Performance of the Survival models in Socioeconomic Phenomena

L.P. Himali ⓘ* **and Zhiming Xia**
Department of Statistics, School of Mathematics, Northwest University, Xi'an, Shaanxi, China

Abstract: Unemployment is still high in Sri Lanka and is considered a severe challenge. Higher unemployment rates are consistently found regardless of the degree of economic activity, in which improvements in the GDP marginalize unemployment. This analysis aims to include statistical tools that help to understand the labour market better and increase the relevance of labour market policies. Survival methods are used, and the Hypertabastic model was found to be more powerful and reliable than other classical models by most researchers who researched it. This research discussed how these observations regarding the Hypertabastic paradigm hold up in the world of economics and social science. This study focuses on developing a survival model for unemployment, and its performance will be compared with the Cox proportional hazard model. Results of the study revealed that unemployed persons with higher education are not advantaged in the labour market, and there are significant differences between women and men concerning the unemployment duration. On average, the study subjects unemployed are young in age and age was not significantly affected by unemployment. According to the fitted models, Hypertabastic Proportional Hazards Model has smaller criterion values than the Cox Proportional Hazards model. Hypertabastic survival models can be an effective survival analysis tool for Socioeconomic Phenomena.

Keywords: Cox Proportional Hazard model, Gender, Hypertabastic Survival model, Unemployment.

1. Introduction

When Graunt (1939) published the first Weekly Bill of Mortality in London, survival research was born, and Healey published the first lifetable. Since then, a growing number of academics have adopted the lifetable approach. Lifetable techniques were first used to measure the reliability of industrial equipment after World War II, then used to research cancer patients' recovery time. That means survival analysis was initially used to research death as a surgical event; however, since the 1970s, these statistical methods have been widely common in economics and social sciences.

Survival data correspond to the time from a well-defined time origin until the occurrence of some particular event or endpoint. Lemeshow et al. (2011), Lee and Wang (2003), Kleinbaum and Klein (2005), and Collett (1994) provide a detailed overview of survival data modelling. Non-parametric and semi-parametric survival models have been widely used to analyze survival data (Cox, 1972). Many applied disciplines use survival modellings, such as medicine, genetics, engineering, epidemiology, and economics. The term "survival time" refers to the time between an incident of significance and its occurrences, such as death, illness relapse, unemployment, or completion of a job.

Survival Analysis has a few unique characteristics. The results are often heavily distorted and non-normal, with few large survival periods, and the incident of significance is rarely found in all subjects due to censored data. This censored observation may be on the right or left.

In mathematical research, censored topics must be used. However, according to Gehan (1969), a large amount of censored data badly affect the accuracy of the statistical tests.

In various areas, non-parametric and semi-parametric survival models have been commonly used to predict survival time and compare survival curves between two or more individuals. The non-parametric Kaplan-Meier method simultaneously ignores the influence of multiple covariates on the result. The semi-parametric Cox model assumes proportional hazards' proportionality and acknowledges the covariates' relative effect on the result. Where the principle of proportionality of hazards is not true and a distribution function for the survival time and baseline hazard function is available, the parametric approach can be a suitable replacement for the Cox model (Lee and Go, 1997). Under the assumed survival distributions, a number of parametric approaches, such as an accelerated failure time model, are also possible. When making a predictive inference, all survival research methods should consider a censoring function (Kalbfleisch and Prentice, 2011).

The baseline hazard function is treated as a nuisance parameter in the Cox model. Parametric or semi-parametric models contribute to more straightforward and precise estimators of the hazard and survival functions, which is helpful in studying change-point hazard rate models in Survival analysis. In recent years, multiple survival analyses and period methods for modelling the length of unemployment spells and strike duration have gained prominence in the Economics and

Corresponding Author's Email: himali@ssl.sab.ac.lk

Social Sciences. Moffitt (1999) discussed new advances in econometric approaches for labour market research. Tansel and Tasci (2004) investigated the determinants of the risk of exiting unemployment in Turkey using non-parametric and parametric prediction approaches when correcting unobserved heterogeneity. To differentiate between exits to a former employer and exits to a current workplace, Nivorozhkin (2006) used a competing risk period model. His model, which considers the relationship between age and tenure, shows that, while being legally covered, older jobs in Sweden are more likely to be unemployed for more extended periods of time if they have the same tenure. D'Agostino and Mealli (2000), on the other hand, used Cox proportional hazard models to investigate the duration of unemployment in several European countries, with mixed results. In Portugal, France, and Denmark, elderly citizens have the most difficulty leaving unemployment, while in Italy, the United Kingdom, and Spain, both the youth and the aged have fewer opportunities for reemployment.

Though standard parametric models are often used in literature, they can have some drawbacks, such as the Exponential, Weibull, or Gamma hazard functions, which either go in the same direction with time-up, time-down, or constant. This means that these models are unable to match the results adequately. Although the hazard functions of the Log-normal and Log-logistic distributions have one turning point, their pose graphs often reflect the versatility poorly. The hazard rate of Log-normal is non-monotonic, while the hazard rate of Log-logistic is monotone decreasing or increasing from 0 to its highest value and then decreasing to 0. This suggests that conventional parametric models are insufficiently scalable to reflect real-world data accurately (Tran, 2014).

New parametric models with rich hazard function properties are often discovered. A Hypertabastic distribution is one of them. Tabatabai et al. (2007) suggested the Hypertabastic proportional hazards model and the Hypertabastic accelerated failure time (AFT) model as a new two-parameter probability distribution. Unlike standard distributions, the Hypertabastic hazard rate function may be monotone increasing, monotone decreasing, or -shaped in the independence of parameter values.

