Waleed DHHAN[*], PhD
E-mail: w.dhhan@gmail.com
Assistant Professor Sohel RANA, PhD
E-mail: srana_stat@yahoo.com
TahaALSHAYBAWEE, PhD
E-mail: taha_hussein2002@yahoo.com
Professor Habshah MIDI, PhD
E-mail: habshahmidi@gmail.com
Faculty of Science and Institute for Mathematical Research,UniversityPutra
Malaysia, 43400 UPM, Serdang, Selangor, Malaysia
[*] Nawroz University (NZU), Duhok, Iraq
Department of Statistics, Al-Qadisiyah University, Al Diwaniyah, Iraq

ELASTIC NET FOR SINGLE INDEX SUPPORT VECTOR REGRESSION MODEL

Abstract: *The single index model (SIM) is a useful regression tool used to alleviate the so-called curse of dimensionality. In this paper, we propose a variable selection technique for the SIM by combining the estimation method with the Elastic Net penalized method to get sparse estimation of the index parameters. Furthermore, we propose the support vector regression (SVR) to estimate the unknown nonparametric link function due to its ability to fit the non-linear relationships and the high dimensional problems. This make the proposed work is not only for estimating the parameters and the unknown link function of the single index model, but also for selecting the important variables simultaneously.Simulations of various single index models with nonlinear relationships among variables are conducted to demonstrate the effectiveness of the proposed semi-parametric estimation and the variable selectionversus the existing fully parametric SVR method. Moreover, the proposed method is illustrated by analyzing a real data set. A data analysis is given which highlights the utility of the suggested methodology.*

Keywords**:***Elastic Net, Single-index model, High-dimensional ,Dimension reduction, Variable selection, Support vector regression.*

JEL Classification: 62J02, 62G08

1..Introduction

The linear regression model $E(y|x) = x^T\beta$is one of the most popular techniques that uses to study the relationship among the dependent variable (output) and the vector of explanatory variables (inputs). Unfortunately,the linear relationship between these variables is too limited which reduces the application of the linear regression model. So the use of linear regression model to describe the relationships in these cases is not appropriate choice. To enhance the elasticity of the model, a single index model $E(y|x) = G(x^T\beta)$was proposed with a smooth unknown link function $G(.)$(Ichimura, 1993). It is an extension of the linear regression model to deal with nonlinear relationships among variables. It is more flexible than the parametric models and at the same time keeps their good properties. The single index model is a semi parametric model and it consists of two parts, parametric and nonparametric. The parametric part β and the nonparametric part $G(.)$of the model need to be evaluated simultaneously.In order to build the single index model, two steps should be implemented: First, is to estimate the parameters, and the other is to estimate the unknown link function. It is well known that the convergence rate of the parametric estimator is much faster than the convergence rate of the nonparametric estimator (Peng & Huang, 2011). Hence, the estimation of the parameters accurately and efficiently leads to get a good estimate for the link function of the single index model. However, if the set of explanatory variables contains some irrelevant variables (noise) or includes hundreds of variables, the precision of the estimation of parameters β will deteriorate by the curse of high dimensionality. As a result, the selecting the most significant variables from a set of inputs is very important issue for the single index model.The other problem that affects the estimate of the parameters of the single index model in addition to the presence of noise variables is when the number of predictors p is greater than the sample size n (less than full rank), but the number of significant variables is typically less than n. The presence of this problem leads to the impossibility of estimating parameters without the process of selecting variables.

Recently, the single index model which is developed by Ichimura (1993),has been widely studied by many researchers and it has gained much attention due to its excellent performance to deal with high dimensional problems in standard mean regression. The single-index model achieves dimension reduction efficiently and avoids the so-called the curse of high dimensionality because, the index $x^T\beta$ aggregates the dimension of covariates x. Hence, $G(\cdot)$in a single-index model can be estimated with the same convergence rate in probability that it would have if the one-dimensional quantity $x^T\beta$ were observable (Horowitz, 2009). Furthermore, the parameters β can be estimated withthe same convergence rate ,$1/\sqrt{n}$, that is achieved in a parametric linear model. Consequently, in terms of convergence rate in probability, the single-index model is considered as accurate as a parametric model to estimate β and as precise as a one dimension nonparametric mean regression to estimate G. The ability of the single-index models for dimension reduction gives them a considerable feature over non parametric techniques in applications where x is multi-dimensional and the structure of single-index is plausible. This advantage is come from the reality that the single index model is a semi-parametric model. It is well known that the parametric model has too hard assumptions and not easy to adapt with high dimensional. In contrast, the nonparametric model features flexible, but it suffers from decrease precision when increase the covariates (Härdle et al. 2004). The combination between parametric and non parametric models is resulting in hybrid assumptions for the single-index models. These hybrid assumptions of the single-index models are weaker than the assumptions of the parametric model and stronger than those of the fully nonparametric model. In other words, the single-index models provide high accuracy and at the same time maintain flexibility of a nonparametric model (Horowitz, 2009). Therefore, the single index models minimize the risk of mis-specification relative to the parametric models and avoid some drawbacks of fully nonparametric models such as the difficulty of interpretation, and lack of capability of extrapolation. Generally, we can say, the single index models have the strengths of the parametric model with interpretability and the flexibility of the nonparametric model.

