Complementary Log-Log with Random Effect Model Using Malaysian Graduate Employability Data

Md Azman Shahadan, Rahmattullah Khan Abdul Wahab Khan, Aimillia Mohd Ramli

Abstract: ***The objective of this research is to investigate the relationship between hazard ratio or survival function of graduate employability and 8 explanatory variables. The 8 explanatory variables are as follows: gender, CGPA, geographic region, English language proficiency, area of study, curriculum satisfaction, carrier guidance services satisfaction and monthly family income. In order to study a survival function of graduate employability, we have developed a sequence of binary numbers (employed (1) or unemployed (0)) at a particular time of being employed among first degree graduate students (N=2228). The data is based on a survival study data, which traces the survival function of graduate students within 12 months of window opportunity. We have used the complementary log-log model in various forms. For the simple complementary log-log model, the results showed that the hazard risk with baseline hazard ratio relate to graduates' gender, CGPA, satisfaction with career guidance, the geographic region from which they came from, their English performance, their area of study and their family income. As for subject-specific (random effects), the hazard ratio also relates to the abovementioned variables.***

Keywords: Modeling discrete time to event, Employability data, Survival function, subject-specific proportional hazards model, Random Effect Complementary Log-Log Regression Model

I. INTRODUCTION

Amongst the common questions posed by the Malaysian public are: How long does it take for Malaysian first-degree graduates to become employed after graduation? Why are a number of well-educated Malaysian graduates unemployed? In order to answer these questions, every year the Malaysian Ministry of Higher Education (Kementerian Pendidikan Tinggi) conducts an employability survey among first-degree graduates from all Malaysian public universities (IPTA). In the survey, a series of questions are asked about the duration and factors that have contributed to their employment. In the context of our research, unemployment refers to the duration or the amount of time that an individual remains unemployed. Clearly, the response of interest is the duration of time that has passed between the time the respondents completed their studies and the time they manage to become employed.

Revised Manuscript Received on April 15, 2019.

Md Azman Shahadan, Faculty of Human Development, Sultan Idris Education University Malaysia

Rahmattullah Khan Abdul Wahab Khan, Faculty of Human Development, Sultan Idris Education University Malaysia

Aimillia Mohd Ramli, Kulliyyah of Islamic Revealed Knowledge and Human Science, International Islamic University, Malaysia

The survey dataset comprises of graduate students whose status of employment are observed from their day of graduation for up to 12 months afterwards from one public university in Klang Valley, Kuala Lumpur, Malaysia. Analysis of this kind of survey dataset requires the use of a good statistical method or model. A rather crude way to model the employability data is to use the logistic model, as demonstrated in a study by Nafi & Ghani (2011) or by using just descriptive statistics as shown in a study by Ayiesah et. al (2010). If we were to use the above method, we would certainly end up losing a lot of information. The loss of crucial information in statistical modeling is highly regrettable. A good statistical model should handle the data on duration models, which can take into account time to event data. In other words, a good statistical model should provide information about the amount of time or transition period from unemployment to employment. In literature concerning unemployment duration time, findings on the relationship between the duration of unemployment and one or more explanatory variables is of the most interest.

Basically, our graduate employability data addresses two important questions: The first is how do the former students manage to get their jobs after they graduate? The second question is what are the factors that are related to their employability? There have been efforts to find or to predict the relationship between employability and several factors, such as gender, CGPA, area of study, job preferences, job position, foreign language proficiency and international experience (Nafi & Ghani, 2011; Bryla, 2015). Furthermore, Ayiesah, Roslizawati, & Maslyn (2010) conducted a study on the employees perception of employability skills among physiotherapy graduates by using the survey method. They found that critical thinking ability, ability to generate hypotheses and linking ideas by applying theory into practice, sharp analytical skills, prioritizing of problems, keeping up-to-date on latest information regarding their professions, and giving clear explanation about problems and treatments are essential skills for employment. On the other hand, Shaharuddin, et. al. (2014) conducted a study on the marketability of UKM graduates, among 273 second-year students using the survey method. They found that academic and social soft skills are the two skills that are required by our graduates. In addition, Stojanova, & Blaskova, (2014) explored the graduates' chances of success in the labour market after finishing their chosen field of study.

