

# Determinants of Unemployment and Unemployment Duration

Himali L P

Department of Statistics, School of Mathematics, Northwest University, Xi'an, Shaanxi, China

Email address: hima099 @ gmail.com

**Abstract**— Unemployment is a global problem and in Sri Lanka has continued to be consistently high. Consistently high unemployment is irrespective of the economic growth but cause to change the Gross Domestic Product. The aim of this study is providing a statistical tool to model unemployment and unemployment duration to enhance the dynamics of the unemployed. Basically used primary data through questionnaire and two stage stratified sampling procedure is adopted to select a sample of 1200 housing units in Kalutara district. Chi square test, Binary Logistic Regression and Cox Regression used as the main analytical tools to find out the determinants of unemployment and unemployment duration. The results revealed that living sector, gender, level of education and language constraints are highly associated with unemployment of an individual. The majority among unemployed are young people who belongs to age group 15 – 29 years and it shows that unemployment rate declines with the age. Highest unemployment rate is reported from educated group. It can be highlight that long-terms unemployment amongst Sri Lankans mainly due to their extended job search.

**Keywords**— Educated group, Global problem, Unemployment, Unemployment duration.

## I. INTRODUCTION

Unemployment rate is a powerful statistic that shapes government policies and personal decisions and key measure of economic health. If persons above a specified age not being in paid employment or self-employment but currently available for work during the reference period can be defined as unemployment. Unemployment and under employment is a global problem which is negatively affected to the economic as well as social development of a country. In 2019 world unemployment rate is 5.395 [worldbank.org] and in the Sri Lankan context it is recorded as 4.2. High unemployment rates are observed irrespective of the level of economic growth, where unemployment responds marginally to changes Gross Domestic Product. This is highly affected to the person's happiness because becoming unemployed is one of the worst experience to a person can have similar in its impact to divorce or bereavement (Ham, and Rea, 1997). In economic terms it enforces a massive problem to the economy of a country and it will most likely present a major social problem.

Unemployment duration can be long term or short term and It is defined as the length of time an individual spent without engaging economically active population. The definition of long-term and short-term unemployment duration differs by country to country. Unemployment duration is a measuring tool which used to estimate the labour market conditions and welfare of the labour market. A healthy labour market is characterised by high employment and low

unemployment, where the later has small incidence of long-term unemployment (South West Observatory, 2008).

There is a queue unemployment in Sri Lanka. According to Isenman (1980) and Glewwe (1987) there is a queues specially for state jobs. This queuing workers want to enter not only for well-paid state jobs but also well paid private sector jobs. That means unemployment in Sri Lanka arise due to the attractiveness of being unemployed while waiting for occupation in the high wage sector. And also there is a significant difference of the unemployment in between rural, urban and estate sectors. Employed rural population is 72.5% but for urban sector only 19%. This is due to the government policy to adoptive rural development by providing high level of public services to the rural sector.

At present it can see that well-educated first time job seekers, young men being unemployed. Many of these unemployed are looking for a public sector job and unemployment duration is higher for those aspiring to a public sector job. Thus the nature of unemployment in Sri Lanka is similar to other developing countries like Ethiopia (Rama 1999), Tunisia (Rama, 1998), and China (Appleton et al., 2001).

Labour market policy, could impact positively on the length of unemployment includes training programmes as in long-term unemployed individuals experience skills depreciation. According to Pissarides (1992) deterioration of Human capital increases the length of stay in unemployment. Ciuca & Matei (2010) considered unemployment duration as an important variable which explains change in the labour market.

In recent years, various survival analysis techniques popular for modelling the unemployment in the Social Science field. Tansel and Tasci (2005) found out the factors of the probability of existing unemployment in Turkey using parametric and nonparametric methods. According to his findings age has a negative effect on the hazard rate. Kupets (2006) proposed that there is a negative association between age and probability of reemployment. D'Agostino and Mealli (2000), developed a Cox proportional hazards models to examine unemployment duration in several European countries, with mixed results. Danacica et al. (2009) used semi parametric Cox regression approach to model the unemployment duration in five Central and Eastern European countries to analyse the Romanian labour market.

