University of Nairobi

School of Mathematics

ISSN: 2410-1397

Master of Science in Social Statistics

Modelling Determinants of Duration of School-to-Work Transition Among Public University Graduates

Research Report in Social Statistics, Number 09, 2020

Japheth Mwamburi MALOMBO

November 2020

Submitted to the School of Mathematics in partial fulfilment for a degree in Master of Science in Social Statistics

Master of Science in Social Statistics **University of Nairobi**
November 2020

Modelling Determinants of Duration of School-to-Work Transition Among Public University Graduates

Research Report in Social Statistics, Number 09, 2020

Japheth Mwamburi MALOMBO

School of Mathematics
College of Biological and Physical sciences
Chiromo, off Riverside Drive
30197-00100 Nairobi, Kenya

Master of Science Project

Submitted to the School of Mathematics in partial fulfilment for a degree in Master of Science in Social Statistics

Prepared for The Director
Graduate School
University of Nairobi

Monitored by School of Mathematics

Dissertations in Mathematics at the University of Nairobi, Kenya.
ISSN 2410-1397: Research Report in Mathematics

DISTRIBUTOR: School of Mathematics, University of Nairobi, Kenya

Declaration and Approval

I, the undersigned declare that this project report is my original work and to the best of my knowledge. It has not been submitted in support of an award of a degree in any other university or institution of learning.

Signature Date

Japheth Mwamburi MALOMBO
Reg No. I56/87252/2016

In my capacity as a supervisor of the candidate, I certify that this report has my approval for submission.

Signature Date

Dr. Chumba Isaac Kipchirchir
School of Mathematics,
University of Nairobi,
Box 30197 - 00100,
Nairobi, Kenya.
E-mail: kipchirchir@uonbi.ac.ke

Acknowledgements

First, I thank my parents for taking me to school; my siblings for their constant support and understanding, as I endeavoured to complete my masters degree. Secondly, I thank all my lecturers at the School of Mathematics, who prepared me well; especially, Dr. Isaac Kipchirchir, for his guidance and patience throughout the period of writing this project. I also wish to acknowledge my classmates for their constant support, especially, for passing along information, ideas and relevant academic material, and all the people who willingly offered to listen to my ideas during the period of writing this report. Above all, I wish to thank The Almighty God for always providing me with good health and thirst for knowledge.

Japheth Mwamburi Malombo

Dedication

This project is dedicated to my parents, Jane and Joseph Malombo.

Abstract

In this study, we analyse Kenya STEP Survey 2013 data by World Bank. We seek to identify significant determinants of duration of school-to-work transition among public university graduates. We analyse a sample of 116 public university graduates, who reported to having taken between 0 to 60 months since graduating, to a acquire a first job. We model the data using the proportional-odds cumulative logit model and observe that type of course taken at the university and reading score are significant determinants of transition from school-to-work among graduates. Graduates who studied science, technology, engineering and mathematics (STEM) courses are 68.6% less likely to remain without a job in the first six months following graduation from the university compared to those who studied non-STEM courses, and those who recorded high reading scores were 63.8% less likely to remain without a job for more than six months after graduating from university. Other factors like sex, parental level of education, university location, among others, were not statistically significance in determining the duration of time it took a public university graduate to acquire a first job following graduation. This study, therefore, recommends to the government and other policy makers to consider creating relevant employment programs for graduates that can keep such graduates, for example those who study non-STEM courses and may take longer to become employed following graduation, engaged in such employment programs, at least for a period not less than six months upon graduating from the university.

Contents

Figures and Tables

Figures

Tables

1 CHAPTER 1: Introduction

1.1 Background

Demand for higher education continues to grow in Kenya. In 2018, the number of universities, both public and private, in Kenya doubled to over 70 universities within a decade. Student enrolments in universities increased by more that two-fold between 2011 and 2017 in both public and private universities within the same period (**Statistical Abstract, 2017**). In 2016, university enrolment rose 3-fold to a high of 504,471 students from a low of 188,774 reported in 2011, in both public and private universities (**Statistical Abstract, 2016**). In 2008, university enrolments stood at 118,239 students (**Statistical Abstract, 2013**).

Like most African countries, Kenya is experiencing a 'youth bulge'; with over 80% of it's population being under 35 years old. The youth, aged 15 - 34 years, continue to face high rates of unemployment; 67% of these youths have no jobs, among them, university graduates, **Oanda and Sifuna (2016)**.

Universities in Kenya continue to review existing degree programs and introduce new market-driven programs. This, a deliberate effort in response to an ever changing labour market environment in the country. For example, many traditional university degrees have since been reviewed to include aspects of computing and information technology —a traditional bachelor of science in statistics degree is now mostly advertised as "bachelor of science degree in statistics with information technology". New market-driven degree programs that were not there before are continually being introduced and offered at the universities. Degree programs in gender studies, monitoring and evaluation, data science and analytics, among other newer fields, are being introduced by different universities in an attempt to address the needs of the current labour market, **Oanda and Sifuna (2016)**.

Between 2011 and 2017, over 300,000 persons graduated with a first university degree in Kenya; translating to an increased supply of university graduates into the labour market, each year. In 2017 alone, universities across the country released well over 50,000 graduates (**Kenya Education Sector Analysis Report, 2018**). However, the demand for labour has grown at a much slower pace that the number of people in search of gainful employment.

The question for most graduands remains: what is next after graduating from the university?

University students are usually hopeful about the prospects of graduating, and would imagine settling into the labour market, shortly after graduation, **Njoku (2011)**. Naturally, any university student should have such expectations —specific expectations, regarding life post-graduation from the university. For most new graduates, the obvious expectation is to become employed.

It follows that, transition from school-to-work is an important and rather complex phase for every new graduate. The choice of the first job in many cases dictate the career trajectory and roles taken by an individual within their societies, in the future, **Mberia and Midigo (2018)**.

A majority of new graduates would usually seek to join the labour market —and begin their career journeys. A few others, who can afford it, prefer to take a voluntary time out and completely stay away from work and/or education, while the rest would consider a further pursuit of higher education. If follows that, a majority of new Kenyan graduates naturally seek to become employed, soon after graduating. In addition, being young adults, between the ages of 18 and 30 years, this group of persons seek to acquire financial independence from their families and becoming productive members of the society.

However, employers have continually prioritised experience over educational attainment in their advertising of job openings. This is attributed to many reasons, one being their utter belief that recent graduates may not posses the skill sets required to perform certain tasks. On the other hand, universities blame employers for not providing enough opportunities for students to gain work-related experience —by not availing enough apprenticeship and internship opportunities, **Oanda and Sifuna (2016)**. It follows that, most recent graduates are thus disadvantaged as potential candidates for jobs available in the labour markets.

The following are some facts and views regarding youth and employment in Kenya, that summarizes the introduction to this report:-

- "Youths aged between 15 and 34 form 35% of the population; those who are unemployed form 67%. This is an extremely high percentage of unemployment," the then Kenya Education Cabinet Secretary, Dr. Fred Matiang'i (**Standard Digital, March 02 2017**).
- Kenya has a high rate of unemployment (at 39.1%) and unemployed youths (at 17.6%) (**UNDP's 2017 HDI; World bank Report 2016; Business Daily, May 03 2017**)
- Employers continue to view fresh graduates as lacking the specialized skills and experience needed to perform certain tasks, **Oanda and Sifuna (2016)**.
- The number of persons graduating from Kenyan universities each year has more than doubled between 2012 (23,523) and 2015 (49,020) (**Standard Digital, March 02 2017**).

