Statistical modeling of unemployment duration in South Africa

by

JEANETTE ZANDILE NONYANA

submitted in accordance with the requirements
for the degree of

MASTER OF SCIENCE

in the subject

STATISTICS

at the

UNIVERSITY OF SOUTH AFRICA

SUPERVISOR: PROF P M NJUHO

DECEMBER 2015

Declaration by student

I declare that **Statistical modeling of unemployment duration in South Africa** is my own work and that all the sources that I have used or quoted have been indicated and acknowledged by means of complete references.

I further declare that I have not previously submitted this work, or part of it, for examination at UNISA for another qualification or at any other higher education institution.

_____ _____

SIGNATURE DATE

Ms Jeanette Zandile Nonyana

Acknowledgements

I am thankful to the department of statistics at the University of South Africa for giving me the opportunity to further my studies with them.

My sincere gratitude goes to my supervisor Peter Njuho, who did not only supervise my work, but also made me believe I can reach for the stars.

I am thankful to the management at Statistics South Africa (Stats SA) for financial support and for instilling a culture of learning throughout the organization. I also acknowledge the usage of Stats SA data in this study and thank the management of the organisation for availing the data for research and other uses.

Most of my gratitude goes to my family and friends for not doubting my abilities and to my son S'busiso for his support and understanding.

Abstract

Unemployment in South Africa has continued to be consistently high as indicated by the various reports published by Statistics South Africa. Unemployment is a global problem where in Organisation for Economic Co-operation and Development (OECD) countries it is related to economic condition. The economic conditions are not solely responsible for the problem of unemployment in South Africa. Consistently high unemployment rates are observed irrespective of the level of economic growth, where unemployment responds marginally to changes Gross Domestic Product (GDP). To understand factors that influence unemployment in South Africa, we need to understand the dynamics of the unemployed population. This study aims at providing a statistical tool useful in improving the understanding of the labour market and enhancing of the labour market policy relevancy. Survival techniques are applied to determine duration dependence, probabilities of exiting unemployment, and the association between socio-demographic factors and unemployment duration. A labour force panel data from Statistic South Africa is used to analyse the time it takes an unemployed person to find employment. The dataset has 4.9 million people who were unemployed during the third quarter of 2013. The data is analysed by computing non-parametric and semi-parametric estimates to avoid making assumption about the functional form of the hazard. The results indicate that the hazard of finding employment is reduced as people spend more time in unemployment (negative duration dependence). People who are unemployed for less than six months have higher hazard functions. The hazards of leaving unemployment at any given duration are significantly lower for people in the following categories - females, adults, education level of lower than tertiary, single or divorced, attending school or doing other activities prior to job search and no work experience. The findings suggest an existence of association between demographics and the length of stay in unemployment; which reflect the nature of the labour market. Due to lower exit probabilities young people spent more time unemployed thus growing out of the age group which is more likely to be employed. Seasonal jobs are not convenient for pregnant women and for those with young kids at their care thus decreasing their employment probabilities. Analysis of factors that affect employment probabilities should be based on datasets which have no seasonal components. The findings suggest that the seasonal components on the labour force panel impacted on the results. According to the findings analysis of unemployment durations can be improved by analysing men and women separately. Men and women have different

challenges in the labour market, which influence the association between other demographic factors and unemployment duration.

Keywords: Unemployment duration, Panel data, Duration dependence, Non-parametric, Semi-parametric, Survival technique, Exit probability.

Contents

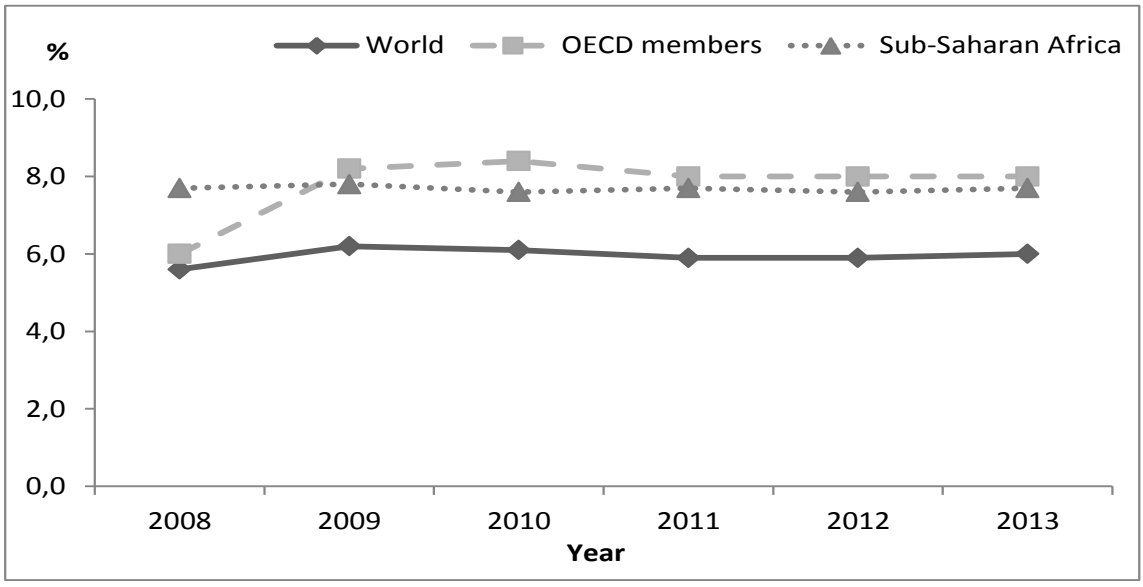
List of tables

List of figures

1. INTRODUCTION

Unemployment is a universal problem, with the world unemployment rate recorded at 6.0% in 2013. According to the World Bank data, the sub-Saharan countries and Organisation for Economic Co-operation and Development (OECD) countries recorded higher unemployment rates (7.7% and 8.0% respectively) for 2013 when compared to rest of the world (World Development Indicators, 2014).

Figure 1 compares unemployment rates for the world, OECD countries and sub-Saharan countries for the years 2008 to 2013. Prior to the economic crises, unemployment rates in the OECD countries were comparable to the world unemployment rate.



Source: World Bank – modeled ILO estimates

Figure 1: Unemployment rate by region

Figure 1 shows that the economic crisis is responsible for the high unemployment rates in OECD countries. The effect of the crisis was not fully realised by sub-Saharan countries, thus marginalising the impact on the world unemployment rate.

The highest unemployment rates among the sub-Saharan countries were recorded in South Africa (24.7%), Lesotho (24.7%) and Swaziland (22.5) in 2013 as per Statistics South Africa, (2014b) and World Development Indicators (2014). For the 2013 period, OECD recorded high unemployment rates in Greece (27.5%) and in Spain (26.1%), (OECD, 2014). The

unemployment rates recorded in these countries were more than three times the average unemployment. During the same period other OECD countries recorded unemployment rates of less than 5% (Korea: 3.1%, Norway: 3.4% and Japan: 4.1%).

Incidence of long-term unemployment[1] for OECD countries increased from 24.9% in 2008 to 35.3% in 2013 on average (OECD, 2014). Increases of more than 20 percentage points were recorded in Greece, Spain and Ireland, which translated to increases of more than double in the unemployment rates. Lancaster & Nickell (1980) modeled the rate of unemployment exit in Britain. The findings suggest that increases in the length of time spent unemployed lead to increases in total unemployment. However a recent British study on long-term unemployment suggests that the rise in British unemployment is related to a continuous decrease in manufacturing and mining jobs rather to lengthy unemployment duration (Webster, 2005).

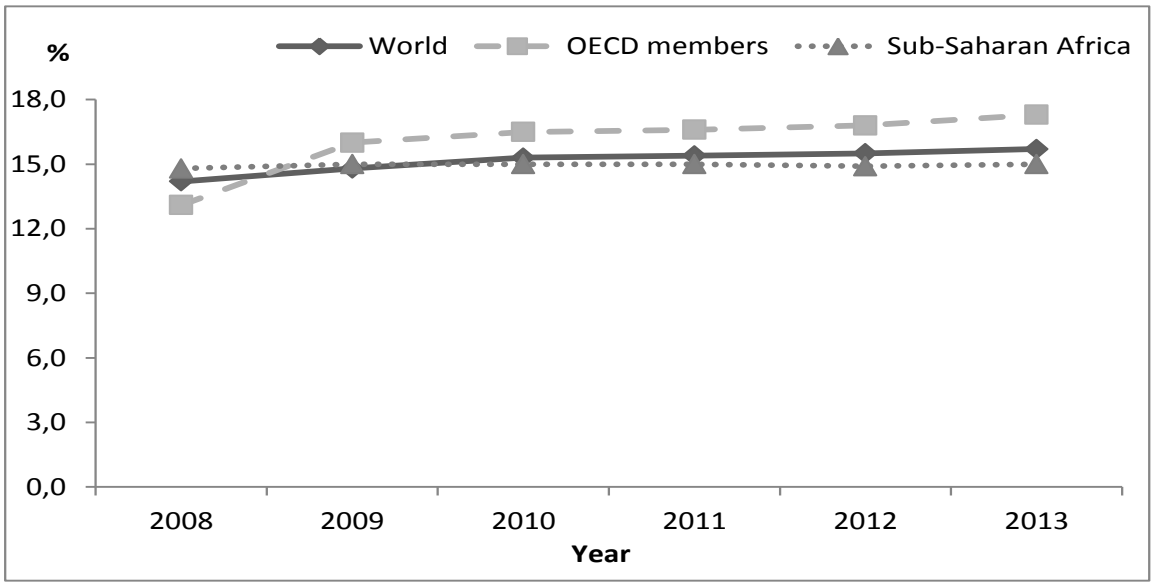
The incidence of long-term unemployment recorded in Greece (67.5%), Slovak Republic (66.6%) and Ireland (60.6%) for 2013 were comparable to that of South Africa (66.0%). However the proportion of women in long-term unemployment (49.3%) in Ireland (OECD, 2014) is lower compared to the proportion of women in long-term unemployment (69.7%) in South African (Statistics South Africa, 2014b).

Studies to determine factors associated with long-term-unemployment were conducted in the United Kingdom (Narendranathan & Stewart, 1993); Germany (Kuhlenkasper & Steinhardt, 2011) and Slovenia (Borsic & Kavkler, 2009) using survival analysis technique. The studies found that the effect of unemployment income and the chances of finding a job are reduced in long-term unemployment. Long-term unemployment is most likely to affect the female and adults (age>34 years old) categories.

Among the unemployed, young people account for the greatest share globally. The world unemployment rate for youth (15–24 years) was 15.7% in 2013. The unemployment rate for this group was 17.3% in OECD countries and 15.0% in sub-Saharan countries during the same period (World Development Indicators, 2014). Among the OECD countries, Greece (58.3%) and Spain (55.5%) recorded higher unemployment rates for youth (15–24 years) compared to South Africa (51.4%), (Statistics South Africa, 2014b).

[1] The incidence of long-term unemployment refers to the proportion of those who were unemployed for a period of a year and over to total unemployment as per OECD.Stat (aged 15 years and above).

Figure 2 compares youth unemployment rates for the world, OECD countries and sub-Saharan countries for the period 2008 to 2013. The figure shows that prior to the economic crises, youth among the OECD countries had lower unemployment rates compared to the world and the sub-Saharan countries. Youth unemployment rate in the OECD region spiked during the recession in 2009 and it remained the highest for the following periods (2009 to 2013).



Source: World Bank – modeled ILO estimates

Figure 2: Youth (15-24 years) unemployment rate by region

According to Altman (2007) youth unemployment is associated with lack of experience, job search capabilities, resources and relevant networks.

1.1. Unemployment Duration

Unemployment duration is defined as the length of time individuals spent unemployed and can be long-term or short-term unemployment duration. The definition of long-term and short-term unemployment duration differs by country. In South Africa, an individual is said to be in long-term unemployment if unemployed for a period of one year and longer, otherwise considered to be in short-term unemployment.

Unemployment duration is a measurement tool used to evaluate the labour market conditions and to determine the welfare of the labour market. A healthy labour market is characterised by high employment and low unemployment, where the later has small incidence of long-term unemployment (South West Observatory, 2008).

The length of time spent in unemployment explains the capability of a labour market to absorb labour (Statistics South Africa, 2015). The release showed that higher incidences of long-term unemployment are associated with lower absorption rates. The design of effective policy welfare depends on identifying causes of variation between unemployed persons in unemployment duration (Lancaster, 1979). Policy makers in South Africa have proposed new labour market policy in the form of employment incentives and subsidies to stimulate labour demand; and this policy is more effective among those who are in short-term unemployment, as it is difficult for individuals in long-term unemployment to be integrated in the labour market (National Treasury, 2011).

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

1.2. Problem statement

Unemployment in South Africa is a salient problem and one of the most critical socio-political problems faced by the South African government. About 14% of the working age population (that is, persons aged 15–64 years) is affected by unemployment (Statistics South Africa, 2014b). Unemployment rate ranged from 21.5% in the fourth quarter of 2008 to 24.1% in the fourth quarter of 2013 (Statistics South Africa, 2009); (Statistics South Africa, 2014a). A large proportion of the unemployed had been on that state for a period of a year and longer (according to the results released in the Labour market dynamics in South Africa, 2013 report).

Statistics South Africa's Labour market dynamics in South Africa is an annual report that summarises the previous year's quarterly reports and analyses the time series data with the previous years since the inception of the Quarterly Labour Force Survey (QLFS) in 2008. The 2013 report showed that long-term unemployment had been increasing over the years (from 59.3% in 2008 to 66.0% in 2013). The most disadvantaged gender is women where those in long-term unemployment ranged from 63.4% in 2008 to 69.7% in 2013.

In South Africa, unemployment is exacerbated by the economic factor. That is there are simply not enough jobs available for the number of people available, thus supply exceeds

demand. The South African labour market is characterised with chronic skills mismatch, where the available labour lacks the skills demanded by the market (Altman, 2007). Skills mismatch in South Africa contributes to the imbalance between the supply and demand of labour. Employment growth has been concentrated in highly skilled sectors, thus leading to a skilled wage premium.

Publications (Statistical releases) by Statistics South Africa (Stats SA) show that a large section of the South African labour force is unskilled with many people who have never worked before. Lack of skills affect individual's employment probabilities; according to Dias & Posel (2007) unemployment is reduced by increased levels of education.

Part of the South African unemployment is structural; that is the supply of labour does not match the quality of labour demanded. Zimmer (2012) used a Beveridge curve to show that unemployment rate and job vacancy rate are negatively related. The length of time spent in unemployment is not always related to scarcity of jobs but rather to other factors (for example, lack of skills).

While unemployment decreases with improved levels of education, the South African graduates' unemployment increased from a rate of 7.6% in 2008 to 9.9% in 2013. Qualification mismatch contributes to graduates' unemployment (Altman, 2007). Zimmer (2012) proposed a macro-economic style approach (skill mismatch index - SMI) to measure the size of qualification mismatch.

$$\mathrm{SMI} = \sum_{j=1}^{n} \left(S_{ijt} - M_{ijt}\right)^2$$

$j =$ educational category
$n =$ number of education level
$S_{ijt} =$ percentage of WAP with education level j at time t and province i
$M_{ijt} =$ percentage of employees with education level j at time t in province i .

The index provides a basis for addressing skills shortage and graduate unemployment. According to Daniels (2007) enterprise training and the education system contributed to skills shortage in South Africa. The quality and quantity of educations received by most South Africans impacted on skill shortage (Erasmus & Breier, 2009).

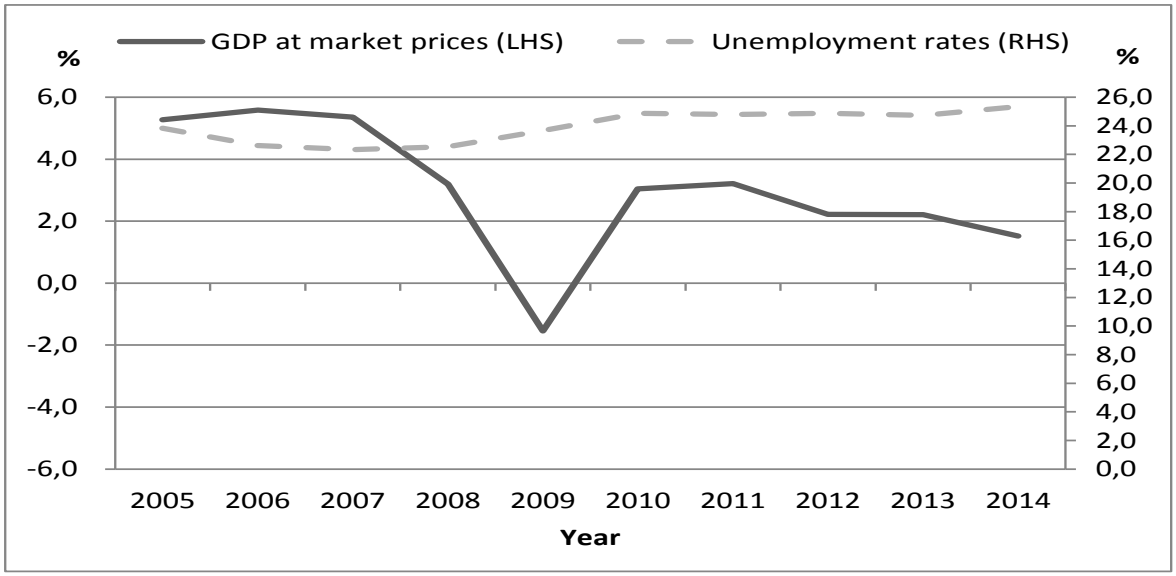
Economic conditions, such as slow growth have also contributed to the lower absorption rates of the low skilled, suggesting that even if all the unemployed are (re)skilled, not all of them will find jobs. As unemployment increases through new entrants and job losses, the individuals in long-term unemployment are further disadvantaged and their chances of finding employment decline.

South Africa is faced with a problem of employment creation and curbing unemployment. Though economic conditions such as slow growth influence employment negatively, the challenge is in understanding the labour market. The QLFS releases are meant to enhance knowledge about the labour market; improve understanding of the labour market and be of policy relevance. Factors like education, age and gender, have emerged as having serious impacts on unemployment; and duration of unemployment is another factor which influences unemployment.

Few studies appear to focus on the length of time that individuals are unemployed as an important factor of the likelihood of exiting from the state of active unemployment. Studies that examine the impact of structural factors in the labour market have focused on membership of the workforce in a trade union, access to social security benefits, employment security, mismatch between job seekers and vacancies etc. (Morgan & Mourougane, 2001). This study intends to provide insight on how the duration of unemployment (as a structural factor) influences unemployment persistence.

Several factors are thought to influence the probability of entering employment; economic factors affect both the demand and the supply side of labour. Higher levels of gross domestic product (GDP) stimulate production which in turn should generate employment (Seyfried, 2005). In South Africa higher unemployment rates are recorded even with the higher GDP level.

Figure 3 plots the South African GDP at market prices and unemployment rates from the year 2005 to 2014. The GDP figures are plotted on the left hand side, while unemployment rate is plotted on the right hand side.



Source: Statistics South Africa.

Note: LHS – read from the left hand side and RHS – read from the right hand side. GDP is expressed in percentages to measure the growth rate from one period to another.

Figure 3: Gross domestic product and unemployment rate

According to Figure 3 the South African economy was performing better for the years 2005 to 2007, with GDP higher than 5%. Between 2008 and 2010 the growth rate declined to a recession level. Unemployment rates on the other hand have been consistently higher (more than 22%) irrespective of the economic growth. The response of unemployment rate to changes in GDP is minimal. This suggests the need for understanding other factors which influence unemployment. Factors such as industrial adjustment, skills mismatches, insufficient labour demand, reservation wages and the underlying dynamics of the unemployed should also be considered when addressing unemployment.

Ranchhod (2009) analysed data from the National Income Dynamic Study (NIDS) and the results suggests that unemployment in South Africa is composed of voluntary (18%) and involuntary (61%) unemployment. The same study showed that there is an association between unemployment and poverty, which impacts on nutrition, schooling, inter-generational persistence and inequality. A study on consequences of long-term unemployment found that periods of long-term unemployment have negative impact on the social welfare of an individual (Nichols, et al., 2013).

In other countries, studies on unemployment duration commenced long ago (e.g. Britain – Nickell, 1979 and United States – Katz, 1986), and in South Africa there is limited literature on this subject. This reason necessitates studies on unemployment duration in South Africa.