The usage and application of the Hypertabastic survival model family have gradually increased over the last ten years. They have a compact form and are resistant to a wide range of fundamental distributions. Clinicians, physicians, and researchers are advised to compare this model to other commonly used survival models before determining which one offers the best match and forecast. This Hypertabastic paradigm is currently used primarily in biomedical and engineering disciplines but was not used by any economics and social science researchers. The unemployment dilemma is not as straightforward as it seems at first glance; it involves more than just a supply and demand imbalance in the labour market. This is especially true in transition countries that have been through the process of tran-

sitioning from a pre-transition economic structure to a market-based economy. A mismatch may be described as the presence of disequilibrium between labour supply and demand; more generally, a mismatch demonstrates the impossibility of linking current unemployment and vacancies at a disaggregated stage. Because of the increased competition in the world market and the rapid technical advances, the current labour supply is finding it more challenging to respond to changing labour demand. Addressing this problem is critical, especially for a developed country like Sri Lanka, because few studies on the unemployment period have been performed in developing countries, and none of the studies has used the Hypertabastic model. Through scientific analysis, this paper will provide keen knowledge for policymakers, economists, and academic researchers about unemployment in Sri Lanka and the extent and severity of the consequences.

Survival analysis is a form of reliability analysis in which "time-to-event" data and their covariates are examined. This is a data-driven method for determining the impact of covariates on the result (Nabizadeh et al., 2018). While still small, this solution to unemployment has grown in popularity in recent years. Recent research efforts have been the development of survival models for unemployment data. Several recent experiments have been methodological in nature. Beamonte and Bermdez (2003) investigated a Bayesian additive model and applied it to graduate first-time work. Carroll (2006) used survey data and introduced search theory into his methods, while Lüdemann et al. (2006) examined the length of unemployment in West Germany using a quantile regression model for administrative data. In Sweden's unemployment results, Carling and Jacobson (1995) applied a competing risk model to find the association between exiting and declining. Jensen and Svarer (2003) compared findings from a multiple phase duration model and a competing risk model to demonstrate that temporary layoffs are correlated with short-term unemployment in Denmark. Baussola and Mussida's (2017) competing risk figures indicate that while males have higher work probabilities, females abandon the labour force due to discouragement. Due to the uncovered results, the Cox Proportional Risk Model will answer heterogeneity. The model enables the data to assess the baseline hazard's physical shape. On the other hand, non-parametric models are built to address unexpected variability and are effective tools for understanding fundamentals and producing descriptive performance (Mills, 2010).

The Weibull (parametric model) is notorious for its benefit since the population is not uniform. However, the model calculates the parameters by identifying values that maximize the function of probability (Haughton and Haughton, 2011).

The longitudinal data structure presents obstacles to parametric inferences. Fan and Li (2004) suggests using semi-parametric modelling to analyze longitudinal dates. Multivariate regression techniques cannot handle the un-

balanced existence of longitudinal data. For administrative data, Wichert and Wilke (2008) proposed basic non-parametric models as administrative data are limited to different ways of censoring. Many scholars have investigated their exertions on modelling growth phenomena and it has become a matter of great interest in various areas of application and exploration. The generalization of standard models and the introduction of new models is one of the research lines most researchers widely practice today. Bürger et al. (2019) recommended generalizations of the logistic and Gompertz models, and Tabatabai et al. (2005) introduced the hyperbolastic curves. These models have been the preliminary themes for the development of others. The hyperbolastic curves are a family of curves included in the (Tabatatai et al., 2005) system to extend behavioural pattering by including hyperbolistic functions, for example, logistic ones (type I) and Weibull (type III). This provides more flexibility, allowing improved modelling of some phenomena of growth. The boost in accuracy compared to cox regression was complimented by the ability to examine the period of illness progression using clearly specified hazards and survival functions.

The function might be monotonous due to the properties of the hazard rate function. They may grow (I), decrease (D) or take form. The hazard properties of the hypertabastic hazard function were expressed by Tabatabai et al. (2007) in the following terms:

1. When the $0 < \beta \leq 0.25$, the hazard rate $\downarrow$ from α to 0.
   (The results for 825 diagnosed patients of primary epithelial ovarian carcinoma were evaluated by Clark et al. (2001). Initially, the risk rate was high after diagnosis and, after that, steadily declined).

2. The hazed rate is unimodal if it has a value of $0.25 < \beta \leq 1$ (Demicheli et al. (2004) found the node-positive postmenopausal women's hazard rate to be unimodal. Schulman et al. (2001) investigated the effects on the development of obliterative bronchiolitis (OB) following lung transplantation by the donor and recipient HLA locus mismatched and evaluated the risk of OB after the transplant. Their risk function for the development of OB has also been calculated to have a unimodal curve).

3. If $1 < \beta \leq 2$, the hazed rate must continue to descend to downward concavity until the inflexion point has been reached.

4. When $\beta > 2$, the hazard rate increase with upward concavity.

5. The hypertabastic baseline hazard increases with time, reaching its horizontal asymptote provided that $\beta = 1$
   Weitz and Fraser (2001) found that hazard-rate plateaus are described as a generalized result of

first-pass death consideration for random drifting processes. In populations of fruit flies, yeast and other species, they studied the plateau risk).

6. If $\beta = 2$, the hypertabastic baseline hazard function increases with upward concavity for a while and then becomes a linear function with slope α.

Tabatabai et al. (2007) compared the overall fit of the Hypertabastic model to Weibull, log-logistic, and log-normal models in simulation analysis. The negative of the log-likelihood was used as a goodness-of-fit test for all distributions. The authors ran 1000 simulations with a sample size of 200 people and a random censoring of around 40%. For 33,000 simulations, 11 different parameter combinations of two-parameter Weibull, log-normal, and gamma distributions were used to produce a time-to-event data set. The log probability was summed over 1000 runs to determine which model best matched the simulation data with the highest overall precision. Samples were created from a two-parameter Weibull distribution in simulations, and the Weibull model matched the data with the greatest precision in all cases. For all parameter variations, the Hypertabastic model came in second, outperforming the Log-normal and Log-logistic models (Singh et al., 2019). The Hypertabastic and Log-logistic models appeared to be comparable, with the Log-logistic model outperforming the Weibull model in eight parameter combinations. Finally, the Weibull model suit with the most accuracy in seven out of eleven configurations when sampling from a two-parameter Gamma distribution. The Hypertabastic model outperformed the Weibull model by a wide margin.

