NORTHWESTERN UNIVERSITY

Ethnicity, Race and Gender in the Labor Market

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Economics

By

Zahra Siddique

EVANSTON, ILLINOIS

June 2008

ABSTRACT

Ethnicity, Race and Gender in the Labor Market

Zahra Siddique

Ethicity, race and gender play an important role in labor markets; labor market outcomes such as hiring and compensation are very different across different social groups. These differentials are partly the result of differences in productivity and preferences and partly the result of discrimination.

Chapter two uses an audit study to determine the existence and extent of caste-based discrimination in the Indian private sector. The study also has policy implications for recent debates regarding introduction of caste-based quotas in Indian private sector jobs. Resumes with caste-specific names are sent to employers for entry-level jobs and callback rates measured. On average, low-caste applicants need to send 20% more resumes than high-caste applicants to get one callback. There is also heterogeneity in callback gaps by recruiter characteristics and firm size which indicates the presence of prejudice against low-caste workers and is consistent with commitments made by large firms to hire actively from among low-caste groups.

In chapter three I find partially identified treatment effect for arrest and other treatments by looking at recidivism for a sample of domestic assault offenders. The treatment effects are not fully identified due to non-compliance with assigned treatment and the possibility of a non-random treatment assignment. Partially identified treatment effects are estimated by making minimal assumptions on the counterfactual probabilities.

Chapter four (based on joint work with Wallace Mok) examines the difference in non-wage compensation between African Americans and whites in the US. Using data from the Current Population Survey (CPS) and National Longitudinal Survey of Youth (NLSY), we find that without controlling for the Armed Forces Qualification Test (AFQT) scores, white men are more likely to receive non-wage compensation and white women are not more likely to get non-wage compensation. With controls for AFQT scores we find that white men are not more likely to receive non-wage compensation but black women are more likely to get non-wage compensation. We also find that the percentage differences in total compensation and the percentage differences in wages across racial groups are essentially the same.

Acknowledgements

I am extremely grateful to Christopher Taber for his patience and invaluable support during my time as a graduate student. Thanks also to Chuck Manski and Dale Mortensen for their constant support and encouragement over the years, to Martin Zelder for his detailed comments on large sections of the thesis, to Wallace Mok for the help and support needed to get chapter four ready in time and to the group of young labor economists in the labor lunches at Northwestern for discussion and comments on numerous ideas which ended up as part of the thesis. I am also grateful Ani Gupta, Siddarth Madhav and Vishy Pingali for their help in the fieldwork for chapter two, to the Centre for the Study of Indutrial Organization at Northwestern for providing access to the Prowess database and to the Centre for the Study of Developing Societies in Delhi for sharing the aggregate National Election Study data.

I am grateful to the friends at Northwestern and Evanston who have made my time here a very happy one; there are too many people to mention but particularly Warefta Hasan, Devi Thota, Samina Yamin and Basit Zafar. Finally I am grateful to my siblings Bilal and Naveera, Sohaib and Saad for their unconditional love and support, to my uncle Javaid Iqbal who has always been an important role model and to my parents, Tasvia and Muhammad Siddique, for their contant 'duas,' love, and encouragement.

To my parents

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

Ethicity, race and gender play an important role in labor markets; labor market outcomes such as hiring and compensation are often found to be very different across different ethnic, racial and gender groups. These differentials are partly the result of differences in productivity and preferences across social groups and partly the result of discrimination (Altonji and Blank (1999) provide detailed discussion for race and gender differentials in the United States). It is generally difficult to directly test for the presence of discrimination in labor markets since we cannot disentangle outcome differentials which arise due to differences from productivity and preferences from those that arise due to discrimination.

In chapter two, 'Caste Based Discrimination: Evidence and Policy,' an innovative dataset is collected and described which tests for the presence of ethnic, caste-based discrimination in the city of Chennai in southern India. Caste-based quotas in hiring have existed in the public sector in India for decades, and there has been recent debate about the introduction of such quotas in private sector jobs. Chapter two uses an audit study to determine the existence and extent of caste-based discrimination in the Indian private sector. Resumes with caste-specific names are sent to employers for entry-level jobs in the white-collar sector and the callback rates measured. Given the design of the study (described in chapter two), differences in productivity across high and low-caste workers which are observable to employers but not observable to the researcher are eliminated, enabling a more direct test of discrimination to be carried out. However differences in productivity across high and low-caste workers which are *unobservable* to both the employer and the researcher are not eliminated. Therefore the disparities in outcomes observed between high and low-caste workers might either be the result of differences in productivity which are unobservable to both the employer and the researcher (statistical discrimination), or the result of employer prejudice. Although the study does not provide a direct test of whether discrimination arises as a result of statistical discrimination or employer prejudice, I argue that the results suggest that at least some of the discrimination observed is the result of prejudice against low-caste applicants. There are two main results that emerge from the audit study: firstly, on average, high-caste applicants need to send 6.2 resumes to get one callback while low-caste applicants need to send 7.4 resumes to get one callback, a difference of approximately 20%. Secondly, the callback gap between high and low-caste applicants is shown to vary across both recruiter and firm characteristics. The effect of low caste on callback is negative for male recruiters and for Hindu recruiters, but it is positive for female recruiters and for non-Hindu recruiters. This finding is interesting since it cannot be easily explained by statistical theories of discrimination, indicating that at least some prejudice might be present against low-caste applicants. The effect of low caste on callback is negative for firms with a larger scale of operations (with multiple domestic offices or with foreign offices) but positive for firms with a smaller scale of operations (without multiple domestic offices or without foreign offices). This observation is consistent with taste-based theories of discrimination which tell us that non-discriminating firms grow faster since they make higher profits than discriminating firms. It is also consistent with commitments made by large firms in the white collar sector to hire more workers from the disadvantaged caste groups.

Chapter four, 'Racial differences in Wages and Non-Wage Compensation' examines racial differences in the United States for a different labor market measure: that of non-wage compensation benefits such as employer provided health insurance and pension coverage. Racial differences in wages are fairly well documented; however, there is less work that examines racial differences in non-wage compensation such as employer-provided health insurance and pension coverage. Thus, chapter four asks several questions: What are the racial differences in health insurance and pension coverage for men and women? What component of the racial difference in non-wage compensation can be explained as the result of racial differences? What are the racial differences in total compensation, and how do these differ from racial differences in wages?

Using data from both the Current Population Survey (CPS) and the National Longitudinal Survey of Youth (NLSY), white men have significantly greater health insurance coverage from their employers and greater pension coverage than do black men. Differences in characteristics favor greater health insurance and pension coverage for black men in the CPS. Therefore, the unexplained racial differences in health insurance and pension coverage are even larger than the observed differences. However, once controls for racial differences in ability (using AFQT test scores) are added in the NLSY data, much of the unexplained racial differences for men disappear. Unexplained differences in non-wage compensation that continue to favor white men could be an indication of discrimination in provision of non-wage benefits to black men; however, these could also be the result of racial differences in preferences.

White women do not always have higher health insurance and pension coverage than black women. Black women have greater coverage of employer-provided health insurance than white women, using CPS data. Racial differences due to differences in characteristics always favor greater coverage for black women in the CPS data. However, once racial differences in AFQT scores are added as controls in the NLSY, racial differences in characteristics always favor greater coverage for white women. The unexplained differences in non-wage compensation favor black women; this is suggestive of reverse discrimination in favor of black women for provision of non-wage benefits (possibly due to affirmative action in jobs that are more likely to provide non-wage benefits).

Total compensation is estimated by including the value of wages, health insurance and pension coverage by use of imputations. The percentage differences in total compensation and the percentage differences in wages across racial groups are found to be essentially the same.

Chapter three examines the effectiveness of arrest as a treatment (in the sense of lowering recidivism rates) for a sample of domestic assault offenders who had been assigned arrest and other treatments such as advice and separation. Domestic assault is a gender specific crime, with serious and debilitating consequences for many low-income women. In the United States there were 5,341,410 victimizations in 2002 with 27.4% of the total being victimizations committed by an intimate (The National Criminal Victimization Survey, Family Violence 2002). In chapter three I find partially identified treatment effects by using data from a randomized experiment, the Minneapolis Domestic Assault experiment. The treatment effects fail to be fully identified due to non-compliance with assigned treatment and due to the possibility of a non-random treatment assignment.

An important contribution of this chapter is the application of the literature on partially identified treatment effects to a substantive problem of interest. It is shown how the estimation of partially identified treatment effects may be carried out very easily. The advantage of the approach is the greater credibility of weaker assumptions than are used in conventional analysis for the estimation of treatment effects.

Partially identified recidivism probabilities associated with the different treatments are estimated first without making any assumptions on the counterfactual probabilities, under the assumption that treatment assignment is not random. The recidivism probabilities associated with the different treatments are also estimated when making the assumption that assigned treatments are perfectly random but without making any assumptions on the counterfactual probabilities due to non-compliance with the assigned treatment. Finally, to improve the no assumptions bounds on the recidivism probabilities associated with different treatments, I use two different models of self selection in treatment assignment which are behaviorally motivated from Manski and Nagin (1998). The models of treatment assignment are the skimming model which assumes officers arrest all high risk offenders and the outcome optimization model which assumes that officers assign treatments to minimize recidivism. I find that arrest is associated with the lowest recidivism probability given that the assigned treatment is perfectly random and no assumption is made on the counterfactual probabilities due to non-compliance. In addition, arrest is also associated with the lowest recidivism in comparison to the other treatments if assigned treatments are not random, provided officers assign treatments by arresting all high risk offenders (skimming). Arrest is not unambigously associated with the lowest recidivism in any of the other cases. The recidivism probabilities are also used to find the average treatment effects from having a mandatory arrest policy for the entire population of domestic assault offenders whose offence gets reported to the police. In the paper, the average treatment effect is the difference in recidivism probability under mandatory arrest and mandatory non-arrest policies (either advice or separation).

CHAPTER 2

Caste Based Discrimination: Evidence and Policy

2.1. Introduction

The caste system in India has existed for thousands of years and operates by dividing society into hierarchical groups by birth, with the hierarchy being defined on a purity scale. The caste functions as a closed group whose members are restricted in their choice of occupation and degree of social interaction in a manner that is reminiscent of European Guilds in the Middle Ages.[1] These restrictions on occupation and social interaction have led to large socioeconomic differentials between different caste groups. Localized affirmative action policies to improve the welfare of low-caste individuals were introduced in the 1930s in individual states such as Tamil Nadu, but nationwide introduction of affirmative action did not occur until after Indian independence in the 1940s. The Indian government initiated national affirmative action policies to improve the status and living conditions of low-caste groups (Scheduled Castes and Scheduled Tribes) by introducing caste-based quotas in political representation, public sector jobs and education. The quotas were later extended to a larger number of disadvantaged caste groups (Other Backward Castes). Current debate centers around whether or not to introduce caste-based quotas in the private sector which would mandate hiring of low-caste employees.[2]

[1] Freitas (2006)

[2] See for instance 'With Reservations' in the Economist, 10/4/2007, 'We have a few Reservations' in the Economist, 5/27/2006, Vol. 379 Issue 8479, p38-38 as well as 'Caste and Cash' in the Economist, 4/29/2006, Vol. 379 Issue 8475, p46-46.

This paper utilizes an audit study to determine the existence and extent of caste-based discrimination in the hiring practices of businesses in the Indian private sector. This study follows the caste categorization conventions used by the Indian government for affirmative action and welfare programs. The categorization scheme (in ascending hierarchical order) is as follows: the untouchable castes are categorized as Scheduled Castes (SC), backward tribes outside the caste system as Scheduled Tribes (ST), disadvantaged castes which do not belong to the untouchable castes as Other Backward Castes (OBC) and the residual category consisting primarily of the high or forward castes as Other Castes. The SC, ST, and OBC consist of the historically disadvantaged groups while the Other Castes consist of groups which have historically been in a strong socioeconomic position.

Despite sixty years of affirmative action programs in India, the socioeconomic divide between high and low-caste groups persists. As shown in figure E.1 the per capita consumption distribution in 2004-05 for the other (high caste) category is positively skewed. In contrast the per capita consumption distributions for all other caste groups are fairly symmetrical.[3] There are large differences in the level of education attained by high versus low-caste groups. Two thirds of those who hold degrees in higher education are members of high-caste groups representing only one-third of the total population.[4]

The economic impact of caste has been studied extensively.[5] Some studies use micro level datasets to analyze caste-based discrimination in urban settings of India. A study

[3,4] Data is taken from the National Sample Survey (NSS), 61st Round, carried out in 2004-05.

[5]See for instance Akerlof (1976) for a theoretical model of caste-based discrimination, Munshi and Rozensweig for a study of caste-based networks and the role of these networks in the workplace as well as Banerjee and Somanathan (2006) and Pande (2003) for a study of the effects of caste-based quotas in political representation.

of factory workers in Poona (Lambert, 1963) finds evidence of substantial wage discrimination against workers belonging to backward caste groups. Other studies, primarily sociological, which find evidence of caste discrimination, use data from cotton mills in Bombay (Morris, 1965) and for shoemakers in Agra (Lynch, 1965). Banerjee and Knight (1985) use survey data to determine wage and occupation discrimination for migrant workers in Delhi by using decomposition techniques. They find wage discrimination to be higher than occupation discrimination and discrimination in formal sector jobs to be higher than discrimination in informal sector jobs.

All of these studies collect data in non-experimental settings. Hence the disparities they report in wages and occupation choice fail to control fully for differences in productivity and differences in preferences between high and low-caste workers. As a result, they do not provide a direct test of the hypothesis that discrimination is present. The resume-based audit study that I carry out uses an experimental design to document the presence and extent of caste-based discrimination in white collar, private sector jobs in the city of Chennai. Given the design of the study (described in section four), differences in productivity across high and low-caste workers which are observable to employers but not observable to the researcher are eliminated, enabling a more direct test of discrimination to be carried out. However differences in productivity across high and low-caste workers which are *unobservable* to both the employer and the researcher are not eliminated. Therefore the disparities in outcomes observed between high and low-caste workers might either be the result of differences in productivity which are unobservable to both the employer and the researcher (statistical discrimination), or the result of employer prejudice. Although the study does not provide a direct test of whether discrimination arises as a result of statistical discrimination or employer prejudice, I will argue that the results suggest that at least some of the discrimination observed is the result of prejudice against low-caste applicants.

The audit study was conducted by using job search websites for white collar jobs. Applications were made for entry level white collar jobs which were based in Chennai and advertised on these websites between March and December of 2006. Two resumes were sent for each job vacancy, one with a high-caste sounding name and the other with a low-caste sounding name. High and low-caste sounding names were assigned randomly to the resumes so that sometimes the same resume was associated with a high-caste sounding name when applying for one job vacancy and a low-caste sounding name when applying for another job vacancy. Resumes depicted applicants of approximately the same level of productivity. One thousand and forty six resumes in customer services and front office/admistration were sent with one hundred and fifty five resumes receiving callback. Resumes which had high-caste sounding names received higher callback in comparison to resumes which had low-caste sounding names. On average, a high-caste applicant had to send 6.2 resumes to get one callback while a low-caste applicant had to send 7.4 resumes to get one callback, a difference of approximately 20%.

The nature of the audit study allows me to look at the variation in callback gaps associated with recruiter and firm characteristics. The effect of low caste on callback is negative for male recruiters and for Hindu recruiters, but it is positive for female recruiters and for non-Hindu recruiters. This finding is interesting since it cannot be easily explained by statistical theories of discrimination, indicating that at least some prejudice might be present against low-caste applicants. The effect of low caste on callback is negative for firms with a larger scale of operations (with multiple domestic offices or with foreign offices) but positive for firms with a smaller scale of operations (without multiple domestic offices or without foreign offices). This observation is consistent with taste-based theories of discrimination which tell us that non-discriminating firms grow faster since they make higher profits than discriminating firms. It is also consistent with commitments made by large firms in the white collar sector to hire more workers from the disadvantaged caste groups. To investigate further the heterogeneity in callback by recruiter and firm characteristics, the average treatment effects on the callback outcome are estimated for different sub-populations of recruiters and firms. The average treatment effect i.e. the average difference in callback between high and low-caste applicants is positive for the overall population and for male recruiters, Hindu recruiters and firms with a small scale of operations. The average treatment effect is negative for female recruiters, non-Hindu recuiters and firms with a large scale of operations. Heterogeneity in the average treatment effects across the different sub-populations persists when confidence intervals constructed around the average treatment effects are compared.

Caste-based affirmative action policy has a long history in India. This paper does not provide a conclusive argument for or against caste-based affirmative action. It does provide convincing evidence on whether or not there exists discrimination in the white collar labor market within India. The existence of large scale discrimination would certainly strenghten the case for a caste-based affirmative action quota. The study finds that particular groups of recruiters and firms discriminate significantly against low-caste workers in comparison to high-caste workers. It is suggested that more and larger of such studies be carried out in other parts of the country before unequivocal policy recommendations may be made.

The paper is organized as follows: the second section gives some background literature on audit studies and the third provides information on the city of Chennai in which the audit study was carried out including the caste affiliation and employment of the city's labor force. The fourth section gives the methodology used in the audit study together with details of the fieldwork. The fifth section provides the results from the audit study. The sixth section gives an interpretation of these results and their policy implications. The last section concludes.

2.2. Related Literature

The audit study method has emerged in social science research as an attractive method to measure discrimination. In general, evidence of discrimination is obtained by using survey data on labor market outcomes (such as wages) together with worker attributes. Regressions of labor market outcomes on attributes of workers that correlate with productivity are run and differences in the outcomes across different groups taken as evidence of discrimination. However survey data does not include all correlates of productivity or information on individual attributes which are generally used by employers to make hiring and wage decisions. This makes it possible that what is taken as discrimination is in fact a difference in productivity across the different groups which is observable to the employer but not observable to the researcher.[6] Also there is little survey data on hiring

[6] For a review of the literature on race differentials in the US see Altonji and Blank, 1999.

decisions made by employers (as opposed to wage decisions) and consequently measurement of discrimination in hiring decisions is difficult to carry out. These problems have led researchers to rely on evidence from either natural experiments[7] or from audit studies.

An audit study is a field experiment in which researchers have the same information on worker attributes as employers so that unobservable (to the researcher but not the employer) correlates of productivity are eliminated; in fact researchers can artificially control these worker attributes depending on what kind of an audit study is being used. There are two different kinds of audit studies which have been used for detecting and measuring the extent of discrimination: matched-pair audits and resume-based audits (or correspondence testing).

In matched-pair audits individuals or auditors are hired and matched on as many observable characteristics as possible. Differential treatment of auditors who vary only in the characteristic of interest (race or gender) is then taken as evidence of discrimination. Such matched-pair audits were initially used in the 1970s in the US to detect discrimination in housing markets and led to implementation of the fair housing laws. They have been used to study discrimination in behavior as diverse as labor market outcomes[8], automobile purchases[9], tipping behavior[10] and home insurance.[11] Matched-pair audits have been criticized due to the absence of double blind procedures, the small number of auditors and the problems in matching pairs of auditors on all characteristics which may potentially correlate with productivity.[12]

[7]Goldin and Rouse, 1996.
[8]Neumark et al, 1995, Turner et al, 1991, Cross et al, 1990, Moreno et al, 2004.
[9]Ayres and Siegelman, 1995.
[10]Ayres et al, 2004.
[11]Wissoker et al, 1997.
[12]Heckman, 1998 and Heckman and Siegelman, 1992.

The second type of audit study is the resume-based audit, known also as correspondence testing. Resume-based audits involve sending resumes of hypothetical workers to employers who have identical productivity but who vary in the characteristic of interest (race or gender). For instance identical resumes with male and female first names may be used in the resumes to test for gender discrimination. Unlike a matched-pair audit, a resume-based audit may be carried out on a larger scale since one does not need to hire and train auditors. In addition there are no problems associated with double blind procedures and in trying to match real individuals on all possible characteristics that might correlate with productivity. However the outcome which is used to measure discrimination in these studies is callback for an interview, not actual hiring, so this method will only be able to detect discrimination in the early stages of the hiring process. Finally resume-based audits (and audits in general) are limited to detection of discrimination in hiring that is done through advertising for vacancies. They cannot study hiring decisions that are made through contacts and social networks. This makes it more practical to carry out the audits in professions where advertising for vacancies is the norm.

The earliest resume-based audits were carried out in the UK. The first such study was done by Jowell and Prescott-Clark (1970) who sent resumes with race specific names to employers to check for discrimination. Later studies include those done in selected white collar professions in the UK by McIntosh and Smith (1974), who look at discrimination by race and by Firth (1981, 1982), who looks at discrimination by race and gender in the accounting profession. Riach and Rich (1991, 1995) carry out a resume-based audit in Victoria, Australia to look at discrimination by race and gender for some white collar professions in which written applications were the norm. Weichselbaumer (2003) carries out an innovative resume-based audit in Austria to find whether there is discrimination by sexual orientation (lesbian or straight) and by gender role (masculine female or feminine female) for female applicants in different occupations. A well known recent resume-based audit is by Bertrand and Mullainathan (2004) which looks at whether there is discrimination by race (white or african-american) in Chicago and Boston. This is an innovative study in that the authors use a randomized design and resumes with both similar and different skill levels to examine how the callback gap changes at higher and lower levels of skills. Bertrand and Mullainathan find in their study that resumes with white-sounding names had 50% higher callback than resumes with black-sounding names and that the callback gap is larger at higher levels of skill.

2.2.1. Background

Chennai is located in the south of India, on the Coromandel Coast of the Bay of Bengal. With a population in 2001 of 4.3 million[13] it is one of the largest metropolitan cities in India. It is also the capital of the state of Tamil Nadu and has served as an important administrative and commercial centre since the time of the British. The majority of the residents of Chennai are Tamilians and speak Tamil although English is widely spoken in the white collar professions. There are also large Telugu, Malayalee and Urdu speaking communities in Chennai. According to the 2001 census, Chennai had a literacy rate of 85.3% and 1.5 million workers.

[13]Census of India, 2001.

2.2.2. Caste Composition and Inequality

Hindus formed 85% of the urban population of Tamil Nadu in 2004-05.[14] Of the Hindus, 15.3% belong to the SC and ST while 76.4% belong to the OBC. This leaves the high-castes in a small minority (at 8.3%). These proportions stand in stark contrast to the overall urban Hindu population of India, of which 20.5% was SC and ST and 36.9% OBC in 2004-05. Possibly due to the large proportions of the low caste groups, Tamil Nadu has a longer history of affirmative action in the public sector than any other part of India. Currently Tamil Nadu is the only state in India where the affirmative action or reservation quota exceeds 50%.

The per capita consumption distributions by caste category for the state of Tamil Nadu in 2004-05 are given in figure E.2.[12] Again, as was also the case for the entire country, the low-caste groups in Tamil Nadu do relatively worse than the high-caste groups in terms of per capita consumption. The other (high-caste) category has a per capita consumption distribution which is skewed to the right, with more people who have a high consumption per capita. On the other hand, the per capita consumption distributions for the SC and ST are skewed to the left, with more people in these caste groups who have a low consumption per capita. For the OBC the per capita consumption distribution is fairly symmetrical. The region of Tamil Nadu is different from the country as a whole in that it has some low-caste groups which have been doing quite well over the years. On the whole this region is characterized by high-caste groups which are a small minority and low-caste groups, which although worse off in per capita consumption compared to high-caste groups, do better than in other areas of the country.

[14,12,13,14,15,16,17] Data taken from the NSS carried out in 2004-05, 61st Round.

2.2.3. Labor Market Statistics by Caste Category

In India as a whole, the labor force participation rate for men (57%) is more than three times as high as it is for women (18%). Low labor force participation rates for women are the characteristic of labor markets across South Asia, so this statistic is not surprising. Across the different caste categories the participation rates are highest among the ST, followed by the SC, OBC, and others (high-caste). About 42% of the population was usually employed with the proportion being 37% in the urban areas. The worker population ratios were highest among the SC and ST. Among urban males the highest fractions of the chronically unemployed were among the SC followed by the others (high-caste) category. Among urban females the fractions chronically unemployed were slightly lower among the SC and ST in comparison to the OBC and others (high-caste) category.[13]

Tamil Nadu is among the most prosperous and urbanized states of the country. Hence, as is shown in the figures, Tamil Nadu does better than the Indian average on every labor market measure. The labor force participation rates for different caste groups by gender for the state of Tamil Nadu in comparison to India for 2004-05 are given in figure E.3.[14] Labor force participation rates are low among females in Tamil Nadu, as in the rest of the country, although they are higher for the lowest-caste women (those belonging to the SC and ST), again as is the case for the country overall. The OBC have the highest labor force participation rates among the men. For every caste group and gender type with the exception of high-caste women, Tamil Nadu has a higher labor force participation rate than the rest of the country. Worker population ratios and unemployment rates (given in figures E.4 and E.5) are also higher and lower, respectively, for every caste and gender type in Tamil Nadu than in India except for women from SC who face a higher unemployment rate in Tamil Nadu.[15] Overall there is greater employment in Tamil Nadu among both men and women. For women employment is high and particularly high among the low-caste women of Tamil Nadu, although they have a higher proportion unemployed than the rest of country.

Figure E.8 gives a breakdown of of employment for different caste categories across occupations.[15] The data is for 2004 and aggregates across the entire South Indian region (including the states of Andhra Pradesh, Kernataka and Kerala as well as Tamil Nadu). The figure shows that the low-caste groups (SC, ST, and OBC) slightly dominate the high-caste groups among service workers and markedly dominate the high-caste groups among skilled, semi skilled and unskilled workers and those working in agriculture. High-caste groups slightly dominate the low-caste groups among administrative, managerial, and clerical workers and markedly dominate the low-caste groups among professionals, government officials, and businessmen/self-employed. In the occupations which form the focus of the audit study, the differences between high and low-caste groups are not very large (service and administrative jobs), therefore it is not a priori obvious if there is discrimination present in hiring which sets a particular caste group at a disadvantage in comparison to another for the kind of jobs in my sample.

2.2.4. Employment and Industry

The main industries of Chennai have traditionally been automobile and automobile parts, but since the late 1990s there has been a high growth in outsourced jobs from the West.

[15]Data taken from the NES 2004.

Industries such as software services, hardware manufacturing, customer services and call centres have become increasingly important over the past decade.

Information on employment by gender is given in figure E.6.[16] Workers are predominantly regular employees with a large number of women not participating in the labor force. A detailed breakdown of employment by industry type for Chennai and all large cities of India (those with a population greater than a million people) for 2004-05 is given in figure E.7.[17] The figure indicates that workers in Chennai are employed primarily in manufacturing, trade, transport, and services, which is no different from other large Indian cities. However relatively more workers are involved in services in Chennai than in other large cities of India. The relatively larger number of service sector jobs in Chennai was an important motivation for carrying out the audit study in this city. Since there are a large number of new job openings in the service sector in Chennai posted on job websites, the location was chosen so as to collect a sufficient number of observations in a relatively short time period.

Chennai is located in a region in which the lower caste groups outnumber the high-castes by a large margin. The low-castes, particularly the low-caste women, have been active participants in the labor market, with labor force participation rates which are higher than the rest of the country. Currently Chennai has seen a boom in outsourced jobs from the West; it has become a centre of growth for a new kind of job and occupation. Outsourced jobs in customer service and other white collar professions have provided jobs to large numbers of workers. This paper looks at whether hiring in some of these new professions discriminates against some groups at the expense of others. Are the new jobs providing an avenue for the lower castes to improve their lot or not? The next sections describe in detail the audit study, the data collection, and the results obtained.

2.3. Audit Study: Methods

The resume-based audit study was carried out over a period of ten months between March 2006 and December 2006 in the city of Chennai. The sample of firms which were audited were firms which posted job vacancies online on job websites. These job vacancies were all located in Chennai and no firm was audited twice. Job websites are a new phenomenon and are used extensively for recruitment into white collar jobs in India. The largest of such sites have as many as 20,000 recruiters and 9 million resume postings. The majority of jobs posted on the job websites are in IT related fields, call centres and customer services, marketing, management, and in human resources.

Recruiters post job vacancies on the website and applicants post resumes. The recruiters can directly get in touch with applicants who have posted publicly available resumes. The applicants can also be the ones to contact the recruiter in response to a particular job vacancy posted by the recruiter. The method used in the study was that the resume of the applicants were not made publicly available and applicants contacted the recruiter in response to specific job vacancies.

An additional feature introduced by the main website that was used in the study (accounting for 70% of the observations) early in the data collection was that individual applicants who belonged to low-caste groups could declare their caste status. Low-caste applicants in the study had their status declared as low income OBC.[16]

[16]After the audit study was completed the callback gaps were checked for heterogeneity across website used (since an important difference was whether or not low caste status had been declared). It was found that there were no differences in the callback gap across the different websites but that the main website

2.3.1. Fieldwork

Prior to carrying out the study, a list of low and high-caste names was constructed that would easily convey caste affiliation. The conventions for Indian names vary across the country. Distinctively high or low-caste Tamil names were used, with high-caste names having Sanskrit roots and low-caste names having Tamil roots.[17] For instance, names such as Iyer or Iyengar belong exclusively to the high Brahmin castes. A partial list of some names that were used is given in Appendix 4. For each of the fictitious identities, an e-mail address was created that was carefully monitored over the course of the audit study.

In order to carry out the audit study, a set of fictitious resumes were needed which were close enough to resumes of actual job seekers so as not to arouse suspicion on the part of employers. Resumes of actual job seekers from cities other than Chennai as posted on different job websites were used. All contact and identity information about the individual applicants was removed from these resumes. Information from the different resumes was mixed so as to obtain a set of resumes which depicted applicants of approximately the same productivity for a particular job category. All resumes for a particular job category depicted applicants who had obtained the same degree (from different educational institutions of the same ranking in the field) and had the same set of skills (basic computer skills and between ten and twelve months of internship experience at firms for which the names were not given).

used had higher callback for all applicants than the other websites, probably due to the higher popularity of this website among recruiting firms.

[17] Once the list of names had been compiled, it was circulated among South Indian students at Northwestern University to check whether they could distinguish whether the name was High or Low caste from the name. The subset of names in the list for which all the students agreed on caste affiliation were used during the actual fieldwork.

Job search websites were used to identify job vacancies to which the applications could be sent; a variety of different job search websites were used for the purpose. Once a particular employer and job vacancy advertisement were identified, two resumes corresponding to the specifications of the vacancy were selected. If the vacancy specified a gender preference (for instance a female for a front office/administration job) then names of the specified gender only were used. The first resume was equally likely to be assigned a high-caste name or a low-caste name. Once the name assignment had been made to the first resume, the second resume was assigned a high-caste name if the first resume was assigned a low-caste name and a low-caste name if the first resume had been assigned a high-caste name. This forced half the resumes to be high-caste and half to be low-caste, with each firm receiving one low-caste resume and one high-caste resume. When assigning names to the two resumes, the name was also equally likely to be a male name or a female name (unless the vacancy specified a gender preference). After the name assignment was made, additional contact information was added to the resumes, a profile of the applicant created on the job website and the resumes e-mailed in response to the job vacancy. The two resumes were e-mailed within a few days of each other.

Callback by employers was measured by monitoring the e-mail addresses of fictitious applicants as well as by monitoring a number of telephone lines which had been obtained for the purpose in India. When a call was made to the telephone lines, it was either taken and the offer of interview rejected or the number and time of the call noted (the telephone numbers were matched with those given by the employer in the advertisement).

2.3.2. Strengths and Weaknesses

By the nature of its design, a resume-based audit eliminates productivity correlates that are observable to the employer but which are not observable to the researcher. All the productivity correlates which are used by the firm in making the callback decision are contained in the resume which is sent in response to the job vacancy advertisement. All the information in the resume is also available to me, which I can use when looking at differential callback. This is an important advantage for using data from an audit study instead of using survey data to look at differential callback, since I can rule out differences in productivity correlates observable to the firm but not to me as a cause of the differential callback. However it is important to note that it is still possible that there are some productivity correlates used by the firm in making the callback decision which are unobservable to both the firm and to me. For instance if the firm considers a good English accent an important productivity correlate for customer services jobs and it infers from the high-caste name that the individual is likely to have a good primary education and a good English accent, then this is a productivity correlate which the firm uses in making the callback decision but one which is not directly observable to either the firm or to me.

Another advantage of using the data from the present study is that the caste-specific names were randomly assigned to the resumes. The same resumes were sometimes associated with a high-caste name and at other times with a low-caste name. This randomization effectively broke the association between resume quality and caste. In other words, the randomization ensured that the low callback rates observed for low-caste applicants were not simply due to the low-caste names being associated with low quality resumes but due to their low-caste. However, since I do not vary the quality of the resumes being sent to the same firm, the randomization step is not crucial for the interpretation of my results.

A resume-based audit was chosen in preference to a matched-pair audit since it was a more cost effective method to sample a larger number of firms. Since the study did not hire or train any auditors there were no problems associated with a lack of double-blinding or of unobserved productivity differentials between auditor pairs. However, the choice of a resume-based audit meant that the set of jobs which form the sample are only white-collar jobs which actually use resumes sent by e-mail and for which there are online advertisements for job openings. Hiring into blue-collar jobs in India is done primarily through contacts and social networks, methods which cannot be subjected to an audit.

The study had initially been designed to include jobs not just in customer services and front office/administration but also in fields such as human resources, finance/accounts, and IT. Unfortunately when the fieldwork was actually carried out, the response rates in these job categories turned out to be extremely low. There could have been many reasons for the lower response rates in human resources, finance/accounts, and IT compared to customer services and front office/administration. One reason could be that employers do not use the job websites to make hiring decisions in these categories. However given the very large numbers of job vacancies posted in these categories on job websites this seems unlikely. The other reason could be that the resumes that had been prepared depicted workers who were underqualified for these jobs. Since all resumes depicted workers with an undergraduate degree only, it is possible that workers who have a masters degree and greater experience are preferred for human resources, finance/accounts, and IT. An attempt was made during the fieldwork to increase the number of projects and experience for applicants in some of these job categories but this strategy also failed to increase the response rates by firms. Due to the high non-response rates for these job categories the analysis makes use only of observations gathered for customer services and front office/administration. Unfortunately this reduces the sample considerably, from 2396 resumes to a total of just 1046 resumes, making the present study under-powered. Given the mean and standard deviation of callback among the groups of high and low-caste applicants, power calculations show that a substantially larger sample size is necessary to find the differences in callback to be statistically significant. This problem arose because it was difficult to estimate response rates by firms in the sample since there are no other published studies that make use of job websites in the white collar sector in India. Hence, some of the results that I find might well have been significant with a larger sample size. However the study, although small-scale, still makes some important observations, and future replications of its design would help improve the power of the results as well as providing more insights regarding external validity.

