499	0.0893	0.1623	0.2513
	100	MAVE-LASSO	0.0085	0.0479	0.0681	0.0589	0.1092	0.1057
		MAVE-EN	0.0085	0.0479	0.0683	0.0589	0.1089	0.1064

Based on the results in Table (1), by comparing the performance of MAVE-LASSO and MAVE-EN based on AMSE and AMAE values, we find that the two methods have very similar performance in most cases, with a slight edge for MAVE-EN in certain instances, especially when ρ=0.9 and for larger n values, where MAVE-EN shows slightly lower AMSE and AMAE values in several cases .Overall, MAVE-EN can be considered the better choice when aiming for the lowest values, though the difference is minor, making both methods highly effective.

Table (2): Results of example 10.2

ρ	n	Methods	AMSE			AMAE		
			$\sigma^2 = 1$	$\sigma^2 = 5$	$\sigma^2 = 9$	$\sigma^2 = 1$	$\sigma^2 = 5$	$\sigma^2 = 9$
0.1	25	MAVE-LASSO	0.0729	0.1784	0.2582	0.1671	0.2605	0.3358
		MAVE-EN	0.0713	0.1971	0.2583	0.1650	0.2776	0.3356
	50	MAVE-LASSO	0.0583	0.1211	0.1465	0.1320	0.2146	0.2304
		MAVE-EN	0.0580	0.1210	0.1467	0.1315	0.2145	0.2302
	100	MAVE-LASSO	0.0118	0.0598	0.1068	0.0653	0.1356	0.1817
		MAVE-EN	0.0117	0.0598	0.1068	0.0652	0.1357	0.1817
0.9	25	MAVE-LASSO	0.0234	0.2122	0.2227	0.1128	0.3312	0.3539
		MAVE-EN	0.0238	0.1387	0.2243	0.1136	0.2593	0.3548
	50	MAVE-LASSO	0.0144	0.1269	0.1382	0.0773	0.2327	0.2145
		MAVE-EN	0.0144	0.0812	0.1351	0.0773	0.1958	0.2067
	100	MAVE-LASSO	0.0088	0.0440	0.0901	0.0556	0.1251	0.1633
		MAVE-EN	0.0088	0.0440	0.0897	0.0557	0.1253	0.1622

Based on the results in Table (2), comparing the performance of MAVE-LASSO and MAVE-EN in terms of AMSE and AMAE values shows that the two methods perform quite similarly overall, with MAVE-EN demonstrating a slight advantage in certain cases. Specifically, MAVE-EN performs slightly better at higher n values and often shows lower AMSE and AMAE values for some combinations of σ^2 when ρ=0.9.

In summary, while both methods are highly effective, MAVE-EN can generally be considered the better choice for obtaining lower values, although the difference remains minor.

Table (2): Results of example 10.3

ρ	n	Methods	AMSE			AMAE		
			$\sigma^2=1$	$\sigma^2=5$	$\sigma^2=9$	$\sigma^2=1$	$\sigma^2=5$	$\sigma^2=9$
0.1	25	MAVE-LASSO	0.0233	0.1045	0.2899	0.0952	0.2185	0.3412
		MAVE-EN	0.0226	0.0971	0.3002	0.0958	0.2161	0.3429
	50	MAVE-LASSO	0.0158	0.0647	0.1100	0.0774	0.1464	0.1980
		MAVE-EN	0.0167	0.0671	0.1101	0.0784	0.1487	0.1981
	100	MAVE-LASSO	0.0105	0.0487	0.0869	0.0571	0.1123	0.1691
		MAVE-EN	0.0105	0.0395	0.0869	0.0571	0.1122	0.1689
0.9	25	MAVE-LASSO	0.0206	0.1277	0.3256	0.1078	0.2599	0.3927
		MAVE-EN	0.0206	0.1268	0.2327	0.1076	0.2581	0.3259
	50	MAVE-LASSO	0.0369	0.0661	0.1256	0.1090	0.1645	0.1946
		MAVE-EN	0.0136	0.0660	0.1218	0.0691	0.1642	0.1943
	100	MAVE-LASSO	0.0088	0.0407	0.0714	0.0536	0.1074	0.1198
		MAVE-EN	0.0092	0.0405	0.0713	0.0583	0.1058	0.1200

Based on the results in the Table (3), MAVE-EN generally performs better in most cases. While MAVE-LASSO shows good performance in some instances, MAVE-EN typically yields lower AMSE and AMAE values, especially when n=50 and ρ=0.9. For example, at ρ=0.9 and n=50, MAVE-EN demonstrates a noticeable reduction in both AMSE and AMAE compared to MAVE-LASSO, making MAVE-EN the better choice overall for achieving lower values.