Bursac et al. (2009) demonstrated a degree of robustness in terms of data propagation variations. The Hypertabastic distribution's feature of changing its form is a more precise reflection of the hazard and survival functions' time course. These novel features of this model become much more relevant in light of scientists' current efforts to improve gene expression variables (Bursac et al., 2009).

The AFT and PH models based on the Hypertabastic distribution were suggested by Tran (2014) and Tahir et al. (2017) as generalized chi-square test statistics. They used simulation experiments to show that the maximum likelihood parameter estimators of the Hypertabastic model have an asymptotically distributed normal distribution, and they confirmed the asymptotic property of their generalized test statistic for the Hypertabastic distribution when the correct censoring probability is between 0% and 20%. The Hypertabastic baseline hazard rate for node-positive postmenopausal women was unimodally formed, according to Demicheli et al. (2004).

Survival models have been studied and applied widely in unemployment data. Much progress was accomplished and played an essential part in survival and reliability analysis throughout the 2nd part of the 20th century.
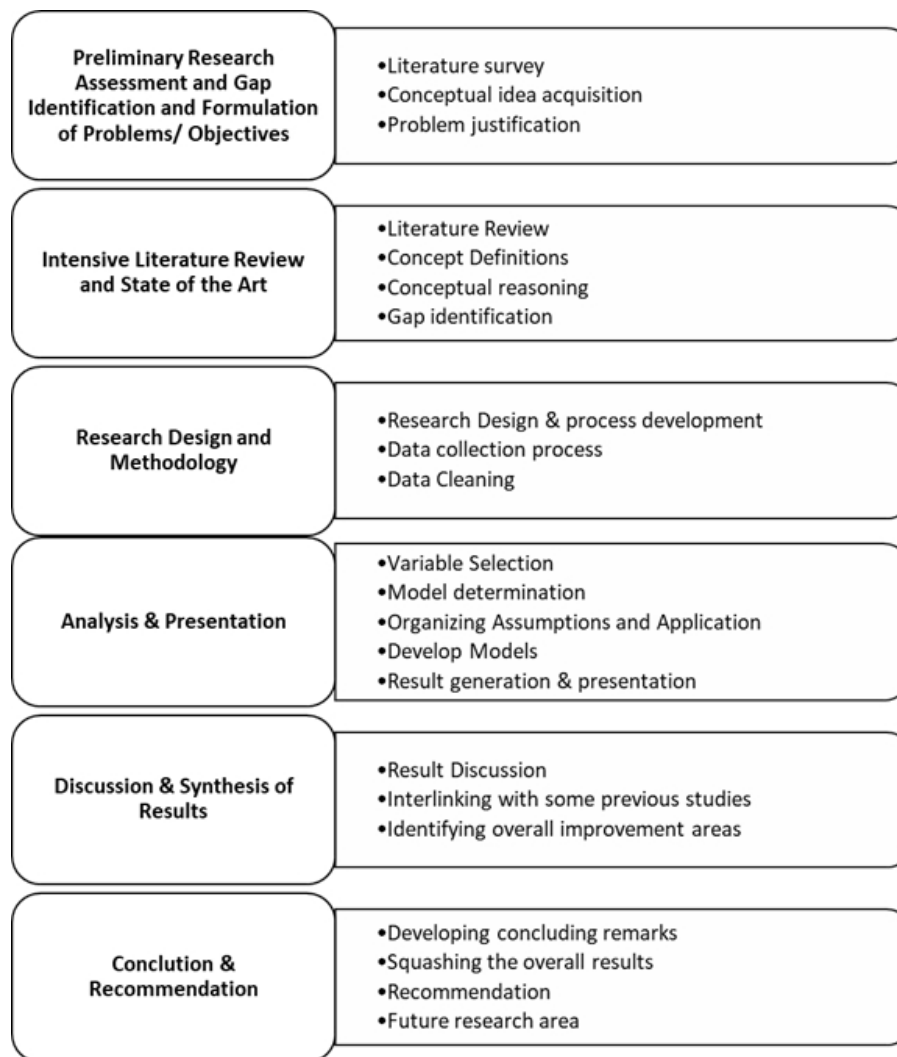
Figure 1: Research Methods and Processes

The models of proportional hazard (PH) and accelerated failure time (AFT) are some of the most often utilized survival models in this discipline. Although the Cox Model was renowned and often used as a logarithm without the requirement for determining the baseline hazard due to the benefit of assessing the influence of variables, the usage of the Cox Model might cause loss or omission of vital information. In survival analysis, a semi-parametric AFT model is employed in a significantly less complex assessment process than the Cox model. Also utilized is a non-parametric AFT model, although frequently employed is a parametric AFT. Unlike Cox, the parametric AFT model is used mainly for failure time and accelerated life testing analysis. In recent years, the study of parametric models for PH or AFT models, such as Exponential, Weibull, Gamma, Lognormal, and Log-logistic for data adaptation, is of concern. However, Exponential, Weibull or Gamma for the principal parametric models are hazardous, with the timeout, time down or consistently always going in the same direction. This indicates that the data cannot fit these

models correctly. The Log-normal log-logical distributions occasionally exhibit the flexibility of their posture graphs, even if they have one turning point with their hazard functions. Therefore, the Hypertabastic model of survival will lead to a higher guarantee of statistical conclusions for covariate-influenced unemployment data.

There has been no recent study in the area of Economics and Social Sciences that has used the Hypertabastic model. This study performed a parametric survival analysis using the Hypertabastic distribution function, which is commonly used in biomedical and engineering applications. The Hypertabastic model was found to be more powerful and reliable than other classical models by most researchers who researched it. This research will see how these observations regarding the Hypertabastic paradigm hold up in economics and social science.

Although survival models have been widely used in unemployment data for a couple of years, new two-parameter probability distribution called the Hypertabas-

tic model includes the Hypertabastic proportional hazards model and the Hypertabastic accelerated failure time (AFT) model has not yet been developed for unemployment data. Thus this study focuses on developing a Hypertabastic survival model for unemployment, and its performance will be compared with the currently used Cox proportional hazard model.

