Jurnal Ilmu Keuangan dan Perbankan (JIKA)
Volume 12 Nomor 1 (Desember 2022)
E-ISSN : 2655-9234 (Online) P-ISSN : 2089-2845 (Print)
Email : jika@email.unikom.ac.id
Website : www.ojs.unikom.ac.id/index.php/jika

Analysis of The Debtor's Endurance using Cox Regression Semiparametric Method

Vitri Aprilla Handayani[1]*, Widya Reza[2],
Andini Setyo Anggraeni [3], Garry Rusmadi[4]

*vitri@iteba.ac.id**

Institut Teknologi Batam, Jl. Gajah Mada Complex Vitka City, Batam, Kepulauan Riau, Indonesia

Received Date : 29.05.2022
Revised Date : 17.06.2022
Accepted Date : 16.11.2022

ABSTRACT

The aim of this research was conducted to determine the factors that influence the resilience of car loan debtors in an area. The research method used is semiparametric Cox regression on secondary data, WAREHOUSE consisting of the customer profile (demography) and historical payment consisting of 25 observed variables. The Cox regression model used in this study is a proportional hazard Cox regression model. Based on the description of the data, it shows that there is a risk of defaulting debtors who have been past due compared to never past due. Cox proportional hazard regression model with the semiparametric method can be used to identify factors that affect debtor resilience. Of the 25 variables observed, there is 1 (one) variable, namely the X16 (Age) variable which does not significantly affect the debtor's endurance.

Keywords : Debtor Resistance; Historical Payment; Semiparametric; Cox Regression; Proportional Hazard

ABSTRAK

Penelitian ini bertujuan untuk mengetahui faktor yang berpengaruh terhadap daya tahan debitur kredit mobil di suatu wilayah. Metode penelitian yang digunakan pada penelitian ini adalah semiparametric regresi cox pada data sekunder, warehouse yang terdiri profil (demographi) customer dan historycal payment terdiri atas 25 varibel yang diamati. Model regresi Cox yang digunakan pada penelitian ini adalah model regresi Cox hazard proporsional. Berdasarkan deskripsi data menunjukkan adanya risiko debitur gagal bayar yang Pernah Past Due dibandingkan Never Past Due. Model regresi cox proportional hazard dengan metode semiparametrik dapat digunakan untuk mengedentifikasi faktor-faktor yang mempengaruhi daya tahan debitur. Dari 25 variable yang diamati ada 1 (satu) variable yaitu peubah X_{16} (Age) yang tidak signifikan berpengaruh terhadap daya tahan debitur.

Kata Kunci : Daya Tahan Debitur; Riwayat Pembayaran; Semiparametrik; Regresi Cox; Proporsi Hazard

INTRODUCTION

Based on the MPP forecast data for Field Collectors, an increase in the contract bucket dpd 1-30 days where in 2018 will increase 34.4% from 2017 and in 2019 will increase by 25% from 2018. This increase in historical data will increase the number of Man the Power Plan (MPP) for Field collectors 1-30 days will increase linearly with the increase in dpd 1-30 buckets, which will have implications for increasing OPEX in the Collection & Recovery Division of a company. Therefore, there are several considerations so as not to increase the number of MPP Field Collectors 1-30, namely, first, only customers with high-risk criteria are billed using field collectors. Second, customers with low-risk criteria can be made using a soft collection pattern. Based on these 2 things, several attempts were made to minimize risk and make it more efficient in risk management. The risk in question is the failure of the debtor to make loan payments to the lender or the so-called creditor. Lending risk can be minimized by means of more selective lenders in conducting a risk analysis of debtors who will apply for loans. The preparation of a model is an alternative that can be done to find out the risk of default on debtors who have been past due compared to never past due and to find out the factors that influence the failure of debtors to make loan payments to lenders.

