International Journal of Statistical Sciences *ISSN 1683-5603*
Vol. 20(2), 2020, pp 133-152

Forecasting of Sale Price of Houses Using Support Vector Regression

Sirajum Munira Khan and A. H. M. Rahmatullah Imon*

Department of Mathematical Sciences, Ball State University,
Muncie, IN47306, USA

*Correspondence should be addressed to A. H. M. Rahmatullah Imon
(Email: rimon@bsu.edu)

[Received June 30, 2020; Accepted August 25, 2020]

Abstract

Support Vector Machine (SVM) is a supervised learning method, with associated learning algorithms, it analyzes data used for classification and regression. The version of SVM used exclusively for regression is called Support Vector Regression (SVR). Kernel methods are a class of algorithms for pattern recognition which have become very popular as they are used in SVM widely. Kernel functions can be used in many applications as they provide a simple bridge from linearity to non-linearity for algorithms. The choice of an appropriate kernel is very crucial and it mostly depends on the problem at hand. Although SVR is generally recommended for high dimensional data, Dhhan *et al*. (2018) pointed out that it may work better for a big data with nonnormality and/or heteroscedasticity. In this paper, we have considered a secondary real estate data from the city Connecticut of the United States. Since the data set is large and some initial results indicate strong evidence of nonnormality, heteroscedasticity, and presence of outliers, we employ SVR technique to fit and forecast the data. We have considered a variety of kernels in our study and compare these models with the linear regression model as well. Several goodness of fit measures reveal that the SVR with Gaussian kernel most adequately fits the data. Finally, we employ a cross validation study to evaluate forecasts generated by different methods and found the SVR with Gaussian kernel generates the most accurate forecasts in comparison with the other methods considered in our study.

Keywords: Nonnormality, Heteroscedasticity, Big data, Kernel, Cross validation.

AMS Classification: 62H30.

1. Introduction

Prediction of sale price is extremely useful in real estate business, especially for the customers who are planning either buying or selling houses. There may be several factors responsible for the price of house. In such a kind of study, our primary task is to find these determining factors and also to study how they influence the sale price. Linear regression (LR) model is the most popular choice for designing this type of relationship and the ordinary least squares (OLS) technique is probably the easiest and the most commonly used technique for studying this type of relationship. But the OLS of a LR model may often suffer a huge set back in the absence of conventional regression assumptions, such as normality, homoscedasticity etc. In a big data, it is very likely that standard regression assumptions may not hold. Fortunately, a number of big data techniques are available in the literature to study this type of relationship. Among them the support vector regression (SVR) has become very popular with the statisticians in the recent years. Hong (2011) has used it to predict traffic volume in Taiwan. Rustam and Kintandani (2019) used SVR for predicting stock price in Indonesia. According to Dhhan *et al*. (2018), SVR is an ideal model when either the data is huge or have high dimensions and there is enough evidence of nonnormality and/or heteroscedasticity. We suspect, the data we are analyzing in our study may suffer from the problems just mentioned. In this paper, we fit the sale price data of Connecticut, the United States of America by a linear regression model. A variety of regression diagnostics techniques such as tests for normality, homoscedasticity, goodness of fit, and the detection of outliers are employed on the linear fit. Because of sufficient evidences of violation of the standard assumptions we fit the data by the SVR model. The SVR fit is not unique, it depends on kernel functions used in it. We considered several popular kernels and among them the Gaussian kernel emerges as the most appropriate for the data under study. Finally we report a cross validation study which is designed to evaluate the performance of the SVR model in forecasting sale prices.

2. Methodology and Data

2.1. Linear Model

We begin with a linear regression model

$$Y_i = b + w_1 X_{1i} + w_2 X_{2i} + ... + w_k X_{ki} + \varepsilon_i , \quad i = 1, 2, \ldots, n \tag{1}$$

where Y is the dependent variable, the X's are the independent variables, w's are the effects of the independent variables and ε is the error term. We can express the multiple regression model(1) in matrix notation as

$$Y = XW + \varepsilon \tag{2}$$

where

$$\boldsymbol{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} \quad \boldsymbol{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{k1} \\ 1 & x_{12} & \dots & x_{k2} \\ \dots & \dots & \dots & \dots \\ 1 & x_{1n} & \dots & x_{kn} \end{bmatrix} \quad \boldsymbol{W} = \begin{pmatrix} b \\ w_1 \\ \dots \\ w_k \end{pmatrix} \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{pmatrix}$$

Writing $x_i^T = (1\, x_{1i}\, x_{2i} \dots x_{ki})$, the ith observation in (1) can be expressed as

$$Y_i = x_i^T W, \quad i = 1, 2, \dots, n \tag{3}$$

The ordinary least squares estimate of $k + 1$ unknown parameters are obtained by minimizing the sum of squares errors $\sum_{i=1}^{n} \varepsilon_i^{\,2} = \varepsilon'\varepsilon = (Y - XW)'(Y - XW)$ yielding

$$\hat{W} = (X'X)^{-1} X'Y \tag{4}$$

We can adopt several measures to check or test the goodness of fit of models. Among them R-square, adjusted R-square, and ANOVA F are very simple and commonly used. Akaike (1974) introduced the Akaike Information Criterion, an information theoretic approach for model/variable selection, via Kullback-Leibler divergence. Another model/variable selection criterion via Kullback-Leibler divergence is the Bayesian information criterion (BIC). A model that corresponds to the lowest AIC or BIC value should be accepted.

