Graphical exploration of robustness of support vector regression for bearing vibration acceleration

Sutawanir Darwis1*, Nusar Hajarisman1, Suliadi1, Achmad Widodo2

1Department of Statistics, Bandung Islamic University, Jl. Ranggagading No 8, 40116, Bandung, Indonesia.
2Department of Mechanical Engineering, Diponegoro University, Jl Prof Sudharto, Tembalang, Semarang 50275, Indonesia.

Abstract. Bearing defect is a major factor in failure of rotating machinery. As a defect is detected the machinery is shut down as soon as possible to avoid catastrophic damage. Therefore, it is important to monitor the condition of bearings and to predict its remaining useful life. Six run to failure datasets are provided by pronostia laboratory to build prognostics models and to estimate the remaining useful life of eleven remaining bearings acceleration vibration signals have been gathered during the experiment. Vertical and horizontal acceleration was measured by two accelerometer mounted on the bearing housing. Support vector regression was proposed as a tool to develop the bearing remaining useful life. It is known that the method is affected by outliers because there are possibilities that outliers were selected as support vectors. Some robust support vector regression have been proposed. However, empirical study of robustness of support vector regression for bearing acceleration vibration has not yet been studied. This paper proposes a graphical exploration to study the robustness of support vector regression for bearing acceleration vibration. The results shows that the graphical approach can be used as a tool for preliminary study of robustness of support vector regression for bearing vibration acceleration.

1. Introduction

Robust regression is an alternative to least squares regression when data are contaminated with outliers or influential observations. It can also be used for the purpose of detecting influential observations. Outliers occur in real data and go unnoticed without careful inspection [1]. There are two types outliers, outlying in the response and outlying in predictor, called leverage points. Both types of influential points may damage a least squares estimations. Statistical procedures have been proposed that are not affected by influential data. The methods are robust, in the sense that the results remain reliable even if a certain amount of data is contaminated. Several families of robust estimator have been developed. [2] advocate the method of least median of squares because it appeal to the intuition and is easy to use. The family of M-estimators minimizes the sum of a function of residuals. The idea of robust regression is to weight the observations differently based on behavior of these data.

The support vector regression main objective is to estimate a relation between response and predictors under assumption that the joint distribution is unknown. The model created by the method depends only on a subset support vector of training data. The objective function of the model ignores all data that are close within a threshold to the prediction. The support vector has similarity with the idea of robust regression. Several robust support vector regression have been developed, however the study of robustness of support vector regression have not been explored. This paper proposed an empirical study of robustness of support vector regression based on some datasets that have been explored for robustness. The datasets used are as follows number of phone calls, monthly interest rate, Hawkin-Bradu-Kass [1], and bearing acceleration [3]. The results shows that the support vector regression has robustness similar to the least squares of median for some datasets and for other datasets needs detailed investigations.
The paper organized as follows: after the introduction, section 2 gives a literature review on least squares time series and support vector regression, the graphical exploration of robustness of support vector regression is presented in section 3, the results and discussion are given in section 4, and section 5 concludes the paper.

2. Method
The methodology of this paper consist of literature review, exploring robustness of support vector regression using some known datasets, comparing the method with least squares, least median of squares and robust fit. Least median of squares and robust fit are known as a robust regression. It is expected that robustness of support vector regression is comparable to robust regression methods, such as least median of squares regression.

The least squares method has been acquired because of ease of computation. However, there is currently awareness posed by the manifestation of outliers. Outliers in real data go unnoticed because these days data is processed by computer. There are two types of outliers, outlying in the response and outlying in the predictor (called leverage). Both outliers may spoil a standard least squares modelling. Such dominant information is hidden because they are hidden in residual plots. Some new statistical methods have been proposed that are not influenced by outliers. These are the robust methods, the results of which remain reliable even if a certain amount of data is contaminated. In this article, the support vector regression fits a regression to the data and discovers the outliers as those having large residuals.