The risk in question is the failure of the debtor to make loan payments to the lender or the so-called creditor. Lending risk can be minimized by means of more selective lenders in conducting a risk analysis of debtors who will apply for loans. The preparation of a model is an alternative that can be done to find out the risk of default on debtors who have been past due compared to never past due and to find out the factors that influence the failure of debtors to make loan payments to lenders. The risk in question is the failure of the debtor to make loan payments to the lender or the so-called creditor. Lending risk can be minimized by means of more selective lenders in conducting a risk analysis of debtors who will apply for loans. The preparation of a model is an alternative that can be done to find out the risk of default on debtors who have been past due compared to never past due and to find out the factors that influence the failure of debtors to make loan payments to lenders.

According to Prasetya (2006), several factors that are considered to influence the debtor's failure to make payments include the demographic characteristics and debtor loans. Modeling the characteristics of this debtor uses endurance analysis to see the effect of debtor default. Durability analysis is one of the statistical analysis techniques used to analyze the endurance time of debtors in making their loan payments.

Hogg *et.al* (2013), some of the method used in the analysis of durability include semiparametric method such as the Cox model. Cox regression is used to estimate comparison ratios without knowing the basic hazard function used. Cox regression can estimate the basic hazard function, hazard function, and endurance function even though the basic hazard function is unknown, the Cox model can give results that are almost the same as those of the parametric model (Kleinbaum and Klein, 2012). The Cox regression model can provide useful information in the form of a hazard ratio. According to Lift and Noeryanti (2018), the comparison ratio is a comparison of the hazard value or risk of default between one debtor and another which is constant and is not affected by the element of time. Therefore,

Survival data is observation data for a certain period of time from the beginning of the observation until an event occurs. These events can be in the form of failure, death, response, onset of symptoms and others (Lee *et.al*, 2002). The analysis of durability is a statistical technique used to describe and measure data on the time of occurrence

(Stevenson, 2007). The application of resilience analysis is used to events such as the death of a person, animal or plant. This is different from the occurrence of failures in the manufacture of goods such as light bulbs, which is known as failure time analysis. The objectives of the resilience analysis according to Kleinbaum & Klein (2012) include: predicting and interpreting survivors and/or hazard functions from survival data,

The survival function gives the value of the life probability of a new birth that will die after t. This is tantamount to saying that a new birth survives to age t, or lives to age t. The survival function is the probability that a randomly selected individual will survive until time t or more which is defined as

$$S(t) = P(T \geq t) = 1 - P(T \leq t) = 1 - F(t) \quad (1)$$

Lee & Wang (2003), stated that the survival function is the basis for survival analysis, because it produces survival opportunities for different t values, and provides an important summary of information from survival data (Kleinbaum & Klein, 2012). The relationship between the probability density function and the survival function is:

$$f(t) = \lim_{\Delta t \to 0} \left[\frac{P(t \leq T < (t + \Delta t))}{\Delta t}\right] = F'(t) = -S'(t) \quad (2)$$

Survival analysis analysis of data by paying attention to the time element of the event to see the variables that influence the observed events (Silmi *et.al*, 2020). In theory, the distance t is from 0 to infinity (infinity). In its use, the survival function can be illustrated by a graph, where the value of t is on the X axis as shown in the following figure:

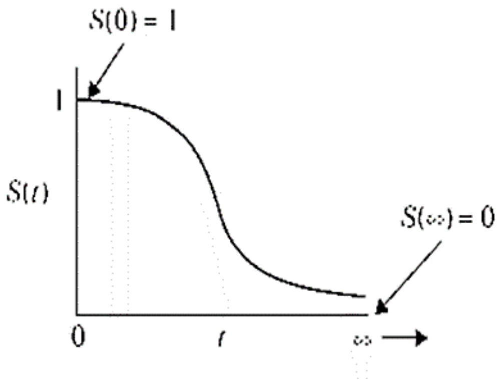

Figure 1. Graph of the Survival Function

The proportional hazard model is one of the nonparametric method available in survival cases which can determine the factors that influence whether or not an event occurs within a certain time. These factors are independent variables known as covariates, which are independent of time (Kleinbaum & Klein, 2005).