Variable selection is very important tool in function approximation and it has become the focus of much research in fields of application for which data sets with tens or hundreds or more of variables are available(Guyon, and Elisseeff, 2003). In most applications, the generalization ability can be deteriorated if redundant predictors are included.Variable selection which solves the problem of deteriorating the generalization ability, directly reduces the number of original predictors by selecting their significant subset that still retains the generalization ability compared with that of the original inputs. Penalization techniques have been proposed as a promising techniques to improve the ordinary least squares method (OLS) since it in most applications does poorly in both prediction and interpretation.Tibshirani (1996) have proposed a new and promising method called the lasso.It is a penalized least squares technique imposing an L_1-penalty on the coefficients of regression. Due to the nature of the L_1-penalty, the lasso technique does both continuous shrinkage and variable selection simultaneously. Further, the lasso is appealing for many researchers owing to its ability to introduce sparse representation. However, the lasso technique has some drawbacks (Zou and Hastie, 2005): First, in the case when p is larger than n, the lasso selects at most n predictors before it saturates, due to the nature of the convex optimization problem. This appears to be a limiting advantage for a variable selection method. Second,if there are very high pairwise correlations among a group of variables, the lasso tends to choose single variable from the group randomly without taking in consideration the importance of variables. To overcome these problems, Zou and Hastie (2005) proposed the Elastic net method to improve the ability of lasso for variable selection when p is greater than n.

The estimation of the nonparametric part of the single index model is important as the estimation of the parameters $\beta$of the fact that this part is complementary to the predictive model, so it should be estimated efficiently. In this paper, we use the support vector machine to evaluate the unknown link function $G(\cdot)$. Support vector machine (SVM) is a set of supervised learning algorithms that can be applied to solve classification and regression problems (Vapnik 1995).It is fully nonparametric approach and considered as an important new methodology in the field of neural networks and non-linear modelling (Suykens et al. 2002).An interesting feature of the SVM model is that one can obtain a sparse solution by the use of dual space , in the sense that some observations in the solution are equal to zero (Suykens et al. 2002). This advantage allows to the SVM model to alleviate the so-called the curse of high dimensionality including the problem when p is larger than n ,which is focused on it in our article. From here, the use of the SVM model to estimate the unknown link function $G(\cdot)$, provides sufficient flexibility to solve the problem in primal space as well as dual space.

In this paper, we suggest an extension of the SIM model of Ichimura et al. (1993) by considering the Elastic Net penalty method for estimation and variable selection. Further, the support vector regression tool has been proposed for estimating the unknown link function $G(\cdot)$,while the non-linear least squares has been considered to estimate the vector of parameters β(Ichimura 1993).

The rest of this article is organized as follows.Penalized SIMwith Elastic Net is introduced in section 2. In section 3, the support vector regression is proposed to estimate the unknown nonparametric link function for the Elastic net single index model (ENSI).Numerical studies are conducted under various settings to evaluate the proposed method in section 4. In section 5, a real data analysis is reported to illustrate the proposed method. Finally, the conclusions are summarized in section 6.

2..Elastic Net Single Index (ENSI)

The Elastic net penalty is proposed by Zou and Hastie (2005) for simultaneous variable selection and parameter estimation.It is an adjusted regression model that combines between two techniques, Lasso (L_1 norm) and Ridge regression (L_2 norm). The penalty of Lasso tends to select individually correlated features and discards the others whereas Ridge penalty shrinks them towards each other. Mixing between these two methods provides the ability to overcome the limitation that faced by Lasso, namely its inability select more predictors than the exist observations in the dataset. Thereupon, the use of this technique helps to get rid of redundant variables. On the other hand it achieves non-zero determinant of the predictors matrix, which allows to apply the single index model. The proposed Elastic Net Single index here is for simultaneous achieving variable selection, parameter estimation and mean regression estimation.In order to illustrate the Elastic Net Single index methodology, we consider the next Single index model.

$$y = G(x^T \beta_{ENSI}) + r_i, \qquad i = 1,2, \dots \dots \dots n \tag{1}$$

where$y$ is the response variable and, x is the p-dimensional vector of predictors,β_{ENSI} is the vector of unknown coefficients, r_i is the residual term, $G(\cdot)$ is an arbitrary smoothnonparametric link function.This method aims to find a suitable estimator for the coefficients vector β that minimizes the next objective function.

$$\arg\min_{\beta_{ENSI}} \{\|y - g(x^T \beta_{ENSI})\|^2 + \lambda_2 \|\beta_{ENSI}\|_2^2 + \lambda_1 \|\beta_{ENSI}\|_1\} \tag{2}$$

Thus, the estimates of the coefficients β_{ENSI} can be computed by using the following minimization problem:

$$\hat{\beta}_{ENSI} = \arg\min_{\beta_{ENSI}} \{\|y - g(x^T \beta_{ENSI})\|^2 + \lambda_2 \|\beta_{ENSI}\|_2^2 + \lambda_1 \|\beta_{ENSI}\|_1\} \tag{3}$$

where$g(x^T \beta_{ENSI})$is the unknown link function estimator of $G(x^T \beta_{ENSI})$,λ_1 and $\lambda_2$correspond to the Lasso and Ridge regression penalties, respectivelywhich control the amount of regularization applied to the estimation. It should be noted that if $\lambda_2 = 0$, leads the Elastic-Net estimationback to the Lasso estimation, while if$\lambda_1 = 0$, leads back to the Ridge estimation.

