SindhUniv. Res. Jour. (Sci. Ser.) Vol Vol.50 (001) 159-164 (2018)

http://doi.org/10.26692/sujo/2018.1.0028

SINDH UNIVERSITY RESEARCH JOURNAL (SCIENCE SERIES)

Machinery Health Prognosis - Data Driven Approach Using Threshold Regression

H. MURTAZA, A. MANSOOR++, A. S. SOOMRO*, H. MUSHTAQ**

Engineering Sciences Department, PNEC, National University of Sciences and Technology, Islamabad, Pakistan

Received 10rd April 2017 and Revised 26th September 2017

Abstract: Machinery health data is the backbone of prognostics. Effective prognostic, from the machinery data, leads towards operational reliability, reduced machinery downtime, cost savings, secondary/catastrophic failures etc. Various methodologies have been adopted by the researchers in an effort to precisely forecast/predict machinery health. In this study, Threshold Regression Methodology has been applied to a machinery vibration data to estimate future health state of machinery. The results show that the proposed method is an effective and reliable approach for data driven prognostics.

Keywords: Outer Race Defect; Prognostics; Regime; Threshold Regression

1. INTRODUCTION

Prognosis is basically a medical term derived from Greek language pro & gnosis with a literal meaning 'forecast of the likely outcome of a situation'. Prognosis is widely applied in medical sciences and enormous advancements have been made in this field owing to the refinement in technology, experience of physicians and modern diagnostics capabilities etc. In the present technological advanced world, where the humans involvement is minimized in the industry through automated machines, robots etc; the concept of prognostics is expediously being adopted in the industry due to requirements of enhanced operational availability, cost savings, reducing downtimes and avoiding catastrophic /secondary failures.

Machinery health prognosis is broadly driven by (i) physical model of machine/component (ii) data driven models (iii) Hybrid (combination of i & ii). Suitability of each method for prognostics application has adequately been deliberated in literature (Heng, *et al.*, 2009) - (Ramasso and Saxena, 2014). Hence, the prognostics methods vary diversely on case to case basis.

Real systems exhibit patterns which are very complicated due to their randomness as well as nonlinear behavior. This aspect poses a great challenge for the development of effective prognostics methodology. To cater aforesaid challenge, the prognostic approach must be flexible enough with respect to its ease of application, user-friendliness vis-à-vis diversified application and effective prediction capability. When all these aspects are considered collectively than data driven approaches seem more practical.

Taking lead from huge usage in the fields of finance and medical, data driven techniques have become popular approaches for prognosis in engineering sector also. A comprehensive review (Lee, *et al.*, 2014) shows that most of the prognostics designs are based on data driven models; even the recognized physical models are driven by historical data of the machinery. Furthermore, it can be deduced from comparative analysis (Lee, *et al.*, 2014) that statistical tools offer an easier and reliable prognostics framework. Similar reviews for different scenarios already presented also support the above argument (Heng, *et al.*, 2009) (Kan, *et al.*, 2015).

Machinery health data is, in fact, time series data in which time is the independent variable and the sensors' values (temperature, vibration, acoustic emission, pressure, flow etc) are dependant variables and show the machinery/system degradation wrt time. This time series is used for feature extraction and for further predictive analysis. A comprehensive review spread over 25 years with regards to time series prediction is presented (Gooijer and Hyndman 2006).

The pattern of machinery health data is nonlinear and non stationary in nature when real time scenario is considered (Kan, *et al.*, 2015).. For the analysis of nonlinear and non stationary time series, by applying statistical tools, regression analysis is gaining much attention (Niu and Yang, 2009).

In order to effectively predict the nonlinear and non stationary time series, this paper discusses a nonlinear time series model, Threshold Autoregressive Model (TAR), for feature extraction and predicting the future

++Corresponding E-mail: asifmansure@gmail.com,
* Institute of Mathematics &Computer Sciences, University of Sindh, Jamshoro, Pakistan E-mail: dr_sattarsoomro@yahoo,com,
** Department of Mathematics, University of Karachi, Karachi, Pakistan

state of machinery. Validation of the proposed model is conducted using standard statistical tests vis-à-vis experimental data.

The remaining part of this paper is divided into further sections. Section 2 contains the description of the model, section 3 discusses its application on machinery health data, section 4 discusses the model stability and finally conclusion have been made in section 5.