2. Methodology

2.1. Research Design

This study employed a mixed type of methods. Primary data was collected through telephone conversation (spend 6-month period and google forms also utilized). The simple random sampling technique was used to select the respondents for this study. Secondary data was gathered from the Census & Statistics Department of Sri Lanka from the 2019 Labour Force Survey. This sample includes individuals residing in housing units and excludes the institutional community. Moreover, randomly selected 3343 subjects belonging to 18 years and over-population in Southern province. Because the highest unemployment rate is reported in the Southern province (6.7%) compared to the other provinces in Sri Lanka, this study employs a descriptive research design to develop an unemployment model using survival analysis. The research's overall design and flow process are depicted in Figure 1.

2.2. Analysis Method

A fundamental task is to pick a simple model of distribution that can better display findings for parametric survival studies. In addition, when considering the results of covariates, a decision must be taken between the models Accelerated Failure Time (AFT) and Proportional Hazard (PH). In this analysis, the best baseline distribution model has been established for the data on unemployment.

Kaplan-Meier (KM) estimators were applied to estimate unemployment survival curves, and the log rank test was used to compare the covariate categories. Next, the hypertabastic proportional hazards model was applied, and this model was compared with the semi-parametric Cox proportional hazards model using the goodness of fit test.

2.3. Kaplan-Meier Estimation

This is a product limit estimation of the survivorship function, developed by Kaplan and Meier (1958) and used by most researchers due to the simplistic step approach. The KM estimator includes information from all available observations, both censored and uncensored, taking any time as steps defined by the observed survival and censored time. When no censorship occurs, the estimator is only the fraction of samples with event timings larger than "not". The process becomes more sophisticated but feasible when filtered periods are added.

This KM estimating is made up of an estimated survival function in the form of a walking function as the outcome of several conditions probabilities. It is a free survival function parameter estimator. In general, the survival function $S(t) = Pr(T > t)$, and it gives the population probability of surviving beyond t, where T is a random variable known as the survival time.

The cumulative distribution of T is $P(t) = Pr(T \leq t)$, and the probability density function is $P(t) = dP(t)/dt$. This means the survival function is $S(t) = 1 - P(t)$. Also, the hazard function is essential in the survival analysis. This function assesses the instantaneous risk of demise at time t, conditional on survival to that time, $h(t) = p(t)/S(t)$.

The Kaplan-Meier estimator allows the computation of an estimated survival function in the presence of right censoring (Dalgaard, 2008). The usual sample survival function for uncensored data is

$$\hat{S}(t) = \prod_{j_j < t} \left(1 - \frac{d_j}{n_j} \right) \tag{1}$$

where d_j is the number of individuals who experienced the event at the time t_j, and n_j is the number of individuals who have not yet experienced the event at that time.

This procedure is widely used in medical studies where "death" and "alive" are used frequently. In this study, "alive" is remaining unemployed, and "death" is finding a job. Thus, the survival function $(S(t))$ is the probability of individual remaining unemployed t units of time from the beginning of the study.

2.4. Cox proportional Hazards Model

Cox proportional method is used to apply several risk factors simultaneously to the survival period and is one of the most common regression methods for survival analyses. The calculation of the influence of a Cox proportional risk regression model is the chance of loss since the participant has been on for a specified period. A likelihood should be between 0 and 1. However, this risk reflects one unit of time in the estimated number of incidents.

The Cox proportional risk model contains vital assumptions, including independent time for different study subjects, a multiplicative relationship between the predictors and the risks and a constant hazard ratio over time. The Cox proportional hazards model for the hazard rate for i^{th} person can be described as

$$h(t) = h_0(t)e^{(b_1 X_1 + b_2 X_2 + b_3 X_3 + \ldots + b_n X_n)} \tag{2}$$

where $h(t)$ at time t is the predicted hazard, $h_0(t)$ is the baseline hazard and is a hazard when all of the X_1, X_2, and X_n predictors are equivalent to zero. The predictors have a proportional effect on the predicted hazard.

The CHF and Survivor function are given by

$$H(t) = H_0(t)e^{(\beta' X)} \text{ and} \tag{3}$$

$$S(t, X, \beta) = [S_0(t)]^{e^{(\beta' X)}}, \tag{4}$$

respectively, where, $S_0(t)$ is a baseline survival function.

2.5. Hypertabastic Survival Model

The hypertabastic distribution is a relatively new type of distribution introduced by Tabatabai et al. (2007). It has been used in several applications, including studying covariates' effect on cancer patients' survival time and engineering applications (Tabatabai et al., 2007; Tran, 2014; Tahir et al., 2017). The most prominent feature of the hypertabastic survival function is its capability to represent various hazard shapes (Tabatabai et al., 2007).

The willingness of the threat mechanism to adopt a wide range of different forms, as opposed to the Weibull, log-normal, and log-logistic, is one essential aspect of the Hypertabastic Survival Model. The hypertabastic distribution function is given by

$$F(t) = \begin{cases} 1 - \text{sech}\left(\frac{\alpha\left(1 - t_i^\beta \coth\left(t_L^\beta\right)\right)}{\beta}\right) & \text{for } t > 0 \\ 0 & \text{for } t \le 0. \end{cases} \tag{5}$$

The baseline hazard function is given by

$$h(t) = \alpha\left(t^{2\beta-1}\text{csch}^2(t^\beta) - t^{\beta-1}\coth(t^\beta)\right)\tanh(W(t)) \tag{6}$$

where $W(t) = \alpha(1 - t^\beta \coth(t^\beta))/\beta$, α and β are the model parameters and both positive, these parameters α and β give the hazard function stability to comply with the data set specified. Then, the hazard function in the type of the hypertabastic proportional hazard model is given by

$$h(t|x, \theta) = h(t)g(\theta|x). \tag{7}$$

The function $g(\theta|x)$ is given by

$$g(\theta|x) = e^{-\theta_0 - \sum_{k=1}^P \theta_k x_k} \tag{8}$$

where x_k are the covariates and the θ_k are the associated parameters. The Hypertabastic survival function $S(t|x, \theta)$ for the proportional hazards model is given by

$$S(t|x, \theta) = [S_0(t)]^{g(\theta|x)} \tag{9}$$

where $S_0(t)$ is the baseline survival function and is given by

$$S_0(t) = \text{sech}\left(\frac{\alpha(1 - t^\beta \coth(t^\beta)}{\beta}\right) \tag{10}$$

2.6. Variable Selection and Model Comparisons

The Cox and hypertabastic models were adapted to the data set to investigate risk factors for unemployment and assess the success of various models. Models were compared using a log-likelihood value (log L) and AIC to evaluate the goodness of fit of models. The AIC is an effective way of trading off the complexity of an estimated model against how well the model fits the data. The AIC is calculated by

$$\text{AIC} = -2\log(likelihood) + 2(p + k) \tag{11}$$

where p is the number of parameters, $k = 1$ for the exponential model, $k = 2$ for the Weibull, log-logistic, and log-normal models and $k = 3$ for generalized gamma.