The main goal is to find efficient estimators for the vector$\beta_{ENSI}$and the nonparametric link function $G(\cdot)$. As β_{ENSI} is inside the link function, the challenge is to find a convenient parametric estimator for β_{ENSI}, provided that reach the $1/\sqrt{n}$ - consistency rate. When the estimator of β_{ENSI} is found with $1/\sqrt{n}$-convergence rate, then we estimate the unknown nonparametric link function $G(.)$. To estimate the coefficients vector β_{ENSI}, we can use the semi-parametric least squares (SLS) or its weighted version (WSLS) that introduced by Ichimura (1993). It is required to set $\|\beta\| = 1$, and the first component is positive $(\beta_1 > 0)$ to satisfy the identification of the single index model (Ichimura, 1993).

3..Estimation of the unknown link function $G(\cdot)$

As already mentioned, the estimation of a single index model $E(y|x) = G(x^T\hat{\beta}_{ENSI})$is carried out in two steps. First, estimating the vector of coefficients β_{ENSI}, then using the resulting index values to estimate the unknown link function $G(\cdot)$ by univariate nonparametric kernel regression technique of y on $x^T\hat{\beta}_{ENSI}$.It should be taken into account when estimating a single index model that the functional shape of the link function $G(\cdot)$is unknown. Further, since the form of $G(\cdot)$will determine the value of conditional expectation$E(y|x) = G(x^T\hat{\beta}_{ENSI})$ for a given index $x^T\hat{\beta}_{ENSI}$, estimation of the index coefficientsβ_{ENSI} will have to adjust to a specific estimating of the link function in order to yield a correct regression value (Härdle et al. 2004).Thus, in single index model both the index and the link function have to be estimatedalthough only the link function has nonparametric feature.

In this part, we suggest to use the support vector regression model to estimate the link function $G(x^T\hat{\beta}_{ENSI})$since it considered one of the most popular kernel regression methods in the machine learning community that used to solve the problems of non-linearity and the curse of dimensionality. For this purpose we can consider the next support vector regression model with the single index $(x^T\hat{\beta}_{ENSI})$.

$$g(x^T\hat{\beta}_{ENSI}) = w, \Phi(x^T\hat{\beta}_{ENSI}) + b \tag{4}$$

where$w$is the weight vector, Φ is a nonlinear function and b is the bias term. According to Vapnik (1995), the function $g(.)$ should be as flat as possible.Flatness of the function $g(.)$ can be achieved by minimizing the Euclidean norm $||w||^2$.

The coefficients w and b should be estimated by minimizing the next ε-tube loss function to optimize the generalization ability (predicted risk), (Andreou, Charalambous, & Martzoukos, 2009; Cristianini & Shawe-Taylor, 2000; Vapnik, 1995):

$$L_\varepsilon(y_i) = \begin{cases} 0 & \text{if} |y_i - g(x^T\hat{\beta}_{ENSI})| \leq \varepsilon \\ |y_i - g(x^T\hat{\beta}_{ENSI})| - \varepsilon & otherwise \end{cases} \tag{5}$$

The problem (4) is equivalent the next convex optimization problem

$$\begin{aligned} &minimize \qquad \frac{1}{2}\|w\|^2 + \mathrm{C}\sum_{i=1}^{n}(\xi_i + \xi_i^*) \\ &subject\ to \quad \begin{cases} y_i - g\left(x^T\hat{\beta}_{ENSI}\right) - b \leq \varepsilon + \xi_i \\ g\left(x^T\hat{\beta}_{ENSI}\right) + b - y_i \leq \varepsilon + \xi_i^* \\ \quad \xi_i, \xi_i^* \geq 0, i = 1,2 \ldots., n \end{cases} \end{aligned} \tag{6}$$

where the parameter C determines the trade-off between the flatness and the number of deviations larger than ε-tube that are tolerated and ξ_i, ξ_i^* are positive slack variables used to measure deviations of the training vectors outside the ε-tube.

The dual formulation is used to solve the optimisation problem in (6). Some of of partial derivatives have been applied to get the final Elastic Net Single Index Support Vector Regression model (ENSI-SVR).

$$g\left(x^T\hat{\beta}_{ENSI}\right) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)k((x^T\hat{\beta}_{ENSI})_i.(x^T\hat{\beta}_{ENSI})) + \quad b \tag{8}$$

where$k(.,.)$is the kernel function thatused to transform the non-linear relationships in the input space to be in a linear form in the high dimensional feature space.

The estimation procedure of the proposed method ENSI-SVR can be summarized by the next algorithm.

