ECO5023W

Masters in Economics (by coursework and dissertation)

Jasmin Jakoet

JKTJAS001

This mini-dissertation entitled "*The Initial Unemployment Duration of Immigrants to Khayelitsha/Mitchell's Plain*" is completed in partial fulfilment of a Masters of Business Science in Economics.

Prepared by Jasmin Jakoet for Professor Murray Leibbrandt

DECLARATION

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.
2. I have used the Harvard Referencing convention for citation and referencing. Each contribution to, and quotation in, this mini dissertation entitled *"The Initial Unemployment Duration of Immigrants to Khayelitsha/Mitchell's Plain"* from the work(s) of other people has been attributed, and has been cited and referenced.
3. This mini dissertation entitled *"The Initial Unemployment Duration of Immigrants to Khayelitsha/Mitchell's Plain"* is my own work.
4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

Jasmin Jakoet

Signature Signed by candidate

The Initial Unemployment Duration of Immigrants to Khayelitsha/Mitchell's Plain

Jasmin Jakoet
University of Cape Town

Abstract

The well-established internal migration streams from the poor neighbouring provinces to Cape Town are driven by the perception of urban opportunities for employment and a better life. This paper serves to broaden the understanding of the assimilation of immigrants to the Cape Town labour market by means of a discrete-time duration analysis of the time taken to find work after arrival. Low exit rates from unemployment mean that the pool of long-term unemployed stays large and puts pressure on government resources that are required to try and prevent people from falling into abject poverty. The initial unemployment spell of migrants is found to exhibit positive duration dependence as the hazard function increases over time, albeit over long time periods. Estimators for the parameters of a discrete-time hazard model are obtained via maximum likelihood methods. They suggest that the reality of the post-migration urban experience depends on gender, the use of household social networks, proficiency in English, possession of a matric qualification and higher reservation wages.

Table of Contents

List of Tables

List of Figures

The Initial Unemployment Duration of Immigrants to Khayelitsha/Mitchell's Plain

1. Introduction

The Khayelitsha/Mitchell's Plain (KMP) 2000 Survey (SALDRU, University of Cape Town) reveals a large proportion of local immigrants living in the magisterial district of Mitchell's Plain in Cape Town, mostly originating from the Eastern Cape province that includes the former homelands, the Transkei and Ciskei. This migration stream was boosted with the abolition of rigid influx controls that regulated the movement of labour by the apartheid government in 1986 and has persisted because people seek a better quality of life in one of the richest provinces, given that they are from one of the poorest provinces in South Africa. The apartheid regime in South Africa ensured that migrants from neighbouring provinces are different to the local-born in Cape Town in terms of all socio-economic bases[1]. Table 1 portrays a vast migrant community in KMP by displaying the period of arrival and the source region for migrant respondents of the individual level KMP 2000 Survey. As can be seen in the table, migration occurred despite influx controls during the apartheid period and squatting in black African Group Areas has resulted in the predominantly black African informal settlements within KMP.

[1] In 1955 a Coloured Labour Preference Policy was introduced that designated the western part of the Cape Province as an area that would give preferential treatment to coloured labour over black labour. This policy was retracted in 1984. (Goldin, 1987: 2).

Table 1: Arrival Cohort by Birthplace of Migrants

Place of Birth	Arrival Cohort						
	1900-69	1970-79	1980-84	1985-89	1990-94	1995-2001	Total
Other areas in Western Cape	34	27	11	6	2	42	122
Ciskei	9	13	23	47	46	80	218
Transkei	42	112	108	187	225	336	1010
Other areas in Eastern Cape	21	25	21	37	48	59	211
Other areas in South Africa	21	14	14	14	22	25	110
Areas outside South Africa	0	1	0	2	0	3	6
Total	127	192	177	293	343	545	1677

Source: Own calculations using KMP data

It is well known that migration to Cape Town from neighbouring provinces is primarily driven by the need for employment and infrastructure. (Bekker, 2002: 3). While the theory of earnings assimilation of immigrants predicts that an initial wage disadvantage dissipates with time spent in the destination labour market, it does not address the process of initial job search, nor the time taken to find employment in the first instance. In addition, the working poor experience low mobility between occupational skill levels that does not facilitate wage increases (Jakoet, 2006: 20). Nattrass (2002: 15) reports the broad unemployment rate to be 46,6 percent in her analysis using the KMP 2000 Survey. Given the depth and breadth of unemployment in the region, informal employment activities are surprisingly modest. Nattrass (*ibid*: 8) finds that people mainly use self-employment and casual jobs to supplement their main wage incomes. Low probabilities of finding employment and low mobility between occupational skill levels suggest that the initial wage job is a crucial accomplishment in accessing the labour market in order to earn a steady income. The systems by which jobs are attained are therefore important.

Evidence that social networks govern labour market links in KMP are well-documented in a paper by Schöer & Leibbrandt (2006), where the use of social networks has been identified as the most popular method of job search for finding wage employment. They (*ibid*: 703) report that two thirds of respondents that were employed at the time of the survey found their jobs using social networks and that the unemployed used search methods that were similar to those with jobs. Kingdon & Knight (2000: 9) state that in places of high unemployment it would not be surprising to find more passive than active job search as the most used method. For KMP residents, Nattrass (2002: 17) finds that seven percent of the total sample is exclusive network-searching unemployed (unemployed individuals who rely solely on networks consisting of friends and relatives to find them work).

The existence of well-developed social networks between the destination region and the main sending areas suggest that location-specific human capital that could hamper progress in terms of labour market assimilation is reduced. This implies that migrants moving into a household with an employed household member should experience shorter durations of unemployment. For those with fewer, weaker, or non-existent social networks, the duration of an episode of unemployment is expected to be longer because they are excluded from the referral mechanism from employed friends and family. Unfortunately, the employment status of household members in the host household at the time of arrival of immigrants is not available in the survey and the concept of network quality cannot be applied in this paper. However, Schöer & Leibbrandt (2006: 720) mention that the local-born unemployed are more likely to rely on social networks exclusively (relative to other search methods) because migrants have to first build their social networks. This implies that migrants are more likely to follow mixed strategies in their search for jobs. Schöer & Leibbrandt (*ibid*: 720) show empirically that the probability of searching either in an exclusively active manner, or with a mixed strategy, is lower for the local-born even though the effects are weakly significant.

An important purpose of the KMP Survey was to capture all the labour market activities of respondents. (Survey Report and Baseline Information: 2). The focus on labour market behaviour provides useful information on job search media for wage employment. It asks the pertinent question: "How did you get this job?" for the first ever wage job, the most recent prior wage job and for the wage job at the time of the survey; and provides a realistic array of possible answers, as detailed in the first column of Table 2. For those that were unemployed at the time of the survey it asks: "What is the best way for someone with your skills and experience to find a job?" and a similar table of answers apply by adjusting the phrasing slightly. For example, for those that found employment through a household member that "told them about the job", the response for the unemployed is adjusted to a preference for "relying on household members to tell them about jobs". These questions have allowed the construction of Table 2 below, where migrants' job search media for their first jobs in Cape Town (whether it is their first ever wage employment or not) are revealed. The second column in Table 2 displays the sample size of migrants that found employment within their first year of arrival in Cape Town for each job search method, while the third column gives the same information for migrants that found wage employment in Cape Town in the period after one year since their arrival. The fourth column shows the job search preferences of unemployed migrants that were never formally employed in Cape Town.

Table 2 shows that both the migrants that found wage jobs within the first year of arrival and those that found wage jobs after one year made use of social networks to get their first jobs. The respondents that did not find employment by the time of the survey expressed their preferred method of search to be this more passive approach. From the table it becomes evident that migrants are able to secure jobs mainly by accessing information about available job opportunities from friends or relatives in other households. Social networks do not assist much more than this, as can be seen by the relatively few migrants actually obtaining employment at the workplaces of their friends and relatives. There are a fair amount of migrants that got jobs by visiting factories. Migrants that did not find employment in Cape Town by the time of the survey also prefer more mixed strategies and, in particular, the more anonymous strategy of responding to newspaper advertisements.