10- Real data:

In order to verify the performance of the methods used in this research, The methods have been applied to the real data of Gestational Diabetes, where the sample of the study consists of 100 pregnant women whose age is among 20-25 collected from Diwaniyah General Hospital, for the purpose of identifying the factors influencing the future incidence of type 2 diabetes among pregnant women. The following variables were selected based on the recommendations of some specialist doctors.

1. x_1: Pregnancies: Number of times pregnant.
2. x_2: Glucose: Plasma glucose concentration a 2-hours in an oral glucose tolerance test.
3. x_3: Blood Pressure: Diastolic blood pressure (mm Hg).
4. x_4: Skin Thickness: Triceps skin fold thickness (mm).
5. x_5: Insulin: 2-Hour serum insulin (mu U/ml).
6. x_6: BMI: Body mass index (weight in kg/(height in m)^2).
7. x_7: Diabetes Pedigree Function: Diabetes pedigree function.
8. x_8: Age: Age (years).
9. y: Outcome: Class variable (0 or 1)

Table (4): Represents estimates of $\widehat{\beta}$ parameters vector based on real data.

Method	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$
MAVE-LASSO	0.2841	0.7437	0.3585	0.2333	0.1061	.	0.2243	.
MAVE-EN	0.3109	0.5211	0.2321	0.2381	0.2127	0. 0.2243	0.2323	.

Table (5): MSE and MAE values based on values for real.

Method	MSE	MAE
MAVE-LASSO	0.0017	0.0216
MAVE-EN	0.0012	0.0209

The results of the real data shown in Table (4 and 5) were reached based on the R code and obtained the estimates of the β parameters vector and the $g(X^T\beta)$ link function of the approved variant (y) and for the MAVE-LASSO and MAVE-EN methods, we noted the following:

- The MAVE-LASSO method excluded the x6 variant of IBM as non-significant and also excluded the x8 variant of age. It selected other variables as significant predictors of the target variable (y).

$$\hat{y}_i = \hat{g}_6(0.2841x_{i1} + 0.7437x_{i2} + 0.3585x_{i3} + 0.2333x_{i4} + 0.1061x_{i5} + 0.2243x_{i6}$$

- The MAVE-EN method excluded one variable, x_8, representing age, as it was deemed non-significant. It selected the other variables, considering them as having a significant effect on the dependent variable (y). Thus, the prediction model is as follows:

$$\hat{y}_i = \hat{g}_i(0.3109x_{i1} + 0.5211x_{i2} + 0.2321x_{i3} + 0.2381x_{i4} + 0.2127x_{i5} + 0.2243x_{i6} + 0.2323x_{i7}$$

Conclusions:

- The results showed that the MAVE-EN method is the best for estimating the parameter vector and the link function, as well as for selecting significant variables for gestational diabetes data, as it provides the lowest values for the MSE and MAE criteria.
- For the MAVE-LASSO and MAVE-EN methods, we observed the following:

The MAVE-LASSO method excluded the IBM variable x6 as statistically insignificant, and also excluded the age variable x8.

The MAVE-EN method excluded one variable, x8, representing age, as it was not statistically significant. It selected the other variables, considering them to have a significant effect on the dependent variable (y).

- It can be concluded that the MAVE-EN method can be used for estimation and variable selection in a single-index semiparametric model.

11- Recommendations:

In light of the theoretical aspect and based on the conclusions reached, the main recommendations are as follows:

- We recommend using the MAVE-EN method for estimating and selecting significant variables for the single-index semiparametric model due to its efficiency in handling cases of high correlation among explanatory variables and in the presence of varying variances.
- I recommend using alternative kernel functions, as well as employing various methods for calculating the bandwidth parameter and utilizing it in semiparametric estimation methods specific to single-index models.
- I recommend using all the indicators used in the real data (Gestational Diabetes) except for the age indicator, as it was excluded by the MAVE-EN method.