Elastic Net Single Index Support Vector Regression Model
• Estimate β_{ENSI} by $\hat{\beta}_{ENSI}$ • Compute index values $x^T\hat{\beta}_{ENSI}$ • Estimate the link function $G(\cdot)$ by using nonparametricmethod (support vector regression model) for the regression of y on$x^T\hat{\beta}_{ENSI}$

4. Simulations examples

In this section, three simulation examples are presented to compare the proposed method , the ENSI-SVR and the standard SVR in case of high-dimensional problem especially when p larger than$n$. The purpose of these simulations is to show the effectiveness of the proposed ENSI-SVR method over the standard SVR technique in terms of dimension reduction and prediction accuracy. The first example combines the non-linear and high-dimensionality problems (Hu et al. 2013; Wu et al. 2010). The next two examples combine the same problems that mentioned in first example in addition to the problem when p larger than n (Peng and Huang, 2011).These methods are evaluated using the mean squared error for testing data (MSE). The prediction risk (MSE) is averaged over 100 replications of random data sets. These data sets have been divided into two groups, the percentage of the training samples, which used to build the model is 70% and the percentage of the testing samples, which used to test the model is 30% . The kernel function that used to transform relationships among variables in this section is the radial basis function (RBF), and all calculations were implemented using R software.

4.1 Simulation I

In this example, we consider the followingsingle-index model with 20 predictor variables:

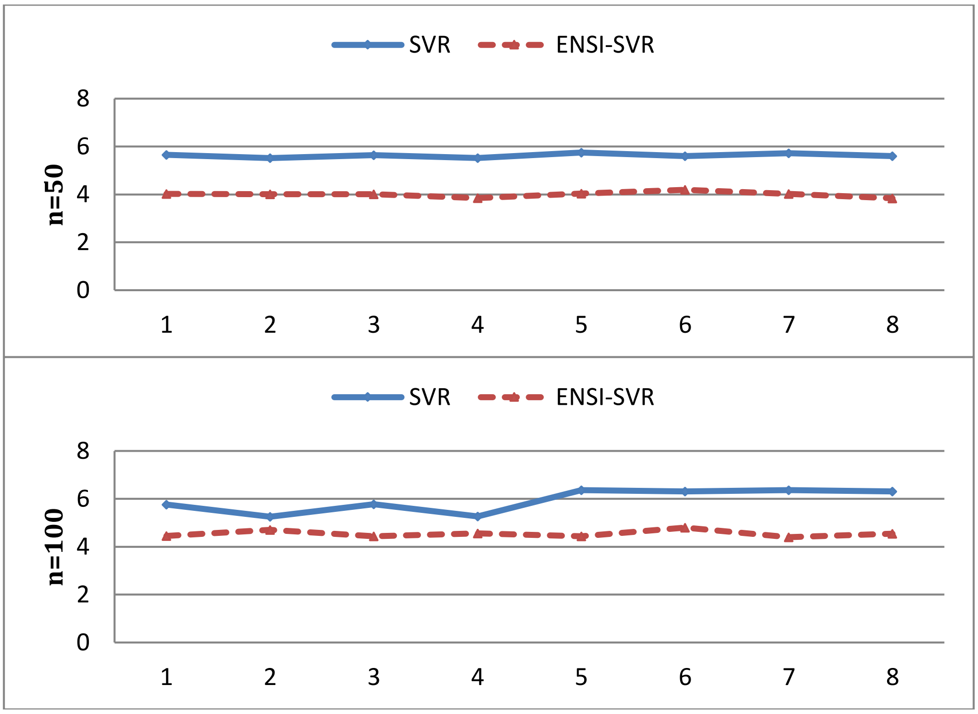
$$y = g(x^T\beta) + r_i, \qquad g(\cdot) = 5\cos(\cdot) + \exp(-\cdot^2)$$

where $\beta = \beta_0/\|\beta_0\|$ with $\beta_0 = (\beta_1, \beta_2, \dots, \beta_{20})^T = (1, 1, 1, 1, 1, 0, \dots \dots 0)$, the set of predictors $x = (x_1, x_2, \dots, x_{20})^T$are generated from uniform distribution with [0,1], while the additive error r_i is distributed from the exponential distribution with parameter [1/2].The models are built using three different sample sizes, n= (50,100,200) with 100 replications.

The results of implementation the SVR and ENSI-SVR are summarized in Table 1. It consists of the prediction risk (MSE) of these methods by using two different values of the parameters C, ε and h (small, and large).All results refer to the superiority of the proposed method over SVR method for different sample sizes and all combinations of the parameters.According to Figure 1, the proposed ENSI-SVR is achieved the minimum values of the MSE compared to the SVR method for all values of samples and parameters. It is clear that the curve of the prediction risk tends to flatness for small sample size (n=50), but with increase of sample sizes we can note some ripples are appeared. On the other hand, there is convergence between the values of the MSE when a small value of the kernel parameter h , while tends to expansion in case of high value of this parameter .However, these ripples and convergence do not affect overall advantage of the proposed method.

