Feasibility of Early Identification of Low Employability Graduates in Malaysia

Hock-Eam LIM*
Universiti Utara Malaysia

Abstract: This paper examines the feasibility of early identification of low employability graduates using statistical profiling models. The applied statistical profiling model successfully predicted correctly 88 per cent of the low employability graduates using in-sample evaluation and 67 per cent correctly using out-of-sample evaluation. It is suggested that university authorities implement programmes for the early identification of low employability graduates for their final year students. It is further recommended that the government consider using statistical profiling as a tool in the allocation of the limited places in the re-training programmes for unemployed graduates.

Keywords: Early identification, graduate employability, proportional hazard model, statistical profiling
JEL classification: J64

1. Introduction

Graduate unemployment is not new. During the recession of the mid-1980s, there was an increase in graduate unemployment in Malaysia. With the recovery of the economy towards the end of the 1980s, graduate unemployment also became less prevalent. However, graduate unemployment that prevailed during the 1998 currency crisis still persists. It has been reported that the number of unemployed graduates had increased from 45,000 in 2000 to 85,000 in 2005 (Sim 2006). According to Lim *et al.* (2008), "*…there is no longer a shortage of graduates in Malaysia with graduate unemployment a persistent problem since the late 1990s.*" Several studies also discuss the relationship between tertiary education and the job market (Morshidi *et al*. 2004a; Morshidi *et al*. 2004b; Ambigapathy and Aniswal 2005).

The Malaysian government has designed several assistance programmes for unemployed graduates. In 2001, the Graduate Training Scheme (GTS) consisting of 13 different re-training programmes ranging from English language proficiency to Information Technology skills was launched to improve the employability of unemployed graduates. GTS is a boon to the unemployed graduates. However, to be eligible for this scheme, graduates must have been unemployed for at least six months. This implies that before they are eligible for GTS, they need to endure the cost of unemployment for a minimum period of six months.

A precept of quality control is that assistance should be provided before a problem occurs. It is preferable to re-train low employability graduates before they enter the labour

* College of Arts and Sciences (Economics), Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia. Email: *lheam@uum.edu.my*

I would like to thank the anonymous referee whose comments have substantially improved this paper. Thanks are also due to Dr. Thi Lip Sam for his constructive comments. I take responsibility for all remaining errors.

market and suffer prolonged unemployment. In this context, early identification of low employability graduates is important as early identification of such graduates could be a useful tool in allocating the limited places in GTS re-training programmes.

As pointed out by Nativel (2004), the design and deliverance of an active labour market policy should target individuals who need it the most. Statistical profiling is suggested as a tool to identify low employability individuals early. Clearly, there is a need to use statistical profiling models that enable early identification of low employability graduates. In Malaysia, due to the persistence of graduate unemployment and its associated costs (financially and psychologically), early identification of low employability graduates is an imperative.

Surprisingly, early identification has been largely overlooked in the literature of graduate unemployment in Malaysia. This paper aims to fill the gap by evaluating the feasibility of applying statistical profiling models for early identification of low employability graduates using panel data and proportional hazard models that take into consideration censoring bias. Feasibility of early identification is evaluated based on the predictive power of the statistical profiling model.

The rest of this paper is organised as follows. A brief review of the literature on Malaysian graduate unemployment and statistical profiling is presented in the Section 2. Section 3 presents the data and methodology while Section 4 discusses data analysis and the findings. The final section concludes the paper.

2. Literature Review

Several studies have attempted to identify the determinants of graduate unemployment in Malaysia. The suggested determinants are English language proficiency, communication skills, mismatch of degrees with market needs, gender, ethnicity and academic attainment (Ball and Chik 2001; Lim and Normizan 2004; Morshidi *et al*. 2004a; Morshidi *et al*. 2004b; Hariyati *et al*. 2006; Lim 2007; Lim *et al.* 2008).

Knowing the determinants enhances our understanding of the nature of Malaysian graduate unemployment. It also facilitates identification of graduates (in groups) who are at high risk of being long-term unemployed. However, it does not assist in early identification of low employability graduates individually. For instance, it has been estimated that the odds of getting employed for Chinese graduates is 4.537 times higher than that of Malay graduates (Morshidi *et al.* 2004a). Clearly Malay graduates are known to be in the high risk group of being unemployed. However, for an individual Malay graduate, we are not able to estimate his risk (individually) of being unemployed. His employability ranking relative to other graduates is also unknown.

Empirically, statistical profiling has been applied to certain fields of studies such as early identification of risk of reconviction (Copas and Marshall 1998) and risk of sudden infant death syndrome (Carpenter 1983). Since the 1990s, there have been studies on statistical profiling of those at high risk of being long-term unemployed. Countries such as United States and Australia, have formally incorporated statistical profiling as part of their efforts on early identification of the long-term unemployed while countries such as the United Kingdom and Canada are experimenting with statistical profiling (OECD 1998).

In the United Kingdom, Payne and Payne (2000) developed a simple statistical prediction model for early identification of the long-term unemployed using logistic regression. Using cross-validation (out-of-sample evaluation) to evaluate the prediction power of the estimated model, they found a large margin of prediction error. Thus they concluded that early identification might not be feasible. This large prediction error might have been partly due to the following factors: small sample size, censoring bias and exclusion of age and ethnicity (which are significant determinants of one's employment outcome).