The proportional hazard model (Cox Proportional Hazard) is used to relate survival time as a dependent variable with a set of covariate-free variables that are independent of time (Amurwani, 2012). The general form of the Hazard Ratio in the proportional hazard model is as follows: The general form of the proportional hazard model

$$h(t, X) = h_0(t) \exp\left(\sum_{i=1}^{p} \hat{\beta}_i X_i\right) \quad (3)$$

Where $h_0\,(t)$ is the baseline hazard function of t, e the exponential of the covariate that includes but does not involve t is $X = (X_{1,}X_{2,}\ldots X_p)$, dan β is the coefficient of the covariate (parameter). The covariates here are time independent, meaning that the covariates have constant values (do not change) all the time or do not depend on time. Hazard Ratio in the proportional hazard model.

$$\widehat{\boldsymbol{H}}\boldsymbol{R} = \frac{\hat{h}(t.X^*)}{\hat{h}(t.X)}\,\mathbf{exp}\left(\sum_{i=1}^{p}\widehat{\boldsymbol{\beta}}_i\,(\boldsymbol{X}_i^* - \boldsymbol{X}_i)\right) \quad \textbf{(4)}$$

Where β̂ is the estimated covariate coefficient (parameter), X^* is the covariate coded by writing 1. In the form X^*=(X_1^*,X_2^*,…X_p^*) and X is the covariate coded by writing 0 (as a comparison), which is X= (X_(1,) X_(2,)…X_(p,)).

The Gini index is a statistical analysis that is commonly used to measure the concentration of positive random variables in the social sciences. However, several researchers have conducted research on the use of the Gini index in evaluating inequality of life chances in the health sector. Subsequent studies have shown that based on simulations, selecting the best model for survival analysis using the Gini index is able to provide a higher power value than other alternatives (Todaro MP and Smith SC, 2006).

The Gini index numbers range from 0 to 1. The highest Gini index is determined as the best method.

$$\boldsymbol{G} = \mathbf{1} - \sum_{1}^{n}\boldsymbol{fp}_i(\boldsymbol{Fc}_i - \boldsymbol{Fc}_{i-1}) \quad \textbf{(5)}$$

Where: [G] Gini Ratio; [fp_i] the frequency of the number who died in class-I; [Fc_i] cumulative frequency of the cumulative sum of the Hazard Ratio Model in class I; [Fc_{i-1}] cumulative frequency of the cumulative sum of the Hazard Ratio Model in class-(i-1).

Therefore, this research was conducted with the aim of knowing the factors that influence the durability of car loan debtors in a region. 25 observed variables. The hypothesis in this study is H_0: All observed factors have a significant effect on debtor endurance, and H_1: Not all observed factors have a significant effect on debtor endurance.

RESEARCH METHOD

Hermanto (2006), A debtor is said to be bad credit if the debtor does not make installment payments within 2 months after the debtor's maturity (month-end cut off). The data in this study used 2 (two) types of data, namely complete and censored data. Complete data is data where the debtor has failed to pay during the observation period. Censored data (right censor) is data where the debtor is able to pay off his loan before the end of the observation period or the debtor is still running credit until the end of the observation period.

The steps of data analysis using the proportional cox hazard regression model were carried out in research by determining the variables contained in the customer profile (demographics) and the history of cal payments consisting of 25 variables. Before data processing is done is the preparation of data. Data preparation is done by changing all continuous variables into classes or categorical (Process Bining) with the help of Ms. software. Excel. Furthermore, descriptive and exploratory analysis was carried out on the observed data by looking at all the independent variables using a nonparametric method, namely the Survival Function. (Pahlevi. 2016) This analysis was conducted to determine the pattern of resistance in each independent variable in modeling with the Cox Proportional Hazard regression model.

$$h(t, X(t)) = h_0(t)\ \exp\left(\sum_{i=1}^{p_1} \hat{\beta}_i\ X_i - \sum_{j=1}^{p_2} \hat{\delta}_j\ X_j(t)\right) \tag{6}$$

Where: $[h_0\ (t)]$ baseline hazard function from t; $[\hat{\beta}_i$ dan $\hat{\delta}_j]$ expected covariate coefficients (parameters); $[X_i]$ time independent covariate; $[X_j]$; time dependent covariate.

