A non-proportional hazards model with time-dependent variables for under-five mortality in Indonesia

Alfensi Faruk\textsuperscript{1,*}, Erlan Syahputra\textsuperscript{1}, Endro Setyo Cahyono\textsuperscript{1}, Sari Mutiara Aisyah\textsuperscript{2}

\textsuperscript{1}Department of Mathematics, Faculty of Mathematics and Natural Sciences, Sriwijaya University
\textsuperscript{2}Department of International Relations, Faculty of Social and Political Sciences, Sriwijaya University

Email: *alfensifaruk@unsri.ac.id

Abstract. Cox proportional hazards (PH) is known as the most popular model for the analysis of multivariate survival data. The main assumption of the model is that the hazard ratio among any two individuals in the population is constant over time. The violation of this assumption, however, may cause serious issues such as overestimation or underestimation of hazard risks and reducing the power of related statistical tests. The main objective of this research, therefore, is to extend the Cox PH model by adding time-dependent variables into the model to cope with the presence of non-PH. The proposed model, moreover, is applied to under-five mortality data based on Indonesian Demographic and Health Survey (IDHS) 2012. The results showed that the Akaike’s Information Criterion (AIC) and Bayesian Information Criterion (BIC) values of the extended Cox model with time-dependent variables are much lower than the Cox PH model. This means that the extended Cox model can increase significantly the goodness-of-fit of the Cox PH model to under-five mortality data in Indonesia.

1. Introduction
Under-five mortality (U5M) is still a serious problem faced by various nations including Indonesia as one of the developing countries in South East Asia. Although U5M rate in Indonesia has declined drastically in recent years, it is still huge and does not achieve yet the target from the Sustainable Development Goals (SDGs). Therefore, in order to achieve this target, the first effort that should be conducted is to analyze the potential risk factors affecting U5M.

The works about the determination of any influential factors affecting U5M rates have been conducted by various authors for many years. Mosley and Chen [1] have proposed socio-economic factors as the determinant of childhood mortality. Their theoretical framework has widely become the basis for many other childhood mortality studies[2,3]. Furthermore, most of the works employed survival analysis approaches to model the U5M. This is because the outcome variable of U5M is the duration until the occurrence of an event (time until death) or survival time.

Cox proportional hazards (PH) is the most popular survival model in order to examine the effects of the potential factors to survival time. Cox PH model has been applied to various demographic and public health problems, such as its application to the birth interval of the first child [4,6] and childhood mortality [2,3]. The model assumes that the hazard ratio of two individuals with the same characteristics is constant over the given period [7]. This PH assumption is usually ignored in the Cox PH model for U5M rates and many works did not check whether this assumption is satisfied by the
Using the U5M dataset. Furthermore, the violation of the PH assumption may bias the final estimation results and decrease the power of the statistical tests used. In order to overcome the violation of the PH assumption in the dataset, one can extend the Cox PH model by assuming any independent variable, which does not satisfy the PH assumption, depends on the function of time. This extension is widely used in practice and known as Cox extended model [8].

In this work, the potential factors affecting the U5M rates in Indonesia are investigated by using the dataset obtained from the results of the most recent Indonesian Demographic and Health Survey (IDHS). The considered independent variables consist of sixteen socio-economic factors based on the theoretical childhood mortality framework by Mosley and Chen [1]. Therefore, the main aims of this research are to check the PH assumption in the U5M and to apply the Cox extended model for time-dependent variables if the assumption is violated in the dataset. The results of this works will provide proper proof of the Cox extended model as a better option than the Cox PH model when U5M data violate the PH assumption.

2. Materials and methods
This work used secondary data based on the 2012 IDHS which were collected between May 2012 and June 2012. This survey used questionnaires from women aged 15-49 as the survey samples to record any information regarding various demographic and health issues.

To achieve the aims of this research, there are several steps that have been conducted. In the first step, the collected raw data were processed so that it can be analyzed by using proposed models. Then, the PH assumption was checked by using both the graphical method and statistical test. After that, the Cox model and the extended Cox model were employed to the dataset. Next, the performance of the proposed model was assessed by using Akaike’s information criterion (AIC) and Bayesian information criterion (BIC). Finally, the backward stepwise variable selection method was applied in order to obtain the best model for U5M data in Indonesia.