United Kingdom (Narendranathan & Stewart, 1993); Germany (Kuhlenkasper & Steinhardt, 2011) and Slovenia (Borsic & Kavkler, 2009) conducted researches to identify the

determinants of long-term-unemployment using survival analysis methods. According to the findings of these researches the effect of unemployment income and the chances of finding a job are decreased in long-term unemployment and this problem remarkably affected to female and adults whose age is greater than 34 years. D’Agostino and Mealli (2000) found that female have lower chances of leaving unemployment in Belgium, Greece, France, Spain, Denmark, and Portugal using Cox proportional hazards models and same result obtained by Tansel and Tasci (2005) with respect to Turkey.

Several previous studies are methodological. Beamonte and Bermúdez (2003) examine a Bayesian additive model and applied that model transitions to first-time employment of graduates. Most of the studies addressed specific research questions. Knut and Zhang (2003) examined the influences of unemployment compensation on the duration of unemployment. Gallo et al. (2006) examined the effect of job loss in older workers on the incidences of heart attacks and strokes. Carol (2006) used survey data and incorporated search theory into his methodology while Ludemann et al. (2005) used quantile regression model for administrative data to examine the length of unemployment duration in West Germany.

II. PROBLEM STATEMENT

Unemployment is an indicator of labour under-utilization and provides information on prevailing labour market drooping. Unemployment in Sri Lanka is a significant problem and one of the most serious socio-political problems faced by the Sri Lankan government. In Sri Lanka available labour lacks the skills demanded by the labour market thus there is a disparity between the supply and demand of labour. Lack of proficiency affect individuals’ employment probabilities. Slow economic growth rate also contributed to the lack of professional skills.

When consider about the annual unemployment rate in Sri Lanka it can see that the rate is stagnant between 4 to 5 percent but the changes are not statistically significant between some consecutive years, when sampling error is considered.

The number of unemployed persons in Sri Lanka is estimated as 372,593 during the year 2018. Out of this total, 44.0 percent are males and 56.0 percent are females. unemployment rate reported for the year 2018 is 4.4 at national level and, the rates for the Urban, Rural and Estate sectors are 4.1, 4.5 and 4.7 percent respectively. It could also be seen that, at the national level, the unemployment rate for females is more than two times higher than that of the male unemployment rate (Sri Lanka labour force Survey,2018). At the same time, female underemployment is higher than that of male.

Dias and Posel (2007) suggested that unemployment rate can be reduced by increasing the education. But, In Sri Lanka according to the labour force Survey, 2018, highest unemployment rate is reported from educated and which is reported as 9.1 percent. It is 5.1 percent and 13.2 percent for male and female respectively. As reported in previous years of

survey the problem of unemployment is more acute in the case of educated females than educated males.

Among the unemployed graduates, about 59.1 percent are Art degree holders while the other 40.9 percent consist with other degree holders (Commerce, Science and Mathematics). Borsic & Kavkler (2009) showed that persons with professional college degrees or bachelor’s degrees are better off than unemployed persons with a master’s degree. A study by Babucea & Danacica (2007) suggest a negative relationship between unemployment and education.

Generally developing countries like Sri Lanka face major challenges regarding the quality of available work for youths (age between 15 – 24 years). In Sri Lanka youth unemployment rate is 21.4 percent in 2018. When consider about the gender both male and female youth unemployment rates observed slight increasing trend over the years 2013 to 2018. Specially, youth unemployment rate is higher among the educated youths.

In Sri Lanka, unemployment rates of youths are significantly higher than adults but, the unemployment problem is not specific to the youth, but is country wide. On the other hand, the proportion of youth unemployment in total unemployment is also high, and this signpost an unequal distribution of the unemployment problem within the country. According to the labour force Survey,2018, youth unemployment to the total youth population is 6.4 and this proportion is 6.7 and 6.2 for males and females respectively.

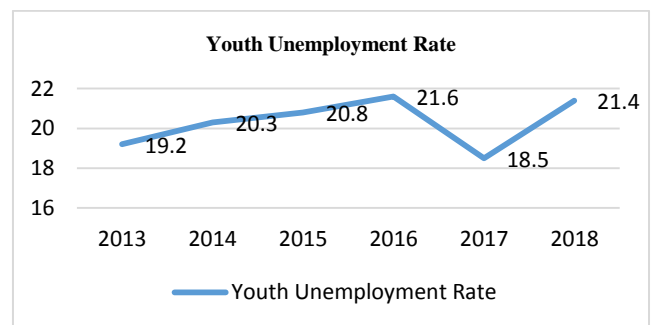


Fig. 1: Distribution of the Youth Unemployment Rate in Sri Lanka  
Source: Labour Force Survey – Annual Report 2018