Least squares approach in time series do not provide tools for handling outliers. The field of robust time series has come into existence related to two types of outliers: innovation outliers and additive outliers. It appears that additive outliers occur more often than innovation outliers. This section describes robust estimation in time series data. A time series \{X_t, t=1,2,\ldots,n\} is a sequence of \(n\) consecutive observations \(X_1,\ldots,X_n\) measured at regular intervals. The observation \(Y_1,\ldots,Y_n\) are fitted by an ARMA model

\[
Y_t = m + Xt + et, \quad t = 1,\ldots,n
\]  
(1)

where \(m\) is the mean of the process, \(et\) is the error, and \(X_t\) follows an autoregressive moving average model, denoted by ARMA(p, q). The ARMA model is given by a model

\[
X_t = \sum_{i=1}^{p} i \cdot X_{t-i} + \sum_{i=1}^{q} X_{t+i} \quad t = 1,\ldots,n
\]  
(2)

where the \(et\) are normally distributed \(N(0, \sigma^2)\). The parameters are

\[
\sigma^2 = \left( \begin{array}{c} p \\ q \end{array} \right), \quad \text{dan } \sigma^2
\]

The ARMA model is difficult to deal with due to the moving part of the model. However, it is sufficient to deal with an AR model. The model reduces to autoregressive part

\[
X_t = \sum_{i=1}^{p} X_{t-i} \quad t = 1,\ldots,n
\]  
(3)

A large value of \(p\) provides an accurate approximation to ARMA model. In terms of the actually observed data \(Y_t = X_t\), the model becomes
\[ Y_t = \sum_{i=1}^{p} (Y_{t-i} + e_i) \]  

(4)

which can be simplified as

\[ Y_t = + \sum_{i=1}^{p} Y_{t-i} + e_t \]  

(5)

There are \( n - p \) sets \( \{Y_t, Y_{t-1}, \ldots, Y_{t-p}\} \) as \( t \) ranges from \( p+1 \) to \( n \) which can be used to estimate \( g, f_1, \ldots, f_p \) by means of regression model \( y = X + \)

\[
\begin{pmatrix}
Y_{p+1} \\
Y_{p+2} \\
\vdots \\
Y_n
\end{pmatrix} = 
\begin{pmatrix}
Y_p & Y_{p-1} & \cdots & Y_1 & 1 \\
Y_p & Y_{p-1} & \cdots & Y_2 & 1 \\
\vdots & \vdots & \cdots & \vdots & \vdots \\
Y_n & Y_{n-1} & \cdots & Y_{n-p} & 1
\end{pmatrix} 
\begin{pmatrix}
\phi_1 \\
\phi_2 \\
\vdots \\
\phi_p \\
\gamma
\end{pmatrix} + 
\begin{pmatrix}
e_{p+1} \\
e_{p+2} \\
\vdots \\
e_p \\
e_n
\end{pmatrix} 
\]

(6)

A simple AR(1) model \( Y_t = Y_{t-1} + e_t \) has linear model representation

\[
\begin{pmatrix}
Y_2 \\
Y_3 \\
\vdots \\
Y_n
\end{pmatrix} = 
\begin{pmatrix}
Y_1 & 1 \\
Y_2 & 1 \\
\vdots & \vdots \\
Y_{n-1} & 1
\end{pmatrix} 
\begin{pmatrix}
\phi_2 \\
\phi_3 \\
\vdots \\
\phi_n
\end{pmatrix} + 
\begin{pmatrix}
e_2 \\
e_3 \\
\vdots \\
e_n
\end{pmatrix} 
\]

(7)

and AR(2) has matrix representation

\[
\begin{pmatrix}
Y_3 \\
Y_2 \\
\vdots \\
Y_n
\end{pmatrix} = 
\begin{pmatrix}
1 & Y_2 & Y_1 \\
1 & Y_3 & Y_2 \\
\vdots & \vdots & \vdots \\
1 & Y_{n-1} & Y_{n-2}
\end{pmatrix} 
\begin{pmatrix}
\gamma \\
\phi_3 \\
\phi_2 \\
\phi_n
\end{pmatrix} + 
\begin{pmatrix}
e_3 \\
e_4 \\
\vdots \\
e_n
\end{pmatrix} 
\]

