Extended Cox Model on Duration Taken to Release Cargo at Kenyan Border Entry Point: 
A Case Study of Malaba Osbp

Kennedy Ekeya Emoru*, Joel Cheruiyot Chelule, 
Herbert Orango Imboga, Ayubu Okango Anapapa

Jomo Kenyatta University of Agriculture and Technology, Department of Statistics and Actuarial Science, Nairobi, Kenya

*Corresponding author: kenniedykeya@gmail.com

Received September 27, 2020; Revised October 29, 2020; Accepted November 06, 2020

Abstract Duration taken to release cargo is one of World Customs Organisation’s most significant instruments of measuring trade facilitation. It is a way of reviewing clearance procedure by measuring the time taken to release cargo at border entry points. This research therefore aimed to model clearance time using the Extended Cox model, which is a modification of the proportional hazard Cox model in which proportional hazard assumptions are not met, and tested the practicability of the model and compared the performance of Extended Cox model with the other conventional methods using secondary data on time clearance that was obtained from the KRA, Malaba OSBP, for the period 1st January 2015 to 31st July 2019. The variables in the study were both time depend (time taken from registration to release) and time independent (customs regimes, origin, destination, customs value and description of goods). The variables used in the research project were tested for proportional hazard assumptions. Extended Cox model was the most suitable model for analysis duration taken to release cargo data has it was found to have a lower AIC and BIC as compared to conventional cox model. Therefore, customs administration should embrace the use of Extended Cox model to analysis duration taken to release cargo data as it is statistically significant and give efficient results that are based on proper statistical methods. It is recommended that further studies to be done on a large data set that encompasses both entry and exit cargo in all the Kenyan border points.

Keywords: cox model, duration, cargo, kenyan border, entry points

Cite This Article: Kennedy Ekeya Emoru, Joel Cheruiyot Chelule, Herbert Orango Imboga, and Ayubu Okango Anapapa, “Extended Cox Model on Duration Taken to Release Cargo at Kenyan Border Entry Point: A Case Study of Malaba Osbp.” American Journal of Applied Mathematics and Statistics, vol. 8, no. 3 (2020): 96-102. doi: 10.12691/ajams-8-3-3.

1. Introduction

The time taken to process release of cargo is a strategic, internationally recognised means for measuring the actual time required for release/clearance of cargo from arrival until the physical release of cargo, with a view to bench mark with other best practices, finding bottlenecks in trade flows process and taking the necessary measures to improve the effectiveness and efficiency of the border operations. In reference [7] defined Time taken to release cargo study as a systematic and standard method to measure the average time taken to release cargoes and for each step or intervention in a border procedure. It measures relevant aspects of the effectiveness of operational procedures that are carried out by Customs and other regulatory actors in the standard processing of imports, exports and in transit movements. Survival analysis is a series of useful statistical practice employed to analyse the timing of events/situations. It illustrates that the survival analysis is a statistical tool that models the duration of time until an event takes place. It is used with more attention on duration of individual survival time until an event occurs that is time to event. Therefore, it is the time until a certain individual event occurs i.e. event data. Also survival analysis can be illustrated as a compilation of statistical methods for data analysis, for which the result variable of concern is time until an event occurs. It is the review of duration between entrance into observed event and a subsequent event.

2. Literature Review

2.1. Review of Previous Research on Extended Cox Model

According to [5] on the study about modelling of credit risks by use of survival approaches where they fitted different survival models into the Kenyan credit data and systematically compare the performance of models. The research attempted to evaluate and give recommendations on the best survival model in credit risk context in relation
to the Kenyan data. The methodology involved modelling Cox Proportional hazard model and its extensions, that is, penalized splines and frailty model as well as the blend of cure and no-cure model to live data set from Kenya. And they assess which is the most effective model in analysing credit risk. The research concluded that penalized spline, they assess which is the most effective model in analysing credit data as compared to Cox PH model.

In their research [2], the research study aims to model the effects of several variables on survival duration of breast cancer patients. They tested the variables for PH assumptions to determine the variables that violate the assumptions hence can be used for Extended Cox model. The result was that extended Cox model malignancy degree and leukocyte level violates the proportional hazard assumptions therefore they added time function on the variables. Therefore, Extended Cox model was used for modelling the data and it was found out that patients with high leukocyte level have high rate of survival time. Hence, extended cox model is an optimal method of modelling breast cancers study.