The South African literature on unemployment duration has not been extensively researched. It is rather skewed towards determination of factors that affect exits probabilities. Brick & Mlatsheni (2008) examined the degree of duration dependence in the Cape Town labour market; the data used for the study was not representative of the Cape Town metropolitan area. This study examines the degree of duration dependence, probabilities of leaving unemployment and factors associated with exit probabilities in the South African context.

1.3. Purpose

The study seeks to model unemployment duration to improve an understanding about the dynamics of the unemployed. The aim is at providing a statistical tool useful in improving the understanding of the labour market dynamics and enhancing of the labour market policy relevancy. A statistical tool is essential for showing empirical evidence on the causes or factors associated with long-term unemployment.

The length of time that individuals are unemployed has considerable policy significance. Labour market policy which could impact positively on the length of unemployment relies in understanding the characteristics of the unemployed. The statistical tool differentiates between people who are more likely to exit unemployment and those who are more likely to remain unemployed.

People in long- and short-term unemployment have different employment probabilities. Those in long-term unemployment experience skills depreciation, which decreases their employability. Labour market policies aimed at reducing unemployment should attend to the varying needs of the unemployed population. The study explores factors which affect the chances of finding employment and they should be considered when designing labour market policies.

1.4. Objectives

The objectives are:

- To investigate the dependency of the prospect of getting a job on time spent in unemployment (duration dependence).
- To determine the probabilities of exiting unemployment in different unemployment durations.

- To determine the association of socio-demographic factors and unemployment duration.
- To determine the extent to which socio-demographics influence the length of stay in unemployment.

The study seeks to gain the perception of the following hypotheses:

1. The length of time spent in unemployment is influenced by socio-demographic factors.
2. The prospect of getting a job is dependent on the time spent in unemployment and
3. The probabilities of getting a job differ with duration of unemployment.

This study is divided into eight sections, section 1 provides an introduction, section 2 is the literature review, section 3 describes the research methodology, section 4 explores the data through non-parametric models, section 5 applies Markov chain to predict transition probabilities, section 6 analyses the data using a semi-parametric model, section 7 discusses the results and section 8 concludes the analysis.

Additional information about the analysis is presented in Appendix A. That is variable description, list of collapsed stratums and sampling units, and the STATA code used for analysis (the analysis was performed on STATA version 13). The questionnaire that was used to collect the data is attached in Appendix B.

2. LITERATURE REVIEW

Studies on unemployment duration are concerned about factors that affect an individual's probability of leaving unemployment. Nickell (1979b) showed that the probability of leaving unemployment depends on personal characteristics, family composition, local labour market demand and income variables. These factors have important implications for policy design and assistance.

The impact of gender on unemployment duration is studied by Tansel & Taşçi (2010) and Mussida (2007) - their findings suggest that women are more prone to longer duration of unemployment. Tansel & Taşçi (2010) found that unemployment exit probabilities are the lowest among married women. According to Mussida (2007) personal characteristics such as marital status and age have a significant impact on women unemployment probabilities.

Landmesser (2011); Tansel & Taşçi (2010); Borsic & Kavkler (2009) and Babucea & Danacica (2007) studied the impact of education on unemployment duration. According to Landmesser (2011) vocational training is important in exiting unemployment. Tansel & Taşçi (2010) found that vocational high school graduates have higher unemployment exit probabilities compared to high school graduates. According to Borsic & Kavkler (2009) persons with professional college degrees or bachelor's degrees are better off than unemployed persons with a master's degree. Babucea & Danacica (2007) suggest a negative relationship between unemployment and education.

Studies on the effect of income variables such as unemployment insurance found that unemployment insurance influences the job search behaviour of the unemployed (Katz, 1986); (Heath & Swann, 1999). According to Nickell (1979a) the impact of income replacement ratio (ratio of income to benefits) was lower for those in long-term unemployment, while high replacement ratio was associated with longer duration of unemployment (Lancaster, 1979). Findings by Caliendo, et al. (2009) suggest that unemployment insurance benefits allow individuals to look for better quality jobs by increasing unemployment duration.

The impact of family composition was outlined in a German study by Kuhlenkasper & Steinhardt (2011). The results show that the likelihood of returning into employment for women is reduced by the presence of young children and older relatives in the household. According to Imbens & Lynch (2006) the number of children living at home has a negative

effect on the chances of women finding a job. Tansel & Taşçi (2010) found that married men have higher opportunity cost of unemployment and thus search more intensively.

South African literature also showed that personal characteristics and labour market demand are the most important determinants of unemployment duration. Potgieter (2012) showed that individuals' labour market conditions are determined by geographical location, race and educational levels. Coulson (2009) concluded that the probability of leaving unemployment is higher among those with job experience; the same conclusion was drawn by Tansel & Taşçi (2010) in Turkey. According to Brick & Mlatsheni (2008) women with children at their care are more prone to long-term unemployment. Dias & Posel (2007) investigated the effect of education on unemployment in South Africa using a Probit regression.

When studying unemployment duration it is important to differentiate exit to employment from other exits. An unemployed person can leave unemployment for employment or for inactivity. Exit to employment comprises of exit to a new job and a recall. A competing risk model is suitable for this type of studies. Narendranathan & Stewart (1993) showed that modeling unemployment duration without distinguishing the nature of the exit biased the estimated results. Their results show that a single-risk model underestimates effects of income on the probability of finding a job.

Carling & Jacobson (1995) applied a competing risk model to Sweden unemployment duration data to conclude that there is a correlation between exit to employment and attrition. Jensen & Svarer (2003) compared results from multiple phase duration model and competing risk model to show that short-term unemployment in Denmark is associated with temporary layoffs. Competing risk estimates by Mussida (2007) suggest that while males have higher employment probabilities, their female counterparts leave the labour force due to discouragement.

The length of stay in unemployment is influenced by both observed and unobserved variables (individual effect). Uncontrolled individual effects lead to incorrect estimates (Lancaster & Nickell, 1980). The Cox proportional hazards model is capable of addressing heterogeneity due to unobserved effects. The model allows the data to determine the functional form of the baseline hazard. Non-parametric models on the other hand are designed to deal with unobserved heterogeneity and are also good methods to understand basics and produce descriptive results (Mills, 2011).

The Weibull (parametric model) is known for its advantage of considering that the population may not be homogeneous, the model estimates parameters by finding values that maximise the likelihood function (Haughton & Haughton, 2011).

Fan & Li (2004) recommends the use of semi-parametric models for analysing longitudinal data, the structure of longitudinal data poses challenges to parametric inferences. Multivariate regression techniques cannot handle the unbalanced nature of longitudinal data. Witchert & Wilke (2008) recommend simple non-parametric models for administrative data, because administrative data come with limitations which include various forms of censoring.

Studies on unemployment duration and the probabilities of leaving unemployment apply different types of datasets. Nickell (1979b) used cross section data, Narendranathan & Stewart (1993) used longitudinal data, Babucea & Danacica (2007) used administrative data and Mussida (2007) used a rotating panel data. Panel data has an advantage of identifying dynamic behaviour and are also able to control for omitted variables.

Heeringa, et al. (2010) recommends the use of sample design feature when analysing complex surveys data, because ignoring the design features can introduce bias in the estimates. Binder (1992) applied the Cox proportional hazards model to discuss the implications of the survey design for large-scale studies. Boudreau & Lawless (2006) used the Cox proportional hazards model to account for the complex survey design.

3. METHODOLOGY

3.1. Data source

The study utilises secondary data from Statistics South Africa (Statistics South Africa, 2014c). Stats SA is a government department in South Africa responsible for the collection and publication of official data. Data collected by Stats SA is cleaned and weighted before they are posted for public usage (Statistics South Africa, 2014c). Stats SA is solely responsible for ethical considerations.

The data used in this study resulted from a Quarterly Labour Force Survey (QLFS), which is a household-based survey conducted on a quarterly basis. It collects data on labour market activities of individuals aged 15 years and older who live in South Africa.

The survey is based on a sample of 30 000 dwelling units (DUs) and 3 080 primary sampling units (PSUs), which covers non-institutional population in the country which excludes workers' hostels. However those who live in private dwelling units within institutions are also enumerated. The sample design follows a complex survey design method – two stages of stratification, sampling of PSUs in the first stage, and sampling of DUs with systematic sampling in the second stage (Statistics South Africa, 2008).

The survey utilises a face-to-face method in collecting data from households. The sampled dwelling units are visited by trained data collectors to collect information using the QLFS questionnaire (Appendix B). Collection happens during the two middle weeks of each month. The first and the last weeks are reserved for survey logistics such as publicity, interview setup and listing maintenance.

Questionnaires are captured using a scanning system, and updates are done manually. The Labour Statistics Unit in Stats SA uses an automated editing and imputation module to ensure that the collected data is clean and complete (Statistics South Africa, 2008). This module has three basic steps - record acceptance; edit, imputation and clean up; deriving variables and preparation for weighting. The collected data is calibrated to the known population by applying survey weights. Stats SA uses regression estimation to calculate survey weights.

The main purpose of the QLFS is to estimate labour market indicators such as the working age population (those aged within 15–64 years), labour force, the employed population, unemployed population and the not economically active population; which are used to

calculate labour market rates. These statistics are published on a quarterly basis at national and provincial levels.

In illustrating the applicability of the technique the study utilises panel data created from the QLFS. The panel data spanned two cross-sectional datasets - the third quarter of 2013 (Q3: 2013) and the fourth quarter of 2013 (Q4: 2013). The data is available for public use and can be accessed from the website http://interactive.statssa.gov.za:8282/webview/ (the dataset is labelled Quarterly Labour Force Survey Panel data - Q3: 2013_Q4: 2013). The technique can also be applied successfully in panels that spanned other quarters.

For the purpose of this study a sub-data is created from the QLFS Panel datasets by keeping only the individuals who were unemployed during Q3: 2013 (the first wave of the panel). An unemployed person is one who is aged between 15 and 64 years and meets the following criteria:

a) Was not employed in the week prior to the survey interview (called reference week).
b) Actively looked for work or tried to start a business in the four weeks preceding the survey interview.
c) Would have been able to start work or would have started a business in the reference week.

The unemployment definition is according to International standards adopted by the 13th International Conference of Labour Statistics (ICLS) (International Labour Organisation, 2003).

The variable of interest is unemployment duration and it is derived using Question 3.6 “for how long have you been without work and trying to find a job or start a business?” from the QLFS questionnaire (Appendix B). It should be noted that the question asks three questions in one to ensure that only people who met the definition of unemployment are captured.

Unemployment duration is a categorical variable measured in months. The considered categories for unemployment duration are: less than three months, three months to less than six months, six months to less than nine months, nine months to less than one year, one year to less than three years, three years to less than five years, more than five years and don’t know. These categories define the different durations.

Those respondents who have been unemployed for less than a year are classified as short-term unemployed, while those who were unemployed for one year and more are classified as long-term unemployed.

Other questions that are used in the analyses include:

1) What have you done to search for work or to start a business/ (job search method)?
2) What was the main activity before you started looking for work?
3) Have you ever worked before?
4) How long was it since you last worked?
5) How do you support yourself?

3.1.1. Panel design

The QLFS sample has features of a longitudinal survey. The sampled 3 080 of PSUs is divided into four rotation groups (panels). In each quarter, a quarter of the sampled dwellings (770 dwelling units) rotates out of the sample and is replaced by new dwellings from the same PSU or next the PSU on the list; this is done to reduce measurement error due to respondent exhaustion and to produce more robust estimates with lower variance. Sampled dwelling units remain in the sample for four consecutive quarters, making it possible to match individuals over the period in which they are present in the sample (Statistics South Africa, 2008).

A panel of two consecutive quarters represent 75% of each cross section. Due to attrition a lesser proportion is matched. The matching process focus on those aged 15 years and above, and the calibrated panel weights are calculated only for those aged 15 to 64 years. Adjustments are made to the QLFS weights to account for the non-overlapping panel and the adjusted panel weights were calibrated to the published aggregates in the initial wave of the panel (Statistics South Africa, 2014c).

The non-overlapping panel represents the 25% sample which is rotated out and the replacing rotation group in the subsequent wave. Weights are applied to calibrate the panel totals to the corresponding cross sectional totals based on age (youth and adults), gender (men and women) and labour market status (employment, unemployment and not economically active).

3.1.2. Quality of the panel

The QLFS panel data is created by linking individuals who responded in two subsequent cross sectional surveys (QLFS). The quality of the QLFS panel is tested by Stats SA and a report on quality diagnostics for the QLFS panel is produced. In testing the QLFS panel the analyst applied the chi square test for independence, focusing on the attrition rate and its association with the respondent's demographic and labour market status. The report showed that the panel is fit for the desired use, which is to track the movement of individuals between labour market status for subsequent quarters.

The QLFS panel data is limiting to analysing for understanding of movements between various labour market categories. Since the data is linked to only those individuals who responded in all two quarters, it is impossible to analyse other aspects of longitudinal data using this data.

The initial dataset had 363 stratums, where 34 of them were having single sampling unit. These strata are collapsed into other stratums. An additional 30 sampling units had missing sampling unit number (PSU numbers). These units were collapsed into other sampling units. The final dataset used in this analysis had 299 stratums.

3.2. Statistical technique

The choice of a statistical technique is dependent on the datasets to be analysed. According to Danacica & Babucea (2010) duration data requires a different statistical analysis compared to quantitative data due to their particularities. Duration data are not normally distributed and often contain incomplete observation or censored subjects.

The QLFS panel data violates the normality assumption which is a common assumption in most statistical models. The probability of leaving unemployment to employment has an instantaneous risk of occurring hence is unreasonable to assume normality (Cleves, et al., 2004). The data spanned two quarters of a year (see panel design in section 3.1.1), which might not be enough to observe most of the subject to fail (get employment). The National Income Dynamic Study (NIDS) shows that even with enough time the rate of failure is still minimal.

NIDS is a panel study which is repeated every two years. The 2008_2010 NIDS panel indicates that 32% of those who were unemployed in 2008 were in employment in 2010

(Cichello, et al., 2012). This shows that South Africans are not only troubled by high unemployment levels but also the difficulty of leaving the state (of unemployment). Ciuca & Matei (2010) refer to labour markets of such condition as damaging.

Censoring is one aspect of duration data which poses difficulty when using traditional statistical model to analyse duration data. Standard methods of data analyses such as t-test and linear or logistic regression cannot be applied to duration data because they do not account for censoring (Rao & Schoenfeld, 2007). According to Lu & Shen (2014) linear regression cannot deal with the influence from censored data correctly and logistic regression does not consider the differences in the timing of event occurrence. The existence of censored subjects complicates estimation of a likelihood function (Fox, 2002). Censoring also leads to biased hazard functions (Kiefer, 1988), and exclusion of censored subjects reduces the sample size thus complicating event analysis (Jakoet, 2007).

Censoring is divided into right and left censoring. A right censored person is one who left their household before the subsequent visit or one who indicated that they are jobless during the follow-up interview. Left censoring is when the event of interest occurred before the study commenced.

Figure 4 illustrates right censoring, where ten persons are randomly selected from the panel data (Q3: 2013_Q4: 2013). Each person is given a number (person number) for identification. All the selected persons were unemployed during the initial wave (Q3: 2013) and their labour market status are observed during the second wave (Q4: 2013). For those whom the event did not occur (that is were still unemployed) during the second wave are said to be right censored.

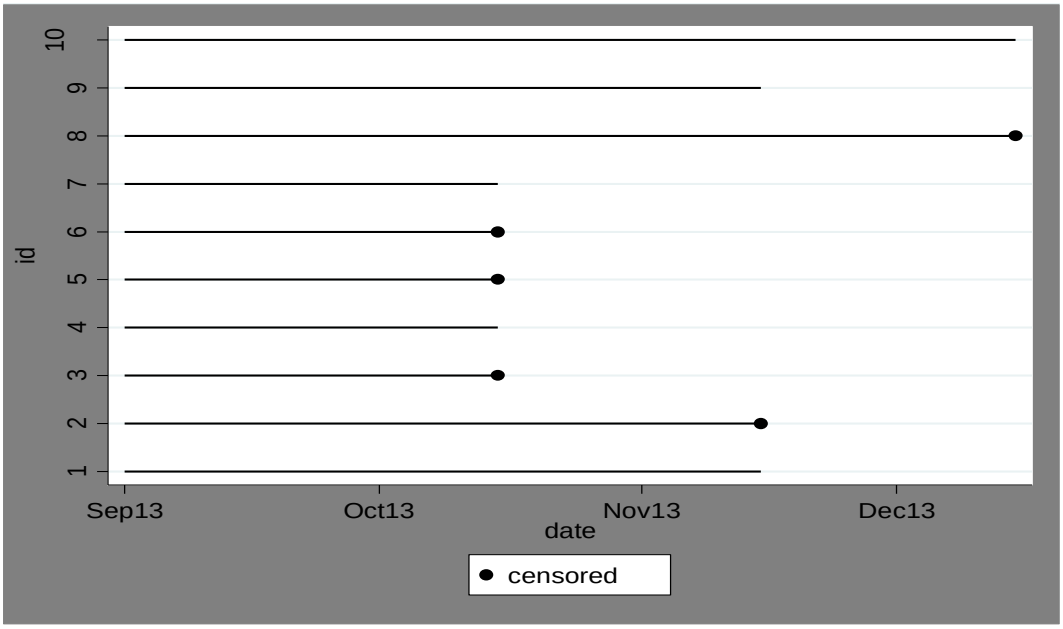


Figure 4: Right censoring

The horizontal axes of Figure 4 present the observation period with survival time of three months (October 2013, November 2013 and December 2013). The observed persons are shown on the vertical axes of Figure 4.

Figure 4 shows that among the five persons who had follow-up interviews in October three persons were still unemployed (right censored) - person 3, person 5 and person 6. Among the three persons who had follow-up interviews in November, two persons found employment and one person was right censored (person 2). Person 8 and person 10 had follow-up interviews in December, person 10 found employment, while person 8 was right censored.

In the QLFS, dwellings are the sampling units, and households are the units of observation. If a household moves out of the dwelling unit the new household will be enumerated (as stated in the Guide to Quarterly labour force survey, 2008 report). Individuals who leave their households are not tracked back to the sample. Enumeration occurs only to those who stayed in their households for at least four nights on average per week during the last four weeks.

Individuals who left their households before the subsequent interview and those who were still jobless during Q4: 2013 interviews are said to be right censored. Right censoring in this analysis refers to people who were did not get jobs during the second wave and were interviewed in both two quarters (Q3: 2013 and Q4: 2013).

This study applied survival analysis to a duration data of 4.9 million unemployed South African to analyse the time it takes an unemployed person to find employment. Survival analysis estimates functions of the elapsed time between the entry of subject in a study until an even of interest occurs (failure time) or until study termination (right censoring time). The theory of survival analysis is based on two main functions that will be defined later, namely the survival function and the hazard function.

Survival analysis is composed of non-parametric, semi-parametric and parametric models. The selection of a survival model depends on the assumptions to be made about the functional form of the hazard (Cleves, et al., 2004). Non-parametric and the semi-parametric models computes estimates using the observed data and do not make any assumption about the distribution of failure times or the baseline hazard. The parametric model requires parameterization of the hazard.

The analysis reported in this dissertation is not making any assumptions about the functional form of the hazard and the distributional function of the data is not known. In addressing the objectives, survival functions and hazard functions are estimated by fitting non-parametric models (Kaplan Meier estimator, Nelson-Aalen methods) and semi-parametric models (Cox regression). The Kaplan Meier and Nelson-Aalen methods compute non-parametric survival and hazard functions.

Statistical techniques such as the Kaplan Meier non-parametric model, Cox proportional hazards semi-parametric model and Weibull hazard parametric models (survival techniques) are widely used in duration analysis studies. According to Kiefer (1988) the application of hazard function models in duration data is proposed to address problems (such as censoring) associated with duration data. Survival techniques were primarily developed in the medical and biological sciences where they were used to study death as the event. The usage of survival techniques in economics and social science became popular in the 1970s (Danacica & Babucea, 2010).

Brick & Mlatsheni (2008) and Jakoet (2007) used hazard functions to determine duration dependence. Tansel & Taşçi (2010) and Mussida (2007) applied hazard functions to determine factors which influence unemployment duration. Ciuca & Matei (2010) used survival curves to determine factors associated with unemployment duration and the Cox model to determine the extent of the relationship.

Non-parametric estimates are important for univariate and preliminary analysis. Heeringa, et al. (2010) suggests that an exploratory analysis be conducted as a first step in model building. This is to get an idea of the covariates that have a significant relationship with the response variable. Lu and Shen (2014) suggest an inclusion of significant covariates when building a model, since covariates which are not significantly associated with survival time will not contribute much to the model. The effect of explanatory variable on hazard rate and on the length of stay in unemployment is computed by Cox regression.