Table 1: Youth Unemployment Duration as a Proportion of Total Unemployment by Gender and Level of Education  
Source: Labour Force Survey – Annual Report 2018

Factor	Proportion of Youth Unemployment
<b>Gender</b>	
Male	61.4
Female	46.4
<b>Level of Education</b>	
Grade 5 & below	16.8
Grade 6 – 10	62.3
GCE (O/L)	58.2
GCE (A/L) & above	46

Fig. 1 illustrate that, the youth unemployment rate is 21.4 percent in 2018 and that rate is considerably higher than the year 2017. Lowest youth unemployment rate is report in the year 2017 and highest rate is reported in the year 2016. The highest proportion of youth unemployment to youth population, 13.8 percent is shown among educated youth

group and it is 10.6 percent and 15.8 percent for male and female respectively.

According to the table 1, males shows a higher proportion of youth unemployment as a proportion to the total unemployment compared to that of female. Further the highest proportion of youth unemployment is reported from Grade 6-10 group in level of education.

There is limited literature with respect to the unemployment compared to the other countries in national level. Thus, there is a necessity to study about this critical problem based on Sri Lankan Scenario. The Sri Lankan researches about unemployment duration has not been extensively researched. It is rather skewed towards determination of factors that affect exits probabilities.

### III. OBJECTIVES

There is a necessity of statistical tool for examine the empirical evidence on the determinants associated with long term and short term unemployment. Thus the objective of this study is providing a statistical tool to model unemployment duration to enhance the dynamics of the unemployed.

### IV. METHODOLOGY

#### Data

Basically used primary data from the individuals in Kalutara district in Sri Lanka. Two stage stratified sampling procedure is adopted to select a sample of 1200 housing units to be enumerated at the survey.

Primary Sampling Units were allocated to each sector (Urban, Rural and Estate) by using the Neymann allocation method which considers the variance of unemployment rate as usually. The allocated sample for each sector then equally distributed for 12 months.

#### Statistical Techniques

Danacica & Babucea (2010) reveal that duration data requires a different statistical analysis compared to quantitative data because these data are not normally distributed and frequently encompass censored subjects. The probability of leaving unemployment to employment has an instantaneous risk of occurring hence is unreasonable to assume normality (Cleves, et al., 2004).

Censoring is one aspect of duration data and difficult to use traditional statistical model to analyse duration data (Rao & Schoenfeld, 2007). According to Lu & Shen (2014) linear regression cannot deal with the censored data correctly and logistic regression does not consider the differences in the timing of event occurrence. Censoring also leads to biased hazard functions (Kiefer, 1988), and exclusion of censored subjects reduces the sample size thus complicating event analysis (Jakoet, 2007).

Survival data correspond to the time from a well define time origin until the occurrence of some particular event or end point. In many areas of application, the response variable is the survival time; life time and duration of machine equipment, time to complete a task and specially use in medical applications. In medical research the time origin will often correspond to the requirement of an individual into an

experiment study. Before '70s, the survival analysis was used to examine the death as an event specific to medical studies (Armitage, 1959, Narendranathan, and Stewart, 1993) and demographical studies (Balakrishnan, 1991, Carroll, 2005, and Foley, 1997), after '70s these techniques have been gradually used in the field of economics and social sciences. In present this Survival analysis technic widely used in socio economic researchers to examine the complex phenomena and Survival analysis, adapted in conventional econometric modelling data, received the title of duration models (Klein, et al, 2005 and Kiefer, 1988).

There are some special features of survival data analysis. Mainly, the data are highly skewed and non-normal with few large survival times and in most cases the event of interest is never observed in all the subjects due to the occurrence of censored data. Censored data may be on the right or left. According to Gehan (1969) large amount of censored data will badly affected to the accuracy of the statistical tests.

There are parametric, semi parametric and nonparametric methods which can be used for analysing survival data. Nonparametric methods and semi parametric methods are utilizing to analyse the data in this research because in the research it is not making any assumption about the functional form of hazard.

In this research 'alive' refers remaining unemployed and 'death' refers finding a job. The survival function  $[S(t)]$  is remaining unemployed 't' unit of time from the beginning of the study. The survival function demonstrates about the proportion of an individual who unemployed remains unemployed as time passes.