- Kenya is ranked at position 146/188 globally on the human development score in 2017.
- In 2016, out of the 832,900 jobs created, only 85,600 were in the formal sector (**Economic Survey, 2017**).

Considering that a significant number of graduates continue to join the labour market each year, for example, the estimated 50,000+ new graduates who joined the Kenyan labour market in 2017 alone, we expect the unemployment problem among graduates to worsen, even further. In addition, with more students enrolling into different universities in the country each year, it naturally follows that, the number of graduates seeking employment will only continue to grow in the future.

Brzinsky-Fay (2007) studied the labour market entry sequences for school levers in Europe. He emphasizes that transition should not just focus on one single point in time but more on transition periods, and hence becomes a sequence of more than one transition types. According to his research, transition is a process and not a single event in time.

Macanu et al. (2012) define the structure of school-to-work transition as referring to length of the process of transition, number and type of transition stages, periods of "overlap" between education and work. They observer that the outcomes of the transition process can be analysed at both macro and micro levels. At micro level, outcomes of the school-to-work transition are related to labour market participation of the graduates, occupational status, education-job mismatch, wage and wage growth, security of employment, job and career mobility, participation to training and job satisfaction.

1.2 Statement of Problem

New graduates continue to face difficulty integrating into the labour market, following graduation from universities in Kenya. More often than not, most end up loosing hope. Some even become objects of ridicule in their communities.

In turn, these graduates turn towards harmful behaviours like drug and alcohol abuse, due to stress. Others turn to crime in order to survive. Some even commit suicide because of depression. In addition, research has shown that their is a causal relationship between increased crime rate and unemployment amongst the youths.

In addition, the problem of delayed employment opportunities for graduates affects the morale of the younger members of the communities who are still pursuing basic education, especially those in secondary school level, who may see no incentive of pursuing higher education —if they continually observe members of their communities, who graduated from universities, languish in joblessness and hopelessness. This creates a 'bad picture' on the intended benefits of higher education; especially, given that other youths in the community, who did not pursue post-secondary education and are more willing to take up any sort of informal work and/or business, thrive.

It takes on average up to five years for a Kenyan graduate to transition from university to gainful employment (**Daily Nation, October 12 2014**). Recent graduates take longer to become employed after graduating from the university, because they have to compete for available jobs with more experienced workers while employers often consider the anticipated higher cost of training that may be associated with hiring a fresh graduates to be of disadvantage, **Mocanu et al. (2012)**.

It should be of interest to the policy markers to put in place measures to deal with the issue of delayed employment among recent graduates - most of whom are youths in the ages of 23-29 years. Firstly, the high rate of youth unemployment in the country poses one of the greatest threats to the successful transition of many university graduates into the workforce, **Gitonga (2014)**. Secondly and of most concern, literature has continued to show that their is a causal relationship between unemployment with crime and other ills in the society, which are largely caused by the idle and desperate youth in the same society.

If the length of time a graduate would remain unemployed (and hence idle) before transitioning into gainful employment in Kenya is reduced and/or replaced with other relevant employment programs, crime and other ills in the society will also reduce. What can policy makers consider to be an acceptable duration for a new gradates to remain unemployed before transitioning into employment in Kenya? What are the major contextual factors that influence such transition? This information should regularly be available and up to date in any progressive society. It generally holds true that a longer period of unemployment affects a person's ability to remain patient and hence may turn to crime and other ills in the society in order to cater for their basic needs and those of their dependants.

1.3 Objectives of the Study

The main objective of this study is

- to model determinants of duration of school-to-work transition among public university graduates in Kenya.

The specific objectives are

- to test for association between the independent variables of the study,
- to test for association between dependent and independent variables of the study,
- to identify significant determinants of the duration of school-to-work transition among public university graduates in Kenya.

1.4 Significance of the Study

In addition to contributing to the literature currently available on statistical modelling of labour market data, more specifically, on transition from university to employment, the findings of this study may deem useful to current university students wishing to become employed in the future and the many recent graduates manoeuvring the transition hurdle, from school-to-work, by being an eye-opener to the realities of the transition process. Policy makers may also find this study useful as it attempts to unpack the transition experiences of different graduates, by exploring the extents to which such factors, both personal and external, contributed to their transition outcomes with regard to duration of transition and hence may influence policy development around this very important issue to many young people who are attending school, especially university students.

We set out to tackle a real social problem facing many young people in developing country and seek to identify the extent to which some of the identified factors influence labour market outcomes among recent graduates, especially with regards to the length in time a new graduate who wishes to become employed may have to wait before securing a first job. As **Gitonga (2014)** notes, as a measure, most college seniors [in the United States] begin planning their transition into work while they are still in school, which is quite commendable. What then can we consider to be a measure that mitigates the effects and duration of school-to-work transition in the context of the study?

At the end of this study, we shall seek to establish a combination of some of such identified factors that significantly contributed to a successful school-to-work transition among new graduates in Kenya. We shall seek to advise on attributes that university students can consider, while still in school, that may improve their school-to-work transition outcomes, following graduation.

2 CHAPTER 2: Literature Review

Most papers studying patterns of labour market entry show that socio-demographic factors shape the structure and outcomes of school-to-work transition. First of all, gender differences in patterns of education attainment produce differences in labour market participation. Also, although education mediates effects of social background on labour market performances, social origin still influences the success in transition to the world of work **D. Hannan at al. (1996)**.

Using three waves (1982, 1986, 1990) of the National Graduate Survey (NGS), **Betts et al. (2000)** focus on the time it takes Canadian graduates to start a full-time job that lasted 6 months or more. They analyzed duration to first job using the Cox proportional hazards model and observed that PhD graduates experienced shorter duration relative to other graduates and married graduates had quicker transitions than non-married ones, but those with children had somewhat longer transitions than those without children. While female graduates had lower hazard ratios than their male counterparts, the effect was statistically significant only in the 1986 cohort.

Lassibille et al. (2001) explained the probability of young Spanish people finding a first job in 6 months or between 6 and 18 months, compared with that of finding a job in more than 18 months, using a multinomial logit specification. Regression results indicated that people with higher education (e.g. university graduates) had, all else being equal, a lower probability of being over-educated and a shorter length of unemployment. Moreover, females were less likely to find their first job in less than 18 months compared with males. However, family background has no significant influence on the duration of initial unemployment.

Survival analysis was used as well by **Biggeri et al. (2001**). With a large dataset from a survey on job opportunities for the 1992 Italian graduates (Italian National Statistical Institute), they demonstrate that information related to academic ability (final marks) has a positive effect on the probability of obtaining the first job after graduation. As regards the social background covariates, the occupational status and educational level of the parents at the time of the degree are both significant; this means that a graduate had a higher probability of obtaining a job if at least one of their parents was currently working or if at least one of their parents had a secondary school certificate or a degree.

In his research paper, **Salas-Velasco (2007)** found out that the individual characteristics of a graduate, such as the field of study, level of studies or the socioeconomic background, and individual job search characteristics bear a significant relationship to the probability of finding a first job among European graduates.