Table 2: Job Search Strategies Employed by Migrants

Successful search method for first job in CT / preferred method if no job in Cape Town yet	Found job within first year	Found job after first year	Did not find job yet	Total
Responded to a newspaper advertisement	8	4	113	125
Household (HH) member told me about job	32	26	38	96
Friend/relative, in a different HH, told me	84	76	46	206
HH member got me job at their workplace	10	13	13	36
Friend/rel.,diff. HH, got me job at their work	17	19	24	60
Went to a factory and waited outside	34	36	54	124
Knocked on factory gates and visited homes	23	18	47	88
Got job through an employment agency	5	9	10	24
Asked previous employer	1	1	5	7
Waited on the side of the road	3	2	1	6
Looking on notice boards	0	0	4	4
Other	9	10	20	39
Total	226	214	375	815

Source: Own calculations using KMP data

The existence of an employed household member increases the probability of finding employment because labour market information may be passed to the unemployed household members. (Schöer & Leibbrandt, 2006: 709). When jobs are attained through a household member, jobs are therefore expected to be found faster than with other network job search methods. Serneels (2004: 31) also notes the importance of social capital in reducing the length of unemployment spells. In describing the unemployment of urban men in Ethiopia he finds evidence that the use of social networks is the most common method of job search, suggesting initially that this may render job search more passive; thereby increasing unemployment duration (*ibid*: 5, 9). Using a probit regression to determine the importance of the job search medium on the incidence of unemployment he shows that job search through social networks has no significant effect (*ibid*: 20). However, the duration analysis proves empirically that using social capital shortens the unemployment duration of those that are unemployed.

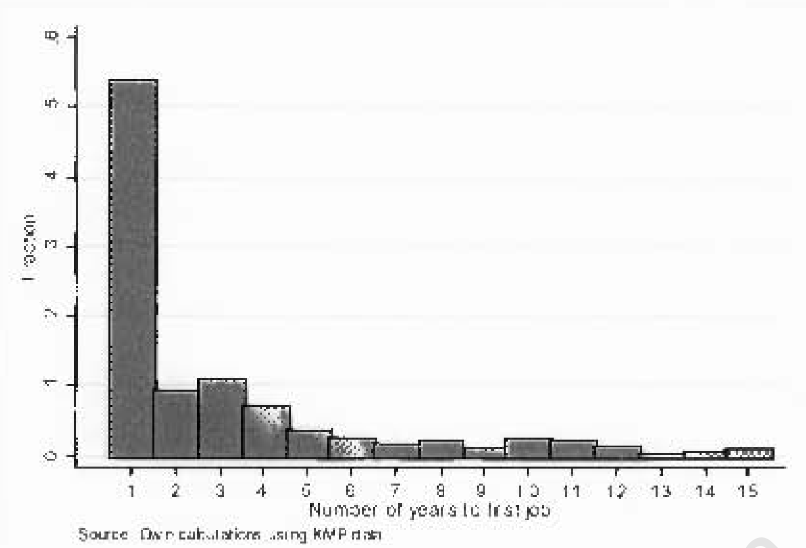
Migrants presumably have existing social networks in Cape Town prior to migration, given the history of strong migration streams from the same towns and villages in the neighbouring provinces over time (Ndegwa, D. Horner, D. & Esau, F. 2004: 20). It is therefore not surprising that there is considerable reliance on social networks; especially since many of the migrants share the same language (Xhosa-speaking people originate from the former homelands of Transkei and Ciskei). As was previously suggested, these social connections should assist migrants in terms of accommodation, access to the labour market upon arrival and general coping strategies, depending on the quality of the network. However, longer durations of unemployment may also be a financial burden on the host household. In addition, longer unemployment spells imply a loss of employed contacts in the job market that presents obstacles to new opportunities for the long-term unemployed.

To understand the post-migration Cape Town experience in this context, it is imperative to examine the initial unemployment of migrants, even though the unemployment durations of all KMP residents has been analysed in a paper by Mlatsheni & Brick (2006) using continuous-time duration modelling techniques. Although their paper provides a valuable starting point to the work on the initial urban unemployment durations of migrants, discrete-time methods are implemented in this paper as the data are observed in discrete units for many respondents.

Even though migration is initiated by the need for employment and migration theories usually attach a positive selection effect to migrants, it appears that migrants stagnate in long spells of unemployment before finding their first jobs in Cape Town. Indeed, Schöer & Leibbrandt (2006: 705) assert that the act of job search requires resources and is not "frictionless or costless", regardless of the level of enthusiasm of the individual. Figure 1 displays the distribution for the duration of initial unemployment that has a long right tail (after dropping quite sharply after the first year of arrival) for those migrants that have been employed in Cape Town. These longer spells of unemployment hint at a discouraged worker[2] effect. The distribution is thus not constant over the period but requires further analysis before declaring the process duration dependent. Given the high unemployment rate in South Africa, and the vast pool of discouraged workers, analysing links to the labour market is important.

[2] Discouraged unemployed are defined as those that are out of work and available for work but making no effort to find work.

Figure 1: Distribution of the Duration of Initial Unemployment Duration



The purpose of this paper is to delve into the complex story of labour market success of migrants in order to contribute to better knowledge on the topic. Using the 2000 KMP Survey, survival analysis estimation methods for discrete data are applied to identify factors that alter the initial duration of unemployment. The data are described in the next section and the theoretical framework and an initial analysis follows. Estimation of the models is then performed. The final section concludes and highlights some policy implications. The appendices detail technical information that supplements the discussion.

2. Data Description

Rich individual and household level[3] data from the first wave of the KMP 2000 Survey are employed. A random sample of adults (aged 18 years and older) from the magisterial district of Mitchell's Plain in Cape Town responded to the questionnaire, resulting in a total sample size of 2644 adults. The empirical investigation features 1682 migrants, of

3 The household questionnaire provided the total monthly household income variable and the size of the household.

distribution of time taken to find the first job after arrival would violate this assumption because durations are always positive and their distributions are often skewed.

Continuous-time models are generally used in discussions of unemployment duration because these data are available monthly, or even weekly, in more developed countries (Serneels, 2004: 2). The rich KMP dataset has been used in a similar fashion by Mlatsheni & Brick (2006) but for the purposes of this paper, simpler discrete-time methods are necessary due to data constraints, and this involves standard maximum likelihood estimation techniques.

The hazard probability is a fundamental statistic of duration analysis because it is the probability that an individual will experience an event in period t whilst he/she is at risk of having an event (i.e. given that no event has occurred in a previous interval) and is described as follows[5]:

$$h_{ti} = \Pr(y_{ti} = 1 \mid y_{si} = 0, s<t)$$

In terms of unemployment duration, it is the probability that an unemployed person will find a job in period t, whereas the survival probability refers to the proportion of respondents that still have not found employment at the end of period t. Therefore, the higher the hazard in a time period, the greater is the possibility of finding work. Although in the unemployment duration sense this is counterintuitive, the survival analysis terminology will be adhered to. The distribution of event occurrence over time for the entire sample of migrants is described by the survivor and hazard functions below and was derived using lifetable methods[6]. The Kaplan-Meier survivor functions plot the number of people that have not found employment at each year after arrival.

---

[5] Steele, F (2005:14).

[6] Jenkins, S. (Lesson 4. EC968. Part 2: Introduction to the Analysis of Spell Duration Data) illustrates how to use Stata to estimate the survivor, failure and cumulative hazard functions using the lifetable method for discrete (grouped) data.

Figure 2a: Kaplan-Meier Survivor Function for Entire Sample of Migrants

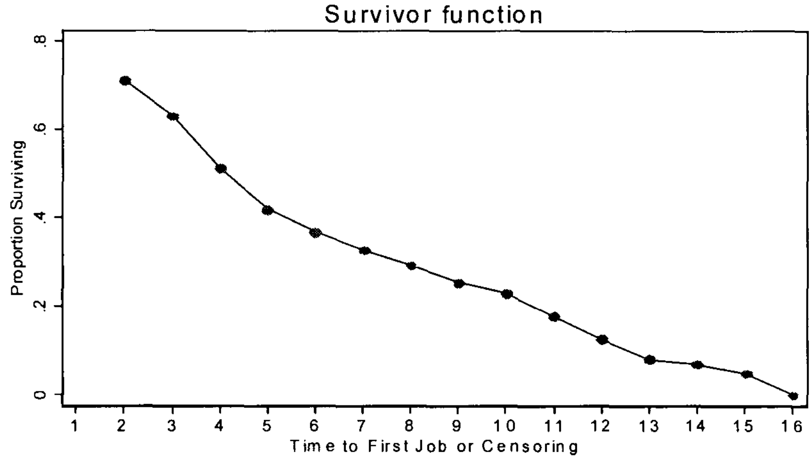


Table 2b: Kaplan-Meier Survivor Function for Entire Sample of Migrants

Interval	Beg. Total	Found Jobs	Censored	Survival	Std. Error	[95% Conf. Int.]
1 2	1114	319	313	0.71	0.014	0.6861 0.7392
2 3	482	55	92	0.63	0.016	0.6003 0.6623
3 4	335	64	41	0.51	0.019	0.4742 0.5473
4 5	230	41	20	0.42	0.020	0.3808 0.4592
5 6	169	21	12	0.37	0.021	0.3279 0.4082
6 7	136	15	14	0.33	0.021	0.2871 0.3683
7 8	107	11	6	0.29	0.021	0.2533 0.3353
8 9	90	12	5	0.25	0.021	0.2145 0.2966
9 10	73	7	2	0.23	0.021	0.1905 0.2722
10 11	64	15	8	0.18	0.020	0.1388 0.2174
11 12	41	12	4	0.12	0.019	0.0905 0.1646
12 13	25	9	3	0.08	0.017	0.0506 0.1174
13 14	13	2	1	0.07	0.017	0.0400 0.1047
14 15	10	3	1	0.05	0.015	0.0235 0.0833
15 16	6	6	0	0.00	.	. .