Most of the standard results of linear regression model are based on the normality assumption and the whole inferential procedure may be subjected to error if there is a departure from this. Violation of the normality assumption may lead to the use of suboptimal estimators, invalid inferential statements and inaccurate predictions. So, for the validity of conclusions we must test the normality assumption. The simplest graphical display for checking normality in regression analysis is the normal probability plot. This method is based in the fact that if the ordered residuals are plotted against their cumulative probabilities on normal probability

paper, the resulting points should lie approximately on a straight line. In recent years the Jarque-Bera (1980) test combining the coefficient of skewness and kurtosis in one test statistic has become very popular. This test is defined as

$$\text{JB} = [n / 6]\ [S^2 + (K-3)^2 / 4] \tag{5}$$

A slight modification of the JB test was done by Imon (2003). His proposed statistic based on rescaled moments (RM) of ordinary least squares residuals is defined as

$$\text{RM} = [n c^3 / 6]\ [S^2 + c(K-3)^2 / 4] \tag{6}$$

where $c = n/(n - p)$, p is the number of independent variables in a regression model including the constant. Both the JB and the RM statistic follow a chi square distribution with 2 degrees of freedom.

One important assumption of a classical regression model is that the error term has a constant variance for all observations. But it does not often hold in reality. If the error variance changes, we call the error *heteroscedastic*. For a model with heteroscedastic errors the least square estimators get unduly large variances. If a graphical display of squared residuals against the fitted values of response shows a funnel shape, it gives an indication of heteroscedasticity. A number of analytical tests are available in the literature for the same. We would employ the White test (1980) in our study because it is very easy to understand and has a wide range of application. Here we use the squared regression residuals to run the following regression:

$$\hat{\varepsilon}_i^2 = x_i' w_i + v_i \tag{7}$$

from which we calculate R^2. The White test is based on the fact that under homoscedasticity,

$$n R^2 \sim \chi^2(p) \tag{8}$$

Here p is the number of explanatory variables in the model including the constant.

In Statistics we often observe that the values of descriptive measures are often much influenced by few extreme observations which are commonly known as outliers. Different aspects of outliers with its consequences are discussed by Barnett and Lewis (1993), and Hadi *et al*. (2009). In a regression problem, observations are judged as outliers on the basis of how unsuccessful the fitted

regression equation is in accommodating them and that is why observations corresponding to excessively large residuals are treated as outliers. In our study, we employ the standardized residuals for the detection of outliers. The *i*-th standardized residual is defined by

$$t_i = \frac{y_i - x_i^T \hat{w}}{\sigma\sqrt{1-h_{ii}}}\ , i = 1, 2, \ldots, n \tag{9}$$

where h_{ii} is the *i*-th diagonal element of the leverage matrix $H = X(X'X)^{-1}X'$. We call an observation an outlier when its corresponding standardized residual value exceeds 3 in absolute value.

2.2. Support Vector Regression Model

Support vector machine (SVM) analysis is a popular machine learning tool for classification and regression, first introduced by Vapnik (1995). A special type of SVM designed for a regression problem is referred to as support vector regression. To understand the features of SVR more clearly we will discuss LR and SVR side by side. In a linear regression, we try to obtain the 'line of best fit' by minimizing the sum of squares of the differences between the observed and fitted observation which we call errors. This fitting method is popularly known as the 'method of least squares (of errors)'. In SVR, our main objective is to fit the best possible 'hyperplane' through the observed points. Observations that lie on the margin of this hyperplane form a vector which is called a support vector. Thus a support vector regression is the method of fitting the best hyperplane through the data points using support vectors.