Cox regression based test is used as a substitution for log rank test because of its advantage to incorporate sample design features and to observe the connection of the survival distribution to covariates. The scale on which linearity is assumed is the log-hazard scale. A linear model for the log-hazard is:

$$\log h_i(t) = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} \tag{1}$$

Where  $i$  is a subscript and 'x' are the explanatory variables. The constant ' $\alpha$ ' denotes log-baseline hazard, since  $\log h_i(t) = \alpha$ , when all of the explanatory variables are zero. The Cox model leaves the baseline hazard function ' $\alpha(t) = \log h_0(t)$ ' unspecified:

$$\log h_i(t) = \alpha(t) + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} \tag{2}$$

This model is semi-parametric because while the baseline hazard can take any form, the covariates enter the model linearly. Two observations 'i' and 'i'' that differ in their x-values, with the corresponding linear predictors.

$$\eta_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} \tag{3}$$

$$\eta_{i'} = \beta_1 x_{i'1} + \beta_2 x_{i'2} + \dots + \beta_k x_{i'k} \tag{4}$$

The hazard ratio for these two observations,

$$h_i(t) / h_{i'}(t) = h_0(t) e^{\eta_i} / h_0(t) e^{\eta_{i'}} = e^{\eta_i} / e^{\eta_{i'}} \tag{5}$$

is independent of time 't' and Cox model is a proportional-hazards model.

The null hypothesis (H0) and alternative hypothesis (H1) for the test are;

H<sub>0</sub>: Unemployment duration is not influenced by covariate Xi.

H<sub>1</sub>: Unemployment duration is influenced by covariate Xi.

The null hypotheses are rejected at a 5% level of significance and if the value of the Wald chi-square is greater than the critical value within each covariate.

To find out the degree of the association between independent variables and the unemployment of Sri Lanka used Binary Logistic Regression by considering the dichotomize nature of the dependent variable whether the individual is unemployed or not).

Taking 'Y<sub>i</sub>' as a binary variable, Logistic Regression can be defined as,

$$Y_i = 1 \text{ with probability } \pi$$

$$Y_i = 0 \text{ with probability } 1 - \pi$$

$$Y_i \sim \text{Bernoulli}(\pi)$$

Where,

$$\pi = \exp(\eta_i) / 1 + \exp(\eta_i) = 1 / 1 + \exp(-\eta_i)$$

The corresponding Logistic Regression model is,

$$\eta_i = \beta_0 + \beta_1\chi_{1i} + \beta_2\chi_{2i} + \beta_3\chi_{3i} + \beta_4\chi_{4i} + \epsilon_i$$

In other words,

$$\text{Logit}(Y) = \text{natural log}(\text{odds}) = \ln[\pi / (1 - \pi)] = \alpha + \beta\chi$$

$\pi$  = Probability (Y=outcome of interest/ X=x, a specific value of X)

$$= [(\alpha - \beta\chi) / (1 + \alpha + \beta\chi)]$$

In this study dependent variable is dichotomize as 'being unemployed' (01) or 'being employed' (0).

### V. SIGNIFICANCE OF THE STUDY

There are few studies of unemployment duration in developing countries. Through a scientific analysis, this study will provide useful information for policy makers, economists and academic researchers about the unemployment and underemployment in Sri Lankan context and also the nature and magnitude of the effects. This will also give directions for academic researchers in their future studies.

### VI. RESULTS OF THE STUDY

Table 2: Distribution of the Unemployed Population by Demographic Factors  
Source: Survey Data, 2019

Demographic Factors	Percentage
<b>Gender</b>	
Male	30
Female	70
<b>Living Sector</b>	
Urban	41
Rural	45
Estate	14
<b>Age Group</b>	
15 - 24	41
25 - 29	30
30 - 39	19
>40	10
<b>Level of Education</b>	
Grade 5 & below	17
Grade 6 - 10	22
GCE (O/L)	28
GCE (A/L) & above	33

According to the table 2, female unemployment rate is 70% and that is more than two times of the male unemployment. The majority among unemployed are young people who are in age group 20 – 24 years and it shows that unemployment rate

declines with the age. Highest unemployment rate is reported from educated group.

Table 3: Distribution of Unemployment by Type of Degree  
Source: Survey Data, 2019

Type of the Degree	Percentage
Art Degree	69
Other Degrees	31

Table 3 shows that, among the unemployed graduates, more than half of unemployed are Art degree holders.