3. Basic concepts
3.1. Cox PH model
In multivariate survival analysis, Cox PH model has become the most popular survival method in order to examine the effect of the potential characteristics toward the survival times which are represented by the hazard function. Formally, the relation between the hazard function and the explanatory variables in the Cox PH model can be written as follows

\[ h(t) = h_0(t) \exp(\beta^T X) \]

where \( t \) is the survival time, \( h(t) \) is the hazard function at time \( t \), \( h_0(t) \) is the baseline hazard function at time \( t \) and \( \beta \) is a vector of regression coefficients with respect to the independent variables in vector \( X \). The parameters in the Cox PH model can be estimated by using partial likelihood estimation (PLE) method. This model assumes that the ratio of the hazard rates between two individuals with similar covariate is constant over time (PH assumption).

3.2. Cox extended model for time-dependent variables
The PH assumption in the Cox PH model can be checked by several approaches, namely graphical approach, time-dependent approach and goodness-of-fit test approach (Klein and Kleinbaum). If PH assumption is not met for one or more of independent variables in the model, the Cox PH model is not appropriate anymore to be used in the dataset. Furthermore, one should use other alternative methods such as Cox extended model for time-dependent variables and a stratified Cox model.

In an extended Cox, the Cox PH model incorporates the time-dependent predictor variables in addition to time-independent predictors. The Cox extended model can be formulated as follows

\[ h(t) = h_0(t) \exp \left[ \sum_{i=1}^{p_1} \beta_i X_i + \sum_{j=1}^{p_2} \delta_j X_j(t) \right]. \]
where $X_i$, $i = 1,2, ..., p_1$ are time-independent variables, $X_j(t)$, $j = 1,2, ..., p_2$ are time-dependent variables and $\delta_j$ are the coefficients of $X_j(t)$. As with the Cox PH model, the regression coefficients in the extended Cox model can be estimated using the PLE procedure.

4. Results

4.1. Data description

In this study, there are 28242 ever-married women aged 15-49 from 33 provinces in Indonesia who are interviewed. The survey recorded any new birth of the respondents from July 2007 until the time of interview (between May 2007 and June 2012). If the newborn babies died between that period, their age became the survival time. Meanwhile, age at the time of interview of the babies became right censored data if they still alive beyond the time of interview. There are a total of 17305 newborn babies in the observation period given which contains 550 (3.18%) under-five who died and 16755 (96.82%) under-five child who survived until the end of the observation period. The response variable in this research is time until the death of under-five children. In Table 1, the first characteristic (status of the child) indicates the living status of the under-five children. The status of the child is not a predictor variable but all of the rest characteristics are the predictor variables of the model.

4.2. Assessing PH assumption

In order to show that the proposed model satisfies the PH assumption, one only needs to demonstrate at least one predictor which does not satisfy the assumption. Although there are three methods that can be used for this task, this research only employed two of them, namely graphical method and goodness-of-fit test. Another method, namely the Cox extended method, is not applied since it is used to overcome the existence of non-PH in this research.

4.2.1. Log cumulative hazard (LCH) plots. The procedure of the LCH plots is employed by plotting log(time) against log(− log(S(time))) for all categories in each predictor variables. Although this research basically applied LCH plots for all independent variables, only four of them are described in this paper. If a plot yields at least two curves which intercept each other, the PH assumption is violated in the related predictor variable. The LCH plots for four predictor variables based on the U5M dataset are shown in Figure1.

According to Figure1, there are two variables (pregnancy complications and time zone) which do not satisfy the PH assumption since there are at least two hazard curves in both variables that intercept each other. Therefore, based on the LCH plots, the PH assumption is not satisfied in the U5M data in Indonesia.

