University of Cape Town

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

The South African State Old Age Pension:

A Reconsideration of the Effects of the State Old Age Pension on the Living Arrangements of the Elderly in South Africa

Lara Bustin

A dissertation submitted to the Faculty of Science, University of Cape Town, in partial fulfilment of the requirements for the degree of Master of Science.

Cape Town, 2006

Supervised by:

Cally Ardington and Murray Leibbrandt

University of Cape Town

CONTENTS

University of Cape Town

University of Cape Town

Declaration

I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.

I have used the Harvard convention for citation and referencing. Each contribution to, and quotation in, this thesis from the work's of other people has been attributed, and has been cited and referenced.

This thesis is my own work.

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.

Signature ______________________

University of Cape Town

Abstract

The South African State Old Age Pension forms the backbone of social assistance in South Africa. The impact of this pension on its recipients has consequently come under much scrutiny. Previous research on the pension has been primarily concerned with estimating the behavioral effects of the pension on the members of pension households. However, because these studies have used cross-sectional data, the findings have been predicated upon the assumption that the composition of pension households does not change on pension receipt. This paper questions the plausibility of this assumption by examining the effect of the South African State Old Age Pension on the living arrangements of recipients. The results support the argument that the living arrangements of pensioners *will* be affected by pension receipt, and that the results from past studies of the pension may consequently have been misinterpreted. In particular, this study finds that the number of children, the number of adults aged 24-29, the number of men aged 30-39 and the number of adults aged 40-49 in a household will change on pension receipt. The study concludes that, by considering the relationship between the pension and household composition, studies based cross-sectional data may be interpreted more accurately. The study employs two different methods to achieve its aim, namely a regression discontinuity approach using 2001 census data and a cohort approach using a series of cross-sectional data from 1994-2003.

University of Cape Town

LIST OF FIGURES

University of Cape Town

University of Cape Town

LIST OF TABLES

University of Cape Town

1. Introduction

The South African State Old Age Pension (SOAP) forms the backbone of social assistance in South Africa. The SOAP is not only the most generous social grant in this country, but it is also the most far-reaching. In 2001, the SOAP amounted to approximately twice the median per capita income of black South African households (South African Population and Housing Census, 2001). In the same year, just less than 80% of all age-eligible Africans received this grant, and consequently, the SOAP reached over 20% of all African households (South African Population and Housing Census, 2001). In a country with an unemployment rate close to 30%, and one of the highest AIDS rates in the world, the size and widespread distribution of the SOAP has made it an important and, in many cases, vital source of income to black South African households. As a result, the interest in the impact of the SOAP on the behaviour of these households has been considerable.

Several researchers have documented the impacts of the SOAP on the behaviour of members of pension households[1]. Amongst other things, the pension has been shown to have a significant effect on the expenditure, health, education, labour supply and the living arrangements of members of a pension household (Maitra and Ray (2003), Case and Deaton (1998), Duflo (2003), Edmonds (2005), Bertrand, Mullainathan and Miller (2003), Posel, Fairburn and Lund (2004) and Edmonds, Mammen and Miller (2005)). This paper focuses on the effect of the pension on living arrangements, where living arrangements are defined in terms of the individuals who reside in the pension household. More specifically, the study aims to determine whether individuals of different ages and gender join or leave the pension household owing to pension receipt. Though some research has dealt with this issue in the past (Edmonds et al (2005)), not enough is known about the dynamics of the living arrangements of pensioners and their extended families around the arrival of the pension and further research is warranted.

University of Cape Town

[1] For the purposes of this study, pension households are defined as households in which one or more members receive(s) the SOAP.

The living arrangements of black South Africans are unusually complex. Black South African families have historically been divided across several households. Under the previous Apartheid regime, the land ownership of Africans (a population group comprising 80% of the South African population) was restricted to no more than 13% of the land area of the country. As a result, few African families could continue to live by subsistence means alone and several prime-age Africans were forced to migrate to urban areas in search of work. Despite the cessation of the Apartheid regime, labour migration in South Africa is still common today (Casale and Posel (2003)) and thus it is normal for families to be dispersed across two or more households. This trend has persisted partly owing to the high unemployment in South Africa and the financial difficulties faced by most black South Africans today.

The AIDS epidemic in South Africa has further complicated the dynamics within African families. The AIDS rate in South Africa is one of highest in the world, with prime-age African adults being the worst hit by this epidemic. Coupled with high unemployment, the AIDS epidemic has severely damaged the labour supply of prime-age Africans. Consequently, several African families are unable to rely on the prime-age adults in the family for financial support. Instead, these families survive on the income generated by the elderly in the household, such as the SOAP.

In light of the above, it seems likely that the living arrangements of Africans should be affected by the receipt of the SOAP. First, the dispersion of African families makes these families more susceptible to a change in living arrangements than most. Second, the existence of labour migration suggests that the living arrangements of African families are sensitive to a need for income. Thus, an inflow of income as large as SOAP will likely affect them. Lastly, the high unemployment and AIDS rates in South Africa have increased the reliance of black South Africans on the SOAP. Therefore, one would expect to see a shift in the living arrangements of black South African households on the arrival of the pension in order that the SOAP income might be expended most effectively.

University of Cape Town

Developing a greater understanding of the dynamics of the living arrangements of pensioners and their families is important for a number of reasons. First, household composition[2] is an important component of well-being. One's household members may impact upon access to resources, living conditions and the ability to influence decisions in the household. Thus, an understanding of how these living arrangements change with pension receipt is central to assessing the impact of the pension on the well-being of the household members. Furthermore, since economic welfare (such as per capita income or expenditure) is generally measured at the household level, identifying any changes that occur in household composition as a result of the SOAP is crucial to understanding the true welfare effects of the pension.

An analysis of the living arrangements of the elderly in South Africa is also relevant on an international scale. Owing to lower fertility and mortality, both developing and developed countries are experiencing substantial growth in the size of their older populations. A better understanding of the living arrangements of the elderly and the factors that affect them is therefore crucial if South Africa, and the rest of the world, is to prepare adequately for these significant demographic changes.

Lastly, a deeper understanding of the dynamics of the living arrangements of pension households may have important implications for the conclusions of past and future studies on the pension. A lack of panel data in South Africa has restricted most previous researchers to using cross-sectional data when examining the effects of the pension. Since cross-sectional data contain information on a group of individuals at one point in time only, these researchers have had to estimate the effect of the pension by measuring the difference between two *different* groups of households, one who receives the pension to one who does not. The use of cross-sectional data has therefore made it impossible for researchers to successfully isolate pension effects without relying on a number of simplifying assumptions regarding the differences between these two groups. One such assumption has been that household composition remains static on pension receipt.

[2] Household composition and living arrangements will be used interchangeably in this study.

University of Cape Town

Recent studies (Posel et al (2004), Edmonds et al (2005) and Hamoudi and Thomas (2005)) have, however, suggested that household composition may respond to pension receipt. If this is true, and if a change in household composition on pension receipt is in turn related to a change in the outcome variable, a pension effect estimated using cross-sectional data will incorrectly measure the changes in the outcome related to pension receipt as well as the changes in the outcome related to the change in household composition. Thus, if household composition does respond to pension receipt, the results from past studies on the pension may have been incorrectly measured and misunderstood. While a shortage of panel studies in South Africa continues to force researchers to use cross-sectional data in their study of the SOAP, a greater knowledge of the expected changes in the composition of households on pension receipt will enable researchers to interpret results more accurately. This study aims to contribute to this knowledge, and in so doing, assist in the interpretation of future studies on the effect of SOAP.

This study estimates the effect of the pension on the living arrangements of elderly *African women*. The analysis focuses exclusively on African women because, as is argued in detail further on, the pension impact on living arrangements is likely to be greatest for this group. It estimates the pension effect on the household composition of African women by determining whether the composition and/or size of an African household changes significantly when a woman in the household becomes age-eligible[3], and it interprets these changes in terms of their effect on the age and gender of the household members. The study uses age-eligibility instead of actual pension receipt because age-eligibility is exogenous while pension receipt is not. Furthermore, given that most age-eligible Africans receive this grant[4], age-eligibility is considered to be a good instrument for pension receipt.

The main challenge of this study comes in isolating those changes in the living arrangements of elderly African women that are attributable to the SOAP alone. The most accurate estimate of the pension effect on the living arrangements of African

[3] Women become age-eligible for the SOAP when they turn 60.
[4] Just under 80% of all age-eligible African females received the SOAP in 2001 (Labour Force Survey in September 2001).

University of Cape Town

women would in theory require one to be able to compare two groups of African women that are identical in all aspects except for their exposure to the pension. In practice, this is impossible to achieve since no one person can be observed in two states (i.e. pre-pension and post pension) at any one time. Since national panel datasets are not available, this study also relies on cross-sectional data and therefore estimates the pension effect on living arrangements by comparing two different groups of individuals, one who receives the pension to one who does not. Therefore, the difficulty facing this study comes in developing a methodology that ensures that all other differences between the two groups that might otherwise affect their living arrangements are adequately controlled for.

The two methodologies employed in this study attempt to compensate for the aforementioned weakness of cross-sectional data. The first method used is a cohort approach. In this approach, the pension effect is estimated by measuring the changes in the living arrangements of *cohorts* of African women as they become age-eligible for the SOAP. In theory, the cohorts should represent the same underlying population groups over time and hence, the cohort data should form a pseudo panel. Therefore, differences between the living arrangements of the cohorts slightly prior to and slightly post age 60 should be largely attributable to the receipt of SOAP.

The second approach uses a number of *regression discontinuity models* to estimate the shift in household composition at age-eligibility. Using these models, the pension effect is measured as the jump in the living arrangements of African women as the women turn 60. Since cross-sectional data are employed for this analysis, the discontinuity at age 60 will be measured by comparing the living arrangements of African who are not yet age-eligible to those that are. If these women are randomly selected into the sample, the primary difference between these two groups will be their average age. Thus, this approach attempts to correct for this difference by only including women who are close to age 60 into the sample and by including a function that measures the relationship between age and household composition into the models.

University of Cape Town

The paper is structured as follows: the following section provides a brief overview of the recent literature on the State Old Age Pension, its effect on living arrangements, and the importance of living arrangements. A descriptive analysis of the living arrangements of elderly African women follows in section 3. The methodology behind each approach is then discussed in section 4. The results are presented in Section 5, and finally, conclusions follow in Section 6.

University of Cape Town

2. Literature Review

2.1 The State Old Age Pension and its Effect on Living Arrangements

The South African State Old Age Pension has had an interesting history. Originally started as a pension exclusively for white South Africans, the pension has evolved into one that is paid equally across all races. The pension was fully deracialized in 1993, one year prior to the first democratic elections. Today, the majority of age-eligible Africans receive this grant[5]. With only a handful of White South Africans becoming eligible for this grant, the social pension now acts as an income transfer from the white population, South Africa's wealthiest population group, to Africans, South Africa's poorest one.

In 2005, the SOAP stood at a maximum level of R780 per month. The size of this grant is significant considering that minimum monthly wages for domestic workers were only R754 in rural areas in this year. All women of ages 60 and above and men of ages 65 and above are eligible to receive the pension if they are citizens of South Africa and pass a certain means test. For an individual, the means test takes into account one's personal income and an income value assigned to assets. For married couples, means are calculated by pooling resources and dividing by two. Relative to average African income, the means test is generous and therefore in practice, most Africans are eligible for this grant.

The pension is a vital source of income to many African families. Møller and Devey (2003) report many older African households being poor. They show that the arrival of the pension greatly reduces the probability of these households falling into the lowest expenditure quintile. Møller and Ferreira (2003) also document the importance of pension income in older households. They show that the pension is often the sole source of income in these households. Their results highlight the relative magnitude of this income in contrast to the extreme levels of poverty faced by most Africans in older households. This problem has undoubtedly been exacerbated by the rising number of working age

[5] Approximately 80% of all age-eligible African received the SOAP in 2001 (South African Population and Housing Census in 2001).

University of Cape Town

adults affected by the AIDS epidemic and/or unemployment who seek support from their elderly parents or family.

In light of the above, it is not surprising that the pension has been shown to have a significant impact on both the behaviour and well-being of members of pension households. Past research has provided evidence that the receipt of the pension is associated with an improvement in the health of granddaughters (Duflo (2003)), a drop in the work activity and rise in the school attendance of children (Edmonds (2005)), a decline in the work activity of prime-individuals (Bertrand, Mullainathan and Miller (2003)), a rise in the number of prime-age female migrants (Posel, Fairburn and Lund (2004)), and finally, a decline in the number of remittances from family members outside the household (Jensen (2004)). However, the results from past studies are predicated upon the assumption that the changes in household composition in response to the pension are negligible. If this assumption does not hold, then the estimated pension effects from past studies will have incorrectly captured both the impact of the receipt of the pension as well as the impact of a change in living arrangements, and consequently the results may have been misinterpreted.

The Bertrand et al (2003) study and a number of papers written in response to this paper can be used to illustrate the above point. This study examines the effect of the receipt of the pension on the number of working hours of prime-age individuals living with the pensioner. When comparing the number of working hours of prime-age individuals living with a pensioner to those not living with a pensioner, Bertrand et al (2003) find that, on average, prime-age individuals living with pensioners work significantly fewer hours than those not living with pensioners. Bertrand et al (2003) conclude that the pension has a negative effect on an adult's willingness to work. However, this conclusion implicitly relies on the assumption that household composition does not change on pension receipt.

If, however, household composition does respond to pension receipt, a number of competing explanations exist for the findings of Bertrand et al (2003). One such

University of Cape Town

explanation is suggested by the results from the paper by Posel et al (2004). Posel et al (2004) reconsider the impact of the pension on the labour supply of working-age adults using the same dataset as Bertrand et al (2003) but examine its impact on the working activity of *non-resident* members of the household as opposed to resident members. She finds that the pension is positively associated with the migrant activity of working-age women and concludes that the pension may encourage working-age women to migrate to work or to look for work. Thus, the results from the study by Posel et al (2004) provide an alternative explanation for why the labour supply of resident household members drops on pension receipt. That is, women who are willing to work leave the household on pension receipt to go work or look for work, and leave only those women who are unwilling to work in the household.

Another possible interpretation of the results presented in Bertrand et al (2003) is suggested by the results from Klasen and Woolard (2005). Klasen and Woolard (2005) consider the relationship between unemployment and household formation. They argue that many unemployed individuals survive unemployment by attaching themselves to pension households. If this is true, then it poses an obvious explanation for the results observed by Bertrand et al (2003).

Lastly, Hamoudi and Thomas (2005) also propose an explanation for the results from Bertrand et al (2003). They show that adults with lower levels of human capital, as measured by height and education, are more likely to reside in pension households. As height and education remains fixed for adults, Hamoudi and Thomas (2005) argue that this cannot be an effect of the pension but rather an indication of the type of individuals who choose to co-reside with pensioners. Thus, Hamoudi and Thomas (2005) conclude that household composition may respond to pension receipt and that by failing to control for the human capital of individuals the results of past studies on the pension (such as those of Bertrand et al (2003)) may be misinterpreted.

Thus, without further knowledge of the dynamics of household composition on pension receipt it becomes very difficult for researchers to rule out competing explanations for

University of Cape Town

differences between pension and non-pension households. This study aims to assist in the interpretation of the SOAP research by examining the changes that occur to the living arrangements of African women on pension receipt. It measures these changes in terms of their effect on the age and gender of household members.

To date, only one other study has considered this exact issue. This study, performed by Edmonds et al (2005), examined the changes that occurred to the living arrangements of pensioners on pension receipt using a regression discontinuity approach and census data from 1996. Edmonds et al (2005) showed that households with female pensioners tend to have more children below the age of 5, more women between the ages of 18 and 23 and fewer women between the ages of 30 and 39 residing in them. With regard to the rise in the number of children below the age of 5 and the number of women between the ages of 18 and 23, Edmonds et al (2005) suggest that the pension may encourage mothers to move back to the pension household with their children so that the children can be looked after by their grandparents while they work. They argue that the number of women aged between 30 and 39 will drop because these women will have the financial freedom to leave the home. Furthermore, the arrival of younger women and children associated with the pension will possibly free the older women of their household duties.

The following study has been based on that of Edmonds et al (2005) in that it identifies the pension effect by measuring the discontinuous changes in household composition at age 60. However, it supersedes the study by Edmonds et al (2005) in that it draws on more recent data and employs a wider range of Regression Discontinuity techniques. Furthermore, it uses a second strategy, a cohort analysis, in which a pseudo panel is created and used to measure the pension effect.

2.2 The Importance of Living Arrangements

Apart from its relevance to past research, a study on living arrangements is important for its insight into the well-being of pensioners. Several studies have found that an older person's living circumstances can have significant effects on his/her health. From their

University of Cape Town

study of Americans aged 51-61, Hughes and Waite (2002) theorize that the health of an older person in a household may worsen if there is an imbalance between the demands placed on the person and the resources made available to them. This balance will be a function of ones living arrangements. Barclay et al (1997) consider the effect of a change in living arrangements on the health of Americans aged 70 and older. They found evidence that living with or changing to living with someone other than a spouse may increase mortality risk of an older person. They also found that living alone will not impact upon the health of an older person.

Not only will the elderly be affected by a change in living arrangements. Other members of the household, especially children, have also been shown to be sensitive to a change in household composition. In their paper on orphans in Africa, Case, Paxson and Ableidinger (2004) showed that the schooling outcomes of orphans (where orphans are defined as a child who has lost one or both parents) depend largely on the how closely-related they are to the household head. Orphans living with a parental head are shown to fare better in school than those living with a non-parental head. Those living in households headed by non-relatives, fare worst of all. The findings of Case, Hosegood and Lund (2003) suggest that a child's living arrangements may also play a role in determining whether or not a child gains access to social assistance. In their study on the South African Child Grant, Case, Hosegood and Lund (2003) show that children living with less educated parents, or parents who are unlikely to be employed, are, on average, more likely to receive the Child Grant. Children whose fathers have died or who live in households with greater numbers of children are also more likely to receive this grant. A child's living arrangements in South Africa may also affect their health. Case, Lin, and McLanahan (2000) found evidence that in households where a child's birth mother is absent, less money is spent on milk, fruit and vegetables, and more on alcohol and tobacco. Finally, living arrangements may also have implications for the fertility of mothers. Moultrie and Timaeus (2001) investigated the relationship between fertility and living arrangements in South Africa. They found that fertility will decline on average when a woman lives with relatives from her own generation, or if a mother is unmarried or separated.

In summary, the results from past studies on the relationship between living arrangements and well-being re-emphasize the importance of understanding the impact of the pension on household composition. The findings of past studies suggest that, through its effect on living arrangements, the SOAP may have an effect on health, survival, education and even fertility of family or household members. Although this paper does not address any of these behavioural outcomes, it seeks to deepen the understanding of the dynamics of household composition at age-eligibility in order that researchers may better understand the welfare effects of the pension.

University of Cape Town

3. Descriptive Statistics

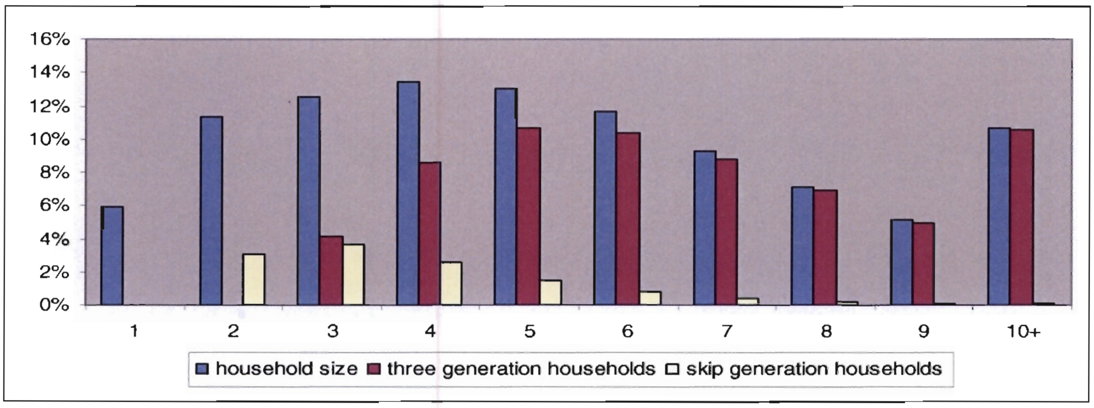
Before estimating the pension effect on household composition, a brief overview of the living arrangements of African women over the age of 60 is provided. A basic understanding of the living arrangements of elderly African women is useful when interpreting the effects of the pension on their living arrangements. The chapter also includes a preliminary investigation into the effect of the pension on the living arrangements of African women. These results are used to provide motivation for the econometric analysis. Finally, the chapter concludes with an analysis of the pension effect on the income and employment of African women. The results from this analysis are particularly valuable in resolving methodological issues described in the next chapter. All results are based on data from the 10% sample of the South African Population and Housing Census in 2001.

3.1 The Living Arrangements of Elderly African Women

The distribution of the size of households of female African pensioners is illustrated in Figure 1 below. The figure also illustrates the proportion of households that are three or skip generation households in each household size group. Three generation households are defined as those with at least one child aged less than 18, at least one prime-age individual aged between 18 and 49, and at least one elder aged 50 and over. Skip generation homes are defined as those with at least one child and one elder, but no prime-age individuals.

University of Cape Town

Figure 1: The Distribution of Household Size, Skip Generation Households and Three Generation Households of African Women aged 60 or older



Source: 2001 South African Population and Housing Census

African women over the age of 60 live in large households. In 2001, only 6% lived on their own and approximately 11% lived only with their spouses. The largest category of African female pensioners lived in households with 4 members. However, with over 10% of these women living in households of 10 members or more, average household size of a female pensioner in 2001 was as much as 5.49. Most African women over the age of 60 living in households of 4 or more lived with three different generations of individuals. In smaller households, a notable proportion of women over 60 lived in skip generation households. Approximately 3% of all African women over the age of 60 lived with children only.

Table 1 below illustrates the average composition of the households of African women over the age of 60, where household composition is defined according to a variety of age and gender groups. The first column presents the percentage of all African women over the age of 60 who live with at least one member from the specific age-gender group. The second column shows the average number of individuals in an age-gender group in households of African women over 60 who live with at least one member from that specific age-gender group.

University of Cape Town

Table 1: The Average Household Composition of African Women over the age of 60

	% living with	average number in households with
children aged 0-5	43.58%	1.5334
children aged 6-17	69.84%	2.3255
men aged 18-23	23.17%	1.1993
women aged 18-23	25.01%	1.1946
men aged 24-29	16.46%	1.1369
women aged 24-29	20.46%	1.1368
men aged 30-39	20.01%	1.1490
women aged 30-39	24.30%	1.1441
men aged 40-49	12.47%	1.0718
women aged 40-49	13.69%	1.0716

Source: 2001 South African Population and Housing Census

From Table 1, it is clear that there is a strong presence of children in the households of African women over the age of 60. Approximately 44% of all African women over 60 live with at least one child under the age of 6, while just fewer than 70% of these women live with children between the ages of 6 and 18. Of those women living with children, a large number will live with more than one child in their households as the average number of children aged 0-5 and the average number of children aged 6-18 is well over one for these households. Approximately a quarter of all African women over 60 will live with young adults between the ages of 18 and 23. Of those women living with young adults, few will live with more than one young male adult and one young female adult in her household as the average number of young male adults and the average number of young female adults in these households is little over one. Two interesting attributes of the living arrangements of African women over 60 are brought to light by the remaining results in Table 1. First, African women over 60 are more likely to live with adults aged 30-39 than adults aged of 24-29 or 40-49. Second, a greater proportion of African women over 60 will live with adult women than adult men. The results show that 20% of all African women over 60 will live with women aged 24-29 however only 16% will live with men aged 24-29. 24% of all African women over 60 will live with women aged 30-39 while only 20% will live with men aged 30-39. Lastly, just under 14% of all African women over 60 will live with women aged 40-49 while just over 12% will live with men

University of Cape Town

aged 40-49. Finally, the averages in column 2 for these age-gender groups show that few African women living with adults will live with more than one adult from each age-gender group.

3.2 The Effect of the Pension on Living Arrangements of African Women

A preliminary investigation into the effects of the pension on living arrangements was conducted. In this analysis, the living arrangements of women between the ages of 50 and 59 and the living arrangements of women aged between 60 and 69 are compared. Women aged 50-59 who already live with a pensioner are excluded from the analysis so that the two groups (women aged 50-59 and women aged 60-69) are differentiated clearly between those women living in pension eligible households and those women who are not. The results from this analysis are presented in Table 2 below.

Table 2: Composition and Size of Households of African Women prior to and post Age-Eligibility

	women aged 50-59		women aged 60-69		difference
household size	5.1154	(0.0114)	5.5039	(0.0127)	0.3885*
no children aged 0-5	0.6277	(0.0035)	0.6993	(0.0039)	0.0716*
no children aged 6-17	1.4327	(0.0057)	1.6040	(0.0065)	0.1714*
no of women aged 18-23	0.3539	(0.0023)	0.3005	(0.0023)	-0.0535*
no of men aged 18-23	0.3211	(0.0022)	0.2763	(0.0022)	-0.0448*
no of women aged 24-29	0.2912	(0.0021)	0.2578	(0.0021)	-0.0333*
no of men aged 24-29	0.2489	(0.0020)	0.2058	(0.0019)	-0.0431*
no of women aged 30-39	0.1908	(0.0017)	0.2868	(0.0022)	0.0961*
no of men aged 30-39	0.1802	(0.0017)	0.2484	(0.0021)	0.0682*
no of women aged 40-49	0.0405	(0.0008)	0.1063	(0.0013)	0.0657*
no of men aged 40-49	0.0638	(0.0010)	0.1094	(0.0013)	0.0456*
% living alone	0.0772	(0.0010)	0.0583	(0.0009)	-0.0189*
% living with spouse only	0.1244	(0.0013)	0.1128	(0.0013)	-0.0115*
% three generation households	0.6222	(0.0019)	0.6506	(0.0019)	0.0284*
% skip generation households	0.1246	(0.0013)	0.1253	(0.0013)	0.0008*
number of observations	68045		63137		

Source: 2001 South African Population and Housing Census

* = significant at a 5% level

The column 1 and column 2 of Table 1 present the means and the standard deviations in parentheses for the various measures of household composition of African women aged 50-59 and 60-69 respectively. The difference between the means of each group is provided in column 3. Asterisks are used in this column to indicate when the difference in means is significant at a 5% level of significance.

The results suggest that the living arrangements of African women who are pension eligible are significantly different to those who are not yet pension eligible. African women between the ages of 60 and 69 will, on average, live in larger households than women aged between 50 and 59. They will, on average, live with more children and with more adults between the ages of 30 and 49 than their younger counterparts. However they will tend to live with fewer young adults, specifically adults between the ages of 18 and 29. The percentage of women living on their own or with a spouse only, drops significantly when a woman becomes eligible for the pension. Furthermore, more women will live in three generation and/or skip generation households once they have become eligible for the pension.

The above results suggest that the pension may play a role in determining the living arrangements of elderly African women. However, it is important to note that there may be a number of differences between the two groups apart from pension receipt (one obvious difference being age) and that these differences may wholly or partly account for the differences between the living arrangements of the two groups. Thus, the challenge facing this study is to isolate those changes in living arrangements that can be attributed to the pension itself. This issue is addressed in the next chapter.

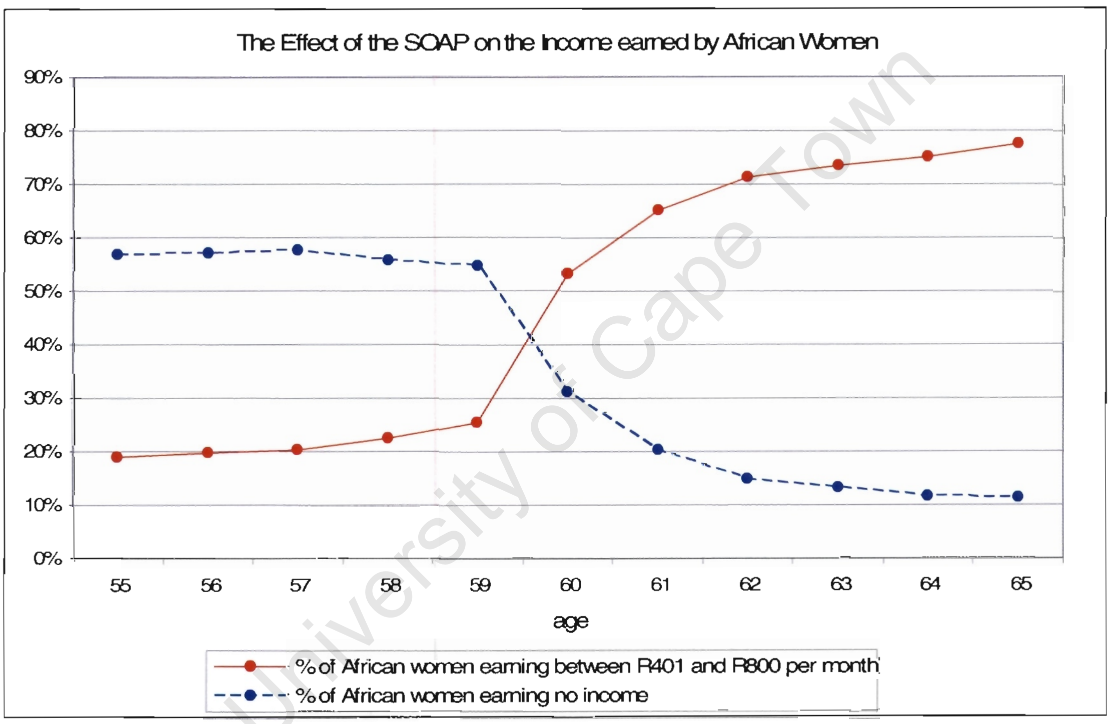
University of Cape Town

3.3 The Effect of the Pension on the Income and Employment of African Women

Before turning to the econometric analysis, the pension effect on the income and employment status of African women is briefly examined. The results from this section are important in that they illustrate the significant impact the pension has on the wealth and work activity of African women. Furthermore, the results provide guidance to the methodology behind the econometric analysis.

Figure 2 below depicts the percentage of African women who received no monthly income and the percentage of African women who received a monthly income between R401 and R800 for various age groups. Since the maximum pension amount in 2001 was R540 per month, women over the age of 60 who earn between R401 and R800 monthly income are likely to do so because they are receiving the pension. This income band is used to approximate pension take-up in this analysis as the 2001 Population Census only reports income in bands and does not differentiate between different sources of income.

Figure 2: The Effect of SOAP on the Income earned by African Women aged between 55 and 65



Source: 2001 South African Population and Housing Census

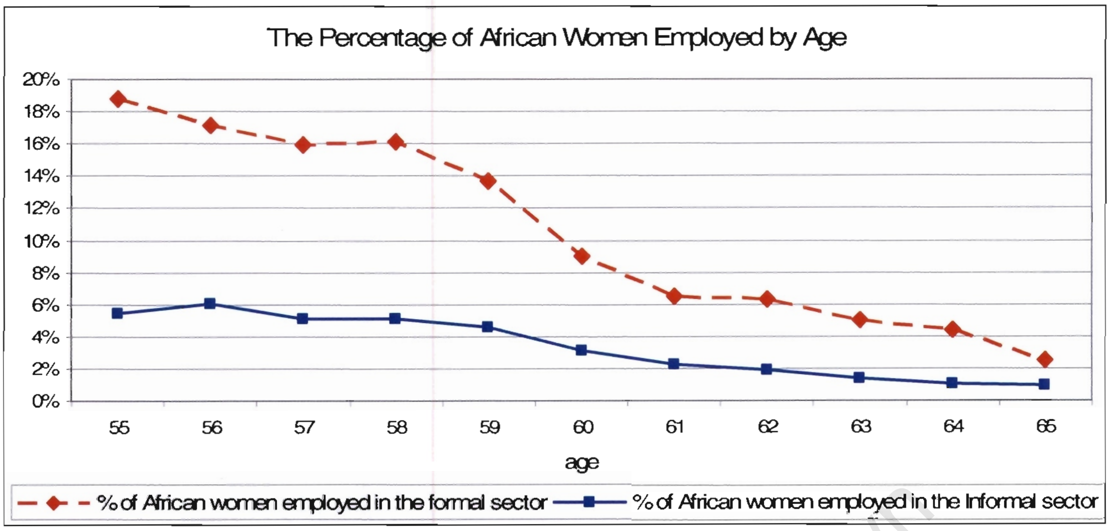
The percentage of African women receiving an income between R401 and R800 per month escalates when women turn 60. Between the ages of 59 and 60, the percentage of women with an income between R401 and R800 per month rises sharply from approximately 26% to approximately 55%. By age 61, almost 70% of all African women appear to receive the pension. The results suggest that most women take up the pension when they become eligible. This in turn implies that age-eligibility acts as an adequate proxy to pension receipt.

The percentage of African women receiving no income is alarmingly high. Between ages 55 and 59, the percentage of African women receiving no income averages around 57%. This figure declines steadily as the women age into pension-eligibility. The percentage of African women receiving no income drops sharply from 55% to just over 30% when these women age from 59 to 60. This percentage drops to 20% by age 61 and to just over 10% by age 65. The results therefore imply that the pension serves as a vital source of income to a large proportion of elderly black South African women.

The effect of the pension on the employment status of African women is then considered. The analysis examines how the percentage of employed African women changes around age-eligibility. The percentage of employed African women is divided into those working in the formal sector and those in the informal sector. A question from the 2001 Census which asked the respondent to describe the type of work she had been engaged in in the last 7 days was used to define whether the respondent worked in the formal or informal sector. The respondent could answer that she had been working in the formal sector, the informal sector, farming or had been absent from work for the past week. All respondents who answered that they worked in the formal sector or in farming were classified as working in the formal sector, and all women who responded that they worked in the informal sector were allocated to the informal sector. Women who were temporarily absent were excluded from this analysis as they comprised a very small percentage of the total number of African women. Figure 3 presents the results from the analysis.

University of Cape Town

Figure 3: The Effect of SOAP on the Employment of African Women aged 55 to 65

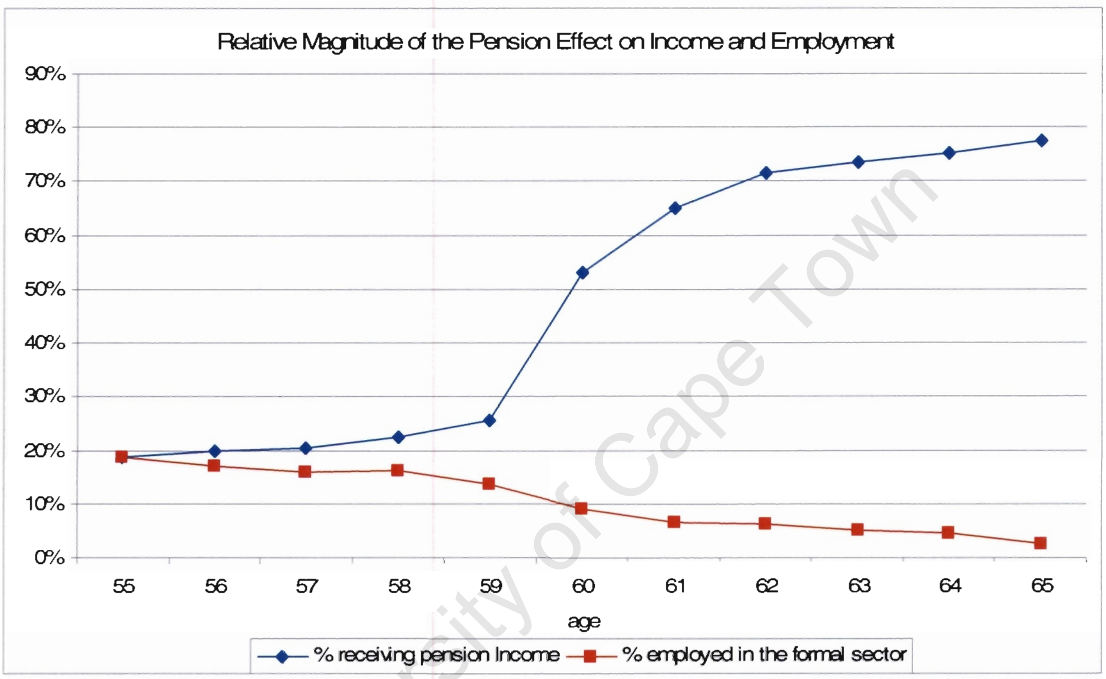


Source: 2001 South African Population and Housing Census

The total percentage of African women between the ages of 55 and 65 who were employed never rises above approximately 25%. Of those employed, a larger percentage is employed in the formal sector. The percentage of African women employed drops steadily as women age into pension eligibility regardless of which sector they are employed in. The percentage of African women employed in the formal sector drops from just under 14% to 9% between age 59 and 60. This amounts to a percentage change of approximately 34%. By age 65, the percentage of African women working in the formal sector has dropped to 2%. The percentage of African women employed in the informal sector falls from approximately 4.5% to 3% between ages 59 and 60. Although the absolute change is small, the percentage change is approximately 32%. Less than 1% of African women are employed in the informal sector by the time they reach 65.

The rapid decline in employment at and around age 60 suggests that the pension encourages African women to retire. However, since the retirement age for women working in the formal work sector often occurs at or near age 60, it is difficult to determine to what extent the drop in employment is attributable to the pension alone. If formal retirement encourages a large proportion of African women to retire at age 60, it is possible that the pension effects estimated in this study will be partly or wholly confounded by the effects of formal retirement. To gain some insight into the potential magnitude of retirement effects, Figure 4 is constructed by superimposing part of Figure 3 onto part of Figure 2. The figure graphically depicts the relative magnitude of the pension effect on income to the change in employment at and around age 60.

Figure 4: Comparison of the Pension Effect on Income versus the Change in Formal Employment at and around age 60



Source: 2001 South African Population and Housing Census

University of Cape Town

From Figure 4, it is clear that the change in the percentage of women working in the formal sector at age-eligibility is small relative to the change in the percentage of women receiving pension income at age-eligibility. Between the ages of 59 and 60, the percentage of African women receiving the pension rises from approximately 26% to 55%. By age 61, this percentage has risen to just less than 70%. The percentage of African women who retire from the formal sector is comparatively smaller. At age 59, less than 14% of all African women are employed in the formal sector. While this figure drops to 9% by age 60, this change is still small next to the 29% rise in the number of women receiving the pension between ages 59 and 60. Thus, the results from the investigation suggest that, while retirement effects may exist, these effects are likely to be outweighed by the effects of the pension. For the purposes of this study, retirement effects are therefore deemed to be negligible.

University of Cape Town

4. Methodology

This study aims to determine the effect of the pension on the living arrangements of African women. It uses two different approaches to achieve this aim, namely a cohort approach and a regression discontinuity approach. This chapter begins with an overview of the methodology used in this study. This overview considers methodological issues common to both approaches. Both approaches are then outlined and motivated in the remainder of the chapter.

4.1 Overview of Methodology

The following analysis focuses exclusively on African women. It does so because the possibility of identifying a pension effect is likely to be greatest for this group. Firstly, the African population group has the highest pension take-up rate. They are also the poorest population group and therefore the most likely to be affected by an inflow of income. Secondly, past research (Edmonds et al (2005), Duflo (2003) and Bertrand et al (2003)) has shown that the possibility of finding pension effects is greater for African women than African men. Edmonds et al (2005) show that the likelihood of pension effects being confounded by the effects of formal retirement (which coincides exactly with age-eligibility for the pension for both men and women), is smaller for African women than African women since fewer African women are affected by formal retirement than men. Duflo (2003) and Bertrand et al (2003) show that African women are also more likely to share their pension income than African men. Thus, one would expect to see greater changes in household composition when an African woman receives her pension than when an African man does.

This study aims to quantify the effect of SOAP on the living arrangements of African women. To achieve this, it estimates the change in the living arrangements of African women as they become *age-eligible* i.e. as they turn 60. The study focuses on age-eligibility instead of pension take-up because age is an exogenous variable while pension

University of Cape Town

receipt is not. However, since take-up is unlikely to be complete[6], any pension effect identified at age 60 will most likely be an underestimate of the true effect.

By considering the changes in household composition occurring between ages 59 and 60 only, the approach is also limited in the type of pension effects it can detect. Some households may adjust their living arrangements in anticipation of the pension whilst others may take some time to make these adjustments. Furthermore, some women may not take up the pension as soon as they are eligible. Again, pension effects measured in this study will be attenuated if adjustments to living arrangements in response to the pension are made prior to or post age 60. Therefore, as a sensitivity analysis, this study also considers the changes in the household composition of African women between ages 58 and 62. Using this approach, the estimated pension effect will not only measure the changes to household composition that occur at age 60, but also those changes that occur a few years prior to (58-60) or post (60-62) pension receipt.

The study uses cross-sectional data to determine the effect of the pension on the living arrangements of African women. Cross-sectional data contain information on a group of individuals at one point in time. Using cross-sectional data, a pension effect on living arrangements can be estimated by comparing two *different* groups of individuals, one who receives the pension to one who does not. Owing to the fact that the two groups comprise of different individuals, a number of differences will exist between the two groups apart from their pension exposure. This makes it difficult to measure to what extent any changes in living arrangements at age-eligibility can be attributed to the pension alone. Pension effects can only be accurately estimated using cross-sectional data if the other factors affecting living arrangements can be identified and adequately controlled for.

Panel data are likely to provide a more accurate estimate of the pension effect. Panel data comprise of information on a group of individuals at various points in time. Using panel

University of Cape Town

[6] Only 80% of all age-eligible African received the SOAP in 2001 (South African Population and Housing Census in 2001).

data, a pension effect can be estimated by calculating the average difference between the household composition of the individuals just before pension receipt and their household composition just after. This estimate is likely to be more accurate than one calculated using cross-sectional data because, unless the panels are very far apart, many of the characteristics of the individuals will remain the same between the two panels of data. (Of course, there will be some characteristics that do change, one obvious one being age, and such changes must be controlled for.) Unfortunately, panel data on a national level are not yet available in South Africa. A few panel datasets are available however these are localized in small areas of the country. National cross-sectional datasets are, however, readily available, and are therefore used in this study.

This study compensates for the weakness of cross-sectional data in two different ways. The first approach uses *cohort data*, constructed using a series of national cross-sectional surveys. In brief, the cohort data consist of information on *groups* of individuals at various points in time, where the groups (or cohorts) are defined by age. Aggregate information is collected on each of the cohorts for every year in the period of study. Provided that (i) the sample cohorts are large enough, (ii) the surveys from which the cohort data are constructed have the same sample frames, and (iii) the underlying population cohorts are not much affected by immigration, emigration or death over the period of study, the sample cohorts will be representative of the same underlying population cohorts over time. In this way, the cohort data form a type of pseudo panel and consequently, the changes in the characteristics of the sample cohorts over the period can be used as an estimate of the actual changes in the characteristics of the underlying population cohorts over this period. For the purposes of this study, an analysis of how the average household composition of a cohort changes as the cohort ages into pension eligibility should therefore provide an accurate estimate of the pension effect.

University of Cape Town

The second method employed by this study is a regression discontinuity design (RD design). Under RD designs, program effects (such as the SOAP program) are estimated as the discontinuous changes in the outcome of interest when a program takes affect. RD models therefore rely on the important assumption that all other changes to the outcome of interest can be modelled as a smooth function over the period of interest. RD models also attempt to control for the difference between treatment (pension households) and non-treatment groups (non-pension households) by only using data very close to the point at which the treatment is received. If individuals are selected randomly into the sample, then the individuals slightly to the left of the boundary are unlikely to differ considerably from those slightly to the right of it. Thus, any differences in the household composition of those receiving pension and those not can be mostly attributed to the impact of the pension.

The specific data and methodology used in the cohort and the regression discontinuity methods are described in the remainder of this chapter. The results from these analyses are presented in Chapter 5.

University of Cape Town

4.2 Cohort Analysis

4.2.1 Cohort Data versus Panel Data

Cohort data can be constructed using data from a series of successive cross-sectional surveys. Unlike panel data which follow individuals over time, cohort data follow *groups* (or cohorts) of people from one survey to another. Individuals are grouped according to some characteristic that each individual shares in common with the rest of the cohort, for example birth year. Although the individuals comprising each sample cohort will change from survey to survey, provided the sample cohorts are large enough, provided the surveys have the same sample frames, and provided the population is not much affected by immigration, emigration or death, the sample cohorts should represent the same underlying population cohorts from year to year. Thus, sample cohort means can be tracked over the observation period as a measure of how certain characteristics of the population cohorts changed over the period of observation. In this way, cohort data share many of the same properties of panel data. For this reason, cohort data are also commonly referred to as a pseudo-panel.

Although cohort data and panel data may be similar, a number of issues arise if one analyzes cohort data exactly as if it were repeated data on individuals (Deaton, 1997). In understanding this, it is helpful to first consider how panel datasets are commonly analyzed. Consider the individual model:

$$y_{it} = \alpha + \beta x_{it} + \theta_i + u_{it} \qquad t = 1, \ldots, T \qquad (1)$$

where i represents the number of individuals in each year t. α is the constant term, x_{it} is the set of explanatory factors, β is the set of parameters to be estimated and u_{it} are the error terms which are normally distributed with mean of 0 and constant variance. The θ_i's represent the unobserved individual effects that are assumed to be constant over time, for example an individual's inherent intelligence or talents. The individual fixed

University of Cape Town

effects are likely to be correlated with the explanatory factors, x_{it}'s, and thus a regression estimate of β would be biased[7]. However, one can counter this problem by taking first differences, and thereby removing any fixed effects. The resulting equation can then be used to solve for β by using Ordinary Least Squares.

Equation (1) can be generalized to the cohort level. Cohort means are constructed by averaging the observations over individuals in each cohort in each year, t. Thus, the following model results:

$$\bar{y}_{ct} = \alpha + \beta\bar{x}_{ct} + \bar{\theta}_{ct} + \bar{u}_{ct} \qquad c = 1, \ldots, C \qquad t = 1, \ldots, T \qquad (2)$$

where c is the cohort that an individual i belongs to. Each individual belongs to exactly one cohort and remains in that cohort for the duration of the observation period. Since (1) includes individual fixed effects, (2) will include cohort fixed effects. However, the cohort "fixed" effects will not be fixed – in fact, they will vary with t – because they comprise the averages of fixed effects of *different* individuals in each year (Baltagi, 2001). Thus, these effects cannot be removed by differencing as would be the case with panel data.

University of Cape Town

The above problem can be solved if it is reasonable to assume that $\theta_{ct} = \theta_c$. Under this assumption:

$$\bar{y}_{ct} = \alpha + \beta\bar{x}_{ct} + \bar{\theta}_{c} + \bar{u}_{ct} \qquad c = 1, \ldots, C \qquad t = 1, \ldots, T \qquad (3)$$

and β can be determined using the within estimator $\hat{\beta}_w$. The above assumption may be plausible if the number of observations in each cohort is large (Baltagi, 2001).

[7] If the individual effects are not correlated with the explanatory factors, then a random effects approach is appropriate and $\theta_i + u_{it}$ form the composite error.

However, Deaton (1997) draws attention to another measurement error problem. The independent and dependent variables in (2) are sample-based means and will therefore estimate the unobserved population means with measurement error. This measurement error will cause the estimates from (2) to be attenuated. Deaton (1997) shows that one can correct for this error by applying errors-in-variable techniques. The variance due to measurement error is equivalent to the standard error of the sample means and therefore can be easily calculated using the sample data. The estimated standard errors can then be removed from the standard fixed effects estimator to yield a consistent estimator.

In practice, many researchers working with cohort data ignore the problem of measurement error and merely compute the Within Cohort Estimator (Baltagi, 2001). Verbeek and Nijman (1993) show that if the average cohort size, n_c, is large enough, in particular, if $n_c = N / C$ tends to infinity (where N is the number of individuals in a sample), then the above-mentioned measurement errors will tend to zero and the Within Cohort Estimator will be asymptotically identical to Deaton's estimator (cited by Baltagi 2001: 191). Thus, in terms of measurement error, it would seem advisable to make n_c as large as possible. Unfortunately, increasing the size of n_c can have a detrimental effect on the variance of the model estimates. By increasing the size of n_c, one also reduces the number of cohorts. Fewer cohorts will mean that the pseudo panel will be smaller and therefore the variance of the estimates will be greater. Thus, when constructing cohorts, there is a direct trade-off between measurement error and variance.

University of Cape Town

Therefore, both the size and the number of cohorts will have a considerable effect on the reliability of results based on cohort data. Verbeek and Nijman (1992) also emphasize the need to select individuals into cohorts carefully so as to improve the precision of estimates (cited by Baltagi 2001: 191). Verbeek and Nijman (1992) state that measurement error can be minimized if the individuals in each cohort are as homogeneous as possible (cited by Baltagi 2001: 191). Furthermore, Verbeek and Nijman (1992) argue that one can maximize the variation in the pseudo-panel and therefore enhance the precision of estimates, by ensuring that the different cohorts are as heterogeneous as possible (cited by Baltagi 2001: 191).

4.2.2 Decomposing Cohort Data into Age, Year and Cohort Effects

It is commonly recognized that certain socio-economic variables follow particular life-cycle patterns. Income is a good example of one such a variable. Income will, on average, be small in one's early working years, increase steadily until middle-age after which it will most likely flatten out and probably decrease (Deaton (1997)). Economists are frequently interested in determining such age profiles however, in the absence of panel data, this exercise is very difficult. This is because age effects measured using cross-sectional data are automatically confounded by cohort effects. Cohort effects measure the extent to which an outcome is a function of year of birth (Deaton (1997)). These effects are also sometimes termed generation effects, which is technically incorrect since one does not generally regard two people born one year apart as coming from different generations. However, the concept is similar. Since cross-sectional datasets only include information on individuals at one point in time, a one-year change in age translates directly into a one-year change in cohort.

In the absence of panel data, pseudo-panels have been used to disentangle the age from the cohort effects. However, since pseudo panels look at observations of individuals over a number of years, when using pseudo panels to isolate age effects from cohort effects it is imperative that one also take into account year effects, where year effects are defined as changes in observations that occur owing to a change in year (Deaton (1997)). Thus, in order to determine the age profile of a variable using cohort data, it is necessary to decompose the data into their constituent age, cohort and year effects. One can think of this decomposition in the following form:

$$y = \beta + A\alpha + C\gamma + Y\psi + u \tag{4}$$

University of Cape Town

where y is a stacked vector of observations arranged in cohort-year pairs and u is vector of error terms that are associated with each y_{ct}. Variables A, C, and Y correspond to the matrix of age, cohort, and year dummies respectively, and α, γ and ψ represent the age, cohort and year effects respectively. Equation (4) illustrates that each observation, y_{ct}, can be separated into its constituent age effect, α_{ct}, cohort effect, γ_c, and year effect, ψ_t. Age effects represent the life-cycle pattern of the variable y for each cohort. Cohort effects measure the extent to which y is a function of time of birth. Year effects represent the general macro-economic changes that occur between each year in the data and that affect the value of y of all cohorts equally (Deaton (1997)).

The estimates of α, γ and ψ cannot be calculated by simply plugging into Equation (4) the values for the variables. As usual with regression using dummy variables, one must first drop one column from each of the matrices because the sum of the columns for the full matrices is a column of ones, which is already included as the constant term. However, even once these columns have been dropped, the parameters cannot be estimated since the age, cohort and year effects are linearly dependent. That is, once one knows the cohort to which an individual belongs and the year in which the same individual was interviewed, one can work out how old he or she is (Deaton (1997)) i.e.

$$\text{age} = \text{cohort} + \text{year} \tag{5}$$

Thus, the matrices of the dummies satisfy

$$Ax_a = Cx_c + Yx_y \tag{6}$$

where x_a, x_c, x_y are vectors of arithmetic sequences of length given by the number of columns of the matrix that pre-multiplies them (Deaton, 1997).

One way to estimate (4) would involve dropping one or more columns from any one the matrices in (6). In dropping one of the columns, one is forced assume that either two age, cohort or year effects are equal. For example, if one drops the first column of matrix A, one assumes that the effect of an age of 1 on household composition is the same as the

University of Cape Town

effect of an age of 2. The necessity to assume two effects equal should not generally be problematic since such an assumption is likely to be only a small distortion of the truth. However, the results obtained from this approach can vary considerably depending on which effects are assumed equal. Consequently, any researcher involved in cohort analysis should have a basic idea of the expected shape of one or more of the distribution of effects prior to analysis. Then the effects assumed equal should be those that provide the most plausible results.

Alternatively, one can tackle the identification problem by assuming year effects are negligible and therefore removing them from equation (4). In this case, one assumes that macro-economic shocks either do not exist over the observation period or do not affect the variable of interest, y, over this period. Where this assumption seems implausible, one could follow a normalization strategy outlined by Deaton (1997). This strategy assumes that year effects have a mean of zero and have no trend. The method therefore implicitly assumes that all trends can be attributed to either age or cohort effects. By imposing this restriction on the year effects, the estimates of α, γ and ψ can be identified. Of course, this method is undesirable if one suspects that the year effects follow a trend.

4.2.3 Advantages and Disadvantages of Cohort Data

Cohort data can in fact have a number of advantages over panel data. While panel studies are likely to suffer from attrition, owing to mortality, refusal and/or mobility, cohort data will not encounter attrition problems as the data are constructed from fresh samples in each year (Deaton, 1997). Furthermore, the way in which cohort datasets are used will often be less prone to measurement error than is the case with panel datasets, the reason being that cohort data follow averages and averaging will almost always reduce the measurement error (ibid). Cohort data can also be preferable to panel data when a long series of cross-sectional surveys are available as panel data are seldom collected over a long period.

University of Cape Town

There are, however, a number of disadvantages to working with cohort data. First, the aggregation of individuals may result in a loss of information. Second, the usefulness of cohort data relies strongly on the assumption that the cohort population is constant. This assumption is crucial in ensuring that successive surveys generate random samples from the same underlying population. As mentioned previously, the levels of immigration, emigration or death in a population may affect the validity of this assumption. Lastly, since each cohort has only one observation per year, the sample sizes of pseudo-panels are often small.

4.2.4 The Use of Cohort Data in this Study

This study is interested in estimating whether the household composition of African women changes when they receive the pension. Since the pension eligibility of African women is mostly determined by age, the study is mainly concerned with whether the household composition of these women changes as they become age-eligible. The study uses cohort data to achieve this aim. First, the cohort data are used in estimating the age profiles of the household composition of African women between the ages of 45 and 75. This is achieved by decomposing data on the average household composition of various cohorts of African women into its constituent age, cohort and year effects. The estimated age profiles are used to provide some insight into how the household composition of African women changes when these women receive the pension. The cohort data are then used in a fixed effects regression to estimate whether the average household composition of the cohorts is related to age-eligibility. If the sample cohort means are representative of the true population means, then the fixed effects regression will provide an accurate estimate of the relationship between age-eligibility and the average household composition of the cohorts. From this, one can draw conclusions on the relationship between pension receipt and the household composition of African women. Finally, it is worth re-emphasizing that the primary reason behind using cohort data in this study is that it allows one to partly overcome the problem of heterogeneity among observation units associated with studies based on cross-sectional data.

University of Cape Town

Construction of Data

The cohort data for this study comprise data on the average household composition of age cohorts of African women for each year between 1994 and 2003, with the exception of 1998. The changes that occur to average household composition of each cohort over the period 1994-2003 can then be used as an indication for how the household structure of African women of a certain age changed, on average, over this period. The cohort data was created using data from Statistics South Africa's national cross-sectional datasets. The datasets used to create the cohort data included two South African national surveys, namely the October Household Survey and the Labour Force Survey, and the 10% samples of the South African Population and Housing Censuses in years 1996 and 2001.

The October Household Survey (OHS) was an annual survey that ran from 1994 to 1999 covering a large range of development indicators. The surveys were expected to be conducted amongst 30 000 households however, owing to financial constraints, fewer households were interviewed in some years. This was the case in 1996 and 1998 and, as a result, the data from these years have not been used in this study.

The Labour Force Survey (LFS) is a biannual rotating panel household survey currently running in South Africa. It was started in 2000 and was specifically designed to collect information on the employment and unemployment status in South Africa. For the purposes of this study, the data from the surveys in the second half of years 2000, 2002 and 2003 are used.

Cohort data for the years 1996 and 2001 were collected from the 10% samples of the South African Population and Housing Censuses in 1996 and 2001. These censuses were conducted on the night of October 9^{th} to 10th in 1996 and 2001 respectively. Individuals living in households as well those residing in hostels and institutions such as prisons or hospitals were interviewed. For the purposes of this study, however, only those individuals living in households are considered. The definition of a household in the censuses is consistent with that in the national surveys. That is, a household is defined as a group of people who resided together under the same roof for 4 or more nights a week.

University of Cape Town

Data from each of the above datasets were restricted to African women and divided into age cohorts. The cohorts were defined by the age of the cohort in 1994. For example, a group of females aged 62 according to the 2001 census would belong to the 55th female cohort. Over the period from 1994 to 2003, the cohorts aged 9 years i.e. cohort 51 aged from 51 to 60, cohort 52 from 52 to 61, and so on. Various measures of household composition were then calculated for each of the cohorts. Specifically, the mean household size, the mean number of children aged 0 to 5 and 6 to 17, and the mean number of males and females aged 18 to 23, 24 to 29, 30 to 39, and 40 to 49 for each of these cohorts were calculated. Person weights supplied in each of the datasets were used to weight the data so that the cohort means were nationally representative. The data from each survey were then combined. A new variable "year" was created which equalled 0 in 1994, 1 in 1995 and so on.

Table 3: The number of Individuals constituting each Cohort in each Year of 1994-2003

	cohort (i)						
year	**45**	**50**	**55**	**60**	**65**	**70**	**75**
1994	384	338	233	351	218	159	84
1995	370	277	309	229	200	148	62
1996	9870	7352	6315	6378	5659	3456	2495
1997	423	300	300	305	258	165	129
1999	321	214	282	225	169	96	68
2000	304	224	212	156	149	94	54
2001	10047	6677	7130	5731	4937	2659	1948
2002	234	225	214	163	133	100	40
2003	257	224	189	170	117	64	42
n_{ci}	2468	1759	1687	1523	1316	771	547

Source: OHS for years 1994, 1995, 1997 and 1999, LFS for years 2000, 2002 and 2003, and Population Censuses for 1996 and 2001

The sizes and average sizes of various cohorts are represented in Table 3 above. Older cohorts tend to be smaller than younger ones. Furthermore, with the exception of the census years, the size of a particular cohort tends to decrease over the observation period. These results are not surprising as one would expect the number of individuals that constitute the true population age cohorts to decline owing to higher mortality rates.

University of Cape Town

However, they do highlight a potential weakness in the cohort data. That is, since the older cohorts are notably small, it is possible that these cohorts do not accurately represent the underlying population cohorts.

Determining Age Effects

The cohort data are used to estimate the age profile of average household composition of African women, which in turn allows an investigation into how average household composition changes over the period, particularly as the cohorts turn 60. To achieve this, the data had to be decomposed into its constituent age, cohort and year effects. As outlined in section 4.2.2, this decomposition requires one to impose a restriction on at least one of the effects in Equation (4). Three possible types of restrictions are explained in section 4.2.2. However all three proved to be problematic in the context of this study.

The first attempt at tackling the identification issue involved assuming two age effects equal. It seemed plausible to assume two age effects equal as long as the ages are close. Unfortunately, the results were extremely sensitive to which effects were assumed equal and therefore the results were deemed unreliable. The procedure was repeated using cohort effects, but the results were no better. Thus, the method of assuming two effects equal was considered unsuitable to this study.

The normalization technique proposed by Deaton (1997) was considered next. However, this approach was also deemed inappropriate when a preliminary analysis of the data revealed that a number of the measures of average household composition showed negative year trends. Table 4 below illustrates this.

University of Cape Town

Table 4: Changes in Average Household Composition of African Women 1994-2003

	1994	1995	1996	1997	1999	2000	2001	2002	2003	total change
household size	5.9953	6.0689	5.9086	6.2467	5.8901	5.6463	5.6074	5.6974	5.3445	-0.6508
no children aged 0-5	0.7808	0.7346	0.7961	0.8521	0.7769	0.6869	0.7014	0.6880	0.6270	-0.1537
no children aged 6-17	1.7526	1.7177	1.7749	1.8629	1.8134	1.7027	1.6365	1.7655	1.5799	-0.1727
no of men aged 18-23	0.3473	0.3595	0.3212	0.3683	0.3417	0.3453	0.3087	0.3344	0.3179	-0.0294
no of women aged 18-23	0.3805	0.3970	0.3738	0.4007	0.3670	0.3461	0.3267	0.3716	0.3519	-0.0286
no of men aged 24-29	0.2376	0.2498	0.2256	0.2434	0.2226	0.2139	0.2135	0.2088	0.1970	-0.0406
no of women aged 24-29	0.2944	0.3164	0.2637	0.2993	0.2761	0.2579	0.2637	0.2512	0.2475	-0.0469
no of men aged 30-39	0.2555	0.2893	0.2374	0.2703	0.2368	0.2578	0.2352	0.2363	0.2173	-0.0382
no of women aged 30-39	0.3629	0.3886	0.3249	0.3542	0.3357	0.3108	0.3174	0.3003	0.3186	-0.0443
no of men aged 40-49	0.1954	0.2010	0.1724	0.1784	0.1826	0.1680	0.1705	0.1756	0.1650	-0.0303
no of women aged 40-49	0.2589	0.2651	0.2369	0.2642	0.2333	0.2476	0.2480	0.2364	0.2247	-0.0343

Source: OHS for years 1994, 1995, 1997 and 1999, LFS for years 2000, 2002 and 2003, and Population Censuses for 1996 and 2001

Table 4 presents the average size of various measures of household composition of African women in each year of the period 1994-2003 apart from 1998. The table also provides the total change that occurred to the average measures of household composition between 1994 and 2003. Apart from the 1994-1995 and 1996-1997 changes, the average number of individuals in each measure of household composition tended to drop for each year that passed. Furthermore, every measure of average household composition dropped between 1994 and 2003.

In light of the above evidence, the application of Deaton's normalization technique was considered inappropriate. Furthermore, since year trends existed, excluding year effects from Equation (4) as a way to identify the age and cohort effects was unsuitable. Owing to these problems, it was decided that one of the other effects apart from the year effects should be excluded from Equation (4). However, this left only the cohort effects and the age effects themselves. Since the age effects are the effects of interest, the cohort effects were excluded from Equation (4), allowing the age and year effects to be identified. Thus average household composition was regressed against all age and year dummies. The particular regression run was:

$$y = \beta + A\alpha + Y\psi + u \tag{7}$$

University of Cape Town

where y represents the average household composition of a certain cohort in a certain year, A and Y correspond to the matrix of age and year dummies respectively, and α, γ and ψ represent the age and year effects respectively.

There are three notable weaknesses in the decomposition model (7). First, by excluding the cohort effects from Equation (4), the study effectively assumes that cohort effects are negligible. This assumption may be implausible since it is very possible that average household composition may depend on what era a cohort of individuals was born. Certainly in Western culture, there has been a trend for households to downscale in the last century (Kobrin (1976), Burch and Matthews (1987)). Thus, a concern of this analysis is that cohort effects may exist and that by failing to account for them in the analysis, they will confound the age and/or year effects. Second, the model in (7) fails to take interactions between age and year effects into account. Interaction terms were excluded primarily because their inclusion would have vastly increased the number of independent variables. The estimates of the model would have then become difficult to interpret and susceptible to problems of over-fitting. However, by excluding interaction terms, the model relies on the assumption that the shape of the age profile of average household composition is the same in every year in the sample. If age and year effects are not purely additive, then the estimated age profiles for each measure of average household composition will be biased. Finally, the size of the older cohorts is notably small and therefore the sample means obtained from older cohorts may not be representative of the true population means. This measurement error will mean that the results produced from the decomposition may be less accurate than hoped.

In light of the above limitations, the results from the decomposition should be interpreted with caution. The decomposition procedure is therefore regarded as exploratory, and the estimated age effects are used to provide some insight into the *possible* life-cycle trends in the average household composition of older African Women, rather than definitive answers.

University of Cape Town

Fixed Effects Model

A fixed effects regression is run to estimate the relationship between the average household composition of the cohorts and the age-eligibility of the cohorts. In essence, the results from the fixed effects regression can be used to test whether there is a shift in the age profiles of the various measures of household composition on receipt of the pension. The particular regression model used was:

$$\bar{y}_{ct} = \beta + \phi age_{ct} + \varphi age_{ct}^{2} + \alpha d_{ct} + \bar{\theta}_{c} + \bar{u}_{ct} \qquad c = 45,\ldots,75 \quad t = 1,\ldots,9 \qquad (8)$$

where $\bar{y}_{ct}$ is the average household composition of cohort c in year t, β is the constant term, age_{ct} is the age of a cohort c in year t, and finally d_{ct} is an indicator variable for whether a cohort is age-eligible (i.e. over the age of 60) or not. d_{ct} equals one if age is greater than 59 and zero otherwise. The terms ϕ, φ, α represent the parameters that are to be estimated by the model. The term, α, represents the pension effect and is therefore of greatest interest. Cohorts 45 to 75 were used for calculating all measures of household composition apart from the average number of men and women aged between 40 and 49. For the latter two measures, only cohorts 50 to 75 were included. This is because the numbers of men and women aged between 40 and 49 in cohorts 45 to 49 are substantially larger than the number of men and women aged between 40 and 49 in cohorts 50 and 75 by design and thus the inclusion of cohorts 45-49 would distort the regression results.

University of Cape Town

The model only includes 3 factors, namely *age*, age^2 and the *dummy for age-eligibility*. This does not mean that age is the only variable apart from age-eligibility assumed to affect average household composition. It means that the relationship between other factors affecting the average household composition of the cohorts and age eligibility is assumed to be accurately captured by either a linear or quadratic function of age. For example, a variable for year is not included in the model although there is preliminary evidence to show that year is likely to be negatively related to average household composition. This is because the relationship between year and the average household composition of the cohorts will be captured by the *age* variable as both age and year are linear. This means, however, that the estimated parameters on *age* and age^2 will not provide an accurate measure of the true relationship between age and the average household composition of the cohorts. Thus, the estimates of the pension effect, α, are the focus of this part of the analysis.

The fixed effects model in (8) relies on two important assumptions. First, it assumes that $\theta_{ct} = \theta_c$. This assumption is deemed plausible since the sizes of the cohorts are mostly large. It estimates α using the standard within estimators and therefore makes no correction for measurement error. It therefore implicitly assumes that the sample sizes are large enough to be representative of the actual population means. While the cohort sizes are large on the most part, the sizes do diminish with age. Thus, the latter assumption may be questionable, and the estimates may be attenuated.

University of Cape Town

4.3 Regression Discontinuity Approach

In addition to the Cohort analysis, the study uses a Regression Discontinuity approach to measure the pension effect on living arrangements. This approach estimates the pension effect as the discontinuous change in the living arrangements of African women at age eligibility, while also accounting for the continuous effect of age on household composition. Three different approaches to measuring a regression discontinuity are used in the analysis, namely a parametric approach, Partial Linear Estimation and Local Linear Regression. Each of the above approaches is outlined in detail at the end of this chapter. This discussion is preceded by a detailed explanation of Regression Discontinuity Designs and their suitability to this study, as well as a description of the data used in the analysis.

4.3.1 Regression Discontinuity Designs

Regression Discontinuity designs (RD designs) refer to a group of models that allow one to measure the effect of a program or treatment, in this case, the pension. In RD designs, participants are assigned to treatment according to a cut-off score or pre-program measure (Trochim (1984)). The assignment to treatment can be either a deterministic function of the cut-off value (Sharp design) or a stochastic one (Fuzzy Design), however it depends on the cut-off value in such a way that the probability of treatment varies discontinuously at this point (Porter (2003)).

RD designs differ from randomized experiments and other quasi-experimental designs by the way in which individuals are assigned to treatment (Trochim (1984)). In the latter designs, treatment is randomly assigned to individuals in a sample. The randomization ensures that the treatment and control groups are similar, and that therefore a comparison between the average value of the outcome variable of the treatment group and the average value of the control group will provide an accurate measure of treatment. Since assignment to treatment is non-random in RD designs, treatment and control groups are inherently different. Therefore the difference between the average values of the outcome variable for each group will not provide an accurate estimate of the treatment effect.

University of Cape Town

This idea can be illustrated formally using regression analysis. Consider the following model of household composition on pension assignment. Having collected data on a random sample of individuals, some having received the pension and some having not, one may consider estimating the pension effect using the following regression model:

$$Y_i = \beta + \alpha d_i + \varepsilon_i \tag{9}$$

where Y_i = household composition of individual i

β = constant term

α = pension effect

d_i = 1 if age >=60 and 0 otherwise

ε_i = error term

When assignment to treatment is not random as in the case of the pension, the variable d and the residuals, ε_i, will be related and the OLS estimate of α will be biased. The dependence between the variable d and the errors arises because the variable d is related to some unobserved characteristic, say X, which is in turn related to the outcome variable Y. In the example of the pension, the probability of receiving the pension, d_i, is dependent on the error term, ε_i, because d_i is dependent on one's age as is one's household composition. Thus, while randomization will ensure that the treatment and control groups are as similar as possible in all aspects apart from treatment, because the assignment of treatment is non-random these groups will be different at least in their average values of X (Van der Klaauw (2002)).

RD designs compensate for the non-random nature of treatment assignment in at least one of two ways. First, RD designs typically only include individuals with values of X very close to the cut off value, say $\bar{x}$, into the sample (Van der Klaauw (2002)). Under these circumstances, the treatment and control groups will be very similar even with regard to their values of X (ibid). Thus, the two groups are likely to have similar average outcomes in the absence of treatment and similar average outcomes after receiving

University of Cape Town

treatment (ibid). The difference in the average outcomes of the two groups would therefore provide an accurate measure of treatment. Of course, since this method only includes individuals with X values within a small band of $\bar{x}$, the method is only advisable if a very large dataset is available.

Alternatively, one can use a slightly larger interval around cut-off and still estimate the treatment effect under a RD design if an assumption is made about the functional form of the relationship between the outcome variable, Y, and the value of X (Van der Klaauw (2002)). If one can model Y as a function of X, then this function can be included in equation (9) so as to control for any dependency between the variable d and ε_i. In particular, equation (9) can be rewritten as:

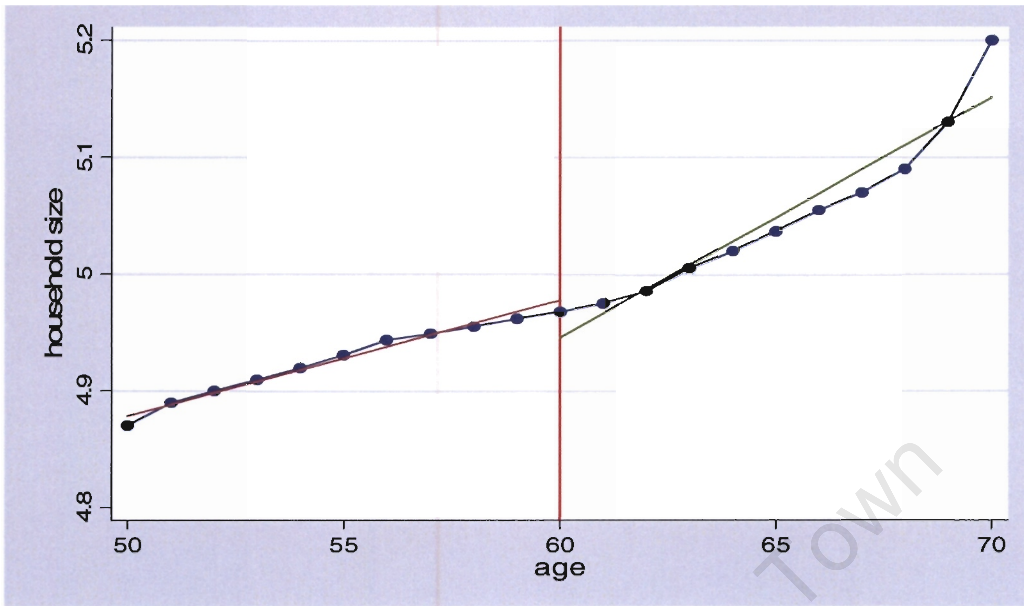
$$Yi = \beta + \alpha d_i + m(x_i) + \varepsilon_i \tag{10}$$

where m(.) is a function that captures the relationship between X and Y. If m(.) is correctly specified, then equation (10) will provide unbiased estimates of the expected treatment effect at cut-off i.e. $E[\alpha_i \mid \bar{x}]$.

From equation (10) above it is clear that the estimate of the pension effect will rely heavily on how m(.) is modelled. Until recently, m(.) has been modelled using parametric techniques. However, because m(.) cannot be observed directly, assuming a functional form for m(.) can often be quite difficult. Furthermore, if the functional form of m(.) is mis-specified, then the estimate of the average treatment effect is likely to be inaccurate. Consider for example estimating a pension effect using the hypothetical data of household size illustrated by Figure 5 below:

University of Cape Town

Figure 5: An Illustration of the Difficulties of Accurately Estimating Program Effects by assuming a Functional Form for m(.)



University of Cape Town

By looking at the values for household size around age 60, it is clear that there is no pension effect at this point. However, if m(.) is incorrectly specified as linear then the regression model (10) above will incorrectly indicate the presence of a pension effect.

In an effort to resolve this issue, researchers have begun modelling m(.) using non-parametric techniques. The groundwork for this approach was laid down by Hahn, van der Klaauw and Todd (2001), who show that treatment effects can be identified by imposing only *two* parametric restrictions on m(.). The first restriction is that the expected value of m(.) is continuous at the cut-off. In other words, it should be plausible to assume that in absence of treatment, individuals in a small interval around cut-off would have similar average outcomes. In terms of this study, it would mean assuming that, in absence of the pension, the household composition of African women changes smoothly as the women turn 60. The validity of this assumption is discussed in due course. Second, their approach assumes that the mean treatment function $E[\alpha_i|X]$ is right continuous at cut-off. Essentially, this means that the effect of the pension on 60 year-old women will be similar in magnitude to its effect on women slightly older than 60. The authors also touch on two different types of treatment effects, namely Constant Treatment

Effects where treatment effects are constant across individuals included in the analysis, and secondly, Variable Treatment Effects where treatment effects are heterogeneous. Under the assumption of constant treatment effects, the model will identify the constant treatment effect α, and under the assumption of variable treatment effects the model will estimate the average treatment effect at the cut-off. Note that if constant treatment effects are assumed then the assumption that $E[\alpha_i | \bar{x}]$ is right-continuous will hold automatically.

Under the above assumptions, Hahn, van der Klaauw and Todd (2001) show that the treatment effect (or the average treatment effect) under a Sharp Design is equivalent to the jump in the conditional expectation of the outcome variable at the cut-off boundary. More specifically,

$$\alpha = \lim_{X \downarrow \bar{x}} E(Y \mid X) - \lim_{X \uparrow \bar{x}} E(Y \mid X) \qquad (11)$$

where α is the treatment effect and $\bar{x}$ is the value of X for which treatment occurs[8]. This result is valuable because it presents another way of estimating the treatment effect. That is, if one can determine the expected value of Y on either side of the boundary, then the pension effect can be estimated as the difference in these two values. Of course, there are a variety of ways in which these limits can be estimated. Perhaps the most simplistic approach would be to take the average values of Y at the value of X just below (or just above) the pension cut-off. Alternatively, one could calculate a weighted average of Y over a few values of X on either side of the cut-off. This is also referred to as Nadaraya-Watson Estimation (Porter (2003)). More sophisticated approaches using a weighted linear regression of Y on X on either side of the boundary have also been used, and these will be discussed more thoroughly in due course[9].

[8] The related proof can be found in Hahn, van der Klaauw and Todd (2001).

[9] Under an additional assumption, a similar result will hold for Fuzzy Designs. This assumption and the related proof can also be found in Hahn, van der Klaauw and Todd (2001).

University of Cape Town

In summary, RD designs provide an accurate estimate of a treatment effect when assignment to treatment is non-random. They attempt to correct for the fact that individuals in the treatment group will be different from those in the control group by only using data close to the treatment cut-off and by controlling for the difference in the X values between the two groups through the function m(.). The function m(.) can be modelled parametrically. This approach will work best if something is known about the functional form of m(.). Alternatively, one can model m(.) non-parametrically, using minimal parametric restrictions as laid out by Hahn, van der Klaauw and Todd (2001). Finally, under the same assumptions set out by Hahn, van der Klaauw and Todd (2001), the treatment effect can be estimated by calculating the difference in the conditional expectation of Y just prior to treatment cut-off and the conditional expectation of Y just after treatment cut-off. Since weighted linear regression is often employed for the latter approach, this approach tends to rely on non-parametric techniques as well.

4.3.2 Suitability of Regression Discontinuity in this study

For the purposes of this study, the effect of the pension can be estimated using RD designs. Treatment assignment is non-random. Furthermore, the probability of treatment (where treatment is defined as age eligibility) will change discontinuously at age 60. Since the probability of being eligible is a deterministic function of age, the model will follow a Sharp Design. The function m(.) will represent the relationship between household composition and age. The aim of the analysis will to be estimate α, the coefficient of the indicator for whether a woman is age-eligible or not.

The study assumes a Constant Treatments Effect Model. Under the Constant Treatment Effects Model, only one parametric restriction on m(.) needs to hold for RD designs to be applicable. This assumption is that, in absence of the pension, household composition is a continuous function of age at age 60. Put more simply, it should be plausible to assume that in absence of pension, individuals in a small interval around age eligibility would have similar average household composition. Without this assumption, an estimated discontinuity in household composition at age 60 could be the attributable to any one of a

University of Cape Town

number of other factors apart from the pension. Consequently, it would be very difficult to discern the existence, let alone the size, of the pension effect.

The smoothness assumption will be violated if there are any other factors that cause household composition to change discontinuously at age 60. Since the retirement age for women working in the formal work sector is 60, one possible concern is that a discontinuity measured at age 60 may partly (or wholly) be a measure of the effect of retirement on household composition and not the effect of the pension. Indeed, if women choose to retire at age 60 regardless of whether the social pension is received or not, the effects of the pension on household structure are likely to be confounded by the effects of retirement. However, the results presented in Figure 4 in Chapter 3 show that only a small proportion of African women may be affected by retirement and thus retirement effects, if they even exist, are likely to be very small. Thus, for the purposes of this study, retirement effects are assumed to be negligible and the smoothness assumption is deemed to hold.

When considering the suitability of the outlined RD method to this study, it is also worth noting the potential advantages and disadvantages of each of the above-mentioned RD designs. While the non-parametric approach is advantageous in that it frees the model from any functional from assumptions of m(.), this method will suffer from the usual biases inherent in non-parametric models. Non-parametric models require the modeller to choose a bandwidth and a kernel for each regression. While one's choice of kernel is rarely of any practical consequence, the choice of bandwidth can affect results significantly. A small bandwidth can produce inaccurate results if very large datasets are not available. On the other hand, a bandwidth that is too large may introduce a bias if individuals further away from the discontinuity point are systematically different from those closest to the discontinuity point. Unfortunately, there is little guidance on how one chooses a bandwidth in the literature. This study follows Edmonds, Mammen and Miller (2005) by using an Epanechnikov kernel with a bandwidth of 2 years. In addition, the results were re-run using a variety of bandwidths and the results were quite robust to a change of bandwidth.

University of Cape Town

Given these potential difficulties, this study uses each of the three approaches outlined above, and collectively pools the results. Each approach is, in a sense, used as a test for the others. For example, results from an approach that are substantially different from the others will be treated with caution. The first approach used is a parametric approach, where m(.) is modelled as a linear and then a quadratic function. The second approach uses non-parametric techniques to model m(.). The specific method used is called the Partial Linear Estimation (PLE), and is outlined in Porter (2003). Finally, the pension effect is also estimated by calculating the jump in conditional expectation at the pension cut-off. The method employed is also described by Porter (2003), and is called the Local Linear Estimation (LLE). Since this method also uses non-parametric regression to estimate the pension effect, it is categorized as the second non-parametric approach in this study. Both of the above non-parametric approaches were chosen because they were recommended by Porter (2003) for having bias-reducing properties[10].

4.3.3 Parametric Approach

The first RD model estimates the pension effect by assuming different polynomial specifications for the function m(.). These specifications are linear and quadratic functions. The model therefore relies on the assumption that all changes in household composition apart from those attributable to the pension can be accurately modelled as a linear or a quadratic function of age. The validity of this assumption can be assessed by observing how well the linear and quadratic functions fit the data.

The linear and quadratic models used are presented in equations (12) and (13) below:

$$Y_i = d_i\alpha + \beta \text{age} + \varepsilon_i \qquad (12)$$

$$Y_i = d_i\alpha + \lambda \text{age}^2 + \varepsilon_i \qquad (13)$$

[10] For details on the bias-reducing properties of the PLE and LLE see Porter (2003). An overview of the bias-reducing properties of the LLE can be found in section 4.3.5 of this study.

University of Cape Town

where β and λ represent the coefficients of the age and age^2 variables respectively and Y_i represents the various measures of household composition. The significance of the pension effect, α, was tested using a t-test. No weights were used in the above regressions.

4.3.4 Partial Linear Estimation

The second RD model estimates the pension effect by modelling m(.) non-parametrically. This would be a straight-forward exercise if one knew the value of α, for then subtraction of the value of α from the outcome variable would produce the new transformed outcome variable which can be used to estimate m(.) (Porter (2003)). In other words, one could transform regression (10) as follows:

$$(Y_i - d_i\alpha) = m(x_i) + \varepsilon_i \tag{14}$$

where $\hat{Y} = (Y - d\alpha)$ is the new dependent variable. Once m(.) was known, the estimate of α could be found by estimating the value of α that minimizes the average squared deviation between the new dependent variable, $\hat{Y}$, and the nonparametric estimate of m(.) (Porter (2003)). Of course, this approach is infeasible because one will not know the value of α to start off with. However, Porter (2003) shows that one can go about estimating α using a similar approach. Porter (2003) points out that the pension effect can be found by solving the following optimization problem:

$$\min \sum_{i=1}^{n} [y_i - d_i\alpha - \sum_{j=1}^{n} w^i{}_j (y_i - d_i\alpha)]^2 \tag{15}$$

where $w^i{}_j = \dfrac{k_h(x_i - x_j)}{\sum_{l=1}^{n} k_h(x_i - x_l)}$, $k_h(u) = k(\frac{u}{h})$

and k is the type of kernel used, h is the bandwidth and n is the number of observations used.

Using this function, the value of α is estimated by minimizing the difference in the new dependent variable, $\hat{Y}$, and the nonparametric estimate of m(.), $\sum_{j=1}^{n} w^{i}{}_{j}(y_j - d_j\alpha)$. The kernel weighting will ensure that points closest to the cut-off value will have the greatest weight in the calculation of $\hat{\alpha}$.

This function has an analytic solution which Porter (2003) shows to be:

$$\hat{\alpha} = \frac{\sum_i [(d_i - \sum_j w^{i}{}_{j} d_j)(y_i - \sum_j w^{i}{}_{j} y_j)]}{\sum_i (d_i - \sum_j w^{i}{}_{j} d_j)^2} \qquad (16)$$

Thus, once a kernel and bandwidth have been chosen for the model, the estimate of α is calculated from equation (16).

The standard error of $\hat{\alpha}$ is important for inference. Porter (2003) finds the standard error of $\hat{\alpha}$ to be:

$$\sigma^2 = \frac{\sigma^{2+}(\bar{x}) + \sigma^{2-}(\bar{x})}{4 f_0(\bar{x}) nh} v_k \qquad (17)$$

where

$$\sigma^{2+}(\bar{x}) = \frac{1}{n}\sum_{i=1}^{n} \frac{\frac{1}{h} k_h(\bar{x}-x_i) d_i \hat{\varepsilon}_i^2}{\frac{1}{2}\hat{f}_0(\bar{x})} \quad \text{and} \quad \sigma^{2-}(\bar{x}) = \frac{1}{n}\sum_{i=1}^{n} \frac{\frac{1}{h} k_h(\bar{x}-x_i)(1-d_i) \hat{\varepsilon}_i^2}{\frac{1}{2}\hat{f}_0(\bar{x})}$$

$$\hat{\varepsilon}_i = y_i - \hat{m}(x_i) - d_i\hat{\alpha}$$

University of Cape Town

$$\hat{m}(x) = \frac{1}{n}\sum_{j=1}^{n}\frac{\frac{1}{h}k_h(\bar{x}-x_j)(y_j-d_i\hat{\alpha})}{\hat{f}_0(x)}$$

$$\hat{f}_0(x) = \hat{f}_+(x) + \hat{f}_-(x)$$

$$\hat{f}_+(x) = \frac{1}{n}\sum_{i=1}^{n} k_h(x - x_i)d_i$$

$$\hat{f}_-(x) = \frac{1}{n}\sum_{i=1}^{n} k_h(x - x_i)(1 - d_i)$$

$$v_k = (\int_0^\infty K_0{}^2(w)dw)^{-2}\int_0^\infty (K_0(w) + L(w) - L(-w))^2 dw$$

$$K_j(w) = \int_w^\infty k(u)u^j du$$

$$L(w) = \int_0^\infty K_0(u)k(u+w)du$$

and $f_o(\bar{x})$ is the density of the distribution of age at the discontinuity, $\bar{x}$.

Thus, using the above formulae, the values of $\hat{\alpha}$ and the standard error were calculated for each measure of household composition. The weights were calculated using an Epanechnikov kernel with a bandwidth of 2 years. The p-values for each value of $\hat{\alpha}$ were calculated and used as a test for significance. Once $\hat{\alpha}$ had been calculated, m(.) could be calculated by first subtracting d$\hat{\alpha}$ from the outcome variable Y and then by running a non-parametric regression of $\hat{Y}$ against m(.). Person weights supplied in the census data were not used in the regressions.

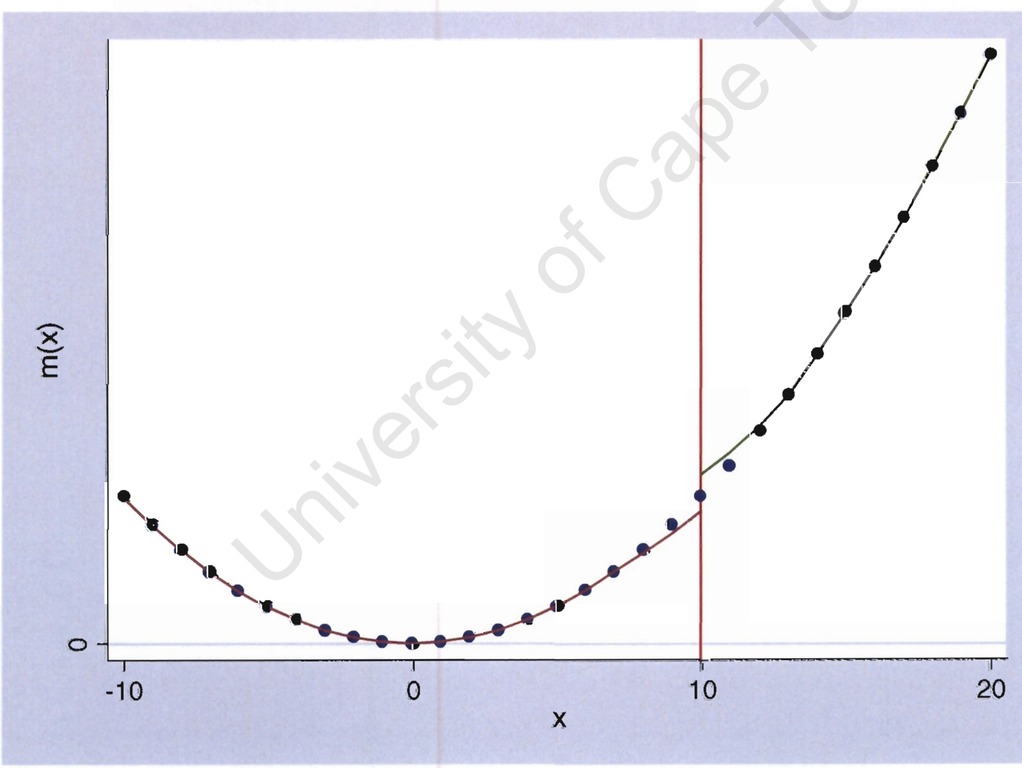
University of Cape Town

4.3.5 Local Linear Estimation

The regression discontinuity can also be estimated by the size of the jump in the conditional expectation at a boundary (in this case, at age 60). The third and final RD model estimates the pension effect in this way.

The simplest kernel estimator of a jump in conditional expectation takes the weighted average of observations on either side of the discontinuity and differences the results. This technique is also referred to as Nadaraya-Watson Estimation (Porter (2003)). However, this estimator is subject to bias problems common to all non-parametric estimation at a boundary (ibid).

Figure 6: An Illustration of the Biases associated with Nadaraya-Watson Estimation at a boundary



University of Cape Town

Consider the hypothetical distribution of m(x) and the Nadaraya-Watson estimates of m(x) above. The circles represent the values of m(x) itself, and the solid line represents the non-parametric estimates of m(x) using Nadaraya-Watson. Owing to the positive slope of m(x) near the discontinuity, observations just to the right of the discontinuity are likely to be greater than the right limit at the discontinuity. Thus, the weighted average of these observations will provide an upward biased estimate of the right limit, as is illustrated by the upward-sloping line at x=10. Similarly, the observations to the left of the discontinuity will likely be less than the left limit, and thus will provide a downward biased estimate of the left limit. Consequently, an estimate of the jump in conditional expectation at x=10, will be biased.

Owing to the biases associated with the Nadaraya-Watson estimator, this study uses a Local Polynomial estimator instead. In Local Polynomial Estimation, the estimate of the right limit is equivalent to the intercept of a weighted regression using data to the right of the discontinuity. Similarly, the estimate of the left limit is equivalent to the intercept of a weighted regression using data on the left side of the discontinuity. As in Nadaraya-Watson estimation, the treatment effect is then calculated as the difference between the two limits. However, by accounting for the polynomial type behaviour of the conditional expectation near the boundary, Local Polynomial Estimation ensures that the estimate of the level at the boundary is free of the biases associated with Nadaraya-Watson Estimation (Porter (2003)).

One can estimate the jump in conditional expectation using Local Polynomial Estimation as follows. If (α_+,β_+) is the solution to the following minimization problem:

$$\min_{a,b_1,\ldots,b_p} \sum_{i}^{n} [\frac{k_h(x - x_i)d_i}{n}][y_i - a - b_1(x_i - \hat{x}) - \dots - b_p(x_i - \hat{x})^p]^2 \tag{18}$$

and if (α_-,β_-) minimizes the analogous criterion with 1 – di replacing di, then the estimate of the jump in conditional expectation (i.e. the estimate of the pension effect, $\hat{\alpha}$) will be:

$$\hat{\alpha} = \alpha_+ + \alpha_- \tag{19}$$

University of Cape Town

The standard error associated with this estimator, as outlined by Porter (2003), is:

$$\sigma^2 = \frac{\sigma^{2+}(\bar{x}) + \sigma^{2-}(\bar{x})}{f_0(\bar{x})nh}\, e_1{}'\Gamma^{-1}\Delta\Gamma^{-1}e_1 \tag{20}$$

$$e_1 = (1,0,....,0)' \quad , \quad \Gamma = \begin{bmatrix} \gamma_0 & \cdots & \gamma_p \\ \vdots & & \vdots \\ \gamma_p & \cdots & \gamma_{2p} \end{bmatrix} \text{ and } \Delta = \begin{bmatrix} \delta_0 & \cdots & \delta_p \\ \vdots & & \vdots \\ \delta_p & \cdots & \delta_{2p} \end{bmatrix}$$

$$\gamma_j = \int_0^{\infty} k(u)\,u^j du \text{ and } \delta_j = \int_0^{\infty} k^2(u)u^j du$$

In this study, p is assigned the value of 1 and hence this estimation is termed Local Linear Estimation. This decision was made in accordance with the recommendation of Fan and Gijbels (1996). They argued that the degree of the polynomial should be restricted to 1 and that the bandwidth should be used to control model complexity. The weights of this model are calculated using an Epanechnikov kernel with a bandwidth of 2 years. Again, person weights were not used in the regressions.

4.3.6 Regression Discontinuity Data

In order for this analysis to have some statistical power, a very large dataset is necessary. Data from the 10 percent sample of the 2001 Population and Housing Census of South Africa were therefore used[11]. This dataset was restricted to include only African headed-households. In those cases where the race of the household head was not indicated, the race of a household was determined by the race of the next closely- related member in the household. All households including one or more individuals who failed to report their age were also excluded. Finally, the dataset was reduced to include only African women who were citizens of South Africa and aged between 50 and 70. The resultant sample size was 148 564.

University of Cape Town

[11] A more detailed description of this dataset can be found in Section 4.2.4.

From this sample, a second dataset was created. This dataset excluded all African women aged below 60 who already lived with a pensioner. 3 061 women are excluded in total. There were two reasons behind the creation of this dataset. First, it was hoped that by restricting the data in this way, all pension households would be removed from the control group (i.e. women below the age of 60) and that as a result, the estimated pension effects would be more accurate. Second, a comparison of the estimated pension effect from the original sample to that from the reduced sample would provide insight into whether the effect of the receipt of the pension on the composition of a household differed if an existing member of the household already received the pension. It is arguable that the pension may have a diminishing effect on the composition of a household as the number of members receiving the pension increases. However, a pitfall of using the reduced sample is that, by virtue of the definition of the sample, the sample excluded women below the age of 60 who lived with men over the age of 65. If women who lived with older men (for example, those who had older husbands) had different living arrangements to all other women, then their exclusion from the sample of women below the age of 60 would affect average household composition of women below the age of 60. Consequently, the RD model would incorrectly attribute this difference to a pension effect.

Two further datasets were then created. The third and the fourth datasets were created by excluding all women aged 59, 60 or 61 from the first and second sample respectively. The idea behind creating these datasets was that they could be used in re-estimating the pension effect so that it included all changes in the living arrangements of African women made slightly prior to or post age 60 in response to the pension. Therefore using these datasets, the pension effect was estimated as the jump in the household composition of African women between ages 58 and 62.

The approach succeeds in taking into account any changes to household composition made in response to the pension slightly prior to or post age 60. However, the approach fails to meet the primary assumption underlying RD designs, that being, in absence of the pension, household composition changes smoothly at cut-off. This assumption is

University of Cape Town

unlikely to hold because the cut-off is defined over a range. Over the range 58 to 62, household composition is likely to change significantly regardless of whether a pension effect exists or not because household composition is a function of age. As a result, the approach will incorrectly attribute some of the changes in household composition between ages 58 and 62 to the pension instead of age and hence the estimated pension effect will be biased. Nonetheless, the results from the approach may still provide some valuable insight into whether pension effects exist slightly prior to or post 60 and so are included in the analysis.

A number of variables were then created in each of the datasets. The number of children aged 0 to 5 and 6 to 17 and the number of males and females aged between 18 and 23, 24 and 29, 30 and 39, and 40 and 49, and household size were calculated for each woman in a sample. These subdivisions of household composition formed the outcome variables, Y. An indicator variable for whether a woman was age-eligible or not, d, was also created. Age, X, constituted the assignment variable (i.e. the variable that determined the value of d) and was expressed in years.

Once these variables had been created, the parametric and non-parametric RD approaches could be applied to each of the datasets. All four RD models were applied to the first and second sample. However, since the PLE estimates based on the third and fourth sample were very sensitive to a change in bandwidth, only the linear, quadratic and the LLE models were applied to the samples. The results for the four samples are presented in chapter 5.

University of Cape Town

5. Findings

5.1 Cohort Findings

Figure 7 illustrates various measures of average household composition for a number of cohorts. To prevent the graphs from becoming too busy, only every fifth cohort was included in these graphs. Each line in the graphs represent the changes that occurred to a particular measure of average household composition (e.g. average household size, the average number of children aged 0 to 5 etc), for a particular cohort over the period 1994-2003. Since some of the cohorts are observed at the same age, the lines do sometimes overlap.

The shape and direction of each line and the position of one line relative to other lines in the graph can provide insight into whether age, year or cohort effects exist. The shape and direction of each line is determined by age and year effects. This is because as one moves along a line, one will move through different ages and years, however the cohort will remain the same. Consider, for example, the changes that occurred to the average household size of cohort 50. Over the period 1994-2003, the average household size of cohort 50 decreased from approximately 6 members in a household to only 5 members. The average household size of cohort 50 may have dropped owing to the fact that over this period, the cohort aged from 50 to 59 and African women who are 59 tend to live in smaller households than African women who are 50. Alternatively, the drop in average household composition may have occurred because, between 1994 and 2003, macro-level shocks forced average household size of all African individuals (and therefore cohorts) to decrease. An example of such a macro-level shock would be if government increased the supply of state housing available to African individuals. Of course, the change in the average household composition of cohort 50 over this period could also have be the result of both age and year effects. One way to begin distinguishing age effects from year effects would be by considering whether each cohort line takes on a common shape. Since year effects will be the same for each and every cohort, each line should take on a common shape if these effects exist. Take for example the graph of Average Number of

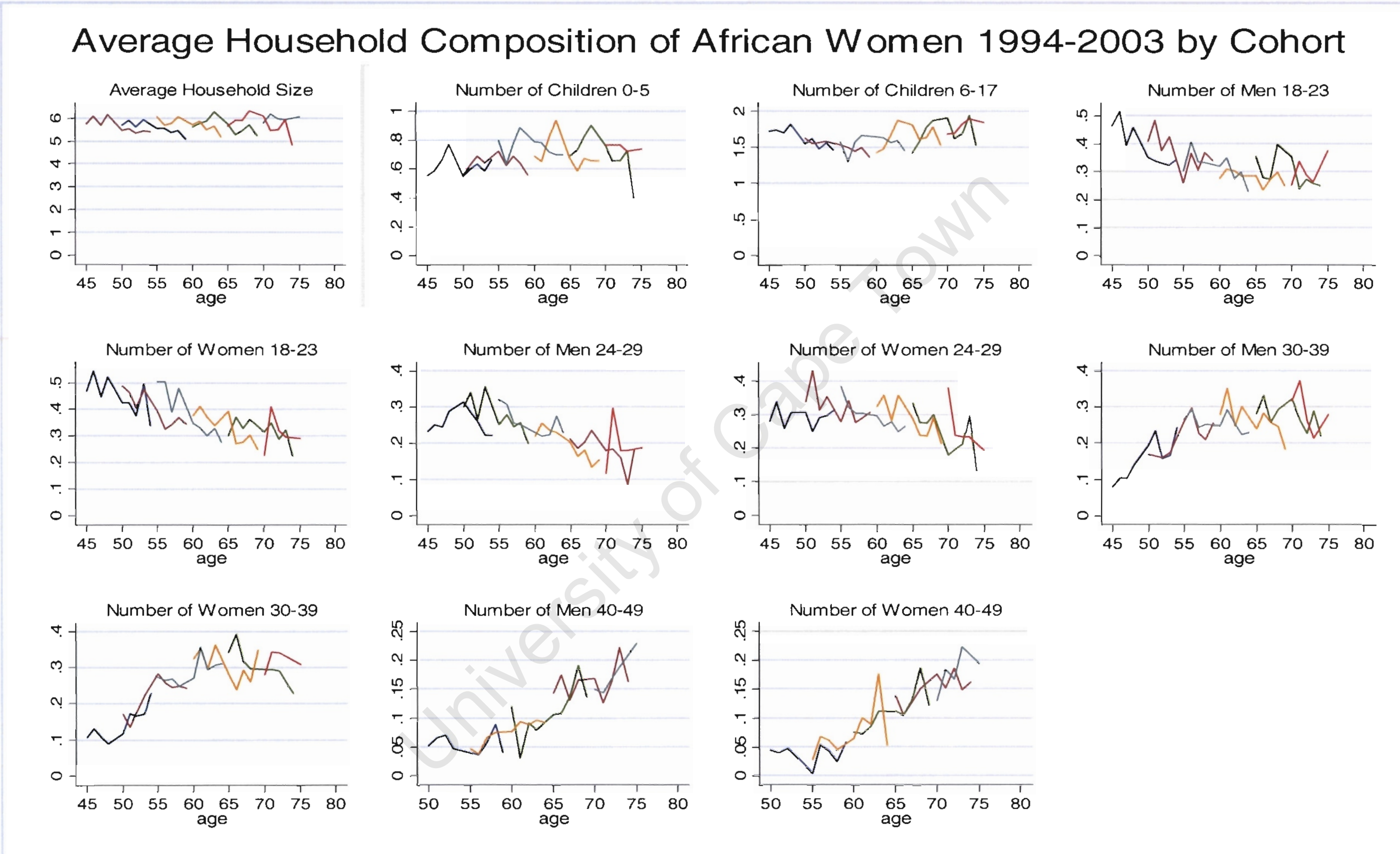
University of Cape Town

Children aged 0-5. The cohort lines in this graph appear to rise sharply in 1997. Since this jump occurs in each and every line of the graph, it is likely that the jump represents a year effect. By removing the common pattern from each of the cohort lines, it would be possible to unveil the hidden age profiles of each cohort.

Cohort and year effects determine the position of a line relative to other lines in the graph. At any given age, the difference in average household composition between two lines exists because the values for the cohorts and years will differ between the lines at this point. Again, this concept can be illustrated more clearly with an example. Consider the average number of women between the ages of 18 and 23 that reside in the households of African women. For African women of age 55, the number of women aged between 18 and 23 that one would expect to reside in their households will be either 0.35 or 0.5 depending on whether the 55 year-old African women belonged to cohort 50 or cohort 55, or whether the observation was made in 2001 or 1994 respectively. The difference of 0.15 may be attributable to the fact that younger cohorts of African women tend to live with fewer women between the ages of 18 and 23 than older ones (a cohort effect), or because in 2001, fewer young women lived with older African women than in 1994 (a year effect). The change could also have been a result of a combination of cohort and year effects. One way to attempt to differentiate cohort effects from year effects would be by considering whether the difference in average household composition between two lines is consistent across the entire length of the lines. Cohort effects will be consistent between two lines because one will be comparing the effect of the same two cohorts at each point along the lines. Year effects may not be as consistent because one will be comparing the effect of two different years at each point along the lines. Note that if neither cohort nor year effects exist, each successive cohort line will lie exactly on top the previous one.

University of Cape Town

Figure 7: The Change in the Average Household Composition of Cohorts of African Women over the period 1994-2003



Source: OHS for years 1994, 1995, 1997 and 1999, LFS for years 2000, 2002 and 2003, and Population Censuses for 1996 and 2001

University of Cape Town

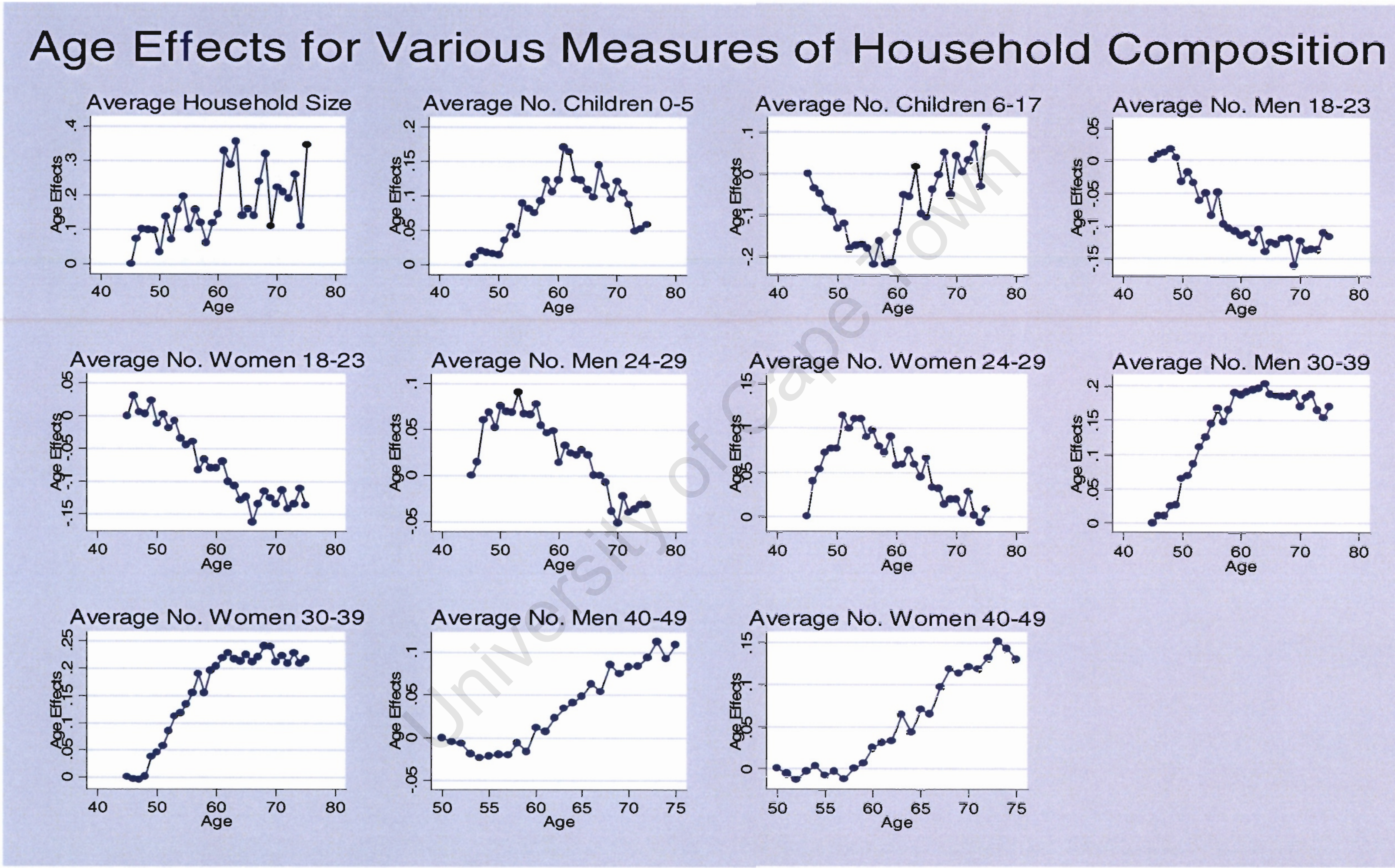
With the exception of 3 graphs, namely the graph depicting average household size, the graph showing average number of women aged 18-23, and the graph showing the average number of women aged 24-29, successive cohort lines do not lie consistently above or consistently below previous ones. This may be because the cohort effects are being masked by year effects. Alternatively, it could mean that cohort effects are negligible. On the other hand, year effects definitely seem to be present. Many lines in each of the graphs appear to jump up and down in *similar* patterns. The presence of year effects would also explain why successive cohort lines do not lie precisely on top of previous ones. Finally, age effects are also likely to be present. The age effects appear to confound the year effects because the lines in any given graph do not follow *identical* patterns.

By analyzing Figure 7, two important points become clear. First, the assumption of no cohort effects seems plausible in at least 8 of the 11 measures of average household composition considered. Second, for the 3 remaining measures of average household composition, namely for average household size, the average number of women aged 18-23, and the average number of women aged 24-29, it is plausible that the differences between successive cohorts are partly attributable to year effects. If negative year effects exist, then successive cohort lines will lie above previous ones as seen in the graphs of the 3 aforementioned measures of average household composition. Furthermore, the fact that the several of the cohort lines in the above 3 graphs are downward-sloping provides further evidence that negative year effects may exist. With these observations in mind, the decomposition of the measures of average household composition may provide more accurate estimates of age profiles than anticipated.

The cohort data are decomposed into its constituent age and year effects by running the regression presented in (7). The estimated coefficients of the age and year dummies in this regression constitute the estimates of the age and year effects respectively. The estimates of the age and year effects are presented in Figure 8 and Figure 9 below.

University of Cape Town

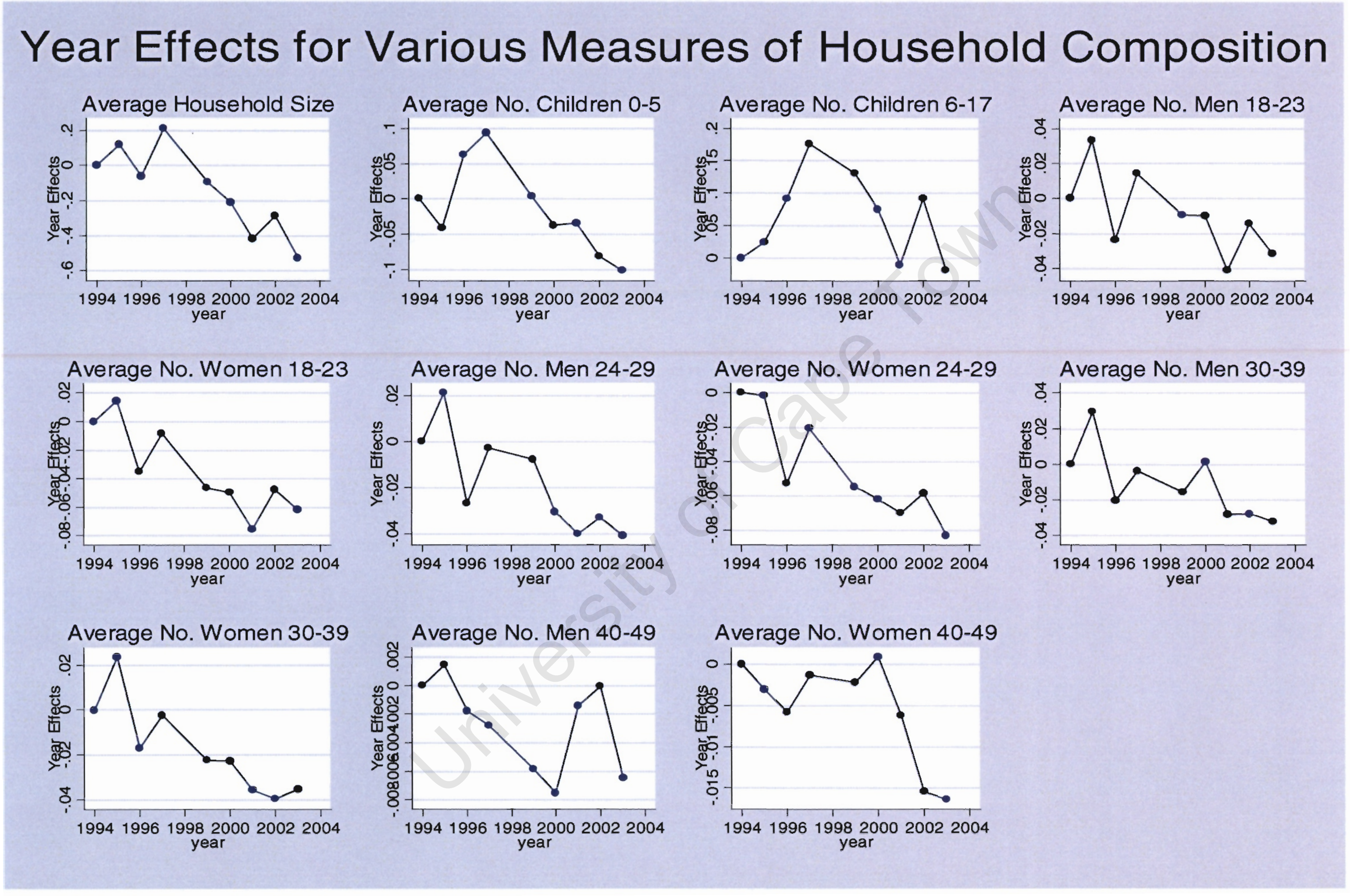
Figure 8: Estimated Age Effect on various measures of Average Household Composition of African Women using Cohort Data



Source: OHS for years 1994, 1995, 1997 and 1999, LFS for years 2000, 2002 and 2003, and Population Censuses for 1996 and 2001.

University of Cape Town

Figure 9: Estimated Year Effect on various measures of Average Household Composition of African Women using Cohort Data



Source: OHS for years 1994, 1995, 1997 and 1999, LFS for years 2000, 2002 and 2003, and Population Censuses for 1996 and 2001.

University of Cape Town

The age effects estimate the life cycle profiles of the various measures of average household composition. According to the results, the average household size of African women tends to rise as these women age between 45 and 75. The results show that average household size grew by approximately 0.3 over this period of the life-cycle. The result is interesting considering that, from the graph on average household size in Figure 7, average household size of the cohorts appeared to decrease over the period 1994-2003. Thus, it illustrates how the age profiles can be masked by the existence of year effects. The decomposition findings also purport that the average number of children aged 0 to 5 in the households of African women rise as African women approach age-eligibility, but drop soon after. This analysis shows that the average number of children aged 0 to 5 in the households of African women rose by approximately 0.2 as African women approached 60, but dropped by approximately 0.15 soon afterwards. Quite the opposite was seen to occur to the average number of children between the ages of 6 and 17 in households of older African women. The average number of children between the ages of 6 and 17 in households of African women dropped by approximately 0.2 as African women aged from 45 to 60. However, it rose again as the women aged from 60 to 75. Either way, the results suggest that the life cycle profiles of the average number of children in households of African women may be associated with age-eligibility.

The results from the decomposition suggest that the average number of men and women between the ages of 18 and 23 in the households of African women will decline by roughly 0.15 as the women age from 45 to 75. The findings also show that the average number of men between the ages of 24 and 29 and the average number of women between the ages of 24 and 29 in the households of older African women is likely to rise by 0.1 and 0.12 respectively as African women age between 45 and 52, and then decline by 0.15 and 0.1 respectively, after this point. Older African women are also expected to live with increasingly more men and women between the ages of 30 and 39 as the women age from 45 to 60. The average number of men will, on average, grow by about 0.2, and the average number of women by approximately 0.25. However, after age 60, the average number of men and women between the ages of 30 and 39 living in their households will more or less stabilize. The results therefore imply that age-eligibility

University of Cape Town

may be associated with the average number of men and women between the ages of 30 and 39 living in the households of age-eligible African women. Finally, the results from the decomposition show that as African women age from 45 to 75, the number of men and women between the ages of 40 and 49 will, on average, increase. The average number of men and the average number of women will increase by 0.1 and 0.15 respectively.

The year effects are presented in Figure 9. Apart from the year effects estimated from the model on the average number of children aged 6-17, the estimated year effects for the various measures of average household composition provide support for the earlier assertion that negative year trends existed over the period 1994-2003. As expected, the average household size of the cohorts of African women dropped in almost every successive year in the observation period. In total, the change from 1994 to 2003 caused average household size of African women to drop by 0.6. Since the year effects are not the primary interest of this study, it is not worth analyzing the estimated year effects in too much detail. However, before concluding, it is important to note that the estimated year effects are susceptible to the different sampling techniques used by each survey in each different year. This is clear from the way in which the year effects estimated for almost all of the 11 measures of average household composition jump up after the census years 1996 and 2001. Thus, the predicted year effects may be slightly confounded by the effect of different methods of sampling. This is not of concern to this study as long as the effects of sampling do not affect the estimated age effects.

Finally, the cohort data were used to estimate to whether the average household composition of the cohorts was associated with the receipt of the pension itself. This was achieved by running a fixed effects regression of average household composition against an instrument for pension receipt - a dummy for age-eligibility. The results of the fixed effects regression are presented in Table 5 below. Table 5 presents the estimate of the pension effect, alpha, for each measure of average household composition as well as the t-statistic associated with it. Asterisks are used to indicate when a pension effect is significant at a 5% level of significance. To aid interpretation, the percentage change in

University of Cape Town

each measure of average household composition at age 60 is calculated using alpha and the average value of the measure of average household composition at age 59. These values are presented in column 4. The average values of each measure of average household composition at age 59 are presented in column 3.

Table 5: Results from the Fixed Effects Regression of Average Household Composition against Age-Eligibility

measure of average household composition	alpha	t statistic	average at age 59	percentage change
average household size	0.1424	1.8185	5.6299	2.63%
average no children aged 0-5	0.0247	0.9895	0.7012	4.02%
average no children aged 6-17	0.1736*	4.1257	1.4737	11.41%
average no of men aged 18-23	0.0038	0.2643	0.3071	1.08%
average no of women aged 18-23	-0.0033	-0.2248	0.3684	-0.90%
average no of men aged 24-29	-0.0259*	-2.0870	0.2702	-9.14%
average no of women aged 24-29	-0.0324*	-2.5885	0.3140	-9.91%
average no of men aged 30-39	-0.0151	-1.0994	0.2734	-4.80%
average no of women aged 30-39	0.0038	0.3041	0.2868	1.78%
average no of men aged 40-49	0.0162	1.5198	0.0600	31.30%
average no of women aged 40-49	0.0120	1.0775	0.0509	25.71%

Source: OHS for years 1994, 1995, 1997 and 1999, LFS for years 2000, 2002 and 2003, and Population Censuses for 1996 and 2001

The results from the fixed effects regression provide strong evidence that 3 of the 11 measures of average household composition are associated with age-eligibility. The 3 measures are the average number of children aged 6 to 17, the average number of men aged 24 to 29, and the average number of women aged 24 to 29. For the remaining measures of average household composition, there is not enough evidence to show that the measures are associated with age-eligibility. The latter result is not surprising given the small sample size of the pseudo panel.

The results show that, for each cohort, the average number of children aged between 6 and 17 living in the households of African women will, on average, rise from approximately 1.47 to 1.65 as the cohorts turn 60. This amounts to a rise of roughly 11.4%. This result implies that African women will, on average, experience a rise in the number of children aged between 6 and 17 living in their households when they turn 60.

University of Cape Town

The regression results also show that for each cohort, the average number of men aged 24 to 29 living in the households of African women will, on average, drop from approximately 0.27 to 0.24. This suggests that, on average, the number of men aged 24 to 29 living in the households of African women drops when these women become age-eligible. Finally, the average number of women between the ages of 24 and 29 living in the households of African women for each cohort is also shown to drop, on average, as the cohorts turn 60. The average drops from 0.31 to approximately 0.28. One would therefore expect the number of women between the ages of 24 and 29 living in the households of African women to drop when the women turn 60.

In conclusion, the results from the cohort analysis suggest that the household composition of African women is likely to change when they receive the state pension. First, the number of children aged between 6 and 17 living in the households of African women will, on average, rise when the women receive the pension. This result gives weight to the earlier assertion that the age profile of the number of children aged between 6 and 17 appears to begin increasing owing to age-eligibility. The results from the fixed effects regression analysis also imply that African women between the ages of 45 and 75 will, on average, live with fewer men and women between the ages of 24 and 29 when they start receiving the pension. The decomposition results suggest that the number of men and women between the ages of 24 and 29 living in the households of African women will decline as the women age from 45 to 75. Thus, the receipt of the pension does not appear to play a role in changing the direction of the age profile of this measure of household composition. Rather, one could argue that the receipt of the pension tends to exacerbate the already declining age trend in the number of men and women between the ages of 24 and 29 in the households of African women.

University of Cape Town

5.2 Regression Discontinuity Findings

The results from the Regression Discontinuity Models are turned to next. Using the Regression Discontinuity approach, the pension effect was estimated by calculating the size of the jump in the living arrangements of African women at age 60. In total, 4 types of RD models were used, namely a linear and quadratic model, and a PLE and a LLE model. These models were used on two different samples of data, where the second sample excluded all women below 60 who already lived with a pensioner.

As a sensitivity analysis, the linear, the quadratic and the LLE model were re-run on two more samples. The third and fourth samples were constructed by excluding all women aged 59, 60 or 61 from the first and second sample respectively. Using these samples, the pension effect was estimated as the jump in the living arrangements of African women between the ages of 58 and 62. In this way, the estimate of the pension effect included all adjustments to living arrangements made slightly prior to or post age 60.

The results based on the first two samples are presented in Table 6 and Table 7 below. Table 6 presents the results from the tests run on the full sample of data while Table 7 presents the results based on the sample excluding all women below 60 who already lived with a pensioner. The tables show the estimated size of the pension effect (alpha) and the t statistic associated with the estimate from each of the 4 RD models. They also present the percentage change in household composition at age 60. In order to calculate the percentage change, the mean of each measure of household composition just before the discontinuity occurs needs to be estimated. This percentage is estimated by using the projected means of household composition at age 59.95 from each model. Estimates are considered to be significant if they are associated with a t-statistic that is greater than or equal to 2.

The RD results are also presented graphically in Figure 10 below and in Figures 11–14 in the appendix. Figure 10 presents the results for the pension effect on household size. These results are also presented in Figures 11-14 in the appendix along with the results for the remaining measures of household composition. These graphs are valuable

University of Cape Town

because they can be used to assess whether m(.) has been accurately specified in a model, and hence whether the estimated pension effects are reliable. Furthermore, they describe the age profile of the various measures of household composition just prior to and post pension receipt which enables the reader to put the estimated pension effects into context. In these figures, the solid dots represent the census means for each age between 50 and 70. The accuracy of the estimate of m(.) can be measured by how well the estimated function passes through these dots.

University of Cape Town

Table 6: Regression Discontinuity Results for the Full Sample

	Parametric Approach				Non-parametric Approach			
	Linear		Quadratic		PLE		LLE	
	alpha	% change	alpha	% change	alpha	% change	alpha	% change
Household Size	0.0739*	1.37%	0.0706*	1.30%	0.0020	0.04%	-0.0079	-0.14%
	(2.3058)		(2.1987)		(0.0462)		(-0.1156)	
Children 0-5	0.0341*	5.23%	0.0305*	4.48%	0.0297*	4.34%	0.0351*	5.15%
	(3.5052)		(3.1312)		(2.7520)		(2.0779)	
Children 6-17	0.0832*	5.53%	0.0874*	5.93%	0.0299	1.97%	0.0155	1.02%
	(5.1567)		(5.4126)		(1.6281)		(0.5408)	
Men 18-23	-0.0062	-2.06%	-0.0054	-1.84%	-0.0022	-0.76%	0.0005	0.17%
	(-1.0487)		(-0.9148)		(-0.3643)		(0.0518)	
Women 18-23	-0.0076	-2.31%	-0.0068	-2.12%	-0.0086	-2.65%	-0.0074	-2.27%
	(-1.2420)		(-1.1196)		(-1.3548)		(-0.7398)	
Men 24-29	-0.0107*	-4.70%	-0.0115*	-4.90%	-0.0176*	-7.53%	-0.0160	-6.94%
	(-2.1053)		(-2.2532)		(-3.3108)		(-1.9287)	
Women 24-29	-0.0019	-0.70%	-0.0034	-1.22%	-0.0014	-0.50%	-0.0040	-1.40%
	(-0.3443)		(-0.6283)		(-0.2376)		(-0.4353)	
Men 30-39	-0.0106*	-4.70%	-0.014*	-5.54%	-0.0195*	-7.66%	-0.0212*	-8.17%
	(-2.1428)		(-2.8202)		(-3.5276)		(-2.4419)	
Women 30-39	-0.0036	-1.43%	-0.0065	-2.39%	-0.0099	-3.61%	-0.0139	-4.92%
	(-0.6904)		(-1.2594)		(-1.6999)		(-1.5177)	
Men 40-49	0.0117*	13.63%	0.0145*	22.80%	-0.0013	-1.85%	-0.0021	-2.86%
	(3.7680)		(4.6738)		(-0.4556)		(-0.4458)	
Women 40-49	0.0140*	18.89%	0.0160*	27.34%	0.0105*	17.27%	0.0086*	14.25%
	(4.8098)		(5.4863)		(3.6673)		(1.9351)	

Source: 2001 South African Population and Housing Census

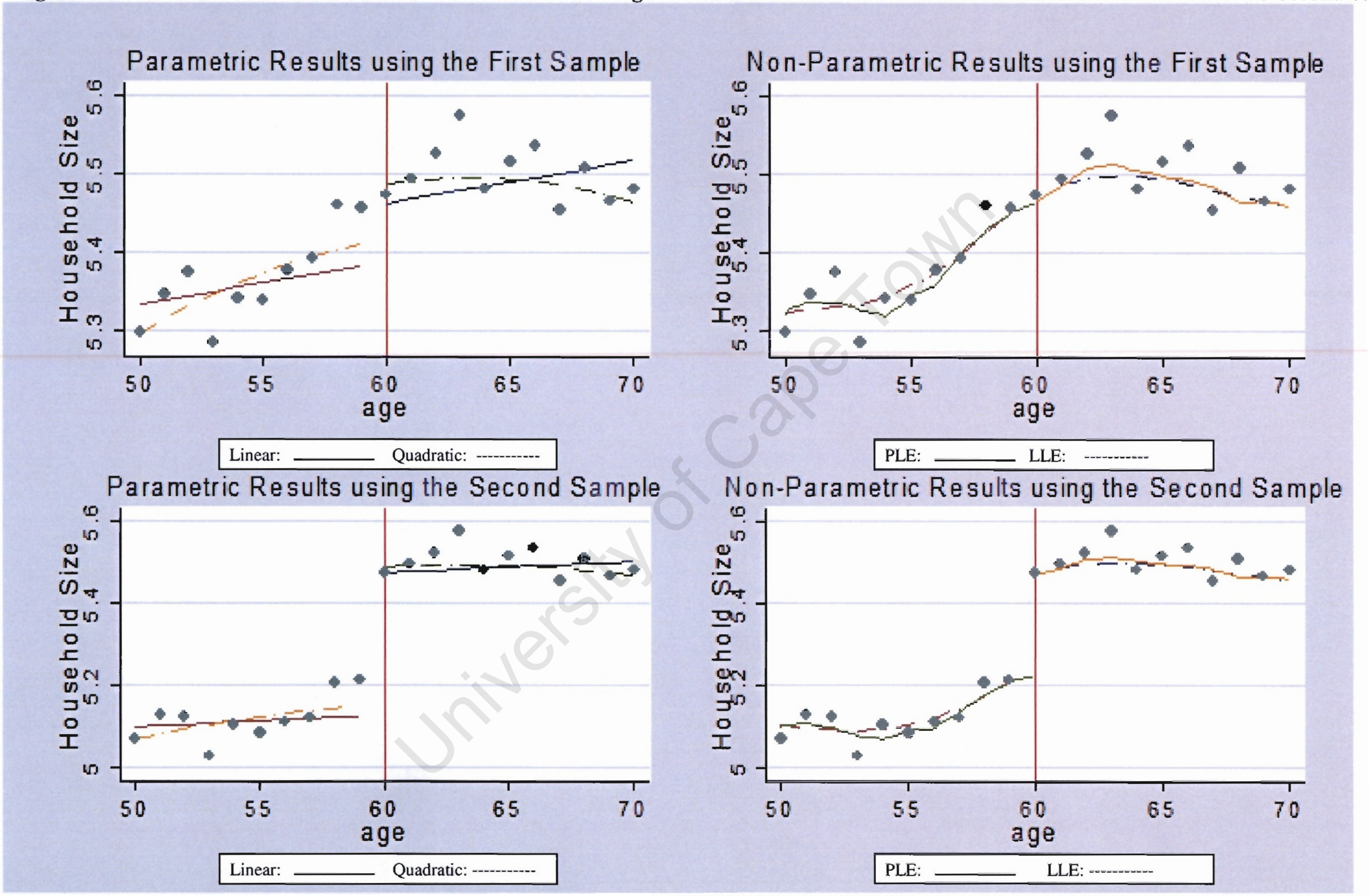
Table 7: Regression Discontinuity Results for the Second Sample

	Parametric Approach				Non-parametric Approach			
	Linear		Quadratic		PLE		LLE	
	alpha	% change	alpha	% change	alpha	% change	alpha	% change
Household Size	0.3457*	6.74%	0.3378*	6.56%	0.2465*	4.72%	0.2199*	4.19%
	(10.4545)		(10.1438)		(5.5394)		(3.1528)	
Children 0-5	0.0598*	9.47%	0.0485*	7.31%	0.0381*	5.65%	0.0351*	5.16%
	(5.8902)		(4.7430)		(3.4011)		(2.0015)	
Children 6-17	0.1244*	8.51%	0.1380*	9.71%	0.0737*	5.00%	0.0510	3.43%
	(7.4032)		(8.1558)		(3.8880)		(1.7175)	
Men 18-23	-0.0010	-0.34%	0.0018	0.63%	0.0031	1.10%	0.0065	2.28%
	(-0.1617)		(0.2931)		(0.4983)		(0.6567)	
Women 18-23	-0.0052	-1.61%	-0.0029	-0.92%	-0.0081	-2.48%	-0.0063	-1.95%
	(-0.8190)		(-0.4569)		(-1.2114)		(-0.6084)	
Men 24-29	-0.0088	-3.89%	-0.0113*	-4.81%	-0.0192*	-8.18%	-0.0188*	-8.04%
	(-1.6559)		(-2.1016)		(-3.4819)		(-2.1708)	
Women 24-29	0.0050	1.86%	0.0006	0.22%	0.0015	0.56%	-0.0030	-1.07%
	(0.8647)		(0.1047)		(0.2538)		(-0.3173)	
Men 30-39	-0.0025	-1.13%	-0.0139*	-5.50%	-0.0172*	-6.78%	-0.0198*	-7.66%
	(-0.4765)		(-2.6550)		(-2.9797)		(-2.1905)	
Women 30-39	0.0052	2.14%	-0.0047	-1.75%	-0.0105	-3.80%	-0.0178	-6.23%
	(0.9514)		(-0.8667)		(-1.7302)		(-1.8703)	
Men 40-49	0.0073*	8.40%	0.0159*	25.68%	-0.0011	-1.53%	-0.0032	-4.44%
	(2.2436)		(4.8689)		(-0.3618)		(-0.6729)	
Women 40-49	0.0127*	17.52%	0.0184*	33.08%	0.0126*	21.58%	0.0107*	18.19%
	(4.1673)		(6.0167)		(4.2327)		(2.2854)	

Source: 2001 South African Population and Housing Census

University of Cape Town

Figure 10: Estimates of the Pension Effect on Household Size using the Parametric and Non-Parametric RD Models and the First and Second Sample



Source: 2001 South African Population and Housing Census

University of Cape Town

The first measure of household composition considered was household size. The results from Table 6 for the two parametric approaches suggest that the receipt of the pension will have a significant effect on household composition. However, from examining the graph of household size in Figure 10, it clear that m(.) has been poorly specified in both of the parametric models. Hence, the estimated pension effects using the parametric approaches on the full sample are likely to be inaccurate. The results from the non-parametric approaches in Table 6 confirm this view as they provide little evidence in favour of the existence of a pension effect. Unlike the linear and quadratic functions, the non-parametric estimates of m(.) appear to fit the data adequately.

In contrast to the results found in Table 6, the results on household size from Table 7 provide overwhelming evidence that household size will change owing to pension receipt. This finding is supported by the graphs in Figure 10 which show a noticeable jump in household size at age 60 when the second sample is used. Furthermore, since the parametric and non-parametric estimates of m(.) fit the data well, the estimated pension effects are likely to be accurate. The results from Table 7 suggest that when a household receives the pension, household size will, on average, increase by at least 4%. Expressed differently, approximately 2 in every 10 households will grow in size by one person when the pension is received.

The results from both tables support the view that the pension will encourage more children aged below 6 to move into the pension household. The estimate of the size of the jump in the number of children aged below 6 varies marginally between each model in Table 6 however the size of the discontinuity appears to be no less than 0.0297. Thus, approximately 3 in 100 households will welcome a young child into their home when the pension is received, and the average number of children aged below 6 in a household will rise by just over 4%. The estimated size of the jump in children aged below 6 is larger using the second sample. According to the results from Table 7, the average number of children aged below 5 will rise by at least 5% owing to pension receipt. Thus, it is possible that the estimated pension effect on the number of children below the age of 5 in household was slightly masked because some of the pension households had not been

University of Cape Town

excluded from the control group. Alternatively, the number of children aged 0-5 that join a household when the household receives a new pension could diminish as the number of existing household members receiving the pension increases. Finally, it is worth noting that from Figures 11-14, it appears as if the linear function estimates m(.) inaccurately. Thus, the estimates of the pension effect on the number of children aged 0-5 using the first parametric approach should be interpreted with caution.

The results from the RD analyses based on the full sample fail to provide sufficient evidence that the pension is associated with a change in the number of children aged 6-17 in a household. Although the estimates of the pension effect from the parametric models are significant, the graph of the number of children aged 6-17 in Figure 11 show that m(.) is poorly specified by both the linear and quadratic functions. Therefore, the estimates are likely to be inaccurate. Furthermore, from the plot of census means in Figure 11, it does not appear as if the average number of children aged 6-17 changes significantly between age 59 and 60. Rather, a noticeable change in the average number of children aged 6-17 is seen to occur between ages 60 and 61. This may mean that the pension has a delayed effect on the number of children aged 6-17 who join the household.

When run on the second sample, the RD analysis produces results that suggest that the number of children between the ages of 6 and 17 in a household will increase when an elder in the household first receives the pension. According to these results, the average number of children aged between 6 and 17 in a household will rise by at least 3.43%. Once again, a closer look at the graphs from the parametric models in Figure 13 reveals that the estimate of the pension effect on the number of children aged 6-17 using the parametric models is likely to overestimate the actual effect. Thus, the estimates from the non-parametric models are likely to be more accurate. The graphs of the census means also show that, while there is an obvious jump in the number of children aged 6-17 between age 59 and 60, there is an even larger jump in the number of children aged 6-17 between ages 60 and 61. Therefore, the results from the models of the number of children aged 6-17 provide two important findings: first, the effect of the pension on the number of children aged 6-17 that join a household diminishes as the number of pensioners in a

University of Cape Town

household increases. Second, there is evidence to suggest the pension may have a partly delayed effect on the number of children aged 6-17 that join the pension household.

In accordance with the cohort results, the RD results also provide strong evidence that the number of men between the ages of 24 and 29 will be affected by the receipt of the pension. According to results from Table 6, the average drop in the number of men between the ages of 24 and 29 on pension receipt will range between 0.0107 and 0.0176. This translates into a percentage drop of 4.70% and 7.53% respectively. The sizes of the jumps and their significance differ marginally when the results are re-run on the second sample. A maximum drop of 0.0192 or 8.18% is predicted using this sample. The graphs in Figure 11-14 show that the m(.) is adequately specified in each of the RD models and that consequently, each of the estimates are likely to accurate. The graph also illustrates how the number of men between the ages of 24 and 29 in the households of older women declines as the women age from 50 to 70. Thus, the receipt of the pension helps in accelerating the downward trend.

The number of men between the ages of 30 and 39 is also shown to be affected by the receipt of the pension. From examining the graphs of the number of men aged 30-39 in Figures 11 and 13, it seems that the estimate of the pension effect calculated using the first parametric approach (with the linear specification for m(.)) may result in the true pension effect being underestimated. The results from the remaining 3 models suggest that number of men between the ages of 30 and 39 will drop on pension receipt, and that the size of this drop will not diminish in households where more pensioners are present. The number of men between the ages of 30 and 39 living with older women is predicted to drop by between 0.0139 and 0.0212 or between 5.50% or 8.17% respectively.

Finally, the results suggest that the number of women between the ages of 40 and 49 in a household will rise in response to pension receipt. The number of women between the ages of 40 and 49 in a household will rise, on average, by no less than 14% according the results found from the full sample, and no less than 18% according the results based on the reduced sample. Apart from the linear specification of m(.), all other specifications

University of Cape Town

for m(.) appear to fit the data well. Thus, it is surprising that the estimate of the pension effect using the second parametric approach is considerably larger than that from the non-parametric approaches. The pension effects estimated using the full sample are, on average, smaller than those estimated using the second sample. Thus, it is possible that the true pension effect was confounded by the presence of pension households in the control group in the first sample. Alternatively, the effect of the pension on the number of women between the ages of 40 and 49 in the pension household could diminish as the number of existing household members receiving the pension increases.

The results based on the third and fourth samples are now considered. Since some African women may adjust their living arrangements slightly prior to or post pension receipt, the estimates of the pension effect presented in Table 6 and Table 7 may be slightly attenuated. Using the third and fourth samples, the pension effect is re-estimated as the jump in the household composition of African women that occurs between ages 58 and 62. Thus, the pension effects estimated using the third and fourth samples should pick up those changes in living arrangements made slightly prior to or post pension receipt as well as those that occur at age 60. Table 8 and Table 9 present the results. Figures 15 - 18 in the appendix illustrate these results graphically.

University of Cape Town

Table 8: Regression Discontinuity Results for the Third Sample

	Parametric Approach				Non-parametric Approach	
	Linear		Quadratic		LLE	
	alpha	% change	alpha	% change	alpha	% change
Household Size	0.0973*	1.81%	0.1125*	2.09%	0.0805	1.48%
	(2.1513)		(2.4672)		(0.9129)	
Children 0-5	0.0110	1.68%	0.0261	3.85%	0.0319	4.65%
	(0.8003)		(1.8881)		(1.4409)	
Children 6-17	0.1423*	9.62%	0.1241*	8.38%	0.0957*	6.49%
	(6.2552)		(5.4068)		(2.5244)	
Men 18-23	-0.0077	-2.49%	-0.0113	-3.70%	-0.0298*	-9.73%
	(-0.9220)		(-1.3321)		(-2.4033)	
Women 18-23	-0.0093	-2.74%	-0.0131	-3.94%	-0.0317*	-9.37%
	(-1.0723)		(-1.5034)		(-2.4591)	
Men 24-29	-0.0058	-2.47%	-0.0014	-0.58%	-0.0124	-5.19%
	(-0.7972)		(-0.1909)		(-1.1609)	
Women 24-29	-0.0076	-2.73%	-0.0003	-0.09%	-0.0087	-3.06%
	(-0.9815)		(-0.0321)		(-0.7435)	
Men 30-39	-0.0224*	-10.50%	-0.0074	-3.12%	0.0095	3.89%
	(-3.2321)		(-1.0613)		(-0.8153)	
Women 30-39	-0.0162*	-6.96%	-0.0033	-1.29%	0.0299*	11.42%
	(-2.2485)		(-0.4493)		(2.4564)	
Men 40-49	0.0414*	57.06%	0.0301*	55.17%	0.0266*	42.65%
	(9.2926)		(6.7124)		(3.8119)	
Women 40-49	0.0324*	52.53%	0.0246*	49.92%	0.0378*	77.82%
	(7.7855)		(5.8558)		(5.4773)	

Source: 2001 South African Population and Housing Census

Table 9: Regression Discontinuity Results for the Fourth Sample

	Parametric Approach				Non-parametric Approach	
	Linear		Quadratic		LLE	
	alpha	% change	alpha	% change	alpha	% change
Household Size	0.3923*	7.69%	0.3962*	7.74%	0.3461*	6.69%
	(8.4115)		(8.4845)		(3.8286)	
Children 0-5	0.0441*	7.01%	0.0494*	7.55%	0.0509*	7.63%
	(3.0867)		(3.4566)		(2.1975)	
Children 6-17	0.1880*	13.12%	0.1808*	12.62%	0.1499*	10.54%
	(7.9112)		(7.6032)		(3.7951)	
Men 18-23	-0.0007	-0.23%	-0.0022	-0.74%	-0.0231	-7.69%
	(-0.0788)		(-0.2519)		(-1.7759)	
Women 18-23	-0.0068	-2.02%	-0.0082	-2.50%	-0.0293*	-8.73%
	(-0.7464)		(-0.9043)		(-2.1710)	
Men 24-29	-0.0003	-0.13%	0.0013	0.53%	-0.0100	-4.22%
	(-0.0401)		(0.1659)		(-0.8922)	
Women 24-29	0.0034	1.23%	0.0058	2.02%	-0.0038	-1.38%
	(0.4163)		(0.7106)		(-0.3155)	
Men 30-39	-0.0142	-6.88%	-0.0083	-3.51%	0.0121	5.03%
	(-1.9550)		(-1.1393)		(0.9955)	
Women 30-39	-0.0047	-2.10%	0.0003	0.12%	0.0344*	13.35%
	(-0.6208)		(0.0387)		(2.6965)	
Men 40-49	0.0369*	51.24%	0.03278*	63.43%	0.0293*	49.14%
	(7.8658)		(6.9880)		(4.0023)	
Women 40-49	0.0290*	49.51%	0.0263*	58.02%	0.0401*	86.44%
	(6.6733)		(6.0541)		(5.5181)	

Source: 2001 South African Population and Housing Census

University of Cape Town

The estimates of the pension effect on household size in Table 8 and Table 9 are notably larger than those in Table 6 and Table 7 respectively. Thus, the results for the pension effect on household size based on the third and fourth sample suggest that a number of households may adjust their living arrangements in response to the pension either slightly prior to or post age 60. According to the results from the LLE model in Table 8, household size will rise by 1.48% between ages 58 and 62 owing to the receipt of the pension. Since the linear and quadratic models for household size based on the third sample do not appear to fit the data well, these estimates are ignored. When applied to the fourth dataset, the RD models estimate a jump in household size of between 6.69% and 7.74%. Based on the second sample, the estimated jump in household size only ranged between 4.19% and 6.74%.

The results from applying the RD models to the third and fourth sample do not suggest that the jump in the number of children aged below 6 between ages 58 and 62 is markedly different from the jump in the number of children aged below 6 at age 60. In Table 6, the estimated pension effect on the number of children aged below 6 ranged between 4.34% and 5.15% (the results based on the linear model are ignored because, as discussed earlier, they are unlikely to be accurate), while in Table 8, the LLE model predicts a jump of 4.65%. Again, the linear and quadratic models do not appear to fit the data well and therefore are ignored. The estimated pension effect on the number of children aged below 6 in Table 7 ranged between 5.16% and 7.31%, which was only slightly smaller than range of 7.01% and 7.63% estimated in Table 9. Thus, the results based on the third and fourth sample suggest that, if children below 6 are sent to live in pension households, they will be sent as soon as this pension is received, and no sooner or later.

The results in Table 8 and Table 9 provide strong evidence that the jump in the number of children aged 6-18 between ages 58 and 62 is significant. Furthermore, the results give weight to the earlier assertion that the pension effect on the number of children aged 6-18 is delayed. The jump in the number of children aged 6-18 between ages 58 and 62 is far greater than the jump in the number of children aged 6-18 at age 60. The pension effect

University of Cape Town

on the number of children aged 6-18 estimated using the first and second sample was expected to be no greater than 2% and 5% respectively. However, when based on the third and fourth samples, this effect was estimated to be at least 6.49% and 10.54% respectively. The graphs of the census means also show that, while there is an obvious jump in the number of children aged 6-17 between age 59 and 60, there is an even larger jump in the number of children aged 6-17 between ages 60 and 61.

The results from the LLE model in Table 8 and Table 9 suggest that the number of men and women between the ages of 18 and 23 living in a household may be affected if the household receives the pension. However, since the parametric models provide no evidence of this effect, the above result is given little weight. The same can be said for the results for the pension effect on the number of women aged 30-39; while there is some evidence to suggest that the number of men aged 30-39 will change significantly between ages 58 and 62, this evidence is not strong enough. In this case, the LLE model finds that a significant pension effect exists, while the quadratic RD model does not (the linear RD model appears to fit the data poorly, and is therefore ignored).

In contrast to the results from Table 6 and Table 7, the results from Table 8 and Table 9 do not provide overwhelming evidence that the pension will affect the number of men aged 30-39 living with an African female pensioner. Only the results from applying the linear RD model to the third sample suggested that the number of men aged 30-39 may be affected by the receipt of the SOAP. The census means in Figures 11-14 provide some explanation as to why the jump in the number of men aged 30-39 is significant at age 60 but not between ages 58 and 62. They show that the average number of men aged 30-39 drops at age 60 and 61, however rises again at age 62. Thus, since the averages at age 60 and 61 are not included in the estimation of the pension effect on the number of men aged 30-39 using the third and fourth sample, these estimates will fail to find evidence of a pension effect. Thus, the results for the number of men aged 30-39 from the third and fourth sample do not suggest that a pension effect on the number of men aged 30-39 does not exist. Rather they show that the pension effect on the number of

University of Cape Town

men aged 30-39 move in the opposite direction to the age trend in the number of men aged 30-39 living with older African women.

Finally, the results from applying the RD models to the third and fourth sample imply a notable increase in the pension effect on the number of men and women aged 40-49. First, the results from Table 6 and Table 7 suggest that the size of the pension effect on the number of women aged 40-49 will be at most 33%. In contrast, the results from Table 8 and Table 9 suggest the size of this effect will be no less than 50%. Second, the results from Table 6 and Table 7 indicate that there is insufficient evidence to conclude that the pension will have an effect on the number of men aged 40-49 in a household. On the other hand, the results from Table 8 and Table 9 provide overwhelming evidence in favour of a positive pension effect on the number of men aged 40-49 in the household of a female pensioner. It is worth noting, however, that the pension effects on the number of men and women aged 40-49 estimated using the third and fourth samples may be slightly inflated owing to the fact that the age trend in the number of men and women aged 40-49 living with elderly African women has not been accounted for. As African women age from 58 to 62, the number of men and women aged 40-49 living with them appears to rise. This can be seen from the census means illustrated in Figures 11-14. If this number rises partly owing to the effect of age, then the pension effects presented in Table 8 and Table 9 may be slightly exaggerated.

In conclusion, the results from the regression discontinuity analysis show that household composition *will* be affected by the receipt of the pension. On average, household size will rise considerably on pension receipt. Furthermore, changes in household size in response to the pension may also occur slightly prior to or post age 60. The number of children in a household will also rise when a household starts receiving SOAP. However, the effect of the pension on the number of children aged 6-18 will be partly delayed. The results also show that the number of men between the ages of 24 and 29, and the number of men between the ages of 30 and 39 will, on average, drop when SOAP is received. Finally, the number of men and women between the ages of 40 and 49 in a household will rise owing to pension receipt. This effect may be slightly premature or delayed.

University of Cape Town

5.3 Comparison of Findings

The results from both the cohort and the regression discontinuity approach imply that the pension is likely to affect household composition. In particular, the results from the cohort analysis suggest that:

- the number of children aged 6 to 17 will rise owing to pension receipt
- the number of men aged 24 to 29 will drop owing to pension receipt
- the number of women aged 24 to 29 will drop owing to pension receipt

The results from the regression discontinuity analysis suggest that:

- the number of children aged 0 to 5 will rise owing to pension receipt
- the number of children aged 6 to 17 will rise owing to pension receipt
- the number of men aged 24 to 29 will drop owing to pension receipt
- the number of men aged between 30 and 39 will drop owing to pension receipt
- the number of men aged between 40 and 49 will rise owing to pension receipt
- the number of women aged between 40 and 49 will rise owing to pension receipt

For clarity, Table 10 below draws a comparison between the results from the cohort and regression discontinuity analyses. Only those measures of household composition associated with a pension effect that was significant in either one of the approaches were included in this table.

University of Cape Town

Table 10: Comparison of the Cohort and Regression Discontinuity Results

		Cohort	First Sample				Second Sample				Third Sample			Fourth Sample		
			linear	quadratic	PLE	LLE	linear	quadratic	PLE	LLE	linear	quadratic	LLE	linear	quadratic	LLE
Household size	alpha	0.14	0.07	0.07	0.00	-0.01	0.35	0.34	0.25	0.22	0.10	0.11	0.08	0.39	0.40	0.35
	% change	3%	1%	1%	0%	0%	7%	7%	5%	4%	2%	2%	1%	8%	8%	7%
Children 0-5	alpha	0.02	0.03	0.03	0.03	0.04	0.06	0.05	0.04	0.04	0.01	0.03	0.03	0.04	0.05	0.05
	% change	4%	5%	4%	4%	5%	9%	7%	6%	5%	2%	4%	5%	7%	8%	8%
Children 6-17	alpha	0.17	0.08	0.09	0.03	0.02	0.12	0.14	0.07	0.05	0.14	0.12	0.10	0.19	0.18	0.15
	% change	11%	6%	6%	2%	1%	9%	10%	5%	3%	10%	8%	6%	13%	13%	11%
Men 24-29	alpha	-0.03	-0.01	-0.01	-0.02	-0.02	-0.01	-0.01	-0.02	-0.02	-0.01	0.00	-0.01	0.00	0.00	-0.01
	% change	-9%	-5%	-5%	-8%	-7%	-4%	-5%	-8%	-8%	-2%	-1%	-5%	0%	1%	-4%
Women 24-29	alpha	-0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.01	0.00	-0.01	0.00	0.01	0.00
	% change	-10%	2%	0%	1%	-1%	-1%	-1%	0%	-1%	-3%	0%	-3%	1%	2%	-1%
Men 30-39	alpha	-0.02	-0.01	-0.01	-0.02	-0.02	0.00	-0.01	-0.02	-0.02	-0.02	-0.01	0.01	-0.01	-0.01	0.01
	% change	-5%	-5%	-6%	-8%	-8%	-1%	-5%	-7%	-8%	-11%	-3%	4%	-7%	-4%	5%
Men 40-49	alpha	0.02	0.01	0.01	0.00	0.00	0.01	0.02	0.00	0.00	0.04	0.03	0.03	0.04	0.03	0.03
	% change	31%	14%	23%	-2%	-3%	8%	26%	-2%	-4%	57%	55%	43%	51%	63%	49%
Women 40-49	alpha	0.01	0.01	0.02	0.01	0.01	0.01	0.02	0.01	0.01	0.03	0.03	0.04	0.03	0.03	0.04
	% change	26%	19%	27%	17%	14%	18%	33%	22%	18%	53%	50%	78%	50%	58%	86%

Source: OHS for years 1994, 1995, 1997 and 1999, LFS for years 2000, 2002 and 2003, and Population Censuses for 1996 and 2001

University of Cape Town

6. Conclusions and Recommendations

This study aims to estimate the effect of the SOAP on the living arrangements of its African, female recipients. It attempts to determine whether the composition and/or size of an African household will, on average, change when an African woman in a household receives her pension. The study interprets the changes in household composition in terms of the pension's effect on the age and gender of the household members. The changes in household composition may be associated with changes in other characteristics of the household members apart from just their age and gender however such analysis is beyond the scope of this study.

The study uses two different methods to achieve its aim. The first approach is a cohort analysis. The living arrangements of cohorts of elderly African women are tracked over the period 1994-2003 and as the women age into pension eligibility. The strength of this analysis lies in the fact that, if the sample cohorts accurately represent the same underlying population cohorts from year to year, the cohort data should form a pseudo panel. Therefore, significant changes in the household composition of the cohorts at age-eligibility should provide an accurate estimate of the pension effect on household composition. There are, however, two weaknesses in this approach. First, although the sample cohorts have been constructed using nationally representative data, because the size of some of the older cohorts is very small, there is a danger that these cohorts may not be nationally representative. Secondly, since only 9 years of data were used in the construction of the cohort data, the size of the pseudo panel may be too small to produce accurate results. Given these weaknesses, the results from the cohort analysis should be seen as suggestive rather than definitive.

Therefore, the study goes onto use a regression discontinuity methodology. Under this method, it is assumed that pension effects can be identified as any discontinuous jumps in household composition at age-eligibility. Thus, all other changes in household composition around age eligibility are assumed to occur smoothly according to some function m(.). The regression discontinuity approach used in this study is rigorous and

University of Cape Town

employs a rich dataset. However, the results rely heavily on the manner in which m(.) is measured. Furthermore, the analysis only employs data from 2001. Thus, it may be argued that the results from this part of the study will not generalize to other years as easily as those from the cohort analysis.

Prior to this analysis, only one study is known to have considered the effect of SOAP on household composition, where household composition has been defined in terms of the age and gender of the household members. This study, conducted by Edmonds et al (2005), showed that (i) the number of children below the age of 5 will rise (ii) the number of women between ages 18 and 23 will rise, and (iii) the number of women between ages 30 and 39 will drop on pension receipt. The effects were measured using a Regression Discontinuity Design that used Partial Linear Estimation and using census data from 1996.

The results from this study also support the view that the pension is likely to affect household composition. The results from the cohort analysis are based on data from 1994-2003. They suggest that:

- the number of men aged 24 to 29 will drop owing to pension receipt
- the number of men aged between 40 and 49 will rise owing to pension receipt
- the number of women aged between 40 and 49 will rise owing to pension receipt

The regression discontinuity analysis employs data from 2001. The results from the analysis imply that:

- the number of children aged 0 to 5 will rise owing to pension receipt
- the number of children aged 6 to 17 will rise owing to pension receipt
- the number of men aged 24 to 29 will drop owing to pension receipt
- the number of men aged between 30 and 39 will drop owing to pension receipt
- the number of men aged between 40 and 49 will rise owing to pension receipt
- the number of women aged between 40 and 49 will rise owing to pension receipt

Thus, in keeping with the findings of Edmonds et al (2005), the results from this study show that the pension will be associated with a rise in the number of children in a

University of Cape Town

household. As argued by Edmonds et al (2005), parents may send their children to live with their grandmothers when their grandmothers receive the pension to allow the parents more time to work. Owing to the arrival of the pension, grandparents will have the financial means and, most likely, the time to look after their grandchildren.

The results from this study also show that the pension is associated with a drop in the number of prime-age men aged of 24-39 and prime-age women aged 24-29 in a household. It is plausible that these individuals will leave the household for similar reasons as those outlined Edmonds et al (2005). That is, they may leave when the pension is first received because the pension provides them with enough initial capital to do so. Alternatively, these individuals may leave in response to the arrival of other individuals on the receipt of the pension, such as the arrival of children.

In contrast to past studies, this study finds strong evidence that women aged 40 to 49 will move into the pension household when the pension is received. It is plausible that these women are the daughters of the pensioner, and arrive in order to look after their mothers. Alternatively, the women may be the wives of the sons of the grandmothers who have be sent to look after their husband's parents. Presumably, these women wait until the pension arrives to move in so that there will be enough capital to support them. There is also some evidence to suggest that the men aged 40 to 49 will move into the pension household when the pension is received. These men are likely to be the husbands of the women aged 40 to 49 who move into the pension household on pension receipt.

The actual sizes of the estimates are deemed unreliable as they differ quite considerably between the two approaches and even within the approaches. Thus, these results might be best used as an indication of the existence and direction of the pension effects rather than a measure of the size of the pension effects. The size of the estimates of the pension effects are also likely to be smaller than the true pension effects because age eligibility is used as the independent variable and not actual pension receipt. The results from the sensitivity analysis presented in section 5.2 also suggest that the initial RD estimates and

University of Cape Town

the cohort estimates are slightly attenuated owing to the fact that they fail to account for pension effects that are premature or delayed.

It is also important to note that the above findings rely on the assumption that it is the other family or household members who move in response to the pension and not the pensioners themselves. This assumption was tested by Edmonds et al (2005) and they found no evidence to suggest that female pensioners will move on pension receipt. However, this study could certainly be improved upon by re-testing this assumption.

A final recommendation to future researchers on this topic would be to consider the pension effect on household composition in terms of its effect on other characteristics of household members apart from age and gender. Not only does such knowledge provide researchers with greater insight into precisely which members are leaving and why, but it also allows researchers to interpret behavioural responses to SOAP more accurately. The example of a study that examines the effect of the pension on the number of leisure hours of prime-age individuals illustrates this concept clearly. If this study finds that the number of leisure hours of prime-age adults is positively related to the receipt of the pension (as Bertrand et al (2003) did), then, without any other information, such a study would most likely conclude that the pension creates a disincentive for prime-age individuals to look for work. However, if one knew that the pension was associated with a drop in the number of prime-age males in the pension household (as this study suggests), then one may be sceptical to conclude that the pension creates a disincentive to work, because the rise in leisure hours may be related to the drop in the number of prime-age men. Knowledge of how employable the men who leave the pension household are (perhaps measured using education or work experience), would however, provide more concrete evidence as to whether the rise in leisure hours is related to the drop in the number of prime-age men in the household or a drop in the willingness to look for work.

In conclusion, this study finds that the living arrangements of African women *will* change when African women receive the South African State Old Age Pension. It finds that the household composition of African women will change when they receive the pension,

University of Cape Town

which will result in a rise in their household size. This paper considers which individuals are leaving and why, but an important step would be to consider how the changes to living arrangements affect the well-being of the existing and the new household members. Finally, it is hoped that the results of this study will provide valuable insight into the behavioural effects of SOAP and pave the way for more accurate estimates of SOAP effects in future research.

University of Cape Town

REFERENCES

ABLEIDINGER, J, CASE, A., & PAXSON, C. 2004. Orphans in Africa: Parental Death, Poverty and School Enrollment. *Demography,* 41(3): 483-508.

ALDERMAN, H. 1999. Safety Nets and Income Transfers in South Africa. World Bank Africa Region Discussion Paper 19335 Washington D.C.: World Bank.

ANGRIST, J., & LAVY, V. 1999. Using Maimonides' Rule to Estimate the Effect of Class Size on Stochastic Achievement. *Quarterly Journal of Economics,* 114: 553-576, May.

BARCLAY, J.D., GEE, L., DAVIS, M. A., MORITZ, D. J., & NEUHAUS, J.M. 1997. Living Arrangements, changes in living arrangements, and survival among community dwelling older adults. *American Journal of Public Health,* 87(3): 371-77.

BALTAGI, B.H. 2001. *Econometric Analysis of Panel Data*. New York: John Wiley.

BERTRAND, M., MULLAINATHAN, S., & MILLER, D. 2003. Public Policy and Extended Families: Evidence from Pensions in South Africa. *World Bank Economic Review*, 17(1): 27-50.

BURCH, T.K., & MATTHEWS, B.J. 1987. Household Formation in Developed Societies. *Population and Development Review*, 13(3): 495-511.

CASE, A., & DEATON, A. 1998. Large Cash Transfers to the Elderly in South Africa. *Economic Journal*, 108(450): 1330-61.

CASE, A., LIN, I. F., & McLANAHAN, S. 2000. How Hungry is the Selfish Gene? *The Economic Journal,* 110(466): 781-804.

CASALE, D., & POSEL, D. 2003. What has been happening to Internal Labour Migration in South Africa, 1993-1999? *DPRU Working Paper* No 03/74.

DEATON, Angus. 1997. *Analysis of Household Surveys: A Microeconometric Approach to Development Policy.* Baltimore: The Johns Hopkins University Press.

University of Cape Town

DUFLO, Esther. 2003. Grandmothers and Granddaughters: Old Age Pension and Intra-household allocation in South Africa. *World Bank Economic Review,* 17(1): 1-25.

EDMONDS, Eric. 2005. Child Labor and Schooling Responses to Anticipated Income in South Africa. *Journal of Development Economics, forthcoming.*

EDMONDS, E., MAMMEN, K., & MILLER, D. 2005. Rearranging the Family? Household Composition Responses to Large Pension Receipts. *Journal of Human Resources,* 40(1):186-207.

ENGELHARDT, G. V., GRUBER, J., & PERRY, C. D. 2002. Social Security and Elderly Living Arrangements. *NBER Working Paper* No 8911.

FAN, J. 1992. Design adaptive non-parametric regression. *Journal of the American Statistical Association,* 87(420): 998-1004.

FAN, J., & GIJBELS, I. 1996. *Local Polynomial Modelling and its Applications.* London: Chapman and Hall.

JENSEN, R. T. 2004. Do Private Transfers 'Displace' the Benefits of Public Transfers? Evidence from South Africa. *Journal of Public Economics,* 88(1-2): 89-112.

HAHN, J., TODD, P., & VAN DER KLAAUW, W. 2001. Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica*, 69(1): 201-209.

HAMOUDI, A., & THOMAS, D. 2005. Pension Income and the Well-being of Children and Grandchildren: New evidence from South Africa. *California Centre for Population Research On-line Working Paper* No 043-05.

HUGHES, M. E., & WAITE, L. J. 1991. Health in household context: living arrangements and health in late middle age. *Journal of Health and Social Behaviour*, 43(1): 1-21

HSIAO, C. 2003. *Analysis of Panel Data*. New York: Cambridge University Press.

KLASEN, S., & WOOLARD, I. 2005. Surviving Unemployment without State Support: Unemployment and Household Formation in South Africa. *CSSR Working Paper* No 129, October.

University of Cape Town

KOBRIN, F.E. 1976. The Fall of Household Size and the Rise of the Primary Individual in the United States. *Demography*, 13(1): 127-138.

LAM, D., LEIBBRANDT, M., & RANCHHOD, V. 2005. Labour Force Withdrawal of the Elderly in South Africa. *DPRU Working Paper* No 05/118, June.

LEMIEUX, T., & MILLIGAN, K. 2004. Incentive Effects of Social Assistance: A Regression Discontinuity Approach. *NBER Working Paper* No 10541, June.

MAITRA, P., & RAY, R. 2003. The Effects of Transfers on Household Expenditure Patterns and Poverty in South Africa. *Journal of Development Economics,* 71(1): 23-49.

McCRARY, J., & ROYER, H. 2003. *Does Maternal Education Affect Infant Health? A Regression Discontinuity Approach Based on School Age Entry Laws.* [On-line]. Available: http://www-personal.umich.edu/~jmccrary/mccrary_and_royer2003.pdf [16 July 2006]

MØLLER, V., & DEVEY, R. 2003. Trends in the Living Conditions and Satisfaction Among Poorer Older South Africans: Objective and Subjective Indicators of Quality of Life in the October Household Survey. *Development Southern Africa*, 20(4): 457-476.

MØLLER, V., & FERREIRA, M. 2003. Getting by…Benefits of Non-contributory Pension Income for Older South African Households. *Report, Institute of Ageing in Africa, University of Cape Town.*

MOULTRIE, T.A., & TIMÆS, I. M. 2001. Fertility and Living Arrangements in South Africa. *Journal of Southern African Studies*, 27(2): 207-223.

PORTER, J. 2003. Estimation in the Regression Discontinuity Model. *Mimeo, Harvard University.*

POSEL, D., FAIRBURN, J., & LUND, F. 2004. Labour Migration and Households: A Reconsideration of the Effects of the Social Pension on Labour Supply in South Africa. *Paper presented at the National Academy of Sciences Conference on Aging in Africa, Johannesburg.*

University of Cape Town

TROCHIM, W.M. 2002. *The Research Methods Knowledge Base, 2nd Edition.* [Online]. Available: http://www.socialresearchmethods.net/kb/quasird.htm [20 August 2005]

VAN DER BERG, S. 2001. Redistribution through the Budget: Public Expenditure Incidence in South Africa, 1993-1997. *Social Dynamics*, 27(1): 140-164.

VAN DER KLAAUW, W. 2002. Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach. *International Economic Review*, 43(4): 1249-1287.

University of Cape Town

GLOSSARY

Acquired Immune Deficiency Syndrome (AIDS) A collection of infections that occur in individuals' whose immune systems have been severely weakened by the onset of the Human Immunodeficiency Virus (HIV). The AIDS virus is crippling the population in Sub-Saharan Africa, and South Africa in particular.

Age Effects The life-cycle pattern of a particular variable. In this study, age effects represent the life-cycle pattern of the living arrangements of black South African women.

Apartheid A system of racial segregation that existed in South Africa until 1991 and politically, economically, and socially discriminated against non-whites.

Assignment Variable The variable used to determine programme assignment. In this study, programme assignment is based on the *age* of the individuals.

Bandwidth A variable used in non-parametric modelling to estimate the density of a variable x, at a specific point, x_0. The density at point x_0 is estimated by calculating the average density across the range $x_0 \pm h$, where h represents the bandwidth. The larger the bandwidth, the less accurate the estimated distribution of x will be. However, a small bandwidth will produce inaccurate results if a large dataset is not available.

Cohort A group of individuals who are studied as a group because they share a common characteristic, such as age.

Cohort Data Aggregate information that has been collected on cohorts of individuals over time. The cohorts need not comprise of the same individuals over time but they must represent the same underlying population cohorts over time. Also referred to as pseudo-panel.

Cohort Effect The effect that the characteristic defining the cohort has on a particular variable. In this study, the cohort effects would describe the effect of an individual's year of birth on their living arrangements.

Constant Treatment Effects Model A type of Regression Discontinuity Model that assumes that treatment effects are constant across all individuals included in the analysis.

University of Cape Town

Cross-Sectional Data Data that contain information on a population sample at one point in time only.

Cut-Off Value The value of the assignment variable that separates the programme participants from the non-participants. In this study, the cut-off value for the pension is age 60.

Exogenous Variable An explanatory variable whose value is determined by factors outside the regression model but whose value influences the values of other variables in the model.

Fixed Effects Regression A regression technique that removes individual fixed effects from the error term of the model that would otherwise create a bias in the estimates of the parameters, by taking first differences.

Fuzzy design A type of Regression Discontinuity Design in which the rule used to assign individuals to treatment is based on a stochastic function of the cut-off value.

Household Composition see *Living Arrangements*

Kernel A variable used in non-parametric modelling to estimate the density of a variable x, at a specific point, x_0. The density at point x_0 is estimated by calculating the average density across the range $x_0 \pm h$, where h represents a variable called bandwidth. The kernel is used to assign weights to the values of x in this calculation.

Living Arrangements The number of males, females and children living in a household. Also referred to as Household Composition.

Local Linear Estimation A method of estimating the jump in the conditional expectation of a variable at a boundary point that uses weighted linear regression to calculate the limits on either side of the cut-off boundary, and then differences the results.

Nadaraya-Watson Estimation A method of estimating the jump in conditional expectation of a variable at a boundary point that takes the weighted average of observations on either side of the boundary and differences the results.

Non-Parametric Regression A type of regression that allows the regression line to take on a shape that is free of any functional or theoretical constraints. As a result,

University of Cape Town

non-parametric models do not have any meaningful parameters associated with them.

Panel Data Data collected on the same sample population at various points in time.

Partial Linear Estimation A method of estimating the jump in the conditional expectation of a variable at a boundary point that uses a semi-parametric regression model whose parametric component is necessarily linear.

Pension Household A household in which one or more members receive(s) the SOAP.

Pseudo Panel see *Cohort Data*

Regression Discontinuity Design A type of quasi-experimental design in which participants are assigned to treatment according to a defined cut-off score or pre-program measure.

Semi-Parametric Regression A regression model that has both a parametric and non-parametric component.

Sharp Design The type of Regression Discontinuity Design that is used for programs in which program assignment is based on a deterministic function of the cut-off value.

Skip Generation Household A household in which no adults between ages 18 and 49 live, but at least one child below the age of 18 and at least one adult over the age of 49 lives.

South African Child Support Grant A social grant awarded to South African citizens over the age of 18 who are the primary care givers of disadvantaged children under the age of 14. Means tests are used to assess whether applicants are eligible to receive the grant.

South African State Old Age Pension (SOAP) The state pension in South Africa, granted to all female South Africans over the age of 60 and all male South Africans over the age of 65 who pass a certain means test. For an individual, the means test takes into account one's personal income and an income value assigned to assets. For married couples, means are calculated by pooling resources and dividing by two. Relative to average African income, the means test is generous and therefore in practice, most Africans are eligible for this grant.

University of Cape Town

Three Generation Households A household in which at least one child below the age of 18, one adult between the ages of 18 and 49 and one elder aged 50 or older lives.

Variable Treatment Effects Model A type of Regression Discontinuity Model that assumes that treatment or program effects are heterogeneous across individuals.

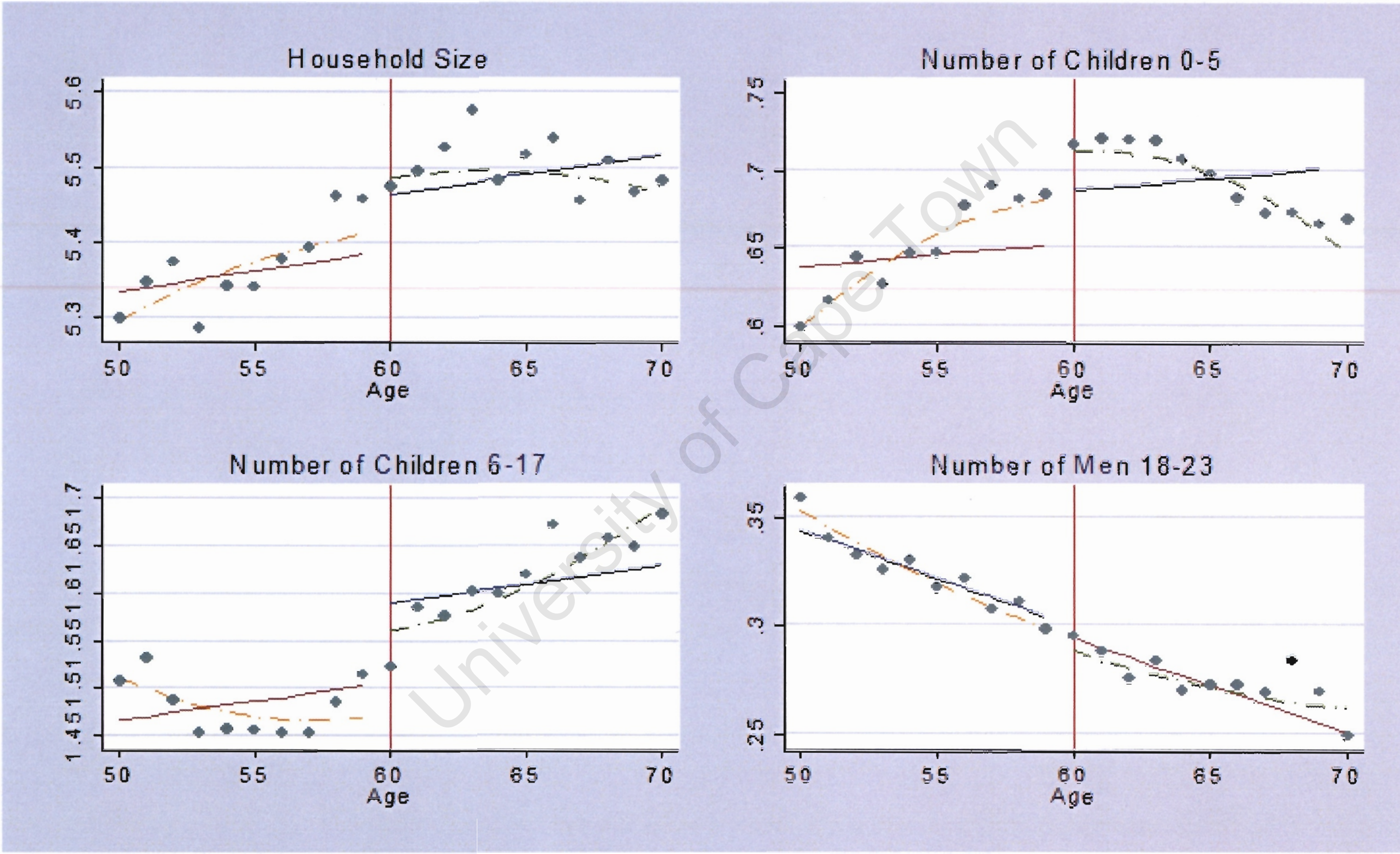
Year Effects The effect of macro-economic changes on a variable from year to year.

University of Cape Town

APPENDIX

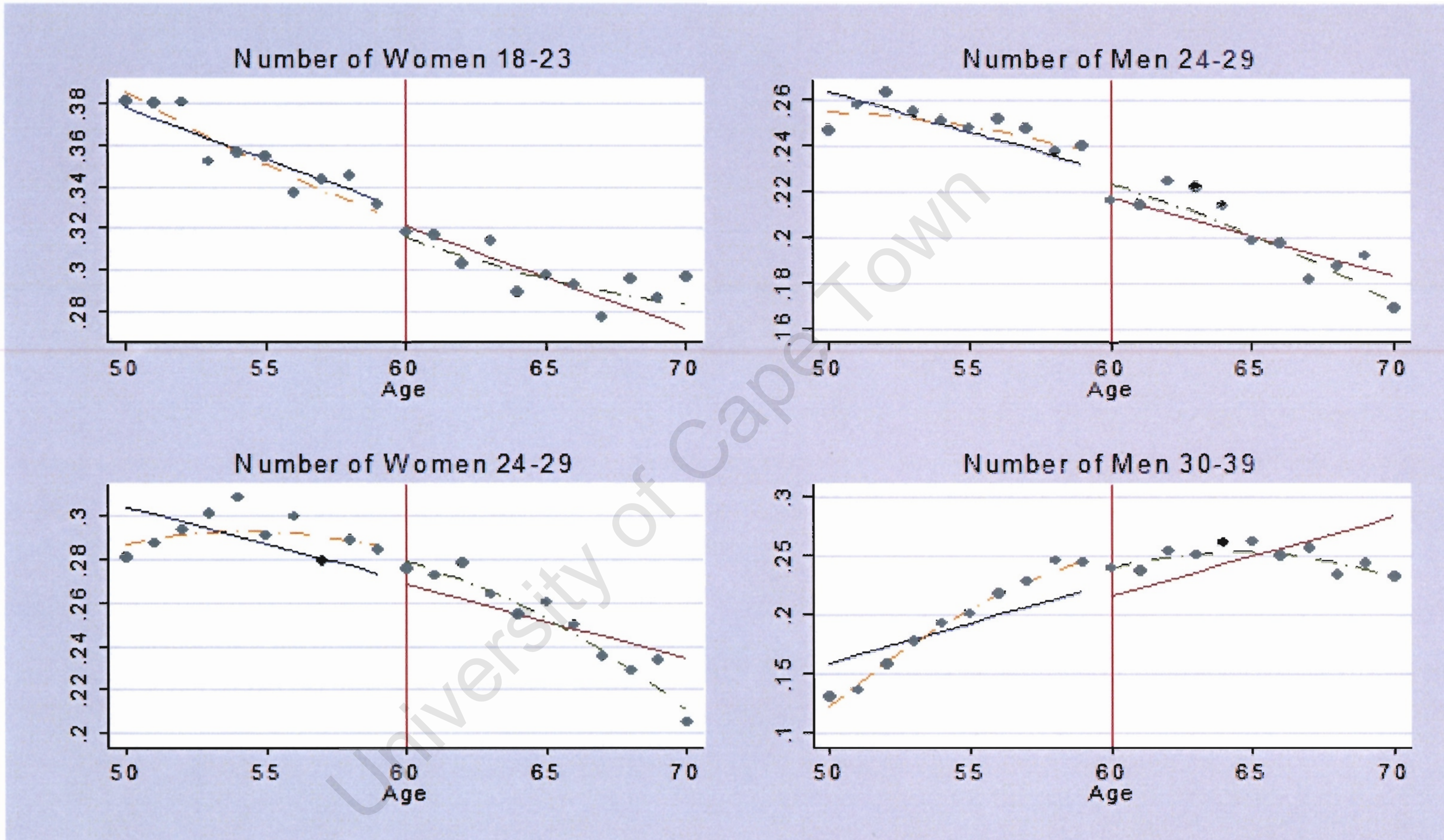
University of Cape Town

Figure 11: Census Means and Parametric Results for the Regression Discontinuity Analysis on the Full Sample of Data



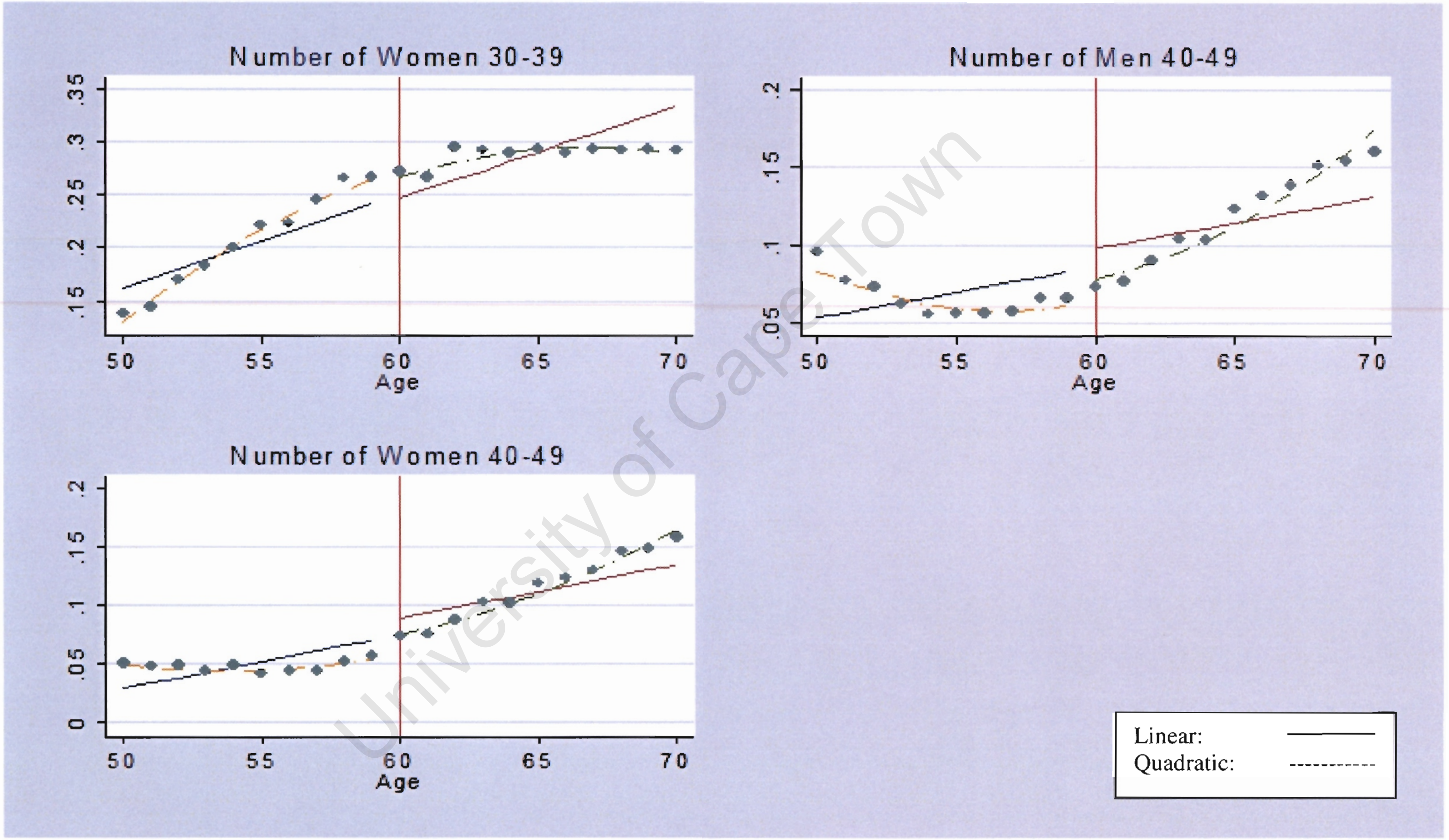
Source: 2001 South African Population and Housing Census

University of Cape Town



Source: 2001 South African Population and Housing Census

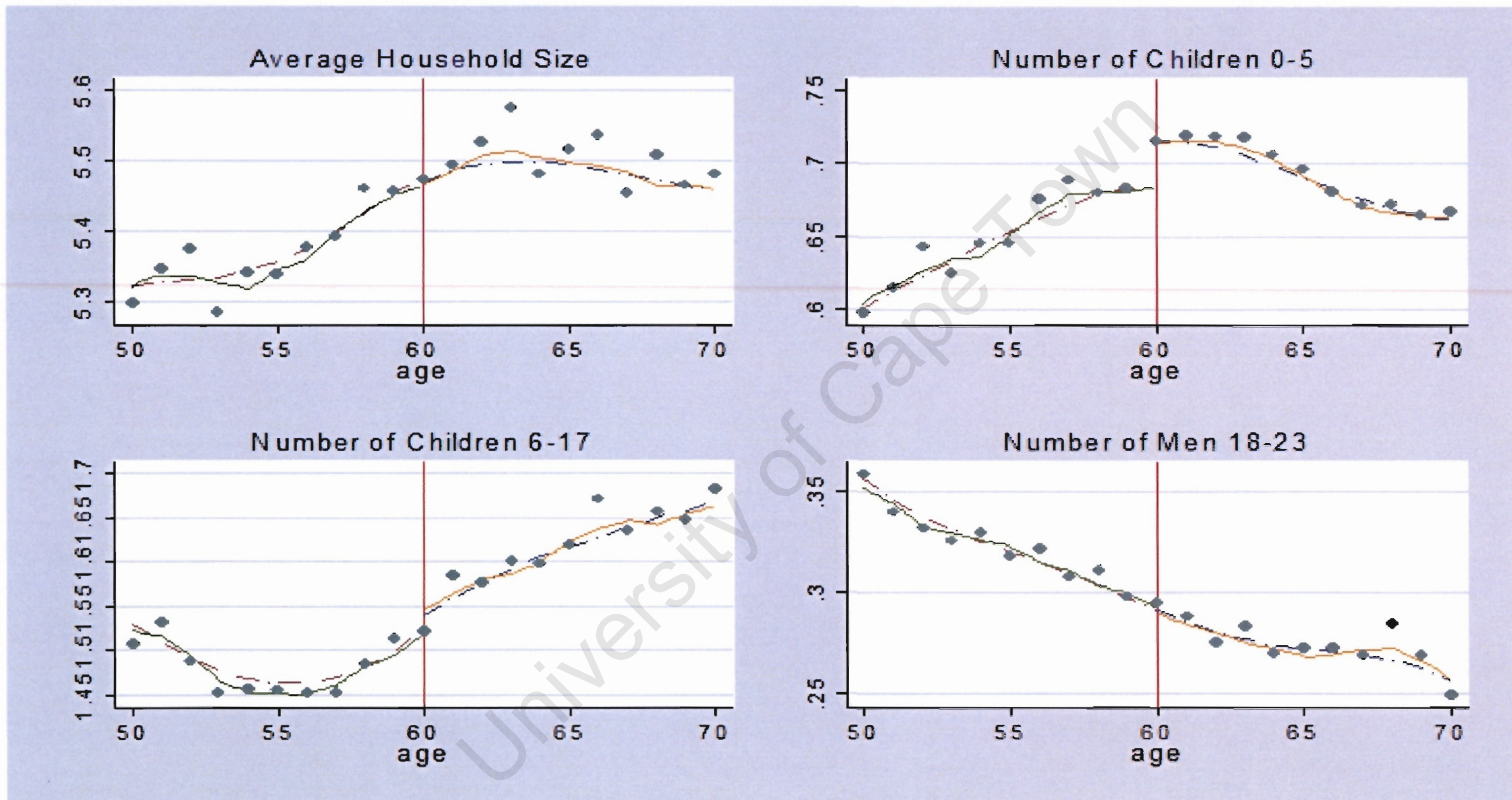
University of Cape Town



Source: 2001 South African Population and Housing Census

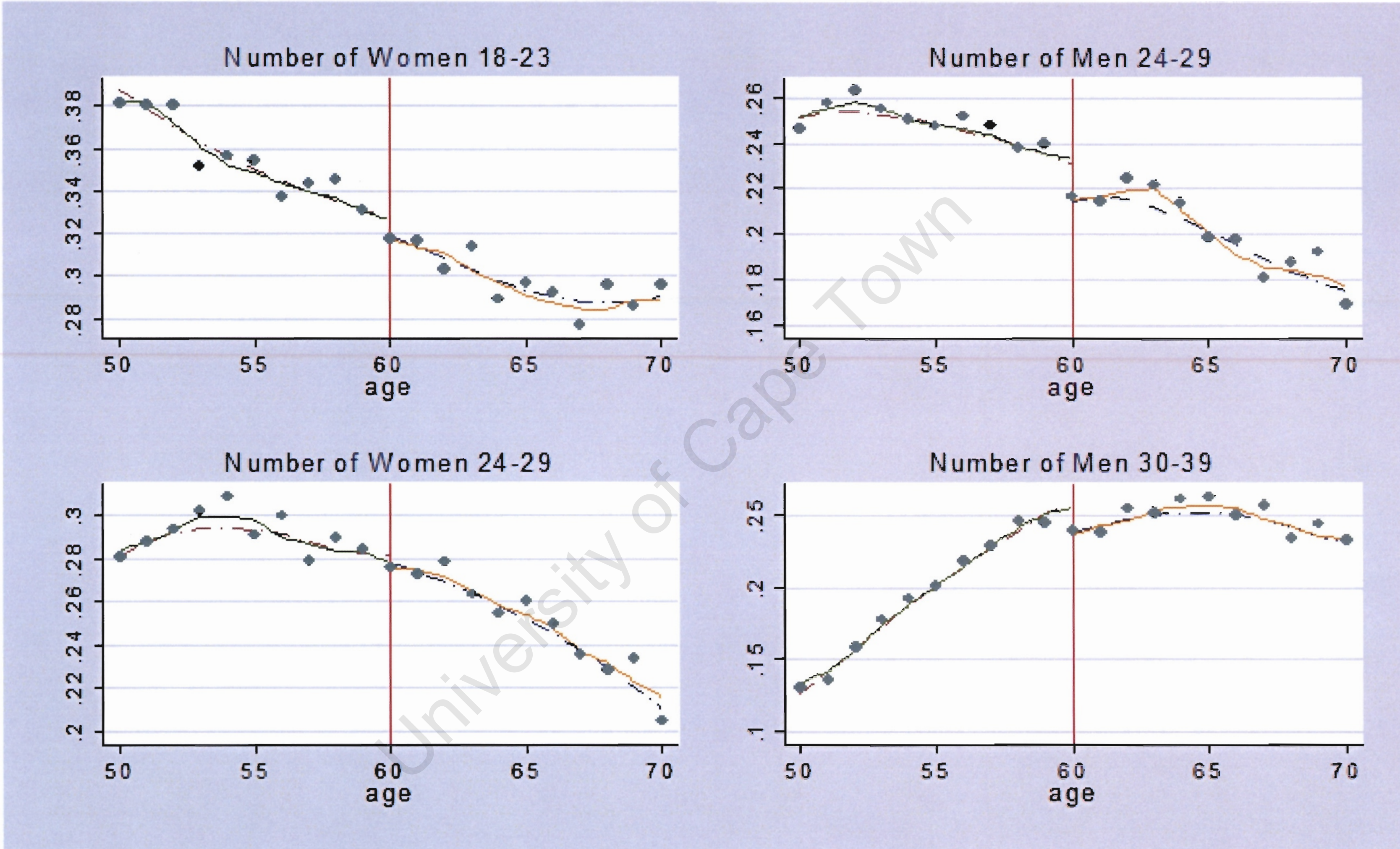
University of Cape Town

Figure 12: Census Means and Non-parametric Results for the Regression Discontinuity Analysis on the Full Sample of Data



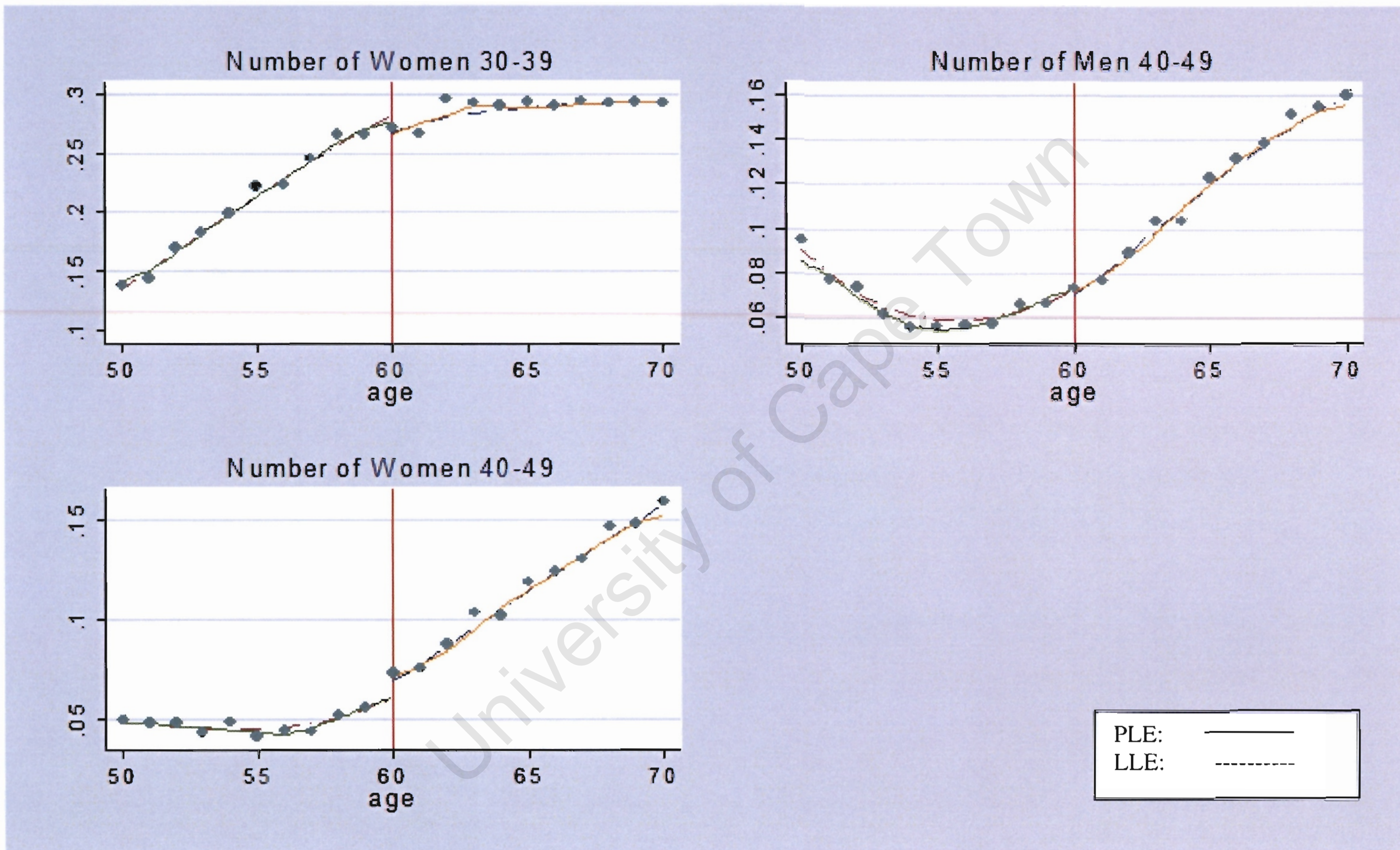
Source: 2001 South African Population and Housing Census

University of Cape Town



Source: 2001 South African Population and Housing Census

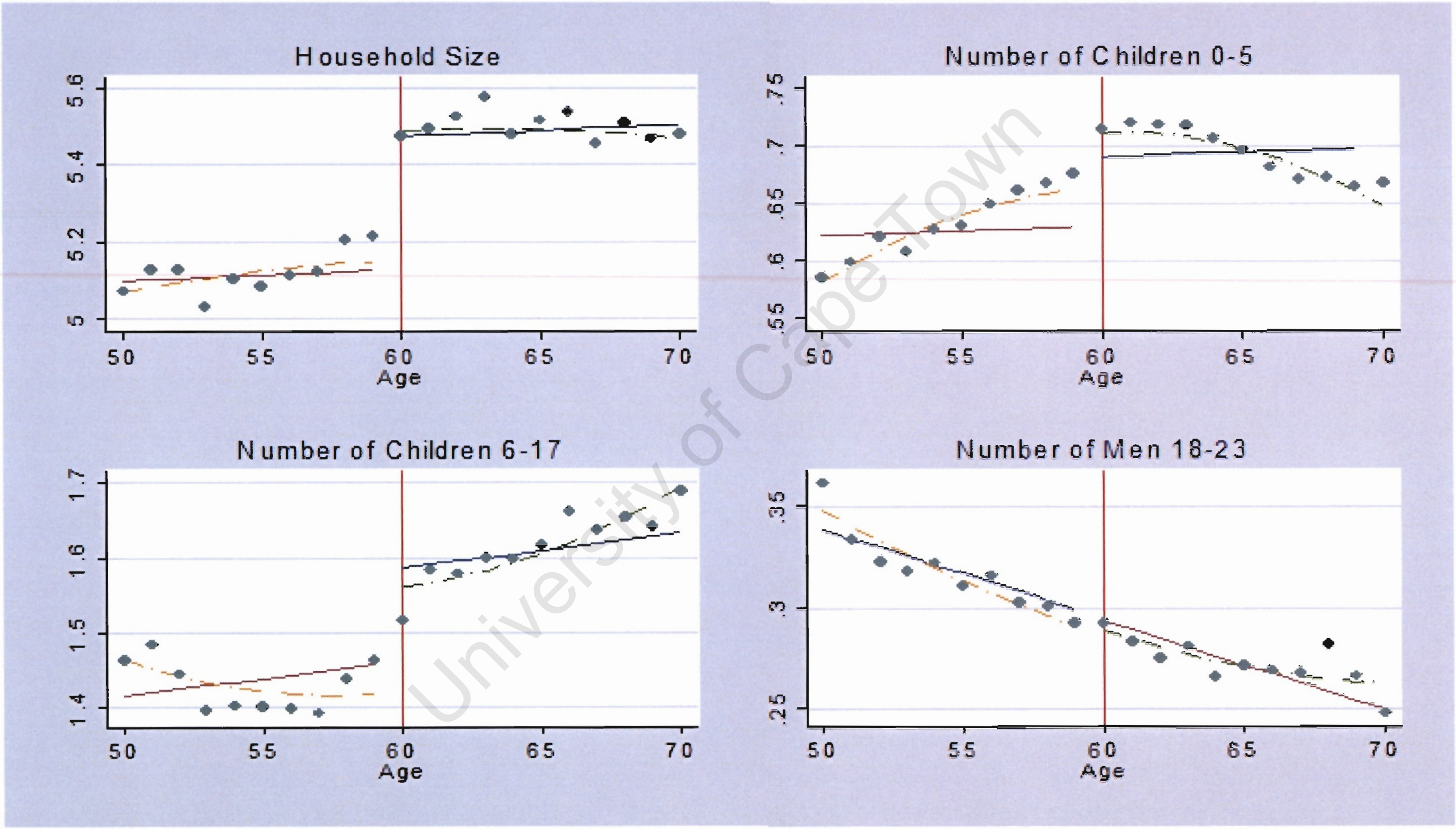
University of Cape Town



University of Cape Town

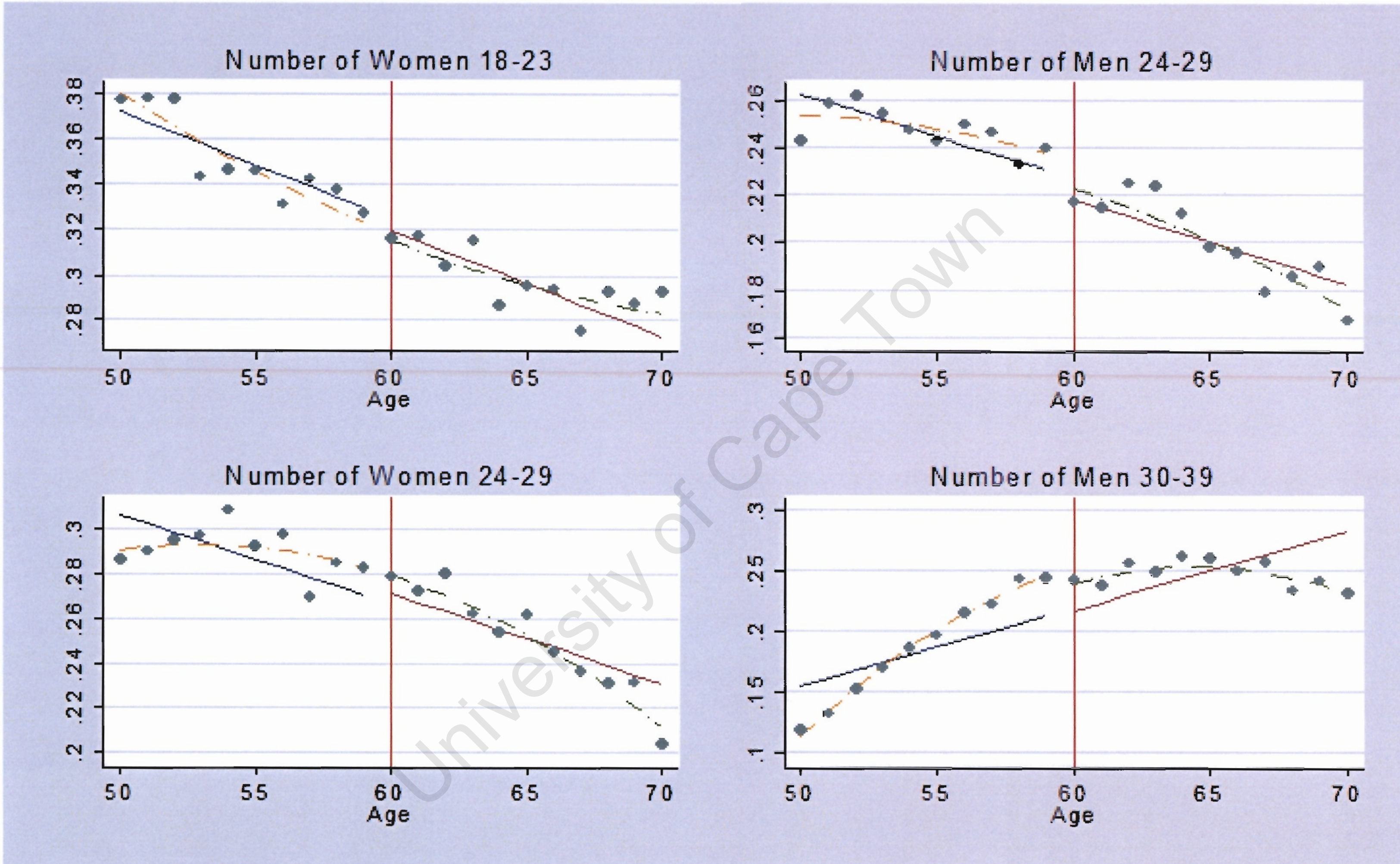
Source: 2001 South African Population and Housing Census

Figure 13: Census Means and Parametric Results for the Regression Discontinuity Analysis on the Second Sample



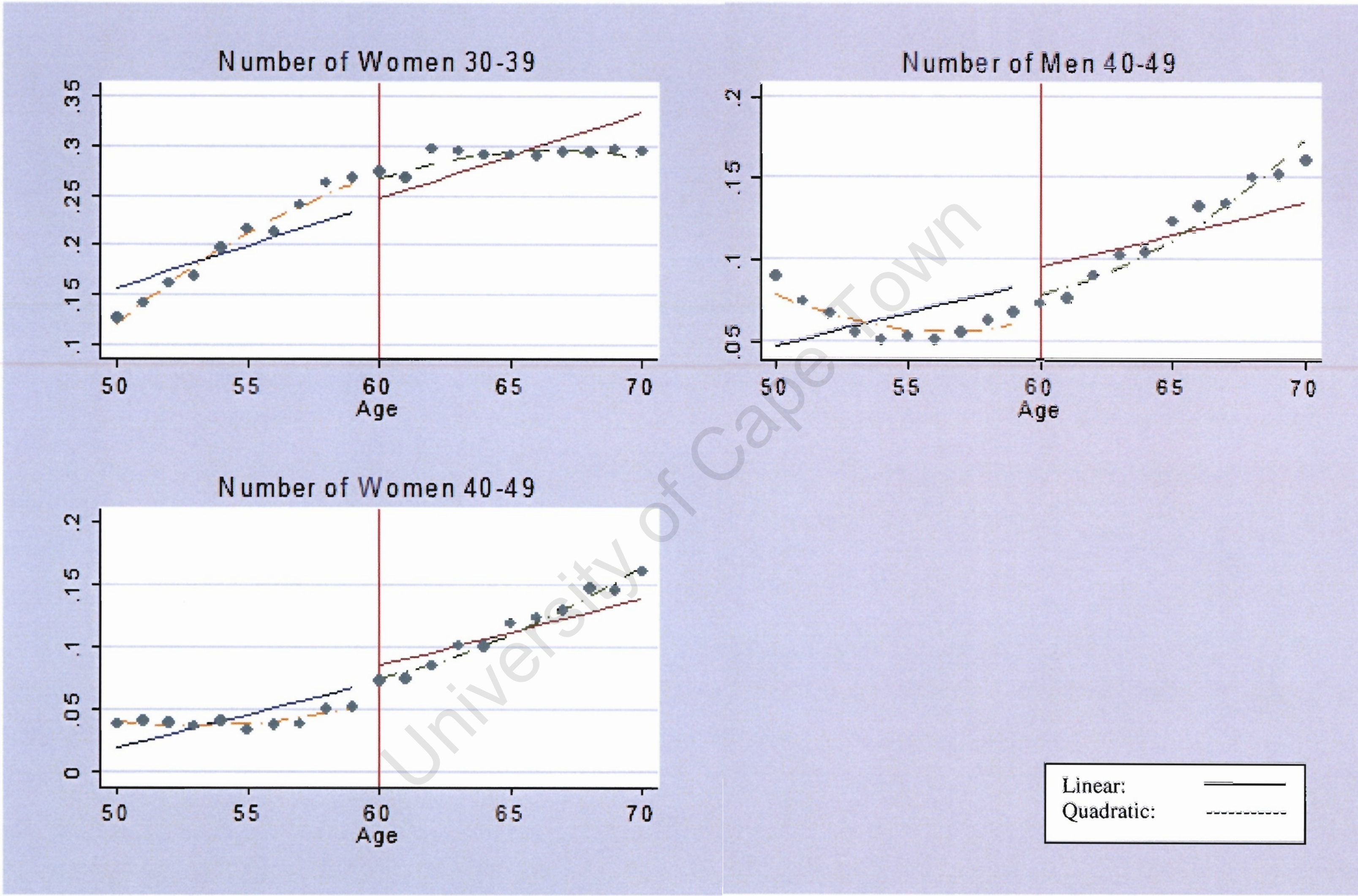
Source: 2001 South African Population and Housing Census

University of Cape Town



Source: 2001 South African Population and Housing Census

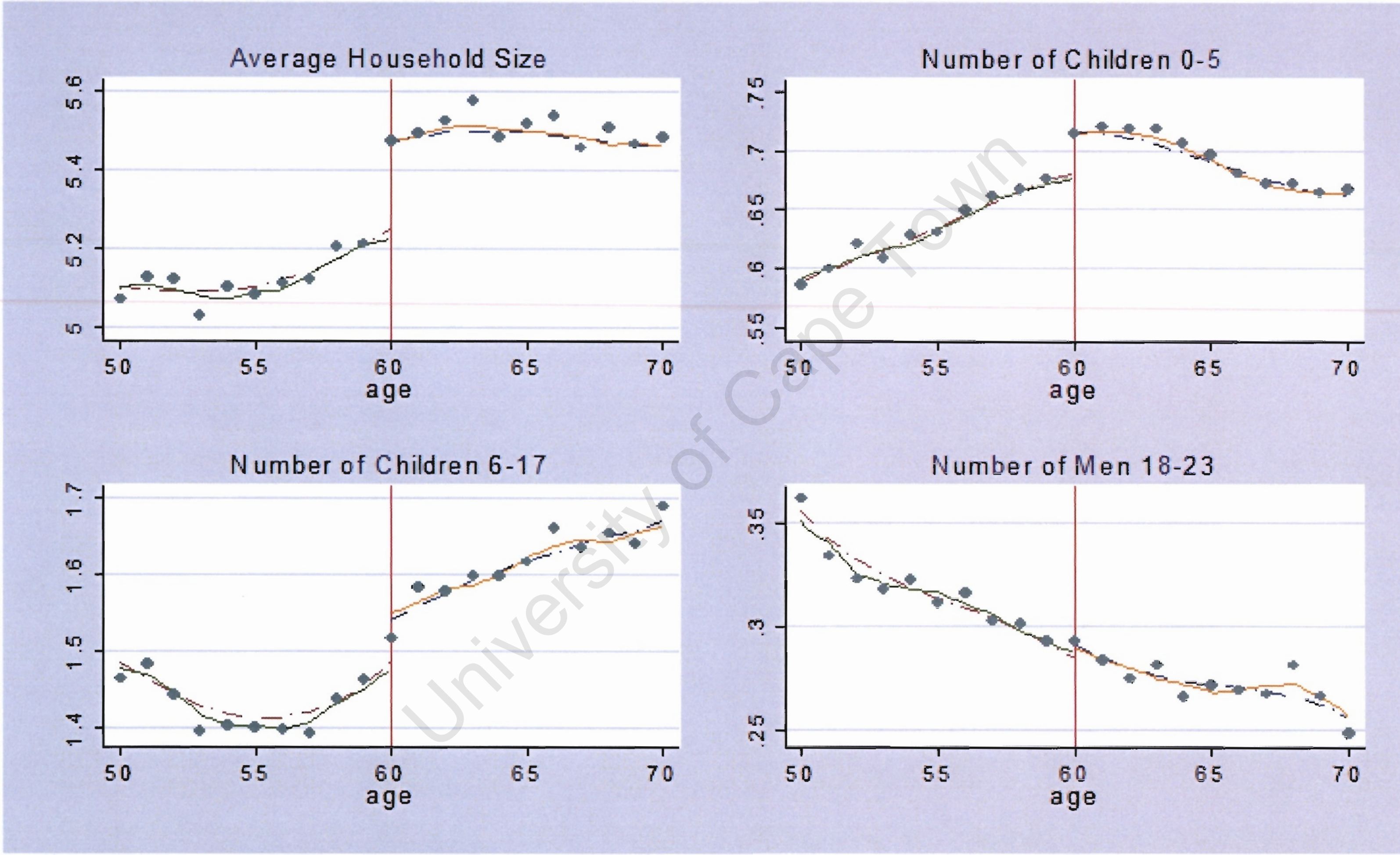
University of Cape Town



University of Cape Town

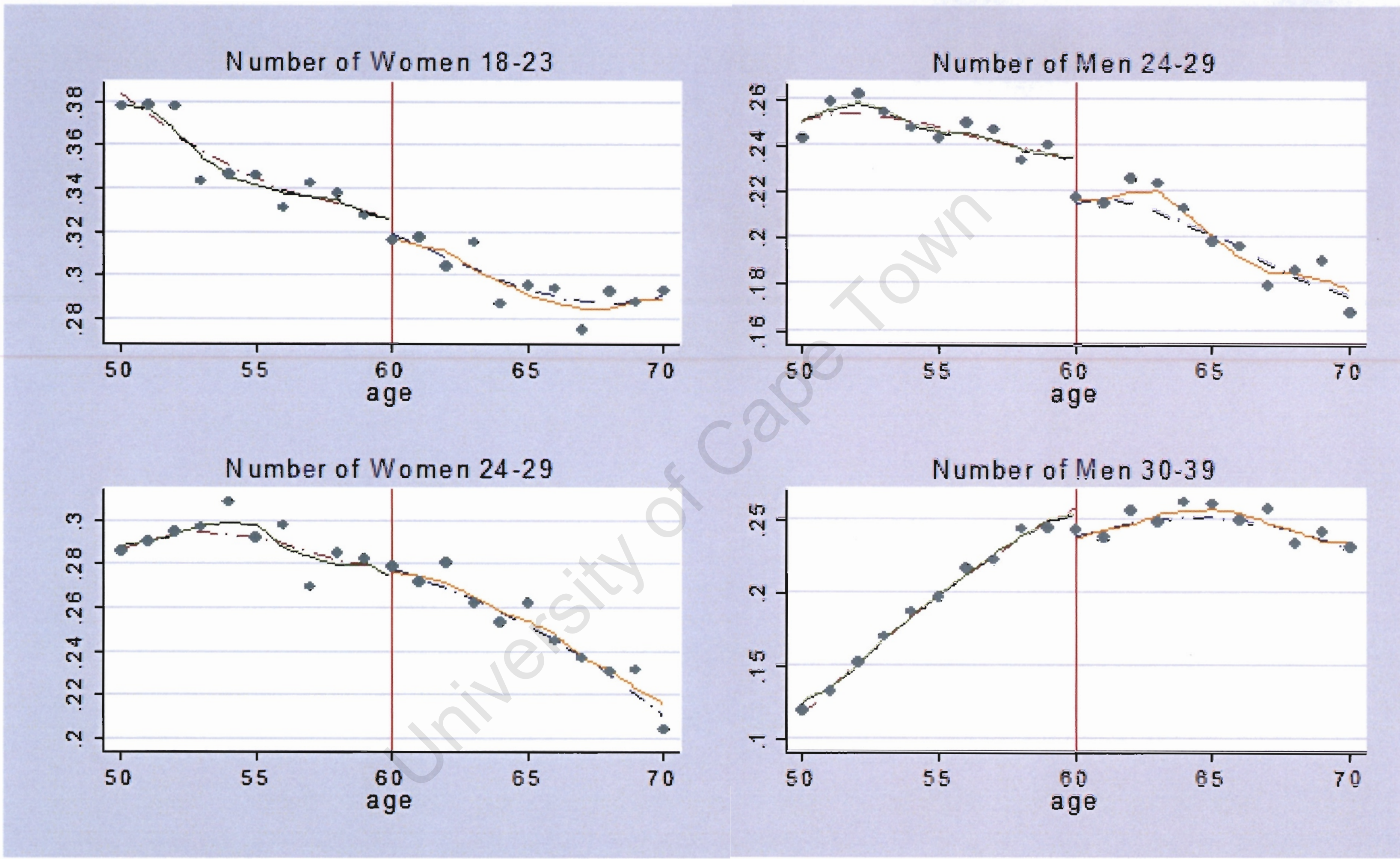
Source: 2001 South African Population and Housing Census

Figure 14: Census Means and Non-parametric Results for the Regression Discontinuity Analysis on the Second Sample



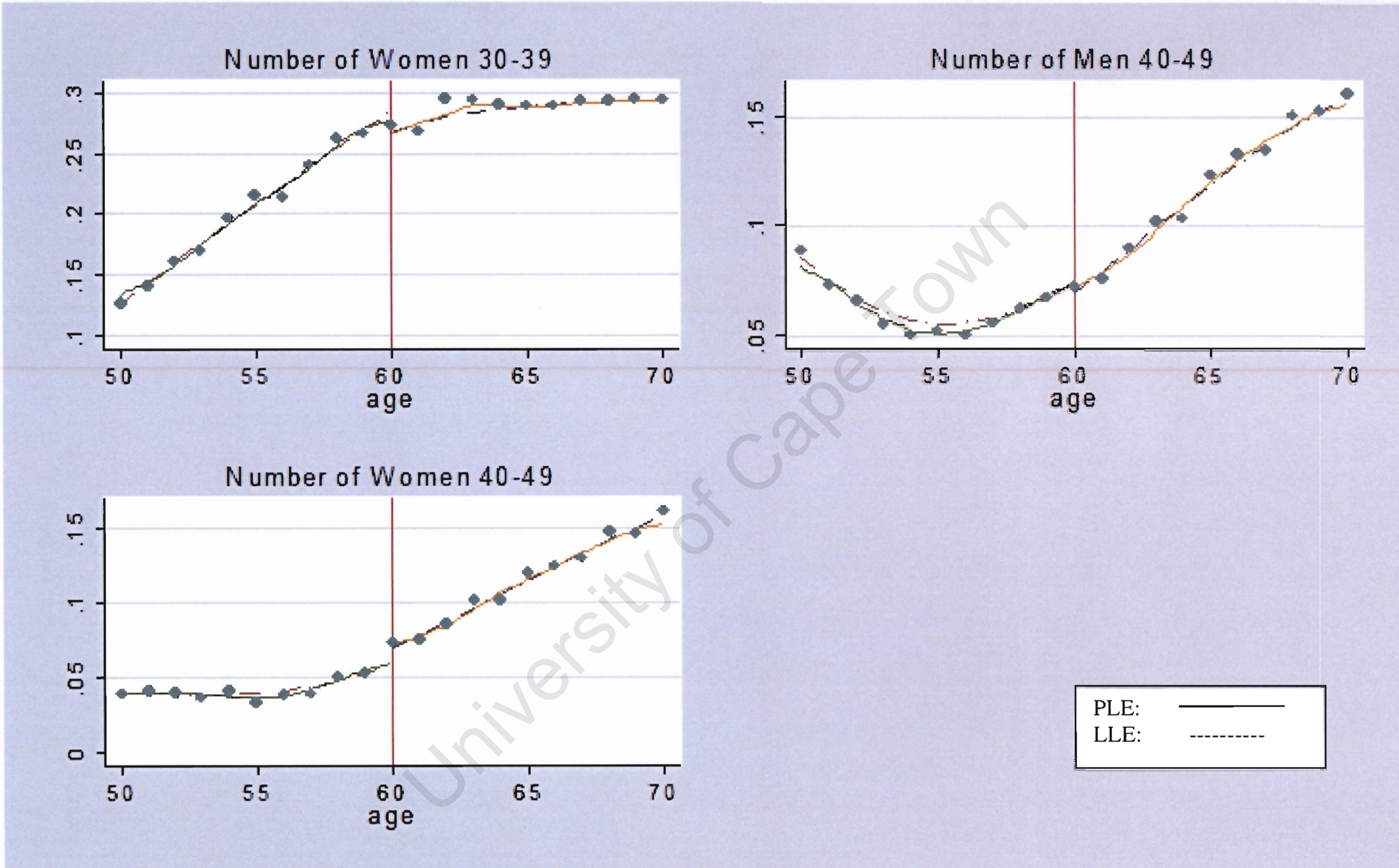
Source: 2001 South African Population and Housing Census

University of Cape Town



University of Cape Town

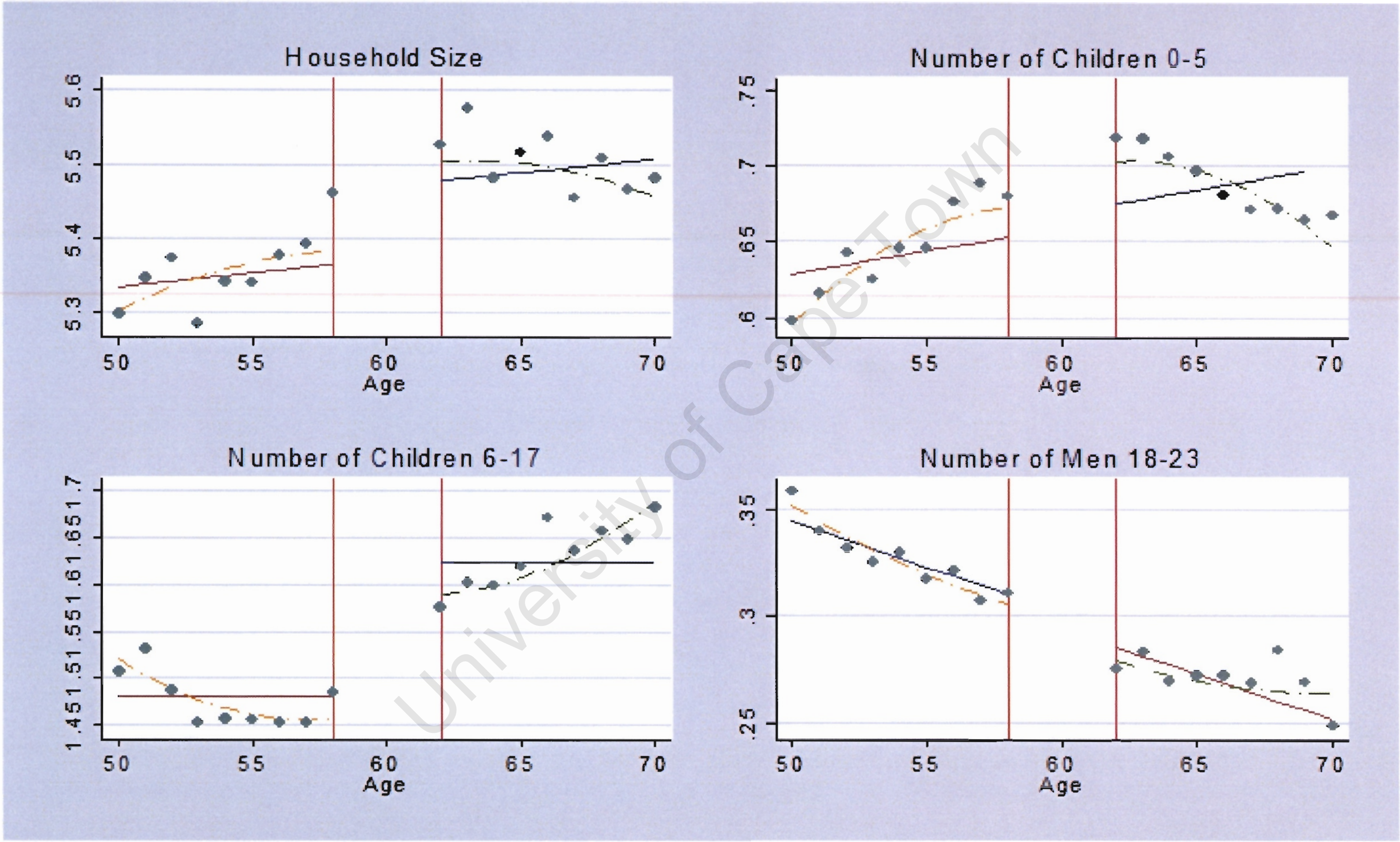
Source: 2001 South African Population and Housing Census



Source: 2001 South African Population and Housing Census

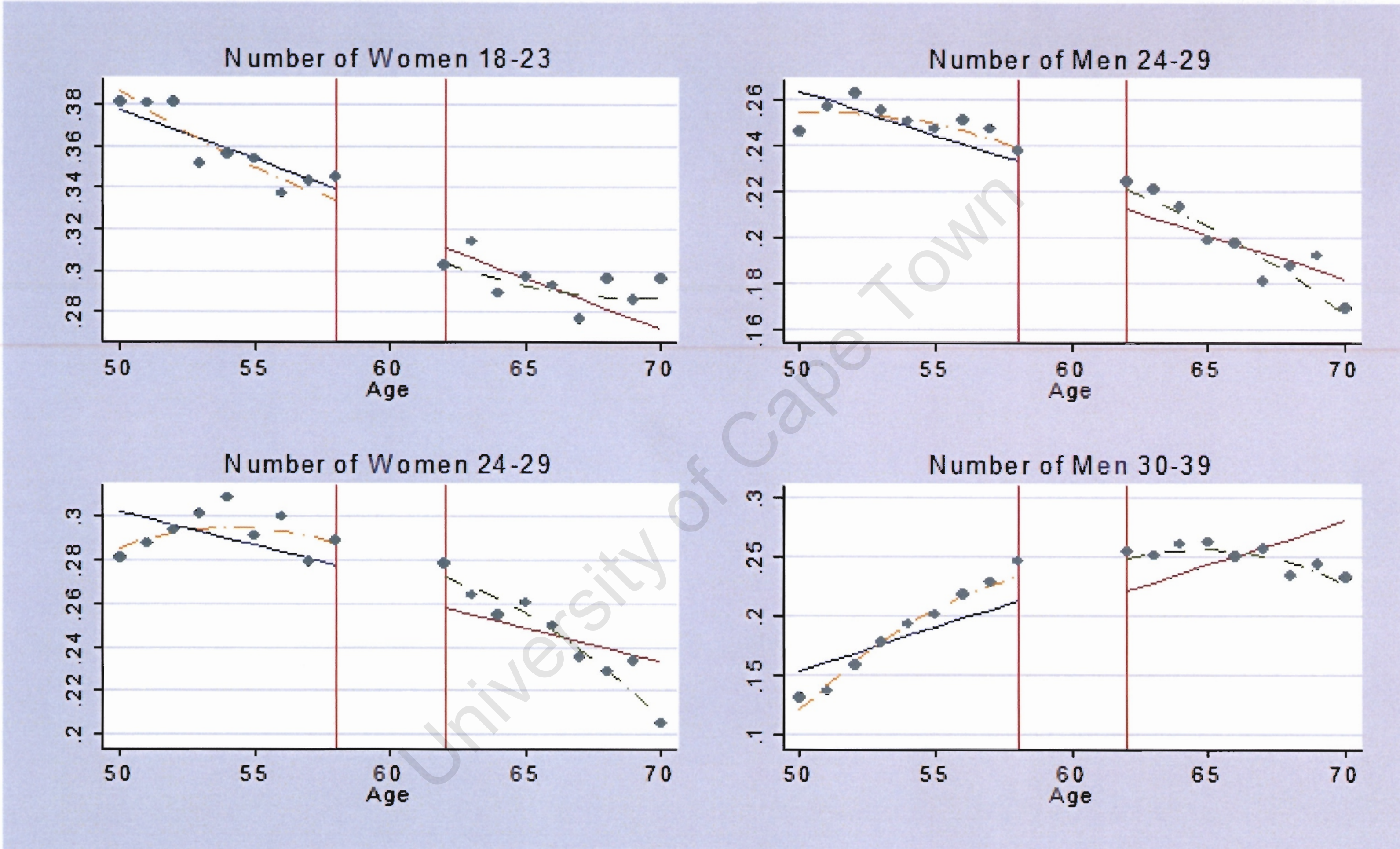
University of Cape Town

Figure 15: Census Means and Parametric Results for the Regression Discontinuity Analysis on the Third Sample



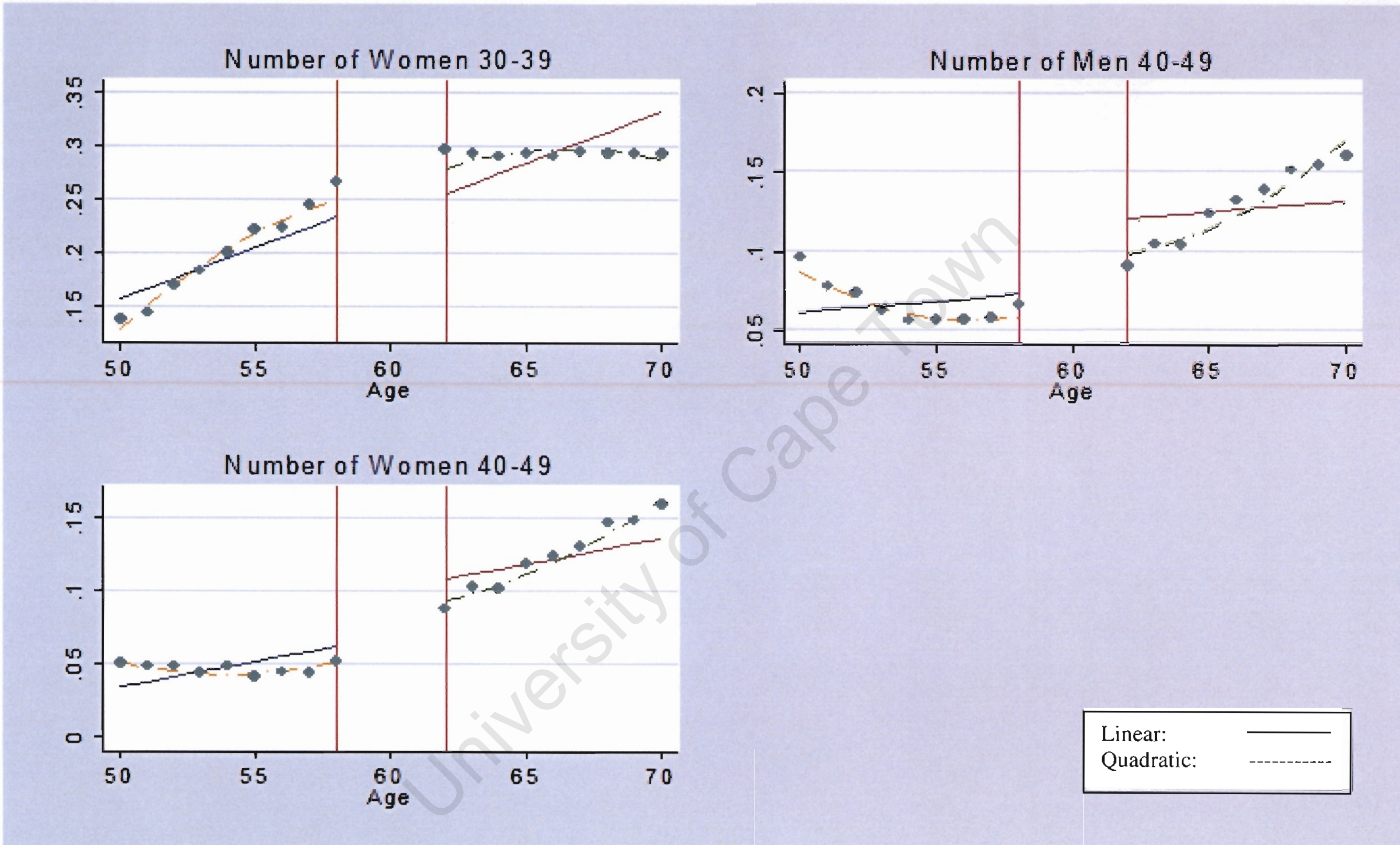
Source: 2001 South African Population and Housing Census

University of Cape Town



Source: 2001 South African Population and Housing Census

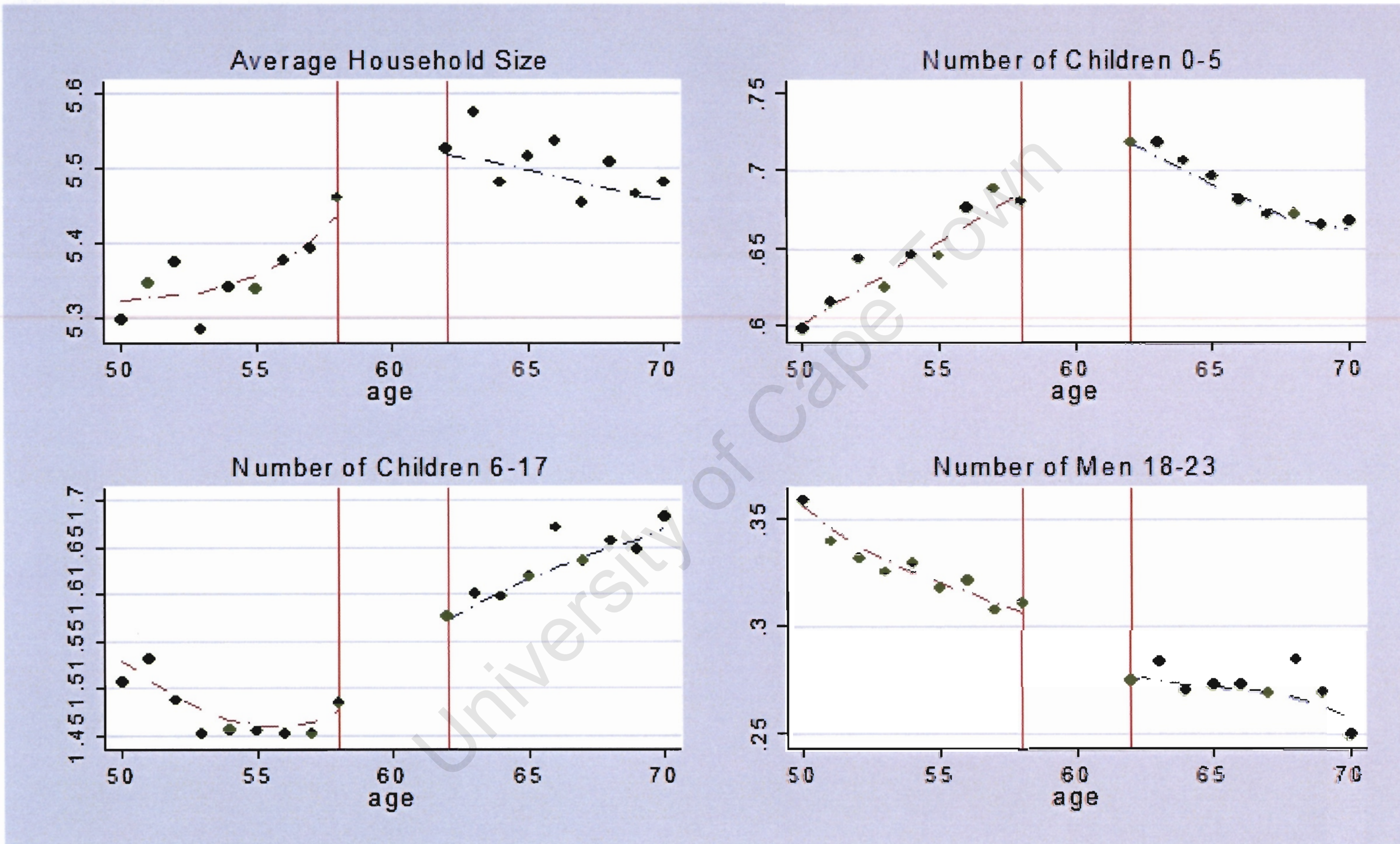
University of Cape Town



Source: 2001 South African Population and Housing Census

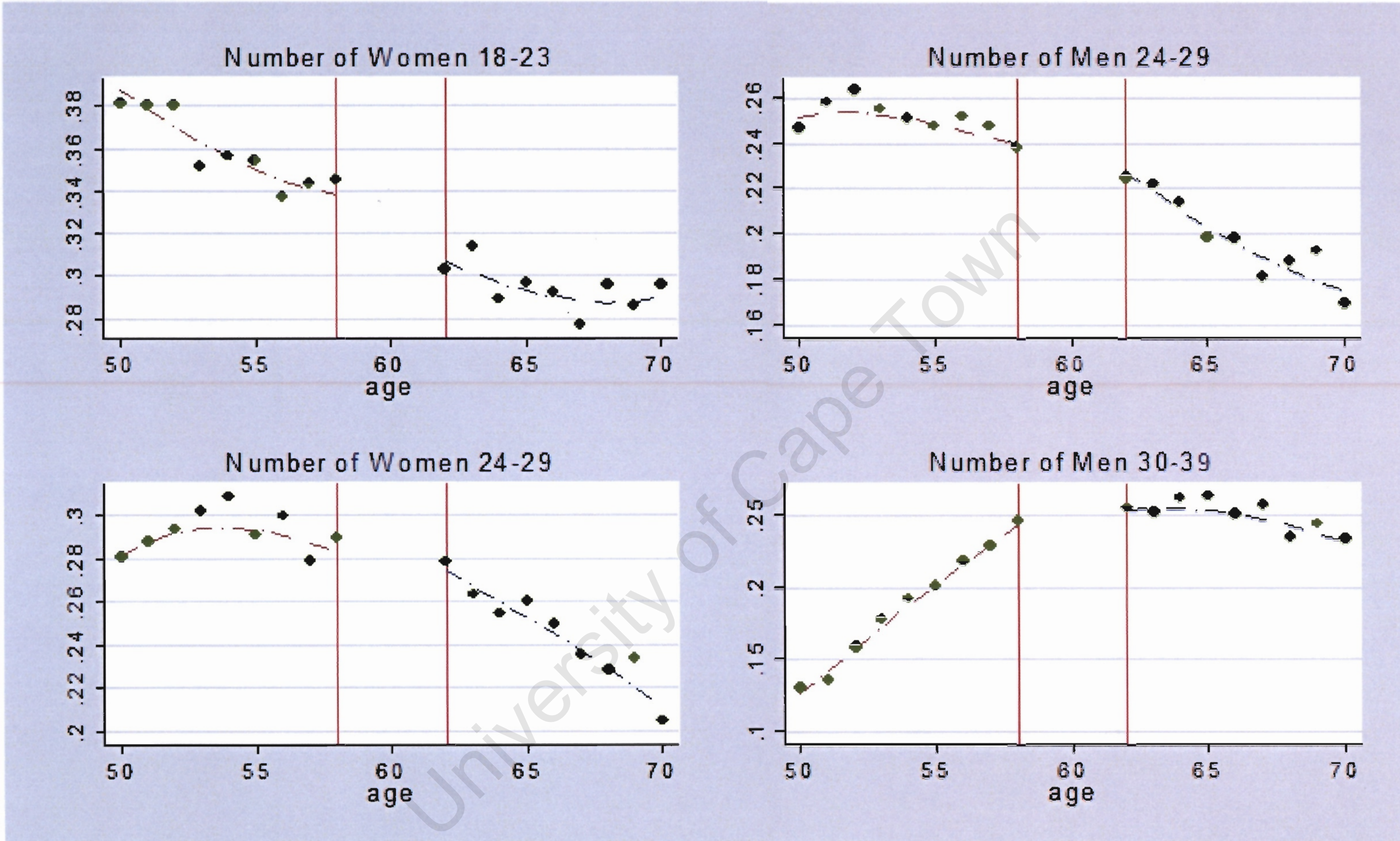
University of Cape Town

Figure 16: Census Means and Non-Parametric Results for the Regression Discontinuity Analysis on the Third Sample



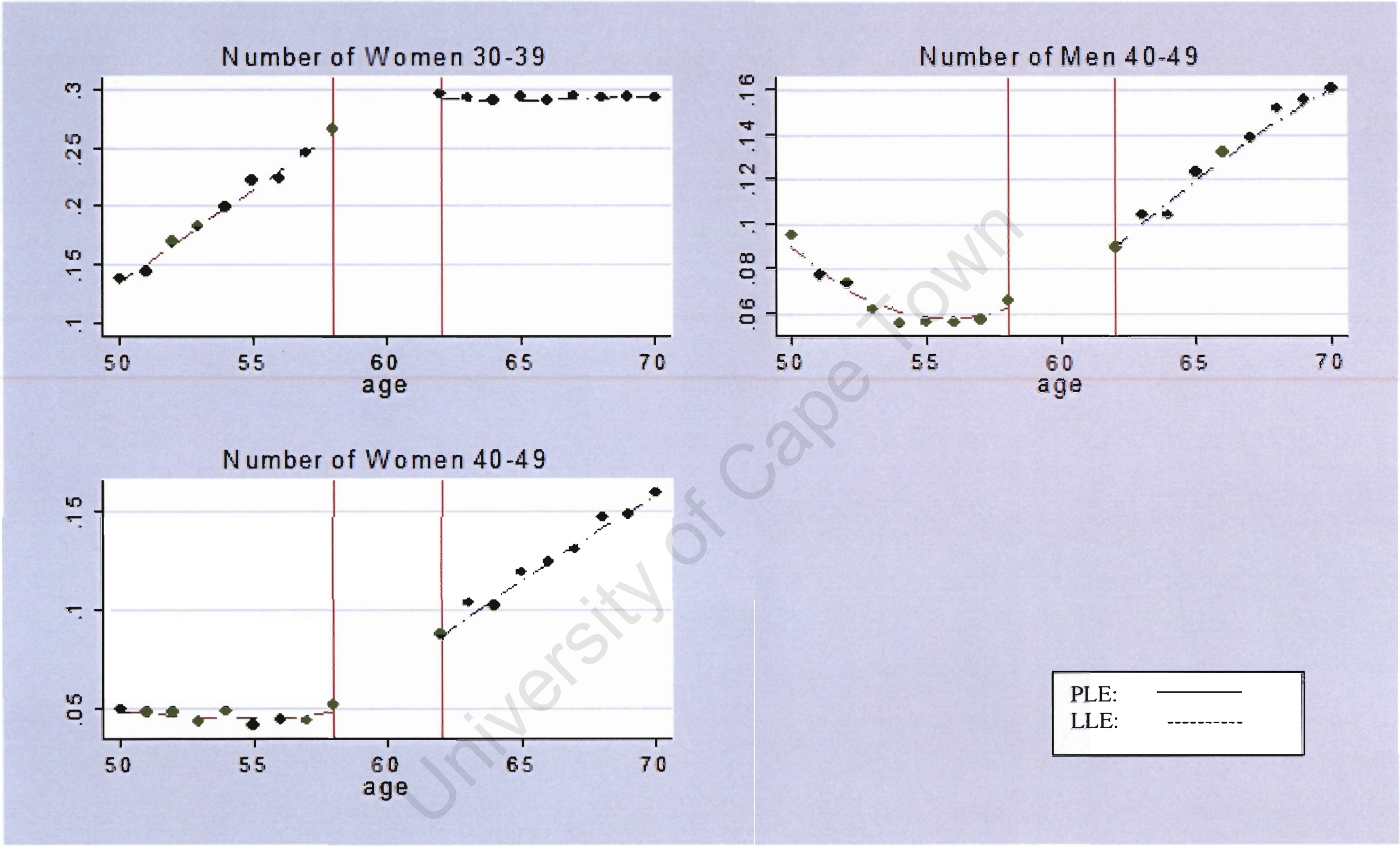
University of Cape Town

Source: 2001 South African Population and Housing Census



Source: 2001 South African Population and Housing Census

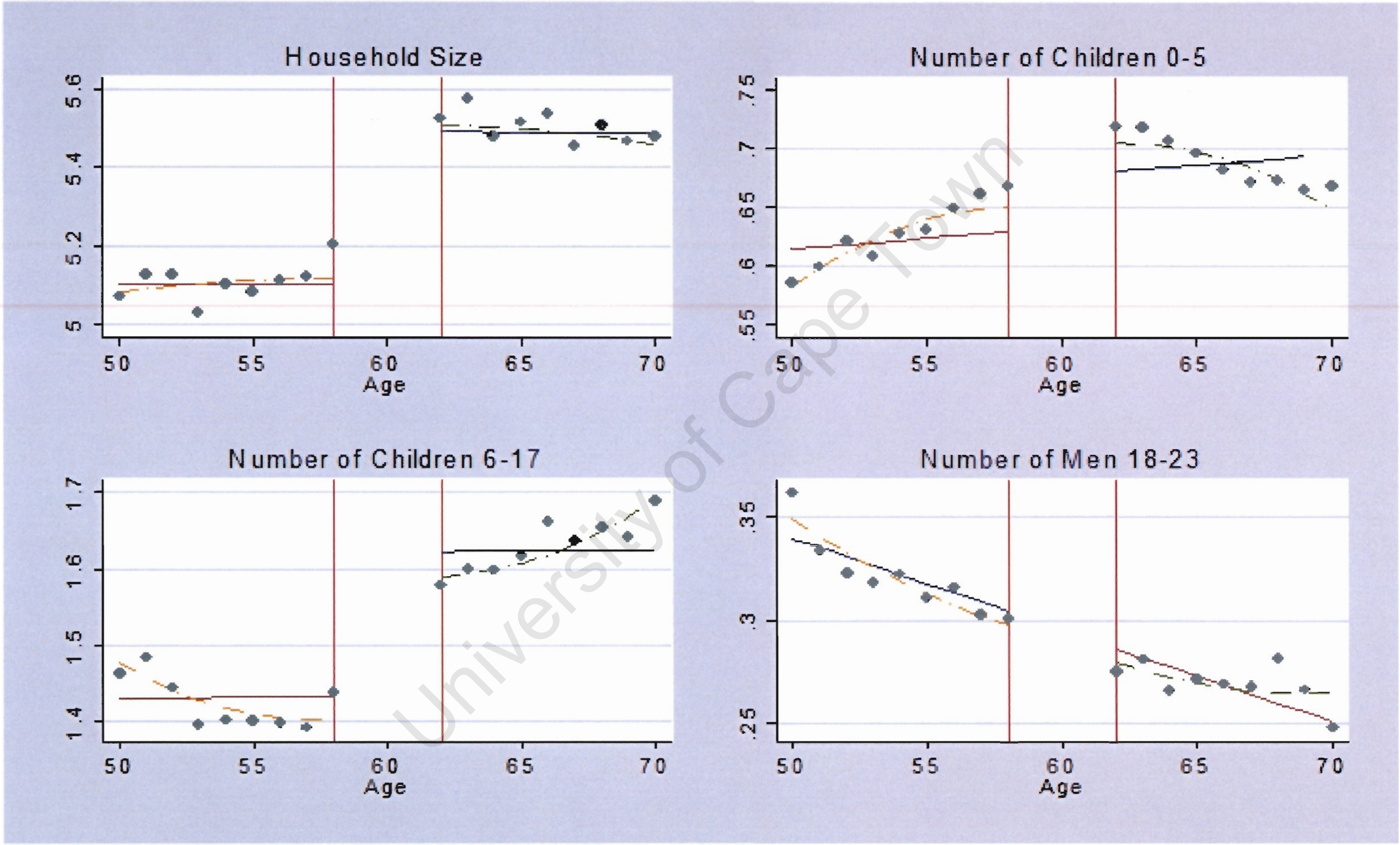
University of Cape Town



Source: 2001 South African Population and Housing Census

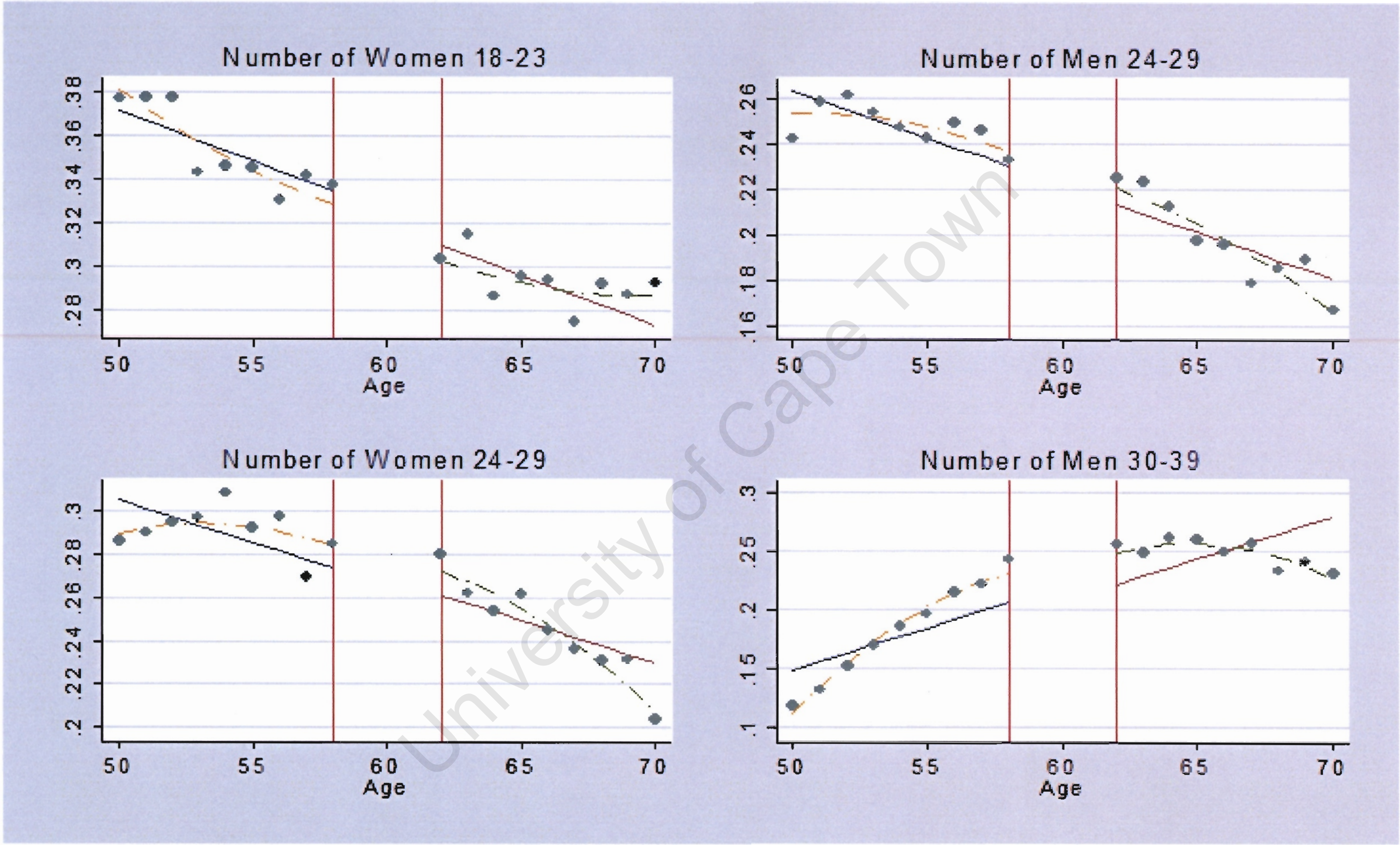
University of Cape Town

Figure 17: Census Means and Parametric Results for the Regression Discontinuity Analysis on the Fourth Sample



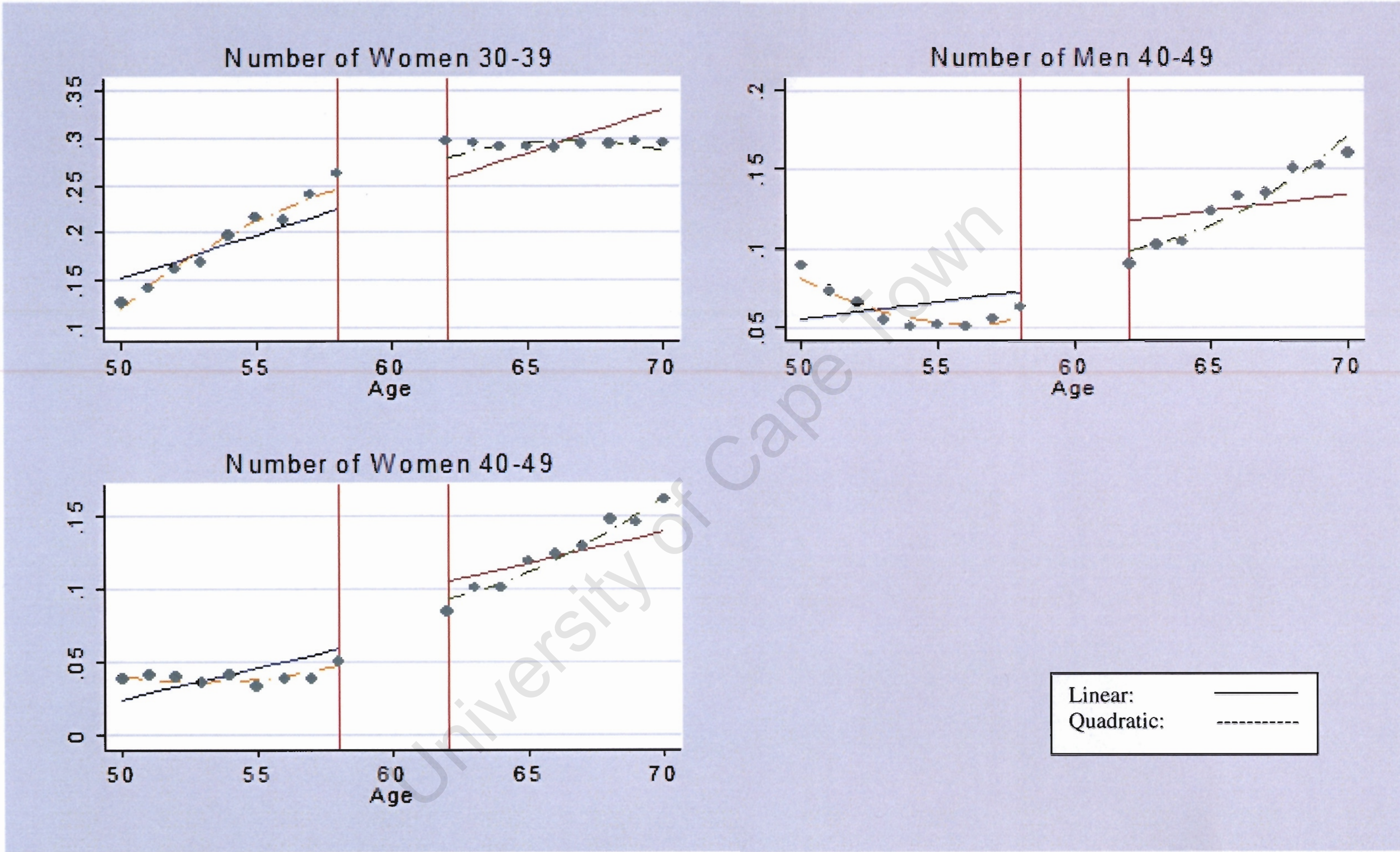
Source: 2001 South African Population and Housing Census

University of Cape Town



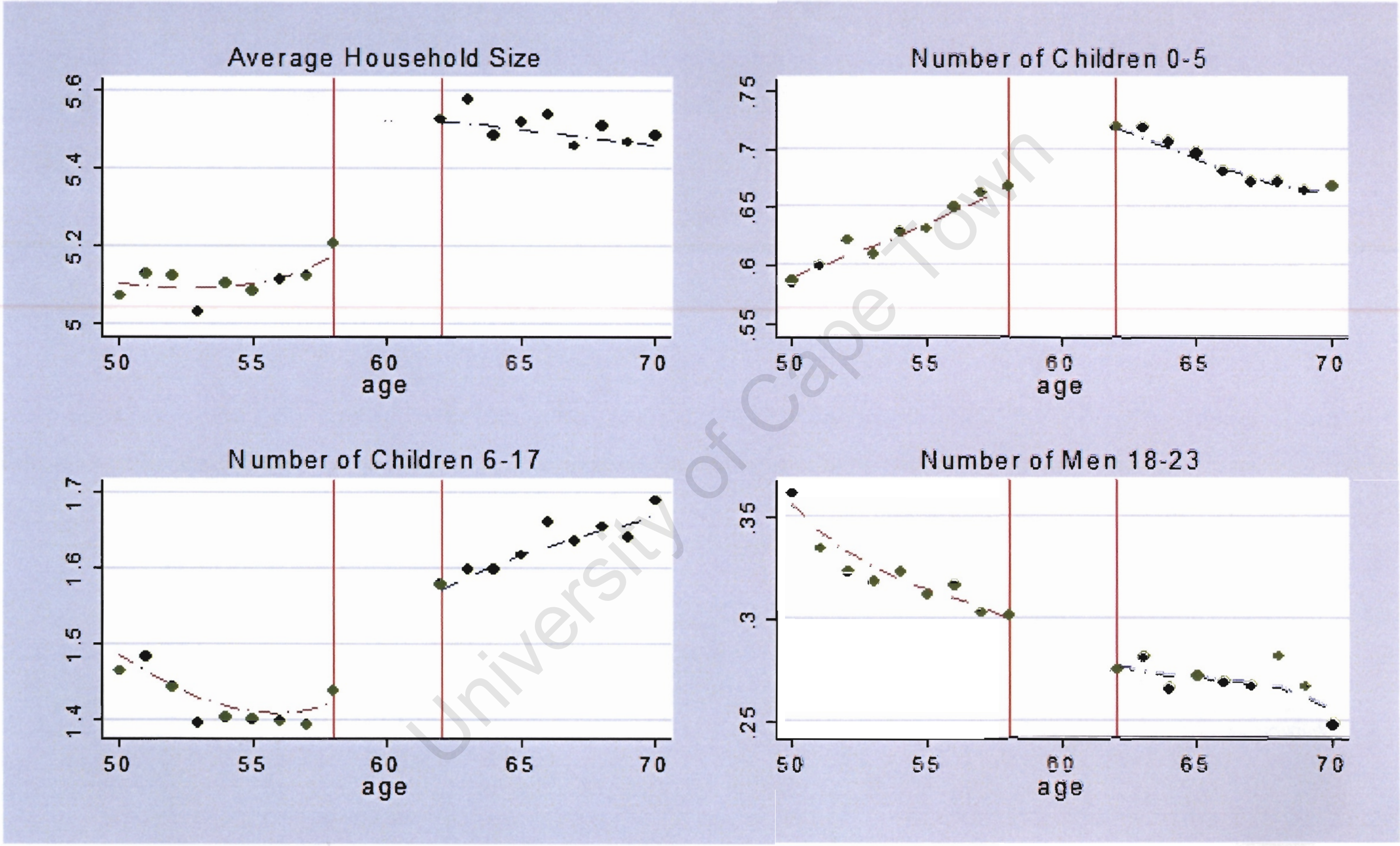
Source: 2001 South African Population and Housing Census

University of Cape Town



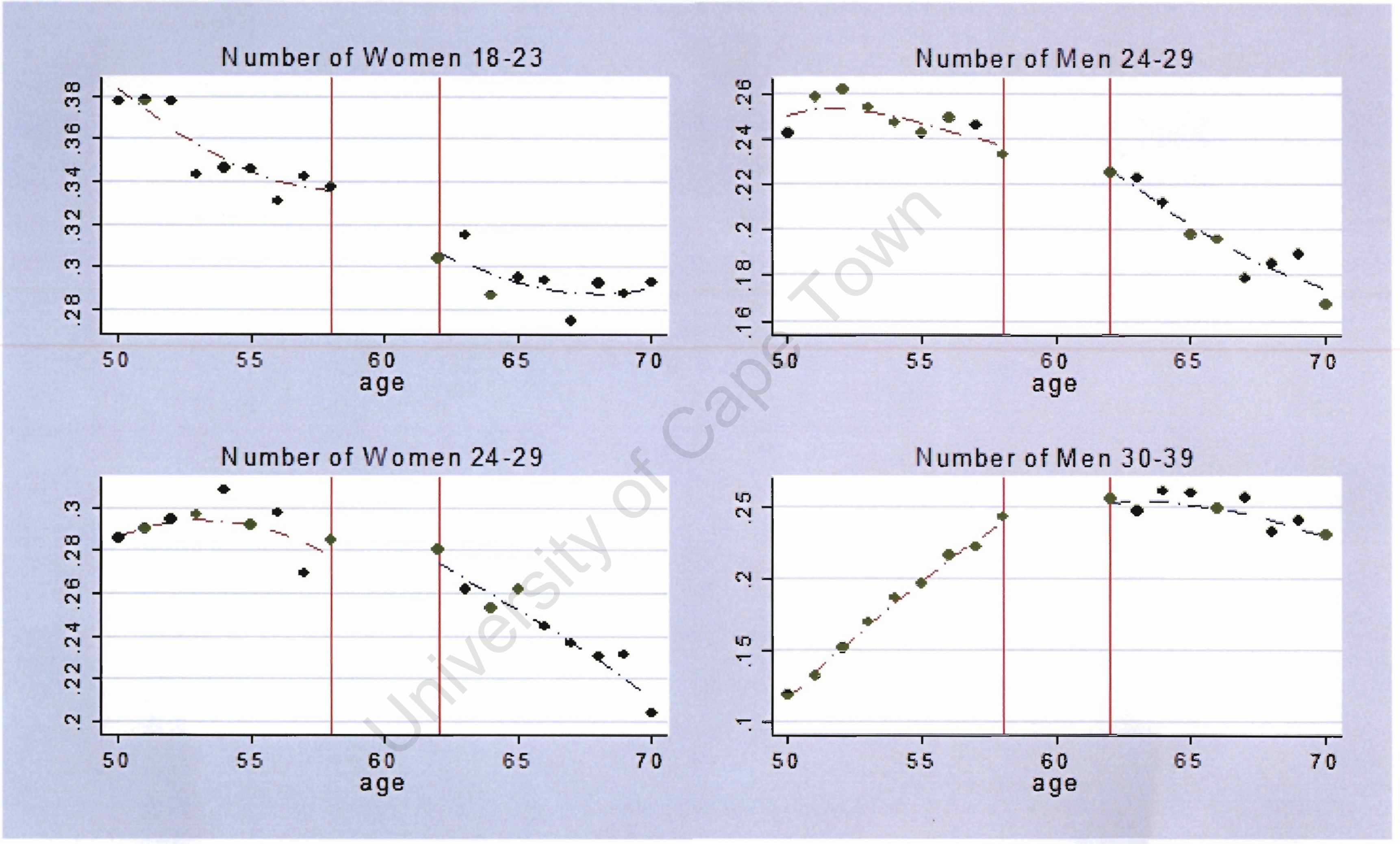
Source: 2001 South African Population and Housing Census

Figure 18: Census Means and Non-Parametric Results for the Regression Discontinuity Analysis on the Fourth Sample



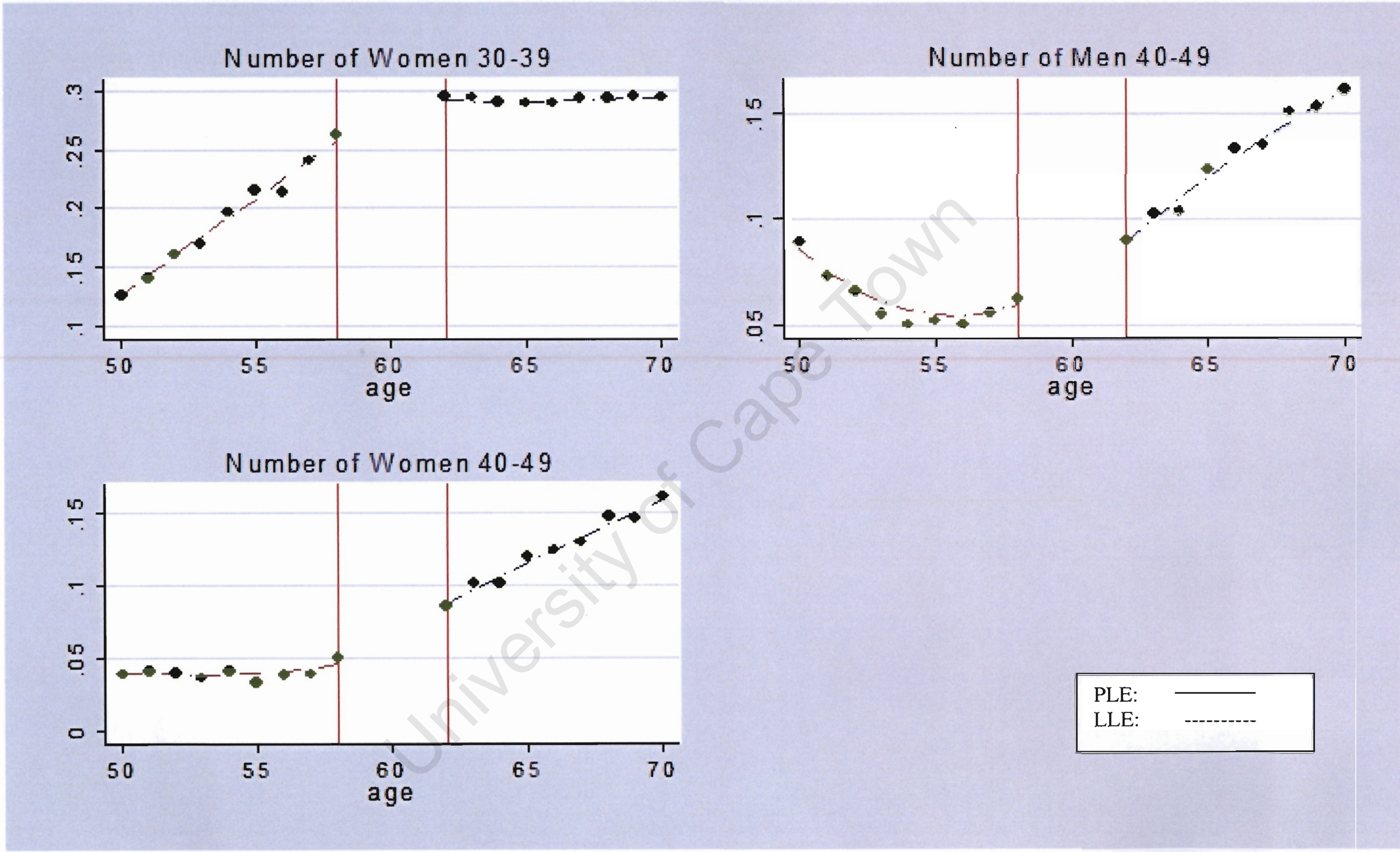
Source: 2001 South African Population and Housing Census

University of Cape Town



Source: 2001 South African Population and Housing Census

University of Cape Town



Source: 2001 South African Population and Housing Census

University of Cape Town