Table 4: Distribution of Unemployment by Nearby Reasons for not Seeking Employment  
Source: Survey Data, 2019

Reason for not Seeking Employment	Male (Percentage)	Female (Percentage)
Sickness	5	3
Discouragement	22	25
Family Responsibility	35	52
Higher Studies	28	12
Entrepreneurial need	10	8

According to the table 4, the main reason for inactivity is 'Family responsibilities', for both the gender because which is publically known as a household duties of females in most of the Asian countries. Apart from that, 'Higher studies' and 'Discouragement' on labour market due to difficulties in finding suitable jobs are key reasons for economic inactivity among both male and female.

Table 5: Distribution of Unemployment by Reason of Job Seeking Sector  
Source: Survey Data, 2019

Reason	Public Sector %	Private Sector %	Foreign Sector %	Self-Employee Sector %
Job Security	78	10	0	12
Pension	100	0	0	0
Personal Ambition	48	22	5	25
Relevance to Qualification	42	48	10	0
Salary	37	41	12	10
Freedom	37	0	6	57
Working Environment	44	18	8	30
Hating Public Sector	0	78	10	12
No Specific Reason	58	35	7	0

Table 5 shows that most of the job seekers have preferred the public sector, mainly based on the 'pension' and 'job security'. In contrast large more than 75% of the individuals' choice was private sector because of 'hating public sector'. More than half of the job seeking individuals had preferred self-employment avenues, owing to their 'freedom'. Relatively less number of individuals' job seeking sector is foreign sector and most of them choose this sector due to salary.

Out of the total 1200 respondents taken for this study, few number of individuals had declared that they had searched for job for less than 3 months. The majority had search for suitable jobs more than 12 months. This illustrates a long-

terms unemployment amongst Sri Lankans mainly due to their extended job search.

Table 6: Distribution of Respondents in terms of Job Search Duration  
Source: Survey Data, 2019

Job Seek Duration	Percentage
Less than 3 months	10
3 – 6 months	16
6 – 12 months	32
More than 12 months	42

*Test the Independence of the Variables*

It is essential to test whether there is a relationship between the dependent variable and independent variables. For that chi-square test was used.

*Hypothesis*

- H1a: unemployment is depending on gender of an individual.
- H1b: unemployment is depending on living sector of an individual.
- H1c: unemployment is depending on age of an individual.
- H1d: unemployment is depending on level of education of an individual.
- H1e: unemployment is depending on language constraints.
- H1f: unemployment is depending on the professional qualifications.
- H1g: Selection of private bank is depending on the job seeking sector.

All the above hypotheses were tested using chi-square test statistic under 5% level of significant. The variables; Gender, living sector, level of education, language constraints, professional qualifications were significant. That means all those determinants are affected to the unemployment.

*Step Wise Selection Procedure*

There are many predictors and the major interest of this report is to identify important predictors. That means, in this report it wishes to identify a small subset that relates significantly to the outcome. Therefore, stepwise selection procedure used to select a good model.

*Logistic Regression Analysis*

In this study it had been consider about five variables based on the results of Chi-Square test. Series of models had been fitted by adding and removing the variables in order to assess whether all the five variables are necessary for prediction or whether any could be dropped. Results are judged using AIC (Akaike's Information Criterion), likelihood ratio test. Also, p-value (used 5% significant level) is used to find out whether the variables are significant or not.

$$\log it(\pi) = \beta_0$$

P value of the intercept term is considerable at 5 % level significant. Though the fitted null model is significant, the mean deviance is 1.2484 thus the current model should be improved. Only the intercept term does not have a meaning when describing the model.

*Results of the Initial Model*

Table 7: Model with Variable Gender  
Source: Survey Data, 2019

Variable	Wald	P value	Exp(B)
Female	4.102	0.043	1.188
Constant	12.591	0.000	0.789

In gender variable, 'male' cannot be identified under the control of last level zero. 'Female' is significant and it is nearly 48% probability. The odds of being female unemployed is 1.2 times than male unemployed.

Table 8: Model with Variable Living Sector  
Source: Survey Data, 2019

Variable	Wald	P value	Exp(B)
Estate	211.562	0.000	4.475
Rural	43.690	0.000	1.502
Constant	261.964	0.000	0.392

The probability of being unemployed in 'estate' sector and 'rural' sector is 63.7%, and 37% respectively. Being an unemployed person in estate sector shows 4.475 times and rural sector shows 1.5 times the odds of urban sector for the response variable.