4.2.2. GOF test. The second method to be used for checking PH assumption is a GOF test based on Schoenfeld residuals. This test may allow the survival time to be transformed into various functional forms of time. However, this study only considers two forms of survival time transformations, namely Kaplan-Meier (KM) transformation and time ranking transformation. The test hypothesis is $H_0: \rho = 0$ (the predictor variable does not significantly correlate with time). The results of the GOF test for the model with no interaction between variable and time and model with time-dependent variables are summarized in Table 2 and Table 3, respectively.
| Characteristics                      | Symbol | Categories          | Quantity | Percentage (%) |
|--------------------------------------|--------|---------------------|----------|-----------------|
| Status of the child                 | state  | Alive               | 16755    | 96.82           |
|                                      |        | Dead                | 550      | 3.18            |
| Sex                                  | sex    | Male                | 8974     | 51.86           |
|                                      |        | Female              | 8331     | 48.14           |
| Mother’s educational level           | edu_f  | No Education        | 549      | 3.17            |
|                                      |        | Primary             | 5331     | 30.81           |
|                                      |        | Secondary & above   | 11425    | 66.02           |
| Father’s educational level           | edu_m  | No Education        | 382      | 2.21            |
|                                      |        | Primary             | 5346     | 30.89           |
|                                      |        | Secondary & above   | 11577    | 66.90           |
| Mother’s employment status           | job_f  | No                  | 7893     | 45.61           |
|                                      |        | Yes                 | 9412     | 54.39           |
| Father’s employment status           | job_m  | No                  | 314      | 1.81            |
|                                      |        | Yes                 | 16991    | 98.19           |
| Household’s wealth index             | wealth | Poor                | 8696     | 50.25           |
|                                      |        | Middle              | 3123     | 18.05           |
|                                      |        | Above               | 5486     | 31.70           |
| Place of residence                   | place  | Rural               | 9440     | 54.55           |
|                                      |        | Urban               | 7865     | 45.45           |
| Time zone                            | time   | West                | 9979     | 57.67           |
|                                      |        | Middle              | 5344     | 30.88           |
|                                      |        | East                | 1982     | 11.45           |
| Mother’s age at giving birth         | age    | Unsafe              | 3805     | 21.99           |
|                                      |        | Safe                | 13500    | 78.01           |
| Mother’s age at the first marriage   | age_m  | Unsafe              | 7984     | 46.14           |
|                                      |        | Safe                | 9321     | 53.86           |
| Mother’s parity                      | par    | Low                 | 13815    | 79.83           |
|                                      |        | High                | 3490     | 20.17           |
| Pregnancy complications              | complex| No                  | 15445    | 89.25           |
|                                      |        | Yes                 | 1860     | 10.75           |
| Household’s water resources          | water  | Bad                 | 5384     | 31.11           |
|                                      |        | Good                | 11921    | 68.89           |
| Household’s sanitation               | sanit  | Bad                 | 6319     | 36.52           |
|                                      |        | Good                | 10986    | 63.48           |
| Type of place of giving birth        | salin  | Nonmedical          | 7759     | 44.84           |
|                                      |        | Medical             | 9546     | 55.16           |
| Person who help of giving birth      | p_salin| Nonmedical          | 3565     | 20.60           |
|                                      |        | Medical             | 13740    | 79.40           |
The individual tests in Table 2 show that there are two predictor variables which reject the null hypothesis. In particular, the category time zone (east) is significantly correlated with time at 5% level ($p$-value = 0.0268 for KM, $p$-value = 0.0164 for time ranking and $p$-value = 0.0382 for no transformation). Meanwhile, the pregnancy complications variable is significantly correlated with time at 5% level for two transformations ($p$-value = 0.0287 for KM and $p$-value = 0.0305 for time ranking) but no transformation category the hypothesis cannot be significantly rejected at 5% level ($p$-value = 0.0822). According to the overall test in Table 2, it can be concluded that the PH assumption is significantly rejected at 5% level for KM and time ranking with $p$-value = 0.0002 and $p$-value = 0.0003, respectively. However, the PH assumption is satisfied in the model for the GOF test with no transformation variable at 5% level ($p$-value = 0.0821).

**Figure 1.** The LCH plots for all predictor variables. The curves in (a) and (b) are not parallel meanwhile curves (c) and (d) are parallel.
Table 2. GOF test of the model with time-independent variables