The first column in Table 2b refers to yearly intervals because the method of estimation makes the valid assumption of underlying continuous data that are only available in grouped form. Although the intervals are labelled to start from one, this should not be confused with the actual year of duration, as the first interval is in fact from time of arrival in Cape Town to the end of the first year in Cape Town. The second column is the total number of respondents that were at risk of failure (respondents that were looking for their first jobs in Cape Town), while the third column shows the number finding employment at each interval. The fourth column reveals the numbers that are censored which 1591 remain after respondents that were not of working age at arrival were discarded[4]. The KMP dataset, although quite informative on the employment histories of respondents, lacks in the necessary question, "how long were you looking for work?" for the first (or recent) wage jobs and only asks respondents how long they searched for their current wage employment. The variable of interest was therefore constructed to enable the analysis. However, the analysis is based on retrospective data and is subject to reliability problems, causing a further 460 migrants to be dropped due to missing data on the main variable. By examining Table A5 in the appendices, there appears to be no worthwhile evidence to suggest that selection bias has resulted from the exclusion of these respondents.

The measure of unemployment duration is *year of first job* less the *year of arrival* (*or year of final full-time education in Cape Town*). The year of the first job is either the starting date of the first wage employment in Cape Town or the first year of self-employment in Cape Town, whichever came first. Although treating these types of employment as though they are equivalent is not ideal, there are too few respondents that got involved in self-employment first to warrant a comparative analysis. For respondents that never ever got a job in Cape Town, the unemployment duration variable is available monthly and these respondents are treated as right censored because they did not find employment by the time of the survey.

The calculation assumes that initial entry (arrival in Cape Town) equals labour market entry and this may not be ideal. Thapa & Gorgens (2006: 6) mitigate this anomaly by restricting their sample to those of working age at initial entry and the same is done here. Also, a few unemployment spells greater than 15 years were dropped in the analysis, as they most probably reflect non-labour force participation in some or all of the years of living in Cape Town. Even after taking these ideas into account, the problem that periods of non-participation in the years after arrival are present still plagues the analysis. For example, a woman may fall pregnant during her second year of search and only re-enter

4 Male respondents that arrived when they were 65 years or older and females that arrived when they were 60 years or older were discarded from the analysis as they were not of working-age.

the labour market in the fourth year, thereby distorting the interval calculation. Full-time education is considered in the calculation but any other absence from the labour market from the time of arrival to the first job cannot be accounted for. This is unfortunate, considering the long and established patterns of temporary (circular) rural-urban migration that prevails amongst migrants in South Africa. (Bekker, 1999: 1). An interesting observation by Thapa & Gorgens (*ibid*: 6) is that many migrants could be the spouses or children of previous migrants. This should reduce the time before finding a job if the aim is indeed to find work. Where the spell is calculated, the time since arrival may be a crude indicator of time spent searching for employment as duration of unemployment is not merely a function of time but rather a function of the information collected throughout that time period. This information (the intensity of search for the first job) is not available in the dataset but the high prevalence of network searchers discussed earlier suggests less active search for a majority of respondents.

Labour market conditions in Cape Town as well as the quality of immigrants may have been different for different cohorts of arrival and this cannot be described adequately with duration spells taken across different time periods. Borjas (1984) warns that these effects may bias results when using a single cross-section of data. Chiswick, Cohen and Zach (1997) use data from three different sources over a 10-year period in their analysis of unemployment ratios and the labour market status of immigrants to minimise these biases. The pooling of data over a long time span in this paper may not be ideal and the year of arrival is included in the estimation to control for these effects. This is noteworthy because the mean initial unemployment duration appears to decrease with later arrival cohorts, as can be seen in the appendices (Table A2).

Discrete-time methods are used, whereby data are organised in "person-year" format so that standard regression techniques may be applied. The numbers of observations are therefore inflated with the number of spell years at risk per person, to create an unbalanced data panel effect. Allison (1982: 82) supports this convenient method, explaining that the estimates and standard errors of the estimates satisfy the necessary properties for maximum likelihood estimation. Where the discrete-time approach is used in duration literature, time-varying covariates are usually included because of the ease of incorporation. Only time-invariant covariates are included here due to data constraints.

Table A1 in the appendices provides some descriptive statistics for the variables used in the unemployment duration analysis and distinguishes between those that have spells of less than one year and those that experienced longer spells, based on the histogram result in Figure 1. Those with shorter spells have more education, larger households in terms of income and size, have greater proficiency in English, comprise of fewer people with rural backgrounds, and are better skilled than those who experience longer spells. Table A2 displays the mean initial unemployment duration by arrival cohort. As mentioned previously, those that arrived earlier have longer spells of initial unemployment. Table A3 summarises the years of education instead of the initial duration of unemployment to show that the unemployment spells are not diminishing because of the higher levels of education of the later cohorts. Table A4 then shows that those that completed their education in Cape Town have lower unemployment spells.

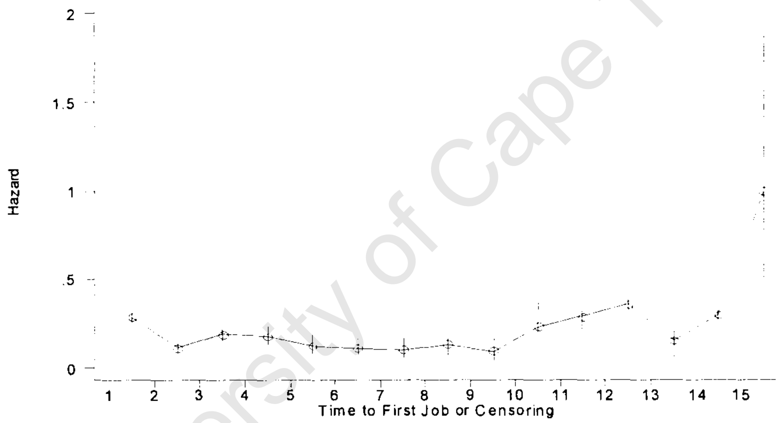
3. Theoretical Framework and Initial Analysis

Application of Discrete-time Survival Analysis to Unemployment Duration

The application of survival analysis techniques to social science issues that involve analysing the time to an event, such as the duration of unemployment, have proliferated because they are able to handle censoring and non-normality in data with ease. Censored observations are defined as those with incomplete information, or observations where the event was not experienced in the time before data collection ceased (Cleves, Gould & Gutierrez: 29). The existence of right-censored observations complicates event analysis because the exclusion of these observations would render shorter mean durations and may reduce the sample size considerably. Even though solutions for censoring can be solved in order to do ordinary least squares (OLS) estimation, the OLS assumption of normality of the residuals is unfounded. (Cleves, *ibid*: 2). It is highly likely that the and are therefore no longer part of the risk set. They are no longer at risk of having an event because they were still unemployed when data collection ended. They are right-censored because we do not know when they actually got employment. The survival function[7] is telling of the length of unemployment as 51% of the sample under consideration had not found employment by the end of their third year in Cape Town.

The hazard function (Figure 3) shows the hazard rate at each duration interval and the increasing probability of finding employment with more time spent in the labour market. For those intervals without any failures, the hazard is equal to zero. This confirms very long initial unemployment spells. The vertical lines show the 95% confidence intervals for the estimates.

Figure 3a: Interval Hazard Function for Entire Sample of Migrants



[7] A simple example for the second row will help to explain the calculation for column 5. The survival rate of 0.63 in the second row is found by: [Beginning total (number at risk = 482) – those that found jobs (55)]/482 and then multiplying this by 0.71. This is the Kaplan-Meier method.