2.2. Review of Comparison of Survival Models

In the study by [7] they identified that the key tenacity of Extended Cox model is its capability to used or accommodate variables that are affected with time. The research focused on modelling customer retention in telecommunication industry by utilizing survival analysis methods with an aim of providing solutions and information about customer retention in telecommunication industry in Rwanda. Therefore, Extended Cox model was chosen as the correct model and conventional Cox model as the incorrect model. The study found out that the median of time independent covariate as larger than the median of time dependent variables and also they found that Cox model had higher bias than Extended Cox model therefore they concluded that Extended Cox model has higher performance hence the its gives the best account on how customer retention data is model.

3. Research Methodology

3.1. Data Sources

The source of data was from the information technology department at KRA. The data ranges from 31st July 2019 to 1st January 2015. It only contained data for cargo cleared for goods from Uganda side of the border to Kenya.

3.2. Sampling and Analysis

The sampling size was data for cargo clearance from 2015 to 2019. R Software version 4.0.0 was used for analysis of the data.

3.3. Basic Concepts on Survival Analysis

Survival technique is a statistical method that is normally utilized to model survival data. Basic concept of survival analysis involves PDF & CDF.

3.4. Review of Cox Model

Cox proportional model is a survival analysis model that examines the relationship between survival of an individual and the covariates utilized in the study. The generalized model was given below;

\[ h_i(t) = \exp(\beta_1 x_{i1} + \ldots + \beta_p x_{ip}) h_0(t) = h_0(t) e^{\beta x_{i1}} \]  (1)

3.5. Proportional Hazard Assumption

Proportional Hazard assumption states that the hazard ratio between to hazard functions should not change over time that is it should be constant over time. The below methods were used to test PH assumptions.

3.5.1. Graphical Method

Cox Proportional Hazard Function is achieved by the relationship between hazard function and survival function

\[ S(t, x_i) = S_0 \exp(\sum_{j=1}^p \beta_j x_{ij}) \]  (2)

Where \( X = (x_{i1}, \ldots, x_{ip}) \). This relationship was very helpful to help us identify situations where we may or may not have proportional hazards. The methodology involved plotting an estimated log (-log (survival)) against survival time for 2 groups then we observed the parallel curves if the hazard was proportional. If from the observation the hazard was proportional, then the proportional hazard assumption was met otherwise they ae not met.

3.5.2. Tests Based on the Schoenfeld Residuals

Schoenfeld residual is one of the statistical tests of the proportional hazards assumption. So this test was accomplished by finding the correlation between the Schoenfeld residuals for a particular covariate and the ranked survival time. The null hypothesis was that the correlation between the Schoenfeld residuals and the ranked survival time was zero. Rejection of null hypothesis concludes that PH assumption was violated.

3.6. Extended Cox Model

Generalised Extended Cox model was given as;

\[ h(t, X) = h_0(t) \exp[\sum_{i=1}^p \beta_i X_i + \sum_{j=1}^P \delta_j X^j(t)] \]  (3)

\( i=1, j=1 \). Taking \( X(t) = (X_1, X_2, \ldots, X_p; X_1, X_2, \ldots, X_p) \) is the all collection of predictors at time t. The extended Cox model contains the following;

\( h_0(t) = \) baseline hazard function \( X^i = \) time independent prognostic variables \( X(t) = \) time varying prognostic variables. The hazard ratio for the extended Cox model is then given as;

\[ HR(t) = \exp \left[ \sum_{i=1}^p \delta X^i \left[ X^i(t) - X_i \right] + \sum_{j=1}^P \delta_j \left[ X^j(t) - X^j_1 \right] \right] \]  (4)

The extended Cox model for duration taken to release cargo utilized in this study is given by;

\[ h(t, X) = h_0(t) \exp[\sum_{i=1}^5 \beta_i X_i + \alpha_6 X_6] \]  (5)

\( i=1 \) then \( h_0 = \) baseline hazard function, \( X_i = (x_1, x_2, x_3, x_4, x_5) = \) the independent variables

The regression coefficients in the extended Cox model are estimated using a maximum likelihood (ML) procedure.
3.7. Test of Performance of the Models

3.7.1. Akaike Information Criterion (AIC)

The models with lower AIC has higher performance and model with higher AIC has a lower performance.