4. ANALYSIS

4.1. Distribution of the unemployed population

A total of 4.9 million individuals were unemployed in the third quarter of 2013. Table 1 indicate the different demographic factors of the unemployed, while table 2 shows how they are engaing in economic ectivities.

Table 1: Unemployment by demographics

	Thousand	**Per cent**
Total unemployed	4 880	
Gender		
Men	2 503	51.29
Women	2 377	48.71
Population group		
Black/African	4 141	84.86
Coloured	545	11.17
Indian/Asian	42	0.86
White	152	3.12
Age group		
Youth	3 267	66.95
Adults	1 613	33.05
Marital status		
Staying with a partner	1 282	26.27
Single	3 598	73.73
Education		
Tertiary	339	6.94
Matric	1 676	34.35
Below matric	2 865	58.71
Province		
Western Cape	641	13.13
Eastern Cape	565	11.58
Northern Cape	123	2.51
Free State	374	7.67
KwaZulu-Natal	666	13.65
North West	306	6.28
Gauteng	1 550	31.76
Mpumalanga	415	8.51
Limpopo	240	4.92

A total of 2.5 million men were unemployed in the third quarter of 2013 compared to 2.3 million women (Table 1). According to the table the unemployed population was concentared among the black African population (84,9%); the youth (67.0%); those who have never married or divorced or widowed (73.7%) and those with education with education levels of below matric. The unemployed population is distributed in all the province, and Gauteng has the largest share of the unemployed at 31.8%.

Table 2: Unemployment by economic activity

	Thousand	**Per cent**
Experience		
Yes	2 964	60.74
No	1 916	39.26
Activity prior to job search		
Working	2 493	51.08
Going to school	915	18.75
Other	1 472	30.17
Type of support[2]		
Household member	3 711	76.04
Non-household member	1 084	22.20
Grants	737	15.10
Job search methods[3]		
Enquiring at work places	2 804	57.46
Placed/answered job	1 870	38.32
Searched the internet	1 361	27.89
Sought assistance from relatives	2 208	45.25

Table 2 shows that a large proportion (60.7) of the unemployed has some work experience. A lesser proportion of the unemployed indicated that they were going to school prior to engaging in job search activities, while a larger proportion indicated that they were working. The most available type of support for the unemployed is by household members and the mostly used job search methods are enquiring at work places and soughting assistance from relatives.

[2] The type of supports are mutually inclusive, hence the percentages do not add to hundred
[3] Job search methods are mutually inclusive, hence the percentages do not add to hundred

4.2. Exploratory analysis based on non-parametric models

Non-parametric models do not make assumptions about how the exit probabilities changes over time. In this section non-parametric models are applied to estimate the distribution of survival time, hazard function, compare survival curves from different groups and test association of survival time with other covariates. This analysis will help in understanding the data and in identifying covariates that have a significant relationship with the response variable.

The Kaplan Meier and Nelson-Aalen methods are applied to compute non-parametric survival and hazard functions for univariate analysis. The Kaplan Meier estimator is discussed in section 4.2.1 while the Nelson-Aalen estimator is discussed in section 4.2.2.

4.2.1. The Kaplan Meier estimator

The Kaplan Meier estimator is a non-parametric estimator of a survival function (Kaplan & Meier, 1958). Kaplan Meier method is useful for preliminary analysis; it is a descriptive method that evaluates one variable at a time (Danacica & Babucea, 2010). The method is well-known for its capability of handling data with incomplete observation or censored objects. According to Jakoet (2007) existence of right censored subjects complicates event analyses.

The analysis of length of time that people spent in unemployment is affected by right censored individuals, the individuals with incomplete observations that is individuals still unemployed at the end of the observation time. In general, individuals or objects are said to be censored if they are lost to follow up for reasons unrelated with the study or have not observed the vent of interest at the end of the study (right censoring) or observed the event before on the onset of the study (left censoring).

Kaplan Meier estimators are considered the best techniques for computing survival function in the presence of censored objects (Goel, et al., 2010). Individuals who are censored during a given time are counted among those who survived, while they are considered as at risk for the next period (Kaplan & Meier, 1958). The Kaplan Meier method sort observation from smallest duration to largest duration and this allows for estimation of survival functions without making any assumption about the form of the function.

Suppose T is the time to employment with probability density function $f(t)$ and cumulative distribution function

$$F(t) = \Pr(T \leq t). \quad (1)$$

The survival function is the reverse cumulative distribution function given by

$$S(t) = \Pr(T > t) = 1 - F(t). \quad (2)$$

Now, consider r unemployment times ordered in ascending order as follows

$t_{(1)} < t_{(2)} < \cdots < t_{(r)}$. The Kaplan Meier estimate of the survival function at time t with

$t_{(1)} \leq t < t_{(k+1)}$ is given by the following product of conditional survival probabilities

$$\hat{S}(t) = \prod_{j=1}^{k}\left(1 - \frac{d_j}{n_j}\right) = \prod_{j=1}^{k}\left(\frac{n_j - d_j}{n_j}\right) \text{ for } k = 1,2,\ldots,r \quad (3)$$

and $\hat{S}(t) = 1$ for $t < t_1$ (in particular at time origin $t = 0$)

where

$n_j = \text{number of unemployed individuals (individuals at risk) at time } t_j$

and

$d_j = \text{number of employed individual (number of events) at time } t_j.$

The conditional survival probability is defined as:

$$P\left(T > t_j | T > t_{j-1}\right) = \frac{n_j - d_j}{n_j} = p_j \quad (4)$$

and, $P(T > T_0 = 0) = 1$.

The Kaplan Meier method estimates survival functions and conditional survival probabilities for different unemployment durations. The survival functions are plotted (survival curves) to determine the potential influence of explanatory variables and to test the difference between levels of predictor variables. Proportionality may be an indication that levels of predictor variables are significantly different.

The Kaplan Meier estimator of the survival function allows for preliminary analysis of survival data, and in this context to explore univariate association between levels of covariates (factors) and unemployment duration.

4.2.1.1. Distribution of survival time in unemployment

During the first wave (Q3: 2013) there were 4.9 million unemployed individuals, with unemployment duration as listed in Table 3. The follow-up interviews were in the fourth quarter of 2013 and 638 000 individuals were employed during that period. Table 3 shows the number of those who found employment per unemployment duration. Among those who found employment 2 000 did not remember their unemployment duration; hence they are not shown in the table. The table indicate that no one found employment during the period 'less than 3 months', this is because the panel started during that period (time origin) where the unemployed were being identified.

The people observed in each unemployment duration are mutually exclusive, that is different people are observed for different unemployment duration. Column c_j indicates the number of those who remained unemployed (censored) given that they were only observed in that particular unemployment duration.

Column (p_j) and column $\hat{S}(t_j)$ indicates the conditional survival probabilities and estimated survival functions respectively. The conditional survival probabilities are estimated per duration, on condition that people were unemployed at the beginning of their unemployment duration.

Table 3: Survival function by duration of unemployment

Unemployment duration (t_j)	$\mathbf{n_j}$	$\mathbf{d_j}$	$\mathbf{c_j}$	$\mathbf{p_j}$	$\hat{\mathbf{S}}(\mathbf{t_j})$
	Thousand			**Proportion**	
Less than 3 months	4 880	0	483	1.0000	1.0000
3 months to less than 6 months	4 397	148	269	0.9663	0.9663
6 months to less than 9 months	3 980	71	310	0.9822	0.9491
9 months to less than 1 year	3 600	44	297	0.9878	0.9375
1 year to less than 3 years	3 258	179	1 093	0.9450	0.8859
3 years to less than 5 years	1 986	65	595	0.9672	0.8569
Over 5 years	1 325	129	1 179	0.9027	0.7735

For an unemployment duration of between three and six months, 148 000 individuals found employment. The number of those who found employment decreased to 71 000 and 44 000 as unemployment duration increased to nine months and one year respectively. During the second wave (Q4: 2013), 4.2 million (87%) individuals were still without jobs, that is they either remained actively engaged in job search activities or they became inactive. Table 3

show that 67.8% of the censored individuals were unemployed for a period of one year and over (long-term unemployment).

According to the conditional survival probability is an increasing function of time. The probability is estimated at (4 397-148)/4 397 =0.9663 for the unemployment duration '3 months to less than 6 months'. When unemployment duration increased to five years and over the estimate of the probability decreased to (1 325-129)/1 325=0.9027.

The survival function is estimated as product of the conditional survival probability for surviving beyond t_j (see equation 3). The survival function decreases as unemployment duration increases. This is one characteristic of a survival function (Collett, 2003).

Figure 5a below plotted the survival function to show a graphical presentation for the different unemployment durations. Figure 5b graphically shows how the survival functions differ between those with experience and those without experience.

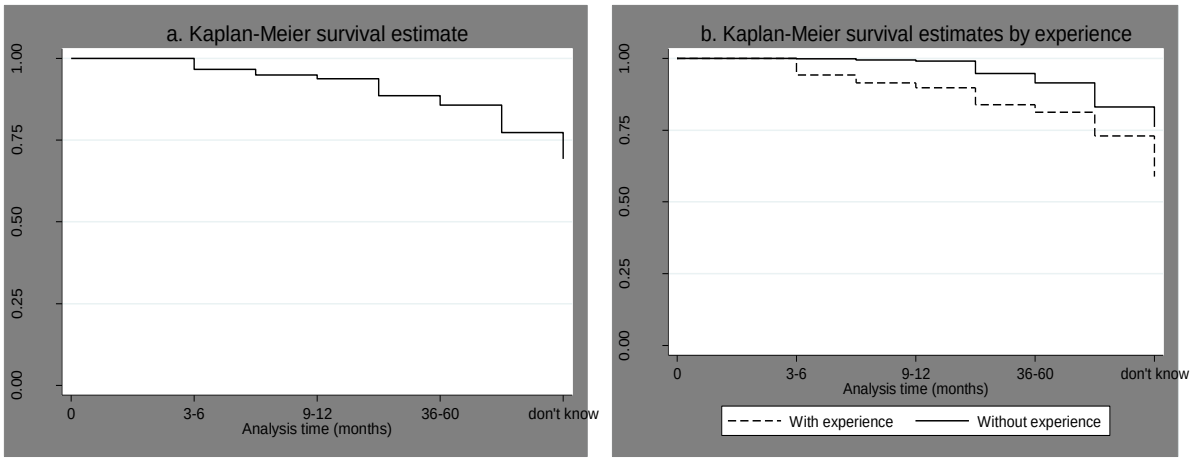


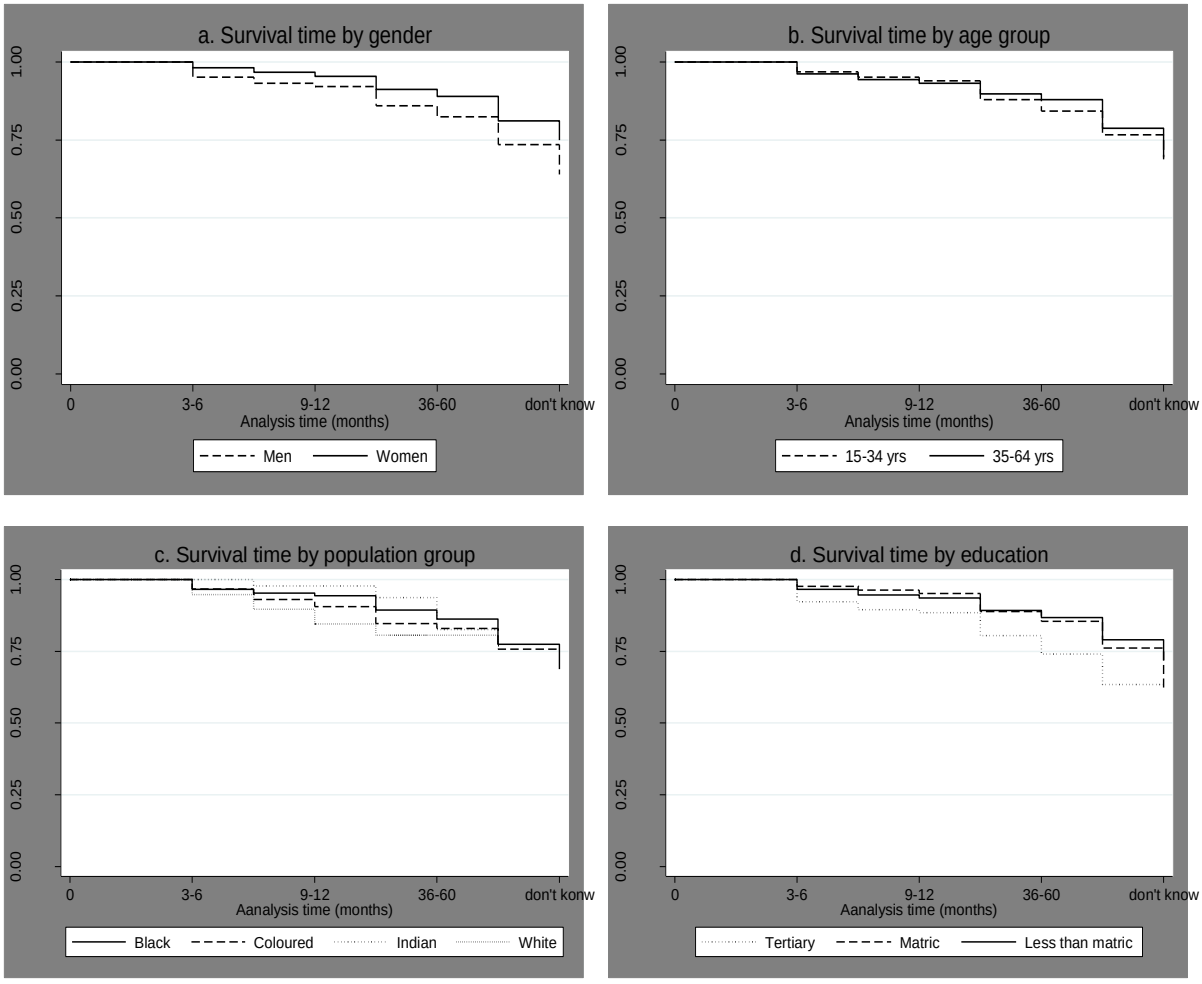
Figure 5: Kaplan Meier survival estimates

The survival plot in Figure 5a shows that the rate of decrease in the survival functions is low between the periods of three and twelve months. This indicates that unemployment exit probabilities were low during those periods, which might be related to lack of experience among individuals in short term unemployment. As unemployment duration increase to one year and over, the rate of decrease in survival function became faster. Suggesting that people in long-term unemployment have higher exit probabilities. Table 3 show that the share of those who exited unemployment for employment is high among those in long-term unemployment (58.6%).

Figure 5b shows that individuals with no work experience have higher survival functions compared to those with work experience. In addition their survival functions are steady during the first twelve months of unemployment. This indicates that a large proportion of those who found employment have work experience. The figure clearly depicts the impact of job experience on the probability of leaving unemployment.

4.2.1.2. Weighted Kaplan Meier survival curve

Survival curves are compared to explore a univariate association between unemployment duration and a set of independent covariates. Figure 6 plots the survival curves for gender, age group, population group, education level, marital status and province. Parallel curves imply that the covariates have a significance influence on unemployment duration.



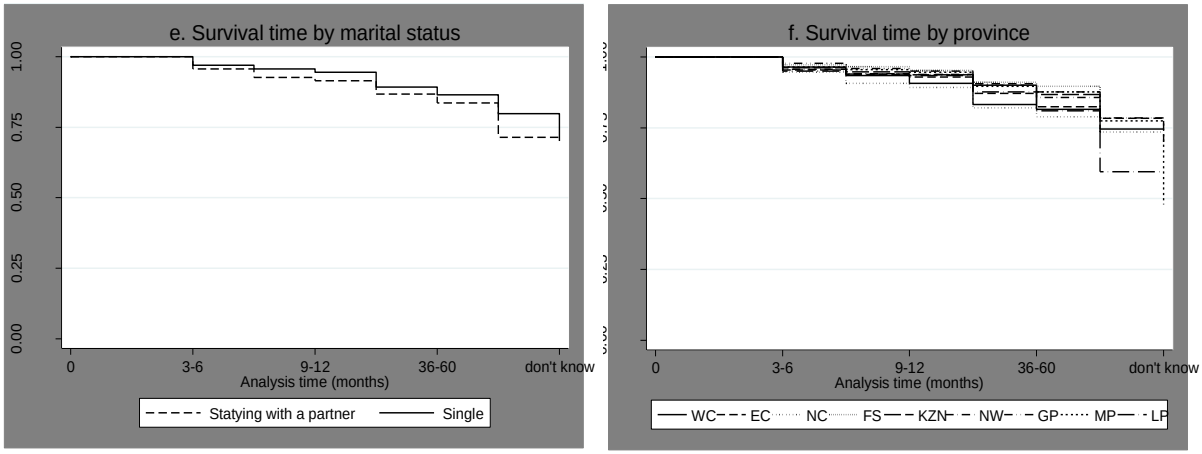


Figure 6: Kaplan Meier survival estimates by covariates

Figure 6a shows that men and women have different survival functions. The survival functions for men are lower compared to that of women, suggesting that unemployment exit rates are higher among men than among women. However both men and women in short-term unemployment have minimal exit probabilities, this is shown by the small decreases in their survival functions during those periods. A similar pattern is shown in Figure 6e where single people (widow/widower, divorced or separated and never been married) record higher survival functions compared to those staying with their partner.

The overlapping lines in Figure 6b suggests that youth and adults in short-term unemployment (unemployed for a period of less than a year) have the same rate of exiting unemployment. The survival functions of young people (aged 15- 34 years) decreased more than that of older people (aged 35-64 years) as their unemployment durations increased to over a year. This suggests that companies prefer to invest in youthful labour once they acquire some experience.

The white population group and those with tertiary education transition to employment at higher rates compared to their counterparts (Figure 6c and Figure 6d). Figure 6f shows that unemployment exit rates are not influenced by location; thus people in different provinces are seen to have undistinguishable rates of leaving unemployment.

Figure 6 shows that in some covariates the survival curves overlapped at some point, while in other covariates the survival curves were parallel throughout the analysis time. Overlapping curves indicates that the survival functions were equal at those points. The association of covariate and unemployment duration is not clearly presented in some covariates. The

difference in survival times for levels of covariates is assessed by a statistical test in the next section.

4.2.1.3. Cox regression based test for equality of survival curves

This section applies the Wald chi-square test statistics based on the Cox regression to statistically test equality of survival curves. Equality of survivor function is usually tested by a log-rank test. The log-rank test does not incorporate design features. The Cox regression based test is used as a substitution for log rank test because of its advantage to incorporate sample design features. Ignoring design effects biases the estimates (standard errors) used to calculate the test statistics (Heeringa, et al., 2010).

The null hypothesis (H_0) and alternative hypothesis (H_1) for the test are defined as follows:

H_0: Unemployment duration is not influenced by covariate A

H_1: Unemployment duration is influenced by covariate A

The null hypotheses are rejected at a 5% level of significance and if the value of the Wald chi-square is greater than the critical value within each covariate. The critical value is a tabulated chi-square value $\{\chi^2_{0.95}(\mathrm{df})\}$ and is found in many statistical books.

Table 4 presents results for testing equality of survival curves on various covariates. The survival curves are generated for each category (level) across covariates. For instance with gender we have a survival curve for women and a survival curve for men.

The other covariates and their levels are as follows:
Age group- youth and adults; race – black, coloured, Indians and white; marital status – staying with partner and single; educational level – less than matric, matric and tertiary; province – Western Cape, Eastern Cape, Northern Cape, Free State, KwaZulu-Natal, North West, Gauteng, Mpumalanga and Limpopo; prior activity – working, going to school and other, experience – yes and no; HH member – yes and no; NHH member – yes and no; grants – yes and no; enquire – yes and no; job ads – yes and no; internet – yes and no and network – yes and no.

For a covariate A with a level a chi-square (χ^2) test with $(\mathrm{a}-1)$ degrees of freedom(df) is constructed as follows):

$$\chi^2_{(\mathrm{a}-1)} = \hat{\beta}'_{\mathrm{A}}\left(\mathrm{Var}\hat{\beta}_{\mathrm{A}}\right)^{-1}\hat{\beta}_{\mathrm{A}} \qquad \text{(Heeringa, et al., 2010).} \qquad (5)$$

where,

$\hat{\beta}_A = \text{a vector of coefficients with dimension } (a-1)$

$\left(\text{Var}\hat{\beta}_A\right)^{-1} = \text{inverse of the variance} - \text{covariance matrix } \left(\text{Var}\hat{\beta}_A\right)$

According to table 4 the survival curves across gender, marital status, educational level, province, prior activity, experience, HH member (support by household member), grants and network are significantly different at a 5% level of significance. The p-values ranged from 0.0000 for gender, prior activity and experience to 0.0370 for grants. The null hypotheses on these variables are rejected – suggesting that the time spent in unemployment is influenced by this covariates.