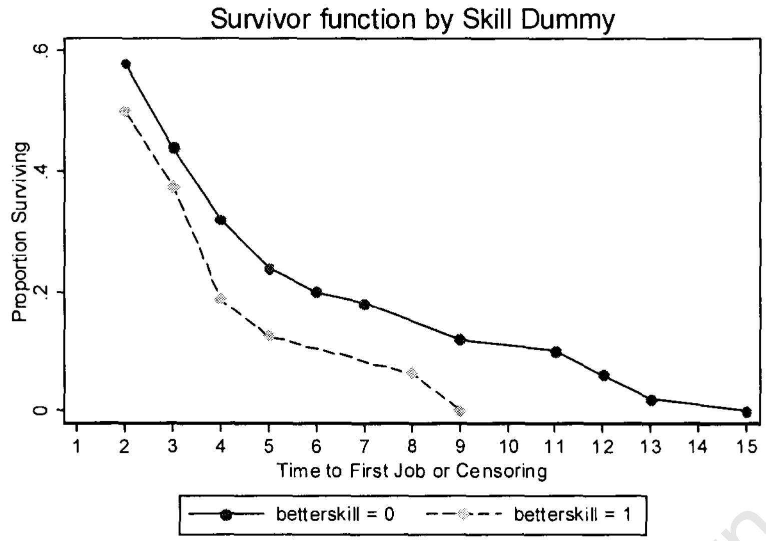
Table 3b: Interval Hazard Function for Entire Sample of Migrants

Interval	Beg. Total	Cum. Failure	Std. Error	Hazard	Std. Error	[95% Conf. Int.]	
1 2	1114	0.28640	.014	0.29	0.016	0.2558	0.3186
2 3	482	0.36780	.016	0.11	0.015	0.0860	0.1462
3 4	335	0.48860	.019	0.19	0.024	0.1471	0.2406
4 5	230	0.57970	.020	0.18	0.028	0.1279	0.2368
5 6	169	0.632	0.021	0.12	0.027	0.0769	0.1828
6 7	136	0.67260	.021	0.11	0.029	0.0617	0.1727
7 8	107	0.70620	.021	0.10	0.031	0.0513	0.1719
8 9	90	0.74540	.021	0.13	0.039	0.0689	0.2187
9 10	73	0.76980	.021	0.10	0.036	0.0386	0.1789
10 11	64	0.82380	.020	0.23	0.061	0.1312	0.3670
11 12	41	0.87530	.019	0.29	0.085	0.1512	0.4800
12 13	25	0.92020	.017	0.36	0.120	0.1646	0.6305
13 14	13	0.93250	.017	0.15	0.109	0.0186	0.4286
14 15	10	0.95270	.015	0.30	0.173	0.0619	0.7225
15 166		1	.	1.00	0.408	0.3670	1.9447

If the probability of getting a job is decreasing with time spent in the labour market (as a hazard for a discouraged worker effect would be), duration dependence is negative and the hazard is falling with time. The hazard function appears to be rising over longer periods of time but not uniformly.

The following analysis uses labour–market characteristics of migrants to ascertain which of these characteristics yield different probabilities of employment over time. Figure 4 shows that survival times fall faster for those that currently work in more skilled occupations. Given the low mobility between occupational skill levels (Jakoet, 2006: 20), this result implies that migrants finding initial jobs in occupations at a higher skill level are absorbed into employment faster than those that find initial jobs in low skill occupations. The corresponding Wilcoxon and Log-rank tests for the statistical significance of the differences between survivor functions are shown below the graph. The null hypothesis that there are no subgroup differences in survivor functions is rejected, as the chi-squared values are sufficiently large (the corresponding p-values are sufficiently small).

Figure 4: Survival Function Estimates Stratified by Occupational Skill Level



Likelihood-ratio test statistic of homogeneity (group=betterskill):
chi2(1) = 1.4654541, P = .22606431

Logrank test of homogeneity (group=betterskill):
Log-rank test for equality of survivor functions

betterskill	Events observed	Events expected	
0	50	56.99	chi2(1) = 7.18
1	16	9.01	Pr>chi2 = 0.0074
Total	66	66.00	

A number of variables are tested in this way in order to inform the decision of whether to include them in the multivariate models. A few relevant graphs are presented in the appendices (Figures A1–A3). Survival functions are found to differ slightly by gender, falling faster for males than for females. In addition to explaining this outcome using the realities of labour market discrimination towards females and the traditional roles for women in the household, we are reminded of the remark by Thapa & Gorgens (2006: 6) that many migrants may be the spouses of previous migrants. This implies less urgency in finding the first job. Figure A2 shows that there are significant subgroup differences in survival functions between those that found their first jobs using social networks and those that did not, while finishing school in Cape Town is shown to separate survival functions in Figure A3. Further examination using multivariate models for an explanation of initial unemployment duration follows in section 4.

A series of functional forms for continuous-time hazard functions are available in the continuous-time case (including Weibull, Log-logistic, Exponential, Gompertz, Lognormal and Generalised Gamma) but there are only two predominant functional forms for the hazard in the discrete-time case, namely the logistic and the complementary log-log (cloglog) forms (Jenkins, S. *op cit*, Lesson 2).

The cloglog model corresponds with an underlying continuous-time Proportional Hazards (PH) model. The logistic model is not a PH model but approximates one if the hazard is sufficiently small. Following Jenkins (*op cit* :11) the logistic and the complementary log-log forms are specified as follows:

Logistic discrete-time hazard function: $p(t)$, where

$$p(t) = [1+\exp(-z(t))]^{-1}$$

Complementary log-log discrete-time hazard function: $p(t)$, where

$$p(t) = 1-\exp[-\exp(z(t))]$$

The baseline hazard is contained within $z(t)$ along with covariates. The baseline hazard function is the hazard that everybody faces modified by covariates. When the hazard rate that an individual faces is multiplicatively proportional to the baseline hazard we call this a proportional hazards (PH) model. For the logistic model, points on the hazard function represent the conditional log-odds of event occurrence in those time periods and the proportional hazards assumption states that the antilogs belonging to each pair of covariates are "magnifications or diminutions of one another". (Singer & Willett, 1993: 185). The authors (*ibid*, 186) also note that violations of the proportionality assumption are the norm, based on studies covering the analysis of different phenomena. Tests for the proportional hazards assumption involve graphical checks using the fitted hazard models stratified by the covariate pairs. A constant vertical separation between the functions implies that the proportional hazards assumption is satisfied. Two relevant graphs are therefore included in the appendices (Figures A4 and A5) and appear to satisfy the assumption. The test is subjective and graphs for all covariates are therefore not shown.

Before proceeding to estimation of models, the functional form of the baseline hazard function must be selected based on an inspection of the hazard function. For the discrete-time baseline hazard, Jenkins (*op cit*: 12) assumes a form that corresponds to the Weibull model.

$$c(t) = (q-1).\ln(t)$$

The logistic and cloglog forms are simplified with the substitution of the baseline hazard function.

Logistic discrete-time hazard function simplifies to:

$$p(t) = [1 + \lambda^{-1}t^{1-q}]^{-1}$$

Complementary log-log discrete-time hazard function simplifies to:

$$p(t) = 1 - \exp[-\lambda t^{q-1}]$$

Singer & Willett (1993: 176) recommend that all estimations begin with a flexible baseline hazard that only consists of all the time-period dummy variables and no intercept term before attempting other representations of time. This provides an introductory analysis of what the risk of event occurrence is in each time period. This initial model is estimated and the results are included in the appendices (Model A1), along with the more parsimonious piece-wise constant model (Model A2) that groups consecutive time periods together according to the similarity in coefficient size. The estimated coefficients on the duration dummy variables generally rises in magnitude as duration increases, with the exception of some spells (especially those duration spells greater than 10 years). This confirms a rising hazard over time and therefore, positive duration dependence. Each coefficient corresponds with the population hazard probability in each time period (Singer & Willett, 1993: 177) and reveals the patterns of the overall population hazard function. Parameterising using dummy indicators for time lacks efficiency and it is worthwhile representing time in a different way, such as the log of time, or linear and polynomial representations. These models are included in the multivariate estimation in section 4.

Because the hazard is small, the differences between the cloglog and the logistic models are small (Jenkins, op cit: 10; and Allison, 1982: 73) and the focus is therefore solely on the cloglog estimation from here onwards. Rodriguez (Lecture Notes: 7.6) explains that cloglog models are preferred when data are inherently continuous but are only observed in grouped form. A logistic (logit) model is included nevertheless for the purposes of comparison.