In the United States, Black *et al*. (2003) examined the effectiveness of the Worker Profiling and Reemployment Services (WPRS) of Kentucky on predicting the probability of exhausting Unemployment Insurance (UI) benefits, that is, the probability of being long-term unemployed. It was found that the statistical profiling model effectively reduced the mean value and number of weeks of UI benefits receipts. Compared to Payne and Payne (2000) who used cross-section data, Black *et al*. (2003) used experimental design data (follow-up surveys). Thus, the type of data used seems to have a significant influence on the prediction power of a statistical profiling model. Specifically, panel data (follow-up surveys) are preferable to cross-section data.

Lechner and Smith (2007) compared the efficacy of Swiss caseworkers in allocating unemployed individuals to various government training programmes. They found that the statistical profiling model performed better than the Swiss caseworkers in terms of post-programmes employment rates. In Malaysia, Lim (2005) applied a statistical prediction model for Universiti Utara Malaysia graduates. His out-of-sample evaluation showed that the estimated model predicted correctly 75 - 83 per cent of the low employability graduates. Thus, this estimated statistical profiling model appears to have an acceptable level of prediction power (since it accurately predicted about 80 per cent of low employability graduates).

In contrast, Saodah *et al*. (2006) attempted to profile the characteristics of low employability graduates in Malaysia. They concluded that it is difficult to predict low employability graduates accurately. This difficulty might have arisen due to their use of cross-section data and choice of statistical profiling model (discriminant analysis) with only four explanatory variables. In short, previous studies offer conflicting evidence on the effectiveness of the statistical profiling model. Estimation method and types of data appear to have significant influence on the prediction power of the statistical profiling model. The studies of Lim (2005) and Saodah *et al*. (2006) appear to be the only studies on early identification of low employability graduates in Malaysia. However, cross-section data was used in these studies and this might have rendered their findings open to censoring bias.

3. Data and Methodology

3.1 Data

The panel data consisted of two surveys. Data collection for the first survey was undertaken between July 2005 and March 2006. A self-administered questionnaire was used to increase the response rate. This approach is feasible since the students are staying in and around the campus. The targeted population was the final year students from Universiti Utara Malaysia (UUM), a public university and Universiti Tunku Abdul Rahman (UTAR), a private university.

The first survey successfully collected a total of 430 responses (304 from UUM and 126 from UTAR). The second survey was conducted between November 2006 and February 2007, targeting the 430 graduates who had responded during the first survey. A mailed questionnaire approach was used as the respondents had graduated and were located in various part of the country. A total of 240 questionnaires was returned.

Due to time and resource constraints, data collection was limited to two universities and included only a small number of long-term unemployed graduates (defined as those who had been unemployed for 12 months or more) (Payne and Payne 2000). Moreover in Malaysia, employers are perceived to prefer foreign university graduates (from English speaking countries such as the United Kingdom) to local university graduates. Due to data limitation, this perception of employers could not be tested in this study.

In addition, the sample covered only a limited number of degree programmes, mostly business-related degree programmes offered by the two local universities. Thus generalisation of this paper is confined only to the business-related degree programmes of these two universities. However, this should not limit the contributions of this paper which aims to illustrate the feasibility of early identification of low employability graduates. The estimation methods used in this paper can be easily extended to a more representative sample.

3.2 Methodology

This study used proportional hazard models (Weibull, Piecewise Constant and Gompertz) as statistical profiling models because of their ability to accommodate censoring bias. Expected unemployment duration (estimated from the proportional hazard models) was used to represent graduate employability. Graduates were ranked based on their expected unemployment duration. Those ranked at the bottom (that is, graduates with the highest value in expected unemployment duration) were identified as being in the low employability group. Hence two information sets were estimated for each graduate: the expected unemployment duration and employability ranking.

Let T_j be a non negative variable that represents unemployment duration of graduate j with covariates of x_j. Then the expected unemployment duration which is conditional on the value of covariates can be expressed as

$$E(T_j \mid x_j) = \int_0^\infty S\left(t \mid x_j\right) \tag{1}$$

where

$S\left(t \mid x_j\right) =$ survival function

$x_j =$ explanatory variables

$S(t|x_j)$ is estimated using the proportional hazard model through maximising the log of the following likelihood function:

$$L_j\left(\beta,\delta\right) = \frac{S\left(t_j \mid x_j\beta,\delta\right)}{S\left(t_{0j} \mid x_j\beta,\delta\right)} h\left(t_j \mid x_j\beta,\delta\right)^{d_j} \tag{2}$$

where

δ = the ancillary parameters

d = the indicative variables, d=1 if failed and 0 if censored

$h\left(t_j \mid x_j\beta,\delta\right) =$ the hazard function

The survival and hazard function can be specified as given below:
Weibull model

$$S(t \mid x_j) = \exp\left\{-\exp\left(\beta_0 + x_j\beta\right)t^p\right\} \tag{3}$$

$$h(t \mid x_j) = pt^{p-1}\exp\left(\beta_0 + x_j\beta\right) \tag{4}$$

Piecewise Constant model

$$S(t \mid x_j) = \exp\left\{-\exp\left(\beta_0 + D_j\alpha + x_j\beta\right)t\right\} \tag{5}$$

$$h(t \mid x_j) = \exp\left(\beta_0 + D_j\alpha + x_j\beta\right) \tag{6}$$

where

D = dummy variables (months)

Gompertz model

$$S(t \mid x_j) = \exp\left\{-\gamma^{-1}\exp\left(\beta_0 + x_j\beta\right)\left(\exp\left(\gamma t\right)-1\right)\right\} \tag{7}$$

$$h(t \mid x_j) = \exp\left(\gamma t\right)\exp\left(\beta_0 + x_j\beta\right) \tag{8}$$

In this paper, the choice of covariates (independent variables) is limited to those available in the existing student records. This implies that university authorities can implement early identification using only their student records. These variables are types of degree, age, gender, ethnicity, hometown location, academic attainment, industrial training, family size, parental education level and parental employment status. These variables are pre-determined (that is, measured at first survey) and hence endogeneity should not pose a problem to the estimated models.