Their study proved that the best employment opportunities are opened to students of technical disciplines. Torres-Machi, Carrion, Yepes, & Pellicer (2013) identified students' perceptions of training gaps that affect their employability. Around 38 and 44 students had enrolled in PMaCE (September 2010 and September 2011) from the third- and fourth-year classes, respectively, and participated in this study. The results indicate that the training gaps that affect the employability are unwillingness to move to another country, their lack of knowledge of a foreign language and communication skills, their preferences for only well-paid and comfortable jobs, economic policy, training gaps, labour market structure, graduate surplus, and setbacks related to business.

Sarfraz, Rajendran, Hewege, & Mohan (2018) conducted a systematic review on employability skills in 43 studies across 17 countries. The results indicated that the skills can be classified into 10 different skill sets. The skills included interpersonal and collaborative skills, relationship management skills, cognitive problem solving skills, productive self-management skills, creative and innovative skills, new technology adaptation skills, personal attributes and individual differences, lifelong learning skills, leadership skills and global citizenship skills.

II. FOCUS OF PRESENT STUDY

The main aim of the present study is to bridge the gap between a substantive theory of employability among graduate students and new developments in statistical modeling. The simple logistic model could be carried out in order to address questions on the relationship between getting employed (response variable) and several factors such as CGPA (exploratory variables). However, the employability data refers to the amount of time that an individual remains unemployed. Studies on transition underwent by graduates from unemployment to employment is important because it can be used to evaluate policies and recommend ways to facilitate these transitions (Ganjali & Baghfalaki, 2012). The times can be assumed as discrete times and then viewed as a series of events over consecutive time periods (i=1, 2,.., T) which can be represented by a binary sequence. This binary sequence is the standard way to estimate discrete time duration models by which we form panel binary data. This sequence can be modeled by binary choice random effect model, with a complementary log-log link. However, it would be easier if we follow the multilevel or hierarchical model framework.

It is important to note that employability dataset typically observed a spell or event over a sequence of interval such as months before the students finally obtain employment. For this kind of data, the more appropriate way to handle the dataset is to assume a discrete time to event. Therefore, the data modeling here concerns analyzing the duration of employment from the start of a spell of unemployment until the start of work among graduate students of 2010. Following Berridge & Crouchley, (2011), suppose we have binary indicators for individual *j*, which takes the value 1 if the spell ends in a particular interval *i* and 0 otherwise. Then individual j's duration can be viewed as a series of binary outcomes (i=1,2,..,). As a result, we only observe a single spell for each subject and this would be a sequence of 0s which would end with a 1, if the spell, is complete and 0, if it is right censored. The probability that Y_{ij} =1 for individual *j* at interval *i*, given that Yij= 0, $\forall i' < i$ is given by

$$\Pr(Y_{ij} = 1|\theta_{ij}) = 1- F(\theta_{ij}) = \mu_{ij.} \quad (1)$$

Here we can propose using complementary log-log link. This link function is more preferable (Ntzoufras, 2009; Congdon, 2010) as it gives,

$$\mu_{ij} = 1- \exp[-\exp(\theta_{ij})]. \quad (2)$$

The random intercept in the linear predictor takes the form of

$$\theta_{ij} = \beta_{0j} + \sum_p \beta_{pj} X_{pij} + K_{i.} \quad (3)$$

Here the K_i are interval-specific constants and the X_{pij} are explanatory variables with complementary log-log link. The challenges to estimate the above model are the likelihood function needed to integrate. The integration becomes analytically intractable, unless for normal distribution cases and necessitates the use of numerical integration techniques. The most famous numerical method is an adaptive-quadrature method. This method was implemented in most software, such as NLME routine in SAS, GLLAMM routine in STATA and lme4 routine in R.

III. THEORETICAL FRAMEWORK AND HYPOTHESES

By exploring all previous studies, we can conclude that our six factors propose models (area of study, backgrounds, languages, academic performances, and industrial training soft skills) were given significant attention by previous researchers. From all previous articles, soft skills was determined as the most crucial factor that is related to graduates' employability while industrial training and area of study were considered as less crucial factors. All these previous studies on employability were really helpful and beneficial to our study since they not only provided us with a handful of information related to graduates' employability but also a useful guidance to our proposed model as well (Rao & Jani, 2012); Nafi & Ghani, 2011; Zaliza & Mohd Safarin,2011).