We have already seen in (1) – (4) that in LR we obtain the line of best fit by minimizing

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - (W.X_i + b))^2 \tag{10}$$

We consider the training data $\{(X_i, y_i)\}$, $i = 1, 2, \ldots, n$. Firstly, we start by describing the case of linear function *f*, as follows:

$$f(x) = \langle w.x \rangle + b \tag{11}$$

where $\langle .,. \rangle$ is defined as the dot product. In a support vector regression, our main objective is to find a function *f*, so that *f*(*X*) can deviate from the target *Y* by at most ε - deviation. In other words, we can estimate *f*(*X*) by minimizing the Euclidean norm $\| w^2 \|$ .The problem can be written as a convex optimization problem where we

$$\text{minimize} \frac{1}{2} \| w^2 \| \quad \text{subject to} \begin{cases} y_i - b - \langle w.x_i \rangle < \varepsilon \\ y_i - b - \langle w.x_i \rangle > -\varepsilon \end{cases} \tag{12}$$

The ε - deviation is also known as a hard margin since we do not allow any value of ε which falls outside of this band. We introduce ξ_i and ξ_i^* which are outside the ε -region of (12). In the regression literature, these points are called outliers. However, in the SVR literature these are called slack variables. Here we aim at finding weights *w* by minimizing

$$\text{minimize } \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i - \xi_i^*) \text{ subject to } \begin{cases} y_i - w.x_i - b \le \varepsilon + \xi_i \\ w.x_i + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases} \tag{13}$$

In this formulation, $C > 0$ is a constant chosen by the user that is used as a parameter that penalizes only those errors which are greater than ε . For Lagrange multipliers α_i, α_i^* , the partial derivatives of

$$\begin{aligned} L = \ & \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) - C\sum_{i=1}^{n}(\eta_i\xi_i + \eta_i^*\xi_i^*) \\ & - \sum_{i=1}^{n}\alpha_i(\varepsilon + \xi_i - y_i + \langle w, x_i \rangle + b) \\ & - \sum_{i=1}^{n}\alpha_i^*(\varepsilon + \xi_i^* + y_i - \langle w, x_i \rangle - b) \end{aligned} \tag{14}$$

to the primal variables (w, b, ξ_i, ξ_i^*) provide optimal solutions [see Smola and Schölkopf (2004)] and then (11) can be rewritten to find the weight vector *w* as

$$w = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)x_i \tag{15}$$

Thus, the regression function is represented as

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)\langle x_i, x\rangle + b \tag{16}$$

We can generalize SVR algorithm for nonlinear case by introducing a function Φ in the form

$$y = w\Phi(x) + b \tag{17}$$

Under this transformation, the equations (15) and (16) can be rewritten as

$$w = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)\Phi(x_i) \tag{18}$$

and

$$f(x) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*)k(x_i, x) + b \tag{19}$$

where $k(x_i, x) = \langle \Phi(x_i), \Phi(x) \rangle$ is the required kernel function.

Thus the SVR algorithm is summarized as follows. At first, the input pattern x_i is mapped into the feature space using the map function Φ .Then we Compute the dot products among the training patterns which are mapped previously by the map function Φ. This corresponds to evaluating kernel functions $k(x_i, x) = \langle \Phi(x_i), \Phi(x) \rangle$ .Next the dot products are inserted using the weights $v_i = \alpha_i - \alpha_i^*$. Finally, adding the parameter b yields the final prediction output for the regression.

2.3. Selection of Kernels

Kernel is a weighting function used in non-parametric statistics for obtaining a smooth estimator. The commonly used smoothing techniques used in regression such as the moving average and the running line, the points in a neighbourhood are all equally weighted. This is not good when we have a big data which may contain several clusters and instability in the neighbourhoods. We can use kernels to overcome this limitation. In the literature the size of the neighbourhood is referred to as the bandwidth. With a kernel smoother the weights for the X_i

depend on their distance from the point of interest x_0. Specifically, the weight assigned to X_i for obtaining the predicted value at x_0 is

$$w_{0i} = \frac{c_0}{\lambda} K\left(\frac{|\, x_0 - x_i \,|}{\lambda}\right) \tag{20}$$

where $K(t)$ is an even function decreasing in t, λ is the bandwidth, and c_0 is a constant that make the weights sum to 1.

A kernel smoother is frequently written in the general form

$$\hat{y}_i = \frac{\sum_1^n K\left[\left(x_i - x_j\right)/\lambda\right] y_j}{\sum_1^n K\left[\left(x_i - x_j\right)/\lambda\right]} = \sum_{j=1}^n w_{ij} y_j \tag{21}$$

For a multiple regression, we can use dot products of vectors to denote kernels as shown in (17). A number of kernels for SVR are proposed in the literature [see Dhhan *et al*. (2017)]. Table 1 presents the most commonly used kernels.

Table 1: Popular kernels for SVR

Kernel	$k(x_i, x) = \langle \Phi(x_i), \Phi(x) \rangle$
Linear	$x_i.x$
Polynomial (degree d)	$\left(x_i.x\right)^d$
Exponential	$\exp\left(-\gamma \parallel x_i - x \parallel\right)$
Gaussian Radial Basis	$\exp\left(-\gamma \parallel x_i - x \parallel^2\right)$

2.4. Cross Validation

Cross-validation is a technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. One round of cross-validation involves partitioning a sample of data into complementary subsets, performing the analysis on one subset (called the training set), and validating the analysis on the other subset (called the validation set or testing set).