2. MODEL AND METHOD

2.1 Model Description

A linear model for a series X_t with the parameters of the model to vary with past values of X_t, first order TAR model is given by:

$$X_t = \begin{cases} a^1 X_{t-1} + e_t^{(1)} & if\ X_{t-1} < d, \\ a^2 X_{t-1} + e_t^{(2)} & if\ X_{t-1} \geq d, \end{cases} \quad (1)$$

Where $e_t^{(1)} \& e_t^{(2)}$ are each white noise terms, $a^1\ and\ a^2$ are coefficients whereas 'd' is called 'delay parameter'. The eq. (1) can be extended for 'l' thresholds as;

$$X_t = a^i X_{t-1} + e_t^{(i)}\ \ if\ X_{t-1} \epsilon\ R^{(i)},\ i = 1,2, \ldots\ldots, l \quad (2)$$

here$R^{(1)}, \ldots\ldots\ldots, R^{(l)}$ are given subsets of real line R^1 and define a partition of R^1 into disjoint intervals (-∞, r_1], $(r_1, r_2], \ldots\ldots (r_{l-1,}\infty]$, with $R^{(1)}$ denoting the interval (-∞, r_1], and $R^{(l)}$ the interval $(r_{l-1,}\infty]$.

The eq. (2) can be approximated to general nonlinear first order model by introducing a nonlinear function'λ',

$$X_t = \lambda(X_{t-1}) + e_t^i \quad (3)$$

k^{th} order threshold model can be written as;

$$X_t = a_0^{(i)} + a_1^{(i)} X_{t-1} + \cdots + a_k^{(i)} X_{t-k} + e_t^{(i)} \quad (4)$$

Considering various sets of parameters to be determined by a single past value,X_{t-d}, eq. (4) can be written as;

$$X_t = a_0^{(j)} + \sum_{i=1}^{k} a_i^j X_{t-i} + e_t^{(j)} \quad (5)$$

$$if\ X_{t-d}\ \epsilon\ R^{(j)},\ \ j = 1,2, \ldots\ldots, l$$

In practice, various threshold regions may give rise to models of different orders. We may assume the largest order of value 'k' and set all redundant coefficients to zero. Eq. (5) then can be specified as;

$$X_t = a_0^{(j)} + \sum_{i=1}^{k_j} a_i^j X_{t-i} + e_j^{(j)} \quad (6)$$

$$if\ X_{t-d}\ \epsilon\ R^{(j)},\ \ j = 1,2, \ldots\ldots, l$$

where$R^{(j)} \subseteq\ R^1$

Where the switching between different sets of parametric values is determined by a past value of an associated process, $\{Y_t\}$, rather than a past value of $\{X_t\}$. The TAR model is of the following form,

$$X_t = a_0^{(j)} + \sum_{i=1}^{m_j} a_i^{(j)} X_{t-i} + \sum_{i=1}^{m_j} b_i^{(j)} Y_{t-i} + e_t^{(j)} \quad (7)$$

$if\ Y_{t-d}\ \epsilon\ R^{(j)},\ j = 1,2, \ldots\ldots, l$ than eq. (7) is called an open loop threshold model.

The identity of the threshold variable and the regressors will determine the type of threshold regression specification. If the threshold variable is lagged dependent then eq. (5) & (6) can be termed as Self Exciting (SE) model otherwise it will be conventional TR model. Furthermore, the model is called Self Exciting Autoregressive (SETAR), if it combines both lagged dependent threshold variable as well as autoregressive components.

2.2Estimation

Tong (Tong and Lim 1980)proposed an algorithm to estimate the parameters of threshold model based on Akaike's AIC criterion which is as follows:
Consider a single threshold model,

$$X_t = \begin{cases} a_0^{(1)} + \sum_{i=1}^{k_1} a_i^1 X_{t-i} + e_t^{(1)} if X_{t-d} \geq r \\ a_0^{(2)} + \sum_{i=1}^{k_2} a_i^2 X_{t-i} + e_t^{(2)} if X_{t-d} < r \end{cases} \quad (8)$$

Where *d*is delay parameter and *r* is threshold region.