4. Model Estimation

Predicting the occurrence of finding the first job will provide answers as to why some migrants find work early and others find their first jobs late. Three multivariate cloglog models and one logit model is displayed below. Covariates include dummy variables for gender (*male*), whether household social networks were used in finding the first job (*hhsocnet*), the number of adults in the household (*adulthhsize*), household total monthly income (*hhtotminc*), English proficiency (*engprof*), whether the person had a rural-based education (*ruraledn*) and whether the individual held a matric qualification (*matric*). A question prompting the reservation wage of individuals allowed the inclusion of a variable indicating a respondent's preference for work over a government grant to the value of R100 per month (*preferwork*). Standard Occupational Codes (SOC) used by Statistics South Africa (STATSSA) for Census 2001 are utilised to create the occupational skill variable. For those that found their first wage employment in Cape Town, the skill level of the first job reveals the level of skill required for that job. According to the SOC codes, those whose first jobs were classified above 7000 and below 9000 had jobs that were less skilled. (KMP Survey Report and Baseline Information, 2003: 82). The dummy variable for those whose first jobs involved higher skills is termed *betterskillf*. Continuous variables include the age at arrival (*aae*), its squared term (*aaesq*) and the year of arrival (*b6num*). The year of entry serves as a control for changes in macroeconomic environment or the changing quality of cohorts that Borjas (1984) refers to.

The baseline hazard was specified with time dummy indicators in model 5a, with a quadratic form for time in model 5b and log form (*lnseqvar*) in the last two models. All explanatory variables that are individual and household characteristics are time-invariant since they cannot be attached to a specific year. Although this would not affect gender, or age at entry, it may not be ideal in terms of household income, for example. Although it may be inappropriate to include household income at the time of the survey rather than at the time of entry, it is not irrelevant and is included to control for differences in wealth.

A sampling weight and a weight used to adjust the standard errors for clustering were added to each of the models. The sample weight adjusts for non-response of adults within a household using the raking ratio method and the clustering weight (psu) ensures that observations are independent across clusters but not necessarily independent within clusters. The KMP Survey Report and Baseline Information (2003: 5) reveal that clusters of households were identified by using the Enumerator Areas as defined by Statistics South Africa for the 1996 Population Census. Enumerator Areas are neighbourhoods of about 50 to 200 households and are designed to be homogeneous with respect to housing type and size.

The estimation results are shown in Figure 5. Each of the models is significant as a whole, as their log pseudo-likelihoods are sufficiently large. All the explanatory variables have a significant effect on the hazard rate excepting *aae*, *aaesq*, *hhtotminc*, *ruraledn* and *betterskillf*. The variable *b6num* is significant and indicates the importance of inclusion in an attempt to control for the quality of the cohorts over time, thereby producing less biased results. The odds ratios (exponentiated coefficients) are reported below for ease of interpretation. For cloglog models the exponentiated coefficients may be interpreted as hazard ratios. (Jenkins, Lesson 6: 20).

Figure 5: Multivariate Models with Different Specifications for the Baseline Hazard

	Complementary log-log			Logit
	5a	5b	5c	5d
event	exp_b	exp_b	exp_b	exp_b
male	1.21 *	1.29 **	1.22 **	1.26 **
	0.10	*0.11*	*0.10*	*0.12*
aae	1.00	0.98	0.99	0.98
	0.03	*0.03*	*0.02*	*0.03*
aaesq	1.00	1.00	1.00	1.00
	0.00	*0.00*	*0.00*	*0.00*
hhsocnet	1.42 **	1.51 **	1.40 **	1.53 **
	0.17	*0.19*	*0.16*	*0.20*
adulthhsize	1.11 **	1.10	1.09 *	1.12 *
	0.05	*0.06*	*0.05*	*0.06*
hhtotminc	1.00	1.00	1.00	1.00
	0.00	*0.00*	*0.00*	*0.00*
engprof	1.39 **	1.37 *	1.35 **	1.40 *
	0.15	*0.18*	*0.15*	*0.19*
ruraledn	0.84	0.76	0.80	0.76
	0.16	*0.17*	*0.15*	*0.19*
matric	2.11 **	2.00 **	1.79 **	2.04 **
	0.24	*0.26*	*0.21*	*0.33*
betterskillf	0.82	0.79	0.83	0.79
	0.16	*0.17*	*0.14*	*0.17*
preferwork	0.53 *	0.40 **	0.47 **	0.40 *
	0.39	*0.40*	*0.34*	*0.50*
b6num1	.00 **	1.03 **	1.02 **	1.03 **
	0.00	*0.01*	*0.01*	*0.01*
dur1	3.43 **			
	0.18			
dur2	2.38 **			
	0.22			
dur3	0.99 **			
	0.25			
dur4	1.94 **			
	0.32			
seqvarquad		1.00 **		
		0.00		
lnseqvar			0.62 **	0.58 **
			0.08	*0.09*
constant		*22.43* **	*17.42* **	*19.11* **
obs	924	924	924	924

Source: Own calculations using KMP data
Significance levels: ** significant at the 5%level, * at the 10% level
(standard errors adjusted for clustering on psu)

The hazard ratios for males suggest that they find employment faster than females do. This result is expected because females traditionally have more domestic duties that prevents more active search. (Schöer & Leibbrandt, 2006: 708). The hazard ratios attached to the age at entry variables are insignificant for all specifications. Respondents that made use of household social networks to find their first jobs have higher hazard ratios than those that found their jobs via other methods. This substantiates the point that employed household members serve to provide labour market information to their households.

English proficiency and the possession of a matric qualification both increase the probabilities of finding jobs earlier, as is to be expected. A matric qualification therefore operates as a useful screening device for potential employees concerning migrants. Those finding initial jobs involving a higher level of skill do not have significantly different outcomes to those that find initial jobs that require less skill. Walker (2003:52) asserts that the reservation wage for KMP residents declines as the unemployment duration increases. An interesting result is presented for the variable *preferwork*, as it implies that those that prefer to work (or carry on with their current activities) over receiving a R100 government grant per month have lower hazard ratios than those that prefer the grant. This may be interpreted to mean that those with higher reservation wages have longer unemployment spells and this is consistent with unemployment duration theories.

When unobserved heterogeneity, or the bias caused by not being able to include particular important explanatory variables in the regression model are taken into account, the results are sensitive to the assumption about whether the frailty term is distributed normally or has a Gamma distribution. The importance of unobserved factors is found to be significant for the Gamma distribution and insignificant for the normally distributed unobserved heterogeneity, based on the likelihood ratio test. The results are nonetheless included in the appendices (Model A3 presents the estimates for gamma distributed frailty and Model A4 shows the results for normally distributed unobserved heterogeneity).

The main limitation of the analysis thus far is the exclusion of the local-born, as factors relevant in determining the initial unemployment spell of migrants could be applicable to this group as well. Initial unemployment spells for the local-born are defined as the number of years it takes for respondents to find employment after completing their full-time education. The inclusion of household income would admittedly be of more use if this could be captured at the time of a respondent's entry into the Cape Town labour market. Data constraints do not permit this luxury and the models in Figure 5 have included household income at the time of the survey to proxy the wealth of the household in earlier times. This is not ideal because the effects are potentially endogenous. The employment status of household members at the time of labour market entry would also be a reasonable proxy for household wealth, as well as for the quality of household social networks but the data does not allow for this either. A complementary log-log model is therefore estimated to take these concerns into account and the results are given in Figure 6 below. As before, the odds ratios are presented for ease of interpretation. Before expanding the dataset for the discrete-time analysis, the total sample size is 1535 respondents, comprising of 1131 migrants and 404 locals. After expansion the number of migrants becomes 3262 and the number of locals is increased to 1220, bringing the total sample size to 4482.