RESULTS AND DISCUSSIONS

Overall, the percentage of smooth applications for car loans is 117455 applications (88.32%) and those that are stuck late +60 days are 15534 applications (11.68%) (Figure 1). Judging from one of the variables Historical Payment Never Past Due for 3 First Installments, those who have Past Due have a 44.79 percent chance that they will default. Conversely, if the first 3 installments are never past due, then the probability of default is 7.91 percent (Figure 1). From this, it can be concluded that the risk of default on debtors who have had past due for the first 3 installments is 5.6 times compared to never past due. (Figure 2).

Figure 2. Stuck-Late Credit Status

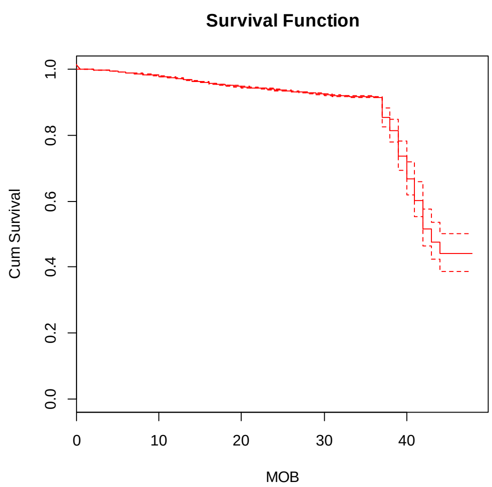

Figure 3. Past Due and Survival Functions

Based on the figure above, the survival graph shows that the chance of survival decreases after 3 years or 36 months. Furthermore, there will be a very drastic decrease in credit resilience. This shows the failure to pay by the debtor occurs over 3 years.

Model candidates are built by looking at the significance of the independent variables in the model, namely by ignoring insignificant variables or by considering other business interests. Therefore, before obtaining a suitable final model, the previous model (which was not suitable) was also built with a combination of variables to see which combination could provide the best possible model. Saving other possible models (model candidates) is intended for model candidates to be able to show how all the variables interact when all of these variables are used as a model, and the candidate model will be used as a benchmark for entering further combinations of variables. The GINI values used in the final model can be seen in table 1.

Tabel 1. GINI Values for the Final Model

	MODEL SJMB CAR
Lots of good Contracts	117,455
Lots of bad Contracts	15534
Total Contracts	132989
Bad rate	11.68 %
GINI	67.50

Source: Analysis output results, 2022

Based on the output results, the results of the Cox Proportional hazard model analysis performed on the SJMB CAR data refer to the significance value (p-Value) generated for each of the observed indicators. The significance value in question can be seen in table 2.

Table 2. Analysis Results Using the Cox Proportional Hazard Model in SJMB CAR

	coef	exp(coef)	se(coef)	z	p
Aset Category Dummy 1	-0.0886	0.9152	0.0214	-4.14	3.50E-05
Aset Category Dummy 2	0.1032	1.1087	0.031	3.33	0.00088
Education	0.1459	1.1571	0.0185	7.89	3.10E-15
Home Location	0.1305	1.1394	0.0163	8.02	1.00E-15
Purpose of Financing Dummy 1	0.1124	1.1189	0.0259	4.33	1.50E-05
Purpose of Financing Dummy 2	0.1237	1.1317	0.0254	4.87	1.10E-06
Customer Profesion	0.121	1.1286	0.0186	6.49	8.40E-11
Customer Status	0.2931	1.3405	0.0188	15.55	2.00E-16
Age of Unit	0.188	1.2069	0.0214	8.8	2.00E-16
LTV dummy 1	0.6536	1.9225	0.0387	16.88	2.00E-16
LTV dummy 2	1.0333	2.8104	0.0367	28.19	2.00E-16
LTV dummy 3	1.2199	3.3869	0.0375	32.51	2.00E-16
Age Dummy 1	0.0794	1.0826	0.0242	3.27	0.00106
Age Dummy 2	0.1928	1.2126	0.0236	8.16	3.30E-16
Age Dummy 3	0.2185	1.2443	0.0296	7.39	1.40E-13
Pernah Telat dalam 6 angsuran pertama	1.7954	6.022	0.0186	96.47	2.00E-16
Keterlambatan pembayaran > 30 hari dalam 3 angsuran pertama (ever 30)	0.8249	2.2816	0.0332	24.86	2.00E-16
Keterlambatan pembayaran > 30 hari dalam 6 angsuran pertama (ever 30)	2.1208	8.3378	0.0227	93.39	2.00E-16