IV. RESEARCH METHOD

The secondary data for this study was obtained from tracer studies (online survey) conducted by the Ministry of Higher Education Malaysia (MOHE) at all Malaysia public universities. The questionnaires were posted online a month prior to the convocation ceremony every year which is normally held in October. For this research, we had used data from a tracer study conducted at one of the universities in Klang Valley from year 2010. The respondents of this study were first-degree graduates' students from various areas of study. The objective of the trace study is to collect the graduate's opinion on their universities' programs, infrastructure and services that are provided over the course of their study, and how they enter and face the challenges of working life. Further information can be obtained from MOHE.

MOHE tracer study questionnaire comprises seven different sections that are aimed at obtaining information related to the graduates, namely, a) background information, b) evaluation of program and services offered by institution, c) effectiveness of study program and self readiness, d) further studies, e) current status, f) employment, and g) unemployed and others. In our study, we are only interested on the part of employment or unemployment status of graduate students.

At year 2010 data, 2341 graduate students were recruited to fill-in the questionnaire online. The respondents submitted their online questionnaires before their convocation ceremony. Participant consists of 756 male and 1588 female graduate students. We also consider only eight explanatory variables, namely; 1) gender, 2) CGPA, 3) geographic region, 4) English language proficiency`1, 5) area of study, 6) curriculum satisfaction, 7) carrier guidance services satisfaction, and 8) monthly family income. Missing process is assumed to be missing at random taxonomy (Little & Rubin, 2002). Furthermore, we treat missing data in exploratory variables as dropout and we used list-wise deletion method to analyse our survival model data. The list wise deletion is accomplished by deleting from the sample any observations that are missing data.

Table. 1 The Mean Score, Standard Deviation and Frequency for Explanatory Variables

Variables(Continuous Variable)	Mean	S.D.
CGPA	2.85	0.695
Curriculum	25.45	4.195
Career guidance	28.15	7.17
Teaching	56.15	8.159
facilities	12.43	2.13
Discrete variable	Frequency	
Gender	Male= 753; female=1588	
Geographical region	West Cost=1738; East Cost=537; Sabah/Sarawak=69	
English performance	Excellence=815; Good=850; Satisfy=612; fail=64	
Area of study	Art and social sciences=1228; sciences=371; technical=691; it=51	
Family income	Low=810; middle income=1164; high income=367	

V. DATA EXPLORATION

Before we start the modelling process, let us explore the data using an exploratory analysis. We have divided the exploratory analysis into two parts: In this section, we explore the correlation matrices for all proposed continuous exploratory variables. This would allow us to explore the extent to which the continuous explanatory variables are related to each other. This is because if there is a strong correlation among explanatory variables, the problem of multicollinearity would occur. The second part comprises techniques to visualize pattern in data, which allows us to uncover patterns that are unexpected and to discover some new clues. We employed Kaplan-Meier (KM) or product limit method to estimate survival functions of getting a first job. However, by using plot of KM, it would be difficult to determine whether two or more survival function curves are identical or different. In our research, we had used log-rank test (non-parametric test based on chi-square distribution (χ^2)) to compare whether there are any significant differences between two or more survival function curves (Kalbfleisch & Prentice, 2002). The null hypothesis being tested is that there is no overall difference between the two survival function curves. Under this condition, the null hypothesis being tested is that all survival curves are the same. We plot five estimates of survival function curve against duration to get employed in month.

Table. 2 Pearson Correlation among Continuous Explanatory Variables

Variables	CGPA	Curriculum	Career guidance	Teaching
CGPA	-			
Curriculum	.044			
Career guidance	.082	.500*		
Teaching	.055	.664*	.537*	
Facilities	.082	.647*	.467*	.537*

**p<.001*

Table *2* show the correlation between all continuous explanatory variables. Results shows that some of correlations appear to highly correlate with one other. According to Menard (1995) high levels of collinearity corresponds to high correlation coefficients among the explanatory variables. In order to prevent the problem of multicollinearity, we have to drop teaching and facilities variables from our analysis.