As before, the hazard ratio for males suggests that they find employment faster than females do, although this variable is only significant at the 10 percent level. The result for the race dummy (*black*) implies that black respondents take longer to find jobs. It should be noted that most migrants are black and most of the local-born are coloured. This could explain why the result for *migrant* is not significant. Age at entry and its squared term have become significant and suggest that older entrants leave unemployment sooner. The household social network dummy was found to be insignificant, but once this was broadened to include all social networks (friends and relatives from households other than the respondent's own), the social network dummy was stable and significant. Those finding jobs through social networks found jobs faster. A variable as a proxy for experience was also included and only affects migrants because the unemployment spells of locals are calculated as the time taken to find their first employment. An interaction term is therefore included called *migfirst* that is a product of *migrant* and whether the respondent got their first-ever wage employment in Cape Town. The result is significant and suggests that migrants that worked in wage jobs for the first time ever in Cape Town have shorter unemployment durations. This result is at odds with theories that experienced workers do better in the labour market. Perhaps it could be attributed to a positive selection effect of migrants that leave their hometowns in search of wage jobs. The only other significant variable is the dummy for whether the respondent finished their education in Cape Town. This implies that migrant children as well as other students that were exposed to an education in Cape Town find jobs faster. Whether one was schooled in a rural setting proved to be insignificant. The possession of a matric qualification also had no significant effect. The constant term in the models in Figure 5 and Figure 6 are disconcertingly significant given the indications of duration dependence in the prior analysis, because it suggests limited duration dependence (Serneels, 2004: 29).

Figure 6: Complementary Log-log Model Including Both Migrants and Locals

event	exp_b	
male	1.18	*
	0.09	
black	0.79	*
	0.13	
migrant	0.69	
	0.68	
aae	1.04	**
	0.02	
aaesq	1.00	*
	0.00	
socialnetworks	1.24	**
	0.09	
migfirst	2.30	**
	0.14	
engprof	1.16	
	0.11	
finaleduCT	1.41	**
	0.14	
ruraledn	0.85	
	0.13	
matric	1.05	
	0.13	
betterskillf	1.10	
	0.10	
preferwork	0.84	
	0.20	
b6num	1.00	
	0.00	
lnseqvar	0.52	
	0.05	
_cons	*0.38*	**
obs	2141	

Source: Own calculations using KMP data
Significance levels: ** significant at the 5%level, * at the 10% level
(standard errors adjusted for clustering on psu)

5. Conclusions and Policy Implications

This paper provides an empirical discrete-time analysis of initial unemployment duration of migrants using the KMP 2000 data set. It finds that the reality of the post-migration urban experience depends on gender, English proficiency, a matric qualification, a higher reservation wage and the use of household social networks. The data seem to be robust for different functional forms of the baseline hazard because the results are similar for different models shown in Figure 5. The significance of household social networks in increasing labour market attachment and reducing unemployment durations for migrants is noteworthy. The initial duration of unemployment becomes dependent on the quality of the household in terms of the number of household members that are employed and on the strength of ties between the source area and the destination household. The period of initial unemployment for most migrants is uncertain given the lack of employment prospects and this poses a burden on the host household. The unemployed migrant may therefore feel a greater sense of urgency in finding employment. When the local-born are included in the analysis (Figure 6) social networks that include friends and relatives in other households also have a significant impact.

The long-term unemployed face an increasing hazard function, even though it appears that those people with such long spells are discouraged. Self-employment does not seem to be a survival strategy since very few migrants as a percentage of total migrants are self-employed. Policies to reduce the large numbers of long-term unemployed need to be addressed. Chiswick et al (1997: 292) explain that immigrants invest in location-specific skills such as English language and other training programs during the initial time in the United States. Because these investments are made in schools and not in a job, they note that employment rates would be lower but unemployment rates would not necessarily increase. Full-time education has been accounted for in the calculation of initial unemployment duration in this paper, but the idea that school enrolment is a step to the desired wage employment resonates with the lack of self-employment in KMP.

Dolton & O'Neill (1996: 387) have evaluated the impact of a programme in the United Kingdom that assists the long-term unemployed to return to the labour market by the provision of a "jobseekers allowance". By utilising a competing risks model, they find that the effect of the programme differs by the type of exit from unemployment. This type of incentive deserves some attention in South Africa since Schöer & Leibbrandt (2006) find that job search is limited by the cost of search, among other factors.

Finding ways to reduce numbers of discouraged workers is beneficial to shrinking the level of unemployment. Longer durations of unemployment mean that these people fall behind because they are unable to advance their human capital through work.

6. References

Allison, Paul D. 1982. Discrete-Time Methods for the Analysis of Event Histories. Volume 13. pp. 61-98. *Sociological Methodology*.

Bekker, Simon. 1999. *Circulatory Migration Linking Cape Town to the Eastern Cape, Some Reflections*. Rhodes University. East London Campus. June.

Borjas, George J. 1984. The Impact of Assimilation on the Earnings of Immigrants: A Reexamination of the Evidence. National Bureau of Economic Research. *Working Paper* No. 1515. December.

Brick, K. Mlatsheni, C. (2006). Examining the Degree of Duration Dependence in the Western Cape Labour Market. Masters thesis, University of Cape Town.

Chiswick, B. Cohen, Y. Zach, T. 1997. The Labour Market Status of Immigrants: Effects of the Unemployment Rate at Arrival and Duration of Residence. Volume 50. No. 2. pp. 289-303. *Industrial and Labor Relations Review*. January.

Cleves, M. Gould, W. Gutierrez, R. 2004. *An Introduction to Survival Analysis Using Stata, Revised Edition*. Stata Press.

Dolton, P. O'Neill, D. Unemployment Duration and the Restart Effect: Some Experimental Evidence. Volume 106. pp. 387-400. *The Economic Journal*. March.

Goldin, Ian. 1987. *Making Race: the Politics and Economics of Coloured Identity in South Africa*. Cape Town: Maskew Miller Longman.

Jakoet, J. 2006. Assimilation of Immigrants to the Cape Town Labour Market. *Working Paper* Number 06/03. Southern Africa Labour and Development Research Unit. University of Cape Town.

Jenkins, S. 1995. Easy Estimation Methods for Discrete-Time Duration Models. Volume 57. pp.129-138. *Oxford Bulletin of Economics and Statistics*.

Jenkins, S. EC968. Part 2: Introduction to the Analysis of Spell Duration Data. [Online].Available:
http://www.iser.essex.ac.uk/teaching/degree/stephenj/ec968/index.php

Khayelitsha/Mitchell's Plain Survey 2000. Southern Africa Labour and Development Research Unit. University of Cape Town.

Khayelitsha/Mitchell's Plain Survey 2000 Survey Report and Baseline Information. 2003. Southern Africa Labour and Development Research Unit, University of Cape Town. Population Studies Center, University of Michigan. March.

Kingdon, G. Knight, J. 2000. Are Searching and Non-Searching Unemployment Distinct States when Unemployment is High? The Case of South Africa. The Centre for the Study of African Economies *Working Paper Series*. 2000-2.

Migration Study in the Western Cape 2001. 2002. Executive Summary. Compiled by S. B. Bekker. June.

Nattrass, Nicoli. 2002. Unemployment, Employment and Labour-Force Participation in Khayelitsha/Mitchell's Plain. *CSSR Working Paper* No. 12. Centre for Social Science Research, University of Cape Town. October.

Ndegwa, D. Horner, D. Esau, F. 2004. The Links Between Migration, Poverty and Health: Evidence from Khayelitsha Mitchell's Plain. Southern Africa Labour and Development Research Unit. University of Cape Town. June.

Rodriguez, G. WWS509. 7.6: Discrete Time Models. [Online]. Available: http://data.princeton.edu/wws509/notes/c7s6.html.

Schöer, V. Leibbrandt, M. 2006. Determinants of job search strategies: Evidence from the Khayelitsha/ Mitchell's Plain Survey. *South African Journal of Economics*. 74(4): 702-24.

Serneels, Pieter. 2004. The Nature of Unemployment in Urban Ethiopia. The Centre for the Study of African Economies *Working Paper Series*. Paper 201.

Singer, J.D. Willett, JB. 1993. It's About Time: Using Discrete-Time Survival Analysis to Study Duration and the Timing of Events. Volume 18. No. 2. pp. 155-195. *Journal of Educational Statistics*. Summer.

Stata Corporation 2003. STATA Release 8.2. STATA Press Publication, STATA Corporation, College Station, Texas.

Steele, F. 2005. Event History Analysis. ESRC National Centre for Research Methods. NCRM Methods Review Papers. NCRM/004. September.

Thapa, P.J. Gorgens, T. 2006. A Duration Analysis of the Time Taken to Find the First Job for Newly Arrived Migrants in Australia. July. *Discussion Paper* No.527. Centre for Economic Policy Research, Australian National University.

Walker, Richard. 2003. Reservation Wages - Measurement and Determinants: Evidence from the Khayelitsha/Mitchell's Plain (KMP) Survey. *CSSR Working Paper* No. 38. Centre for Social Science Research, University of Cape Town.

