Southern Africa Labour and Development Research Unit

SALDRU

Examining the Degree of Duration Dependence in the Cape Town Labour Market

by
Kerri Brick and Cecil Mlatsheni

Working Papers can be downloaded in Adobe Acrobat format from www.saldru.uct.ac.za.
Printed copies of Working Papers are available for R15.00 each plus vat and postage charges.

Orders may be directed to:

The Administrative Officer, SALDRU, University of Cape Town, Private Bag, Rondebosch, 7701,

Tel: (021) 650 5696, Fax: (021) 650 5697, Email: brenda.adams@uct.ac.za

Examining the Degree of Duration Dependence in the Cape Town Labour Market

Kerri Brick and Cecil Mlatsheni

ABSTRACT

This paper investigates the effects of long bouts of unemployment on the employment probabilities of the unemployed, and the factors affecting unemployment duration. This is achieved through the estimation of the hazard rate. The parametric analysis was conducted using a Weibull regression model. The Weibull estimates indicate that long term unemployment will be more prevalent for young women who care for dependents aged below 16 years, have no grade 12 or tertiary qualifications and have limited proficiency in English. This effect is compounded when the unemployed are marginalised and live with few or no wage earners. Slight positive duration dependence was found, indicating that employment probabilities do not decrease as unemployment duration increases. A graphical depiction of the Weibull hazard indicates that the hazard is initially upward sloping, reaching a maximum at approximately 13 months of unemployment, at which point it declines steeply. At the 13 month turning point, merely 17 percent of respondents have exited from unemployment. Even at its highest, the hazard remains below 8 percent. A piece-wise model was estimated which confirmed the Weibull estimates, but indicated that the hazard follows a more flexible shape and is not monotonic in nature.

1.   <u>INTRODUCTION</u>

This paper examines the factors affecting the unemployment durations of cohorts residing in the magisterial district of Mitchell's Plain (which includes the African townships of Khayelitsha, Gugulethu and Langa). Examining the effects of explanatory variables on the exit rate from unemployment enables the identification of particular demographic groups that are more likely to suffer from longer bouts of unemployment (van den Berg 1999). Furthermore, this paper examines to what degree duration dependence is observed. Duration dependence occurs when the exit rate out of unemployment is not constant over time. In other words, if the expected future duration of an unemployment spell is partially dependent on the amount of time spent in unemployment thus far, the process is duration dependent (Bennet 1999). Duration dependence will be examined with the use of the hazard rate (out of unemployment), which examines how the conditional probability of leaving unemployment changes as unemployment duration increases (Serneels 2002b). The hazard rate is the probability that an individual will make a transition into employment, in a given interval, conditional on having been in a state of unemployment at the start of that interval. The hazard thus represents the exit rate from unemployment (Cleves et al 2004). The hazard rate will be discussed in greater detail in the theoretical overview of section 3.3.1. Throughout the unemployment spell, the hazard can "increase, decrease, remain constant, or even take on more serpentine shapes" (Cleves 2004, p. 8). If observed duration dependence is negative, the conditional probability of leaving unemployment falls as unemployment duration rises. Conversely, a positive hazard rate means that a short duration of the unemployment can be expected (Kalb 2001).

As discussed by van den Berg and van Ours (1999), economic theory would suggest that duration dependence is negative. The loss of human capital that is synonymous with long bouts of unemployment might produce a stigma effect, thereby reducing the number of job-opportunities available to the unemployed. Furthermore, the discouraged worker effect, whereby workers decrease their search intensity, or resort to more passive means of job search, could serve to decrease the exit probability of the unemployed (van den Berg and van Ours 1999). Although duration dependence has historically been acknowledged to be negative, an overview of the literature presented by Serneels (2002a) shows this assumption

to be empirically unfounded[1]. Specifically, more recent literature implies non-negative duration dependence. Serneels (2002a) suggests that this counter-intuitive finding may be due to "improved modelling techniques as well as the richer information on the unemployed, both of which improve the ability to control for unobserved heterogeneity." (Serneels 2002b, p. 3)

High unemployment rates in South Africa have been well documented. Bhorat and Oosthuizen (2005) provide a comprehensive overview of the South African labour market in their paper: *The Post-Apartheid South African Labour Market* (Bhorat and Oosthuizen 2005). Specifically, the paper analyses changes in employment, unemployment and labour market participation for the period 1995 to 2002 with the use of two datasets, namely the October Household Survey (OHS) and the Labour Force Survey (LFS). A significant finding is that the period between 1995 and 2002 has been synonymous with rising broad and narrow rates of unemployment. This finding is supported by Kingdon and Knight (2000), who show that both searching and non-searching unemployment have increased steadily between 1995 and 1997 (Kingdon and Knight 2000). Although these two unemployment definitions will be discussed in detail in a subsequent section, it bears mentioning at this point that the broadly defined unemployed encompasses discouraged work-seekers, who no longer actively seek employment, and work-seekers who use more passive job search methods i.e. social networks.

The official unemployment rate in 2002 was 30.5, percent while the broad unemployment rate for the same year was 41.8 percent. Between 1995 and 2002 the narrowly defined unemployed increased by 12.9 percent, while broadly defined unemployment increased by 11 percent (Bhorat and Oosthuizen 2005). While the *conservative* estimate of 30.5 percent is in itself, alarmingly high, the broad definition estimate of 41.8 percent is unacceptably high, especially when viewed in relation to international standards.

Dinkenlman (2004) investigates transition rates between different labour market states of Black cohorts living in KwaZulu-Natal between the periods 1993-1998[2]. The paper finds that respondents classified as non-economically active (NEA) (non-labour market participants

---

[1] The reader is referred to an outline of the literature on duration dependence in OECD countries provided by Serneels (2002b, p. 4, Table 1).
[2] The data used was from the 1993 Project for Statistics on Living Standards and Development (PSLSD) household survey, and the KwaZulu-Natal Income Dynamics Study (KIDS), which is a re-survey of those Black and Indian households living in KwaZulu-Natal who were surveyed in 1993 under the PSLSD survey.

which includes the non-searching unemployed and discouraged workers) in 1993 were very likely to still be in this same labour market state in 1998. Furthermore, fewer than 10 percent of those reported to be NEA in 1993 had joined the ranks of the searching unemployed by 1998. In addition, approximately 50 percent of men and women identified as part of the searching unemployed had exited from this state by 1998, either into employment or out of the labour force. The implication is that during their unemployment spell, the searching unemployed either find employment or alternatively stop searching (Dinkelman 2004).

Bhorat et al (2005) indicate that Black labour force participants experience the highest rate of unemployment, with 50 percent of Black labour force participants broadly unemployed in 2002. The unemployment rates of Coloured and White work-seekers are significantly lower at 29.3 and 9 percent respectively. When decomposing these effects further along gender lines, it becomes evident that within each racial group, females experience higher unemployment rates than do males. Overall however, Black females experience the highest rates of unemployment with more than 55 percent of Black females unemployed in 2002. Black male labour force participants experience higher unemployment rates than any other racial group with almost 43 percent unemployed in 2002. The lowest unemployment rates were experienced by White males and females with unemployment rates of 7.2 and 11.2 percent respectively (Bhorat and Oosthuizen 2005).

Individuals without a matric certificate represent 41 percent of the population. In 2002, 50 percent of these work-seekers were unemployed. However, it is not just those with incomplete secondary education that are experiencing exceedingly high unemployment levels, approximately 40 percent of matric graduates were unemployed in 2002. The high and increasing rate of unemployment of matriculants is disturbing in light of the traditional view that completing secondary education is considered to be a way out of poverty and unemployment. It appears that tertiary education provides a greater safeguard against unemployment with 14.6 percent of those with a tertiary education unemployed (Bhorat and Oosthuizen 2005).

As worrying as the high levels of unemployment prevalent in the South African labour market, is the degree of permanence of unemployment experienced by the unemployed. Bhorat and Oosthuizen (2005) show that almost 90 percent of unemployed individuals have unemployment durations exceeding three years, or alternatively, have never had a job at all.

Kingdon and Knight (2000) report similar findings, and suggest that a high proportion of unemployed Sough Africans face very long unemployment durations. Kingdon and Knight (2000) found that in 1997, 37 percent of the searching unemployed experienced unemployment durations of more than three years, whilst a further 30 percent of the unemployed were without work for between 1 and 3 years. Kingdon and Knight (2000) thus suggest that the unemployment problem in South Africa is characterised by both "depth and breadth" (Kingdon and Knight 2000, p. 5).

Historically, much of the empirical literature on duration dependence has been conducted using OECD data, where unemployment durations are less than one year (Serneels 2002b). Although unemployment duration in developing countries is expressed in years as opposed to months, analysis of duration dependence in this context is "virtually non-existent" (Serneels 2002b, p.3) [3].

The paper will continue as follows: Section 2 begins with a brief description of the data and proceeds with a discussion on the relevance of the broad and narrow unemployment definitions in the context of this analysis. Section 3 includes a concise theoretical overview, and proceeds with an examination of duration dependence using non-parametric, parametric and semi-parametric modelling techniques. A non-parametric analysis, which makes no assumptions about the functional form of the hazard, is a useful point of departure as it provides a useful empirical overview of the data. This is followed by a parametric investigation, in which a Weibull regression model is estimated. This allows for the effects of covariates on the hazard to be estimated. The Weibull has been extensively used in the empirical literature and is thus useful as a point of comparison with previous studies. As the assumption of a monotonic hazard is very restrictive, however, the analysis will proceed with the estimation of a piece-wise constant hazard model. Conclusions are drawn in section 4. The reader is referred to the CD at the back of this paper which includes the Do-file used in this analysis, as well as an accompanying list of some of the relevant commands.