Table 1: Results of applying SVR and ENSI-SVR methods for 20 predictors

n	Parameters	SVR				ENSI-SVR			
		ε =0		ε =0.2		ε =0		ε =0.2	
		C=1	*C=100*	*C=1*	*C=100*	*C=1*	*C=100*	*C=1*	*C=100*
50	*h=1*	5.6617	5.5174	5.6398	5.5191	**4.0205**	**4.0144**	**4.0036**	**3.8498**
	h=5	5.7493	5.5955	5.7259	5.5990	**4.0349**	**4.1925**	**4.0275**	**3.8417**
100	*h=1*	5.7606	5.2534	5.7718	5.2670	**4.4500**	**4.7066**	**4.4362**	**4.5639**
	h=5	6.3634	6.3064	6.3646	6.3106	**4.4353**	**4.7992**	**4.4002**	**4.5484**
200	*h=1*	5.2930	4.9759	5.2872	4.9668	**4.6369**	**4.8669**	**4.6344**	**4.8026**
	h=5	6.1728	6.0852	6.1648	6.0861	**4.5953**	**4.8190**	**4.5716**	**4.7357**



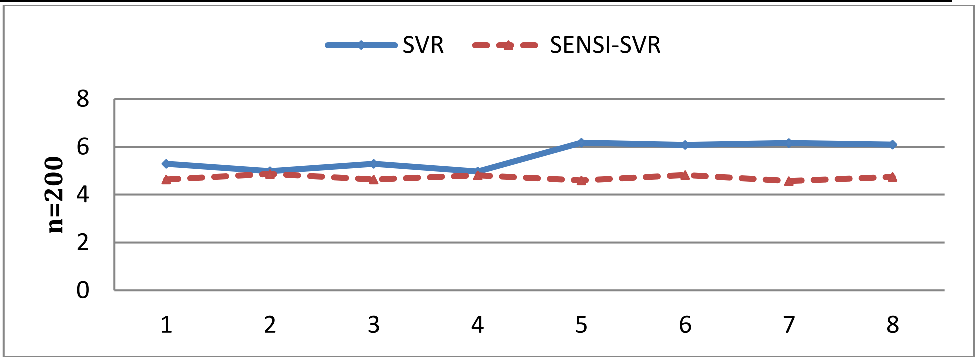


Figure 1: The mean square error of SVR and ENSI-SVR for for 20 predictors

4.2 Simulation II

In the nextexample , we simulated 100 data sets consisting of 40 predictors and two sample sizes (30, 100) based the next single index model. The first sample size is less than full rank since *p* larger than *n*.

$$y = g(x^T\beta) + 0.1r_i, \qquad g(\cdot) = \sin(\cdot)$$

where $\beta = \beta_0/\|\beta_0\|$ with $\beta_0 = (\beta_1, \beta_2, \dots, \beta_{40})^T = (3, 1.5, 0, 0, 2, 0, \dots \dots 0)$, the set of predictors $x = (x_1, x_2, \dots, x_{40})^T$ and the additive error r_i are independent identically distributed N(0,1).

The values of the prediction risk of the compared methods (SVR and ENSI-SVR) are presented in Table 2. Two different values of the parameters C, ε and *h* (small, and large) have been used to find of these values (MSE).Figure 2 illustrates these results graphically for two sample sizes.According to Table and Figure 2, the proposed method is succeeded to achieve values of the MSE lower than the SVR method for all values of parameters in case of full and less than full rank. Further, percentage of the MSE of the SVR method to the proposed ENSI-SVR amounted to more than 10 times in most of cases . All this reflects the superiority of the proposed method over SVR method.

Table 2: Results of applying SVR and ENSI-SVR methods for 40 predictors

n	Parameters	SVR				ENSI-SVR			
		ε =0		ε =0.2		ε =0		ε =0.2	
		C=1	*C=100*	*C=1*	*C=100*	*C=1*	*C=100*	*C=1*	*C=100*
30	*h=1*	0.4295	0.4271	0.4239	0.4237	**0.0273**	**0.0644**	**0.0391**	**0.0382**
	h=5	0.4296	0.4272	0.4240	0.4237	**0.0596**	**0.1225**	**0.0875**	**0.0874**
100	*h=1*	0.3996	0.3986	0.3968	0.3967	**0.0135**	**0.0191**	**0.0180**	**0.0179**
	h=5	0.3996	0.3986	0.3968	0.3967	**0.0206**	**0.0333**	**0.0298**	**0.0299**