| Predictor Variables | Categories                  | Kaplan-Meier | Time Ranking | No Transformation |
|---------------------|-----------------------------|--------------|--------------|-------------------|
| Sex                 | Male                        | -            | -            | -                 |
|                     | Female                      | 0.1291       | 0.1046       | 0.5452            |
| Mother’s educational level | No                         | -            | -            | -                 |
|                     | Primary                     | 0.9790       | 0.9300       | 0.9323            |
|                     | Secondary & above           | 0.6783       | 0.7524       | 0.7833            |
| Father’s educational level | No                         | -            | -            | -                 |
|                     | Primary                     | 0.0795       | 0.0618       | 0.0620            |
|                     | Secondary & above           | 0.0711       | 0.0483       | 0.0828            |
| Mother’s employment status | No                         | -            | -            | -                 |
|                     | Yes                         | 0.1994       | 0.2333       | 0.7826            |
| Father’s employment status | No                         | -            | -            | -                 |
|                     | Yes                         | 0.1641       | 0.1895       | 0.1658            |
| Household’s wealth index | Poor                      | -            | -            | -                 |
|                     | Middle                      | 0.3733       | 0.3792       | 0.4208            |
|                     | Above                       | 0.3829       | 0.6237       | 0.3829            |
| Place of residence | Rural                       | -            | -            | -                 |
|                     | Urban                       | 0.9045       | 0.9851       | 0.3414            |
| Time zone           | West                        | -            | -            | -                 |
|                     | Middle                      | 0.2069       | 0.2273       | 0.2181            |
|                     | East                        | 0.0268       | 0.0164\(^a\) | 0.0382\(^a\)     |
|                     |                             |              |              |                   |
| Mother’s age at giving birth | Unsafe                     | -            | -            | -                 |
|                     | Safe                        | 0.1335       | 0.1291       | 0.4019            |
| Mother’s age at the first marriage | Unsafe                   | -            | -            | -                 |
|                     | Safe                        | 0.9502       | 0.9334       | 0.1844            |
| Mother’s parity    | Low                         | -            | -            | -                 |
|                     | High                        | 0.2532       | 0.2481       | 0.4339            |
| Pregnancy complications | No                        | -            | -            | -                 |
|                     | Yes                         | 0.0287\(^a\) | 0.0305\(^a\) | 0.0822            |
| Household’s water resources | Bad                       | -            | -            | -                 |
|                     | Good                        | 0.4606       | 0.5642       | 0.5648            |
| Household’s sanitation | Bad                       | -            | -            | -                 |
|                     | Good                        | 0.9931       | 0.9580       | 0.6634            |
| Type of place of giving birth | Nonmedical                | -            | -            | -                 |
|                     | Medical                     | 0.2898       | 0.3202       | 0.9535            |
| Person who help of giving birth | Nonmedical              | -            | -            | -                 |
|                     | Medical                     | 0.6258       | 0.6542       | 0.4328            |
| Global test        |                             | 0.0002\(^a\) | 0.0003\(^a\) | 0.0821            |

\(^a\) Significant at 5% level
\(^b\) Reference category

The different results are obtained for the GOF test of the model with time-dependent variables. Table 3 shows that there is no evidence to significantly reject the null hypothesis including for two time-dependent variables (time zone * time and pregnancy complications * time) at 5% level of significance. Consequently, the null hypothesis in global test for the model with time-dependent variables cannot significantly rejected at 5% level (\(p\)-value= 0.0609, \(p\)-value= 0.0903 and \(p\)-value= 0.571 for KM, time ranking and no transformation, respectively). This means that adding time-dependent predictor variables into the original model can effectively make the extended model to satisfy the PH assumption.
### Table 3. GOF test of the model with time-dependent variables