Macanu et al. (2012) employed Kaplan-Maier estimator to determine the speed of entry and the stability in the first job. The results by country were presented in a plot of the cumulative share of higher education of graduates that ever entered a first job in a country. In three out of the four studied, namely Slovenia, Lithuania and Hungary, the pattern of labour market entry was similar. In these countries, three quarters of higher education graduates entered the first job within 10-12 months from graduation. However, Poland followed a different pattern; the speed of labour market entry being considerably higher. Three quarters of higher education graduates in Poland entered in their first job after less than 6 months, while more than 90% of the graduates entered the labour market within 12 months after leaving education.

In their contribution to a paper by British Council titled: *Universities, employability and inclusive development: repositioning higher education in Ghana, Kenya, Nigeria and South Africa. International Higher Education*, **Oanda and Sifuna (2016)** observe that students considered non-academic factors like institutional status and reputation, gender, availability of jobs for certain specialisations, family networks and low economic growth in the country as factors that influenced their entry into the labour market.

3 CHAPTER 3: Methodology

3.1 Data

The data is taken from World Bank Step Survey 2013 tittled *Kenya STEP Skills Measurement Household Survey 2013 (Wave 2)*, Ref: KEN_2013_STEP-HH_v02_M. Dataset downloaded from `https://microdata.worldbank.org/index.php/catalog/2226/download/37417` on 01/01/2020

The sample size for the STEP survey (2013) is 3894 households. The Kenya sample design is a stratified 3 stage sample design. The sample was stratified by 4 geographic areas: 1-Nairobi, 2-Other Large Cities (over 100,000 households), 3- Medium cities (60,000 to 100,000 HHs) and 4-Other Urban Areas.

For this study, we shall consider respondents who responded to having graduated from a public university and subsequently made a successful attempt to become employed. We have 116 such respondents of the STEP survey 2013.

The dependent variable Y is the "time it took a graduate to get a first job since graduating from a public university in Kenya" (*duration*). This is an ordinal categorical variable with three categories, that is

$$Y = \begin{cases} 1, & \text{if } 0 \leq t \leq 6 \\ 2, & \text{if } 7 \leq t \leq 12 \\ 3, & \text{if } 13 \leq t \leq 60 \end{cases} \tag{3.1}$$

where t is time in months.

From equation 3.1 above, we observe that all the respondents in the study sample got a first job within 5 years.

The table below shows a list of fourteen (15) independent variables that we shall consider in this study (note: we create two dummy variables - *ses(middle)* and *ses(high)* for *ses* variable).

Table 3.1.1. List of Independent Variables

Variable name	Variable label	Variable type	Value label	Categories
sex	Sex of respondent	Categorical	0 1	Male Female
field	Highest fied of study	Categorical	0 1	Non-STEM Courses STEM Courses
age	Age you left formal education (graduated)	Numerical	$18 < x < 32$	N/A
uni_loc	Location of university attended	Categorical	0 1	In a different location In the same city
pedu	Maximum of parents' education	Categorical	0 1	None to secondary Tertiary
pinv	Parental involvement in education	Categorical	0 1	Low High
read	Length of material read overall score	Categorical	0 1	Medium to low High
write	Length of material written overall score	Categorical	0 1	Medium to low High
num	Numeracy overall score	Categorical	0 1	Medium to low High
comp	Frequency of computer use score	Categorical	0 1	Medium to low High
esh	Economic shocks before age 15	Categorical	0 1	No shocks At least on shock
ses	Socio-economic status (SES) at age 15	Categorical	1 2 3	Low Middle High
cert	An industry-recognised or government certificate	Categorical	0 1	No Yes
apprent	Has completed an apprenticeship	Categorical	0 1	No Yes

3.2 Logistic Regression Model

Logistic regression model is a model in the class of generalised linear regression models, which helps us to be able to estimate the probability of falling into a certain level of a categorical dependent variable, given a set of independent variables. The independent variables may be a mixture of categorical (ordinal and/or nominal) and numerical/ratio variables.

The polynomial logistic regression model is expressed as follows

$$\ln\left(\frac{Pr(Y=c_j)}{1-Pr(Y=c_j)}\right)=\beta_{0j}+\beta_{1j}x_1+\beta_{2j}x_2+\cdots+\beta_{kj}x_k \tag{3.2}$$

where β_{0j} is the intercept of the model and $\beta_{1j},\beta_{2j},..,\beta_{kj}$ are the regression coefficients for the k independent variables in the model and $Pr(Y=c_j)=p_j$, is the probability of belonging to c_j (category *j*) of the dependent variable.

There are three general variations of the logistic regression model that may be considered

- **Binary logistic regression model**: this is used to model data when the dependent variable contains two categories against a set of independent variables.
- **Nominal logistic regression model**: this is used to model data when the dependent variable contains three or more *unordered* categories against a set of independent variables.
- **Ordinal logistic regression model**: this is used to model data when the dependent variable contains three or more *ordered* categories against a set of independent variables.

3.2.1 Odds Ratio (OR)

Odd ratio is a measure of the odds that an event will occur in one group (category) compared to the same event happening in another group (category).

$$\text{Odds Ratio (OR)}=\frac{\text{Odds of event for group A}}{\text{Odds of event for group B}} \tag{3.3}$$

where $0\leq OR<\infty$.

Referring to equation (3.3), the odds ration (OR) is referred to when interpreting results of the logistics regression, and is given by

$$\text{Odds ratio (OR)} = \exp(\beta_j) \tag{3.4}$$

where β_j is the regression coefficient.

Referring to equation (3.3), we interpret odds ratio (OR) as follows

- $OR < 1$: the event of interest is less likely to occur in group A compared to group B.
- $OR = 1$: the event of interest is equally likely to occur in group A and group B.
- $OR > 1$: the event of interest is more likely to occur in group A compared to group B.

3.2.2 Ordinal Logistic Regression

In this study, we shall use ordinal logistic regression model, sometimes referred to as **cumulative logit model**. This is so because the dependent variable, *duration*, is an ordinal categorical variable and the independent variables are a mixture of categorical (nominal and ordinal) and numerical variables.

Description of the Model

Let, Y be a dependent variable with *m* categories, $c_1, c_2, ..., c_j, .., c_m$, where the categories have a natural ordering. The associated probabilities are $p_1, p_2, .., p_m$ and the cumulative probability of an event of interest occurring into or below category j of the dependent variable is given by

$$Pr(Y \leq c_j) = \sum_{k=1}^{j} p_k = p_1 + p_2 + \cdots + p_j \tag{3.5}$$

It follows that, we have a model expressed in the form

$$\ln\left(\frac{Pr(Y \leq c_j)}{1 - Pr(Y \leq c_j)}\right) = \beta_{0j} + \beta_{1j}x_1 + \beta_{2j}x_2 + \cdots + \beta_{kj}x_k \tag{3.6}$$

where β_{0j} is the intercept of the model and $\beta_{1j}, \beta_{2j}, .., \beta_{kj}$ are the regression coefficients for the k independent variables in the model and $Pr(Y \leq c_j)$ is the cumulative probability of an event of interest occurring into or below category j of the dependent variable, and

$$logit[Pr(Y \leq c_j)] = \ln\left(\frac{Pr(Y \leq c_j)}{1 - Pr(Y \leq c_j)}\right) = \ln\left(\frac{p_1 + p_2 + \cdots + p_j}{p_{j+1} + \cdots + p_m}\right) \tag{3.7}$$

is the logit link function.