In figure 1, we plot five estimates of survival function curve against duration to get employed in months. Each plot produces individual survival function curve of remaining unemployed in the unit of months.

The first plot is the survival function for gender and this is followed by a survival function for geographical region. The third plot is survival function for English language proficiency, followed by survival function for area of study. The last survival function is for monthly family income. For plot one, the Kaplan-Meier survival probability (not get employed) estimates at 12 months were around 0.23 for males and 0.40 for females. The plot also indicates that female respondents appear to enter the job market at a much later time compared to male respondents. We tested the significant difference between the genders using the log rank test and the results showed a significant difference in survival times between the genders ($\chi^2(1)=13.41$, $p\leq.05$).

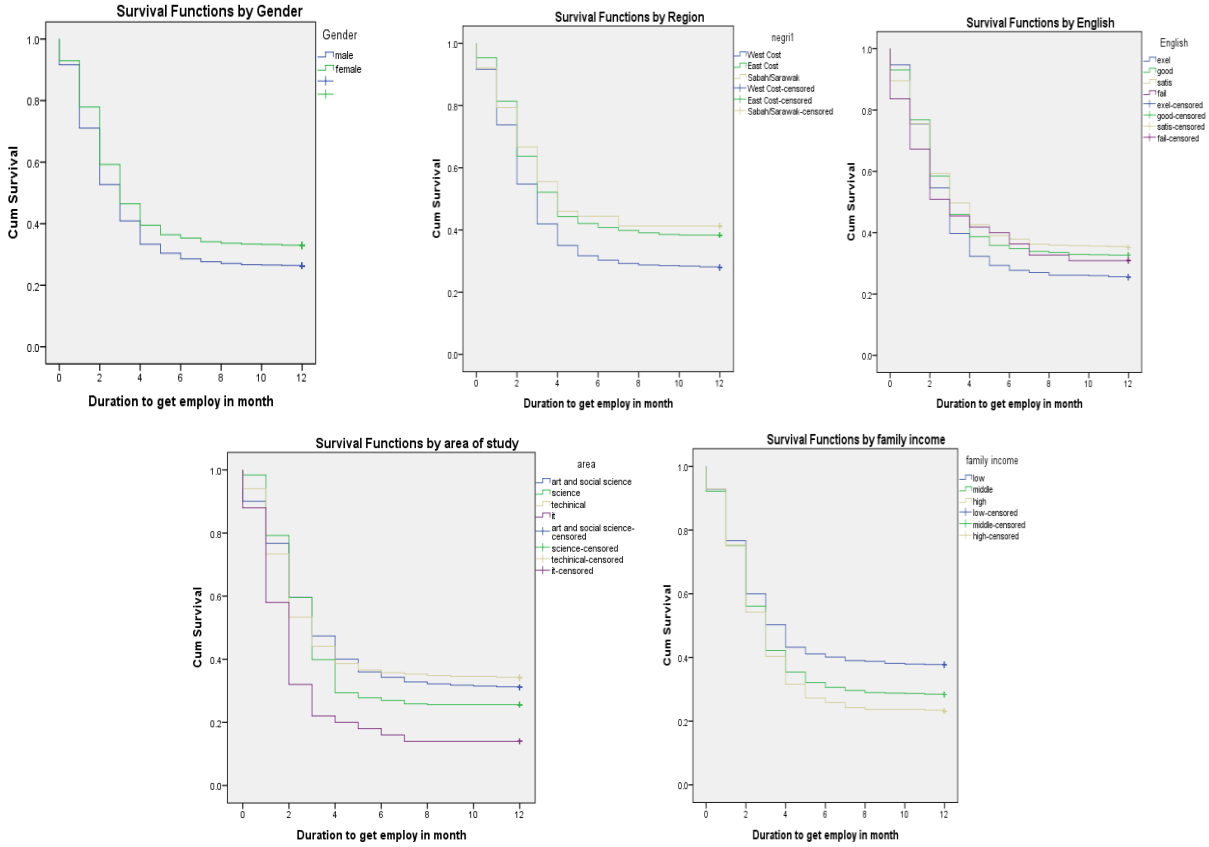

Fig. 1 Kaplan-Meier Survivor Functions Estimates for Five Explanatory Variable (Categories variables)