2.4. Audit Study: Results

The audit study was carried out between March, 2006 and December, 2006. A total of 523 job vacancies in customer services and front office/administration were applied to, and 1046 resumes were sent (two for each job). Job vacancies were selected from different online job search websites and applications were made via e-mail. All jobs were entry-level jobs and all respondents had an undergraduate degree in the same field (from colleges which were ranked the same in that field of study) as well as ten to twelve months of experience. The study did not vary the quality of resumes across the applicants. This meant that although the two resumes used for a particular job vacancy were not identical, they were nevertheless perfectly comparable in terms of education, skills, and experience. Callback was measured via e-mail and through the telephone numbers provided to recruiters.

The callback rate for high-caste applicants was 16.1% while the callback rate for low-caste applicants was 13.6%, with a 20% higher chance that a high-caste applicant gets called back for an interview. In other words a high-caste applicant had to respond on average to 6.2 job vacancies in order to get a single callback while the low-caste applicant had to respond on average to 7.4 job vacancies in order to get a single callback. The breakdown of resumes that were sent by job type and job website used are given in table F.1. Job type is a category created to simplify the discussion of the results. Applications were made to a variety of industries and occupations. For each of these different industries and occupations resumes were used that satisfied the specifications of the industry and occupation. However, the job-type category was created because all of the job vacancies could easily be put into a few well defined groups. The job categories used in the paper are customer services and front office/administration.

Of the resumes, 64% were used to apply for jobs in customer services and the remaining 36% for jobs in front office/administration. The response rate from firms was 17% for resumes sent to customer services jobs and 22% for resumes sent to front office/administration jobs. Four different job websites were used for the audit study: Naukri, Monster India, JobsAhead, and the Times of India. The Naukri website was used for approximately 70% of the resumes that were sent while the other three were used for the rest. All job vacancies posted in the designated job categories were applied to during the time period in which the audit study was in progress. The main constraint in choosing job vacancies was the frequency with which new firms posted vacancies. Naukri was used more often than any of the others due to the large number of postings on it by different firms, as it is the most popular job website in India at present. Not only did it have the highest number of job postings but also the highest response rates by the firms that were contacted (15% instead of the 13% response rate by firms posting vacancies on other websites).

2.4.1. Symmetry of Treatment by Job Vacancies

In this section I carry out tests on the null hypothesis of symmetry in treatment by recruiters for high and low caste applicants. Specifically, the tests determine whether the number of applicant pairs in which the high caste applicant is favored is significantly different from the number of applicant pairs in which the low caste applicant is favored. This is a weaker test than a test which tests for zero differences in callback but the advantage of this test is that it allows for the possibility of race neutral chance or randomness in hiring. Of the 523 job vacancies that were applied to in customer services and front office/administration, there were 28 applicant pairs for which the low-caste applicant was called back and the high-caste applicant was not, and 41 applicant pairs for which the high-caste applicant was called back and the low-caste applicant was not (as given in table F.2). For 43 applicant pairs both the high and low-caste applicants received callback, while for 411 applicant pairs neither of the applicants received callback. The symmetry tests which are used in this paper are the likelihood ratio test and the conditional sign test. [18]

For the likelihood ratio test the null hypothesis to be tested is that the number of job vacancies (firms) in which the high-caste applicant is favored is equal to the number of job vacancies (firms) in which the low-caste applicant is favored. Given that the outcomes follow a multinomial distribution, the constrained and unconstrained likelihood may be estimated and the chi square statistic estimated from these. It is then simple to use the chi square distribution (with a single degree of freedom) to determine whether to accept the null hypothesis. Given the data available from the audit study, the likelihood ratio test gives a p-value of 0.1168.

Another method to test for symmetry is to run the conditional sign test, which is a small sample test. Conditional on just one applicant receiving callback suppose it is recorded as a plus sign when a firm favors a high-caste applicant, the total number of plus signs (say Y) is then a binomial variable with a distribution $b(p, n)$ where n is the number of firms that respond to one applicant only and p is the probability that the high-caste applicant is favored. Then the sign test will test the null hypothesis that $p = 0.5$ (or there is symmetry in callback across caste) against the (one-sided) alternative $p > 0.5$. The sign test gets rejected when $|Y - 0.5n|$ is too large. Given the data from the audit study the p-value is estimated as 0.0740.

I find that high-caste applicants are more likely to receive callback than low-caste applicants and the number of cases in which the high-caste applicants are favored is higher than the number of cases in which the low-caste applicants are favored. The p-values when

[18] A detailed discussion of the tests is found in Heckman and Siegelman (1992) and Lehman (1986). Monte Carlo simulations are given in Appendix 1 to check the size and power of the two tests.

running tests of symmetry between high cand low-caste applicants are consistently either close to 0.10 or less, indicating there is a case to be made for the presence of caste-based discrimination. The next section gives the results when carrying out parametric analysis on the aggregate data collected. I find there are some important differences in the mean callback across different groups, although in all groups high caste applicants have higher callback than low caste applicants.

2.4.2. Job Types and Gender Pairs

The binary outcome for applicant j who faces firm i is given as

$$y_{ij} = \begin{cases} 1 \text{ if applicant receives callback} \\ 0 \text{ if applicant does not receive callback} \end{cases}$$

for $j = 1, 2$ applicants and $i = 1, ..., N$ firms. Then the binary outcome, y_{ij} follows a Bernoulli distribution with parameter $p_{ij} = P[y_{ij} = 1|x_{ij}, \beta, \alpha_i]$. Assuming a probit specification for the parameter gives the following

$$P[y_{ij} = 1|x_{ij}, \beta, \alpha_i] = \Phi(\alpha_i + x'_{ij}\beta)$$

where $\Phi(.)$ is the standard normal cdf, x_{ij} is the set of regressors for applicant j (including caste) when facing firm i, and α_i is the individual firm effect.[19] Of specific interest is determining the effect of a change in caste on the change in the probability of callback by the firm. Given the nature of the audit study, the assignment of caste and other regressors is random conditional on the firm (at least for firms that do not specify a gender requirement since gender is also included as a regressor). Therefore for

[19] A linear probability specification with fixed effects and robust standard errors on the entire sample of applicants gave similar results.

the subset of firms that do not specify a gender requirement, the individual firm effect may be treated as a random effect which is independent of the regressors. Assume that, together with the probit specification, the individual firm specific effects are normally distributed, $\alpha_i \sim N[0, \sigma_\alpha^2]$. Then the random effects maximum likelihood estimate of β and σ_α^2 maximizes the log likelihood $\sum_{i=1}^{N} \ln f(y_i|X_i, \beta, \sigma_\alpha^2)$, where

$f(y_i|X_i, \beta, \sigma_\alpha^2) = \int f(y_i|X_i, \alpha_i, \beta) \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} \exp(\frac{-\alpha_i}{2\sigma_\alpha^2})^2 d\alpha_i$

This random effects probit specification is carried out for the set of observations for which the job vacancy did not specify a gender requirement. Table F.3 gives the characteristics of the resumes which were sent in response to hiring firms, both in the complete sample and in the sub-sample on which the random effects probit was carried out. All resumes are pooled so that there are two resumes for every hiring firm with a total of 1046 across the complete sample. There are more women than men for the complete sample, with 55% of the complete sample of applicants being female. As already mentioned, a large number of the job vacancies in front office/administration requested a female, so that females form a higher proportion of applicants than do males. There are fewer resumes which apply for jobs in front office/administration jobs. Again this is the result of the fact that most of the job vacancies available on the job websites were in customer services. In the sub-sample in which gender assignment was random and on which the probit specification was done, there are a total 906 observations. In this sub-sample there are a higher proportion of resumes which apply for jobs in customer services than in front office/administration in comparison to the overall sample (69% instead of 64%).

Table F.4 gives the results from estimation of a probit model on the callback dummy with random effects at the firm level. In column (1), the regression result is reported for when caste is the only regressor. The effect of low-caste is a reduction by 0.19 in callback probability. In column (2) the regression result is reported when callback probability is regressed on gender and caste of the applicant together with interactions of applicant caste and gender. The effect of low caste for a female applicant is a reduction by 0.37 in callback probability. Being a male applicant also reduces callback, but at 0.01 the effect is more than ten times smaller than it is for female applicants and the reduction is not significant. In column (3) the regression result is reported when callback is regressed on interactions of job type and caste as well as interactions of job type and caste. The effect of low-caste in customer services is smaller than the effect of low-caste in front office/administration. Low-caste applicants have lower callback in both customer services and in front office/administration jobs. Column (4) gives the full specification when callback is regressed on the interaction of caste, applicant gender and job type. The effect of low-caste on male applicants in both customer services and front office/administration jobs are small. In fact the effect of being a low-caste male applicant in customer services jobs is actually positive, but insignificant. The effects of being a low-caste applicant are stronger for female applicants. In particular female applicants who are low-caste and who apply for jobs in front office/administration face significantly lower callback. The effect of low-caste is large and significant at the 5% level.

An important result that is obtained from carrying out the multivariate regression analysis is that low-caste reduces callback for both male and female applicants and for both customer services and front office/administration jobs. However the effect of low-caste on callback varies across both applicant gender and job type. For instance female applicants face higher reductions in callback due to lower caste than do male applicants.

Similarly resumes sent in response to job vacancies in front office/administration face higher reductions in callback due to lower caste than do resumes sent in response to job vacancies in customer services. The highest reductions in callback due to low-caste are observed for female applicants who respond to jobs in front office/administration. While a very detailed disaggregation is affected by the small number of observations, it is important to note that high-caste applicants always do at least as well as low-caste applicants. There is no job type or gender disaggregation in which low-caste is associated with higher callback. The next sections examine the callback gaps between high and low-caste applicants when information on recruiter and firm characteristics is incorporated in the analysis.

2.4.3. Recruiter Characteristics

Each employer advertisement for a job vacancy that was used in the study had the name of the contact person in the firm. It was possible to find some characteristics of the recruiter by looking at the list of names that were compiled from these advertisements (it should be noted that inference of such characteristics is purely subjective and there is no rigorous method by which to check the inferences other than directly surveying the recruiters themselves). Names were available for 379 recruiters. The callback decision is probably made by these individuals.

50% of the names were definitely male and 41% were definitely female. About 74% of the names were typical Hindu names and 16% were definitely not Hindu names (of which 55% were typical Christian names and 17% were typical Muslim names). This provides some information regarding the sample of recruiters who are making the hiring decision and were the subjects of my study.

In table F.5 the callback gaps by caste and recruiter name characteristics are given. There are heterogeneities in the callback gaps across the recruiter characteristics. The ratio of number of cases in which high-caste applicants are favored to number of cases in which low-caste applicants are favored is 2.1 for male recruiters and 1.5 for Hindu recruiters. It is 0.9 for female recruiters and 0.4 for non-Hindu recruiters. Tests of symmetry (the likelihood ratio and conditional sign tests) are performed on the different sub-samples with the results given in table T4. Tests of homogeneity (the one-sided Fisher exact test) across recruiter characteristics are also carried out.[20] The p-value for the test across recruiter gender is 0.136 and the p-value across recruiter religion is 0.118.

The effect of low-caste on callback across the different recruiter characteristics can also be obtained. Table F.6 gives the characteristics for sub-samples for which recruiter characteristics are available and for which the probit regressions are carried out. Table F.7 presents the estimation results from probit regressions with random effects at the firm level. Dummies for female applicants, job types and recruiter characteristics are included in the specifications. Again, low-caste reduces callback, as given in column (1). The effect of low-caste on callback is low, at just 0.08. Column (2) shows the effects of low-caste on callback seperately for male and female recruiters. The effect of low-caste on callback is negative for male recruiters but positive for female recruiters. These effects are larger than in column (1), with low-caste reducing callback by 0.34 for male recruiters and increasing callback by 0.22 for female recruiters. Column (3) lists the effects of low-caste on callback

[20]For details of the Fisher exact test and how it was estimated in the present case see Appendix 2 of the paper.

seperately for Hindu and non-Hindu recruiters. The effects of low-caste are negative for Hindu recruiters but positive for non-Hindu recruiters. Again these effects are larger in magnitude than those in column (1). For Hindu recruiters low-caste reduces callback by 0.21 while for non-Hindu recruiters low-caste actually increases callback by 0.62. Finally column (3) gives the effects of low-caste when recruiter gender and recruiter religion are interacted with low-caste. From this specification the largest effects associated with low-caste are found among Hindu recruiters who are male. Low-caste reduces callback by as much as 0.51 among this group of recruiters, and this reduction in callback is statistically significant. For all other groups of recruiters the effects are positive and the largest in magnitude for female recruiters who are non-Hindus.

The differences in callback gaps across recruiter gender and religion are interesting because recruiter characteristics are seldom available for analysis in empirical studies. This may potentially have implications for an interpretation of why discrimination arises in the first place. Heterogeneity in callback across recruiter characteristics such as recruiter gender is more consistent with a taste based theory of discrimination rather than an asymmetric information theory of discrimination. This paper suggests that inclusion of recruiter characteristics is important to understand fully how discrimination gets played out in the labor market.

2.4.4. Firm Characteristics

To analyze how the callback gaps varied by firm characteristics it was important to obtain more information on firms that advertised for vacancies on job search websites. In most cases the firms that advertised for vacancies also included a website address in the job vacancy. Some of the job search websites also had publicly available information on the website addresses of their clients. In short it was possible to find the websites of most firms that were present in the sample (around 53% of the total). These websites had some information that may be used to determine the scale of operations of these firms. The location information of the branch offices for different firms was a source of information which was utilized in this regard. Using the websites, it was found that 30% of these firms had offices in foreign locations outside of India and that 44% of the firms had offices in more than one city within India. These measures were used as firm characteristics to compare the callback gaps across the different firms (serving as a measure of large and small firms respectively).

Table F.8 gives the callback gaps across caste and across firm characteristics. There is substantial heterogeneity in callback gaps across the different firm characteristics. High-caste applicants are favored by firms without foreign offices and without multiple domestic offices in a larger number of cases than are low-caste applicants favored. On the other hand low-caste applicants are favored by firms with foreign offices and with multiple domestic offices in a larger number of cases than are high-caste applicants favored. The callback gaps (ratios) vary from 1.8 for firms without multiple domestic offices to 0.8 for firms with multiple domestic offices and 2.1 for firms without foreign offices to 0.6 for firms with foreign offices. When data is disaggregated by firm characteristics the symmetry tests reject the null hypothesis of symmetry for firms without multiple domestic offices and without foreign offices. These firms significantly favor the high-caste applicants in more cases than they favor the low-caste applicants. Firms with multiple domestic offices and foreign offices favor low-caste applicants over high-caste applicants. Tests of homogeneity of callbacks across firm characteristics (the one-sided Fisher exact test) give a p-value of 0.057 when the test is carried out for homogeneity in callback across firms with and without foreign offices. The p-value is 0.154 when the test is carried out for homogeneity of callback across firms with and without multiple domestic offices.

Parametric analysis of the callback dummy is carried out with the introduction of firm characteristics along the same lines as the previous two sections. Instead of recruiter characteristics, firm characteristics are added to the set of regressors. The sub-sample for which firm characteristics are available is of 279 firms and 558 applicants. The sub-sample in which the gender assignment was non-random (and on which the probit specification is run) consists of 478 observations.

Table F.9 gives detailed characteristics of the sample of resumes for which firm characteristics are available and the sub-sample in which gender assignment is random and firm characteristics are available. As in the previous regressions there are more resumes which apply for jobs in front office/administration in the entire sample for which firm characteristics were available than in the sub-sample for which the probit regressions are carried out. Approximately 44% of the firms in both sub-samples had multiple domestic offices. The proportion of firms with a foreign office was slightly larger in the probit sub-sample than in the overall sample for which firm characteristics were available, 33% instead of 30%.

Table F.10 presents the estimation results from running a random effects probit regression. The regression is run for the sub-sample of applicants who apply to firms for which the firm characteristics are available and for which gender assignment is random. In column (1) the effect of low-caste is given for the sub-sample of resumes for which firm characteristics are available. Low-caste reduces callback by 0.12. Column (2) gives the effects of low-caste on callback among firms with and without multiple domestic offices. In the sub-sample for which the probit is carried out the effect of low-caste on callback is negative for firms without multiple domestic offices but positive for firms with multiple domestic offices. The magnitudes for these effects are larger than in column (1). For firms with multiple domestic offices, low-caste increases callback by 0.47, but for firms without multiple domestic offices low-caste reduces callback by 0.36. Column (3) gives the effects of low-caste on callback among firms with and without a foreign office. Again the effects of low-caste are stronger in these groups than they were overall in column (1). Low-caste increases callback for firms with foreign offices by 0.42 but it reduces callback for firms without foreign offices by 0.37. Column (4) gives the effects when low-caste is interacted with both the presence of multiple domestic and foreign offices. In this specification, low-caste increases callback for firms which have both multiple domestic offices and foreign offices and this increase in callback is statistically significant. For all other firm types low-caste reduces callback, with the largest reductions in callback ocurring among firms without multiple domestic offices and without a foreign office.

Although caste in insignificant in explaining callback in the overall sample there is heterogeneity in callback gaps across the different kinds of firms. Given that firms without multiple domestic offices and foreign offices have a smaller scale of operations, it may be argued that among firms with a small scale of operations caste continues to play an important role in setting the low-caste applicant at a disadvantage in comparison to a high-caste applicant. The results for firms with larger scale of operations (multiple domestic offices and foreign offices) are not too surprising. In response to the government's perceived support for introduction of quotas for low-caste workers in private sector jobs, many large firms recently committed themselves to recruit more actively from among low-caste workers. The reason that the main job search website used (Naukri) introduced the feature allowing applicants to declare their caste status was precisely because they expected caste status to be beneficial to applicants given the commitments made. The results from the study indicate that although low-caste applicants are more actively sought by firms which have a larger scale of operations, this is not the case for firms with a smaller scale of operations. Low-caste applicants still face a serious disadvantage in callback when facing these firms. These results also follow naturally from taste-based theories of discrimination. These theories argue that non-discriminating firms do not incur the costs associated with hiring less productive workers from preferred groups and such firms should be making higher profits and growing faster than non-discriminating firms. For this interpretation, firm size is a consequence of firm practices (discriminate or not) rather than the other way around. Either or both of these interpretations would be consistent with the evidence.

2.4.5. Effects on Differential Callback of Recruiter and Firm Characteristics

Given the data from the study, an important finding is that the callback gap is higher among particular groups of recruiters and firms than among others. This may be seen more clearly in table F.11 which gives the effect of recruiter and firm characteristics on differential callback. I find, for instance, that male recruiters increase the callback gap for low-caste applicants by 0.53 and that Hindu recruiters increase the callback gap for low-caste applicants by 0.83. The callback gap for low-caste applicants also increases by 0.9

when they face firms without multiple domestic offices and this increase is significant at the 5% level. Finally the callback gap for low-caste applicants increases by 0.83 when they face firms without foreign offices and this increase in the callback gap is significant at the 10% level. In order to look at differences in callback across the different sub populations of firms and recruiters the next section looks at the average treatment effects on callback.

2.4.6. Average Treatment Effects

An important finding of the paper, that the gap in callback between high and low-caste applicants is higher in some sub populations than in others, also holds when looking at the average treatment effect across the population and in the different sub populations. Let the outcome be the callback probability which takes the value $Y(1)$ among high-caste applicants and the value $Y(0)$ among low-caste applicants. Then the average treatment effect is given by

Average Treatment Effect $= E[Y(1) - Y(0)]$

Given the randomization carried out during the field experiment, the Average Treatment Effect simplies considerably to

Average Treatment Effect $= E[Y|\text{high-caste}] - E[Y|\text{low-caste}]$

The average treatment effect may be found simply by comparing the sample averages and a confidence interval constructed around the treatment effects by using the bootstrap. I use 200 bootstrap replications to find the 90% confidence intervals around the Average Treatment Effects. In addition I also find the 90% confidence intervals around the Average Treatment Effects for specific sub-populations. These confidence intervals are given in table F.12.

The average treatment effect for the entire population is 0.02. This is the expected difference in callback probability between all applicants being high-caste and all applicants being low-caste. For the population, high-caste applicants have higher callback than do low-caste applicants. The average treatment effects for the different sub populations of recruiters and firms also give interesting results. The average treatment effect is positive for male recruiters and for Hindu recruiters. It is 0.05 for male recruiters and 0.03 for Hindu recruiters, higher than it is in the overall population. The average treatment effect is negative for female recruiters and for non-Hindu recruiters. It is -0.01 for female recruiters and -0.05 for non-Hindu recruiters. For these groups, the average difference in callback favors low-caste applicants over high-caste applicants. The confidence intervals on the average treatment effects take into account the sampling variation. These confidence intervals, although they contain zero, are still fairly informative, being skewed positively for both male and Hindu recruiters. The average treatment effects across different firms are also given in the table. For firms without multiple domestic offices or without foreign offices, the average treatment effects are positive and larger than for the overall population, at 0.06. For firms without foreign offices the confidence interval also does not include zero and is entirely positive. For firms with multiple domestic offices and with foreign offices, the average treatment effect is negative, being -0.01 for firms with multiple domestic offices and -0.04 for firms with a foreign office, indicating that callback is higher for low-caste applicants than it is for high-caste applicants when they are facing these firms. Although this section makes the important identifying assumption that the distribution of outcomes (callbacks) is the same for recruiters and firms which are part of my sample but for which I do not observe the recruiter and firm characteristics, it does show that the main results of the paper do not rely simply on the parametric specifications used in the previous sections.

2.5. Discussion

The results from this audit study indicate that there are more firms which favor high-caste applicants over low-caste applicants rather than the other way around. There is considerable heterogeneity found in the callback gap by caste when information on recruiter and firm characteristics is incorporated in the analysis. I found that male and Hindu recruiters have larger callback gaps favoring the high-caste applicant than female and non Hindu recruiters. In fact, female and non-Hindu recruiters are more likely to call back low-caste applicants than high-caste applicants. I also found that firms without multiple domestic offices and without foreign offices have larger callback gaps, favoring high-caste applicants over low-caste applicants. Firms with multiple domestic offices and with foreign offices tend to favor low-caste applicants over high-caste applicants.

2.5.1. Taste Based vs. Statistical Discrimination

How does the data collected during the audit study relate to the theoretical literature on labor market discrimination? The study of discrimination within the labor economics literature goes back to the early 1960s when Becker first described such discrimination as a result of prejudice or taste. The employer was modelled as willing to forego some money income in order to avoid associating with people of a certain race in comparison to others. Alternatively, employees or consumers might be willing to forego some money income in order to avoid associating with people of a certain race in comparison to others. The resulting disparity in outcomes which results from this prejudice is referred to as employer, employee, or consumer discrimination. An unattractive feature of the early taste based model of discrimination as discussed by Becker and others was that it failed to explain the persistence of discriminating firms in the long run, since such firms should be making lower profits than would competitive non-discriminating firms. Later models either introduced search frictions into the taste-based models[21] or modelled discrimination as an information problem.[22] The latter class of models, referred to as models of statistical discrimination, assume that firms have incomplete information about the actual productivity of a worker when the hiring decision is being made so they use either racial stereotypes[23] or signals which might be more informative about some racial groups as compared to others.[24] Groups of prospective workers which have identical productivity ex-ante may turn out to have different productivity levels ex-post due to the information problem and differing incentives for workers in different groups to invest in human capital. It seems plausible to conclude that both types of discrimination– taste based and statistical– may be present in the labor market although empirical researchers have found it difficult to disentangle the two.

All the productivity correlates which are observable and which get used by the firm in making the callback decision are contained in the resumes which are sent in response to the job vacancy advertisements. I can therefore rule out as a cause of differential callback any differences in observable productivity between high and low-caste applicants, since all

[21]See for instance Borjas and Bronars 1989, Black 1995 and Bowlus and Eckstein 1998.

[22]See Arrow 1973 and Phelps 1972 for early examples. In Foster and Vohra 1992 group disparities arise as a result of co-ordination failure.

[23]Coate and Loury 1993 and Moro and Norman 2004 which extends the Coate and Loury framework to endogenize the wage rate and carry out general equilibrium analysis.

[24]Aigner and Cain 1977, Lundberg and Startz 1983 and Lundberg 1991.

the resumes which were used had the same level of observable productivity. However, it is possible that the hiring firm infers more from the resumes than observable productivity. Suppose for instance firms associate high-caste names with good primary education and better English accents for jobs in customer service. In this case the differential callback could be due to differences in unobservable productivity which I cannot observe in the study. This is an example of statistical discrimination. In other words, differential callback observed in the study could arise not just as a result of prejudice of hiring firms against low-caste applicants, but it may also arise as a result of differences in unobservable productivity across the different groups of applicants. Therefore the audit study method is unable to distinguish between taste-based discrimination and statistical discrimination explicitly. However, the results that are observed from the study in this paper make it likely that at least some of the callback gap is due to employer prejudice.

One would expect rational and informed recruiters to statistically discriminate against low-caste applicants if they believed that the expected productivity of low-caste applicants was less than the expected productivity of high-caste applicants. However, the randomization in the audit study implies that expected productivity should not depend on the background of the recruiter, thus the callback gap should not vary across recruiter background. Therefore the variation in callback gap across recruiter background that I find has to be coming, at least partly, from prejudice and not differences in expected productivity (statistical discrimination). This reasoning assumes that recruiter background is not related to the type of skills demanded by the job vacancy.[25]

[25]See Anwar and Fang (2004) for development of statistical tests which test for whether troopers of different races are monolithic in their search behavior and whether they exhibit relative racial prejudice. One problem when testing for statistical discrimination in labor markets instead of mortgage lending or

Suppose X is the set of productivity correlates used by the recruiter in making the callback decision and that $X = \{X_u, X_o\}$. Also assume that X_u are productivity attributes unobservable to both the employer and the researcher and X_o are productivity attributes which are observable to both the employer and the researcher. Assume further that X is additively separable in X_u and X_o, then

$E[X|X_o,\text{high-caste}] = X_o + E[X_u|X_o,\text{high-caste}]$

and

$E[X|X_o,\text{low-caste}] = X_o + E[X_u|X_o,\text{low-caste}]$

since observable productivity is constant and known for all applicants by the nature of the audit study design. Then it is also true that

$E[X|X_o,\text{high-caste}] - E[X|X_o,\text{low-caste}] = E[X_u|X_o,\text{high-caste}] - E[X_u|X_o,\text{low-caste}]$

Assume for simplicity there are two groups of recruiters, $R = \{m, f\}$. Given the presence of statistical discrimination, there is a population expectation of unobservable productivity and all rational recruiters are aware of this expectation (or at least form expectations the same way). Then

$E[X_u|X_o,\text{high-caste}, R = m] = E[X_u|X_o,\text{high-caste}, R = f]$

and

$E[X_u|X_o,\text{low-caste}, R = m] = E[X_u|X_o,\text{low-caste}, R = f]$

Given this assumption,

$E[X_u|X_o,\text{high-caste}, R = m] - E[X_u|X_o,\text{low-caste}, R = m]$

$= E[X_u|X_o,\text{high-caste}, R = f] - E[X_u|X_o,\text{low-caste}, R = f]$

or

racial profiling (as in Anwar and Fang (2004)) is that the outcome of interest, actual worker productivity, is not available. Therefore the tests from Anwar and Fang (2004) cannot be applied in this context.

$$E[X|X_o,\text{high-caste}, R = m] - E[X|X_o,\text{low-caste}, R = m]$$
$$= E[X|X_o,\text{high-caste}, R = f] - E[X|X_o,\text{low-caste}, R = f]$$

So if there is no prejudice the callback gap should be the same for both groups of recruiters. This is not the case, implying there is some prejudice present. There are some important caveats concerning the above argument; it assumes implicitly that recruiter background is not related to the type of skills required by the job. For instance higher callback gaps for male recruiters would be observed if male recruiter carry out more recruiting in jobs for which unobserved components of productivity are more important. However this is not true for the sample of jobs in my sample. The distribution of recruiter background among the kinds of jobs is fairly similar: 63% of the male recruiters and 62% of the female recruiters in the sample were recruiting for jobs in customer services. This is in comparison to 64% of the job vacancies in customer services for the entire sample of recruiters in the dataset and 63% of jobs in customer services for the sub-sample for which recruiter characteristics were available. As regards recruiter religion, 63% of the Hindu recruiters were recruiting for jobs in customer services while 61% of the non-Hindu recruiters were recruiting for such jobs.

The heterogeneity in the callback gaps across different kinds of firms also follows naturally from taste-based theories of discrimination, although one cannot rule this out as being the result of information asymmetries. According to taste-based theories of discrimination, non-discriminating firms make higher profits than discriminating firms since they do not incur the costs of hiring high-caste workers of low productivity. Since they make higher profits they also grow faster than discriminating firms. Therefore non-discriminating firms are larger than discriminating firms as a result of their non-discriminatory practices.

2.5.2. Policy Implications

What are the implications of the present study on policy? Affirmative action quotas for low-caste workers in Tamil Nadu have been in operation for decades. These quotas had been introduced in public sector jobs, education, and political representation for the Scheduled Castes and Scheduled Tribes only. They were later extended to the Other Backward Castes. Recently there has been some debate as to whether these quotas should also be introduced in private sector white-collar jobs. The study described in this paper can help inform somewhat policy analysis regarding affirmative action but it gives no definitive evidence one way or the other.

The efficacy of affirmative action policy, at least theoretically, is supect. In the Indian context, there have been cases made both for and against affirmative action. In general, the policies which have been in place have not had led to large scale improvements for low-caste individuals relative to other groups. Often, the places which are reserved for low-caste individuals remain unfilled and the ones to benefit are the most advantaged among the low-caste. On the other hand, affirmative action policy is a relatively costless policy for the government (although not for the society) to try to reduce inequalities between different socio-economic groups. The question that this paper can help with is whether large-scale discrimination exists in the private sector, since this would strenghten the case for the introduction of affirmative action.

The main result of the audit carried out in Chennai was that, in the overall sample of hiring firms, low-caste applicants had lower callback than did high-caste applicants, but the differences in callback were not statistically significant. However the magnitude of the difference in callback for applicants of different castes is non-trivial. If an average applicant gets one job after interviewing at ten different places, then a high-caste applicant will need to send her resume to 62 different job vacancies to get a job while a low-caste applicant will need to send her resume for 74 different job vacancies to get a job. This assumes that there is no discrimination at the interview stage of the hiring process, an assumption which need not hold in the real world. It is financially costless for the low-caste applicant to send an additional ten resumes using the job website, but there are costs associated with a longer wait time while new vacancies become available and these costs are likely to be higher the fewer the alternatives to job websites in searching for a white collar job. While carrying out the study, ten months were required to find a little more than five hundred vacancies in customer service and front office/administration for entry level positions by distinctive firms. This suggests that to get one callback, the low-caste applicant has to wait a little more than half a day compared to a high-caste applicant, if both are applying only for entry level jobs in customer service and never apply to the same firm twice. When applying for entry level jobs in front office/administration, the low-caste applicant has to wait for five-and-a-half days more to get one callback compared to a high-caste applicant, if both never apply to the same firm twice.

Another important result from the study was the heterogeneity in the callback gap across recruiter and firm characteristics. The heterogeneity across recruiter characteristics indicates the presence of prejudice, and heterogeneity across firm characteristics suggests that although firms with a larger scale of operations seek out low-caste workers, this is not the case for firms with a smaller scale of operations. Low-caste applicants still face a significant disadvantage when applying for jobs at firms with a smaller scale of operations.

The presence of discrimination in hiring would lead low-caste workers to invest less in their human capital skills than they otherwise would. A caste-based quota in hiring would lead to a larger number of low-caste workers being hired and reduce the inequality they face in hiring. This would also improve their incentives to make human capital investments.[26] Given the disparities in human capital between the different caste groups, a caste-based quota is a relatively costless policy (for the government but not for the society) which might lead to a more equitable outcome. It might also be justifiable given that low-caste applicants face some disadvantage in the private sector. However, at the same time it would be premature to unequivocally support caste-based quotas. The collection of more and larger of such datasets would be an important prelude to providing a context for the design of the best policy for the welfare of low-caste workers.

2.6. Conclusion

What do the results from the audit study tell us and what is their impact on the current debate regarding the introduction of caste-based affirmative action quotas in private sector jobs? The resume-based audit study reveals that low-caste applicants receive lower callback than high-caste applicants irrespective of job type (customer service or front office/administration) or gender (female applicant or male applicant). Low-caste reduces callback more for jobs in front office/administration than it does for jobs in customer services. Low-caste also reduces callback more for female applicants than for male

[26]It should be noted that the theoretical literature on discrimination provides us with ambigous results on the effects of affirmative action policy. In Coate and Loury (1993) and Moro and Norman (2003) affirmative action may lead to patronizing equilibria in which discriminated groups find it easier to get jobs in high skill sectors and this leads them to invest less in their human capital rather than more.

applicants. The effect of low-caste for female applicants who applied for jobs in front office/administration is significantly negative.

Incorporation of recruiter and firm characteristics into the analysis reveals substantial heterogeneities in callback gaps across these characteristics. I find that the effect of low-caste on callback is negative for resumes sent to male recruiters and to Hindu recruiters but that the effect of low-caste on callback is positive for resumes sent to female recruiters and to non-Hindu recruiters. The effect of low-caste on callback is negative for resumes sent to firms that have a smaller scale of operations (absence of multiple domestic offices or any foreign offices) but the effect of low-caste on callback is positive for resumes sent to firms that have a larger scale of operations (presence of multiple domestic offices or foreign offices). The variation of the effect of low-caste on callback across the different recruiter characteristics indicates that at least some of the discrimination observed in favor of the high-caste applicants is taste based rather the result of information asymmetries. The variation of the effect of low-caste on callback across the different firm characteristics is important since it indicates that low-caste significantly disadvantages the applicant when applying for jobs with firms who have a smaller scale of operations. This variation is consistent with taste-based theories of discrimination and also with commitments made by large firms to recruit more actively from amongst low-caste groups. The heterogeneities in callback across different groups of recruiters and firms may also be seen by looking at the average treatment effects across the different sub-populations of recruiters and firms. The average treatment effects support the earlier analysis carried out by using multivariate regressions.