Making $Pr(Y \leq c_j)$ subject of the formula in equation (3.6), we get the alternative expression of the model, as follows

$$Pr(Y \leq c_j) = \frac{exp(\beta_{0j} + \beta_{1j}x_1 + \cdots + \beta_{kj}x_k)}{1 + exp(\beta_{0j} + \beta_{1j}x_1 + \cdots + \beta_{kj}x_k)} \tag{3.8}$$

From equation (3.6) and (3.7), and for a dependent variable with *m* categories, we fit a sequence of *m-1* cumulative logits, L_j, as follows;-

$$\begin{aligned} L_1 &= \ln\left(\frac{p_1}{p_2 + \cdots + p_m}\right) \\ L_2 &= \ln\left(\frac{p_1 + p_2}{p_3 + \cdots + p_m}\right) \\ &\vdots \\ L_{m-1} &= \ln\left(\frac{p_1 + p_2 + \cdots + p_{m-1}}{p_m}\right) \end{aligned} \tag{3.9}$$

We shall make a no multicollinearity assumption and test for it in Chapter 4, to ensure that at least two or more of the independent variables are not highly correlated with each other. We shall use Pearson's chi-square test of independence to test the following hypothesis

H_0: "X_i is not associated with X_j, $i \neq j$" *vs.* H_1: "X_i is associated with X_j, $i \neq j$"

The hypothesis shall be tested using the following test statistic at $\alpha = 0.05$ level of significance;-

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{3.10}$$

which has $(r-1)(c-1)$ degrees of freedom (*df*). Also, we can use the p-value and reject H_0 when p-value generated is less than 0.05.

In this study, we shall simplify the model in equation (3.6) by requiring that the coefficients $\beta_{1j}, \beta_{2j}, .., \beta_{kj}$ be the same for all the *m-1* logits (outcomes). Therefore, for a dependent variable with *m* categories, we shall have *m-1* logit models, which are like parallel lines, since only the intercepts (β_{0j}) are different. All the other $\beta_1, \beta_2, .., \beta_k$ are the same for the m-1 equations. This simplification implies that the effects of the independent variables in the models do not change as we move from one category of the dependent variable to another. This model simultaneously uses all cumulative logits and is referred to as the **proportional-odds cumulative logit model**, given below

$$\ln\left(\frac{Pr(Y \le c_j)}{1 - Pr(Y \le c_j)}\right) = \beta_{0j} + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k \tag{3.11}$$

For example, for a continuous independent variable X, the following chart depicts the model when m=4.

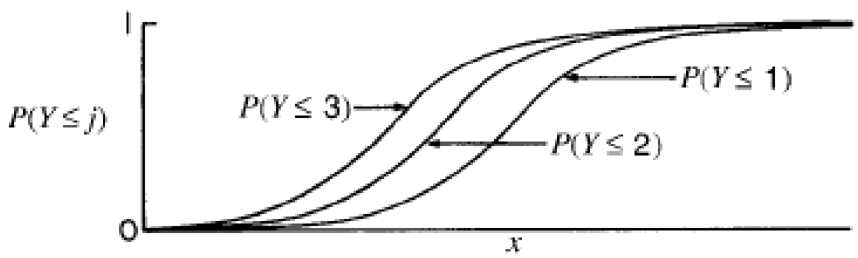

Figure 3.2.1. Cumulative logit model with effect independent of cutpoint

For fixed *j*, the response curve is a logistic regression curve for a binary response with outcomes $Y \le j$ and $Y > j$. The response curves for $j = 1, 2$, and 3 have the same shape.

For an ordinal dependent variable with *m* categories, the following are the *m-1* logit equations derived from equation (3.8) and (3.11)

$$p_1 = \frac{exp(\beta_{01}+\beta_1 x_1+...+\beta_k x_k)}{1+exp(\beta_{01}+\beta_1 x_1+...+\beta_k x_k)}$$

$$p_j = \frac{exp(\beta_{0j}+\beta_1 x_1+...+\beta_k x_k)}{1+exp(\beta_{0j}+\beta_1 x_1+...+\beta_k x_k)} - \frac{exp(\beta_{0j-1}+\beta_1 x_1+...+\beta_k x_k)}{1+exp(\beta_{0j-1}+\beta_1 x_1+...+\beta_k x_k)}, \text{ for } j=2,3,..,m-1$$

$$\vdots$$

$$p_m = 1-\sum_{j=1}^{m-1} p_j \tag{3.12}$$

Now, considering this study, we have a dependent variable *duration* with *m=3* categories and *k=15* independent variables. We have 2 parallel lines given by the following

$$p_1 = \frac{Pr(Y = 1)}{Pr(Y = 2 \text{ or } Y = 3)}$$

$$p_2 = \frac{Pr(Y = 1 \text{ or } Y = 2)}{Pr(Y = 3)} \tag{3.13}$$

$$p_3 = 1-(p_1+p_2)$$

3.3 Fitting the Model

In this study, we have a dependent variables with *m=3* categories and *k=15* independent variables. We shall seek to estimate all the intercepts and coefficient, β_{01}, β_{02} for the null model and $\beta_{01}, \beta_{02}, \beta_1, \beta_2, ..., \beta_{15}$ for the saturated model, using the maximum-likelihood (ML) estimation procedure.

The likelihood function is $L(\underline{\beta})$ and the log-likelihood function is $\ell(\underline{\beta}) = ln[L(\underline{\beta})]$.

We obtain the ML estimates for the logit models by solving the likelihood equations

$$\frac{\partial \ell(\underline{\beta})}{\partial \beta_{01}} = 0, \textit{ for the first intercept}$$

$$\frac{\partial \ell(\underline{\beta})}{\partial \beta_{02}} = 0, \textit{ for the second intercept} \tag{3.14}$$

$$\frac{\partial \ell(\underline{\beta})}{\partial \beta_{j}} = 0, \textit{ for model coefficients (j=1,2,..,k)}$$

We shall first fit the saturated model to the data. We shall then step-wisely fit more models to the data, first by dropping variables that did not satisfy the no multicollinearity assumption, followed by dropping variables with high p-values —we shall keep dropping independent variables until we arrive at the most sufficient model. We shall them test the adequacy of all fitted models by comparing such models, including the saturated model, to the null model.

To determine the best model, we shall choose the most optimal model that is complex enough to fit the data well, but simple to interpret; smoothing rather than over-fitting the data.

Testing for Significance of Model Fit

To select the best model for the data, we can use the methods of the **deviance test statistic** (D) and/or alternatively the **likelihood ratio statistic** (G^2). We can also use the method of Akaike Information Criterion (AIC), which shows the closeness of model's fitted values to the true value in terms of a certain expected value.

The **deviance test statistic (*D*)** describes a lack of fit in this case, the larger the value, the poorer the fit. When using p-values, a significant fit is achieved where the test gives a non significant result; i.e., p-value > α. We shall seek to test the following hypotheses

H_O: "The fitted model is better fit" *vs.* H_1: "The saturated model is better fit"

The deviance test statistics is calculated as follows

$$D = 2[\ell(\underline{y}, \underline{y}) - \ell(\underline{y}, \underline{\hat{\mu}})] \tag{3.15}$$

and D has an approximate chi-square distribution

$$D \sim \chi^2{}_{(n-p)} \tag{3.16}$$

where n is the number of observations and p is the number of parameters estimated.

