August 2003

What Is Known about Testing for Discrimination: Lessons Learned by Comparing across Different Markets

Stephen L. Ross
University of Connecticut

Department of Economics Working Paper Series

What Is Known about Testing for Discrimination: Lessons Learned by Comparing across Different Markets

Stephen L. Ross
University of Connecticut

Working Paper 2003-21

August 2003

341 Mansfield Road, Unit 1063
Storrs, CT 06269–1063
Phone: (860) 486–3022
Fax: (860) 486–4463
http://www.econ.uconn.edu/

Abstract

The paper provides a fairly comprehensive examination of recent empirical work on discrimination within economics. The three major analytical approaches considered are traditional regression analysis of outcomes, paired testing or audits, and finally analysis of performance where higher group performance suggests that a group has been treated disfavorably. The review covers research in the labor, credit, and consumption markets, as well as recent studies of discrimination within the legal system. The review suggests that the validity of interpreting observed racial differences as discrimination depends heavily on whether the analysis is based on a sample that is representative of a population of individuals or households or based on a sample of market transactions, as well as the analyst?s ability to control for heterogeneity within that sample. Heterogeneous firm behavior and differentiated products, such as those found in labor and housing markets, also can confound empirical analyses of discrimination by confusing the allocation of individuals across firms or products with disparate treatment or by ignoring disparate impacts that might arise based on that allocation.

An earlier version of this paper was sponsored by the National Resource Council of the National Academy and presented to the Panel on Methods for Assessing Discrimination on July 2, 2002. Any opinions, findings, conclusions, or recommendations expressed in this article are those of the author and do not necessarily reflect the views of the panel, the National Academy, or any agency of the U.S. Government.

1. Introduction

This literature review was written to accompany and provide supporting information for a paper called “Detecting Discrimination” that was coauthored with John Yinger and was commissioned by the National Resource Council Panel on Methods for Assessing Discrimination (Ross and Yinger, 2002). The paper provides a fairly comprehensive examination of recent empirical work on discrimination within economics, see Lundberg and Startz (2000) and Ross (In Press) for recent reviews of theoretical work on discrimination in labor and housing markets, respectively. The three major analytical approaches considered are traditional regression analysis of outcomes, paired testing or audits, and finally analysis of performance where higher group performance suggests that a group has been treated disfavorably. The review covers research in the labor, credit (primarily mortgage), and consumption markets (primarily housing), as well as recent studies of discrimination within the legal system. A substantial literature also exists on discrimination in education, see Holzer and Ludwig (2002) for a recent review.

This review provides new insights into the methodological dilemmas facing researchers who study discrimination by comparing the literature both across markets and across methodologies. The across market comparison illustrates two major differences between literatures. First, studies of the labor market tend to use nationally or regionally representative samples of individuals due to the availability of numerous high quality, publicly available data sets that report labor market outcomes while analyses of credit market discrimination, especially in the mortgage market, focus on representative samples of transactions, e.g. mortgage or credit applications. Analyses based on actual transactions often result in more homogenous samples with better controls, but raise issues concerning selection into the sample.

The second major difference arises between the interpretation of observed price differences in traditional studies of the labor and housing markets. Observed wage differences by race, ethnicity or gender are often interpreted as direct evidence of discrimination in the setting of prices while in the housing market racial differences in housing prices have been interpreted as indirect evidence of discrimination in terms of access to neighborhoods that lead to higher prices in predominantly minority neighborhoods. On the surface, these differences in interpretation seem appropriate given the large spatial variation in housing markets, but upon further consideration both markets are differentiated products markets characterized by substantial segregation of the majority and minority across product types, either neighborhood or occupation.

Comparisons across methodologies suggest that the methodologies are complementary under many circumstances. Paired testing directly addresses many of the limitations of regression analysis. Paired testing controls for many variables that might be unobserved in administrative samples by assigned detailed profiles to testers and providing training to eliminate behavioral differences. Testers are also sent to the same establishment or firm and make very similar requests thereby eliminating heterogeneity arising from differences across firms or across product requests. On the other hand, paired testing only tests for disparate treatment and cannot capture disparate impact discrimination. Furthermore, regression analysis may provide a better indication of the overall impact of discrimination on minority groups because such an analysis capture economy wide patterns of outcome differences. Performance analyses of price discrimination, i.e. discrimination in the setting of wages, interest rates, or bail bonds, also complements direct regression analysis because the bias arising from omitted variables tends to bias the estimated coefficients on race, ethnicity or gender in opposite directions for the two methodologies. On the other hand, the review concludes that performance analysis of screening discrimination, such as analyses of mortgage default to test for underwriting discrimination, suffers from a wide variety of biases that are not present in direct regression analysis.

The paper is structured as follows. Each of the next three sections cover the one of the three major analytical approaches with the discussion proceeding from traditional regression analysis to paired testing to performance analysis. Each section contains individual subsections devoted to specific markets where the ordering of the subsections is based upon the relative size and importance of the literature that considers this market. Finally, section five summarizes the major conclusions and lessons learned.

2. Regression Analysis of Group Outcomes

This section reviews empirical papers that use traditional regression analysis to test for across group differences in individual outcomes in a specific market. This test is direct in the sense that outcomes are used as the dependent variable and the estimated coefficient on race, gender or some other variable to represent a protected class is intended to capture the effect of discrimination in the market. The section is organized by market and presented based on the size of the contemporary literature in order labor, credit, housing, and legal markets. The first major issue arising from these related literatures is the sample frame. Research in the labor market tends to focus on samples of individuals in the economy who might face discrimination while research in credit markets focus on samples of market transactions, such as mortgage applications, in which discrimination might have taken place. The second major difference occurs in how observed price differences between groups are interpreted in labor and housing markets. These differences are interpreted as price discrimination in the labor market while in the housing market they are interpreted as premium that most likely arise due to the exclusion of minorities from certain segments of the market. Both of these markets, however, are complex, differentiated product markets and neither interpretation provides a complete description of the implication of such findings.

2.1. The Labor Market

The largest single area of research on discrimination or racial differences in the labor market is the study of hourly wages or labor market earnings. Substantially smaller literatures examine racial differences in hiring and occupational segregation. The following two subsections discuss analyses of wages and employment in order. For a review of the literature on occupational segregation, see Blau, Simpson, and Anderson (1998).

2.1.1. Wage Regressions

The traditional approach to examining wage discrimination is a simple regression of log earnings or wages on individual worker attributes with the inclusion of a group dummy variable or using a more complex specification where the return of specific attributes vary by group status (Blinder-Oaxaca decomposition). Four major nationally represented data sets have typically been used to estimate these wage equations: the decennial census, the Census Population Survey (CPS), the National Longitudinal Survey of Youth (NLSY), and the Panel Survey of Income Dynamics (PSID). For example, Altonji and Blank (1999) using the 1995 CPS find a 9 percentage point differential for blacks, a 10 point difference for Hispanics, and a 22 percent difference for women after controlling for education, potential experience, region, occupation, industry, and region. Darity, Guilkey, and Winfrey (1996) examine racial differences in wages by gender using the Decenial census and find no unexplained differences for black women, but substantial differences for men. Rodgers and Spriggs (1996) find similar results using the NLSY.

The most generally accepted application of wage and earnings regressions is the examination of changes or trends in the relative economic status of woman and minorities in the labor market. This type of analysis provides strong evidence that the 1964 Civil Rights Act lead to large reductions in racial and gender wage differences, on race see Card and Kruger (1992) and Donohue and Heckman (1991) and on gender, see Leonard (1989). These results suggest that substantial wage discrimination based on race and gender existed prior to the passing of this legislation. Using CPS data, the unexplained African-American wage gap decreased from approximately 40% to 15% between 1964 and 1975, but has not decreased since that time. On the other hand, the unexplained gender gap continually decreased from 60% to about 35% by the early 1990's (Gottschalk, 1997).

Many studies extend this type of analysis by decomposing the wage or earnings gaps based on the changing attributes of the group's population and the changing character of the labor market. Juhn, Murphy, and Pierce (1991), Card and Lemieux (1996), and Chay and Lee (2000) find that the wage gap narrowed post 1975 due to a narrowing of the racial gap in education and experience gap, but that this gain was counteracted by the negative effect of an increase in the returns to education, experience, and unobserved ability. Couch and Daly (2002) provide more recent evidence on the Black-White wage gap suggesting that the gap started narrowing again in the 1990's leading to the smallest gap in history at the end of their sample period in 1998. Alternatively, Card and Lemieux (1996) using the Panel Survey of

Income Dynamics find evidence of a re-emerging racial gap in the wages of younger black women especially college educated black women.

2.1.1.1. Ability and Work Experience

These national level differences in wages are often put forth as evidence of wage discrimination by race, ethnicity or gender. Estimates of wage differentials from the census and the CPS, however, have been criticized as measures of discrimination because these samples do not contain any variables to capture ability or actual experience. O'Neill (1990) and Maxwell (1994) use the NLSY to estimate wage regressions and include the individual's test scores on the Air Force's Qualifying Test (AFQT) and self-reported experience as additional control variables. After simply controlling for years of schooling, industry, and region, O'Neill and Maxwell find unexplained racial differences of between 17 and 20 percentage points. These differences fall to between zero and five percentage points and are statistically insignificant after adding the controls for ability and experience. Later results by Johnson and Neal (JPE 1996), who do not control for actual experience, suggest that most of the effect found by O'Neill and Maxwell arises from including the AFQT rather than actual experience. Similarly, White (1997) controls for the results of a 10 item cognitive skills test called Wordsum using data from the General Social Survey (GSS) and finds that minorities earn a wage premium after controlling for age, sex, parents education, occupational prestige, and religious affiliation. Blank and Altonji (1999) confirm the effect of controlling for ability (AFQT), family background, and experience on the unexplained black-white wage gap, but find that these factors can only explain about 3 percentage points of the 24 point male-female wage gap in the NLSY. They find that occupation, industry and job characteristics can explain another 1.5 percentage points of the gender gap.

The ability of the AFQT to explain racial differences in wages has been challenged on many fronts. Rodgers, Spriggs, and Waller (1997) find a statistically significant wage differential after adjusting the AFQT scores for age and education at the time of taking the exam. Goldsmith, Veum and Darity (1997) also find a statistically significant wage differential after controlling for psychological variables. Finally, Rodgers and Spriggs (1996) argue that the AFQT is racially biased. They estimate a race neutral AFQT score and find that the entire effect of AFQT in explaining racial differences in the mean of AFQT scores.

In order to assess these claims and counter-claims, it is important to distinguish between the two competing issues: the direct impact of discrimination on minority wages and the overall impact of prejudice on minority wages. If we wish to focus on the impact of discrimination, an important question arises that goes far beyond whether the AFQT is a measure of productivity. This question is Aare the racial differences that are captured by the AFQT actually observable by firms when determining wages especially among the young workers and the short time horizons considered in these studies?@ If observable, the AFQT may indeed provide a proxy for the unobserved productivity that is considered by firms when setting wages, and the question of racial bias in the AFQT becomes central to the debate of whether wage discrimination exists in the U.S. labor market. If firms do not observe proxies for ability that capture the information in the AFQT, however, any ability of the AFQT to explain wage differentials simply captures an incentive that employers face to use race as a signal for productivity and statistically discriminate. Naturally, such behavior should not be factored out of wage regressions that are intended to capture discrimination. A similar issue arises for the Wordsum skills test used by White. In this context, the criticisms above concerning the AFQT are really irrelevant; administrative variables that are not observed by firms should not be included in wage or earnings regressions. In fact, the same problem of variables not being observed by the firms almost certainly arises for the psychological variables considered by Goldsmith, Veum and Darity.

In this context, it is worthwhile to revisit Johnson and Neal's analysis. They argue that experience and education may be influenced by wage discrimination where wage discrimination causes a reduction in investment by African-American males and as such should not be included in a model of wages. Therefore, in their analysis, they only include the AFQT as a measure of pre-market ability. As with the inclusion of the AFQT score, the omission of education and experience variables makes little sense if the analysis is intend to measure discrimination in the labor market. Regardless of the reasons behind racial differences in education and experience, firms are clearly allowed to consider these variables when setting wages. Moreover, education and experience differences may have arisen from the choices made by individuals, and so the omission of these variables may overstate racial differences. Therefore, Johnson and Neal's analysis should be considered a measure of the maximum possible effect of prejudice on racial differences in earnings, which they find to be 7 percentage points for young black males. In this context, however, one of the above critiques is quite important. Johnson and Neal control for the age of the individual when he or she took the AFQT, but they did not control for the education of the individual when he or she took the AFQT. While education should not be included in the wage specification, it is certainly important to remove the influence of educational differences on AFQT performance if one is to obtain a measure of pre-market ability. When this correction is made the influence of prejudice on earnings is 11 percentage points.

In addition to attempts to control for experience in wage regressions, many studies examine the return to experience and test for racial or gender differences in this return. On Race, Bratsberg and Terrell (1998) find 10 to 12.5 percentage point lower return for experience for blacks relative to whites over a six year period depending on the methodology used, but the very similar returns to seniority using the NLSY. Alternatively, D'Amico and Maxwell (1994) examine a sample of men from the NLSY who did not continue schooling past high school and estimate a model of earnings in the fifth year following the individual's leaving school while controlling for AFQT score as a measure of ability. They find that return to experience in year five of the labor market is quite similar for whites and blacks. However, they ignore the selection process that leads to a work history with few if any interruptions in work experience. If there is discrimination in hiring, only the highest quality African-Americans will have such a smooth transition. This unobserved high quality of black workers who have an uninterrupted work history may offset the influence of discrimination on wage growth.

Recent work by Altonji and Pierret (2001) on statistical discrimination is directly relevant to this debate. They estimate a model in which they control for the standardized AFQT and education and allow the influence of these variables to evolve with experience and find that the return to education falls over time and the return to AFQT rises over time. They conclude that firms use education as a signal for ability over time as firms gain more direct information about ability the return to education falls and the return to actually ability increases. In fact, the AFQT has no influence on earnings at very low levels of experience since as we discussed earlier the ability differences captured by the AFQT are probably not observable at that time. After controlling for education, ability, and experience in this manner, the average unexplained effect of race on earnings is between 13 and 15 percentage points and highly statistically significant. Altonji and Pierret decompose the racial difference into an initial difference in wages and a difference in wages that grows over time suggesting that a decrease in the racial wage difference over time would represent statistical discrimination where employers use race as a signal for low ability and eventually replace this signal with information on actual ability. However, they find no evidence of a wage differences at low levels of experience and rather find that racial differences in wages grow by between 0.8 and 2.7 percentage points per year either using potential experience or an instrumental variable specification for actual experience where potential experience is used as an instrument. In our opinion, this represents strong evidence of a lower return to years of experience for blacks in a model that does the best job of controlling the return to ability and education over time.

In terms of gender, it is important to focus on samples that have information on actual experience. Estimates using potential experience based on the census and CPS understates the return to experience for both men and women and potentially overstates gender differences in the return to experience due to the higher rate of job interruptions among women, see Altonji and Black (1999) and the papers discussed below. The following studies examine the male-female wage gap while controlling for actual experience. Blau and Kahn (1997) using the PSID find that increases in female experience levels explain approximately one-third to one half of the closing wage gap depending on whether the specification controls for industry and occupation. Wellington (1993) also uses the PSID to examine wage differentials by gender including sample selectivity controls for labor force participation. She finds that 50 percent of the narrowing of the wage gap between 1976 and 1985 is do to changes in work history among women. Filer (1993) works with the NLS-YW/YM (young women and mature women) and estimates separate earnings equations for each occupation. Filer finds that the return to experience is 20 percent higher in a model that controls for actual experience. Of note, Filer finds a substantial black-white gap among women after controlling for the fact that black women spend more time in the labor market than white women on average. Finally, Light and Ureta (1995) examine a model that uses disaggregated dummy variables describing work experience and work interruptions in order to provide non-parametric controls for these factors. Their results show that work timing and interruptions are very important explaining 12 percent of the 40 percentage point gender wage gap that exists after controlling only for experience and near 20 percent of the gap for low levels of experience. They also find rate of loss due to interruptions is smaller for women and they recover more quickly (i.e. have a higher rate of return to experience after such interruptions). They suggest that this difference arises due to the occupational choices made by women.