A likelihood ratio test (LRT) is also used to compare the fit of the two models. The LRT test statistic is twice the difference in the log-likelihoods of the models considered for comparison. The lowest AIC matches and is chosen for the optimal suit. The experiment was equipped with Hypertabastic proportionate hazards using the forward selection process and partial likelihood ratio test, identifying the same range of risk factors as the model Cox with the exception of diameter. As the diameter of the Cox model is chosen, each reference model has been used. The probability of correlations between variables is also taken into account. By applying the key impact model, the importance of each interaction is evaluated, and no relevant associations were found by the selection procedure using the partial likelihood ratio test.

3. Results and discussions

The number of unemployed persons was 411,318 during the year 2019 in Sri Lanka, and the highest unemployment reported was in the Southern province (6.7%). Of this total, 45 % are males, and 55% are females. This study used 3343 unemployed persons who lived in Southern Province. The women's unemployment rate is more than twice as high at the national level (7.4%) as the men's unemployment rate (3.3%) (see Figure 2).
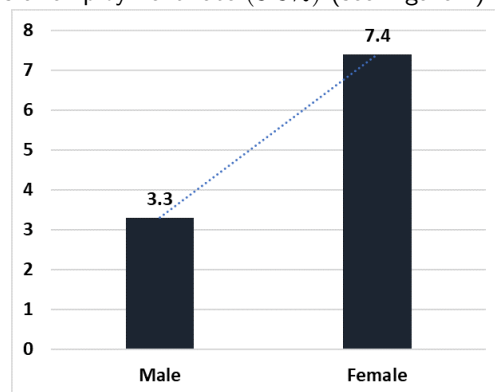
Figure 2: Unemployment rate by gender (Source: Labour force survey, 2019)

According to the world bank report, young women, in comparison with young men, in the whole of Sri Lanka are facing growing barriers to jobs. Much female unemployment correlates with discriminatory workplaces and lower working networks for women than men. Given this low results outlook in the workplace, a few women never

want to take part or are forced to withdraw entirely from the industry.
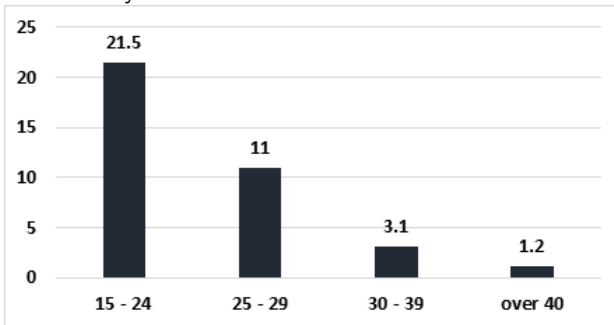
**Figure 3:** Unemployment rate by age group (Source: Labour force survey, 2019)

Overall unemployment for young people (aged 15-24) is 21.5% (see Figure 3). Most unemployed are young job seekers or those in the 20-24-year age range. With age, the unemployment rate falls. Many of these unemployed people are looking for work in the public sector, and their unemployment duration is longer. As a result, the nature of unemployment in Sri Lanka is comparable to that of other developing countries such as Ethiopia (Hosmer and Lemeshow, 1999), Tunisia (Rama, 1998), and China (Appleton et al., 2006).
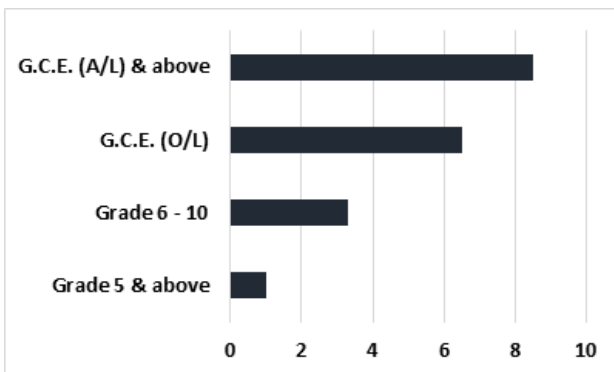
Figure 4: Unemployment rate by level of education (Source: Labour force survey, 2019)

The skilled GCE (A/L) category reported the highest unemployment rate, 8.5% (see Figure 4). Previous surveys stated that unemployment in educated women is more severe than that in educated men. According to Isenman (1980) and Glewwe and Bhalla (1987), there are long lines for government jobs. Queuing workers want to enter not only for well-paid jobs but also for well-paid government jobs. Unemployment in Sri Lanka arises due to the favourability of being unemployed while waiting for a job in the high-paying public sector.

Unemployment lasts an average of 6.25 months, and the histogram is not normally distributed (see Figure 5).
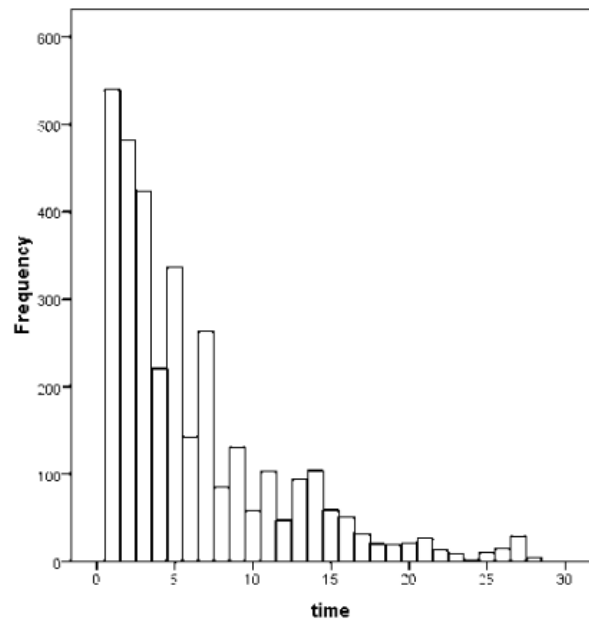

Figure 5: Histogram of the unemployment duration

3.1. Survival Probabilities

Table 1 shows the surviving probabilities over time.