Source: Analysis output results, 2022

Plot of the GINI value based on the selected final model can be seen in Figure 4.

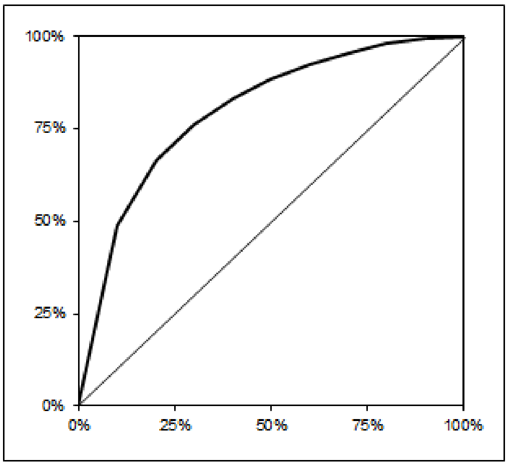

Figure 4. SJMB CAR GINI Curve in the Final Model

Model validation is the most important step in modeling, because it will show how credible the resulting model is. Model validation is carried out to measure the extent to which the model results are close to actual conditions.

In cases where the model has the potential for overfitting, the model will show weak separating power when used for other data, for example the GINI value of the validation data decreases compared to the GINI value of the development data.

Matrix on the overall durability of debtors on car loans which are predicted to make good payments in terms of smooth payments (Good) and late payments of more than 2 months (Bad) can be seen in table 4.

Table 4. Confusion Matrix for Data Jan 2012 – Dec 2014

Pre-Scoring		**Actual**		**Grand Total**
		Good	**Bad (60+)**	
Predicted	**Green**	102,298	5,408	107,706
	Not Green	16,789	10,760	27,549
Grand Total		119,087	16,168	135,255
Scoring		**Actual**		**Grand Total**
		Good	**Bad**	
Model	Excelent	21492	236	21728
	Good	41832	1490	43322
	Normal	38974	3682	42656
	Warning	9833	2281	12114
	Bad	6956	8479	15435
	Grand Total	119087	16168	135255

Source: Analysis output results, 2022

Overall, in the percentage of smooth applications, there were 88.32% of applications, and 11.68% of applications were late for more than 2 months. According to Hermanto (2006), A debtor is said to be bad credit (stuck-late) if the debtor does not make installment payments within 2 months after the debtor's maturity (month-end cut off). Judging from one of the variables Historical Payment Never Past Due 3 for First

Installments, those who have Past Due have a 44.79 percent chance that they will default. Conversely, if the first 3 installments are never past due, then the probability of default is 7.91%. This shows that there is a risk of default for debtors who have been past due compared to never past due. The cox proportional hazard regression model with the semiparametric method can be used to identify factors that influence debtor resilience.

CONCLUSIONS

From the results of the analysis of determining the model with cox proportional hazard regression, it can be concluded that several things are related to the overall durability of debtors on car loans, on average, most debtors are fluent in making payments compared to those who are late paying. Based on one of the variables, namely Historical Payment Never Past Due 3 First Installments, having Past Due indicates there is a risk of default for debtors who have never past due compared to never past due.