(8)

Scatter plot \( \{Y_t, Y_{t-1}\} \) can be used to explore the contaminated data, outlier, and leverage data. For uncontaminated data, the gradient and the intercept can be estimated appropriately. Two types of outliers may occur in time series: innovation outliers, and additive outliers [4]. The first type may be modeled using ARMA with error heavy tails distributed. The second type of outlier occur when the contamination is added to the response

\[ Y_t = Y_{t-1} + e_t, P(V_t 0) = \]  

(9)

For an AR(p) models, the least squares corresponds to:
\[
\min \sum_i \left( Y_i - \sum_{i=1}^p Y_{i-1} \right)^2
\]

Its generalization to \( M \)-estimator is:

\[
\min \sum_i \left( \frac{Y_i - \sum_{i=1}^p Y_{i-1}}{r} \right)
\]

where \( r \) is a symmetric function with a unique minimum at zero.

The estimate of the parameter is given by:

\[
\hat{\nu} = \left( 1 + \sum_{i=1}^n \right)^{1/n}
\]

The idea of support vector regression is mapping the data into a high dimensional space and performing regression on this space [5]. The method is adaptive to complex system and robust in dealing with nonlinear data. The application to time series modeling has shown many breakthrough.

The equation of support vector regression can be written as

\[
f(x) = \beta_0 + \beta^T x
\]

where \( x = (x_1, \ldots, x_p)^T \). The parameter estimation can be formulated as convex optimization:

\[
\min \left\| y - \beta^T x \right\|^2
\]

subject to

\[
\begin{align*}
(1 \cdots p)^T, x &= (x_1, \ldots, x_p)^T. \\
&\quad \text{subject to}
\end{align*}
\]

Introduce slack variables, the formulation described by [6]:

\[
\min \left\| y - \beta^T x \right\|^2 + C \sum_{i=1}^n \left( \omega_i + \omega_i \right)
\]

subject to

\[
\begin{align*}
(1 \cdots p)^T, x &= (x_1, \ldots, x_p)^T. \\
&\quad \text{subject to}
\end{align*}
\]

The formulation correlated with insensitive loss
\[
| x | = \begin{cases} 
0 & \leq x \leq \epsilon \\
> & \\
< & 
\end{cases} .
\]

This can be solved using QUADPROG with objective:

\[
z^T H z + f^T z
\]

where 
\[
z = \begin{bmatrix} + \end{bmatrix}^T, 
H = \text{diag}\left(\begin{bmatrix} \text{ones}(1,p) \ 0 \times \text{zeros}(1,2n) \end{bmatrix}\right), 
f = \begin{bmatrix} \text{zeros}(1,p) \ 0 \times \text{ones}(1,2n) \end{bmatrix}^T,
\]

The model can be expressed as

\[
f(x) = \bar{y} \sum_{j=1}^{n} \left( \begin{bmatrix} x_j^T \end{bmatrix} \right) + \sum_{i=1}^{n} \left( \begin{bmatrix} x_i^T \end{bmatrix} \right) (x^T x_i).
\]

This expression is a linear combination of the support vector, and is called support vector expansion. Only nonzero values of \( x_i \) are useful in predicting the regression model.