Table 1: Surviving probabilities for time

Month	Survival
1	0.91
2	0.83
3	0.77
4	0.73
5	0.66
6	0.64
7	0.57
8	0.56
9	0.52
10	0.51
11	0.48
12	0.47
13	0.44
14	0.39
15	0.35
16	0.32
17	0.30
18	0.28
19	0.27
20	0.26
21	0.24
22	0.22
23	0.21
24	0.21
25	0.20
26	0.18
27	0.12
28	0.07

According to Table 1, the probability of survival in the
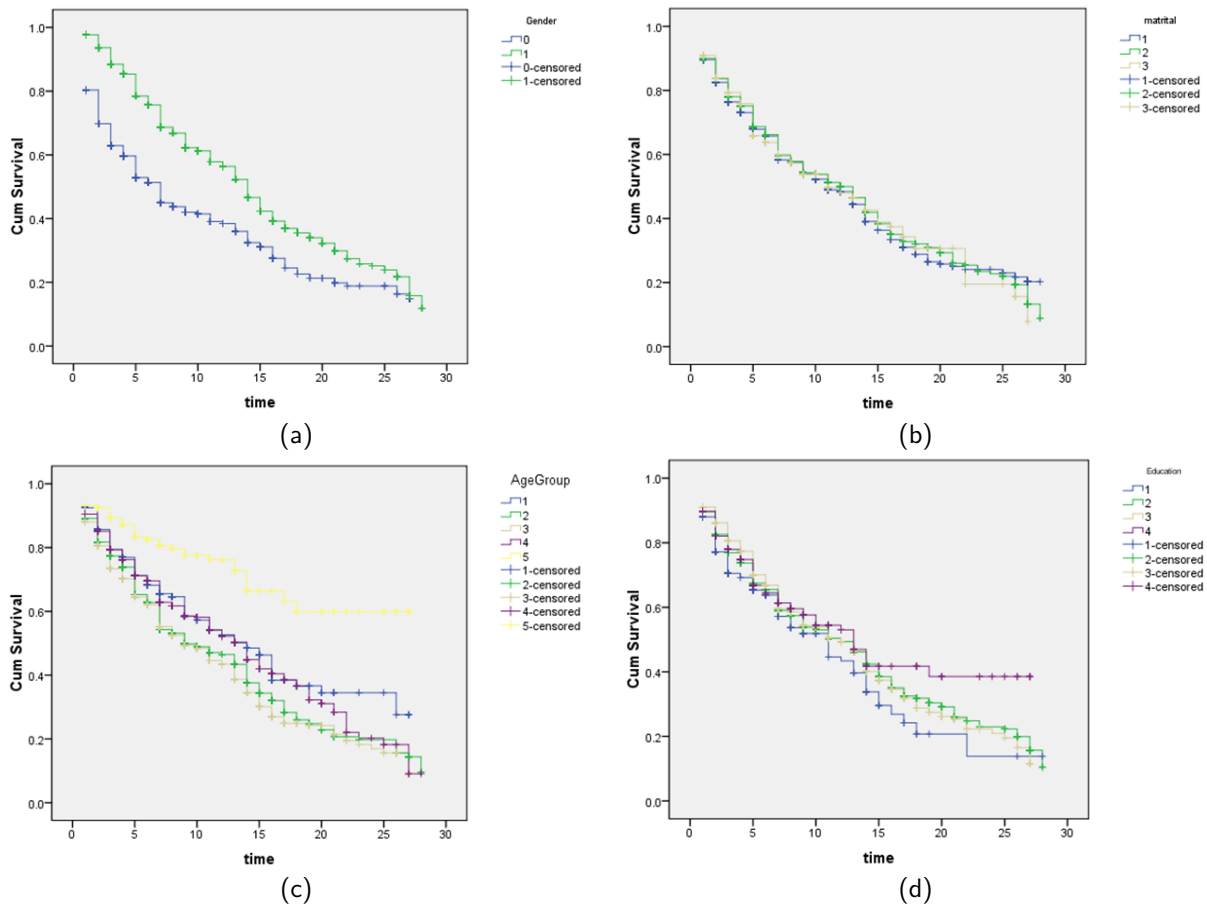
Figure 6: Survival functions for (a) Gender, (b) Marital Status (c) Age group (d) Level of education

first month of the study is 0.91. This probability falls to 0.57 in the seventh month. After 26 months, the survival probability falls below 20%, and after 27 months, it falls below 10%. During the 28-month study, the probability of survival approaches zero.

The graphical representation of surviving probabilities over time is given in Figure 6.

According to Figure 6(a), it can be noted that there are significant differences between women and men.

The Log-rank test indicates that the survival functions are not the same for women and men ($p < 0$).

The Kaplan Meier estimators of the survival curves on marital status are shown in Figure 6(b). In this study, three categories were considered: level 1 is single, level 2 is married, and level 3 is widowed. After the twentieth month of unemployment, the survival curves show significant disparities. Individuals in the unmarried group have the best chance of finding work.Individuals in the widowed group had the lowest chance of finding work.

According to the Log Rank test for age group ($p < 0$), the age influences are indicated, and the null hypothesis has been rejected. This point is shown in Figure 6(c),

presenting the Kaplan Meier estimators of the survival curves in unemployment by age group. This study considered five Age group categories: 1 for 15-25 age group, 2 for 26-35,3 for 36-45, 4 for 46-55 and 5 for 56-65 age group. It can notice a big difference between the 56-65 age group and other age groups. It seems that for the individuals in the age groups 56-65 years and 15-25 years is the most difficult to find a job and that it is easier for those in the age group 36-45 years.