2.   __DATA DESCRIPTION__

---

[3] Serneels (2002a) gives a brief description of the empirical literature on unemployment duration in developing countries. The reader is referred to Dickens and Lang (1996), Kingdon and Knight (2000), Tunali and Assaad (1992), Appleton, Knight, Song and Xia (2001), and Serneels (2002a, 2002b).

The data used in this analysis has been sourced from the 2000 Khayelitsha/Mitchell's Plain (KMP) Survey. The survey was administered to a random sample of adults, aged 18 years and older, living in the Mitchell's Plain magisterial district, which includes the African townships of Khayelitsha, Gugulethu and Langa (Crankshaw and Welch 2003 and Nattrass 2002). Thus, the survey contains labour market data associated with Black and Coloured South Africans residing in Cape Town. As mentioned by both Nattrass (2002) and Walker (2003), the survey does not comprise a "representative sample of the Cape Town metropolitan area" (Nattrass 2002, p.1), but rather, reflects a working class contingent of Black and Coloured South Africans (Nattrass 2002 and Walker 2003). An upper boundary of age 65 has been imposed, reflecting the conventional retirement age, furthermore, there were very few observations for those aged 66 years and above.

As noted by Dinkelman (2004), the discussion regarding which unemployment definition is more relevant in a South African context has been much debated in the empirical literature.[4] In 1998, Statistics South Africa adopted the narrow unemployment measure as the "official" definition of unemployment (Kingdon and Knight 2000). It is debatable whether such a perfunctory application of this definition to the South African labour market underestimates the unemployment problem, as illustrated by the significant discrepancy between the broad and the narrow unemployment rates mentioned earlier. The distinction between the ILO narrow definition and the broad definition is important in terms of which approach to use in this paper when measuring unemployment and defining who is unemployed. To be considered unemployed under the narrow definition, the jobless person must have looked for work in the reference period, and be available for work during weekdays. The reference period is usually the week or month prior to the inquiry (Nattrass 2002). Active job search includes any of the following: looked for a job in the newspaper, went to factories and waited outside, knocked on factory gates and/or visited private homes and shops, visited employment agencies, phoned up or visited old employers and asked for jobs, waited on the side of the road for jobs, or looked on notice boards in community centres, shopping centres, shops etc. (Nattrass 2002 and KMP Survey 2000).

The broadly defined unemployed consists of the narrowly defined unemployed, and in addition, includes all those jobless persons who wanted work but did not look for it in the

---

[4] See Kingdon and Knight (2000), Nattrass (2002) and Dinkelman (2004) for a more complete analysis of this debate.

specified time period. Nattrass (2002) describes the exclusively network-searching unemployed as unemployed individuals who rely on networks consisting of friends and relatives to find them work. Such individuals have reported that they desire to work and are available to work but have not taken active steps to find work within the reference period. The network-searching unemployed have indicated that they have done one or more of the following in the past month, namely; relied on household members for information about jobs, relied on friends/family members in different households for information about jobs, relied on household members to get them a job at their workplace, and relied on friends/family members in a different household to get them a job at their workplace (Nattrass 2002). Table 1 on the following page indicates that active search methods do not constitute the dominant search method of the unemployed in the KMP data set, and furthermore, that cohorts relied on more passive job-search (network-searching) methods.

Table 1: Job search methods of the unemployed[5]

Search Strategy	
Looked in newspapers	.43
Relied on household members to tell you about jobs	.54*
Relied on family/friends in different households to tell you about jobs	.56*
Relied on household members to get you a job at their workplace	.51*
Relied on family/friends in different households to get you a job at their workplace	.54*
Went to factories and waited outside	.30
Knocked on factory gates and/or visited private homes and shops	.26
Visited employment agencies	.09
Phoned up or visited old employers and asked for jobs	.09
Waited on the side of the road for jobs	.08
Looked on notice boards in community centers, shopping centers, shops etc	.21

*indicates methods used by the network-searching unemployed

Discouraged workers are those that do not qualify as either the actively searching unemployed or the network-searching unemployed. These are individuals who are willing to and available for work, but have become discouraged by a fruitless job-search. Such respondents answered yes to the question "Do you want a job?", and are available to work during weekdays, but did not perform any of the job-search activities listed in the survey (and above) in the month prior to the interview (Nattrass 2002). Almost all respondents that had

---

[5] Table 2 illustrates the percentage of respondents using each job-search method one month prior to the KMP interview. Note that respondents were not restricted to one search method, and thus the total is over 100 percent.

not undertaken any job-search activities one month prior to the interview indicated that they *do* in fact want a job and are able to work during weekdays.

Thus, the broad definition of unemployment incorporates the actively seeking unemployed, the network-searching unemployed and discouraged workers. The narrow definition excludes both the network-searching unemployed and discouraged workers, and so does not permit passive job search methods (Nattrass 2002). Kingdon and Knight (2000) and Nattrass (2002) provide a discussion as to whether passive work-seekers should be represented in the official unemployment statistics. As discussed by Serneels (2002a), and argued by Kingdon and Knight (2000), given the high unemployment rates prevalent in South Africa, passive job-search may constitute a viable job-search method, and thus, it is likely that the unemployed are waiting as opposed to actively searching for jobs (Kingdon and Knight 2000 and Serneels 2002). Dinkelman (2004) notes that as the probability of successful job-search in high unemployment areas is low, whilst information (and transport) costs are high, the unemployed rely on friends and household members to relay information back to them (Dinkelman 2004). Reliance on passive job-search methods, such as waiting for information from family and friends, is evident from Table 1 above, where passive job-search methods constituted the greatest proportion of all job search methods undertaken by the unemployed. Nattrass (2002) uses the KMP dataset to demonstrate that a sizeable difference exists between the two unemployment definitions. Using the strict unemployment definition, which only includes the actively-searching unemployed, Nattrass (2002) finds a strict unemployment rate of 28.4 percent. However, when the passively-searching unemployed are included, Nattrass (2002) generates a broad unemployment rate of 46.3 percent (Nattrass 2002). Thus, in this analysis, the definition of unemployment is extended to include both the network-searching unemployed and discouraged workers.

Having established what constitutes being unemployed or employed at the time of the interview, it is necessary to determine the duration of the unemployment spell. Respondents who are currently in wage employment (at the time of the survey) were asked "were you unemployed and wanting a job before you got your current job?" If the respondent answered yes to this question, they were then asked "how long where you unemployed before starting your current job?" Thus, for the wage employed, there is a direct observation of their unemployment spell in months. The unemployed were asked "how long have you been wanting work (and been without any paid employment)?" Thus once again, a direct

observation of the length of the unemployment spell is possible. Unfortunately, the same questions were not asked of those that are currently self-employed, meaning that these respondents had to be excluded from the analysis. Furthermore, those cohorts who were currently employed, but were not unemployed before their current job (which represents 42 percent of all those currently in wage employment) have also been excluded, owing to the fact that they did not make a transition from unemployment to employment. In the analysis, observed unemployment spells ending without a transition into a state of employment have been treated as right censored. The sample contains 1084 uncensored spells plus 458 right-censored spells.

3.   EXAMINING DURATION DEPENDENCE

3.1   NON-PARAMETRIC ESTIMATION

Parametric analysis requires the specification of a particular functional form for the hazard. The hazard, as previously mentioned, is the exit rate from unemployment. The hazard is thus the probability that an individual will make a transition into employment, in a given interval, conditional on having been in a state of unemployment at the start of that interval (Cleves et al 2004). The implications of this will be discussed in the following sub-section. The advantage of a non-parametric estimation, however, is that no prior assumptions about the functional form of the hazard are required (Cleves et al 2004). Thus, a non-parametric analysis is useful at the beginning of an empirical analysis as it provides a summary of the data (Cleves et al 2004), the results of which may be compared to the parametric analysis that follows. A limitation of non-parametric estimation is that it does not explore the impact of explanatory variables on the hazard. Thus, in order to analyse the effect of covariates on the hazard, a parametric estimation will follow in section 3.3.

The Kaplan-Meier survival function, depicted in Figure 1 on the following page, illustrates the number of cohorts that remain in unemployment through time (Serneels 2002b). The survival function starts with 100 percent of the respondents unemployed (at analysis time zero), and drops towards zero as respondents make the transition to employment. Whenever a failure occurs, the survival curve drops to a new level, and to accommodate this, it is plotted as a step function (Jenkins 2004).
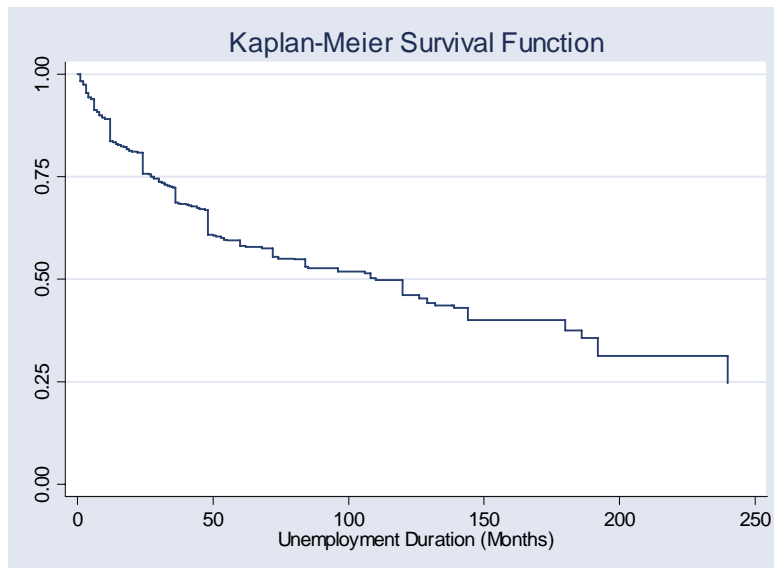
Figure 1: Kaplan Meier Survival Function

Figure 2 depicts the cumulative (integrated) hazard function, which has been estimated using the Nelson-Aalen method[6]. The cumulative hazard function is the integral of the continuous time hazard rate (Cleves et al 2004). The cumulative hazard function depicts the number of people exiting from unemployment, divided by the total number unemployed, at each point of the unemployment spell (Serneels 2002b). Because the continuous time hazard is the derivative of the cumulative hazard function, one might attempt to estimate the hazard from the slope of the cumulative hazard function. However, as the cumulative hazard is a step function, it cannot be directly differentiated (Cleves et al 2004). The hazard rate can still be derived, however, by smoothing the steps of the Neslon-Aalen cumulative hazard function using a kernel smoother (Cleves et al 2004 and Jenkins 2004). The smoothed hazard rate is depicted in Figure 3 on the following page.