| Predictor Variables                        | Categories     | Kaplan-Meier | Time Ranking | No Transformation |
|--------------------------------------------|----------------|--------------|--------------|-------------------|
| Sex                                        | Male           | 0.9169       | 0.8546       | 0.783             |
|                                            | Female         |              |              |                   |
| Mother’s educational level                 | No             | 0.9730       | 0.9924       | 0.851             |
|                                            | Primary        | 0.6085       | 0.6126       | 0.580             |
|                                            | Secondary & above |            |              |                   |
| Father’s educational level                 | No             |              |              |                   |
|                                            | Primary        | 0.6733       | 0.6232       | 0.310             |
|                                            | Secondary & above |          |              | 0.321             |
| Mother’s employment status                 | No             |              |              |                   |
|                                            | Yes            | 0.3825       | 0.4196       | 0.814             |
| Father’s employment status                 | No             |              |              |                   |
|                                            | Yes            | 0.2546       | 0.2604       | 0.929             |
| Household’s wealth index                   | Poor           |              |              |                   |
|                                            | Middle         | 0.4234       | 0.4117       | 0.362             |
|                                            | Above          | 0.6862       | 0.7328       | 0.564             |
| Place of residence                         | Rural          |              |              |                   |
|                                            | Urban          | 0.9306       | 0.9830       | 0.538             |
| Time zone                                  | West           |              |              |                   |
|                                            | Middle         | 0.6586       | 0.7161       | 0.382             |
|                                            | East           | 0.5486       | 0.4634       | 0.476             |
| Mother’s age at giving birth               | Unsafe         |              |              |                   |
|                                            | Safe           | 0.8987       | 0.9213       | 0.883             |
| Mother’s age at the first marriage         | Unsafe         |              |              |                   |
|                                            | Safe           | 0.8880       | 0.7944       | 0.238             |
| Mother’s parity                            | Low            |              |              |                   |
|                                            | High           | 0.1200       | 0.1268       | 0.398             |
| Pregnancy complications                    | No             |              |              |                   |
|                                            | Yes            | 0.6232       | 0.6654       | 0.965             |
| Household’s water resources                | Bad            |              |              |                   |
|                                            | Good           | 0.6908       | 0.7898       | 0.747             |
| Household’s sanitation                     | Bad            |              |              |                   |
|                                            | Good           | 0.8076       | 0.8192       | 0.899             |
| Type of place of giving birth              | Nonmedical     |              |              |                   |
|                                            | Medical        | 0.0647       | 0.0882       | 0.884             |
| Person who help of giving birth            | Nonmedical     |              |              |                   |
|                                            | Medical        | 0.9080       | 0.8996       | 0.601             |
| Time zone * time                          | West * time    |              |              |                   |
|                                            | Middle * time  | 0.3551       | 0.3266       | 0.631             |
|                                            | East * time    | 0.1706       | 0.1430       | 0.016             |
| Pregnancy complications * time             | No * time      |              |              |                   |
|                                            | Yes * time     | 0.6004       | 0.6264       | 0.692             |
| Global test                                |                | 0.0609       | 0.0903       | 0.571             |

* Significant at 5% level  
  b Reference category

### 4.3. Fitting the best model

This subsection is addressed to find the best model to the dataset used. This research used two criterions in assessing the goodness-of-fit of the proposed models, i.e. AIC and BIC which are very popular in statistical modelling. The results of the computation of the two criterion values in both the Cox PH model and the extended Cox model (model with time-dependent variables) are depicted in Table 4.
It can clearly be seen that in Table 4 that the AIC and BIC values of the Cox PH model are much higher than the extended Cox model with time-dependent variables. This means that the goodness-of-fit of the extended Cox model to the U5M dataset is much better than the Cox PH model. Afterwards, the best model based on the extended Cox model can be derived by using backward stepwise as described in Table 5. This final model is obtained from the results of the function 'step()' in R software.

5. Conclusions
In this research, the Cox PH model and extended Cox model were applied to U5M data in Indonesia. The extended Cox model is obtained by adding two time-dependent variables which violated the PH assumption in the Cox PH model, namely time zone and pregnancy complications. The violation of PH assumption in U5M data in Indonesia was examined by using LCH plots and GOF test based on Schoenfeld residuals. Furthermore, the results showed that the AIC and BIC values of the extended Cox model for time-dependent variables were much lower than the Cox PH model. This means that the Cox extended model can increase significantly the goodness-of-fit of the Cox PH model in under-five mortality data in Indonesia.
References

[1] Mosley W H and Chen L C 1984 An analytical framework for the study of child survival in developing countries Population and development review 10 25-45

[2] Kayode G A, Adekanmbi V T and Uthman O A 2012 Risk factors and a predictive model for under-five mortality in Nigeria: evidence from Nigeria demographic and health survey BMC Pregnancy and Childbirth 12 1-11

[3] Abir T, Agho K E, Page A N, Milton A H and Dibley M J 2015 Risk factors for under-5 mortality: evidence from Bangladesh Demographic and Health Survey, 2004–2011 BMJ Open 5 1-9

[4] Faruk A 2017 Aplikasi regresi Cox pada selang kelahiran anak pertama di provinsi Sumatera Selatan J. Mat. 7 19-29 Another reference

[5] Faruk A 2018 The comparison of proportional hazards and accelerated failure time models in analyzing the first birth interval survival data J. Phys.: Conf. Ser. 974 012008

[6] Faruk A, Cahyono E S and Eliyati N 2018 Fitting the first birth interval in Indonesia using weibull proportional hazards model Cauchy-jurnal matematika murni dan aplikasi 5 55-63

[7] Cox D R 1972 Regression models and life-tables J. R. Stat. Soc. B 34 187–220

[8] Kleinbaum D G and Klein M 2005 Survival analysis a self-learning text(second edition) (New York: Springer)