The **likelihood ratio test statistic (G^2)** describes goodness of fit. When using p-values, a significant fit is achieved where the test gives a significant result; i.e., p-value < α. We shall seek to test the following hypotheses

H_0: "The null model is a better fit" vs H_1: "The fitted model is a better fit"

The likelihood ratio test statistics is calculated as follows;-

$$G^2 = 2[\ell(\underline{y}, \underline{\hat{\mu}}) - \ell(\underline{y}, \underline{\hat{\beta}})] \tag{3.17}$$

and G^2 has an approximate chi-square distribution,

$$G^2 \sim \chi^2{}_{(k)} \tag{3.18}$$

where k is the number of independent variables included in the model.

For this study, we shall use both the likelihood ratio test statistic and deviance test statistics to test the significance of the fitted model.

Significance of Independent Variables in the Model

We shall seek to test the statistical significance of the ML estimates obtained for the model after solving equation (3.14) above. We shall test the following hypotheses

H_0: $\beta_j = 0$ *vs* H_1: $\beta_j \neq 0$, for $j = 1,2,..,k$

We use the test statistics

$$Z_{calc} = \frac{\hat{\beta}_j}{s.e(\hat{\beta}_j)} \sim N(0,1) \tag{3.19}$$

We reject H_0 if Z_{calc} in equation (3.19) above is greater that $Z_{\frac{\alpha}{2}}$ and conclude that the independent variable X_j is statistically significant in the model.

Alternatively, referring to equation (3.4), we can test for significance of the independent variable by using the Odds Ration (OR), and test

H_0: $OR = 1$ vs H_1: $OR \neq 1$

for all the $\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k$. In this case, the independent variable X_j is statistically significant in the model if the confidence interval for *OR* does not include the value **one**, i.e., ether the lower confidence limit is greater than one or the upper confidence limit is lesser than one.

Note: statistical significance of an independent variable in a model may not be used as the only criterion to determine inclusion of an independent variable in a model. Other theoretical consideration may be taken into account when including independent variables in a model, even if they are not statistically significant. This helps in reducing biases on the effects of the other independent variables included in a model.

We shall use $pseudo - R^2$ statistic to determine the measure of strength of association between dependent and independent variables fitted in the models.

Interpretation of Results of the Model

We shall interpret the results by means of odds ratio discussed in section 3.2.1.

- For $0 \leq OR \leq 1$: we interpret that an even of interest is $100(1-OR)\%$ less likely to occur for every unit increase in the dependent variable.
- For $1 < OR \leq 2$: we interpret that an even of interest is $100(OR-1)\%$ more likely to occur for every unit increase in the dependent variable.
- For $OR > 2$: we interpret that an even of interest is $OR - times$ more likely to occur for every unit increase in the dependent variable.

3.4 Analysis software

We shall analyse the data by using the statistical and data (Stata) software. Stata is a general-purpose statistical software package created in 1985 by StataCorp. Stata's capabilities include data management, statistical analysis, graphics, simulations, regression, and custom programming.

4 CHAPTER 4: Data Analysis and Results

4.1 Introduction

We shall seek to model determinants of duration of time to first job among university graduates —both personal and external, using the proportional-odds logit model. Subsequently, we shall seek to identify determinants that may be most favourable to achieving a shorter school-to-work transition duration among graduates. The dependent variable is ordinal with 3 categories: *0 - 6 months*, *0 - 12 months* and *0 - 60 months*; the independent variables are a mixture of categorical (ordinal and nominal) and numerical variables.

4.2 Exploratory Data Analysis

We shall seek to summarize and explore the assumptions of the model using various statistical procedures, more specifically, for the following

- Contingency tables —to help summarize the data and explore possible association between variables, by generation counts and frequencies.
- Charts and Graphs —to help visualize the data.
- Chi-square tests of independence —to test the assumption of no multicollinearity and explore association between variables, both dependent and independent variables.
- Summarize numerical data using measures of central tendency like mean, range, et cetera.

4.2.1 Exploring Depending Variable

The dependent (response) variable is the time it took a graduate who attended a public university to get a first job since graduating, *duration*. This is a count variable indicating the number of months a graduated had to wait before getting a first job. The table below shows the frequencies

Table 4.2.1. Summary of dependent variable

Time to find first job since graduating (in months)	
Category	*Frequency (%)*
0 - 6 months	61 (52.6)
0 - 12 months	37 (31.9)
0 - 60 months	18 (15.5)

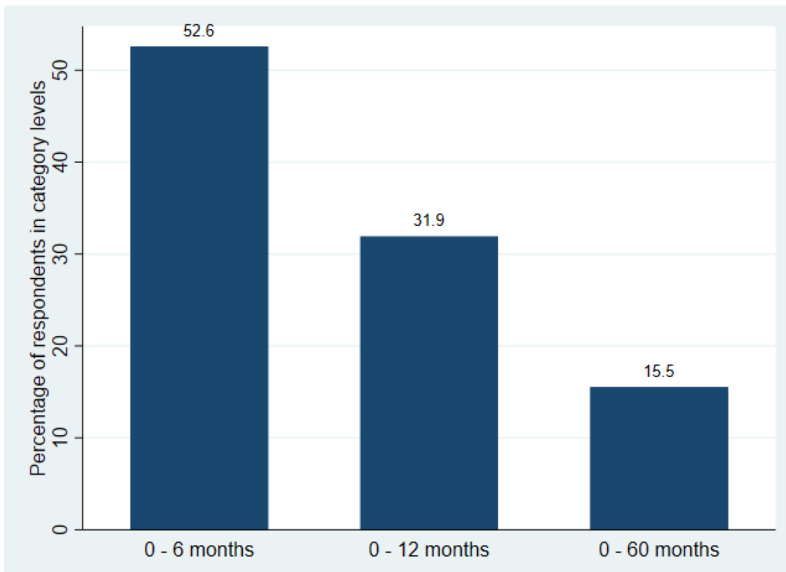

Figure 4.2.1. Bar chart for time to first job (in months)

We observe that 52.6% (n=116) of the respondents got a first job within 6 months since graduating.

4.2.2 Exploring Independent Variables

Tests of Association between Independent Variables

We check the no multicollinearity assumption by testing for association for all combinations of the independence variables by using Pearson's chi-square tests. We have

$$^{13}C_2 = \frac{13!}{2!(13-2)!} = 78$$

chi-square tests of association that we shall conduct between the independent variables in this study.

We shall test the following hypothesis, for each combination of variables

H_0: "X_i is not associated with X_j *vs* H_1: "X_i is associated with X_j, $i \neq j$"

The p-values generated from the chi-square tests of association are presented in the table below

Table 4.2.2. Test of association between independent variables

Independent variables	*sex*	*field*	*uni_loc*	*read*	*write*	*num*	*comp*	*pedu*	*esh*	*pinv*	*ses*	*cert*	*apprent*
Sex (sex)	0												
Highest field of study (field)	0.227	0											
Location of university attended (uni_loc)	0.618	0.237	0										
Length of material read overall score (read)	0.921	0.642	0.341	0									
Length of material written overall score (write)	0.921	0.642	0.341	0.000	0								
Numeracy overall score (num)	0.618	0.008	0.191	0.043	0.029	0							
Frequency of computer use score (comp)	0.036	0.228	0.035	0.321	0.468	0.815	0						
Max. parent education (pedu)	0.117	0.270	0.064	0.552	0.411	0.004	0.712	0					
Economic shocks before age of 15 (esh)	0.223	0.292	0.250	0.039	0.539	0.008	0.662	0.000	0				
Parental involvement in education (pinv)	0.941	0.866	0.102	0.635	0.471	0.718	0.198	0.051	0.025	0			
Socio-economic status at age 15 (ses)	0.098	0.279	0.784	0.187	0.351	0.149	0.079	0.001	0.001	0.111	0		
Industry-recognised or government certificate (cert)	0.451	0.150	0.476	0.468	0.735	0.666	0.114	0967	0.006	0.619	0.076	0	
Has completed an apprenticeship (apprent)	0.315	0.866	0.102	0.212	0.823	0.174	0.153	0.342	0.474	0.346	0.111	0.166	0

We observe that most of the independent variables in this study indicate no statistically significant association between them. However, we observe p-values that are less than the level of significance of $\alpha = 0.05$ for a combination of some of the independent variables (*presented in red-font*). For example, social economic status at age 15 (*ses*) and maximum parent education (*pedu*) show a statistically significantly associated at $\alpha = 0.05$.