Figure 2: Cumulative Hazard Function

---

[6] The relationship between the survivor function $S(t)$ and the cumulative hazard function $H(t)$ is characterised as $H(t) = -\ln\{S(t)\}$ where the Kaplan-Meier survivor function could be used as an estimate of $S(t)$. However, as an alternate approach, the nonparametric Nelson-Aalen method for estimating $H(t)$ is advocated in the literature, due to the superior small-sample properties of the method (Cleves et al 2004).

Nelson-Aalen cumulative hazard estimate
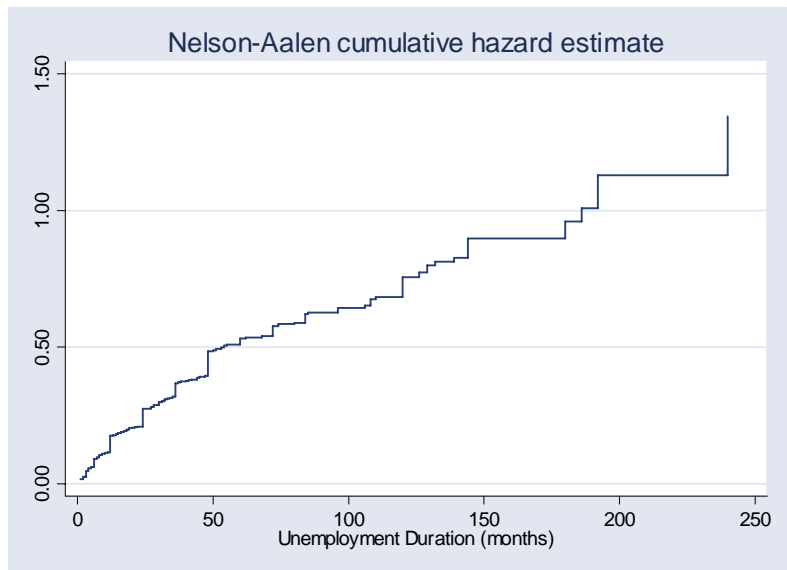
Figure 3: Smoothed Hazard Function

Smoothed hazard estimate
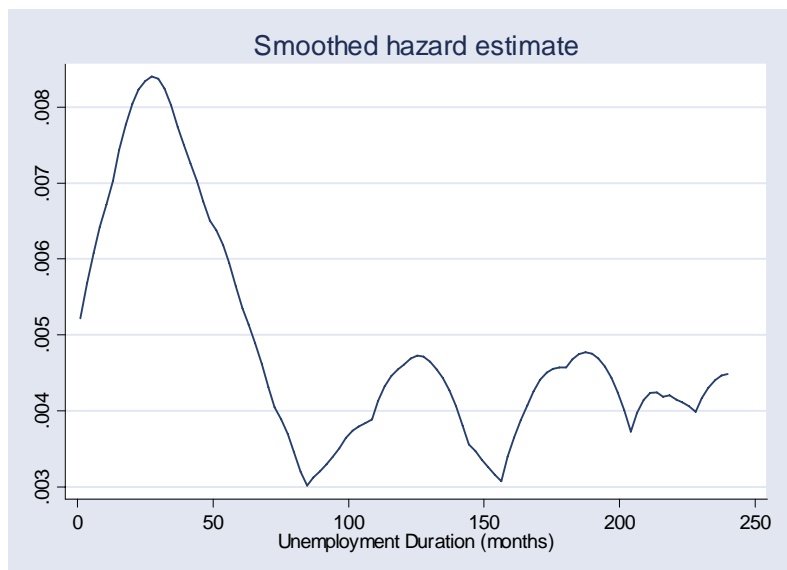
As previously mentioned, the hazard rate is the probability that an individual will leave unemployment, in a given interval, conditional on having been unemployed at the start of that interval. The smoothed hazard in Figure 3 above indicates that, at its maximum, the hazard remains below 9 percent. Furthermore, the hazard does not appear to be monotonic, but rather, follows a more flexible course.

11

3.2   PARAMETRIC ESTIMATION

3.2.1   Theoretical Overview

The Hazard Rate

The hazard rate is defined as:

$$h(t/x) = \lim_{dt \to 0} \frac{P(t \leq T < t + dt \mid T \geq t, x)}{dt},$$

where T is the duration of an unemployment spell and x is a vector of observed explanatory variables. The hazard rate is the probability that an individual will make a transition out of unemployment in the interval [t, $t+dt$], conditional on being unemployed at t (van den Berg 2000)[7] More specifically, the hazard rate is the probability that a certain event will occur i.e. a transition into employment, given that the event has not yet occurred (Box-Steffensmeier and Jones 1997). Duration dependence is observed if the value of the hazard rate, $h(t/x)$, changes over time t. Positive duration dependence occurs when the hazard function or the exit rate increases with time, whilst negative duration dependence occurs when the hazard function decreases with time (van den Berg 2000).

Proportional Hazard Models

Proportional hazard models are defined as:

$$h(t, X_j) = h_0(t) \exp(X_j \beta_x),$$

The function $h_0(t)$ is the baseline hazard which is a function of t but not of X. The baseline hazard determines the shape of the hazard when the effects of the explanatory variables are

---

[7] Other references include: Cleves et al (2004), Korpi (1995), Box-Steffensmeier and Jones (1997), Jenkins (2004), Lancaster (1990), and Yoon (1985).

zero, and thereby depicts the pattern of duration dependence (Box-Steffensmeier and Jones 1997, Jenkins 2004). The exponential function, $\exp(X_j\beta_x)$, is a function of a vector of person-specific explanatory variables *X*. This function illustrates the effects of differing personal characteristics on exit rate out of unemployment (Jenkins 2004). This provides for the identification of certain characteristics that predispose individuals to longer unemployment durations. Although the baseline hazard summarises the shape of the hazard which is common to all, it is assumed that the level of the hazard may differ across individuals (Kalb 2001, van den Berg 2000, Jenkins 2004, Cleves et al 2004). Furthermore, it is assumed that the determinants of the hazard are multiplicative. Thus, if an elapsed unemployment spell impacts negatively on the hazard i.e. decreases the employment probabilities of cohorts, then this negative effect will be compounded for individuals with personal characteristics that impact negatively on the hazard (van den Berg 2000).

Weibull Model

In order to analyse duration dependence parametrically, a distributional form for the baseline hazard must be specified (Cleves et al 2004). Although there are numerous distributions available[8], this paper considers a Weibull model in the PH metric[9].

The baseline hazard, $h_0(t)$, is specified as

$$h_0(t) = pt^{p-1}.$$

Thus the hazard is defined as

$$h(t \mid X_j) = h_0(t)\exp(X_j\beta_x)$$
$$= pt^{p-1}\exp(X_j\beta_x)$$

The shape parameter, *p,* provides information about the shape of the hazard function and indicates the direction of duration dependence. If $p > 1$, the hazard is monotonically rising

---

[8] As previously mentioned, other parametric distributions include the exponential, the gompertz, log-normal, log-logistic and the generalised gamma model.
[9] The reasons behind this are discussed in the preceding sub-section.

with time, indicating positive duration dependence; conversely if $p < 1$, the hazard is monotonically falling with time, thereby indicating negative duration dependence. Finally, if $p = 1$, the hazard is flat, thus implying an exponential distribution and no duration dependence (Korpi 1995, Box-Steffensmeier and Jones 1997, and Cleves et al 2004).

Proportionality Assumption

The proportionality assumption states that differences in the explanatory variables imply proportional differences in the hazard at each survival time t (Serneels 2002a, Jenkins 2004 and Cleves et al 2004). As discussed by Serneels (2002a), the implication is that a matric qualification has the same impact on the hazard after any number of years of unemployment as it did after 1 year. Serneels (2002a) provides a useful discussion as to when the proportionality assumption would be theoretically justified, namely when the "optimal strategy of the individual is myopic[10]" (Serneels 2002a, p. 22). Myopic behaviour could arise because of high-discount rates and/or the opportunity for repeated-job search i.e. if cohorts are aware that there will be other opportunities to secure employment, there will be a greater tendency to behave myopically. Serneels (2002a) suggests that the unemployed will often have high discount rates as they live in households with low levels of welfare. Furthermore, it is assumed that the youth have both high discount rates and myopic strategies (Serneels 2002a).

Because the model is characterised by this assumption, it is important to assess its validity. If the assumption does not hold, more appropriate modelling choices should be considered. An empirical test of whether the proportional hazards assumption holds is through an analysis of the Schoenfeld residuals.[11] After estimating both the Schoenfeld and scaled Schoenfeld residuals, and performing both the global and variable-by-variable tests, it emerges that there is little evidence to indicate that the proportional hazards assumption has been violated (Stata

---

[10] See van den Berg (2000) for a full discussion of the theoretical justification of the proportional hazards assumption.