This leads us to a discussion about what can be learned from these studies about gender differences in the return education. Based on their detailed specification of job experience and tenure, Light and Ureta provide compelling evidence concerning gender differences in the return to experience. They provide standard estimates based on actual and potential experience and find statistically significant gender differences in the return to experience. Their results are quite robust to the use of actual rather than potential experience. Using their more sophisticated work history model, they also find substantial differences in the return to experience. The most compelling result is for workers with at least 13 years of work experience over the 14 year period. The raw gender gap for these workers was 35 percentage points of which over 75 percent of this gap is explained by differences in the return to experience even though these workers had nearly uninterrupted work histories. Again, these differences may be the result of occupational differences where women chose careers where the reward to experience is lower, but the penalty for interruptions is also lower and less persistent.

A number of occupation specific studies, however, suggest that the occupation explanation is not sufficient to account for the lower return to experience. Wood, Corcoran, and Courant (1993) use detailed data on graduates of the University of Michigan Law School to examine the male-female earnings gap. They find that working part-time to care for children over a three year period, which was the average in the female sample, lowered earnings by 17 percent. Only four men worked part time, but they lost 3.4 percent of earnings for every month not worked, which is consistent with the more severe penalty for men found by Light and Ureta. Wood et. al.'s specification controls for occupation by construction, as well as child care, work history, school performance, and job setting measures by regression, but still finds unexplained earnings gap of between 12 and 17 percentage points. Finally, they decompose the earnings gap into first year gap and the change in the gap over the next 15 years. As with other studies, they find that almost all of the gap is due to gender differences in the return to experience. In addition, Kahn and Shere (1988) examine the determinants of salary in professional basketball. They have incredibly detailed information on performance, and the standard problems of controlling for labor supply tend not to arise in this sample. They find that black ball players make 20% less after controlling for performance. They also find evidence that fan attendance is related to the racial composition of basketball teams and suggest that the observed salary discrimination may be driven by customer prejudice.

Finally, for both race and gender, there exist a literature that considers the influence of on the job training and whether this can explain differences in the return to experience. Veum (1996) finds no racial differences in training, and Altonji and Spletzer (1991) find that blacks are more likely to receive training using the NLSHS72 and women are less likely to receive training, but the effects are not large after controlling for labor supply factors. In no cases, however, can differences in on the job training explain racial or gender differences in the return to experience. Both of these studies, however, use the AFQT, which may or may not explain ability differences that are observable by employers.

2.1.1.2. Education

Card and Kruger (1992) use state level information on student-teacher ratios as a measure of school quality and compare the return to education between blacks born and raised in different states and at different times who worked in Northern states during the census years. The focus on northern states controls for regional differences and allow them to treat school quality in the state of an exogenous treatment. They find that black men born in states like Delaware, West Virginia, and Missouri between 1910 and 1939 earned 3 percent higher earnings for each year of education then black men born in states like Mississippi, Georgia, and South Carolina. Examining changes over time, they find that approximately half of the 20 percentage point decline in the black-white wage gap between the 1910-19 and the 1940-49 cohorts of southern blacks can be explained by differences in school quality as measured by state-wide student-teacher ratios.

Given the importance of education quality in explaining changes in the black-white wage gap, it seems quite reasonable to consider the possibility that the current unexplained black-white wage gap might be in part due to differences in the quality of education. For example, Heckman (1998) argues that an interpretation of racial differences in wages from census and CPS data as discrimination is not credible given the solid evidence on the inferior inner city schools attended predominantly by minority children. Altonji and Blank (1999) go further suggesting that the AFQT score's ability to predict racial differences in earnings might arise in part due to the score's ability to detect differences in school quality.

One possible way to gain insight into this issue is to follow Card and Kruger and focus on differences the return to education across groups. Both Card and Lemieux (1996), and Chay and Lee (2000) find an increase in the unexplained black-white wage gap among college educated males, but no such increase among less education black males. One interpretation is that Blacks experienced lower rates of return from education potentially due to poor school quality. This basic argument appears in Altonji and Blank and was first made by Juhn, Murphy and Pierce (1991). This interpretation, however, appears to be a bit of a stretch to us. First, the majority of evidence on poor quality education for blacks concerns primary and secondary, not higher, education, and yet there is no evidence of lower returns to education in high school. So either the poor primary and secondary education causes minorities to attend inferior colleges or lowers the return to higher education, but at the same time has no effect on the actual return to education obtained in those schools. Second, as noted by Card and Lemieux, the pattern differs for black women who lost ground for all education levels. It seems strange to us that the effect of poor quality education is having dramatically different effects on the black-white wage gap by gender. A more reasonable interpretation is that these changes in the black-white wage gap are not explained by racial differences in school quality.

Unfortunately, we lack studies that perform a direct comparison of school quality to wages at the individual level. One might consider linking panel data sets such as the NLSY or PSID to the limited national education data that is available on schools and school districts from the National Center for Education Statistics.

A second important educational issue involves the inability to distinguish between high school graduates and receipt of a general equivalence degree in many national samples, such as the census or the CPS. Cameron and Heckman (1993) examine the pattern of high school completion and GED receipt. They find that 79 percent of blacks are high school certified of which 14 percent hold GED's. Among the white population, 88 percent are high school certified and only 8 percent of those hold GED's. Moreover, using the NYSY, they show that the earnings of GED are comparable to the earnings of high school drop-outs and substantially below the earnings of high school graduates. Since blacks have a much higher rate of GED usage, wage models that simply control for educational attainment and ignore the mechanism through which this level was accomplished are likely to overstate the black-white wage gap. This observation is especially relevant for studies using census or CPS data where the researcher cannot distinguish between GED and high school graduation. Heckman (1998) claims that the GED factor can explain 1 to 2 percentage points of the black-white wage gap or on the order to 10 to 15 percent of the unexplained gap.

Finally, many studies examine the influence of educational content on wages. Brown and Corcoran (1997) use data from the SIPP and the NLSY to examine the return to education content by educational attainment. Their estimations confirm earlier results that years of education cannot explain the gender wage gap, but they find that for college graduates details on course work and major can explain between a 4 and 10 percentage point gap in wages. They do not find a sizable effect of course work for individuals who did not graduate from college, also see Altonji (1995). The inclusion of these variables, however, only increases the explanatory power of the model by 1 or 2 percentage points for the model that controls for job variables and percent female in occupation. Therefore, while college major is very important as a potential causal factor of wage differences by gender, the omission of these variables does not appear to substantially bias estimates of unexplained wage differences in models that already include controls for occupation and job attributes. It should be noted, however, that these variables reduce unexplained wage differences by about 7 percentage points in a model that does not include occupational variables. An alternative interpretation is offered by Brown and Corcoran that after controlling for educational content information on occupation and job attributes explains very little of the male-female wage gap. The final residual wage gap for college graduates is about 10 percentage points for the SIPP and 6 percentage points for the NLS72 in the most complete specifications. Similar results are available from Altonji (1993) and Grogger and Eide (1995).

Paglin and Rufolo (1990) find a much stronger effect by focusing on mathematical ability, which may lead to major choice, rather than the mathematical content of the major. They find that mathematical ability as measured by Graduate Record Examination scores can explain a 20 percentage point male-female wage gap. Altonji and Blank (1999) suggest that this result might be biased due to the selection process associated with taking the GRE. A second possible explanation for the contradiction between Paglin and Rufolo and other studies is that Paglin and Rufolo focus on starting wages and other research discussed in this review suggests that a substantial portion of the wage gap arises due from a differential return to experience. So Paglin and Rufolo's effect may simply represent a larger percentage of a smaller initial gap.

2.1.1.3. Other Issues: Selection into Employment, Occupational Segregation, and Skin Color

Both Maxwell (1994) and Johnson and Neal (1996) consider the possibility that selection into employment biases the estimated wage gap. Specifically, selection into employment shifts up the distribution of workers on unobserved quality relative to the pool of all potential workers. Since employment is lower among African-Americans, the shift in average unobserved quality is larger for employed African-Americans than for employed whites. This selection process potentially biases the racial wage gap downwards. A sample selection correction increases the unexplained wage gap to 17 and 13 percentage points for Maxwell and Johnson and Neal, respectively. Note that unlike the AFQT these unobserved quality factors are observed by the firm by definition because the firm must observe the factor for it to enter into the hiring process. In addition, Juhn and Murphy (1997) using the CPS finds that approximately one-third of the 12 percentage point decline in the racial wage gap between 1969 and 1989 can be explained by decreases in the labor force participation and employment rate of black men. Blau and Beller (1988) and Wellington (1993) examine selection into employment and the male-female wage gap. Blau and Beller find that the 1981 male-female gap in the CPS increases from 30 to 40 percent after controlling for selection. Wellington does not find any evidence that selection influence the male-female gap using the Panel Survey of Income Dynamics.

Blau and Kahn (1997) find that when controlling for occupation, industry, and unionism in the Panel Survey of Income Dynamics the unexplained mail female earnings gap fell from 20 percent to only 12 percent in 1988. Note that a wide array of studies, see the previous section on work experience, examine the types of models used by Blau and Kahn, but most do not examine the effect of occupation on earnings differences. This omission is especially concerning given the implication of that work, especially Light and Ureta (1995), that occupation choice might explain part of the gender differences in return to experience.

Finally, Darity and Mason (1998) provide a detailed summary of the literature on Skin Color and wage differences. They argue that the studies use skin shade/color as a natural experiment in which skin shade provides a treatment that can influence wages and yet has no link to productivity. For example, Johnson, Bienenstock, and Stoloff (1995) compared light and dark skinned black males from the same neighborhoods in Los Angeles using Multi-City Study of Urban Inequality (MCSUI) data and found that having a dark skin increased chances of working by 52% in a model that controlled for education, age, criminal record, English proficiency, parent's education, and cultural background. Similarly, Telles and Murguia (1990) examines data from the 1979 National Chicano Study and finds that 79% or $1,262 of earnings differences between the dark phenotypic group and non-dark Mexican Americans cannot be explained by traditional labor market variables.

A major concern about using skin color as a natural experiment is that skin color may be correlated with specific within group ethnic and cultural differences that in turn are correlated with unobserved productivity variables. What is especially compelling about these two studies is that they provide comparisons that dramatically mitigate problems on non-comparability across skin color. Johnson, Bienenstock and Stoloff explicity control for parent's education, English proficiency, immigration status and many other cultural factors. Telles and Murguia explicitly compare Mexican Americans with different skin color as opposed to a comparison across Hispanics of all national origins with different skin color, which would encompass a much more diverse group. As an example, Darity, Guilkey, and Winfrey's (1996) comparison of black and white Hispanics using Census data is much less compelling because these groups most likely differ on many important factors that may be important to employers.

2.1.2. Employment

The literature on racial and gender differences in employment outcomes and labor force participation is much smaller that the literature on wage differentials. Altonji and Blank (1999) examine labor force participation using the CPS. For the Anglo-Hispanic comparisons, all of the eight percent difference in labor force participation can be attributes to observable across group differences in education, experience, other personal characteristics, and location. On the other hand, for the male-female comparison, none of the 15 percent gap is attributable to differences in across group characteristics. For the black-white comparison, a little more than half of the 8 percent gap is attributable to observable characteristics. Juhn (1992) finds an 8.5 percent gap in annual employment and a 13 percent gap in weekly employment in 1989 using the CPS and strong evidence that the employment gap increased during the 70's and 80's.

Bound and Freeman (1992) examine data from the CPS and find that industrial shifts during the 80's provides an important explanation for the increase in the black-white employment gap during that period. They calculate shifts in demand for whites and black by region and by education by calculating the share of black employees in each industrial category by region and by education and comparing these shares to the employment changes in those industrial categories. They found that demand factors shifted heavily against black workers, especially college educated blacks which not surprisingly also suffered the largest earnings losses during this period. Similarly, Bound and Holzer (1993) examine micro-census data for 52 metropolitan areas finding that demand shifts away from manufacturing explained 40-50 percent of the decline in less-educated black youths in the 1970s. These findings suggest that occupational segregation may play an important role in black-white employment differentials, but it is uncertain whether this segregation is the result of discrimination or other factors. Bound and Freeman also use the NLSY to examine alternative explanations for these employment differentials. They find that criminal activity played an important role explaining differences in employment, but their results do not indicate whether racial differences in employment remain after controlling for past criminal activity.

Juhn (1992) also examines employment using the CPS. She documents falling Labor force participation rates for both white and black males since the early 70's. She concludes that the entire decline in white labor force participation can be explained by declines in wages for this group, but that only half of the decline for black men can be explained by wage declines. She examines entry and exit rates by race. She finds a disproportionately large increase in the likelihood of not working at all during the last year, and she also finds that the increase was much larger for blacks (9 percentage points) than whites (4 percentage points). The fact that the increase in racial differences is concentrated at the very low end of the labor force attachment distribution seems to suggest that this increase has little to do with discrimination. Unfortunately, June does not provide an analysis of racial differences in employment when this extreme category is dropped from the analysis so we do not know whether unexplained employment differences exist in groups with a stronger degree of attachment to the labor market. Also, of interest, black-white differences in employment exit rates has converged to near equality while the gap in entry rates has increased to more than offset the convergence in exit rates.

In a unique study, Holzer and Ihlanfeldt (1998) examine employment outcomes using the Multi-City Study of Urban Inequality (MCSUI) data for the Atlanta, Boston, and Detroit metropolitan areas. They find strong evidence that the likelihood of blacks being hired at a particular establishment depends upon the racial composition of the establishment's customers and conclude that employers in these cities are practicing customer-based discrimination. They regress race of the last hired worker on the self-reported racial composition of customers, as well as occupational dummies at the one digit level, dummy variables representing cognitive and social tasks performed on the job, industry dummies (one digit level), presence of collective bargaining, and geographic locations both within and between the three metropolitan areas. The problem faced by Holzer and Ihlanfeldt is that they have no direct information on the pool of applicants from which the firms were choosing. While it is difficult to disentangle racial differences in supply side factors from unobserved racial differences in employment for many of the studies discussed above, this problem is much more severe for Holzer and Ihlanfeldt because all candidate quality variables are unobserved. They address this problem by controlling for the location, job attribute, and firm characteristic dummy variables under the assumption that a comparison between similar jobs with similar firms in the same location should result in quite similar pools of applicants for both whites and blacks. This idea seems reasonable at least when considering the influence of the customers' racial composition, but might be less compelling if they were only considering the overall level of racial differences in employment.

In examining the labor force participation of women, Blau (1998) finds continually declining gender differences over the last three decades, but that unexplained differentials on the order to 5 to 15 percent remained in 1995 among non-high school drop-outs after controlling for age and education in the CPS. Juhn and Murphy (1997) and Leibowitz and Klerman (1995) examine the labor supply of married females using the CPS, but do not perform a comparison across gender. Of note, Juhn and Murphy find that employment has increased mostly among women with high levels of education arguing that the shifts in employment have been influenced heavily by changes in real wages. Leibowitz and Klerman find that the negative influence of family status on labor supply has declined in magnitude. They also find that black women have higher labor supply than white women after controlling for observable factors that should affect labor supply. In this context, it is difficult to judge whether the limited evidence concerning gender differentials in employment are in part the result of discrimination or whether these differences are entirely do to unobserved supply factors.