4.2.3 Relationships between Dependent and Independent Variables

The following table shows the results of cross-tabulations and chi-square tests of association between the categorical dependent variables (time to first job after graduating, duration) and all the categorical independent variables of this study.

Table 4.2.3. Crosstabulation between the dependent and independent variables(s)

		Time to first job after graduating *(Ordinal dependent variable)*				
		0 - 6 months	*0 - 12 months*	*0 - 60 months*		
Independent Variables	***Categories***	*n (%)*	*n (%)*	*n (%)*	χ^2	**P-value**
Gender	*Male*	40 (65.6)	29 (78.4)	13 (72.2)	1.847	0.397
	Female	21 (34.4)	8 (21.6)	5 (27.8)		
Highest field of study	*Non-STEM Courses*	39 (63.9)	27 (73.0)	17 (94.4)	6.410	0.041
	STEM Courses	22 (36.1)	10 (27.0)	1 (5.6)		
Location of university attended	*In a different location*	25 (41.0)	14 (37.8)	6 (33.3)	0.364	0.834
	In the same city	36 (59.0)	23 (62.2)	12 (66.7)		
Maximum of parents' education	*No tertiary*	23 (37.7)	14 (37.8)	10 (55.6)	1.999	0.368
	Tertiary	38 (62.3)	23 (62.2)	8 (44.4)		
Parental involvement in education	*Low*	8 (13.1)	8 (21.6)	4 (22.2)	1.539	0.463
	High	53 (86.9)	29 (78.4)	14 (77.8)		
Length of material read overall score	*Low*	9 (14.8)	13 (35.1)	6 (33.3)	6.208	0.045
	High	52 (85.2)	24 (64.9)	12 (66.7)		
Length of material written overall score	*Low*	34 (55.7)	21 (56.8)	12 (66.7)	0.703	0.702
	High	27 (44.3)	16 (43.2)	6 (33.3)		
Numeracy score	*Low*	33 (54.1)	25 (67.6)	10 (55.6)	1.805	0.406
	High	28 (45.9)	12 (32.4)	8 (44.4)		
Frequency of computer use score	*Low*	8 (13.1)	5 (13.5)	5 (27.8)	2.446	0.294
	High	53 (86.9)	32 (86.5)	13 (72.2)		
Economic shocks before age 15	*No shocks*	35 (57.4)	26 (70.3)	11 (61.1)	1.634	0.442
	At least one shock	26 (42.6)	11 (29.7)	7 (38.9)		
Socio-economic status (SES) at age 15	*Low*	13 (21.3)	7 (18.9)	2 (11.1)	3.358	0.500
	Middle	30 (49.2)	22 (59.5)	13 (72.2)		
	High	18 (29.5)	8 (21.6)	3 (16.7)		
An industry-recognised or government certificate	*No*	47 (71.1)	31 (83.8)	16 (88.9)	1.535	0.464
	Yes	14 (22.9)	6 (16.2)	2 (11.1)		
Has completed an apprenticeship	*No*	48 (78.7)	33 (89.2)	15 (83.3)	1.785	0.410
	Yes	13 (21.3)	4 (10.8)	3 (16.7)		

From the above table, we observe that highest field of study (*field*) and literacy scores (*lit*) have significant statistical association with the dependent variable, *duration*, at $\alpha = 0.05$

level of significance. The mean age at graduation is 24.3 years, with a range of 18 to 32 years.

4.3 Fitting the Proportional-Odds Cumulative Logit Model

Now, we shall seek to fit the model as provided in equation (13). First, we shall fit a saturated model, which will include all the variables in this study. Then, we fit other models to the data, by considering significance of the independent variables and other theoretical considerations —regression results are presented in the table below

Table 4.3.1. Saturated and Fitted Models

duration		*Saturated Model*		*Model I*		*Model II*		*Model III*		Model IV	
		$\hat{\beta}$	P>z	$\hat{\beta}$	P>z	$\hat{\beta}$	P>z	$\hat{\beta}$	P>z	$\hat{\beta}$	P>z
cons_1		-2.948		-3.188		-1.670		-1.129		-0.820	
cons_2		-1.072		-1.362		0.136		0.645		0.883	
sex	*Female*	-0.625	0.173	-0.605	0.160	-0.539	0.201	-0.525	0.213		
field	*STEM courses*	-1.234	0.009	-1.145	0.011	-1.124	0.012	-1.160	0.009	-0.918	0.031
age		-0.048	0.682	-0.048	0.670						
uni_loc	*In the same city*	0.702	0.100	0.618	0.131	0.676	0.095	0.611	0.124		
read	*High*	-1.117	0.031	-1.049	0.016	-1.056	0.012	-1.017	0.015	-0.885	0.030
write	*High*	0.102	0.823								
num	*High*	0.189	0.673								
comp	*High*	-0.243	0.669								
pedu	*Tertiary*	-0.487	0.279	-0.678	0.112	-0.534	0.180	-0.601	0.124		
esh	*At least one shock*	-0.196	0.688	-0.321	0.478						
pinv	*High*	-0.684	0.192	-0.668	0.191	-0.628	0.196				
ses	*Middle SES*	0.531	0.340								
	High SES	-0.237	0.733								
cert	*Yes*	-0.419	0.445	-0.287	0.589	-0.455	0.372				
apprent	*Yes*	-0.528	0.364	-0.575	0.304						

To test for statistical significance of the independent variables in fitted models above, we test the hypothesis

H_0: $\beta_j = 0$ vs H_1: $\beta_j \neq 0$

We observe that, independent variables: field of study (*field*) and length of material read overall score (*read*), are statistically significant across all the fitted models, i.e., $p - values < \alpha = 0.05$ level of significance. However, all the other independent variables considered in this study are not statistically significantly in the models fitted at $\alpha = 0.05$ level of significance.

Next, we shall look at the significance of the models —saturated and fitted model. Table below shows the model statistics

Table 4.3.2. Saturated and Fitted Model Statistics

Model Statistics	**Saturated**	**Model I**	**Model II**	**Model III**	**Model IV**
Sample size	116	116	116	116	116
Likelihood ratio statistic (G^2)	24.17	20.6	18.62	16.15	10.15
P-value	**0.062**	0.0241	0.0095	0.0064	0.0062
Pseudo R2	0.1051	0.0895	0.081	0.0702	0.0441
Log likelihood	-102.938	-104.725	-105.711	-106.95	-109.947

We shall test the following hypotheses for model significance

H_0: "The null model is a better fit" *vs* H_1: "The fitted model is a better fit"

We observe that, for the saturated model, the calculated likelihood ratio statistic for the saturated model, $G^2 = 24.17$, is less than $\chi^2_{(15)} = 24.996$, hence we fail to reject H_0 and conclude that the saturated model is not statistically significant at $\alpha = 0.05$ level of significance (also, $p - value > 0.05$ for the saturated model); we conclude that the null model is a better fit.