Table 9: Model with Variable Level of Education  
Source: Survey Data, 2019

Variable	Wald	P value	Exp(B)
GCE (A/L) & above	89.744	0.000	7.089
GCE (O/L)	55.090	0.000	4.603
Grade 6 – 10	9.351	0.002	1.935
Constant	104.293	0.000	0.124

All the variables in the level of education is significant at 5% significance level. The odds of 'GCE (A/L) & above' is 7.08 times, 'GCE (O/L)' is 4.603 and 'Grade 6 – 10' is 1.935 times comparatively the odds of 'Grade 5 & below'. Further, probability of being unemployed comes under 'GCE (A/L) & above', GCE (O/L), and 'Grade 6 – 10' are 46.8%, 36.4%, and 19.4% respectively. All the terms in this model are significant.

Table 10: Model with Variable Language Constraint  
Source: Survey Data, 2019

Variable	Wald	P value	Exp(B)
Yes	4.972	0.026	0.710
Constant	7.668	0.006	0.675

In the table 10, 'language constraints are not affected for unemployment' category is not defined according to the constraint of last level zero. 'language constraints are affected for unemployment (yes)' shows 32% probability for unemployment.

Table 11: Model with Variable Professional Qualifications  
Source: Survey Data, 2019

Variable	Wald	P value	Exp(B)
No	4.815	0.028	1.385
Constant	5.614	0.018	0.714

Table 11 shows that probability of being unemployed without professional qualification (No) is 49.7%.

According the results of the table 7-11 the variables Gender, living sector, level of education, language constraints, professional qualifications are most important determinant for

being unemployed. Then a series of models fitted in order to assess whether all the five variables are necessary for prediction or whether any could be dropped. For that, it should identify the most important determinant among these five significant determinants. The variable ‘living sector’ is the most important variable among others because that variable has the lowest AIC. Then the fitted model is,

$$Logit(\Pi_i) = \beta_0 + \beta_i^{Living\ sector}$$

Keeping the variable ‘living sector’ in the model, the remaining four factors were added one by one to the model and checked the significance of additional variable. The results were judged using likelihood ratio test and Akaike’s information criterion. After the variable ‘gender’ added to the current model it suggested that ‘gender’ is needed to the model in addition to ‘living sector’. Here, p-value of the variable ‘living sector’ is significant after adjusting for the variable ‘gender’. The fitted model is shown in bellow.

$$Logit(\Pi_{ij}) = \beta_0 + \beta_i^{LivingSector} + \beta_j^{Gender}$$

When ‘living sector’ and ‘gender’ are already in the model, the significance of other variables was checked. Additional variable was checked using likelihood ratio and test whether those variables are necessary for prediction or any could be removed. Among significant models, variable ‘level of education’ showed the lowest AIC value when ‘living sector’ and ‘gender’ were already included in the model.

$$Logit(\Pi_{ijk}) = \beta_0 + \beta_i^{Living\ sector} + \beta_j^{Gender} + \beta_k^{Educationlevel}$$

To investigate, further important variables which should be included in the fitted model and to exclude the unnecessary variables, it was tried to develop model when the variables ‘living sector’, ‘gender’, and ‘level of education’ exist already in the model. Then the variable ‘language constraints’ was significant.

$$Logit(\Pi_{ijkl}) = \beta_0 + \beta_i^{Living\ sector} + \beta_j^{Gender} + \beta_k^{Educationlevel} + \beta_l^{Language}$$

All the main effects are considered and it can be concluding that, none of the models significant after the model with four variables.

Then, check the significance of the models with interaction terms.

Table 12: Model with Interaction Terms  
Source: Survey Data, 2019

$Logit(\Pi_{ijkl}) = \beta_0 + \beta_i^{Living\ sector} + \beta_j^{Gender} + \beta_k^{Educationlevel} + \beta_l^{Language}$	Difference Of (-2log likelihood)	P Value
Living sector * Gender	6.758	0.343
Living sector * Education Level	6.782	0.662
Living sector * Language	13.415	0.144
Gender * Education Level	6.218	0.399
Gender * Language	3.39	0.758
Education Level*Language	4.328	0.632

According to table 12, It is clear that none of the models is significant. It indicates that model with two-way interaction terms not significant. Furthermore, it can be concluded that two-two way, two-three way or more are not significant. Associated p value for each interaction term is not significant. All the interaction terms should be removed from the model

and should be considered only the complete independent model with four main effects.