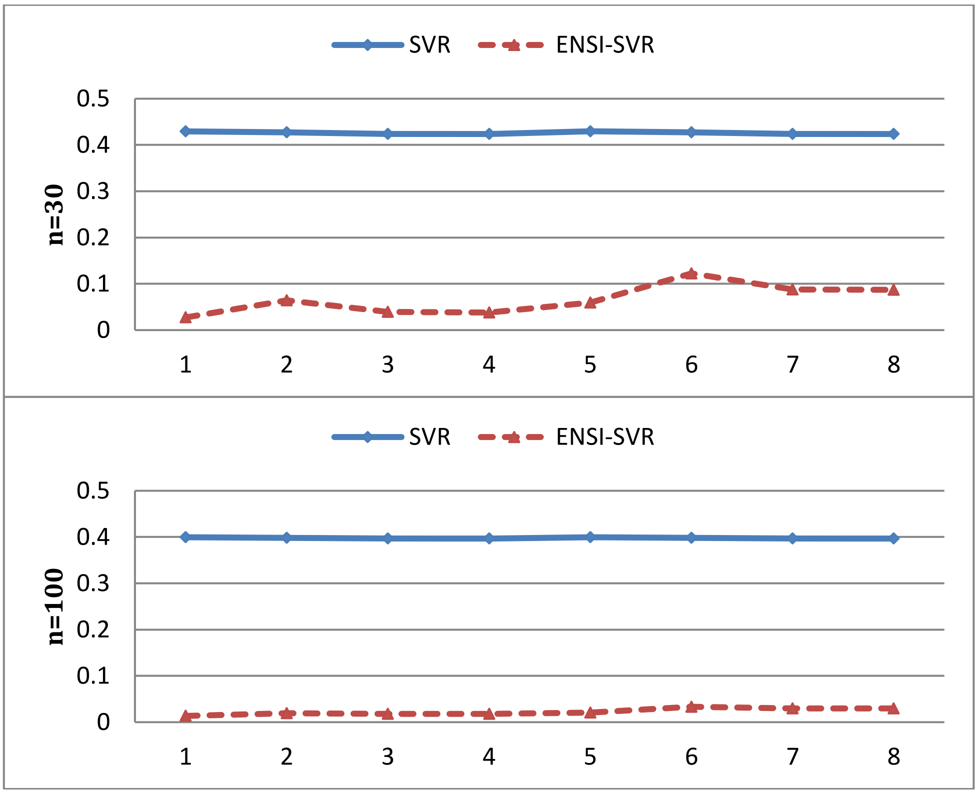


Figure 2: The mean square error of SVR and ENSI-SVR for for 40 predictors

4.3 Simulation III

This example is the same as previous example (Simulation II), except that the data sets consisting of 50 predictors and two sample sizes (40, 100).However, the number of predictors is still larger than the sample size for first data set.We let $\beta = \beta_0/\|\beta_0\|$with $\beta_0 = (\beta_1, \beta_2, \dots, \beta_{50})^T = (3, 1.5, 0, 0, 2, 0, \dots \dots 0)$, the set of predictors $x = (x_1, x_2, \dots, x_{50})^T$and the error term $r_i$are sampled from standard normal distribution.

The comparison results of ENSI-SVR and SVR are displayed in the Table and Figure 3.The results of this example do not differ significantly from the results of the previous example with the exception of the number of the predictors is 50 instead of 40 variable, which means of an increase in the number of noise variables by 10 variables.Despite the significant increase of the number of noise variables, but the method of ENSI-SVR has maintained low levels of prediction error, which indicates the importance the tool of variable selection in the model.In general, the results point out to the importance of the proposed method especially in case when p larger than n.

Table 3: Results of applying SVR and ENSI-SVR methods for 50 predictors

n	***Parameters***	**SVR**				**ENSI-SVR**			
		ε =0		**ε =0.2**		**ε =0**		**ε =0.2**	
		C=1	*C=100*	*C=1*	*C=100*	*C=1*	*C=100*	*C=1*	*C=100*
40	*h=1*	0.4907	0.4871	0.4837	0.4836	**0.0182**	**0.0313**	**0.0314**	**0.0314**
	h=5	0.4908	0.4871	0.4838	0.4837	**0.0365**	**0.0528**	**0.0660**	**0.0659**
100	*h=1*	0.3993	0.3984	0.3978	0.3977	**0.0128**	**0.0177**	**0.01664**	**0.0172**
	h=5	0.3992	0.3985	0.3978	0.3978	**0.0176**	**0.0284**	**0.0269**	**0.0267**