3.7.2. Bayesian Information Criterion (BIC)

The model with lower BIC has higher performance but model with higher BIC has lower performance.

4. Results and Analysis

4.1. Descriptive Analysis

| Variables | Coef | exp(coef) | se(coef) | Z | Pr(>|z|) |
|-----------|------|-----------|----------|---|---------|
| Regime E2 | 0.418369 | 1.519481 | 0.716538 | 0.584 | 0.559304 |
| R3 | -1.865719 | 0.154785 | 0.547667 | -3.407 | 0.000658 |
| S6 | -0.520366 | 0.594303 | 0.715858 | -0.727 | 0.467280 |
| T8 | -0.064597 | 0.937126 | 0.105564 | -0.615 | 0.538803 |
| Origin | Rwanda | 0.347171 | 1.415059 | 0.141964 | 2.445 | 0.014465 |
| Kenya | -1.106165 | 0.330825 | 0.293913 | -3.764 | 0.000167 |
| S.Sudan | -0.263526 | 0.768338 | 0.113825 | -2.315 | 0.020662 |
| Uganda | 0.250654 | 1.284865 | 0.084299 | 2.973 | 0.002945 |
| Product | Coffee | 0.154579 | 1.167167 | 0.153202 | 1.009 | 0.312981 |
| Non Agri | -0.034375 | 0.966209 | 0.161483 | -0.213 | 0.831428 |
| Other Agri | -0.176965 | 0.837809 | 0.149138 | -1.87 | 0.032591 |
| Tea | -0.002282 | 0.997720 | 0.191914 | -0.012 | 0.990512 |
| Teak Logs | -0.239084 | 0.787349 | 0.169199 | -1.413 | 0.157646 |
| Customs Value | 0.020546 | 1.020759 | 0.064428 | 0.319 | 0.749802 |
| Destination America | 0.107154 | 1.113106 | 0.186587 | 0.574 | 0.565775 |
| M. East | -0.067283 | 0.934931 | 0.162579 | -0.414 | 0.678985 |
| Europe | 0.439490 | 1.551915 | 0.139638 | 3.147 | 0.001648 |
| Asia | 0.421464 | 1.524191 | 0.142385 | 2.960 | 0.003076 |

4.2. Cox Model

Cox Model was fitted to model the time for clearance of cargo from Malaba border of Kenya. The results as shown in Table 2 suggested that the data had 1957 observations and 1957 events. This shows that all the cargoes were released hence no censored data. The findings show that Regime variable had R3 with statistically significant Wald Statistic value (p-value=0.006<0.05) and others factors such as E2, S6 and T8 regimes had statistically insignificant Wald Statistic value (p-values>0.05). This show that the variable regime has R3 with statistically significant coefficients and others factors with statistically insignificant coefficients including E2, S6 and T8 regimes. The results further show that those cargoes with E2 regime had positive beta coefficients. This indicates that those cargoes with E2 regime are more likely to be released or cleared earlier as compared to those with E1 regime. Moreover, the hazard ratio of those cargoes with E2 regime was found to be greater than one suggesting that those with this regime increases the hazard of being cleared earlier by a factor of 1.5195 (52%) times as compared to those with E1 regime. However, the beta coefficients of those cargoes with R3, S6 and T8 were found to be negative. This shows that those cargoes with R3, S6 and T8 regimes are less likely to be released or cleared earlier as compared to those with E1 regime. Furthermore, the hazard ratio for those cargoes with R3, S6 and T8 regimes were found to be less than 1. This indicates that those with R3, S6 and T8 regimes reduces the hazard of being cleared or released earlier by a factor of 0.1548 (85%), 0.5943 (41%) and 0.94 (7%) times respectively as compared to those with E1 regime. The results further show that the Wald Statistic for the country of origin variable was statistically significant (p-values < 0.05). This indicates that country of origin variable has statistically significant coefficients.