Table 4: Wald chi-squared test for equality of survival curves

Covariate	Degrees of freedom	P-value	Wald Chi-Square	Critical value
Gender	1	0.0000	24.16	3.84
Age group	1	0.1796	1.80	3.84
Race	3	0.1989	4.65	7.81
Marital status	1	0.0005	11.97	3.84
Educational level	2	0.0003	16.28	5.99
Province	8	0.0015	25.06	15.5
Prior activity	2	0.0000	81.57	5.99
Experience	1	0.0000	60.89	3.84
HH member	1	0.0016	9.92	3.84
NHH member	1	0.5997	0.28	3.84
Grants	1	0.0354	4.43	3384
Enquire	1	0.2446	1.35	3.84
Job ads	1	0.1550	2.02	3.84
Internet	1	0.9006	0.02	3.84
network	1	0.0098	6.67	3.84

The results on table 4 suggest that the covariates age, race, NHH member (support by non-household member), enquires, job ads and internet have no univariate effect on the time spent unemployed. The p-values for these covariates ranged from 0.1501 for job ads to 0.9294 for internet.

The purpose of a univariate analysis is to get an indication of variables that could be relevant for the model. According to the rule of thumb variables with p-values of more than 0.25 in a

univariate analysis are less likely to contribute to the model (Institute for digital research and education, 2006).

4.2.2. Nelson-Aalen estimator

The Nelson-Aalen estimator is a non-parametric estimator of the cumulative hazard function (Cleves, et al., 2004). According Klein & Moeschberger (2003) Nelson-Aalen method provides an efficient means of estimating a cumulative hazard function.

The hazard function $h(t)$ for unemployment is the probability of exiting unemployment in an interval $[t, t+h]$, for h very small positive number, given were unemployed until time t is given by

$$h(t) = \lim_{\Delta t \to 0} \frac{Pr(t+\Delta t > t | T > t)}{\Delta t} \quad (6)$$

$$= \frac{f(t)}{S(t)}$$

The cumulative hazard function is defined as:

$$H(t) = \int_0^t h(u)du \; = \int_0^t \frac{f(u)}{S(u)} du = -\int_0^t \frac{dS(u)}{S(u)} \quad (7)$$

$$= -\ln\{S(t)\}$$

The hazard function measures the risk of employment occurring at time t, and the total accumulated risk up to time t is measured by a cumulative hazard function $H(t)$.

Hazard functions are estimated by applying smoothing techniques to the estimated cumulative hazard. The Nelson-Aalen estimator uses Kernel smoothing technique to estimate hazard functions $\left(\hat{h}(t)\right)$, which can be used to measure the probabilities of exiting unemployment in different unemployment durations.

The smoothed estimate of the hazard function at time t is given by

$$\hat{h}(t) = b^{-1} \sum_{j=1}^{D} K\left(\frac{t-t_j}{b}\right) \Delta\hat{H}(t_j) \quad (8)$$

where,

$\hat{h}(t) =$ estimator of the hazard function

$b = \text{bandwidth}$

$K\left(\frac{t-t_j}{b}\right) = \text{kernel density}$

$\Delta\hat{H}(t_j) = \hat{H}(t_j) - \hat{H}(j_{j-1})$ is the estimated hazard contribution

$D = \text{number of times at which employments occur}$

In determining duration dependence, the rate of change is calculated as $(\frac{dh(t)}{dt})$, assuming that $h(t)$ is differentiable,

when, $\frac{dh(t)}{dt} = 0$, $\frac{dh(t)}{dt} > 0$ and $\frac{dh(t)}{dt} < 0$, the hazard is constant, increasing and decreasing, respectively.

The prospect of getting a job is said to be dependent on time spent unemployed if the rate of change is different for different times (unemployment duration). That is if $\frac{dh(t)}{dt} > 0$ or $\frac{dh(t)}{dt} < 0$ for all $t > 0$ then duration dependence holds (Wooldridge, 2002).

Hazard functions $h(t)$ are calculated for the different unemployment durations, and the values of the hazard over time are used to calculate the rate of change $(\frac{dh(t)}{dt})$. That is duration dependence holds if the hazard rate changes over time.

The method of Nelson-Aalen is applied to estimate hazard function for the different unemployment durations t_j.

The Kernel smoothing techniques for estimating hazard function are inappropriate for the QLFS panel data. The Gaussian kernel applied in smoothing hazard has an exponential distribution.

$$K(t) = \frac{1}{2\pi} e^{\frac{t^2}{2}} \quad (9)$$

The time variable (unemployment durations) in the dataset is categorical. The number of failures in each duration are mutually exclusive (different people are observed per duration).

This study estimates hazard functions using the Kaplan Meier type estimate (Collett, 2003).

The estimated hazard function in the interval t_j to t_{j+1} is given by

$$\hat{h}(t) = \frac{d_j}{n_j \tau_j} \qquad (10)$$

where

$\tau_j = t_{j+1} - t_j$.

The probability of failure (finding employment) at duration t_j is estimated as the ratio of those who found employment at a given time to the number of individuals at risk (unemployed).

$$e_j = \frac{d_j}{n_j} \qquad (11)$$

Table 5 presents estimate of the probability of failure (e_j) and the estimated hazard function ($\hat{h}(t)$) per unemployment duration (time interval). For the unemployment duration of sixty months and over, there is no estimate of the hazard function. This time interval is open ended, thus it is impossible to estimate the hazard function on that interval (Collett, 2003).

Table 5: Distribution of employment hazard function

Time interval (t_j)	$\boldsymbol{\tau_j}$	$\boldsymbol{n_j}$	$\boldsymbol{d_j}$	$\boldsymbol{e_j}$	$\boldsymbol{\hat{h}(t)}$
Months	**Number**	**Thousand**		**Proportions**	
0-3	3	4 880	0	0.0000	0.0000
3 -6	3	4 397	148	0.0337	0.0112
6 -9	3	3 980	71	0.0178	0.0059
9 -12	3	3 600	44	0.0122	0.0041
12-36	24	3 258	179	0.0550	0.0023
36-60	24	1 986	65	0.0328	0.0014
60 and over		1 325	129	0.0973	

Table 5 shows that the hazards of finding employment are low in all the time intervals and that the hazards decrease as the duration lengthens. The hazard function is a decreasing function of time implying duration dependence. The highest hazard is seen among those who were unemployed for a period of three to six months at 148/(3x4397)=0.0112. The probability of failure shows minimal unemployment exit probabilities in all time intervals and this explains the low hazard rates.

The probability of finding employment is observed at 148/4397=0.0337 among those who were unemployed for a period of three to six months. The probability decreased to 44/3600=0.0122 as unemployment duration increases to between nine and twelve months. The results in Table 5 suggest that probabilities of finding employment in South Africa are minimal. Exit probabilities of more than 5% are seen among those in long-term unemployment, while those in short-term unemployment have exit probabilities of less than 5%.

The slow exit probabilities increase the length of stay in unemployment, which translates to higher incidence of long-term unemployment. The time spent unemployed thus became a significant factor that influences unemployment.

5. MARKOV CHAIN

Markov chain is a stochastic process, where the outcome of an experiment depends only on the outcome of the previous experiment (Kemeny & Snell, 1976). In a Markov chain the process moves from one state to another and the next state is predicted by only using the previous experiment and neglecting any other information about the past.

In this study Markov chain is applied to predict transition probabilities for other quarters starting with Q1: 2014. The resulting transition matrices indicate changes in transition probabilities as time increases.

A transition matrix for a Markov chain is a matrix P with p_{ij} entries.

$$\mathrm{P} = \begin{pmatrix} \mathrm{p}_{11} & \mathrm{p}_{12} & \mathrm{p}_{13} \\ \mathrm{p}_{21} & \mathrm{p}_{22} & \mathrm{p}_{23} \\ \mathrm{p}_{31} & \mathrm{p}_{32} & \mathrm{p}_{33} \end{pmatrix}$$

$$= \begin{pmatrix} 0.929 & 0.032 & 0.039 \\ 0.131 & 0.680 & 0.189 \\ 0.041 & 0.059 & 0.900 \end{pmatrix}$$

The transition probabilities in the matrix P are calculated from the Quarterly Labour Force Survey Panel data - Q3: 2013_Q4: 2013. The matrix(P) shows labour market movement between Q3: 2013 and Q4: 2013.

where,

$\mathrm{p}_{11} = 0.929$ is the probability of remaining employed

$\mathrm{p}_{12} = 0.032$ is the probability of leaving employment for unemployment

$\mathrm{p}_{13} = 0.039$ is the probability of leaving employment for inactivity

$\mathrm{p}_{21} = 0.131$ is the probability of leaving unemployment for employment

$\mathrm{p}_{22} = 0.680$ is the probability of remaining unemployed

$\mathrm{p}_{23} = 0.189$ is the probability of leaving unemployment for inactivity

$\mathrm{p}_{31} = 0.041$ is the probability of leaving inactivity for employment

$\mathrm{p}_{32} = 0.059$ is the probability of leaving inactivity for unemployment

$p_{33} = 0.900$ is the probability of remaining inactivity

The next section applies Markov chain to the matrix *P* to predict transition probabilities.

5.1. Transition probability prediction

The transition probability for Q1: 2014 is the power of the transition matrix *P* (Durrett, 2012).

$$p_{ij}^{(n+m)} = \sum_{k=1}^{3} p_{ik}^{n} p_{kj}^{m} \quad (12)$$

where,

$p_{ij}^{(n+m)}$ = estimate the probability of leaving states i for state j in quarter $(n + m)$

p_{ik}^{n} = probability of leaving state i and being in state k at quarter n

p_{kj}^{m} = probability of leaving state k and being in state j at quarter m

for $i = 1,2,3;\ j = 1,2,3;\ k = 1,2,3$

let $1 =$ employed (E), $2 =$ unemployed (U)and $3 =$ inactivity (I) .

The probability of leaving employment for unemployment (p_{12}) in Q1: 2014 is estimated from the matrix *P* as follows:

$$\begin{aligned} p_{12} &= p_{11}p_{12} + p_{12}p_{22} + p_{13}p_{32} \\ &= (0.929)(0.032) + (0.032)(0.680) + (0.039)(0.059) \\ &= 0.054 \end{aligned}$$

The other transition probabilities are calculated the same way, and the resultant transition probability matrix for labour market movement between Q4: 2013 and Q1: 2014 is:

$$P_{Q1:2014} = \begin{pmatrix} 0.868 & 0.054 & 0.077 \\ 0.218 & 0.478 & 0.303 \\ 0.082 & 0.095 & 0.824 \end{pmatrix}$$

Transition probabilities for the second quarter of 2014 are predicted using the matrix *P* and the matrix $P_{Q1:2014}$.

Such that,

$$P_{Q2:2014} = P.P_{Q1:2014}$$

$$= \begin{pmatrix} 0.929 & 0.032 & 0.039 \\ 0.131 & 0.680 & 0.189 \\ 0.041 & 0.059 & 0.900 \end{pmatrix} \begin{pmatrix} 0.868 & 0.054 & 0.077 \\ 0.218 & 0.478 & 0.303 \\ 0.082 & 0.095 & 0.824 \end{pmatrix}$$

$$= \begin{pmatrix} 0.817 & 0.070 & 0.114 \\ 0.277 & 0.351 & 0.372 \\ 0.122 & 0.116 & 0.763 \end{pmatrix}$$

Transition probabilities for the third quarter of 2014 are predicted by solving the square of the matrix$P_{Q1:2014}$.

such that,

$$P_{Q3:2014} = P_{Q1:2014}.P_{Q1:2014}$$

$$= \begin{pmatrix} 0.868 & 0.054 & 0.077 \\ 0.218 & 0.478 & 0.303 \\ 0.082 & 0.095 & 0.824 \end{pmatrix} \begin{pmatrix} 0.868 & 0.054 & 0.077 \\ 0.218 & 0.478 & 0.303 \\ 0.082 & 0.095 & 0.824 \end{pmatrix}$$

$$= \begin{pmatrix} 0.772 & 0.081 & 0.148 \\ 0.319 & 0.269 & 0.412 \\ 0.159 & 0.128 & 0.713 \end{pmatrix}$$

The transition matrices for each quarter show that people who lost their jobs are more likely to be inactive than being actively looking for employment. Among the 22.8% ($p_{12} + p_{13}$) of those who lost their jobs between the fourth quarter of 2013 and the third quarter of 2014, about 15% (p_{13}) became inactive.

The rate of transitioning from unemployment to employment was high between the first six months. The probability of leaving unemployment for employment increased from 13.1% (in Q4: 2013) to 21.8% (in Q1: 2014). The matrix $P_{Q3:2014}$ show that jobless people stop job search activities and became inactive as unemployment duration increases.

Figure 7 use a transition diagram to illustrate the probabilities in the matrix $P_{Q3:2014}$. The three labour market states are indicated by the numbers 1, 2, and 3, where 1=employed, 2=unemployed and 3=inactivity.

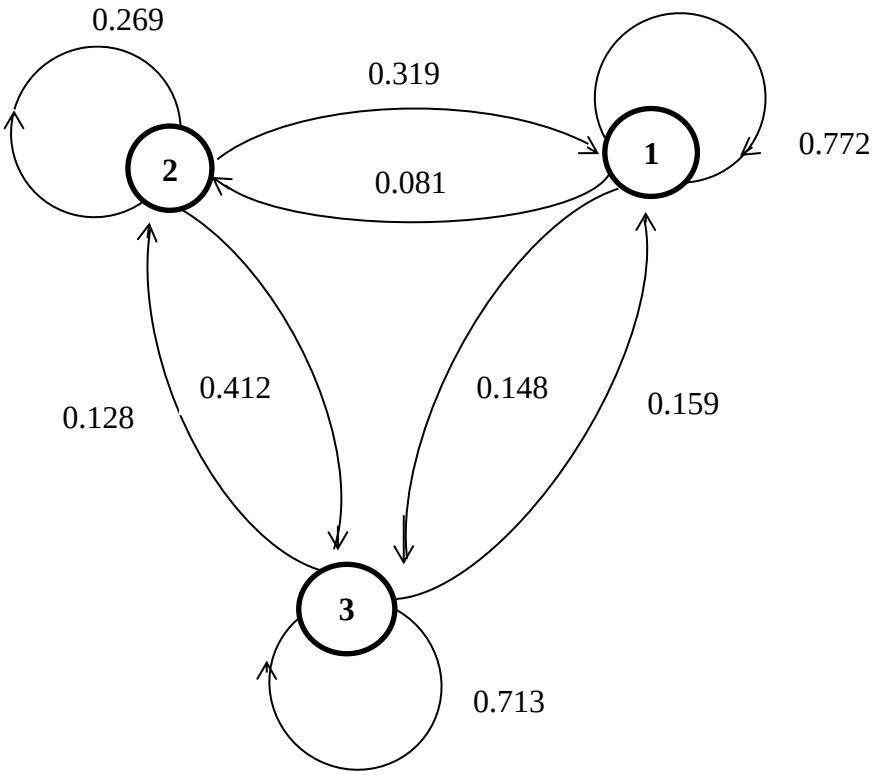


Figure 7: Transition diagram

Figure 7 show that over year an unemployed person has 26.9% chances of remaining unemployed, 31.9% chances of getting a job and 41.2% chances of transitioning to inactivity. The probability of remaining in employment decreased to 77.2% over four quarters, where 14.8% became inactive and 8.1% joined the unemployed population (actively engaged in job search activities). An inactive person has 71.3% chances of remaining in that state over a year, 15.9% chances of transitioning to employment and 12.8% chances of transitioning to unemployment.

6. ANALYSIS BASED ON SEMI-PARAMETRIC MODEL

Non-parametric analysis does not model the effects of covariates on the hazard. To explore the relationship between unemployment duration and explanatory variables semi-parametric models are computed.

Semi-parametric models examine relationships between a response variable and variables thought to have an impact on it. These models incorporate non-linear functional relationships in regression analyses. The choice of the regression models depends on the data to be analysed. It is thus important that diagnostic tests are performed on the data first to ensure that the assumptions related to the regression model are not violated.

A traditional simple linear regression analyses the relationship between a response variable with only one covariate.

Simple regression model: $Y = \beta_0 + \beta_1 X + \varepsilon$ (13)

Y: Response variable (unemployment duration)

β_0: Constant term (the equivalent of unemployment duration when the covariates is zero)

β_1: Measures the change in unemployment duration per change in the covariate

X: Covariate

ε: Error term (reflecting other factors which influence unemployment duration)

Traditional regression models assume normality, thus they are inappropriate for modeling duration data. The QLFS panel data are survival data collected according to complex survey design, these data are also characterised with censoring and non-normality.

6.1. Cox proportional hazards model

Cox proportional hazards model is a semi-parametric model which was proposed by Cox in 1972 as an approach to model relationship between survival time and covariates (Cleves, et al., 2004). The main use of Cox regression is to determine variables which affect the hazard rate. Survival predictions are difficult if some of the factors that had an influence are not taken into account. According to Danacica & Babucea (2010), Cox Proportional hazards regression is able to identify variables that affect survival.

The Cox proportional model has an advantage of incorporation sample design features such as complex survey design[4] (Boudreau & Lawless, 2006). Analysing data with complex design features requires a statistical technique which will account for the design features, because ignoring complex design factors bias estimates of the standard error.

The Cox proportional hazards regression model states that the hazard rate for the i^{th} person in the data is:

$$h(t_i|X_i) = h_0(t_i)\exp\left(X_i^T\beta\right) \qquad (14)$$

where,

β = vector of unknown parameters to be estimated from the data

$h_0(t_i)$ = baseline hazard function at time t_i,

that is hazard function when all predictors are equal to zero

X_i = independent predictor variables

Potential covariates are determined through the Kaplan Meier analysis. A Cox regression model with the potential covariates is fitted to determine the impact (multivariate effect) of socio-demographic factors and unemployment duration. Covariates with $p-values$ that are less than the critical value (0.05) are said to have a significance influence on unemployment duration. The extent of the association is measured using the resulting hazard ratios.

The parameters β in equation (14) are estimated by maximising the partial likelihood function given by

$$L(\beta) = \prod_{i=1}^{N}\left[\frac{h(t_i|X_i)}{\sum_{j=1}^{N} Y_j h(t_i|X_j)}\right]^{\sigma_i} \text{ (Collett, 2003)} \qquad (15)$$

where

t_i is the failure time of the $i-th$ unit for $t_1, \ldots t_N$;

$\sigma_i = 1$ if the $i-th$ unit is an observed failure and 0 if the $i-th$ unit is censored;

$Y_j(t) = 1$ if $t \leq t_j$, 0 if $t > t_j$.

[4] A complex survey design is a probability sample developed using sampling procedures such as stratification, clustering and weighing. The main idea behind this sampling method is to reduce cost and to improve precision of subpopulation estimates.

The partial likelihood is maximised such that

$$\sum_{i=n}^{N} \delta_i \left[X_i(t_i) - \frac{S^{(1)}(t_i,\beta)}{S^{(0)}(t_i,\beta)} \right] = 0 \qquad (16)$$

where

$$S^{(0)}(t,\beta) = \frac{1}{N}\sum_{i=1}^{N} Y_i \exp\left(X_i^T \beta\right)$$

$$S^{(1)}(t,\beta) = \frac{1}{N}\sum_{i=1}^{N} Y_i X_i \exp\left(X_i^T \beta\right)$$

Equation (16) can be used to estimate the parameters β when a finite population is observed.

To estimate the parameters $\hat{\beta}$ for a subset of the population (sample of size n), Binder (1992) suggests maximising a weighted partial likelihood.

The weighted partial likelihood is given by

$$L(\hat{\beta}) = \prod_{i=1}^{N} \left[\frac{h(t_i|X_i)}{\sum_{j=1}^{N} Y_j h(t_i|X_j)} \right]^{w_i \sigma_i} \qquad (17)$$

where

$w_i =$ sampling weights for unit i and $\sum w_i = 1$.