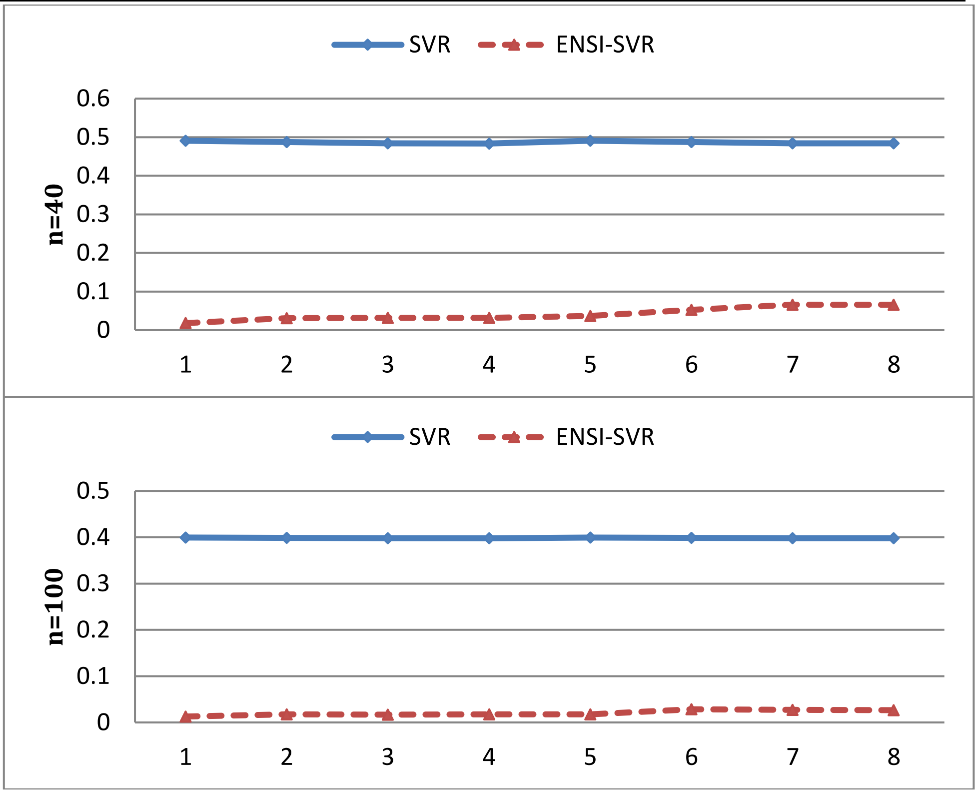


Figure 3: The mean square error of SVR and ENSI-SVR for for 50 predictors

5. Real case study

In this part, the comparison methods, ENSI-SVR and the SVR are illustrated through an analysis of the Body dimensions data (Heinz et al., 2003). The mean square error of the testing data (MSE) have been used to evaluate of these methods. The the radial basis function (RBF)Kernel function is used to transform the inputs into high dimensional feature space and all computations were performed by the use of R software.

5.1 Body dimensions data

In this part, Body dimensions data have been utilized to evaluate the proposed method (SI-SVR) under condition of high-dimensionality. The measurements were initially taken by Heinz et al. (2003). This data set consists of 21 predictors in addition to four measurements as dependent variables, such as age, gender, height, and weight on 507 individuals. In this paper, we fit the weight with all 21 explanatory variables: biacromial diameter (BiacSk), biiliac diameter (BiilSk), bitrochanteric diameter (BitrSk), chest depth among spine and sternum at the level of nipple (CheDeSk),chest diameter at nipple level (CheDiSk), elbow diameter (ElbowSk), wrist diameter (WristSk), knee diameter (KneeSk), ankle diameter (AnkleSk), shoulder girth over deltoid muscles (ShoulGi), chest girth (ChestGi), waist girth (WaistGi), navel or abdominal girth (NavelGi), hip girth at level of bitrochanteric diameter (HipGi), thigh girth below gluteal fold (ThighGi), bicep girth (BicepGi), forearm girth (ForeaGi), knee girth over patella (KneeGi), calf maximum girth (CalfGi), ankle minimum girth (AnkleGi), and wrist minimum girth (WristGi). All of these variables (inputs in addition to output) are standardized to achieve$\|\beta\| = 1$. This data has been divided to 0.75 as training samples and 0.25 as testing samples.

Tableand Figure 4, summarizes the results of applying SVR and ENSI-SVR with set of parameters (C, ε and *h*) by three values for each parameter (small, moderate and large). According to Figure 4, there are some of sharp leaps appear in the curve of MSE especially the SVR method for variousvalues of the parameters, whether small, moderate or large. While the MSE curve of the ENSI-SVR method suffers from leaps less. On the other hand, the proposed method achieved low levels of MSE are almost near zero for different values of the parameters,except when the kernel parameter is high (*h*=5), whereas the SVR method achieved of high levels of MSE for all of the parametersvalues which again reflects the superiority of the proposed ENSI-SVR method.

Table 4: Results of applying SVR and SI-SVR methods for Body dimensions data

Parameters		**SVR**			**SI-SVR**		
		C=1	*C=50*	*C=100*	*C=1*	*C=50*	*C=100*
h =0.5	*ε =0.0*	0.5833	0.5581	0.5581	**0.0870**	**0.0526**	**0.0552**
	ε =0.1	0.6017	0.5816	0.5817	**0.0854**	**0.0550**	**0.0521**
	ε =0.2	0.6272	0.6108	0.6109	**0.0923**	**0.0540**	**0.0525**
h =1	*ε =0.0*	0.9913	0.9749	0.9750	**0.1234**	**0.0598**	**0.0647**
	ε =0.1	0.9970	0.9851	0.9851	**0.1220**	**0.0757**	**0.0750**
	ε =0.2	1.0042	0.9954	0.9955	**0.1254**	**0.0824**	**0.0741**
h =5	*ε =0.0*	1.0972	1.0946	1.0946	**0.2288**	**0.2234**	**0.2770**
	ε =0.1	1.0962	1.0946	1.0947	**0.2321**	**0.2147**	**0.2486**
	ε =0.2	1.0955	1.0947	1.0948	**0.2360**	**0.1871**	**0.1927**