3. Result and Discussion

The monthly series (Table 1) describes the number of installation of telephone extensions from January 1966 to May 1973. It contains eighty nine observations. [7] perform seasonal differencing: \( Y_t = X_{t+12} - X_t \). The autoregressive AR(1) and AR(2) will be explored as a model for the residential telephone data. The robustness of linear model will be compared with support vector regression. Support vector regression gives the flexibility to define how much error is acceptable in the model and will find an appropriate line to fit the data. Outlier is an observation with large residual. It is an observation whose predictor is unusual. Leverage is a measure of how far a predictor deviate from its mean. High leverage points can have a great amount of effect on the estimate. Figure 1 displays the histogram of the telephone installation data. The data for November and December are extremely large. The seasonal differencing is listed in Table 3.1 and exhibited in Figure 2. The point \( (Y_{70}, Y_{71}) \) is an outlier in the vertical direction, \( (Y_{71}, Y_{72}) \) is a leverage that does not lie far from the linear shape of the majority, and \( (Y_{72}, Y_{73}) \) is a leverage point. The models are: \( \text{AR(1)} : Y_t = 1857.12 + .4052 Y_{t-1}, \) \( \text{lm} : Y_t = 608.36 + .6003 Y_{t-1}, \) \( \text{svr} : Y_t = 729.594 + .5101 Y_{t-1} \)
Table 1. Eighty nine observations of residential telephone extensions from January 1866 to May 1973, \(X_t\) and seasonal differencing \(Y_t = X_{t+12} - X_t\). The values for November and December 1972 are exceptionally large.

|       | jan   | feb   | march  | apr   | may   | june  | july   | aug   | sept  | oct   | nov   | des   |
|-------|-------|-------|--------|-------|-------|-------|--------|-------|-------|-------|-------|-------|
| 1966  | 10165 | 9279  | 10930  | 15876 | 16485 | 14075 | 14618  | 14535 | 15367 | 13936 | 12606 | 12932 |
| 1967  | 10545 | 10120 | 11877  | 15876 | 16485 | 14075 | 14618  | 14535 | 15367 | 13936 | 12606 | 12932 |
| 1968  | 12689 | 11791 | 12771  | 16952 | 21854 | 17028 | 16988  | 14943 | 14943 | 16573 | 15548 | 15838 |
| 1969  | 13335 | 12395 | 15450  | 19092 | 22301 | 18260 | 19427  | 18974 | 20180 | 18395 | 15596 | 14776 |
| 1970  | 13453 | 13086 | 14340  | 19714 | 20796 | 18183 | 17981  | 17706 | 20923 | 18380 | 17343 | 15416 |
| 1971  | 12465 | 12442 | 15448  | 21402 | 25437 | 20814 | 22066  | 21528 | 24418 | 20853 | 20673 | 18746 |
| 1972  | 15637 | 16074 | 18422  | 27326 | 32883 | 24998 | 25996  | 20923 | 20673 | 18746 | 17343 | 15416 |
| 1973  | 18115 | 15184 | 19832  | 27326 | 32883 | 24998 | 25996  | 27583 | 22068 | 75344 | 47365 | 43897 |

(ii) 12-differencing \(Y_t = X_{t+12} - X_t\), \(t = 1,\ldots,77\)

|       |       |       |       |       |       |       |       |       |       |       |       |       |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 380   | 841   | 947   | -1124 | 444   | 48    | 609   | 408   | 1206  | 2152  | 3232  | 1227  | 1227  |
| 2144  | 1671  | 894   | 2200  | 4922  | 2905  | 2211  | 3854  | 1453  | 2497  | 680   | 266   | 266   |
| 646   | 604   | 2679  | 2140  | 447   | 1232  | 2439  | 177   | 2154  | 350   | -922  | 353   | 353   |
| 118   | 691   | -1110 | 622   | -1505 | -77   | -1446 | -1268 | 743   | -15   | 1747  | 638   | 638   |
| -988  | -644  | 1108  | 1688  | 4641  | 2631  | 4085  | 3822  | 3495  | 2473  | 3330  | 3330  | 3330  |
| 3172  | 3672  | 2974  | 5924  | 7446  | 3495  | 2932  | 4468  | 3165  | 1215  | 54671 | 28619 | 28619 |
| 2478  | -890  | 1410  | 271   | 1373  | 1373  | 1373  | 1373  | 1373  | 1373  | 1373  | 1373  | 1373  |

Source: [8], [1].

Figure 1. Histogram of number of telephone installation from January 1966 to May 1973. The data for November and December 1973 are exceptionally large. These outliers represents free installation period.