4.3. Test of Proportional Hazard Assumption

Table 3. P-values

| Variables | Chiq | DF | p |
|-----------|-----|----|---|
| Regime | 10.65 | 5 | 0.0588 |
| Origin | 13.96 | 4 | 0.0074 |
| Product | 12.01 | 5 | 0.0346 |
| Destination | 54.80 | 4 | 3.6e-11 |
| Value | 4.96 | 1 | 0.0259 |
| Global | 99.55 | 19 | 6.5e-13 |

We reject H0 since p-value<0.05.
From the above comparison only regime 0.0588 > 0.05 therefore we fail to reject H0 and concluded that PH assumption are met. P-value for origin was 0.0074 < 0.05 hence we rejected H0 and concluded that PH assumptions were not met. P-value for Product was 0.0346 < 0.05 hence we rejected H0 and concluded that PH assumptions were not met. P-value for destinations was 3.6e-11 < 0.05 hence we rejected H0 and concluded that PH assumptions were not met. P-value for customs value was 0.0259 < 0.05 hence we rejected H0 and concluded that PH assumptions were not met. The Global p-value was 6.5e-13 < 0.05. Therefore, the PH assumptions are not met by the other variables except the variable regime which met the assumptions.

4.4. Graphical Method

Graphical method of testing PH assumptions involved plotting the variables. The plots below show separate plots for all variables, Regime, Origin, products, Destination and Customs value. The plot (residuals line) is surrounded by lower and upper limits of 95% confidence interval.
The results as shown in Figure 1, Figure 2, Figure 3, Figure 4 and Figure 5 suggested that the lines h=0 was not entirely within the lower and upper limits of 95% confidence interval except for the Regime variable. Therefore, we concluded that all other variables except Regime violated the PH assumption because both the statistical test method and Schoenfeld residual graphical method shows that the PH assumption are not met except regime variable which met the PH assumptions.

4.5. Extended Cox Model

Extended Cox Model was performed with the extensions of Cox model to incorporate variable which vary over time. One of the simplest extensions is a step function for $\beta(t)$, i.e., different coefficients over different time intervals. As such, in this study the Extended Cox model used comprises of six variables including release time as the dependent variable and independent variables regime, origin, destination, customs value and products. The total number of events was 1957 hence all the data were released therefore no censored. The results as shown in Table 4 below suggested that the p-values for the three alternative tests including the likelihood-ratio test, Wald test and the Score Logrank statistics for the overall model significance was less than 0.05 indicating that the model was statistically significant. Concordance level was 0.602 that is 60%. Therefore, the Extended Cox model was appropriate for modelling this data in analysing duration taken to release cargo. Moreover, the findings as presented in Table 4 below still indicated that the Wald Statistic value for the regime variable had some factors which were not statistically significant (p-values>0.05) including E2, S6 and T8 regimes hence these variables had statistically insignificant coefficients. However, R3 regime was found to be statistically significant (p-value<0.05) hence has statistically insignificant coefficients. Notably, cargoes originating from Rwanda had statistically insignificant (p-value>0.05) Wald Statistic value hence has statistically insignificant coefficients while cargoes originating from Rwanda, South Sudan and Uganda had statistically significant (p-values<0.05) Wald Statistics values indicating that they have highly statistically significant coefficients. Notably, cargoes from Kenya and South Sudan had negative beta coefficients hence these cargoes were less likely to be released or cleared earlier as compared to those originating from DRC. Moreover, the beta coefficients of cargoes made up of coffee and tea products were found to be positive indicating that these cargoes have higher chances of being cleared or released earlier as compared to those made up of animal products. As such, the hazard ratios of those cargoes made up of coffee and tea products were found to be greater than one indicating that these cargoes increases the hazard of clearance by a factor of 1.5575 and 1.2847 times respectively as compared to those originated from DRC.

Accordingly, the beta coefficients of cargoes originating from Rwanda and Uganda were found to be positive suggesting that these cargoes were more likely to be released or cleared earlier as compared to those originating from DRC. As such, the hazard ratios of those cargoes from Kenya and South Sudan were found to be less than one showing that these cargoes decreases the hazard of clearance by a factor of 0.4563 and 0.4116 times respectively as compared to those originating from DRC. Moreover, the beta coefficients of cargoes originating from Rwanda and Uganda were found to be less than one indicating that these cargoes decreases the hazard of clearance by a factor of 0.1515, 0.5862 and 0.9458 times respectively as compared to those with E1 regime.