2.2. Discrimination in Credit Markets

The literature on discrimination in the mortgage market differs from the labor market literature in two important ways. First, unlike the labor market literature, which focuses heavily on wages and earnings, the literature on mortgage discrimination has historically focused on the decision to provide credit as opposed to the price of credit. This difference may become less important over time because the traditional credit rationing approach to underwriting is in the process of being replaced with systems of risk-based pricing. The second major difference arises because the data available for scholarly studies of the mortgage market often differs dramatically from the type of data available for labor market studies. Almost all major national surveys gather some information on the labor market attributes and outcomes the individual's surveyed, but very few gather any information on mortgage outcomes. At the same time, however, researchers on mortgage lending discrimination have successfully obtained detailed records from firms concerning the attributes and final outcomes for large samples of mortgage applications. The availability of such data for labor market research has been much more limited. The first subsection reviews studies of nationally or regionally representative samples that are comparable to the majority of studies of labor market discrimination. The second and third subsections review studies that use samples of actual loan applications or approved loans with information on the cost of credit, respectively.

2.2.1. Access to Credit: Population Samples

The first set of studies considered here are quite comparable to the large number of labor market studies discussed in the previous subsection. National or regional survey data is used to examine racial differences in access to credit or mortgage market outcomes. These studies tend to examine the link between credit markets and racial differences in homeownership using data on outcomes in the housing market. Duca and Rosenthal (1994) examines racial differences in homeownership using the survey of consumer finance and a self-report variable on whether a household had been unsuccessful in obtaining desire credit. They find that the denied credit variable explained a substantial portion of the racial differences. Gyourko, Linneman and Wachter (1999) and Deng, Ross and Wachter (In Press) estimate a proxy for whether households face a downpayment constraint and find that the effect of facing a downpayment constraint is substantially larger for African Americans than whites. Finally, Charles and Hurst (2002) examine both the transition to homeownership and the success of households in obtaining mortgage credit. They also find that race plays a substantial role in the various decisions and outcomes that lead to homeownership including the mortgage application decision.

Taken together, these studies provide pretty compelling evidence that differential outcomes in credit markets have a substantial market impact on minority homeownership. However, they do not provide much compelling evidence to suggest that discrimination exists in the mortgage market. Gyourko et. al. and Deng et. al. have no information about whether households even applied for credit let alone whether minority applications were differentially denied. Duca and Rosenthal and Charles and Hurst do have information on differential rates of application denial, but do not have access to even the most basic information on the quality of the credit application, such as borrower's capacity to repay or credit history. These last two studies are directly comparable to labor market studies using the Current Population Survey where the wage regressions are only able to control for human capital using years of potential experience, years of education, industry, and occupation, as opposed to information on actual experience, content of education, and actual skills possessed.

2.2.2. Access to Credit: Transaction Samples

On the other hand, there exist a long history of mortgage lending studies that use firm provided data containing both application outcomes and detailed application characteristics. These studies include Black, Schweitzer, and Mandell (1978), King (1980), Schafer and Ladd (1981), and Maddala and Trost (1982). All of these studies found evidence of racial differences in mortgage lending, but none of these studies contained information on the credit history of the borrower. Moreover, a number of the studies were missing other crucial information, such as loan to value ratio, housing expense to income ratio, or information to proxy for the risk associated with lending in a given neighborhood. These data omissions raise the concern that the results cited above might be attributable to omitted variable bias rather than racial discrimination. Munnell, Tootell, Browne and McEneaney (1996) attempted to address these concerns by collecting a sample of loans from the Boston metropolitan with detailed information on the borrower including credit history, loan terms, and unit and neighborhood attributes. They found an eight percentage point racial difference in the likelihood of loan denial after controlling for the key underwriting information. The typical white denial rate was ten percentage points leading to the well-known claim that blacks in Boston were 80 percent more likely to have their mortgage application denied.

There have been many competing claims concerning the results of the Boston Fed study. The major concerns raised involve omitted variables, data errors in control variables, misclassification of withdrawn or counter-offer loans as denials, incorrect specification, and endogeneity between lender underwriting decisions and submitted loan terms, Horne (1997) and Day and Liebowitz (1996).

Alternatively, Carr and Megboluge (1993) and Glennon and Stengel (1994) reanalyzed the Boston Fed data concluding that the results were robust to many of the criticisms. Ross and Yinger (1999a) review all of the criticisms and reanalyses of the Boston Fed data and find that the central result is only sensitive to one of the major complaints. Namely, the inclusion of a self-reported lender variable on whether the application meets the lenders credit guidelines lowers racial differences from 7.7 to 4.1 percentage points, but the result is still statistically significant.

The problem with simply including this variable is that it is endogenous. The loan files do not contain a report that concluded whether the application met guidelines at the time of application, but rather at the time of the Boston Fed study loan officers reviewed the files and gave their opinion about whether the application met the lenders credit history guidelines at the time of the underwriting decision. At the very least, this problem creates a measurement error bias, but more importantly the problem creates an endogeneity bias because the loan officer knows whether the application was denied when answering the meets guidelines question, see Browne and Tootell (19).

Ross and Yinger (1999a) estimate a simultaneous equations model that controls for meets guidelines in the underwriting model and find that racial differences fall to 6.5 percentage points. They find no evidence of omitted underwriting variables, which would be captured as a correlation between the error terms in the underwriting and meets guidelines equations. Rather, they conclude that the effect of controlling for meets guidelines may be the result of either racial bias in the assignment of meets guidelines or across lender differences in the criteria for whether a loan application meets guidelines. The potential role of across lender underwriting differences in creating racial differences in outcomes raises a serious question concerning whether market level underwriting analyses can truly capture disparate treatment discrimination. Analyses of loan applications that are pooled across lenders may capture also disparate impact discrimination or even the effect of legitimate lender variation in guidelines that are motivated by a business purpose.

Ross and Yinger (2002) merge raw HMDA data with the Boston Fed data in order to obtain lender identifies for the Boston Fed sample and to examine whether racial differences in lending can be explained by differences in lender underwriting. Ross and Yinger examine a series of model in which the underwriting weights used by lenders vary based on the characteristics of the lenders applicant pool, such as the average loan to value or debt to income ratio of applications in the pool. They find substantial evidence that attributes of a lender=s applicant pool is related to its underwriting even in a model that controls for both lender fixed effects and allows underwriting weights to vary by the value of other underwriting variables (so for example in the model that allows underwriting weights to vary with the average debt to income ratio of the lender's pool, Ross and Yinger also allow weights to vary with the applicant's actual debt to income ratio). None of the estimated models suggest, however, that these differences can explain the racial underwriting differences in the sample. Ross and Yinger also examine a model for the 10 largest lenders in the sample and allow the weights placed on credit history, loan to value ratio, and debt to income ratio to vary freely across lenders. Again, while substantial differences in underwriting weights were identified, there was no evidence that these differences could explain the racial underwriting differences in the sample.

An alternative approach has been followed by two major regulatory agencies: the Office of the Comptroller of the Currency and the Federal Reserve Board of Governors. These agencies have developed procedures for estimating lender specific underwriting models. Corchane, Nebhut, and Nickerson (2000) presents the results of OCC estimations for ten lenders including the only two lenders for which racial or ethnic differences in lending were identified. Their lender specific analyses have two key features that distinguish them from market level analyses such as the Boston Fed study. First, all of the applications come from a single lender so that the weights place on underwriting variables, such as credit history or loan to value ratio, can vary across lenders. Secondly, they allow the actual specification and variable construction to vary between lenders. For example, some lenders simply include debt to income in their considerations while others only consider debt to income ratio when it exceeds a lender specific threshold. These models can provide quite strong evidence of disparate treatment discrimination when lenders have sufficient minority applications to support statistical analysis.

Using the same ten lenders as Corchane et. al., Blackburn and Vermilyea (In Press) explicitly examine the differences between the estimation results in the individual lender models and a pooled model similar to the Boston Fed analysis. They find that the unexplained racial differences in the pooled model are much larger than the average unexplained racial differences in the lender specific models. The apparent contradiction between Blackburn and Vermilyea's findings and the analysis by Ross and Yinger is actually quite easy to resolve. Further analysis by Blackburn and Vermilyea suggests that the elimination of racial differences with the lender specific models arises almost entirely from the move to lender specific underwriting variables rather than from the restriction that underwriting weights be the same across lenders. Ross and Yinger could only relax the assumption of equal weights across lenders.

The total body of evidence seems to suggest that the Boston Fed studies as well as the market level underwriting studies that preceded it have not isolated disparate treatment discrimination. Rather, these studies have identified evidence of a disparate impact in the mortgage market. The key question that remains is whether the lender specific variation in the construction of underwriting variables can be defended with a business necessity argument. Ross and Yinger (2002) deal with this issue in detail. They found no evidence that racial differences in underwriting are attributable to observable portfolio differences between the lenders. As a result, they conclude that there is no evidence to support the idea that the across-lender differences in underwriting are justified by legitimate differences between lenders that might affect the relationship between default risk and profitability. In fact, it seems quite unlikely that these lenders have validated their specific models against performance data in a way that would allow them to reject the market level model that seems to treat minorities much more favorably because most lenders do not even have access to the data necessary to perform such a validation.

Finally, a recent reanalysis of the Boston Fed data (Han, 2002) suggests an alternative explanation for the racial differences uncovered. Han distinguishes between applicants who have a consumer credit history regardless of whether the credit history is good or bad and applicants who had no consumer credit history in the major credit bureau files at the time of the application. Han finds very large racial differences in underwriting for the subsample with no credit history information and no racial differences for the subsample where information on credit history is available. Han interprets this finding as evidence of statistical discrimination. On average, African-American applicants had substantially worse credit history than white applicants in the Boston mortgage market of the early 1990's. Lenders can directly control for these differences when credit history is observed, but if these racial differences in creditworthiness are replicated in the subsample of applicants with no credit history lenders have an incentive to discriminate against African-American in this subsample. These results are similar to the findings of Altonji and Pierret (2001) in the labor market except that Altonji and Pierret have the advantages offered by a longitudinal sample so they can examine the relationship between changes in the information available to firms and changes in the effect of education or race on wages.

2.2.3 The Price of Credit

To our knowledge, little if any research on credit pricing has been conducted using traditional survey samples like the Survey of Consumer Finance or the American Housing Survey. A number of studies, however, examine the pricing of credit by an individual lender. For example, Crawford and Rosenblatt (1997) examine differences between the final interest rate and the interest rate to which the borrower committed early in the mortgage process using a sample of 1988 and 1989 loans from a major mortgage lender. The argument being that minority borrowers may be less able to renegotiate a locked-in rate when interest rates decline. They find that the average decline below the lock-in rate is 6 basis points and that the average unexplained racial differential decline was 3 basis points or half of the overall average decline. Alternatively, Courchane and Nickerson (1997) examine mortgage overages where an individual is charged more than the standard rate for their type of mortgage at a given point in time. They examine the results at three mortgage companies and find some evidence of racial or ethnic discrimination in the setting of overages at each company.

Finally, Ayres (2001a) examines a similar phenomenon in the price of credit for new automobile loans at the Nisson Motor Assurance Corporation (NMAC). He examines the mark-up charged by dealerships over the interest rate at which NMAC is willing to provide financing. He finds that African-Americans on average pay a mark-up equal to $970 as compared to $462 for white borrowers. This differential disappears when a regression model is estimated that controls for the credit history of borrowers. Ayres argues that default risk factors should not be included in an analysis of mark-up because the dealer faces no cost associated with borrower default and NMAC has already set an interest rate based on the default risk posed by each borrower. The dealer mark-up is over and above the NMAC credit history specific interest rate. Given that the dealer actually faces no cost associated with default risk, it appears that credit history operates as a signal that informs the dealer about some other characteristic of the borrower. NMAC has argued that credit history predicts the higher labor costs dealers face when they must approach multiple credit sources to obtain a loan for a client. Ayres suggests that credit history is associated with a borrower's ability to obtain alternative financing and that the dealer uses this information to price discriminate based on the borrowers elasticity of credit demand. He notes that the high credit history mark-up persists even when the sample is restricted to loans that are financed directly by NMAC, which is the dealers' first choice for financing. A similar issue arises in the interaction between the primary and secondary mortgage markets where lenders' credit history mark-ups on interest rates for loans that are sold on the secondary market are often substantially larger than the mark-up implied by the price paid by the secondary market for these loans.

Naturally, given that overages and release from locked-in rates result from a negotiation between a loan officer and seller, it is not always possible to determine whether these results arise from adverse treatment discrimination or whether the overage process in setting interest rates has an adverse impact on minority borrowers. Moreover, if it has an adverse impact on minority borrowers, the question arises concerning whether this impact is justified by business necessity. If business necessity is defined based on profit maximization, these systems may stand up to legal scrutiny. If on the other hand a cost basis for business necessity is used (see Ayres NMAC report), it seems unlikely that such non-competitive pricing could be legally defended if it has an adverse impact on minority borrowers. However, in car dealership financing, some dealers have argued in court that price discrimination on non-racial factors is legitimate when business necessity is defined on a cost basis because it is necessary to cover their fixed cost. This argument is similar to the one used by airlines to defend price discrimination against business travelers. Ayres counters that there is no evidence that the finance charge mark-ups have been based on fixed costs or any other factor beyond the ability to extract a surplus.

2.3. Discrimination in Consumption

The research in this section is divided between research on discrimination in markets for non-housing goods and services and discrimination in housing markets. The research on non-housing goods and services is mixed in structure with some studies focusing on national samples of individuals obtained from publically available data and others using administrative data from individual firms. The research on housing market discrimination focuses on price differences using representative samples of housing units, but interprets the results in a manner that is quite distinct from the interpretations imposed on racial and gender differences identified in labor market studies.

2.3.1 Goods and Services

Goldberg (1996) using the Consumer Expenditure Survey creates a data set of approximately 1279 new automobile purchases and constructed measures of the discount from the dealers sticker price. Goldberg finds a $129 racial differential for white female, $274 differential for black males, and $426 differential for black females, but none of these differences are statistically significant.

Aryes (2001b) reconsiders these results. He notes that the point estimates from Goldberg's and his work are quite close. Goldberg estimates $129 for white women, and Aryes (1995) finds $216 while an earlier testing study (Aryes, 1991) finds $192. For black women, Goldberg and Ayres differences are all just above $400. The only instance where there are large differences in the point estimates are for black males where Aryes (1995) find differences over $1000, but it should be noted that Aryes' (1991) pilot study found statistically significant differences of $283, which is very close to Goldberg's $274 figure. The primary difference between Goldberg and Ayres results involves the precision of the estimates. Only five percent of Goldberg's sample were minorities in an extremely heterogenous sample of automobile purchases. In addition, Goldberg has to estimate the dealer mark-up that includes constructing a sticker price and imputing the value of trade-ins creating a huge measurement error problem. As a result, the standard errors on estimates are hundreds of dollars. The upper bound of the black male confidence interval is almost $800 and the upper bound for black females is over $1,400. On the other hand, paired testing controls for differences across dealers and manufacturers, differences in market demand, and differences in the season of purchases, and collects actual information on offers to compare to publicly available information of dealer costs for a specific vehicle.

In addition, Aryes (2001b) conducts his own regression analysis with data on sales and commissions from an Atlanta area car dealer. The sample contains actual dealer profit on each vehicle limiting measurement error and naturally by focusing on a single car dealership eliminates the increased variation caused by across-dealer and across-manufacturer heterogeneity. He finds no evidence of discrimination against white females ($9 gender differential), but finds very large racial differentials when comparing minority treatment to the experiences of white males, $837 for black males and $1,018 for black females.

Graddy (In Press) examines neighborhood variation in restaurant pricing. He collected a sample of 356 fast food restaurants in New Jersey and Eastern Pennsylvania in 1992. He finds that the price of a basic entree in each store is about 5% higher in a neighborhood that has a 50 percentage point rise in percent African-American after controlling for starting wage, number of employees, housing prices, and restaurant chain. These results raise the standard concern about omitted variable bias and would be much more convincing if a second wave of prices were added and compared to information on neighborhood racial composition in the 2000 census.

2.3.2 The Housing Market

Current research on discrimination in housing markets focuses on two types of discriminatory behaviors: exclusion from specific residential locations by either sellers or real estate agents, and the provision of inferior services by real estate agents. The only direct data available on these types of behavior arise from paired testing studies, see section 3.1.1. A substantial literature exists, however, that examines racial differences in housing prices as a test for whether minorities face discrimination in the housing market that excludes them from certain neighborhoods. A finding that African-Americans pay a higher price for housing than whites suggests that they face substantial constraints on their residential location choices potentially due to housing discrimination. Studies from the 1960's tend to find evidence that African-Americans pay more for equivalent housing (King and Mieszkowski, 1973, Yinger, 1978), while studies from the 1970's (Schnare, 1976, Follain and Malpezzie, 1981) tend not to find evidence of a housing price premium.

Cutler, Glaeser, and Vigdor (1999) confirm this pattern finding that the African-American rent premium fell dramatically between 1940 and 1970 and had reversed entirely by 1990. They argue that today segregation in America is enforced by a decentralized racism where whites outbid African-Americans for houses in white neighborhoods and therefore pay more for housing than African-Americans. Alternatively, Schafer (1979), Chambers (1992), and Kiel and Zabel (1996) argue that earlier work fails to find an African-American price premium because that work does not control for neighborhood quality and did not account for housing submarkets within metropolitan areas. Cutler et. al. also find that the white price premium in the 1990's is highest in the most segregated cities. This additional finding is consistent with their decentralized racism hypothesis that whites pay a premium to live in segregated neighborhoods, but it is also consistent with African-Americans being segregated and steered into the worst neighborhoods in the metropolitan area that are the most highly segregated.

In fact, we are not sure that an analysis of racial differences in housing prices can ever answer this question. As many earlier studies recognize and Cutler et. al.'s work makes very clear, racial differences in housing prices track very closely with the influx of southern rural blacks into major U.S. cities especially into the industrial north. Moreover, the decline of both racial segregation and the

African-American housing price premium followed the conclusion of this great migration. The observed housing price premium is as much a feature of this period of migration as it is an indication of discrimination. As Cutler et. al. discuss, ghetto's can operate as a mechanism to help groups assimilate into new environments. If housing markets adjust slowly relative to the speed of migration, price spikes will arise in these ghettos whether or not housing discrimination is practiced. Similarly, as migration slows, the housing market will adjust allowing price premiums to disappear even if housing discrimination is practiced. Such discrimination is likely to force the growing minority group to expand into specific regions as opposed to preventing expansion, which is required for the price premium to persist. For example, Yinger (1995, p. 123) discusses the role played by real estate agents in the creation of new predominantly African-American neighborhoods in the Mattapan neighborhood of Boston, MA.

This debate raises the more general concern about studying discrimination in differentiated product markets. As noted by Berry, Levinsohn, and Pakes (1995), a product's position in attribute space relative to other substitutes, whether the product is a housing unit, automobile, or worker, can influence the price paid for the product in equilibrium. If minorities are concentrated in certain segments of the market, minorities may experience different prices on average than whites. The key to avoiding this problem is to search for differences in prices that arise within a relatively well-defined submarket, which suggests the use of submarket fixed effects. In the case of the housing market, the use of neighborhood fixed effects may be most appropriate. If minorities feel like they have less housing options over space or face higher search costs than whites, they are likely to be at a disadvantage relative to whites when negotiating the price of similar housing.

Moreover, this interpretation stands in start contrast to interpretations applied to racial differences identified in wage regressions where the racial differences are often interpreted as wage discrimination. The labor market, however, is also a differentiated product market where the market is characterized by large racial differences in labor market attributes as well as substantial segregation across occupations. In this context, it is difficult to see how traditional regression analyses of prices in either the labor or housing market can distinguish between price discrimination and the effect of occupational segregation where occupational segregation may be caused by discrimination or across group differences in preferences.

In addition, Epple (1987) and Bayer, McMillan and Rueben (2002) explicitly model housing as a differentiated product using hedonic and multinomial choice approaches, respectively. Their models also suggest that price alone provides little guidance concerning the preferences of and constraints faced by households and individuals in markets. For alternative approaches to studying segregation that attempt to model preferences, see Ihlanfeldt and Scafidi (2002) and Ross (In Press).

2.4 The Legal System

There are only two publicly available studies that examine the relationship between race and police conduct. The first study, Ayres (2000), examines police department records for the Tulsa police department between 1995 and 2000. The files typically include the defendant=s race and age, the time, date, location and nature of the offense, and the police officer's name and identification number. Ayres compares the racial composition of citation, arrest and field incidents to the racial composition of Tulsa=s population, which was 14% African-American. Blacks were 74 percent more likely to be cited than whites and 109 percent more likely to be subject to a multiple citation stop than whites over the period. Plus the ratios increase over time peaking in 2000 at 1.95 and 2.20.

Ayres admits that this behavior might be consistent with factors other than racial profiling because the population share may not correctly describe the share of African-Americans observed in violation of a law or ordinance, but he argues that the use of the overall population of Tulsa may in fact overstate the share of African-Americans in an offender class. A substantial number of citations were issued to suburban commuters which are predominantly white, a larger percentage of the African-American population is below 16 and therefore is unlikely to be guilty of an automobile offense, and finally African-Americans are less likely to have access to a motor vehicle or to use that motor vehicle for commuting. Similarly, Aryes finds black-white ratios of 4.0, 4.1 and 3.5 for all charges, all arrests, and multi-charge arrests. These ratios grew throughout the 1990's peaking in 1999.

Ayres also examines the disparities associated with the 30 most common citation types. In this case, he actually used the age distribution to determine relevant population shares of 13 percent for automobile violations, 20 percent for juvenile offenses, and 14 percent for all others. Twenty-five citation types had statistically significant racial disparities, and none of these types showed any evidence of a higher incidence of white citation. Similarly, Ayres examines disparities associated with the 30 most common arrest charges, and 27 of the racial disparities were statistically significant with none favoring African-Americans. Ayres argues that it is quite unlikely that behavioral factors would lead to higher incidences of citation and arrest on almost all types of police actions and never favor minorities on any action.

The major limitation of Ayres study is the inability to control for the share of African-Americans engaged in behavior that might lead to a citation or arrest or to control for any observable correlates for that share. Ayres arguments that the population share may understate the population of potential violators, e.g. drivers or juveniles, are probably correct, but his arguments do not in anyway address the concern that the population share may dramatically overstate the population of actual violators. The most convincing part of Ayre=s paper arises from the disaggregate analysis in which behavioral differences never lead to a higher rate of citation or arrest for whites. At the same time, there are some disturbing results in those tables. Namely, the black-white ratios for no drivers license or suspended drivers license fall between 5 and 8. Every stop should lead to a license check by the police officer, and the severity of this type of offense leads me to believe that it would almost always lead to a citation. The black-white ratio of traffic stops is only 1.6. If most stops lead to a license check and citation if no valid license (above 90%) and the distribution of driving without a license is the same across race, the black-white ratio for no license citations should be below 2, but of course it is between 5 and 8 suggesting substantial behavioral differences between white and black drivers.

Donohue and Levitt (1998) examine a panel of 134 U.S. cities with populations over 100,000 in 1975. For each city, they collect data on the racial composition of municipal police forces (EEOC), arrests by crime category and race (FBI), and total crimes per capita (FBI). They estimate models of white and non-white arrests as a function of the number of white and non-white police officers, a vector of time varying city and metropolitan area attributes, and year, city, and region-year interaction fixed effects. They test whether adding a non-white police officer has a different effect on the likelihood of white or non-white arrest as compared to adding a white police officer. The substitution of a non-white police officer for a white officer lowers the likelihood of non-white arrest for property crimes and raises the likelihood of white arrest for violent and drug arrests.

Donohue and Levitt take the analysis one step further by examining whether the hiring of white and non-white police officers have differential effects on overall crime rates. In other words, they ask whether the effectiveness of a police force in reducing crime rates depends upon race. They interact the number of white and non-white police with the percent of the population that is white and non-white and as before examine the influence of white and non-white police on crime for a city with a given percent of white or non-white citizens. Given the inclusion of city fixed effects, their analysis tests whether a shift from white to black police officers reduces or increases crime rates in a predominantly white city or in a predominantly non-white city. For property crime, they find strong results suggesting that white police officers are more effective in policing predominantly white cities, and that non-white police officers are more effective in policing predominantly non-white cities. The results are mixed and only weakly significant for violent crimes.

The study provides strong evidence that the effect of hiring a police officer on the racial composition of arrests depends upon the race of that police officer and suggests that white and black police officers behavior differently with regard to white and black crime suspects. Typically, one concludes that discrimination is prejudice-based when the extent of disparate treatment depends upon the race of the economic actor. In this case, however, it is impossible to know whether the observed effect is the result of white officers who on average are prejudiced against black citizens relative to white citizens or black officers who are prejudiced against white citizens relative to black citizens.

In other words, does the hiring of non-white police officers reduce racial profiling or lead to unduly favorable treatment of non-white, criminal suspects. One potential answer to this question may be gleaned from Donohue and Levitt's analysis of crime rates. If the analysis of arrest patterns truly represents the influence of racial prejudice, the traditional performance analysis logic of Becker may apply, see section four, and a shift from white to non-white officers may affect crime rates. For violent crime, there appears to be some evidence that crime rates increase in all types of cities with a shift towards non-white officers, which is not consistent with racial profiling by predominantly white police departments. For property crime, the results are mixed in that the addition of non-white officers appears to increase crime rates in predominantly white cities and decrease crime rates in predominantly non-white cities. These results might be consistent with racial profiling by white police departments in predominantly non-white cities. See the discussion below on performance tests for racial profiling.

2.5 Conclusions

This section starts with a survey of the labor market literature and suggests that direct statistical tests for discrimination are likely to omit important variables for explaining outcomes, such as experience, analytical ability, and education quality. Moreover, given the substantial inequality in our society, these omitted variables are likely to be negatively correlated with minority status biasing direct statistical tests towards finding discrimination. This simple omitted variable bias becomes much more complex as we consider the lessons learned from the mortgage, automobile, and housing markets. The form of the data used in statistical tests varies dramatically across research studies and markets, and these differences have dramatic implications for the interpretation of the findings. In addition, many of the markets that have been most extensively studied are very complex, differentiated product markets in which prices may vary dramatically across submarkets.

Most labor market studies use self-reported labor market outcomes from nationally representative samples. In contrast, the studies in the mortgage market have focused on samples of loans or loan applications drawn from lenders' or even a single lender's files. These files contain the lender's treatment of the applicant, as well as the actual information submitted on the loan application. Single lender studies of underwriting or credit pricing have the potential to directly demonstrate disparate treatment discrimination, and market underwriting studies based on applications to many lenders can accurately describe the experiences of minorities in that market. On the other hand, the typical labor market study, even when using a high quality panel like the PSID or NLSY, contains almost no information about the firm where the employee works, the employee's job responsibilities, or if unemployed the firms and jobs to which the worker has submitted applications.

The limitations of research based on nationally representative samples of households or individuals become even more obvious when the possibility of heterogeneous firm behavior is considered. Even in the mortgage market where the product (credit) is fairly homogenous, strong evidence exists that loan underwriting models differ dramatically across lenders. It seems very unlikely that simple controls for industry and occupation or even those controls interacted with experience and education could capture actual differences in firm compensation and hiring policies. Using mortgage underwriting data, researchers often can gather detailed information about the lenders' portfolios of loans or the lenders' financial characteristics. In national samples of households or individuals, however, the firm where an individual works is almost never identified and at best only the most limited information is available about the worker's firm or position within the firm. Therefore, it is virtually impossible to assess the impact of across-firm differences in compensation and hiring policies on minorities, and even if this impact could be assessed it would be impossible to examine whether these differences could be explained by reasonable differences between the firms.

Only a few studies outside of analyses of the mortgage market can legitimately interpreted as strong evidence of discrimination after considering the points above. For example, in the labor market, Wood, Corcoran, and Courant's (1993) analysis of the University of Michigan School of Law's alumni files and Holzer and Ihlanfeldt's (1998) analysis of the McSui data are both quite compelling because they are able to control for the type of job activities performed by the different alumni or entry-level workers, respectively. Similarly, for automobile sales, Aryes is able to directly control for the cost and attributes of the vehicles being sold. Moreover, Arye's study avoids the problems of differences in behavior across firms because it uses data from a single dealership.

Even in studies that do a very good job of controlling for heterogeneity in the job or product, the issue of heterogeneity of firm behavior makes interpretation very difficult. If across firm differences in behavior might explain observed racial differences in outcomes, the classification of these behavioral differences suddenly depends upon the nature of those differences. Specifically, one must ask whether these differences are justified by business necessity. Moreover, the law concerning business necessity is not settled. For example, in the case of non-competitive profits, courts have not yet ruled whether lenders must adhere to a cost basis for establishing business necessity as argued explicitly by Aryes (2001a) and implicitly by the Justice Department (Courchane and Nickerson, 1997), or whether firms with market power can justify a policy with an adverse impact based on profit maximization. Ross and Yinger's (2002) view that a firm policy that maximizes profits solely by changing the racial or ethnic mix of sales should not satisfy the business necessity standard is similarly untested.

The analysis of differentiated product markets, such as labor, automobile, and housing markets, raise even more fundamental problems. As we first discussed in the context of the housing market, differentiated product markets usually contain many interrelated submarkets, and isolated shortages and surpluses can arise in those submarkets. Even a study as well designed as Wood, Corcoran, and Courant's (1993) can be severely biased by these shortages and surplus. In this context, their analysis is simply a case study and a shortage in one field of the law, such as patent or tax law, could dramatically bias estimates if women were either over or under represented in those fields. These insights suggest that the simple inclusion of dummy variables for occupation or industry is insufficient for addressing this problem. Holzer and Ihlanfeldt's (1998) use of neighborhood and job activity dummy variables are a step in the right direction, but in fact they should in principle have estimated racial differences by comparing the same type of job in the same neighborhood, which is equivalent to removing fixed effects associated with the interaction of neighborhood and job activity dummy variables. The bias in regression analysis caused by heterogeneous firm policies and/or heterogeneous products have seen very little discussion in the literature relative to the debate over omitted variable bias, and in our opinion this has been a serious omission. The little evidence available suggests that these biases may be quite large.

One major lesson arises from the past research on residential segregation and on other related topics. Information on price and choice alone cannot be used to separate the effect of preferences and constraints when modeling patterns of residential location. Potentially, this problem is less important in the labor market because workers and employers can adjust their patterns of behavior more quickly based on price differencials than households adjust residential location and neighborhoods evolve, but at present no empirical evidence has been offered to support such a notion.

Finally, the research on discrimination in the legal setting is still at a very early stage of development. On a positive note, statistical analyses are sometimes based on administrative data from a single police department, as in Aryes (2000), which limits some of the problems discussed above. However, these analyses are often forced to focus on aggregate data because the administrative files only contain information on those individuals who were actually stopped or cited. Specifically, the researchers must analyze racial or ethnic composition because they do not actually observed the police decision to stop or cite that may be discriminatory, and as a result these analyses may be biased due to racial differences in the pool of potential offenders.

3. Paired Testing

As many readers are already aware, a paired test or audit consists of sending two testers, one belonging to a majority group and one belonging to a minority group, to the same firm, agency or institution. These two testers are matched on key characteristics that might influence the treatment received and follow a common protocol at the firm or agency. After their visits, each tester independently completes a detailed survey designed to capture the treatment experienced during his or her visit. The paired testing approach is specifically designed to measure or detect the likelihood that an individual will encounter adverse treatment based on their race, ethnicity, or other protect group membership (adverse treatment discrimination). The key advantage of this approach is that it can explicitly control for many variables that are often omitted in regression studies by assigning detailed profiles to testers and implicitly control for other omitted variables through the use of testing protocols and tester training that limit differences in tester behavior during the study.

The paired testing methodology, however, has specific limitations that arise from the basic structure of the methodology. Paired testing focuses on differences in treatment observed when two testers of different race, ethnicity or gender make the same inquiry to the same firm and therefore cannot identify disparate impact discrimination where discrimination arises from systems of equal treatment that place minorities at a substantial and unnecessary disadvantage in the market place. Second, paired testing often cannot reach the later stages of a market transaction due to the fictitious nature of the tester's assigned attributes, and therefore testing may miss discrimination that is only practiced near the completion of the transaction. Paired testing also requires some mechanism or portal for accessing the market and requires that testers follow consistent protocols for their market inquiry. As a result, the estimates of disparate treatment identified in a paired testing study do not reflect the average treatment experienced by minorities in the economy, but rather a prediction of the treatment that would be experienced by a minority who followed a search strategy that paralleled the structure used for the testing study.

In this context, Heckman (1998) distinguishes between discrimination against a person at a randomly selected firm from what he defines market discrimination, which takes place in the market as a whole and argues that findings of adverse treatment in audit studies are completely consistent with no market discrimination at the margin or no market effect of discrimination. While Heckman's claim is possible in principle given his definition of market discrimination, the problem with his argument is that his definition of market discrimination has no standing in the law. The law defines discrimination in terms of the individual and the potential damaged suffered by that individual regardless of whether the market adjusts to mitigate the overall effects on a racial, ethnic, or other protected class. Moreover, while market level analyses can assess overall racial differences observed in the marketplace, they ignore the potential transaction costs imposed upon minorities who may have been required to search extensively for non-discriminatory firms.

This section continues with the discussion of specific studies that used paired testing to study discrimination in different markets and concludes with a general assessment of the strengths and weakness of the methodology.

3.1. Consumer Markets

3.1.1. Housing and Real Estate Agents

Paired testing has the longest history in the housing market where three major national studies housing discrimination studies were conducted in 1977, 1989 and 2000 and where small scale enforcement audits have been conducted by both the Justice Department and local housing groups for decades. The first major study in 1977 performed a national set of tests for black-whites and piloted Anglo-Hispanic tests. Both the 1989 and 2000 studies conducted a full-scale national testing program for both blacks and Hispanics. Phase I of the 2000 study piloted tests for Asians and Native Americans. In phase II of the 2000 study, which was conducted in 2001, a national study of Asians was conducted. These national studies are constructed as two stage samples of tests in which first a set of representative metropolitan areas are chosen based on the distribution the minority population across metropolitan areas and in each selected site tests are conducted based on a random sample of advertisements from the major metropolitan newspaper, see Yinger (1993), Smith (1993), and Boggs, Sellers, and Bendick (1993) for a history of testing in the housing market.

Heckman and Siegelman (1993) raise an important issue concerning the use of paired testing to study discrimination. The key issue raised concerns whether average differences in treatment between races are driven by discrimination or unobservable differences in the distribution of white and minority testers, i.e. an imperfect match between testers. The observed incidence of favorable treatment will vary by group if minority testers differ in the average quality of unobservables or even differ on the second or third moments (variance and skewness) of the unobservable distribution. While it is not possible to empirically distinguish between discrimination and the influence of racial differences in the distribution of unobservables, Heckman and Siegelman propose a necessary condition for racial differences in tester unobservables to bias the net incidence of adverse treatment. The tester unobservables can only bias the net incidence if they influence tester treatment overall. This condition can be tested by examining the experiences of the individual pairs used in audit studies. If tester unobservables influence treatment, systematic differences in patterns of treatment may exist between the pairs.

At present some evidence is available on this issue from the 1989 study, Ondrich, Ross and Yinger (1999) using a bivariate probit with tester pair random effects found no evidence that tester unobserved characteristics influenced treatment such as the whether the advertised unit or similar units were available when the tester inquired. Specifically, the vast majority of the variation in treatment is associated with across visit variation rather than across tester pair variation. Similarly, Ondrich, Ross and Yinger (In Press) found no evidence of tester heterogeneity in an analysis of the likelihood of real estate agents showing specific units to testers using a multinomial logit analysis with Heckman-Singer heterogeneity.

The 2000 Housing Discrimination Study explicitly collects information on actual tester characteristics in order to examine whether these characteristics influence treatment and whether the distribution of these characteristics differ by group. Two different methodological approaches are used to assess the impact of observable tester attributes on estimates of discrimination. A fixed effect logit model of racial or ethnic differences in treatment is estimated while controlling for the effect of actual tester characteristics where the fixed effect captures the common effect of the advertised unit and real estate agency. Second, a multinomial logit model without fixed effects is estimated also controlling for actual tester characteristics and then predicting differences in treatment after eliminating differences in tester characteristics. While both analyses suggest that actual tester characteristics affect treatment, neither analysis found any evidence that racial or ethnic differences in these attributes lead to a systematic overstatement of the level of discrimination in the marketplace, see Turner, Ross, Galster, and Yinger (2002) and Turner and Ross (2003).

The two other major issues that are addressed by the 2000 study are 1. Is the sampling of random advertisements from a set of major metropolitan newspapers the appropriate portal into the market place? and 2. Is gross adverse treatment (incidence of white favored) or net adverse treatment (difference between white and minority favored) the appropriate measure of adverse treatment in U.S. society? The 1989 study demonstrated that many regions of the metropolitan area were underrepresented in the major newspaper and other research suggests that other more localized sources of information may play a very important role in the housing market. In addition, many members of the National Research Council workshop on Measuring Housing Discrimination (Foster, Mitchell, and Fienberg, 2002) raised many concerns about whether the traditional sampling methodology for housing audits provides a meaningful measure of the national incidence of discrimination. Some of the issues raised at the workshop included the relevance of a target population of housing units as compared to a population of individuals who might be discriminated against and the mapping from a sample of paired events to a population.

Phase I and Phase II of HDS2000 followed two different approaches to addressing concerns about the sample frame. Phase I was a replication of HDS1989 and therefore followed the 1989 sampling frame of selecting advertisements from the major metropolitan newspaper. In order to address this limitation, the study identified under-sampled regions of the metropolitan area and administered additional tests in those locations by first over-sampling those regions and second by collecting test locations from alternative sources within those regions. A comparison of tests from over and under represented regions of the metropolitan area indicated that treatment varied across the two regions. On average, adverse treatment tended to be higher against African-Americans in underrepresented neighborhoods, but for Hispanics adverse treatment tended to be higher in overrepresented neighborhoods. This patterns was stronger for rental than for owner-occupied housing markets. Phase II, which was conducted in 2001, used a sampling protocol that rotated between the various sources available to homebuyers and renters, such as the internet, weekly newspapers, and local homebuyer or rental guides that cover a region of the metropolitan area. A comparison of tests based on advertisements drawn from major metropolitan newspapers to tests based on alternative sources did not yield a consistent or strong pattern of differences across advertisement sources.

The debate over the use of net and gross adverse treatment arises from two concerns. First, the gross measure overestimates discrimination because sometimes a minority tester is treated disfavorably for reasons that have nothing to do with race. Second, the net may under estimate discrimination because sometimes minorities are favored for systematic reasons, such as a white not being shown units in minority neighborhoods. If such race-based exclusion of whites occurs, the incidence of minority favored treatment will overstate the frequency of adverse treatment that arises due to factors other than race and the net will be based on the subtraction of a number that is too large. Ondrich, Ross, and Yinger (2000) uses the estimates a parametric model to correct for these problems and develop upper and lower bounds of discrimination, but the bounds are often quite far apart.

The most direct approach to this problem is to conduct sandwich or triplet audits in which three testers participate, two testers of the same race. With this structure, which is piloted in two sites in phase II of HDS2000, the frequency of adverse treatment that is not related to race can be calculated by comparing the treatment of the same race testers. Analysis of these results yields little evidence that the incidence of minority favored treatment systematically overstates the frequency that differences arise due to random variation. In fact, the few treatment variables where a difference was found between the incidence of minority favored and same race differential treatment were treatment variables where no systematic adverse treatment against minorities was observed, see Turner and Ross (2003). These findings suggest that the net measure of adverse treatment is an appropriate measure of discrimination in the market. It should be noted that the two piloted sites had substantially lower levels of adverse treatment than many of the sites in the Phase I study and that the results might have been different in metropolitan areas with higher levels of adverse treatment.

Note that some studies have offered an alternative measure of discrimination referred to as systematic discrimination, see Yinger (1995), based on predicted racial differences from a regression analysis, these predicted racial differences are simply net measures of discrimination conditional upon the variables included in the regression. Experience with those measures do not offer any insights into the debate over the appropriateness of the net and gross measures.

Finally, a recent study of HDS 1989, Ondrich, Ross, and Yinger (In Press), adds insights into our earlier concern that product heterogeneity could bias estimates of discrimination in a regression context. Ondrich, Ross, and Yinger examine the decision of real estate agents to show specific homes to white and black testers during the 1989 study of the market for owner-occupied housing. They control explicitly for the initial request made by the tester, i.e. the attributes of the advertised unit and the neighborhood in which the unit is located, and find that the initial request and its relationship to the available housing stock has a very large impact on the treatment of white and minority testers. This heterogeneity of treatment does not lead to bias in audit studies because the request is assigned to pairs and there is no correlation between race and request, but the study does suggest that preferences might have a large impact on estimated differences in traditional regression analysis. Also see Ondrich, Ross, and Yinger (2002) for a related analysis of the same sample.

3.1.2 Other Consumer Markets

Two major studies of retail automobile sales have been conducted. The first study (Aryes, 1991), which was a pilot for the second study, conducted 180 tests in the Chicago area, and the second study (Ayres and Siegelman, 1995; Ayres, 1995) conducted 306 tests in Chicago. The major difference between the two studies involved the selection of testers. The pilot study used only six testers: three white males, one white female, one black male, and one black female, which raises concerns that the results may be influenced by the idiosyncratic characteristics of these six testers. The second study hired 38 testers in all: 18 white males, 8 white females, 5 black males, and 7 black females. The two studies provide comparable results. The first study found that white females paid $242 more on average than white males, while the second study found a difference of $129. Both studies found that black females paid a premium of just over $400 relative to white males. The only major difference arose for black males where the first study found a difference of $274 and the second study found a difference of $1068. Ayres (2002) attempts to test various hypotheses concerning the causal factors behind discrimination. Ayres finds evidence that the seller' bargaining behavior is at least partially driven by statistical discrimination based on analyzing the bargaining times and sellers' concessions from the pilot study.

The major issue concerning the pilot study, namely the breadth of the tester pool, relates directly to Heckman and Siegelman's concern that the testers may differ by race on unobservables that influence treatment. It is unclear to me, however, how broadening the pool of testers solves the problem. Murphy (2002) argues that broadening the pool solves the tester heterogeneity problem as long as testers are matched on all variables for which the distribution varies by race and suggests that the pool of testers should be made as broad as is feasible. Under her scenario, the random sample of tester pairs form the basis for comparison and testers need not be matched on the presumably larger set of variables for which the distribution does not vary by race. In my opinion, however, Murphy misses the point of Heckman and Siegelman that in a society with great inequality it is likely to be very difficult to control for all variables that vary by race and potentially influence treatment. As a result, broadening the pool may do little to mitigate tester heterogeneity bias, but rather simply assures that racial differences in treatment have converged to the racial differences in unobservables.

Finally, an Urban Institute pilot study used paired testing to examine neighborhood discrimination in the home insurance market (Wissoker, Zimmermann, and Galster, 1998). The testers and their homes were matched on a wide range of attributes. The homes within a matched pair, however, were located in neighborhoods with different ethnic compositions. Specifically, neighborhoods were divided into predominantly white and Hispanic neighborhoods, and then pairs of neighborhoods, one white and one Hispanic, were created by matching neighborhoods on standard demographic and housing stock attributes. In Phoenix, substantially higher premiums were quoted for homes in Hispanic neighborhoods, but Hispanic neighborhoods tended to be in different insurance rating territories so the study could not determine whether the differences were due to legitimate differences in loss. Of course, since the neighborhoods were matched on a broad array of attributes, one potential explanation for these differences is discrimination in the setting of rating territories.

3.2 The Labor Market

A number of paired testing studies of employment were conducted during the 1990's. The Urban Institute conducted paired testing studies of employment in Chicago and Washington D.C. using both black-white (Turner, Fix, and Struyk, 1991) and Anglo-Hispanic (Cross, Kenny, Mell, and Zimmerman, 1990) pairs. They find net differences, the difference between frequency of white and minority favored treatment, in the likelihood of receiving a job offer of between 5.1 and 16.2 percentage points. James and DelCastillo (1991) conducted similar audits in Denver finding net differences in the likelihood of a job offer of 6.3 and -3.5 for Anglo-Hispanics and black-whites, respectively. Neumark (1996) conducted a small audit study of the restaurant industry using two male-female pairs finding that men were more likely to obtain job offers.

Heckman and Siegelman (1993) are quite critical concerning the use of paired testing to study discrimination in labor markets. Heckman and Siegelman argue that the problem of tester heterogeneity may be especially severe for employment tests because the determinants of productivity within a firm are not well understood and very hard to measure. As a result, perfect matching between testers is nearly impossible. In fact, Heckman and Siegelman point out that pairing on a small subset of easily observable or assigned attributes, such as age or education, might exacerbate the problem of unobservable tester characteristics because those characteristics are the only ones on which the competing testers differ. Ross (2002) counters that the key is not matching as much as training in order to assure that the tester's true attributes, as opposed to the attributes assigned during the test, have little influence on the tester's behavior and therefore cannot be detected by the employer being tested.

As discussed earlier, while it is not possible to empirically distinguish between discrimination and the influence of racial differences in the distribution of unobservables, Heckman and Siegelman suggest that the tester unobservables can only bias the net incidence if they influence tester treatment overall. Heckman and Siegelman examine differences in outcomes across tester pairs for the Urban Institute audits and the Denver audits conducted by James and DelCastillo using a non-parametric test for homogeneity across pairs. For the Urban Institute, the test for homogeneity is rejected only for black-white tests in Chicago, which had by far the smallest incidence of discrimination. For the other three testing efforts, the overall net incidence fell between 13 and 16 percentage points, and there was no evidence that tester unobservables matter. For example, the pair specific white-favored net incidence fell between 16 and 26 percentage points and black-favored fall between 4 and 7 percentage points for the five black-white pairs in Washington D.C. The Denver tests appear more susceptible to bias caused by tester unobservables. For example, for one pair in the black-white tests, the white tester was treated favorably 28 percent of the time and the minority tester was never treated favorably, and for another pair the white tester was never treated favorably and the minority tester was treated favorably 19 percent of the time. Similarly, Neumark's study has been questioned because one of the male-female pairs contained an Asian female tester who had much less success in medium priced restaurants than the other female tester, see Altonji and Black (1999).

3.3. Mortgage Lending

A grow body of evidence suggests that testing can be used to examine discrimination in the pre-application stage of the mortgage lending process. Fair housing groups have conducted enforcement-oriented testing of mortgage lenders, and many of these testing efforts lead to court cases and legal settlements under the Fair Housing Act, see Lawton (1996) and Smith and Cloud (1996). In addition, the Urban Institute recently completed a pilot testing study for blacks and Hispanics in the Chicago and Los Angeles metropolitan areas.

The Urban Institute (Smith and Delair, 1999) obtained the 1993 pre-application testing data from enforcement efforts of the National Fair Housing Alliance in five cities and reanalyzed the data. There was tremendous heterogeneity in terms of the products and information provided to testers. In the end, they focused on whether the tester received a quote and the number of quotes received where quote was defined as information about a loan product with an estimate of monthly mortgage payments and closing costs. Statistically significant net differences of 13 and 25 percentage points were found for Chicago and Atlanta, respectively. In terms of number of quotes, significant differences existed for Chicago, Atlanta, and Denver. These results must be considered with care because the tests were designed for enforcement purposes and little is known concerning the process of selecting and training testers and the procedures followed by testers during their visit to a mortgage lender.

The urban institute pilot study of pre-application mortgage testing, called the Homeownership Testing Program (HTP) see Turner, Godfrey, Ross and Smith (2003), provides the first paired testing evidence arising from a cohesively designed and carefully documented process. HTP focuses on a single scenario in which a first-time homebuyer visits a mortgage lender and requests information about obtaining a mortgage including help in figuring out a price range, a reasonable loan amount, and suitable loan products. A population of lenders for which testing is viable was constructed as all lenders listed in the Home Mortgage Disclosure Act (HMDA) data that received at least 90 applications the previous year and had a local office in the metropolitan area that provided information on mortgage products directly to consumers. The population of testable lenders represented 56% of mortgage activity in Los Angeles and 62% of activity in Chicago. Lenders were sampled from the population with a probability proportional to each lenders HMDA loan volume.

The testers were assigned similar personal and financial characteristics in order to assure comparability across testers and across tests. All testers were assigned to be married, first-time homebuyers. Their income and assets were assigned to assure that all testers were downpayment rather than income constrained. A set of financial profiles were developed, which were assigned randomly to tester pairs. These profiles started with a home price near the median house price for the metropolitan area and set assets based on a five percent downpayment, income based on a 28 percent debt to income ratio, and debts based on a 32 percent total debt expense to income ratio. Based on these values, the applicant faces a binding downpayment constraint is binding, and income ratios based on the implied maximum loan amount that are well within guidelines for conventional conforming mortgages. All testers were randomly assigned A- credit profiles, and all testers had time at current residence and at current employment set at 3 years or greater since tenure in job and residence usually does not matter as long as it exceeds 2 years. Approximately 75 tests were conducted for each group in each site.

The HTP was designed to allow the analysis of multiple factors over a variety of loan products. In terms of total supply of credit, Hispanic testers were provided products that offered a smaller maximum loan amount in Chicago. The net difference was 33 percentage points and the difference in average maximum loan amount across tests was $10,000 for an average loan amount of about $190,000. In Chicago, both blacks and Hispanics received information about less products with net differences in treatment of 27 and 28 percentage points, respectively. Blacks in both Los Angeles and Chicago received less information and assistance where assistance is defined as suggestions concerning paying down or consolidating debts, obtaining a downpayment, information on points or closing costs, provision of a pre-qualification letter, or information about homebuying seminars. The net differences were 32 percentage points for Los Angeles and 31 points for Chicago. Black testers in Chicago were also less likely to receive a follow-up phone call where only the white tester received a phone call for 13 percent of tests and only the minority tester received a phone call for 1 percent of tests. Finally, blacks in Chicago were more likely to face higher closing costs with a net difference of 25 percentage points.

These results point to substantial and important racial and ethnic differences between the pre-application experiences of first-time homebuyers especially for blacks in Chicago. Admittedly, the form of adverse treatment varied across markets and groups, and there were many treatment variables, such as being quoted a loan amount, encouraged to pursue FHA financing, suggestions concerning variable rate mortgages, products requiring PMI, and average interest rates, where no differences in treatment were found. Adverse treatment against minorities, however, is observed on some variables for both groups in both sites and is sizable when present. Furthermore, favorable treatment of minorities is never observed for any treatment variables for any group in any site.

The complexity of the mortgage market suggests that Heckman's concerns about tester heterogeneity may also be important for mortgage testing efforts. Preliminary investigations by Godfrey, Ross, and Turner (2003) find some evidence that tester identity can explain differences in treatment between testers of the same race in different tests, but the treatment variables for which tester identify can explain treatment are few and do not coincide with the variables where statistically significant adverse treatment against minorities was observed. Therefore, it seems unlikely that the results reported above can entirely be explained by racial differences in the pool of testers.

3.4. Conclusions

It is important to recognize the limitations of testing. Testing can only detect explicit differences in treatment (disparate treatment discrimination). Testing does not detect or account for firm behaviors that may have an unjustified, disparate impact on minorities while the regression evidence from the mortgage market suggests that the disparate treatment discrimination may only be a small contributor to racial differences in lending. Testing usually can only detect discrimination at the early stage of a market transaction, such as an initial visit or visits to a real estate agency or the pre-application stage in the mortgage process, and simply cannot be used for situations where the firm or institution has an on-going relationship with individuals. This limitation may be quite problematic in the case of the labor market where regression evidence suggests that most racial, ethnic, and gender differences accumulate over time due to differential returns to experience. Moreover, when substantial enforcement testing is taking place, discriminating firms or individuals may move discrimination until later stages that cannot be reached by testing. Finally, as suggested by Heckman (1998), testing does not offer any insights concerning the equilibrium impact of discrimination on minorities. Testers follow common protocols and visit randomly selected firms or institutions, and as a result testing does not consider the adjustments that minorities may make in order to limit the impact of discrimination on their well being. As a counter-argument, however, it must be noted that such adjustments are unlikely to be costless, and so any discrimination that requires individuals to adjust their behavior almost certainly has an impact on welfare.

In addition, a number of important methodological issues have been raised concerning testing. Heckman and Siegelman's critique concerning tester heterogeneity that varies by race is potentially the most important because it suggests that testing may provide a biased test for the existence of discrimination. Murphy raises a related issue regarding the number of testers arguing that each tester only provides one heterogeneous observation from the white or minority population and that accurate estimates of discrimination require a large number of testers. However, the best evidence available at this point suggests that tester training and protocols can prevent testers' actual characteristics from having an undue affect on observed tester treatment. Moreover, with regards to Murphy's comment, a smaller number of testers provides more observations for each tester and increases the statistical power of tests for tester heterogeneity. Nonetheless, given the importance of these issues, future testing studies should collect data on actual tester characteristics in order to allow more direct tests for the influence of tester characteristics.

The second two issues, adjusting estimates to represent a relevant population and using the appropriate measure of adverse treatment, suggest that testing may not provide an accurate measure of the level of discrimination in a market or submarket. The most reasonable approach to addressing the first question is to examine whether estimates are sensitive to different weighting schemes, such as weights to the population of available housing units or weights to the population of relevant minority households or individuals. The second issue concerns whether the frequency of adverse treatment against the minority tester in a sample of tests (gross measure) or the difference between minority and white adverse treatment (net measure) provides the best measure of discrimination. This important issue remains unresolved, and at present the most reasonable and conservative approach is to focus on the net measure as providing a lower bound and a robust test for discrimination in the market place.

In spite of these limitations, we believe that testing provides a powerful tool for examining discrimination in a variety of markets. Only Heckman and Siegelman's heterogeneity critique raises any questions concerning the validity of a finding that discrimination exists based on testing data, and most of the empirical evidence suggests that well designed studies are robust to this critique. A well designed testing study provides a very simple solution for the omitted variable problem that plagues regression analysis. All relevant information that a tester might provide during a visit can be assigned to the tester, and the influence of tester characteristics is mitigated by tester training and test protocols. Note that matching is not crucial in these studies, but rather omitted variable bias is avoided because the assigned characteristics are orthogonal to race or ethnicity. In a simple competitive market, the matching on assigned characteristics simply reduces variance and increases the statistical power of the test.

Finally, paired testing is robust to the heterogeneity of firm policies and product attributes, which was first discussed in the previous section on regression analysis. The combination of matching testers on observable and assigned attributes and assigning testers to the same firm and product is very powerful. The attribute matching essentially provides a non-parametric control for the variables on which firms or institutions base decisions. Since the testers approach the same firm and request the same product, a non-discriminating firm should apply the same model to the two testers and treat the testers the same. Obviously, shortages or surpluses in different segments of a differentiated product market should affect white and minority testers equally (presuming there is not discrimination in the market) since the testers always enter the same submarket in randomized order. While product and firm heterogeneity may increase the importance of a well defined sample frame for providing accurate measurements of discrimination, this difficulty is minor in our opinion when compared to the problems heterogeneity creates for regression analysis.

4. Performance Analyses

The performance analysis of discrimination arises from the simple logic that prejudiced firms must be compensated in order to be willing to do business with minorities (Becker, 1971). In the case of the mortgage market, prejudiced lenders may discriminate setting a higher underwriting threshold for minority borrowers than white borrowers. On the margin, the prejudiced firms are indifferent between approving a white application that lies at the white underwriting threshold and approving a higher quality minority application at the minority threshold. Empirically, the higher threshold might be detected by observing lower default rates among marginal minority borrowers. Similarly, in the labor market, the prejudiced firm might be compensated for hiring minority workers if they paid a lower wage to minority workers as compared to equally skilled white workers. In this case, wage discrimination in the labor market would be consistent with a situation where minorities had higher job performance after controlling for the worker's wage rate.

The discussion below, however, highlights a key difference between the performance analyses of discrimination in the mortgage and labor market. Performance analyses in the mortgage market are intended to capture discrimination in the underwriting process where mortgage applications are approved or denied, or screening discrimination where a firm imposes a higher threshold upon minority candidates. On the other hand, the labor market analyses test for wage or price discrimination where a firm's prejudice leads to lower pay for a minority employee when compared to an equally qualified white. Performance analyses that test for price discrimination essentially test for whether the individual is performing beyond their level of compensation, i.e. under compensated after controlling for their ability. For performance analyses for screening discrimination, however, the test arises from the selection bias that is created by the screening process. The reliance of performance tests on selection bias is quite problematic because it is difficult to see how an analysis can be both reasonably insulated against omitted variable bias and yet allow selection bias to remain in the estimated parameters. The section also discusses recent performance analyses of discrimination in a legal setting.

4.1 Default Analysis in the Mortgage Market

The default approach to studying mortgage discrimination has received a great deal of attention in recent years. This attention may reflect the fact that this approach has great intuitive appeal. The basic argument was expressed in a widely read magazine column by Gary Becker (1993a) and in the lecture Becker delivered when he was awarded the 1992 Nobel Prize in Economics (1993b). In Becker's words:

> If banks discriminate against minority applicants, they should earn *greater* profits on the loans actually made to them than on those to whites. The reason is that discriminating banks would be willing to accept marginally profitable whites who would be turned down if they were black (1993b, p. 389; emphasis in the original).

In other words, if minority applicants are held to a higher standard of creditworthiness, the marginal black applicant will be less likely to default than the marginal white applicant. As with much of Becker's work, this approach is developed based on a prejudice-based model of discrimination in which firms sacrifice profits in order to discriminate and is not intended to detect instances of statistical discrimination.

The two best know studies that have directly apply the default approach as described by Becker are Peterson (1981) and Berkovec, Canner, Gabriel, and Hannan (1994). Peterson examines outcomes for 30,000 commercial bank consumer loans across 30 banks between 1966 and 1971. They find no gender differences in the performance of these loans either unconditionally or after controlling for a wide array of variables that might influence the likelihood of default. Berkovec et. al. estimate a logit model of default for a sample of FHA mortgages from 1987 through 1989 controlling for standard underwriting variables, such as loan to value ratio, debt to income ratio, housing expense to income ratio. They find that minorities are more likely than whites to default after controlling for all the underwriting variables that are available in their data set. On the basis of these findings, both studies conclude that their samples provide no evidence of discrimination against women or blacks, respectively.

Peterson, however, refers to his analysis as a comparison of average defaults and states that his conclusion requires the assumption that the underlying distribution of unexplained or unobservable creditworthiness does not vary by gender. Berkovec et. al. refer to their analysis as a marginal comparison of default because they use a regression approach, but nonetheless they require the same assumption concerning the distribution of the unobservable. Neither Peterson's sample nor the FHA sample of Berkovec et. al. contain a full set of variables present on the loan application. Peterson's sample is missing a number of important variables, and the FHA variable does not contain any information on credit history, which is well known to both vary by race and be an important determinant of default risk. Moreover, the default unobservable may contain both factors observed by the lender and factors that are not observed by the lender. The factors that are not observed by the lender cannot bias the estimates arising from an underwriting model, but still influence the estimated coefficients in a default model.

Essentially, neither study explicitly identifies marginal applicants, but rather attempts to take advantage of the fact that denial of poor quality applications creates a selection bias in the sample of loans and that discrimination against minorities leads to a larger selection bias for the sample of minority loans. The problem with this approach is that the value of the race coefficient prior to any selection bias is unknown, and these studies simply assume that the race coefficient without selection is zero, which will not be the case if any variables are omitted that are both correlated with race and influence default, Ross (1996, 1997). In principle if the omitted-variable bias arose entirely from variables that lenders do not observe, one might interpret these results as finding that discrimination does not exceed the level expected by a profit maximizing statistical discriminator. The problem of omitted variables, however, is complicated by the fact that the use of a complete underwriting specification will eliminate any borrower specific variation in the underwriting error term. As a result, the error term in the underwriting equation will be uncorrelated with the error term in the default equation, and the default equation will not suffer from selection bias (exogenous stratification), Ross (1996, 1997). In Ross's (1997) empirical application, he finds that the omitted variable bias is larger than the selection bias reversing the implications of the default approach. Therefore, default estimations can only detect discrimination if valid underwriting variables are omitted, but is not an unbiased test for prejudice-based discrimination unless those omitted variables are uncorrelated with race.

Two alternative approaches have been developed to address these concerns. Berkovec, Canner, Gabriel, and Hannan (1998) develop an alternative version of the default approach that may be insulated from the omitted variable bias discussed above. Rather than test directly for differences in default based on minority status, they identify a proxy variable that should be related to the level of discrimination but not to minority status, and then test whether group-based differences in default are affected by this proxy variable. Instead of focusing on the coefficient of a minority-status variable, this version of the default approach focuses on the coefficient of an interaction between minority status and the proxy variable C a coefficient that, under certain assumptions, will not be subject to omitted variable bias. Specifically, they develop a Herfindahl index of market concentration in each metropolitan area and re-estimate the FHA default model controlling for both the market concentration and the interaction between race and market concentration. They suggest that prejudice-based discrimination should be higher in more concentrated markets, but they find no evidence that the race coefficient varies by market concentration.

Although it is novel, the Berkovec et al. article also has a fundamental flaw, namely, that it is built on two assumptions that are essentially contradictory. The first assumption, which is explicit, is that discrimination motivated by prejudice is stronger in locations where the lending industry is more concentrated, that is, when there are fewer lenders to compete against each other. We find this assumption to be plausible; it says, in effect, that lenders are in a better position to allow prejudice to affect their underwriting procedures when they face less competition; as a result, a lack of competition will be particularly hard on minority customers. The second assumption, which is implicit in the Berkovec et al. argument, is that underwriting standards are not affected by the degree of competition among lenders. If underwriting standards are affected by market concentration, the distributions of the unobservables for white and minority loans are likely vary with market concentration due to the correlation between race and other variables that capture credit worthiness. Regardless of the degree of concentration, in other words, lenders must always set the same credit standard and place the same weight on each underwriting variable. This assumption all but contradicts the first one: How can a lack of competition cause lenders to be more aggressive in rationing credit to minorities without causing lenders be more aggressive in rationing credit on other grounds, as well?

Ironically, Berkovec et al. actually test this second assumption. As noted earlier, they include in their default equation the concentration variable itself, not interacted with anything. They find that the level of market concentration lowers the likelihood of default. They interpret this result as support for the view that lenders ration credit more aggressively in more concentrated markets. However, they fail to see that this result undermines their interpretation of their key interaction variable. Given the shift in lender credit rationing, the relationship between market concentration and racial differences in default may not be zero in a world without prejudice.

Han (2003) generalizes the default approach by theoretically deriving the relationship between prejudice-based or statistical discrimination and loan performance measures such as default rates or lender loss from default. Based on his theoretical derivations and specific calibrations in the data, the author concludes that prejudice-based discrimination leads to higher loan performance. This relationship is based on the behavioral responses of borrowers and does not arise from a sample selection bias, which is a major advantage relative to other studies. The model also, however, requires the use of a reduced form underwriting model that includes the predictors of key underwriting variables, such as loan to value or debt to income ratio. The need for a reduced form model leads to new and serious concerns about omitted variable bias because lenders clearly care about loan to value and debt to income ratios and most samples including the one used by Han do not contain sufficient controls for the individual attributes that influence loan terms.