For the plot two, the KM survival probability (not get employed) estimates at 12 months were about .23 for West Coast of Malaysia subjects, .40 for East Coast of Malaysia subjects and 0.48 for Sabah and Sarawak subjects. The survival function for West Coast subjects appears to decline faster than the East Coast and Sabah and Sarawak subjects. It is clear that the Sabah and Sarawak subjects appear to enter the job market at the latest time. The statistical results of the log rank test showed a significant difference in survival times between the geographic regions ($\chi^2(2)=26.39$, $p\leq.05$).

For the third plot, the KM survival probability (not get employed) estimates at 12 months were around .30 among excellent English proficiency, while .38 for poor (failed) English proficiency, .40 for good English proficiency and .41 for satisfactory English proficiency. The results indicated that among the satisfactory English proficiency group appear to to enter the job market at a later time. The statistical results of the log rank test showed a significant difference in survival times between English performance ($\chi^2(3)=13.17$, $p\leq.05$).

For the fourth plot, the KM survival probability (not get employed) estimates at 12 months were around .19 for respondents specializing in information technology subjects, while .30 for respondents specializing in science subjects, .38 for respondents specializing in art and social science subjects and .40 for respondents specializing in technical subjects. The results indicated that respondents specializing in the technical subjects appear to enter the job market at the latest time. The statistical results of the log rank test showed a significant difference in survival times between the respondents' English language proficiency ($\chi^2(3)=16.73$, $p\leq.05$).

For the last plot, the KM survival probability (not get employed) estimates at 12 months were around .24 for respondents with high family income, .30 and .40 for respondents with moderate and low family income. The survival function for the low family income appears to decline slower then those for the other two groups of subjects, which appear broadly similar. The results indicated that the low family income subject appears to enter the job market at the latest time. The statistical results of the log rank test showed a significant difference in survival times between family income group among subjects ($\chi^2(2)=23.44$, $p\leq.05$).

VI. RESULTS FOR THE DISCRETE TIME TO EVENT DATA: SUBJECT-SPECIFIC PROPORTIONAL HAZARDS MODEL

In this subsection, we present likelihood-based estimation procedures that are parametric model for random intercepts and hazards ratio for parametric model as shown in Table 3. The deviances, likelihood ratio test are also reported in the same table, where we have used the level-two subjects (observations) for N (14359) and the number of estimated parameters for d. Furthermore, the number of subjects who took part in this research is 2341 (year 2010 dataset).

The interpretation of odd ratio in complementary log-log model is not straightforward like in the logit link model. Infact, the CLL link is direct on a proportional hazards function. It is discrete time to event version of proportional hazards model. The complementary log-log link function does not model in term of log odds, like the logit and probit link do (Kalbfleisch & Prentice, 2002). Furthermore, it is important to note - *time* that the model to predict is the probability of an event, not the absence of the event. The CLL function is not symmetric (asymmetric) around $\pi = .5$, since the CLL function approaches zero much slower than it approaches one.

Let $Y_{ij} = 1$ if level-2 unit, *j*th succeeds on getting a job or a risk of getting a job on the *i*th occasions (i=0,...,1,2), and $Y_{ij} = 0$ otherwise. The following random effects (mixed effects) complementary log-log regression model for Y_{ij} was fitted to the data: $\log(-\log(1-\pi_{month})) = \beta_0 + \beta_{1gender}: female_{ij} + \beta_2 CGPA_{ij} + \beta_3 curriculum + \beta_4 careerguidance_{ij} + \beta_5 region2_{ij} + \beta_6 region3_{ij} + \beta_7 English2_{ij} + \beta_8 English3_{ij} + \beta_9 English4_{ij} + \beta_8 area2_{ij} + \beta_9 area3_{ij} + \beta_{10} area4_{ij} + \beta_{11} Income2_{ij} + \beta_{12} Income_{ij} + \zeta_j$.