[11] "The actual test is based on the findings of Grambsch and Therneau (1994) that the Schoenfeld residuals should have a slope of zero for each covariate." (Serneels 2002a, p. 23) Using the test outlined in Cleves et al (2004) and the Stata Reference Manuel [ST], the researcher is able to test, both globally and for individual covariates, the null hypothesis of zero slope. This is equivalent to testing that the log-hazard-ratio function is constant over time. If the null hypothesis of zero slope is rejected, the proportional-hazards assumption has been violated (Stata 9 Reference Manuel [ST]). For a more theoretical overview, the reader is referred to Lancaster (1990).

14

Reference Manuel [ST]). According to the global test, the null hypothesis can only be rejected at the 39 percent level. Furthermore, according to the variable-by-variable test, the null hypothesis is rejected by only two covariates; the number of employed household members at the 5 percent level, and number of dependents in the household at the 15 percent level. Because the data do not fail the proportionality assumption, we proceed with the Weibull model. To see the results of the proportionality test, the reader is referred to the appendix.

3.2.2   Model Selection

Parametric estimation is useful because it provides a formal means of testing the degree of duration dependence, estimates the impact of explanatory variables on the hazard, and enables the graphical depiction of the hazard rate. As discussed by Serneels (2002b), the problems associated with parametric estimation arise from the specific distributional conditions that are imposed on the data. Because the underlying distributional assumptions made by each model are different, the estimation results are often sensitive to *which* duration model is used, and selecting the wrong parametric specification could result in biased estimates (Serneels 2002b). As there are several parametric models from which to choose[12], it is important to ascertain which model is preferable (Serneels 2002b). In order to discriminate between the different parametric models, two methods will be used. In the case of nested models, both the likelihood-ratio test and the Wald test will be used to determine the most suitable specification. In the case of non-nested models, like the gompertz and log-logistic models, the Akaike information criterion (AIC) [13] will be used (Cleves 2004 and Serneels 2002b).

Following the tests outlined by Cleves et al (2004) and performed by Serneels (2002b), a generalised gamma model is estimated with inverse Gaussian heterogeneity[14]. The gamma distribution includes, as special cases, the Weibull, exponential, and log-normal distributions.

---

[12] Parametric distributions include the exponential, Weibull, gompertz, log-normal, log-logistic and the generalised gamma distribution.
[13] The AIC test penalises the log likelihood of each model to reflect the number of parameters being estimated. The preferred model is the one with the lowest value of the AIC. AIC= -2 ln L + 2(k+c) where k is the number of covariates and c the number of ancillary parameters (Cleves et al 2004). The reader is referred to a footnote discussion by Serneels (2002b) where the relevance of the AIC test is questioned. However, as done by Serneels (2002b), the AIC test will be used in this paper to discriminate between non-nested models.
[14] The significance of controlling for heterogeneity will be discussed in full in section 3.2.3.

Thus, the generalized gamma distribution is used for selecting the most appropriate parametric model (Cleves et al 2004). Using the Wald test, the log-normal model is rejected in favour of the gamma model, at the 10 percent level.[15] In addition to the Wald test, the asymptotically similar likelihood-ratio test is also performed. Since sampling weights and robust standard errors have not been specified at this point, the likelihood-ratio test is considered to be more appropriate than the Wald test (Cleves 2004). Using this test, the log-normal can be rejected in favour of the gamma model at the 1 percent level.[16] This compares favourably against the Wald test. The model is again estimated, and using the Wald test, the Weibull is rejected against the gamma at the 1 percent level.[17] Table 2 below shows that, according to the AIC test, the generalised gamma model scores best, followed by the Weibull model. The exponential model comes third.

Table 2: Preferred model specification according to Akaike Information Criterion

Distribution	Log Likelihood	K	C	AIC	Rank
Exponential	-1021.8363	12	1	2069.673	3
Weibull	-1018.843	12	2	2065.686	2
Gompertz	-1021.6628	12	2	2071.326	5
Log-normal	-1022.5367	12	2	2073.073	6
Log-logistic	-1020.8807	12	2	2069.761	4
Generalized gamma	-1017.7512	12	3	2065.502	1

It is useful to continue with the Weibull specification (which is ranked the second most preferred model) since it has been widely used to model duration dependence in OECD counties, and will thus allow comparison between this paper and previous studies (Serneels 2002b). Furthermore, proportional hazard (PH) models have been more commonly used in the empirical literature where the researcher is attempting to gain insight into the hazard function, and how covariates impact on the hazard. Models in the accelerated failure time (AFT) metric, such as the generalised gamma model, focus more on analysis time (van den Berg 2000). Furthermore, PH models allow for a more straightforward interpretation of the

---

[15] The null hypothesis is: H_0: k=0. If the null hypothesis cannot be rejected, then the log-normal model is preferred to the gamma model. Model estimates: z=1.74 with significance level of 0.081.
[16] Model estimates: LRchi2(1)=9.57with prob=0.0020
[17] The null hypothesis is: H_1: k=1. If the null hypothesis cannot be rejected, then the Weibull model is preferred to the gamma model. Model estimates: chi2(1)=7.36 with prob=0.0067

coefficients (Serneels 2002b). The Weibull is also the simplest parametric form that directly estimates duration dependence (Bennet 1999). In addition, the Weibull encompasses the exponential model (Serneels 2002b), which has been ranked third according to the AIC test.[18]

Despite the "mathematical simplicity of the Weibull" (Arulampalam and Stewart 1995, p.323), the assumption of a monotonic hazard is very restrictive (Korpi 1995). Thus, in subsequent sections this assumption will be relaxed by estimating a piece-wise constant hazard in order to enable for a more flexibly shaped hazard. This will allow for a closer inspection of the course of the hazard (Serneels 2002b).

3.2.3 WEIBULL REGRESSION MODEL

The dependent variable is the duration of unemployment of cohorts measured in months. The covariates are: male, age, age squared, level of education (education has been modelled as a series of dummy variables, namely, primary school, junior secondary school, senior secondary school and matric[19]), tertiary qualification[20], number of children aged 16 years or younger living in the household, number of household members in wage employment, receipt of financial support, participation in unemployment training programs, and finally, proficiency in English.

The binary variable describing whether a cohort is the recipient of financial support requires further discussion. In the KMP survey (2000), respondents where asked to indicate whether they were recipients of various government grants, including the unemployment insurance fund (UIF). Approximately 1.5 percent of respondents were in receipt of UIF. The survey also asked respondents if they benefitted from financial support (from friends or family) that did not have to be repaid. Approximately 31 percent of respondents received financial support of this nature. Woolard and Klasen (2000) estimate that only 3 percent of the unemployed receive unemployment support at any given time, and furthermore, that the

---

[18] Two diagnostics tests based on the deviance and Cox-Snell residuals can be found in the appendix. The tests are diagnostic tools which aid in assessing the appropriacy of a particular specification. Although the test based on the deviance residual confirms that the Weibull model is an appropriate choice, the plot of the Cox-Snell residuals indicates that the gamma distribution fits the data better.

[19] Note that the categories are as follows: primary school: grade 1 – grade 6
Junior secondary school: grade 7 – grade 9
Senior secondary school: grade 10 – grade 11
Matric: Received a Matric certificate

[20] This includes a university degree, diploma, or a technical qualification.

unemployed are largely supported by a network of family and household members. Thus, unemployment assistance is represented in the model in terms of receipts from family and friends, as opposed to the more conventional government unemployment grants used in studies utilising OECD data.

A priori expectations are as follows: The probability of exiting from unemployment is expected to increase with age, albeit at a decreasing rate. Eventually the hazard is expected to decline with age. Men are expected to find jobs sooner than women given the gender bias inherent in the labour market. Those with a higher level of education are expected to exit from unemployment sooner. In addition, the more dependents cohorts are responsible for, the longer the duration of unemployment is expected to be. However, the greater the number of wage earners residing in the household, the sooner is the cohort expected to exit from unemployment. This is largely due to the fact that employed household members could finance the various financial costs associated with job-seeking behaviour. Although it would seem likely that unemployment benefits could also aid the job-search process and thus hasten the exit from unemployment, a number of studies have indicated that unemployment benefits are associated with longer bouts of unemployment (Belzil 1995). In addition, it is anticipated that training programs for the unemployed and proficiency in English will enable a faster transition in employment.

Thus far when discussing the model, it has been assumed that all personal characteristics of individuals were captured by the explanatory variables. However, one of the major issues in this literature concerns the distinction between duration dependence of the hazard rate and unobserved heterogeneity (van den Berg 1996). Unobserved heterogeneity may occur because of omitted variables and/or measurement errors (Jenkins 2004). If unobserved heterogeneity is ignored, there is a tendency for the duration dependence estimate to be biased downwards (Jenkins 2004). For example, assume that a group of individuals with differing personal characteristics are in the preliminary stages of an unemployment spell. Those individuals who posses the characteristics most favoured by employers will leave unemployment in the early stages of the unemployment spell, leaving behind those with less favorable characteristics and lower employment prospects. Thus, it would appear that exit probabilities are negatively related to unemployment duration, and that hazard rates decline throughout the unemployment spell. However, this effect actually "represents changes in the

distribution of unobserved characteristics in the population yet to exit from unemployment" (Kalb 2001), and does not exhibit true duration dependence (Kalb 2001).

The heterogeneity term has a multiplicative effect in the hazard (Serneels 1999). In this case the hazard is defined as

$$h(t/X_j, \alpha_j) = \alpha_j h(t/X_j)$$

where α_j represents an unobserved heterogeneity term indicating that "individuals in the population are heterogeneous due to factors that remain unobserved" (Cleves et al 2004, p.279).