Given the results from the audit study, is there a case to be made for the introduction of caste-based affirmative action qoutas in private sector jobs? I find that there is strong evidence of discrimination among particular groups of recruiters and firms. A caste-based quota would potentially force all recruiters and firms to hire more low-caste workers. Therefore the results of this study provide some support for the introduction of hiring quotas by caste in the private sector. However, more and larger studies need to be carried out before more definite policy recommendations may be made. Replications of the design given is this paper in different urban centres of India would be useful in carrying out more definitive policy analysis.

Another important issue concerning the study is that of external validity: how far can the results of the study be generalized to other labor market settings within India? The study was carried out for white collar jobs in the city of Chennai. It is not clear whether the patterns observed in Chennai are the same as would be found in other large urban areas of India. The inter-caste dynamics vary in the different areas of the country, although it would be fair to generalize the results from Chennai to other large urban areas in the South of the country (such as, for instance, Bangalore and Hyderabad) in which most of the new white collar jobs of the country are located. Also, the study was carried out for entry-level jobs in specific white-collar professions. It is not clear whether the same patterns would be observed in white-collar jobs that require higher skills (for instance jobs in IT) or in the blue collar professions. For white-collar jobs that require higher skills, caste may play an important role. Professional occupations are heavily dominated by high-caste groups within India. Although the white-collar jobs in this study requiring greater skills (such as for instance in IT, finance and human resources) were dropped from the analysis due to high non-response, the few observations which were obtained indicated even larger gaps in callback in favor of high-caste applicants. As regards blue collar professions in India, it has been noted in the past that networks play a very important role in these professions.[27] Hiring is done in very different ways within the blue-collar professions in comparison to the kind of hiring which is the subject of the audit in this paper. It is possible that low-caste actually provides advantages to applicants in certain blue collar professions. This would be the case, for instance, if a particular low-caste group dominates a blue collar profession and has strong networks in that profession.

The audit study as a tool to measure discrimination has gained more importance recently. In particular, the use of resume-based audits is now recognized by economists as an innovative technique to gather clean and reliable evidence of discrimination in different labor market settings. It has proved to be useful in order to study more carefully employer's hiring decisions on which there existed little data. Variations on the resume-based audit methods may be used to gather information on how employers make hiring decisions when faced with applicants who vary across different dimensions. In the present context of caste-based discrimination, it is very important to have more and larger of such audit studies conducted in different parts of the country. It would also be useful to have audit studies which use channels other than job search websites to apply for jobs, as well as matched-pair audits which would measure discrimination at the interview stages of the hiring process as well as in callback.

[27]Munshi and Rozensweig

CHAPTER 3

Treatment Effects in the Absence of Complete Randomization: The Minneapolis Domestic Assault Experiment

3.1. Introduction

The usefulness of randomly assigning treatments to determine treatment effects can be traced to the works of Fisher (1935). Random assignment of treatments ensures that the distribution of outcomes experienced by the treatment group is the same as the distribution of outcomes that would result if the treatment were received by the entire population. As a result, average treatment effects in an experimental setting are fully identified, and may be estimated trivially. It is for this reason that field experiments in the social sciences randomly assign treatments to subjects in an experiment. Unfortunately, how far the random assignment of treatments is actually carried out in practise is in many cases outside the control of the researcher. There may be non-compliance with the randomly assigned treatment or there may even be some self selection of treatment in cases when the administration of treatment is not done directly by the researcher. In such cases there are again problems that arise with identification of the treatment effects since there are important counterfactual probabilities which cannot be observed. In the case of non-compliance, the counterfactual probability is the outcome probability associated with the assigned treatment among the group of subjects who did not comply with the assigned treatment. In the case of non-random assignment of treatment, the counterfactual probability is the outcome probability associated with a treatment among subjects who did not receive the treatment. In the second case the problem is essentially the same as would arise if we had data from a non-experimental setting. It is possible to find partially identified treatment effects that make no or minimal assumptions on the unknown counterfactual probabilities. The advantage of using this approach is that it makes minimal assumptions. The disadvantage is that instead of finding a specific estimate for the treatment effect one may only be able to estimate a bound on the average treatment effect.

I use data from the Minneapolis experiment carried out in 1983 to determine the effect of different treatments (arrest, advice and separation) on the repeat incidance of domestic assault. The randomization was carried out by officers instead of the researchers. Although the treatment assignment was done randomly, there were possibilities that the randomization did not proceed as expected. The specific problems were raised by the original authors of the experiment as well as by subsequent researchers (see section two for more discussion). In this paper, I estimate the partially identified treatment effects under different assumptions on the treatment assignment, without taking a position on which assumption is the best one to make. The treatment effects which are found to proceed from assumptions which are judged the most credible may then be used.

An important contribution of this paper is the application of the literature on partially identified treatment effects to a substantive problem of interest.[1] It is shown how the estimation of partially identified treatment effects may be carried out very easily. The advantage of the approach is the greater credibility of weaker assumptions than are used in conventional analysis for the estimation of treatment effects.

[1] See Manski (1990) and Manski (2003).

In the paper, partially identified recidivism probabilities associated with the different treatments are estimated first without making any assumptions on the counterfactual probabilities, under the assumption that treatment assignment is not random. The recidivism probabilities associated with the different treatments are also estimated when making the assumption that assigned treatments are perfectly random but without making any assumptions on the counterfactual probabilities due to non-compliance with the assigned treatment. Finally, to improve the no assumptions bounds on the recidivism probabilities associated with different treatments, I use two different models of self selection in treatment assignment which are behaviorally motivated from Manski and Nagin (1998). The models of treatment assignment are the Skimming model which assumes officers arrest all high risk offenders and the Outcome Optimization model which assumes that officers assign treatments to minimize recidivism. I find that arrest is associated with the lowest recidivism probability given that the assigned treatment is perfectly random and no assumption is made on the counterfactual probabilities due to non-compliance. In addition, arrest is also associated with the lowest recidivism in comparison to the other treatments if assigned treatments are not random, provided officers assign treatments by arresting all high risk offenders (Skimming). Arrest is not unambigously associated with the lowest recidivism in any of the other cases. The recidivism probabilities are also used to find the average treatment effects from having a mandatory arrest policy for the entire population of domestic assault offenders whose offence gets reported to the police. In the paper, the average treatment effect is the difference in recidivism probability under mandatory arrest and mandatory non-arrest policies (either advice or separation).

In the next section the background of the experiment is discussed, the third section describes the data and the fourth carries out an analysis of the data from the experiment. The fifth section concludes.

3.2. Background

Domestic violence is an important problem; in the US there were 5,341,410 victimizations in 2002 with 27.4% of the total being victimizations committed by an intimate (The National Criminal Victimization Survey, Family Violence 2002). How should the government and law enforcement agencies deal with the problem of domestic violence? Some of the suggested treatments for offenders of domestic violence are arrest, advice or separation of offender and victim. Which of the three treatments is the most effective in reducing the incidence of domestic violence?

The issues mentioned above are relevant to the question of whether punishment reduces crime. In the context of the domestic violence the question is whether arrest is the most effective treatment. Sociologists have opposing theories regarding the effects of punishment on behavior: according to specific deterrence punishment deters people from repeating crime whereas the labeling school of deviance says that punishment makes people commit more crimes due to the negative consequences that result from labeling an individual as deviant.

Randomized experiments are a common method employed in empirical social science literature to assess the impact of different treatments. Randomized experiments to determine the impact of different treatments on repeat incidence of domestic violence have been carried out and had a large impact on public policy over the 1980s and 1990s. The first and most influential experiment was the Minneapolis experiment carried out in 1983 over a period of eighteen months. Repeat incidance (recidivism) of domestic assault against the same victim was measured over a six month follow up period using criminal justice reports and victim interviews for offenders who were randomly assigned the treatments of arrest, advice (informal mediation at the officer's discretion) and an order to the offender to leave for eight hours.

Parametric analysis of the experiment (in Sherman and Berk (1984)) was carried out by analyzing linear probability, logit and proportional hazard specifications. The first two specifications used a dummy variable outcome which indicated recidivism while the third specification used a time to failure measure as the outcome. The three specifications were run separately for outcome data from official reports and outcome data from victim interviews. The analysis showed that arrest resulted in significantly lower recidivism than separation when using official data and that arrest resulted in significantly lower recidivism than advice when using victim interview data.

Since arrest was associated with the lowest recidivism from the parametric analysis that was carried out, the experiment ended up playing an important role in adoption of a mandatory arrest policy nationwide. The decision to impose mandatory arrest came under criticism due to concerns regarding external and internal validity of the experiment (see Binder and Meeker (1988 and 1992), Lempert (1989), and Buzawa and Buzawa (1996)).

The concerns regarding internal validity of the experiment centered around how the randomization was carried out (see Berk and Sherman (1988) and Gelles (1993)). The method of treatment assignment in the Minneapolis experiment was as follows: for each officer in the program there was a pile of randomly arranged color coded forms with the color of the form representing the treatment (arrest, advice or separate) that was to be given to the offender. When a call reporting a case of domestic violence came in the officer determined whether the case was eligible for the experiment and then applied the treatment that was topmost in the pile of color coded forms. Once a case was made eligible for the experiment the officer could still choose to deviate from the randomly assigned treatment when applying the treatment, provided sufficient reason for deviation was recorded. There were two possible problems that could potentially occur with random assignment of treatment. If the officers systematically made cases ineligible for the study (the possibility of 'differential attrition' in Berk and Sherman (1984)) then the treatment assignment would be non-random. If the officers did not apply the randomly assigned treatment for an eligible case then there would be non-compliance with the randomly assigned treatment. Both of these possibilities have been raised by the original authors and others. Under either of the two possibilities, however, treatment effects may be estimated which would be an improvement over the no assumptions case but which would give weaker results than if neither of the two possibilities existed. This paper gives treatment effects under all the different assumptions separately, allowing an evaluation of the results on the basis of whether the assumptions that are used to obtain them are credible.

The experiment showed that arrest was aasociated with the lowest recidivism for the population of domestic violence offenders in the two Minneapolis precincts in which the field experiment was carried out but could the results be generalized to other populations and time periods? Replications of the Minneapolis experiment carried out in different districts and time periods were unable to find conclusive evidence that arrest was associated with the lowest recidivism making external validity suspect. Replications in Milwaukee, Charlotte and Omaha produced evidence that arrest increased recidivism rates whereas replications in Colorado Springs and Metro-Dade produced victim report data according to which arrest reduced recidivism rates. There were important differences among the replications: they used different sample sizes, different treatments and better treatment assignment methods as compared to the Minneapolis experiment. Treatment assignment was improved by requiring the police officers to determine eligibility on the scene and then call the dispatcher for a random assignment of treatment. However it was still possible for police officers to fail compliance with the randomly assigned treatment if they strongly felt that a particular treatment is best for a particular case. Also recidivism was measured in replications as repeat offenses by the offender not only against the original victim but against any victim.

3.3. Data

The experiment was carried out in 1981-82, initially by a group of 33 police officers in the two Minneapolis precincts with highest density of domestic violence crime reports and arrests. table G. 1 is reproduced from Sherman and Berk (1984) and gives characteristics of part of the sample (205 of the total 312) for whom initial interviews were obtained. The couples which reported domestic violence and which formed part of the sample were disproportionately unmarried couples with high unemployment rates (a rate of 60% in a community for which the average was 5%), who were likely to have had past incidents of domestic assault and arrest as well as facing intervention by the police. They were also likely to have lower education levels and to belong to a minority race or a mixed race couple. The high proportion of Native Americans is the result of Minneapolis' proximity to many Indian reservations. The data from table G. 1 indicates that except for the high representation of Native Americans the sample is likely to be fairly representative of the kind of domestic assault cases that get reported to the police. Unfortunately it is likely that domestic assault cases among other groups (high income or higher educated) are not as widely reported, in which case the results of the experiment should not be extrapolated to such groups.

The experiment was designed to analyze the effect of arrest, advice, and separation on repeat incidance of domestic assault. Officers were given a pad of report forms with each form color coded for the different treatments. In order to ensure random assignment of treatments the forms were numbered and arranged in random order for each officer. Once the officers had dealt with a particular case they made a brief report and gave it to the research staff for follow-up. The research staff then followed up on the cases by detailed face to face interviews followed by telephone follow up interviews every two weeks for twenty-four weeks. Criminal justice reports mentioning the suspects name during the six month follow up period were also obtained. Recidivism was measured as repeated domestic violence against the same victim. The data from Official reports is given in table G.2.

For the outcomes using victim interviews, recidivism was measured as cases in which the victim reported new violence during follow-up interviews. Due to sampling attrition the recidivism outcomes from victim interviews could be obtained on just 161 of the 312 cases. Without strong assumptions regarding the missing data due to sampling attrition, this source of data does not provide much information. Nevertheless the outcome data from Victim Interviews is given in table G.3.

3.4. Analysis

The notation used throughout the paper is the following: t represents the treatment and equals 1 for arrest, 2 for advice and 3 for separation, z is the received treatment which equals 1 for arrest, 2 for advice and 3 for separation, r is the assigned treatment which equals 1 for arrest, 2 for advice and 3 for separation and $y(t)$ is the outcome which equals 1 if the offender commits another act of violence (recidivates) against the same victim and 0 if the offender does not recidivate. Finally m is a binary variable which is 1 if the outcome from victim interviews is observed and 0 if the outcome from victim interviews is not observed due to sampling attrition.

The probability that the offender recidivates under treatment 1 (arrest) $P(y(1) = 1)$ is given by

$P(y(1) = 1) = \sum_{i=1}^{3} P(y(1) = 1|z = i)P(z = i)$

where the equality follows from the law of total probability. In the above formulation $P(y(1) = 1|z = 1)$ is the probability that the offender would recidivate if he were given treatment 1 (arrest) given that the treatment he received is also 1; this can be observed from the data. $P(y(1) = 1|z = 2)$ and $P(y(1) = 1|z = 3)$ are the counterfactual probabilities; these are the probabilities that the offender would recidivate if he were given treatment 1 given that he is actually given either treatment 2 (advice) or 3 (separate). In the absence of identifying assumptions we do not know these counterfactual probabilities.

$P(z = 1), P(z = 2)$ and $P(z = 3)$ are the probabilities that the received treatment is $1, 2$ and 3 with all three known from the experiment.

Similarly the probability that the offender recidivates under treatment 2 (advice) is given by

$P(y(2) = 1) = \sum_{i=1}^{3} P(y(2) = 1|z = i)P(z = i)$

where now $P(y(2) = 1|z = 1)$ and $P(y(2) = 1|z = 3)$ or the probabilities that the offender would recidivate if he were given treatment 2 when he is actually given treatments 1 and 3 are the unknown counterfactual probabilities. The probability that the offender would recidivate if he were advised or given treatment 2 when the received treatment is also advice or treatment 2 ($P(y(2) = 1|z = 2)$) is known from the experiment as are the probabilities of received treatment.

The probability that the offender recidivates after he is separated is given by

$P(y(3) = 1) = \sum_{i=1}^{3} P(y(3) = 1|z = i)P(z = i)$

where $P(y(3) = 1|z = 1)$ and $P(y(3) = 1|z = 2)$ or the probabilities that the offender would recidivate if he were given treatment 3 when he is actually given treatments 1 or 2 are the unknown counterfactual probabilities. The probability that the offender would recidivate if he were separated from the victim for a short time or given treatment 3 when the received treatment is also separation or treatment 3 ($P(y(3) = 1|z = 3)$) is known from the experiment as are the probabilities of received treatment.

In the absence of any assumptions regarding the counterfactual probabilities I can obtain an interval of values or bounds for the recidivism probabilities by setting the unknown counterfactual probabilities equal to zero (giving me a lower bound) and one (giving me an upper bound). Doing so gives the following bounds for recidivism from treatment $i = 1, 2, 3$

$P(y(i) = 1|z = i)P(z = i) \leq P(y(i) = 1) \leq P(y(i) = 1|z = i)P(z = i) + \sum_{k \neq i} P(z = k)$

The no assumption bounds on recidivism probabilities were introduced in Manski (1989, 1990) and developed further in Manski (1994). Using the experimental data these bounds are $P(y(1) = 1) \in [0.06, 0.63]$, $P(y(2) = 1) \in [0.05, 0.77]$, and $P(y(3) = 1) \in [0.07, 0.79]$. Given the bounds on treatment effects, arrest is not a better treatment than either advice or seperation in the absence of any assumptions at all regarding counterfactual probabilities since the upper bound on recidivism from arrest is greater than the lower bounds on recidivism from either advice or from separation.

In the experiment, received treatments sometimes differed from assigned treatments provided sufficient reasons to deviate from assigned treatment were documented by the officers. Data on the deviations that occurred during the experiment is available. Given this data the recidivism probabilities under the assumption that the assigned treatments were random may be found. It should be noted that assigned treatments need not be random if cases of domestic assault were made systematically ineligible for the study, another criticism which has been made of the experiment. The next section looks in greater detail at how the results would change if the assumption of self-selection in treatment assignment is made.

Assuming randomly assigned treatments, the recidivism probability from arrest is given by

$P(y(1) = 1) = P(y(1) = 1|r = 1) = \sum_{i=1}^{3} P(y(1) = 1|r = 1, z = i)P(z = i|r = 1)$

Similarly the recidivism probability from advice is given by

$P(y(2) = 1) = P(y(2) = 1|r = 2) = \sum_{i=1}^{3} P(y(2) = 1|r = 2, z = i)P(z = i|r = 2)$

and the recidivism probability from separation by

$P(y(3) = 1) = P(y(3) = 1|r = 3) = \sum_{i=1}^{3} P(y(3) = 1|r = 3, z = i)P(z = i|r = 3)$

The bounds on recidivism probabilities for treatments $i = 1, 2, 3$ may be obtained by setting the unknown probabilities to their maximum and minimum possible values, which gives

$P(y(i) = 1|r = i, z = i)P(z = i|r = i) \leq P(y(i) = 1)$

$\leq P(y(i) = 1|r = i, z = i)P(z = i|r = i) + \sum_{k \neq i} P(z = k|r = i)$

Using data from the experiment gives the bounds on recidivism probabilities as $P(y(1) = 1) \in [0.11, 0.12]$, $P(y(2) = 1) \in [0.14, 0.36]$ and $P(y(3) = 1) \in [0.18, 0.45]$. From these bounds arrest is the best treatment since the upper bound on recidivism from arrest is less than the lower bounds on recidivism from either advice or separation. Therefore as long as the received treatment is random, experimental data indicates that arrest is the most effective treatment in reducing repeat incidence of domestic violance against the same victim.

The discussion till now does not consider sampling variation and focuses on the identification of the treatment effects. However the data from the experiment is taken from a sample of the overall population of interest. In order to take the sampling variation into account I construct the 90% confidance intervals around the bounds on the treatment effects using 1000 bootstrap replications. The resulting confidance intervals are given in table G.4. The confidance intervals are wide and the upper confidance interval on arrest is greater than the lower confidance interval on either advice or seperation, not just in case of the worst case bounds but also when the assigned treatments are random. Therefore taking into account sampling variation, arrest is no longer unambigously better than advice or separation as a treatment.

Table G.5 gives the bounds on the average treatment effects of using a mandatory arrest policy instead of a mandatory non-arrest policy under different assumptions. In the absence of distributional assumptions on the misssing counterfactual probabilities, we can estimate only bounds on the treatment effects and these bounds are reported in the table, together with 90% confidance intervals on the bounds. From the worst case bounds a mandatory arrest policy relative to either a mandatory advice policy or a mandatory separation policy may be substantially beneficial (with lower bounds -0.71 and -0.73) but it may also be substantially harmful (with upper bounds 0.58 and 0.56). However, looking at the bounds on the treatment effects when the treatment assignment is random and there is no assumption on treatment effects among compliars, a mandatory arrest policy is always beneficial relative to either a mandatory advice policy or a mandatory separation policy (the upper bounds are negative).

To look at treatment effects using outcome data from victim interviews, consider first the recidivism probability if all offenders of domestic violence were to be arrested or $P(y(1) = 1)$. This is given by

$$P(y(1) = 1) =$$

$$\sum_{i=1}^{3} \sum_{j=0}^{1} P(y(1) = 1|z = i, m = j)P(m = j|z = i)P(z = i)$$

The data from the experiment gives values for the recidivism probability of arrest given that the actual treatment was arrest and if recidivism was reported during follow-up interviews (or $P(y(1) = 1|z = 1, m = 1)$). The data from the experiment also gives the probability that recidivism was reported during follow-up interviews given the different treatments $P(m = 1|z)$ and the probability of actual treatment assignment $P(z)$. The data does not give the recidivism probability from arrest given that the actual treatment was arrest but recidivism was not reported due to sampling attrition or $P(y(1) = 1|z = 1, m = 0)$. The data also does not give the counterfactual event probabilities $P(y(1) = 1|z = 2)$ or $P(y(1) = 1|z = 3)$. So the best one can hope for is a bound on $P(y(1) = 1)$ in the absence of any other identifying assumptions. From the data obtained in the experiment the bounds on $P(y(1) = 1)$ are $[0.06, 0.78]$.

Now consider the recidivism probability if all offenders of domestic violence are advised or $P(y(2) = 1)$. This can be written as:

$P(y(2) = 1) =$

$\sum_{i=1}^{3} \sum_{j=0}^{1} P(y(2) = 1|z = i, m = j)P(m = j|z = i)P(z = i)$

The data from the experiment gives values for $P(y(2) = 1|z = 2, m = 1)$, $P(m = 0|z)$ and $P(z)$ but not $P(y(2) = 1|z = 2, m = 0)$ (due to sampling attrition) and $P(y(2) = 1|z = 1)$ or $P(y(2) = 1|z = 3)$ (counterfactual event probabilities). In the absence of any identifying assumptions on the unknown probabilities the bounds on $P(y(2) = 1)$ as given by the data obtained in the experiment are $[0.03, 0.89]$.

The third probability of interest is the recidivism probability from separation or $P(y(3) = 1)$. This is given by

$P(y(3) = 1) =$

$\sum_{i=1}^{3} \sum_{j=0}^{1} P(y(3) = 1|z = i, m = j)P(m = j|z = i)P(z = i)$

The data from the experiment gives values for $P(y(3) = 1|z = 3, m = 1)$, $P(m = 0|z)$ and $P(z)$ but not $P(y(3) = 1|z = 3, m = 0)$ (due to sampling attrition) and $P(y(3) =$

$1|z = 1)$ or $P(y(3) = 1|z = 2)$ (the counterfactual event probabilities). So the best one can hope for is a bound on $P(y(3) = 1)$ in the absence of any other identifying assumptions. From the data obtained in the experiment the bounds on $P(y(3) = 1)$ are $[0.04, 0.87]$.

As is the case with recidivism probabilities using official data the bounds on the recidivism probabilities for the different treatments have considerable overlap in the absence of any identifying assumptions when using the victim interview data, although the bounds using outcome data from victim interviews are wider than those using outcome data from official reports as a result of sampling attrition. From the victim interview data I cannot conclude that arrest results in lower recidivism as compared to other treatments in the absence of any identifying assumptions.

The victim interview data in table G.2 gives useful information about assigned treatments and deviations from assigned treatments. Suppose assigned treatments were random; in that case $P(y(i) = 1|r = i) = P(y(i) = 1)$ for $i = 1, 2$ and 3. So the recidivism probability from arrest $P(y(1) = 1)$ is given also by $P(y(1) = 1|r = 1)$ so that

$P(y(1) = 1) =$

$\sum_{i=1}^{3} \sum_{j=0}^{1} P(y(1) = 1|r = 1, z = i, m = j)P(m = j|r = 1, z = i)P(z = i|r = 1)$

Data from the experiment, as given in table G.2, gives values for the recidivism probability of arrest given that both assigned and received treatments are arrest and there is no attrition or $P(y(1) = 1|r = 1, z = 1, m = 1)$. The data also gives the probability that the received treatment is z given that the assigned treatment is arrest or $P(z|r = 1)$ and the probability that there is attrition given the received and assigned treatments or $P(m|z, r)$. The data does not give the recidivism probability of arrest given that the assigned treatment is arrest and the received treatment is advice or separation. These are

$P(y(1) = 1|r = 1, z = 2)$ and $P(y(1) = 1|r = 1, z = 3)$. It also does not give the recidivism probabilities when there is sampling attrition or $P(y(1) = 1|m = 0, z = 1, r = 1)$. Without making any assumptions about the unknown probabilities and using the data from the experiment, the bounds on $P(y(1) = 1)$ are given by $[0.10, 0.51]$.

Similarly the recidivism probability from advice is

$P(y(2) = 1) =$

$\sum_{i=1}^{3} \sum_{j=0}^{1} P(y(2) = 1|r = 2, z = i, m = j)P(m = j|r = 2, z = i)P(z = i|r = 2)$

Using the data from the experiment the bounds on $P(y(2) = 1)$ are $[0.09, 0.68]$.

The recidivism probability from separation is

$P(y(3) = 1) =$

$\sum_{i=1}^{3} \sum_{j=0}^{1} P(y(3) = 1|r = 3, z = i, m = j)P(m = j|r = 3, z = i)P(z = i|r = 3)$

Using the data from the experiment the bounds on $P(y(3) = 1)$ are $[0.10, 0.63]$.

Although the bounds in this case are an improvement over the ones obtained without making any assumptions, there is still considerable overlap between the recidivism from different treatments. Therefore using victim interview data and without making any assumptions on the missing data due to sampling attrition, I cannot conclude that arrest results in lower recidivism as compared to the other treatments.

Suppose I make the assumption of ignorable selection of missing data due to sampling attrition in addition to the assumption of random assignment of treatments. As before, the assumption of random assignment of treatments gives $P(y(i) = 1|r = i) = P(y(i) = 1)$ for $i = 1, 2$ and 3. The assumption of ignorable selection of missing data due to attrition gives, additionally,

$P(y(t) = 1|z, r, m = 0) = P(y(t) = 1|z, r, m = 1)$

which holds for all values of t, z and r.

Together the assumptions give the following expression for recidivism probability from arrest

$P(y(1) = 1) = \sum_{i=1}^{3} P(y(1) = 1|r = 1, z = i, m = 1)P(z = i|r = 1)$

The data from the experiment gives $P(y(1) = 1|r = 1, z = 1, m = 1)$ and $P(z|r = 1)$ but not $P(y(1) = 1|r = 1, z = 2, m = 1)$ or $P(y(1) = 1|r = 1, z = 3, m = 1)$ which are the counterfactual probabilities. Using data from the experiment, bounds on the recidivism probability from arrest are given by $[0.16, 0.18]$.

Similarly the recidivism probability from advice is

$P(y(2) = 1) = \sum_{i=1}^{3} P(y(2) = 1|r = 2, z = i, m = 1)P(z = i|r = 2)$

The data from the experiment gives $P(y(2) = 1|r = 2, z = 2, m = 1)$ and $P(z|r = 2)$ but not $P(y(2) = 1|r = 2, z = 1, m = 1)$ or $P(y(2) = 1|r = 2, z = 3, m = 1)$ which are the counterfactual probabilities. Using data from the experiment, bounds on the recidivism probability from advice are given by $[0.18, 0.39]$.

The recidivism probability from separation is

$P(y(3) = 1) = \sum_{i=1}^{3} P(y(3) = 1|r = 3, z = i, m = 1)P(z = i|r = 3)$

The data from the experiment gives $P(y(3) = 1|r = 3, z = 3, m = 1)$ and $P(z|r = 3)$ but not $P(y(3) = 1|r = 3, z = 1, m = 1)$ or $P(y(3) = 1|r = 3, z = 2, m = 1)$ which are the counterfactual probabilities. Using data from the experiment, bounds on the recidivism probability from separation are given by $[0.15, 0.42]$.

The bounds on recidivism when using victim interview data tell us that arrest results in lower recidivism as compared to advice but not when compared to separation. This is because there is overlap in recidivism probabilities from arrest and separation in spite of making very strong assumptions about the missing data from sampling attrition.

Table G.5 gives the 90% confidance interval on the bounds on treatment effects obtained by using the victim interview data. The confidance intervals are constructed using 1000 bootstrap replications. They show that if sampling variability is also incorporated into the analysis then arrest is no longer an unambigously better treatment than either advice or separation. This is the case even when the strongest assumptions regrading treatment assignment (randomly assigned treatment) and missing data due to sampling attrition (ignorable selection) are made. The upper confidance interval on recidivism from arrest is higher than the lower confidance interval on recidivism from either advice or separation.

Table G.7 gives the bounds on the average treatment effects of using a mandatory arrest policy instead of a mandatory non-arrest policy under different assumptions when victim interview data is used. Bounds on the treatment effects are reported in the table, together with 90% confidance intervals on the bounds. From the worst case bounds a mandatory arrest policy relative to either a mandatory advice policy or a mandatory separation policy may be substantially beneficial (with lower bounds -0.83 and -0.81) but it may also be substantially harmful (with upper bounds 0.75 and 0.74). Looking at the bounds on the treatment effects when the treatment assignment is random and there is no assumption on treatment effects among subjects for whom victim interviews could not be obtained, a mandatory arrest policy relative to either a mandatory advice policy or a mandatory separation policy may again be substantially beneficial (with lower bounds -0.58 and -0.53) but it may also be substantially harmful (with upper bounds

0.42 and 0.41). When treatment assignment is random and in addition the missing data due to sample attrition is ignorable, then using a mandatory arrest policy instead of a mandatory advice policy or a mandatory separation policy may be substantially beneficial (the lower bound is -0.23 and -0.26) and can at most be marginally harmful (the upper bound is close to zero and 0.03).

3.5. Self Selection: Skimming and Outcome Optimization

Suppose that instead of following the randomization process as dictated by the field experiment the officers carrying out the randomization systematically made cases ineligible for the study if they believed the randomly assigned treatment was not the correct treatment for a particular case. This was also among the criticisms faced by the study since it would affect the internal validity of the experiment and the interpretation of the results. Suppose that this criticism is correct. Can the treatment effects of arrest, advice and separation be improved from the no assumptions case? In this section I show that they can, provided certain assumptions are made on how the received treatments were selected.

If the process of self selection is known, a specific model of self selection may be used to get identifying restrictions which would tighten the bounds on recidivism probabilities compared to the no assumptions case. I use two different models of self selection which are behaviorally motivated and taken from Manski and Nagin (1998). These are the 'Skimming' and the 'Outcome Optimization' models used originally to determine effects on recidivism of different sentencing options used by judges in the juvenile justice system.

They may equally well be applied to the present case of a randomized experiment where there is a possibility of self selection.

Consider first the Skimming model of self selection. This model assumes that when officers self select they will arrest all high risk offenders (those who have a high probability of recidivism under all treatments) and not arrest the low risk offenders (those who have a low probability of recidivism under all treatments). Suppose the information available to the officers when making the decision of which treatment to assign is s (this would be the information that the officers get from the dispatch call). The officers know the recidivism probabilities for different treatments based on their information or they know $P(y(1) = 1|s)$, $P(y(2) = 1|s)$and $P(y(3) = 1|s)$. Type A offenders are high risk (they have a high probability of recidivism under all treatments) and Type B offenders are low risk (they have a low probability of recidivism under all treatments). There exist thresholds Π_1, Π_2 and Π_3 such that the Type A offenders are characterized by

$P(y(i) = 1|s) \geq \Pi_i, i = 1, 2, 3$

and Type B offenders are characterized by

$P(y(i) = 1|s) < \Pi_i, i = 1, 2, 3$

According to the Skimming model the officers assign treatment 1 to high risk offenders (Type A) and treatments 2 and 3 to the low risk offenders (Type B). Therefore

$P(y(i) = 1|s) \geq \Pi_i, i = 1, 2, 3 \Rightarrow z = 1$

$P(y(i) = 1|s) < \Pi_i, i = 1, 2, 3 \Rightarrow z = 2$ or $z = 3$

If self selection occurs according to the Skimming model then on observing treatment 1 (arrest) we know that the offender was high risk (Type A) or $P(y(1) = 1|s) > \Pi_1$, $P(y(2) = 1|s) > \Pi_2$ and $P(y(3) = 1|s) > \Pi_3$. On observing either treatments 2 or 3 (advice or separation) we know that the offender was low risk (Type B) or $P(y(1)=1|s) < \Pi_1$, $P(y(2)=1|s) < \Pi_2$ and $P(y(3)=1|s) < \Pi_3$. Therefore

$z = 1 => P(y(i) = 1|s) \geq \Pi_i, i = 1, 2, 3$

and

$z = 2$ or $z = 3 => P(y(i) = 1|s) < \Pi_i, i = 1, 2, 3$

The treatment z is a function of s so the above equations are equivalent to

$P(y(i) = 1|s, z = 1) \geq \Pi_i, i = 1, 2, 3$

and

$\sum_{j=1,2,3,j\neq i} P(y(i) = 1|s, z = j) < \Pi_i, i = 1, 2, 3$

From the law of iterated expectations

$P(y(i) = 1|z = 1) \geq \Pi_i, i = 1, 2, 3$

and

$\sum_{j=1,2,3,j\neq i} P(y(i) = 1|z = j) < \Pi_i, i = 1, 2, 3$

These inequalities can be rewritten as

$P(y(i) = 1|z = 1) > \sum_{j=1,2,3,j\neq i} P(y(i) = 1|z = j), i = 1, 2, 3$

Using the above inequalities allows me to tighten bounds on recidivism probabilities as compared to the no assumptions bounds. Consider first the recidivism probability of arrest $P(y(1) = 1)$. The lower bound is unchanged from the no assumptions bounds but consider the upper bound which can be tightened as follows

$P(y(1) = 1) = \sum_{i=1}^{3} P(y(1) = 1|z = i)P(z = i)$

$< P(y(1) = 1|z = 1) - P(y(1) = 1|z = 2)P(z = 3) - P(y(1) = 1|z = 3)P(z = 2)$

$\leq P(y(1) = 1|z = 1)$

The bounds on $P(y(1) = 1)$ are therefore

$P(y(1)=1|z=1)P(z=1) \leq P(y(1)=1) \leq P(y(1)=1|z=1)$

Now consider the recidivism probability of advice $P(y(2)=1)$. The upper bound remains unchanged from the no assumptions bounds but the lower bound is now given as follows

$P(y(2)=1) = \sum_{i=1}^{3} P(y(2)=1|z=i)P(z=i)$

$> P(y(2)=1|z=2)[P(z=1)+P(z=2)] + P(y(2)=1|z=3)[1-P(z=2)]$

$\geq P(y(2)=1|z=2)[P(z=1)+P(z=2)]$

The bounds on $P(y(2)=1)$ are therefore

$P(y(2)=1|z=2)[P(z=2)+P(z=1)] \leq P(y(2)=1)$

$\leq P(y(2)=1|z=2)P(z=2)+P(z=1)+P(z=3)$

The bounds on recidivism probability of separation can be found in the same way as those from advice to get the following

$P(y(3)=1|z=3)[P(z=3)+P(z=1)] \leq P(y(3)=1)$

$\leq P(y(3)=1|z=3)P(z=3)+P(z=1)+P(z=2)$

From the experimental data the bounds are $P(y(1)=1) \in [0.06, 0.13]$, $P(y(2)=1) \in [0.13, 0.77]$, and $P(y(3)=1) \in [0.18, 0.79]$. In this case there is no overlap between the recidivism due to arrest and that due to separation. Arrest unambiguously results in lower recidivism as compared to separation, even if there is self selection provided that self selection is such that high risk offenders are always arrested and low risk offenders are not. However note that there is still some overlap between recidivism from arrest and recidivism from advice. This means that arrest is not better at lowering recidivism as compared to advice if there is self selection such that all high risk offenders are arrested and low risk offenders are not.

Another way in which the officers may assign treatments is according to the Outcome Optimization model of treatment selection. This model makes the assumption that the officers, knowing the recidivism probabilities of different treatments, select the treatment with the lower probability of recidivism. Suppose as with the Skimming model the information the officers have when making their decision is s (this would be the information available from the dispatch call). Then the Outcome Optimization model says that if I observe treatment 1 then the recidivism probability from treatment 1 is less than that from treatment 2 or 3, if I observe treatment 2 then the recidivism probability from treatment 2 is less than that from treatment 1 or 3 and if I observe treatment 3 then the recidivism probability from treatment 3 is less than that from either treatment 1 or treatment 2. This can be written as

$z = i => P(y(i) = 1|s) < P(y(j) = 1|s), \forall i = 1, 2, 3, \forall j = 1, 2, 3 \neq i$

The above can be written as

$P(y(i) = 1|s, z = i) < P(y(j) = 1|s, z = i), \forall i = 1, 2, 3, \forall j = 1, 2, 3 \neq i$

Using the Law of Iterated Expectations I can write the above as

$P(y(i) = 1|z = i) < P(y(j) = 1|z = i), \forall i = 1, 2, 3, \forall j = 1, 2, 3 \neq i$

Using the above inequalities I can increase the lower bounds on the recidivism probabilities as compared to the no assumption bounds. Consider for instance the probability of recidivism from arrest $P(y(1) = 1)$. Using the inequalities allows me to tighten the lower bounds as follows

$P(y(1) = 1) = \sum_{i=1}^{3} P(y(1) = 1|z = i)P(z = i)$

$> P(y(1) = 1|z = 1)[P(z = 1) + P(z = 2) + P(z = 3)]$

$= P(y(1) = 1|z = 1)$

Similar results hold for the recidivism probabilities from advice and separation so the new bounds on recidivism probabilities are the following for treatments $j = 1, 2, 3$:

$\sum_{i=1}^{3} P(y(i) = 1|z = i)P(z = i) < P(y(j) = 1) < P(y(j) = 1|z = j)P(z = j) + \sum_{k \neq j} P(z = k)$

The data from the experiment gives me the numerical bounds on the recidivism probabilities $P(y(1) = 1) \in [0.18, 0.63]$, $P(y(2) = 1) \in [0.18, 0.77]$, and $P(y(3) = 1) \in [0.18, 0.79]$. Given these bounds, if the officers assigned treatments such that the treatment with the lowest recidivism probability was assigned, then I cannot conclude whether any of the treatments is better than the others. This is because there is considerable overlap in the recidivism probabilities from different treatments.

If there is self-selection then the effectiveness of arrest as a treatment depends crucially on which model of self-selection is used. If the officers assign treatments according to the Skimming model, arresting all high risk offenders and not arresting low risk offenders, then the experimental data suggest that arrest is a better treatment than is separation but not is not unambigously better than advice. If the officers assign treatments according to the Outcome Optimization model, assigning treatment so as to minimize recidivism, then experimental data does not tell me whether arrest is better than either of the two treatments. The model which follows closely the actual decision-making process of police officers would be the best. Ofcourse this assumes that the officers actually did not follow the randomization procedures laid out in the field experiment. While this need not have been the case, I show that even if the criticism of self selection is correct arrest would still lead to lower recidivism than separation, provided the officers arrested all high risk offenders and did not arrest the low risk offenders.

Both the Skimming and Outcome Optimization models assume that the officers are concerned only with recidivism and that the officers correctly perceive how treatment affects recidivism. Neither of the assumptions may hold in the real world. The Skimming model would seem to fit the behavior of officers better; in the Skimming model the officers are assumed to increase the apparent effectiveness of the program or act according to the normative view of punishing the high risk offender and going easy on the low risk offender. On the other hand the Outcome Optimization model assumes that officers assign treatments in order to minimize recidivism probability. It is also possible that the officers assign treatments in a way which is different from either of these models; for instance if the officers discriminate by race and gender. However as long as the treatment assignment by the officers has some conformity with the models described there is reason to believe in the results. Sociologists may then use the results from the model that they believe are closer to the actual decision making of the police.

Table G.6 gives the 90% confidance interval on the bounds on treatment effects obtained under both the Skimming and Outcome Optimization models. The confidance interval is constructed using 1000 bootstrap replications. They show that if sampling variability is also incorporated into the analysis then arrest is no longer an unambigously better treatment than advice or separation under either the Skimming or the Outcome Optimization models.

Table G.9 gives the bounds on the average treatment effects of using a mandatory arrest policy instead of a mandatory non-arrest policy under different assumptions when official data is used and different identifying assumptions regarding self selection are made. Bounds on the treatment effects are reported in the table, together with 90% confidance intervals on the bounds. If self selection occurs and is according to the skimming model then a mandatory arrest policy is always beneficial (with upper bounds which are either very close to zero or negative) in comparison to a mandatory advice policy or a policy of mandatory separation. If self selection occurs and is according to the outcome optimization model then a policy of mandatory arrest in comparison to mandatory advice or mandatory separation may be substantially beneficial (with lower bounds 0.59 and 0.61) but may also be substantially harmful (with upper bounds 0.44).

Now consider the bounds on treatment effects when officers assign treatments according to either the Skimming or Outcome Optimization models and victim reported recidivism measures are used. If the Skimming model is used to model self selection, then without any assumptions on the missing data due to sampling attrition the bounds on the recidivism probability from arrest are given by

$P(y(1) = 1|z = 1, m = 1)P(m = 1|z = 1)P(z = 1) \leq P(y(1) = 1) \leq$

$P(y(1) = 1|z = 1, m = 1)P(m = 1|z = 1) + P(m = 0|z = 1)$

Similarly the bounds on the recidivism probability from advice are given by

$P(y(2) = 1|z = 2, m = 1)P(m = 1|z = 2)[P(z = 2) + P(z = 1)]$

$\leq P(y(2) = 1) \leq$

$[P(y(2) = 1|z = 2, m = 1)P(m = 1|z = 2) + P(m = 0|z = 2)]P(z = 2)+$

$P(z = 1) + P(z = 3)$

and the bound from the recidivism probability from separation by

$P(y(3) = 1|z = 3, m = 1)P(m = 1|z = 3)[P(z = 3) + P(z = 1)]$

$\leq P(y(3) = 1) \leq$

$[P(y(3) = 1|z = 3, m = 1)P(m = 1|z = 3) + P(m = 0|z = 3)]P(z = 3)+$

$P(z = 1) + P(z = 2)$

From the data the bounds are $P(y(1) = 1) \in [0.06, 0.48]$, $P(y(2) = 1) \in [0.08, 0.89]$ and $P(y(3) = 1) \in [0.10, 0.87]$. Due to considerable overlap between the recidivism probabilities from different treatments, if officers assign arrest to all high risk offenders and the other treatments to all low risk offenders and the victim interview data is used then I cannot conclude that any one treatment is better than the others.

Now consider the bounds on recidivism probabilities if officers assign treatments according to the Outcome Optimization Model. If no assumptions about the missing data due to sampling attrition are made then the bounds on the recidivism probability from arrest are given by

$\sum_{i=1}^{3} P(y(1) = 1|z = i, m = 1)P(m = 1|z = i)P(z = i) \leq P(y(1) = 1) \leq$

$[P(y(1) = 1|z = 1, m = 1)P(m = 1|z = 1) + P(m = 0|z = 1)]P(z = 1)+$

$P(z = 2) + P(z = 3)$

The bounds on the recidivism probability from advice are given by

$\sum_{i=1}^{3} P(y(2) = 1|z = i, m = 1)P(m = 1|z = i)P(z = i) \leq P(y(2) = 1) \leq$

$P(y(2) = 1|z = 2, m = 1)P(m = 1|z = 2) + P(m = 0|z = 2)]P(z = 2)+$

$P(z = 1) + P(z = 3)$

and the bounds on the recidivism probability from separation by

$\sum_{i=1}^{3} P(y(3) = 1|z = i, m = 1)P(m = 1|z = i)P(z = i) \leq P(y(3) = 1) \leq$

$P(y(3) = 1|z = 3, m = 1)P(m = 1|z = 3) + P(m = 0|z = 3)]P(z = 3)+$

$P(z = 1) + P(z = 2)$

From the victim data obtained through the experiment the bounds on the recidivism probabilities are $P(y(1) = 1) \in [0.13, 0.78]$, $P(y(2) = 1) \in [0.13, 0.89]$ and $P(y(3) = 1) \in [0.13, 0.87]$. Due to considerable overlap between the recidivism probabilities from different treatments, if officers assign the treatment with the lowest recidivism probability and the victim interview data is used then I cannot conclude that any one treatment is better than the others.

Does the victim interview data give stronger results when the assumption of ignorable selection on the missing data due to sampling attrition is made? Assume that in addition to the assumption of ignorable selection I also assume that officers assign treatments according to the Skimming model. Then the bounds on the recidivism probability from arrest are given by

$P(y(1) = 1|z = 1, m = 1)P(z = 1) \leq P(y(1) = 1) \leq P(y(1) = 1|z = 1, m = 1)$

The bounds on the recidivism probability from advice are given by

$P(y(2) = 1|z = 2, m = 1)[P(z = 2) + P(z = 1)] \leq P(y(2) = 1) \leq$

$P(y(2) = 1|z = 2, m = 1)P(z = 2) + P(z = 1) + P(z = 3)]$

and the bounds on the recidivism probability from separation by

$P(y(3) = 1|z = 3, m = 1)[P(z = 3) + P(z = 1)] \leq P(y(3) = 1) \leq$

$P(y(3) = 1|z = 3, m = 1)P(z = 3) + P(z = 1) + P(z = 2)]$

From the data these bounds are $P(y(1) = 1) \in [0.09, 0.21]$, $P(y(2) = 1) \in [0.16, 0.78]$ and $P(y(3) = 1) \in [0.16, 0.78]$. In this case there is again overlap between the recidivism probabilities due to different treatments. So using the victim interview data, arrest is not unambigously better than either of the other treatments if the officers assign treatments according to the Skimming model even if the assumption of ignorable selection is made on the missing data due to sampling attrition.

Now assume that in addition to the assumption of ignorable selection of missing data due to sampling attrition, I also assume that officers assign treatments according to the Outcome Optimization Model. Then the bounds on the recidivism probability from arrest are

$\sum_{i=1}^{3} P(y(i) = 1|z = i, m = 1)P(z = i) \leq P(y(1) = 1) \leq$

$P(y(1) = 1|z = 1, m = 1)P(z = 1) + \sum_{i=2,3} P(z = i)$

The bounds on the recidivism probability from advice are given by

$\sum_{i=1}^{3} P(y(i) = 1|z = i, m = 1)P(z = i) \leq P(y(2) = 1) \leq$

$P(y(2) = 1|z = 2, m = 1)P(z = 2) + \sum_{i=1,3} P(z = i)$

and the bounds on the recidivism probability from separation are given by

$\sum_{i=1}^{3} P(y(i) = 1|z = i, m = 1)P(z = i) \leq P(y(3) = 1) \leq$

$P(y(3) = 1|z = 3, m = 1)P(z = 3) + \sum_{i=1,2} P(z = i)$

From the data these bounds are $P(y(1) = 1) \in [0.22, 0.66]$, $P(y(2) = 1) \in [0.22, 0.78]$ and $P(y(3) = 1) \in [0.22, 0.78]$. In this case there is again overlap between the recidivism probabilities due to different treatments. So using the victim interview data, arrest is not unambigously better than either of the other treatments if the officers assign treatments according to the Outcome Optimization model if, in addition, the assumption of ignorable selection is made on the missing data due to sampling attrition.

Table G.10 gives the 90% confidance interval on the bounds on treatment effects obtained under both the Skimming and Outcome Optimization models when data from victim interviews is used. The confidance interval is constructed using 1000 bootstrap replications. They show that arrest does not result in lower recidivism in comparison to either advice or separation, either with or without the assumption of ignorable selection on the missing data due to sampling attrition under the Skimming and the Outcome Optimization models. The upper confidance interval on recidivism from arrest is higher than the lower confidance interval on recidivism from either advice or from separation.

Table G.11 gives the bounds on the average treatment effects of using a mandatory arrest policy instead of a mandatory non-arrest policy under different assumptions when victim interview data is used and different identifying assumptions regarding self selection are made. Bounds on the treatment effects are reported in the table, together with 90% confidance intervals on the bounds. If self selection occurs according to the skimming model and the missing data due to sampling attrition is ignorable then a mandatory arrest policy may be substantially beneficial (with lower bounds -0.69) and at most marginally harmful (with upper bounds 0.05) in comparison to either a mandatory advice policy or a mandatory separation policy. If self selection occurs according to the outcome optimization model and the missing data due to sampling attrition is ignorable then a policy of mandatory arrest in comparison to mandatory advice or mandatory separation may be substantially beneficial (with lower bounds 0.56) but may also be substantially harmful (with upper bounds 0.44).

3.6. Conclusion

Prior to the 1970s the police enjoyed considerable discretion in treatment assignment for cases of minor domestic violence. During the 1970s there were increasing concerns that non-arrest did not provide enough deterrence to offenders of minor domestic violence from moving on to battery. Feminist groups were particularly vociferous in speaking against non-arrest treatment assignment by the police force which was predominantly male and often accused of sympathizing with the offender rather than the victim. It was in 1983 that the first randomized experiment that looked at the treatment effects of arrest and non-arrest on domestic assault was undertaken with the co-operation of the Minneapolis police department. Treatment assignment was done by the officers assigning treatments according to a pad of color-coded report forms. Not only was there non-compliance with the assigned treatment but there was also the possibility that the assigned treatments were non-random. Parametric analysis of the data suggested that arrest was more effective as compared to the other treatments. The results were very well-publicized and resulted in many states requiring mandatory arrest.

Non-parametric analysis of the data from the Minneapolis experiment also suggests that mandatory arrest is the most effective in reducing recidivism although the conclusions can be more clearly seen as resting on certain assumptions. Arrest has lower recidivism probability than advice or separation when using the recidivism measures from Official Data, provided there assigned treatment is random. If sampling variation is also taken into account, then arrest is no longer unambigously better than either advice or separation as a treatment even in this case. If the assigned treatment is non-random, then the effectiveness of arrest as a treatment depends on how the officers carry out treatment assignment: if officers arrest all high risk offenders then arrest is still associated with lower recidivism than advice or separation but not if officers assign treatments with the lowest recidivism. The data from victim reports does not suggest that arrest is associated with lower recidivism in comparison to either advice or separation, even under the assumption of ignorable selection for the missing data due to sampling attrition. If sampling variability is incorporated into the analysis, arrest is not associated with lower recidivism than either advice or separation under any identifying assumptions.

Replications of the Minneapolis experiment suggested that a mandatory arrest policy might not be the best under all circumstances. For instance, it was found that mandatory arrest increased domestic violence among the unemployed but that no-arrest increased domestic violence among the employed. In fact Sherman (1992) states the following 'Jurisdictions with large populations living in concentrated ghetto poverty areas should strongly consider repealing a mandatory arrest policy.' As in much of policy analysis there does not seem to be any simple solution.

CHAPTER 4

Racial Differences in Wages and Non-Wage Compensation (joint with Wallace Mok)

4.1. Introduction

Racial gaps in the labor market have persisted, and in some cases, have even increased recently (for examples see Altonji and Blank (1999) and Neal (2004)). The racial differences in wages are fairly well documented; however, there is less work that examines racial differences in non-wage compensation such as employer-provided health insurance and pension coverage. Thus, this paper asks several questions: What are the racial differences in health insurance and pension coverage for men and women? What component of the racial difference in non-wage compensation can be explained as the result of racial differences? What are the racial differences in total compensation, and how do these differ from racial differences in wages?

Very few works in the labor economics literature examine the role of non-wage compensation. Even and Macpherson (1994) look at gender differences in pension coverage. Solberg and Laughlin (1995) find that inclusion of fringe benefits reduces the gender wage gap. Our work is similar in spirit to these studies; we look at racial differences in health insurance and in pension coverage for men and women using data from both the Current Population Survey (CPS) and the National Longitudinal Survey of Youth (NLSY). In the data we find that white men have significantly greater health insurance coverage from their employers and greater pension coverage than do black men. Differences in characteristics favor greater health insurance and pension coverage for black men in the CPS. Therefore, the unexplained racial differences in health insurance and pension coverage are even larger than the observed differences. However, once we control for racial differences in ability (using AFQT test scores) in the NLSY data, much of the unexplained racial differences for men disappear. Unexplained differences in non-wage compensation that continue to favor white men could be an indication of discrimination in provision of non-wage benefits to black men; however, these could also be the result of racial differences in preferences.

For women, we find that white women do not always have higher health insurance and pension coverage than that of black women. Black women have greater coverage of employer-provided health insurance than that of white women, using CPS data. Racial differences due to differences in characteristics always favor greater coverage for black women in the CPS data. However, once we control for racial differences in AFQT scores in the NLSY, we find that racial differences in characteristics always favor greater coverage for white women. The unexplained differences in non-wage compensation favor black women; this is suggestive of reverse discrimination in favor of black women for provision of non-wage benefits (possibly due to affirmative action in jobs that are more likely to provide non-wage benefits).

We estimate total compensation by including the value of wages, health insurance and pension coverage by use of imputations. Using the CPS data, we find that the racial difference in total compensation is sometimes higher and sometimes lower than the racial difference in wages for men. For women, racial differences in total compensation are always smaller than the racial differences in wages. Racial differences in total compensation are always higher than racial differences in wages in the NLSY79 cohort. They are also always higher for men in a sub-sample of CPS, which has the same age range as the NLSY79 cohort. Our main conclusion from examining measures of total compensation is that racial differences in wages are not very different from racial differences in total compensation, with the percentage differences being less than or around two percent. This result is somewhat surprising given our earlier finding of large racial differences in health insurance and pension coverage, particularly for men.

The paper is organized as follows: We begin with a discussion of the data that was used in the paper, which describes the unconditional racial differences in health insurance, pension coverage, and median wages for different cohorts and sub-samples in the data. Next, we decompose the racial differences in health insurance and pension coverage into two components: one that may be explained by differences in characteristics, and one that cannot be explained by a difference in characteristics across racial groups. We then examine how total compensation differs between blacks and whites when the value of non-wage compensations is included. The last section concludes.

4.2. Data

We use two different datasets in our analysis. First we use the Current Population Survey (CPS) dataset across 1996 to 2006 cohorts to examine non-wage compensation and wage differences by race for men and for women. Second we use the National Longitudinal Survey of Youth (NLSY) dataset which allows us to control for differences in ability (AFQT score) across the different racial groups.

The CPS is a monthly survey of about 60,000 nationally representative households (the sample increased to about 100,000 after 2001). During March, a further supplement questionnaire was administered. This supplement, known as the Annual Social and Economic Supplement (ASEC),[1] provides additional data on work experience, income for the previous year, non-cash benefits received, and employment situations. We use the 1996-2006 ASEC in this study.

Our analysis of non-wage compensation makes use of the health insurance coverage information available in the CPS March supplements. Information about whether an individual is covered by employer-provided health insurance, how the plan was paid (in part or full), and how much the contribution the employer made, is available fairly consistently in the CPS March Supplements. Beginning the 1996 survey, a more detailed set of health insurance questions is administered, these questions further inquire the type of plan (self or family), whether the individual is covered due to dependency (e.g. as a spouse) or as the policyholder. Note that these questions do not address the issue of "Take-Up" - an individual may be offered employer-provided health insurance but elects to purchase it privately, and thus he would answered that he is not covered by employer-provided health insurance.[2]

[1]The ASEC was called Annual Demographic Supplement (ADF) prior to 2003.

[2]We also use the Survey of Income Program Participation (SIPP) to investigate take-up of health insurance by blacks and whites. The results suggest that the take-up rates of black married women and single black men are very similar to their white counterparts. Married white men and single white women are more likely to take up employer-provided health insurance than their black counterparts. Individuals are also asked the reasons for not taking up. Results suggest that among the married black men and single black women who refuse employer provided health insurance, most of them (about 70%) did so because that they had other health insurance plans. These results suggest that the take-up problem may be of minor concern.

The pension recipiency indicator that we use in our analysis comes from two questions asked of all CPS interviewees: 1) Other than social security did the employer or union that the interviewee worked for in (the previous year) have a pension or other type of retirement plan for any of the employees? 2) Was he/she included in that plan? We treat a person as covered by employer-provided pension if he gives an affirmative answer to both questions above.

We also use the 1979 cohort of the National Longitudinal Survey of Youth (NLSY79). The questions in the NLSY regarding health insurance and pension are phrased as follows: Did the employer MAKE AVAILABLE to you (type of benefit)? The NLSY79 is a panel study of a sample of 12686[3] young men and women who were 14-22 years old when they were first interviewed in 1979.[4] Since then, they have been re-interviewed yearly from 1979 to 1994, and bi-annually since 1996. The NLSY79 documents each respondent's experience, mainly with the labor market- such as labor market attachment, training and education. A particularly attractive feature of using the NLSY79 to analyze wage differentials is that it provides a proxy for the individual's ability – the Armed Forces Qualification Test (AFQT) score. In 1980, over 90% of the NLSY79 respondents were given a set of 10 tests from the Armed Services Vocational Aptitude Battery (ASVAB)[5] and a subset of 4 of these tests constitutes the AFQT. The AFQT is used by the military

[3]The sample size diminished over time due to funding limitation. Sample size of the survey dropped from 12686 respondents in 1979 to 10436 respondents in 1990.

[4]The NLSY79 is formally constituted by 3 subsamples: 1) A cross sectional sample of 6111 young people residing in US in 1979. 2) A supplemental sample of 5295 young people. This sub-sample is designed to over-sample hispanics, blacks and economically disadvantaged whites. 3) A sample of 1280 young people who were enlisted in one of the 4 branches of military as of 30th September 1978.

[5]Formally, the tests in the ASVAB consist of (1) general science, (2) arithmetic reasoning, (3) word knowledge, (4) paragraph comprehension, (5) numerical operations, (6) coding speed, (7) auto and shop information, (8) mathematical knowledge, (9) mechanical comprehension, and (10) electronics information

services to screen applicants and thereby assigning various jobs within the military. The use of the AFQT score as a measure of the individual ability has been fairly widespread in economics and sociology, most notably the controversial study by Herrnstein and Murray (1994).

We mainly focus the 1996-2004 periods, because many variables of interest, such as labor union status, employer pension provision, employer health insurance provision, are not available in the early waves of the survey. It is also important to point out that the results generated by the NLSY79 are not comparable with those of the CPS, because the sample in the NLSY79 is not nationally representative.

4.3. Racial Differences in Health Insurance and Pension Coverage

We examine two different fringe benefits that are provided to workers from their employers: health insurance and pension. Figures L.1 to L.4 give the fraction of white and black workers who get employer provided fringe benefits (health insurance and pension) across different cohorts from 1996 to 2006 when we use CPS data. Figures L.5 to L.10 give median wages for workers who do/do not get employer provided health insurance/pension coverage when we use CPS data. In addition supplementary tables K.1 to K.8 give the fractions of black and white workers covered by health insurance and pension, from CPS data and NLSY79 data.

From figures L.1 to L.4, higher fractions of white men receive health insurance than do black men. Also, higher fraction of white men receive a pension than do black men in all cohorts. For women, white women always get higher pension coverage than black women but higher fractions of black women get employer provided health insurance than do white women. The higher coverage of employer provided health insurance for black women is due to higher coverage for black married women. Black women who are unmarried have a lower fraction with employer provided health insurance than white women.

Figures L.5 to L.10 provide the variation in median weekly wages across jobs which provide health insurance/pension and those that do not across different cohorts of black and white workers. Median wages are higher for individuals receiving a pension, either black or white. In addition median wages are higher for whites than for blacks, both for jobs that give health insurance benefits and those that do not. The difference in median wages across black and white workers is larger for jobs that provide health insurance benefits for jobs that do not, for both men and women. Median wages are also always higher for jobs providing pensions than for jobs that do not. Whites always have a higher median wage than blacks in either jobs with pension or jobs without pension benefits. Racial gaps in median wages are higher for jobs which provide pension benefits than for jobs that do not provide pension benefits.

As mentioned before, we use the NLSY79 in addition to the CPS to examine racial differences in health insurance and pension coverage. Detailed means and standard errors for benefit coverage for each race, gender and year cohort from 1996 to 2006 are provided in: supplementary tables K.1 and K.2 for employer provided health insurance using CPS data, supplementary tables K.3 and K.4 for employer provided health insurance using NLSY79 data, supplementary tables K.5 and K.6 for pension coverage using CPS data and supplementary tables K.7 and K.8 for pension coverage using NLSY79 data. Overall trends are captured well in the figures just described. Examination of the sample averages reveals that, as in the CPS, white men have higher health insurance coverage and higher pension coverage than do black men. In both the CPS and NLSY79 racial differences in health insurance and pension coverage are significant for men. For women, the results are less clear. We find that health insurance coverage is higher for black women than for white women in the CPS, but that health insurance coverage is not always higher for black women than for white women in the NLSY79. The differences in health insurance coverage across black and white women in the NLSY79 are not significantly large, however. Pension coverage for white women is higher than pension coverage for black women in the CPS, but pension coverage for black women is higher than pension coverage for white women in the NLSY79. However, black women do not have significantly higher pension coverage than do white women in the NLSY79.

The racial gap in health insurance and pension coverage is positive and significant for men. For women, racial gaps in health insurance and pension coverage do not always favor white women. How are the racial gaps in health insurance and pension coverage affected when we control for racial differences in characteristics across black and white workers? The next sections answer this question.

4.4. Decomposing Racial Differences in Health Insurance and Pension Coverage

There are important differences in characteristics across black and white workers. How much of the difference in employer provided health insurance and pension can be explained by racial differences in characteristics asross black and white workers? Before answering this question we specify a model of who gets employer provided health insurance and pension and who does not.

More specifically we use the following models:

$$C_i^* = Z_i \delta_C + \epsilon_{C,i} \tag{4.1}$$

$$C_i = \begin{cases} 1 \text{ if } C_i^* > 0 \\ 0 \text{ if } C_i^* \leq 0 \end{cases} \tag{4.2}$$

where $C \in \{HI, P\}$ is a dummy variable taking the value one if individual i has non-wage compensation (health insurance HI or pension P) from the employer and the value zero if not, C_i^* is a latent variable that determines whether or not an individual gets non-wage compensation (health insurance or pension) and Z_i is a vector of characteristics that determines whether or not an individual gets health insurance and pension. We assume the error term $\epsilon_{C,i}$ is distributed normally so we carry out probit estimation of the model given in (1)-(2), separately for men and women.

As a next step we decompose the gap in non-wage compensation across black and white men (women) separately into a component which may be explained by a difference in covariates across black and white men (women) and a component which may not be explained by a difference in covariates across black and white men (women). For a linear regression the decomposition in the average value of the dependent variable C across black and white workers may be expressed as follows

$$\overline{C}^W - \overline{C}^B = [(\overline{Z}^W - \overline{Z}^B)\widehat{\delta_C}^W] + [\overline{Z}^B(\widehat{\delta_C}^W - \widehat{\delta_C}^B)]$$

This is also referred to as the Blinder-Oaxaca decomposition. $\overline{Z}^j$ is the row vector of average values of the independent variables and $\widehat{\delta_C}^j$ is the vector of coefficient estimates for benefit type C and race j. However the dependent variable in our case is a dummy variable denoting whether or not a worker has employer provided health insurance or pension coverage. Therefore we need to use a modification of the method that decomposes a non-linear equation, $C = F(Z\widehat{\delta_C})$. The issue arises since $\overline{C}$ does not necessarily equal $F(Z\widehat{\delta_C})$. Following the method proposed in Fairlie (2005), we use the following decomposition

$$(4.3)\ \ \overline{C}^W - \overline{C}^B = [\sum_{i=1}^{N^W} \frac{F(Z_i^W \widehat{\delta_C}^W)}{N^W} - \sum_{i=1}^{N^B} \frac{F(Z_i^B \widehat{\delta_C}^W)}{N^B}] + [\sum_{i=1}^{N^B} \frac{F(Z_i^B \widehat{\delta_C}^W)}{N^B} - \sum_{i=1}^{N^B} \frac{F(Z_i^B \widehat{\delta_C}^B)}{N^B}]$$

In both the Blinder-Oaxaca decomposition and the Fairlie decomposition, the first term represents the part of the racial gap which is due to group differences in distribution of Z while the second term represents the part of the gap which is due to differences in group processes determining the level of $C \in \{HI, P\}$. The interpretation of the second component which we use in the context of our paper is as the part of the racial gap in non-wage compensation which cannot be explained by the racial difference in characteristics which determine non-wage compensation.

The equations give the contribution of the overall racial difference in characteristics; it is also possible to carry out estimation of the contribution of racial differences in individual characteristics using Fairlie (2005) with the standard errors associated with these estimated by the delta method. The decomposition method involves a one to one matching between the black and white groups. Since there are fewer black workers than white workers, samples are drawn randomly from the white sample. Fifty different samples are drawn, racial differences estimated using each sample and the mean results from across the fifty different samples reported.

Table J.1 gives the estimation results when we run the probit regressions on whether or not an individual is covered by health insurance with a race dummy included in the set of Z_i variables in section A and when we carry out non-linear decompositions by running probit regressions separately by race in section B. Table J.1 uses CPS data for full time workers from 1996 to 2006 cohorts, which is combined together, with year dummies included in the probit regressions. In (I), estimation results are reported when we use education, age, region, children, and spouse salary only as the control variables. We use a set of dummy variables for education; whether the worker has no education, some high school education, high school education or college/grad school education. In addition we use four region dummies. In (II), estimation results are reported when a more full set of controls is included which includes union membership,[6] firm size, occupation, industry and work type. We use five dummy variables for employer's firm size (whether number of employees in the firm are less than 25, between 25 and 99, between 100 and 499, between 500 and 999 or greater than 1000), eight dummies for occupation and thirteen different dummy variables for industry. Inclusion of these controls reduces the coefficient and marginal effect associated with the black dummy so that it is significantly negative,

[6]Since 1983, questions on union/employment association membership are asked only to a quarter of the sample (the outgoing rotation groups) in each month (Hirsch and Macpherson, 2003). To obtain information of union membership for the remaining three quarters of the sample in each year, we make use their responses to the Basic CPS survey in the following months. Specifically, we look at their responses to the questions on union membership during their outgoing interviews. We also restrict to those who do not experience unemployment between the ASEC and their outgoing interview. Doing so essentially eliminate those who changed jobs during this period, which will contaminate our data (i.e. the employer that offers pension may not be the employer the interviewee worked for during the month when he answered the union membership questions).

for both the male and female sub-samples. In the decomposition results, inclusion of the latter set of control variables changes the component of the explained difference from positive to negative. In other words, racial differences in union membership, firm size, occupation, industry and work type favor health insurance coverage for black men and women. The unexplained differences in health insurance coverage across race, however, favor white men and women.

Tables J.2 and J.3 gives the estimation results when we run the probit regressions on whether or not an individual is covered by health insurance with a race dummy in section A and when we carry out non-linear decompositions in section B. Tables J.2 and J.3 uses NLSY79 data which allows us to add controls for tenure and for racial differences in AFQT scores. We use data for full time workers from 1996 to 2006, the data being combined with year dummies included in the probit regressions. In (I), estimation results are reported when we use education, age, region, children, and spouse salary only as the control variables. In (II) estimation results are reported with the addition of tenure to the set of control variables. In (III) estimation results are reported with the further addition of the AFQT test score, standardized by age. In (IV), estimation results are reported when we include controls for union membership, firm size, occupation, industry and work type but none for tenure or AFQT. In (V) we again add tenure to the set of controls from (IV) and in (VI) we also add standardized AFQT scores. Addition of tenure for men, with and without job characteristics included as controls, increases the coefficient and marginal effect associated with the black dummy, making it less negative. However, it continues to be significantly negative. With the addition of AFQT test scores it is still negative but no longer significantly so. Addition of tenure for women also increases the coefficient and marginal effect associated with race. However, for women the race dummy is not significant until we add AFQT scores as a control and it is significantly positive. In other words, controlling for racial differences in AFQT scores, we find that black women get significantly higher health insurance benefits. This can also be seen in the non-linear decompositions; inclusion of the AFQT score for men reduces the difference between rows (7) and (8), which is the difference in health insurance coverage which cannot be explained by differences in characteristics. Inclusion of the AFQT score for women increases row (8) in comparison to row (7), so the unexplained difference in health insurance coverage favors black women over white women.

Table J.4 gives the estimation results when we run the probit regressions on whether or not an individual gets pension coverage with a race dummy included in the set of regressors in section A and when we carry out non-linear decompositions by running probit regressions separately by race in section B. Table 3 uses CPS data for full time workers from 1996 to 2006 cohorts, which is combined together, with year dummies included in the probit regressions. In (I), estimation results are reported when we use education, age, region, children, and spouse salary only as the control variables. In (II), estimation results are reported when a more full set of controls is included which includes union membership, firm size, occupation, industry and work type. Inclusion of these controls reduces the coefficient and marginal effect associated with the black dummy so that it is significantly negative, for both the male and female sub-samples. In the decomposition results, inclusion of the latter set of control variables changes the component of the explained difference from positive to negative. In other words, racial differences in union membership, firm size, occupation, industry and work type favor health insurance coverage for black men and women. The unexplained differences in health insurance coverage across race, however, favor white men and women. These results are very similar to the results for health insurance coverage using CPS data.