Since we have a large number of data set, we prefer the data splitting technique [Montgomery *et al*. (2006)] for cross-validation of the fitted model. In data splitting technique we can take subset that contains 80 to 90% of the original data, develop a prediction equation using the selected data, and apply this equation to the samples set aside. These actual and predicted output (for the samples set aside) help us to compute the mean squared error if the response variable is quantitative or misclassification probability if the response is class variable.

We may determine the accuracy of the prediction model by computing the following accuracy measures.

Mean Absolute Error (MAE) of the prediction, which is measured as the mean absolute squared error defined as

$$MAE = \frac{1}{m}\sum_{t=1}^{m}|y_t - \hat{y}_t| \tag{22}$$

where y_i equals the actual value, $\hat{y}_i$ equals the fitted value, and m equals the number of observations in the test set.

Root Mean Squares Error (RMSE) is the square root of the mean square error and computed as

$$RMSE = \sqrt{\frac{1}{m}\sum_{t=1}^{m}(y_t - \hat{y}_t)^2} \tag{23}$$

Mean Absolute Percentage Error (MAPE) measures the accuracy of fitted time series values. It expresses accuracy as a percentage.

$$MAPE = \frac{\sum_{i=1}^{m}|(y_i - \hat{y}_i)/y_i|}{m} \times 100 \tag{24}$$

For all three measures, the smaller the value, the better the fit of the model. Use these statistics to compare the fits of the different methods.

2.5. Data

In our study we have used a ‘Real State Sales’ data taken from Connecticut Open Data using the link **data.ct.gov**. The Office of Policy and Management maintains a listing of all real estate sales of the city Connecticut of the United States with a sales price of $2,000 or greater that occur between October 1 and September 30 of

each year. For each sale record, the file includes: town, property address, date of sale, property type (residential, apartment, commercial, industrial or vacant land), sales price, and property assessment. This data base provides a listing of all real estate sales in the years between 2001 and 2017. This data set contains six variables and 509,548 observations.

3. Fitting of Linear and SVR Models

3.1. Fitting of Sales Data by Linear Regression Model

At first we fit a linear regression model for the sales amount data. The response variable 'Sales Amount' is a quantitative variable. Among the explanatory variables, 'Assessed Value', 'List Year', and 'Year of Sale' are quantitative variables. The variables 'Property Status', and 'Residential Status' are qualitative variables. In property status, we have two categories, Residential and Others. For this reason, we have used one dummy variable defined as

$$P = \begin{cases} 1 & \text{Residential} \\ 0 & \text{Otherwise} \end{cases}$$

In residential status, we have two categories, Single Family and Others and for this reason we have used one dummy variable defined as

$$R = \begin{cases} 1 & \text{Single Family} \\ 0 & \text{Otherwise} \end{cases}$$

We fit the model by using the method of least squares for this linear regression model and the fitted equation is

$$\text{Sales Amount} = 12800 + 1.243\ \text{Assessed Value} - 6.350\ \text{Year}$$
$$- 26.01\ \text{Property Status} + 24.57\ \text{Residential Status}$$

Table 2: ANOVA table for sales data with linear regression model

Source	SS	DF	MS	*F*	*p*-value
Regression	4495460207	4	1123865052	163924	0.000000
Error	3493311987	509543	6856		
Total	7988772194	509547			

The regression ANOVA table as given in Table 2 shows that F value for regression is highly significant (p-value is 0.000) which confirms that the overall regression is significant. From the results presented here we obtain the value of R^2 for this fit as 0.5627 which is satisfactory for a sample of size 509548. The value of the adjusted R^2 for this model is also 0.5627. The AIC value for this model is 4500764 and the BIC is 4500820.

Table 3: Linear regression coefficients for sales data

Variables	Coefficients	St. Errors	t-value	p-value
Constant	12800	48.6766	262.960	**0.00000**
Assessed Value	1.243	0.0016	793.698	**0.00000**
Year	–6.350	0.0239	–266.045	**0.00000**
Property Status (1)	–26.01	9.3159	–2.792	**0.00524**
Res Status (1)	24.57	0.4033	60.925	**0.00000**

The table of the coefficients (Table 3) show that all of the explanatory variables considered in this fit are highly significant. Assessed value of a house has a positive impact on the sale amount of the house. We also observe that the sales amount of house is decreasing over the years. The price of residential houses are significantly less than houses used in commercial activities. The price of single family residences are significantly higher than other types of residences.