The weighted partial likelihood is maximised such that

$$\sum_{i=n}^{N} w_i \delta_i \left[X_i(t_i) - \frac{\hat{S}^{(1)}(t_i,\hat{\beta})}{\hat{S}^{(0)}(t_i,\hat{\beta})} \right] = 0 \qquad (18)$$

where

$$\hat{S}^{(0)}(t,\hat{\beta}) = \frac{1}{N}\sum_{i=1}^{N} w_i Y_i \exp\left(X_i^T \hat{\beta}\right)$$

$$\hat{S}^{(1)}(t,\hat{\beta}) = \frac{1}{N}\sum_{i=1}^{N} w_i Y_i X_i \exp\left(X_i^T \hat{\beta}\right)$$

The design-based variance is derived by applying Taylor series linearization to a weighted partial likelihood score vector (Binder, 1992) given by

$$\hat{U}(\hat{\beta}) = \sum_{i=1}^{n} w_i u_i(\hat{\beta}) = 0 \qquad (19)$$

where

$u_i(\hat{\beta})$ is the contribution to the score vector from ith cluster.

The design-based variance of $\hat{\beta}$ is given by

$$\mathrm{Var}(\hat{\beta}) = J^{-1}\mathrm{Var}\{\hat{U}(\beta)\}J^{-1} \quad (20)$$

where

$J = \frac{\partial \hat{U}(\beta)}{\partial \beta}$.

This section applies Cox proportional model to determine the effect of explanatory variable on hazard rate and on the length of stay in unemployment.

Semi-parametric models do not make prior assumptions about the baseline function; however Cox proportional assumes proportional hazards (constant relative hazards) between two groups. This assumption must be satisfied prior to application of the Cox model, since violation will results to invalid results.

Consider two individual with covariate values X_i and X_i^*, with a ratio of their hazards as:

$$\frac{h(t_i,X_i)}{h(t_i,X_i^*)} = \frac{h_0(t_i)\exp(X_i)\beta_x}{h_0(t_i)\exp(X_i^*)\beta_x}$$
$$= \exp[\beta_x(X_i - X_i^*)] \quad (21)$$

The expression $\exp[\beta_x(X_i - X_i^*)]$ does not depend on time t_i, implying proportionality of the two hazards.

The Cox model also assumes non-informative censoring. The non-violation of this assumption was considered during the matching process. The proportionality assumption is tested in section 6.1.2.

6.1.1. Model building

Prior to a statistical analysis the predictive power of a model is being determined through the process of model building. The model building process selects suitable variables for the Cox proportion model to ensure that the fitted model defines the objective of the study.

Collett (2003) discourages analysts from using automatic routines for variables selection because they have a limited role in model selection and do not account for the hierarchic

principle. A general strategy for model selection is recommended. To identify variables in which the hazard function depends on, the recommended approach assumes that all variables have an equal footing. This approach has four model selection steps.

This section implements the Collett's model selection approach to select covariates which significantly influence unemployment duration (Collett, 2003). The first step requires that a univariate model be fitted to identify predictors with small $p-values$ (lower than 0.2). Covariates with $p-values$ that are smaller than 0.2 are assumed to be relevant in the model.

The second step requires the analyst to fit a multivariate model with all significant univariate predictors, and use a backward selection and eliminate non-significant covariates (covariates with $p-values$ that a greater than 0.1). The recommended significance level in this selection approach is 10%. Step three considers the variables which were excluded in step two, and check for their multivariate significance. This step assumes that relevancy of other variables might depend on other variables in the model. Step four involves a stepwise selection as a final check of important variables; this step also considers interaction effects.

A univariate analysis in section 4 (Table 4) identified twelve significance covariates at 0.2 level of significance. The covariates are gender, age group, race, province, education, marital status, prior activity (activity prior to job search), experience, HH member (supported by person in the household), grants (receive child/foster care grants), job ads (searched through job advertisement) and network (sought assistance from relatives or friends) – their p-values range from 0.000 to 0.1989. These covariates are included the second step of model selection.

The identified covariates are categorical with different factors (categories) and a reference factor is defined in each variable. The reference factor for gender is 'male', for age group is 'youth' (15-34 years), for race is 'black,/African', for province is 'Western Cape', for education is 'tertiary', for marital status is 'staying with a partner' and for prior activity is 'working'. The effect of the indicator variables (experience, HH member, grants, job ads and network) is measured by using the reference category 'yes'.

Table 6 presents results on the Cox regression model for multivariate effect. In each covariate the reference factors are omitted.

Table 6: Cox regression for variable selection

Covariates	Hazard Ratio	Standard error	t-statistic	P-value	95% Confidence	
Gender (female)	0.6903	0.0723	-3.54	0.0000	0.5622	0.8477
Age group (adults)	0.5224	0.0545	-6.22	0.0000	0.4256	0.6411
Race (coloured)	0.8934	0.1659	-0.61	0.5440	0.6205	1.2862
Race (Indian/Asian)	0.7009	0.3835	-0.65	0.5160	0.2396	2.0509
Race (white)	1.2445	0.3912	0.7	0.4870	0.6716	2.3061
Province (Eastern Cape)	0.9377	0.1807	-0.33	0.7390	0.6424	1.3687
Province (Northern Cape)	0.9699	0.2123	-0.14	0.8890	0.6313	1.4903
Province (Free State)	0.5387	0.1181	-2.82	0.0050	0.3504	0.8281
Province (KwaZulu-Natal)	0.7205	0.1563	-1.51	0.1310	0.4707	1.1028
Province (North West)	0.9207	0.2105	-0.36	0.7180	0.5879	1.4420
Province (Gauteng)	0.7372	0.1369	-1.64	0.1010	0.5121	1.0612
Province (Mpumalanga)	0.9270	0.2038	-0.34	0.7300	0.6022	1.4271
Province (Limpopo)	1.0489	0.2479	0.2	0.8400	0.6597	1.6677
Education (matric)	0.5761	0.1001	-3.17	0.0020	0.4097	0.8101
Education (below matric)	0.4664	0.0839	-4.24	0.0000	0.3277	0.6638
Marital status (single/divorced)	0.7162	0.0760	-3.14	0.0020	0.5816	0.8820
Prior activity (scholar)	0.5535	0.0954	-3.43	0.0010	0.3947	0.7762
Prior activity (other)	0.6002	0.1295	-2.37	0.0180	0.3930	0.9166
Experience (no)	0.5797	0.1187	-2.66	0.0080	0.3879	0.8665
HH member (no)	1.3827	0.1482	3.02	0.0030	1.1204	1.7064
Grants (no)	1.1710	0.1975	0.94	0.3500	0.8410	1.6304
Job ads (no)	1.2484	0.1397	1.98	0.0480	1.0023	1.5550
Network (no)	1.1693	0.1164	1.57	0.1170	0.9617	1.4216

Table 6 shows that the covariates gender, age group, education, marital status, prior activity, experience, household member and job ads are significant at a 0.1 level of significance. The factors of gender, age group and education (below matric) recorded p-values of 0.000 (the lowest). P-values of the other significant factors ranged from 0.001 (scholar) to 0.048 (job ads). The final model will be fitted with these eight covariates.

The factors for race, province, grants and network have p-values of more than the 0.1 significance level with exception for Free State (with p-value of 0.005). Covariates with higher p-values (higher than the significance level) have no multivariate effect on the model that is their contributions to the model are not significant when they are fitted with other variables. These covariates are excluded from the final model.

6.1.2. Testing the proportional hazards assumption

In every regression analysis it is crucial that the analyst perform model diagnostics to ensure non-violation of the model assumption. The Cox proportional hazards model assumes that the hazards of two observations are proportional. This assumption is verified prior to application of the Cox proportional hazards model. Schemper (1992) argues that the relative risk of covariates with non-constant hazard ratios is either overestimated or underestimated depending on the direction of change.

Well-known techniques for testing the proportional hazards assumption include the log-minus-log plots and the Schoenfeld residuals. Bellera, et al. (2010) recommends Schoenfeld residual plots over the log-minus-log plots, because the Schoenfeld residual plots are independent of time. The log-minus-log plots are said to be misleading because of their non-reaction to the structure of the data (Schemper, 1992).

This analysis applies a test based on Schoenfeld residuals to test the proportional hazards assumption on the QLFS data. With the Schoenfeld residual test proportionality is tested using both p-values and a graphical display. The test calculates weighted and scaled Schoenfeld residuals by fitting a Cox model. The weighted Schoenfeld residuals used to calculate p-values and the scaled residuals are plotted for a graphical test.

Schoenfeld residuals for covariate x_k, $k = 1, \dots, z$, and observation j are defined as the difference between the explanatory variable x_{kj} and the mean of the other persons in the risk set, weighted by their estimated relative hazard (Cleves et al., 2004).

The Schoenfeld residuals are given by

$$r_{kj=}\, x_{kj} - \frac{\sum_{i \epsilon R_j} x_{kj} \exp(X_i \hat{\beta}_x)}{\sum_{i \epsilon R_j} \exp(X_i \hat{\beta}_x)} \qquad (22)$$

where

$R_j = \text{risk set (subject at risk of failure)}.$

The Schoenfeld residual based test assumes homogeneity of variance across risk sets, which does not hold for the QLFS panel data (the QLSF panel data has complex design features). Cleves et al. (2004) recommend that the proportional hazards assumption be tested for each covariate separately in cases where the assumption does not hold.

Table 7 presents the results for testing proportionality of eight covariates – gender, age group, education, marital status, prior activity, experience, HH member and job ads. The proportionality assumption is violated if a covariates has significant p-values (p-values which are smaller than 0.05).

Table 7: Model for testing the proportional-hazards assumption

Covariates	rho	Chi-square	Degrees of freedom	P-value
Gender (female)	0.10009	8.04	1	0.0046
Age group (adults)	0.04535	1.45	1	0.2284
Education (matric)	0.05771	2.22	1	0.1358
Education (below matric)	0.00720	0.04	1	0.8462
Marital status (single/divorced)	0.00527	0.02	1	0.8921
Prior activity (scholar)	0.02312	0.43	1	0.5114
Prior activity (other)	0.06511	3.32	1	0.0683
Experience (no)	0.08832	5.87	1	0.0154
HH member (no)	-0.10588	8.12	1	0.0044
Job ads (no)	-0.07409	4.41	1	0.0358
global test		104.01	10	0.0000

According to Table 7 the proportionality assumption is not violated in four of the eight covariates (age group, education, marital status and prior activity). The factors of these covariates have p-values that are greater than 0.05 –adults (0.2284); matric (0.1358); below matric (0.8462); single/divorced (0.8921); scholar (0.5114) and other (0.0683).

The results indicate that the p-values for the factors of gender; experience, HH member and job ads are significant at 0.05 level of significance – implying violation of the proportionality assumption. The p-values for these covariates ranged from 0.0044 for the factor of HH member to 0.0358 for the factor of job ads.

The graphical method is applied to assess the slope of the scaled Schoenfeld residuals against time. Grambsch & Therneau (1994) suggest a graphical display when an investigator has no hypotheses about the nature of the non-proportionality.

Figure 8 shows the plots for testing proportionally of gender; experience, HH member and job ads. Plots which is centred about zero (zero slope) indicate that the proportionality assumption is not violated (Schoenfeld, 1982). The assumption is violated when a non-random pattern against time is depicted from the plots.

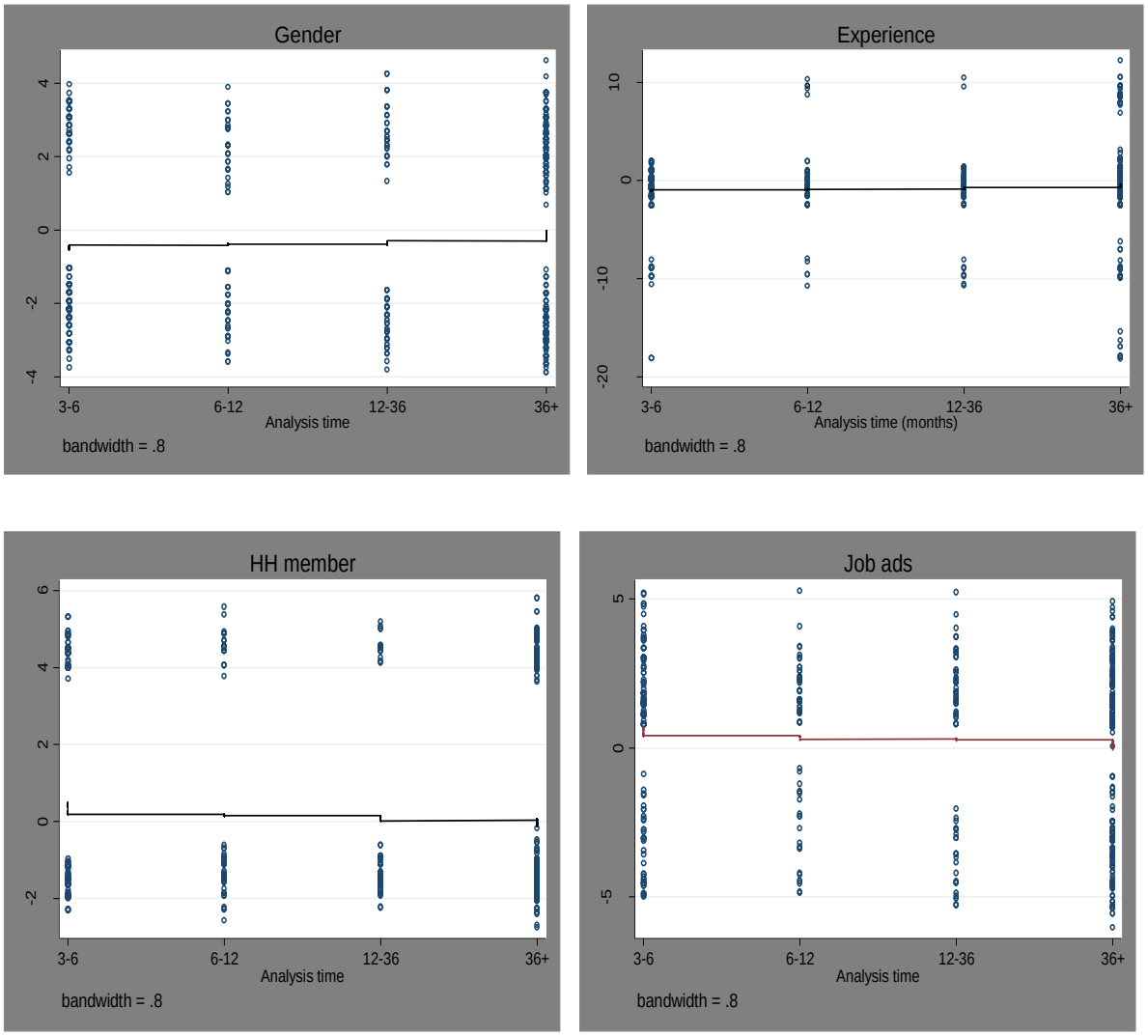


Figure 8: Schoenfeld plots

The graphical test as indicated in figure 8 shows that the proportionality assumption on gender, experience, HH member and job ads is not violated. The figure shows that the scaled Schoenfeld residuals plots are approximately centred around zero.

6.1.3. The fitted model

Table 8 presents results for examining effects of covariates on unemployment duration. Hazard ratios are estimated by fitting eight covariates to a Cox proportional hazards model. Significant p-values suggest a significance impact on the hazard of leaving unemployment at any given duration. A hazard ratio of less than one indicates that the factor (category) of the covariates has a lower probability of exiting unemployment compared to the reference category.

The covariates include demographic (gender, age group, education level and marital status), economic activity (prior activity) and indicator variables (experience, support by household member and searched job adverts).

Table 8: Cox proportional hazards model for estimating hazard ratios

Covariates	Estimated hazard ratio	Estimated standard error	t-statistic	P-value	95% confidence interval	
Gender (female)	0.6582	0.0623	-4.42	0.0000	0.5466	0.7926
Age group (adults)	0.5218	0.0532	-6.38	0.0000	0.4272	0.6373
Education (matric)	0.5525	0.0937	-3.50	0.0000	0.3961	0.7706
Education (below	0.4386	0.0751	-4.81	0.0000	0.3135	0.6137
Marital status (single/divorced)	0.7236	0.0755	-3.10	0.0020	0.5896	0.8880
Prior activity (scholar)	0.5345	0.0916	-3.65	0.0000	0.3818	0.7482
Prior activity (other)	0.5935	0.1307	-2.37	0.0180	0.3852	0.9144
Experience (no)	0.5862	0.1204	-2.60	0.0090	0.3917	0.8772
HH member (no)	1.3639	0.1413	3.00	0.0030	1.1130	1.6713
Job ads (no)	1.2860	0.1409	2.29	0.0220	1.0371	1.5945

Table 8 shows that the hazard of leaving unemployment at any given duration is significantly lower for most of the categories within the covariates relative to their reference categories. Holding the other covariates fixed the hazard of leaving unemployment at any given duration is:

- 34.2% lower for females than for men.
- 47.8% lower for adults than for youth
- 44.7% and 56.1% lower for those with education level of matric and below matric (respectively) than those with tertiary education.
- 27.6% lower for single and divorced persons than for those who are married or living together as husband and wife.
- 46.6% and 40.6% lower for those who were attending school and those who were doing other activities prior to job search (respectively) than those who were working.
- 41.4% lower for those with no work experience than those with work experience.

People who are supported by household member have a 36.4% more chances of exiting unemployment compared to those who use other means of support. The probability of exiting unemployment is 28.6% higher among those who searched job adverts when looking for employment compared to those who used other search methods.

7. DISCUSSION

The study applied survival analysis (Non-parametric and semi-parametric) techniques to South African labour market data. Non-parametric and semi-parametric models are fitted to a QLFS panel data to address the following study objective:

- Investigating dependence of the prospect of getting a job on time spent in unemployment (duration dependence).
- Determining probabilities of exiting unemployment in different unemployment durations.
- Determining association of socio-demographic factors and unemployment duration and the extent of the association.

Duration dependence can be positive or negative and is determined by hazard functions. When the hazard of finding employment increases with unemployment times, positive duration dependence is observed. Negative duration dependence happens when the hazard of finding employment decreases as unemployment times increases. Duration dependence does not hold when the probabilities of exiting unemployment remain the same for all the unemployment times.

Hazard functions are estimated in section 4 (column $\hat{h}(t)$ of Table 5) of this analysis. The results suggest negative duration dependence, where the hazard of finding employment decreased from 0.0112 to 0.0014 (as the unemployment times increased from three months to 60 months). These findings do not coincide with the Weibull analysis by Brick & Mlatsheni (2008). Their findings suggests an increasing hazard among those in long-term unemployment and a decreasing hazard as unemployment time increases to over 13 months.

The probabilities of exiting unemployment for different unemployment times show the effect of lengthy duration on unemployment. According to Lancaster & Nickell (1980) unemployment is a probabilistic process, where the probability of leaving unemployment for employment varies over time. Employment transition probabilities are presented in column e_j of Table 5. The transition probabilities are low for all the unemployment times, which results to higher incidence of long-term unemployment. According to Ciuca & Matei (2010) a labour market is damaging if the unemployed stay unemployed for a long time.

7.1. Strength and critics of the survival techniques

Section 4 of the dissertation uses Kaplan Meier estimate to estimate survival function of the unemployed. Kaplan Meier assumes that at any unemployment time individuals who are censored (not observed in the next duration) have the same survival prospects as those who continue to be followed. According to Van Den Berg, et al. (1994) a portion of individuals who finds jobs are lost to follow-up. These suggest censored individuals have different survival prospects compared to those who continue to be followed.

Survival analysis techniques are inadequate for analysis of mean time to failure or median time to failure. Survival data is characterized with censored objects and subjects has multiple entries. Cleves, et al. (2004) suggest calculation of survival time to estimate the median (using the point where the survival probability is 0.5 as the median).

Survival functions can also be estimated using life table method; this method has an advantage of estimating mean survival time and median time. However the life table method requires larger samples where the time intervals are large enough to be divided into smaller units. The Kaplan Meier estimates yield better estimates compared to the life table method because it uses exact survival times to make time stratification.

The Cox proportional hazards model in section 6 assumes proportional hazards and linearity of the exponential argument. The model yield invalid results when these assumptions are violated. Binder (1992) argues that when a sample has complex design features, design parameters may relate to the true hazard function, but not part of the model fitted. Linearity of covariates can be violated for many of the quantitative factors.

Possible techniques for modeling the relationship between unemployment duration and covariates include Weibull, linear and logistic regression models. The Weibull analysis requires parametric specifications of the hazard function. Narendranathan & Stewart (1993) found that such specification were not satisfactory for the Britain unemployment data. Abbring & Van Den Berg (2007) argues that misspecification of the functional form leads to biased estimators. Linear and logistic regression cannot handle censored objects, where the former assumes normality and the later does not allow for analysis of survival time.