The cox proportional hazard regression model with the semiparametric method can be used to identify factors that affect debtor resilience.

RECOMMENDATIONS

Based on the results of the analysis conducted on the 25 observed variables, there is one variable that does not significantly influence debtor resilience. In this research, only 1 research object was carried out, namely car loans. For this reason, it is necessary to analyze other types of credit for goods such as houses, etc. Also, it is necessary to carry out further analysis regarding the findings of the problem of debtor endurance during the Covid-19 pandemic, whether the age factor is also an obstacle to debtor endurance.

REFERENCES

Amurwani Jatu Herlina. 2012. Analisis Regresi Cox Proporsional Dengan Hazard Dasar Weibull Pada Data Tersensor Tipe II. Surabaya. SKRIPSI: FSAINTEK UNAIR

Angkat & Noeryanti. (2018). Analisis Survival dengan Model Regresi Cox Proportional Hazard Terhadap Kesembuhan Tubercolosis di RSU PKU Muhammadiyah Gamping. Jurnal. Institut Sains dan Teknologi AKPRIND Yogyakarta.

Aini Isna Nur. 2011. Extended Cox Model untuk time-independent covariateyang tidak memenuhi asumsi proportional hazard pada model cox proportional hazard. Depok. SKRIPSI: FMIPA UI.

Faisal, A. R., Bustan, M. N., & Annas, S. (2020). Analisis Survival dengan Pemodelan Regresi Cox Proportional Hazard Menggunakan Pendekatan Bayesian (Studi Kasus: Pasien Rawat Inap Penderita Demam Tifoid di RSUD Haji Makassar). Journal of Statistics and Its Application on Teaching and Research Vol. 2 No. 2, 62-70.

Hermanto. 2006. Faktor-faktor Kredit Macet pada PD. BPR BKK. Ungaran Kabupaten Semarang. Tugas Akhir. Semarang: Fakultas Ekonomi.

Hogg, R. V., McKean, J. W., & Craig, A. T. (2013). Introduction to Mathematical Statistics. USA: 7rd Ed. USA: Pearson.

Kleinbaum, David G. & Klein, Mitchel. (2005). Survival Analysis a Self-Learning Text. 2nd. ed. New York: Springer.

Kleinbaum, D. G., & Klein, M. (2012). Survival Analysis: A Self Learning Text. London: Springer.

Lee, E. T. & Wang, J. W., 2003. Statistical Method for Survival Data Analysis. Third ed. Oklahoma: John Wiley & Sons, Inc., Hoboken, New Jersey.

Lian, P. Z, C.M. Lee, and K.H. Chung. 2002. Textural and Physical Properties of Acid-Induced and Potassium-Substituted Low-Sodium Surimi Gel, J. Food Sci,67(1)109-112.

Pahlevi, M. R., Mustafid, & Wuryandari, T. (2016). MODEL REGRESI COX STRATIFIED PADA DATA KETAHANAN. JURNAL GAUSSIAN, Volume 5, Nomor 3, 455-464.

Setiawan, A., Siregar, H., & Maulana, A. T. (2014). Model Behavioural Scoring pada Bisnis Pembiayaan Konsumen Menggunakan Analisis Daya Tahan (Studi Kasus: PT Karya Besar Cabang Bogor). Jurnal Manajemen Teknologi, 40-52.

Silmi AL, Sudarno, Kartikasari P. 2020. Perbandingan Model Regresi Kegagalan Proporsional Dari Cox Menggunakan Metode *Efron* Dan *Exact*. JURNAL GAUSSIAN, Volume 9, Nomor 4, Tahun 2020, Halaman 474 – 485. ISSN: 2339-2541.

Todaro, M.P. dan Smith, S.C. 2006. Pembangunan Ekonomi. Jilid I Edisi Kesembilan. Haris Munandar (penerjemah). Erlangga, Jakarta

William J, Stevenson. 2007. Manajemen Operasi. Edisi 9. McGrawHill/Irwin.