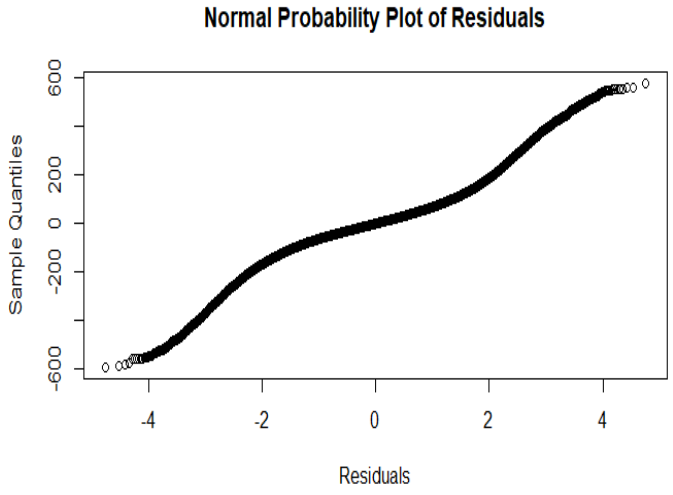

Figure 1: Normal probability plot of residuals

However, the normal probability plot as shown in Figure 1 indicates non-normal behavior of errors. But for an analytical confirmation we perform the Jarque-Bera, and rescaled moment test of residuals. According to R output the value of the Jarque-Bera and RM statistics are

$$\text{JB} = 326370 \qquad \text{and} \qquad \text{RM} = 326371$$

At a 5% level, the cut-off value for a Chi-square (2) is 5.99. Hence there exists a very strong evidence against the normality of the errors and that supports our justification of fitting this model by the SVR.

In addition to normality test, we test error for possible heteroscedasticity. Figure 2 presents a squared-residual versus fits plot for the sales data Instead of a scatter plot, it clearly shows a funnel shape. The variances of the houses with sale prices below 3K get increase with the lower price of houses.

On the other hand, the variances of the houses with sale prices above 3K get increase with the higher price of houses. The value of the White test statistic is

$$W = 45732$$

At a 5% level, the calculated value of Chi-square (5) is 11.07, so we must reject the null hypothesis of homoscedasticity and acknowledge the fact that the errors are heteroscedastic. It strengthens our argument of applying the SVR model for this data.

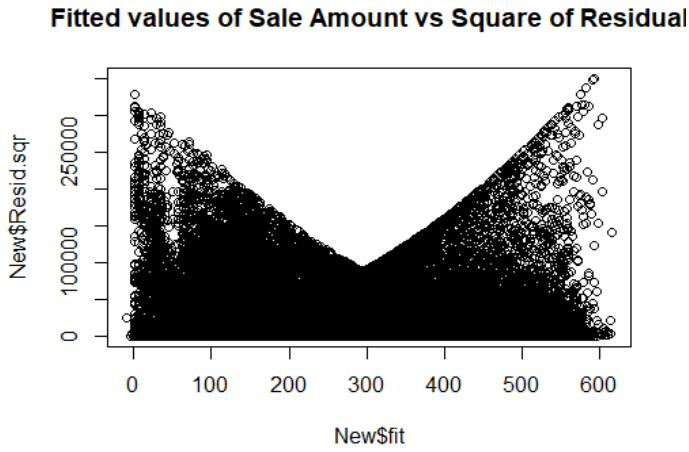

Figure 2: Squared-residual versus fits plot for the sales data

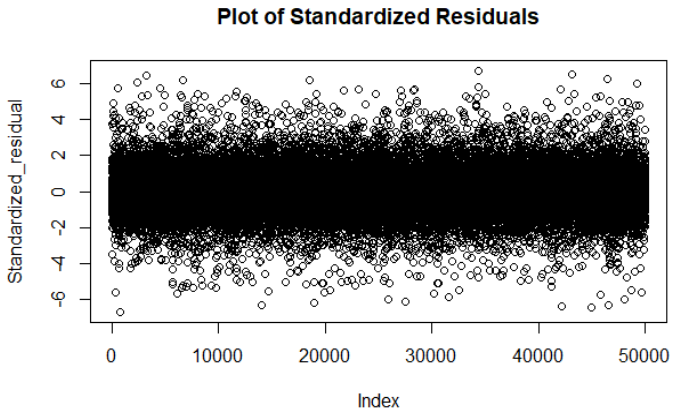

Figure 3: Index plot of standardized residuals

Finally, we employ an outlier detection test to identify possible outliers in this data. We use the standardized residuals to identify outliers. We observe from Figure 3 that a huge number of standardized residuals fall outside the cut-off range ± 3. Hence there is clear evidence of the existence of outliers in this data.