Moreover, the Wald Statistics values for products were found to be statistically insignificant (p-values>0.05) hence have statistically insignificant coefficients. The beta coefficients for cargoes made up of coffee and tea products were found to be positive. This shows that these cargoes have higher chances of being cleared or released earlier as compared to those transporting animal products. As such, the hazard ratios of those cargoes made up of coffee and tea products were found to be greater than one. This suggested that these cargoes increased the hazard of being cleared or released earlier by a factor of 1.1719 and 1.0336 times respectively as compared to those made up of animal products. However, the beta coefficients of those cargoes made up of teak logs, non-agricultural and other agricultural products were found to be less than one indicating that these cargoes decreases the hazard of clearance by a factor of 0.9951, 0.8594 and 0.8008 times respectively as compared to those made up of animal products.
Furthermore, the Wald Statistic value for custom value variable was not statistically significant (p-value>0.05) suggesting that this variable had statistically insignificant coefficients. The beta coefficient for cargoes with custom value below 3 million was found to be positive meaning that these cargoes were more likely to be cleared or released earlier as compared to those cargoes with custom value above 3 million. The hazard ratios of those cargoes with custom value below 3 million was found to be positive meaning that these cargoes increase the hazard of clearance by a factor of 1.0132 times as compared to those cargoes with custom value above 3 million. Therefore, the Wald Statistics value of those cargoes with more than 3 million custom value. The beta coefficient for cargoes with custom value below 3 million was found to be positive meaning that these cargoes are more likely to be cleared or released earlier as compared to those cargoes heading to Africa. In addition, the Wald Statistics value of those cargoes with Middle East and Asia as their destinations were not statistically significant (p-values >0.05) hence have statistically insignificant coefficients. The beta coefficient of those cargoes with Middle East as their destination was found to be negative indicating that these argoes are less likely to be cleared or released earlier as compared to those with Africa as their destinations. As such, the hazard ratio of these cargoes was found to be less than one meaning that these cargoes decrease the hazard of being cleared or released earlier by a factor of 0.4799 (52%) times as compared those with Africa as their destinations. Moreover, the beta coefficients of those cargoes heading to America, Asia and Europe were found to be positive suggesting that these cargoes are more likely to be cleared or released earlier as compared to those heading to Africa. In addition, the hazard ratios of these cargoes were found to be greater than one meaning that these cargoes heading to America, Asia and Europe increase the hazard of clearance by a factor of 1.5106 (51%), 1.5996 (60%) and 1.8149 (81%) times respectively as compared to those heading to Africa.

### Table 4. Extended Cox model findings

| Variables | Coef | exp(coef) | se(coef) | Z | Pr(>|z|) |
|-----------|------|-----------|----------|---|---------|
| **Regime** | | | | | |
| E2 | 0.418369 | 1.478513 | 0.717401 | 0.545 | 0.585702 |
| R3 | -0.187227 | 0.151491 | 0.556873 | -3.89 | 0.000702 |
| S6 | -0.534100 | 0.586197 | 0.76114 | -1.04 | 0.445570 |
| T8 | -0.055680 | 0.945842 | 0.105803 | -0.52 | 0.598705 |
| **Origin** | | | | | |
| Rwanda | 0.443078 | 1.557493 | 0.164459 | 2.69 | 0.007057 |
| Kenya | -0.784707 | 0.456253 | 0.588358 | -1.33 | 0.182294 |
| S.Sudan | -0.887812 | 0.411555 | 0.152768 | -5.81 | 0.049719 |
| Uganda | 0.250654 | 1.28465 | 0.084299 | 2.97 | 0.002945 |
| **Product** | | | | | |
| Coffee | 0.15861 | 1.117189 | 0.153359 | 1.03 | 0.301002 |
| Non Agri | -0.004078 | 0.995136 | 0.161686 | -0.03 | 0.975942 |
| OtherAgri | -0.151530 | 0.859392 | 0.148753 | -0.10 | 0.308360 |
| **Customs** | | | | | |
| Value | 0.013067 | 1.013152 | 0.064759 | 0.20 | 0.840093 |
| **Destination** | | | | | |
| America | 0.412478 | 1.510556 | 0.237890 | 1.73 | 0.082936 |
| M. East | -0.734135 | 0.479290 | 0.236960 | -3.18 | 0.001461 |
| Europe | 0.596003 | 1.814850 | 0.240231 | 2.45 | 0.13103 |
| Asia | 0.469753 | 1.599599 | 0.163335 | 2.87 | 0.004027 |

Concordance= 0.602 (se = 0.008 )
Likelihoodratio test= 289.6 on 35 df, p=<2e-16
Wald test = 247.7 on 35 df, p=<2e-16
Score (logrank) test = 269.8 on 35 df, p=<2e-16.