The survival techniques are suggested in this analysis because of their strength in handling problems associated with survival data. The QLFS panel data is a survival data characterised

with censoring and violates the normality assumption. Survival techniques accounts for censored subjects and do not impose any specification on the baseline hazard. These techniques allow the data to determine the functional form of the hazard. The length of stay in unemployment is influenced by both observed and unobserved variables. Application of Cox proportional hazards to a panel data controls for unobserved variables. Panel data improves analysis by controlling for omitted variables and is able to identify dynamic behaviour.

7.2. Limitations

The QLFS panel data was designed such that only individuals who responded in all two quarters are included in the data set. These data can only be used to analyse movements between labour market categories. It is impossible to analyse other aspects of longitudinal data (for example effect of attrition on survival functions and on hazards) using this data.

The use of survey data is subject to measurement error. To determine the length of unemployment, those who are looking for employment are asked how long they have been searching for employment. There is however no way to check the validity of the answer to this question. There is considerable amount of rounded answers, which can lead to overestimating or underestimating unemployment time.

Estimation of hazard functions using Kernel smoothing techniques did not suffice. The time variable in the QLFS panel data is categorical and the Gaussian kernel applied in smoothing hazard has an exponential distribution.

8. CONCLUSION

The study analyses unemployment duration, with a focus on factors associated with long-term unemployment, duration dependence and employment transition probabilities. The study aims to improve labour market understanding and enhance labour market policy relevancy. Reports from the quarterly labour force survey showed that unemployment rate ranged from 21.8% since Q1: 2008 to 24.1% in Q4: 2013 where an average of 65.7% has been unemployed for a period of one year and longer. The analysis in this report shows that the hazard of finding employment decreases when the time spent unemployed lengthens.

The Kaplan Meier survival estimates indicate the survival functions decreases at a slow rate for all unemployment times, which suggest that unemployment exit rates are low. The analysis showed that exit probabilities are better for people with work experience compared to those with no work experience. Similar results were also found by Coulson (2009) and Tansel & Taşçi (2010).

The estimated hazard functions indicate higher hazards among those in their first six months of unemployment. However the rate is not enough to reduce pressure among those in long-term unemployment. The slow exit rates among those in short-term unemployment translate to long-term unemployment and eventually to increased level of unemployment. Lancaster & Nickell (1980) found that increases in the length of time spent unemployed lead to increases in total unemployment.

The semi-parametric analysis is performed by fitting a Cox proportional model. This model assumes proportional hazards between groups. A Schoenfeld based test and scaled Schoenfeld plots attested a non-violation of this assumption for eight covariates (gender, age group, education level, marital status, activity prior to unemployment, experience, household member and job ads).

The fitted model suggest that there is an association between socio-demographic factors and unemployment duration – where the time spent in unemployment is significantly (at a 5% level of significance) influenced by gender, age group, education level, marital status, activity prior to unemployment, experience, household member and job ads.

The hazards of leaving unemployment at any given duration are significantly lower for people in the following categories - females, adults, education level of lower than tertiary, single or divorced, attending school or doing other activities prior to job search and no work experience. Brick & Mlatsheni (2008) and Tansel & Taşçi (2010) found that women are more likely to be in long-term unemployment. Studies by Babucea & Danacica (2007) and Dias & Posel (2007) showed a negative relationship between unemployment and education.

Improved levels of education and work experience decreases unemployment duration. Other factors which affect unemployment duration includes – living arrangements, activity prior to unemployment, means of support while unemployed and method used to look for employment. According to Mussida (2007) unemployment probabilities are significantly influenced by marital status and age.

The QLFS panel data spanned the third and the fourth quarter of 2013 and is affected by seasonality – suggesting that the transition from unemployment to employment resulted from seasonal jobs. The section on Markov chains indicates a large increase among those who transition from employment to unemployment between Q4:2013 and Q1: 2014.

To improve on this analysis a study which will analyse men and women separately is necessary and it should be based on datasets which are seasonally adjusted or have no seasonal components.

8.1. Further work

- Repeat the study using a panel which spans more quarters of a year so that the unemployed are allowed enough time to search for employment and to address seasonal effects. A panel that spanned two consecutive quarters is influenced by seasonal factors.
- Expand the analysis by fitting parametric models and kernel smoothing techniques, and this can be achieved by using a different type of dataset.
- Expand the analysis by fitting separate models for men and women. Men and women have different challenges in the labour market, which influence the association between other demographic factors and unemployment duration.

References

Abbring, J. H. & Van Den Berg, G. J., 2007. The unobserved heterogeneity distribution in duration analysis. *Biometrika,* 94(1), pp. 87-99.

Altman, M., 2007. *Youth Labour market challanges in South Africa,* Pretoria: Human Science Research Council.

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

Bellera, C. A., MacGrogan, G., Debled, M., de Lara, C., Brouste, V., Mathoulin-Pe´lissier, S., 2010. *Variables with time-varying effects and the Cox model: some statistical concepts illustrated with a prognostic factor study in breast cancer.* BMC Medical Research.

Binder, D. A., 1992. Fitting Cox's proportional hazards models from survey data. *Biometrika,* 79(1), pp. 139-147.

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

Boudreau, C. & Lawless, J. F., 2006. Survival analysis based on the proportional hazards model and survey data. *The Canadian Journal of Statistics,* 34(2), pp. 203-216.

Brick, K. & Mlatsheni, C., 2008. *Examining the degree of duration dependence in the Cape Town labour market.* Cape Town: Southern Africa Labour and Development Research Unit, working paper no. 10.

Caliendo, M., Tatsiramos, K. & Uhlendorff, A., 2009. *Benefit duration, unemployment duration and job match quality: a regression-discontinuity approach.* IZA Discussion paper no 4670.

Carling, K. & Jacobson, T., 1995. Modelling unemployment duration in a dependent competing risks framework: Identification and estimation. *Lifetime data analysis,* 1(1), pp. 111-122.

Cichello, P., Leibbrandt, M. & Woolard, I., 2012. *Labour Market: Analysis of the NIDS Wave 1 and 2 Datasets.* Cape Town: Southern Africa Labour and Development research Unit, working paper no 78.

Ciuca, V. & Matei, M., 2010. *Survival analysis for the unemployment duration.* Proceedings of the 5th WSEAS international conference on economy and management transformation.

Cleves, M. A., Gould, W. W. & Gutierrez, R. G., 2004. *An inroduction to survival analysis.* Revised ed. College Station: Taylor & Francis.

Collett, D., 2003. *Modelling survival data in medical research.* Second ed. Chapman and Hall, CRC.

Coulson, L., 2009. *An exploration of the correlates of long-term unemployment in South Africa.* Dissertation submitted to the school of development studies, University of KwaZulu-Natal.

Danacica, D. & Babucea, A., 2010. Using survival analysis in economics. *Scientific Annals of the Alexandru Ioan, Cuza University of Labsi,* 2010(November), pp. 439-450.

Daniels, R. C., 2007. *Skills shortages in South Africa: A literature review.* Development Policy Research Unit, working paper no. 07/121.

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

Durrett, R., 2012. *Essentials of Stochastic Processes.* Second ed. New York: Springer-Verlag.

Erasmus, J. & Breier, M., 2009. *Skills shortages in South Africa: Case studies of key professions.* Cape Town: Human Science Research Council Press.

Fan, J. & Li, R., 2004. New estimation and model selection procedures for semi-parametric modelling in longitudinal data analysis. *Journal of the American statistical association,* 99(467), pp. 710-723.

Fox, J., 2002. *Cox proportional-hazards regression for survival data. Appendix to an R and S-plus companion to applied regression.* Sanford Weisberg.

Goel, M. K., Khanna, P. & Kishore, J., 2010. Understanding survival analysis: Kaplan-Meier estimate. *International Journal of Ayurveda Research,* 1(4), pp. 274-278.

Grambsch, P. M. & Therneau, T. M., 1994. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81(3), pp. 515-526.

Haughton, D. & Haughton, J., 2011. *Living standards analytics.* New York: Statitics for social and behavioural sciences.

Heath, A. & Swann, T., 1999. *Reservation wages and the duration of unemployment.* Reserve bank of Australia, research discussion paper 1999-02.

Heeringa, S. G., West, B. T. & Berglud, P. A., 2010. *Applied survey data analysis.* Boca Raton: Taylor and Fancis group.

Imbens, G. W. & Lynch, L. M., 2006. Re-employment probabilities over the business cycle. *Springer-Verlag,* 5(2), pp. 111-134.

Institute for digital research and education, 2006. *Statistical computing seminars: Survival analysis with STATA.* [Online]
Available at: http://www.ats.ucla.edu/stat/stata/seminars/stata_survival/
[Accessed 1 December 2014].

International Labour Organisation, 2003. *International training compendium on labour statistics, module 1: Statistics of employment, unemployment, underemployment: economically active population.* Turin: International Labour Organisation.

Jakoet, J., 2007. *The initial unemployment duration of immigrants to Kayelitsha/Mitchell's Plain.* Mini-Dissertation for a Master of Business Science in economics, University of Cape Town.

Jensen, P. & Svarer, M., 2003. Short- and Long-term unemployment: How do temporary layoffs affect the distinction?. *Emperical economics,* 28(1), pp. 23-44.

Kaplan, E. L. & Meier, P., 1958. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association,* 53(283), pp. 457-481.

Katz, L., 1986. *Layoffs, recall and the duration of unemployment.* s.l.:National bureau of economic research, working paper series no.1825.

Kemeny, J. G. & Snell, J. L., 1976. *Finite Markov Chains.* First ed. New York: Springer-Verlag.

Kiefer, N. M., 1988. Economic duration data and hazard functions. *Journal of economic literature,* 26(2), pp. 464-679.

Klein, J. P. & Moeschberger, M. L., 2003. *Survival analysis techniques for censored and truncated data.* Second ed. New York: Springer-Verlag.

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

Lancaster, T., 1979. Econometric methods for the duration of unemployment. *Econometrica,* 47(4), pp. 939-956.

Lancaster, T. & Nickell, S., 1980. The analysis of re-employment probabilities for the unemployed. *Journal of the royal statistical society. Series A (general),* 143(2), pp. 141-165.

Landmesser, J., 2011. The impact of vocational training on the unemployment duration. *International Advances in Economic Research,* 17(1), pp. 89-100.

Lu, J. & Shen, D., 2014. *Survival analysis approaches and new developments using SAS.* PharmaSUG, paper PO02.

Mills, M., 2011. *Introducing Survival and Event History Analysis.* illustraded ed. London: SAGE publications.

Morgan, J. & Mourougane, A., 2001. *What can changes in structural factors tell us about unemployment in Europe?.* European central bank, working paper series no. 81.

Mussida, C., 2007. *Unemployment duration and compelling risks: a regional investigation.* Piacenza: Universita Cattolicca Del Sacro Coure.

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

National Treasury, 2011. *Confronting youth unemployment: policy options for South Africa,* Pretoria: Discussion paper for public comments.

Nichols, A., Mitchell, J. & Lindner, S., 2013. *Consequances of long-term unemployment.* Washington D.C: The urban institute.

Nickell, S. J., 1979a. The effect of unemployment and related benefits on the duration of unemployment. *The Economic Journal,* 89(353), pp. 34-49.

Nickell, S. J., 1979b. Estimating the probability of leaving unemplpyment. *Econometrica,* 47(5), pp. 1249-1266.

OECD, 2014. *Incidence of unemployment by duration.* [Online]
Available at: https://stats.oecd.org
[Accessed 23 February 2015].

Pissarides, C. A., 1992. Loss of skill during unemployment and the persistence of employment shocks. *The quarterly Journal of economics,* 107(4), pp. 1371-1391.

Potgieter, K., 2012. *The waiting game: a survival analysis of unemployment duration in South Africa 2002-2004.* Dissertation submitted to the school of Economics, University of Cape Town.

Ranchhod, V., 2009. *Labour market: analysis of wave 1 dataset.* Cape Town: Southern Africa Labour and Development Research Unit, discussion paper no. 12.

Rao, S. R. & Schoenfeld, D. A., 2007. *Statistical primer for cariovascular research, Survival methods.* [Online]
Available at: http://circ.ahajournals.org/content/115/1/109.full
[Accessed 28 August 2014].

Schemper, M., 1992. Cox analysis of survival data with non-proportional hazards function. *Journal of the royal statistical society, Series D (the statistician),* 41(4), pp. 455-465.

Schoenfeld, D., 1982. Partial residuals for the proportional hazards regression model. *Biometrika,* 69(1), pp. 239-241.

Seyfried, W., 2005. Examining the relationship between employment and economic growth in the ten largest States. *Southwestern Economic Review,* pp. 13-24.

South West Observatory, 2008. *The South West Healthy Labour Market Review.* [Online]
Available at: http://www.swslim.org.uk/research/hlmr/
[Accessed 6 November 2013].

Statistics South Africa, 2008. *Guide to Quarterly Labour Force Survey.* [Online]
Available at: http://www.statssa.gov.za

Statistics South Africa, 2009. *P0211 - Quarterly Labour Force Survey.* [Online]
Available at: http://www.statssa.gov.za

Statistics South Africa, 2014a. *P0211 - Quarterly Labour Force Survey.* [Online]
Available at: http://www.statssa.gov.za

Statistics South Africa, 2014b. *Labour Market Daynamics in South Africa.* [Online]
Available at: http://www.statssa.gov.za

Statistics South Africa, 2014c. *Quartely Labour Force Survey Panel Data.* [Online]
Available at: http://interactive.statssa.gov.za:8282/webview
[Accessed 10 08 2014].

Statistics South Africa, 2015. *National and provincial labour market: Long-term unemployment.* [Online]
Available at: http://www.statssa.gov.za

Tansel, A. & Taşçi, H. M., 2010. *Hazard analysis of unemployment duration by gender in a developing country: The Case of Turkey.* IZA discussion paper no. 4844.

Van Den Berg, G. J., Lindeboom, M. & Ridder, G., 1994. Attrition in longitudinal panel data and the empirical analysis of dynamic labour market behaviour. *Journal of applied econometrics,* 9(4), pp. 421-435.

Webster, D., 2005. Long-term unemployment, the invention of 'hysteresis' and the misdiagnosis of structural unemployment in the UK. *Cambridge Journal of economics,* 29(6), pp. 975-995.

Witchert, L. & Wilke, R. A., 2008. Simple non-parametric estimators for unemployment duration analysis. *Journal of the royal statistical society, Series C (applied statistics),* 57(1), pp. 117-126.

Wooldridge, J. M., 2002. *Econometric analysis of cross section and panel data.* illustrated, reprint ed. Cambridge: The MIT Press.

World Development Indicators, 2014. [Online]
Available at: http://data.worldbank.org/indicator/sl.uem.totl.zs
[Accessed 1 May 2015].

Zimmer, H., 2012. Labour market mismatches. *Economic review,* 2012(September), pp. 55-68.

Appendix A

A1. Variable description

Gender

This variable defines the sex of a person; it is a categorical variable with two categories (male or female). It is derived using question 1.3 on the QLFS questionnaire. This variable is coded as either 1 or 2 on the data set, where 1 represents males and the 2 represent females.

Race

This variable defines the population group of a person; it is a categorical variable with four categories (black/African, coloured, Indian/Asian and white). It is derived using question 1.5 on the QLFS questionnaire. This variable is coded using 1,2,3 and 4 on the data set, where 1 represents black/African, 2 represents coloured, 3 represents Indian/Asian and 4 represents white.

Marital status

This variable defines the marital status of a person, it is a categorical variable with five categories (married, living together like husband and wife, widow/widower, divorced or separated and never married). It is derived using question 1.6 on the QLFS questionnaire. In this analysis some categories are collapsed into other categories to form two categories. The two new categories are represented by 1 and 2 in the data set, where 1 combines the married and living together like husband and wife categories and 2 represents the other categories (widow/widower, divorced or separated and never married).

Educational level

This variable defines the educational attainment of a person, it is a categorical variable with three categories (tertiary, matric and below matric). It is derived using question 1.7 on the QLFS questionnaire. This variable is coded using 1,2 and 3 on the data set, where 1 represents tertiary, 2 represents matric, 3 represents below matric.

Province

This variable defines the province in which a person resides; it is a categorical variable with nine categories (Western Cape, Eastern Cape, Northern Cape, Free State, Kwa-Zulu Natal, North West, Gauteng, Mpumalanga and Limpopo). It is derived using information of section

a (particulars of the dwelling) on the QLFS questionnaire. This variable is coded using values 1-91 on the data set, where 1 represents WC, 2 EC, 3 represents NC, 4 represents FS, 5 represents KZN, 6 represents NW, 7 represents GP, 8 represents MP and 9 represents LP.

Activity prior to job search

This variable defines the activities a person that a person was doing prior to being unemployed, it is a categorical variable with four categories (working, managing a home, going to school, other). It is derived using question 3.7 on the QLFS questionnaire. This variable is coded using 1,2 and 3 on the data set, where 1 represents working, 2 represents going to school and 3 represents other (combining managing a home and other).

Experience

This variable indicates whether a person had worked in the past, it is a categorical variable with two categories (yes or no). It is derived using question 3.12 on the QLFS questionnaire. This variable is coded using 1 or 2 on the data set, where 1 represents yes and 2 represents no.

Type of support variables

- Household member (HH member)
- Non-household member (NHH member)
- Grants

These variables indicate how the unemployed supported themselves during their unemployment spells; they are categorical variables with two categories (yes or no). They are derived using question 3.19 on the QLFS questionnaire. This variable is coded using 1 or 2 on the data set, where 1 represents yes and 2 represents no.

Job search methods

- Enquiring at work places (enquire)
- Placed/answered job advertisements (job ads)
- Searched the internet (internet)
- Sought assistance from relatives or friends (network)

These variables indicate the job search methods used to look for employment; they are categorical variables with two categories (yes or no). they are derived using question 3.2 on the QLFS questionnaire. This variable is coded using 1 or 2 on the data set, where 1 represents yes and 2 represents no.