3.3 Construction and Validation Sample

A total of 35 graduates were randomly selected as a validation sample for the purpose of out-of-sample evaluation (cross-validation). The remaining (205 graduates) made up the construction sample (to estimate the three statistical profiling models). Among these 35 graduates (validation sample), 5 graduates (14 per cent) were unemployed; 13 graduates (37 per cent) were in full time employment commensurate with qualifications; 15 graduates (43 per cent) were in full time employment, not commensurate with qualifications; 1 graduate (3 per cent) was self-employed and 1 graduate (3 per cent) was in part-time employment.

4. Analysis and Results

4.1 Descriptive Statistics

Table 1 presents the labour market outcomes and unemployment duration for the 240 graduates. It is observed that a quarter (25 per cent) of the graduates were unemployed. About forty per cent of them were in full-time employment commensurate with their qualifications (FT1) and around thirty per cent were in full-time employment, not commensurate with their qualifications (FT2). The remaining were self-employed (SE) and in part-time employment (PT).

Relating to the mean of unemployment duration, graduates in the FT1 category have the lowest mean (49.5 days) unemployment duration while graduates in the SE category have the highest mean unemployment duration (131.25 days). Those who were unemployed and in the FT2 and PT categories appear to have almost a similar mean of unemployment duration.

Table 2 presents the respondents' characteristics and mean of unemployment duration. Based on ethnicity, Chinese graduates have the lowest mean unemployment duration of 38.31 days compared to 67.31 and 53.72 days for Malays and other ethnic groups respectively.

By type of degree, UTAR Information Technology/Computer Sciences was found to have the lowest mean unemployment duration of 23.92 days while UUM International Business/Issues Management had the highest mean of 70.61 days.

It is important to note that the duration of all degree programmes is three years except for the UUM accounting degree which is a 4-year programme. Some degree programmes might be able to attract more able students than other degree programmes. Thus these mean differences across various types of degrees might reflect a self-selection bias. Caution should be taken in interpreting the effects of types of degree on graduate's unemployment duration.

Referring to continuous or discrete variables, it can be seen from Table 2, two variables have a correlation of more than 0.10 with unemployment duration: family size (0.14) and mother's education level (0.13). In short, descriptive statistics show that the burden of unemployment duration varies across different social demographic groups.

Table 1: Labour market outcomes

Labour market outcomes	%	Mean unemployment duration (days)
Unemployed	25.00	67.75
FT1	40.63	49.50
FT2	28.13	67.25
Self-employed (SE)	1.79	131.25
Part-time employment (PT)	4.46	62.60

Note:
1. FT1 = full-time employment commensurate with qualifications; FT2 = full-time employment not commensurating with qualifications.
2. Number of respondents = 240.

Table 2: Respondents' characteristics and mean unemployment duration

(a) Categorical variables	Proportion	Mean
Type of degree		
UUM Economics	0.09	55.27
UUM Public Mgt/Devep Mgt	0.05	55.58
UUM Business Admin	0.11	46.60
UUM Accounting	0.08	64.37
UUM Info Technology	0.12	63.45
UUM Other degrees[2]	0.07	48.30
UUM Human Resource/Soc Work	0.06	27.36
UUM Int Business/Issues Mgt	0.05	70.61
UUM Finance	0.07	63.06
UTAR Business Admin	0.08	43.44
UUM Communication	0.04	54.46
UTAR Accounting	0.08	36.47
UTAR Info Tech/Computer Sc	0.05	23.92
UTAR Other degrees[3]	0.04	37.82
Other variables		
Father is eco-inactive:		
Yes	0.10	41.95
No	0.90	49.06
Mother is eco-inactive:		
Yes	0.58	47.41
No	0.42	44.73
English as main communication language:		
Yes	0.07	69.63
No	0.93	47.98
Industrial training:		
Yes	0.46	47.13
No	0.54	50.92
Gender:		
Male	0.28	48.57
Female	0.72	49.78
Ethnic group:		
Chinese	0.58	38.31
Malay	0.34	67.31
Others	0.08	53.72
Hometown:		
Rural	0.58	48.71
Big cities or state capital	0.42	50.45
(b) Continuous/discrete variables	**Sample mean**	**Correlation with unemployment duration**
Father's education level	4.33	0.08
Mother's education level	3.97	0.13
Family size	6.32	0.14
Academic attainment	3.08	-0.08
Age (in years)	23.33	0.05

Notes:
1. Please refer to Appendix 1 for definition and measurement of the variables.
2. UUM Other degrees: Bachelor of Tourism, Education, Technology Management and Decision Sciences.
3. UTAR Other degrees: Bachelor of Chinese Studies, Journalism and Public Relations.
4. Number of respondents = 240.