To proceed with the final choice of the model following calculation method for AIC criterion was suggested (Tong and Lim 1980):

Step 1: Fit separate AR models for given values of *d and r*;

$$AIC\ (d,r) = AIC(\grave{k}_1) +\ AIC\ (\grave{k}_2) \quad (9)$$

where$\grave{k}_1 and \grave{k}_2$ are the values which minimize AIC criteria.

Step 2: Keeping *d* at a fixed value, choose $\grave{r}$ so that the criterion of AIC attains its minimum value;

$$AIC(d) = AIC(d, \grave{r}) \quad (10)$$

Step 3: Search for the best value of d over a range of possible values so that AIC reaches minimum value.

2.3Testing

Upon finalization of best fitted model, diagnostic tests were conducted to assess whether a fitted model shares the main characteristics of the data. In this regard, test for whiteness, normality and homo sedasticity in innovations were performed. Furthermore, to evaluate the forecasting performance, suitable test were also performed.

3. APPLICATION

Test rig is a simple arrangement driven by a 2 Hp variable speed induction motor. A customized designed shaft is then supported by two bearings. The test bearing UC-203 (deep groove ball bearing) follows the support bearings which will be subjected to various radial and axial loading conditions, under modified housing. A screw type loading mechanism is installed for static load in radial direction and for axial loading a load is mounted on shaft see **(Figure 1)**(Murtaza and Mansoor 2017).

Fig.1: Experimental Setup

Testing is done on the basis of varying loading and variable speed conditions. The concept of accelerated testing is adopted in order to perform the failure analysis with varying static load from 0N-75N and a constant axial load of 100N. The analysis is confined to the bearing only, in view of its criticality. The data collection was conducted using NI DAQ system.

The asymptotically unbiased AIC estimator is computed by utilizing eq (9) & (10)which describe the superiority of threshold variable in particular regime. Here it is apparent that threshold variable (X_{t-3}) in regime four is vital accurate predictor as shown in (**Table 1)**.

Table 1: Model Selection Criteria

Threshold Variable	AIC	Regimes
X_{t-2}	10.6	1
X_{t-1}	2.56	2
X_{t-3}	1.71	4
X_{t-4}	3.39	5
X_{t-5}	3.1	3

After the identification of threshold variable and number of regimes, we approximate univariate threshold autoregressive model with four regimes. The suitable order model equation fitted to the vibration parameter $(X_t)$for different regimes is as follows:

$$X_t = \begin{cases} -0.26 - 0.04\,X_{t-1} + 2.46\,X_{t-2} - 1.917\,X_{t-3} + e_t^{(1)} \\ \quad if\;\; X_{t-2} < -0.065 \\ 0.04 - 0.434\,X_{t-1} - 0.03\,X_{t-2} + 0.373\,X_{t-3} + e_t^{(2)} \\ \quad if\;\; -0.065 \le X_{t-2} < -0.017 \\ 0.05 - 0.742\,X_{t-1} - 0.246\,X_{t-2} + 0.378\,X_{t-3} + e_t^{(3)} \\ \quad if\;\; -0.017 \le X_{t-2} < 0.029 \\ -0.03 + 0.098\,X_{t-1} - 0.22\,X_{t-2} - 0.24\,X_{t-3} + e_t^{(4)} \\ \quad if\;\; 0.029 \le X_{t-2} \end{cases}$$

4. STABILITY ANALYSIS

The model was validated for its stability using various standard statistical tests. The test statistics and standard error of estimated parameters in different regimes are illustrated in **(Table 2)**:

Table 2: Estimated test statistics and standard error

Necessary test statistics of the projected model are represented in **(Table 3)**.

Regime	X_{t-1}	X_{t-2}	X_{t-3}	Std Erra$_i^j$
1 *t-ratio*	-0.15	6.4	-5.64	0.095
2 *t-ratio*	-2.6	0.02	3.3	0.03
3 *t-ratio*	-15.2	1.8	2.78	0.03
4 *t-ratio*	1.29	4.1	-1.7	0.03

Table 3: Model Statistics

Test Statistics	Computed Values
R Squared	0.9303
SE of Regression	0.08
Mean Dependent Var	0.019
SD Dependant Var	0.309
Hannan-Quin Criterian	-1.84
Durbin Watson	2.04

The innovations represented by $e_t^{(j)},\ for\ j = 1,2,3,4$ are checked to be identically and independently distributed by plotting histogram. The disturbance histogram in **(Fig. 2)** shows that the residuals are normal, whereas DW in table 3 confirms no serial correlation between the residuals.

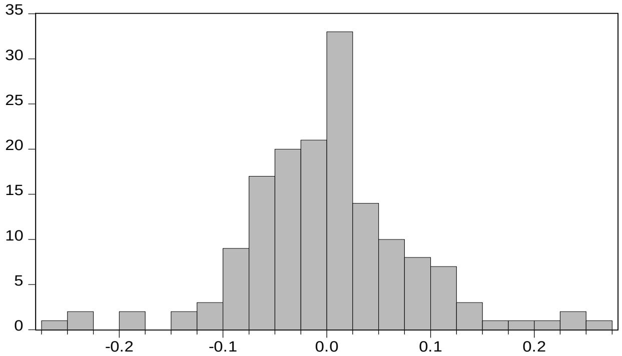

Fig 2: Histogram for residual distribution

CUSUM test shows that estimated coefficients using threshold classifier are in control & does not deviate significantly from the bench mark as evident from the illustrate graph in **(Fig. 3)**.

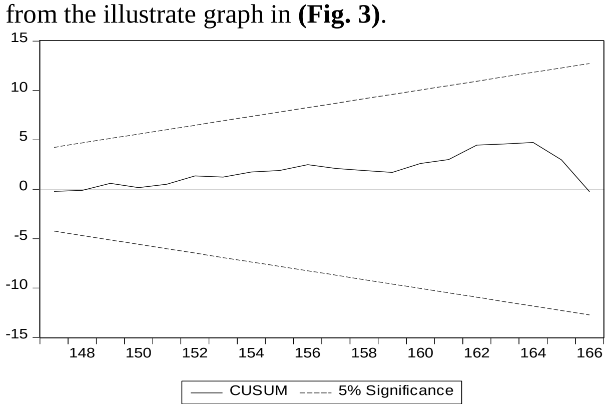

Fig 3. CUSUM test confirm the control of Threshold classifier.

The computed threshold regression equation with its necessary statistics authenticate the assessed model. The estimated model effectively calculated threshold values of vibration which were significantly effecting on the health of bearing. The defect features extracted by the proposed model in the form of threshold variable and regimes were found almost similar with those extracted by other tools. Furthermore, the proposed model not only defines a single value of significance, rather defines a region of importance, which has significant effect on the state of machinery component.

5. RESULTS AND DISCUSSION

The vibration data of the bearing was collected during accelerated testing as mentioned in para 3 under application. Only the significant values related to bearing defect were taken into consideration rather the whole broad band frequency spectrum. During the modeling, the model was initially trained on 80% of the randomly selected data and subsequently it was tested on the remaining 20% data to ascertain its prediction capability.

(Fig. 4) shows the real time instability in actual and estimated health indicator which demonstrates inconsistency in machinery health after its usage of about 150 running hours and it keeps on growing afterwards. The warning and failure zones are depicted in figure below. Machine continuous in present state will be a complete failure after 164 running hours

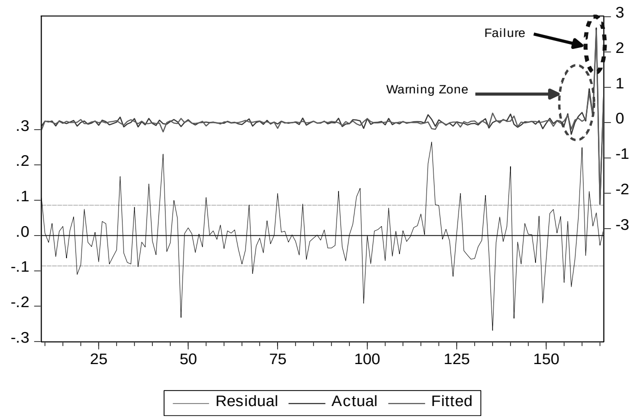

Fig 4. Actual & Estimated health indicator with error plot

For threshold autoregressive models with lagged endogenous variables, *n*-step ahead nonlinear dynamic forecasting is used which is considerably difficult (Potter, 1999). The forecasts were computed by stochastic simulation with the forecast and forecast standard errors obtained from sample average and standard deviation of the simulated values. The purpose of forecast is to provide the best estimate of what will happen at specified point in time in future. Based upon the model fitted on vibration data and the most recent observation, we can obtain minimum mean square error forecast of future observation.

The forecasted plots and test statistics signifies the predicting performance of the estimated model as confirmed by depicted graph and predicting output below **(Fig. 5)**:

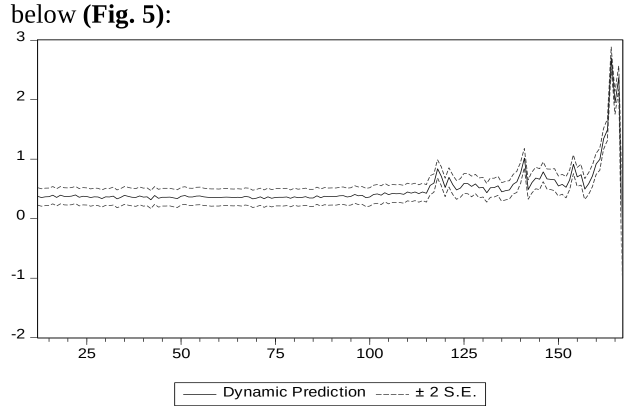

Fig 5. Dynamic Forecast of X_t with ± 2 S. E

The plots and test statistics for the forecasting as evident from Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) of the forecast signifies the predicting performance of the estimated model as mentioned in **(Table 4)**.

Table 4. Forecast Output

Forecast Parameter	Forecast Statistics
RMSE	0.016
MAE	0.041
Theil Inequality	0.052

6. CONCLUSION

In this study data driven machinery health prognosis using threshold regression with its multistep ahead prediction is presented. The adequate model is selected on the basis of AIC criteria and then the parameter of the suitable model were computed. The proposed model is designed for data driven machinery health condition and prognosis which could efficiently used for analyzing machinery health and subsequent cost savings. The estimated output of the constructed model and prediction, authenticate the prediction performance of the computed model. The proposed model possess feature extraction capability as well as prediction efficiency. As a future work the concept can be applied for calculating remaining useful life W.R.T time introducing various variables involve. The concept can be extended on different health indicators provided by available condition monitoring tools.

REFERENCES:

Bai J. and P. Perron, (1998). "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, vol. 66, no. 1, 47–78.

De Gooijer J. G. and R. J. Hyndman, (2006). "25 years of time series forecasting," *Int. J. Forecast.*, vol. 22, no. 3, 443–473.

Hussain, M., A. Mansoor and Q. Ali, (2017) "A Novel Approach for Machinery Health Prognostics Using Statistical Tools", *Proceedings of 50th International Conference of Machinery Failure Prevention Technologies*

Heng, A., S. Zhang, A. C. C. Tan, and J. Mathew, (2009). "Rotating machinery prognostics: State of the art, challenges and opportunities," *Mech. Syst. Signal Process.*, vol. 23, no. 3, 724–739,

Kan, M. S., A. C. C. Tan, and J. Mathew, (2015). "A review on prognostic techniques for non-stationary and non-linear rotating systems," *Mech. Syst. Signal Process.*, vol. 62, 1–20.

Lee, J., F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, (2014). "Prognostics and health management design for rotary machinery systems - Reviews, methodology and applications," *Mech. Syst. Signal Process.*, vol. 42, no. 1–2, 314–334,.

Niu G. and B. S. Yang, (2009). "Dempster–Shafer regression for multi-step-ahead time-series prediction towards data-driven machinery prognosis," *Mech. Syst. Signal Process.*, vol. 23, no. 3, 740–751,.

Potter, S. (1999) "Nonlinear Time Series Modelling: An Introduction," *J. Econ. Surv.*, vol. 13, no. 5, pp. 505–528,

Qiu, H., J. Lee, J. Lin, and G. Yu, (2006.). "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics," *J. Sound Vib.*, vol. 289, no. 4–5, 1066–1090,

Ramasso E. and A. Saxena, (2014). "Performance Benchmarking and Analysis of Prognostic Methods for CMAPSS Datasets," *Int. J. Progn. Heal. Manag.*, no. ISSN2153–2648, 1–15.

Tong H. and K. S. Lim, (1980). "Threshold autoregression , limit cycles and cyclical data," *J. R. Stat. Soc. Ser. B*, vol. 42, no. 3, 245–292,.