A2. Collapsing stratum and sampling units

Table A1: Collapsed stratums and sampling units

Old stratum	New stratum
103102	103101
103201	101201
103501	104501
104103	102103
210101	210102
210501	212404
212501	212404
213101	213102
214201	212201
215101	212101
215407	215405
244401	244403
244404	244403
309501	308501
420501	419501
521102	521101
523101	524101
543402	543401
572102	572103
572108	572103
572114	572113
637201	640201
637402	637401
638101	637102
638102	637102
639501	640501
640103	640101
746101	742101
774101	774102
774202	774102
832408	832409
933101	934401
933201	933401
936501	947401

Old sampling unit number	New sampling unit number	stratum
Missing	10600010	102101
Missing	17102073	171109
Missing	17100803	171115
Missing	22700159	214401
Missing	27520537	275104
Missing	27520537	275104
Missing	68100038	345401
Missing	40500269	417105
Missing	41500034	417404
Missing	41400113	419101
Missing	41600017	419103
Missing	41800156	420101
Missing	41900054	420103
Missing	51100659	522102
Missing	52400103	525101
Missing	57203493	572201
Missing	60200333	637404
Missing	77401051	774105
Missing	77402524	774111
Missing	77402485	774115
Missing	77600261	776105
Missing	67600433	776203
Missing	80700046	830104
Missing	80700046	830104
Missing	81500791	832403
Missing	81500200	832403
Missing	81700392	832409
Missing	91200746	935403
Missing	91000007	935406
Missing	11200045	103101

A3. STATA Code (do file)

```
use "C:\Users\zandileno\Desktop\unisa\dessertation\Assignment\wide_25122014.dta", replace

***checking stratum_psu combination for the whole sample

gen wgt=full_calwgt/1000

svyset psuno_q32013 [pweight=wgt], strata(stratum_q32013)

svydes

*********selecting those who where looking fo employment in Q3:2014

keep if status_q32013=="2"

gen id =_n

replace dur_q3 = "1" if dur_q3=="01"

replace dur_q3 = "2" if dur_q3=="02"

replace dur_q3 = "3" if dur_q3=="03"| dur_q3=="04"

replace dur_q3 = "4" if dur_q3=="05"

replace dur_q3 = "5" if dur_q3=="06"| dur_q3=="07"| dur_q3=="08"

gen wgt=full_calwgt/1000

destring dur_q3, replace

***renaming variables

rename q13gender_q32013 gender

rename q14age_q32013 age

rename q15population_q32013 race

rename province_q32013 province

rename edu_q3 education

rename q16maritalstatus_q42013 mstatus

rename q37actpriorjobseek_q32013 prioractivity

rename q313timeunemploy_q42013 lastsincework

rename q312everwrk_q32013 experience

rename q319binhhpers_q32013 hhmember
```

```
rename q319cnothhpers_q32013 nhhmember

rename q319hgrants_q32013 grants

rename q319dcharity_q42013 other_1

**********

destring gender, replace

destring age, replace

destring race, replace

destring province, replace

destring education, replace

destring mstatus, replace

destring prioractivity, replace

destring lastsincework, replace

destring experience, replace

destring hhmember, replace

destring nhhmember, replace

destring grants, replace

****Generating new variables

gen enquire = .

replace enquire = 1 if q3202enquire_q32013=="1"

replace enquire = 2 if q3202enquire_q32013!="1"

gen jobads = .

replace jobads = 1 if q3203jobads_q32013=="1"

replace jobads = 2 if q3203jobads_q32013!="1"

gen internet = .

replace internet = 1 if q3204jobsearch_q32013=="1"

replace internet = 2 if q3204jobsearch_q32013!="1"
```

```
gen network = .

replace network = 1 if q3205assistance_q32013=="1"

replace network = 2 if q3205assistance_q32013!="1"

gen employed=.

replace employed = 1 if status_q42013=="1"

replace employed = 0 if status_q42013!="1"

gen mstatus2 =.

replace mstatus2 = 1 if mstatus==1| mstatus==2

replace mstatus2 = 2 if mstatus==3| mstatus==4| mstatus==5

gen age_group = .

replace age_group = 1 if age >=15 & age <=24

replace age_group = 2 if age >=25 & age <=34

replace age_group = 3 if age >=35 & age <=44

replace age_group = 4 if age >=45 & age <=54

replace age_group = 5 if age >=55 & age <=64

gen agegrp = .

replace agegrp = 1 if age >=15 & age <=34

replace agegrp = 2 if age >=35 & age <=64
```

*****Collapsing categories**

```
replace prioractivity = 2 if prioractivity ==4

replace education = 3 if
education==1|education==2|education==3|education==4|education==7

replace education = 2 if education==5

replace education = 1 if education==6

**************************

drop if wgt ==.
```

```
sort stratum_q32013 psuno_q32013

sort person_id

save  "C:\Users\zandileno\Desktop\unisa\dessertation\Assignment\wide_sort_stratum.dta", replace

use "C:\Users\zandileno\Desktop\unisa\dessertation\Assignment\Stratum_psu.dta", replace

gen id =_n

*now collapsed stratum data set and initial data set

drop psuno_q32013

drop stratum_q32013

drop psuno_q42013

drop stratum_q42013

sort person_id

merge 1:1 person_id  using "C:\Users\zandileno\Desktop\unisa\dessertation\Assignment\wide_sort_stratum.dta"

save  "C:\Users\zandileno\Desktop\unisa\dessertation\Assignment\atl_two_psu_str.dta", replace

use "C:\Users\zandileno\Desktop\unisa\dessertation\Assignment\atl_two_psu_str.dta", replace

gen dur_q4 = dur_q3

destring status_q42013, replace

replace dur_q4 = 2 if dur_q3==1 &status_q42013==1

replace dur_q4 = 3 if dur_q3==2 &status_q42013==1

replace dur_q4 = 4 if dur_q3==3 &status_q42013==1

replace dur_q4 = 5 if dur_q3==4 &status_q42013==1

svyset psuno [pweight=wgt], strata(stratum)

svydes

tab lastsincework
```

svyset psuno [pweight=wgt], strata(stratum)

stset dur_q4 [pweight=wgt], failure(employed=1)

*****collett's model selection approach**

delimit;

xi:svy: stcox i.gender i.agegrp i.race i.province i.education i.mstatus2 i.prioractivity i.experience

i.hhmember i.grants i.jobads i.network;

****

delimit;

xi:svy: stcox i.gender i.agegrp i.education i.mstatus2 i.prioractivity i.experience

i.hhmember i.jobads;

*****************Cox proportional model**

delimit;

xi:svy: stcox i.gender i.agegrp i.education i.mstatus2 i.prioractivity i.experience

i.hhmember i.jobads;

delimit cr;

*****proportional hazards test**

*****exclude internet and hhmember , ware insignificant on bivariate test**

use "C:\Users\zandileno\Desktop\unisa\dessertation\Assignment\atl_two_psu_str.dta", replace

gen dur_q4 = dur_q3

destring status_q42013, replace

replace dur_q4 = 2 if dur_q3==1 &status_q42013==1

replace dur_q4 = 3 if dur_q3==2 &status_q42013==1

replace dur_q4 = 4 if dur_q3==3 &status_q42013==1

replace dur_q4 = 5 if dur_q3==4 &status_q42013==1

svyset psuno [pweight=wgt], strata(stratum)

stset dur_q4 [pweight=wgt], failure(employed=1)

```
# delimit;
xi:stcox i.gender i.agegrp i.education i.mstatus2 i.prioractivity i.experience i.hhmember
i.jobads, schoenfeld(sch*) scaledsch(sca*);
# delimit cr;
estat phtest, detail
```

************fit the below model before the plots**

```
# delimit;
stcox gender agegrp education mstatus2 prioractivity experience hhmember
jobads, schoenfeld(sch*) scaledsch(sca*);
stphtest, plot(gender) msym(oh)
stphtest, plot(age) msym(oh)
stphtest, plot(race) msym(oh)
stphtest, plot(province) msym(oh)
stphtest, plot(agegrp) msym(oh)
stphtest, plot(education) msym(oh)
stphtest, plot(mstatus2) msym(oh)
stphtest, plot(prioractivity) msym(oh)
stphtest, plot(experience) msym(oh)
stphtest, plot(hhmember) msym(oh)
stphtest, plot(nhhmember) msym(oh)
stphtest, plot(grants) msym(oh)
stphtest, plot(enquire) msym(oh)
stphtest, plot(jobads) msym(oh)
stphtest, plot(network) msym(oh)
```

Appendix B

Quarterly Labour Force Survey (Questionnaire)

A. Particulars of the dwelling

Unique no.

A1. PSU number

A2. Dwelling unit Number

A3. Assignment number

A4. Survey 1 2 0 1 3

A5. Physical identification of the dwelling unit

A6. Telephone number for enumerated household (if any)

A7. Total number of persons in the household

A8. Total number of persons aged 15 years and above in the

A9. Questionnaire no. for this household (for person no. 01-10=1,

B. Households at selected dwelling unit

B1. Household number for this household

B2. Total number of households at selected dwelling unit

C. Response details

Visit no.	Date (actual)	Result code	Next visit (planned)
C1.			
C2.			
C3.			
C4.			

C5. FINAL RESULT

C6. Comments and full details for result code 2-11

RESULT CODES

01	**Completed**		
02	**Non-contact**	**0**	**Listing error**
03	**Refused**	**0**	**Demolished**
04	**Partly completed**	**0**	**Change of status**
05	**No usable information**	**1**	**Other non response**
06	**Vacant/unoccupied**	**1**	**Ended at question 1.2**

Comment in C6 giving full details for

d d m m y y y y

D. Field staff

D2. DSC **Assignment number** **Date**

D3. PQM **Assignment number** **Date**

SECTION 1 This section covers particulars of each person in the household
The following information must be obtained for every person who has stayed in this household for at least four nights on average per week during the last four weeks.
Do not forget babies. If there are more than 10 persons in the household, use a second questionnaire.

		Person number		
1.0	**Who is the <u>head</u> (or the <u>acting head)</u> of the household?** *(record that person in column 01)*			
1.1	**Record first name and surname** **First name:** **Surname:**			
1.2	**Has ..... stayed in this household for at least four nights on average per week during the last four weeks?** 1 = Yes 2 = NO → ***End of questions for this person***	1 2	1 2	1 2
1.3	**Is ..... a male or a female?** 1 = MALE 2 = FEMALE	1 2	1 2	1 2
1.4	**What is …….'s date of birth and age in completed years?** **Day (dd)** **Month (mm)** **Year (yyyy)** **Age** (less than 1 year = 000)			
1.5	**What population group does ..... belong to?** 1 = African/Black 2 = Coloured 3 = INDIAN/ASIAN 4 = WHITE 5 = OTHER, *specify in the box at the bottom*	**1** 2 **3** 4 **5**	**1** 2 **3** 4 **5**	**1** 2 **3** 4 **5**
1.6	**What is ….'s present marital status?** 1 = MARRIED 2 = Living together like husband and wife 3 = Widow/widower 4 = Divorced or separated 5 = Never married	**1** 2 **3** 4 **5**	**1** 2 **3** 4 **5**	**1** 2 **3** 4 **5**

1.7	**What is the highest level of education that …… has successfully completed?** 98 = NO SCHOOLING 00 = GRADE 0 01 = GRADE 1/ SUB A 02 = GRADE 2 / SUB B 03 = GRADE 3/STANDARD 1 / ABET 1 (KHA RI GUDE, SANLI) 04 = Grade 4/ STANDARD 2 05 = GRADE 5/ STANDARD 3/ABET 2 06 = GRADE 6/STANDARD 4 07 = GRADE 7/STANDARD 5/ABET 3 08 = GRADE 8/STANDARD 6/FORM 1 09 = Grade 9/Standard 7/Form 2/ABET 4 10 = GRADE 10/ STANDARD 8/ FORM 3 11 = GRADE 11/ STANDARD 9/ FORM 4 12 = GRADE 12/STANDARD 10/FORM 5 ***If code 98 or 00-12, Go to Q1.9*** 13 = NTC I/N1/ NIC/(V) LEVEL 2 14 = NTC II/N2/ NIC/(V) LEVEL 3 15 = NTC III/N3/ NIC/(V) LEVEL 4 16 = N4 /NTC 4 17 = N5/NTC5 18 = N6/NTC 6 19 = CERTIFICATE WITH LESS THAN GRADE 12/STD 10 20 = Diploma with less than Grade 12/Std 10 21 = Certificate with Grade 12/Std 10 22 = Diploma with Grade 12/Std 10 23 = Higher Diploma 24 = Post Higher Diploma (Masters, Doctoral Diploma) 25 = BACHELORS DEGREE 26 = BACHELORS DEGREE AND POST GRADUATE DIPLOMA 27 = HONOURS DEGREE 28 = HIGHER DEGREE (MASTERS/PHD) 29 = OTHER, *specify in the box at the bottom* ***If code 13-28, Go to Q1.8*** ***If code 29, Go to Q1.9*** ***Diploma or certificate should have been at least six months study duration full-time (or equivalent)*** ***Write the appropriate code in the boxes***			

1.8	**If diploma, certificate or degree (code 13-28 in Q1.7): In which field is ..... highest post-school qualification?** **UNIVERSITY/TECHNIKON/COLLEGE** 01 = AGRICULTURE OR RENEWABLE NATURAL RESOURCES 02 = ARCHITECTURE OR ENVIRONMENTAL DESIGN 03 = ARTS, VISUAL OR PERFORMING 04 = BUSINESS, COMMERCE OR MANAGEMENT SCIENCES 05 = COMMUNICATION 06 = COMPUTER SCIENCES 07 = EDUCATION,TRAINING OR DEVELOPMENT 08 = ENGINEERING OR ENGINEERING TECHNOLOGY 09 = HEALTH CARE OR HEALTH SCIENCES 10 = Home Economics 11 = INDUSTRIAL ARTS, TRADERS OR TECHNOLOGY 12 = LANGUAGES, LINGUISTICS OR LITERATURE 13 = LAW 14 = Libraries or Museums 15 = LIFE SCIENCES OR PHYSICAL SCIENCES 16 = MATHEMATICAL SCIENCES 17 = MILITARY SCIENCES 18 = PHILOSOPHY, RELIGION OR THEOLOGY 19 = PHYSICAL EDUCATION OR LEISURE 20 = Psychology 21 = PUBLIC ADMINISTRATION OR SOCIAL SERVICES 22 = SOCIAL SCIENCES OR SOCIAL STUDIES 23 = OTHER **FURTHER EDUCATION AND TRAINING (FET)** 24 = MANAGEMENT 25 = MARKETING 26 = INFORMATION TECHNOLOGY AND COMPUTER SCIENCE 27 = FINANCE, ECONOMICS AND ACCOUNTING 28 = OFFICE ADMINISTRATION 29 = ELECTRICAL INFRASTRUCTURE CONSTRUCTION 30 = CIVIL ENGINEERING AND BUILDING CONSTRUCTION 31 = ENGINEERING 32 = PRIMARY AGRICULTURE 33 = HOSPITALITY 34 = TOURISM 35 = SAFETY IN SOCIETY 36 = MECHATRONICS 37 = EDUCATION AND DEVELOPMENT 38 = OTHER			

		Person number		
1.9	**Does ... currently attend any educational institution?** 1 = Yes 2 = No → Go to Section 2	☐ 1 ☐ 2	☐ 1 ☐ 2	☐ 1 ☐ 2
1.10	**Which educational institution does … currently attend?** 1 = PRE-SCHOOL (INCLUDING DAY CARE, CRÈCHE, GRADE R AND PRE-GRADE R IN AN ECD CENTRE) 2 = ORDINARY SCHOOL (INCLUDING GRADE R LEARNERS WHO ATTEND A FORMAL SCHOOL, GRADE 1-12 LEARNERS AND LEARNERS IN SPECIAL CLASS) 3 = SPECIAL SCHOOL 4 = FURTHER EDUCATION AND TRAINING COLLEGE (FET) 5 = OTHER COLLEGE 6 = HIGHER EDUCATION INSTITUTION (UNIVERSITY O UNIVERSITY OF TECHNOLOGY) 7 = ADULT BASIC EDUCATION AND TRAINING CENTRE (ABET CENTRE) 8 = LITERACY CLASSES (E.G. KHAI RI GUDE, SANLI) 9 = HOME-BASED EDUCATION OR HOME SCHOOLING	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9

+ **INTERVIEW START TIME** ☐☐☐☐ **Person no.** ☐☐ **Age** ☐☐☐ +

SECTION 2

This section covers economic activities in the last week for persons aged 15 years and above

INDIVIDUAL **1**

2.0	*Interviewer to answer* ***Is the person him/herself responding to questions?*** 1 = YES → ***Go to Q 2.2*** 2 = NO	☐ 1 ☐ 2
2.1	*Give person number for the proxy respondent*	☐☐
2.2	**Do you have a landline or cellular telephone where you can be contacted?** 1 = YES 2 = NO → ***Go to Q 2.4*** 3 = DON'T KNOW	☐ 1 ☐ 2 ☐ 3
2.3	**May I please have a number where I can contact you at a later stage?**	
2.4	**In the last week (Monday to Sunday) ….**	**YES NO**
	(a) Did you work for a wage, salary, commission or any payment in kind (including paid domestic work), even if it was for only one hour? *Examples: a regular job, contract, casual or piece work for pay, work in exchange for food or housing, paid domestic work.*	☐ 1 ☐ 2
	(b) Did you run or do any kind of business, big or small, for yourself or with one or more partners, even if it was for only one hour? *Examples: Commercial farming, selling things, making things for sale, construction, repairing things, guarding cars, brewing beer, collecting wood or water for sale, hairdressing, crèche businesses, taxi or other transport business, having a legal or medical practice, performing in public, having a public phone shop, etc.*	☐ 1 ☐ 2
	(c) Did you help without being paid in any kind of business run by your household, even if it was for only one hour? *Examples: Commercial farming, help to sell things, make things for sale or exchange, doing the accounts, cleaning up for the business, etc.*	☐ 1 ☐ 2
	If yes to any part of Q 2.4 go to Section 4, otherwise go to Q 2.5	

2.5	**In the last week (Monday to Sunday), even though you did not do any work for pay, profit or did not help without pay in a household business, ….**	**YES NO**
	(a) Did you have a paid job that you would definitely return to? → ***If yes, go to Q 2.7, otherwise continue*** *Examples: a regular job, contract, casual or piece work for pay, work in exchange for food or housing, paid domestic work.*	☐ 1 ☐ 2
	(b) Did you have a business that you would definitely return to? → ***If yes, go to Q 2.7,*** ***otherwise continue*** *Examples: Commercial farming, selling things, making things for sale, construction, repairing things, guarding cars, brewing beer, collecting wood or water for sale, hairdressing, crèche businesses, taxi or other transport business, having a legal or medical practice, performing in public, having a public phone shop, etc.*	☐ 1 ☐ 2
	(c) Did you have an unpaid job in any kind of business run by your household that you would definitely return to? → ***Go to Q 3.1*** *Examples: Commercial farming, help to sell things, make things for sale or exchange, doing the accounts, cleaning up for the business, etc.*	☐ 1 ☐ 2

123456789101112

2.7	**What was the main reason you were absent from your job/business in the last week (Monday to Sunday)?** 01 = HEALTH REASONS 02 = VACATION LEAVE 03 = CARING FOR FAMILY OR OTHERS (EXCEPT MATERNITY / PATERNITY LEAVE) 04 = MATERNITY OR PATERNITY LEAVE 05 = OTHER FAMILY/COMMUNITY OBLIGATIONS (FUNERALS, MEETINGS) 06 = STRIKE / STAY-AWAY / LOCKOUT 07 = PROBLEMS WITH TRANSPORT 08 = BAD WEATHER 09 = STUDY OR TRAINING LEAVE 10 = UNREST (VIOLENCE) 11 = TEMPORARILY LAID OFF / REDUCTION IN ECONOMIC ACTIVITY 12 = SEASONAL WORK → ***Go to Q 3.1*** 13 = START A NEW JOB/BUSINESS AT A DEFINITE DATE IN THE FUTURE → ***Go to Q 3.1*** 14 = OTHER REASON, *specify* ***For all reasons, except options 12 & 13, go to Section 4***	☐ 01 ☐ 02 ☐ 03 ☐ 04 ☐ 05 ☐ 06 ☐ 07 ☐ 08 ☐ 09 ☐ 10 ☐ 11 ☐ 12 ☐ 13 ☐ 14

123456789101112
123456789101112

+ Person no. ☐☐ Age ☐☐☐ +

SECTION 3

This section covers unemployment and economic inactivity for persons aged 15 years and above

3.1	**In the last four weeks, ….** **a) Were you looking for any kind of work?** 1 = YES → ***Go to Q 3.2*** 2 = NO **b) Were you trying to start any kind of business?** 1 = YES 2 = NO → ***Go to Q 3.3***	☐ 1 ☐ 2 ☐ 1 ☐ 2
3.2	**In the last four weeks what have you done to search for work or to start a business?** *Mark all applicable options* 01 = WAITED/REGISTERED AT EMPLOYMENT AGENCY/TRADE UNION 02 = ENQUIRED AT WORKPLACES, FARMS, FACTORIES OR CALLED ON OTHER POSSIBLE EMPLOYERS 03 = PLACED/ANSWERED ADVERTISEMENT(S) 04 = SEARCHED THROUGH JOB ADVERTISEMENT(S) / SEARCHED THE INTERNET 05 = SOUGHT ASSISTANCE FROM RELATIVES OR FRIENDS 06 = LOOKED FOR LAND, BUILDING, EQUIPMENT OR APPLIED FOR PERMIT TO START OWN BUSINESS OR FARMING 07 = WAITED AT THE STREET SIDE WHERE CASUAL WORKERS ARE FOUND 08 = SOUGHT FINANCIAL ASSISTANCE TO LOOK FOR WORK OR START A BUSINESS 09 = OTHER, *specify* 10 = NOTHING → ***Go back to Q 3.1*** → ***Go to Q 3.6***	☐ 01 ☐ 02 ☐ 03 ☐ 04 ☐ 05 ☐ 06 ☐ 07 ☐ 08 ☐ 09 ☐ 10
3.3	**Was this because you had already arranged to take up a job or to start a business at some later date?** 1 = YES → ***Go to Q 3.6*** 2 = NO	☐ 1 ☐ 2
3.4	**Would you have liked to work last week (Monday to Sunday)?** 1 = YES → ***Go to Q 3.8*** 2 = NO	☐ 1 ☐ 2

3.5	**What was the main reason you did not want to work last (Monday to Sunday) week?** 1 = SCHOLAR OR STUDENT 2 = HOUSEWIFE/HOMEMAKER (FAMILY CONSIDERATIONS/CHILD CARE) 3 = HEALTH REASONS 4 = RETIRED OR TOO OLD TO WORK 5 = NO DESIRE TO WORK 6 = TOO YOUNG TO WORK 7 = PREGNANCY 8 = DISABLED OR UNABLE TO WORK 9 = OTHER, *specify* → ***Go to Q 3.12***	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9
3.6	**For how long have you been without work and trying to find a job or start a business?** 01 = LESS THAN 3 MONTHS 02 = 3 MONTHS – LESS THAN 6 MONTHS 03 = 6 MONTHS – LESS THAN 9 MONTHS 04 = 9 MONTHS – LESS THAN 1 YEAR 05 = 1 YEAR – LESS THAN 3 YEARS 06 = 3 YEARS – 5 YEARS 07 = MORE THAN 5 YEARS 08= DON'T KNOW	☐ 01 ☐ 02 ☐ 03 ☐ 04 ☐ 05 ☐ 06 ☐ 07 ☐ 08
3.7	**What was your main activity before you started looking for work?** 1 = WORKING 2 = MANAGING A HOME 3 = GOING TO SCHOOL 4 = OTHER, *specify* → ***Go to Q 3.9***	☐ 1 ☐ 2 ☐ 3 ☐ 4