Table 3: Estimated proportional hazard models

(a) Estimated models	Weibull Coefficient	Piecewise Coefficient	Gompertz Coefficient
Type of degree			
UUM Economics	-0.4023 (0.5624)	0.6542 (0.3661)	-0.4206 (0.5242)
UUM Public/Devep Mgt	-0.8968 (0.6722)	0.3857 (0.2580)	-0.8566 (0.6277)
UUM Buss Admin	-0.5223 (0.5557)	0.5582 (0.3135)	-0.5039 (0.5135)
UUM Accounting	-1.6068 (0.7068)**	0.2050 (0.1466)**	-1.5778 (0.6624)**
UUM Info Tech	-1.8136 (0.6655)***	0.1599 (0.1055)***	-1.7780 (0.6249)***
UUM Other degrees	-1.1300 (0.6647)*	0.3089 (0.2062)*	-1.0748 (0.6223)*
UUM Human Res/SocWork	0.2994 (0.6788)	1.3204 (0.9026)	0.1368 (0.6081)
UUM Int Buss/Issues Mgt	-1.0066 (0.6058)*	0.3684 (0.2207)*	-1.0004 (0.5663)*
UUM Finance	-1.5726 (0.7479)**	0.2012 (0.1514)**	-1.5041 (0.7001)**
UUM Communication	-0.2763 (0.6344)	0.7121 (0.4498)	-0.2976 (0.5793)
UTAR Buss Admin	-0.5263 (0.5205)	0.5697 (0.2969)	-0.4574 (0.4771)
UTAR Info Tech/CompSc	-1.6867 (0.9682)*	0.2106 (0.2168)	-1.7511 (1.0111)*
UTAR Other degrees	-0.5078 (0.5780)	0.5549 (0.3267)	-0.4484 (0.5310)
Socio-demographics			
Age	0.1856 (0.0495)***	1.1957 (0.0597)***	0.1854 (0.0467)***
Male	0.0030 (0.2657)	0.9996 (0.2672)	0.0162 (0.2489)
Chinese	1.2071 (0.4110)***	3.2714 (1.3524)***	1.1398 (0.3852)***
Other ethnic group	0.3192 (0.5377)	1.3774 (0.7506)	0.3155 (0.5133)
Rural	0.0336 (0.2390)	1.0446 (0.2518)	0.0018 (0.2173)
Academic attainment	0.0767 (0.3736)	1.0826 (0.4111)	0.1129 (0.3557)
Industrial training	0.3806 (0.3210)	1.4299 (0.4575)	0.3768 (0.3052)
Family background			
Family size	-0.1895 (0.0734)***	0.8301 (0.0621)**	-0.1792 (0.0692)**
Father's education level	0.0463 (0.0788)	1.0531 (0.0846)	0.0379 (0.0734)
Mother's education level	-0.0144 (0.0781)	0.9728 (0.0788)	-0.0022 (0.0730)
Father eco inactive	-0.1371 (0.3935)	0.8890 (0.3503)	-0.1444 (0.3690)
Mother eco inactive	0.7342 (0.2771)***	2.0337 (0.5714)**	0.7128 (0.2611)**
Constant	-9.8444 (1.7495)***	-	-8.9610 (1.6012)***
Dummy variables (months)			
Month 2	-	1.8547 (0.4826)**	-
Month 3	-	2.2791 (0.7242)**	-
Month 4	-	1.9872 (0.8987)***	-
Month 5	-	1.5511 (0.8984)	-
Month 6	-	2.2816 (1.1654)	-
Month 7	-	2.1125 (1.5875)	-
Month 8	-	1.6739 (1.8194)	-
Month 9	-	3.9336 (4.7575)	-
Month 10	-	0.0000 (0.0000)***	-
Month 11	-	0.0000 (0.0000)***	-
Month 12	-	0.0000 (0.0000)***	-
Month 13 and above	-	16.2606 (13.5479)***	-

Table 3 continued

Ancillary parameters			
P	1.2709 (0.1052)***	-	-
Gamma	-	-	0.0035 (0.0019)*
(b) Goodness-of-fit tests:	Weibull (*P*-value)	Piecewise (*P*-value)	Gompertz (*P*-value)
1. Overall fit test	0.0000	0.0000	0.0000
2. Link test (General specification test)	0.8540	0.2400	0.7990
3. Martingale residual plots (functional form test)	No evidence[a]	No evidence[a]	No evidence[a]

Notes:
*, **, and ***, represent significance at 10, 5 and 1 per cent respectively.
Please refer to Appendix 1 for explanation and measurement of variables.
Reference group for dummy variable is: (a) Type of degree: UTAR Bachelor of Accounting; (b) Ethnicity: Malay; (c) Months: Month 1.
Figures in perentheses are standard errors
[a.] There is no evidence of inappropriate functional form.

4.2 Estimated Proportional Hazard Models

Table 3 presents the estimated models (Weibull, Piecewise Constant and Gompertz) and also the results of goodness-of-fit tests. The overall fit test is found to be significant with a *p*-value of almost zero (Table 3). The link test shows that there is no evidence of general specification error while Martingale residual plots reveal no evidence of insufficient linear functional form of the covariates in the estimated models. Thus, statistically, these estimated models appear to have an acceptable goodness-of-fit level.