However, fitted models; *model I, model II, model III* and *model IV* are statistically significance at $\alpha = 0.05$ level of significance because the p-values<0.05, for all the four fitted models. We conclude that the fitted models are better fits than the null model.

Choosing the Model

In this study, we shall choose the model that best fits the data. Firstly, such a model must be significant at $\alpha = 0.05$ level of significance. *Model I*, *model II, model III* and *model IV* pass this test.

In *model I*, we have dropped all independent variables that did not satisfy the no multicollinearity assumption tested earlier. *Model I* has the highest log-likelihood among the four significant models fitted.

In *model II*, we further drop independent variables that are not statistically significant in *model I* and/or do not hold sufficient theoretical relevance to this study.

In *model III*, we further drop independent variables that we have established may hold theoretical relevance to the study but have high p-values and we keep only independent variables that have a p-value of less than 0.25.

Model IV includes only the two significant independent variables of the study *field* and *read*.

Now, the saturated (full) model is not statistically significant. Hence, we shall choose *model I* to be the saturated model, since it satisfies the no multicollinearity assumption and is also statistically significant at $\alpha = 0.05$ level of significance. In addition, we shall choose *model III* as the best model that fits the data —firstly, *model III* is statistically significant at $\alpha = 0.05$ level of significance. Secondly, it contains all the significant variables *read* and *field* as well as variables that we have considered to hold theoretically relevance in the study of duration of school-to-work transition —*sex, uni_loc* and *pedu* (also, considering that the variables have small p-values, although not statistically significant at the $\alpha = 0.05$ level of significance that we choose in this study).

Next, we shall now compare the **fitted model** —*model III* and the **saturated model** —*model I*, using the deviance test statistic (D). We shall test the hypothesis

H_0: "The fitted model is a better fit" vs H_1: "The saturated model is a better fit"

Referring to log-likelihoods for *model I* (-104.725) and *model III* (-106.95) and equation (3.15) and (3.16), we calculate the deviance statistic as follows

$$D_0 - D_1 = 2[-104.725 - (-106.95)] = 4.45$$

$$\chi^2{}_{(n-p_0)-(n-p_1)} = \chi^2{}_{(116-7)-(116-12)} = \chi^2{}_{(5)} = 11.07$$

We observe that $D_0 - D_1 < \chi^2_{(5)}$ value at $\alpha = 0.05$, hence we fail to reject H_0 and conclude that the *model III* is a better fit than *model I*.

The table below shows the fitted model of choice

Table 4.3.3. The Fitted Model

duration		**Coef.**	**Std. Err.**	**z**	**P>z**	**[95% Conf. Interval]**	
cons_1		-1.129	0.492			-2.094	-0.165
cons_2		0.645	0.487			-0.309	1.599
sex	*Female*	-0.525	0.422	-1.24	0.213	-1.351	0.302
field	*STEM courses*	-1.160	0.443	-2.62	0.009	-2.028	-0.292
uni_loc	*In the same city*	0.611	0.397	1.54	0.124	-0.168	1.389
read	*High*	-1.017	0.418	-2.43	0.015	-1.837	-0.198
pedu	*Tertiary*	-0.601	0.391	-1.54	0.124	-1.367	0.165

Fitted model equation

$$\begin{aligned}\ln\left[\frac{Pr(Y=1)}{Pr(Y=2 \text{ or } Y=3)}\right] &= -1.129-0.525*sex(Female)\\ &-1.160*field(STEM\ courses)\\ &+0.611*uni\_loc(In\ the\ same\ city)\\ &-1.017*read(High)\\ &-0.601*pedu(Tertiary)\end{aligned} \tag{4.1}$$

$$\begin{aligned}\ln\left[\frac{Pr(Y=1 \text{ or } Y=2)}{Pr(Y=3)}\right] &= 0.645-0.525*sex(Female)\\ &-1.160*field(STEM\ courses)\\ &+0.611*uni\_loc(In\ the\ same\ city)\\ &-1.017*read(High)\\ &-0.601*pedu(Tertiary)\end{aligned}$$

Referring to equation (3.12) and (3.13), we now calculate p_1, p_2 and p_3 using the fitted model in equation (4.1), as follows

$$p_1 = \frac{\mathit{exp(\text{-}1.129\text{-}0.525{*}sex\text{-}1.160{*}field + 0.611{*}uni\_loc\text{-}1.017{*}read\text{-}0.601{*}pedu)}}{1+\mathit{exp(\text{-}1.129\text{-}0.525{*}sex\text{-}1.160{*}field + 0.611{*}uni\_loc\text{-}1.017{*}read\text{-}0.601{*}pedu)}}$$

$$\begin{aligned} p_2 = {} & \frac{\mathit{exp(0.645\text{-}0.525{*}sex\text{-}1.160{*}field + 0.611{*}uni\_loc\text{-}1.017{*}read\text{-}0.601{*}pedu)}}{1+\mathit{exp(0.645\text{-}0.525{*}sex\text{-}1.160{*}field + 0.611{*}uni\_loc\text{-}1.017{*}read\text{-}0.601{*}pedu)}} \\ & - \frac{\mathit{exp(\text{-}1.129\text{-}0.525{*}sex\text{-}1.160{*}field + 0.611{*}uni\_loc\text{-}1.017{*}read\text{-}0.601{*}pedu)}}{1+\mathit{exp(\text{-}1.129\text{-}0.525{*}sex\text{-}1.160{*}field + 0.611{*}uni\_loc\text{-}1.017{*}read\text{-}0.601{*}pedu)}} \end{aligned} \tag{4.2}$$

$$p_3 = 1-(p_1+p_2)$$

Odds Ratios

Next, we calculate the odds ratios for the fitted model and interpret results.

Table 4.3.4. The Fitted Model OR

duration		**Odds Ratio**	**Std. Err.**	**z**	**P>z**	**[95% Conf. Interval]**	
sex	*Female*	0.592	0.250	-1.24	0.213	0.259	1.352
field	*STEM courses*	0.314	0.139	-2.62	0.009	0.132	0.747
uni_loc	*In the same city*	1.841	0.732	1.54	0.124	0.845	4.012
read	*High*	0.362	0.151	-2.43	0.015	0.159	0.820
pedu	*Tertiary*	0.548	0.214	-1.54	0.124	0.255	1.179

From the odds ratio of the chosen model, we interpret as follows: holding all other factors constant

- a female graduate was 40.8% less likely to remain without a job for more than 6 months following graduation from a public university in Kenya compared to a male graduate;
- a graduate who studied a STEM course at a public university in Kenya was 68.6% less likely to have to wait for more that 6 months before getting a first job compared to one who studied non-STEM courses;
- a graduate who went to university in the same city they lived in was 84.1% more likely to remain without getting a first job in the first 6 months following graduation compared to a graduate who went to a public university elsewhere;
- compared to a graduate who recorded a low reading score, a public university graduate in Kenya who scored high in reading score was 63.8% less likely to have to stay for more than 6 months before getting a first job, following graduation;
- a graduate who has at least one parent having tertiary education is 45.2% less likely to remain without a first job 6 months following graduation compared to one who did not have at least a parent having tertiary education.