Reference

[1] J. L. Powell, J. H. Stock, and T. M. Stoker, 'Semiparametric Estimation of Index Coefficients', *Econometrica*, vol. 57, no. 6, p. 1403, Nov. 1989, doi: 10.2307/1913713.

[2] W. Hardle, P. Hall, and H. Ichimura, 'Optimal Smoothing in Single-Index Models', *The Annals of Statistics*, vol. 21, no. 1, Mar. 1993, doi: 10.1214/aos/1176349020.

[3] A. Alkenani and K. Yu, 'Penalized single-index quantile regression', *International Journal of Statistics and Probability*, vol. 2, no. 3, p. 12, 2013.

[4] M. Y. Hammoud, 'On the Semi-Parametric Single-Index Model', *Arab Journal of Statistical Research*, 2013.

[5] R. Bellman, 'Adaptive control processes', *Princeton, NJ: Princeton University Press.*, 1961.

[6] P. A. Naik, 'Single-index model selections', *Biometrika*, vol. 88, no. 3, pp. 821–832, Oct. 2001, doi: 10.1093/biomet/88.3.821.

[7] J. S. Racine, L. Su, A. Ullah, L. Su, and Y. Zhang, 'Variable Selection in Nonparametric and Semiparametric Regression Models', in *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, Oxford University Press, 2014. doi: 10.1093/oxfordhb/9780199857944.013.009.

[8] R. Tibshirani, 'Regression Shrinkage and Selection Via the Lasso', *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 58, no. 1, pp. 267–288, Jan. 1996, doi: 10.1111/j.2517-6161.1996.tb02080.x.

[9] H. Zou and T. Hastie, 'Regularization and variable selection via the elastic net', *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, no. 2, pp. 301–320, 2005.

[10] H. Zou, 'The Adaptive Lasso and Its Oracle Properties', *Journal of the American Statistical Association*, vol. 101, no. 476, pp. 1418–1429, Dec. 2006, doi: 10.1198/016214506000000735.

[11] X. Liu, 'Penalized Variable Selection for Semiparametric Regression Models', 2011.

[12] H. Peng and T. Huang, 'Penalized least squares for single index models', *Journal of Statistical Planning and Inference*, vol. 141, no. 4, pp. 1362–1379, Apr. 2011, doi: 10.1016/j.jspi.2010.10.003.

[13] J. Fan and R. Li, 'Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties', *Journal of the American Statistical Association*, vol. 96, no. 456, pp. 1348–1360, Dec. 2001, doi: 10.1198/016214501753382273.

[14] A. E. Hoerl and R. W. Kennard, 'Ridge Regression: Biased Estimation for Nonorthogonal Problems', *Technometrics*, vol. 12, no. 1, pp. 55–67, Feb. 1970, doi: 10.1080/00401706.1970.10488634.
[15] S. X. Chen, 'Local Linear Smoothers Using Asymmetric Kernels', *Annals of the Institute of Statistical Mathematics*, vol. 54, no. 2, pp. 312–323, 2002, doi: 10.1023/A:1022422002138.
[16] H. Wang, R. Li, and C.-L. Tsai, 'Tuning parameter selectors for the smoothly clipped absolute deviation method', *Biometrika*, vol. 94, no. 3, pp. 553–568, Aug. 2007, doi: 10.1093/biomet/asm053.
[17] O. Akkus, 'XploRe Package for the Popular Parametric and the Semiparametric Single Index Models', *Gazi University Journal of Science*, vol. 24, no. 4, pp. 753–762, 2011.
[18] J. F. Thomas, 'Simulation study for single index model', *submitted to the Department of Mathematical sciences of Clemson university, in partial fulfilliment for The requirements for The degree of Master of science in Mathematical sciences*, 2006.
[19] Y. Xia, H. Tong, W. K. Li, and L.-X. Zhu, 'An adaptive estimation of dimension reduction space', *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 64, no. 3, pp. 363–410, 2002.
[20] H. Zou and T. Hastie, 'Regularization and Variable Selection Via the Elastic Net', *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, no. 2, pp. 301–320, Apr. 2005, doi: 10.1111/j.1467-9868.2005.00503.x.