The study findings reveal that the significant determinants of graduate unemployment duration are type of degree, age, ethnicity and family background variables. In particular, compared to Malay graduates, Chinese graduates are found to have a significantly higher probability of leaving unemployment. This is consistent with the finding from descriptive statistics whereby Chinese graduates are found to have a substantially lower mean of unemployment duration than Malay graduates (see Table 2).[1]

UUM Human Resource/Social Work and UTAR Information Technology/Computer Sciences degree programmes (found to have the lowest mean of unemployment duration in descriptive statistics (Table 2)), do not have significantly higher probability of leaving unemployment. This might be due to the influence of other variables (such as ethnicity that might correlate with types of university and degree programmes) which are controlled in econometric models but not in descriptive statistics. This highlights the importance of controlling the influences of other variables in evaluating the effects of type of degree programmes.

[1] As pointed out by a referee, graduate labour markets in Malaysia might be segregated by ethnicity: bumiputera (mostly Malay) and non bumiputera (mostly Chinese). This suggests that ethnicity is an important predictor in early identification of low employability graduates.

Estimation results (Table 3) are typical for a study on determinants of graduate unemployment duration. To enable early identification, these estimated models were used as statistical profiling models to estimate the expected unemployment duration (that is conditional on the value of covariates). The feasibility of this early identification of low employability graduates will be determined by the predictive power of these statistical profiling models.

4.3 Early Identification of Low Employability Graduates

Expected unemployment duration (conditional on the value of covariates) is estimated for each graduate in the constructed sample. Based on this expected unemployment duration, graduates are ranked with graduates at the bottom being identified as low employability graduates. If actual employment status of these graduates is that of unemployment, they are correctly predicted. A high percentage of graduates correctly predicted shows high feasibility of this early identification. For comparison purpose, the sample percentage of unemployed (29 per cent) in the constructed sample was used (the comparison model is the Naïve prediction model that uses the sample percentage of unemployed for early identification of low employability graduates).

Table 4 presents the percentages correctly predicted using Weibull, Piecewise Constant, Gompertz and Naïve models. Assuming that graduates who are at the bottom 5 per cent are identified as low employability graduates (8 graduates), the Weibull model correctly predicted 88 per cent of these 8 graduates (that is, 7 graduates identified as low employability graduates were actually unemployed). The percentage correctly predicted was 100 per cent for both the Piecewise Constant and Gompertz models. These results are impressive compared to only 29 per cent for the Naïve model.

For the bottom 10 per cent (16 graduates) and 15 per cent (24 graduates), the percentages correctly predicted for the three hazard models were also substantially higher than that for the Naïve model. For example Weibull and Gompertz correctly predicted 75 per cent of low employability graduates.

The above correct predictions of low employability are for the in-sample evaluation. There is the potential for bias as the evaluations tend to over-fit. This in-sample evaluation is also less inappropriate because practically early identification of low employability is an extrapolation from the estimated model, instead of being an interpolation. Thus, percentages correctly predicted based on an out-of-sample evaluation (cross-validation) is also presented using the validation sample (consists of 35 graduates).

Table 4: Percentage correctly predicted (in-sample evaluation)

	Bottom 5%: 8 graduates	Bottom 10%: 16 graduates	Bottom 15%: 24 graduates
Weibull	88% (7)	81% (13)	75% (18)
Piecewise	100% (8)	88% (14)	63% (15)
Gompertz	100% (8)	81% (13)	75% (18)
Naïve (% unemployed)	29% (2)	29% (5)	29% (7)

Note: Figures in parentheses are the number of graduates correctly predicted.

Table 5: Percentage correctly predicted (out-of-sample evaluation)

	Bottom 5%: 2 persons	Bottom 10%: 3 persons	Bottom 15%: 5 persons
Weibull	50% (1)	67% (2)	60% (3)
Piecewise	50% (1)	67% (2)	40% (2)
Gompertz	50% (1)	67% (2)	60% (3)
Naïve (% unemployed)	16% (0)	16% (0)	16% (1)

Note: Figures in parentheses are the number of persons correctly predicted.

Using the estimated proportional hazard models, expected unemployment duration of these 35 graduates was estimated. Due to missing values in covariates, 3 graduates (id=130, 184 and 339) were dropped. Among the remaining 32 graduates, 5 graduates (16 per cent) were unemployed. Application of the Naïve model found the percentage correctly predicted to be 16 per cent. Appendix 2 presents the early identification of each graduate in the validation sample.

Table 5 presents the percentage correctly predicted based on the out-of-sample evaluation. Using the Weibull model, the expected unemployment duration ranged from 29 to 849 days (Appendix 2). Assuming that those at the bottom 5 per cent are identified as low employability graduates, the Weibull model was used to identify graduates with id 257 and 60. Their expected unemployment duration was 637 and 849 days respectively. Graduate with id 257 was actually unemployed. Thus the percentage correctly predicted was 50 per cent. The Piecewise Constant and Gompertz models also predicted correctly similar percentages (50 per cent).

For the bottom 10 per cent, the percentage correctly predicted was 67 per cent for all three proportional hazard models. For the bottom 15 per cent, Weibull and Gompertz models predicted 60 per cent correctly whereas the Piecewise Constant model was only able to predict 40 per cent correctly. This indicates that the Weibull and Gompertz models outperformed the Piecewise Constant model. Overall, the percentages correctly predicted by the proportional hazard models were substantially higher than that of the Naïve model (16 per cent).