Finally, consider about the overall result, the final significant model can be shown as bellow and there are no any interaction terms with regarding to this model.

$$Logit(\Pi_{ijkl}) = \beta_0 + \beta_i^{Living\ sector} + \beta_j^{Gender} + \beta_k^{Educationlevel} + \beta_l^{Language}$$

In this model there are 1200 binary observations and success is happening 795 times. The deviance of the resulting model is 0.9282. For that reason, it is possible to conclude that the model is good because, the value of the deviance is close to one. In addition to that, the P-values related to the parameters in the model are significant at 5% level. This summarized that the model is in the condition of significant.

### Logistic Regression Diagnostic

After selecting the preliminary model, goodness of fit of the model has to be tested. With categorical predictor, it is useful to form residuals to compare observed and fitted counts. To test the goodness of test was used Standardized Pearson’s Residuals. Absolute value larger than roughly two or three are provide evidence of lack of fit. (if  $|e_i| > 2$  lack of fit).

To test the goodness of fit for the selected model from the former step and results was shown in following plot.

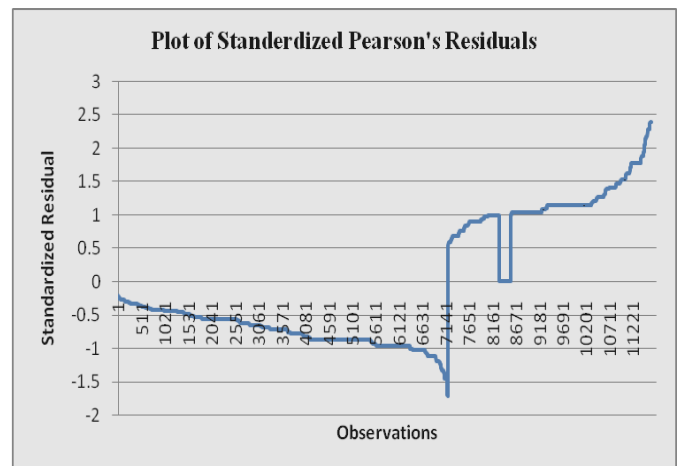


Figure 02: Plot of the Standardized Pearson’s Residuals  
Source: Sample Survey, 2019

According to figure 02, it can be illustrating that residuals are approximately lie in the exception are except nearly 40 observations. That observation holds nearly 0.03% out of the region when consider of the whole observation. Undoubtedly, to come conclusion that the fitted logistic regression model for the data is matched well.

### Survival Functions and Cox Models

In this section, examine the relationship of the explanatory variables with unemployment duration using Cox regression. The selected explanatory variables are gender, living sector, level of education, language constraints and professional qualifications. These factors regressed against the ‘unemployment’ so as to determine to what extent such associations affect the unemployment of Sri Lanka.

Table 13: Results of the Cox Regression Model  
Source: Sample Survey, 2019

Covariate	Coef	Exp (Coef)	SE (Coef)	P
Age	-0.02	0.98	0.006	0.0007
Gender	0.021	1.02	0.02	0.14
Level of Education	-0.27	0.76	0.1	0.00

According to the results in table 13, all the three explanatory variables are statistically significant. The exponential coefficients are interpreted the multiplicative effects on the hazard. When age increased by one will cause to decrease the monthly hazard of employment by 0.98 on average remaining all the other covariates constant. Similarly, education level is higher will cause to decrease the corresponding hazard by 0.76 on average.

Checking the Proportional Hazards Assumption

The mean of scaled Schoenfeld residuals from the final model at a particular time is measure how far the time-varying coefficient differs from the one fitted by the proportional hazards model. A plot of this scaled residual against time for each parameter estimate should vary randomly about a horizontal line. Though here are many graphs to show them all the variable age is typical and it shows a more or less horizontal fit. Consequently, the coefficient does not depend on time.

VII. DISCUSSION

The results of this study may be considered as a picture of the factors shaping unemployment and unemployment duration, which may be helping policy makers as well as for the researches who expected to research about this global problem specially in Sri Lankan context. From the results of the Binary Logistic Regression and Cox Regression analysis conclude that the variables living sector, gender, level of education and language constraints are significant with unemployment and age, gender and level of education are associated with unemployment duration of a person. It could highlight that, unemployed persons with higher education are not effectively participate to the labour market; may be it is more difficult for them to engage with a job which is not related to their qualifications. It is important to noticed that, there are strong differences between women and men, concerning the unemployment. The unemployed persons from the age group 15 – 29 years have a higher probability of remaining unemployed and this may be due to the difficult for them to adapt to the new labour market conditions. The future researchers can investigate which factors are causing for these differences.