7. Appendices

Table A1: Descriptive Statistics

	Initial Unemployment Duration Greater Than One Year				
Variable	**Obs**	**Mean***	**Std. Dev.**	**Min**	**Max**
Initial Unemployment Duration	499	5.14	4.50	1.1	30
Age	499	35.34	12.02	18	78
Age at entry	499	21.95	8.85	1	60
Male Dummy	499	0.41	0.49	0	1
Years Completed Education	453	8.59	3.34	0	14
Household Total Monthly Income	421	2578	7184	20	114400
Number of Adults in the Household	455	2.96	1.39	1	8
Total Household Size	488	4.58	2.39	1	13
Social Network Dummy (1 if used)	202	0.65	0.48	0	1
Low Skill Occupation Dummy	39	0.79	0.41	0	1
Censor (1 if found job)	486	0.58	0.49	0	1
English Proficiency Dummy (1 if speak well)	499	0.18	0.38	0	1
Final Education completed in CT Dummy	383	0.43	0.50	0	1
Rural Education Dummy	455	0.76	0.42	0	1
Rural Born Dummy	488	0.77	0.42	0	1
	Initial Unemployment Duration Less Than or Equal to One				
Variable	**Obs**	**Mean***	**Std. Dev.**	**Min**	**Max**
Initial Unemployment Duration	632	0.34	0.43	0	1
Age	632	34.81	15.37	18	86
Age at entry	632	22.15	10.56	1	59
Male Dummy	632	0.37	0.48	0	1
Years Completed Education	574	8.94	3.12	0	15
Household Total Monthly Income	519	3095	8450	5	114400
Number of Adults in the Household	583	3.04	1.34	1	8
Total Household Size	623	4.71	2.23	1	15
Social Network Dummy (1 if used)	210	0.66	0.48	0	1
Low Skill Occupation Dummy	29	0.72	0.45	0	1
Censor (1 if found job)	499	0.64	0.48	0	1
English Proficiency Dummy (1 if speak well)	632	0.24	0.43	0	1
Final Education completed in CT Dummy	424	0.45	0.50	0	1
Rural Education Dummy	581	0.70	0.46	0	1
Rural Born Dummy	622	0.74	0.44	0	1

Source: Own calculations using KMP data

*Proportion if dummy variable

Table A2: Initial Unemployment Duration by Arrival Cohort

Arrival Cohort	Mean	Std. Dev.	Freq.
2000	0.66	0.44	87
95-99	1.31	1.29	305
90-94	2.37	2.82	239
85-89	2.90	3.74	198
80-84	4.15	4.87	112
70-79	5.65	7.39	105
1900-1969	1.45	3.38	85
Total	2.46	3.84	1131

Source: Own calculations using KMP data

Table A3: Mean Years of Education by Arrival Cohort

Arrival Cohort	Mean	Std. Dev.	Freq.
2000	9.73	2.88	83
95-99	9.70	2.96	279
90-94	8.83	3.11	215
85-89	8.35	3.23	181
80-84	8.69	3.33	102
70-79	7.63	3.43	95
1900-1969	6.76	2.93	72
Total	8.79	3.22	1027

Source: Own calculations using KMP data

Table A4: Mean Initial Duration of Unemployment for those Educated in Cape Town

Educated in CT	Mean	Std. Dev.	Freq.
Not Educated in CT	3.06	4.49	451
Educated in CT	2.11	2.79	356
Total	2.64	3.86	807

Source: Own calculations using KMP data

Table A5: Sample Selection

	Sample Selected				
Variable	**Obs**	**Mean***	**Std. Dev.**	**Min**	**Max**
Age	1131	35.05	13.99	18	86
Age at entry	1131	22.06	9.84	1	60
Male Dummy	1131	0.38	0.49	0	1
Years Completed Education	1027	8.79	3.22	0	15
Household Total Monthly Income	940	2863.45	7908.07	5	114400
Number of Adults in the Household	1038	3.01	1.36	1	8
Total Household Size	1111	4.65	2.30	1	15
Social Network Dummy (1 if used)	1131	0.24	0.43	0	1
Low Skill Occupation Dummy	68	0.76	0.43	0	1
Censor (1 if found job)	985	0.61	0.49	0	1
English Proficiency Dummy (1 if speak well)	1131	0.21	0.41	0	1
Final Education completed in CT Dummy	807	0.44	0.50	0	1
Rural Education Dummy	1036	0.73	0.44	0	1
Rural Born Dummy	1110	0.75	0.43	0	1
	Sample Dropped due to Missing Data				
Variable	**Obs**	**Mean***	**Std. Dev.**	**Min**	**Max**
Age	460	36.77	10.32	18	67
Age at entry	460	21.21	8.15	1	48
Male Dummy	460	0.54	0.50	0	1
Years Completed Education	432	9.41	3.34	1	15
Household Total Monthly Income	397	3145.62	7297.42	5	114400
Number of Adults in the Household	420	2.91	1.37	1	10
Total Household Size	452	4.53	2.33	1	21
Social Network Dummy (1 if used)	460	0.29	0.45	0	1
Low Skill Occupation Dummy	432	0.71	0.46	0	1
Censor (1 if found job)	382	1.00	0.00	1	1
English Proficiency Dummy (1 if speak well)	460	0.27	0.45	0	1
Final Education completed in CT Dummy	372	0.52	0.50	0	1
Rural Education Dummy	434	0.76	0.42	0	1
Rural Born Dummy	448	0.70	0.46	0	1

Source: Own calculations using KMP data

*Proportion if dummy variable

Figure A1: Survival Function Estimates by Gender

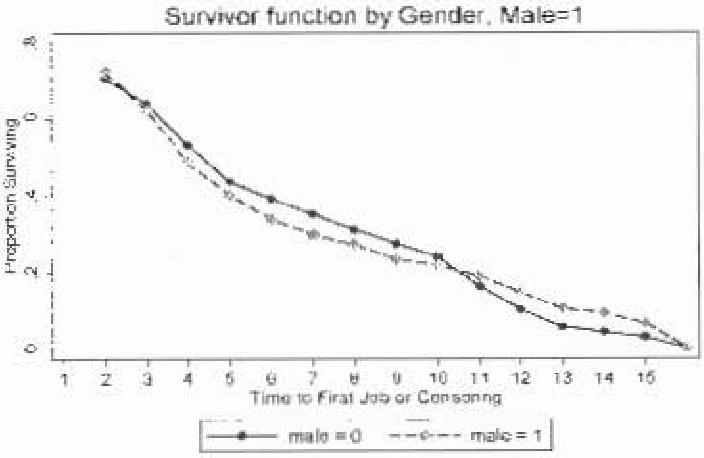


Figure A2: Survival Function Estimates by Social Network Dummy variable

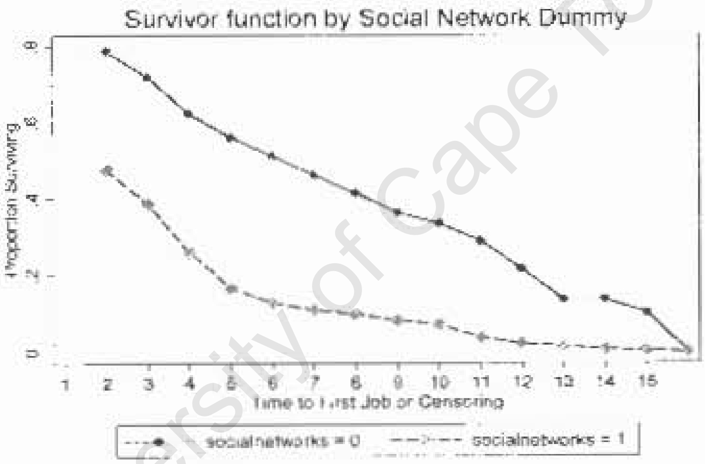


Figure A3: Survival Function Estimates by Dummy Variable for Completing Education in Cape Town

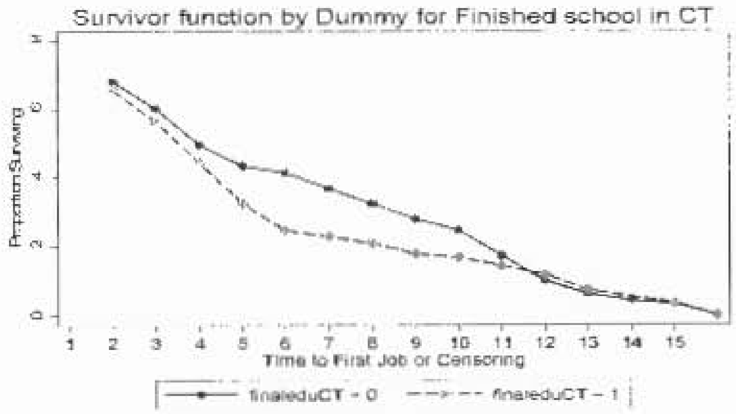


Figure A4: Graphical check of Proportional Hazards Assumption-Fitted Hazard by Gender