3.2. Fitting of Sales Data by the SVR Model

The results in the previous section show that when a linear regression model is employed to the sales data, there are strong evidences of violation of standard assumptions, such as nonormality, heteroscedasticity, and existence of outliers. According to Dhhan (2017) this is an ideal situation to fit the data by support vector regression models. We consider four different types of kernels for SVR, they are linear, Gaussian radial basis, polynomial and sigmoid. But the results of SVR with polynomial and sigmoid kernels are very poor and hence are omitted for brevity. At first we report results of SVR with linear kernel. The results are presented in Tables 4 and 5.

Table 4: ANOVA table for sales data with SVR linear kernel

Source	SS	DF	MS	*F*	*p*-value
Regression	3476794406	4	869198602	98159.6	0.000000
Error	4511977788	509543	8855		
Total	7988772194	509547			

Table 5: SVR linear kernel coefficients for sales data

Variables	Coefficients	St. Errors	*t*-value	*p*-value
Constant	12966	63.7372	203.429	**0.00000**
Assessed Value	1.156	0.0021	559.400	**0.00000**
Year	–6.430	0.0313	–205.722	**0.00000**
Property Status (1)	–6.000	11.9985	–0.500	0.30854
Res Status (1)	6.000	0.5335	11.246	**0.00000**

We observe from Table 4 that the SVR with linear kernel produces an overall significant fit, however, its ANOVA F is lower than that of a linear regression model. The value of R^2 for this fit is 0.4352 which is not quite satisfactory and even less than that of linear regression model (0.5627). For this model, the AIC is 4631150 and the BIC is 4631166. Both of them are less than their corresponding linear regression values 4336371 and 4336387 respectively. This result confirms that SVR with linear kernel produces a fit even worse than a linear regression model. But the most interesting finding of this model is one of the explanatory variables 'Property Status' which was significant in a linear regression model became insignificant in SVR with linear kernel as reported in Table 5. The rest of the coefficients agree with the linear regression coefficients.

Now we report results of SVR with Gaussian kernel in Tables 6 and 7. The SVR with Gaussian kernel output shows that the value of R^2 for this fit is 0.6833 which is quite satisfactory and much better than the linear regression counterpart (0.5627). For this model, the AIC is 4336371 and the BIC is 4336387. Both of them are less than those values produced by the LR and SVR linear kernels. The ANOVA F is highly significant and higher than the values yielded by the two other methods.

Table 6: ANOVA table for sales data with SVR Gaussian kernel

Source	SS	DF	MS	*F*	*p*-value
Regression	5458760794	4	1364690199	274848	0.000000
Error	2530011400	509543	4965		
Total	7988772194	509547			

Table 7: SVR Gaussian kernel coefficients for sales data

Variables	Coefficients	St. Errors	*t*-value	*p*-value
Constant	12990	49.3489	263.228	**0.00000**
Assessed Value	1.232	0.0016	779.150	**0.00000**
Year	−6.396	0.0242	−264.477	**0.00000**
Property Status (1)	−115.6	9.2899	−12.444	**0.00000**
Res Status (1)	25.27	0.4131	61.177	**0.00000**

All the four explanatory variables considered in our study have significant impacts on the fit. The explanatory variable 'Property Status' emerges as an insignificant predictor by the SVR with linear kernel. But for the SVR Gaussian kernel 'Property Status' not only shows a significant impact on the response its corresponding *p*-value is much less than the value produced by the linear regression model.

Now we offer a comparison to evaluate how SVR models fit the sales data. We consider SVR with linear and Gaussian kernels and compare them with the linear regression and transformed linear regression models.

Table 8: Measures of adequacy of different fits for sales data

Models	MSE	Adj R^2	ANOVA *F*	AIC	BIC
LR	6856	0.563	163924	4500764	4500820
TLR	6739	0.568	166192	4481060	4481094
SVR (Linear)	8855	0.435	98160	4631150	4631166
SVR (Gaussian)	**4965**	**0.683**	**274848**	**4336371**	**4336387**

Results presented in Table 8 clearly shows the merit of using the SVR with Gaussian kernel to fit the data. It outperforms the other three competitors in every respect. SVR with Gaussian kernel produces the maximum adjusted *R*-square and ANOVA *F*, and the smallest MSE, AIC, and BIC. This method is followed by the transformed linear regression model and the ordinary linear regression method. The performance of the SVR with linear kernel is the least in this comparative study.

3. Forecasting with Linear and SVR Models

In this section we report a cross validation study which is designed to evaluate forecasts of sales amounts generated by different models. For both linear regression and SVR, we have used roughly 80% observations as a training set and the remaining 20% as a test set. At first, we fit a linear regression model for the sales amount data. But this time we used roughly the first 80% of the original data. We made a forecast for the last 20% observations using the same fit.

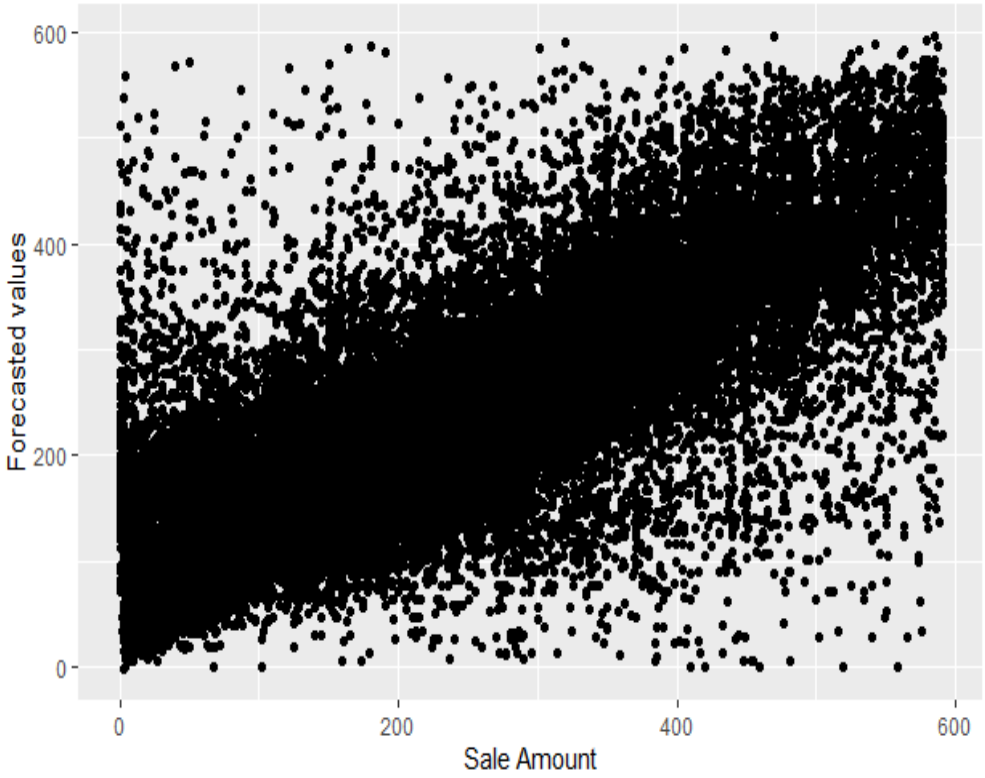

Figure 4: Scatter plot of linear regression forecasts vs original sales

A scatter plot of the forecasted sales amount are plotted against the original sales amount in Figure 4. We observe a linear positive relationship between observed and forecasted sales amounts. The correlation coefficient between them is 0.7475.

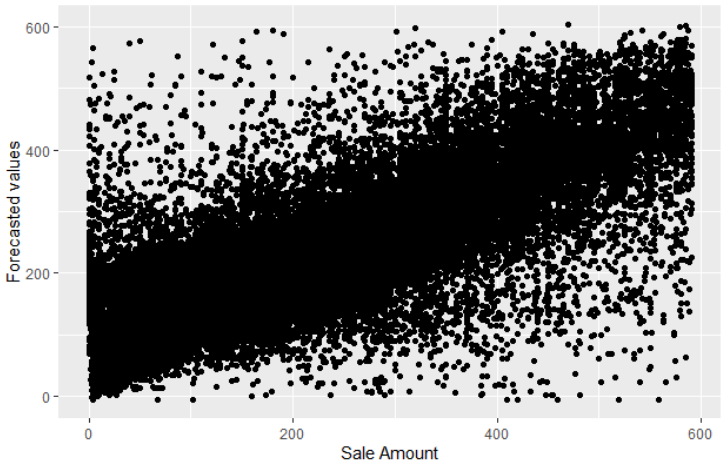

Figure 5: Scatter plot of SVR linear kernel vs original sales

Then we fit the first 80% of the data by SVR with linear kernel and use this fit to forecast the last 20% sales amount. In Figure 5, we observe a linear positive relationship between observed and forecasted sales amounts for the SVR linear kernel. The correlation coefficient between them is 0.7271.