In order to estimate the unconditional hazard function, whereby the hazard is not conditional on the value of α_j (Jenkins 2004), the unobservable heterogeneity term is integrated out of the hazard function (Cleves et al 2004, Serneels 1999). In order to integrate out α_j, a distribution for the unobserved heterogeneity term must be specified. (Cleves et al 2004). The most common specifications for heterogeneity are the Gamma and inverse Gaussian distributions (Jenkins 2004 and van den Berg 2000) [21]. In this paper, the inverse Gaussian specification is used in order to make the estimation results comparable to those of Serneels (2002a and 2002b). Thus, the estimated models control for unobserved heterogeneity, using an inverse Gaussian distribution.[22]

The parameter estimates and standard errors of the Weibull regression model are presented in Table 3 on the following page. Note that in all models estimated, sampling weights have been used and standard errors have been adjusted for clustering.

---

[21] For a more in-depth and technical discussion of heterogeneity models, the reader is referred to Cleves et al (2004).

[22] For a more complete and technical discussion, the reader is referred to Cleves et al (2004), Jenkins (2004), van den Berg (2000), and the Stata 9 Reference Manuel [ST]. Further helpful discussion may be found in Serneels (2002b).

Table 3: Weibull Model Estimates of the Rate of Transition Out of Unemployment

Variable	Beta Coefficient	s.e.
Constant	-8.6441***	1.064
Male	0.7665***	0.220
Age	0.1246***	0.046
Age squared	-0.0016***	0.001
Primary schooling	-0.7651**	0.325
Junior Secondary	-0.5242 **	0.245
Senior Secondary	-0.2837	0.258
Tertiary Qualification	0.5660**	0.231
Dependents aged 16 or less	-0.2387***	0.083
Number of wage earners in household	1.3082***	0.232
Recipient of financial support	-1.1710***	0.303
Training program for unemployed	-0.2863	0.269
English proficiency	0.5231**	0.236
Duration Dependence		
P	1.2814	0.206
Ln(p)	1.54	
Probablity	0.123	
Log pseudo-L[23]	-176179.54	
Number of observation	1362	

Source: Khayelitsha/Mitchell's Plain Survey
  *** significant at 1% level
  **   significant at 5% level
  *   significant at 10% level

The Weibull coefficient estimates are presented in a proportional hazard metric. Accordingly, a negatively signed coefficient implies a decrease in the hazard, while a positively signed coefficient implies an increase in the hazard (Box-Steffensmeier and Jones 1997).

From the results, it is apparent that men find jobs far more easily than women. The latter therefore have a higher permanency of unemployment. Casale (2004) explains this occurrence by noting that, since 1995, female labour market participation has increased dramatically, outstripping the increase in male participation rates. For the period 1995 to 2001, the working age female population increased by 13.2 percent, while the broadly

---

[23] Note that there is a log pseudo-likelihood as opposed to a log likelihood because sample weights and adjustments for clustering were included during model estimation.

defined female labour force increased by 52.2 percent (3.2 million). Thus, a large number of women who were not economically active in 1995 had joined the labour force by 2001. During the same period, the male working age population increased by 12.2 percent, while the broadly defined male labour force increased by 22.9 percent. Males and females thus experienced similar growth rates in the working age population, but the relative growth in female labour force participation was twice that of men. The increased labour market participation of women has translated into increased levels of unemployment, with 65 percent of the additional female labour force participants broadly unemployed in 2001 (Casale 2004).

Statistics South Africa (2002) analysed the labour market experiences of men and women for the period 1995 to 2001. Their results confirm the Weibull estimate in that, within each population group, a lower proportion of working aged women is employed, and a higher portion is classified as NEA, relative to working aged males. In 2001, employment statistics for males were reported as 73, 58, and 43 percent for White, Coloured, and African males, respectively. For the same period, in each population group, females exhibited lower employment rates than their male counterparts. White females displayed the highest rates of employment followed by Coloured and then African females, at 54, 45, and 36 percent respectively. In addition, within each population group, women experience higher rates of unemployment than do males (Statistics South Africa 2002).

The age and age-squared variables impact significantly on the hazard and their coefficients have the expected sign. The probability of employment increases with age, but at a decreasing rate. At some point over the life-cycle, the probability of employment will begin to decline with age. Wittenberg (2001) notes that the transition from school to labour market participant takes longer for Black males relative to other racial groups. "In their mid-twenties approximately 30 to 40 percent of Black males are still classified as "not economically active". The net flows out of this category only cease at about age thirty." (Wittenberg 2001, p. 3) Furthermore, Wittenberg (2001) notes that, once classified as economically active, Black males are slower to find employment relative to other racial groups. "As a result one sees a build up in the proportion of the unemployed (whether searching or not)" (Wittenberg 2001, p. 3). Wittenberg finds that the highest levels of unemployment occur at approximately age 27. Comparable results are found using the KMP data, where the greatest proportions of unemployment were found for those aged between 20 and 25, peaking at 25, after which the rate was found to decline significantly. Wittenberg (2001) suggests that the subsequent

decline in unemployment levels beyond age 27 occurs because cohorts are being absorbed into work (Wittenberg 2001). Moreover, the narrow unemployment rate declines significantly beyond age 40. Wittenberg (2001) suggests that the tapering off of the narrow rate occurs because a portion of the searching unemployed stop actively searching for employment beyond age 40, indicating that search intensity declines with age (Wittenberg 2001). Using the KMP data, this result is implied by the fact that only 2.3 percent of the searching unemployed are aged 40. 50 year olds represent 0.69 percent of the searching unemployed, while 55 year olds represent just 0.2 percent of the searching unemployed.

With respect to the education variables, the reference category is matric. It is evident that cohorts not holding a matric certificate will experience longer bouts of unemployment. Using the hazard ratio[24], the hazard rate for those with a primary level of schooling is 47 percent of the hazard rate of matric graduates, while the hazard rate for those with a junior secondary level of schooling is 59 percent of the hazard rate of matric holders. Finally, the hazard rate for those with incomplete senior secondary education is 75 percent of the hazard rate of matric graduates, although this result is insignificant. A tertiary qualification increases the rate of entry into employment. The hazard rate for those with a tertiary qualification is 76 percent higher relative to those without a tertiary qualification. Thus, individuals with a tertiary level education are more likely to leave unemployment sooner. Burger et al (2005) find that schooling has a significant impact on both urban and rural labour market participation and employment. More specifically, both participation and employment were found to be highest for those with a tertiary qualification. Those in possession of a secondary level qualification faired second best, followed by those with less than a secondary education (Burger et al 2005). Bhorat and Leibbrandt (1999b) find that cohorts with lower levels of education have lower employment probabilities relative to those with high levels of education, specifically a tertiary qualification. This reflects the growing demand for more highly skilled labour in the South African labour market (Bhorat and Leibbrandt 1999b).

An additional dependent aged 16 years or less living in the household decreases the transition rate. The hazard rate for those responsible for an additional dependent is 79 percent of the hazard rate of those that are not. Bhorat and Leibbrandt (1999a) find that women with one or more children aged under six are less likely to be employed than women with no children in

---

[24] The hazard ratio is calculated as an exponentiated coefficient i.e. to calculate the hazard ratio for those whose highest level of schooling is primary school: exp (-0.7384).

this age bracket (Bhorat and Leibbrandt 1999b). Mlatsheni and Rospabe (2002) find that cohorts aged between fifteen and thirty are less likely to be employed, and more likely to be self-employed, when taking care of children aged six or less (Mlatsheni and Rospabe 2002). Bhorat and Leibbrandt (1999b), not unexpectedly, find a gender bias in child rearing. The number of children living in the household has no significant effect on male labour market participation decisions. For females, however, an additional child under the age of seven, or aged between eight and fifteen decreases their probability of participating in the labour market (Bhorat and Leibbrandt 1999b).

Living with an additional wage earner will shorten unemployment duration, as an additional employed household member significantly increases the probability of exiting from unemployment. Dinkelman (2004) comments that the household acts as a "privileged source of information" (Dinkelman 2004, p. 489) for the unemployed, as employed household members serve as links to the labour market. Wittenberg (2001) indicates than an older, employed, household member serves to increase the probability of employment in excess of 5 percent for men, and 10 percent for women (Wittenberg 2001). This result is important in light of the fact that a significant proportion of the unemployed (broadly defined) live in households where there are few or no wage-earners. In 2002, 48.6 percent of the unemployed lived in households with no wage-earners. This proportion has increased from 42 percent in 1995. In 2002, less than 12 percent of the unemployed lived in households with two or more wage-earners. Alarmingly, this proportion has decreased from 16.6 percent in 1995. The unemployed are therefore becoming increasingly marginalised; residing in households with little or no access to wage-income (Bhorat and Oosthuizen 2005), and thus, according to the above results, will experience a greater permanency of unemployment.

The receipt of financial support (that does not have to be repaid) serves to decrease the transition rate. The hazard ratio implies that those who are receiving financial support have a hazard rate that is only 31 percent of the hazard faced by a job-seeker that does not. This result conforms to the findings of OECD countries whereby unemployment benefits are found to impact negatively on the exit rate (Serneels 2002b). In a South African analysis, Bhorat and Leibbrandt (1999b) find that pension and other income sources reduce the probability of labour market participation, for both males and females (Bhorat and Leibbrandt 1999b). Serneels (2002b) provides an explanation for this seemingly counterintuitive result. When unemployment benefits are of a limited duration, the unemployed recipient will

increase their probability of finding employment as the termination period of the benefit approaches. For example, they could increase their job-search and lower their reservation wage. Conversely, the more financial support an unemployed individual receives from family and friends, the longer the job-search process can be postponed. This translates into longer unemployment durations.

