International Journal of Applied Mathematics and Statistics,
Int. J. Appl. Math. Stat.; Vol. 53; Issue No. 5; Year 2015,
ISSN 0973-1377 (Print), ISSN 0973-7545 (Online)
Copyright © 2015 by CESER PUBLICATIONS

Cox Proportional Hazard with Multivariate Adaptive Regression Splines to Analyze the Product Sales Time in E-Commerce

E. Irwansyah[1] , D. A. N. O. Stefani[2] and **R. D. Bekti[3]**

[1]Department of Computer Science, Bina Nusantara University,
Jl. K.H. Syahdan no. 9, Palmerah, Jakarta Barat 11480, Indonesia;
Email : edirwan@yahoo.com

[2]Department of Statistics, Bina Nusantara University,
Jl. K.H. Syahdan no. 9, Palmerah, Jakarta Barat 11480, Indonesia;
Email : dewaoctalia@yaho.com

[3]Department of Mathematics IST Akprind,
JlKalisahak No 28 KomplekBalapan Yogyakarta 55222, Indonesia;
Email : groo_jgroo@yahoo.com

ABSTRACT

Cox Proportional Hazard (Cox PH) model is a survival analysis method to perform model of relationship between independent variable and dependent variable which shown by time until an event occurs. This method compute residuals, martingale or deviance, which can used to diagnostic the lack of fit of a model and PH assumption. The alternative method if these not satisfied is Multivariate Adaptive Regression Splines(MARS) approach. This method use to perform the analysis of product selling time in e-commerce. The samples were collected by survey on website. The results areMARS model with martingale residuals has good performance than residual deviance. MARS modeling with martingale residuals have GCV minimum 0.502 with a combination of BF = 10, MI = 1, and MO = 2 with information number of products sold (X_6) that contribute. Variables significant effect on α = 5% were $BF_2 = (X_6\text{-}135)_+$, $BF_3 = (X_6\text{-}170)_+$, and $BF_5=(X_6\text{-}196)_+$.

Keywords: Cox PH Model, MARS, product selling time, e-commerce.

Mathematics Subject Classification: 62N02, 62H86

Journal of Economic Literature (JEL) Classification:C10

1. INTRODUCTION

Survival analysis use for data analysis when the outcome variable of interest is time until an event occurs. Cox Proportional Hazard (Cox PH) is one method of semi parametric survival analysis to estimate effect of independent variable in survival data (Kleinbaum and Klein, 2011). Survival analysis also includesmodelingthe relationshipbetweenthe independent variableswiththe function ofthe model calledMultivariateAdaptiveRegressionSplines(MARS) which developed byFriedman, 1991.This method is flexiblefor thehigh-dimensionaldata, which therelationships betweenindependent and

dependent variable scan be linear or non-linear pattern. In addition, MARS modeling can involve a lot of interaction between in dependent variables and able to detect these interactions (Kreiner, 2007).

Research of Kreiner, 2007, used Cox PH with MARS approach to analyse the survival time of patients after a heart attack and the factors that influence the survival time of patients. The advantage of this study is that researchers can compare with modeling survival analysis with MARS approach. This research concludes that survival MARS presents a feasible and powerful alternative to Cox PH and even superior in the case of nonlinear effects and interactions. MARS provides simple data-driven and hands-off modeling of these types of functional forms. Research of Javier et al, 2011, uses survival analysis with MARS to predictingbankruptcyof a companyinSpainin theconstruction sector.Then Nisa and Budiantara, 2012 use Cox PH with MARS approach in dengue fever.Cox PH model is compute residuals which named as martingale and deviance residuals.The analyses of these residual is important for diagnostic the lack of fit of a model to a given subject and assumption model (Gharibvand et al, 2008). Since the curve of residuals and independent variable was close to zero line, it can be concluded that the PH assumption was not satisfied. So, alternative method is needed to use.

E-Commerce is a marketing system with electronic media. E-Commerce includes distribution, sales, purchasing, marketing and service of a product that made in an electronic system via internet. Nielsen research firm states that 70% of e-commerce users in Indonesia use the internet with the purpose of purchasing products online (Nielsen, 2013). E-Marketer, a company engaged in the research record that Indonesia has increased the number of online shoppers, from 5.2 million people in 2011 increased to 10.6 million in 2013 and spent $ 1.8 billion USD costs arising in the purchase of products online. Indonesia itself was up in 2013 has had an online store of more than 3,000 to more than 550,000 and 8.5 million online sellers of products sold online (Pricearea, 2013).Armesh et al, 2010,show thatthe valueof a product, designandease ofaccess tothe websiteandproductadjustmentsaffectconsumer satisfactioninmaking purchasesvia the Internet. Then, Wang and Zhang, 2013, also use survival analysis to get a new opportunity model to explicitly incorporate time in an e-commerce recommender system.

This research perform the Cox PH survival analysis with MARS approach to get the relationship between marketing products (price, type of products, and information in website) and selling time. It also compares the results of MARS model with martingale and deviance residuals.

2. GENERATION OF THE DATA

The data source is primarydata by survey at February 2014 until March 2014. The survey was conductedby periodic monitoringabout products soldon the websiteof PT. OnlinePertamainwww.lakupon.com. The total sample was 279 samples. The data analysis use dependent and independent variables. The dependent variable is time of product from displayed on website until was sold. The time unit is days. The status of that time was categorized into censored and not censored data. The independent variables (x) were type of products in dummy variable Gadget and Electronics (X_1), Fashion and Beauty (X_2), Baby and Kids (X_3), Home Appliances (X_4), and Food and Beverages. Other independent variables were price (X_5), number of products soldinformation(X_6), and status of selling time (X_7) which limited or unlimited.

2.1. Cox PH Model with MARS

The Cox PH model used was:

$$h(t,\mathbf{X}) = h_0(t)e^{\beta_1 x_1+\beta_2 x_2+\ldots+\beta_7 x_7}$$

Where $h_0(t)$ is baseline hazard, β is parameter model, and x is independent variable.Then the Cox PH model was combined with MARS approach model. It uses the martingale and deviance residuals as dependent variable.

$$r_{Mi} = N_t - \hat{h_0}(t)e^{\beta_1 x_1+\beta_2 x_2+\ldots+\beta_7 x_7}$$

$$r_{Di} = \text{sgn}(r_{Mi})[-2\{r_{Mi} + N(t)\log(N(t) - r_{Mi})\}]^{1/2}$$

Wherer_{Mi} ismartingale residual *i-th*, r_{Di} isdeviance residual *i-th*at time *t-th*, and $N_t$is1 for non-censored data and 0 for censored data. $\hat{h_0}(t)$ isbaseline hazard estimation at time *i-th*which computed from Cox PH Model.The residual martingale use to develop MARS model, as follows:

$$f(x) = a_0 + \sum_{m=1}^{M} a_m \prod_{k=1}^{K_m} [S_{km}(x_{v(k,m)} - t_{km})]_+$$

Where a_0 is main basis function, a_m is coeefficients of m-basis function, M is the maximum basis function, $K_m$is the degree of interaction, S_{km}= ± 1, $x_{v(k,m)}$is independent variables, and $t_{km}$is knot value for independent variable$x_{v(k,m)}$.

3. RESULTS

From 279 product samples,there are 167 products which these times not censored and 112 products which censored. For not censored data, the average time of product from displayed on website until sold was 21.661 days. The fastest time was 2 days and the longest time was 39 days.Based on the type of product, a product which has the fastest selling time is type home and appliances that during 7.066 days. Then, the product which have fastest selling time based on price was at a price less than Rp. 50,000,-, based on status of time sold was a product with limited sales status, and based on the number of products sold information wasthe product which the information was less than 50 pieces.

3.1. Cox PH Modeling

Cox PH model was done to determine the factors that affect for selling time. It also used stepwise Cox PH to get the best result (see Table 1). At α = 5%, the factors which influence in the model were type fashion and beauty products (X_2), the type of baby and kids products (X_3), the type of food product and beverages (X_4), number of products soldinformation (X_6), and the status of selling time (X_7). Hazard Ratio (HR) performs that risk of selling products fashion and beauty was 1.51times largerthanothers. Product babyandkidshavesoldrisk1.67times greaterthanothers. Product foodandbeverageshas soldrisk1.77times greaterthanothers. Product with high number of products soldinformation has high opportunities selling fast. Then product with limited status of selling time has high opportunities selling faster than unlimited status.

The Cox PH assumption was not satisfied, especially for variable type of food product and beverages and status of selling time. It isalso shown by scatterplot of residual and independent variable, especially number of products soldinformation in Figure 1.It shows that there were pattern of them. So,this research use MARS approach to perform it.

*Table 1:*Stepwise Cox PH Modeling

Variables	Coefficients	Hazard Ratio	Z-Value	P-Value
X_2	0.414	1.51	2.05	0.040
X_3	0.512	1.67	2.43	0.015
X_4	0.571	1.77	2.37	0.018
X_6	0.004	1.22	3.72	0.000
X_7	2.624	13.79	10.67	0.000

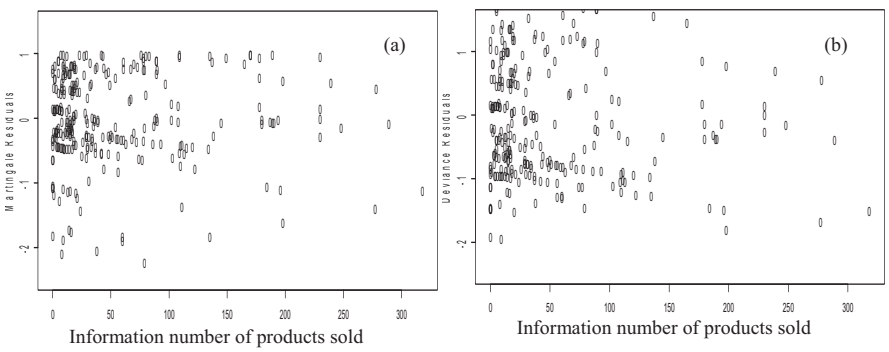

Figure 1. Scatterplot of number of productssoldinformation and martingale residual (a) and devianceresidual (b)

3.2. MARS Approach

Modelling analysis of survival with MARS approach was done by trial and error combined value of BF (BF), Minimum Observation (MO), and Maximum Interaction (MI). After that, get the best model to determine the lowest GCV value. The dependent variable is martingale and deviance residuals. MARS model use combinations BF = 10, MI = 1, and MO=2, then get the best model with GCV value = 0.502. Estimation parameter and significantly basis function or independent variable of MARS approach can be seen on Table 2. All basis function in MARS with martingale residuals and BF_2, BF_3, and BF_4in MARS with deviance residuals are significantly influence on model at α=5%. These basis functions consist of independent variable information number of products sold (X_6) and status of selling time (X_7). MARS with martingale and deviance residual are different. Mean Square Error (MSE) in MARS with martingale residual is lower than MARS with residual deviance. Therefore, it can be concluded that the MARS modelling with martingale residuals better than the residual deviance.

Table 2: MARS Model with Martingale and Deviance Residual as Dependent Variable

Basis Function	Coefficients	t – value	P-Value
Martingale Residuals (MSE=0,0025)			
Constant	-0.2712		
$BF_1 = (X_6 – 89)_+$	-0.0120	-1.690	0.092**
$BF_2 = (X_6 – 135)_+$	0.0440	2.759	0.006*
$BF_3 = (X_6 – 170)_+$	-0.0728	-2.739	0,018*
$BF_4 = (170 - X_6)_+$	-0.0018	-1.096	0.274
$BF_5 = (X_6 - 196)_+$	0.0401	1.987	0,048*
Deviance Residuals (MSE=0.0039)			
Constant	0,6225		
X_7	0.2477	1,873	0,062**
$BF_1 = (X_6 – 89)_+$	-0.0274	-1,908	0,057**
$BF_2 = (X_6 – 122)_+$	0.0508	2,398	0,017*
$BF_3 = (X_6 – 170)_+$	-0.0381	-2,881	0,004*
$BF_4 = (170 - X_6)_+$	-0.0049	-1,81	0,007*

Note: *) significantly at α=5% and **) significantly at α=10%

Plots of BF and variable number of sold product information are presented in Figure 2 and 3. In martingale residuals, the increased of risk of selling product occurs when information number of products sold (X_6) between 0 to 89 pieces. Furthermore, when the information number of products sold between 89 to 135 pieces, the risk of selling product has decreased. When the number of products sold between 135 to 170 pieces, the risk increased. However, when the number of products sold is more than 170 pieces sold, the risk has decreased dramatically to 196. In fact, itcontinuesdecreaseduntil300 pieces.

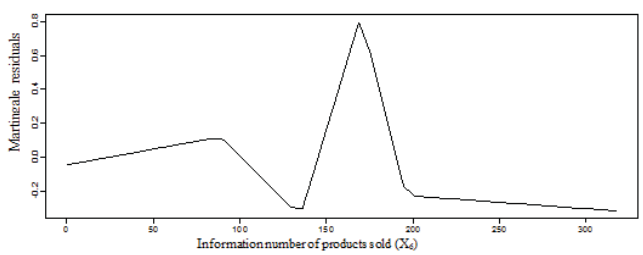

Figure 2.Plot of MARS basis functionswith martingale and number of productssold (X_6)

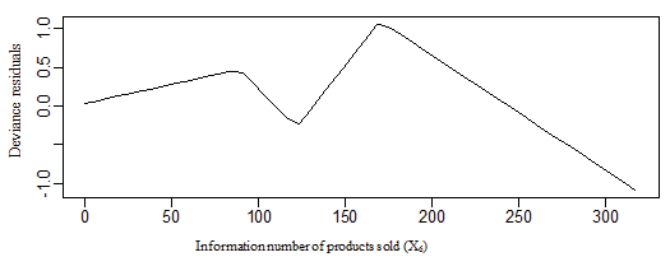

Figure 3.Plot of MARS basis functionswithdevianceresiduals and number of productssold (X_6)

4. DISCUSSION AND CONCLUSION

Based on the analysis, it can be concluded that stepwise Cox PH model with α = 5%, the variables that significantly affect the selling timewas type of fashion and beauty products (X_2), the type of baby and kids products (X_3), the type of food products and beverages (X_4), information number of products sold (X_6) and the status of selling time (X_7). The Cox PH assumption was not satisfied and there is pattern between residual and information number of products sold (X_6), so MARS approach is alternative. MARS model with martingale residuals has good performance than deviance residual. MARS modelling with martingale residuals have GCV minimum 0.502 with a combination of BF = 10, MI = 1, and MO = 2 with 1 variables that contribute. Variables significant effect on α = 0.05 were $BF_2 = (X_6\text{-}135)_+$, $BF_3 = (X_6\text{-}170)_+$, and $BF_5=(X_6\text{-}196)_+$.

This model give much information about marketing e-commerce, such as when is the product will be high sold. The most important variable which effect on selling is information about the number of products sold. This refers to the buyer's number displayed on the web. If the number of buyers which written in the web was high, it means that demand is high. Consumers always see whether other consumers interested or not. If there are more consumers interested so it means the product has a good quality or special attraction. It will motivate consumers to compete buy the product. Therefore, the one way to increase selling is always display the product sold information and updated.

Information when the products sold up and down is also generated from this model, which is shown on a plot basis function MARS. Plots with martingale residuals shows that the Increased of risk of selling product information occurs when number of products sold between 0 to 89 pieces and between 135 to 170 pieces. Furthermore, when the information number of products sold between 89 to 135 pieces and more than 170 pieces, the risk of selling the product has decreased.

In contrast to the results of the Cox PH models, it is only provide risk information sold on the hazard ratio. Variable X_6 has a hazard ratio of 1.22. It means that the product with high number of products sold information has high opportunities selling fast. It gives information that if consumers are high then the product has high risk to sold. But, in fact it is not always the case, where the number of consumers can go up and down as generated in the MARS approach.

5. REFERENCES

Armesh, H., Salarzehi, H., Yaghoobi, N.M., Heydari, A., Nikbin, D., 2010, Impact of Online/Internet Marketing on Computer Industry in Malaysia in Enhancing Consumer Experience. *International Journal of Marketing.* **2,** 75-86.

Friedman, J.H., 1991, Multivariate Adaptive Regression Splines, *The Annals ofStatistics*. **19**, 1– 67.

Gharibvand, L., Jeske, D.R.,Liao, S., 2008. *Evaluation of a Hospice Care Referral Program Using Cox Proportional Hazards Model, WUSS 2008 Annual Conference Proceedings***.**

Javier, DE. A., Fernando, S., Pedro, L.,Francisco, DE.C., 2011,*A Hybrid Device of Self Organizing Maps (SOM) and Multivariate Adaptive Regression Splines (MARS) for The Forecasting of Firms Bakruptcy*. University of Oveido.

Kleinbaum, G. D., Klein, M., 2011.*Survival Analysis*.3rd ed.Springer Science + Business Media.

Kreiner.M., 2007.*Survival Analysis With Multivariate Adaptive Regression Splines.* Ph.DDissertation, Munchen University Germany.

Nielsen, 2013.*Indonesian Online User*. Accessed 23 October 2013 at http://www.acnielsen.co.id/news/NEWS14072010. hmtl.

Nisa', S.F., Budiantara, I.N., 2012, Analisis Survival denganPendekatan Multivariate Adaptive Regresi Splines pada Kasus Demam Berdarah Dengue (DBD). *Jurnal SainsdanSeniITS*.**1,**D319-D320.

Pricearea. 2013. *The State of eCommerce Indonesia*. Accessed 27 October 2013 athttp://www.techinasia.com/keynote-state-ecommerce-indonesia-live-blog/

Wang, J., Zhang, Y., 2013.*Opportunity model for e-commerce recommendation: right product; right time***,** Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval, Pages 303-312.