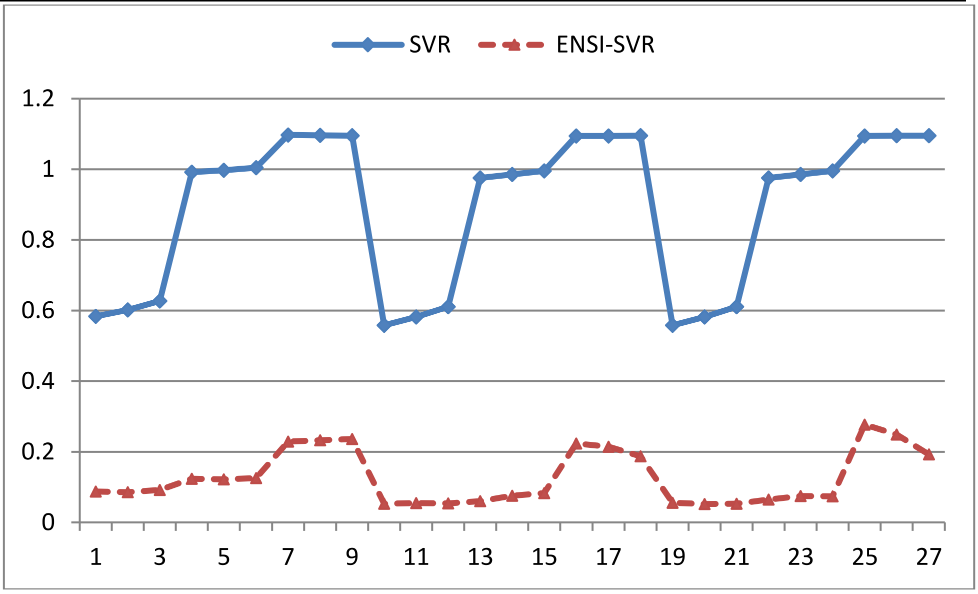


Figure 5: The mean square error of SVR and ENSI-SVR for Body dimensions data

6. Discussion and conclusion

In this paper, we have proposed the variable selection for the single-index model (SIM) in order to achieve the dimensional reduction. The SIM consists of two parts, parametric and non-parametric, which make it combines the high accuracy and the flexibility. The key to the success of our proposed method is the use of the Elastic Net penalty in obtaining the significant parameters, which helps to get rid of noise variables. Further, without the use of this penalty technique, the SIM model can not be applied when the number of predictor variables p larger than sample size n. on the other hand, the unknown link function of the SIM is estimated using the support vector regression tool. The proposed method, ENSI-SVR is compared with fully parametric SVR method using three simulation examples and a real data set. Morever, same combination of free parameters and sample sizes have been used in order to provide same conditions for comparison.Finally, the comparative results are yielded on the superiority of the proposed method, ENSI-SVR over existing SVR method to dispose of noise variables and reduce the curse of high dimensionality.

REFERENCES

[1] **Andreou, P. C., Charalambous, C. & Martzoukos, S. H. (2009),** ***European Option Pricing by Using the Support Vector Regression Approach.****Artificial neural Networks–ICANN 2009* (pp. 874-883) *Springer;*

[2]**Cristianini, N. & Shawe-Taylor, J. (2000),** ***An Introduction to Support Vector Machines and other Kernel-Based Learning Methods****.* Cambridge university press;

[3]**Guyon, I. & Elisseeff, A. (2003),** ***An Introduction to Variable and Feature Selection.*** *The Journal of Machine Learning Research, 3,* 1157-1182;

[4]**Härdle, W., Werwatz, A., Müller, M. & Sperlich, S. (2004),** ***Single Index Models*****.** *Nonparametric and semiparametric models* (pp. 167-188) *Springer;*

[5]**Horowitz, J. L. (2009),** ***Semiparametric and Nonparametric Methods in Econometrics****. Springer;*

[6]**Hu, Y., Gramacy, R. B. & Lian, H. (2013),** ***Bayesian Quantile Regression for Single-index Models*****.** *Statistics and Computing, 23*(4), 437-454;

[7]**Ichimura, H. (1993),** ***Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models****. Journal of Econometrics, 58*(1), 71-120;

[8]**Peng, H. & Huang, T. (2011),** ***Penalized Least Squares for Single Index Models.*** *Journal of Statistical Planning and Inference, 141*(4), 1362-1379;

[9]**Suykens, J. A., De Brabanter, J., Lukas, L. & Vandewalle, J. (2002),** ***Weighted Least Squares Support Vector Machines: Robustness and Sparse*** *Approximation. Neurocomputing, 48*(1), 85-105;

[10]**Tibshirani, R. (1996),** ***Regression Shrinkage and Selection via the Lasso.*** *Journal of the Royal Statistical Society.Series B (Methodological),* 267-288;

[11]**Vapnik, V. (1995),** ***The Nature of Statistical Learning Theory****. Data Mining and Knowledge Discovery*, 1-47;

[12]**Wu, T. Z., Yu, K. & Yu, Y. (2010),** ***Single-index Quantile Regression.*** *Journal of Multivariate Analysis, 101*(7), 1607-1621;

[13]**Zou, H. & Hastie, T. (2005),** ***Regularization and Variable Selection via the Elastic Net****. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67*(2), 301-320.