Maximum Likelihood Estimates (MLE) for the random intercept model using 12-point adaptive quadrature are given in Table 3. If we do not assume unobserved heterogeneity (random effect), the gender and the CGPA would be erroneously assessed to be not statistically significant (Proportional Hazards Model for discrete time event).

The deviance are 25,025.38 with the degree of freedom at 26. The estimated coefficients on the duration interval dummies (time 0 to time 11) tell us about the shape of the baseline hazard. Less negative values are associated with higher hazards ratio. Our model indicate that the baseline hazards ratio increases through time in a non-monotonous fashion. The results clearly showed that the increases follow a seasonal pattern. However, it is beyond this research to explore the seasonal effect of baseline hazard ratio.

The survival risk or hazards ratio of getting a job is related to gender (male: exp(0.138) = 1.148), the curriculum (exp(-0.022) = 0.978 (borderline significant), the career (exp(0.042) = 1.043, the region (West Coast exp(0.608) = 1.837, the English (Good exp(-0.189) = 0.828, satisfactory exp (-0.168) = 0.846, failure exp(1.017) = 2.762, the area of study (science exp(-1.38) = 0.251, technical exp(-0.237) = 0.78 , ICT exp(0.654) =1.923, and family income(medium exp (0.356) = 1.427, high exp (0.208) = 1.231.

The random intercept variance is estimated as $\sigma^2 = 9.314$, which is significantly different from zero. Alternatively, we can interpret σ^2 by appealing to the notion of a latent variable distribution of proportional hazard. Moreover, the scale parameter estimate of the random intercept is 3.052 (standard error = 0.134). It is significantly different from zero and indicates considerable residual heterogeneity. It is clear that the unobserved heterogeneity should not be ignored. Moreover, the ρ indicate the proportion of the total variance contributed by the subject level variance component. The significant of ρ in our model suggested that the subject level variance component is very important. Moreover, it clearly stated that our random effect model is significantly better then the fixed effect model (pooled estimator Complementary log-log). The estimated variance clearly showed that the variance parameter and standard error are well-behaved and there is no evidence of a high standard error for this estimate.

A test of the random effects with the 12 explanatory variables of the parametric models against fixed effects models is statistically significant, $\chi^2(1,N=2228) = 545.76$, $p \leq .0001$.

Table. 3 Estimates for Parametric Random Effect Complementary Log-Log Regression Model for Employability Data

Parameter	Parametric Est. (SE)	Hazards ratio	95% CI
Gender: male	0.138 (0.083)	1.148	(-0.024, 0.300)
CGPA	0.137 (0.089)	1.146	(-0.037, 0.311)
Curriculum	-0.022 (0.011)	0.978	(-0.044, 0.001)
Career	0.042 (0.007)	1.043	(0.028, 0.056)
region: East Coast	0.608 (0.096)	1.837	(0.420, 0.795)
region: West Coast	0.317 (0.202)	1.373	(-0.796, 0.713)
English: Good	-0.189 (0.095)	0.828	(-0.376, -0.004)
English: Satisfactory	-0.167 (0.104)	0.846	(-0.371, 0.036)
English: Failure	1.016 (0.186)	2.762	(0.651, 1.382)
area: science	-1.381 (0.155)	0.251	(-1.685, -1.078)

area: technical	-0.237 (0.087)	0.788	(-0.407, -0.066)
area: ICT	0.654 (0.155)	1.923	(0.349, 0.957)
income: Medium	0.356 (0.084)	1.427	(0.191, 0.521)
income: High	0.208 (0.119)	1.231	(-0.026, 0.443)
time 0	-6.875 (0.356)	0.001	(-7.590, -6.159)
time 1	-3.417 (0.281)	0.032	(-3.967, -2.867)
time 2	-1.114 (0.246)	0.328	(-1.596, -0.631)
time 3	0.126 (0.237)	1.134	(-0.339, 0.592)
time 4	0.859 (0.246)	2.361	(0.376, 1.340)
time 5	1.547 (0.269)	4.697	(1.019, 2.075)
time 6	1.324 (0.272)	3.758	(0.790, 1.857)
time 7	1.619 (0.267)	5.048	(1.095, 2.143)
time 8	1.085 (0.287)	2.959	(0.522, 1.648)
time 9	0.516 (0.321)	1.675	(-0.112, 1.145)
time10	-0.687 (0.453)	0.503	(-1.576, 0.201)
time11	-0.196 (0.392)	0.822	(0.964, 0.572)
σ_1^2 (subject)	9.314(0.092)		
Scale			
Variance component	0.803(0.0876)		(2.060, 2.404)
Parameter	3.052(0.134)		(2.801, 3.326)
ρ	0.849(0.011)		(0.826, 0.870)
Null model	32,467.26		
Deviance	25,025.38		