The second approach is to combine information on underwriting and performance using the information on racial differences in one equation as a proxy for unobserved racial differences in the second equation. Van Order and Zorn (1995) examine loan rejection rates from HMDA and FreddieMac loan performance data using cross-sectional data. In the southeastern U.S., they find that default rates are either unaffected or fall as the share of blacks in a census tract increases, but that rejection rates increase with the share of black. They argue that higher rejection rates for blacks imply that the quality of the underlying application pool is lower on average for blacks, and therefore a lower rate of default for blacks for the same pool provides evidence prejudice-based discrimination in the southeastern U.S. Ross (2000) estimates an underwriting model using the Boston Fed data and a sample selection corrected model of default using FHA data. By imposing the assumption that certain key underwriting variables only enter the underwriting process through their influence on default, Ross is able to back-out an estimate of the ratio of racial differences in underwriting and racial differences expected if lenders were perfect statistical discriminators. Ross estimates a ratio near one, which suggests that the average underwriting behavior of lenders is consistent with statistical rather than prejudice-based discrimination. Ross's results are also consistent with Han's (2002) finding that the Boston Fed results can be explained by statistical discrimination against minority borrowers without information on credit history.

Both Van Order and Zorn and Ross provide a mechanism for controlling for unobserved racial differences in application quality and therefore these approaches have the potential to provide more robust results than the earlier single equation studies. At the same time, each study has substantial limitations. Van Order and Zorn focus on aggregate data, with the corresponding potential for aggregation bias. While Ross's results are robust to a number of alternative identifying assumptions, his simultaneous equations methodology requires strong exclusion and cross-equation parameter restrictions that cannot be independently verified. Regardless, these papers illustrate the importance of controlling for unobserved racial variation in performance. Ross (2000) and Berkovec et. al. (1994) use the same FHA sample and reach quite different conclusions. Berkovec et. al. finds that minorities default more often suggesting that the magnitude of discrimination is not large enough to offset racial differences in performance, and Ross finds that racial differences in performance are completely captured by racial differences in underwriting.

4.2 The Labor Market

4.2.1 Job Performance as Evidence of Labor Market Discrimination

Szymanski (2000) examines racial differences in the performance of soccer teams as a market-based test for discrmination in hiring. Following the standard logic, Symanski defines evidence of discrimination as a situation in which firms can earn higher profits by hiring an above average proportion of one group of workers. Symanski develops a simple model in which owners maximize a utility function that depends upon a weighted average of profits and the share of white players. Based on this model, Symanski concludes that a taste for discrimination acts as a tax on team success. Symanski assumes away the possibility of customer- based discrimination by specifying revenues as a function of performance only, but the same results could have be generated by a model in which the owner trades-off performance and share of white players and both owner and customer prejudice act as a tax on performance.

Naturally, this logic is also derived from Becker's theory of firm manager or owner prejudice-based discrimination. In this case, however, the logic leads to a test for wage or price discrimination; with prejudice, a black player will be paid less than an equally qualified white player. Specifically, Symanski estimates a model of league standing using data from the English soccer league between 1978 and 1993 using a balanced panel of 39 clubs controlling for club fixed effects, wage bill, turnover, and share of black players on the club. He find that clubs hiring a below average number of black players have lower standings after controlling for the teams wage bill. In other words, black players offer greater performance for a given salary level. He concludes that this result provides evidence that team owners are prejudiced and as a result discriminate against black players, which leads to lower salaries for black players.

One of the standard critiques of default analyses also applies to performance analyses in the labor market. Unobserved player variables that affect compensation are likely to bias direct estimation of racial differences in compensation. Moreover, if racial differences in compensation are biased towards finding discrimination because minorities have unobservables that lead to lower wages after controlling for performance, the performance approach will be biased away from finding discrimination because the firm will have a lower wage bill overall if they have a higher fraction of black players. This analysis, however, is much more compelling than traditional default tests in the mortgage market because the author has a panel of club outcomes that allow him to control for club fixed effects, as well as division-year fixed effects. These fixed effects explain a substantial amount of the systematic variation in club performance and should provide an effective control for level differences between the distribution of white and black players over unobservables that influence salary. Specifically, the racial composition coefficient in the fixed effect model actually captures the effect of the marginal worker added by comparing changes in racial composition to changes in performance.

It should be noted that the standard selection issue arises here, as well. If discrimination in hiring leads to a higher quality pool of black players, the performance of black players overall should be higher than the performance of white players, and of course the standard criticisms involving unobservable differences in the distribution are maintained. This issue, however, is much less relevant in this context because the author controls for wages in club performance. So, assuming that there is no wage discrimination and that wages reflect true performance, the author's estimation should not capture the resulting selection bias since after controlling for wages there are no omitted variables to create a selection bias.

Hellerstein, Neumark, and Troske (1999) and Hellerstein and Neumark (1999) use establishment level data to compare the relative productivity of workers by race and/or gender using a production function that controls for the demographic composition of the workforce. In one set of estimates for U.S. data, they find that women are 15% less productive, but paid 32% less, and they find that blacks are 9% more productive and paid 7% more with these differences being statistically insignificant. In Israel, they find no differences between the gender wage and productivity gaps. These results are thrown into some doubt by the findings for education. Workers with a college education are estimated to be 74% more productive, but only earn 27% more. The wage expense and productivity returns to education should be the same. Hellerstein, Neumark, and Troske (2002) perform a similar analysis using firm profits. They find that a 10% increase in percentage female increase profit by 4.6% suggesting prejudiced based discrimination.

These analyses suffer from the usual problem of performance analyses. They make average instead of marginal comparisons, even if sometimes this average comparison is conditional on a set of observables. Such analyses can suffer from severe biases if the unobservable distribution over race, ethnicity or gender differs on any of its moments. Firms most likely vary in the distribution of applicants and therefore firms that make identical decisions on the margin may have a very different pool of hired workers. In our opinions, these problems are exacerbated by the uses of aggregate data. Even if the data is only aggregated to the firm level, the omitted variable bias concern is quite serious because these type of aggregate data set do not contain information on the characteristics of the individual employees for which firms are making compensation decisions.

Our earlier discussion of Holzer and Ihlanfeldt on employment seems relevant here. In their estimation of minority hiring, they include detailed controls for location, job attributes, and firm attributes in order to assure that the pool of white and minority candidates were comparable across the firms that were being examined, and tested whether customer racial composition affected the racial hiring decisions of firms with similar applicant pools, e.g. same labor market pool based on location, job type, and firm type. These controls are sufficient for Holzer and Ihlanfeldt because they are examining differences in racial hiring disparities and they do not need the pool of white and minority applicants to have similar unobservable distributions. On the other hand, the three analyses above examine the level of racial and gender disparities. These analyses would be much more compelling if a panel could be developed in which changes in the racial composition could be compared to changes in productivity or profitability.

At this point, a more careful discussion of the effect of omitted variable bias on wage regression and performance analysis of wage discrimination is helpful. In a world where wage and one performance measure is observed, the omitted variable bias may be reinterpreted as an endogeneity bias. Specifically, unobservables that are correlated with race influence both wages and the performance measure, and single equation models of wages that include the performance measure or performance that include wages in the specification are biased due to the endogeneity of performance and wages, respectively. The bias in the race coefficient in these two equations depends upon whether these unobservables have a larger effect on performance or wages. If we assume a negative correlation between minority status and unobservables that positively influence wages and performance and we assume that the effect of unobservables are larger on wages, the wage coefficient is biased towards finding discrimination and the direct performance analysis is biased away from finding discrimination. Regardless of the assumptions, the two estimates are likely to bracket the true level of discrimination.

The interpretation of performance results are complicated by the different types of relationships between firms and households that might be modeled. Some firms, such as mortgage lenders, have isolated transactions with customers in which the price is set at one point and performance is only observed at a later stage. In this case, the price cannot be influenced directly by performance; in fact, price only depends upon variables observed by the firm at the time of the transaction while performance may depend upon a much wider array of individual or household variables. As a result, no simple rule arises in comparing a traditional price regression based upon variables observed during the transaction to a performance analysis controlling for price. The notion that price and performance analyses bracket discrimination will hold only for in a very specific context. If the price regression includes performance or a group specific proxy for performance, see Ross (2000), the price and performance regressions will provide bounds on a measure of discrimination attributable to prejudice or more specifically the amount by which discrimination exceeds the level expected if firms engage in statistical discrimination.

The interpretation of performance analyses change when firms have an on-going relationship with individuals or households and have the opportunity to observe performance and adjust prices. Even though performance may be influenced by variables that are unobservable to the firm, the firm observes performance and can implicitly account for those unobservables by directly rewarding that performance. In this case, performance analyses deserve equal standing with standard wage regressions and might be preferable to simple wage regressions, which do not control for key performance measures that are observed by firms when wages are set. The circumstances clearly match the situation being considered in Symanski and in the work by Hellerstein, Neumark, and Troske. The interpretation of their results, however, is complicated by the fact that they are examining aggregate firm level data instead of the performance and wages of individual workers, which is why we believe that Symanski's results are and Hellerstein et. al.'s results would be dramatically strengthened by the use of a panel.

4.2.2. Reverse Regression of Labor Market Outcomes

Another alternative to the standard regression analyses is the reverse regression approach. In the first stage, an initial model of the relationship between observables and wages is estimated first and those estimates are used to create a qualification index. In the second stage, this index is treated as the dependent variable and regressed on wages and protected class in order to test whether members of the protected class are more qualified than equally paid whites. Many studies, see White and Piette (1998), find that racial and gender differences in wages actually reverse sign when reverse regression is employed suggesting reverse discrimination in favor of the protected class. Goldberger (1984) examines the bias in standard and reverse regression under a number of errors in variables specification. In the simple case with one qualification covariate in the first stage, Goldberger finds that standard regression is biased upwards and the reverse regression is unbiased when the researcher measures the employers productivity indicator with error, but he finds that standard regression is unbiased and that reverse regression is biased downwards when the employer relies on a productivity proxy that is observed by the researcher plus a second unobserved indicator of productivity. He finds that the theoretical implications are even less clear in a multivariate context where worker productivity may be described on multiple dimensions. In fact, in the full model, both standard and reverse regressions are biased and the direction of the bias is indeterminate.

The focus of the debate on errors in variables is quite interesting given that much of the debate concerning wage discrimination concentrates on omitted variables that may be correlated with protected class. In a simple world with no errors in variables and only omitted variable bias in the coefficient on the protect class, i.e. estimates of other behavioral parameters are unbiased, omitted variable bias in the protect class coefficient is in the same direction for standard and reverse regression. If a positive correlation between white and unobservables raises white salaries controlling for qualifications, the same correlation will lower white qualifications controlling salary. The main implication of reverse regression results is that errors-in-variable problems may be quite large and the focus on omitted variables in many studies may be misplaced. Racine and Rilstone (1995), however, examine standard and reverse regression using a non-parametric specification for potential experience and focusing only on childless adults under the assumption that potential experience is a better measure for that group. They find substantial differences between the parametric and non-parametric reverse regression results and quite comparable results for the parametric regression, non-parametric regression, and non-parametric reverse regression. Their findings suggest the importance of measurement error may be minimal in well specified equations that contain quality control variables for experience.

For analyses using firm provided data, the regresssion/reverse regression debate is resolved by Goldberger's (1994) analysis at least for the case of one productivity attribute. With firm provided data, the analyst has access to at least a subset of the exact variables firms used to make decisions. Goldberger shows that in instances where the some of actual variables used by the firm are measured without error, but other factors considered by the firm are unobserved, regression analysis is unbiased, and reverse regression is biased. In this type of data, the measurement error in the provided information is likely to be quite small. Rather, the main concern is that some factors that influence compensation may not be present in the firm=s database. If these factors are uncorrelated with race, regression analysis provides unbiased estimates. Of course, if these unmeasured factors are correlated with race, the regression analysis will suffer from omitted variable bias, but as discussed earlier reverse regression suffers from the same bias and that bias operates in the same direction.

In fact, the debate over reverse regression bears striking resemblance to the debate over the performance approach for situations where firms have an on-going relationship with their customers or employees. When performance can be measured on one dimension, reverse regression and regression estimates will bracket the actual level of discrimination when the model suffers from measurement error and does not suffer from omitted variable bias, and performance analysis and regression estimates will bracket discrimination when the model only suffers from omitted variable bias. The existing evidence suggests, especially for firm provided data, that omitted variable bias is the more important of the two problems, and accordingly we believe that reverse regression is probably less important than performance analyses as a approach for improving our understanding of discrimination.

4.3. Race and the Legal System

Knowles, Persico, and Todd (2001) develop a model of police and motorist behavior where the police officer observes race and an additional set of motorist information (C) that together determine the return to carrying contraband, probability of being searched, and the penalty if caught. Motorists will carry contraband if the expected payoff is positive and will randomize between carrying and not carrying if the expected payoff is zero. Police calculate the probability of contraband being carried based on race and C and compare this to the cost of motorist search. If the probability exceeds motorist search costs for a given value of C*, the police will search all motorists with C greater than C* and randomize when C equals C*. The authors construct an equilibrium in which the probability of carrying contraband equals the cost of motorist search for all C. They argue that no other equilibrium can exist when motorists are free to choose whether to carry contraband. Motorist who receive a high payoff from carrying contraband (high value of C) will realize that they will be stopped by police with certainty and as a result will choose not to carry contraband. This creates a negative relationship the likelihood of carrying contraband and C, which leads to police not to search high C motorists who under those circumstances will choose to carry contraband.

The features of this equilibrium are quite useful in order to establish the empirical implications of police racial prejudice on the arrest and conviction outcomes arising from searches for contraband. They conclude that the probability of guilt conditional on being searched does not vary by race implying that a simple comparison of conviction rates arising from arrests, i.e. the effectiveness or performance of arrests, by race is sufficient to test for prejudice. It should be noted that this is true for any observed variable. Unless police are prejudice against a group or people with a given attribute, conviction rates should not vary across any socio-economic or demographic class, or even by classic indicators of criminal activity like presence of cell-phones, tinted windows, or leased vehicles. They further extend the model to allow for a random, unobserved motorist component in the return to carrying contraband that creates uncertainty in whether any given motorist is carrying contraband, a random utility model. Their results hold. If the return to carrying contraband increases with C, police will still establish a threshold on C above which they search all motorists. In equilibrium, the return to carrying contraband must be independent of C for all motorists, and the probability of being searched is the same for all values of C. In fact, the independence of guilt from C is the key to their principle result because if guilt depended on C the omission of C from the racial comparison would create an omitted variable bias. This second implication creates interesting prospects for validating the predictions of their theoretical model. See Han (2003) for a similar exercise in the context of the mortgage market.

While the addition of a stochastic error to the motorist's problem is nice, the key problem with the model stems from the deterministic nature of the likelihood of police search. Is it reasonable to believe that police officers are able to observe all potential criminals and systematically search with probability one each one who possesses observable characteristics above a given threshold? This scenario seems quite unlikely to me. The simplest modification to Knowles, Persico, and Todd's is one where the police only see some fraction X of all motorists. In this case, police can still set the probability of search in order to assure randomization by low C households, but at some level of C the probability of search required for randomization is X, which can only be reached by setting the probability of search upon contact or detection equal to one. For values of C above this level, motorists carry contraband with certainty. In equilibrium, the racial distribution of guilt will depend upon the within race distribution of C, and if the minority distribution is more concentrated over high values of C minority conviction rates will be higher. Therefore, the simple conviction test proposed in the paper is invalid for this model. Their result can also be overturned if there is random variation in the likelihood of search or if the model that determines search varies across different police officers in the same department.