Figure 2. Time plot of the seasonal differencing installation data. The observations 71 and 72 are outliers. Except for the outliers the series is stationary with mean zero.
Figure 3. Scatter plot \( (Y_{t-1}, Y_t) \). Three points \( (Y_{t0}, Y_{t1}), (Y_{t1}, Y_{t2}), (Y_{t2}, Y_{t3}) \) are unusual in x values or outliers among y values. The point \( (Y_{t0}, Y_{t1}) \) is a regression outlier, \( (Y_{t1}, Y_{t2}) \) is a leverage point, \( (Y_{t2}, Y_{t3}) \) is a leverage point.

Figure 4. Scatter plot \( (Y_{t-1}, Y_t) \), autoregressive AR(1), robust linear model, and support vector regression of residential phone data. Support vector regression lies between autoregressive and robust linear model.

Figure 5. Standardized residual associated with autoregressive, robust linear model, and support vector regression. The three residual plots confirms that observation 70 has large residual.
Figure 5 shows the result from fitting an AR(2) $Y_t = +_1 Y_{t-1} + _2 Y_{t-2} + e_t$ to the 75 triples $(Y_1, Y_2, Y_3), \ldots, (Y_{75}, Y_{76}, Y_{77})$. The outliers $Y_{71}$ and $Y_{72}$ influence four cases: $(Y_{69}, Y_{70}, Y_{71})$ is a regression outlier, $(Y_{70}, Y_{71}, Y_{72})$, $(Y_{71}, Y_{72}, Y_{73})$ and $(Y_{72}, Y_{73}, Y_{74})$ are leverage points [1].

**Figure 6.** Plot $(Y_{1}, Y_{1})$ shows that the two outliers $Y_{71}, Y_{72}$ effect four regression cases of residential phone data.

Monthly interest dataset is a time series data from January 1965 to December 1982. Figure 2 shows index plot of the observations. The plot shows a positive trend of monthly interest rate. There is no outliers in the data. The support vector regression is close to least squares.

**Figure 7.** Monthly interest rate from January 1965 to December 1982. Observation 200 considered as extreme value. There is positive trend of the data.
Figure 8. Scatter plot of $\left( Y_{t-1}, Y_t \right)$, autoregressive AR(1), robust fit and support vector regression of Monthly interest rate. The equations are AR 1(1): $Y_t = 0.1896 + 0.9769Y_{t-1}$, robust fit: $Y_t = 0.2033 + 0.9741Y_{t-1}$, support vector regression $Y_t = 0.8889 + 0.9667Y_{t-1}$.

The dataset Hawkins-Bradu-Kass [9] consist of 75 observations in three predictor and one response. The first ten observations are bad leverage data, and the next four are good leverage ($X_i$ are outlying, but the corresponding response fit the model). The $M$-estimators do not produce the expected results [1]. From least squares residual plot, the observations 11, 12, and 13 are outliers. However these are good observations. The first ten points have small standardized residual (Figure 9). Robust fit regression (Figure 10) yields similar results with least squares regression. Index plot support vector regression (Figure 11) shows similar result with least squares and robust fit.

Figure 9. Index plot related to least squares regression of Hawkins-Bradu-Kass data. The observations 11, 12, 13 are outliers, however from the generation these are good observations. The leverage points have small standardized residuals.
Figure 10. Index plot related to robust fit regression of Hawkins-Bradu-Kass data.

Figure 11. Index plot related to support vector regression of Hawkins-Bradu-Kass data