4.6. Test of Performance of the Models

In order to compare the appropriate model for analysing duration time for the release or clearance of cargoes, Extended Cox model and Conventional Cox model was compared using use Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Extended Cox model had the least AIC (25540.71<25611) value as compared to Cox model i.e. (25540.71<25611). Also, the results further show that BIC for Extended Cox model was smaller as compared to BIC for Conventional Cox model i.e (25705.98<25717.01) hence Extended Cox model was more efficient in analysing data for duration taken to release cargo.

5. Conclusions and Recommendations

5.1. Conclusions

The aim of the study was to analyse duration taken to release cargo data by modelling both the Extended Cox model and the Cox model. The results of the Cox model were tested for the Proportional Hazard assumption where the findings show that all the variables except regime violated this assumption hence Cox model was not appropriate for this dataset. Test of performance methods including AIC and BIC was used in comparing the fitted Extended Cox model and the Conventional Cox model with an aim of presenting scientist with a better and statistically efficient way of analysis time taken to release cargo because it encompassed variable which change with time. The variable regime was the only variable which met the proportional hazard assumptions hence it was not
significant in analysis duration taken to release cargo with
the Extended Cox model. The other variables violated
the proportional hazard assumption hence they were
statistically significant in analysis of duration taken to
release cargo data by use of Extended Cox model. We
concluded that BIC and AIC were the most suitable
method of measuring the performance of models when
using data that are made of factors that is data that are not
numbers. The results show that cargoes with R3 regime,
those heading to Middle East and Asia and those
originating from Rwanda, South Sudan and Uganda
significantly influences duration taken to release cargo. In
 nutshell, Customs administration, policy makers, business
stakeholders, governments and World customs organisation
should adapt the use of Extended Cox model in analysing
duration taken to release cargo.

5.2. Recommendations

From this study we recommend that further study
should be done incorporating data from cargo heading
from Kenya to Great Lakes region through Malaba OSBP.
Also, research should be done on data for all other Kenya
border points. Also, we recommended more research on
other intervening variables at the border that were not
included as variable in our research such as weight,
container number, truck number, clearing and forwarding
agents, consignor and consignee details. Moreover, further
research should use other parametric survival methods
such as Weibull and exponential models.

References

[1] Bellera, C. A., MacGrogan, G., Debleed, M., de Lara, C. T.,
Brouste, V., & Mathoulin-Pélissier, S.. Variables with time-varying
effects and the cox model: some statistical concepts illustrated
with a prognostic factor study in breast cancer. BMC medical
research methodology, 2010, 10(1), 20.

[2] Husain, H., Thamrin, S. A., Tahir, S., Mukhlisin, A., & Apriani, M.
M.. The application of extended cox proportional hazard method
for estimating survival time of breast cancer. In Journal of physics:
Conference series 2018 (Vol, 979)

[3] Klein, J. P., & Moeschberger, M. L.. Survival analysis: techniques
for censored and truncated data. Springer Science & Business
Media 2006.

[4] Kleinbaum, D. G., & Klein, M. Extension of the cox proportional
hazards model for time-dependent variables. In Survival analysis
Springer 2012(pp. 241-288).

[5] Mungasi, S., & Odhiambo, C. Comparison of survival analysis
approaches to modelling credit risks. American Journal of
Theoretical and Applied Statistics, 2019 8(2), 39-46.

[6] Njuguna, A. Modelling survival in hiv/aids treatment: Cox
proportional hazard versus accelerated failure time models.
(Unpublished doctoral dissertation). University of Nairobi 2013.

[7] Shingo Matsuda. The Time Release Study as a performance
measurement tool for a supply chain and an international corridor.
WCO Journal (2011).

[8] Orwa, A. O., Orwa, G. O., & Odhiambo, R.. Deriving penalized
splines for estimation of time varying effects in survival data
American journal of Science 2014.