The estimated result for respondents that have received some training while unemployed is contrary to what was expected and is likely the result of a sample selection bias. Respondents that have undergone such training have a lower probability of exiting from unemployment, and thus will experience longer unemployment durations. This result is insignificant, which may be due to the small sample of respondents, under 7 percent, who have participated in training programmes while unemployed. Those respondents that completed unemployment training programmes represent a selection of individuals with lower employment prospects relative to those individuals that have not participated in an unemployment training programme and therefore have the longest duration of unemployment. Of the respondents that completed an unemployment training programme, none has a tertiary qualification, whereas approximately 15 percent of the respondents who did not participate in any training programmes have attained a tertiary qualification. The Weibull estimates regarding education indicate that those with a tertiary qualification are substantially more likely to exit unemployment relative to those without a qualification. In addition, it is questionable whether some of programmes undertaken by respondents are really beneficial in terms of developing the skills and human capital necessary for employment in the labour market. Such programmes include drama training, job interview skills, drawing and art, and first aid training programmes. Less than 40 percent of respondents indicated that they got a job using the skills they learned on their respective courses.

Proficiency in English increases the rate of entry into employment. Respondents who are proficient in English have a hazard rate that is 69 percent greater than those who do not possess a workable level of fluency in the language. Wittenberg (2001) suggests that the level of literacy within the household impacts on the ability of the household member to access information about available jobs and employment opportunities (Wittenberg 2001). This result is important when viewed in the light of the 2001 Census results published by Statistics South Africa, which indicate that just over 8 percent of respondents spoke English.

As for duration dependence, the parameter p, being slightly greater than 1, implies slight positive duration dependence[25]. By testing whether $p=1$ (ln (p) = 0), it can be determined whether the exponential model is in fact a more appropriate specification. As the test statistic is 1.54 the null hypothesis can be rejected at the 15 percent level. Thus the Weibull distribution is preferred to that of the exponential at p=0.123 (Cleves et al 2004).

The hazard rate is plotted in Figure 4 on the following page. The Weibull model predicts a hazard rate that increases, reaches a maximum, and then, decreases sharply. The maximum occurs at approximately 13 months of unemployment. At this maximum, merely 17 percent of the unemployed have left unemployment. This result is more startling when compared to the hazard rate estimated by Serneels (2002b). The hazard "first rises and then falls" (Serneels 2002b, p. 24), but reaches a maximum at approximately 6 years of unemployment. Serneels notes that by the time the 6 year turning point has been met, more than 80 percent of the unemployed have already left unemployment. This means that the vast majority of the unemployed "face an increasing hazard" (Serneels 2002b, p. 24). The hazard rate in Figure 4 reaches a turning point at 13 months, at which time 83 percent of respondents are still in a state of unemployment. Thus, it is evident that the majority of the unemployed face a decreasing hazard whereby the probability of transitioning into employment decreases as unemployment duration increases.

Although the monotonic nature of the Weibull hazard rate makes it is less flexible than the non-parametric estimate plotted in Figure 3, the Weibull hazard rate compares favourably to the smoothed hazard function. The Weibull hazard rate reaches a maximum at 13 months, at which the probability of leaving unemployment is below 8 percent. The smoothed hazard function indicates that the hazard reaches a maximum at approximately 2 years, at which point the probability of exiting from unemployment is below 9 percent. Thereafter the smoothed hazard function follows a more complex course of rising and falling.

---

[25] The frailty model, controlling for Inverse Gaussian heterogeneity, is preferred to that of the no-frailty model. The Likelihood-ratio test of $H_0 : \theta = 0$ can be rejected at the 1 percent level. The importance of controlling for unobserved heterogeneity is important considering that the non-frailty model indicated that duration dependence was slightly negative (Stata 9 Reference Manuel [ST] and Cleves et al 2004).

Figure 4: Weibull hazard rate
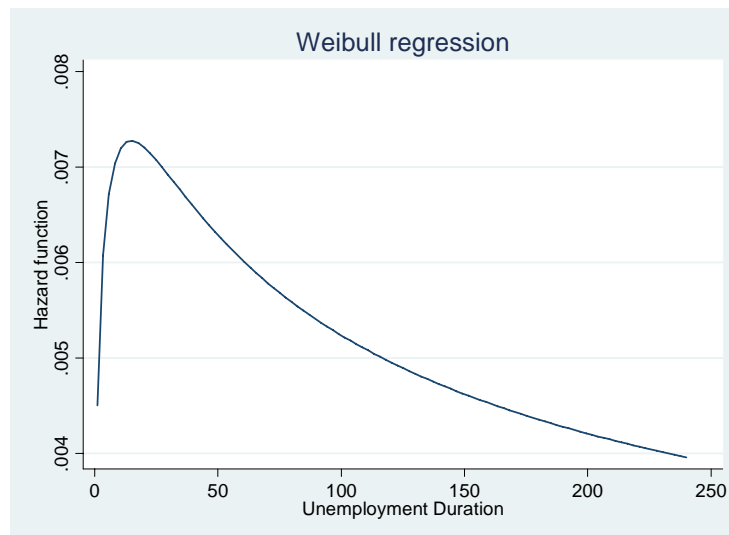
3.3 SEMI-PARAMETRIC ESTIMATION

3.3.1 Piece-wise Constant Regression Model

The disadvantages of using a Weibull specification have been previously mentioned. In particular, the monotonic hazard imposed by the Weibull model is fairly restrictive (Korpi 1995), and biased estimates may arise from specifying an inappropriate distribution (Serneels 2002b). In the following sub-section a semi-parametric model is estimated, namely, the flexible piece-wise constant model, which allows for a "closer analysis of the development of the hazard rate" (Korpi 1995, p.357). An advantage of the piece-wise constant hazard model is that it does not require the shape of the hazard function to be imposed in advance (Jenkins 2004). For example, whether the hazard increases or decreases with unemployment duration is ascertained from the data, rather than specified *a priori*, enabling a closer analysis of the hazard rate (Jenkins 2004).

The piece-wise constant hazard model allows the hazard to shift over time by incorporating dummy variables for different time intervals (Serneels 2002a). The hazard rate is assumed to be constant within each interval but may vary between intervals (Jenkins 2004).The time intervals have been grouped into eight categories: that is 1-12 months, 1-2 years, 2-4 years, 4-6 years, 6-8 years, 8-10 years, 10-12 years, and, beyond 12 years. Note that the point of reference for the time intervals is 1-12 months. Thus all estimates and p-levels reported in Table 4 below are relative to this initial period. The estimates of the piece-wise constant hazard model are presented in Table 4 on the following page.

Table 4: Piece-wise Constant Hazard Estimates of the Rate of Transition out of Unemployment

Variable	Beta Coefficient	s.e.
Constant	-6.8966***	0.614
Male	0.4995***	0.111
Age	0.0948***	0.029
Age squared	-0.0012***	0.000
Primary schooling	-0.4787**	0.110
Junior Secondary	-0.3121**	0.139
Senior Secondary	-0.1761	0.156
Tertiary Qualification	0.3797***	0.123
Dependents aged 16 or less	-0.1679***	0.048
Number of wage earners in household	0.8252***	0.056
Recipient of financial support	-0.8379***	0.172
Training program for unemployed	-0.1335	0.170
English proficiency	0.4013**	0.171
Piece-Wise Constant Hazards		
One to Two Years	-0.5234***	0.155
Two to Four Years	-0.3704***	0.134
Four to Six Years	-1.0691***	0.226
Six to Eight Years	-1.3540***	0.329
Eight to Ten Years	-0.8185***	0.272
Ten to Twelve Years	-0.3771	0.327
Beyond Twelve Years	-0.9169**	0.354
Log pseudo-L[26]	-173756.26	
Number of observation	3473	

Source: Khayelitsha/Mitchell's Plain Survey

---

[26] note that there is a log pseudo-likelihood as opposed to a log likelihood because probability weights have been included and adjustments have been made for clustering

The effects of the variables included in the Weibull model have remained similar, although the variables have lower p-values and smaller standard errors in the piece-wise model[27]. The effects of the new dummy variables will therefore be the primary focus of this analysis. The effects of the time intervals are more immediately meaningful when examined in terms of the hazard ratio.[28] The probability of becoming employed in years 1-2 is 59 percent of that of becoming employed in months 1-12, whereas the probability of becoming employed in years 2-4 is 69 percent of that of becoming employed in months 1-12. The hazard is thus at a maximum in the first 12 months of unemployment. This result is consistent with the predictions of the initial Weibull model, which indicated that the hazard reaches a maximum at approximately 13 months.

It is already apparent that the hazard depicted by the piece-wise model follows a more flexible course than the monotonic hazard prescribed by the Weibull model. The piece-wise hazard initially falls in years 1-2 and then rises by 10 percent in the 2-4 year period. Thereafter the hazard follows a more flexible pattern of rising and falling. The chance of finding a job after 4-6 years of unemployment is 66 percent less than that of the probability found during the first 12 months of unemployment, whilst the probability of becoming employed in years 6-8 is 74 percent lower than the probability during the first 12 months into an unemployment spell. Employment probabilities in years 8-10 are 56 percent lower than employment probabilities in the first year of unemployment, while the exit rate in years 10-12 is 31 percent less than that of the exit rate corresponding to the first year of unemployment. This result is insignificant, however. Beyond 12 years, the probability of finding work is 60 percent lower than that of the probability during the first year of unemployment.