To ascertain whether the inclusion of additional covariates will increase the predictive power of these estimated models, language proficiency (English, Malay, Chinese and Tamil), use of English as the main language of communication, overall happiness in life, university life, health condition, university holiday working experience and number of working members in family, were added to the models.

As these additional covariates were not available in the existing student records, the information was collected by distributing a questionnaire to the students. The percentage correctly predicted is presented in Appendix 3. Results reveal that although the additional covariates improved slightly, the percentage correctly predicted for the in-sample evaluation, the percentage dropped for the out-of-sample evaluation. Thus no substantial improvement was found. This finding suggests that the use of covariates available in the existing student records is sufficient for implementing the early identification of low employability graduates.

5. Conclusion

Using proportional hazard models of Weibull, Piecewise Constant and Gompertz, this paper illustrates that the implementation of early identification of low employability graduates is feasible. Percentages correctly predicted from these proportional hazard models are substantially higher than that for the Naïve model that uses sample proportion. Specifically for the bottom 10 per cent of low employability graduates, the percentages correctly predicted were 88 per cent (in-sample) and 67 per cent (out-of-sample) for the proportional hazard model compared to 29 per cent (in-sample) and 16 per cent (out-of-sample) for the Naïve model. It is also found that the use of additional covariates (that are not available in the student records and need to be collected) does not improve substantially the percentage correctly predicted. This suggests that early identification can be implemented by university authorities using information that is available from their existing student records. Thus early identification of low employability graduates is feasible.

It is suggested that for ease of implementation, estimation of expected unemployment duration be automated by the use of a specially written computer program. As such, students' expected unemployment duration could be estimated and ranked using information technology. Final year students who are identified as having low employability, that is, the bottom 5 per cent, will be alerted and informed of their estimated unemployment duration and ranking (for instance through email) as confidential information. This is similar to the results of the GRE (Graduate Record Examination) test. GRE reports the candidates' scores and also their relative ranking among all candidates. In this context, early identification will serve as a warning to the students that there is a need for them to improve (on) their employability. It also helps them to have a realistic expectation of their employability status.

It is suggested that government authorities consider using statistical profiling as a complementary tool in allocating the limited places for their re-training programmes such as the Graduates Training Schemes (GTS). In a similar vein, early identification of those in high risk of being long-term unemployed can be implemented using different sets of covariates that are available from the existing database of the GTS participants (or other government re-training programmes). Further studies are suggested to explore the feasibility of statistical profiling of low employability graduates.

Further studies are also proposed to include long-term unemployed graduates, types of universities (especially foreign universities) and types of degree programmes. Control of self-selection bias (for instance, some degree programmes such as Accounting might attract more 'able' students than other degree programmes) is also an important area for future studies.

Finally, it is important to note that statistical prediction models need to be re-estimated and revised continuously to reflect the rapid changes that occur in graduate labour markets. This requires continuous efforts and resources. Hence, further development of the statistical profiling model for early identification of low employability largely rests on the shoulders of government authorities such as the Ministry of Higher Education. It is hoped that this study serves as the first step in exploring the feasibility of early identification of low employability graduates. It is hoped that this paper will open the discussion on early identification of low employability graduates using statistical models and shed more light in the literature concerning graduate unemployment in Malaysia.

References

Ambigapathy, Pandian and Aniswal Abd. Ghani. 2005. *University Curriculum: an Evaluation on Preparing Graduates for Employment*. USM IPPTN Monograph 5/2005. Penang.

Ball, Rob and Chik, Razmi. 2001. Early employment outcomes of home and foreign educated graduates: the Malaysian experience. *Higher Education* **42(2)**: 171-189.

Black, A.D., A.J. Smith, M. Plesca and S. Shannon. 2003. Profiling UI claimants to allocate reemployment services: evidence and recommendations for States. ETA Occasional Papers. USA: U.S. Department of Labor.

Carpenter, R.G. 1983. Scoring to provide risk-related primary health care: evaluation and updating during use. *Journal of the Royal Statistical Society Series C* **42**: 315-331.

Copas, J. and P. Marshall. 1998. The offender group reconviction scale: a statistical reconviction score for use by probation officers. *Journal of the Royal Statistical Society Series C* **47**: 159-171.

Hariyati Shahrima Abdul Majid, Shukran Abdul Rahman, Saodah Wok, Noor Azlan Mohd Noor, Ainul Madziah Zubairi and Danial Mohd Yusof. 2006. Gender differences in graduate employability: the role of inclusive higher education in Malaysian universities. Paper presented at *International Higher Education Policy Research and Management Forum 2006*, 8-11 Nov 2006, USM, Penang, Malaysia.

Lechner, M. and A. Jeffrey Smith. 2007. What is the value added by caseworkers? *Labour Economics* **14(2)**: 135-151.

Lim, H.E. 2005. Early identification of low employability graduates in Malaysia: the use of proportional hazard model. *Jurnal Teknologi Maklumat dan Sains Kuantitatif* **7(1)**: 41-48.

Lim, H.E. 2007. Estimating the employment performance indicator: the case of Universiti Utara Malaysia graduates. *Singapore Economic Review* **52(1)**: 73-91.