123456789101112
123456789101112

3.8	**What was the main reason why you did not try to find work or start a business in the last four weeks?** 01 = AWAITING THE SEASON FOR WORK 02 = WAITING TO BE RECALLED TO FORMER JOB 03 = HEALTH REASONS 04 = PREGNANCY 05 = DISABLED OR UNABLE TO WORK (HANDICAPPED) 06 = HOUSEWIFE/HOMEMAKER (FAMILY CONSIDERATIONS/CHILD CARE) 07 = UNDERGOING TRAINING TO HELP FIND WORK 08 = NO JOBS AVAILABLE IN THE AREA 09 = LACK OF MONEY TO PAY FOR TRANSPORT TO LOOK FOR WORK 10 = UNABLE TO FIND WORK REQUIRING HIS/HER SKILLS 11 = LOST HOPE OF FINDING ANY KIND OF WORK 12 = NO TRANSPORT AVAILABLE 13 = SCHOLAR OR STUDENT 14 = RETIRED OR TOO OLD TO WORK 15 = TOO YOUNG TO WORK 16 = OTHER REASON, *specify*	☐ 01 ☐ 02 ☐ 03 ☐ 04 ☐ 05 ☐ 06 ☐ 07 ☐ 08 ☐ 09 ☐ 10 ☐ 11 ☐ 12 ☐ 13 ☐ 14 ☐ 15 ☐ 16
3.9	**If a suitable job had been offered, would you have been able to start work last week (Monday to Sunday)?** 1 = YES → ***Go to Q 3.12*** 2 = NO 3 = DON'T KNOW	☐ 1 ☐ 2 ☐ 3
3.10	**If circumstances had allowed, would you have started a business last week (Monday to Sunday)?** 1 = YES → ***Go to Q 3.12*** 2 = NO 3 = DON'T KNOW	☐ 1 ☐ 2 ☐ 3
3.11	**What was the main reason why you were not available for work last week (Monday to Sunday)?** 1 = SCHOLAR OR STUDENT 2 = HOUSEWIFE/HOMEMAKER (FAMILY CONSIDERATIONS/CHILD CARE) 3 = HEALTH REASONS 4 = RETIRED OR TOO OLD FOR WORK 5 = NO DESIRE TO WORK 6 = TOO YOUNG TO WORK 7 = PREGNANCY 8 = DISABLED OR UNABLE TO WORK 9 = OTHER, *specify*	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9

3.11.b	**How soon can you start work or a business?** 1 = WITHIN A WEEK 2 = WITHIN TWO WEEKS 3 = WITHIN FOUR WEEKS 4 = LATER THAN FOUR WEEKS FROM NOW 5 = NO DESIRE TO WORK 6 = DON'T KNOW	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6
3.12	**Have you ever worked for pay or profit or helped unpaid in a household business?** *Work could be:* *Formal work for salary, wage, profit or unpaid work in household business; informal work such as making things for sale; selling things or providing a service for payment; work on a farm or land for a wage or as part of the household's farming business* **1 = Yes** **2 = No** → ***Go to Q 3.19***	☐ 1 ☐ 2
3.13	**How long ago was it since you last worked?** 01 = LESS THAN 3 MONTHS 02 = 3 MONTHS – LESS THAN 6 MONTHS 03 = 6 MONTHS – LESS THAN 9 MONTHS 04 = 9 MONTHS – LESS THAN 1 YEAR 05 = 1 YEAR – LESS THAN 3 YEARS 06 = 3 YEARS – 5 YEARS 07 = MORE THAN 5 YEARS → ***Go to Q 3.19*** 08 = DON'T KNOW	☐ 01 ☐ 02 ☐ 03 ☐ 04 ☐ 05 ☐ 06 ☐ 07 ☐ 08
3.14	**What was the main reason you stopped working in your last job/business?** 01 = HEALTH REASONS 02 = CARING FOR OWN CHILDREN/RELATIVES 03 = PREGNANCY 04 = OTHER FAMILY/COMMUNITY RESPONSIBILITIES 05 = GOING TO SCHOOL 06 = LOST JOB/JOB ENDED /LAID OFF /BUSINESS SOLD/CLOSED DOWN. 07 = CHANGED RESIDENCE 08 = DISSATISFIED WITH THE JOB 09 = RETIRED 10 = OTHER, *specify*	☐ 01 ☐ 02 ☐ 03 ☐ 04 ☐ 05 ☐ 06 ☐ 07 ☐ 08 ☐ 09 ☐ 10

123456789101112
123456789101112

3.15.a	**What kind of work did you do in your last job/business?** *Work includes all the activities mentioned earlier* *Record at least two words: Car sales person, Office cleaner, Vegetable farmer, Primary School teacher, etc.*
3.15.b	**What were your main tasks or duties in this work?** *Examples: Selling fruit, repairing watches, keeping accounts, feeding and watering cattle, teaching children.*
CODE BOXES FOR OFFICE USE	
3.16.a	**What was the name of the establishment/ institution/ business/ organisation that worked for (the one that paid your salary)?** *For government or large organisations, give the name of the establishment and branch or division: e.g. Education Dept – Rapele Primary School; Harmony Gold Mining – Maintenance Div.* *For individuals who worked from home and their businesses don't have names write 'Own house'* *For individuals who worked in private households write "Private household"* *For individuals who worked for businesses without names write "No name"*
3.16.b	**What were the main goods or services produced at your place of work or its main functions?** *Examples: Repairing cars, Selling commercial real estate, Sell food wholesale to restaurants, Retail clothing shop, Manufacture electrical appliances, Bar/ restaurant, Primary Education, Delivering newspapers to homes. For domestic workers write "private household"*
CODE BOXES FOR OFFICE USE	

3.17	**In your last job/business, were you ….** 1 = Working for someone else for pay? (including paid domestic workers, gardeners or security guards) *Payment in cash or in kind (e.g. food, accommodation).* *Option 1 includes all employees: Full time, part-time, casual work and piecework.* 2 = An employer (employing one or more employees) 3 = Own-account worker (not employing any employees)? 4 = Helping without pay in a household business?	☐ 1 ☐ 2 ☐ 3 ☐ 4
3.19	**How do you support yourself?** *Mark all applicable options* a = Did you do odd jobs during the last week **(Monday to Sunday)**? ***→ If yes, go back to Q 2.4,*** ***otherwise continue*** b = Are you supported by persons in the household? c = Are you supported by persons not in the household? d = Are you supported by charity, church, welfare, etc.? e = Do you receive any money from unemployment Insurance Fund (UIF)? f = From savings or money previously earned? g = Do you receive old age or disability pension? h = Do you receive child support/ foster care grant? i = Do you receive any other welfare grants? j = Do you have other sources of support, e.g. bursary, study loan? ***→ Go to Q5.9***	**YES NO** ☐ 1 ☐ 2 ☐ 1 ☐ 2 ☐ 1 ☐ 2 ☐ 1 ☐ 2 ☐ 1 ☐ 2 ☐ 1 ☐ 2 ☐ 1 ☐ 2 ☐ 1 ☐ 2 ☐ 1 ☐ 2 ☐ 1 ☐ 2

123456789101112

123456789101112

SECTION 4

This section covers main work activity in the last week for persons aged 15 years and above

4.1	**In the last week (Monday to Sunday) did you have more than one job/business?** 1 = YES 2 = NO 3 = DON'T KNOW	1 2 3

If "yes", in Q4.1 read out: **The questions that follow refer to your main job/business. That is the one where you usually work the most hours per week, even if you were absent from it in the last week.**

4.2.a	**What kind of work do you usually do in the main job/business that you had during the last week (Monday to Sunday)?** *Work includes all the activities mentioned earlier* *Record at least two words: Car sales person, Office cleaner, Vegetable farmer, Primary school teacher, etc*
4.2.b	**What are your main tasks or duties in this work?** *Examples: Selling fruit, repairing watches, keeping accounts, feeding and watering cattle, teaching children*
CODE BOXES FOR OFFICE USE	

4.3.a	**What is the name of the establishment/ institution/ business/ organisation that you work for (the one that pays your salary)?** *For government or large organisations, give the name of the establishment and branch or division: e.g. Education Dept – Rapele Primary School; Harmony Gold Mining – Maintenance Div.* *For individuals who work from home and their businesses don't have names write 'Own house'* *For individuals working in private households write "Private household"* *For individuals who work for businesses without names write "No name"*
4.3.b	**What are the main goods or services produced at your place of work or its main functions?** *Examples: Repairing cars, Selling commercial real estate, Sell food wholesale to restaurants, Retail-clothing shop, Manufacture electrical appliances, Bar/ restaurant, Primary Education, Delivering newspapers to homes. For domestic workers write "private household"*
CODE BOXES FOR OFFICE USE	
4.4	**When did you start working for this employer or started running this business? Give year and month.** *State year in **four** figures, e.g. 2001* *State month in **two** figures, e.g. 08 for August*

123456789101112
123456789101112

4.5	**In the job/business that you had during last week (Monday to Sunday), were you …..** 1 = Working for someone else for pay? (including paid domestic workers, gardeners or security guards) *Payment in cash or in kind (e.g. food, accommodation).* *Option 1 includes all employees: Full-time, part-time, casual work and piecework.* → ***Go to Q 4.6*** 2 = An employer (employing one or more employees)? → ***Go to Q 4.13*** 3 = Own-account worker (not employing any employees)? → ***Go to Q 4.13*** 4 = Helping without pay in a household business? → ***Go to Q 4.13***	☐ 1 ☐ 2 ☐ 3 ☐ 4

For employees only (option 1 in Q 4.5)

4.6	**Does your employer contribute to any pension/retirement fund for you?** 1 = YES 2 = NO 3 = DON'T KNOW	☐ 1 ☐ 2 ☐ 3
4.7	**Are you entitled to any paid vacation leave?** 1 = YES 2 = NO 3 = DON'T KNOW	☐ 1 ☐ 2 ☐ 3
4.7.b	**Are you entitled to any …** 1 = paid sick leave? 2 = maternity / paternity leave?	**YES NO** ☐ 1 ☐ 2 ☐ 1 ☐ 2
4.7 c	**In the last twelve months, did you take.....** 1 = vacation leave? 2 = sick leave? 3 = maternity / paternity leave?	**YES NO** ☐ 1 ☐ 2 ☐ 1 ☐ 2 ☐ 1 ☐ 2
4.8	**Does your employer pay UIF contributions for you?** 1 = YES 2 = NO 3 = DON'T KNOW	☐ 1 ☐ 2 ☐ 3
4.9	**Are you entitled to medical aid benefits from your employer?** 1 = YES 2 = NO 3 = DON'T KNOW	☐ 1 ☐ 2 ☐ 3

4.10	**Does your employer deduct income tax (PAYE / SITE) from your salary/ wage?** 1 = YES 2 = NO 3 = DON'T KNOW	☐ 1 ☐ 2 ☐ 3
4.11	**Are you employed on the basis of …** 1 = A written contract? 2 = A verbal agreement?	☐ 1 ☐ 2
4.12	**Is the contract/agreement of a …..** 1 = Limited duration? 2 = Permanent nature? 3 = Unspecified duration?	☐ 1 ☐ 2 ☐ 3
4.12.b	**Are you a member of a trade union or other workers' organisation?** 1 = YES 2 = NO 3 = DON'T KNOW	☐ 1 ☐ 2 ☐ 3
4.12.c	**Who determines your annual salary increase?** 1 = NEGOTIATION BETWEEN MYSELF AND EMPLOYER AT COMPANY 2 = NEGOTIATION BETWEEN UNION AND EMPLOYER 3 = BARGAINING COUNCIL OR OTHER SECTOR BARGAINING ARRANGEMENT 4 = EMPLOYER ONLY 5 = NO REGULAR ANNUAL SALARY INCREASE 6 = OTHER, SPECIFY	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6

FOR EMPLOYERS, OWN ACCOUNT WORKERS AND PERSONS HELPING UNPAID IN HOUSEHOLD BUSINESSES (Options 2, 3 and 4 in Q 4.5)

4.13	**Is your business (or household business where you work) registered for VAT?** 1 = YES 2 = NO 3 = DON'T KNOW	☐ 1 ☐ 2 ☐ 3
4.14	**Is the business (or household business where you work) registered for income tax?** 1 = YES 2 = NO 3 = DON'T KNOW	☐ 1 ☐ 2 ☐ 3

123456789101112
123456789101112

4.14.a	**Does your business or the business where you work belong to any organisation/ association that protects your business interests?** 1 = YES 2 = NO 3 = DON'T KNOW	1 2 3

FOR ALL EMPLOYED PERSONS (employees, employers, own account workers and persons helping unpaid in household businesses)

4.15	**Is the institution/ establishment / business/ organisation you work for (the one that pays your salary) classified as.....** 1 = National/Provincial/Local government? 2 = Government controlled business (e.g. Eskom/Telkom) ? 3 = A private enterprise ? 4 = Non-profit organisation (NGO/CBO) ? 5 = A private household? 6 = DON'T KNOW	1 2 3 4 5 6
4.16	**How many employees are there at your place of work?** 01 = 0 02 = 1 03 = 2 – 4 04 = 5 – 9 05 = 10 – 19 06 = 20 – 49 07 = 50 OR MORE 08 = DON'T KNOW	01 02 03 04 05 06 07 08

FOR ALL EMPLOYED PERSONS
Ask for those with one job (Options 2 and 3 in Q4.1)

4.18	**How many hours do you <u>usually</u> work each week (Monday to Sunday)?**	
4.19	**Thinking of each day last week (Monday to Sunday), how many hours did you <u>actually</u> work ….**	
	Day	**Hours**
	On Monday?	
	On Tuesday?	
	On Wednesday?	
	On Thursday?	
	On Friday?	
	On Saturday?	
	On Sunday?	
	Total hours actually worked	
	→ ***Go to Q 4.22***	

Ask for those with more than one job (Option 1 in Q 4.1)

4.20	**How many hours do you <u>usually</u> work each week (Monday to Sunday)….**	
	1. In your first job/business?	
	2. In your second job/business?	
	3. In all other jobs/businesses?	
	Total hours for all jobs/businesses	

4.21	**Thinking of each day last week (Monday to Sunday), how many hours did you <u>actually</u> work ….**			
	Day	**Hours in …**		
		First job/ business	**Second job/ business**	**All other jobs/ businesses**
	Monday?			
	Tuesday?			
	Wednesday?			
	Thursday?			
	Friday?			
	Saturday?			
	Sunday?			
	Total hours			

123456789101112
123456789101112

4.22	**Last week (Monday to Sunday), would you have liked to work more hours than you actually worked, provided the extra hours had been paid?** 1 = YES, in the current job 2 = YES, in taking an additional job 3 = YES, in another job with more hours 4 = NO → ***Go to Section 5*** 5 = DON'T KNOW	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5
4.23	**How many additional hours could you have worked last week (Monday to Sunday)?**	☐☐

4.24	**Do you want to work more hours at your current rate of pay?** 1 = YES 2 = NO 3 = DON'T KNOW	☐ 1 ☐ 2 ☐ 3
4.25	**If extra work became available, would you be able to start such work in the next four weeks?** 1= YES 2= NO 3= DON'T KNOW	☐ 1 ☐ 2 ☐ 3

123456789101112
123456789101112

SECTION 5

This section covers earnings in the main job for employees, employers and own-account workers aged 15 years and above

5.1	***Copy response from Q4.5*** 1 = *Working for someone else for pay* → ***Go to Q 5.2*** 2 = *An employer (employing one or more employees)* → ***Go to Q 5.6*** 3 = *Own-account worker (not employing any employees)* → ***Go to Q 5.6*** 4 = *Helping without pay in a household business* → ***Go to Q5.9***	☐ 1 ☐ 2 ☐ 3 ☐ 4

FOR EMPLOYEES

5.2	**In your main job, what is the easiest way for you to tell us your wages or salary before taxes or any other deduction? Would it be …** 1 = Monthly? 2 = Weekly? 3 = Fortnightly (every two weeks)? 4 = Daily? 5 = Hourly? 6 = Annually? 7= REFUSED → ***Go to Q 5.8***	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7
5.3	**Do you usually receive any tips or commission?** 1 = YES 2 = NO	☐ 1 ☐ 2
5.4.a	**What is your (*choose one*) annual/ monthly / weekly / daily / hourly wage or salary before deductions?** *(Include tips and commissions)* **R** ☐☐☐☐☐☐☐☐ → ***Go Q5.9*** ***If amount not stated*** 1 = DON'T KNOW } → ***Go to Q 5.8*** 2 = REFUSED **5.4.b**	☐ 1 ☐ 2

FOR EMPLOYERS AND OWN-ACCOUNT WORKERS

5.6	**What is the easiest way for you to tell us your earnings after expenses? Would it be …** 1 = Monthly? 2 = Weekly? 3 = Fortnightly (every two weeks)? 4 = Daily? 5 = Hourly? 6 = Annually? 7= REFUSED → ***Go to Q 5.8***	☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7
57.a	**What are your (*choose one*) annual/ monthly / weekly / daily earnings after expenses?** **R** ☐☐☐☐☐☐☐☐ → ***Go to Q5.9***	
5.7.b	***If amount not stated*** 1 = DON'T KNOW } → ***Go to Q 5.8*** 2 = REFUSED	☐ 1 ☐ 2

123456789101112

For those who don't know, refused or prefer to provide ranges, please use the Prompt Card. Indicate earnings using the weekly, monthly or annual figures as indicated on the Prompt Card (do not forget to include tips and commission).

5.8	Weekly	Monthly	Annually	
01	NONE	NONE	NONE	01
02	R1 - R46	R1 - R200	R1 - R2 400	02
03	R47 - R115	R201 - R500	R2 401 - R6 000	03
04	R116 - R231	R501 – R1 000	R6 001 - R12 000	04
05	R232 - R346	R1 001 - R1 500	R12 001 - R18 000	05
06	R347 - R577	R1 501 - R2 500	R18 001 - R30 000	06
07	R578 - R808	R2 501 - R3 500	R30 001 - R42 000	07
08	R809 - R1 039	R3 501 - R4 500	R42 001 - R54 000	08
09	R1 040 - R1 386	R4 501 - R6 000	R54 001 - R72 000	09
10	R1 387 - R1 848	R6 001 - R8 000	R72 001 - R96 000	10
11	R1 849 - R2 540	R8 001 - R11 000	R96 001 - R132 000	11
12	R2 541 - R3 695	R11 001 - R16 000	R132 001 - R192 000	12
13	R3 696 - R6 928	R16 001 - R30 000	R192 001 - R360 000	13
14	R6 929 – R8 654	R30 001 – R37 500	R360 001 – R450 000	14
15	R8 655 – R12 500	R37 501 – R54 167	R450 001 – R650 000	15
16	R12 501– R14 423	R54 168 – R62 500	R650 001 – R750 000	16
17	R14 424 – R16 346	R62 501– R70 800	R750 001 – R850 000	17
18	R16 347 – R19 231	R70 801 – R83 300	R850 001 – R1 000 000	18
19	R19 232 OR MORE	R83 301 OR MORE	R1 000 001 OR MORE	19
20	DON'T KNOW	DON'T KNOW	DON'T KNOW	20
21	REFUSE	REFUSE	REFUSE	21

123456789101112

123456789101112

Ask for all persons aged 15 years and above

5.9	In the last week (Monday to Sunday), ….	YES NO
	(a1) Did you do any work on your own or the household's plot, farm, food garden, cattle post or kraal or help in growing farm produce or in looking after animals for the household's own consumption? *Examples: ploughing, harvesting, looking after livestock*.	1 2
	(a2) *If YES*, for how many hours?	
	(b1) Did you fetch water or collect wood/dung for household use?	1 2
	(b2) *If YES,* for how many hours?	
	(c1) Did you produce any other goods for household use? *Examples: clothing, furniture, clay pots, etc.*	1 2
	(c2) *If YES,* for how many hours?	
	(d1) Did you do any construction or major repair work on your own home, plot, cattle post or business or those of the household?	1 2
	(d2) *If YES*, for how many hours?	
	(e1) Did you catch any fish, prawns, shells, wild animals or other food for household consumption?	
	(e2) *If YES,* for how many hours?	