5 CHAPTER 5: Conclusions and Recommendations

5.1 Conclusions

Considering the challenges graduates continue to face in their quest to become employed upon graduating from the university, we observe that certain factors significantly contribute towards reducing the duration of time a graduate may take to transition from school into gainful employment.

In this study, one such determinant is reading score (length of material read overall score). We observe that, holding all other factors constant, a graduate who scored high in the reading score was 63.8% less likely to remain without employment for more than 6 months upon graduating from the university compared to those who recorded a low reading score. Therefore, it is of importance for university students wishing to become employed upon graduation to consider working on their reading skills. Such a skill is of significance importance in determining duration of transition from school-to-work among public university graduates.

Another significant determinant of duration of transition from school-to-work among public university graduates is the type of course studied at the university. We observe that a graduate who studied science, technology, engineering and mathematics (STEM) courses was 68.6% less likely to remain without employment in the first six months following graduation from a public university, compared to those who studied non-STEM courses.

Although not statistically significant at $\alpha = 0.05$ level of significance, we find factors like sex, location of the university and maximum of parents' education, as factors that may become significant with a larger sample size (n).

5.2 Recommendations

The problem of youth unemployment continues to grow in Kenya, and it includes unemployment among university graduates. Parents continue to invest in the higher education of their children, with more and more students enrolling in the first-expanding sector. Unfortunately, some graduates remain without gainful unemployment for longer periods of time after graduation.

Following the results of this study, we recommend to policy makers, especially the government, to consider introducing national employment programs that may specifically target graduate employment sector. Such a program may require a graduate to voluntarily register to participate for a period not less than 6 months beginning, say two weeks after graduating from a university in Kenya. Such a program's objective may be to keep graduates gainfully engaged as they gain useful skills that may be relevant once when they become employment, and to help such graduates transition from school-to-employment. In addition, such a program for graduates will help keep such graduates away from the many ills in the society that such unemployed graduates eventually become engaged in because of idleness, from example, crime and drug abuse.

Such programs, as a recommendation of this study, may include a stipend; to help such graduates take care of themselves as they serve the country in their different capacities. In addition, some of the best performing graduates that may come out of such programs may become absorbed and remain employees of the government while others may graduate and/or leave such programs after some minimum duration, say, a first and compulsory phase of 6 months upon registration, and become employed in the private sector or take up any other gainful income generating activities, like self-employment and entrepreneurship, or continue with higher education. Such programs will ensure that graduates, for example those who study non-STEM courses and may usually take longer to become employed, remain engaged in gainful activities as they transition from university to employment.

5.3 Future Research

Results of this study may leave the reader with some more questions on the topic, that can be considered for further research, by myself or any other person who may come across this study report.

For me, and given the limitation of available secondary data on this topic, I expect, while observing the p-values of the significance of the independent variables in the saturated model, that, given a larger sample size, factors considered in this study, such as sex of the respondent (p-value, 0.213), maximum parental education (p-value, 0.124), location of the university (p-value, 0.124), among others not included in this study because of limitation to the available secondary data, such as whether the university attended is public or private, the name and/or reputation of the university, et cetera, may become significant.

References

[1] K. W. Gitonga (2014). *Navigating the transition from college to work: a study of baccalaureate graduates of a private university in Kenya.* Graduate College, University of Illinois at Urbana-Champaign.

[2] Economic Survey (2017). Kenya National Bureau of Statistics. P.O Box 30266-00100, Herufi House, Nairobi.

[3] C. Brzinsky-Fay (2007). *Lost in transition? Labour market entry sequences of school leavers in Europe.* European Sociological Review, 23(4), 409–422. https://doi.org/10.1093/esr/jcm011

[4] C. Brzinsky-Fay (2011). *School-to-Work Transitions in International Comparison.* Academic Dissertation, University of Tampere, School of Social Sciences and Humanities, Finland.

[5] D. Richard and L. Paolo (2013). *Visualising the School-To-Work Transition: an Analysis using Optimal Matching* National Institute of Economic and Social Research, 2 Dean Trench Street, London SW1P 3HE , Issue 114.

[6] B. B. Cobus (2016). *A Structural Approach to Modelling South African Labour Market Decisions.* Dissertation presented for the degree of Doctor of Philosophy (Economics) in the Faculty of Economic and Management Sciences at the University of Stellenbosch

[7] F. Rosso (2004). *The School-to-Work Transition of Canadian Post-secondary Graduates: A dynamic analysis.* Journal of Higher Education Policy and Management, Volume 26, Issue 1, pp. 35-58.

[8] G. Nora (2013). *School-to-Work Transitions — Findings from Quantitative and Qualitative Approaches in Youth Transition Research 1. Qualitative and/or Quantitative Transition Research?* Qualitative Social Research, Volume 14, No. 2, Art. 12 – May 2013.

[9] D. McVicar and M. Anyadike-Danes (2002). *Predicting successful and unsuccessful transitions from school to work using sequence method.* Journal of the Royal Statistical Society, Volume A 165, Issue 2, pp. 317-334.

[10] B. Nilsson (2017). *The School-to-work transition in developing countries.* Universite Paris-, PSL Research University, IRD, LEDa, DIAL, 75016 Paris, France.

[11] M. Ramos, P. Parente and M. Santos (2017). *A typology of professional situations in the analysis of graduate transition from higher education to the labor market.* Sociologia: Revista da Faculdade de Letras da Universidade do Porto, Vol. XXXIII, pp.47-72.

[12] M. Salas-Velasco (2007). *The transition from higher education to employment in Europe: The analysis of the time to obtain the first job.* Journal of Higher Education, Volume 54, Issue 3, pp. 333-360.

[13] G. Tchibozo (2004). *The dynamics of school-to-work transition processes of university graduates: An analysis of french data.* British Journal of Guidance and Counselling Volume 32, Issue 1, pp. 93-105.

[14] G. J. Van den Berg (2000). *Duration Models: Specification, Identification, and Multple Durations.* MPRA Paper No. 9446.

[15] I. Oanda and D. Sifuna (2016). *Divergent narratives on graduate employability in Kenya: dysfunctional institutions or dysfunctional labour markets?* Universities, Employability and Inclusive Development project (2013–16), British Council, Chapter 3: pp. 39-56.

[16] I. M. Wilson and R. D. Stern (2001). *Approaches to the Analysis of Survey Data.* Statistical Services Centre, The University of Reading, UK.

[17] Human Development Report (2016). *Human Development for Everyone.* United Nations Development Programme. 1 UN Plaza, New York, NY 10017 USA.

[18] A. O. Njoku (2008). *Transition from higher education to employment: A case study of graduates of faculty of social sciences University of Botswana.* Educational Research and Review, Volume 3, Issue 8, pp. 262-274.

[19] A. Mberia and R. Midigo (2018). *Understanding Career Choice Dilemma in Kenya: Issues of Informed Choices and Course Availability.* Journal of Education and Practice, Vol.9, No.9, 2018.

[20] C. Mocanu, A. M. Zamfir, E. O Lungu and E. Militaru (2012) *School-To-Work Transition Of Higher Education Graduates In Four Eastern European Countries* Working Paper No. 2012/15

[21] A. Agresti (2002). *Categorical Data Analysis.* Second Edition. ISBN 0-471-36093-7. John Wiley Sons, Inc., Hoboken, New Jersey.