The Kaplan Meier estimators of the survival curves on educational levels are shown in Figure 6(d). This study examined four educational levels: level 1 is primary education, level 2 is secondary education, level 3 is collegiate education, and level 4 is university education. Individuals with an elementary education have the highest chance of finding work. Individuals with a high school tertiary education have the lowest chance of finding work.

3.2. Models based on Unemployment and Socio - demographic Factors

This study used three survival analysis techniques, hypertabastic, Kaplan Meier and Cox proportional model, to determine the effective model for unemployment data. -2 log-likelihood score and the Akaike Information Cri-

Table 2: Predictors in the Individual Models based on Cox Proportional Hazards

Covariates	Hazard ratio	Standard error (SE)	p-Value	Lower Confidence Level (95%)	Upper Confidence Level (95%)
Gender (Female)	0.478	0.054	0.000	0.430	0.532
Age	0.994	0.002	0.018	0.989	0.999
Marital Status	0.982	0.043	0.670	0.903	1.068
Level of Education	0.981	0.009	0.037	0.963	0.999
English Literacy	1.014	0.054	0.790	0.913	1.127
Professional Training	1.025	0.104	0.810	0.836	1.258

Table 3: Final Model Summary for Cox Proportional Hazards Model for Unemployment

Explanatory Variable	β	SE	Wald	df	p-Value	e^{β}	Lower Confidence Level (95%)	Upper Confidence Level (95%)
Gender	-0.739	0.056	177.097	1	0	0.478	0.429	0.533
Age	0	0.003	0.005	1	0.941	1	0.995	1.005
Education	-0.019	0.009	4.094	1	0.043	0.981	0.963	0.999

Table 4: Final Model Summary for Hypertabastic Proportional Hazards Model for Unemployment

Explanatory Variable	Estimate	SE	t value
Gender	-0.765	0.674	-1.1249
Age	0.043	0.022	2.1595
Education	-0.657	0.710	-0.9253

Table 5: Model Comparison for the Hypertabastic Proportional Hazards and Cox Proportional Models for Unemployment

Criterion	Cox Proportional Hazard Model		Hypertabastic proportional Hazard Model	
	Without Covariates	With Covariates	Without Covariates	With Covariates
-2 LOG L	162.903	143.041	162.903	135.005
AIC	162.903	175.007	162.903	168.207

terion were used to compare these models. The Akaike Information Criterion is commonly used when selecting among several competing models, with the smallest value corresponding to the best fit model. The covariates included in the model were gender, age, marital status, level of education, English literacy and whether or not they completed a professional training programme. A correlation analysis was first performed among covariates to determine if they correlated. The correlation analysis indicated that the selected covariates are not correlated and are independent of each other. Therefore, in this study, all the selected covariates are included as a parameter for the survival analysis of unemployment. The model-building process selects suitable variables for the Cox proportion model as the first step fits univariate models to identify predictors with small p—values (lower than 0.5). Covariates with p—values smaller than 0.5 are assumed to be relevant in the model. In the second step, fit a multivariate model with all significant univariate predictors, and use a stepwise selection procedure and significance level in this selection approach is 5%. Step three considers the variables excluded in step two and checks for their multivariate significance. This step assumes that the relevancy of other variables might depend on other variables in the model. Step four

involves a stepwise selection as a final check of important variables; this step also considers interaction effects.

The result in Table 2 indicates that the model factor approximates of coefficients for the predictors in the individual models based on Cox proportional hazards. The predictors such as age, gender, level of education, marital status, English literacy and whether or not they completed professional training used the forward variable selection method to choose the important covariates. To fix whether a covariate is important, the p-value related with each model parameter has been estimated. Variables with a p-value less than or equal to the 0.05 cut point are considered important variables and included in the final model. According to the results of the above table, the variables gender, age and level of education significantly affect unemployment.

Table 3 examines the survival distribution relationship to an explanatory variable, using a Cox regression of time to employment on age, gender and level of education. According to the final model of Cox Proportional hazards results, the survival moment of unemployment of the individual was significantly affected due to gender

and level of education. The values of the Wald value for every coefficient hold that the values were significantly dissimilar with zero at $\alpha = 5\%$ significant leval.

Based on Table 4, gender and education level significantly contribute to the risk of unemployment in this study, which is similar to the Cox Proportional Hazards Model result.

Table 5 provides a clear picture of the comparison of the two models based on the comparison criterion -2 LOG L and AIC values. The model with the lower value fits the data the best. As a result of the findings of this investigation, the Hypertabastic Proportional Hazards Model has lower criteria values than the Cox Proportional Hazards Model and better matches the data.

4. Conclusion

According to the results of the study, it can be concluded that there are significant differences between women and men concerning unemployment duration. The male gender has a great deal of lesser effect on unemployment than the female gender. Most unemployed are young job seekers or those aged 20-24-year age range. With age, the unemployment rate falls. The unemployed persons with higher education are not advantaged in the labour market; maybe it is more difficult for them to re-orientate than the unemployed with low levels of education. The survival moment of unemployment of the individual was significantly affected due to gender and level of education. Based on the studies, the Hypertabastic survival models are an effective survival analysis tool for analyzing unemployment data.

The results of the estimations of this study give a picture of the factors shaping unemployment duration, which may help policymakers.

References

Appleton, S., Knight, J., Song, L. and Xia, Q., 2006. Labour retrenchment in China: Determinants and consequences. In *Unemployment, inequality and poverty in urban China* (pp. 35-58). Routledge.

Baussola, M. and Mussida, C., 2017. Regional and gender differentials in the persistence of unemployment in Europe. *International Review of Applied Economics*, 31(2):173-190.

Beamonte, E. and Bermúdez, J.D., 2003. A bayesian semiparametric analysis for additive hazard models with censored observations. *Test*, 12(2):347-363.

Bürger, R., Chowell, G. and Lara-Díaz, L.Y., 2019. Comparative analysis of phenomenological growth models applied to epidemic outbreaks. *Mathematical Biosciences and Engineering*, 16(5):4250-4273.

Bursac, Z., Tabatabai, M.A., Williams, D.K., Singh, K.P., 2009. Simulation Study of Performance of Hypertabastic Survival Models in Comparison with Classic Survival models. In *the 2008 ASA Proceedings of the Joint Statistical Meetings*, Biometrics Section, Alexandria, VA.

Carling, K. and Jacobson, T., 1995. Modeling unemployment duration in a dependent competing risks framework: Identification and estimation. *Lifetime Data Analysis*, 1(1):111-122.

Carroll, N., 2006. Explaining unemployment duration in Australia. *Economic Record*, 82(258):298-314.

Clark, T.G., Stewart, M.E., Altman, D.G., Gabra, H. and Smyth, J.F., 2001. A prognostic model for ovarian cancer. *British journal of cancer*, 85(7):944-952.

Collett, D., 1994. *Modelling survival data in medical research*. CRC press.

Cox, D.R., 1972. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187-202.

D'Agostino, A. and Mealli, F., 2000. *Modelling short unemployment in Europe* (No. 2000-06). ISER Working Paper Series.

Dalgaard, P.,2008. *Introductory statistics with R*.

Demicheli, R., Bonadonna, G., Hrushesky, W.J., Retsky, M.W. and Valagussa, P., 2004. Menopausal status dependence of the timing of breast cancer recurrence after surgical removal of the primary tumour. *Breast Cancer Research*, 6(6):1-8.

Fan, J. and Li, R., 2004. New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of the American Statistical Association*, 99(467):710-723.

Gehan, E.A., 1969. Estimating survival functions from the life table. *Journal of chronic diseases*, 21(9-10):629-644.

Glewwe, P. and Bhalla, S.S., 1987. A response to comments by Graham Pyatt and Paul Isenman. *The World Bank Economic Review*, 1(3):533-536.

Graunt, J., 1939. *Natural and political observations made upon the bills of mortality* (No. 2). Johns Hopkins Press.

Haughton, D. and Haughton, J., 2011. *Living standards analytics: Development through the lens of household survey data*. New York: Springer.

Hosmer, D.W. and Lemeshow, S., 1999. *Applied Survival Analysis: Regression Modelling of Time to Event Data*. New York: Wiley.

Isenman, P., 1980. Basic needs: the case of Sri Lanka. *World Development*, 8(3):237-258.

Jensen, P. and Svarer, M., 2003. Short-and long-term unemployment: How do temporary layoffs affect this distinction?. *Empirical Economics*, 28(1):23-44.

Kalbfleisch, J.D. and Prentice, R.L., 2011. *The statistical analysis of failure time data*. John Wiley & Sons.

Kaplan, E.L. and Meier, P., 1958. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457-481.

Kleinbaum, D.G. and Klein, M., 2005. Survival analysis. Statistics for biology and health. *Survival*, 510.

Lee, E.T. and Go, O.T., 1997. Survival analysis in public health research. *Annual review of public health*, 18:105.

Lee, E.T. and Wang, J., 2003. *Statistical methods for survival data analysis* (Vol. 476). John Wiley & Sons.

Lemeshow, S., May, S. and Hosmer Jr, D.W., 2011. *Applied survival analysis: regression modeling of time-to-event data*. John Wiley & Sons.

Lüdemann, E., Wilke, R.A. and Zhang, X., 2006. Censored quantile regressions and the length of unemployment periods in West Germany. *Empirical Economics*, 31(4):1003-1024.

Mills, M., 2010. *Introducing survival and event history analysis*. Sage.

Moffitt, R.A., 1999. New developments in econometric methods for labor market analysis. In *Handbook of labor economics* (Vol. 3, pp. 1367-1397). Elsevier.

Nabizadeh, A., Tabatabai, H. and Tabatabai, M.A., 2018. Survival analysis of bridge superstructures in Wisconsin. *Applied Sciences*, 8(11):2079.

Nivorozhkin, A., 2006. New estimates of the risk and duration of registered unemployment in urban Russia. *International Journal of Manpower*, 27(3):274-289.

Rama, M., 1998. How bad is unemployment in Tunisia? Assessing labor market efficiency in a developing country. *The World Bank Research Observer*, 13(1):59-77.

Singh, K.P., Ndetan, H.N., Bursac, Z., Eby, W.M., Tabatabai, M.A., Eum, M., Bertolucci, A.A., Bae, S., 2019. A flexible family of hypertabastic models. In *Proceedings of 23rd International Congress on Modelling and Simulation*, Australia.

Schulman, L.L., Weinberg, A.D., McGREGOR, C.C., Suciu-Foca, N.M. and Itescu, S., 2001. Influence of donor and recipient HLA locus mismatching on development of obliterative bronchiolitis after lung transplantation. *American journal of respiratory and critical care medicine*, 163(2):437-442.

Tabatabai, M., Williams, D.K. and Bursac, Z., 2005. Hyperbolastic growth models: theory and application. *Theoretical Biology and Medical Modelling*, 2(1):1-13.

Tabatabai, M.A., Bursac, Z., Williams, D.K. and Singh, K.P., 2007. Hypertabastic survival model. *Theoretical Biology and Medical Modelling*, 4(1):1-13.

Tahir, M.R., Tran, Q.X. and Nikulin, M.S., 2017. Comparison of hypertabastic survival model with other unimodal hazard rate functions using a goodness-of-fit test. *Statistics in Medicine*, 36(12):1936-1945.

Tansel, A. and Tasci, H.M., 2004. Determinants of unemployment duration for men and women in Turkey. *Available at SSRN 512222*.

Tran, X.Q., 2014. *Dynamic regression models and their applications in survival and reliability analysis* (Doctoral dissertation, Université de Bordeaux).

Weitz, J.S. and Fraser, H.B., 2001. Explaining mortality rate plateaus. *Proceedings of the National Academy of Sciences*, 98(26):15383-15386.

Wichert, L. and Wilke, R.A., 2008. Simple non-parametric estimators for unemployment duration analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 57(1):117-126.