The piece-wise constant seems to follow more closely the smoothed non-parametric hazard function depicted in Figure 3. When looking at the estimates of the dummy variables, although there is a general pattern that shows that the hazard initially reaches a maximum

---

[27] With the exception of the variable denoting whether a respondent has been on a training programme during unemployment. This variable has a higher p-value in the piece-wise model, but is insignificant in both models.
[28] Once again, the hazard ratio is calculated as exp (coefficient).

turning point and then decreases, it does not appear to be monotonic in nature, as suggested by the Weibull model.

4.   <u>CONCLUSION</u>

As much of the literature examining duration dependence has been conducted using OECD data, there are few empirical studies on duration dependence in a developing country context (Serneels 2002b). This analysis investigates whether duration dependence is prevalent in the Western Cape labour markets using data from the Mitchell's Plain Magisterial district. The analysis began with a discussion of unemployment definitions. It became clear, when looking at the data, that a perfunctory application of the ILO unemployment definition underestimates the unemployment problem. Thus the unemployment classification was assumed to incorporate both the broad and narrowly defined unemployed. Duration dependence was examined using non-parametric, parametric, and semi-parametric methods.

The parametric analysis was conducted with the use of a Weibull regression model, which enabled for the identification of demographic groups more prone to long bouts of unemployment. The Weibull estimates indicate that long term unemployment will be more prevalent for females who have low levels of schooling, limited proficiency in English, and are responsible for the care of children and dependents aged less than 16. This effect is compounded when the unemployed are marginalised and live with few or no wage earners.

As the shape parameter is greater than 1, slight positive duration dependence was found. The implication is that as unemployment duration increases, the probability of finding employment (at the least) will not decrease. This result, when taken out of context however, was found to be misleading. A graphical depiction of the Weibull hazard rate indicates that although the hazard is initially upward sloping, it reaches a maximum at approximately 13 months of unemployment. At this point, merely 17 percent of respondents have exited from unemployment. Thus, the majority of the unemployed face a decreasing hazard. More disturbing, is that even at its highest, the hazard remains below 8 percent, while the long-term unemployed face an exit probability that is substantially lower.

Due to the restrictive assumption of a monotonic hazard, a piece-wise constant model was estimated, which enabled a more thorough analysis of the course of the hazard (Korpi 1995).

The piece-wise model confirmed much of the results obtained from the Weibull. The coefficient estimates remained significantly similar (although p-values and standard errors were smaller), and the hazard rate was found to peak within the first 12 months. Thereafter, the hazard followed a more flexible shape of increasing and decreasing. The dummy intervals also reflected that the probability of employment for the long-term unemployed is low, relative to the first 12 months of unemployment.

The aim of this paper has been to provide meaningful commentary on the employment probabilities of the long-term unemployed. It has been shown that the long-term unemployed have low and decreasing employment probabilities, irrespective of the model used. Furthermore, this analysis has enabled for the identification of demographic groups who are most prone to long bouts of unemployment, making it clear where government policy should be directed. Government labour market policies must take a two-tiered approach. Firstly, government policies must be directed towards identifying individuals in the "high-risk" demographic who are most at risk of lapsing into long-term unemployment. For example, young women with relatively short unemployment durations and unfavourable employment prospects (such as incomplete secondary schooling and limited English skills) could be signalled out by employment agencies. Secondly, once high-risk candidates are identified, policy should be directed towards preventing such individuals from becoming long-term unemployed by providing them the opportunity to participate in practical and meaningful unemployment training programmes. As previously mentioned, less than 40 percent of respondents who received training whilst unemployed indicated that they got a job using the skills they learned on their respective courses. Furthermore, it seems questionable whether some of the courses undertaken by the respondents were beneficial in terms of developing the skills and human capital necessary to improve their employment prospects.

5.   BIBLIOGRAPHY

- Arulampalam W., Stewart M. (1995), The Determinants of Individual Unemployment Durations in an Era of High Unemployment, The Economic Journal, Vol. 105:321-332

- Belzil C. (1995), Unemployment Insurance and Unemployment Over Time : An Analysis with Event History Data, The Review of Economics and Statistics, Vol. 77:113-126

- Bennet D. (1999), Parametric Models, Duration Dependence, and Time-Varying Data Revisited, American Journal of Political Science, Vol. 43:256-270

- Bergstrom R., Edin P. (1992), Time Aggregation and the Distributional Shape of Unemployment Duration, Journal of Applied Econometrics, Vol. 7:5-30

- Bhorat H, Leibbrandt M (1999a), Correlates of Vulnerability in the South African Labour Market, Development Policy Research Unit, Working Paper 99/27 ISBN 0-7992-1967-3

- Bhorat H, Leibbrandt M (1999b), Modelling Vulnerability and Low Earnings in the South African Labour Market, Development Policy Research Unit, Working Paper 99/32 ISBN 0-7992-1977-0

- Bhorat H., Oosthuizen M. (2005), The Post-Apartheid South African Labour Market, Development Policy Research Unit, Working Paper 05/93 ISBN 1-920055-09-6

- Box-Steffensmeier J., Jones B. (1997), Time is of the Essence: Event History Models in Political Science, American Journal of Political Science, Vol. 41:1414-1461

- Box-Steffensmeier J., Zorn C. (2001), Duration Models and Proportional Hazards in Political Science, American Journal of Political Science, Vol. 45:972-988

- Burger R., Leibbrandt M., Mlatsheni C., van der Berg S. (2006), Migration and the changing rural-urban interface in South Africa: What can we learn from census and survey data, unpublished manuscript

- Casale D. (2004), What has the Feminisation of the Labour Market "Bought" Women in South Africa? Trends in Labour Force Participation, Employment and Earnings, 1995-2001, Development Policy Research Unit, Working Paper 04/84 ISBN 0-7992-2239-9

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

- Crankshaw O., Welch M. (2003), Sample Method: The Khayelitsha/Mitchell's Plain 2000 Survey, CSSR

- Dinkelman T. (2004), How Household Context Affects Search Outcomes of the Unemployed in KwaZulu-Natal, South Africa: A Panel Data Analysis, South African Journal of Economics, Vol. 72:3

- Elbers C., Ridder G. (1982), True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model, The Review of Economic Studies, Vol. 49:403-409

- Henderson R., Oman P. (1999), Effect of Frailty on Marginal Regression Estimates in Survival Analysis, Journal of Royal Statistical Society, Vol. 61:367-379

- Jenkins S. (2004), Survival Analysis, Unpublished Manuscript, Institute for Social and Economic Research, University of Essex, Colchester, UK

- Kalb G. (2001), The impact of social policy initiatives on labour supply incentives: A review of the literature, Policy Research Paper No. 18, Melbourne Institute of Applied Economic and Social Research, The University of Melbourne
  Available at the following website:

http://www.facs.gov.au/research/prp18/PRP_No_18.pdf

- Kasper H. (1967), The Asking Price of Labor and the Duration of Unemployment, The Review of Economics and Statistics, Vol. 49:165-172

- Korpi T. (1995), Effects of Manpower Policies on Duration Dependence in Re-employment Rates: The Example of Sweden, Economica, Vol. 62:353-371

- Lancaster T. (1990), The Econometric Analysis of Transition Data, Cambridge University Press

- Mlatsheni C., Rospabe S. (2002), Why is youth unemployment so high and unequally spread in South Africa? Working paper, Development Policy Research Unit, University of Cape Town, DPRU Working Paper No. 02/65

- Natrass N. (2003), Unemployment, Employment and Labour-Force Participation in Khayelitsha/Mitchell's Plain, CSSR Working Paper No. 12

- Natrass N. (2002), Do-file, Data First Resource Unit

- Serneels P. (2002a), The Nature of Unemployment in Urban Ethiopia, Centre for the Study of African Economies, Oxford University

- Serneels P. (2002b), Explaining Non-Negative Duration Dependence Among the Unemployed, The Centre for the Study of African Economies Working Paper Series, Centre for the Study of African Economies, Oxford University, Paper 172

- Statistics South Africa (2002), Women and Men in South Africa Five Years On

- Trivedi P., Hui W. (1987), An Empirical Study of Long-Term Unemployment in Australia, Journal of Labour Economics, Vol. 5: 20-42

- van den Berg J. (2000), Duration Models: Specification, Identification, and Multiple Durations in *Handbook of Econometrics*, *Volume V*

- van den Berg J., Grogan L. (1999), The Duration of Unemployment in Russia

- van den Berg G., van Ours J. (1994), Unemployment Dynamics and Duration Dependence in France, the Netherlands and the United Kingdom, The Economic Journal, Vol. 104:432-443

- van den Berg G., van Ours J. (1996), Unemployment Dynamics and Duration Dependence, Journal of Labour Economics, Vol. 14:100-125

- van den Berg G., van Ours J. (1999), Duration dependence and heterogeneity in French youth unemployment durations, Journal of Population Economics, Vol. 12:273-285

- Walker R. (2003), Reservation Wages- Measurement and Determinants: Evidence from the Khayelitsha/Mitchell's Plain (KMP) Survey, CSSR Working Paper No. 38

- Walker R., Do-file, Data First Resource Unit

- Wittenberg M. (2001), Spatial Dimensions of Unemployment, Paper presented at the Development Policy Research Unit/Friedrich Ebert Stiftung (FES) Conference on Labour Markets and Poverty in South Africa

- Yoon B. (1985), A non-stationary hazard function of leaving unemployment for employment, Economics Letters, Vol. 17:171-175

6. <u>APPENDICES</u>

6.1 ASSESSING MODEL FIT

A diagnostic test that is used to assess the appropriateness of a distribution is to plot the deviance residuals. Plots of the deviance residuals against unemployment duration can be used to assess the fit of the model. Table 5, on the following page, depicts plots of the deviance residuals obtained after fitting the four most preferred models according to the AIC test. In all plots, the deviance residuals are large for short durations but decrease with analysis time, suggesting that the probability of exiting from unemployment is underestimated for cohorts with short survival durations, and is overestimated for cohorts with longer unemployment spells. The deviance residuals obtained from the Weibull model are comparable to the residuals obtained from the exponential, gamma and log-logistic models and the Weibull does not appear to fit the data more poorly than any of the other distributions. (Stata 9 Reference Manuel [ST] and Serneels 2002b)

Table 5:   Deviance Residuals Diagnostic Test
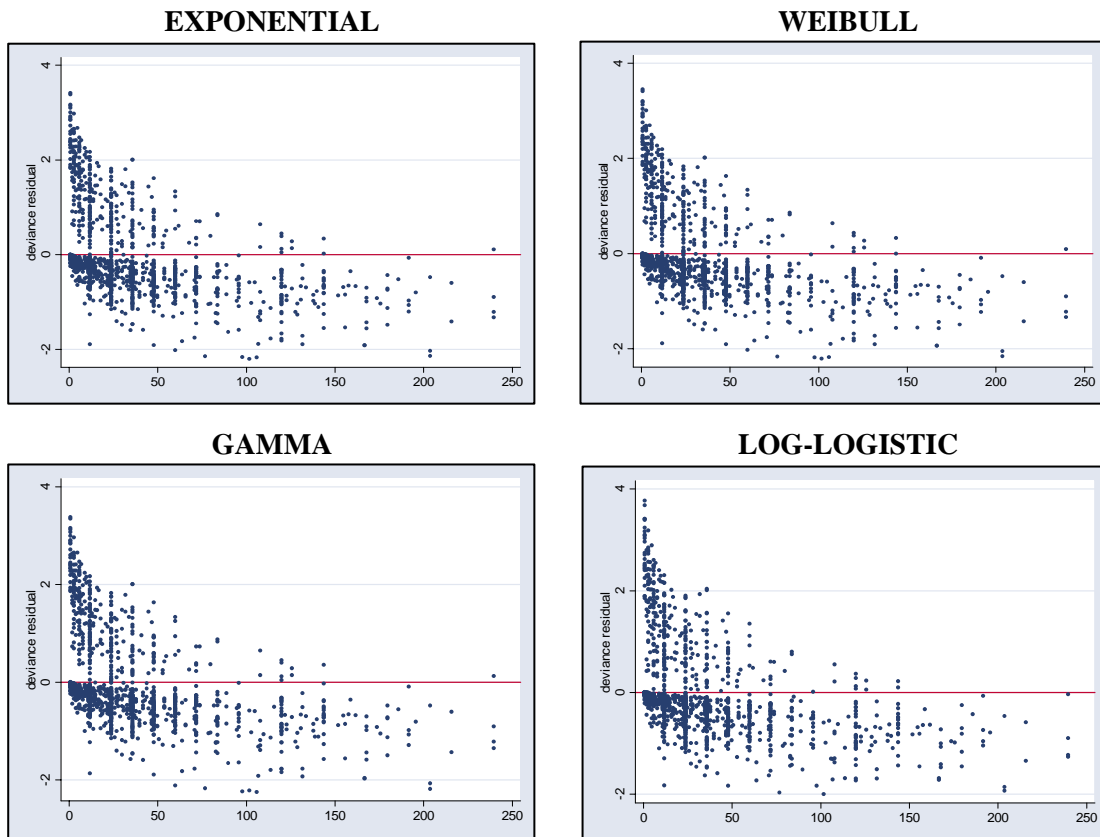
The Cox-Snell residuals, when plotted against the cumulative hazard rate i.e. based on the Kaplan-Meier survival estimates), can also be used as a diagnostic tool to assess model fit. The Cox-Snell residuals are defined as the estimated cumulative hazard function obtained from the fitted model (Serneels 2002b quoting Cox and Snell 1968). If the estimated model fits the data, the plot should be a straight line with a slope of 1 (Stata 9 Reference Manuel [ST] and Serneels 2002b).
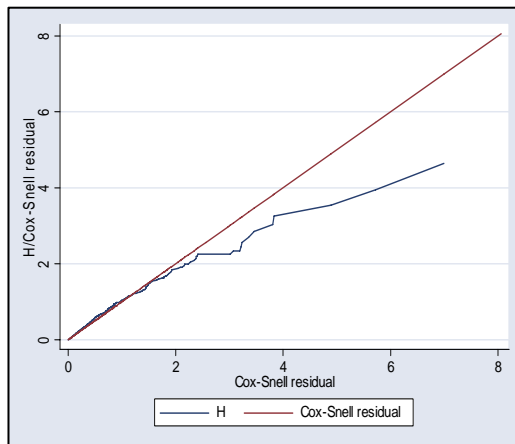
In Table 6, on the following page, it is clear that the Weibull does not fit perfectly for long durations. Although the Weibull fits the data as well as the exponential distribution, the gamma and log-logistic distribution appear to fit better.

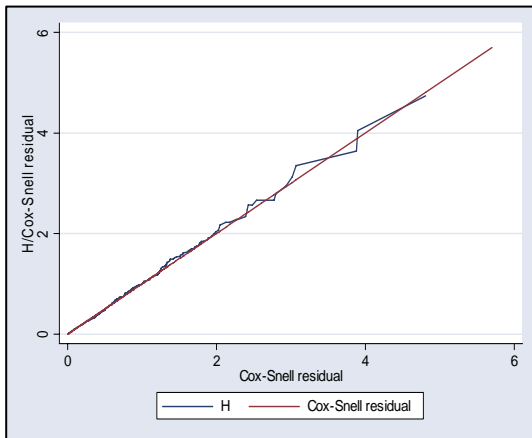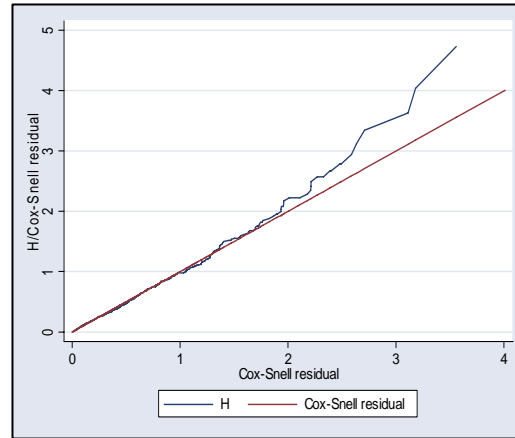**Table 6:**   Cox-Snell Residuals Diagnostic Test
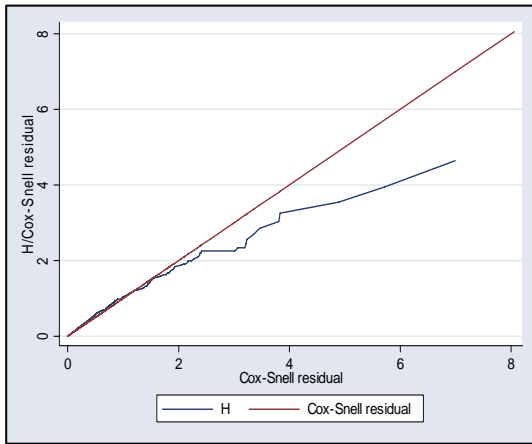
6.2   TESTING THE PROPORTIONAL HAZARDS ASSUMPTION

Table 7:   Proportionality test based on the Schoenfeld residuals

Variable-by-variable test	Prob
Male	0.81
Age	0.48
Age squared	0.55
Primary schooling	0.20
Junior Secondary	0.31
Senior Secondary	0.33
Tertiary Qualification	0.47
Dependents aged 16 or less	0.13
Number of wage earners in household	0.05
Recipient of financial support	0.21
Training program for unemployed	0.64
English proficiency	0.23
Global test	**0.38**

According to the global test, depicted in Table 7 above, the null can only be rejected at the 38 percent level. Thus the proportionality assumption does hold.

The Southern Africa Labour and Development Research Unit

The Southern Africa Labour and Development Research Unit (SALDRU) conducts research directed at improving the well-being of South Africa's poor. It was established in 1975. Over the next two decades the unit's research played a central role in documenting the human costs of apartheid. Key projects from this period included the Farm Labour Conference (1976), the Economics of Health Care Conference (1978), and the Second Carnegie Enquiry into Poverty and Development in South Africa (1983-86). At the urging of the African National Congress, from 1992-1994 SALDRU and the World Bank coordinated the Project for Statistics on Living Standards and Development (PSLSD). This project provide baseline data for the implementation of post-apartheid socio-economic policies through South Africa's first non-racial national sample survey.

In the post-apartheid period, SALDRU has continued to gather data and conduct research directed at informing and assessing anti-poverty policy. In line with its historical contribution, SALDRU's researchers continue to conduct research detailing changing patterns of well-being in South Africa and assessing the impact of government policy on the poor. Current research work falls into the following research themes: post-apartheid poverty; employment and migration dynamics; family support structures in an era of rapid social change; public works and public infrastructure programmes, financial strategies of the poor; common property resources and the poor. Key survey projects include the Langeberg Integrated Family Survey (1999), the Khayelitsha/Mitchell's Plain Survey (2000), the ongoing Cape Area Panel Study (2001-) and the Financial Diaries Project.

Southern Africa Labour and Development Research Unit
School of Economics
University of Cape Town
Private Bag, Rondebosch, 7701
Cape Town, South Africa

Tel: +27 (0) 21 650 5696
Fax: +27 (0) 21 650 5697

Email: brenda.adams@.uct.ac.za
Web: www.saldru.uct.ac.za