Tables J.5 and J.6 gives the estimation results when we run the probit regressions on whether or not an individual gets pension coverage when we run probit regressions with a race dummy as given in section A and when we carry out non-linear decompositions as given in section B. Tables J.4 and J.5 use NLSY79 data which allows us to add controls for tenure and for racial differences in AFQT scores. We use data for full time workers from 1996 to 2006, the data being combined with year dummies included in the probit regressions. In (I), estimation results are reported when we use education, age, region, children, and spouse salary only as the control variables. In (II) estimation results are reported with the addition of tenure to the set of control variables. In (III) estimation results are reported with the further addition of the AFQT test score, standardized by age. In (IV), estimation results are reported when we include controls for union membership, firm size, occupation, industry and work type but none for tenure or AFQT. In (V) we again add tenure to the set of controls from (IV) and in (VI) we also add standardized AFQT scores. Addition of tenure for men and women increases the coefficient and marginal effect associated with race, making it less negative for men and more positive for women. However, the race dummy is not significant for men. For women it is significantly positive. Inclusion of AFQT scores makes the race dummy for men positive, but insignificant. For women, inclusion of AFQT scores makes the race dummy significantly positive. As for health insurance, once we control for racial differences in AFQT scores, we find that black women get significantly higher pension benefits. This can also be seen in the non-linear decompositions; inclusion of the AFQT score for men reduces the difference between rows (7) and (8), which is the difference in pension coverage which cannot be explained by differences in characteristics. Inclusion of the AFQT score for women increases row (8) in comparison to row (7), so the unexplained difference in pension coverage favors black women over white women.

4.5. Racial Differences in Wages and in Total Compensation

We define total compensation as the combined value of wages, health insurance and pension. While both the NLSY and the CPS ASEC ask about the magnitude of the wage the individual gets, the values of health insurance and pension are not asked. A plausible reason is that they are extremely difficult to measure from the perspective of the employee. The value of employer provided health insurance depends on the individual's health status, the nature of the plan, the coverage particulars. Similarly, the value of pension to the employee depends on the current and future interest rates, the individual's assessment of future inflation, the self assessed probability of death before retirement etc. From the perspective of the employer, however, the values of these non-wage compensation items are not simple to assess either. Provision of such non wage compensation may improve the productivity of employees and increase the retention rates, which are both beneficial to the employer, making the value of non-wage compensation depend on more than just the costs involved in providing them.

To abstract from the complexity in modeling the value of non wage compensation, we assume their values are just the direct costs to provide them. In the CPS ASEC, individuals who are covered by employer-provided health insurance are also asked about the amount of contribution of the employer. For pension, however, the CPS ASEC does not ask about the amount of employer's contribution. To estimate the amount paid by the employer, we use the Survey of Consumer Finances (SCF) which ask about the amount of employer's contribution towards the employee's pension (as a percent of the employee's wage).[7] We then apply these contribution rates to the CPS data and estimate the employer's contribution.

For NLSY, however, the issues involved in imputing values of non wage compensation schemes are more complicated. First, as we discussed previously, the NLSY asks whether the individual's employer makes a certain type of non-wage compensation available to him, rather than whether the individual is covered by such compensation. Second, for health insurance, we do not know the amount paid by the employer as well as the type of health insurance (such as whether it is a single or family plan). Thus we assume that in the NLSY, individuals who are offered non-wage compensation schemes always accept them. To impute the value of employer-provided pension, we again use the contribution rates estimated using the SCFs. For health insurance, we use the CPS ASEC to estimate the amount paid by the employer as a percentage of the employee's wages (separately for single and family plans).[8] We additionally assume that the interviewee chooses a single plan if he is single, and chooses a family plan if he is married.

[7]The SCF is a triennial cross-sectional survey containing detail data about the interviewees' income, assets and investment portfolios. We use the 1995, 1998, 2001, and 2004 SCF to estimate the average pension contribution rates (employers). We use the averages of the 1995 and 1998 rates, 1998 and 2001 rates, 2001 and 2004 rates as the rates in 1996-1997, 1999-2000, and 2002-2003 respectively.

[8]Employers pay more if the individual elects to have a family health insurance plan – for instance, in 1996, the average contribution rate for a single plan is about 7%, while it is 13% for those who choose a family plan.

We find and report in table J.7 the difference in mean hourly wages and in total compensation for full time working blacks and whites using CPS data and NLSY79 data.[9] The columns represent the male and female sub-samples from the CPS and NLSY79 datasets. Row (1) in section (A) gives the mean hourly wage for white workers, row (2) gives the mean hourly wage for black workers, row (3) gives the difference in mean hourly wages of white and black men and row(4) gives the difference as a percentage of the black hourly wage. Section B of the table gives the hourly total compensation for whites in row (5), for blacks in row(6), the difference in total compensation across white and black workers in row (7) and the difference as a percentage of the black hourly total compensation in row (8). We are interested in a comparison of rows (4) and (8) for each column; in words, how does the percentage difference in wage compare with the percentage difference in total compensation? We find the percentage differences to be essentially the same, particularly for the CPS dataset. For men, racial differences in total compensation are larger than racial differences in wages, for both the CPS and NLSY79 datasets. For women, racial differences in total compensation are larger than racial differences, for the NLSY79 but not for the CPS.

4.6. Conclusion

We find that white men have significantly higher employer-provided health insurance and pension coverage than do black men. Of the unexplained racial differences that favor white men, a large component disappears when we control for racial differences in AFQT scores. Possible reasons for the differences that persist may be possible discrimination

[9] Table J.7 aggregates the data from all cohorts of CPS (1996 to 2006) and of NLSY79. Supplementary tables K.9 to K.12 give the same results for each cohort separately. The results are essentially the same as those from aggregated data, as reported in this paragraph.

against black men in the provision of health insurance and pension, or due to racial differences in preferences.

On the other hand, white women have lower health insurance coverage than black women. The racial differences in characteristics favor black women when we do not control for AFQT scores, but favor white women when we control for AFQT scores. Unexplained differences in non-wage compensation favor black women, once we control for AFQT scores. An interesting research question is: Why do we find such results for women? The higher health insurance coverage for black women is the result of higher coverage for married black women. One possible explanation for this could be reverse discrimination in favor of black women, possibly due to affirmative action in jobs that provide health insurance.

When we examine racial differences in total compensation, we find them to be similar to racial differences in wages. This results is somewhat surprising since we find non-wage compensation is different for different racial groups. Possible reasons for this could be that there is substitution across wage and non-wage characteristics (compensating differentials), which are different for different racial groups.

Several questions remain unanswered. For instance, it is not clear what causes racial differences in non-wage compensation to be so different for men and women. In particular, why do married black women have such high rates of employer-provided health insurance coverage? Why do unexplained differences in both health insurance and pension benefits favor black women instead of white women? Future studies should look at a greater variety of non-wage compensation benefits provided to employees in addition to health insurance and pensions. A better understanding of the differences in non-wage compensation across racial groups is important in proper measurement of the extent of racial inequalities in labor markets.

References

[1] Aigner, Dennis J. and Glenn G. Cain (1977), 'Statistical Theories of Discrimination in Labor Markets', *Industrial and Labor Relations Review*, 30: 175-187.

[2] Akerlof, George (1976), 'The Economics of Caste and the Rat Race and Other Woeful Tales', *Quarterly Journal of Economics*, 90: 599-617.

[3] Altonji, Joseph G. and Rebecca Blank (1999), 'Race and Gender in the Labor Market', in: O. Ashenfelter and D. Card eds. Handbook of Labor Economics, Vol 3, Elsevier Science B.V.

[4] Anderson, Patricia M. and Bruce Meyer (1997), 'Unemployment Insurance Takeup Rates and the After-Tax Value of Benefits,' Quarterly Journal of Economics, CXII, 913-938

[5] Anwar, Shamena and Hanming Fang (2004), 'An Alternative Test of Racial Profiling in Motor Vehicle Searches: Theory and Evidence,' *American Economic Review*, 96 (1): 127-151.

[6] Ayres, Ian and Peter Siegelman (1995), 'Race and Gender Discrimination in Negotiating for the Purchase of a New Car', *American Economic Review*, 85 (3): 304-321.

[7] Banarjee, A. and R. Somanathan (2006), 'The Political Economy of Public Goods: Some Evidence from India,' Working Paper.

[8] Banerjee, B. and J. B. Knight (1985), 'Caste in Indian urban labor market,' *Journal of Development Economics*, 17: 277-307

[9] Becker, Gary S. (1961), The Economics of Discrimination, 2nd edition (The University of Chicago Press, Chicago, IL).

[10] Berk, Richard A. and Lawrence W. Sherman (1984) 'The Specific Deterrent Effects of Arrest for Domestic Assault." American Sociological Review 49:261-272.

[11] Berk, Richard A. and Lawrence W. Sherman (1988) 'Police Responses to Family Violence Incidents: An Analysis of an Experimental Design with Incomplete Randomization.' Journal of the American Statistical Association 83:70-76.

[12] Bertrand, Marrianne and Sendhil Mullainathan (2004), 'Are Emily and Greg more employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination.'

[13] Cain, Glenn G. (1986), 'The Economic Analysis of Labor Market Discrimination', in: O. Ashenfelter and R. Layard eds. Handbook of Labor Economics, Vol 1 (North Holland, Amsterdam) pp 693-785.

[14] Coate, Stephen and Glenn Loury (1993), 'Will Affirmative Action Policies eliminate negative Stereotypes?' *American Economic Review*, 83 (5): 1220-1240.

[15] Cook, Thomas D. and Donald T. Campbell (1979) Quasi Experimentation: Design and Analysis Issues for Field Settings.. Chicago: Rand McNally.

[16] Corcoran, Mary and Gregory Duncan (1979), 'Work History, Labor Force Attachment and Earnings Differences Between the Races and Sexes,' Journal of Human Resources, 14: 3-20

[17] Duncan, Gregory (1976), 'Earnings Functions and Nonpecuniary Benefits,' Journal of Human Resources, 11(4): 462-483

[18] Even, William E. and David A. Macpherson (1990), 'The Gender Gap in Pensions and Wages,' Review of Economics and Statistics, 72(2): 259-265

[19] Even, William E. and David A. Macpherson (1994), 'Gender Differences in Pensions,' Journal of Human Resources 29 (2): 555-587

[20] Fairlie, Robert (1999), 'The Absence of African-American Owned Business: An Analysis of the Dynamics of Self-Employment,' Journal of Labor Economics, 17(1): 80-108

[21] Fairlie, Robert (2005), 'An Extension of the Blinder-Oaxaca Decomposition Technique to Logit and Probit Models,' Journal of Economics and Social Measurement, 30(4): 305-316

[22] Firth, M. (1981), 'Racial Discrimination in the British Labor Market,' *Industrial and Labor Relations Review*, 34 (2): 265-272.

[23] Firth, M. (1982), 'Sex Discrimination in Job Opportunities for Women,' *Sex Roles* 8(8): 891-901.

[24] Fix, Michael and Raymond J. Struyck, eds. (1992), Clear and Convincing Evidence: Measurement of Discrimination in America (Urban Institute Press, Lanham, MD).

[25] Foster, V. and R. Vohra (1992), 'An Economic Argument for Affirmative Action,' *Rationality and Society* 4(2): 176-188.

[26] Frietas, Kripa (2007), 'The Indian Caste System as a Means of Contract Enforcement,' Working Paper, Northwestern University.

[27] Goldin, Claudia and Cecilia Rouse (2000), 'Orchestrating Impartiality: the impact of Blind Auditions on Female Musicians,' *The American Economic Review*, 90(4): 715-741.

[28] Government of India, National Sample Survey Organization. 'Employment and Unemployment Situation among Social Groups in India, 2004-05' Report No. 516

[29] Government of India, National Sample Survey Organization. 'Employment and Unemployment Situation in Cities and Towns in India, 2004-05' Report No. 520

[30] Heckman, James J. (1998) 'Detecting Discrimination', *The Journal of Economic Perspectives* 12(2): 101-116.

[31] Heckman, James (1979), 'Sample Selection Bias as a Specification Error,'Econometrica 47: 153-161

[32] Heckman, James J. and Seigelman, Peter. (1992) 'The Urban Institute Audit Studies: Their Methods and Findings.' in Micheal Fix and Raymond J. Struyk, eds. Clear and Convincing Evidence: Measurement of Discrimination in America (Urban Institute Press, Lanham, MD).

[33] Herrnstein, Richard, and Charles Murray (1994), 'The Bell curve: Intelligence and Class Structure in American Life,' New York: Free Press

[34] Hirsch, Barry T. and David A. Macpherson (2003), 'Union Membership and Coverage Database from the Current Population Survey: Note,' Industrial and Labor Relations Review, Vol. 56, No. 2, January 2003, pp. 349-54

[35] Horowitz, Joel and Charles F. Manski (1999) 'Nonparametric analysis of Randomized Experiments with Missing Covariate and Outcome data,' Journal of the American Statistical Association.

[36] Hu, Luojia and Christopher Taber (2007) 'Displacement, Asymmetric Information and Heterogenous Human Capital,' Working Paper, Northwestern Unievrsity.

[37] Ippolito, Richard (1985), 'The Labor Contract and True Economic Pension Liabilities,' American Economic Review, 75: 1031-1043

[38] Ippolito, Richard (1987), 'The Implicit Pension Contract: Developments and New Directions,' Journal of Human Resources, 22: 441-467

[39] Jowell, R. and P. Prescott-Clark (1970) 'Racial Discrimination and White Collar Workers in Britain' *Race* 11: 397-417.

[40] Kenney, G. M. and D. A. Wissoker (1994) 'An Analysis of the Correlates of Discrimination Facing Young Hispanic Job-Seekers,' *The American Economic Review* 84(3): 674-683.

[41] Kotlikoff, Laurence J. and David Wise, 'Labor Compensation and the Structure of Private Pension Plans: Evidence for Spot versus Contractual Labor Markets,' in David Wise (ed.) Pensions, Labor and Individual Choice (Chicago: University of Chicago Press, 1985)

[42] Lambert, Richard D. (1963), Workers, Factories and Social Change in India (Princeton University Press, Princeton, NJ)

[43] Lehman, E.L. (1986) Testing Statistical Hypotheses, 2nd edition (Wiley, New York).

[44] Lundberg, Shelley J. and Richard Startz (1983) 'Private Discrimination and Social Intervention in Competitive Labor Markets' *American Economic Review* 73: 340-347.

[45] Lundberg, Shelley J. and Richard Startz (1991) 'The Enforcement of Equal Opportunity Laws under Imperfect Enforcement: Affirmative Action and Alternatives', *Quarterly Journal of Economics* 106(1): 309-326.

[46] Lynch, Owen M. (1969) 'The politics of untouchability. social mobility and social change in a city of India (Columbia University Press, New York)

[47] Maddala, G. S. and Forest Nelson (1985), 'Switching Regression Models with Exogenous and Endogenous Switching,' Proceedings of the American Statistical Association, 423-426

[48] Manski, Charles F. (2006) 'Search Profiling with Partial Knowledge of Deterrence,' *Economic Journal*, Vol. 116, No. 515, pp. F385-F401

[49] Manski, Charles F. (2005) Social Choice with Partial Knowledge of Treatment Response. Princeton University Press.

[50] Manski, Charles F. (2003) Partial Identification of Probability Distributions. Springer-Verlag.

[51] Manski, Charles F. (1990) 'Non-parametric Bounds on Treatment Effects,' American Economic Review (Papers and Proceedings) 80: 319-323.

[52] Manski, C. and D. Nagin (1998) 'Bounding Disagreements about Treatment Effects: A Case Study of Sentencing and Recidivism,' Sociological Methodology, Vol. 28, pp. 99-137.

[53] Manski, Charles F., Gary Sandefur, Sara McLanahan, and Daniel Power (1992) 'Alternative Estimates of the Effect of Family Structure during Adolescence on High School Graduation.' Journal of the American Statistical Association 87:25-37.

[54] McIntosh, N. and D. Smith (1974) 'The Extent of Racial Discrimination' Political and Economic Planning Broadsheet, No.547 (London: Political and Economic Planning).

[55] Mitchell, Olivia (1982), 'Fringe Benefits and Labor Mobility,' Journal of Human Resources, 17: 286-298

[56] Moro, Andrea and Peter Norman (2003) 'Affirmative Action in a Competitive Economy' *Journal of Public Economics*, 87:3-4: 567-594.

[57] Moro, Andrea and Peter Norman (2004) 'A General Equilibrium Model of Statistical Discrimination' *Journal of Economic Theory*, 114(1): 1-30.

[58] Morris, David (1965) The emergence of an industrial labor force in India. A study of the Bombay Cotton mills, 1854-1947 (University of California Press, Berkeley/Los Angeles, CA).

[59] Munshi, Kaivan and Mark Rozensweig 'Traditional Institutions meet the Modern World: Caste, Gender and Schooling Choice in a Globalizing Economy,' *American Economic Review*, 96(4): 1225-1252.

[60] Neal, Derek (2004), 'The Measured Black-White Wage Gap among Women is too Small,' Journal of Political Economy, 112

[61] Neumark, David (1996), 'Sex Discrimination in Restaurant Hiring: An Audit Study' *Quarterly Journal of Economics*, 111(3): 915-942.

[62] Newman, Jerry M. (1980), 'Discrimination in Recruitment: An Empirical Analysis,' *Industrial and Labor Relations Review*, 32(1): 15-23.

[63] Oaxaca, Ronald, 'Male-Female Wage Differentials in Urban Labor Markets,' International Economic Review 14: 693-709

[64] Pager, Devah (2007), 'The Use of Field Experiments for Studies of Employment Discrimination: Contributions, Critiques and Directions for the Future,' Annals of the American Academy of Political and Social Science 609: 104-133.

[65] Pande, Rohini (2003), 'Can Mandated Political Representation increase Policy Influence for Disadvantaged Minorities? Theory and Evidence from India,' BREAD Working Paper No. 024.

[66] Phelps, E. (1972) 'The Economic Theory of Racism and Sexism' *American Economic Review*, 62: 659-661.

[67] Riach, P.A. and Rich, J. (1991) 'Testing for Racial Discrimination in the Labor Market,' *Cambridge Journal of Economics*, 15: 239-259.

[68] Riach, P.A. and Rich, J. (1995) 'An Investigation of Gender Discrimination in Labor Hiring,' *Eastern Economic Journal*, 21(3): 343-356.

[69] Schiller, Bradley R. and Randall D. Wise (1980), 'The Impact of Private Pensions on Firm Attachments,' 61: 369-380

[70] Sherman, Lawrence W. with Janell D. Schmidt and Dennis P. Rogan. (1992). Policing Domestic Violence. New York, NY: The Free Press

[71] Solberg, Eric and Teresa Laughlin (1995), 'The Gender Pay Gap, Fringe Benefits, and Occupational Crowding,' Industrial and Labor Relations Review, 48(4): 692-708

[72] Somanathan, Rohini (2006) 'The Assumptions and The Arithmetic of Caste-Based Reservations,' *Economic and Political Weekly*, 2436-2438.

[73] Weichselbaumer, Doris (2003) 'Sexual Orientation Discrimination in Hiring,' *Labor Economics*, 10: 629-642.

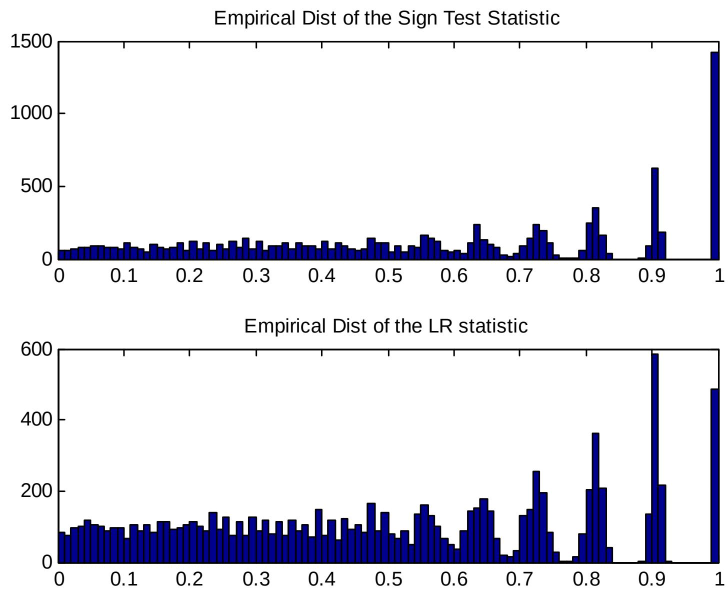
[74] Woodbury, Stephen A (1983), 'Substitution between Wage and Nonwage Benefits,' American Economic Review, 73(1): 166-182

APPENDIX A

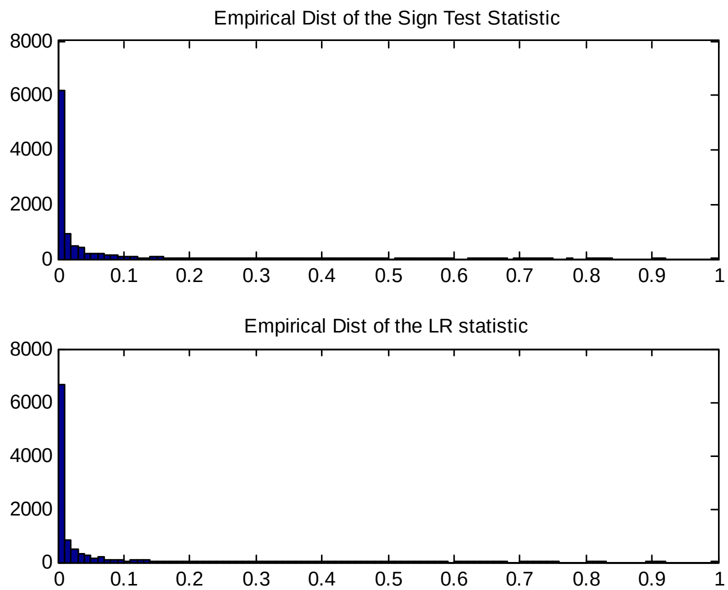
Chapter One- Symmetry Tests

In order to test symmetry between high and low-caste applicants two different tests were used. These were the likelihood ratio test and the conditional sign test. The likelihood ratio test uses the chi square statistic to test the null hypothesis of symmetry between high and low-caste applicants. The null hypothesis being tested is that the probability that a high-caste applicant is called back and a low-caste applicant is not is the same as the probability that a low-caste applicant is called back and a high-caste applicant is not. The alternative to the likelihood ratio test is the conditional sign test, which is an exact test more suited to small samples. The test conditions on the event that just one of the two applicants gets called back. In order to test the performance of the two symmetry tests, a Monte Carlo exercise is undertaken in this appendix to check the size and power of the two tests. The results of the exercise are given in the following paragraphs and figures. It was found that the two tests give fairly similar results in samples the same size as the one used in the paper.

In the first set of simulations the data-generation process is multinomial with the probability of the different events being the following: Probability (high-caste Callback, low-caste Callback)= $f(x, y) = (0, 0) = 0.79, f(x, y) = (0, 1) = 0.07, f(x, y) = (1, 0) = 0.07, f(x, y) = (1, 1) = 0.07$. The sample size is 523 and the sign test statistic and likelihood ratio statistics estimated from $10,000$ simulations. The results from the simulations are given in the Figure below.



In the second set of simulations the data generation process is again multinomial but with the probability of the different events being the following: Probability (high-caste Callback, low-caste Callback) $= f(x,y) = (0,0) = 0.75, f(x,y) = (0,1) = 0.05, f(x,y) = (1,0) = 0.10, f(x,y) = (1,1) = 0.10$. The sample size is 523 and the sign test statistic and likelihood ratio statistics estimated from $10,000$ simulations. The results from the simulations are given in the Figure below.


Empirical Dist of the Sign Test Statistic
8000
6000
4000
2000
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Empirical Dist of the LR statistic
8000
6000
4000
2000
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

APPENDIX B

Chapter One- Fisher Exact Test

To test whether differences in callback gaps between high and low-caste workers are significantly different across job types and gender pairs or across recruiters and firms, Fisher's Exact test is performed on the data. In general the test is used to determine the significance of the association between two categorical variables. With large samples a chi-square test can be performed to determine this significance level but for small sample sizes (as the ones in this paper) the Fisher test is used since the chi-square approximation is inappropriate.

Suppose the two variables are X and Y, with X taking on m different values and Y taking on n different values. Let α_{ij} be the number of observations in which $m = i$ and $n = j$ in an $m \times n$ matrix. Let the row and column sums be R_i and C_j and let N be the total sum of the matrix. Then the conditional probability of getting the actual matrix given $R_i, \forall i$ and $C_j, \forall j$ is given by

Conditional Probability $= \frac{(R_1!R_2!...R_m!)(C_1!C_2!...C_n!)}{N!\sum_{i,j}\alpha_{ij}!}$

which is a multivariate version of the hypergeometric distribution. In order to obtain the p-value from the Fisher test all possible matrices consistent with the row and column totals are constructed and their conditional probabilities estimated. Conditional probabilities of all matrices which exhibit equal or greater independence than the original matrix are then added together to get the p-value. Independence may be measured using either the Pearson chi-square or differences in proportions. In the paper the p-values obtained considered matrices as exhibiting equal or greater independence if the conditional probability of obtaining them was less than or equal to the conditional probability of getting the actual matrix.

APPENDIX C

Chapter One- Secondary Data Sources

There are a number of secondary data sources used in the paper which outline the status of low-caste applicants vis a vis high-caste applicants as well as firm level activity in the region of Chennai. This appendix gives a brief background on these data sources, the methods used in their collection, the actual data contained in these datasets and their uses. In the absence of Census level data on Caste[1] these datasets are an important source of information regarding the status of the different caste groups in the country.

C.1. National Sample Survey

The National Sample Survey (NSS) is a nation wide, quinquennial survey on employment and unemployment conducted by the National Sample Survey Organization (NSSO) of the Ministry of Statistics and Program Implementation of the Government of India. The data used in the paper is taken from publicly available reports[2] using the most recent and seventh survey. This is the 61st round of the National Sample Survey conducted between July 2004 and June 2005.

The first quinquennial survey on employment and unemployment was carried out by the NSSO between September 1972 and October 1973 (the 27th round). Since then six more such surveys have been undertaken by the NSSO. The 61st round survey covers the

[1]Since 1941 the collection of individual data on caste was discontinued in the Census (except for affiliation with the SC and ST categories).

[2]The reports are available online at http://mospi.nic.in/mospi_nsso_rept_pubn.htm

whole of Indian Union except Leh and Kargil districts of Jammu and Kashmir, the interior villages in Nagaland and inaccessible villages in Andaman and Nicobar islands. The entire survey period of twelve months was divided into four sub-rounds of three months each with an equal number of sample villages/blocks allotted in each of the four sub-rounds. The survey was conducted in the form of face to face interviews and the sample chosen by stratified multistage sampling.

In the multistage sampling, the first stage units were the census villages for the rural areas and the NSSO urban frame blocks for urban areas. The final stage units were households for both urban and rural areas. Hamlet groups/sub-blocks formed the intermediate stage whenever these were found in the sampled first stage units. Of a total of 12,788 first stage units selected (8128 villages and 4660 urban blocks) 12,601 first stage units ended up being included in the survey. The final sample included 7,999 villages and 4,602 urban blocks covering 124,680 households and enumerating 602,833 individuals. The survey includes detailed information on employment and caste category which, in the publicly available aggregate form, provides a rich source of descriptive data.

C.2. National Election Study

The National Election Study (NES), 2004 is a post election survey in India conducted by the Center for Studies in Developing Economies (CSDS). The study is comparable to the National Election Studies conducted in the US and Britain. The single wave of the post poll survey of 2004 was carried out in all twenty eight states of the Indian Union as well as the three Union territories. The main objective of the survey was to determine the behavior and opinions of indian voters and to explain electoral outcomes. The background information collected by the survey included self-reported data on caste category and jati which has been used in the paper to examine the breakdown by caste group for different occupations (see figure E.8).

The first such surveys were conducted by the CSDS during the 1970s. No survey was then conducted till the mid 1990s. In 1996 three waves of surveys (pre election, mid campaign and post poll) were conducted on a panel of respondents selected by multistage stratified random sampling. The same panel of respondents was used in 1998 for two waves of pre election and post poll surveys. In 1999 again the same panel formed the sample of a post poll survey. The survey carried out in 2004 is the fourth general election for which the survey has been conducted in a row. It has a substantially larger sample of respondents than did the previous surveys and more state level variables.

The sample for the NES, 2004 survey was collected by using a four stage stratified random sampling design. In first stage 420 of the 543 parliamentary constituencies were sampled. In the second stage sampling of assembly contituencies within the parliamentary constituencies was done to get a set of 932 assembly constituencies. In the third stage sampling of polling station areas within the assembly constituencies was carried out by using systematic random sampling to get a set of 2,380 polling station areas. Finally respondents were drawn randomly from the electoral rolls of the selected polling station areas which provided a target of 35,360 names. Of these face to face interviews were conducted for 76.9%, to get a sample size of 27,189 respondents. The numbers of variables on which data was collected were 160. In comparison to official data the sample which was eventually used had a slight over representation of men, significant over representation of rural areas, slight under representation of Muslims, and slight over representation of SC.

C.3. Prowess

Prowess is a firm level database which consists of approximately $10,000$ large and medium size Indian firms. The database is maintained by the Center for the Monitoring of the Indian Economy (CMIE). In order to be included in the database the firm needs to be either a listed company or if it is a public limited company it needs to satisfy the condition $\frac{\text{Income + Liabilities}}{2} = 200$ million Indian Rupees. The database covers most of the organised industrial activity, banking, organized financial and other services sectors in India. The firms in the database account for 75% of all corporate taxes collected by the Indian government, more than 95% of excise duty and 60% of all savings of the Indian corporate sector.

The company level data in Prowess is gathered from annual accounts of the companies and from other sources such as stock exchanges, roc associations, etc. The number of indicators per company is close to two thousand and the information is usually available for ten years. Prowess also provides a normalization of the firm level data across companies and over time. For industry analysis, Prowess has 140 industry groupings with industry wide income and expenditure statements, balance sheets, ratio analysis, benchmarking of industry averages as well as inter and intra industry comparisons. It is also possible to create user defined industry groups or any set of groups within Prowess to carry out the industry analysis. Prowess also generates numerous reports for the firms in the database, as well as providing querying and charting facilities and a textual search tool.

Financial analysis in Prowess contains the income, expenditure, profit and loss summaries, liabilities, assets, cash flow, cost analysis, investment schedule, growth indicators, inventory/working capital cycle analysis, sources and uses of funds, profitability ratios, liquidity ratios, asset utilisation ratios, structure of assets/liabilities, banking disclosure, notes to accounts and auditors qualifications, accounting policies and audited segment wise results for the individual firms. Interim results (quarterly and half yearly) are also available. The database contains information on the products manufactured and traded by the firms such as product details, trends in capacity, trends in production, trends in capacity utilization and trends in sales. It also contains information on raw materials consumed, plant location, total energy consumption, product wise energy consumptions, capital history details, bonus issues, dividend issues and news abstracts. Concerning share prices and indicators it has information on unadjusted daily share prices, adjusted daily share prices, trading volumes, daily stock indicators, investment indicators and stock return analysis. The database also includes background information for the firm such as contact and basic information, investors information, chairmans speech, energy conservation note, technology absorption note and the corporate governance note. Each firm's auditors, bankers, registrars and ratings, its board of directors, equity holding patterns and details, directors report and events are also provided. In short the Prowess dataset has very comprehensive information regarding the firms in its database and this information may be combined with the firm's hiring practices to examine how hiring gaps vary across differences in firm characteristics.

APPENDIX D

Chapter One- Selected Names from Names Database

D.1. High-Caste Names

High-Caste Male	High-Caste Female
Ramesh Iyer	Geetha Iyer
Yogesh Viswanathan	Priyadarshini Santhanam
Madhav Rajagopalan	Mallika Sunderarajan
Anand Seshadri	Seetha Iyengar
Ravi Krishnamurthy	Priya Rajagopalan
Anand Iyer	Veena Radhakrishnan
Ramesh Gopalakrishnan	Vaidehi Rangarajan
Dinesh Raghavan	Gayathri Padmanabhan
Narayan Iyengar	Devaki Iyer
Shiva Kalyanaraman	Lalitha Iyer

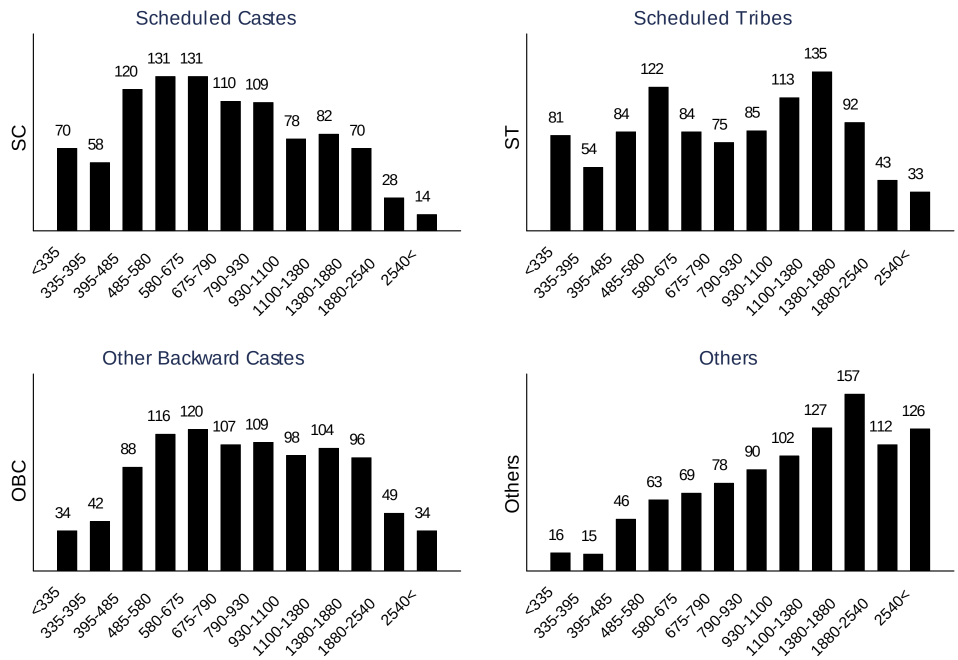
D.2. Low-Caste Names

Low-Caste Male	Low-Caste Female
P Tamizharasan	P Kanmani
K Muthukaruppan	R Kalaimagal
P Inbamani	P Arulmozhi
Veerachami	P Tamarai
P Selvamani	P Ezhil
A Anbarasan	L Kumutha
M Pazhani	Thangam
R Ilango	Shanmugavadivu
Nesamani	P Tamilarasi
G Alagesan	T Poonguzhali

APPENDIX E

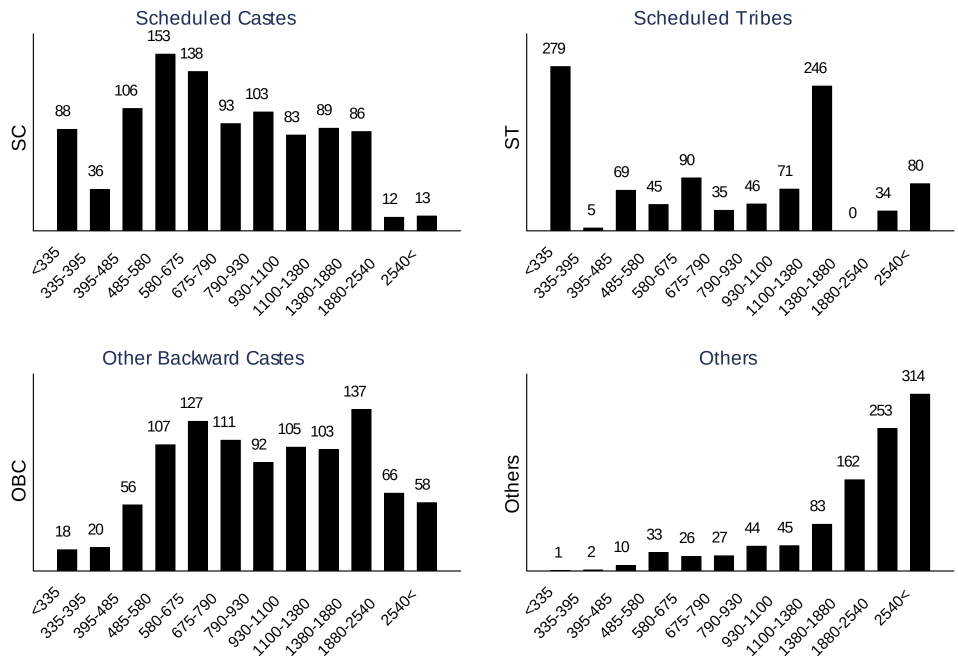
Chapter One- Figures

Figure E.1: Per Capita Consumption Distributions by Caste Category for India



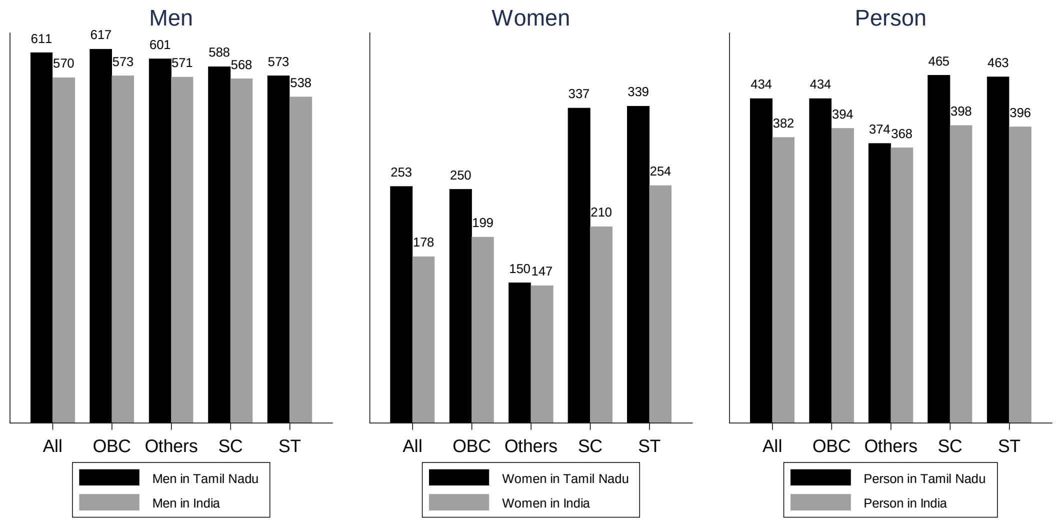
Notes: NSS 61st Round Data (2004-05), Report No. 516. Per Capita Consumption is in Indian Rupees and the Distributions are given for every 1000 people

Figure E.2: Per Capita Consumption Distributions by Caste Category for Tamil Nadu



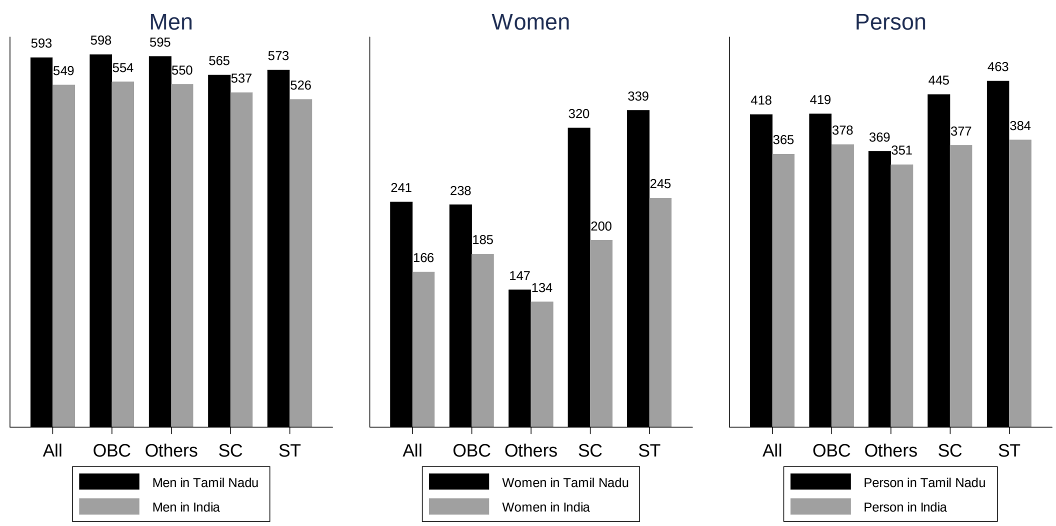
Notes: NSS 61st Round Data (2004-05), Report No. 516. Per Capita Consumption is in Indian Rupees and the Distributions are given for every 1000 people

Figure E.3: Labor Force Participation Rates for different Caste Groups



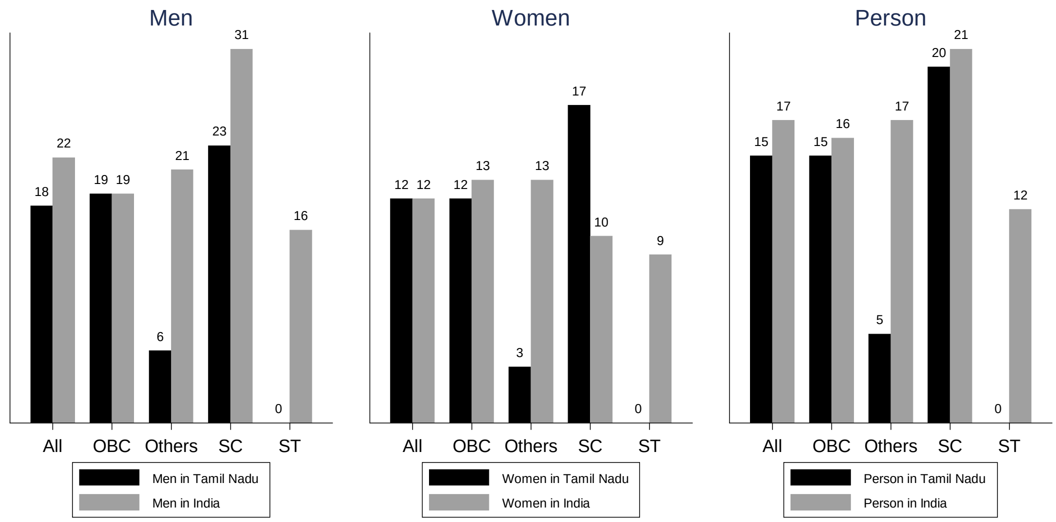
Notes: NSS 61st Round Data (2004-05), Report No. 516. SC stands for Scheduled Castes, ST for Scheduled Tribes, OBC for Other Backward Castes and Other for High Caste groups. Labor Force Participation is per 1000 people.

Figure E.4: Worker Population Ratio for different Caste Groups



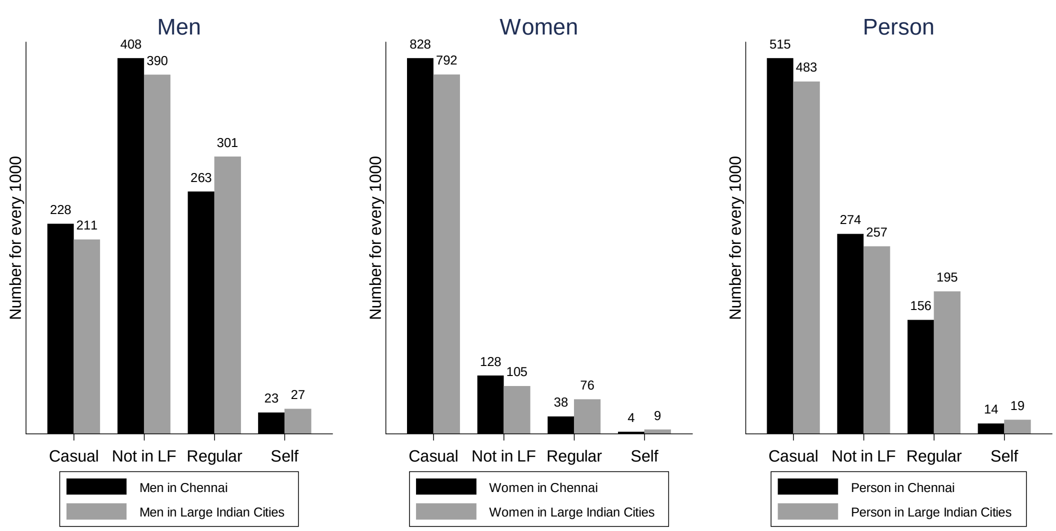
Notes: NSS 61st Round Data (2004-05), Report No. 516. SC stands for Scheduled Castes, ST for Scheduled Tribes, OBC for Other Backward Castes and Other for High Caste groups. Worker Population Ratio is per 1000 people.

Figure E.5: Proportion Unemployed for different Caste Groups



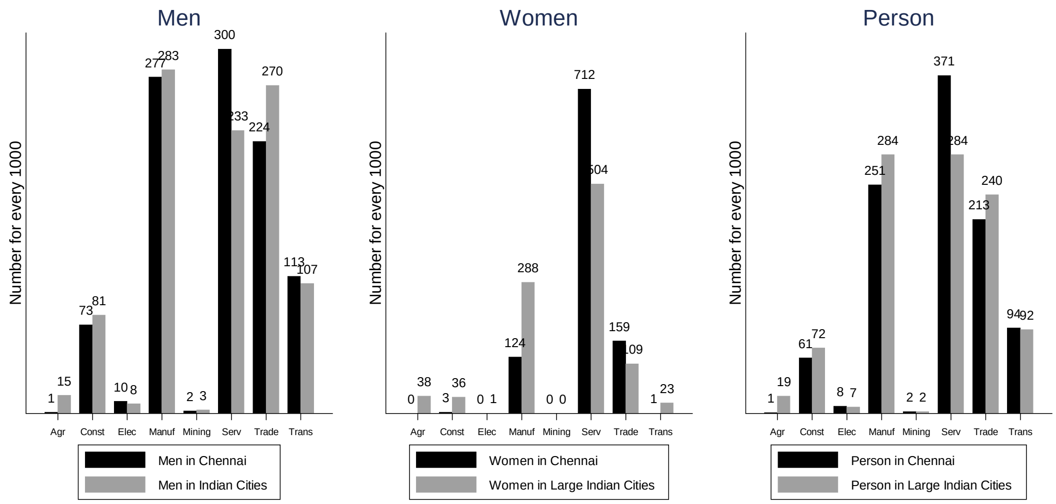
Notes: NSS 61st Round Data (2004-05), Report No. 516. SC stands for Scheduled Castes, ST for Scheduled Tribes, OBC for Other Backward Castes and Other for High Caste groups. Proportion Unemployed is per 1000 people.

Figure E.6: Employment- Chennai vs. Large Indian Cities



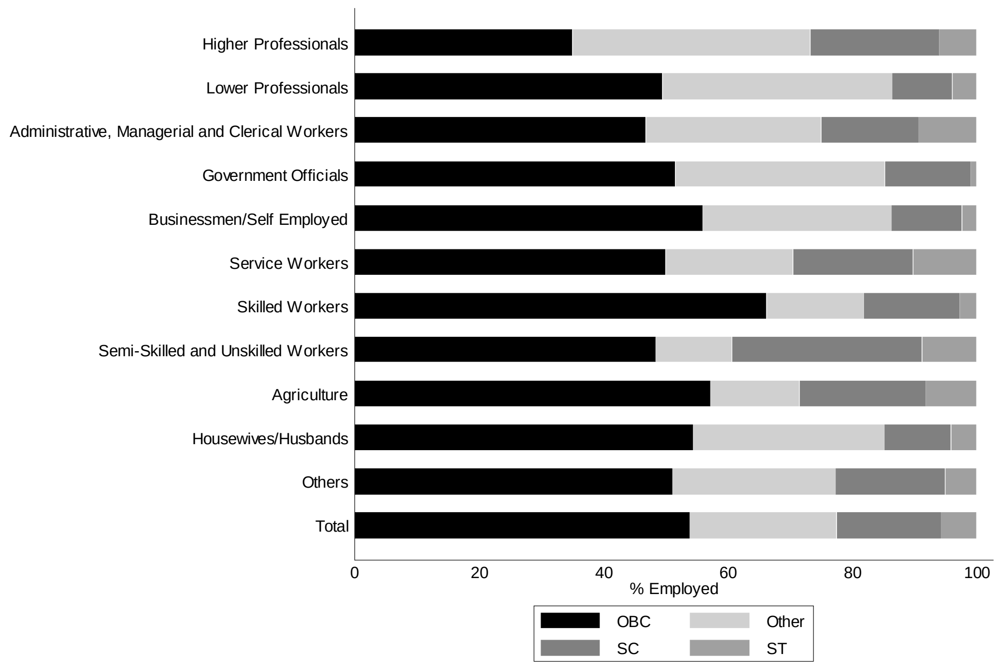
Notes: NSS 61st Round Data (2004-05), Report No. 520. Large Indian cities are cities with a population of more than one million people. Categories are Casual Workers, those not in the Labor Force, Regular Employees and those who are Self-Employed.

Figure E.7: Industry Distribution- Chennai vs. Large Indian Cities



Note: NSS 61st Round Data (2004-05), Report No. 520. Large Indian cities are cities with a population of more than one million people. Categories (in order) are Agriculture, Construction, Electricity and Water, Manufacturing, Mining and Quarrying, Other Services, Trade Hotels and Restaurent and Transportation.

Figure E.8: Employment by Occupation and Caste in South India



Note: Data taken from the NES, 2004 for the States of Tamil Nadu, Andhra Pradesh, Kernataka and Kerala. N=4,653.

APPENDIX F

Chapter One- Tables

Table F.1: Job Types and Websites Used in the Study

	Total	Callback Rate
Number of Resumes	**1046**[**100**%]	**155**[**15**%]
by Type: Customer Service Jobs	674[64%]	114[17%]
Front Office/Administration	372[36%]	41[22%]
by Job Website: Naukri.com	732[70%]	113[15%]
Others	314[30%]	66[13%]

Notes:
1. Other job search websites used in the study included MonsterIndia JobsAhead and Times of India.
2. % in Column 2 are out of the total number of resumes.
3. % in Column 3 are out of number of resumes in a particular category.

Table F.2: Testing Symmetry of Treatment Between High and low-caste Applicants

(High-Caste, Low-Caste)	
(Received Callback, Received Callback)	43
(Received Callback, Did Not Receive Callback)	41
(Did Not Receive Callback, Received Callback)	28
(Did Not Receive Callback, Did Not Receive Callback)	411
Testing Symmetry: p- value	
Likelihood Ratio Test	0.1168
Conditional Sign Test	0.0740

Notes:
H_0 : (Received Callback, Did Not Receive Callback) = (Did Not Receive Callback, Received Callback)

Table F.3: Sub-Sample Characteristics

	Sample, N=1046	**Probit Sub-Sample, N=906**
Applicant Gender		
Female	0.5507	0.4967
	(0.4977)	(0.5003)
Job Type		
Front Office/Administration	0.3556	0.3068
	(0.4789)	(0.4614)

Notes:
1. Columns 2 and 3 give the means of the variables. Standard deviations are given in parentheses.
2. Probit Sub-Sample exludes observations for which gender assignment is non-random.

Table F.4: Effect of Low-Caste on Callback

X	(1) $\frac{\partial y(\text{Prob Callback})}{\partial X}$	(2) $\frac{\partial y(\text{Prob Callback})}{\partial X}$	(3) $\frac{\partial y(\text{Prob Callback})}{\partial X}$	(4) $\frac{\partial y(\text{Prob Callback})}{\partial X}$
L	−0.1886			
	(0.1451)			
L×F		−0.3677		
		(0.2272)		
L×M		−0.0103		
		(0.2251)		
L×CS			−0.1248	
			(0.1662)	
L×FA			−0.4008	
			(0.3059)	
L×M×FA				0.0703
				(0.3954)
L×M×CS				−0.0569
				(0.2493)
L×F×FA				−1.0041**
				(0.4932)
L×F×CS				−0.1989
				(0.2508)
Observations	**906**	**906**	**906**	**906**

Notes:
1. L is a dummy for low caste, F is a dummy for female applicants, M is a dummy for male applicants, CS is a dummy for resumes sent to customer services jobs, FA is a dummy for resumes sent to front office/administration jobs.
2. The marginal effect is for a discrete change in the variable X. Standard deviations are given in parentheses.
3. The effect of low caste is obtained by using a probit specification, controlling for job vacancy level random effects. Controls for applicant gender and job type are included.
4. * Significant at 10% level,** Significant at 5% level,*** Significant at 2.5%, **** Significant at 1%.

Table F.5: Callback Gaps across Applicant Pairs by Recruiter Characteristics

		(I)	(II)	(III)	(IV)	Callback Gaps		Testing Symm	
(H, L)	Total	$(0,0)$[1]	$(1,0)$	$(0,1)$	$(1,1)$	Diff	Ratio	LR[2]	CS[3]
Male	191	78%	9%	4%	9%	5	2.1	0.07*	0.05**
Female	155	78%	6%	7%	8%	−1	0.9	0.83	0.50
Hindu	281	77%	9%	6%	9%	3	1.5	0.16	0.11
Non Hindu	60	80%	3%	8%	8%	−5	0.4	0.25	0.23
(H, L)	**379**	**78%**	**8%**	**6%**	**8%**	**2**	**1.3**	**0.27**	**0.17**

Notes:
For columns (I) through (IV) the percentages are out of total number of applicant pairs in that category (as given in the first column). Diff is the difference between columns (II) and (III), while ratio is the ratio of (II) to (III)
[1] 0 if did not receive callback,1 if received callback, [2] p-value using the likelihood ratio test, [3] p-value using the conditional sign test
* significant at 10% level,** significant at 5% level,*** significant at 2.5%,
**** significant at 1%.

Table F.6: Sub-Sample Characteristics with Recruiter Characteristics

	Entire Sample	**Probit Sub-Sample**
Applicant Gender		
Female	0.5377	0.4836
	(0.4990)	(0.5002)
Job Type		
Front Office/Administration	0.3774	0.3309
	(0.4851)	(0.4710)
Recruiter Characteristics		
Male Recruiters	0.5692	0.5636
	(0.4956)	(0.4964)
Hindu Recruiters	0.8208	0.8291
	(0.3839)	(0.3768)
N	**636**	**550**

Notes:
1. Columns 2 and 3 give the means of the variables. Standard deviations are given in parentheses.
2. Probit Sub-Sample exludes observations for which gender assignment is non-random.

Table F.7: Effect of Low-Caste on Callback with Recruiter Characteristics

X	(1) $\frac{\partial y(\text{Prob Callback})}{\partial X}$	(2) $\frac{\partial y(\text{Prob Callback})}{\partial X}$	(3) $\frac{\partial y(\text{Prob Callback})}{\partial X}$	(4) $\frac{\partial y(\text{Prob Callback})}{\partial X}$
L	-0.0819			
	(0.1848)			
L×MR		-0.3398		
		(0.2556)		
L×FR		0.2202		
		(0.2754)		
L×HR			-0.2061	
			(0.2026)	
L×NHR			0.6181	
			(0.4970)	
L×MR×HR				-0.5090^{*}
				(0.2831)
L×MR×NHR				0.5334
				(0.6146)
L×FR×HR				0.1336
				(0.2942)
L×FR×NHR				0.7290
				(0.6658)
Observations	**550**	**550**	**550**	**550**

Notes:
1. L is a dummy for low caste resumes, MR is a dummy for male recruiter, FR is a dummy for female recruiter, HR is a dummy for hindu recruiter and NHR is a dummy for non-hindu recruiter.
2. The marginal effect is for a discrete change in the variable X. Standard deviations are given in parenthesis.
3. The effect of low caste is obtained by using a probit specification, controlling for job vacancy level random effects. Controls for applicant gender, job type and recruiter characteristics are included.
4. * significant at 10% level,** significant at 5% level,*** significant at 2.5%, **** significant at 1%.

Table F.8: Callback Gaps across Applicant Pairs by Firm Characteristics

		(I)	(II)	(III)	(IV)	Callback Gaps		Symm Tests	
(H, L)	Total	$(0,0)^1$	$(1,0)$	$(0,1)$	$(1,1)$	Diff	Ratio	LR	CS
FO	85	80%	5%	8%	7%	-4	0.6	0.36	0.27
NFO	195	74%	11%	5%	10%	6	2.1	0.05^{**}	0.04^{**}
MDO	124	86%	4%	5%	6%	-1	0.8	0.76	0.50
NMDO	155	68%	13%	7%	12%	6	1.8	0.10^{*}	0.07^{*}
(H, L)	**280**	**76%**	**9%**	**6%**	**9%**	**3**	**1.5**	**0.22**	**0.14**

Notes:
H stands for high caste, L stands for low caste, FO stands for firms with foreign offices NFO stands for firms without foreign offices, MDO stands for firms with multiple domestic offices, NMDO stands for firms without multiple domestic offices.
For columns (I) through (IV) the percentages are out of total number of applicant pairs in that category (as given in the first column). Diff is the difference between columns (II) and (III), while ratio is the ratio of (II) to (III)
[1] 0 if did not receive callback,1 if received callback, [2] p-value using the likelihood ratio test, [3] p-value using the conditional sign test
* significant at 10% level,** significant at 5% level,*** significant at 2.5%,
**** significant at 1%.

Table F.9: Sub-Sample Characteristics with Firm Characteristics

	Entire Sample	**Probit Sub-Sample**
Applicant Gender		
Female	0.5771	0.5188
	(0.4944)	(0.5002)
Job Type		
Front Office/Administration	0.3620	0.3096
	(0.4810)	(0.4628)
Firm Characteristics		
Firms with Multiple Domestic Offices	0.4444	0.4435
	(0.4973)	(0.4973)
Firms with Foreign Offices	0.3011	0.3264
	(0.4591)	(0.4694)
N	**558**	**478**

Notes:
1. Columns 2 and 3 give the means of the variables. Standard deviations are given in parentheses.
2. Probit Sub-Sample exludes observations for which gender assignment is non-random.

Table F.10: Effect of Low-Caste on Callback with Firm Characteristics

X	(1) $\frac{\partial y(\text{Prob Callback})}{\partial X}$	(2) $\frac{\partial y(\text{Prob Callback})}{\partial X}$	(3) $\frac{\partial y(\text{Prob Callback})}{\partial X}$	(4) $\frac{\partial y(\text{Prob Callback})}{\partial X}$
L	-0.1198			
	(0.1909)			
L×MDO		0.4679		
		(0.3715)		
L×NMDO		-0.3560		
		(0.2303)		
L×FO			0.4151	
			(0.3458)	
L×NFO			-0.3739	
			(0.2366)	
L×MDO×FO				1.0565^{**}
				(0.4893)
L×MDO×NFO				-0.1696
				(0.5179)
L×NMDO×FO				-0.1283
				(0.4546)
L×NMDO×NFO				-0.4277
				(0.2624)
Observations	**550**	**550**	**550**	**550**

Notes:
1. L is a dummy for low caste resumes, FO is a dummy for firms with a foreign office, NFO is a dummy for firms without a foreign office, MDO is a dummy for a firm with multiple domestic offices, NMDO is a dummy for a firm without multiple domestic offices.
2. The marginal effect is for a discrete change in the variable X. Standard deviations are given in parenthesis.
3. The effect of low caste is obtained by using a probit specification, controlling for job vacancy level random effects. Controls for applicant gender, job type and firm characteristics are included.
4. * significant at 10% level,** significant at 5% level,*** significant at 2.5%, **** significant at 1%.

Table F.11: Effect of Recruiter and Firm Characteristics on Differential Callback

X	Sample Mean	Effect
Recruiter Characteristics ($N = 550$):		
Male Recruiter	0.5636	-0.5336
	(0.4964)	(0.3726)
Hindu Recruiter	0.8291	-0.8277
	(0.3768)	(0.5328)
Firm Characteristics ($N = 478$):		
Multiple Domestic Offices Absent	0.5565	-0.8968^{**}
	(0.4973)	(0.4556)
Foreign Offices Absent	0.6736	-0.8349^{*}
	(0.4693)	(0.4327)

Notes:
1.* significant at 10% level,** significant at 5% level,*** significant at 2.5%, **** significant at 1%.
2. Marginal effects are estimated by running a probit with random firm effects on the recruiter or firm characteristic, low-caste and the interaction of the two. The effect of recruiter and firm characteristics on differential callback is the effect due to discrete changes in the interaction term. Standard deviations are given in parentheses.

Table F.12: Average Treatment Effects on Callback Outcome

	ATE	90% CI on ATE
Entire Population	0.02	$[0.0000, 0.0535]$
Male Recruiters	0.05	$[0.0000, 0.0904]$
Female Recruiters	-0.01	$[-0.0548, 0.0440]$
Hindu Recruiters	0.03	$[-0.0038, 0.0673]$
Non-Hindu Recruiters	-0.05	$[-0.1190, 0.0172]$
Firms with Multiple Domestic Offices	-0.01	$[-0.0526, 0.0301]$
Firms without Multiple Domestic Offices	0.06	$[0.0000, 0.1220]$
Firms with Foreign Offices	-0.04	$[-0.0851, 0.0267]$
Firms without Foreign Offices	0.06	$[0.0148, 0.1095]$

Notes:
1.Average Treatment Effect is $E[y(1) - y(0)]$ where $y(1)$ is callback for high-caste resumes and $y(0)$ is callback for low-caste applicants.
2.confidence intervals are constructed by using 200 bootstrap replications.

APPENDIX G

Chapter Two- Tables

Table G.1: Victim and Suspect Characteristics

Unemployment	Victims	61%	
	Suspects	60%	
Relationship	Divorced or separated husband	3%	
	Unmarried male partner	45%	
	Current husband	35%	
	Wife or girlfriend	2%	
	Son, brother, roommate, other	15%	
Prior assaults	Victim assaulted by suspect, last 6 months	80%	
	Police intervention in domestic dispute, last 6 months	60%	
	Couple in counseling program	27%	
Prior arrests	Arrested for any offense	59%	
	Arrested for crime against person	31%	
	Arrested on domestic violence statute	5%	
	Arrested on an alcohol offense	29%	
Mean Age	Victims	30 years	
	Suspects	32 years	
		Victims	Suspects
Education	< High school	43%	42%
	High School only	33%	36%
	> High school	24%	22%
		Victims	Suspects
Race	White	57%	45%
	Black	23%	36%
	Native American	18%	16%
	Other	2%	3%

Notes: This information was available for cases in which initial interviews were obtained, $N = 205$

\

Table G.2: Official Data

Assigned Treatment	Received Treatment	Number who Recidivated
1	1	10 (of 91)
1	2	0 (of 0)
1	3	0 (of 1)
2	1	3 (of 18)
2	2	15 (of 84)
2	3	1 (of 5)
3	1	5 (of 26)
3	2	1 (of 5)
3	3	20 (of 82)

Notes: 1 stands for Arrest, 2 stands for Advice, 3 stands for Separation.
Recidivism is measured as repeat violence against the same victim.
Data from Official Police Reports, $N = 312$.

Table G.3: Victim Interview Data

Assigned Treatment	Received Treatment	Number who Recidivated
1	1	9 (of 54)
1	2	0 (of 0)
1	3	0 (of 0)
2	1	2 (of 15)
2	2	10 (of 44)
2	3	1 (of 1)
3	1	8 (of 20)
3	2	0 (of 1)
3	3	11 (of 53)

Notes: 1 stands for Arrest, 2 stands for Advice, 3 stands for Separation.
Recidivism is measured as repeat violence against the same victim.
Data from Victim Interviews, $N = 188$.

Table G.4: Recidivism Probabilities using Official Data

	Treatment	Bounds on Recidivism	90% CI on Recidivism Bounds
Worst Case Bounds	1	$[0.06, 0.63]$	$[0.04, 0.66]$
	2	$[0.05, 0.77]$	$[0.03, 0.80]$
	3	$[0.07, 0.79]$	$[0.04, 0.82]$
Randomly Assigned Treatment	1	$[0.11, 0.12]$	$[0.06, 0.18]$
	2	$[0.14, 0.36]$	$[0.09, 0.43]$
	3	$[0.18, 0.45]$	$[0.12, 0.53]$

Notes: 1 stands for Arrest, 2 for Advice and 3 for Separation.

Table G.5: Treatment Effects using Official Data

	Treatment Effect	Bounds	90% CI
Worst Case Bounds	$P[y(1)] - P[y(2)]$	$[-0.71, 0.58]$	$[-0.76, 0.63]$
	$P[y(1)] - P[y(3)]$	$[-0.73, 0.56]$	$[-0.78, 0.62]$
Randomly Assigned Treatment	$P[y(1)] - P[y(2)]$	$[-0.25, -0.02]$	$[-0.37, 0.09]$
	$P[y(1)] - P[y(3)]$	$[-0.34, -0.06]$	$[-0.47, 0.06]$

Notes: 1 stands for Arrest, 2 for Advice and 3 for Separation.

Table G.6: Recidivism Probabilities using Victim Interview Data

		Treatment	Bounds on Recidivism	90% CI on Recidivism Bounds
Worst Case Bounds		1	$[0.06, 0.78]$	$[0.04, 0.86]$
		2	$[0.03, 0.89]$	$[0.02, 0.95]$
		3	$[0.04, 0.87]$	$[0.02, 0.93]$
Randomly Assigned Treatment	Ignorable Selection	1	$[0.16, 0.18]$	$[0.09, 0.27]$
		2	$[0.18, 0.39]$	$[0.10, 0.50]$
		3	$[0.15, 0.42]$	$[0.08, 0.51]$
	No Ignorable Selection	1	$[0.10, 0.51]$	$[0.04, 0.63]$
		2	$[0.09, 0.68]$	$[0.05, 0.76]$
		3	$[0.10, 0.63]$	$[0.05, 0.70]$

Notes: 1 stands for Arrest, 2 for Advice and 3 for Separation.

Table G.7: Treatment Effects using Victim Interview Data

		Treatment Effects	Bounds	90% CI
Worst Case Bounds		$P[y(1)] - P[y(2)]$	$[-0.83, 0.75]$	$[-0.91, 0.84]$
		$P[y(1)] - P[y(3)]$	$[-0.81, 0.74]$	$[-0.89, 0.84]$
Randomly Assigned Treatment	Ignorable Selection	$P[y(1)] - P[y(2)]$	$[-0.23, 0.00]$	$[-0.41, 0.17]$
		$P[y(1)] - P[y(3)]$	$[-0.26, 0.03]$	$[-0.42, 0.19]$
	No Ignorable Selection	$P[y(1)] - P[y(2)]$	$[-0.58, 0.42]$	$[-0.72, 0.58]$
		$P[y(1)] - P[y(3)]$	$[-0.53, 0.41]$	$[-0.66, 0.58]$

Notes: 1 stands for Arrest, 2 for Advice and 3 for Separation.

Table G.8: Recidivism Probabilities when there is Self-Selection using Official Data

	Treatment	Bounds on Recidivism	90% CI on Recidivism Bounds
Skimming	1	$[0.06, 0.13]$	$[0.04, 0.18]$
	2	$[0.13, 0.77]$	$[0.08, 0.80]$
	3	$[0.17, 0.79]$	$[0.11, 0.82]$
Outcome Optimization	1	$[0.18, 0.62]$	$[0.08, 0.66]$
	2	$[0.18, 0.77]$	$[0.08, 0.80]$
	3	$[0.18, 0.79]$	$[0.08, 0.82]$

Notes: 1 stands for Arrest, 2 for Advice and 3 for Separation

Table G.9: Treatment Effects when there is Self-Selection using Official Data

	Treatment Effect	Bounds	90% CI
Skimming	$P[y(1)] - P[y(2)]$	$[-0.71, 0.00]$	$[-0.76, 0.10]$
	$P[y(1)] - P[y(3)]$	$[-0.73, -0.04]$	$[-0.78, 0.07]$
Outcome Optimization	$P[y(1)] - P[y(2)]$	$[-0.59, 0.44]$	$[-0.72, 0.58]$
	$P[y(1)] - P[y(3)]$	$[-0.61, 0.44]$	$[-0.74, 0.58]$

Notes: 1 stands for Arrest, 2 for Advice and 3 for Separation

Table G.10: Recidivism Probabilities when there is Self-Selection using Victim Interview Data

		Treatment	Bounds on Recidivism	90% CI on Recidivism Bounds
Skimming	Ignorable Selection	1	$[0.09, 0.21]$	$[0.06, 0.29]$
		2	$[0.16, 0.78]$	$[0.09, 0.84]$
		3	$[0.16, 0.78]$	$[0.09, 0.84]$
	No Ignorable Selection	1	$[0.06, 0.48]$	$[0.04, 0.58]$
		2	$[0.08, 0.89]$	$[0.04, 0.95]$
		3	$[0.10, 0.87]$	$[0.05, 0.93]$
Outcome Optimization	Ignorable Selection	1	$[0.22, 0.66]$	$[0.17, 0.72]$
		2	$[0.22, 0.78]$	$[0.17, 0.84]$
		3	$[0.22, 0.78]$	$[0.17, 0.84]$
	No Ignorable Selection	1	$[0.13, 0.78]$	$[0.10, 0.86]$
		2	$[0.13, 0.89]$	$[0.10, 0.95]$
		3	$[0.13, 0.87]$	$[0.10, 0.93]$

Notes: 1 stands for Arrest, 2 for Advice and 3 for Separation

Table G.11: Treatment Effects when there is Self-Selection using Victim Interview Data

		Treatment Effects	Bounds	90% CI
Skimming	Ignorable Selection	$P[y(1)] - P[y(2)]$	$[-0.69, 0.05]$	$[-0.78, 0.20]$
		$P[y(1)] - P[y(3)]$	$[-0.69, 0.05]$	$[-0.69, 0.05]$
	No Ignorable Selection	$P[y(1)] - P[y(2)]$	$[-0.83, 0.40]$	$[-0.91, 0.54]$
		$P[y(1)] - P[y(3)]$	$[-0.81, 0.38]$	$[-0.89, 0.53]$
Outcome Optimization	Ignorable Selection	$P[y(1)] - P[y(2)]$	$[-0.56, 0.44]$	$[-0.67, 0.55]$
		$P[y(1)] - P[y(3)]$	$[-0.56, 0.44]$	$[-0.67, 0.55]$
	No Ignorable Selection	$P[y(1)] - P[y(2)]$	$[-0.76, 0.65]$	$[-0.85, 0.76]$
		$P[y(1)] - P[y(3)]$	$[-0.74, 0.65]$	$[-0.83, 0.76]$

Notes: 1 stands for Arrest, 2 for Advice and 3 for Separation

APPENDIX H

Chapter Three- Non-Wage Compensation and Wage Gaps

To study racial differences in non-wage compensation and wages, we also carry out estimation of the following model:

$$IP_i^* = Z_i\delta + \epsilon_i \tag{H.1}$$

$$IP_i = \begin{cases} 1 \text{ if } IP_i^* > 0 \\ 0 \text{ if } IP_i^* \leq 0 \end{cases} \tag{H.2}$$

where IP_i is a dummy variable taking the value one if individual i has both health insurance and pension from the employer and the value zero if not, IP_i^* is a latent variable that determines whether or not an individual gets non-wage compensation (health insurance or HI and pension or P) and Z_i is a vector of characteristics that determines whether or not an individual gets non-wage compensation (health insurance and pension). We assume the error term ϵ_i is distributed normally so we carry out probit estimation of the model given in (H.1)-(H.2), separately for men and women.[1]

[1]Probit estimates of regressions on whether or not the individual received both health insurance and pension for each year from 1996 to 2006 and separately for men and women are not given in the paper but are available from the authors on request.

The set of Z_i variables specified in (H.2) include education, age, occupation, region and number of children less than age six. When using the CPS data for estimation, we introduce a set of dummy variables indicating whether the worker has no education, some high school education, high school education or college/grad school education. We also use eight occupation dummies and four region dummies to determine the impact of occupation and region on non-wage compensation. These characteristics influence the worker's demand for health insurance and pension. Education, occupation and age have an impact on demand for health insurance and pension because of their impact on wages. The number of children has an impact on demand by changing the family's saving behavior. We also control for spouse income which may have an impact on demand for non-wage compensation by changing the family's marginal tax rate. In addition we also use in the set of Z_i variables, union membership,[2] employer's firm size and industry. We use five dummy variables for employer's firm size (whether number of employees in the firm are less than 25, between 25 and 99, between 100 and 499, between 500 and 999 or greater than 1000) and thirteen different dummy variables for industry.

In order to examine the consequences of differences in non-wage compensation on wages, we specify a model of wage formation. Assume that whether or not a worker gets both health insurance and a pension from the employer, or IP_i is determined according to

[2]Since 1983, questions on union/employment association membership are asked only to a quarter of the sample (the outgoing rotation groups) in each month (Hirsch and Macpherson, 2003). To obtain information of union membership for the remaining three quarters of the sample in each year, we make use their responses to the Basic CPS survey in the following months. Specifically, we look at their responses to the questions on union membership during their outgoing interviews. We also restrict to those who do not experience unemployment between the ASEC and their outgoing interview. Doing so essentially eliminate those who changed jobs during this period, which will contaminate our data (i.e. the employer that offers pension may not be the employer the interviewee worked for during the month when he answered the union membership questions).

equations (H.1)-(H.2). For workers who receive both health insurance and pension from their employers, wages W_{IPi} are determined according to:

$$W_{IPi} = X_i \beta_{IP} + u_{IPi} \tag{H.3}$$

For workers who do not receive both health insurance and pension from their employers, wages W_{NIPi} are instead determined according to:

$$W_{NIPi} = X_i \beta_{NIP} + u_{NIPi} \tag{H.4}$$

where X_i is the set of characteristics of worker i which determine wages. We assume that the error terms u_{IPi} and u_{NIPi} are distributed normally.

The expected wage conditional on observed characteristics and whether or not the individual gets health insurance and pension is given by

$$E(W_{IPi}|X_i, IP_i = 1) = X_i \beta_{IP} + \omega_{IP} \lambda_{IPi} \tag{H.5}$$

$$E(W_{NIPi}|X_i, IP_i = 0) = X_i \beta_{NIP} - \omega_{NIP} \lambda_{NIPi} \tag{H.6}$$

where $\lambda_{IPi} = \frac{f(Z_i \delta)}{F(Z_i \delta)}$ and $\lambda_{NIPi} = \frac{f(Z_i \delta)}{1 - F(Z_i \delta)}$, $f(.)$ and $F(.)$ being the standard normal density and distribution functions, $\omega_{IP} = cov(u_{IPi}, \epsilon_i)$ and $\omega_{NIP} = cov(u_{NIPi}, \epsilon_i)$.

The above is an endogenous switching regression model, which we estimate by using the Heckman two stage procedure. This involves first the estimation of probit regressions of whether or not the individual has non-wage compensation (health insurance and pension) and then the estimation of the two wage equations with selectivity corrections.

To examine the wage effect of differential racial representation in non-wage compensation, we use the following method: we estimate the predicted wage for each black worker with estimates of the switching regression model,

(H.7)

$$\widehat{W}_i(\widehat{\delta}_b, \widehat{\beta}_b, \widehat{\omega}_b) = F(Z_i\widehat{\delta}_b)[X_i\widehat{\beta}_{IPb} + \widehat{\omega}_{IPb}\lambda_{IPi}(\widehat{\delta}_b)] + [1 - F(Z_i\widehat{\delta}_b)][X_i\widehat{\beta}_{NIPb} - \widehat{\omega}_{NIPb}\lambda_{NIPi}(\widehat{\delta}_b)]$$

Then the predicted mean wage is given by

$$\overline{W}_b(\widehat{\delta}_b, \widehat{\beta}_b, \widehat{\omega}_b) \equiv (\frac{1}{N_b}) \sum_{i=1}^{N_b} W_i(\widehat{\delta}_b, \widehat{\beta}_b, \widehat{\omega}_b) \tag{H.8}$$

The consequence of the unexplained racial gap in non-wage compensation for black wages is determined by computing the expected wage when black non-wage compensation is determined by the white model. The computation is the same as in equation (H.7)-(H.8) for black wages except that $\widehat{\delta}_w$ replaces $\widehat{\delta}_b$ everywhere. The effect of racial differences in wage structure, holding determination of non-wage compensation constant, is done by replacing $(\widehat{\beta}_b, \widehat{\omega}_b)$ with $(\widehat{\beta}_w, \widehat{\omega}_w)$ in (H.7)-(H.8). The combined effect of differences in non-wage compensation and wage determination may be obtained by replacing $(\widehat{\delta}_b, \widehat{\beta}_b, \widehat{\omega}_b)$ by $(\widehat{\delta}_w, \widehat{\beta}_w, \widehat{\omega}_w)$ in (H.7)-(H.8). Differences in the observed characteristics explain that part of the wage gap which cannot be explained by the difference in non-wage compensation and wage structure.

We carry out estimation of the switching regression model by using the CPS data on men and women from different cohorts. The independent variables that are used to determine wages, the set of X_i variables specified above, we use worker's education, age, occupation, region, union membership, firm size, industry and tenure. For education we use a set of dummy variables indicating whether the worker has no education, some high school education, high school education or college/grad school education. To determine the impact of occupation and region on wages, we use eight occupation dummies and four region dummies. We use five dummy variables for employer's firm size (whether number of employees in the firm are less than 25, between 25 and 99, between 100 and 499, between 500 and 999 or greater than 1000) and thirteen different dummy variables for industry. Data on tenure is available only for the years 1996, 1998, 2000, 2002, 2004 and 2006.[3] Due to this we carry out estimation of the switching model and wages only for these years. Predicted wages under different assumptions, as outlined in the previous paragraph, are given in tables Appendix Table 9 and 10.

Supplementary tables K.13 and K.14 gives the predicted logarithim of the weekly wage when we make different assumptions on non-wage compensation and wage determination for the male samples in different cohorts of the CPS. The predicted average of the logarithm of weekly wage for white men is always higher than that for black men, in evey

[3]Information about tenure with current employer is available in the CPS "Occupational Mobility and Job Tenure" supplements every two years beginning 1996. These supplements are administered in January or February. Precisely, the question asks "How long have you been working CONTINUOUSLY working for the present employer?" Given the structure of the CPS, there will be a fraction of interviewees who responded to both the Occupation Mobility and Tenure Supplement and the March Annual Supplement.

cohort. The difference in the predicted average of the logarithm of weekly wages across black and white men actually increases if black men get non-wage compensation according to the white model (while wages are determined as in the black model), or alternatively if white men get non-wage compensation according to the black model (while their wage is still determined by the white model). The difference in the predicted average of the logarithm of weekly wages across black and white men decreases if black men get wages according to the white model and non-wage compensation according to the black model. The difference also decreases if white men get wages according to the black model and non-wage compensation according to the white model. If black men get both wages and non-wage compensation according to the white model then again the racial gap in predicted logarithm of the weekly wage is reduced, but does not disappear. This racial in wages is due to differences in characteristics across black and white men. Similarly if white men get both wages and non-wage compensation according to the black model then the racial gap is reduced.

From the simulation exercises carried out and reported in supplementary tables K.13 and K.14, we find that the racial gap in wages for men increases if black men get non-wage compensation according to the white model. We had found that coverage of non-wage compensation (health insurance and pension) is higher for black men than it is for white men. Wages are higher in jobs with non-wage compensation than jobs without non-wage compensation; therefore the racial gap in wages goes up if non-wage compensation for blacks is according to the white model. Alternatively the same result holds (racial gap in wages increases) if non-wage compensation for white men is according to the black model. Higher coverage of non-wage compensation for black men is probably because black men are more likely to work in larger firms which offer greater non-wage compensation. Also we find that the racial gap in wages persists even if both non-wage compensation and wage determination for black men is according to the white model, indicating that part of the racial gap in wages is due to differences in characteristics across black and white men.

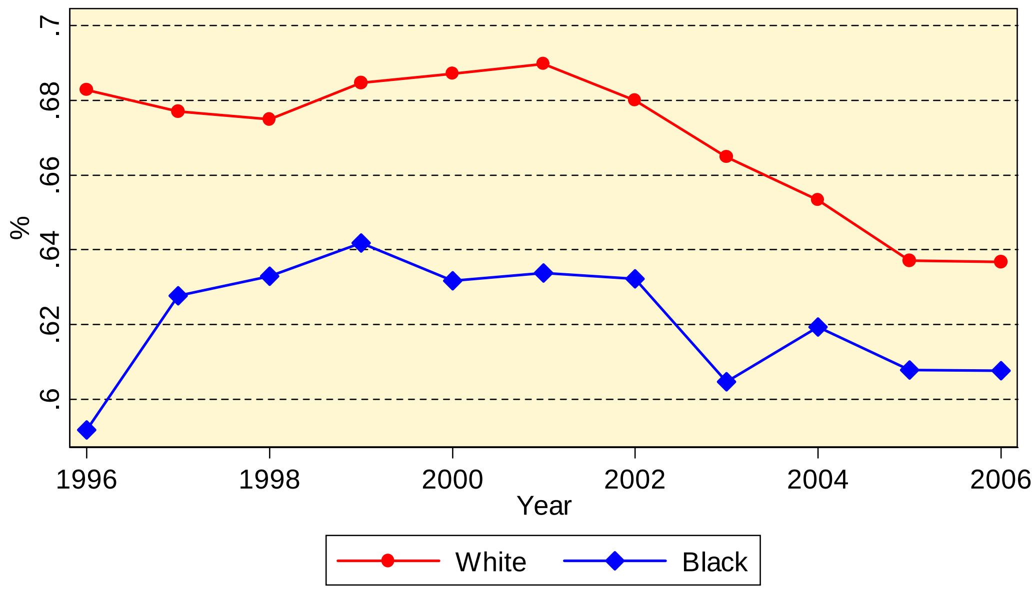
In supplementary tables K.15 and K.16 is given the predicted logarithm of the weekly wage when different assumptions on non-wage compensation and wages are made for the different cohorts of female samples in the CPS. Except for 2006, the predicted logarithm of weekly wage for white women is always higher than that for black women. The difference in the predicted logarithm of weekly wage across black and white women does not always increase if black women get non-wage compensation according to the white model and wages according to the black model. Neither does it always increase if white women get non-wage compensation according to the black model and wages according to the white model. Instead there is an increase in some cohorts and a decrease in the racial gap in predicted wage for other cohorts. The difference in the predicted average of the logarithm of weekly wages across black and white women does not always decrease if black women get wages according to the white model and non-wage compensation according to the black model. Neither does it always decrease if white women get wages according to the black model and non-wage compensation according to the white model. As previously the racial gap in predicted wage decreases for some cohorts but increases for others. If black women get both wages and non-wage compensation according to the white model then the racial gap in wages is reduced but persists due to differences in characteristics between black and white women. Similarly if white women get wages and non-wage compensation according to the black model, then again the racial gap in wages is less.

From the simulation exercises carried out and reported in supplementary tables K.15 and K.16 for women, we find inconclusive results on how the racial gap in wages for women changes when non-wage compensation for black women is according to the white model. For some cohorts the racial gap increases when non-wage compensation for black women is according to the white model. For other cohorts it increases. Black women have higher non-wage compensation than white women in some cohorts and lower non-wage compensation in others, therefore the inconclusive results vis a vis wage gaps are expected. We also find that racial gaps continue to favor white women over black women when both non-wage compensation and wage determination are reversed either for black or for white women. This indicates the importance of differences in characteristics across black and white women in explaining the racial gaps in wages.

APPENDIX I

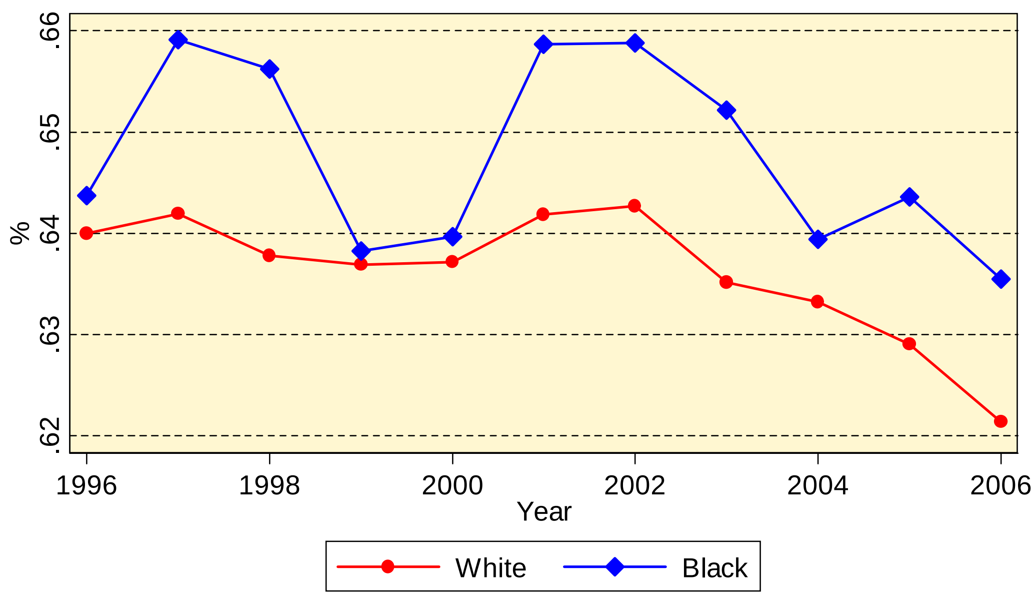
Chapter Three- Figures

Figure I.1: Proportion of men with employer provided health insurance, 1996 to 2006



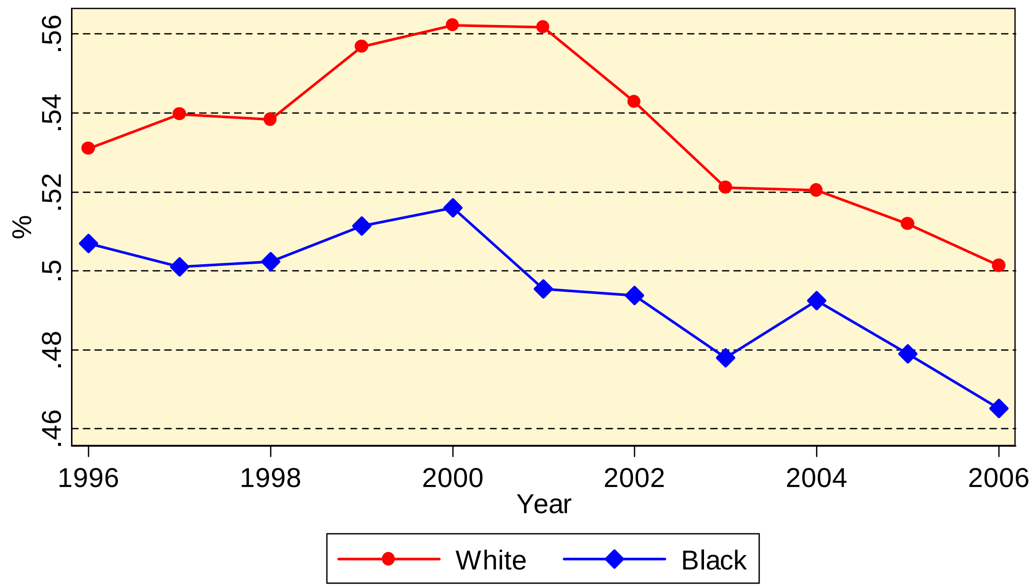
Notes:
Data from Current Population Surveys- March Annual Supplements.
The diagram shows, by calendar year, the percentage of black and white men who receive employer provided health insurance.
We restrict to those men who work more than 35 hours per week and ages 25-64.

Figure I.2: Proportion of women with employer provided health insurance, 1996 to 2006



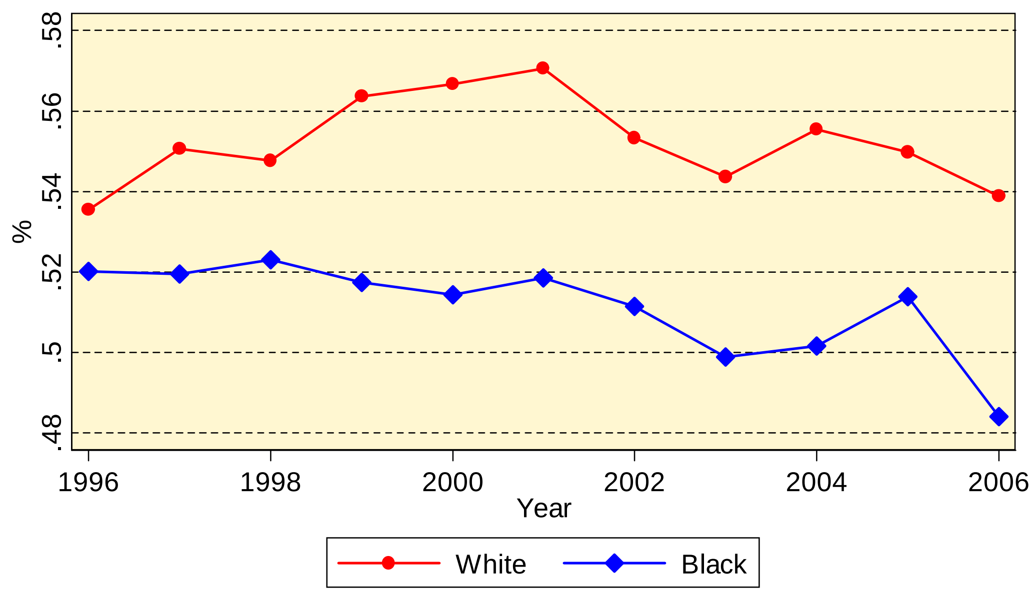
Notes:
Data from Current Population Surveys- March Annual Supplements.
The diagram shows, by calendar year, the percentage of black and white women who receive employer provided health insurance.
We restrict to those women who work more than 35 hours per week and ages 25-64.

Figure I.3: Proportion of men with pension coverage, 1996 to 2006



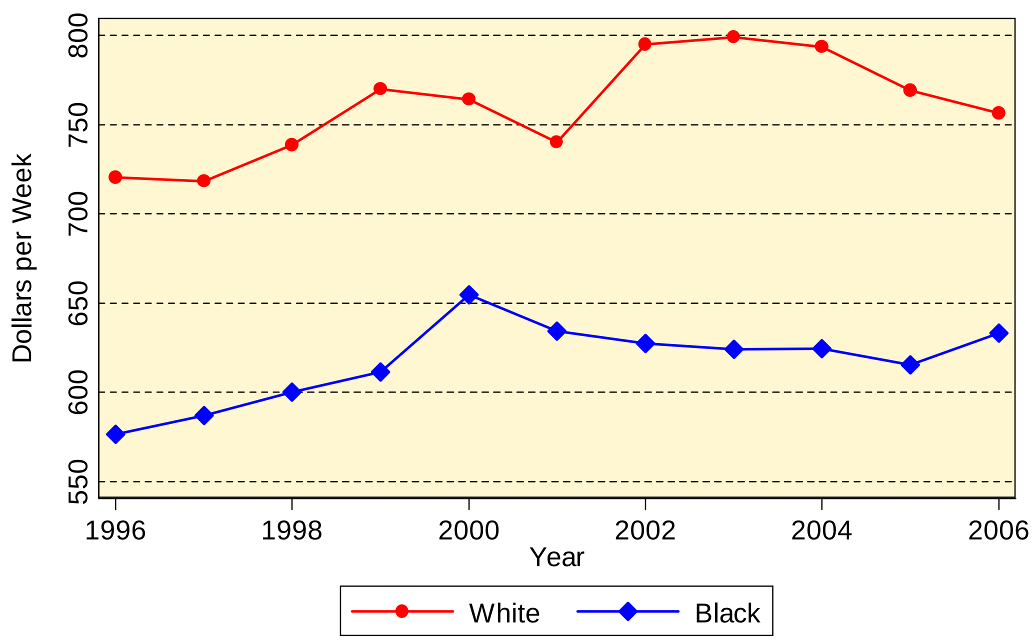
Notes:
Data from Current Population Surveys- March Annual Supplements.
The diagram shows, by calendar year, the percentage of black and white men who are covered by pension.
We restrict to those men who work more than 35 hours per week and ages 25-64.

Figure I.4: Proportion of women with pension coverage, 1996 to 2006



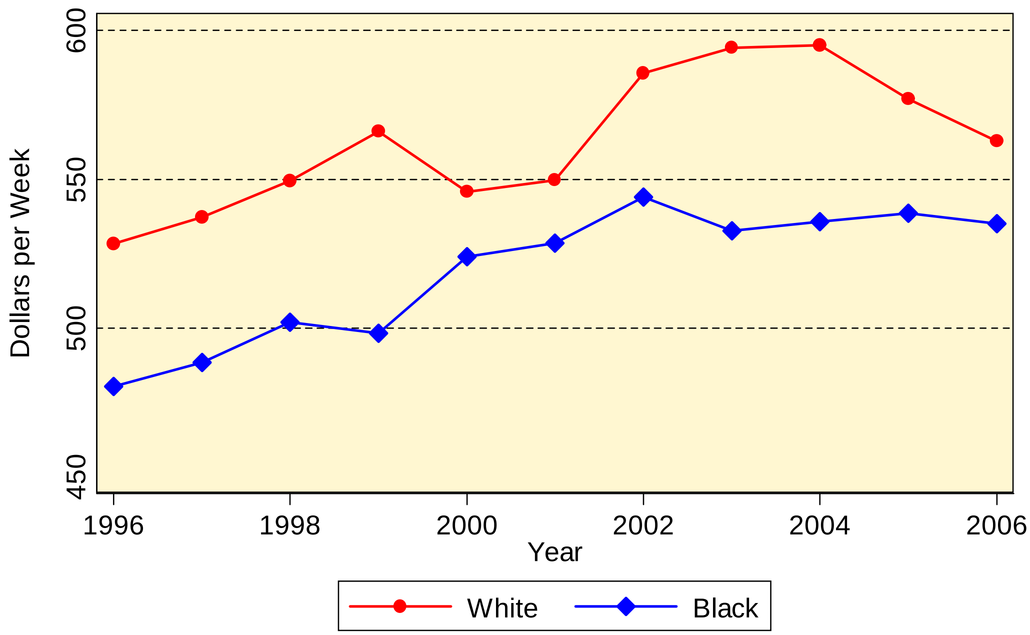
Notes:
Data from Current Population Surveys- March Annual Supplements.
The diagram shows, by calendar year, the percentage of black and white women who are covered by pension.
We restrict to those women who work more than 35 hours per week and ages 25-64.

Figure I.5: Median weekly salary for men, 1996 to 2006



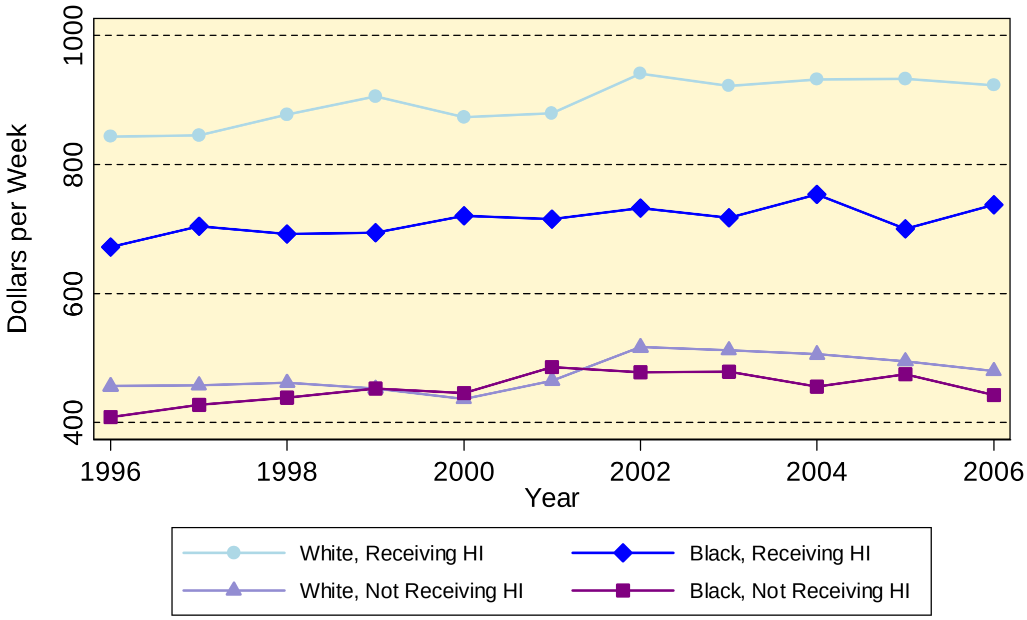
Notes:
Data from Current Population Surveys- March Annual Supplements.
The diagram shows, by calendar year, the median salary of black and white men.
We restrict to men who work more than 35 hours per week and ages 25-64.

Figure I.6: Median weekly salary for women, 1996 to 2006



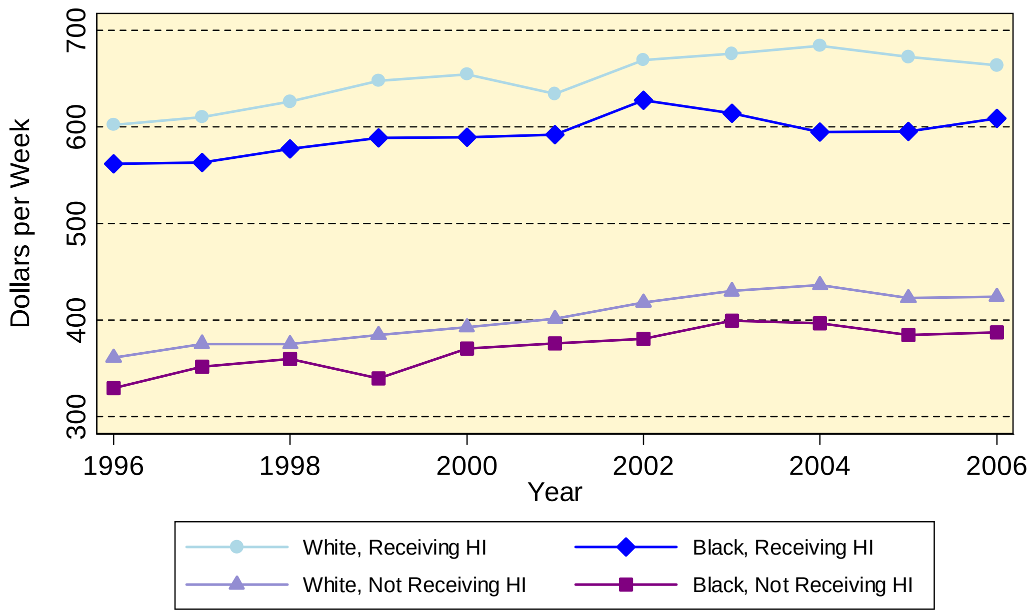
Notes:
Data from Current Population Surveys- March Annual Supplements.
The diagram shows, by calendar year, the median salary of black and white women.
We restrict to women who work more than 35 hours per week and ages 25-64.

Figure I.7: Median weekly salary for men with/without health insurance, 1996 to 2006



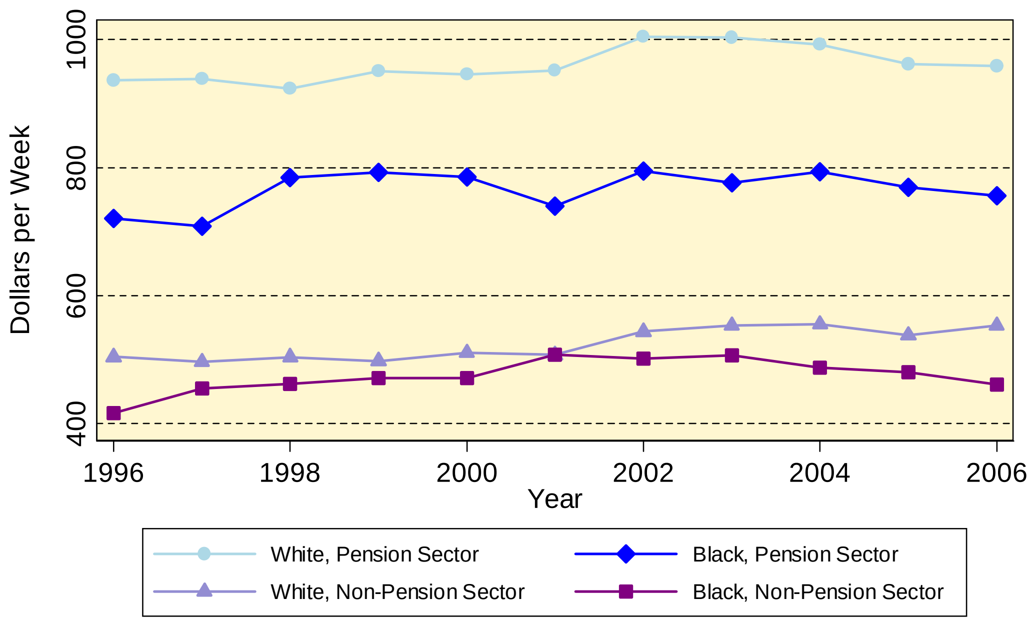
Notes:
Data from Current Population Surveys- March Annual Supplements.
The diagram shows, by calendar year, the median salary of black and white men who receive/do not receive health insurance separately.
We restrict to men who work more than 35 hours per week and ages 25-64.

Figure I.8: Median weekly salary for women with/without health insurance, 1996 to 2006



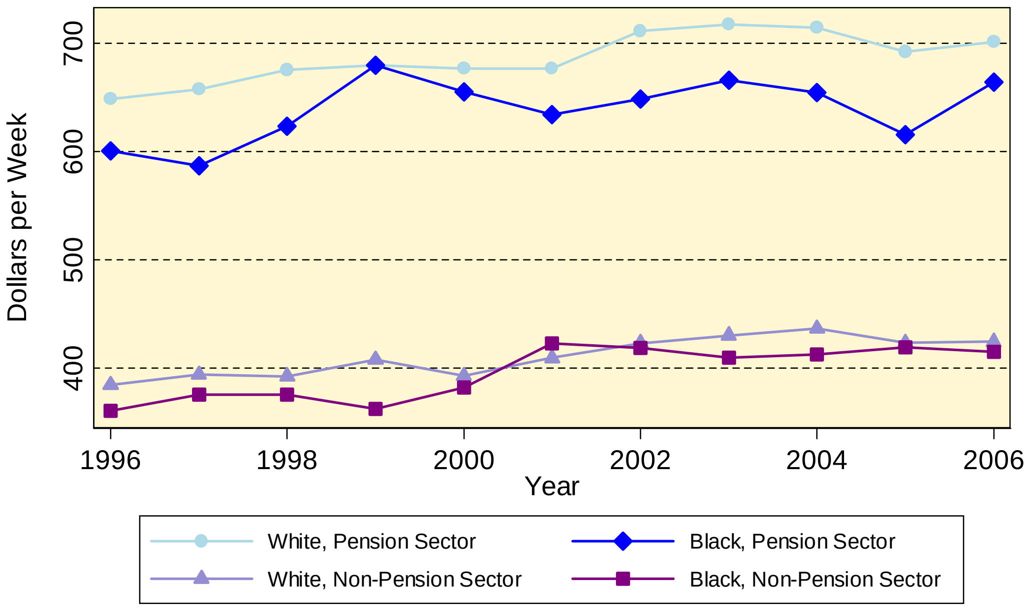
Notes:
Data from Current Population Surveys- March Annual Supplements.
The diagram shows, by calendar year, the median salary of black and white women who receive/do not receive health insurance separately.
We restrict to women who work more than 35 hours per week and ages 25-64.

Figure I.9: Median weekly salary for men with/without pension coverage, 1996 to 2006



Notes:
Data from Current Population Surveys- March Annual Supplements.
The diagram shows, by calendar year, the median salary of black and white men in the pension and non-pension sectors separately.
We restrict to men who work more than 35 hours per week and ages 25-64.

Figure I.10: Median weekly salary for women with/without pension coverage, 1996 to 2006



Notes:
Data from Current Population Surveys- March Annual Supplements.
The diagram shows, by calendar year, the median salary of black and white women in the pension and non-pension sectors separately.
We restrict to women who work more than 35 hours per week and ages 25-64.

APPENDIX J

Chapter Three- Tables

Table J.1: Coefficients and decompositions of race differentials in health insurance, CPS data

	Male Sample		Female Sample	
	(I)	(II)	(I)	(II)
Controls:				
(1) Education, age, region, children, spouse salary	Yes	Yes	Yes	Yes
(2) Add union membership, firm size, occupation, industry, work type (private, public or self-employed)	No	Yes	No	Yes
(A) Combined sample with race dummies				
(3) Black coefficient	−0.0956	−0.2104	0.0384	−0.0394
	(0.0141)	(0.0150)	(0.0150)	(0.0155)
(4) Black marginal effect	−0.0296	−0.0640	0.0142	−0.0147
	(0.0045)	(0.0048)	(0.0055)	(0.0058)
(B) Non-linear decompositions				
(5) Total difference	0.0483	0.0483	−0.0140	−0.0140
(6) Explained by characteristics	0.0186	−0.0085	0.0001	−0.0265

Notes:
Combined dataset from 1996 to 2006 CPS cohorts, full time workers only. In all regressions and decompositions, year dummies are included. Standard errors in parentheses.

Table J.2: Coefficients and decompositions of race differentials in health insurance, male sample of NLSY79 data

	Male Sample					
	(I)	(II)	(III)	(IV)	(V)	(VI)
Controls:						
(1) Education, age, region, children, spouse salary	Yes	Yes	Yes	Yes	Yes	Yes
(2) Add tenure	No	Yes	Yes	No	Yes	Yes
(3) Add standardized AFQT score	No	No	Yes	No	No	Yes
(4) Add union membership, firm size, occupation, industry, work type (private, public or self)	No	No	No	Yes	Yes	Yes
(A) Combined sample with race dummies						
(5) Black coefficient	−0.1370	−0.0771	−0.0049	−0.1374	−0.0839	−0.0203
	(0.0362)	(0.0371)	(0.0407)	(0.0391)	(0.0397)	(0.0433)
(6) Black marginal effect	−0.0301	−0.0156	−0.0010	−0.0272	−0.0154	−0.0037
	(0.0082)	(0.0077)	(0.0081)	(0.0080)	(0.0075)	(0.0078)
(B) Non-linear decompositions						
(7) Total Difference	0.0650	0.0650	0.0650	0.0650	0.0650	0.0650
(8) Explained by Characteristics	0.0294	0.0430	0.0588	0.0313	0.0390	0.0504

Notes:
Combined dataset from 1996 to 2006 NLSY cohorts, full time workers only. In all regressions and decompositions, year dummies are included. Standard errors in parentheses.

Table J.3: Coefficients and decompositions of race differentials in health insurance, female sample of NLSY79 data

	Female Sample					
	(I)	(II)	(III)	(IV)	(V)	(VI)
Controls:						
(1) Education, age, region, children, spouse salary	Yes	Yes	Yes	Yes	Yes	Yes
(2) Add tenure	No	Yes	Yes	No	Yes	Yes
(3) Add standardized AFQT score	No	No	Yes	No	No	Yes
(4) Add union membership, firm size, occupation, industry, work type (private, public or self)	No	No	No	Yes	Yes	Yes
(A) Combined sample with race dummies						
(5) Black coefficient	0.0027	0.0064	0.1389	−0.0116	−0.0035	0.1162
	(0.0398)	(0.0411)	(0.0451)	(0.0423)	(0.0432)	(0.0471)
(6) Black marginal effect	0.0006	0.0012	0.0259	−0.0023	−0.0006	0.0202
	(0.0086)	(0.0080)	(0.0082)	(0.0084)	(0.0078)	(0.0080)
(B) Non-linear decompositions						
(7) Total Difference	0.0134	0.0134	0.0134	0.0134	0.0134	0.0134
(8) Explained by Characteristics	0.0145	0.0129	0.0360	0.0139	0.0133	0.0327

Notes:
Combined dataset from 1996 to 2006 NLSY cohorts, full time workers only. In all regressions and decompositions, year dummies are included.
Standard errors in parentheses.

Table J.4: Coefficients and decompositions of race differentials in pensions, CPS data

	Male Sample		Female Sample	
	(I)	(II)	(I)	(II)
Controls:				
(1) Education, age, region, children, spouse salary	Yes	Yes	Yes	Yes
(2) Add union membership, firm size, occupation, industry, work type (private, public or self-employed)	No	Yes	No	Yes
(A) Combined sample with race dummies				
(3) Black coefficient	−0.0487	−0.1916	−0.0026	−0.1357
	(0.0136)	(0.0147)	(0.0150)	(0.0160)
(4) Black marginal effect	−0.0177	−0.0697	−0.0010	−0.0502
	(0.0050)	(0.0055)	(0.0055)	(0.0061)
(B) Non-linear decompositions				
(5) Total Difference	0.0442	0.0442	0.0241	0.0241
(6) Explained by Characteristics	0.0278	−0.0108	0.0237	−0.0158

Notes:
Combined dataset from 1996 to 2006 CPS cohorts, full time workers only. In all regressions and decompositions, year dummies are included. Standard errors in parentheses.

Table J.5: Coefficients and decompositions of race differentials in pensions, male sample of NLSY79 data

	Male Sample					
	(I)	(II)	(III)	(IV)	(V)	(VI)
Controls:						
(1) Education, age, region, children, spouse salary	Yes	Yes	Yes	Yes	Yes	Yes
(2) Add union membership, firm size, occupation, industry, work type (private, public or self)	No	No	No	Yes	Yes	Yes
(3) Add tenure	No	Yes	Yes	No	Yes	Yes
(4) Add standardized AFQT score	No	No	Yes	No	No	Yes
(A) Combined sample with race dummies						
(5) Black coefficient	−0.0695	−0.0127	0.0675	−0.1118	−0.0639	0.0215
	(0.0327)	(0.0333)	(0.0366)	(0.0355)	(0.0359)	(0.0392)
(6) Black marginal effect	−0.0219	−0.0039	0.0204	−0.0333	−0.0186	0.0062
	(0.0104)	(0.0102)	(0.0110)	(0.0108)	(0.0106)	(0.0112)
(B) Non-linear decompositions						
(7) Total Difference	0.0687	0.0687	0.0687	0.0687	0.0687	0.0687
(8) Explained by Characteristics	0.0440	0.0588	0.0785	0.0370	0.0459	0.0647

Notes:
Combined dataset from 1996 to 2006 NLSY cohorts. In all regressions and decompositions, year dummies are included.
Standard errors in parentheses.

Table J.6: Coefficients and decompositions of race differentials in pensions, female sample of NLSY79 data

	Female Sample					
	(I)	(II)	(III)	(IV)	(V)	(VI)
Controls:						
(1) Education, age, region, children, spouse salary	Yes	Yes	Yes	Yes	Yes	Yes
(2) Add union membership, firm size, occupation, industry, work type (private, public or self)	No	No	No	Yes	Yes	Yes
(3) Add tenure	No	Yes	Yes	No	Yes	Yes
(4) Add standardized AFQT score	No	No	Yes	No	No	Yes
(A) Combined sample with race dummies						
(5) Black coefficient	0.1215	0.1235	0.2378	0.1012	0.1080	0.2154
	(0.0358)	(0.0367)	(0.0402)	(0.0382)	(0.0389)	(0.0422)
(6) Black marginal effect	0.0367	0.0359	0.0678	0.0289	0.0299	0.0585
	(0.0107)	(0.0105)	(0.0111)	(0.0108)	(0.0106)	(0.0111)
(B) Non-linear decompositions						
(7) Total Difference	−0.0198	−0.0198	−0.0198	−0.0198	−0.0198	−0.0198
(8) Explained by Characteristics	0.0189	0.0155	0.0415	0.0098	0.0089	0.0315

Notes:
Combined dataset from 1996 to 2006 NLSY cohorts. In all regressions and decompositions, year dummies are included.
Standard errors in parentheses.

Table J.7: Racial differences in wages and total compensation

	Male Sample		Female Sample	
	CPS	NLSY79	CPS	NLSY79
(A) Wages				
(1) Mean of hourly white wages	23.3933	23.1765	17.3590	16.6749
	(0.0494)	(0.2655)	(0.0984)	(0.1631)
(2) Mean of hourly black wages	18.1075	15.9995	15.5545	14.4160
	(0.1079)	(0.2042)	(0.0853)	(0.1674)
(3) Difference, (1)-(2)	5.2859	7.1770	1.8044	2.2588
	(0.1187)	(0.3350)	(0.1302)	(0.2338)
(4) Percentage difference, $\frac{(1)-(2)}{(2)} \times 100$	29.1916	44.8573	11.6007	15.6689
(B) Total compensation				
(5) Mean of hourly white compensation	25.4480	26.3577	18.8628	18.8415
	(0.0522)	(0.3041)	(0.1047)	(0.1868)
(6) Mean of hourly black compensation	19.6977	17.9424	16.9501	16.1368
	(0.1141)	(0.2353)	(0.0905)	(0.1892)
(7) Difference, (3)-(4)	5.7503	8.4153	1.9127	2.7047
	(0.1255)	(0.3845)	(0.1384)	(0.2659)
(8) Percentage difference, $\frac{(3)-(4)}{(4)} \times 100$	29.1930	46.9016	11.2843	16.7612

Notes:
Combined dataset from 1996 to 2006 CPS cohorts and 1996 to 2006 for the NLSY79 dataset, full time workers only. Standard errors in parentheses.

APPENDIX K

Chapter Three: Supplementary Tables

Table K.1: Racial differences and decompositions in employer provided health insurance, by year, using male sample of CPS data

	Male Sample				
Year	White (I)	Black (II)	Diff A (III)	Diff B (IV)	Charac (V)
1996	0.7055	0.6410	0.0645	0.0579	−0.0055
	(0.0029)	(0.0107)	(0.0111)		(0.0032)
1997	0.7019	0.6452	0.0567	0.0692	0.0076
	(0.0030)	(0.0107)	(0.0111)		(0.0030)
1998	0.7035	0.6605	0.0430	0.0464	−0.0111
	(0.0029)	(0.0104)	(0.0108)		(0.0034)
1999	0.7099	0.6703	0.0396	0.0480	−0.0067
	(0.0030)	(0.0105)	(0.0109)		(0.0033)
2000	0.7062	0.6635	0.0427	0.0425	−0.0182
	(0.0029)	(0.0098)	(0.0102)		(0.0030)
2001	0.7080	0.6485	0.0595	0.0688	−0.0090
	(0.0030)	(0.0103)	(0.0107)		(0.0034)
2002	0.7099	0.6618	0.0481	0.0546	−0.0161
	(0.0023)	(0.0072)	(0.0075)		(0.0030)
2003	0.6972	0.6434	0.0538	0.0468	−0.0139
	(0.0024)	(0.0078)	(0.0082)		(0.0034)
2004	0.6855	0.6568	0.0287	0.0191	−0.0088
	(0.0024)	(0.0077)	(0.0081)		(0.0032)
2005	0.6766	0.6438	0.0328	0.0468	−0.0016
	(0.0025)	(0.0080)	(0.0084)		(0.0034)
2006	0.6721	0.6483	0.0239	0.0480	0.0000
	(0.0025)	(0.0079)	(0.0082)		(0.0033)

Notes: Full time workers only.
Diff A (III) is difference in raw means, (I)-(II).
Diff B (IV) is difference for sub-sample used for decomposition analysis.
Characteristics (V) is part of the diff B (IV) which is explained by racial differences in characteristics.
Standard errors in parentheses.

Table K.2: Racial differences and decompositions in employer provided health insurance, by year, using female sample of CPS data

	Female Sample				
Year	White	Black	Diff A	Diff B	Charac
	(I)	(II)	(III)	(IV)	(V)
1996	0.6516	0.6625	−0.0108	−0.0157	−0.0358
	(0.0036)	(0.0097)	(0.0103)		(0.0048)
1997	0.6470	0.6674	−0.0204	0.0028	−0.0367
	(0.0037)	(0.0099)	(0.0105)		(0.0046)
1998	0.6449	0.6713	−0.0263	−0.0422	−0.0382
	(0.0036)	(0.0095)	(0.0101)		(0.0048)
1999	0.6447	0.6648	−0.0201	−0.0480	−0.0223
	(0.0037)	(0.0097)	(0.0104)		(0.0048)
2000	0.6495	0.6516	−0.0022	0.0084	−0.0261
	(0.0035)	(0.0091)	(0.0098)		(0.0042)
2001	0.6488	0.6683	−0.0195	−0.0181	−0.0226
	(0.0037)	(0.0095)	(0.0102)		(0.0047)
2002	0.6469	0.6828	−0.0359	−0.0213	−0.0241
	(0.0028)	(0.0065)	(0.0071)		(0.0039)
2003	0.6369	0.6655	−0.0286	−0.0132	−0.0194
	(0.0030)	(0.0070)	(0.0076)		(0.0042)
2004	0.6379	0.6584	−0.0205	0.0271	−0.0283
	(0.0029)	(0.0070)	(0.0076)		(0.0045)
2005	0.6366	0.6649	−0.0283	0.0069	−0.0245
	(0.0030)	(0.0073)	(0.0079)		(0.0043)
2006	0.6231	0.6525	−0.0293	−0.0421	−0.0274
	(0.0030)	(0.0071)	(0.0077)		(0.0043)

Notes: Full time workers only.
Diff A (III) is difference in raw means, (I)-(II).
Diff B (IV) is difference for sub-sample used for decomposition analysis.
Characteristics (V) is part of the diff B (IV) which is explained by racial differences in characteristics.
Standard errors in parentheses.

Table K.3: Racial differences and decompositions in employer provided health insurance, by year, using male sample of NLSY79 data

	Male Sample				
Year	White	Black	Diff A	Diff B	Charac
	(I)	(II)	(III)	(IV)	(V)
1996	0.8285	0.7571	0.0713	0.0867	0.0549
	(0.0085)	(0.0158)	(0.0180)		(0.0123)
1998	0.8453	0.8057	0.0396	0.0550	0.0522
	(0.0084)	(0.0145)	(0.0168)		(0.0128)
2000	0.8624	0.7961	0.0663	0.0631	0.0431
	(0.0076)	(0.0136)	(0.0156)		(0.0112)
2002	0.8806	0.8035	0.0771	0.0808	0.0334
	(0.0076)	(0.0141)	(0.0160)		(0.0110)
2004	0.8875	0.8305	0.0570	0.0155	0.0210
	(0.0076)	(0.0141)	(0.0160)		(0.0167)

Notes: Full time workers only.
Diff A (III) is difference in raw means, (I)-(II).
Diff B (IV) is difference for sub-sample used for decomposition analysis.
Characteristics (V) is part of the diff B (IV) which is explained by racial differences in characteristics.
Standard errors in parentheses.

Table K.4: Racial differences and decompositions in employer provided health insurance, by year, using female sample of NLSY79 data

	Female Sample				
Year	White	Black	Diff A	Diff B	Charac
	(I)	(II)	(III)	(IV)	(V)
1996	0.8462	0.8537	−0.0075	0.0062	0.0370
	(0.0099)	(0.0138)	(0.0170)		(0.0127)
1998	0.8810	0.8447	0.0364	0.0348	0.0187
	(0.0090)	(0.0139)	(0.0166)		(0.0126)
2000	0.8430	0.8369	0.0061	−0.0006	0.0331
	(0.0093)	(0.0127)	(0.0157)		(0.0124)
2002	0.8628	0.8790	−0.0162	−0.0140	0.0569
	(0.0091)	(0.0119)	(0.0150)		(0.0129)
2004	0.8771	0.8593	0.0178	−0.0045	0.0248
	(0.0089)	(0.0129)	(0.0156)		(0.0163)

Notes: Full time workers only.
Diff A (III) is difference in raw means, (I)-(II).
Diff B (IV) is difference for sub-sample used for decomposition analysis.
Characteristics (V) is part of the diff B (IV) which is explained by racial differences in characteristics.
Standard errors in parentheses.

Table K.5: Racial differences and decompositions in pension coverage, by year, using the male sample of CPS data

	Male Sample				
Year	White	Black	Diff A	Diff B	Charac
	(I)	(II)	(III)	(IV)	(V)
1996	0.5535	0.5398	0.0137	0.0179	−0.0182
	(0.0032)	(0.0111)	(0.0115)		(0.0035)
1997	0.5620	0.5220	0.0400	0.0643	−0.0036
	(0.0033)	(0.0112)	(0.0116)		(0.0031)
1998	0.5659	0.5274	0.0385	0.0355	−0.0113
	(0.0032)	(0.0110)	(0.0115)		(0.0036)
1999	0.5844	0.5345	0.0499	0.0578	−0.0024
	(0.0032)	(0.0111)	(0.0116)		(0.0035)
2000	0.5771	0.5514	0.0258	0.0220	−0.0252
	(0.0031)	(0.0103)	(0.0108)		(0.0031)
2001	0.5792	0.5130	0.0662	0.0906	−0.0056
	(0.0033)	(0.0108)	(0.0113)		(0.0036)
2002	0.5837	0.5261	0.0576	0.0506	−0.0091
	(0.0025)	(0.0076)	(0.0080)		(0.0033)
2003	0.5649	0.5153	0.0495	0.0272	−0.0167
	(0.0026)	(0.0082)	(0.0086)		(0.0036)
2004	0.5659	0.5229	0.0429	0.0317	−0.0039
	(0.0026)	(0.0081)	(0.0085)		(0.0034)
2005	0.5593	0.5199	0.0395	0.0418	−0.0134
	(0.0027)	(0.0083)	(0.0088)		(0.0034)
2006	0.5456	0.5046	0.0410	0.0549	0.0005
	(0.0026)	(0.0082)	(0.0086)		(0.0034)

Notes: Full time workers only.
Diff A (III) is difference in raw means, (I)-(II).
Diff B (IV) is difference for sub-sample used for decomposition analysis.
Characteristics (V) is part of the diff B (IV) which is explained by racial differences in characteristics.
Standard errors in parentheses.

Table K.6: Racial differences and decompositions in pension coverage, by year, using the female sample of CPS data

	Female Sample				
Year	White	Black	Diff A	Diff B	Charac
	(I)	(II)	(III)	(IV)	(V)
1996	0.5448	0.5344	0.0104	−0.0021	−0.0345
	(0.0038)	(0.0102)	(0.0109)		(0.0046)
1997	0.5565	0.5316	0.0249	0.0159	−0.0225
	(0.0038)	(0.0104)	(0.0111)		(0.0044)
1998	0.5573	0.5280	0.0293	0.0035	−0.0311
	(0.0037)	(0.0101)	(0.0107)		(0.0047)
1999	0.5718	0.5411	0.0307	0.0058	−0.0200
	(0.0038)	(0.0103)	(0.0109)		(0.0044)
2000	0.5729	0.5277	0.0452	0.0249	−0.0203
	(0.0036)	(0.0096)	(0.0102)		(0.0039)
2001	0.5795	0.5228	0.0567	0.0381	−0.0179
	(0.0038)	(0.0101)	(0.0108)		(0.0045)
2002	0.5738	0.5311	0.0427	0.0496	0.0007
	(0.0029)	(0.0070)	(0.0076)		(0.0038)
2003	0.5609	0.5148	0.0461	0.0196	−0.0091
	(0.0031)	(0.0074)	(0.0080)		(0.0040)
2004	0.5676	0.5222	0.0454	0.0236	−0.0222
	(0.0030)	(0.0074)	(0.0080)		(0.0044)
2005	0.5711	0.5430	0.0281	0.0212	−0.0103
	(0.0031)	(0.0077)	(0.0083)		(0.0041)
2006	0.5532	0.5049	0.0483	0.0554	−0.0053
	(0.0031)	(0.0074)	(0.0080)		(0.0040)

Notes: Full time workers only.
Diff A (III) is difference in raw means, (I)-(II).
Diff B (IV) is difference for sub-sample used for decomposition analysis.
Characteristics (V) is part of the diff B (IV) which is explained by racial differences in characteristics.
Standard errors in parentheses.

Table K.7: Racial differences and decompositions in pension coverage, by year, using the male sample of NLSY79 data

	Male Sample				
Year	White	Black	Diff A	Diff B	Charac
	(I)	(II)	(III)	(IV)	(V)
1996	0.6733	0.6377	0.0356	0.0584	0.0400
	(0.0106)	(0.0177)	(0.0207)		(0.0134)
1998	0.7355	0.6721	0.0635	0.0833	0.0748
	(0.0102)	(0.0173)	(0.0201)		(0.0140)
2000	0.7526	0.6816	0.0711	0.0780	0.0461
	(0.0095)	(0.0158)	(0.0184)		(0.0129)
2002	0.7937	0.7009	0.0928	0.0916	0.0766
	(0.0094)	(0.0162)	(0.0188)		(0.0131)
2004	0.7910	0.7232	0.0679	0.0168	0.0140
	(0.0097)	(0.0168)	(0.0194)		(0.0159)

Notes: Full time workers only.
Diff A (III) is difference in raw means, (I)-(II).
Diff B (IV) is difference for sub-sample used for decomposition analysis.
Characteristics (V) is part of the diff B (IV) which is explained by racial differences in characteristics.
Standard errors in parentheses.

Table K.8: Racial differences and decompositions in pension coverage, by year, using the female male sample of NLSY79 data

	Female Sample				
Year	White (I)	Black (II)	Diff A (III)	Diff B (IV)	Charac (V)
1996	0.7229	0.7485	−0.0255	−0.0200	0.0369
	(0.0122)	(0.0170)	(0.0209)		(0.0142)
1998	0.7465	0.7751	−0.0286	−0.0314	0.0312
	(0.0121)	(0.0161)	(0.0201)		(0.0149)
2000	0.7414	0.7447	−0.0033	−0.0131	0.0240
	(0.0112)	(0.0150)	(0.0187)		(0.0135)
2002	0.7769	0.7939	−0.0170	−0.0177	0.0521
	(0.0110)	(0.0148)	(0.0184)		(0.0140)
2004	0.8032	0.8128	−0.0096	−0.0447	0.0351
	(0.0108)	(0.0144)	(0.0180)		(0.0187)

Notes: Full time workers only.
Diff A (III) is difference in raw means, (I)-(II).
Diff B (IV) is difference for sub-sample used for decomposition analysis.
Characteristics (V) is part of the diff B (IV) which is explained by racial differences in characteristics.
Standard errors in parentheses.

Table K.9: Racial differences in wages and total compensation, male samples from the CPS

	Wages				Total Compensation			
Year	White	Black	Difference	% Diff	White	Black	Difference	% Diff
	(I)	(II)	(III)	(IV)	(I)	(II)	(III)	(IV)
1996	917	713	204	28.55	1014	789	225	28.53
	(7.26)	(18.76)	(20.11)		(7.72)	(20.08)	(21.51)	
1997	928	720	207	28.76	1025	795	230	28.97
	(6.70)	(16.40)	(17.72)		(7.18)	(17.69)	(19.10)	
1998	946	701	244	34.83	1034	769	265.57	34.55
	(7.02)	(10.87)	(12.93)		(7.47)	(11.92)	(14.07)	
1999	964	736	228	30.95	1051	802	248.78	31.01
	(6.83)	(14.73)	(16.24)		(7.28)	(15.54)	(17.16)	
2000	936	746	190	25.40	1020	812	207.73	25.58
	(5.69)	(11.48)	(12.82)		(6.10)	(12.25)	(13.69)	
2001	993	762	231	30.37	1078	824	254.05	30.82
	(10.19)	(14.25)	(17.52)		(10.77)	(15.13)	(18.57)	
2002	1060	771	290	37.60	1152	837	314.80	37.60
	(6.52)	(10.09)	(12.02)		(6.90)	(10.76)	(12.78)	
2003	1051	802	249	31.04	1146	871	275.19	31.59
	(9.16)	(15.42)	(17.93)		(9.39)	(16.06)	(18.60)	
2004	1031	766	265	34.52	1129	840	289.13	34.42
	(6.35)	(10.41)	(12.19)		(6.66)	(11.19)	(13.02)	

Notes: Full time workers only, wages are mean weekly wages
Difference (III)=(I)-(II), % Diff (IV)=$\frac{(I)-(II)}{(II)} \times 100$
Standard errors in parentheses.

Table K.10: Racial differences in wages and total compensation, female samples from the CPS

	Wages				Total Compensation			
Year	White	Black	Difference	% Diff	White	Black	Difference	% Diff
	(I)	(II)	(III)	(IV)	(I)	(II)	(III)	(IV)
1996	653	557	96	17.17	724	622	102	16.42
	(28.64)	(8.59)	(29.90)		(30.53)	(9.40)	(31.94)	
1997	637	599	38	6.34	704	662	42	6.34
	(4.79)	(14.99)	(15.74)		(5.13)	(15.66)	(16.48)	
1998	653	585	68	11.55	714	643	71	11.03
	(5.28)	(8.35)	(9.88)		(5.61)	(9.01)	(10.62)	
1999	676	600.38	75	12.57	736	658	78	11.80
	(5.67)	(11.34)	(12.67)		(6.02)	(12.26)	(13.66)	
2000	658	609.72	49	8.00	716	664.62	52	7.77
	(4.29)	(9.31)	(10.25)		(4.56)	(9.96)	(10.95)	
2001	672	615	57	9.19	729	669	60	9.01
	(4.56)	(10.92)	(11.83)		(4.83)	(11.55)	(12.52)	
2002	714	640	73	11.47	776	698	78	11.19
	(4.09)	(8.15)	(9.12)		(4.34)	(8.64)	(9.67)	
2003	715	661	54	8.22	780	723	58	7.97
	(4.51)	(9.38)	(10.40)		(4.76)	(9.95)	(11.03)	
2004	727	653	74	11.31	797	717	81	11.29
	(4.96)	(10.09)	(11.24)		(5.24)	(10.67)	(11.89)	

Notes: Full time workers only, wages are mean weekly wages
Difference (III)=(I)-(II), % Diff (IV)=$\frac{(I)-(II)}{(II)} \times 100$
Standard errors in parentheses.

Table K.11: Racial differences in wages and total compensation, male samples from the NLSY79

	Wages				Total Compensation			
Year	White	Black	Diff	% Diff	White	Black	Diff	% Diff
	(I)	(II)	(III)	(IV)	(I)	(II)	(III)	(IV)
1996	20.8565	14.7360	6.1205	41.5345	23.6275	16.4872	7.1403	43.3084
	(0.4196)	(0.3740)	(0.5621)		(0.4564)	(0.4272)	(0.6252)	
1998	22.5473	15.8326	6.7148	42.4111	25.3373	17.5640	7.7733	44.2568
	(0.5360)	(0.4949)	(0.7295)		(0.6148)	(0.5587)	(0.8307)	
2000	23.0369	15.9455	7.0914	44.4723	25.7697	17.5927	8.1770	46.4798
	(0.3371)	(0.3437)	(0.4814)		(0.3835)	(0.3847)	(0.5432)	
2002	24.6392	16.5486	8.0906	48.8898	27.5046	18.2234	9.2812	50.9304
	(0.4867)	(0.5417)	(0.7282)		(0.5472)	(0.5799)	(0.7973)	
2004	24.2569	16.9292	7.3276	43.2838	27.4996	18.8564	8.6432	45.8371
	(0.4401)	(0.6903)	(0.8187)		(0.5077)	(0.7221)	(0.8827)	

Notes: Full time workers only, wages are mean weekly wages

Diff (III)=(I)-(II), % Diff (IV)=$\frac{(I)-(II)}{(II)} \times 100$

Standard errors in parentheses.

Table K.12: Racial differences in wages and total compensation, female samples from the NLSY79

	Wages				Total Compensation			
Year	White	Black	Diff	% Diff	White	Black	Diff	% Diff
	(I)	(II)	(III)	(IV)	(I)	(II)	(III)	(IV)
1996	15.6582	13.6651	1.9930	14.5848	17.7042	15.3243	2.3800	15.5307
	(0.3036)	(0.3442)	(0.4590)		(0.3414)	(0.3788)	(0.5100)	
1998	15.9062	13.0875	2.8186	21.5367	17.7637	14.5567	3.2070	22.0313
	(0.2751)	(0.2481)	(0.3705)		(0.3087)	(0.2830)	(0.4188)	
2000	17.1058	14.0067	3.0991	22.1259	19.0884	15.4976	3.5907	23.1695
	(0.3039)	(0.2828)	(0.4151)		(0.3428)	(0.3155)	(0.4659)	
2002	17.2926	14.4827	2.8099	19.4020	19.2300	16.0418	3.1881	19.8739
	(0.2978)	(0.3033)	(0.4251)		(0.3328)	(0.3400)	(0.4758)	
2004	17.5790	14.2129	3.3662	23.6838	19.8113	15.9450	3.8662	24.2472
	(0.5599)	(0.3296)	(0.6497)		(0.6428)	(0.3712)	(0.7423)	

Notes: Full time workers only, wages are mean weekly wages
Diff (III)=(I)-(II), % Diff (IV)=$\frac{(I)-(II)}{(II)} \times 100$
Standard errors in parentheses.

Table K.13: Simulations for Wages, White Characteristics, Male Sample, CPS Data

Year	**White Characteristics**			
	$(\widehat{\delta}_w, \widehat{\beta}_w, \widehat{\omega}_w)$	$(\widehat{\delta}_b, \widehat{\beta}_w, \widehat{\omega}_w)$	$(\widehat{\delta}_w, \widehat{\beta}_b, \widehat{\omega}_b)$	$(\widehat{\delta}_b, \widehat{\beta}_b, \widehat{\omega}_b)$
1996	6.751	6.759	6.563	6.574
1998	6.782	6.795	6.586	6.569
2000	6.818	7.164	6.704	6.976
2002	6.878	6.910	6.694	6.737
2004	6.845	6.861	6.670	6.688
2006	6.845	6.852	6.481	6.472

Notes:
1. Data from CPS, March Annual Supplements from 1996 to 2006.
2. The sample consists of men who are married, currently working, who work in the private sector, are between the ages 24 and 65 and who work at least 35 hours per week.
3. The coefficients are from the switching regression model, estimated using the Heckman two stage procedure.

Table K.14: Simulations for Wages, Black Characteristics, Male Sample, CPS Data

Year	Black Characteristics			
	$(\widehat{\delta}_w, \widehat{\beta}_w, \widehat{\omega}_w)$	$(\widehat{\delta}_b, \widehat{\beta}_w, \widehat{\omega}_w)$	$(\widehat{\delta}_w, \widehat{\beta}_b, \widehat{\omega}_b)$	$(\widehat{\delta}_b, \widehat{\beta}_b, \widehat{\omega}_b)$
1996	6.634	6.642	6.434	6.443
1998	6.650	6.663	6.484	6.479
2000	6.718	7.061	6.618	6.871
2002	6.741	6.758	6.575	6.591
2004	6.751	6.765	6.585	6.596
2006	6.707	6.718	6.548	6.565

Notes:
1. Data from CPS, March Annual Supplements from 1996 to 2006.
2. The sample consists of men who are married, currently working, who work in the private sector, are between the ages 24 and 65 and who work at least 35 hours per week.
3. The coefficients are from the switching regression model, estimated using the Heckman two stage procedure.

Table K.15: Simulations for Wages, White Characteristics, Female Sample, CPS Data

Year	White Characteristics			
	$(\widehat{\delta}_w, \widehat{\beta}_w, \widehat{\omega}_w)$	$(\widehat{\delta}_b, \widehat{\beta}_w, \widehat{\omega}_w)$	$(\widehat{\delta}_w, \widehat{\beta}_b, \widehat{\omega}_b)$	$(\widehat{\delta}_b, \widehat{\beta}_b, \widehat{\omega}_b)$
1996	6.288	6.282	6.241	6.250
1998	6.353	6.369	6.388	6.399
2000	6.377	6.373	6.378	6.367
2002	6.448	6.474	6.418	6.473
2004	6.454	6.514	6.365	6.426
2006	6.409	6.411	6.375	6.393

Notes:
1. Data from CPS, March Annual Supplements from 1996 to 2006.
2. The sample consists of men who are married, currently working, who work in the private sector, are between the ages 24 and 65 and who work at least 35 hours per week.
3. The coefficients are from the switching regression model, estimated using the Heckman two stage procedure.

Table K.16: Simulations for Wages, Black Characteristics, Female Sample, CPS Data

Year	**Black Characteristics**			
	$(\widehat{\delta}_w, \widehat{\beta}_w, \widehat{\omega}_w)$	$(\widehat{\delta}_b, \widehat{\beta}_w, \widehat{\omega}_w)$	$(\widehat{\delta}_w, \widehat{\beta}_b, \widehat{\omega}_b)$	$(\widehat{\delta}_b, \widehat{\beta}_b, \widehat{\omega}_b)$
1996	6.221	6.239	6.248	6.235
1998	6.282	6.281	6.264	6.261
2000	6.280	6.261	6.303	6.282
2002	6.382	6.390	6.325	6.328
2004	6.348	6.369	6.264	6.272
2006	6.391	6.382	6.433	6.423

Notes:
1. Data from CPS, March Annual Supplements from 1996 to 2006.
2. The sample consists of men who are married, currently working, who work in the private
sector, are between the ages 24 and 65 and who work at least
35 hours per week.
3. The coefficients are from the switching regression model,
estimated using the Heckman two stage procedure.