VII. SUBSTANTIVE DISCUSSION

In the psychological literature review, we raised the issue about the risk factor of getting a job among graduate students. Generally, for the MLE method, our best fit model is the parametric random effect model which consists of variables concerning gender, curriculum, career, region, English (Good and satisfactory groups), area (sciences and ICT) and family income. The present study thus provides strong evidence that male subjects are more likely to get a job compared to female students. In fact, the gender factor is the most robust explanatory variable in our study, because the gender factor is statistically significant in all our models. This study does not seem to fully support findings by previous research that mentioned the insignificant relationship between the time taken to secure employment and the respondents' gender (Rao & Jani, 2012).

In our study, the results showed that the respondents who came from families with high and median monthly incomes were highly likely to enter the labor market sooner compared to students who come from families with low monthly income. These findings were consistent with findings by Nafi & Ghani (2011) which indicate family income as a significant predictor of graduate employment. It could be the case that students who come from families with high and median monthly incomes might be exposed to more materials resources, supports and job opportunities that are provided by their parents or other family members.

In contrast, students who come from families with low monthly income may have limited resources.

The present study also provides strong evidence that geographical region is a significant factor to determine the risk of survival function in our graduate employability. The result from all the models showed that the graduate students who originated from West Coast of Malaysia and East Coast of Malaysia were more likely to enter the job market sooner than those who had graduated from Sabah and Sarawak. One possible cause for this is that most of the major cities in Malaysia, such as Kuala Lumpur, Putrajaya, Penang, Shah Alam and Johore Baharu, are located in West Coast of Malaysia. Hence, more job opportunities and job resources are available in this region than in the other two regions.

Our model also showed that the curriculum and carrier guidance services explanatory variables are significant factors in determining the risk of survival function in our graduate employability. Although the curriculum and carrier explanatory variables are not robust enough to be the significant explanatory variables in all of the models, these are significant in most of the models. The significance of the curriculum and the carrier guidance services explanatory variables indicate that students are concerned with the quality of our education.

The quality of education is an important factor and has been highlighted as a lead factor for the problem of unemployablity amongst Malaysian graduates (Zaliza & Mohd Safarin, 2011).

As far as English is concerned, the MLE parameters estimate are as follows. We fixed the excellent groups as reference group, the English performance is highly significantly different between the good, the satisfactory and the failed groups. However, the failed group is 2.762 times more likely to enter the job market earlier, compared to the excellent group students. It can be argued that English proficiency is important in getting employed, as highlighted by many researchers (Nora et al., 2012; Krish et al., 2012; Noor, 2011). However, this research indicates a mixed result between the respondents' levels of English proficiency. On one hand, it seems to support the theory of the importance of acquiring a high level of English proficiency as mentioned by previous studies, but on the other hand, it seems to challenge this assumption. It is important to note that our findings seem to suggest two theories regarding English proficiency, the first theory is the importance of English language that is true among students generally. The second theory is the insignificant role that English proficiency has among the failed group. It is important to note that, our second theory seem to support the previous finding that language skills (including English) are not emphasized by most organizations (Shaharuddin et al., 2014). Our current position state that both results are true in our study because we not to discriminate or control between respondents who are employed and those who are underemployed. It could be that the group with the lowest level of English proficiency just took on any available job in the labour market, without caring about the low wages. The question that remains unanswered is how English proficiency influences employability among graduate students, not the survival risk of employed (or unemployed) graduate students.