Pronostia is an experimental plan dedicated to test methods related to bearing prognostics and its main purpose is to provide real data related to bearing degradations [10]. Figure 12 shows the scatter plot \((Y_{t-1}, Y_t)\) of bearing vibration acceleration. It compares the plot of three regression methods: autoregressive \(\text{ar}(1)\), robust fit and support vector regression. The equation of autoregressive model is \(Y_t = 0.0363 + Y_{t-1}\), for robust regression is \(Y_t = 0.011 + 1.007Y_{t-1}\), and for support vector regression is \(Y_t = 0.637 + 623Y_{t-1}\). Autoregressive and robust fit have same gradient, and to random walk model. Support vector regression has different gradient coefficient compared to autoregressive and robust fit regression. Figure 3.13 shows the index plot of the root mean squares of bearing acceleration of vibration. The plot shows nonlinearity pattern of the data. Figure 14 shows scatter plot \((Y_{t-1}, Y_t)\) of root mean squares of bearing acceleration of vibration of pronostia dataset. The plot shows two outlying points which influences the support vector regression.
Figure 12. Scatter plot \( (Y(t-1), Y(t)) \) of bearing vibration acceleration from pronostia dataset. Robust fit regression is close to autoregressive model influenced by the majority of the data. Support vector regression was influenced by some outlying points.

Figure 13. Index plot of root mean squares bearing vibration acceleration. The plot shows a nonlinearity pattern.

Figure 14. Scatter plot \( (Y_{t-1}, Y_t) \) of root mean squares of bearing vibration acceleration. The shows a linear pattern except some data are outlying points. The autoregressive and robust fit not influenced by these two points, and two models close to random walk. These two points influenced significantly the support vector regression, and the gradient is significantly different from random walk.

4. Conclusion
This paper proposes a graphical approach to study the robustness of support vector regression. The method is compared with least squares of median and robust fit methods. These two methods are known as robust method. The robustness of support vector regression is studied on three well known
datasets, and on bearing acceleration dataset. For residential phone data, support vector regression lies between autoregressive and robust linear model. Support vector regression has robustness similar with robust regression. For monthly interest dataset, support vector regression is close to least squares regression. This dataset has increasing trend and contains no outliers. For Hawkins-Bradu-Kass data, the residual plot of support vector regression is similar to least squares and robust fit regression. The method failed to detect ten bad leverage points and failed to accommodate four good leverage points. The graphical exploration of robustness of support vector regression for bearing acceleration data shows different gradient compared to least squares and robust fit regression. The support vector regression can be used as a tool in robust regression. It needs more study to improve the robustness of support vector regression to attain well known robust regression methods.

Acknowledgments
This work was supported by Unisha research grant no: 138/B.04/LPPM/XI/2019. The authors thank the reviewer for valuable comments.

References
[1] Rousseeuw P and Leroy A M 1987 Robust Regression & Outlier Detection (New York: John Wiley & Sons)
[2] Rousseeuw P J 1984 Least Median of Squares Regression, J. Am. Stat. Assoc. 79 871-880
[3] Nectoux P, Gouriveau R, Medjaher K, Ramasso E, Morello B, Zerhouni N, and Varnier C 2012 Pronostia: An Experimental Platform for Bearing Accelerated Degradation Tests IEEE Int. Conf. on Prognostics and Health Manag. 1-8
[4] Fox A J 1972 Outliers in Time Series J. R. Stat. Soc. Ser. B, 34, 350-363
[5] Widodo A 2009 Application of Intelligent System for Machine Fault Diagnosis and Prognosis (Semarang: Badan Penerbit Universitas Diponegoro)
[6] Vapnik V N 1995 The Nature of Statistical Learning (New York: Springer)
[7] Brubacher S R 1974 Time Series Outlier Detection and Modelling with Interpolation (Murray Hill :Bell Laboratories Technical Memo)
[8] Martin R D, Samarov A, and Vandale W 1983 Robust Methods for ARIMA Models, in Applied Time Series Analysis of Economic Data, edited by Zellner A (Washington DC: Economic Research ER-5, Bureau of the Census)
[9] Hawkins D M, Bradu D, and Kass G V 1984 Location of Several Outliers in Multiple Regression Data Using Elemental Sets, Technometrics 26 197-208
[10] Benkedjouh T, Medjaher K, Zerhouni N, and Rechak S 2013 Remaining Useful Life Estimation Based on Nonlinear Feature Reduction and Support Vector Regression Engineering Applications of Artificial Intelligence 26 1751-1760