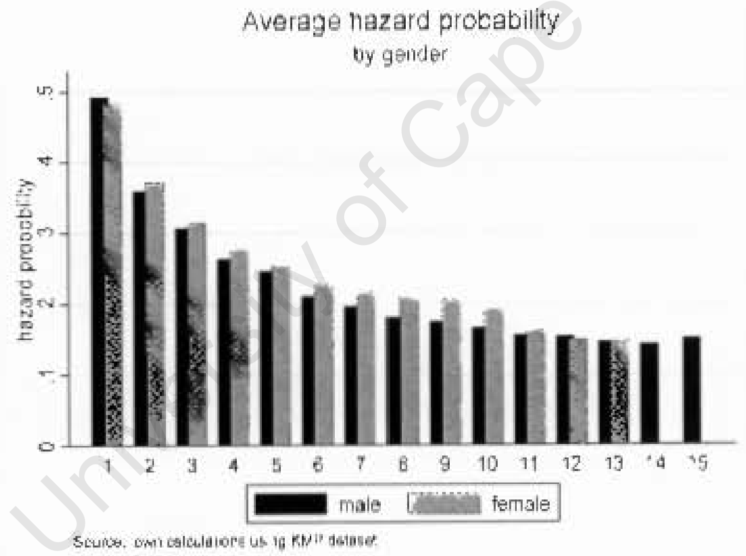
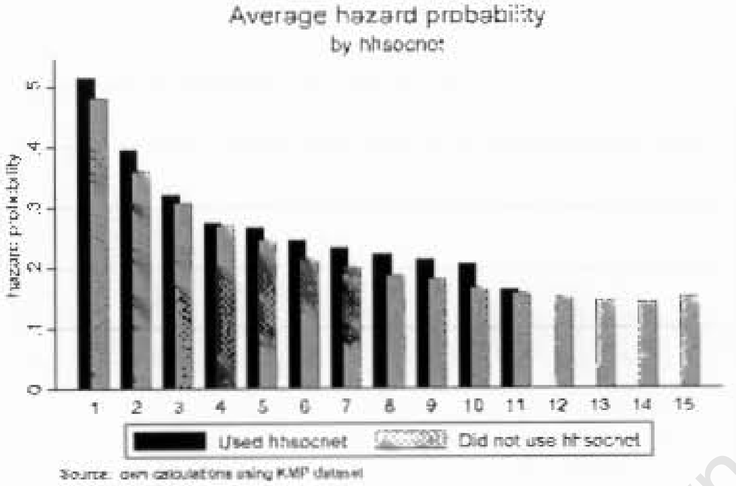


Figure A5: Graphical check of Proportional Hazards Assumption-Fitted Hazard by Household Social Network Dummy Variable



Model A1: Initial Discrete-Time Cloglog Model Including 15 Time Dummy Variables - Main Effects of the Time Indicators

Complementary log-log regression
Number of obs = 3262
Zero outcomes = 2659
Nonzero outcomes = 603
Wald chi2(15) = 1265.95
Prob > chi2 = 0.0000

Log pseudo-likelihood =-212801.23
(standard errors adjusted for clustering on psu)

event	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
durat1	-1.13	0.06	-17.91	0.0000	-1.25	-1.00
durat2	-2.08	0.14	-15.33	0.0000	-2.34	-1.81
durat3	-1.53	0.13	-12.12	0.0000	-1.77	-1.28
durat4	-1.64	0.17	-9.89	0.0000	-1.97	-1.32
durat5	-2.03	0.20	-9.98	0.0000	-2.43	-1.63
durat6	-2.25	0.27	-8.33	0.0000	-2.78	-1.72
durat7	-2.34	0.30	-7.85	0.0000	-2.92	-1.75
durat8	-2.07	0.32	-6.47	0.0000	-2.70	-1.44
durat9	-2.42	0.38	-6.44	0.0000	-3.15	-1.68
durat10	-1.62	0.27	-5.96	0.0000	-2.16	-1.09
durat11	-1.41	0.28	-5.06	0.0000	-1.95	-0.86
durat12	-1.39	0.33	-4.28	0.0000	-2.03	-0.76
durat13	-2.66	0.73	-3.65	0.0000	-4.09	-1.23
durat14	-2.12	0.57	-3.74	0.0000	-3.24	-1.01
durat15	-1.10	0.40	-2.76	0.0060	-1.89	-0.32

Model A2: Constant Piece-wise Cloglog Model Including 4 Time Dummy Variables - Main Effects of the Time Indicators

Complementary log-log regression
Number of obs = 3262
Zero outcomes = 2659
Nonzero outcomes = 603
Wald chi2(4) = 1198.2
Prob > chi2 = 0.0000

Log pseudo-likelihood = -217109.76
(standard errors adjusted for clustering on psu)

event	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
dur1	-1.35	0.06	-21.83	0.0000	-1.47	-1.23
dur2	-1.57	0.10	-15.68	0.0000	-1.77	-1.38
dur3	-2.19	0.10	-20.87	0.0000	-2.39	-1.98
dur4	-1.59	0.15	-10.68	0.0000	-1.89	-1.30

Model A3: Hazard Model with Gamma Frailty

PGM hazard model with gamma frailty

event	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
hazard						
male	0.19	0.19	0.98	0.3280	-0.19	0.56
aae	-0.01	0.04	-0.32	0.7520	-0.09	0.07
aaesq	0.00	0.00	0.14	0.8910	0.00	0.00
hhsocnet	0.40	0.25	1.62	0.1050	-0.08	0.88
adulthhsize	0.19	0.07	2.54	0.0110	0.04	0.33
hhtotminc	0.00	0.00	-0.83	0.4080	0.00	0.00
engprof	0.52	0.25	2.10	0.0360	0.03	1.01
ruraledn	-0.40	0.22	-1.82	0.0690	-0.83	0.03
matric	0.79	0.34	2.31	0.0210	0.12	1.46
betterskillf	-0.20	0.23	-0.86	0.3900	-0.64	0.25
preferwork	-1.26	0.41	-3.05	0.0020	-2.07	-0.45
b6num	0.02	0.01	2.16	0.0300	0.00	0.03
_cons	-35.63	16.58	-2.15	0.0320	-68.12	-3.13
ln_varg						
_cons	-0.65	0.22	-2.95	0.0030	-1.09	-0.22
Gamma var.	0.52	0.12	4.51	0.0000	0.34	0.80

LR test of Gamma var. = 0: chibar2(01) = 37.6627 Prob.>=chibar2 = 4.2e-10

Model A4: Hazard Model with Normal Frailty

Gaussian

event	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
male	0.16	0.13	1.19	0.2350	-0.10	0.41
aae	-0.02	0.03	-0.89	0.3720	-0.07	0.03
aaesq	0.00	0.00	0.65	0.5180	0.00	0.00
hhsocnet	0.31	0.17	1.84	0.0660	-0.02	0.64
adulthhsize	0.09	0.05	1.97	0.0490	0.00	0.18
hhtotminc	0.00	0.00	-0.60	0.5490	0.00	0.00
engprof	0.32	0.16	1.99	0.0460	0.01	0.63
ruraledn	-0.22	0.14	-1.50	0.1330	-0.50	0.07
matric	0.59	0.22	2.63	0.0080	0.15	1.02
betterskillf	-0.25	0.16	-1.59	0.1120	-0.56	0.06
preferwork	-0.75	0.26	-2.89	0.0040	-1.26	-0.24
b6num	0.02	0.01	2.81	0.0050	0.01	0.03
lnseqvar	-0.51	0.08	-6.43	0.0000	-0.67	-0.36
_cons	-35.61	12.68	-2.81	0.0050	-60.46	-10.75
/lnsig2u	-14	502.018			-997.9372	969.9372
sigma_u	0.0009	0.2289			2.00E-217	4.20E+210
rho	5.06E-07	0.0003			0	.

Likelihood-ratio test of rho=0: chibar2(01) = 0.00 Prob >= chibar2 = 1.000