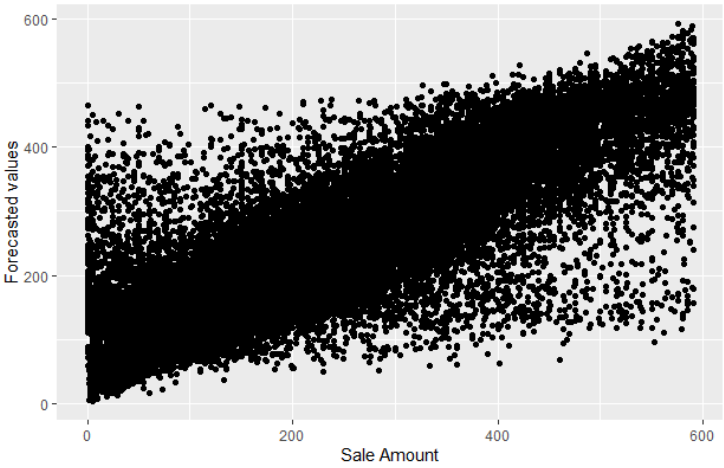

Figure 6: Scatter plot of SVR Gaussian kernel vs original sales

Finally, we fit the first 80% of the data by SVR with Gaussian kernel and generate forecasts for the next 20% based on this fit. In Figure 6, we observe a linear positive relationship between observed and forecasted sales amounts and clearly it

produces the best fit. It is also reflected in their corresponding correlation coefficient which gives the highest value 0.8266.

Table 9: Accuracy measures of different forecasts for sales data

Models	MAE	RMSE	MAPE	Correlation
Linear Regression	58.7725	82.7992	852.07	0.7475
SVR (Linear)	60.4215	94.1003	875.98	0.7271
SVR (Gaussian)	**50.0169**	**70.4642**	**725.13**	**0.8266**

Table 9 offers a numerical comparison of different accuracy measures to evaluate how different models preform in cross validation to forecast the original sales amount. These results clearly show the merit of using the SVR with Gaussian kernel in forecasting the sales data. It outperforms the other two competitors in every respect. SVR with Gaussian kernel produces forecasts that yield the minimum mean absolute error (MAE), the root mean (sum of) squares (of) errors (RMSE), and the mean absolute percentage error (MAPE) and also gives the maximum correlation coefficient with the true sales amount values. This method is followed by the linear regression method. The SVR with linear kernel performs very similar to the linear regression model although they perform the least in this comparative study.

4. Conclusions

The main objective of our research is to determine the most appropriate model and method to fit the data and generate forecasts for future sales. Conventionally the linear regression model and the ordinary least squares method are used in this type of study, but since the data size is large and there are strong evidences of nonnormality, heteroscedasticity and the presence of outliers we employed the SVR model with different kernels. We consider the four most commonly used kernels- linear, Gaussian, polynomial and sigmoid. The most important finding of our study is that the SVR with Gaussian kernel most adequately fits the data and generates forecasts. It produces the maximum adjusted *R*-square and ANOVA *F*, and the smallest MSE, AIC, and BIC while fitting the model. A cross validation study also shows that the SVR with Gaussian kernel produces the minimum MAE, RMSE, and MAPE and gives the maximum correlation coefficient with the true sales amount values. This method is followed by the linear regression, and the SVR with linear kernel methods.

Acknowledgements: The authors express their thanks and gratitude to the reviewer for giving some useful suggestions that led to considerable improvement in the methodology and presentation of the results.

Reference

[1] Akaike, H. (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19,716–723.

[2] Barnett, V. and Lewis, T. B. (1994). Outliers in statistical data, 3rd ed., Wiley, New York.

[3] Das, K. R. and Imon, A. H. M. R. (2016). A brief review of tests for normality. American Journal of Theoretical and Applied Statistics, 5**,** 5–12.

[4] Dhhan, W., Rana, S., Alshaybawee, T., and Midi, H. (2017). The single-index support vector regression model to address the problem of high dimensionality. Communications in Statistics – Simulation and Computation, 47, 2792-2799.

[5] Hadi, A. S., Imon, A. H. M. R. and Werner, M. (2009). Detection of outliers. Wiley Interdisciplinary Reviews: Computational Statistics, 1, 57–70.

[6] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). Robust statistics: The approach based on influence functio*n*, Wiley, New York.

[7] Hastie, T. J. and Tibshirani, R. J. (1990). Generalized additive models, Chapman and Hall, New York.

[8] Hong, W. (2011). Traffic flow forecasting by seasonal SVR with chaotic simulated annealing algorithm. Neurocomputing, 74, 2096-2107.

[9] Imon, A. H. M. R. (2003). Regression residuals, moments, and their use in tests for normality, Communications in Statistics—Theory and Methods, 32, 1021–1034.

[10] Jarque, C. M. and Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Economics Letters, 6, 255–259.

[11] Montgomery, D., Peck, E., and Vining, G. (2006). Introduction to linear regression analysis, 4th ed., Wiley, New York.

[12] Rustam, Z. and Kintandani, P. (2019). Application of support vector regression in Indonesian stock price prediction with feature selection using particle swarm optimization. Modelling and Simulation in Engineering, 2019, https://doi.org/10.1155/2019/8962717.

[13] Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14, 199-222.

[14] Vapnik, V. (1995). The nature of statistical learning theory, Springer, New York.

[15] White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. Econometrica, 48**,** 817–838.