The employability dataset is very complex and huge, therefore, we need a subject matter expert or psychologist who can help us clean the dataset. Before we can model the data, it is better for us to clean the data and to make sense of it from a theoretical or psychological perspective. Cleaning the employability dataset requires a lot of effort and time. From the literature review we found that soft skills and communication skills are very crucial for employment (Mohamad Sattar et al., 2013). However, in our model we did not include either soft or communication skills as explanatory variables because we are not able to clean the variables properly. Moreover, it is worth noting that the communication between the researcher, the data collector (MOHE) and the psychologist are important in order to manage, clean and make sense of the dataset. For future research, we would like to suggest possible collaborations between the stake holders of these data and the statistician in order to disentangle the issue of graduate employability once and for all.

ACKNOWLEDGEMENTS

The researchers would like to acknowledge the Research Management Centre (RMC), UPSI for Research University Grant (2015-0033-106-01).

REFERENCES

1. Abdul Hamid, M.S., Islam, R. & Noor Hazilah (2014). Malaysian graduates, employability skills enhancement: an application of the important performance analysis. *J. global business advancement, 7(3),* 181-197.
2. Ayiesah, R., Roslizawati, N., Maslyn, P. (2010). Employees' perception of employability skills needed in today's workforce among physiotherapy graduates. *Procedia social and behavioral sciences, 7(2).* 455-463.
3. Berridge, D. & Crouchley, R. (2011). *Multivariate generalized linear mixed models using R.* Boca Raton: CRC Press.
4. Bryla, P. (2015). The impact of international student mobility on subsequent employment and professional career: a large-scale survey among Polish former Erasmus students. *Procedia social and behavioral sciences, 176,* 633-641
5. Congdon, P. (2001). *Bayesian statistical modelling*. Chichester: Wiley.
6. Ganjali, M. & Baghfalaki, T. (2012). Bayesian analysis of unemployment duration data in the presence of right and interval censoring. *Journal of reliability and statistical studies, 5(1).* 17-32.
7. Kalbfleisch, J. & Prentice, R. (2002). *The statistical analysis of failure time data.* New York: Wiley.
8. Lim, H. (2011). The determinants of individual unemployment duration: The case of Malaysian graduates. *Journal of global management, 2(1).* 184-202.
9. Little, R. J. A. & Rubin, D. B. (2002). *Statistical analysis with missing data: Wiley series in probability and mathematical statistics*. New York: John Wiley & Sons.
10. Menard, S. (1995). *Applied logistic regression analysis*. London: Sage.
11. Mohamad Sattar, R., Rose, A., Azlin, N. M., Ruhizan, M. & Zamri, M. (2013). Graduate employability for manufacturing industry. *Procedia social and behavioral sciences, 102*, 242-250.
12. Nafi, M. N. & Ghani, I. A. (2011). Modelling employability of graduates using logistic regression. *Journal of statistical modeling and analytics, 2*. 45-51.
13. Ntzoufras, I. (2009). *Bayesian modeling using WinBUGS*. New York: John Wiley.
14. Pouratashi, M. & Zamani, A. (2018). Employment issues and its effect on academic burnout (case: agricultural students). *Int. J. education economic and development, 9(3)*,236- 247.
15. Rao, R, & Jani, R. (2012). Exploring employment of graduates. *World applied sciences journal, 17(2).*184-188.
16. Shaharuddin, A., Khaidzir, I.,Azian, A., Kadir, A., Zainul Ariffin, A., Khairil, A. & Wan Mazlina, W. (2014). The social and academic skills and the marketability of UKM's graduates. *Procedia social and behavioral sciences, 131.* 118-123.
17. Sarfraz, I., Rajendran, D., Hewege, C., & Mohan, M.D. (2018). An exploration of global employability skills: a systematic research review. *Int. J. work organization and emotion, 9(1),* 63-88.
18. Stojanova, H. & Blaskova, V. (2014). The role of graduates' field of study and its impact on the transition to working life. *Procedia economics and finance, 123*, 636-643.
19. Torres-Machi, C., Carrion, A., Yepes, V. & Pellicer, E. (2013). Employability of graduate students in construction management. *Journal of professional issue in engineering education practice, 139(2),* 163-170.
20. Zaliza, H. & Mohd Safarin, N. (2011). Unemployment among Malaysia graduates: Graduates' attributes, lecturers' competency and quality of education. *Procedia social and behavioral sciences, 112.* 1056-1063.