REFERENCES

[1] A.G. Babucea, and D. Danacica, "Using Kaplan-Meier curves for preliminary evaluation the duration of unemployment spell", *Annals of the University "Constantin Brancusi" of Targu Jiu*, Issue 2, pp. 33-38, 2007.

[2] Kavkler, D. E. Danacica, A.G. Babucea, I. Bicanic, B. Bohm, D. Tevdovski, K. Tosevska, and D. Borsic, "Cox regression models for unemployment duration in Romania, Austria, Slovenia, Croatia and Macedonia", *Romanian Journal of Economic Forecasting*. (9), 2009.

[3] D'Agostino, Antonella, and Fabrizia Mealli, "Modelling Short Unemployment in Europe", *Institute for Social & Economic Research Working Paper 06*, 2000.

[4] D.Borsic, and A. Kavkler, "Modeling unemployment duration in Slovenia using Cox regression models", *Transition studies review*, 16(1), pp. 145-156, 2009.

[5] E. A. Gehan, "Estimating Survival Function for the Life Table", *Journal of Chronic Diseases*, 21 629-44, 1969.

[6] E. L. Kaplan, and P. Meier, "Nonparametric estimation from incomplete observations", *Journal of the American Statistical Association*, Vol. 53, No 282, pp 457-481, 1958.

[7] E. Lüdemann, R. A. Wilke, and X. Zhang, "Censored Quantile Regressions and the Length of Unemployment Periods in West Germany", Discussion Paper No. 04-57, 2005.

[8] G. J. Johnson, and W. R. Johnson, "Perceived over qualification, positive and negative affectivity, and satisfaction with work", *Journal of Social Behaviour and Personality*, 15(2), 167-184, 2000a.

[9] J. P. Klein, and M. L. Moeschberger, "Survival Analysis: Techniques for Censored and Truncated Data", New York: Springer Verlag, 2005.

[10] J. C. Ham, and J.R. Rea, "Unemployment Insurance and Male Unemployment Duration in Canada". *Journal of Labour Economics*, 5(3): 325-53, 1997.

[11] Kupets Olga, "Determinants of Unemployment Duration in Ukraine", *Journal of Comparative Economics*, 34 (2): 228-247, 2006.

[12] M. C. Foley, "Determinants of Unemployment Duration in Russia", *Yale Economic Growth Centre Discussion Paper 779:39*, 1997.

[13] M. Rama, "How Bad is Unemployment in Tunisia? Assessing Labour Market Efficiency in a Developing Country", *The World Bank Observer* 13(1): 59-77, 1998.

[14] M. Rama, "The Sri Lanka Unemployment Problem Revisited", Washington DC,

[15] The World Bank, Development Research Group: 45, 1999.

[16] N. Carroll, "Explaining Unemployment Duration in Australia", 2005.

[17] N. Carroll, "Explaining Unemployment Duration in Australia", *Economic Record*, 82 (258), 298-314, 2006.

[18] N. M. Kiefer, "Economic Duration Data and Hazard Functions", *Journal of Economic Literature* 26: 646-679 1988.

[19] P. B. G. Armitage, "Statistical Methods in Medical Research", Blackwell, 1959.

[20] R. Dias, and D. Posel, "Unemployment, education and skills constraints in post-apartheid South Africa", *Development Policy Research Unit, working paper no. 07/120*, 2007.

[21] S. Appleton, and J. Knight, "Labour Retrenchment in China: Determinants and Consequences", mimeo, 2001.

[22] T. Korpi, "Effects of Manpower Policies on Duration Dependence in Reemployment Rates: The Example of Sweden", *Economica* 62(247): 353-71. [5], 1995.

[23] T. Kuhlenkasper, and M. F. Steinhardt, "Unemployment duration in Germany - A comprehensive study with dynamic hazard models and P-splines", *Discussion paper no. 2011- 18 Norface: Norface migration*, 2011.

[24] Tansel, Aysit, and H. Mehmet Tasci, "Determinants of Unemployment Duration for Men and Women in Turkey", *IZA Discussion Paper no. 1258*, 2005.

[25] W. Narendranathan, and M. B. Stewart, "Modelling the probability of leaving unemployment: Competing risks models with flexible baseline hazards", *Journal of the royal statistical society, Series C (applied statistics)*, 42(1), pp. 63-83, 1993.

[26] <http://www.statistics.gov.uk/LabourForce/StaticInformation/AnnualReports/2018>

[27] <https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS>