Essays in Education and Labor Economics

by

Vimal Ranchhod

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in The University of Michigan
2007

Doctoral Committee:

Professor David A. Lam, Co-Chair
Professor James A. Levinsohn, Co-Chair
Professor Charles C. Brown
Professor Justin Reed McCrary

To my parents and sisters.

ACKNOWLEDGEMENTS

There are so many people who contributed to my being here, and whose assistance and inspiration I gratefully acknowledge.

First and foremost to my parents, who made lifelong sacrifices so that myself and all of my sisters are free to be masters of our destinies. I have had privileges and opportunities that very few people from my background have enjoyed. My mom and dad provided a stable and safe home environment, a family that forms a strong part of my identity and one that I am extremely proud of, and inculcated in us a love of learning that motivates me far beyond the strategic and economic payoffs associated with a formal educational qualification.

To my sisters, Maya, Jayshree and Jasvina, who have always been supportive and generous and full of encouragement. Knowing that I could always count on them for anything allowed me the freedom to spend years in what at times seemed like a failing endeavour. The same goes for my brothers-in-law, Jits, Ajen and Raj. And my lovely niece and nephews. At certain times, the only reason I had to finish the thesis was to graduate and go home to be with them.

A very special expression of gratitude is due to Taryn Dinkelman. Taryn never stopped caring despite the pain, and never withdrew even during my darkest hours. She was always there to celebrate the small victories and commiserate in defeats, and helped me to keep going when I simply did not have the energy to do so myself. More than anyone, she might understand the cost this process has exacted on me. I shall always be in her debt.

My advisors have all been wonderful and giving of their time, advice and energy. David, Jim, Charlie and Justin formed a committee par excellence, and my only regret is that I didn't take more advantage of their incredible skills. I am also in awe of the generosity both David and Jim have shown to me in connecting me to people and resources in South Africa, in arranging things so that I could go home at least once a year, and in helping me to return with a set of options that seem truly incredible to me.

There have been other academic mentors as well. The first to come to mind are Martin Wittenberg and Johannes Fedderke, who literally bullied me into continuing my studies in economics. I was not always pleased with them for the intervention. Indeed, I could have made an excellent auditor. Murray Leibbrandt, in his quiet and unassuming way, has provided a difficult role model to try to emulate. I've enjoyed and benefited from talking to Steve DesJardins, John Dinardo, Jeff Smith and Gary Solon.

Then there are all the other members of my huge extended family, especially Bhavna, Jayeshbhai, Jeetash, my Narsikaka and my Manimasi. With all of these people, my concept of home has never wavered.

I have enjoyed the warmth of several friends along the way. The 'boys' I did most of my growing up with; Bhaiya, Mits, Fish, Naans, Raj, Piney and Dhrups. The friends who became family while in the US. Rodney, Heidi, Coul, Herman, Mato, Waldery, Yo, Chris, Mine, and Owen from my first stint here. Clau and Jorge, Leticia, Seb, JB, Estelle, Nzinga, Yaz and Sara from my second.

Finally, there were people with whom the daily interactions were always pleasant and positive at the Population Studies Center, the School of Public Policy and Lorch Hall. Mary Braun deserves a medal for being as sweet and helpful as she is. I am also grateful for financial support from the Mandela Economics Scholarships Program, the Fogarty International Center and the Spencer Foundation.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Table

CHAPTER I

Introduction

The first chapter of the dissertation investigates how poor households in developing countries respond to adverse income developments. I use nationally representative longitudinal data to investigate behavioral responses to the loss of an Old Age Pensioner in South Africa. I find that household composition adjusts, with an outflow of school aged children and an inflow of middle-aged females and older adults. The household, on aggregate, also has more people employed. Conditional on compositional stability within demographic groups, I find large and significant increases in both labor supply and employment. Policy makers might be concerned with the impact of cash transfers on the labor supply of non-recipients.

The second chapter examines the effect of the Old Age Pension on retirement behavior of elderly South Africans. I make use of the rules on age eligibility to measure changes in various dimensions of labor supply that occur when people reach the pensionable age. I find significant decreases in employment rates and labor supply. Those who remain employed beyond the pensionable age are more likely to work in jobs with flexible hours of work, and work even fewer hours than people in similar jobs who are not pension age-eligible.

The final chapter investigates the impact of changes in the probability of being admitted into a selective college on students' SAT score sending behavior. We capture this using a student's class rank combined with the Texas Top Ten Percent Rule. In response to dwindling minority enrollment rates, the University of Texas at Austin and Texas A&M College

Station embarked on targeted recruitment programs at previously under-represented high schools. We evaluate the effectiveness of these programs using individual level SAT data. We find that score sending is affected by the legal change, and that both targeted recruitment programs were successful in attracting scores. In each case, the effects were manifest most strongly amongst the students in the top decile of the class. This suggests that students from poor schools face multiple barriers to obtaining postsecondary schooling at selective colleges.

CHAPTER II

Household responses to adverse income shocks: Pensioner out-migration and mortality in South Africa

2.1 Introduction

How do households respond to the cessation of cash transfers in developing countries? I estimate the magnitude of changes in household composition and household labor supply that occur when a pensioner leaves the household, either due to out-migration or death. The data used is nationally representative household level matched data from South African Labour Force Surveys (LFS) from September 2001 to March 2004.

The non-contributory South African Old Age Pension (OAP) forms the backbone of the South African social security system. Recipiency rates are high amongst the elderly, and over 77% of Africans who are age-eligible report receiving the pension. In addition, a means test ensures that the pension disproportionately reaches poorer households. Not only is coverage widespread, but its value is sufficiently high to generally make the pensioner the main breadwinner in their households. Case and Deaton (1998) note that in 1993, the value of the pension was "twice the median household's per capita income" amongst African households. Based on the September 2002 LFS, 19.28% of all households report "pensions and grants" as their main source of income. Amongst households with a member who is old enough to be eligible, this percentage rises to 63.67% for all households, and 70.17% for African headed households.

Given the importance of the pension, investigating how household's cope with its loss is of interest for at least two policy related reasons. First, it can inform us as to how relatively poor families act to mitigate against the effects of adverse economic developments. One dimension of this involves household responses in terms of household composition. If the group that constitutes a 'household' is itself endogenously determined, then careful consideration for policies targeted at the household level is warranted. Second, sharing of pension income within households might lead to non-recipients deciding not to work. On the other hand, if poor households are liquidity constrained and the pension eases these constraints, we might observe that employment rates actually increase as a result of the pension. Estimating the magnitude of the net effect is the primary contribution of this paper.

To summarize my results, I find significant evidence that the pension does indeed affect both household composition and labor force participation of non-recipients. The largest and most robust effects are observed for older adults. Households re-organize such that they have more adult time in the labor force, more non-pension aged adult residents who are not employed, and a decrease in the number of school-aged children. I find significant increases in labor force participation and employment amongst resident non-pensioners.

2.2 Background

Lund (1993) provides an introduction to the OAP as we see it today. As stated previously, the pension is means tested, and provides a relatively generous cash transfer to recipients. Eligibility depends only on age, nationality and satisfying the means test. The age-eligibility threshold is 60 for women and 65 for men. The level of the means test is set fairly high, so that most of the elderly receive the grant. Moreover, it is based on individual income for the unmarried elderly, or joint spousal income for married couples, and hence should not have distortionary 'implicit taxation' effects for other household members.

The value of the pension is adjusted periodically, usually on an annual basis, to adjust for inflation. In 2002 and 2003, the value of the pension was set at 620 and 700 rands per month respectively. Adjusting for consumer inflation[1], and converting using the prevailing

[1] The deflator used is the official Consumer Price Index released by Statistics South Africa.

exchange rate of 6 rand : 1 US dollar, these numbers equate to approximately $125 per month. This is a large transfer relative to potential wage income, and continues for as long as the pensioner remains alive and continues to satisfy the means test.

2.3 Related Literature

Several researchers have investigated the effects of pension recipiency on various dimensions of household welfare. Case and Deaton (1998) find that the pension is an effective tool for redistribution and that the households it reaches are predominantly poor. Furthermore, the prevalence of three-generation households, as well as 'skip generation'[2] households, results in the pension disproportionately reaching children in poverty.

Some authors have looked at whether the OAP impacts on the health of recipients or their household members. Duflo (2000) finds a discontinuous increase in girls' height for age for children living with pension eligible persons. This increase is significant and is realized on average only when the pension recipient is a woman. Duflo (2003) reports similar evidence that the pension is shared between members of the household. Moreover, the sharing of recipients' income from pensions is differentiated by gender. Case (2001) finds that the health of all household members is improved as a result of the pension.

Others have asked the question; 'How do other members of the household respond when a member becomes pension eligible'? Bertrand *et al* (2003) analyse cross sectional data and find that having a pension eligible person in the household has a statistically significant and negative impact on the labor supply of prime aged resident males in the household. Edmonds (2003) considers the impact of the OAP on child labor supply and schooling attendance. Ranchhod (2006) finds that the pension causes retirement amongst the recipients themselves. Posel *et al* (2006) find that the pension actually increases the labor supply of non-resident household members by financing labor migration towards areas with better employment prospects.

Jensen (2003) questions whether household disposable income increases by the full value

[2]Households with grandparents and grandchildren but non-resident parents.

of the pension. He estimates that crowding out of remittances by pensions is large and significant. On average, every rand of pension income received by the elderly is met with a 0.25 to 0.30 rand decrease in remittances received from the pensioner's children. Pension income is thus *de facto* shared with family members even when they do not reside with the pensioner.

Edmonds *et al* (2005) find that household composition itself is affected by someone becoming pension age-eligible. They find a decrease in the number of prime working-age women, and an increase in the number of children younger than five and young women of childbearing age. Hamoudi and Thomas (2005) go further, and find evidence that the OAP results in compositional changes consistent with sorting on the basis of unmeasured personal characteristics. This result poses a challenge to much of the prior research conducted on cross-sectional data.

Given that the pension seems to be so important in sustaining the poor and the elderly in South Africa, a natural question to consider is how do these households cope when the pension income stops. A paper which asks a similar question, *inter alia*, is presented by Ardington *et al* (2007). Using individual level longitudinal data from a poor rural area in the KwaZulu-Natal province, they investigate changes in labor supply amongst both migrant and resident household members as a function of the pension. In their dataset, after controlling for individual level fixed effects, they find that both the migration decision as well as labor supply are positively affected by pension receipt. Moreover, they find asymmetries in the effects of the pension on household members who are already migrant laborers, as compared to those who are potential migrants.

In this paper, I answer the question: 'How does household composition and labor market activity of resident household members change at the same time that we observe the departure of the pensioner', either due to out-migration or death. This question has not been thoroughly investigated in the literature to date. A second major contribution is that it is the first study in the OAP literature that does so using nationally representative and longitudinal data.

2.4 Theory

The most basic model of household formation assumes that households form for the production of some non-tradeable good in which there are economies of scale (Becker, 1973). In this paper, I assume that household composition and labor supply of household members are both endogenous outcomes to changes in non-labor income. Various authors have commented on the fact that inter-household migration occurs in response to the pension; see Keller (2004), Edmonds *et al* (2005) and Posel *et al* (2006).

Economic theory is fairly clear on the effect of a loss of outside income on labor supply in a non-credit constrained household. Assuming that leisure is a normal good, we would expect people to be more likely to work when the pensioner leaves the household. This could be manifest in terms of home production or market based work.[3] In this context however, an increase in a member's willingness to participate in market based work must depend on their time available to increase their work hours. For example, if all 30 years olds are already engaged in market related work, then we cannot observe an increase in their labor force participation (LFP). We would thus expect the response to be greatest amongst those groups who have time to work and whose wages are relatively high (amongst those household members not currently working).

The story becomes more complicated when one considers the role of the pensioner within the household. Suppose, for example, that the pensioner looked after household children. Then there exists the possibility that a resident adult has to leave the labor market to assist with child care. On the other hand, if the pensioner is ill and requires care within the household, their departure would free up some other member's time and possibly allow for greater labor market activities.[4]

Economic theories of the family and household formation are also unclear about what

[3]Data limitations preclude me from considering the effects on home production.

[4]This latter scenario seems unlikely given that the amount of time reportedly spent by adults on care giving to the elderly is almost zero. See Ranchhod and Wittenberg (2007) for details.

would happen to household composition. If the pensioner provided child care, we might expect non-resident family adults to take up residence in the household. On the other hand, the household may have to send adult members out of the household to become migrant laborers in other regions, which is consistent with the model by Rosenzweig and Stark (1989). In contrast, the model presented by Ardington *et al* (2007) incorporates liquidity constraints and assumes that migrant laborers initially need to draw resources from the original sending household. In this scenario, the loss of pension income would lead to a decrease in out-migration of adults, and a return of some migrants into the household. A different compositional response could be for the household to send children to live with members of their kin network in other households.[5]

The prediction of the effects on household composition is thus by no means unambiguous. Ultimately, the question remains to be informed by empirical analysis.

2.5 Data, Sample Selection, and Related Issues

2.5.1 Sample selection

The data I use comes from the South African Labour Force Surveys (LFS). These are nationally representative household level surveys that are conducted with a biannual frequency, in March and September of each year. They contain a complete household roster, demographic information such as age, race, gender and education for each respondent, and detailed information on labor force participation, employment, occupation, hours worked and earnings for all household residents aged 16 and above. In some waves there is also basic information about the quality of the household's physical dwelling structure, home ownership, and the relative importance of various forms non-labor related earnings. These latter are household level variables.

I make use of data from wave 4 through wave 9 of the LFS in this paper (i.e. Sept 2001 - March 2004). Since most of the analysis I conducted is at the household level, I collapsed all the relevant information to the household level. Table 2.1 shows the initial sample sizes

[5]Fostering of African children is not uncommon in South Africa. See for example, Beittel (1992) and Sagner and Mtati (1999).

in the cross-sections, and the subsequent sample after each additional restriction discussed below is imposed. Initially, there are 163197 unique household numbers (within waves) across all of the waves combined.

The question on whether a person receives the pension or not is asked only of those who are not currently employed. Since the means test is relatively generous, a non-trivial proportion of the working elderly could also be receiving the pension. Moreover, the LFS is structured to classify a broad range of activities as 'employment', which exacerbates the problem.[6] I therefore decided to make use of the legal age requirements as proxies for pension income, which is consistent with what almost all researchers investigating the effects of the OAP have done. For this reason, I excluded all households which had any household members' age as unknown. I also focus exclusively on African headed households.[7] Africans comprise the majority of the population, are disproportionately poor, and conditional on age-eligibility, are highly likely to be receiving the pension. In all of the September waves combined (i.e. wave 4, 6 & 8), 88.5% of African headed households that included at least one pension-aged member reported that someone in the household receives the old age pension.

From waves 4 to 9, the LFS contained a 20% out-rotation component of dwellings.[8] Thus, theoretically at least, 80% of dwellings were revisited between any two six month periods. The essence of my analysis is to identify households that we observe in two subsequent waves of the LFSs, identify those which had a pensioner in the 'first' wave and 'lost' that pensioner by the next one, and measure the magnitude of other changes that occur in such households as well. In its most basic form, this is simply a 'before and after' comparison.

[6] In September waves, there is a household level module which asks, 'Does any person in this household receive an Old Age Pension?', but this is not present in the subsequent March wave.

[7] Technically, I included households in which the eldest member is an African. Given that the eldest member is generally the household head, and the infrequency with which multi-racial households are observed in the data, this captures the race of the household head accurately in almost all cases.

[8] Source: Statistics South Africa Labour Force Survey metadata documents.

2.5.2 Measurement Error

Measurement Error arising due to false matches

One challenge to the analysis is measurement error arising due to the possibility of false matches. Since identification is based on the idea that a pensioner's departure from a household will lead to additional responses from the remaining household members, it is essential that I do, in fact, observe the same household in each of the two waves. However, this is a dwelling level panel, and is thus not necessarily the same household over time.

To minimize this potential problem, I included only those dwellings where at least one resident member was included in the Statistics South Africa[9] (StatsSA) individual level panel, and has the same race, gender, and similar ages in wave t and $t+1$.[10] StatsSA recently invested considerable resources to extract an individual level panel from this rotating dwelling level panel. The match quality is likely to be good, since they use the dwelling identifier information, demographic characteristics, as well as the confidential first and last name of the respondent to identify the person level matches.

On the other hand, the individual level match rate was far from complete, even after accounting for migration and the rotation pattern.[11] McLaren Z (2007) reports that between 46.67% and 66.40% of individuals in the cross-sections survive into the matched panel. This is considerably less than the expected upper limit of 80%. The author presents evidence that a major reason for the attrition is due to household non-response in a subsequent wave. This is supported by the observation that the distribution of match rates within households is strongly bimodal, with large spikes at zero and one, and a relatively sparse density in between. It is also possible that the dwelling that was 'revisited' was in fact a different physical dwelling to the original one surveyed in the prior wave. In shanty towns in urban areas, and mud huts in rural areas, dwellings could well be impermanent structures.

I excluded dwellings where there was more than one household on the property, since

[9] Statistics South Africa is the official national statistics organization, and is responsible for conducting the LFS.

[10] By 'similar' age I required that the $0 \leq age_{t+1} - age_t \leq 1$.

[11] Approximately 5% of residents in waves 5 - 9 reported that they were not living there 6 months ago.

this represented a greater risk of false matches than single household properties.[12] The impact of each of these criteria on the sample size is shown in Table 2.2. Of African headed households living in single dwelling properties, I match about 76% of households across waves, which is close to the 80% I would expect from the rotation pattern.[13] Restricting the sample further to those households with at least one 'good' individual level match reduces the sample size by roughly one third, from 68,413 to 45,582 households.[14] Finally, I included only households with at least one person who was pension age-eligible in wave t.

I am thus able to identify 12,342 households that had a pension-aged member in wave t, where the 'pensioner'[15] is absent in wave $t+1$ for some subset of these households.

There are also 444 households that show a net increase in the number of resident pensioners by the subsequent wave. These I excluded from the analysis for various reasons. First, the substantive question is about how households cope with the loss of the pensioner and the related cash transfer. The question of the impact of pension receipt has been widely studied already, albeit using cross-sectional data. Second, the date of pension eligibility should be fully anticipated, and so only liquidity constrained households are likely to respond. In contrast, the date of the departure of the pensioner, particularly if due to death, is a stochastic variable. Third, it is difficult to separate between actual aging into the pension and age misreporting, which is discussed below. Finally, the sample size is much smaller, which results in limited statistical power.

Measurement Error arising due to 'Age Heaping'

Of the remaining 11,898 households, there are 10266 'Keeper' households which have no change in the net number of pensioners between wave t and wave $t+1$. There are also

[12] StatsSA indicated that the 'hhid' were maintained by property across waves, but not necessarily for dwellings within properties. Of the households that satisfied every other requirement for inclusion in the sample, this excluded 6.25% of households.

[13] Some of the remaining slippage is likely to occur if a property experiences a change in the number of dwellings between waves.

[14] This is slightly better than the individual match rate, and occurs since I only require one individual in the household to be matched for the household to remain in the sample.

[15] For the remainder of the paper, I use the word 'pensioner' to refer to a person who is age-eligible to receive the old age pension.

1632 'Loser' households, which show a decrease in the net number of pensioners between waves.

A potentially serious measurement error problem arises if people report age imperfectly. Figure 2.1 shows the 'Age Heaping' phenomenon, whereby people tend to round ages to focal points of multiples of 5 or 10. This results in spikes in the observed age distribution in the data, and is particularly pronounced amongst the elderly. Suppose that a person's reported age in an initial wave is such that I would classify them as pension age-eligible. If, in a subsequent wave, a person's age is then reported as below the pensionable age, I would end up classifying the household as a Loser. In reality, however, the economic environment of such a household has remained the same. This is also not of the classical measurement error form, which necessarily results in attenuation bias, since the composition would simultaneously reflect an increase in the number of older adults who are not yet pensioners.

In response, I only included a subset of Loser households where I can be reasonably confident that a pension-aged individual did indeed leave the residence. I classified a household as a 'strict Loser' if:

$$(\text{\# residents aged} \geq ((\text{pension-age})+2)) \text{ in } Wave_t > (\text{\# residents aged} \geq \text{pension-age}) \text{ in } Wave_{t+1}$$

Households with someone reported as having age ≥ ((pension-age)+2) are unlikely to include households that are not actually pensioner households but get incorrectly classified as such due to the age-heaping. On the other hand, households that are actually Keeper households with a pensioner aged ≥ ((pension-age)+2)) are unlikely, even with age-heaping, to report the person's age as being strictly less than the pensionable age in a subsequent wave.[16]

All subsequent analysis in this paper includes only the 1220 'strict Loser' and 10266 'Keeper' households.[17] The final sample, then, has 11,486 households, observed once in the

[16] I did not impose a similar restriction on Keeper households. Preliminary analysis revealed that Keeper households are incredibly stable, such that including only 'strict Keeper' households would only limit the sample size but not impact on the magnitude of the estimated coefficients.

[17] For the remainder of this paper, the term 'Loser' implies a 'strict Loser' household.

'before' period ($Wave_t$) and once in the 'after' period ($Wave_{t+1}$) each. Unless specified otherwise, the unit of observation is thus a household-panel, and some households are included more than once if they meet all the criteria and appear in more than one panel.

2.5.3 Selection Correction on Observables

The fact that the match rate was relatively poor introduces the possibility of selection bias. In order for the analysis to be a valid description of what happens to pensioner households on average, I need to assume that the households that are included are representative of pensioner households in general. This assumption is unlikely to be true.

The first three columns in Table 2.3 show how the households that meet all the other criteria[18] but did not feature in the panel compare to those that were included in the panel.[19] I compare these two groups for a host of composition and activity variables. For most variables, t-tests for differences in the means reject the null hypothesis that those included and those that attrited were drawn from the same underlying population.

To the extent that such attrition arises for observable reasons, we can correct for this by reweighting our matched sub-sample. For example, if shack dwellers are more likely to move and are thus less likely to be matched, we can adjust the weighting of those shack dwellers who we do manage to match. Thus, non-random matching on observables is not an insurmountable problem per se, as we can use the 'inverse probability weighting' (IPW) method to obtain unbiased estimates.(see Wooldridge 2001, pp 587-590).

I estimated probit models and reweighed the panel sample using the IPW method. The probits were estimated separately for each wave. The probits were estimated only using those observations for which I had a corresponding household that I could potentially match to. This is not too problematic, since the objective is purely a statistical rebalancing one - I want the group in the panel to look more like that from the full cross-section.

The variables I included were original household composition and location variables,

[18]That is, they had a pensioner, the eldest person was African, there were no observations with age missing, and there was only one dwelling on the property.

[19]All pensioner households from the panel were included in this process, including 'Gainers' and Losers who were not 'strict Losers'.

employment data of various demographic groups within the household, and information about the ownership and characteristics of the physical home. Since the 'wall type', 'home ownership' and 'dwelling type' questions are only asked in the September waves, for waves 5 & 7 I used the information from the matched household in the subsequent wave.

The regression results suggest that the panel over-represents larger households, urban households, as well as households whose residents owned their home, none of which is surprising.[20] I then predicted the probability of inclusion in the panel, and all results in the analysis are weighted by the inverse of this probability, multiplied by the relevant sampling weights.

A glance at columns III to VI of Table 2.3 suggests that the process was reasonably successful at achieving its objective of re-balancing the panel to look like the cross-section.

2.5.4 Non-random Selection on Unobservables

A more difficult potential problem occurs if we have non-random matching based on unobservable characteristics, which persists even after the selection correction on observable characteristics. If these characteristics are orthogonal to the variables we are interested in, the estimates will still be unbiased in expectation. If, however, an entire household migrates in search of better economic opportunities upon the death of a pensioner, I cannot control or adjust for this. I thus need to qualify my findings to those Loser households where at least one member stays in the same residence.[21]

2.6 Summary statistics

2.6.1 Dependent Variables

Table 2.4 presents mean household characteristics for Keeper and Loser households in the initial $Wave_t$ period, for each of the major dependent variables that I consider. Columns 1 & 2 shows the mean household composition for each age-group. The age classification was somewhat arbitrary, with 'kids young' being aged 7 or lower, 'kids school'

[20] These are omitted for brevity, but are available from the author upon request.
[21] It is impossible to determine whether this is a large or small problem in this context.

aged 8 - 15, 'youth' aged 16 - 20, 'young adults' aged 21 - 35, 'middle adults' aged 36 - 50, and 'older adults' ages 51 - 59 if female, and 51 - 64 if male. [22]

Columns 3 & 4, 5 & 6, and 7 & 8 show the mean proportion of people within these age groups that are working, in the labor force using the broad definition, and in the labor force using the narrow definition, respectively. In the 'broad' category is included anyone currently employed or willing to work. In the 'narrow' category are the employed, and only those unemployed who are willing to work and had been actively searching for employment in the past month.

To begin with, Loser households are considerably larger, by more than 1 person on average. The differences are most pronounced for the younger age groups, up to and including the young adults. For most groups, the proportion in each of the LFP categories are somewhat similar. The exceptional category is the middle-aged adult males in the soon to be Loser households, who are about 10% points more likely to be in the labor force using either definition, and 8% points more likely to be employed. Also of note is that employment rates for each age-group are relatively low, for both Keepers and Losers. Fewer than 1 in 5 young adults, and 1 in 3 middle and older adults are working, in both Keeper and Loser households.

2.6.2 Income Sources

The survey instrument only captures non-labor income in a crude fashion by asking "What is the main source of income for this household?". One possible response is "remittances". It should be stressed that there is no information on the value of remittances. Moreover, the question is only asked in waves 4, 6, and 8 (i.e. the September waves). I calculate the distribution of the responses for the Keeper and Loser households in the relevant panels. In order for this comparison to be valid, one needs to believe that Loser (and Keeper) households in waves 5 and 7, were similar to Loser (and Keeper) households

[22] At age 7, children should legally be enrolled at school, but enrollment become almost universal by age 8 only. Similarly, 16 is the legal age at which a person may drop out of school or enter employment, while at 21, a person becomes a legal adult.

in waves 4 and 6.[23] For these reasons, the summary statistics presented in Table 2.5 are only suggestive.

The Keeper households remain fairly stable, which lends credibility to the aforementioned assumption. About 78% of Keeper households in both time periods report 'Pensions and Grants' as their main income source.[24] Loser households look different from Keepers even in the period prior to their loss. Losers in $Wave_t$ are more likely to report 'Salaries and Wages' as their main income source (26.3% vs. 15.6%). This would be expected if people are anticipating the coming departure of the pensioner.

In $Wave_{t+1}$, this distribution changes remarkably in Loser households. Pensions and grants decreases as the main income source from 61.7% in the 'before' period, to 35.0% in the 'after' period. This is accounted for mostly be a massive increase in the proportion that report remittances as their main income source, which increases from 7.3% to 26.0%. Almost 1 in 5 Loser households experience this transition. Somewhat smaller changes are observed in the fraction of Losers that report salaries and wages as their main income, which increases from 26.3% to 31.0%. Given this, and previous research by Jensen (2003), it seems plausible that there might be offsetting increases in remittances in Loser households.

Changes in both household composition and labor force participation amongst residents may be muted if there is an offsetting increase in remittances to the household to compensate for the loss of pension income. The observed changes in the distribution of main income sources lends credibility to the subsequent analysis and interpretation; that Loser households are indeed experiencing changes that correlate with the loss of pension income.

2.7 Empirical Specification

I next employ multivariate regression techniques to control for additional factors, and test for the statistical significance of changes in household composition and labor force participation. I regress the difference in the 'dependent variable' for households between

[23] i.e. Losers at time T_0 in waves 5 and 7 were similar to Losers at time T_0 in waves 4, 6 and 8.

[24] These are simply weighted means. South Africa also has several other grants, the most common being the Child Support Grant.

$Wave_t$ and $Wave_{t+1}$, on an indicator for whether the household was a Keeper or a Loser. The regression that I fit is of the form:

$$D_{j,t+1} - D_{j,t} = \beta_0 + \beta_1 losepen_{j,t} + \beta_2 X_{j,t} + \epsilon_{j,t}$$

where j, t denotes a household j in $Wave_t$, and D is the dependent variable of interest.

'$losepen_{j,t}$' is an indicator variable that equals one if household j is a strict Loser between $Wave_t$ and $Wave_{t+1}$, and 0 if household j is a Keeper between $Wave_t$ and $Wave_{t+1}$. Additional X variables include an indicator variable for urban areas, provincial dummy variables, wave dummies, household size [25] and a count variable for the number of pensioners in the household in the initial period. I include this last one since losing one of two pensioners potentially has smaller effects than losing the only pensioner in the household.

This specification nets out any unobservable but time invariant characteristics that are specific to a particular household. Moreover, by comparing the change in Loser households relative to Keeper households, I also expect to net out any effects that arise due to the aging of the underlying population, as well as changes in economic conditions that affect members of both groups of households equally.

One mis-specification of the above regression is that I am implicitly assuming that the households are independent across panels. However, with the rotation policy discussed, this cannot be true. To correct for this, I estimate robust standard errors which are clustered at the household level.

2.8 Changes in composition and aggregate household labor force participation

The coefficient on the *losepen* variable is presented for each of the dependent variables in Table 2.6. Each coefficient and corresponding standard error is obtained from a separate regression. The coefficients measure the difference in the mean changes in the dependent variables between Keepers and Losers, after controlling for all the other X variables. As stated previously, exploratory analysis revealed that the Keeper households are incredibly stable, so the coefficients reported are identified primarily using variation within the Loser

[25] Household size was not included in the regression where the dependent variable was the difference in household size itself.

households.

I observe large and significant changes in household composition. Not surprisingly, aggregate household size in Loser households goes down. There is a reduction in the number of school aged children in the household of 0.063, and an inflow of 0.052 middle aged females. The largest change in composition occurs amongst the older adults, with a net increase of 0.279. Thus, more than 1 in 4 losing households get an additional older adult on average. This in-migration is approximately equally comprised of men and women.

In terms of the numbers employed, almost all the groups considered experience a significant increase in the number of employed persons in that group.[26] The largest of these is experienced by the older adults, with a coefficient of 0.102. These are comprised equally of men and women.

With respect to the number in the labor force, only for the older adults is the increase significant. This is also the only group for whom the increase in number employed is exceeded by the increase in the number in the labor force. The increase in number employed is less than the increase in the number of residents for the adult groups that show a significant change in composition. Thus, there is also a considerable amount of more adult time for home production activities in Loser households.

In sum, I find significant evidence that the household re-organizes itself in conjunction with the departure of a pensioner. Aggregate household level labor supply increases primarily amongst the older adults, while the number of employed adults increases for most categories. That said, the change in composition makes it difficult to identify an important policy concern. Are the changes in labor supply (employment) arising due to an increase in the labor supply (employment) of household members who were previously not in the labor force (employed), or are they the result of the in-migration of people who were already in the labor force (employed) and maintained their status. I partially inform this question in the next two sections.

[26] The exception, once again, are the middle-aged men.

2.9 Changes in labor force participation conditional on group level compositional stability

Regression results presented in Table 2.7 are estimated using the same specification as before, but using a restricted estimation sample. Specifically, only Loser households in which the net number of residents in the relevant age-group (or gender specific age-group) remained unchanged across the two waves were included. All Keeper households were included in the regressions. In these regressions, estimated coefficients are almost surely the result of behavioral responses amongst household members who were already residing in the household prior to the departure of the pensioner.[27] At the same time, since the results are obtained off a selected subset of Loser households, one should be cautious about generalizing these results to the entire set of Losers.

Table 2.7 presents the coefficient and standard error for each group and each dimension of LFP. I then divide the coefficient estimate by the mean number of Loser households included in the regression, to measure the size of the coefficient relative to the average size of the underlying population in the relevant Loser households. Youth seem to be unaffected by the departure of the pensioner. Young adults are significantly affected, with each household getting on average 0.059 more active searchers and 0.047 more employed young adults. In proportionate terms, these are fairly large. The probability of a young adult in such a Loser household finding employment increases by 5.6% points, and the increases are larger in terms of labor supply. Of these young adults, we see large proportionate increases for both genders, but the females seem to be considerably more successful at finding employment. Indeed, female employment increases by a statistically significant 0.029 (or 6.5% points), whereas the corresponding coefficient for males is 0.009 (or 2.4% points) and is not significant at any reasonable level.

Amongst middle-aged adults, there are large and significant employment effects, but

[27] It remains mathematically possible that there is a circular and perfectly offsetting migratory pattern which could explain these results. However, if I were to impose the additional constraint that the number in the group in $Wave_t$ exactly equals the number in the group who report being resident in the household six months prior to the survey in $Wave_{t+1}$, this would reduce the set of Loser households by at most 8 observations in any of the regressions, a further reduction in the number of Losers of at most 1.5%.

insignificant, smaller and negative effects for labor supply. There are 0.025 and 0.014 more middle-aged female and male workers respectively, which represents an increase in the proportion working of 9.0% points and 9.6% points respectively.[28] These are particularly large proportions. It is also interesting to observe that this is not met with a corresponding increase in labor supply. The most plausible explanation is that people who were not employed but in the labor force found employment between waves.

The older adults, particularly the women, seem to be the most responsive in terms of labor supply.[29] Each household gets a significant 0.016 and 0.02 more older women who are employed or in the labor force using the narrow definition. These correspond to increases of 14.5% points and 18.1% points respectively.

2.10 Changes in labor force participation conditional on maintained individual residency

An alternative in similar vein is to analyze LFP responses at the individual level, using observations who are matched in the StatsSA panel. Advantages of this method are that we can effectively difference out person specific yet time invariant unobservable characteristics, interpretation is somewhat simpler, and we are not requiring that the entire demographic group in the household remain unchanged. A disadvantage is that some people who were indeed resident in both waves get omitted, since the match rate in not perfect, as discussed earlier. It is probable that both methods are informative.

Table 2.8 presents individual level regression results for the same set of dependent variables. The estimation sample is the set of matched individuals in the Loser and Keeper households already identified. The exact specification is:

$D_{i,g,j,t+1}$ - $D_{i,g,j,t} = \beta_0 + \beta_1 losepen_{j,t} + \beta_2 X_{j,t} + \epsilon_{i,g,j,t}$

where i, g, j, t denotes individual i in demographic group g in household j in $Wave_t$, and D is the dependent variable of interest. Additional X variables include an indicator variable

[28]The coefficient for the group with both genders does not need to be a convex combination of the male and female only coefficients, as the households that are included are not necessarily the same.

[29]The coefficients for men are not significant, but the % point increase is also fairly large.

for urban areas, provincial dummy variables, wave dummies and the number of pensioners in the initial $Wave_t$ period. Group status was determined by age in the $Wave_t$ period.

Youth and young adults generally show small and statistically insignificant effects on labor supply. However, there is a 5.7 percentage point decrease in the probability that youth are in the labor force using the broad definition, and a 7.4 percentage point decrease in the probability that young female adults are actively searching for employment. This is in stark contrast to the corresponding estimate in Table 2.7, which showed a 7.3 percentage point increase in said probability, although that estimate was not statistically significant. On aggregate, the conflicting results suggests that caution is warranted in making strong conclusions regarding the effects on youth and young adults.

The results for the middle aged adults, both male and female, are indeed similar to those obtained using the alternative analysis. We see no significant effects on labor supply, but large and positive increases in the probability of employment. Middle aged females and males in Loser households are 9.3 and 8.1 percentage points more likely to be employed respectively, relative to the comparison group in Keeper households.

Again, the largest results are obtained for the group of older adults. Older adults of either gender are 10.3 percentage points more likely to be employed, and the increase is statistically significant at the 5% level for the group with both genders combined. These are approximately matched by increases in labor supply, which are also large in magnitude. The increases are larger for women than for men, 12.2 percentage points vs. 10.6 percentage points using the broad definition, and 13.1 percentage points vs. 8.3 percentage points using the narrow definition. The coefficients for women are also statistically significant at the 5% level, which is not the case for men.

In summation then, for these selected observations, I find significant and large increases in employment rates for middle aged and older adult groups. This is accompanied by increases in labor supply for the older adults, which is not suprizing since these groups tend to have relatively low LFP rates to begin with. For the middle-aged adults, there is no corresponding increase in labor supply on average. Overall, the finding is all the more striking given that South Africa's unemployment rate is about 30% amongst adults, that most of the unemployed seem to experience chronic long term unemployment[30] and that this occurs within the relatively short space of time between waves.

2.11 Caveats and Robustness checks

It is important to stress that the correlations presented cannot be interpreted as causal estimates. For example, the out-migration of a pensioner might be a consequence of the in-migration of other household members, or a change in their employment status. Alternatively, there may be other factors that simultaneously change the pensioners choice of residence as well as that of other household members. There are other limitations, partly due to the data available. I cannot observe why the pensioner left or where she went to. I also have no information regarding where the new household members came from, where the out-migrating children go to, nor the activities of any non-resident members. There is also only limited information about remittances and resource sharing within families but across physical households. A complete analysis would be able to observe all of these in order to gain a full understanding of the effects of the pensioner's departure.

There is one case, however, where the departure of the pensioner is plausibly exogenous, namely the death of the pensioner. This is still not a panacea, for the family may anticipate the death of the pensioner and start rearranging the family prior to his death. In this case, I would be biased away from finding any results, which implies that my 'death' estimates are biased towards zero, and should thus be interpreted as lower bounds of the true effect.

2.11.1 Identification using deaths

One might be concerned about endogenous out-migration of the pensioner. As a robustness check, I further restricted the sample to include only Losers who experience the plausibly exogenous event of the death of the pensioner.[31] However, only in the wave 5

[30] See Kingdon and Knight (2002), Banerjee *et al* (2006)

[31] Ardington *et al* note that in the dataset that they used, 77% of Loser households lost their pensioner due to death.

(March 2003) module were respondents asked about recent deaths in the household. I use this data to generate an indicator variable for whether an elderly member died recently in a Loser household.[32] This variable is called '$Death1$', which includes 47 Loser observations.

In the remaining waves, I can only infer deaths indirectly and probabilistically. I do so using the marital status variable, in combination with the question on who the persons' spouse is. To do this, I used the 'good' individual level matches from the StatsSA panel, and identified who was married to a pensioner. This is only possible for the subset that were married in $Wave_t$ and lived with their spouse at the time. I then infer death by identifying those who transitioned to become a widow or widower in $Wave_{t+1}$. I classify the variable '$Death2$'=1 if the above criteria are satisfied in a Loser household. This yields a subset of Losers who lost a pensioner through death. The '$Death2$' sub-sample has 60 observations.

Regression results for the $Death1$ and $Death2$ samples are presented in Tables 2.9 and 2.10 respectively. All Keeper households were included in the regressions. The samples are necessarily smaller, with fewer statistically significant effects. The 'treatment' is also different. In the case of $Death2$, I am also selecting on marital status and co-residency of spouses, which might also have a bearing on the coefficient estimates.

From Table 2.9, there is some evidence that there is an inflow of middle aged adults, but these are not significant. We do observe a statistically significant increase in the number of older adults, of 0.151 persons. In general, all the LFP coefficients for the number of middle and older adults are positive, but are usually not significant. There are marginally significant and positive LFP coefficients for the numbers of young female adults, middle aged adults and middle adults males. The category which clearly experiences some change are the older adults, where there is a marginally significant increase in the number of employed older adults of 0.065, and a significant increase in the number in the labor force of about 0.09.

[32]The 'age at death' variable in the 'deaths' file in LFS 5 is corrupted, in that the last digit of the variable is missing. This implies that I only observe the age at death in 10 year intervals. I included all deaths where the age at death was non-missing and greater than or equal to 60.

Table 2.10 presents the same regressions results using the spouse-widow identification of deaths. In this case, very few of the coefficients are significant. We do still observe that the coefficients for middle aged adults and older adults seem broadly consistent with the previous estimates. There are more middle aged adults in the household and more in each category of LFP. For the middle-aged women, there is a marginally significant increase in the number in the labor force using the broad definition of 0.121. The strongest results are once again observed for the older adults. The coefficients are all positive, and there is a marginally significant increase in the number of older adults of both genders combined. The is also a significant increase in the number of older adults either working or actively searching for employment.

On aggregate, the death results lends support to a causal interpretation of our aggregate results. The results, broadly speaking, were similar to those observed for the full sample in Table 2.6. Despite the small number of observations, we observed positive and significant effects on both residency patterns and the numbers in the labor force for the older adult group in particular.

2.12 Discussion

At this point, it is worthwhile to place these findings in the context of the broader OAP literature. In terms of compositional changes, the results seem consistent with the papers by Edmonds *et al* (2006), who analyze only pension receipt, and Ardington *et al* (2007), who analyze both pension receipt and loss.

A bit more effort is required to reconcile the various findings regarding labor supply. Bertrand *et al* (2003) found that the pension reduces the labor supply of prime aged individuals in three generational households, using cross-sectional data. Ardington *et al* (2007) find quite the opposite using longitudinal data on both resident and non-resident family members. By considering the effects on both labor migrants as well as residents, and controlling for person specific unobservable characteristics, they find that prime aged household members are significantly more likely to be employed following pension gain. When considering the loss of a pensioner, and restricting to members of either gender who were resident in both waves of their study, they find that prime aged residents were 1.1 percentage points less likely to be employed, although the estimate was not statistically significantly different from zero.

In contrast, this study finds large and significant increases in employment probabilities for middle-aged adults, and even larger increases in both labor supply and employment rates amongst older adults.

There are various possible explanations for the difference. First, this paper is limited by the data to analyzing the effects on people who are resident in the household. I can make no inference on changes that are manifest for those who out-migrate, in-migrate or are somehow attached to the household but not resident in either wave.

Second, the age groups considered differ in a way that is likely to be important. Both Bertrand *et al* and Ardington *et al* focus on 'prime-aged' adults, which they classify as individuals aged (16 to 50) and (18 to 50) respectively. This effectively ignores the 'older adult' category, which in this paper seems to be the most sensitive in terms of labor supply. In addition, pooling together the groups of youths, young adults and middle aged adults effectively assumes that the responses amongst these groups are the same. In this study, pooling would lead to an estimate that would be some convex combination of the effects for youth, young adults and middle aged adults, which would reduce the estimate on employment for the middle adults and raise it for the other two groups. However, given that South African pensioners live with considerably more youth and young adults than middle aged adults on average, the pooled estimate would be much closer to that of the younger age groups, which was small in magnitude and not statistically different from zero.

A third point of departure from Ardington *et al* is in the geographic scope of the study. Whereas this study is based on limited but nationally representative data, theirs is based on incredibly detailed data obtained from a rural and extremely poor part of the country.

It is certainly possible that regional differences in local economic conditions might result in different responses.

To investigate the possibility, I replicated the individual level analysis for the subset of Keepers and Losers in the rural parts of the KwaZulu Natal province.[33] Table 2.11 presents these estimates. For the first time, we find that youth are significantly more likely to be employed, by 3.4 percentage points. With the exception of female young adults, none of the LFP categories for young and middle aged adults are significant even at the 10% level.[34] The coefficient estimates for employment are also much smaller here than for the national sample for middle aged adults, at 0.028 vs. 0.088 for middle aged adults of either gender, 0.031 vs. 0.093 for middle aged females, and -0.032 vs. 0.081 for middle aged males. The geographic restriction thus results in estimates much closer to those of Ardington *et al.* Despite the very small sample sizes, we do still find large and statistically significant increases in labor supply and employment amongst the older adult females.

2.13 Conclusion

How do poorer households adapt in response to the loss of a valuable economic member? The results presented were consistent with most of the prior empirical literature. Household composition and household labor supply both adjust, with an outflow of dependents and an increase in the number of potentially valuable economic contributors. There is some evidence that the relative importance of remittances increase as well.

Conditional on compositional stability or maintained residency within demographic groups, I find large and significant increases in labor force participation and employment amongst older adults of either gender. I also find large increases in employment rates amongst middle aged men and women, but no corresponding increase in their labor supply. While the proportionate increases are large, the base population within these households are relatively small, and so the average number of people within households that find

[33]The LFS data does not easily allow for more specific areas of analysis, and even if possible, I would likely run in even more serious problems of small sample sizes.

[34]That said, the lack of statistical significance might be due to considerably smaller sample sizes.

employment is also small.

On the other hand, there are over 2.1 million Old Age Pension recipients in South Africa, 1.3 million Disability Grant recipients and almost 6.9 million Child Support Grant recipients.[35] Relatively small labor supply and employment elasticities may have a considerable bearing on national employment levels. Policy makers need to consider the effects of cash grants on the labor supply of non-recipients, while simultaneously being aware that there are significant other positive outcomes that arise from such grants.

[35] See Pauw and Mncube (2007) for details.

2.14 Tables

Table 2.1: **Sample Sizes - Cross Sections**

Wave	# HHID	+ No Age missing	+ African 'headed'	+ only 1 HH on property
4	27,356	27,253	21,138	17,549
5	29,011	28,931	22,093	17,775
6	26,474	26,393	20,073	18,378
7	26,702	26,653	20,282	18,301
8	26,825	26,792	20,373	18,347
9	26,829	26,791	20,397	18,312
Total	163,197	162,813	124,356	108,662

Table 2.2: **Sample Sizes - Matched HH Data**

Panel	# Matched HHs	& $\geq$ 1 'good' indiv. Match	& $\geq$ 1 $pens_t$	Gainers	Keepers	Losers	Losers (strict)
4-5	12,634	8,947	2,395	91	1,967	337	236
5-6	14,143	9,560	2,672	81	2,239	352	268
6-7	14,380	9,282	2,524	94	2,097	333	254
7-8	13,554	8,887	2,385	87	1,973	325	242
8-9	13,702	8,906	2,366	91	1,990	285	220
Total	68,413	45,582	12,342	444	10,266	1,632	1,220

Notes:
1. Sample restricted to African headed households with only 1 dwelling per property, in both periods.
2. 'Strict' definition of a Loser household:
(# residents aged $\geq$ ((pension-age)+2)) at $Time_t$ > (# residents aged $\geq$ pension-age) at $Time_{t+1}$

Table 2.3: **Selection (on observables): Means in $Wave_t$**

Col	I	II	III	IV	V	VI
Variable	X-sect only	Panel	Diff (II-I)	Full X-Section	Panel - Reweighted	Diff (V-IV)
urban	0.372	0.379	0.007	0.376	0.373	-0.003
hhsize	4.915	5.647	0.732***	5.350	5.401	0.051
# kids young (0-7)	0.738	0.886	0.148***	0.826	0.830	0.004
# kids school (8-15)	0.971	1.186	0.215***	1.099	1.120	0.021
# youth (16 - 20)	0.525	0.633	0.108***	0.589	0.599	0.010
# young adults (21 - 35)	0.910	1.076	0.166***	1.009	1.022	0.014
# middle adults (36 - 50)	0.419	0.480	0.062***	0.455	0.456	0.000
# older adults (51 - pension age)	0.193	0.208	0.014**	0.202	0.202	0.000
# pension aged	1.177	1.197	0.020***	1.189	1.191	0.002
# young adults work	0.190	0.213	0.023***	0.203	0.205	0.001
# middle adults work	0.137	0.152	0.015***	0.146	0.144	-0.002
# older adults work	0.052	0.056	0.004	0.055	0.056	0.001
# young adults in LF (broad)	0.741	0.885	0.144***	0.827	0.841	0.014
# middle adults in LF (broad)	0.321	0.367	0.045***	0.348	0.347	-0.001
# older adults in LF (broad)	0.088	0.094	0.006	0.092	0.092	0.000
# young adults in LF (narrow)	0.517	0.602	0.085***	0.567	0.575	0.007
# middle adults (narrow)	0.244	0.278	0.033***	0.264	0.263	-0.001
# older adults (narrow)	0.070	0.074	0.005	0.072	0.073	0.000

Notes:
1. Sample is all African headed households, with a single dwelling on the property, with no member's age missing, and at least one 'pensioner' in the household
2. Data corresponds to $Wave_t$ - i.e. from Waves 4 - 8
3. Means are unweighted, except in column V
4. The 'single dwelling' requirement excludes 6.25% of the sample, when all the other constraints are satisfied
5. * denotes statistical significance at the 5% level, ** denotes the same at the 1% level
6. The pension age is 60 or above for women, and 65 or above for men

Table 2.4: **Summary statistics in $Wave_t$ (Mean, and ratio of means)**

	Composition number		# Work proportion		# in LF (broad) proportion		# in LF (nar) proportion	
	Keep	Lose	Keep	Lose	Keep	Lose	Keep	Lose
HH size	5.36	6.39						
kids young	0.86	0.97						
kids school	1.11	1.32						
youth	0.59	0.77	0.035	0.023	0.223	0.240	0.112	0.103
young adults	1.02	1.20	0.193	0.185	0.822	0.816	0.551	0.548
young adult:F	0.55	0.64	0.171	0.155	0.812	0.794	0.507	0.501
young adult:M	0.47	0.56	0.219	0.219	0.834	0.842	0.603	0.601
mid-aged adults	0.46	0.58	0.312	0.326	0.764	0.779	0.578	0.588
mid-aged adult:F	0.25	0.35	0.326	0.291	0.766	0.715	0.570	0.520
mid-aged adult:M	0.21	0.23	0.296	0.378	0.761	0.871	0.588	0.688
older adults	0.17	0.25	0.274	0.317	0.485	0.512	0.372	0.397
older adult:F	0.09	0.13	0.264	0.310	0.445	0.454	0.339	0.349
older adult:M	0.08	0.13	0.286	0.325	0.530	0.571	0.410	0.445
# pens age	1.17	1.33						
N	10,266	1,220						

Notes:
1. Means are weighted by [pweight x IPWeight]
2. The 'proportions' are the ratio of the mean number in a particular labor market category and demographic group, to the mean number in that demographic group.
3. Age-groups: kids young (0-7), kids school (8-15), youth (16-20), young adults (21-35), middle adults (36-50), older adults (51 - pension-age)

Table 2.5: **Summary statistics: Main Income Source**

Distribution of Main Income Source in Household (%)

	T_0		T_1	
	Keeper	**Loser**	**Keeper**	**Loser**
Salaries and/or wages	15.6	26.3	15.4	31.0
Remittances	4.5	7.3	4.2	26.0
Pensions and grants	77.8	61.7	77.9	35.0
Sales of farm product	0.4	0.5	0.4	1.3
Other non-farm income	1.5	2.8	1.8	5.4
no income	0.2	1.3	0.2	1.4
Unspecified	0.0	0.1	0.1	0.0
N	6,054	710	4212	510

Notes:
1. The T_0 data relates to observations in Panels 4-5, 6-7 and 8-9, the T_1 data relates to observations in Panels 5-6 & 7-8
2. Means are weighted by [pweight x IPWeight]

Table 2.6: **Regression Results: Composition and Activity (in # of people)**

	Composition		# Work		# in LF (broad)		# in LF (nar)	
Outcome variable	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
△ HH size	-0.86	[0.079]***						
△ # kids young	0.031	[0.033]						
△ # kids school	-0.063	[0.032]**						
△ # youth	-0.007	[0.026]	0.008	[0.007]	-0.027	[0.020]	0.007	[0.014]
△ # young adults	0.039	[0.034]	0.056	[0.019]***	0.04	[0.032]	0.055	[0.035]
△ # young adult:F	0.025	[0.025]	0.032	[0.013]**	0.025	[0.025]	0.031	[0.024]
△ # young adult:M	0.014	[0.025]	0.024	[0.013]*	0.015	[0.023]	0.023	[0.023]
△ # mid-adults	0.063	[0.026]**	0.05	[0.018]***	0.01	[0.025]	0.03	[0.025]
△ # mid-adult:F	0.052	[0.019]***	0.035	[0.011]***	0.021	[0.018]	0.027	[0.016]
△ # mid-adult:M	0.011	[0.016]	0.015	[0.012]	-0.012	[0.016]	0.003	[0.014]
△ # older adults	0.279	[0.022]***	0.102	[0.015]***	0.118	[0.017]***	0.118	[0.016]***
△ # older adult:F	0.144	[0.015]***	0.051	[0.010]***	0.06	[0.011]***	0.058	[0.011]***
△ # older adult:M	0.135	[0.016]***	0.051	[0.011]***	0.059	[0.013]***	0.06	[0.012]***
△ # pens age	-1.064	[0.007]***						

Notes:
1. Robust Std. Errors, clustered at the 'hhid' level are reported
2. Omitted coefficients on variables included in the regression for variables:
Province dummies, wave dummies, urban dummy and no. of pensioners in the '$Wave_t$' period.
3. With the exception of the regression on household size, all the other regressions also controlled for initial household size.
4. N = 11486 in each of the regressions
5. * denotes statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level
6. Each coefficient and corresponding standard error is obtained from a separate regression.
7. Age-groups: kids young (0-7), kids school (8-15), youth (16-20), young adults (21-35), middle adults (36-50), older adults (51 - pension-age)

Table 2.7: **Regression results: △ Activity conditional on within group stability**

Group	Outcome Variable	Losing Obs	Coef.	Std. Err.	Mean # in HH	% pt. △
Youth	Work	769	0.008	[0.007]	0.540	1.5
	In LF broad	769	-0.006	[0.015]	0.540	-1.1
	In LF Narrow	769	0.004	[0.012]	0.540	0.7
Young Adult	Work	601	0.047	[0.020]**	0.841	5.6
	In LF broad	601	0.055	[0.020]***	0.841	6.5
	In LF Narrow	601	0.059	[0.030]**	0.841	7.0
Young Adult:F	Work	793	0.029	[0.013]**	0.452	6.4
	In LF broad	793	0.004	[0.015]	0.452	0.9
	In LF Narrow	793	0.033	[0.022]	0.452	7.3
Young Adult:M	Work	819	0.009	[0.012]	0.375	2.4
	In LF broad	819	0.027	[0.011]**	0.375	7.2
	In LF Narrow	819	0.032	[0.015]**	0.375	8.5
Middle Adult	Work	800	0.03	[0.016]*	0.446	6.7
	In LF broad	800	-0.021	[0.014]	0.446	-4.7
	In LF Narrow	800	-0.018	[0.015]	0.446	-4.0
Middle Adult:F	Work	922	0.025	[0.010]**	0.277	9.0
	In LF broad	922	-0.006	[0.011]	0.277	-2.2
	In LF Narrow	922	-0.002	[0.011]	0.277	-0.7
Middle Adult:M	Work	985	0.014	[0.008]*	0.145	9.6
	In LF broad	985	-0.005	[0.007]	0.145	-3.4
	In LF Narrow	985	-0.003	[0.007]	0.145	-2.1
Older Adult	Work	796	0.015	[0.010]	0.193	7.8
	In LF broad	796	0.025	[0.011]**	0.193	13.0
	In LF Narrow	796	0.022	[0.011]**	0.193	11.4
Older Adult:F	Work	958	0.016	[0.007]**	0.110	14.5
	In LF broad	958	0.016	[0.008]*	0.110	14.5
	In LF Narrow	958	0.02	[0.008]**	0.110	18.1
Older Adult:M	Work	995	0.007	[0.006]	0.106	6.6
	In LF broad	995	0.005	[0.006]	0.106	4.7
	In LF Narrow	995	0.006	[0.006]	0.106	5.7

1. Outcome variables in units of △ in '(# of people')
2. Reported coefficient corresponds to dependent variable *losepen*
3. Robust Std. Errors, clustered at the 'hhid' level are reported
4. Omitted coefficients on variables included in the regression for variables:
Province dummies, wave dummies, urban dummy, initial household size
and no. of pensioners in the $Wave_t$ period.
5. N = 10266 Keepers + # Losing Observations in each of the regressions
6. * denotes statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level
7. '% pt. △' column is obtained by: 100*coefficient/(Mean # in HH)
8. The Mean # in HH contains information only from Loser HHs included in the regression,
and observations are weighted by [pweight x IPWeight]
9. Each coefficient and corresponding standard error is obtained from a separate regression.
10. Age-groups: kids young (0-7), kids school (8-15), youth (16-20), young adults (21-35),
middle adults (36-50), older adults (51 - pension-age)

Table 2.8: **Individual level regressions for changes in Labor Market status**

		Dependent variable		
Group	N	△ work	△ in LF br	△ in LF nar
youth	4391	0.015	-0.057	0.005
		[0.011]	[0.026]**	[0.021]
young adults	6856	0.002	-0.027	-0.035
		[0.021]	[0.019]	[0.027]
young adult: F	3765	-0.012	-0.021	-0.074
		[0.027]	[0.029]	[0.036]**
young adult: M	3091	0.016	-0.032	0.007
		[0.030]	[0.024]	[0.035]
mid-aged adults	3316	0.088	0.016	0.039
		[0.030]***	[0.031]	[0.031]
mid-aged adult: F	1850	0.093	0.025	0.025
		[0.037]**	[0.042]	[0.043]
mid-aged adult: M	1466	0.081	0.000	0.058
		[0.041]**	[0.041]	[0.040]
older adults	1537	0.103	0.111	0.107
		[0.043]**	[0.043]**	[0.044]**
older adult: F	856	0.103	0.122	0.131
		[0.059]*	[0.062]**	[0.059]**
older adult: M	681	0.103	0.106	0.083
		[0.053]*	[0.055]*	[0.059]

Notes:
1. Outcome variables in units of △ in labor market status, values of -1, 0 & 1 in data.
2. Reported coefficient corresponds to dependent variable *losepen*
3. Robust Std. Errors, clustered at the 'hhid' level are reported
4. Omitted coefficients on variables included in the regression for variables:
Province dummies, wave dummies, urban dummy and no. of pensioners in the $Wave_t$ period.
5. * denotes statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level
6. Observations are weighted by [pweight x IPWeight]
7. Each coefficient and corresponding standard error is obtained from a separate regression.
8. Age-groups: youth (16-20), young adults (21-35), middle adults (36-50), older adults (51 - pension-age)

Table 2.9: **Regression results using 'Deaths1' for identification: Composition and Activity (in number of people)**

	Composition		# Work		# in LF (broad)		# in LF (nar)	
Outcome variable	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
△ HH size	-0.91	[0.261]***						
△ # kids young	-0.017	[0.123]						
△ # kids school	-0.02	[0.093]						
△ # youth	0.003	[0.119]	-0.007	[0.015]	-0.026	[0.071]	-0.109	[0.075]
△ # young adults	-0.047	[0.150]	0.073	[0.089]	-0.168	[0.148]	0.001	[0.153]
△ # young adult:F	0.114	[0.093]	0.112	[0.069]	-0.033	[0.085]	0.161	[0.093]*
△ # young adult:M	-0.161	[0.112]	-0.04	[0.065]	-0.135	[0.105]	-0.16	[0.108]
△ # mid-adults	0.132	[0.108]	0.145	[0.111]	0.2	[0.114]*	0.153	[0.113]
△ # mid-adult:F	0.066	[0.081]	0.058	[0.056]	0.089	[0.082]	0.073	[0.088]
△ # mid-adult:M	0.065	[0.078]	0.087	[0.064]	0.111	[0.063]*	0.08	[0.080]
△ # older adults	0.151	[0.066]**	0.065	[0.037]*	0.094	[0.043]**	0.089	[0.044]**
△ # older adult:F	0.092	[0.063]	0.037	[0.028]	0.036	[0.028]	0.044	[0.031]
△ # older adult:M	0.058	[0.036]	0.028	[0.025]	0.058	[0.033]*	0.045	[0.031]
△ # pens age	-1	[0.000]***						

Notes:
1. Death 1 is obtained from the Deaths file in wave 5
2. Robust Std. Errors, clustered at the 'hhid' level are reported
3. Omitted coefficients on variables included in the regression for variables:
Province dummies, wave dummies, urban dummy and no. of pensioners in the '$Wave_t$' period.
4. With the exception of the regression on household size, all the other
regressions also controlled for initial household size.
5. N = 10313 in each of the regressions, 47 Losers, 10266 Keeper
6. * denotes statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level
7. Each coefficient and corresponding standard error is obtained from a separate regression.
8. Age-groups: kids young (0-7), kids school (8-15), youth (16-20), young adults (21-35),
middle adults (36-50), older adults (51 - pension-age)

Table 2.10: **Regression results using 'Deaths2' for identification: Composition and Activity (in number of people)**

	Composition		# Work		# in LF (broad)		# in LF (nar)	
Outcome variable	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
△ HH size	-0.651	[0.233]***						
△ # kids young	0.016	[0.067]						
△ # kids school	0.048	[0.107]						
△ # youth	0.134	[0.094]	0.041	[0.038]	0.026	[0.048]	0.038	[0.075]
△ # young adults	-0.012	[0.108]	0.097	[0.096]	0.014	[0.124]	-0.052	[0.105]
△ # young adult:F	-0.044	[0.077]	0.081	[0.066]	0.03	[0.074]	-0.058	[0.067]
△ # young adult:M	0.032	[0.075]	0.016	[0.056]	-0.016	[0.093]	0.007	[0.085]
△ # mid-adults	0.12	[0.098]	0.069	[0.069]	0.088	[0.087]	0.058	[0.088]
△ # mid-adult:F	0.114	[0.088]	0.103	[0.063]	0.121	[0.070]*	0.104	[0.079]
△ # mid-adult:M	0.005	[0.058]	-0.034	[0.026]	-0.033	[0.035]	-0.046	[0.050]
△ # older adults	0.077	[0.041]*	0.022	[0.024]	0.025	[0.024]	0.078	[0.036]**
△ # older adult:F	0.025	[0.026]	0.001	[0.005]	0.001	[0.005]	0.04	[0.024]*
△ # older adult:M	0.052	[0.032]	0.021	[0.023]	0.024	[0.023]	0.038	[0.027]
△ # pens age	-1.026	[0.019]***						

Notes:
1. Death2 is obtained using the 'spouse - widow' algorithm described in the paper
2. The sample in $death2$ conditions on marital status and co-residency, so it is not entirely comparable
to the other two samples.
3. Robust Std. Errors, clustered at the 'hhid' level are reported
4. Omitted coefficients on variables included in the regression for variables:
Province dummies, wave dummies, urban dummy and no. of pensioners in the '$Wave_t$' period.
5. With the exception of the regression on household size, all the other
regressions also controlled for initial household size .
6. N = 10326 in each of the regressions, 60 Losers, 10266 Keepers
7. * denotes statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level
8. Each coefficient and corresponding standard error is obtained from a separate regression.
9. Age-groups: kids young (0-7), kids school (8-15), youth (16-20), young adults (21-35),
middle adults (36-50), older adults (51 - pension-age)

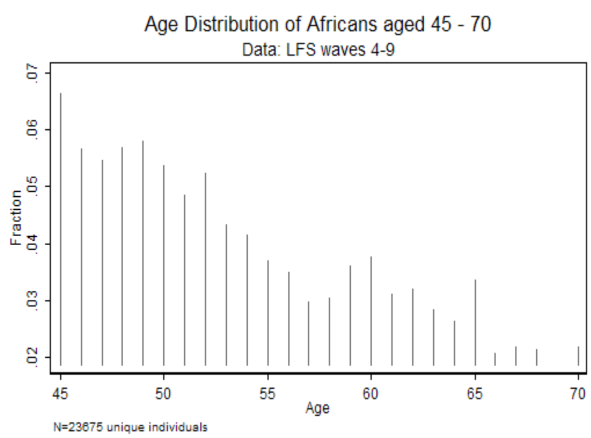
Table 2.11: **Individual level regressions for changes in Labor Market status: Rural KwaZulu Natal**

		Dependent variable		
Group	N	△ work	△ in LF br	△ in LF nar
youth	636	0.034	-0.031	-0.007
		[0.018]*	[0.062]	[0.049]
young adults	840	0.042	-0.008	-0.075
		[0.049]	[0.048]	[0.064]
young adult: F	471	0.036	0.029	-0.166
		[0.068]	[0.088]	[0.080]**
young adult: M	369	0.043	-0.049	0.003
		[0.074]	[0.037]	[0.090]
mid-aged adults	342	0.028	-0.099	-0.039
		[0.059]	[0.079]	[0.066]
mid-aged adult: F	205	0.031	-0.142	-0.086
		[0.068]	[0.097]	[0.090]
mid-aged adult: M	137	-0.032	-0.073	0.004
		[0.109]	[0.138]	[0.071]
older adults	170	0.19	0.145	0.17
		[0.093]**	[0.100]	[0.110]
older adult: F	86	0.283	0.247	0.304
		[0.115]**	[0.126]*	[0.117]**
older adult: M	84	0.021	-0.059	-0.053
		[0.132]	[0.126]	[0.154]

1. Outcome variables in units of △ in labor market status, values of -1, 0 & 1 in data.
2. Reported coefficient corresponds to dependent variable *losepen*
3. Robust Std. Errors, clustered at the 'hhid' level are reported
4. Omitted coefficients on variables included in the regression for variables:
Wave dummies and no. of pensioners in the $Wave_t$ period.
5. * denotes statistical significance at the 10% level, ** at the 5% level, and *** at the 1% level
6. Observations are weighted by [pweight x IPWeight]
7. Each coefficient and corresponding standard error is obtained from a separate regression.
8. Age-groups: youth (16-20), young adults (21-35), middle adults (36-50), older adults (51 - pension-age)

2.15 Figures

Figure 2.1: **Age Heaping in the LFS Data**



Bibliography

Ardington C, Case A & Hosegood V, "Labor supply responses to large social transfers: Longitudinal evidence from South Africa", Paper presented at the Twelfth Annual Conference of the African Econometric Society, Cape Town

Banerjee A, Galiani S, Levinsohn J & Woolard I, "Why Has Unemployment Risen in the New South Africa?", CID Working Paper No. 134, October 2006

Becker G, "A Theory of Marriage", Journal of Political Economy 81: 813-846, 1973

Beittel M, "The Witwatersrand: Black Households, White Households", in Creating and Transforming Households: The constraints of the world economy, edited by Smith J. & Wallerstein I. Cambridge University Press 1992

Bertrand M, Mullainathan S & Miller D, "Public policy and extended families: Evidence from pensions in South Africa", World Bank Economic Review 17 (1): 27-50 2003

Case A, "Does money protect health status? Evidence from South African pensions", NBER working paper 8495, October 2001

Case A & Deaton A, "Large cash transfers to the elderly in South Africa", Economic Journal 108 (450): 1330-1361, Sep 1998

Duflo E, "Grandmothers and granddaughters: Old-age pensions and intrahousehold allocation in South Africa", World Bank Economic Review 17 (1): 1-25 2003

Duflo E, “Child health and household resources in South Africa: Evidence from the Old Age Pension program”, American Economic Review 90 (2): 393-398 MAY 2000

Edmonds EV, Mammen K & Miller DL, “Rearranging the family? Income support and elderly living arrangements in a low-income country”, Journal of Human Resources 40 (1): 186-207 WIN 2005

Edmonds EV, “Child labor and schooling responses to anticipated income in South Africa”, forthcoming in Journal of Development Economics, accepted May 2005

Hamoudi A, & Thomas D, “Pension Income and Well-Being of Children and Grandchildren: New Evidence from South Africa”, California Center for Population Research On-Line Working Paper Series, CCPR-043-05, October 2005.

Jensen RT, “Do private transfers ’displace’ the benefits of public transfers? Evidence from South Africa”, Journal of Public Economics 88 (1-2): 89-112 JAN 2004

Kingdon G & Knight J, “What have we learnt about unemployment from microdatasets in South Africa?”, Social Dynamics, 27, No. 1: Spring, 2002.

Keller S, “Household formation, poverty and unemployment - The case of rural households in South Africa”, South African Journal of Economics 72 (3): 437-483 SEP 2004

Lund F, . “State social benefits in South Africa.”, International Social Security Review 46, 1993

McLaren Z, “Why is Unemployment so high in South Africa? The Effect of household job loss on labor market outcomes of the unemployed.", Unpublished Manuscript, April 2007

Pauw K. & Mncube L., "Expanding the Social Security Net in South Africa: Opportunities, Challenges and Constraints", International Poverty Centre, Country Study:Cash Transfer Research Programme, Country Study number 8, July 2007

Posel D, Fairburn JA & Lund F, "Labour migration and households: A reconsideration of the effects of the social pension on labour supply in South Africa", Economic Modelling 23 (5): 836-853, 2006

Ranchhod V, "The effect of the South African old age pension on labour supply of the elderly", South African Journal of Economics 74 (4): 725-744 Dec 2006

Ranchhod V, and Wittenberg M, "Income Effects of Time Allocation in Poor South African Households: Evidence from South Africas Old Age Pension", presented at the IZA Topic Week on Nonmarket Time in Economics, Bonn, May 2007

Rosenzweig R. & Stark O., "Consumption Smoothing, Migration and Marriage: Evidence from Rural India.", Journal of Politial Economy, 97: 905-926, 1989

Sagner A. & Mtati R., "Politics and Pension Sharing in Urban South Africa", Ageing and Society 19: 393-416, 1999

Wooldridge J. M.,"Econometric Analysis of Cross Section and Panel Data", MIT Press, 2002

CHAPTER III

Estimating the responsiveness of college applications to the likelihood of acceptance and financial assistance: Evidence from Texas

3.1 Introduction

This paper addresses the question: To what extent does the South African Old Age Pension (OAP) reduce the labor supply of the elderly? The pension is large relative to the earning capacity of many elderly people. Recipiency rates are high and all South Africans could potentially receive it. However, the existence of a means test excludes wealthier people from receiving the pension. Our identification is obtained using the discontinuity in age that determines eligibility for receiving the pension. The paper is structured as follows. Section 2 provides an introduction to the current state pension, its history in brief, and the rules that govern the scheme. Section 3 provides a review of some of the literature that has already been written about the pension. Section 4 presents a simple theoretical model that underlies our a priori expectations for labor supply. Section 5 describes the data used to answer the research question. Section 6 discusses the empirical strategy and presents results and interpretation of our analysis. Section 7 discusses caveats to the results and interpretation, and provides some further evidence and robustness checks in response to these caveats. Section 8 concludes.

3.2 Background

3.2.1 History of the old age pension

The South African Old Age Pension involves a relatively large cash transfer to elderly South Africans. The pension is subject to both a means test and an assets test. In terms of the statute, the pension is decreased in value over some interval as other non-pension income increases [1], until it is reduced to zero. What makes it almost unique in the developing world is its value and coverage, both of which are relatively high.[2]

Lund (1993) provides an introduction to the OAP as we see it today. The author documents a number of historical characteristics of the pension, up to the early 1990's. A similar documentation of the evolution of the scheme is provided by van der Berg (1997). Case and Deaton (1998) provide a detailed description of the OAP system at the time just prior to the first democratic elections in 1994. They note that in the 1993 dataset that they used, the value of the pension was"twice the median household's per capita income" of African households. They also note that the pension value was not, in practice, reduced as other income increased. In fact, it seemed more likely that the means test was applied in a binary fashion, i.e. the means test precluded some people from receiving the pension, but those who did receive it tended to receive the maximal value.[3]

At present, the structure of the pension remains largely unchanged from 1993. The value has periodically been adjusted in response to inflation, and in 2000/2001 it stood at R540 per month, or R6480 per annum.

3.2.2 Pension Rules

Information gathered from the government's Dept. of Social Welfare website [4] indicates that for someone to be eligible to receive the old age pension, they need to satisfy the following criteria:

[1] However, empirical evidence suggests that this is not the case in practice. This is discussed later in the section.
[2] See Case and Deaton (1998) for details.
[3] The value of the pension was 370 rand in 1993. From April 2005, this value will be R780 per month, as per the national budget.
[4] http://www.welfare.gov.za. Information obtained from the document"Social Assistance Procedural Manual 2003"

- They must be resident South African citizens
- They must be 60 years or older if female, 65 years or older if male
- Their income must not exceed R16 920 if single, or R31 320 (jointly) if married
- Their assets must not exceed R252 000 if single, or R504 000 (jointly) if married.[5]

If a person is eligible to receive the social grant, the value of the grant is to be calculated as follows:

Grant = 1.5 x [Maximum Grant per Annum] - 0.5*[Annual Net Income] if single;

or Grant = 1.5 x [Maximum Grant per Annum] - 0.5*[Total Annual Net Income of Applicant and Spouse/2] if married.

The maximum annual grant value in 2003 was R8400.[6] Thus, as per the formula for calculating the value of the grant, there is some low level of income that does not affect the value of one's pension, beyond which there is effectively a 50% tax rate on other income. This taxation is maintained until no pension is provided. [7]

Following Case and Deaton (1998), we check whether the pension value is de facto adjusted in relation to other income using data from the IES 2000.[8] We ran regressions to test whether the pension value is correlated with other income, conditional on the pension value being strictly positive.[9] We also test whether the probability of receiving a pension is decreasing in other income. Our findings are consistent with the hypothesis that other income is relevant only in determining whether a person receives a pension or not, but not relevant in determining the value received. Conditional on receiving some pension, the value of the pension tends to be the maximal value. It is also worth noting that a vast majority of respondents who are pension age-eligible would probably be able to continue working and satisfy the means test. In the IES 2000, more than 85% of African women

[5]The value of a personal home is excluded from the 'asset' criterion.

[6]This paper was written in the SA fiscal year 2003/2004, so some of the values are outdated. However, the important salient features, such as the means test and the age threshold, remain in place.

[7]Note, however, that the amount of other income that would drive the pension value to zero is R21 000 per annum, considerably more than the income threshold of R16 920.

[8]We discuss the data in detail in Section 3.

[9]Since this is not crucial to our research, we have not included these regression results. Nor have we corrected for the 'joint' means test as applied to couples.

aged 55 - 59 who report positive non-pension income, earned less than the income threshold of R16 920 per annum. The corresponding figure for African men aged 60 - 64 is 78%. The proportion that could work and receive the pension is likely to be even higher for those who are pension age eligible, as they have less education on average. For poorer households, it may thus be desirable for the elderly to continue working even as they start to receive the pension.

3.3 Literature Review

As stated previously, the OAP is unique in the developing world due to its coverage and value. This uniqueness has attracted the attention of several economists.

One of the first mainstream economics paper published on the OAP was by Case and Deaton (1998). The authors analyze the redistributive consequences of the OAP, as well as the expenditure patterns of households that receive the pension using the 1993 SALDRU dataset. They find that the OAP is an effective transfer to the poor and poverty stricken in general. Furthermore, the prevalence of three-generation households, as well as 'skip generation' [10] households, results in the pension potentially benefiting children as well. In particular, it disproportionately reaches children in poverty.

Various authors have also analyzed other dimensions of household behavior in response to the pension. Some have looked at whether the OAP impacts on the health of recipients or their household members. Duflo (2000) finds a discontinuous increase in children's height for age for children living with pension eligible persons. This increase is significant for female children, and is realized on average only when the pension recipient is a woman. Duflo (2003) reports similar evidence that the pension is shared between members of the household. Moreover, the sharing of recipients' cash from pensions is differentiated by gender.

In similar vein, others have asked the question; 'How do other members of the household respond when a member becomes pension eligible?' Bertrand et al (2003) find that having

[10] Households with grandparents and grandchildren but no parents.

a pension eligible person in the household has a statistically significant and negative impact on the labor supply of prime aged individuals[11] in the sample. Edmonds (2003) considers the impact of the OAP on child labor supply and schooling attendance. He finds that when a household member who is male[12] becomes pension eligible, there is a sizable decline in child labor, coupled with an increase in schooling attendance and attainment.

Jensen (2003) questions whether household disposable income increases by the full value of the pension. He estimates that crowding out of remittances by pensions is large and significant. On average, every rand of pension income received by the elderly is met with a 0.25 to 0.30 rand decrease in remittances received from the pensioner's children. This results in an over-estimation of the poverty alleviation achieved by the pension. Jensen also investigates the effects of the OAP on household labor supply. He finds no evidence that the pension induced large levels of labor force withdrawal for the household on average. However, he does find a small, though not statistically significant decrease in other household income when a pension is received.

Most recently, Edmonds et al (2005) find that household composition itself is affected by someone becoming pension age-eligible. They find a decrease in the number of prime working-age women, and an increase in the number of children younger than five and young women of childbearing age.

By and large, however, most authors have ignored the potential effects on the labor supply of the actual recipients, namely the elderly. Case and Deaton (1998), and Alderman (1999) mention the disincentives involved in a means tested cash transfer program, but also cite high levels of unemployment as a mitigating factor.

While it certainly is possible that the impact on the labor supply of the elderly is negligible, this cannot be assumed a priori. We shall add to the existing literature by rigorously investigating whether the OAP has significant labor supply effects for the elderly or not. Moreover, this informs us about the sensitivity of retirement behaviour with

[11]Prime aged individuals are defined as individuals between 16 and 50 years of age in their study.
[12]As consistent with other papers, there is a gender differential in the impact of the pension.

respect to income. It also has potential implications for labor supply effects of non-elderly recipients in response to other welfare related transfers, such as the Child Support Grant.

3.4 Theoretical Framework

One can think of the impact on labor supply (L^s) within the framework of a static model of a consumer optimizing over (X, l) where X is the consumer's bundle of consumption goods and l is the consumer's consumption of leisure. Leisure is assumed to be a normal good. The consumption of leisure has an impact not only on utility directly via consumption, but also indirectly via its impact on the budget set. We assume that the consumer needs to decide how many hours of labor to supply for a given wage rate, where h = hours = $K - l$. In this case K is the endowment of possible work hours a person has.[13] We also make the standard assumptions that utility is twice differentiable and strictly concave in all of its arguments.

The consumer's problem then, is to:

$$\max_{(X,l)} U(X, l) \qquad s.t. \qquad p.X \leq w.(K - l) + Y + P(age, Y, l)$$

where p is a price vector corresponding to the set of consumption goods, w is the consumer's wage rate, Y is the consumer's non - labor, non-pension income and P is the consumer's pension income, if any. Pension income is a function of age and other income as described previously.

In this simple static model, it is straightforward to show that an exogenous increase in P, *ceteris paribus*, would lead to an increase in consumption of all goods including l (or equivalently, a decrease in h.) However, this expectation of a decrease in h applies only to that subset of the population who were willing to supply positive amounts of labor prior to the hypothetical increase in P.

In addition to the income effect, the OAP provides incentives to reduce earned income via the means test. Whether the pension value is reduced in response to other income or

[13] An upper limit on this would be 24 hours per day, although this is clearly unreasonable.

not remains uncertain. However, if the rules are strictly enforced, then the reduction in pension value provides further incentives for people to reduce their L^s when they become eligible to receive the pension. Put together, the income effect of the pension is to consume more leisure, and the means test (as well as the potential implicit tax on earned income) also encourages the consumer to substitute away from labor towards leisure.

Figure 3.1 depicts the movement of the budget set for a consumer just before and just after they become pension age-eligible.[14] 'Diagram a)' represents the case where the pension value is gradually adjusted to zero as other income increases, while 'Diagram b)' represents the case where the income test is used to determine whether a person receives any pension or not, and everyone who receives a pension receives the maximal value of the pension. As stated previously, regardless of which of these two models is more correct, the implication of both of them is for labor supply to (weakly) decrease when a person becomes pension age-eligible.

Thus, a priori, we expect that becoming pension age-eligible would cause people to stop working or reduce their hours worked, which is a testable hypothesis. Note that for people with high levels of non-pension income, or for people with sufficiently high wage rates, the pension is unlikely to have any effect since these people are unlikely to satisfy the means test.[15] That said, our analysis does measure the average effect of the pension on the population.

3.5 Data

For the core section of this paper, we use data contained in the South African Labor Force Surveys (LFS). These are nationally representative household level surveys conducted biannually by Statistics South Africa. The sampling methodology involves a stratified clustered design.

We pooled two waves of the LFS data, namely LFS 2000:2 and LFS 2001:2, which

[14]The solid line is their original budget line, and the dashed line shows how the original budget line shifts outwards upon reaching the pension eligible age.

[15]It is probably more correct to say that the incentive to reduce their earned income in order to receive the pension is too weak, and they prefer to forego the pension.

were conducted in September 2000 and September 2001 respectively. We treat the data from the two waves as information coming from a single survey.[16] The advantage of doing so is that we have more data and therefore, are likely to get better and more accurate estimates.[17] The data contained in the LFS datasets includes demographic information, current employment status and related information about respondents. This focus makes these surveys ideal for analyzing labor force participation.

We restrict our attention to the sub-sample of elderly Africans. Africans are the majority population group in the country[18], and are disproportionately likely to be receiving the OAP.[19] We define elderly as being 50 years old or greater. We thus exclude all persons aged below 50, and also exclude all persons aged 75 or older, as there is very little variation in their labor force participation behavior. After cleaning the data as mentioned above, we have a sample of 18951 observations. Of these, 7774 are men and 11177 are women. Mean pension recipiency rates are shown in Fig. 3.2.

It is reassuring to observe that self-reported recipiency rates are low for women younger than 60 and men younger than 65, and to observe the sharp rise in recipiency rates thereafter, as this is a crucial requirement for the interpretation of our results to be valid. Fig. 3.3 shows the mean years of education, by gender, for each age group in the data. This is in general a poorly educated subpopulation, with an average education level not exceeding the primary school level. On average, older cohorts are less educated than younger ones, and women are less educated than men.

In addition to the LFS data, we use the Income and Expenditure Survey 2000 (IES 2000), which is the most recent IES released by Statistics South Africa. This survey has a cross-sectional design and is conducted every five years. It contains detailed information on

[16]However, we do identify the clusters separately. This is because we are not certain that the same cluster number in different waves represent the same districts.

[17]A completely new sample was drawn starting with LFS 2001:2, so there are no repeat observations in the two pooled datasets. We also have no reason to believe that there were significant changes in labor force participation rates during this time period.

[18]72% of respondents aged 50 - 74 were African in our dataset, and 79% of the population was African in 2001. (Source: Census 2001, Stats SA)

[19]Approximately 80% of age-eligible Africans report receiving the pension as compared to 30% of age-eligible Whites in LFS 2000:2.

income and expenditure. We use this data for additional analysis relating to our research that contains an income component. Another reason to do the analysis using data from this time period is that LFS 2000:2 coincides with the IES 2000, which allows us to combine information from both datasets for some of our investigation.

3.6 Empirical Strategy

In order to estimate the average effects of the pension, we analyze various dimensions of labor supply as well as type of job and hours worked. The identifying assumptions involve the discontinuity in age-eligibility, and that people who are almost but not quite age-eligible are not 'too different' from those who have 'just become' age eligible for the pension. Thus, for example, we assume that 59 year old African women are not significantly different from 60 year old women in any material respects except for the eligibility of the 60 year old women to receive the pension.

3.6.1 Labor supply analysis

We use five different measures related to labor supply. All of these are indicator variables and are constructed as follows:

- '$works$': =1 if in the past seven days the person was engaged in paid work, or unpaid work in a family business, or employed as a domestic worker, or was self employed, or worked on a family owned plot of land or has a job but was temporarily absent; = 0 otherwise.

- '$works2$': which is the same as 'works' except that it equals 0 if the only working activity the person engaged in was on a family owned plot of land.

- '*LFP_broad*': (labor force participation: broad definition) =1 if the person 'works' or if the person is 'willing to accept a reasonable offer'; =0 otherwise.

- '*retired*': =1 if a person is not 'working', and the reason they are not working is that they are either 'retired' or 'too young or too old to work' [20] ; =0 otherwise.

- '*accept*': is a variable only generated for those not currently working. 'accept'=1 if the person is 'willing to accept a reasonable offer'; =0 if not; and is undefined for those currently 'working'.

We then fit probit models to these binary variables.[21] Our set of covariates is simply: $(age-50)$, $(age-50)^2$, years of education, and a dummy variable to indicate pension age-eligibility. We also interact the pension age-eligibility dummy with all the other covariates in the regression[22], and take into account the complex sampling framework. Since the relevant age thresholds differ for men and women, we estimate these probits separately by gender. Education is used primarily as a 'control' variable, which we comment on more in a later section.

We predict the probability of a particular outcome for each individual and find the mean probability for each age. We also 'project' the probability for those exactly at the age threshold, by predicting their probability while ignoring the coefficients on the dummy and dummy-interacted variables. We then calculate the mean of these projected probabilities and interpret the difference between the 'predicted' and 'projected' means, evaluated at the relevant age, as the average effect on the outcome variable due to the pension potentially becoming available at this age.

For example, suppose our dependent variable is *retired* and our subsample is African women, aged 50 to 74. Our specification is equivalent to estimating two separate regressions. The first is for the women aged 50 to 59. We use these coefficients to then project the probability of being retired for the women aged 60 in our relevant sample[23], and calculate the mean of the projected probability of being retired. This is effectively an extrapolation of the trend for the women aged 50 to 59, conditional on education. We interpret this as

[20]Since we have restricted the sample to those aged 50 - 74, we expect that this means that they are 'too old to work' rather than 'too young to work'.

[21]For a discussion on binary dependent variables, see Greene, WH , *Econometric Analysis*, Ch 19.

[22]Note that this is equivalent to estimating the same model separately on either side of the pension age-eligibility threshold.

[23]'Project' in this instance is simply an out of estimation sample prediction.

the counterfactual retirement rate if women aged 60 behaved identically to the younger women, except for the fact that they are simply a year older.

The second regression involves fitting the identical specification to the data for women aged 60 to 74. Using this second set of coefficients, we again predict the probability of retirement, and calculate the mean of the predicted probability for each age. Our estimate of the treatment effect is then the difference between the projected mean and the predicted mean, evaluated for the women who report being aged exactly 60.

In order to measure the accuracy of our estimates of the treatment effect, we bootstrap the results over 1000 replications for each regression. In performing the bootstrap, we explicitly replicate the stratification that was used to draw the original sample. We also test for the joint significance of the dummy and dummy-interacted variables by means of adjusted Wald tests.

3.6.2 Labor supply analysis: Results

Tables 3.1 and 3.2 present the regression results for men and women respectively. The individual coefficients relating to the dummy and dummy interacted variables are not always significant.[24]

However, the results are most elegantly presented graphically. Figures 3.4a - 3.4e depict the average predicted probability of the relevant measure being equal to one for African males. Figures 3.5a - 3.5e depict the average predicted probability of the relevant measure being equal to one for African females. The square on the vertical line through the pension age threshold represents the projected trend from our regression, after controlling for education. Graphically, the difference between the 'square' and the 'dot' on the vertical line represents our estimate of the impact of the OAP.

We reject the hypothesis that the dummy and dummy-interacted coefficients are simultaneously equal to zero in all but one of our regressions. We reject these at any reasonable level of significance. The only regression where we cannot claim joint significance of the

[24]In fact, they are never individually significant for men.

dummy and dummy interacted variables at the 5% conventional level, but can still claim it at the 10% level, is the regression relating to males with dependent variable *accept*.

Insofar as general statements can be made about individual covariates, we observe the following: For men, the $(age - 50)^2$ coefficient is usually significant, while it is not so for women. 'Years of education' has a clearly significant effect on labor force participation for both genders. As expected, an increase in education is associated with an increase in the probability of working or being in the labor force[25], and a decrease in the probability of being retired for any given age.

For women the coefficient on the dummy variable is sometimes significant, but is of the 'wrong' sign in these cases. However, after accounting for the significant and negative coefficients corresponding to the dummy interacted with $(age - 50)$, we observe a net negative marginal effect from the set of dummy and dummy interacted variables.

That said, our stated objective was to measure the impact of becoming pension age-eligible, and not only to identify whether labor force behavior is affected by the pension or not. Table 3.3 presents the difference in the average predicted probabilities generated as described above. We also present our bootstrapped mean, 5th centile, 95th centile and standard deviation of the difference.

In general, our estimates of the difference are both large and significant. From Table 3.3, for males: we observe that becoming pension-age eligible coincides with a drop of 7.6% points in the probability of working and a drop of 9.5% points in the probability of working if we exclude subsistence farming. This larger drop for the latter is consistent with a migratory labor supply system, whereby workers leave their formal employment and return to homesteads in rural areas where they are likely to engage in some agricultural activities . We also see that broadly defined labor force participation drops by 8.4% points and the proportion retired increases by 9.3% points. The difference in willingness to accept is the smallest at 3.4% points.

[25] This is expected because better educated people have higher wages on average and therefore have a higher opportunity cost of leisure. We assume away the possibility of a backward bending labor supply function for this relatively poor population.

For women, turning 60 coincides with a drop in the average probability of working of 5.7% points. This drop rises to 8.8% points if we exclude family plot based agriculture. Broadly defined labor supply drops by 12.6% points and the probability of being retired rises by 20.6% points. The average probability of being willing to accept a job drops by 11.7% points.

This rise in the proportion of women 'retired' seems implausibly large in relation to all of the other estimates. Bear in mind, however, that 'retired' is a combination of those not working and those not willing to accept a job. Since both of these measures exhibit large decreases at age 60, the effect on the retired group is bigger than the effect in either single dimension.

Our bootstrap results reinforce our belief that becoming pension age eligible is correlated with discrete changes in labor supply. We focus on the sign of the 5th centile of each simulated variable except for 'retired', where we focus on the 95^{th} centile. For example, the 5^{th} centile for the 'works' variables for men is 0.015 which is positive. This means that we estimate a reduction in the proportion of males working in at least 95% of our bootstrapped samples. With the exception of the 'accept' dimension for males, we conclude that there is sufficient statistical evidence to believe that there is a significant discontinuous reduction in labor supply in every dimension for both genders.

3.6.3 Hours Worked Analysis

The second theoretical prediction that we have is that a person's hours worked, conditional on them still working, should decrease once they become pension age-eligible. In order to identify this effect, we analyze the reported 'usual total hours worked' of elderly African people. We separate our data into two types of workers, those in 'fixed-time' jobs where the employer fully determines the hours worked by the employee, and 'flextime' jobs, those where the employee fully determines the hours worked.[26] The type of job accepted

[26]The question reads, "Can 'respondent' decide on the number of hours per week during which he/she works?" (Q4.22). We define flextime=1 if option 1 was selected and flextime=0 if option 3 was selected.

is almost surely an endogenous choice of the worker[27], and we examine the possibility that workers may switch from fixed-time employment towards flextime employment upon becoming pension age-eligible.

We test by means of probit regressions whether being pension age-eligible coincides with a significant increase in the probability of being in a flextime job relative to being in a fixed-time job. As before, this is done separately by gender. We also estimate the difference in hours worked for workers in each type of job for the group of people who are pension age-eligible, and compare this to the difference for those who are not pension age-eligible.

3.6.4 Hours Worked Analysis: Results

Fig. 3.6 depicts the ratio of people in flextime relative to fixed-time employment by gender. This ratio goes above one for the first time, and remains above one thereafter[28], at exactly the gender-specific age at which people first become pension age-eligible.[29] These ratios both seem relatively stable for ages below the relevant age threshold, and rises very sharply for men between the ages of 64 and 66, and women between the ages of 59 and 61.

Table 3.4 reports the results of probit estimates under slightly different assumptions about the data, with $flextime$ as our dependent variable. In all cases, we find that conditional on being employed, being in the older group of workers is coincident with a greater probability of being in flextime employment. From the marginal effects column, we observe that this probability is estimated to be 14.7% points for men and 13.9 % points for women. The coefficient on the dummy variable is significant at the 5% level for both men and women. This is true regardless of whether we adjust for the complex sample design or not.

Table 3.5 reports the results of the male and female regressions of hours worked[30] on age,

[27]Subject to the availability of such jobs to that worker.

[28]We cannot distinguish whether this is only due to differential attrition from the two types of jobs or whether people switch types of jobs at these ages.

[29]The sharp drop in the ratio for women aged 65 is possibly due to age heaping, although there is no way to test this. The proportion of this sample who remain employed is also very small.

[30]We disregarded observations where reported hours worked was greater than 100. We believe this to be a plausible cut off point given that the mean and modal hours worked in the data are between 40 and 50 hours per week. Moreover, we only

type of job, a dummy for pension age-eligibility and an interaction term for the dummy and 'type of job' variables. If people do want to consume more leisure upon receiving the pension, we would expect there to be a drop in the usual hours worked for people who are pension age-eligible, in addition to a possible general trend effect of working less as one gets older. We expect that this drop would be more pronounced for those in flextime positions relative to those in fixed-time jobs.

Interpreting the coefficient estimates as averages, we see that men in a flextime positions work 7.74 fewer hours per week on average. Those who are over 65 years old and in a flextime position work an additional 4.86 fewer hours than their similarly aged counterparts. This latter difference is statistically significant at the 5% level. For women, a similar pattern is observed. Women in flextime positions work 6.32 fewer hours per week on average. Amongst those who are pension age-eligible, this difference increases by an additional 5.57 hours . This latter difference is statistically significant at the 1% level. These results are all the more compelling given that we find no statistically significant difference in the average hours worked between those in fixed-time jobs on either side of the relevant age thresholds.

3.7 Caveats and Robustness Checks

In this section we discuss some of the potential reasons to doubt our results. We address them to the extent possible, and discuss their impact in the cases where we cannot.

3.7.1 Age Heaping

Our first caveat relates to the "age-heaping" phenomenon. By this, we refer to the tendency of people to round their reported age to the nearest multiple of ten or possibly five. Fig. 3.7 shows the age profile for men and women in our sample. We note that the heaping is more acute for women than for men. The heaping is problematic for us since our identification is achieved off of the discontinuity in age-eligibility at exactly the ages

disregard 22 out of 7089 observations, and hence this is unlikely to substantively affect our results.

at which the heaping occurs. Moreover, it is not clear what the effect of the heaping is on our estimates.[31] We also have no way to verify a person's true age.

We do assume, however, that education is negatively correlated with the heaping phenomenon.[32] We thus hope that by controlling for education in our regressions we are also, at least in part, controlling for some of the non-random effects of heaping.

3.7.2 Empirical Specification

One major reason to be suspicious of our results lies in the possibility that the estimated 'discontinuity' at the relevant ages is simply a product of trends for people of those ages. In order to check whether this is the case, we estimate the labor supply regressions discussed previously, but change the age at which the dummy variable equals one. We thus estimated the difference for men assuming that the discontinuity occurred at ages 63, 64, 66 and 67, and compared these estimates to the results discuss previously.[33] For women we perform a similar adjustment and estimate the difference assuming that the discontinuity occurs at ages 58, 59, 61 and 62.

Figures 3.8a - 3.8e) provide the results in graphic form for male labor force participation. Fig. 3.9a - 3.9e) show the corresponding graphs for female labor force participation. By looking at the graphs and the size of the estimated difference for the various ages, we conclude that if there is a discontinuous change in labor supply at these ages, it seems most likely that it occurs for men at age 65 and for women at age 60.[34] This is because the largest estimated difference in the expected direction occurs at these particular ages.

[31] At a stretch, one could assume an equal probability of heaping for people close to a particular age, say 60. In this case, more people below sixty would erroneously report their age as 60 than the number of people heaping to 60 whose true ages are above sixty, due to the shape of the population pyramid. One could make the conjecture that women below 60 are more likely, on average, to be in the labor force than women above 60. In this specific case we would underestimate the magnitude of the true discontinuity for women at age 60.

[32] This seems more plausible when one observes the 'dip' in average years of education at the ages of 60 for women and 65 for men. See Figure 3.3).

[33] We ran separate regressions for each different age tested.

[34] We have also generated the corresponding graphs for the other measures of labor supply. The general pattern is consistent with the belief that discontinuities occur for men at age 65 and for women at age 60.

3.7.3 Consumption Smoothing

The third critique of our findings arises when we change from our static theoretical framework into a dynamic rational expectations environment. Applying the permanent income hypothesis[35], people should optimally 'smooth' their consumption of all goods, leisure included, in response to the anticipated receipt of the pension.[36]

To the extent that such smoothing occurs, the pension will induce a decrease in labor force participation prior to the pensionable age. This will bias our estimates downward, and our estimates are thus likely to be lower bounds of the effects of the pension.

3.7.4 Mandatory Retirement

A particularly problematic issue is the one of mandatory retirement, which is legal in South Africa. There is no way to control for the potential problem that people could involuntarily lose their jobs. Moreover, the conventional retirement ages for firms are 60 or 65.[37] However, this does not fully explain the estimated changes in all of our many dimensions of labor supply. Self-reported retirement rates increase for both genders, and the willingness to accept employment decreases significantly for women at the 5% level and for men at the 10% level. These are unlikely to be consistent with an increase in involuntary unemployment.

Furthermore, the majority of men aged 60 - 64 and women aged 55 - 59 that are employed are not employed in organizations registered for VAT[38], which is an indicator of whether they are employed in the formal sector or not. It seems unlikely that people employed in informal sector jobs would be forced into retirement by corporate rules. In addition, very few people receive employment related pensions, which makes it less likely that mandatory retirement is coupled with employment related retirement funds that reduce one's willingness to work.

[35] See Friedman, M. (1957) A Theory of the Consumption Function.

[36] This assumes that they are not too risk averse, have faith in the government continuing to provide the pension and are not liquidity constrained.

[37] Although firms may not legally discriminate by gender, which applies to retirement ages as well.

[38] Of the employed men aged 60 - 64, this proportion is 58% in our data. For women aged 55 - 59, the proportion is 81%.

Using data from the LFS 2000:2, we find that the percentage of women in the 60 - 64 year age group who receive employment related pensions was only 2.6%. For men, the percentage aged 65 - 69 who receive employment related pension was 7.2%, which we consider to be fairly small. In addition, this is identical to the proportion of men aged 60 - 64 who report receiving an employment related pension. Thus it seems improbable that the employment related pension and mandatory retirement explain the discrete drops in all of our measures of labor supply at the relevant age threshold.

3.8 Conclusion

We conclude with the finding that the income effects, and possible substitution effects, created by the Old Age Pension do have significant effects on labor supply. We conservatively estimate that the pension causes a reduction in labor supply of 8.4% points for men and 12.6% points for women in our sample. This is equivalent to labor force withdrawal of 1 in 12 African men at age 65, and 1 in 8 African women at age 60, and we believe that this is a lower bound of the true effect of the pension.

The second part of our analysis shows that people are more likely to be in flextime positions relative to fixed-time positions if they are of pension-eligible ages and continue to work. Moreover, the difference in average hours worked between flextime and fixed-time jobs increases for people who are pension age-eligible.

One policy implication is that the state needs to carefully consider the incentives it provides under its various welfare programs. In particular, the recently resurrected "Unemployment Insurance Fund" as well as the Basic Child Grant, will also provide disincentives to work, and appropriate mechanisms need to be put into place so as to reduce similar labor force distortions.

3.9 Tables

Table 3.1: **Regression Results for African Men**

	Works	LFP_broad	Works2	Retired	Accept
D_65	**-1.405**	**-1.707**	**-0.973**	**0.075**	**-1.101**
	1.796	*1.753*	*2.093*	*1.669*	*3.351*
(age-50)	**0.014**	**-0.024**	**0.011**	**0.066****	**-0.059***
	0.016	*0.018*	*0.016*	*0.024*	*0.026*
D_65*(age-50)	**0.074**	**0.087**	**0.026**	**0.125**	**0.029**
	0.191	*0.186*	*0.223*	*0.177*	*0.359*
$(age-50)^2$	**-0.004****	**-0.004****	**-0.004****	**0.004****	**-0.003**
	0.001	*0.001*	*0.001*	*0.002*	*0.002*
D_65*$(age-50)^2$	**0.001**	**0.001**	**0.001**	**-0.007**	**0.002**
	0.005	*0.005*	*0.006*	*0.005*	*0.01*
Years of Education	**0.019****	**0.028****	**0.026****	**-0.001**	**0.032****
	0.005	*0.006*	*0.005*	*0.007*	*0.008*
D_65*(Yrs of Educ.)	**-0.016**	**-0.015**	**-0.017**	**-0.013**	**0.009**
	0.011	*0.011*	*0.012*	*0.012*	*0.018*
Constant	**0.246****	**0.961****	**0.121***	**-1.943****	**0.186***
	0.053	*0.059*	*0.052*	*0.086*	*0.085*
Observations	7624	7558	7624	7624	3832
Adjusted Wald Test for joint significance of dummy and dummy interacted coefficients					
test - statistic	6.02	8.23	6.45	12.12	2
Pr(F > t-stat)|Ho is true	0.000	0.000	0.000	0.000	0.092

Standard errors in italics below the coefficient estimates
* significant at 5%; ** significant at 1%
D_65 is a dummy, =1 if age>=65, =0 otherwise
The sample size for LFP_broad differs from the other samples as some people who are not employed respond 'Don't know' to the accept question, and are excluded from this regression

Table 3.2: **Regression Results for African Women**

	Works	LFP_broad	Works2	Retired	Accept
D_60	**1.031****	**0.65**	**1.253****	**-0.994****	**-0.496**
	0.358	*0.356*	*0.422*	*0.322*	*0.741*
(age-50)	**-0.031**	**-0.051***	**-0.028**	**0.065***	**-0.06**
	0.025	*0.026*	*0.025*	*0.033*	*0.035*
D_60*(age-50)	**-0.166****	**-0.151****	**-0.220****	**0.261****	**-0.03**
	0.053	*0.053*	*0.061*	*0.053*	*0.106*
$(age-50)^2$	**-0.001**	**-0.002**	**-0.002**	**0.004**	**-0.001**
	0.003	*0.003*	*0.003*	*0.003*	*0.004*
D_60*$(age-50)^2$	**0.006**	**0.006**	**0.007***	**-0.011****	**0.002**
	0.003	*0.003*	*0.003*	*0.004*	*0.005*
Years of Education	**0.058****	**0.059****	**0.068****	**-0.042****	**0.038****
	0.005	*0.005*	*0.005*	*0.007*	*0.008*
D_60*(Yrs of Educ.)	**-0.027****	**-0.021****	**-0.015**	**0.016**	**0.011**
	0.008	*0.008*	*0.008*	*0.009*	*0.013*
Constant	**-0.223****	**0.345****	**-0.380****	**-1.360****	**-0.322****
	0.05	*0.051*	*0.05*	*0.071*	*0.069*
Observations	11038	10903	11038	11038	7577
Adjusted Wald Test for joint significance of dummy and dummy interacted coefficients					
test - statistic	12.14	19.76	16.29	45.3	11.67
Pr(F > t-stat)|Ho is true	0.000	0.000	0.000	0.000	0.000

Standard errors in italics below the coefficient estimates
* significant at 5%; ** significant at 1%
D_60 is a dummy, =1 if age>=60, =0 otherwise
The sample size for LFP_broad differs from the other samples as some people who are not employed respond 'Don't know' to the accept question, and are excluded from this regression

Table 3.3: **Regression and Bootstrap Results**

	Regression Output			Bootstrap Output of Differences			
	Ave Projected	Ave Predicted	Diff.	Mean	5^{th} %	95^{th} %	Std Dev.
Male							
$Works$	0.343	0.267	0.076	0.075	0.015	0.14	0.038
$Works2$	0.286	0.192	0.095	0.094	0.038	0.151	0.034
LFP_broad	0.411	0.327	0.084	0.083	0.018	0.152	0.041
Retired	0.493	0.586	-0.093	-0.094	-0.164	-0.025	0.043
Accept	0.112	0.079	0.034	0.035	-0.014	0.086	0.03
Female							
$Works$	0.326	0.269	0.057	0.058	0.005	0.11	0.032
$Works2$	0.285	0.197	0.088	0.09	0.039	0.14	0.03
LFP_broad	0.447	0.321	0.126	0.126	0.069	0.185	0.035
Retired	0.315	0.522	-0.206	-0.205	-0.262	-0.148	0.035
Accept	0.185	0.068	0.117	0.117	0.068	0.17	0.03

Note: All data on the '*Accept*' variable are not to be interpreted as population proportions, since they relate to the subpopulation of people who are not employed.

Table 3.4: **Regression Results: Dependent Variable is 'flextime'**

	Male			Female		
	Mfx	**Probit**	**Svy-Probit**	**Mfx**	**Probit**	**Svy-Probit**
$(age - 50)$	**0.007**	**0.02**	**0.012**	**-0.004**	**-0.01**	**-0.015**
	0.005	*0.013*	*0.015*	*0.005*	*0.013*	*0.015*
$(age - 50)^2$	**0**	**0.001**	**0.002**	**0.001****	**0.003****	**0.003****
	0	*0.001*	*0.001*	*0*	*0.001*	*0.001*
D_65	**0.147****	**0.387****	**0.320***			
	0.054	*0.138*	*0.158*			
D_60				**0.139****	**0.349****	**0.439****
				0.036	*0.092*	*0.098*
Constant		**-0.714****	**-0.652****		**-0.360****	**-0.347****
		0.046	*0.055*		*0.047*	*0.055*
Observations	3597	3597	3597	3137	3137	3137

Standard errors in italics below coefficient estimates
* significant at 5%; ** significant at 1%
D_65 is a dummy variable. =1 if age≥65, =0 otherwise
D_60 is a dummy variable. =1 if age≥60, =0 otherwise
The marginal effects correspond to the coefficient estimates of the probit estimates evaluated at X_bar
The svy-probit coefficient estimates adjust for the complex survey design
'flextime' = 1 if person works in a job with flexible hours
'flextime' = 0 if a person works in a 'job with fixed hours
'flextime' is undefined if a person does not work

Table 3.5: **Regression Results: Dependent Variable is 'usual total hours per week worked'**

	Male	Female
(age-50)	**-0.09**	**-0.25**
	0.08	*0.13*
flextime	**-7.74****	**-6.32****
	0.93	*0.98*
D_65	**1.31**	
	1.83	
D_65*flextime	**-4.86***	
	2.34	
D_60		**0.82**
		1.66
D_60*flextime		**-5.57****
		1.86
Constant	**49.86****	**44.79****
	0.51	*0.63*
Observations	3551	3104

Standard errors in italics below coefficient estimates
* significant at 5%; ** significant at 1%
D_65 is a dummy variable. =1 if age$\geq$65, =0 otherwise
D_60 is a dummy variable. =1 if age$\geq$60, =0 otherwise
'flextime' = 1 if person works in a job with flexible hours
'flextime' = 0 if a person works in a job with fixed hours
'flextime' is undefined if a person does not work

3.10 Figures

Figure 3.1: **Effect of Pension on Budget Constraint**

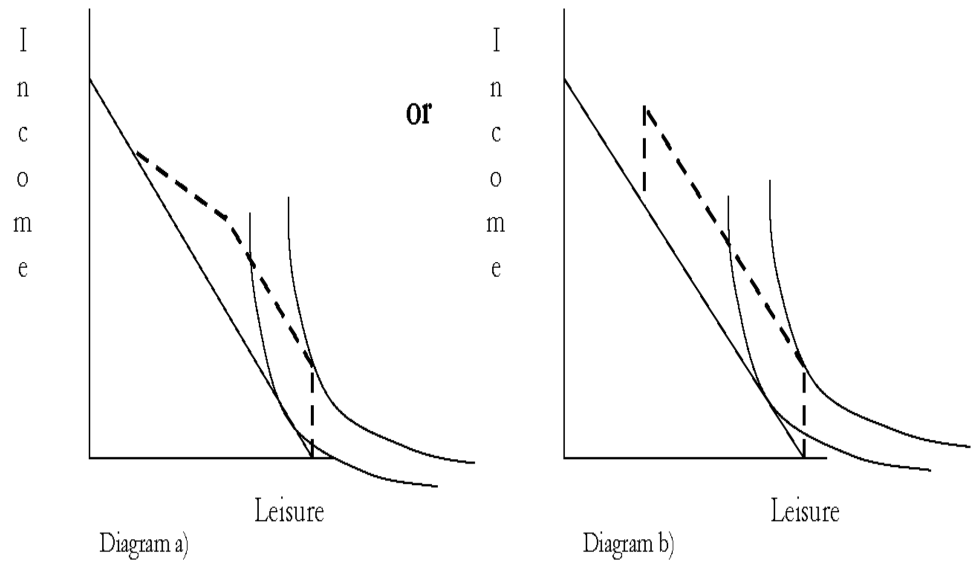


Figure 3.2: **Proportion Receiving Old Age Pension by Gender**

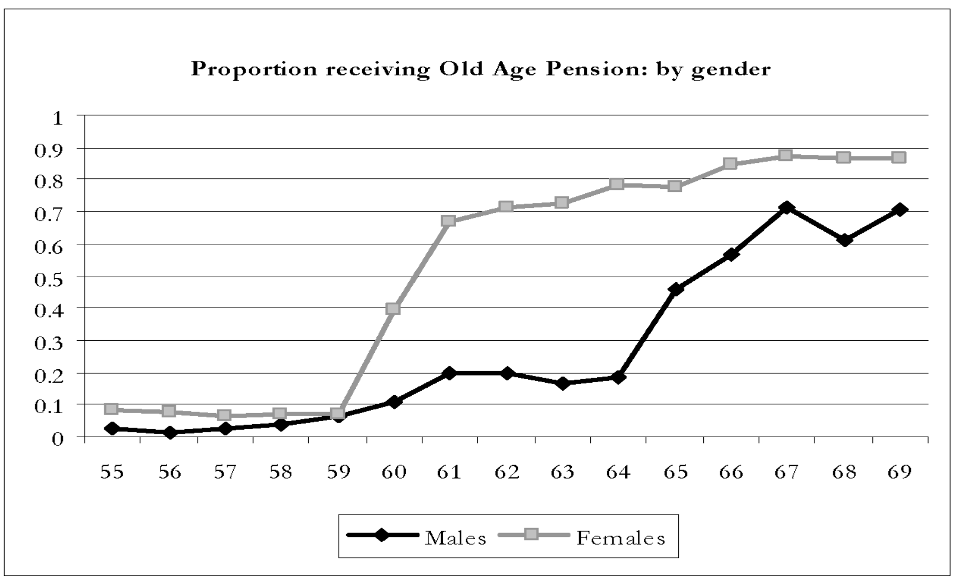


Figure 3.3: **Average Years of Education by Gender**

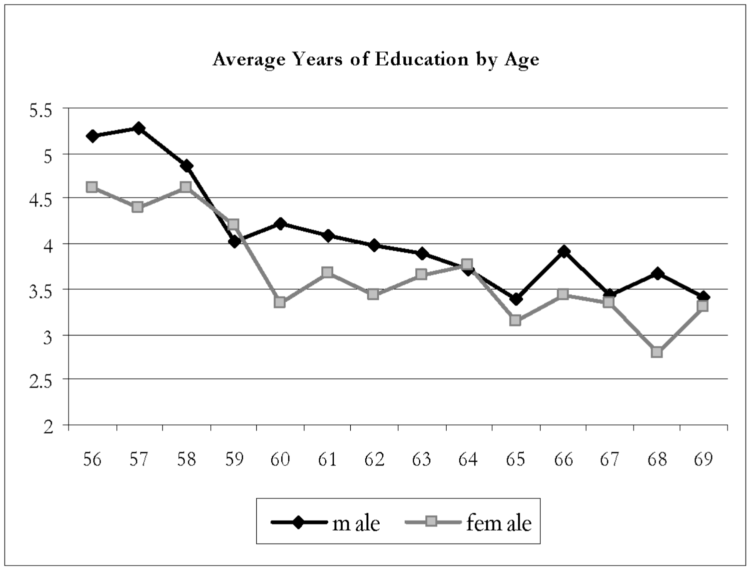


Figure 3.4a: **African Males: Works**

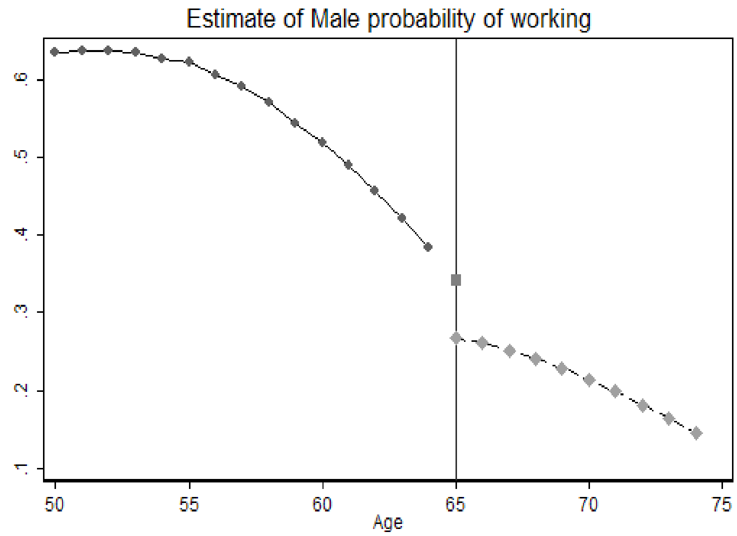


Figure 3.4b: **African Males: Works, excluding subsistence agriculture**

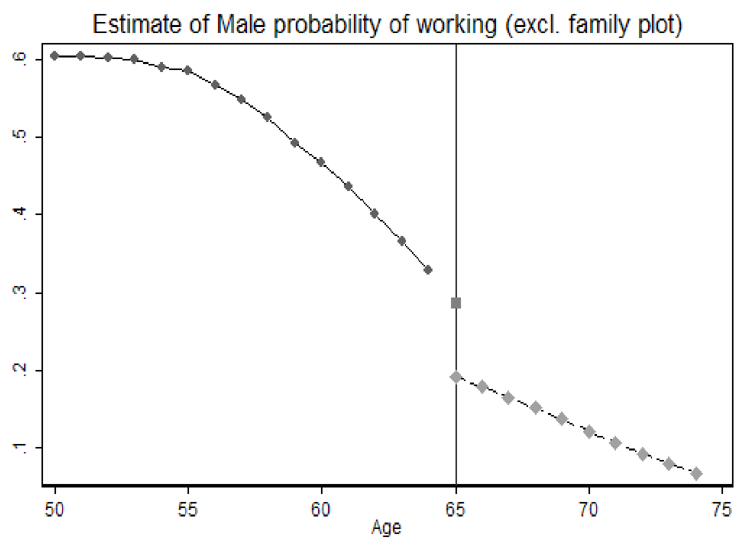


Figure 3.4c: **African Males: In Labor Force (broad definition)**

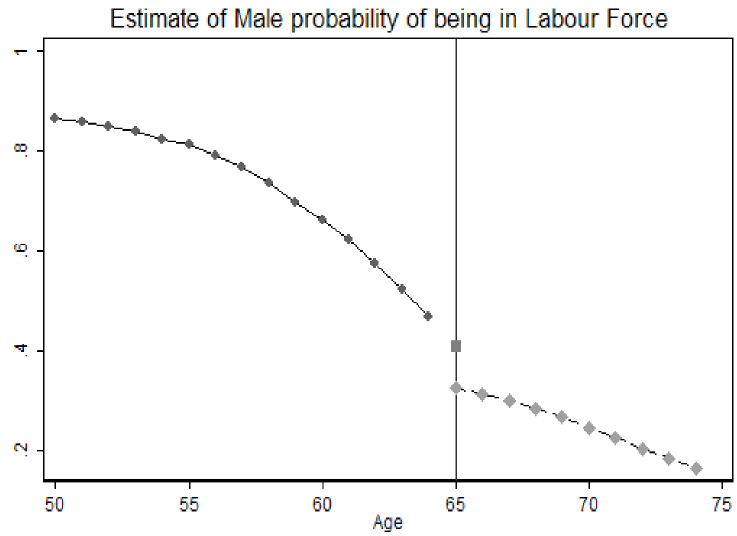


Figure 3.4d: **African Males: Retired**

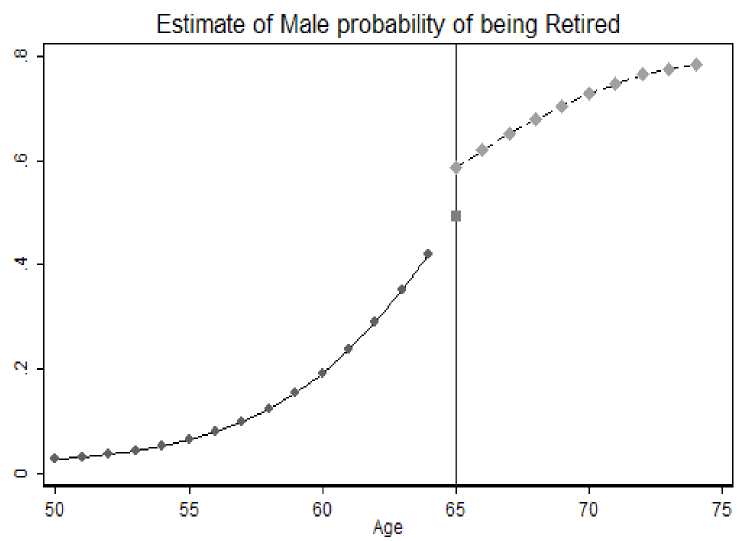


Figure 3.4e: **African Males: Accept 'Reasonable Offer'**

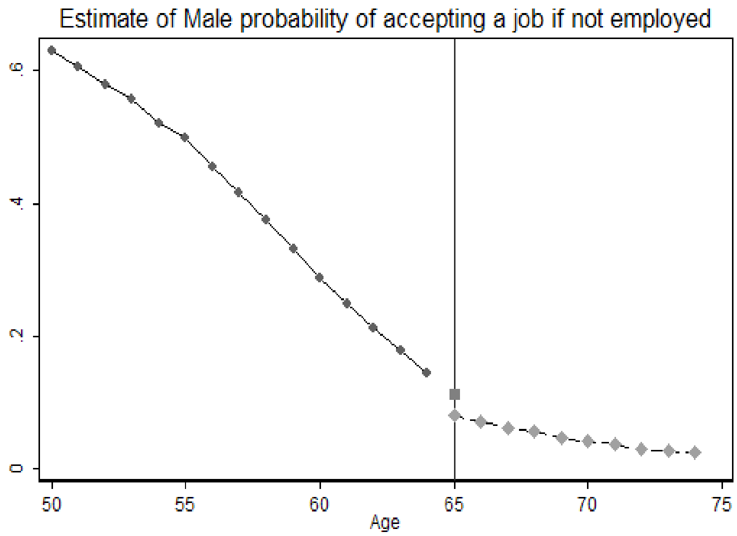


Figure 3.5a: **African Females: Works**

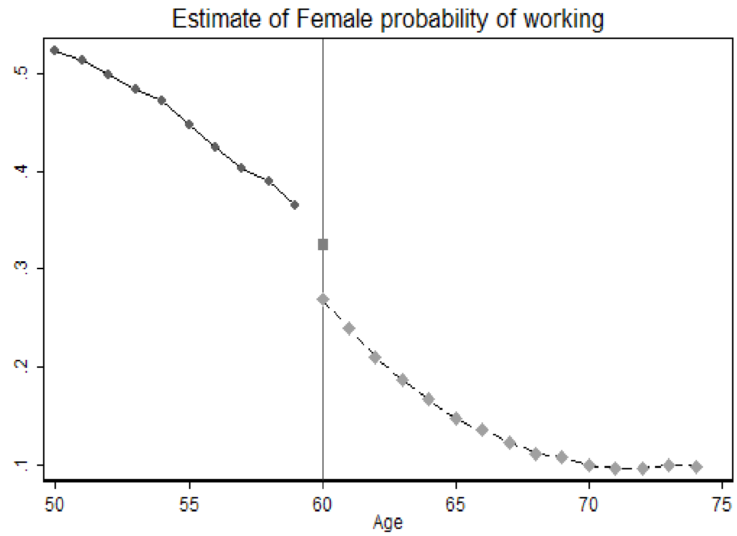


Figure 3.5b: **African Females: Works, excluding subsistence agriculture**

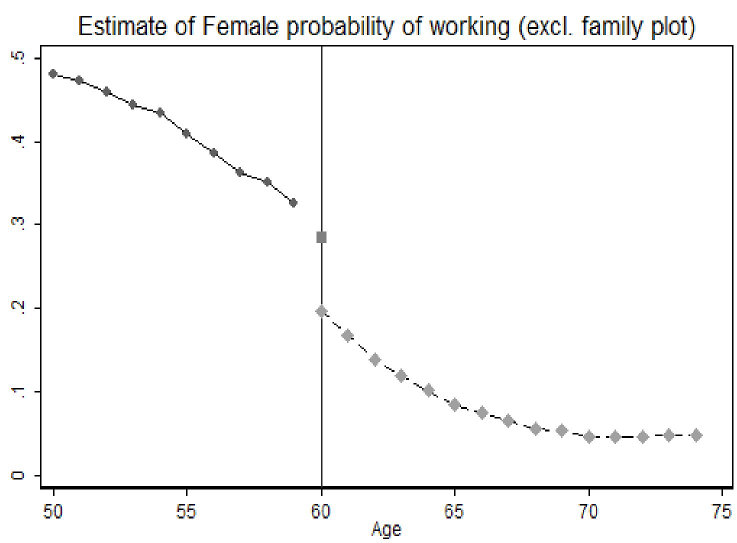


Figure 3.5c: **African Females: In Labor Force (broad definition)**

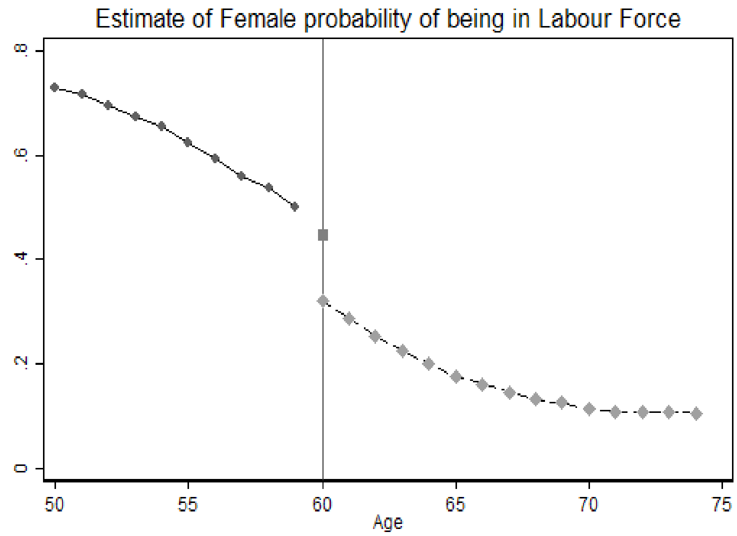


Figure 3.5d: **African Females: Retired**

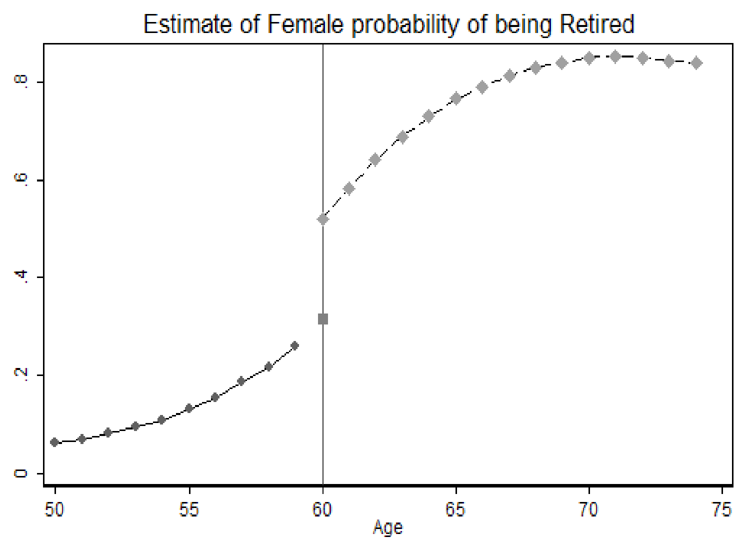


Figure 3.5e: **African Females: Accept 'Reasonable Offer'**

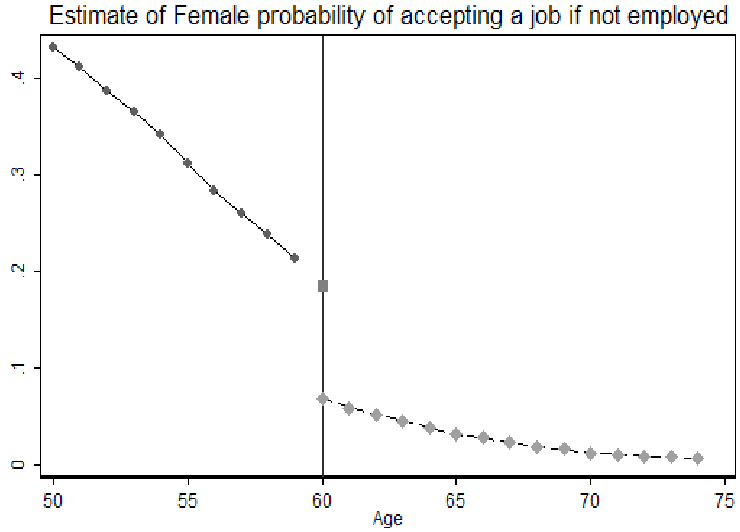


Figure 3.6: **Ratio in flexi-time employment to full time employment**

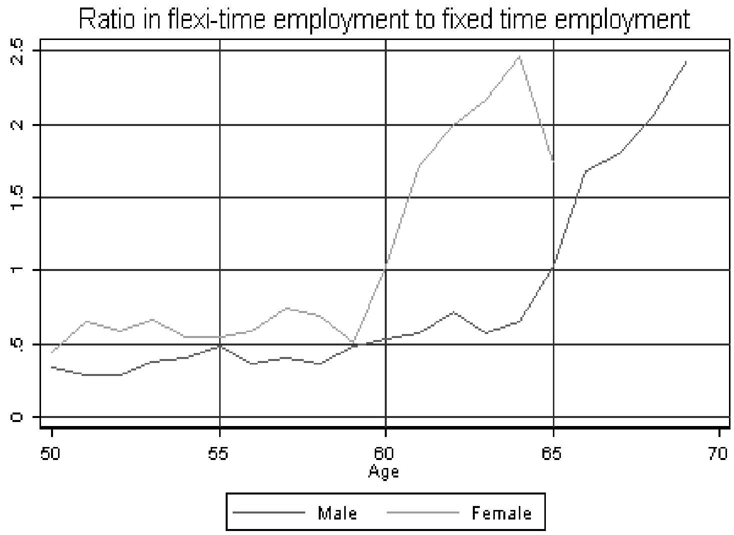


Figure 3.7: **Age Distribution of African Sub-population**

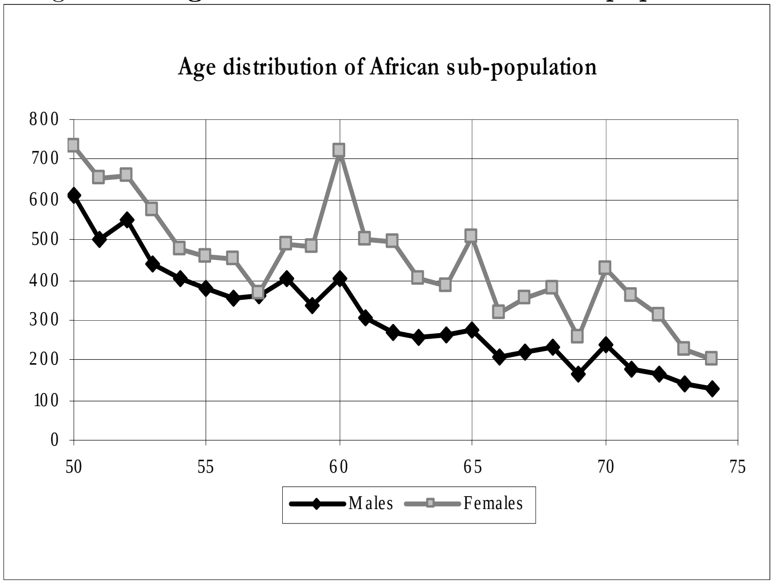


Figure 3.8a: **African Males in Labor Force: Discontinuity at 63**

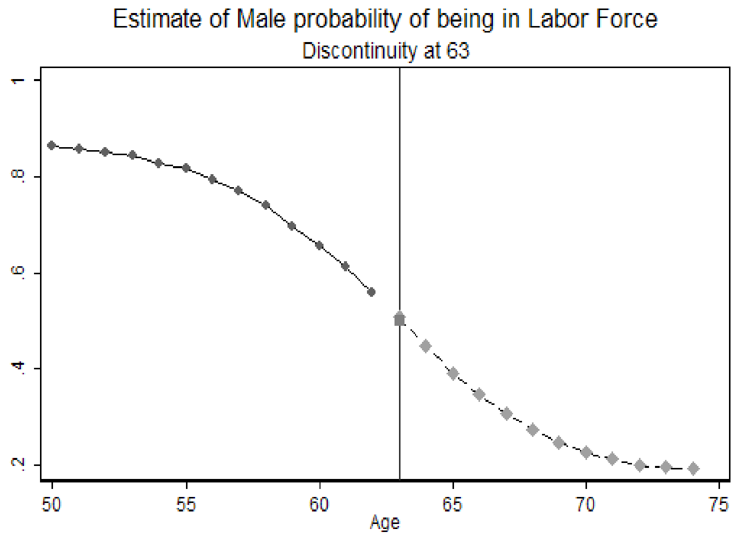


Figure 3.8b: **African Males in Labor Force: Discontinuity at 64**

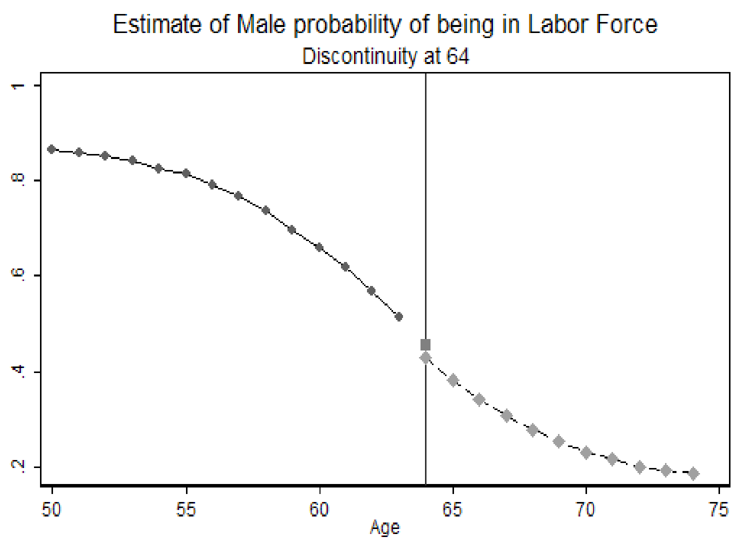


Figure 3.8c: **African Males in Labor Force: Discontinuity at 65**

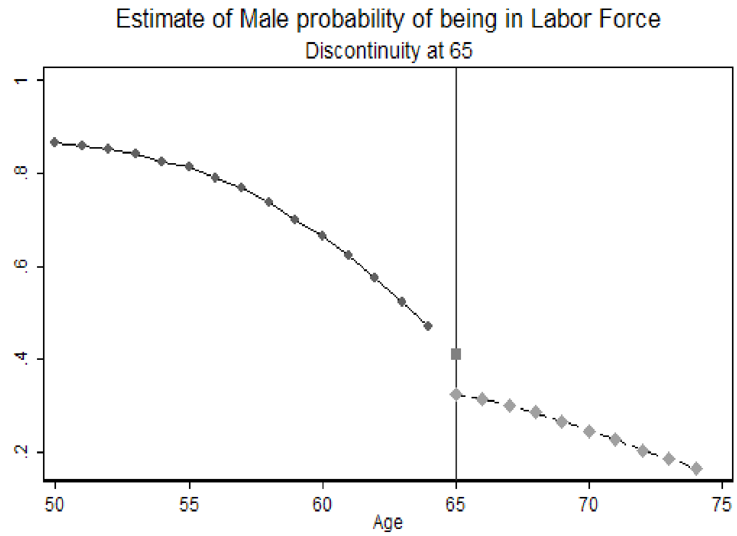


Figure 3.8d: **African Males in Labor Force: Discontinuity at 66**

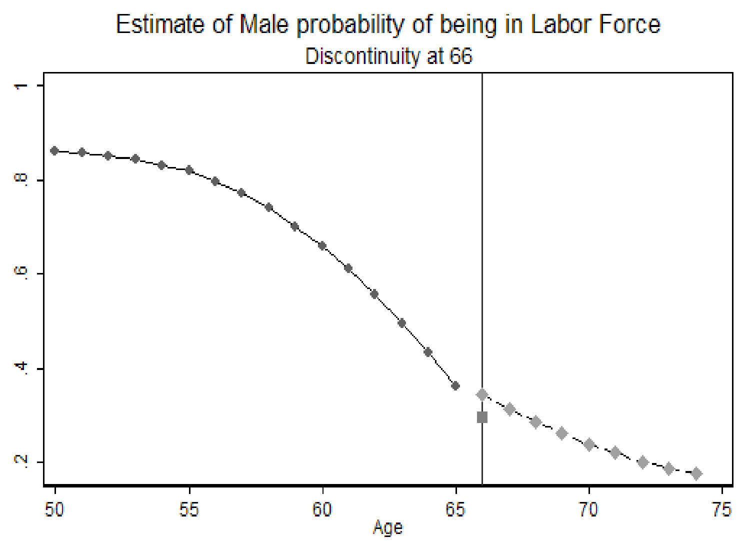


Figure 3.8e: **African Males in Labor Force: Discontinuity at 67**

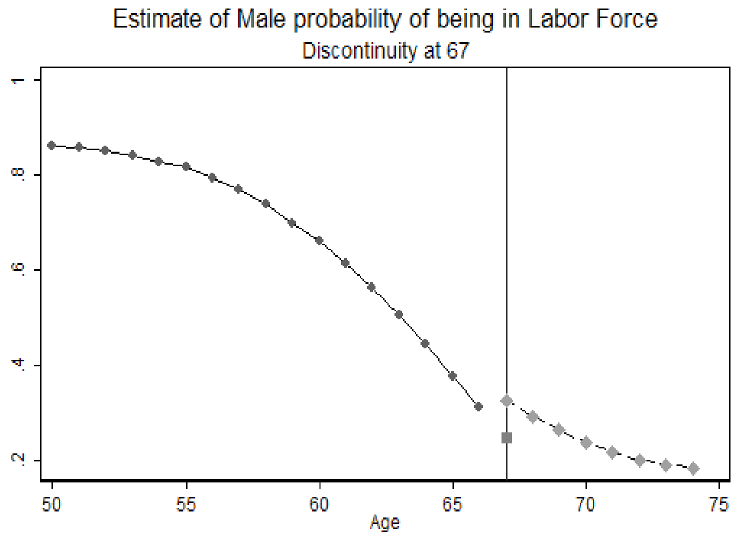


Figure 3.9a: **African Females in Labor Force: Discontinuity at 58**

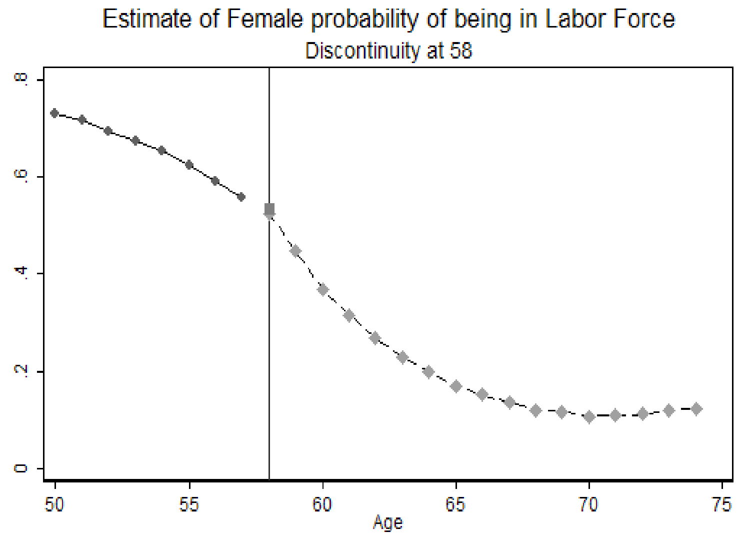


Figure 3.9b: **African Females in Labor Force: Discontinuity at 59**

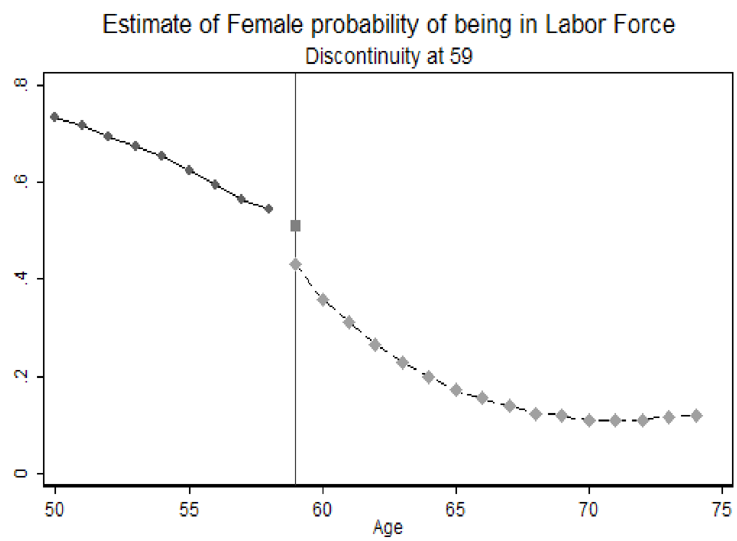


Figure 3.9c: **African Females in Labor Force: Discontinuity at 60**

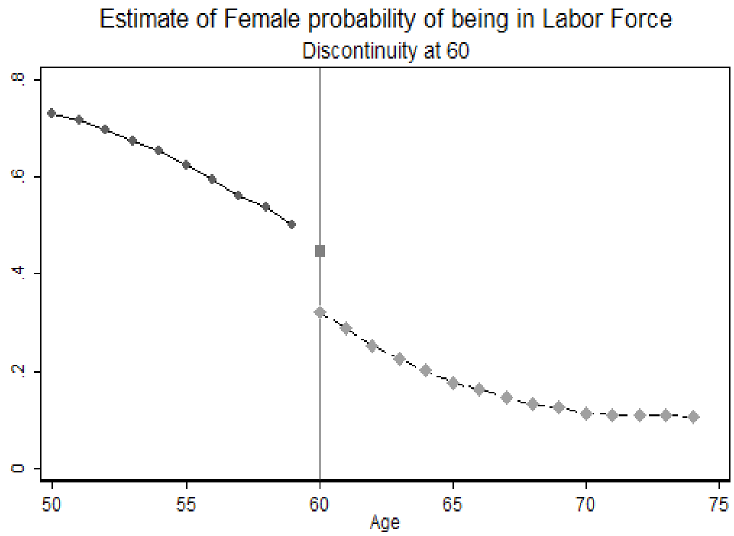


Figure 3.9d: **African Females in Labor Force: Discontinuity at 61**

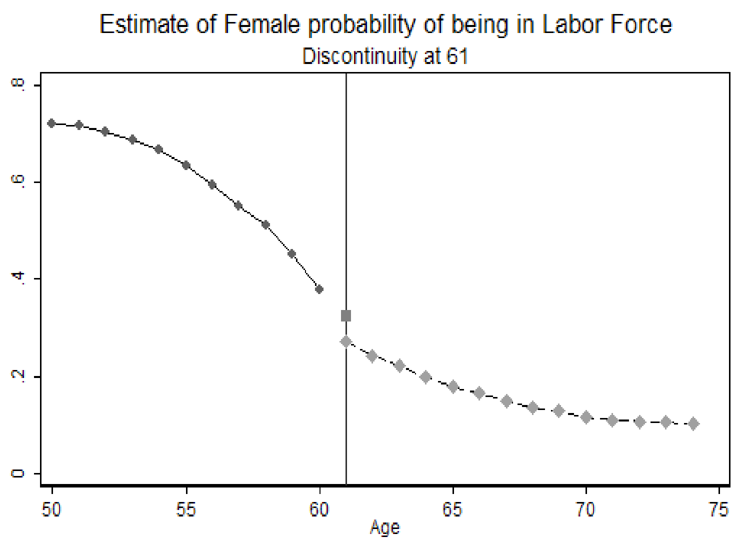
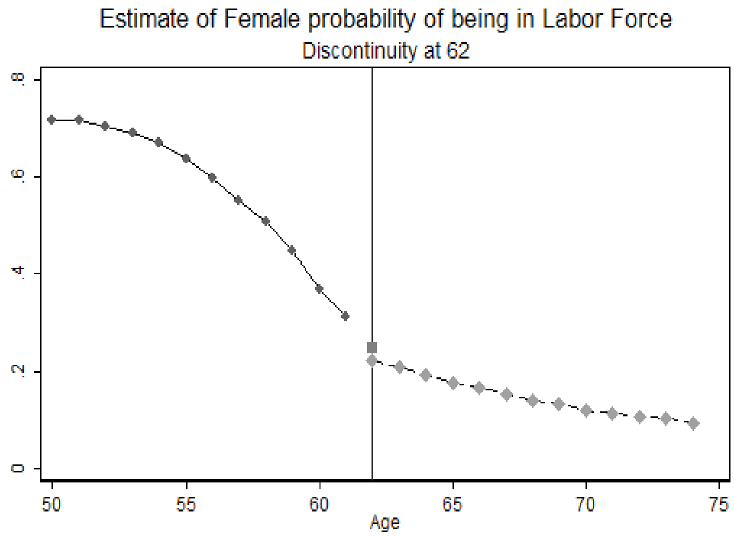


Figure 3.9e: **African Females in Labor Force: Discontinuity at 62**



Bibliography

Alderman, H, “Safety Nets and Income Transfers in South Africa”, Informal Discussion Paper Series 19335. Washington, D.C.: World Bank, 1999.

Bertrand M, Mullainathan S and Miller D, “Public policy and extended families: Evidence from pensions in South Africa”, WORLD BANK ECON REV 17 (1): 27-50 2003

Case, A and Deaton, A, “Large cash transfers to the elderly in South Africa”, ECON J, 108 (450): 1330-1361 SEP 1998

Duflo, E, “Child health and household resources in South Africa: Evidence from the Old Age Pension program”, AM ECON REV, 90 (2): 393-398 MAY 2000

Duflo, E, “Grandmothers and granddaughters: Old-age pensions and intrahousehold allocation in South Africa”, WORLD BANK ECON REV, 17 (1): 1-25 2003

Edmonds, EV, “Does Illiquidity Alter Child Labor and Schooling Decisions? Evidence from Household Responses to Anticipated Cash Transfers in South Africa”, NBER Working Paper No. w10265

Edmonds, EV., Kristin M and Douglas LM. “Rearranging the Family? Income Support and Elderly Living Arrangements in a Low-Income Country.” Journal of Human Resources 40(1): 186-207 Winter 2005.

Friedman, M. ”A Theory of the Consumption Function”, Princeton University Press, 1957

Greene, WH , “Econometric Analysis, 3rd edition”, Ch 19., Prentice Hall, 1997

Jensen, RT, “Do private transfers displace the benefits of public transfers? Evidence from South Africa”, J PUBLIC ECON 88 (1-2): 89-112 JAN 2004

Lam, D., Leibbrandt, M. and Ranchhod, V. “Labour Force Withdrawal of the Elderly in South Africa”. Paper presented at the Workshop on Ageing in Africa at the National Academy of Sciences, WITS University, 2004.

Lund, F, “State social benefits in South Africa”, International Social Security Review, 46 1993

van der Berg, S, “South African social security under apartheid and beyond”, Development Southern Africa: 14 (4): 481-503 DEC 1997

CHAPTER IV

The Impact of Colorblind Admissions on the Educational Expectations of Texas High School Graduates

4.1 Introduction

The end of affirmative action in Texas due to the 1996 *Hopwood v. Texas* decision led to precipitous drops in minority enrollment at the University of Texas at Austin and Texas A&M University-College Station. In order to reverse the decline in minority enrollment at Texas's elite public institutions, the Texas legislature passed *House Bill 588* or the Top Ten Percent Rule (TTPR) which was signed into law on May 20, 1997 by then governor George W. Bush. The Top Ten Percent Rule grants automatic admission to any public college or university in Texas for Texas high school graduates who both finish in the top decile of their graduating cohort and submit a completed application for admission to a qualifying postsecondary institution within two years of graduating.[1]

The Top Ten Percent Rule did not have a large enough effect on minority applicants to restore minority enrollment to the levels obtained prior to the *Hopwood v. Texas* decision. Texas's Attorney General interpreted the decision handed down in *Hopwood v. Texas* by the Fifth Circuit Court of Appeals to mean that race or ethnicity could factor into neither the decision to admit a student nor the decision to provide financial aid. Two selective institutions in Texas, The University of Texas at Austin and Texas A&M University-College Station, identified a set of high schools with student bodies that were,

[1] *House Bill 588* also allows each public college or university in Texas to annually determine if it will offer automatic admission to graduates in the top quartile and provides each institution with a list of eighteen factors that can be used in making admissions decisions if a student does not qualify for automatic admissions

on average, socioeconomically disadvantaged and had not had many students matriculate at the respective institutions.

The University of Texas at Austin identified 70 Texas high schools that were both poor and, which historically, had not had many students matriculate at the University of Texas at Austin. These schools, on average, consisted of student bodies that were more than ninety percent minority (Tienda and Niu, 2006). The most deserving graduates at these high schools were offered scholarships, smaller classes, and tutoring if they were admitted to the University of Texas at Austin. The funding is not exclusively for Top Ten Graduates. This program, which was introduced to selected Texas High schools in 1999, 2000, and 2001, is known as the Longhorn Opportunity Scholarship (LOS) Program.[2]

Texas A&M followed suit with its Century Scholars program which offers scholarships to the graduates of 40 high schools located in Houston, Dallas, and San Antonio. The Century Scholars (CS) Program began in the fall of 2000. The high schools were selected based on high poverty rates of their students and the low number of applications that these schools sent to Texas A&M University-College Station.[3]

The Top Ten Percent rule and the institution of the targeted financial aid and recruitment programs offer the opportunity to examine two questions:

1. What is the impact of both the increased emphasis and transparent use of class rank on the score-report sending behavior of Texas's High School Graduates? The sending of a score report to a college or university is our proxy for an application.

2. How does targeted financial aid and recruitment affect the application behavior of potential recipients?

The use of race in the admissions process remains a contentious issue across this country. California, Washington State, and, most recently, Michigan have passed bans that forbid

[2]The Longhorn Opportunity Scholarship information was obtained from Dr. Lawrence W. Burt, former associate vice president and director of student financial services at the University of Texas at Austin.

[3]The Century Scholar program information was obtained from correspondence with Myra Gonzalez, Associate Director Office of Honors Programs and Academic Scholarships at Texas A&M - College Station.

the use of race in both the admissions and financial aid processes. Knowledge of how prospective students respond to race neutral admissions regimes is vital to institutions of higher learning that are interested in maintaining a diverse student body.

Two of the selective institutions in Texas responded by both increasing recruitment efforts and offering financial aid to high schools that were likely to yield students who were members of under-represented minority groups. In addressing the second question, we explore the efficacy of such programs by asking: Do students from the targeted high schools increase the rate that they send score reports to the university that targeted the high school relative to the rate that would have prevailed had there been no intervention? By exploring the behavior of potential applicants and their responsiveness to targeted recruitment and financial aid, we will enhance our understanding of the determinants of student application behavior.

4.2 Previous Research

Card and Krueger (2004, 2005) examine the impact of ending affirmative action in California and Texas on the score report sending behavior of highly qualified minority applicants. Card and Krueger have information on all SAT takers in California and Texas from 1994–2001. Using the sending of a score report as a proxy for applying to a college, they find that highly qualified minorities—minorities with an A/A- grade point average or a score of at least 1150 on the SAT—did not substantially alter the set of institutions that they choose to send their score reports in response to the changes in admissions policies.

Long (2004), using a random sample of ten percent of all SAT I takers from 1996–2000, finds that the gap between minorities and non-minorities in the number of score reports sent to in-state public institutions widened. He simulates the effect of the change in state policies on the number of score reports of minority and non-minority students and compares the results of the simulation to the actual outcomes. The simulation predicts both a decrease in the number of score reports sent to top-tier public colleges by minorities and an increase in the number of score reports sent to top-tier colleges by non-minorities due to changes in the relative probability of admissions between minority and non-minority applicants.

Dickson (2006) analyzes the impact of the change in the admissions regime in Texas on the percent of graduates from Texas's public high schools who attempt an admissions examination, either the SAT or SAT. Using data from the Texas Education Agency's Academic Excellence Indicator System, Dickson constructs a balanced panel of high schools and estimates weighted fixed effects models which include high school level covariates—for example, the percentage of the high school that is black and the percentage of a the high school that receives free or reduced price meal. Dummies for the various admissions regimes as well as an indicator variable that assumes a value of one if the school is eligible for the Longhorn Opportunity Scholarship program are also included in the empirical specification. Dickson finds a significant decrease in the percentage of graduates taking admissions examinations after the implementation of the Top Ten Percent Rule. In addition, she finds that high schools that were selected as Longhorn Opportunity Scholarship schools experienced an increase in the percentage of graduates who attempted admissions examinations.

Niu, Tienda, and Cortes (2006) use a representative sample of Texas high school seniors in 2002 who were re-interviewed one year later to discern the effects of the Top Ten Percent Rule. First, they find that Texas seniors—and top decile graduates in particular—are sensitive to institutional selectivity. That is, Texas seniors prefer more selective institutions all else equal. Second, they find that graduates from affluent high schools are more likely than their counterparts from less-affluent high schools to apply to selective institutions. Third, they find that while there are disparities in the selectivity of colleges that blacks and Hispanics within the top decile apply to, these differences do not carry over into the actual matriculation decision.

Our paper builds on this existing literature in various ways. First, we are interested in evaluating the transition from an admissions regime where class rank was merely one factor that was used in differing degrees with respect to admissions to a regime where class rank is the primary factor in admissions. Moreover, this fact is known to all interested parties.[4] In contrast to Card and Krueger (2004, 2005), we do not limit our sample to highly qualified test-takers as measured by SAT performance or grade point average, as the Top Ten Percent rule likely impacts students who aren't highly qualified, especially those students who are from high schools with relatively low test scores.[5]

Second, our data allows us to investigate in detail the set of schools a student chose to send their scores to. Thus, whereas Long (2004) is able to only observe classifications of the colleges that are designated to receive score reports, we observe the student's full choice set.

Third, only limited evidence has been gathered in the evaluation of targeted recruitment programs. The only other paper that considers such programs is Dickson (2006). The analysis in Dickson (2006) is conducted at the high-school level and focuses on the extensive margin of test taking, the percent of graduates that attempt an admissions examination. In addition, Dickson (2006) focuses only on the LOS program. Our data allows us to examine the impact of both the LOS program and the CS program on the actual score report sending behavior of individual students, conditional on a student having attempted the SAT I examination. We believe that this provides a more direct measure of the effectiveness of such programs. If the programs worked, then the probabilities that students from either a LOS school or students from a CS school sent score reports to the University of Texas at Austin and Texas A&M-College Station, respectively, should increase relative to the probabilities for similar students from non-Longhorn high school and non-Century Scholar schools.

In addition, our empirical method for evaluating the LOS differs considerably from

[4]For example, Bucks (2004) provides evidence that the University of Texas relied heavily on class rank prior to *House Bill 588*; however, this policy was not explicit.

[5]In our data set we identify 74,472 students who self-identify as being in the top ten percent of their class and have SAT scores less than 1150, one of the benchmarks they used to identify highly qualified students.

that of Dickson (2006). In analyzing the Longhorn Opportunity Scholarship, Dickson (2006) draws inference from the dummy associated with Longhorn Opportunity Scholarship status. The fixed effects specification means that the variation used to identify the effect of the Longhorn Opportunity Scholarship comes from inter-temporal variation in Longhorn status. The counterfactual high schools are the set of Longhorn Schools prior to the schools obtaining Longhorn status and the set of high schools that never obtained Longhorn status. Later in the paper, we show that Longhorn Opportunity Scholarship Schools have a higher percentage of minority students, score lower on the SAT, and are less likely to to have graduates attempt a college admissions examination. Because the Longhorn Opportunity Scholarship schools are so different from other high schools in Texas, using the other high schools in Texas as a comparison group could lead to biased estimates of the impact of the Longhorn Opportunity Scholarship program.

We take a number of steps to reduce the bias in estimating the impact of the targeted recruitment and financial aid programs. We impose a common support condition to identify a set of non-treated schools that are "similar" to the Longhorn Opportunity Scholarship Schools. We compute the propensity score and use inverse probability weighting to identify the average effect of treatment on the treated for students who attend LOS high schools and CS high schools. Selection into treatment for these programs depended on the rate at which the high school's students sent score reports to either the University of Texas at Austin or Texas A&M and the socioeconomic status of the high school's students. We directly include measures of the first factor and proxies for the second in calculating the propensity scores. This is likely to provide better estimates of the true effects of the LOS program and CS program than estimation procedures that ignore the problem of non-random selection into treatment.

4.3 Theory

We adopt verbatim the theoretical model presented in Card and Krueger (2004). In this framework, a student needs to decide whether to apply to a particular college or not. Relevant factors are the utility of attending the school being considered, the chance of being admitted to that school, the cost of attending the school, the costs of applying, and the corresponding factors in the set of available alternative colleges.

4.3.1 The Model

A student assigns a net utility level $U_i(Q_i, C_i)$ to attending college i. Utility increases with the quality of the institution (Q_i) and decreases with the cost of attending the institution (C_i). The student estimates the probability of being admitted to school i with probability p_i, which is, for simplicity, assumed to be independent across schools. The cost of applying to any school is d and the utility of not attending college is U_0.

Optimizing behavior on the part of the student generates an application set C consisting of an ordered list of J schools with $U_1 \leq U_2 \leq \ldots \leq U_J$. A necessary condition for applying to a given school i is that $U_i > U_0$, and that the student's subjective estimate of the probability of admission is strictly positive. Let π_j represent the probability that school j is the best school in C that admits the student. Then

$$\pi_j = p_j \times \prod_{i=j+1}^{J} (1 - p_i) \tag{4.1}$$

Let $C(\sim k)$ denote the optimal choice set when school k is excluded, and $J(\sim k)$ represent the number of schools in this set. School k will be included in the final choice set if and only if

$$p_k \{ \prod_{j=0}^{J(\sim k)} \pi_j \; max[0, U_k - U_j] \} \; - \; d \; > \; 0 \tag{4.2}$$

The above equation defines the condition for including school k in the choice set. It states that college k will be included in the optimal application set if the expected return of attending college k exceeds the cost of applying.

4.4 Some predictions from the model

This simple framework is useful to clarify the effects we expect to identify empirically.[6] The introduction of the Top Ten Percent Rule changes the subjective estimation of the probability of admission conditional on rank and the type of institution under consideration. In particular, for students in the top decile, the probability of admission to any of Texas's public colleges or universities is one. For Texas high school graduates who aren't in the top decile, the impact of change in the admissions regime is not clear. Perhaps students not in the top decile perceive that their probability of admission, p_i, has declined at elite public institutions. If this is the case, then students in the lower rank classifications are more likely to submit scores to both less selective institutions within Texas and selective institutions outside of the state of Texas. The perceived change in the probability of admission will cause some colleges to be added and other colleges to be deleted relative to the optimal choice set that would have prevailed absent the change in the admissions regime. Conversely, students in the top decile should be more likely to submit score reports to more selective institutions in Texas and less likely to submit to less selective institutions.

The introduction of the Longhorn Opportunity Scholarship at UT Austin and the Century Scholars Program at Texas A&M, on the other hand, changed the expected costs of attending these two colleges for students at the eligible schools. This effectively increases the expected utility, U_i from attending one of these schools, which should increase the likelihood that the applicant places either the University of Texas at Austin or Texas A&M University in the optimal college choice set.

4.5 Data

The data we use for this paper was obtained from two sources. First, we downloaded Academic Excellence Indicator System (AEIS) data from the Texas Educational Agency (TEA). This is high school level data, available publicly on the internet, which provides

[6] A more complete discussion of the model and its implications is presented in the original paper by Card and Krueger (2004).

a wide range of information on the performance of students in each school and district in Texas for each academic year. Indicators include Texas Assessment of Academic Skills (TAAS) performance, attendance rates, dropout rates, completion rates, and SAT/ACT test results." [7]

Second, we used extremely rich student level data from the College Board. This data set contains the SAT verbal and math scores of every high school senior in the state of Texas who took the SAT exam, for the 1996–2004 cohorts. Student demographic information was also available including age, race, and gender.[8] We also utilized information on the number of test scores each student sent to colleges, as well as the name of the destination college.

Both data sets contained the name of the student's high school, enabling us to merge the two data sets and conduct analyses that include both high school and student level information. In addition, we used information provided on the SAT Questionnaire (formerly known as the Student Descriptive Questionnaire). Most students taking either the SAT Reasoning Test or any of the SAT Subject Tests also complete the optional SAT Questionnaire when they register to take the SAT Program tests, providing valuable contextual information to aid in interpreting and understanding individual and group scores. The questionnaire asks students about their family background, high school courses and performance, college aspirations, and most importantly for this study, a student's class rank. It should be stressed that the SAT Questionnaire, as well as any particular question, is voluntary. Moreover, the responses are all self-reported.

Table 4.1 provides some descriptive statistics of our data set. We include only students for whom we are able to merge these data sets. This excludes primarily students in private and charter high schools. In total, we have 916,348 remaining unique student observations

[7]The URL is "http://www.tea.state.tx.us/"

[8]We only have racial information for students from the 1999 cohort onwards

across all years combined, with a mean number of test takers of about 102,000 per year. [9] Of these, approximately 11 percent are Black, 17 percent are Hispanic and 46 percent are White, with a sizable proportion for whom we do not have racial information. Roughly 54.5 percent of takers are female, and the average performance of these Texas graduates is 458 points and 466 points on the verbal and mathematical component of the SAT, respectively.

4.6 Empirical Specifications

4.6.1 Score Report Sending

We use the inter-temporal variation in the importance of rank with respect to the probability of admissions to identify the impact of the change in the admissions regime on the score report sending behavior of Texas's public high school graduates. We estimate linear probability models of the following form:

$$Y_{ist} = \alpha_s + \varphi\alpha_t + \beta_1' X_{st} + \beta_2' X_{ist} + \pi_0 Pre + \pi_1 P + \sum \Gamma^j R_{ist}^j + \sum \delta_P^j (R_{ist}^j \times P) + \varepsilon_{ist} \quad (4.3)$$

Y_{ist}, our outcome, is an indicator variable that assumes a value of one in the event that student i in high school s in year t engages in a particular behavior with respect to score report sending. Examples of the outcomes that Y_{ist} represent include the following: does a student send a score report to a selective institution in Texas, does the student send a score report to one of the schools that must abide by the Top Ten Percent rule, does the student send at least one score report to a less-selective institution in Texas, does the student send more than four score reports in total.[10]

We include both high-school fixed effects, α_s, and a linear time trend, α_t. We include year specific high school level variables, X_{st}. X_{ist} are individual level characteristics; we include the individual's verbal SAT score and the math SAT score. Pre is an indicator variable for the 1996 cohort that applied during the pre-Hopwood regime. P is an indicator

[9] The original sample size was 1,068,071 individuals.

[10] The first four score reports are free. A student must pay a fee for each additional score report that he or she chooses to send.

variable that assumes a value of one for cohorts that are under the Top Ten Percent regime. The R^j_{ist} are a series of six rank dummies. We include dummies that assume a value of one if student i in school s in year t identifies as being in one of the following categories: the first decile, the second decile, the second quintile, the fourth quintile, the fifth quintile, and non-reported rank . The excluded category consists of students who are in the third quintile. We cluster at the high school level to produce standard errors that are both robust to using variables that are at a higher level of aggregation than the micro-units (Moulton, 1990) and to allow for arbitrary temporal correlation of the ε_{ist}'s within a cluster (Bertrand et al., 2004).

The coefficients of interest are δ^j_P, the coefficients associated with the interaction terms, $R^j_{ist} \times P$. To flesh out the interpretation of the δ^j_P consider the following example. Let Y_{ist} represent the event that a student sends to a selective university in Texas and consider the coefficient associated with the interaction term between the indicator for the top decile and the indicator dummy for the Top Ten Percent Rule admissions regime. This coefficient can be interpreted as the difference in the probability that a student in the top decile sends a score report to a selective institution relative to the probability that she would have sent a score report to a selective university in Texas prior to the advent of the Top Ten Percent Rule. Similar interpretations apply to the coefficient estimates with respect to both different rank categorizations and different outcomes.

4.6.2 Targeted Recruitment and Financial Aid

We are interested in determining the average effect of the treatment on the treated with respect to both the Longhorn Opportunity Scholarship program and the Century Scholar program. If we naively compared schools that were chosen to receive the scholarships with the schools that were not selected, then we would likely obtain biased estimates of the impact of the Longhorn Opportunity Scholarship. The schools that are selected to receive the treatment are likely to be systematically different in important ways that affect the outcome of interest. Therefore, we need to select an appropriate set of non-treated high schools that are "similar" to the set of schools that were selected to receive either the Longhorn Opportunity Scholarship program or the Century Scholar program.

The LOS program and the CS programs were designed to boost minority enrollment at the University of Texas at Austin and Texas A&M-College Station, respectively. As such, high schools where the student body was majority white were unlikely to be selected to receive either the LOS program or the CS program. We trimmed the sample of high schools with students bodies that had high concentrations of white students in the year 1996. For the LOS program, we trimmed high schools with a percentage of white students that exceeded 66 percent of the student body in 1996. For the CS program we trimmed the sample of high schools with a percentage of white students that exceeded 75 percent of the student body in 1996. The maximum value of the percentage of the student body that is white in 1996 for LOS schools and CS schools is 41.7 percent and 60.3 percent, respectively. In evaluating the LOS program, the trimming resulted in 886 high schools being dropped from consideration. In evaluating the CS program, the trimming resulted in 757 schools being dropped from consideration. The total number of high schools in our base sample is 1440.

Using the trimmed samples, we estimate the probabilities that these particular Texas high schools are selected to receive either the Longhorn Opportunity Scholarship or the Century Scholar Program, $\mathcal{P}(Longhorn|X)$ or $\mathcal{P}(Century|X)$. That is, we estimate the propensity score (Rosenbaum and Rubin, 1983) of treatment by either of the programs on the appropriate sample.[11] The common support condition means that we are interested in comparing treated and non-treated schools with similar values of $\mathcal{P}(\widehat{Longhorn}|X)$ or $\mathcal{P}(\widehat{Century}|X)$. The common support condition is implemented as follows:

$$\max(\min[\mathcal{P}(\widehat{T=1}|X), \mathcal{P}(\widehat{T=0}|X)]) \ , \ \min(\max[\mathcal{P}(\widehat{T=1}|X), \mathcal{P}(\widehat{T=0}|X)])$$

where T is either the Longhorn Opportunity Scholarship program or the Century Scholar

[11] We use a probit specification that include as regressors high school level characteristics prior to the start of the Longhorn Opportunity Scholarship program and the Century Scholar Program for the years $1996-1998$.

program. The above condition establishes the set of schools that will be used in the analysis. The common support condition for the Longhorn Opportunity Scholarship program includes high schools with estimated propensity scores in the interval .02 to .93 inclusive. The common support condition for the Century Scholar program includes high schools with estimated propensity scores in the interval .01 to .74 inclusive.

Histograms of the propensity scores for the LOS schools and CS schools are shown in Figure 4.1 and Figure 4.3, respectively. Figure 4.2 and Figure 4.4 are adjusted histograms of the propensity scores. Essentially, Figure 4.2 and Figure 4.4 remove the huge spikes near zero. This readjusts the scale of the graphs to better show the full distribution of the propensity scores.

Table 4.2*a* and Table 4.2*b* demonstrate the effects of the balancing. The LOS program and CS program schools are remarkably different from non-LOS schools and non-CS schools.

In Table 4.2*a*, Column I contains the average of a particular characteristic for non-LOS schools minus the average of the same characteristic for LOS schools. Examining the tables we see that LOS schools have student bodies with higher percentage of Hispanic students, black students, students on free or reduced price lunch, limited English proficient students and higher student-to-teacher ratios. The LOS and non-LOS schools are clearly different. Column II and Column III demonstrates how trimming and trimming in conjunction with inverse probability weighting using the propensity score, respectively, reduces the differences in the observables. Table 4.2*b* shows the effects of balancing with respect to the difference between CS and non-CS schools.[12]

We make a conditional independence assumption and employ inverse probability weighting (Horvitz and Thompson, 1952; DiNardo et al., 1996) to estimate models of the following form:

[12]We only show the balancing results for the year 1996. Similar results are obtained for the other years. This is expected as these measures are highly correlated over time.

$$
\begin{aligned}
Y_{ist} &= \alpha_s + \varphi\alpha_t + \beta_1' X_{st} + \beta_2' X_{ist} + \rho D_T + \pi_0 Pre + \pi P \\
&\quad + \sum \Gamma_j R^j_{ist} + \sum \delta^j_P (R^j_{ist} \times P) + \sum \delta^j_T (D_T \times R^j_{ist}) + \varepsilon_{ist} \qquad (4.4)
\end{aligned}
$$

Y_{ist} is an indicator variable that assumes a value of one if the student submits a score report to the University of Texas at Austin. The terms α_s, α_t, X_{st}, X_{ist}, Pre, P, R^j_{ist}, and $R^j_{ist} \times P$ are the same as the previous specification. D_T is an indicator variable that assumes a value of one the year a school becomes a Longhorn Opportunity Scholarship School when analyzing the LOS program. D_T is an indicator variable that assumes a value of one the year that a high school becomes a Century Scholar school when we analyze the CS program. $D_T \times R^j_{ist}$ is an interaction term. The coefficient associated with the interaction term, δ^j_T is the difference-in-differences estimate. However, in this case we limit the sample to students in schools that survive the trimming procedure.

We weight the students in the non-treated schools by $\frac{\mathcal{P}}{1-\mathcal{P}} \times \frac{1-\Pi}{\Pi}$ where $\mathcal{P}$ is equal to propensity score for either the LOS program or the CS program for the particular school that the student attends and Π is equal to the unconditional probability of being either a LOS school or a CS school. This weighting scheme re-weights students in non-treated schools so that, on average, they are similar to students in treated schools. With a defensible set of counterfactuals, we are able to reduce the bias in the estimates of the impact of the Longhorn Opportunity Scholarship program and the Century Scholar programs. We cluster at the level of the high school to deal with the aggregation issues raised by Moulton (1990) and the serial correlation issues raised by Bertrand et al. (2004).

4.7 Results and Discussion

Our analysis comprises of three sections. First, we summarize our data in terms of the outcome variables of interest. This provides a broad overview of trends and aggregates for high school seniors in Texas. Second, we estimate how student behavior changed in the TTPR period relative to the prior two years using multivariate regressions. We next investigate the effects of the LOS program and CS on score report sending behavior. For the TTPR the LOS program, and CS, we estimate standard OLS models as well as models that include school level fixed effects.

4.7.1 Summary Statistics

4.7.1.1 Score Sending

Table 4.3 presents the mean values of various score sending outcomes. We see that the mean number of scores sent for each SAT taker declined between 1996 and 2004, from 4.29 scores to 3.86 scores, an 11 percent decline. This is also reflected in the mean numbers sent to Texas based institutions. In terms of the selective schools we consider, the proportion of SAT takers sending to the University of Texas at Austin drops by 4 percentage points from 35.5 percent to 31.5 percent, the proportion to the University of Texas at Dallas grew by 2.7 percentage points from 3.1 percent to 5.8 percent, while the proportion sending to Texas A&M decreased by 7 percentage points from 31.2 percent to 23.3 percent. On average, the proportion sending to any of these three schools decreased by 7.1 percentage points, or about 14 percent relative to the 1996 level. Thus, we observe that both the number of scores sent, and the proportion sending to selective schools decreased markedly in the period of our study.

In Table 4.4*a*, we observe an increasing trend in the proportion who refuse to answer the entire Student Descriptive Questionnaire (SDQ), although this occurs largely in 2003 an 2004. At the same time, there is a gradual and persistent upward trend in the proportion who answer parts of the SDQ, but refuse to respond to the question on class rank. The increase is large, from 8651 in 1996 to 33829 in 2004, which is almost a quadrupling in absolute number. Indeed, by 2004, more than a third of takers do not respond to the class rank question. This is problematic for our analysis, as a significant portion of our analysis is related to the class rank variable. Thus, all results making reference to the class rank variable pertains only to those students who provided us with such information.

Table 4.4*b* captures the mean proportion taking the SAT by class rank. This was computed for each school by dividing the total number of SAT takers reporting a particular rank in a given school in a given year, by the number of twelfth graders in that particular rank in that same school in that same year. The number of twelfth graders was obtained from Texas's Academic Excellence Indicator System (AEIS) data set. The mean was calculated by averaging across schools within a year, where each school received equal weighting in the calculation. We observe that the mean proportion in the first decile exceeds one for 1996 and 1997. Therefore, there has to be some misreporting in the data, with students systematically overstating their rank. Within each rank, the proportion taking the SAT decreased, although this is possibly due to the increased non-response rate for the class rank question. Nevertheless, we do observe sustained downward trends in the proportions taking an SAT conditional on them answering the class rank question. There are also large differences in test taking across rank classifications, which is to be expected. In 2004, the ratio of students who reported being in the top decile relative to one tenth of the number of twelfth graders is 0.90, compared to an analogous ratio of 0.376 for the second quintile, and only 0.051 for students in the lowest quintile.

Tables 4.4*c* and 4.4*d* provide the mean performance on the SAT verbal and mathematics components respectively, in each rank category for each year. Somewhat reassuring is that the performances are monotonically decreasing in rank, amongst those who do report a rank. We also observe that 2002 seems to be an outlier year, in that performance on both the verbal and mathematics component of the SAT greatly exceed those in other years. These tables also indicate that the students for whom we have no class rank information are not likely to be randomly drawn from the population of SAT takers in terms of their scholastic abilities in the period from 2002 to 2004. Prior to 2002, these students were on average likely to be placed somewhere between the 2nd and 3rd quintiles, whereas thereafter, their mean performance places them between the 2nd decile and the 2nd quintile. [13]

Tables 4.5a – 4.5i summarize the various dimensions of score sending behavior within each class rank category. This allows us to observe any heterogeneity in the trends that occur as a function of a student's relative rank within their high school cohort. The mean number of scores sent decreases for most categories, although this is largest for those in the first decile (Table 4.5a). The mean number of scores sent to colleges within Texas decreased as well, most markedly in the first and second deciles, but was stable in the fourth quintile and actually increased for those in the fifth quintile (Table 4.5b). Table 4.5c shows the mean number of scores sent to schools that are legally compelled to follow the Top Ten Percent rule. For all students who do report a valid class rank variable, this number remains fairly stable. The next table concerns score sending to Texas two year colleges. The mean numbers here are also relatively stable. [14]

Tables 4.5e – 4.5g summarize the proportion sending to each of the three selective schools we consider, namely UT Austin, UT Dallas and Texas A&M. Table 4.5h presents the mean number of scores sent to this set of selective schools on aggregate, and Table 4.5i contains the proportion of students who send to at least one of these selective schools. Students in the top decile in 2004 are 2.2 percentage points more likely to send to UT Austin than similar students in 1996, while students in the other reported ranks in 2004 are generally less likely to send to Austin than their 1996 counterparts. [15] Score sending to UT Dallas increases in every category, although this occurs off of a rather small initial base. In all reported class ranks we observe a clear decrease in the proportion sending to Texas A&M, with large decreases of about nine percentage points in the first two deciles

[13] We use deciles for the first and second deciles, and quintiles for the remainder, as this is how the question that asks a student to report her rank is structured.

[14] Two year colleges are not the primary focus of this paper. Moreover, the SAT is taken primarily for applications to selective colleges. Thus, we do not comment in depth on the application decision for two year colleges.

[15] An exception is students in the 5th quintile, although these students represent a very small percentage of the population of SAT takers in any year.

as well as the second quintile. Table 4.5*i* indicates that approximately seventy percent of students in the first decile send to at least one selective school, and that this remains stable during the period of analysis. However, students in the other reported class rank categories generally show marked declines in the probability of sending to a selective school. This is most pronounced for students in the 2nd quintile, where the proportion decreases by about 10 percentage points, or almost 20 percent of the 1996 base proportion. This is in line with the model provided above, as those far from the first decile will need to compete aggressively in order to gain admission into selective schools compelled to admit anyone in the top 10 percent of their cohort.

4.7.1.2 Longhorn and Century Scholars Programs

Tables 4.6*a* – 4.6*c* show the number of test takers, the proportion of twelfth graders taking an SAT, and the proportion of SAT takers who send a score to UT Austin respectively. Table 4.6*a* shows that regardless of the scholarship program being administered at a school, if any, the number of takers increased with time. We also observe that more than 10 percent of all public school SAT takers attend a school where the LOS was subsequently implemented. The proportion taking an SAT increases dramatically in the LOS schools, by more than 20 percent, actually decreases in the CS only schools, and increases by a few percentage points in the non-scholarship programs. Moreover, the large increases observed in the LOS schools coincides with the time period when the LOS was introduced. Further suggestive evidence that LOS caused these responses is seen in Table 4.6*c*. Whereas the proportion of score senders sending to UT Austin actually decreased with time by about 5 percentage points between 1996 and 2004, the proportion from LOS schools increased by about 3 percentage points. These patterns are observed in Table 4.7 as well, where we summarize the schools using a binary classification of LOS or non-LOS schools, ignoring the particular potential school's CS status.

4.7.2 Top Ten Percent Rule

4.7.2.1 Score Sending

Table 4.8 presents individual level regression results for score sending behavior. Of primary interest to us are the coefficients on the post and post interacted with class rank variables. The dependent variable in the second column is an indicator variable for whether the student sent more than four SAT score reports. In the post period, students in the 1st decile were 5.6 percentage points less likely to send out more than four scores. Those in the 2nd decile and second quintile were also significantly less likely to send out more than four reports, but the decrease is less than 2 percentage points. In the third column, we model the actual number of scores sent. Most students send fewer scores on average in the post period. However, this again is most pronounced amongst students who are in the 1st decile, who send 0.353 [16] fewer scores out. This is as predicted by the theory, since these students are guaranteed admission to any of the top public colleges in Texas. In the fourth column, the dependent variable is an indicator variable that takes a value of one if a student sends a score to one of the non-selective four year colleges in Texas. In the post period, students in the 1st decile are significantly less likely to apply to a non-selective school. At the same time, those in the 4th and 5th quintiles who do take an SAT have a significantly increased probability of sending to a non-selective college. This suggests that students of all ranks are being forced by the law to pre-select themselves into a particular type of college more strongly than prior to the law. The estimates in the sixth column shows that students in the first decile have an increased probability of sending to UT Austin in the post-TTPR period, of 3.8 percentage points, while students in the 2nd quintile have a small, but significant reduction in this probability. Interestingly, students in the 5th quintile show a large increase in the probability of sending there, although we observed that this group is small and likely to be selected in terms of other unobservables. Very little changes for the probability of sending to a school in-state, or a school to which

[16] $-0.105 + -0.248$

the TTPR applies.

Table 4.9 presents models with the same dependent variables and same set of covariates, but includes school level fixed effects. Of note in comparison to Table 4.8 is that the coefficients are very similar, both in terms of magnitude and significance. By and large, the theoretical framework predictions of the impact of the law are substantiated by the empirical analysis.

4.7.3 Longhorn Opportunity Scholarship

This component of our analysis looks at the effects of the targeted recruitment and financial aid packages offered by the University of Texas at Austin under the LOS program. This program targeted high schools with both socioeconomically disadvantaged students and graduates who send standardized test scores to UT Austin at a rate below the average rate at which Texas high school graduates send score reports to UT Austin.

Table 4.10 presents difference-in-differences (DD) results for various score sending behaviors as a function of the LOS program. We further allow for these effects to differ depending on the class rank of the students. Interpreting the coefficients on the interaction terms, we observe that most of the effects manifest in the first decile of students. The probability that a student in the first decile at a Longhorn school sends to Austin increases by 9.7 percentage points in the DD model. In 1999, 20.0 percent of test takers from LOS schools sent at least one score report to the University of Texas at Austin. Relative to the 1999 level, the change in the first decile is a 48.5% increase. The coefficients associated with the interaction of being LOS and self-identifying as being in the second decile increase the probability of increasing of sending to to the University of Texas at Austin by 3.4 percentage points. The coefficient is statistically significant at the one percent level.

The second dependent variable is the total number of score reports sent. The coefficients associated with the interaction terms are all small and statistically insignificant.

The third dependent variable is an indicator variable that assumes a value of one if a student sends a score report to a non-selective four year institution in Texas.[17] Surprisingly, the coefficient associated with the interaction term between being in a LOS school and a self-reported rank of being in the top decile is 4.9 percentage points and is statistically significant at the one percent level.

The fourth dependent variable is a set of selective schools in Texas.[18] The coefficient associated with the interaction term between LOS and a self-reported rank of the top decile is 8.2 percentage points and is statistically significant at the one percent level.

The fifth dependent variable is a dummy variable that assumes a value of one if the test-taker chooses to send at least one score report to the set of colleges in Texas that must abide by the Top Ten Percent Rule. The difference-in-differences estimate of the effect of the LOS program on the first decile is 5.4 percentage points, and it is statistically significant at the one percent level.

The sixth dependent variable is an indicator variable that assumes a value of one if a test-taker sends at least one score report to the a school in Texas. Students in the top decile at a LOS school are 1.2 percentage points more likely to send to an in-state college or university. This effect is relatively small as a large proportion of test takers send at least one score report to an in-state school.

The final dependent variable is an indicator variable that assumes a value of one if a test-taker chooses to indicate more that four colleges to receive score reports. SAT test takers in the first decile at LOS schools are 3.5 percentage points more likely to send more than four score reports. We find small and statistically insignificant effects of the LOS program on the probability of designating more than four colleges to receive score reports for the remaining interaction terms.

Table 4.11 presents the results from the difference-in-differences model with fixed effects. The results we obtain are fairly similar. Test-takers who self-identify as being in the top decile at schools under the LOS program are 9.9 percentage points more likely to submit

[17] Colleges are classified as non-selective according to the system used by Barron's Guide to Colleges and Universities

[18] Selective schools include the University of Texas at Austin, Texas A&M-College Station, and the University of Texas at Dallas.

a score report to the University of Texas at Austin; the coefficient is significant at the one percent level. Students who self-identify as being in the second decile at LOS schools are 3.3 percentage points more likely to send a score report to the University of Texas at Austin; the coefficient is significant at the one percent level. Students who self-identify as being in the second quintile at LOS schools are 2.4 percentage points more likely to send a score report to the University of Texas at Austin.

With respect to non-selective institutions, the difference-in-differences estimate of the impact of the LOS program on the probability of a test-taker in the top decile submitting to a non-selective dropped to 3.9 percentage points, roughly a 20.0 percent decline relative to the estimate from the model without fixed effects. Still, the estimate is statistically significant at the five percent level.

The fixed effects difference-in-differences estimate of the impact of the LOS program on the probability of a test taker in the top decile submitting to a selective institution in Texas is 8.7 percentage points an increase of .5 percentage points relative to the estimate from the model without school fixed effects; the estimate is statistically significant at the one percent level. The estimate of the difference-in-differences estimate of the impact of the LOS program on the probability of students in the top decile submitting to a TTPR school is 5.5 percentage points and is significant at the one percent level. The finding that students in the top decile at LOS schools are 1.1 percentage more likely to remain in state is significant at the five percent level.

Overall, this is a remarkable finding. The impact of the LOS is large and significant. It also has the largest impact on the best ranked students in these schools. Indeed, the effects on weaker students are small and often statistically indistinguishable from zero. This is consistent with the hypothesis that student application decisions from under-represented schools are subject to multiple constraints. These include both the likelihood of admission as well as funding and post-enrollment support.

4.7.4 Century Scholars Program

This part of the analysis will focus on the Century Scholars Program. Table 4.12 contains the results of the difference-in-differences models without school fixed effects.

The first dependent variable is an indicator variable that assumes a value of one if the student sends a score report to Texas A&M-College Station. The difference-in-differences estimate of the impact of the Century Scholar program on the likelihood that a test-taker who self identifies as being in the first decile sends a score report to Texas A&M College Station is 3.9 percentage points, and is significant at the five percent level. This is considerably smaller than the analogous estimate for the LOS program. None of the other interaction terms are significant.

The second dependent variable is the total number of scores sent. The difference-in-differences estimates of the impact of the CS program on number of score reports sent by the first and second decile are $-.287$ and $-.251$ score reports. The estimates are both significant at the one percent level.

The third dependent variable is an indicator variable that assumes a value of one if a test-taker sends to a non-selective school. For students who self-identify a class ranking in the third and fourth quintiles, the difference-in-differences estimates take on values of -5.7 and -8.2 percentage points respectively. The estimates are significant at the one percent level. The CS program seems to reduce the probability that lower ranked students send score reports to non-selective institutions.

The Century Scholar program increases the probability that a student who chooses not to indicate rank sends a score report to a selective institution by 3.3 percentage points. This estimate is significant at the one percent level. This result is not surprising given the positive impact of the CS program on the likelihood that a student who does not indicate rank applies to Texas A&M. Texas A&M is one of the selective schools. The difference in differences estimates for the impact of the CS program on the probability that a student who identifies as being in ranked in the top decile is 4.0 percentage points; the estimate is significant at the one percent level.

Students in the top decile at CS high schools are 3.6 percentage points more likely to apply to a school with admissions that are subject to the Top Ten Percent rule. The difference-in-differences estimates for the impact of the CS program on the likelihood that test-takers of various ranks send score reports to Texas are small and statistically insignificant.

The CS program does seem to impact the probability that students from in the top decile and second deciles send more than four score reports. The difference-in-differences estimate for the test-takers who identify as being in the top decile is -4.2 percentage points and is significant at the one percent level. The estimate of the impact of the program on the probability of sending more than four score reports for a student whose self-reported rank is the second decile is -5.0 percentage points; the estimate is statistically significant at the one percent level.

Table 4.13 contains the results for the difference-in-differences models that include high school level fixed effects. Students in the top decile are 4.2 percentage points more likely to send a score report to Texas A&M-College Station, and the estimate is statistically significant at the five percent level. This is similar to the estimate we obtained without fixed effects.

The difference-in-differences estimates of the impact of the CS program on number of score reports sent by the first and second decile are $-.270$ and $-.229$ score reports. The estimates are both significant at the one percent level.

The difference-in-difference estimates for the test-takers in the third quintile and test-takers who identify as being in the fourth quintile remain remain statistically significant, with values of -4.2 percentage points and -6.2 percentage points, respectively.

Test-takers in CS schools who do not indicate rank or self-report being in the top decile are more likely to send score reports to selective institutions. The estimates are similar to the estimates obtained from the model without school fixed effects. The estimate for students who don't report rank is 3.4 percentage points and is statistically significant at the five percent level. The difference-in-differences estimate for students in the top decile is 4.5 percentage points; the estimate is significant at the one percent level. The other differences-in-differences estimates are similar in magnitude relative to the model with no fixed effects, and are also not statistically significant.

The estimate of the impact of the CS program on the likelihood that a test-taker who self-reports as being in the top decile sends a score report to a TTPR schools remains nearly the same across specifications. The estimate obtains a value of 3.7 percentage points and is significant at the one percent level. The addition of school fixed effects does not change appreciably change the estimates of impact of the CS program on the likelihood that test-takers send score reports to in-state schools. The estimates remain small and statistically insignificant.

Test takers in the top decile and the second decile at CS schools are less likely to send out more than four score reports, estimates for these ranks take on values of −4.2 percentage points and −4.8 percentage points; both estimates are statistically significant at the one percent level. The sign of the difference-in-differences estimate for the second quintile is of similar magnitude to the estimates without fixed effects.

The Century Scholar program, while not as successful as the Longhorn Opportunity Scholarship program, does appear to attract highly ranked students to Texas A&M. The magnitude of the impact of the LOS program on the likelihood that a student who self reports as being in the top decile at a LOS school sends a score report to the University of Texas at Austin is 136 percent larger than the estimate of the impact of the CS program on the likelihood that a student who self-reports as being in the top decile in a CS school sends to Texas A&M-College Station.

There is evidence that the CS program reduces the probability that low ranked students send score reports to non-selective schools.We find evidence that the CS program increases the probability that both students who don't report rank and students who report being in the top decile send score reports to selective institutions. The CS program increases the probability that test-takers in the top decile send a score report to colleges that must abide by the Top Ten Percent rule; although the Longhorn Opportunity Scholarship Program has a larger impact on the probability that students in the top decile send score reports to Top Ten Percent colleges. The CS program also appears to reduce the probability that highly ranked students send more than four score reports.

4.8 Conclusion

Our analysis and results highlighted some important and useful new insights. First, the theoretical framework we adopted was useful in understanding the changes that were observed in student college applications behavior. Second, as predicted, there exists significant heterogeneity in the effects of the law, depending on a student's class rank. Previous research has not investigated this.

Third, the targeted recruitment program implemented by UT Austin was extremely successful. This success was limited mostly to the best students in these schools. The targeted recruitment program implemented by the Century Scholar program was also effective at increasing the likelihood that students send score reports to Texas A&M, but slightly less so than the LOS program.

Our evidence suggests that students from poor schools face multiple barriers to obtaining postsecondary schooling at selective colleges. Using the combination of the TTPR and the LOS and CS programs, we find strong evidence that it is possible for targeted recruitment programs to attract students from such schools.

4.9 Tables

Table 4.1: **Sample Size and composition**

Year	# of Obs	% Black	% Hisp	% Cauc.	% Female	Mean SAT-V	Mean SAT-M
1996	83,769	–	–	–	54.7	460	466
1997	87,750	–	–	–	54.73	459	467
1998	94,136	–	–	–	54.84	456	463
1999	98,730	11.45	16.54	51.95	54.73	453	459
2000	103,367	11.11	16.57	49.46	54.41	454	461
2001	105,015	11.22	16.73	47.87	54.57	452	459
2002	110,097	11.23	16.58	45.07	54.57	488	499
2003	115,260	10.84	16.37	40.48	54.02	452	460
2004	118,224	12.07	18.28	43.53	54.19	450	458

Table 4.2a: **Difference in Means between LOS and non-LOS schools: 1996**

	I		II		III	
	Full Sample		Trimmed on % White		Trimmed + IPW	
	Diff	Std. Error	Diff	Std. Error	Diff	Std. Error
SAT Verbal 1996	84.2	11.3	60.6	12.9	-76.0	41.0
SAT Math 1996	78.7	11.5	56.9	13.2	-76.6	42.5
% Male1996	0.03	0.02	0.02	0.02	0.07	0.07
% LEP 1996	-9.08	1.06	-5.46	1.45	-1.92	2.91
% Poor 1996	-21.61	2.47	-11.18	2.69	-4.55	6.26
% White 1996	56.14	3.32	31.38	2.54	-0.61	1.67
% Black 1996	-28.67	1.96	-24.30	2.67	-4.35	10.69
% Hispanic 1996	-27.49	3.47	-7.45	3.86	5.57	10.35
% Twelfth Grade 1996	1.29	0.82	1.78	0.85	-1.77	0.89
Teacher Exper. 1996	-1.52	0.30	-1.31	0.31	-0.85	0.73
Stud to Teach. ratio 1996	-3.49	0.40	-2.97	0.46	0.89	0.99
% Taking Exam 1996	9.45	2.30	6.95	2.45	-11.09	6.61
Sent to Austin 1996	0.06	0.02	0.09	0.02	-0.04	0.04
Sent to UTD 1996	-0.01	0.01	-0.01	0.01	-0.02	0.01
Sent to A&M 1996	0.14	0.02	0.13	0.02	0.10	0.10

Notes:
IPW is our acronym for Inverse Probability Weighting.

Table 4.2b: **Difference in Means between CS and non-CS schools: 1996**

	I		II		III	
	Full Sample		Trimmed on % White		Trimmed + IPW	
	Diff	Std. Error	Diff	Std. Error	Diff	Std. Error
SAT Verbal 1996	68.2	15.0	53.0	16.3	-07.0	27.7
SAT Math 1996	59.7	15.1	46.6	16.5	-14.2	28.0
% Male 1996	0.02	0.03	0.01	0.03	-0.01	0.02
% LEP 1996	-8.32	1.40	-5.86	1.72	-3.77	2.43
% Poor 1996	-8.94	3.32	-1.42	3.44	6.12	3.86
% White 1996	52.03	4.59	33.75	3.83	6.91	2.96
% Black 1996	-43.78	2.47	-39.99	3.00	-14.09	8.19
% Hispanic 1996	-6.46	4.64	7.79	4.89	8.45	6.96
% Twelfth Grade 1996	1.97	1.07	2.36	1.07	0.97	1.11
Teacher Exper. 1996	-1.79	0.40	-1.49	0.39	-0.88	0.66
Stud. to Teach. ratio 1996	-4.25	0.52	-3.82	0.56	-0.49	0.54
% Taking Exam 1996	7.14	2.96	5.65	3.04	-3.99	3.57
Sent to Austin 1996	0.06	0.03	0.08	0.03	0.13	0.04
Sent to UTD 1996	-0.02	0.01	-0.02	0.01	-0.02	0.01
Sent to A&M 1996	0.10	0.03	0.10	0.03	0.03	0.02

Notes:
IPW is our acronym for Inverse Probability Weighting.

Table 4.3: **Score Report Sending Behavior**

	Mean Number of Scores Sent:				Proportion of Scores Sent:			
Year	In Total	In State	To TTPR schools	To TX 2yr college	To UT Austin	To UT Dallas	To Texas A&M	To selective TX college
1996	4.29	3.11	1.75	0.34	0.355	0.031	0.312	0.514
1997	4.27	3.10	1.74	0.35	0.343	0.034	0.296	0.498
1998	4.08	2.97	1.67	0.34	0.334	0.034	0.281	0.480
1999	3.98	2.91	1.64	0.33	0.330	0.034	0.278	0.475
2000	4.05	2.97	1.70	0.34	0.352	0.043	0.288	0.499
2001	4.06	2.95	1.69	0.33	0.337	0.046	0.281	0.484
2002	4.02	2.95	1.70	0.32	0.328	0.053	0.265	0.470
2003	3.97	2.93	1.72	0.31	0.332	0.056	0.253	0.466
2004	3.86	2.88	1.69	0.31	0.315	0.058	0.233	0.443

Table 4.4a: **Number of Taker within class rank**

year	Q not ans	top decile	2nd decile	2nd quintile	3rd quint	4th quint	5th quint	SDQ not ans	Total
1996	8,651	15,678	16,286	21,162	16,553	3,064	604	1,771	83,769
1997	9,528	16,182	17,093	21,674	16,983	3,308	630	2,352	87,750
1998	11,019	16,838	17,582	23,204	18,727	3,569	695	2,502	94,136
1999	13,229	17,311	18,483	23,776	19,133	3,730	773	2,295	98,730
2000	16,677	17,912	19,051	23,192	19,232	3,797	823	2,683	103,367
2001	22,258	18,202	18,623	21,438	17,792	3,495	729	2,478	105,015
2002	31,548	17,817	18,218	19,310	16,724	3,381	763	2,336	110,097
2003	27,176	17,693	17,510	17,567	15,338	3,098	704	16,174	115,260
2004	33,829	18,933	18,846	17,622	15,858	3,406	807	8,923	118,224

Note the increasing proportion that do not provide rank

Table 4.4b: **Proportion Taking conditional on rank reported**

year	top decile	2nd decile	2nd quintile	3rd quint	4th quint	5th quint
1996	1.02	0.93	0.52	0.43	0.13	0.06
1997	1.01	0.93	0.52	0.42	0.12	0.06
1998	0.98	0.90	0.51	0.43	0.13	0.06
1999	0.95	0.91	0.50	0.42	0.12	0.05
2000	0.97	0.89	0.50	0.42	0.13	0.07
2001	0.96	0.89	0.46	0.40	0.12	0.05
2002	0.92	0.87	0.43	0.37	0.13	0.06
2003	0.90	0.81	0.38	0.34	0.10	0.05
2004	0.90	0.83	0.38	0.34	0.11	0.05

Table 4.4c: **Mean of SAT Verbal by class Rank**

year	Q not ans	top decile	2nd decile	2nd quintile	3rd quint	4th quint	5th quint	SDQ not ans	Overall
1996	412.59	560.67	491.05	451.29	400.73	354.39	323.00	404.24	460.04
1997	409.32	558.26	489.88	451.78	401.79	359.40	312.10	408.78	458.91
1998	414.18	557.35	489.49	450.18	396.57	351.26	309.63	398.29	455.64
1999	420.49	558.94	486.29	445.38	395.49	345.47	304.42	371.25	453.35
2000	424.51	556.36	486.25	446.74	396.64	347.87	303.03	377.37	453.53
2001	430.66	555.35	482.66	441.31	393.78	342.59	312.73	363.80	452.09
2002	483.85	567.01	503.02	468.48	433.83	403.11	391.88	426.55	487.75
2003	432.26	553.92	478.71	431.74	382.32	331.86	297.32	463.76	452.16
2004	426.76	552.46	480.72	430.55	385.40	336.66	304.36	464.98	449.96

Table 4.4d: **Mean of SAT Math by class Rank**

year	Q not ans	top decile	2nd decile	2nd quintile	3rd quint	4th quint	5th quint	SDQ not ans.	Overall
1996	409.78	579.89	501.58	455.39	399.39	354.20	323.39	410.01	466.28
1997	409.55	578.17	501.47	458.20	402.92	358.52	318.43	411.77	466.77
1998	413.80	576.98	500.83	456.63	397.77	350.07	313.60	406.12	463.25
1999	420.04	575.89	496.49	449.57	394.06	343.22	301.51	376.52	458.92
2000	426.37	575.69	498.25	453.16	397.10	343.45	303.78	383.75	460.93
2001	432.98	574.87	494.29	447.49	393.59	339.02	309.14	370.52	459.27
2002	490.49	588.62	518.06	479.01	436.91	405.01	400.19	437.22	498.54
2003	435.62	573.47	491.72	437.26	383.36	331.02	296.58	472.41	460.10
2004	429.93	568.76	492.44	437.97	386.93	335.48	308.20	478.64	457.68

Table 4.5a: **Number of Scores Sent**

year	Q not ans	top decile	2nd decile	2nd quintile	3rd quint	4th quint	5th quint	SDQ not ans	Overall
1996	3.36	5.39	4.63	4.24	3.83	3.43	3.26	2.80	4.29
1997	3.39	5.39	4.55	4.25	3.81	3.51	3.22	2.77	4.27
1998	3.22	5.24	4.43	4.05	3.61	3.26	3.08	2.75	4.08
1999	3.21	5.06	4.3	3.98	3.59	3.24	3.12	2.53	3.98
2000	3.32	5.06	4.41	4.11	3.71	3.39	3.16	2.51	4.05
2001	3.32	5.09	4.45	4.14	3.77	3.49	3.33	2.63	4.06
2002	3.43	5.00	4.41	4.17	3.81	3.53	3.35	2.74	4.02
2003	3.50	4.94	4.36	4.12	3.80	3.57	3.34	3.380	3.97
2004	3.29	4.77	4.22	3.98	3.67	3.41	3.39	3.73	3.86

Table 4.5b: **Number of Scores Sent within Texas**

year	Q not ans	top decile	2nd decile	2nd quintile	3rd quint	4th quint	5th quint	SDQ not ans	Overall
1996	2.58	3.38	3.33	3.21	3.00	2.72	2.58	2.15	3.11
1997	2.59	3.40	3.30	3.22	2.98	2.74	2.47	2.15	3.10
1998	2.45	3.34	3.21	3.07	2.81	2.56	2.46	2.15	2.97
1999	2.41	3.19	3.14	3.03	2.81	2.56	2.41	1.97	2.91
2000	2.53	3.22	3.22	3.11	2.91	2.67	2.47	1.99	2.97
2001	2.52	3.17	3.23	3.11	2.93	2.71	2.69	2.05	2.95
2002	2.58	3.16	3.26	3.19	2.98	2.81	2.67	2.15	2.95
2003	2.68	3.16	3.24	3.16	2.98	2.81	2.62	2.52	2.93
2004	2.54	3.11	3.17	3.13	2.92	2.72	2.71	2.61	2.88

Table 4.5c: **Number of Scores Sent to TTPR schools**

year	Q not ans	top decile	2nd decile	2nd quintile	3rd quint	4th quint	5th quint	SDQ not ans	Overall
1996	1.42	1.82	1.90	1.88	1.70	1.41	1.26	1.16	1.75
1997	1.41	1.83	1.90	1.87	1.68	1.43	1.16	1.17	1.74
1998	1.32	1.80	1.85	1.80	1.59	1.32	1.19	1.14	1.67
1999	1.34	1.75	1.81	1.76	1.60	1.36	1.19	1.02	1.64
2000	1.43	1.80	1.88	1.83	1.66	1.43	1.23	1.07	1.70
2001	1.44	1.78	1.90	1.83	1.67	1.46	1.40	1.14	1.69
2002	1.49	1.80	1.94	1.87	1.68	1.50	1.31	1.15	1.70
2003	1.56	1.83	1.95	1.90	1.70	1.50	1.24	1.53	1.72
2004	1.47	1.82	1.93	1.86	1.68	1.46	1.31	1.57	1.69

Table 4.5d: **Number of Scores Sent to TX 2yr colleges**

year	Q not ans	top decile	2nd decile	2nd quintile	3rd quint	4th quint	5th quint	SDQ not ans	Overall
1996	0.39	0.17	0.28	0.35	0.48	0.62	0.67	0.36	0.34
1997	0.41	0.17	0.29	0.37	0.48	0.61	0.68	0.37	0.35
1998	0.38	0.18	0.27	0.35	0.44	0.58	0.63	0.40	0.34
1999	0.36	0.16	0.27	0.35	0.44	0.54	0.56	0.39	0.33
2000	0.36	0.16	0.28	0.36	0.47	0.59	0.62	0.40	0.34
2001	0.33	0.15	0.27	0.35	0.46	0.55	0.64	0.43	0.33
2002	0.29	0.14	0.27	0.36	0.46	0.59	0.68	0.49	0.32
2003	0.32	0.14	0.28	0.37	0.48	0.62	0.70	0.25	0.31
2004	0.32	0.14	0.26	0.37	0.47	0.59	0.69	0.24	0.31

Table 4.5e: **Number of Scores Sent to UT Austin**

year	Q not ans	top decile	2nd decile	2nd quintile	3rd quint	4th quint	5th quint	SDQ not ans	Overall
1996	0.25	0.51	0.42	0.36	0.25	0.17	0.12	0.22	0.35
1997	0.24	0.49	0.40	0.35	0.25	0.18	0.12	0.20	0.342
1998	0.24	0.49	0.41	0.33	0.24	0.16	0.12	0.20	0.33
1999	0.26	0.49	0.40	0.32	0.24	0.17	0.14	0.18	0.33
2000	0.29	0.51	0.42	0.35	0.26	0.18	0.15	0.18	0.35
2001	0.27	0.52	0.41	0.33	0.23	0.17	0.17	0.17	0.34
2002	0.28	0.53	0.40	0.30	0.21	0.16	0.14	0.16	0.33
2003	0.27	0.54	0.40	0.31	0.21	0.15	0.13	0.31	0.33
2004	0.24	0.53	0.39	0.28	0.20	0.13	0.15	0.34	0.32

Table 4.5f: **Number of Scores Sent to UT Dallas**

year	Q not ans	top decile	2nd decile	2nd quintile	3rd quint	4th quint	5th quint	SDQ not ans	Overall
1996	0.03	0.04	0.03	0.03	0.03	0.02	0.02	0.02	0.03
1997	0.03	0.04	0.03	0.04	0.03	0.03	0.03	0.01	0.03
1998	0.03	0.04	0.04	0.03	0.0	0.03	0.04	0.01	0.03
1999	0.03	0.04	0.04	0.04	0.03	0.03	0.03	0.01	0.03
2000	0.04	0.04	0.05	0.05	0.04	0.04	0.04	0.02	0.04
2001	0.04	0.05	0.05	0.05	0.05	0.05	0.05	0.02	0.05
2002	0.05	0.05	0.06	0.06	0.05	0.05	0.06	0.04	0.05
2003	0.05	0.06	0.06	0.06	0.05	0.04	0.04	0.05	0.06
2004	0.05	0.07	0.07	0.06	0.06	0.05	0.04	0.06	0.06

Table 4.5g: **Number of Scores Sent to Texas A&M**

year	Q not ans	top decile	2nd decile	2nd quintile	3rd quint	4th quint	5th quint	SDQ not ans	Overall
1996	0.21	0.45	0.38	0.31	0.21	0.13	0.12	0.18	0.31
1997	0.20	0.43	0.37	0.30	0.20	0.13	0.10	0.16	0.30
1998	0.19	0.41	0.36	0.28	0.19	0.13	0.10	0.18	0.28
1999	0.20	0.41	0.35	0.28	0.19	0.12	0.11	0.15	0.28
2000	0.22	0.42	0.36	0.29	0.20	0.13	0.12	0.15	0.29
2001	0.22	0.42	0.36	0.28	0.20	0.13	0.12	0.13	0.28
2002	0.23	0.40	0.34	0.26	0.17	0.12	0.07	0.12	0.27
2003	0.20	0.39	0.33	0.24	0.16	0.116	0.07	0.24	0.25
2004	0.18	0.36	0.30	0.22	0.15	0.10	0.10	0.26	0.23

Table 4.5h: **Number of Scores Sent to a selective school**

year	Q not ans	top decile	2nd decile	2nd quintile	3rd quint	4th quint	5th quint	SDQ not ans	Overall
1996	0.49	0.99	0.83	0.71	0.49	0.33	0.27	0.41	0.70
1997	0.47	0.96	0.81	0.69	0.48	0.34	0.26	0.38	0.67
1998	0.45	0.95	0.80	0.65	0.46	0.32	0.26	0.40	0.65
1999	0.48	0.94	0.78	0.64	0.46	0.32	0.28	0.34	0.64
2000	0.55	0.98	0.82	0.69	0.51	0.34	0.31	0.35	0.68
2001	0.53	0.99	0.82	0.66	0.47	0.35	0.33	0.32	0.66
2002	0.56	0.99	0.81	0.63	0.43	0.32	0.28	0.32	0.65
2003	0.53	0.99	0.79	0.61	0.43	0.30	0.25	0.60	0.64
2004	0.47	0.96	0.76	0.57	0.41	0.28	0.29	0.66	0.61

Table 4.5i: **Proportion who send to at least one selective school**

year	Q not ans	top decile	2nd decile	2nd quintile	3rd quint	4th quint	5th quint	SDQ not ans	Overall
1996	0.38	0.71	0.60	0.52	0.38	0.27	0.23	0.31	0.51
1997	0.36	0.69	0.59	0.51	0.37	0.28	0.21	0.29	0.50
1998	0.35	0.68	0.58	0.48	0.35	0.26	0.22	0.30	0.48
1999	0.37	0.68	0.57	0.47	0.36	0.26	0.22	0.27	0.48
2000	0.41	0.70	0.59	0.50	0.39	0.27	0.26	0.27	0.50
2001	0.39	0.71	0.58	0.48	0.36	0.27	0.27	0.26	0.48
2002	0.40	0.70	0.57	0.46	0.34	0.26	0.23	0.24	0.47
2003	0.39	0.70	0.56	0.45	0.33	0.25	0.20	0.43	0.47
2004	0.35	0.68	0.54	0.42	0.32	0.23	0.23	0.47	0.44

Table 4.6a: **Number of Takers by LOS x CS schools**

	LOS=0 CS=0	LOS=1 CS=0	LOS=0 CS=1	LOS=1 CS=1
1996	73574	5369	1802	3024
1997	77113	5704	1845	3088
1998	82870	6221	1829	3216
1999	86909	6612	1774	3435
2000	91058	6981	1796	3532
2001	92824	6898	1774	3519
2002	97596	7242	1663	3596
2003	102307	7425	1880	3648
2004	104945	7746	1909	3624

Table 4.6b: **Proportion of 12th Graders taking an SAT by LOS x CS schools**

	LOS=0 CS=0	LOS=1 CS=0	LOS=0 CS=1	LOS=1 CS=1
1996	0.47	0.45	0.47	0.47
1997	0.47	0.44	0.47	0.45
1998	0.48	0.46	0.45	0.46
1999	0.48	0.48	0.45	0.50
2000	0.49	0.51	0.45	0.51
2001	0.49	0.53	0.44	0.52
2002	0.50	0.54	0.42	0.55
2003	0.49	0.53	0.43	0.54
2004	0.49	0.56	0.42	0.54

Table 4.6c: **Proportion of SAT Takers sending to UT Austin by LOS x CS Schools**

year	LOS=0 CS=0	LOS=1 CS=0	LOS=0 CS=1	LOS=1 CS=1
1996	0.36	0.23	0.31	0.20
1997	0.35	0.21	0.30	0.18
1998	0.34	0.22	0.27	0.17
1999	0.34	0.21	0.29	0.18
2000	0.36	0.24	0.28	0.20
2001	0.34	0.26	0.30	0.20
2002	0.33	0.26	0.29	0.21
2003	0.32	0.26	0.27	0.22
2004	0.31	0.27	0.27	0.23

Table 4.7: **Proportion Taking SAT and Sending to UT Austin**

	Prop. SAT		Prop. Austin—SAT	
Year	LOS=0	LOS=1	LOS=0	LOS=1
1996	0.39	0.44	0.24	0.23
1997	0.42	0.43	0.22	0.21
1998	0.37	0.47	0.20	0.20
1999	0.33	0.49	0.22	0.20
2000	0.45	0.52	0.23	0.24
2001	0.45	0.53	0.18	0.26
2002	0.42	0.55	0.19	0.25
2003	0.40	0.53	0.24	0.25
2004	0.41	0.57	0.24	0.26

Table 4.8: **Individual level regressions for score-report sending w/o fixed effects**

	Dependent Variables						
	> 4 scores	# scores	non-select	selective	Austin	TTPR	In-state
Trend	0.001	-0.01	0.009	-0.004	-0.003	0.001	0
	[0.001]	[0.004]*	[0.001]**	[0.001]**	[0.001]**	[0.000]	[0.000]
Pre-period	0	0.007	0.006	0.011	0.008	0.004	0
	[0.003]	[0.015]	[0.003]*	[0.003]**	[0.003]**	[0.002]*	[0.001]
NoRank	-0.089	-0.499	-0.071	-0.009	-0.007	-0.031	-0.008
	[0.004]**	[0.024]**	[0.006]**	[0.005]	[0.004]	[0.005]**	[0.002]**
1st decile	0.186	0.984	-0.109	0.201	0.157	-0.006	0.003
	[0.005]**	[0.027]**	[0.006]**	[0.006]**	[0.005]**	[0.004]	[0.002]
2nd decile	0.084	0.442	-0.043	0.152	0.112	0.012	0.004
	[0.004]**	[0.019]**	[0.005]**	[0.004]**	[0.004]**	[0.003]**	[0.002]**
2nd quintile	0.052	0.241	-0.008	0.097	0.073	0.016	0.004
	[0.003]**	[0.018]**	[0.004]	[0.004]**	[0.004]**	[0.003]**	[0.001]**
4th quintile	-0.035	-0.208	-0.024	-0.071	-0.045	-0.042	-0.008
	[0.005]**	[0.032]**	[0.007]**	[0.007]**	[0.006]**	[0.007]**	[0.003]*
5th quintile	-0.049	-0.343	-0.08	-0.095	-0.076	-0.074	-0.017
	[0.011]**	[0.063]**	[0.015]**	[0.014]**	[0.010]**	[0.014]**	[0.007]*
Post	0.001	-0.105	-0.022	0.002	-0.006	-0.001	-0.001
	[0.004]	[0.022]**	[0.005]**	[0.004]	[0.004]	[0.003]	[0.002]
Post×NoRank	0.036	0.014	-0.004	0.007	0.017	0.006	0.002
	[0.004]**	[0.026]	[0.006]	[0.005]	[0.004]**	[0.005]	[0.002]
Post×1st decile	-0.056	-0.248	-0.02	0.019	0.038	0.005	0
	[0.004]**	[0.026]**	[0.005]**	[0.005]**	[0.004]**	[0.004]	[0.002]
Post×2nd decile	-0.019	-0.093	-0.003	-0.001	0.012	0.002	0
	[0.004]**	[0.021]**	[0.005]	[0.004]	[0.004]**	[0.003]	[0.002]
Post×2nd quintile	-0.016	-0.038	-0.002	-0.016	-0.01	-0.005	-0.002
	[0.004]**	[0.020]	[0.004]	[0.004]**	[0.004]*	[0.003]	[0.002]
Post×4th quintile	0.005	0.054	0.015	0.013	0.007	0.016	0.004
	[0.006]	[0.035]	[0.007]*	[0.007]	[0.006]	[0.007]*	[0.004]
Post×5th quintile	0.008	0.123	0.031	0.044	0.046	0.022	0.015
	[0.012]	[0.069]	[0.016]*	[0.014]**	[0.011]**	[0.015]	[0.008]
Constant	0.314	3.849	0.792	0.353	0.431	0.596	0.97
	[0.072]**	[0.442]**	[0.121]**	[0.057]**	[0.066]**	[0.053]**	[0.019]**
Observations	865490	865490	865490	865490	865490	751339	751339
R^2	0.1	0.1	0.08	0.13	0.11	0.01	0

Notes:
1. Robust standard errors in brackets
2. * significant at 5%; ** significant at 1%
3. Omitted coefficients include SAT math and verbal scores, racial composition of the school, % poor, numbers in 12th grade, student-to-teacher ratio, teacher experience, and whether the school subsequently gets LOS or CS.
4. All dependent variables are 0 - 1 indicators, except for Col 2 which is a count variable

Table 4.9: **Individual level regressions for score-report sending with fixed effects**

	Dependent Variables						
	> 4 scores	# scores	non-select	selective	Austin	TTPR	In-state
Trend	0.003	0.007	0.012	-0.004	-0.002	0	0
	[0.001]**	[0.005]	[0.001]**	[0.001]**	[0.001]**	[0.000]	[0.000]
Pre-period	0	0.014	0.008	0.01	0.008	0.003	0
	[0.003]	[0.015]	[0.003]**	[0.003]**	[0.003]**	[0.002]	[0.001]
NoRank	-0.09	-0.505	-0.07	-0.004	-0.003	-0.026	-0.008
	[0.004]**	[0.022]**	[0.005]**	[0.004]	[0.004]	[0.004]**	[0.002]**
1st decile	0.213	1.104	-0.113	0.214	0.176	-0.007	0
	[0.004]**	[0.025]**	[0.006]**	[0.005]**	[0.004]**	[0.003]*	[0.002]
2nd decile	0.1	0.515	-0.044	0.158	0.122	0.012	0.002
	[0.004]**	[0.019]**	[0.005]**	[0.004]**	[0.004]**	[0.003]**	[0.002]
2nd quintile	0.06	0.279	-0.008	0.099	0.077	0.016	0.002
	[0.003]**	[0.018]**	[0.004]*	[0.004]**	[0.004]**	[0.003]**	[0.001]
4th quintile	-0.046	-0.268	-0.021	-0.072	-0.048	-0.04	-0.006
	[0.005]**	[0.032]**	[0.007]**	[0.007]**	[0.006]**	[0.006]**	[0.003]
5th quintile	-0.067	-0.439	-0.074	-0.092	-0.073	-0.069	-0.013
	[0.011]**	[0.062]**	[0.014]**	[0.014]**	[0.011]**	[0.013]**	[0.007]
Post	0.002	-0.107	-0.026	0.001	-0.006	-0.002	-0.002
	[0.004]	[0.022]**	[0.004]**	[0.004]	[0.003]	[0.003]	[0.002]
Post×NoRank	0.036	0.016	0.001	0.001	0.011	0.004	0.002
	[0.004]**	[0.024]	[0.006]	[0.005]	[0.004]*	[0.004]	[0.002]
Post×1st decile	-0.059	-0.261	-0.018	0.018	0.036	0.007	0
	[0.004]**	[0.026]**	[0.005]**	[0.004]**	[0.004]**	[0.003]	[0.002]
Post×2nd decile	-0.021	-0.1	-0.002	-0.002	0.01	0.003	0
	[0.004]**	[0.022]**	[0.004]	[0.004]	[0.004]*	[0.003]	[0.002]
Post×2nd quintile	-0.017	-0.044	-0.003	-0.016	-0.01	-0.004	-0.002
	[0.004]**	[0.021]*	[0.004]	[0.004]**	[0.004]**	[0.003]	[0.002]
Post×4th quintile	0.007	0.061	0.012	0.011	0.005	0.015	0.004
	[0.005]	[0.034]	[0.007]	[0.007]	[0.006]	[0.007]*	[0.004]
Post×5th quintile	0.013	0.144	0.023	0.038	0.039	0.016	0.014
	[0.012]	[0.069]*	[0.015]	[0.014]**	[0.011]**	[0.015]	[0.008]
Constant	0.028	1.854	0.447	-0.041	-0.094	0.775	0.933
	[0.067]	[0.570]**	[0.069]**	[0.103]	[0.105]	[0.074]**	[0.020]**
Observations	865490	865490	865490	865490	865490	751339	751339
R^2	0.12	0.12	0.15	0.15	0.14	0.04	0.02

Notes:
1. Robust standard errors in brackets
2. * significant at 5%; ** significant at 1%
3. Omitted coefficients include SAT math and verbal scores, racial composition of the school, % poor, numbers in 12th grade, teacher student ratio, teacher experience and whether the school subsequently gets LOS or CS.
4. All dependent variables are 0 - 1 indicators, except for Col 2 which is a count variable
5. Estimates of high school level fixed effects have been suppressed.

Table 4.10: **Difference-in-Differences estimates for LOS program w/o fixed effects**

	Dependent Variables						
	Austin	# scores	non-select	selective	TTPR	In-State	> 4 scores
Trend	0.006	-0.014	0.011	0.004	0.006	0.001	-0.004
	[0.002]**	[0.014]	[0.004]**	[0.002]*	[0.002]**	[0.001]	[0.002]
No Rank	0.006	-0.526	-0.062	-0.013	-0.028	0.007	-0.109
	[0.015]	[0.058]**	[0.013]**	[0.013]	[0.015]	[0.005]	[0.010]**
1st decile	0.184	1.079	-0.1	0.2	0.02	0.013	0.237
	[0.012]**	[0.070]**	[0.016]**	[0.019]**	[0.012]	[0.010]	[0.015]**
2nd decile	0.101	0.514	-0.009	0.127	0.036	0.014	0.112
	[0.020]**	[0.060]**	[0.012]	[0.023]**	[0.011]**	[0.007]*	[0.012]**
2nd quintile	0.036	0.325	0.007	0.039	0.019	0.014	0.084
	[0.010]**	[0.080]**	[0.012]	[0.013]**	[0.009]*	[0.009]	[0.018]**
4th quintile	-0.02	-0.195	-0.026	-0.049	0.005	0.012	-0.035
	[0.020]	[0.057]**	[0.027]	[0.016]**	[0.028]	[0.015]	[0.009]**
5th quintile	-0.034	-0.367	-0.14	-0.095	-0.08	0.003	-0.055
	[0.035]	[0.087]**	[0.026]**	[0.026]**	[0.037]*	[0.022]	[0.017]**
LOS×No Rank	-0.009	-0.033	-0.015	-0.002	-0.013	0.003	-0.001
	[0.014]	[0.094]	[0.014]	[0.013]	[0.015]	[0.004]	[0.013]
LOS×1st decile	0.097	0.12	0.049	0.082	0.054	0.012	0.035
	[0.013]**	[0.085]	[0.016]**	[0.013]**	[0.013]**	[0.005]**	[0.015]*
LOS×2nd decile	0.034	0.115	0.019	0.021	0.022	0.008	0.023
	[0.011]**	[0.076]	[0.013]	[0.013]	[0.009]*	[0.004]	[0.013]
LOS×2nd quintile	0.022	0.109	0.018	0.034	0.002	0.012	0.02
	[0.009]**	[0.097]	[0.019]	[0.013]**	[0.012]	[0.006]	[0.019]
LOS×3rd quintile	0.004	0.149	0.008	-0.004	0.001	-0.006	0.019
	[0.012]	[0.098]	[0.016]	[0.015]	[0.013]	[0.007]	[0.017]
LOS×4th quintile	-0.002	0.059	0.025	0.011	0.01	0.007	-0.006
	[0.014]	[0.119]	[0.025]	[0.015]	[0.016]	[0.007]	[0.017]
LOS×5th quintile	-0.035	0.061	0.064	-0.03	-0.037	-0.011	-0.019
	[0.031]	[0.231]	[0.036]	[0.028]	[0.043]	[0.013]	[0.042]
LOS School	-0.034	-0.076	-0.003	-0.035	-0.015	-0.008	-0.004
	[0.012]**	[0.097]	[0.027]	[0.015]*	[0.014]	[0.004]	[0.018]
Constant	-11.918	27.432	-22.176	-7.302	-12.066	-0.696	7.636
	[3.925]**	[27.447]	[8.036]**	[3.817]	[3.146]**	[1.244]	[4.187]
Observations	189354	189354	189354	189354	167646	167646	189354
R^2	0.09	0.1	0.01	0.1	0.02	0.01	0.11

Notes:
1. Robust standard errors in brackets
2. * significant at 5%; ** significant at 1%
3. Omitted coefficients include SAT math and verbal scores, racial composition of the school, % poor, numbers in 12th grade, student-to-teacher ratio, teacher experience, ,whether the school subsequently gets CS, the 'pre' dummy, the post dummy, and interactions between the post dummy and rank categories.
4. All dependent variables are 0 - 1 indicators, except for Col 2 which is a count variable.
5. Estimates of high school level fixed effects have been suppressed.
6. The common support condition has been imposed.

Table 4.11: **Difference-in-Differences estimates for LOS program with fixed effects**

	Dependent Variables						
	Austin	# scores	non-select	selective	TTPR	In-State	> 4 scores
Trend	0.008	0.011	0.014	0.005	0.005	0.001	0.001
	[0.002]**	[0.016]	[0.003]**	[0.002]*	[0.002]**	[0.001]*	[0.002]
No Rank	0.01	-0.537	-0.054	-0.012	-0.018	0.006	-0.112
	[0.014]	[0.052]**	[0.012]**	[0.012]	[0.013]	[0.006]	[0.010]**
1st decile	0.184	1.101	-0.075	0.191	0.022	0.012	0.246
	[0.011]**	[0.066]**	[0.016]**	[0.020]**	[0.011]*	[0.010]	[0.014]**
2nd decile	0.101	0.518	0.007	0.12	0.036	0.013	0.116
	[0.019]**	[0.062]**	[0.012]	[0.022]**	[0.010]**	[0.007]*	[0.011]**
2nd quintile	0.037	0.33	0.011	0.038	0.02	0.014	0.087
	[0.010]**	[0.082]**	[0.011]	[0.014]**	[0.009]*	[0.009]	[0.018]**
4th quintile	-0.021	-0.201	-0.025	-0.049	0.002	0.011	-0.036
	[0.020]	[0.053]**	[0.027]	[0.016]**	[0.028]	[0.015]	[0.009]**
5th quintile	-0.034	-0.376	-0.128	-0.098	-0.081	0.002	-0.06
	[0.035]	[0.089]**	[0.027]**	[0.026]**	[0.038]*	[0.022]	[0.018]**
LOS×No Rank	-0.005	-0.037	-0.021	0.004	-0.009	0.002	-0.002
	[0.013]	[0.083]	[0.013]	[0.012]	[0.011]	[0.004]	[0.011]
LOS×1st decile	0.099	0.06	0.039	0.087	0.055	0.011	0.024
	[0.013]**	[0.074]	[0.015]*	[0.013]**	[0.014]**	[0.005]*	[0.015]
LOS×2nd decile	0.033	0.075	0.009	0.025	0.018	0.007	0.015
	[0.011]**	[0.064]	[0.011]	[0.013]	[0.009]*	[0.004]	[0.012]
LOS×2nd quintile	0.024	0.089	0.009	0.037	-0.001	0.011	0.019
	[0.008]**	[0.082]	[0.016]	[0.013]**	[0.008]	[0.006]	[0.016]
LOS×3rd quintile	0.007	0.181	-0.002	0.004	-0.004	-0.006	0.026
	[0.011]	[0.082]*	[0.014]	[0.014]	[0.010]	[0.006]	[0.014]
LOS×4th quintile	0.002	0.123	0.021	0.023	0.008	0.007	0.007
	[0.013]	[0.106]	[0.018]	[0.015]	[0.013]	[0.006]	[0.014]
LOS×5th quintile	-0.028	0.196	0.042	-0.01	-0.043	-0.01	0.002
	[0.030]	[0.224]	[0.030]	[0.029]	[0.042]	[0.013]	[0.042]
Constant	-16.174	-18.633	-27.595	-9.731	-9.352	-1.765	-2.206
	[4.340]**	[32.106]	[5.988]**	[4.234]*	[3.423]**	[1.360]	[4.581]
Observations	189354	189354	189354	189354	167646	167646	189354
R^2	0.1	0.13	0.06	0.11	0.04	0.02	0.13

Notes:
1. Robust standard errors in brackets
2. * significant at 5%; ** significant at 1%
3. Omitted coefficients include SAT math and verbal scores, racial composition of the school, % poor, numbers in 12th grade, student-to-teacher ratio, teacher experience, ,whether the school subsequently gets CS, the ‘pre’ dummy, the post dummy, and interactions between the post dummy and rank categories.
4. All dependent variables are 0 - 1 indicators, except for Col 2 which is a count variable
5. Estimates of high school level fixed effects have been suppressed.
6. The common support condition has been imposed.

Table 4.12: **Difference-in-Differences estimates for CS program w/o fixed effects**

	Dependent Variables						
	A&M	# scores	non-select	selective	TTPR	In-State	> 4 scores
Trend	-0.005	0.031	0.006	0.002	0.001	0	0.005
	[0.002]**	[0.016]	[0.004]	[0.002]	[0.002]	[0.001]	[0.002]**
No Rank	-0.015	-0.61	-0.091	-0.03	-0.046	-0.004	-0.105
	[0.007]*	[0.059]**	[0.016]**	[0.011]**	[0.012]**	[0.007]	[0.012]**
1st decile	0.144	1.048	-0.091	0.198	0.022	0.009	0.211
	[0.016]**	[0.095]**	[0.025]**	[0.018]**	[0.016]	[0.007]	[0.014]**
2nd decile	0.081	0.459	-0.029	0.108	0.018	0.008	0.103
	[0.011]**	[0.060]**	[0.016]	[0.014]**	[0.012]	[0.006]	[0.012]**
2nd quintile	0.042	0.231	-0.002	0.054	0.013	0.005	0.063
	[0.008]**	[0.050]**	[0.013]	[0.013]**	[0.010]	[0.006]	[0.008]**
4th quintile	-0.037	-0.164	-0.052	-0.08	-0.041	-0.023	-0.024
	[0.017]*	[0.090]	[0.015]**	[0.020]**	[0.021]*	[0.015]	[0.015]
5th quintile	-0.065	-0.235	-0.121	-0.13	-0.124	-0.023	-0.031
	[0.023]**	[0.146]	[0.039]**	[0.030]**	[0.047]**	[0.036]	[0.028]
CS×No Rank	0.023	-0.241	-0.046	0.033	0.001	0.003	-0.024
	[0.014]	[0.112]*	[0.026]	[0.014]*	[0.012]	[0.005]	[0.017]
CS×1st decile	0.039	-0.287	0.016	0.04	0.036	0.004	-0.042
	[0.019]*	[0.085]**	[0.022]	[0.015]**	[0.016]*	[0.006]	[0.014]**
CS×2nd decile	0.014	-0.251	-0.015	-0.001	-0.005	0.006	-0.05
	[0.014]	[0.090]**	[0.016]	[0.015]	[0.011]	[0.005]	[0.017]**
CS×2nd quintile	0.012	-0.147	-0.023	0.025	-0.009	0.008	-0.028
	[0.013]	[0.095]	[0.020]	[0.015]	[0.013]	[0.006]	[0.015]
CS×3rd quintile	0.02	-0.127	-0.057	0.018	-0.009	-0.003	-0.02
	[0.011]	[0.096]	[0.021]**	[0.015]	[0.014]	[0.007]	[0.016]
CS×4th quintile	0.015	-0.004	-0.082	0.021	-0.057	0.005	-0.019
	[0.017]	[0.157]	[0.029]**	[0.026]	[0.032]	[0.011]	[0.017]
CS×5th quintile	0.019	-0.31	-0.097	0.062	-0.018	0.016	-0.059
	[0.033]	[0.256]	[0.047]*	[0.041]	[0.048]	[0.011]	[0.054]
CS School	0.001	-0.148	-0.01	-0.035	-0.054	-0.004	-0.02
	[0.012]	[0.089]	[0.034]	[0.014]*	[0.014]**	[0.005]	[0.015]
Constant	11.311	-56.196	-11.322	-4.115	-1.365	1.348	-10.304
	[3.363]**	[31.809]	[7.282]	[4.049]	[3.617]	[1.102]	[4.192]*
Observations	265849	265849	265849	265849	233562	233562	265849
R^2	0.07	0.1	0.04	0.13	0.03	0.01	0.11

Notes:
1. Robust standard errors in brackets
2. * significant at 5%; ** significant at 1%
3. Omitted coefficients include SAT math and verbal scores, racial composition of the school, % poor, numbers in 12th grade, student-to-teacher ratio, teacher experience, ,whether the school subsequently gets LOS, the 'pre' dummy, the post dummy, and interactions between the post dummy and rank categories.
4. All dependent variables are 0 - 1 indicators, except for Col 2 which is a count variable.
5. Estimates of high school level fixed effects have been suppressed.
6. The common support condition has been imposed.

Table 4.13: **Difference-in-Differences estimates for CS program with fixed effects**

	Dependent Variables						
	A&M	# scores	non-select	selective	TTPR	In-State	> 4 scores
Trend	-0.006	0.02	0.014	-0.003	0.002	0.001	0.003
	[0.002]**	[0.017]	[0.003]**	[0.002]	[0.002]	[0.001]	[0.002]
No Rank	-0.015	-0.602	-0.08	-0.031	-0.042	-0.004	-0.105
	[0.007]*	[0.061]**	[0.014]**	[0.011]**	[0.012]**	[0.007]	[0.012]**
1st decile	0.146	1.086	-0.08	0.2	0.022	0.007	0.223
	[0.016]**	[0.093]**	[0.024]**	[0.018]**	[0.015]	[0.007]	[0.014]**
2nd decile	0.082	0.478	-0.022	0.108	0.02	0.007	0.11
	[0.011]**	[0.062]**	[0.016]	[0.014]**	[0.011]	[0.006]	[0.013]**
2nd quintile	0.042	0.245	0.002	0.053	0.015	0.005	0.067
	[0.008]**	[0.048]**	[0.013]	[0.012]**	[0.009]	[0.006]	[0.008]**
4th quintile	-0.037	-0.187	-0.064	-0.079	-0.042	-0.023	-0.028
	[0.017]*	[0.089]*	[0.014]**	[0.020]**	[0.021]*	[0.015]	[0.015]
5th quintile	-0.058	-0.314	-0.097	-0.14	-0.128	-0.021	-0.045
	[0.024]*	[0.139]*	[0.042]*	[0.030]**	[0.046]**	[0.036]	[0.029]
CS×No Rank	0.02	-0.153	-0.038	0.034	-0.001	0.004	-0.015
	[0.012]	[0.100]	[0.021]	[0.012]**	[0.011]	[0.004]	[0.014]
CS×1st decile	0.042	-0.27	0.004	0.045	0.037	0.006	-0.042
	[0.019]*	[0.084]**	[0.022]	[0.014]**	[0.014]**	[0.005]	[0.012]**
CS×2nd decile	0.016	-0.229	-0.018	-0.002	-0.006	0.008	-0.048
	[0.015]	[0.080]**	[0.014]	[0.014]	[0.011]	[0.005]	[0.014]**
CS×2nd quintile	0.011	-0.111	-0.022	0.021	-0.008	0.009	-0.024
	[0.013]	[0.088]	[0.016]	[0.015]	[0.011]	[0.005]	[0.012]
CS×3rd quintile	0.017	-0.096	-0.042	0.009	-0.009	-0.001	-0.016
	[0.011]	[0.087]	[0.018]*	[0.014]	[0.013]	[0.007]	[0.014]
CS×4th quintile	0.005	0.014	-0.062	0.011	-0.053	0.007	-0.018
	[0.016]	[0.144]	[0.025]*	[0.025]	[0.032]	[0.011]	[0.015]
CS×5th quintile	0.002	-0.31	-0.099	0.06	-0.002	0.013	-0.058
	[0.034]	[0.248]	[0.046]*	[0.040]	[0.046]	[0.011]	[0.054]
Constant	12.598	-37.769	-28.606	6.149	-3.742	-0.063	-6.191
	[3.296]**	[34.452]	[5.481]**	[4.379]	[3.571]	[1.430]	[4.563]
Observations	265849	265849	265849	265849	233562	233562	265849
R^2	0.08	0.13	0.1	0.14	0.05	0.01	0.13

Notes:
1. Robust standard errors in brackets
2. * significant at 5%; ** significant at 1%
3. Omitted coefficients include SAT math and verbal scores, racial composition of the school, % poor, numbers in 12th grade, student-to-teacher ratio, teacher experience, ,whether the school subsequently gets LOS, the 'pre' dummy, the post dummy, and interactions between the post dummy and rank categories.
4. All dependent variables are 0 - 1 indicators, except for Col 2 which is a count variable
5. Estimates of high school level fixed effects have been suppressed.
6. The common support condition has been imposed.

4.10 Figures

Figure 4.1: **Propensity Score Histograms for LOS Schools**

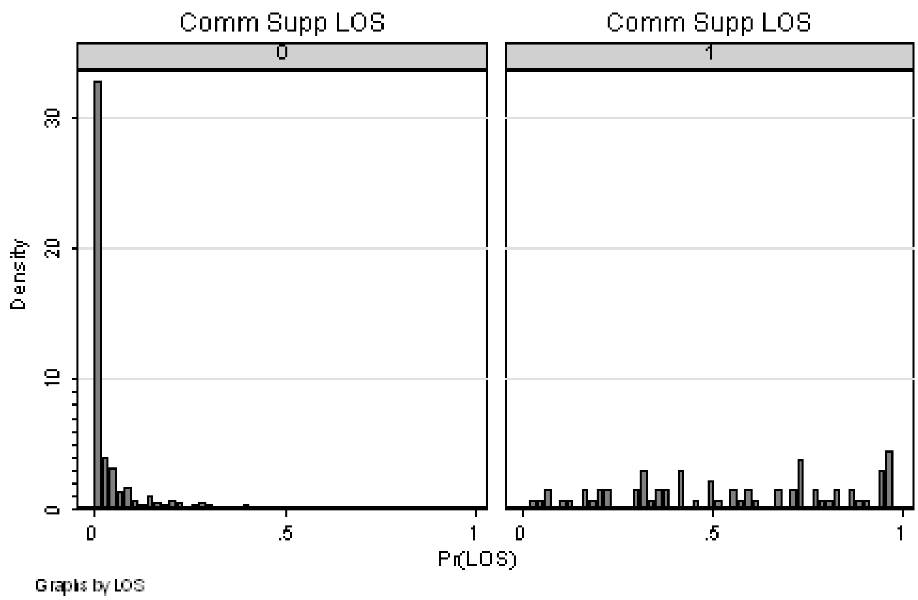


Figure 4.2: **Adjusted Propensity Score Histograms for LOS Schools**

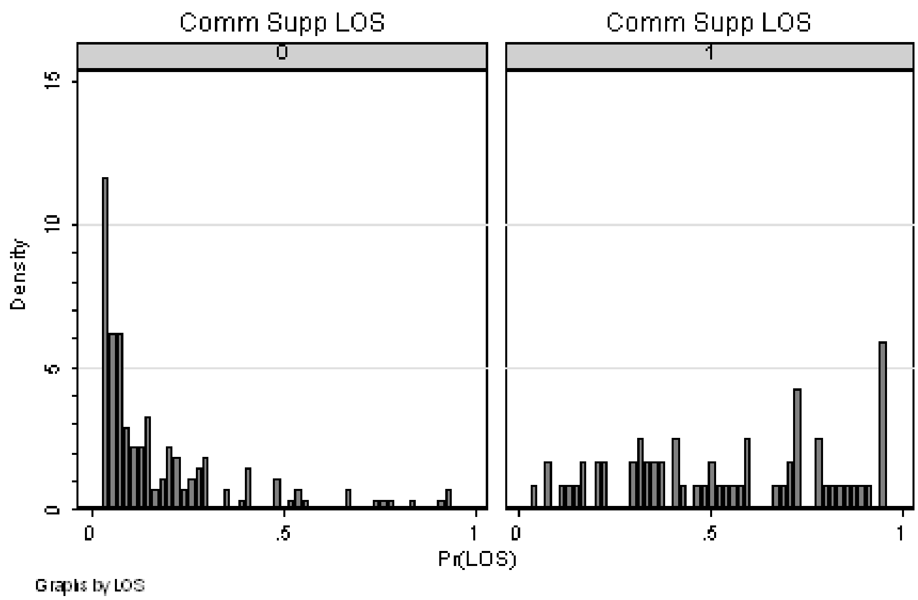


Figure 4.3: **Propensity Score Histograms for CS Schools**

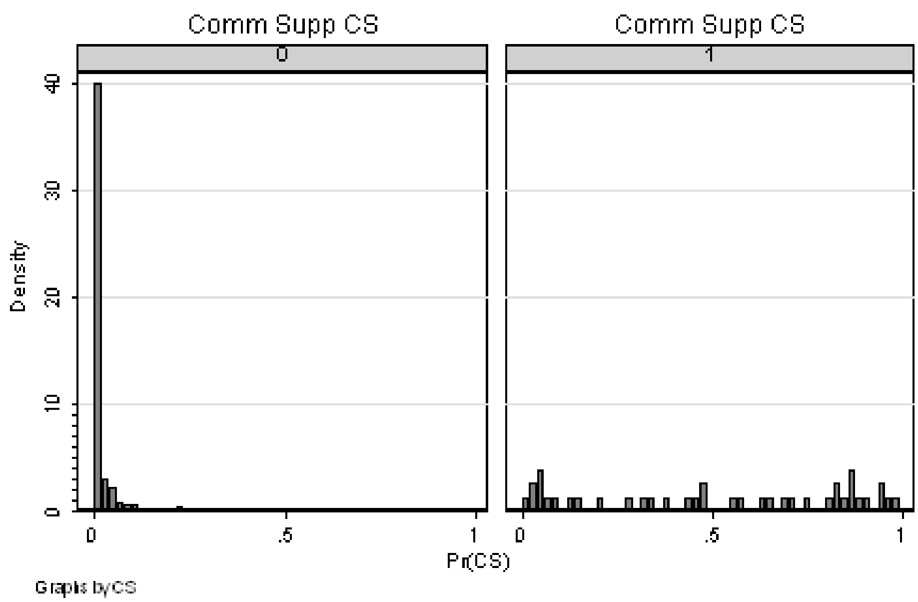
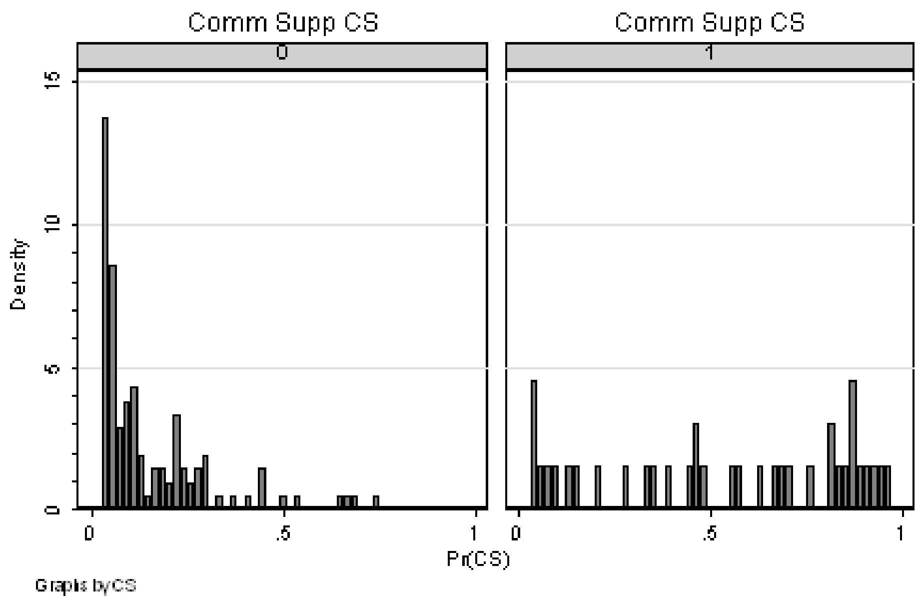


Figure 4.4: **Adjusted Propensity Score Histograms for CS Schools**



Bibliography

Bertrand, M., E. Duflo, and S. Mullainathan (2004). "How much should we trust difference in differences estimates?", Quarterly Journal of Economics 119 (1), 249 - 275

Bucks, B. (2004). "Affirmative access versus affirmative action: How have texas's race-blind admissions policies affected college outcomes?", Texas Schools Project Working Paper, Dallas, Texas

Card, D. and A. Krueger (2005). "Would the elimination of affirmative action affect highly qualified minority applicants? evidence from California and Texas", Industrial and Labor Relations Review 56, 416 - 434

Card, D. and A. B. Krueger (2004, March). "Would the elimination of affirmative action affect highly qualified minority applicants? evidence from California and Texas", Working Paper 10366, National Bureau of Economic Research

Dickson, L. M. (2006, February). "Does ending affirmative action in college admissions lower the percent of minority students applying to college?", Economics of Education Review 25, 109 - 119

DiNardo, J., N. M. Fortin, and T. Lemieux (1996). "Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach", Econometrica 64 (5), 1001 - 1044

Horvitz, D. and D. Thompson (1952). "A generalization of sampling without replacement from a finite universe" Journal of the American Statistical Association 47 (260), 663 - 685

Long, M. C. (2004). "College applications and the effect of affirmative action", Journal of Econometrics 121 (1-2), 319 - 342

Moulton, B. R. (1990). "An illustration of a pitfall in estimating the effects of aggregate variables on micro units", The Review of Economics and Statistics 72 (2), 334 - 338

Niu, S., M. Tienda, and K. Cortes (2006). "College selectivity and the texas top 10%

law", Economics of Education Review 25 (3), 259 -272

Rosenbaum, P. R. and D. B. Rubin (1983). "The central role of the propensity score in observational studies for causal effects", Biometrika 70 (1), 41 - 55

Tienda, M. and S. Niu (2006). "Flagships, feeders, and the texas top 10% law: A test of the brain drain hypothesis" Journal of Higher Education 76 (4), 712 - 739

CHAPTER V

Conclusion

The first chapter of the dissertation investigates how poor households in developing countries respond to adverse income developments. I use nationally representative longitudinal data to investigate behavioral responses to the loss of an Old Age Pensioner in South Africa. I find that household composition adjusts, with an outflow of school aged children and an inflow of middle-aged females and older adults. The household, on aggregate, also has more people employed. Conditional on compositional stability within demographic groups, I find large and significant increases in both labor supply and employment. Policy makers might be concerned with the impact of cash transfers on the labor supply of non-recipients.

The second chapter examines the effect of the Old Age Pension on retirement behavior of elderly South Africans. I make use of the rules on age eligibility to measure changes in various dimensions of labor supply that occur when people reach the pensionable age. I find significant decreases in employment rates and labor supply. Those who remain employed beyond the pensionable age are more likely to work in jobs with flexible hours of work, and work even fewer hours than people in similar jobs who are not pension age-eligible.

The final chapter investigates the impact of changes in the probability of being admitted into a selective college on students' SAT score sending behavior. We capture this using a student's class rank combined with the Texas Top Ten Percent Rule. In response to dwindling minority enrollment rates, the University of Texas at Austin and Texas A&M College

Station embarked on targeted recruitment programs at previously under-represented high schools. We evaluate the effectiveness of these programs using individual level SAT data. We find that score sending is affected by the legal change, and that both targeted recruitment programs were successful in attracting scores. In each case, the effects were manifest most strongly amongst the students in the top decile of the class. This suggests that students from poor schools face multiple barriers to obtaining postsecondary schooling at selective colleges.