Lim, H.E. and Normizan Abu Bakar. 2004. Unemployment duration of graduates of Universiti Utara Malaysia: the impact of English language proficiency. *Malaysia Journal of Economic Studies* **41(1)**: 1-20.

Lim, H.E., J. Rich and M. Harris. 2008. *Employment outcomes of graduates: the case of Universiti Utara, Malaysia. Asian Economic Journal **22(3)**: forthcoming.*

Morshidi Sirat, Rosni Bakar, Lim, H.E, and Mohamed Nasser Katib. 2004a. *Pencapaian Akademik dan Kebolehgunaan Tenaga Siswazah Institusi Pengajian Tinggi*. USM IPPTN Monograf 3/2004. Penang.

Morshidi Sirat, Abd Aziz Buang, Abd Majid Mohd Isa, Ambigapathy Pandian, Moha Asri Abdullah, Mohamed Dahlan Ibrahim, Mohd Haflah Piei, Molly N.N. Lee, Munir Shuib, Rosni Bakar, Rujhan Mustafa, Shukran Abdul Rahman, Siti Zubaidah A. Hamid, Susie See Ching Mey, and Wan Ahmad Kamil Mahmood. 2004b. *Masalah Pengangguran di Kalangan Siswazah*. USM IPPTN Monograf 2/2004. Penang.

Nativel, Corrine. 2004. *Economics Transition, Unemployment and Active Labour Market Policy: Lessons and Perspectives from the East German Bundeslander*. UK: University of Birmingham Press.

OECD. 1998. Early identification of jobseekers at risk of long-term unemployment: the role of profiling. Paris: Organization for Economic Co-Operation and Development.

Payne, C. and J. Payne. 2000. Early identification of the long-term unemployed. PSI Research Discussion Paper 4. London: PSI.

Saodah Wok, Shukran Abdul Rahman, Hariyati Shahrima Abdul Majid, Noor Azlan Mohd Noor, Ainul Madziah Zubairi and Danial Mohd Yusof. 2006. On profiling the first-degree graduates' employability: a longitudinal analysis. Paper presented at *International Higher Education Policy Research and Management Forum* 2006, 8-11 Nov 2006, USM, Penang, Malaysia.

Sim, Leoi Leoi. June 3, 2006. *Abdullah: fill the vacancies*. The Star Online Newspaper. Retrieved June 3, 2006 from url: *http://thestar.com.my/news/story.asp?file=/2006/6/3/nation/14434625&sec=nation*

Appendix 1: Definition and measurement of variables

Variable abbreviation	Definition	Measurement
Type of degree		
UUM Economics	Dummy variable for UUM Bachelor of Economics (UBec) with comparison group of UTAR Bachelor of Accounting	1 if UBec and 0 otherwise
UUM Public/ Devep Mgt	Dummy variable for UUM Bachelor of Public Mgt (UBPM) and Devep Mgt (UBDM) with comparison group of UTAR Bachelor of Accounting	1 if UBPM/UBDM and 0 otherwise
UUM Buss Admin	Dummy variable for UUM Bachelor of Business Admin (UBBA) with comparison group of UTAR Bachelor of Accounting	1 if UBBA and 0 otherwise
UUM Accounting	Dummy variable for UUM Bachelor of Accounting (UBACC) with comparison group of UTAR Bachelor of Accounting	1 if UBACC and 0 otherwise
UUM Info Tech	Dummy variable for UUM Bachelor of Info Tech (UBIT) with comparison group of UTAR Bachelor of Accounting	1 if UBIT and 0 otherwise
UUM Other degrees	Dummy variable for UUM Others degree (Tourism(TOU)/Education(EDU)/ Technology Mgt(TECH)/Decision Sciences (DECS)) with comparison group of UTAR Bachelor of Accounting	1 if TOU/EDU/TECH/ DECS and 0 otherwise
UUM Human Res/SocWork	Dummy variable for UUM Bachelor of Human Res Mgt (UBHR) or Soc Work Mgt (UBSW) with comparison group of UTAR Bachelor of Accounting	1 if UBHR/UBSW and 0 otherwise
UUM Int Buss/ Issues Mgt	Dummy variable for UUM Bachelor of International Business (UBIBM) or Issues Mgt (UBISM) with comparison group of UTAR Bachelor of Accounting	1 if UBIBM/UBISM and 0 otherwise
UUM Finance	Dummy variable for UUM Bachelor of Banking (UBBank) or Finance (UBFin) with comparison group of UTAR Bachelor of Accounting	1 if UBBank/UBFin and 0 otherwise
UUM Communication	Dummy variable for UUM Bachelor of Communication (UComm) with comparison group of UTAR Bachelor of Accounting	1 if UComm and 0 otherwise

UTAR Buss Admin	Dummy variable for UTAR Bachelor of Business Admin (TBBA) with comparison group of UTAR Bachelor of Accounting	1 if TBBA and 0 otherwise
UTAR Info Tech/CompSc	Dummy variable for UTAR Bachelor of Info Sys (TBIS) or Info Sys Eng (TBISE) or Comp Sc (TBCS) with comparison group of UTAR Bachelor of Accounting	1 if TBIS/TBISE /TBCS and 0 otherwise
UTAR Other degrees	Dummy variable for UTAR Bachelor of Chinese Stud(TBChS) or Journalism (TBJ) or Pub Relations (TBPR) with comparison group of UTAR Bachelor of Accounting	1 if TBChS/TBJ/TBPR and 0 otherwise
Socio-demographics		
Age	Age	In non negative discrete numbers (years)
Male	Dummy variable for male	1 if male and 0 if female
Chinese	Dummy variable for ethnic group of Chinese with comparison group being Malay	1 if Chinese and 0 if otherwise
Other ethnic group	Dummy variable for ethnic group of Indian or others (OthEthn)with comparison group being Malay	1 if OthEthn and 0 if otherwise
Rural	Dummy variable for home town of other than big cities or state capital (cityO)	1 if cityO and 0 if otherwise
Academic attainment	Cummulative Grade Point Average	Continuous scale: from 2 to 4
Industrial training	Dummy variable for having practicum /industrial training	1 if yes and 0 if no
Family background		
Family size	Family size	In non negative discrete number
Father's education level	Father's education level from 1 being no formal schooling to 7 A level & above	1=no formal schooling; 2=do not complete primary; 3=complete primary; 4=do not complete secondary; 5=complete secondary;6=O level or equivalent; 7=A level & above

Mother's education level	Mother's education level from 1 being no formal schooling to 7 A level & above	1=no formal schooling; 2=do not complete primary; 3=complete primary; 4=do not complete secondary; 5=complete secondary;6=O level or equivalent; 7=A level & above
Father eco-inactive	Dummy variable father's employment	1 if eco inactive and 0 otherwise
Mother eco-inactive	Dummy variable mother's employment status of eco inactive status of eco inactive	1 if eco inactive and 0 otherwise
Dummy variables: Month		
Month 1 – 13 & above	Dummy variables for each month of unemployed (comparison group: month 1)	

Appendix 2: Validation sample

Weibull				Piecewise				Gompertz			
id	Emp[1]	Une[2]	Rank	id	Emp[1]	Une[2]	Rank	id	Emp[1]	Une[2]	Rank
353	2	29	1	393	1	20	1	393	1	20	1
393	1	31	2	353	2	21	2	353	2	20	2
407	2	38	3	177	1	27	3	407	2	28	3
242	2	40	4	321	1	34	4	226	2	31	4
226	2	41	5	226	2	38	5	242	2	31	5
177	1	42	6	466	1	46	6	177	1	31	6
243	2	42	7	294	3	49	7	243	2	31	7
345	1	45	8	407	2	53	8	345	1	34	8
294	3	52	9	242	2	56	9	294	3	39	9
321	1	54	10	55	1	58	10	321	1	41	10
466	1	58	11	243	2	62	11	466	1	47	11
55	1	65	12	345	1	62	12	55	1	50	12
160	2	82	13	160	2	72	13	160	2	63	13
11	2	84	14	459	1	72	14	11	2	68	14
459	1	87	15	11	2	77	15	459	1	71	15
313	0	89	16	219	2	114	16	313	0	72	16
366	2	111	17	102	1	139	17	366	2	99	17
102	1	133	18	313	0	152	18	102	1	107	18
360	1	137	19	455	1	178	19	219	2	117	19
219	2	138	20	366	2	179	20	360	1	122	20
454	1	142	21	119	4	184	21	454	1	122	21
119	4	171	22	175	2	191	22	119	4	136	22
175	2	172	23	101	0	210	23	175	2	146	23
455	1	193	24	9	2	212	24	455	1	154	24
16	0	193	25	360	1	238	25	16	0	158	25
9	2	217	26	454	1	264	26	73	1	173	26
73	1	218	27	16	0	411	27	9	2	176	27
101	0	220	28	73	1	498	28	101	0	177	28
209	2	377	29	209	2	514	29	209	2	277	29
207	0	491	30	207	0	692	30	207	0	337	30
257	0	637	31	257	0	1172	31	257	0	419	31
60	2	849	32	60	2	2491	32	60	2	511	32

Notes:
1. Emp = actual employment status where 0: unemployed; 1: full-time employment that commensurates with qualification; 2: full-time employment that does not commensurate with qualification; 3: self-employment; 4: part-time employment.
2. Une = estimated expected unemployment duration

Bottom 5 per cent: 2 graduates. (for instance, Weibull: id=257 and 60)
Bottom 10 per cent: 3 graduates. (for instance, Weibull: id=207, 257 and 60)
Bottom 15 per cent: 5 graduates. (for instance, Weibull: id=101, 207, 257 and 60)

Appendix 3: Percentage correctly predicted using more covariates

Percentage correctly predicted (in-sample evaluation)

	Bottom 5%: 8 graduates	Bottom 10%: 16 graduates	Bottom 15%: 24 graduates
Weibull	100% (8)	75% (12)	79% (19)
Piecewise	100% (8)	88% (14)	79% (19)
Gompertz	100% (8)	75% (12)	79% (19)
Naïve (% unemployed)	29% (2)	29% (5)	29% (7)

Note: Figures in parentheses are the number of graduates correctly predicted.

Percentage correctly predicted (out-of-sample evaluation)

	Bottom 5%: 2 graduates	Bottom 10%: 3 graduates	Bottom 15%: 5 graduates
Weibull	50% (1)	33% (1)	40% (2)
Piecewise	50% (1)	33% (1)	40% (2)
Gompertz	50% (1)	33% (1)	40% (2)
Naïve (% unemployed)	16% (0)	16% (0)	16% (1)

Note: Figures in parentheses are the number of graduates correctly predicted.