The paper applies data that was collected as part of a federal lawsuit files in February 1993 by the ACLU against the Maryland State Police. The data set consists of all motor vehicle searches on a section of Interstate 95 in Maryland between 1995 and 1999. They find little evidence of differential guilty rates between white and African-Americans for their baseline definition of guilty, but they do find that Hispanics and white women are less likely to be guilty suggesting prejudice against these groups of motorists. They also check whether type of vehicle or time of day affect the likelihood of being guilty as a validation of the underlying model. They do not find any evidence that guilt varies by these variables except for luxury model car, which is only significant at the 6% level. Except for the finding of discrimination against white women, these results provide validation of their underlying model and insulate their findings somewhat against the alternative theoretical model presented here.

Many more differences arise, however, for more stringent definitions of guilt, e.g. dropping cases involving less than 2 grams of marijuana. There is evidence of discrimination against white males and against third party vehicles in terms of these more serious offenses. In principle, their model might be extended to multiple acts over which the returns and costs vary, and in equilibrium motorists are randomizing over all acts. In such a model, however, the sum of different offense probabilities should be weighted by severity of offense. This implication is violated by the finding that guilty rates for the unweighted sum are equal across groups, but unequal as severity of offense increases. In fact, for African-Americans, the racial difference increases monotonically with offense severity. These results are consistent with the model in which police do not observe and therefore do not have the opportunity to stop all motorists. In that model, motorists and police only randomize over a set of intermediate values of C, where a probability of search upon detection less than one is sufficient for a motorist to be indifferent between carrying and not carrying contraband in equilibrium.

Dharmapala and Ross (2003) develop such a model of offender and police behavior and analyze the empirical implications of the extended model. As discussed above, their model assumes that police only observe potential offenders with some probability, allows for two levels of offense severity, and for simplicity restricts offender heterogeneity to simply two types. In this model, three equilibria exist where outcomes are symmetric by race and both types of offenses are observed: one type randomizes over no offense and committing the severe offense and the other randomizes over no offense and committing the mild offense, one type randomizes over no offense and committing the severe offense and the other type commits the mild offense with certainty, and one type randomizes over no offense and committing the mild offense and the other type commits the severe offense with certainty. The Knowles et. al. test only applies for the case where both types randomize. Dharmapala and Ross reanalyze the Maryland ACLU data and find that it is consistent with no discrimination for pure randomization equilibrium, with reverse discrimination for the equilibrium where minor offenses are committed with certainty and with discrimination for the equilibrium where severe offenses are committed with certainty. There is no way to reject one model in favor of the other, but it seems relevant that such a simple change to the model can reverse the empirical findings.

Borooah (2001) performs a similar analysis using aggregate data on stops in England. He finds that traffic stops are substantially more frequent in predominantly African-American neighborhoods, but also finds that the resulting arrests rates are similar between African-Americans and whites. He concludes that police are practicing statistical discrimination searching African-Americans more frequently because African-American searches are more productive (unconditionally rather than on the margin) in terms of leading to arrests. His empirical results, however, are no different than the findings of

Knowles et. al. which as discussed earlier cannot be used to rule out the existence or prejudiced-based discrimination. Moreover, his findings also might be consistent with no discrimination of any type because police may be responding to non-race based variables that Borooah cannot observe rather than using race as a signal for the unobserved likelihood of being guilty.

In this context, the analysis of crime rates by Donohue and Levitt (NBER 1998) should be reconsidered. The appealing feature of their analysis is the use of a panel that controls for police department/city fixed effects, which controls for the time-invariant component of the distribution of observable predictors of offense/violation and time-invariant racial differences in that distribution. Their study seems to suggest that hiring more non-white police officers shifts arrests away from non-whites towards whites, and at least for violent crime this shift is also associated with an overall increase in crime rates. Such results would be consistent with rational racial profiling or statistical discrimination by white police in terms of minimizing crime rates. If the hiring of non-white police officers reduces the level of discrimination, the resulting shift in arrest patterns leads to higher crime rates. Under this interpretation, their results are consistent with Knowles, Persico, and Todd in that they find no evidence that prejudice underlies racial profiling for violent crimes. On the other hand, Donohue and Levitt do have evidence that prejudice may be the underlying cause of racial profiling against non-whites for property crime in predominantly non-white cities. Again, the strength of their analysis arises from the fact that they were able to use a cross-city panel in order to eliminate biases that arise from racial differences in the distribution of suspect attributes that are observed by the police. The key limitation of this analysis is that these interpretations rely on a link between arrests and crime rates that is not explicitly considered in their analysis. Regardless, their analysis suggests the unrealized potential that might exist in the collection of national level data that is relevant to studying racial profiling.

Ayres and Waldfogel (2001b) examine the setting of bail amounts and bail bondsman rates in the state of Connecticut for racial disparities. They argue that the major factor underlying both bail amount and bail bondsman rates is risk of flight: bail amount based on statute and bail bondsman rates due to the increased monitoring costs and increased risk of flight. They note that direct regression analyses of bail amounts may be biased by the omission of key variables that explain flight risk, are observed by the court, and are not observed by the researcher. They suggest that an examination of bail bondsmen rates might provide information concerning how the market evaluates the flight risk of individuals and somehow be insulated against omitted variable bias. They find that bail is set 35 percent higher and bondsman rates are set 19 percent lower for African-American males after controlling for observed information on offense severity and category. It is important to note that this models differs from the traditional performance approach in that is based on a selection process, e.g. lower mortgage defaults for a groups due to higher underwriting standards. In this case, the vast majority of applicants to bail bondsmans are covered by a bond and any bias in the amount of a bond has no effect on the individual=s underlying flight risk because they have no money at stake once they pay their non-refundable fee to the bondsman. Rather, the logic behind this approach is that the same unobserved propensity to default underlies the behavior of both the court and the bondsman.

As should be clear from earlier discussions, I disagree with the assertion that the analysis of bail bondsman rates avoids the omitted-variable bias. It suffers from exactly the same omitted variables bias as the bail regression. Unobservables that imply that African-Americans have higher flight risk and are observed and considered by both courts and bail bondsman will result in a positive bias of the race coefficient in both the bail and rate regressions. The value of their analysis arises from the estimation of both the bail and the rates equations when most performance analyses do not examine the first stage process. If bail bondsmans are pure profit maximizers, the 19 percent lower rate for African-Americans suggests that the underlying propensity for African-Americans to flee lower than whites after controlling for observables, and yet the courts set bail 35 percent higher for this group even though they are less likely to flee after controlling for observable characteristics. As long as both courts and bail bondsman observe the same set of information, this combined analysis provides fairly compelling evidence that bail is set too high for African-American relative to white defendants.

Of course, interpretation of these results may be complicated by the fact that the races differ in flight risk even after controlling for all variables that are observed by the court and bondsman. For example, as suggested by Ayres and Waldfogel, African-Americans may be much less likely to flee after conditioning on all variables observed by the court possibly because African-Americans on average have fewer contacts outside their current city of residence. If the court treats white and African-Americans exactly the same, the 35 percent racial difference must be due to omitted variable bias. Finally, if this is the case, the bail bondsman being a pure profit maximizer statistically discriminates against whites and the discrimination effect dominates the omitted variable effect in the rate equation. Ayres and Waldfogel suggest that even with this possible interpretation the results demonstrate that the current bail setting system in Connecticut has a disparate impact on African-Americans.

As discussed earlier in the section on regression analysis, a major difficulty in the study of discrimination within in the legal system is that administrative files usually do not contain any information about the populations from which police decided selected individuals to stop or arrest. As a result, it is impossible to directly model the decision making process of police officers. This limitation of administrative data appears to make performance analysis an attractive alternative to direct regression because all information is available for modeling performance, but in truth it simply forces both direct regression and performance analysis to consider the patterns found in aggregate data. Both Aryes (2000) and Knowles, Persico, and Todd (2001) are forced to examine the racial share of offenses using a direct approach and performance analysis, respectively. Both studies suffer from a basic omitted variable problem in that they do not know the underlying distribution from which this share arose. Knowles, Persico, and Todd attempt to solve this problem by examining the implications of a specific theoretical model, but this survey shows how simple modifications to the model can change the implications of the model. Aryes and Waldfogel (2001b) provide an example in which information is available on explicit treatment of minorities in a legal setting, but differential bail rates is a dramatically less pressing social issue than racial profiling, which Aryes and Knowles et. al. consider.

4.4. Conclusion

As discussed above, performance approaches have received the most attention in the mortgage market. These papers, e.g. Berkovec, Gabriel, Canner and Hannan (1994, 1998), argue that if minorities are held to a higher standard in mortgage underwriting then minority loans will have lower rates default. Three key concerns arise with the default approach (Ross, 1996, 1997). First, some variables that are observed by the lender may be omitted from the default model. This concern is relatively minor since it simply puts the default model on the same footing as direct regression, which may also omit important underwriting variables. Second, borrower variables that are unobserved by both the lender and the researcher may be correlated with default leading to an additional omitted variable bias in performance studies that do not arise for direct regression. Finally, the default approach is based on bias caused by the selection of applications into a sample of loans, but selection bias will only exist when some underwriting variables are omitted. Since these omitted underwriting variables are likely correlated with race, the default approach can only be insulated against omitted variables bias when it has no power to detect discrimination. As a result, we recommend against the use of single equation performance approaches for studying discrimination in screening or any other discrete firm choice.

Performance analysis appears to be a much more viable approach to studying discrimination in the price stage. A performance analysis of price or wage discrimination regresses performance on the observed price and other relevant control variables. Higher performance by minorities after controlling for price suggests that minorities are under compensated for their performance, i.e. discriminated against. The selection problem does not arise because discrimination directly influences the price. Of course, when prices are set prior to the observation of performance, as is the case in the mortgage industry, racial differences in price would only capture racial differences in performance if price was in fact adjusted to control for unobserved risk factors that are correlated with race, i.e. statistical discrimination. Under these circumstances, performance based studies can only detect discrimination if it exceeds the level expected based on firms practicing statistical discrimination.

In markets like the labor market where the firm has an on-going relationship with the individual or household, however, the firm has the opportunity to adjust prices in response to observed performance. This adjustment eliminates the concern about omitted variables that are not observed by the firm because the firm can implicitly account for those variables, to the extent that they matter, simply by basing price on observed performance. As a result, this type of performance analysis is really on an equal footing with direct regression. In fact, the direct and regression analyses essentially are the same model with performance and wages switched between the right and left hand sides of the equations. In this way, these models are comparable to the models examined by Ross (2000) and Van Order and Zorn (1995) that combine information on firm decisions and individual performance. Also, while it has not been formally evaluated, the estimation of both direct price and performance regressions may bound the true level of discrimination in the presence of omitted variable bias, much as direct and reverse regression bound the true effect in the presence of measure measurement effort. As with reverse regression, however, these approaches cannot be interpreted as offering bounds when performance is described by multiple measures.

The interpretation of performance studies in the labor market is complicated by the fact that most rely on data aggregated up to the firm rather than data on outcomes for individual workers. For example, Szymanski (2000) estimates a model of the performance of English soccer teams controlling for both the racial composition of the team and the wage bill of these teams. Studies of this type have the potential too dramatically increase the omitted variable problem relative to a sample composed of individual workers at specific firms. Not only might there be racial differences in the unobservables, but also firms may draw from very different applicant pools with different racial compositions and different racial differences in the separate pools. Essentially, this concern builds on our earlier discussion in the regression section on heterogeneity in behavior or product to heterogeneity in applicant pool. The labor market is a differentiated product market, and different firms may draw from different segments of this market. In our opinion, the key approach for addressing the concern is the use of fixed effects over space and/or preferably over time under the assumption that the heterogeneity in a firms applicant pool does not vary systematically with minority status over time or over space, see Holzer and Ihlanfeldt (1998) for a study that uses spatial fixed effects.

These lessons concerning aggregate data and the performance approach are very relevant to studyings of racial profiling. The biggest obstacle in providing statistical evidence of racial profiling is the inability of researchers to describe the underlying pool of potential offenders from which citations, motor vehicle stops, searches, arrests and convictions arise. Due to basic practical constraints, all administrative data on law enforcement provides only a sample of those who were actually stopped, searched, or arrested. This limitation of administrative data appears to make performance analysis an attractive alternative to direct regression because all information is available for modeling performance, but in reality it simply forces both direct regression and performance analysis to consider the patterns found in aggregate data. Due to these limitations, the choice of a base for comparing racial differences in arrests or stops is crucial to studying racial profiling no matter which statistical approach is used and very little work has been done to search for a proper solution to this question.

In my opinion, the use of panel techniques may be very valuable for such analyses. For example in Knowles et. al. (2001), a division of the sample by roadway section and time of day and the use of fixed effects to control for these factors is likely to provide substantial controls for underlying variation in the likelihood of carrying drugs. Similarly, with Aryes (2000), the use of neighborhood/time of day fixed effects would make his analysis much more convincing. Alternatively, courts might require more aggressive plans for documenting racial differences in treatment by police officers. For example, police departments might be required collect detailed records of the pattern of citations from checkpoint and other events that involve universal pull-overs separate from overall citation rates.

5. Summary and Conclusions

This survey has reviewed the empirical literature within economics on discrimination in four major markets: the labor, housing, and goods markets, as well as discrimination in law enforcement and criminal justice. The survey also reviewed three major methodological approaches for the four markets: direct regression analysis, paired testing, and performance analysis. A number of important conceptual issues arose from this review that are important for evaluating all three methodologies. First, the validity of interpreting observed racial differences as discrimination depends heavily on the type of sample used in the analysis and the analyst's ability to control for characteristics of that sample. Heterogeneous firm behavior also confounds empirical analyses of discrimination both in terms of providing an accurate measure of disparate treatment and in terms of missing the possibility for disparate impact discrimination. Finally, many markets that are very important in the study of discrimination, such as the housing and labor markets, are differentiated product markets. In these types of markets, short term surplus or shortages in different regions of attribute space can lead to substantial differences in prices across groups that may have nothing to do with discrimination.

Most empirical research on discrimination uses either samples of households or individuals that are nationally or regionally representative, samples of firms that are national or regionally representative, or samples of market transactions. Analyses based on representative samples of households do not contain direct information about the behavior of a potentially discriminating firm or institution and usually only contain noisy proxies for the variables actually considered by firms. As a measure of discrimination, estimated racial differences in outcomes from individual based samples, as are often used in analyses of the labor market, are likely to suffer from both severe omitted variable and measurement error bias.

Analyses of firm or transaction level data have the advantage that they focus on the potential discriminator and on the transactions in which discrimination may take place. Estimations based on a sample of firms are inherently aggregate in nature examining average racial differences in compensation or racial share of employment. These samples tend to contain very little information about the individual and so may suffer from omitted variable bias if the firms being compared draw from different submarkets and as a result have pools of customers, applicants, or employees that differ on unobservable. One effective solution to this problem is to control for the pool by including submarket fixed effects, such as spatial and job type controls or firm fixed effects over time. Transaction level data, which has been used extensively in the mortgage market, appears to offer the best opportunity to directly test for discrimination. The data usually contain the actual firm screening or price decision, as well as at least a subset of the exact variables used by the firm during its decision making process.

Firms in the same market are often quite heterogeneous and as a result may differ from each other in their policies and their treatment of similar applicants, customers, clients, and/or employees. Analyses that pool observations across individuals who work at or do business with different firms, either explicitly for firm and transaction data or implicitly for household and individual data, do not have the ability to distinguish between disparate treatment of minorities and the disparate impact created by across firm differences in behavior. Moreover, in many data sets, especially those based on samples of individuals or households, the research has no information available to assist in determining whether any disparate impact is justified by business necessity. Within firm comparisons are required to identify disparate treatment discrimination. In additions, analyses that pool across firms are dramatically strengthened by the incorporation of firm fixed effects as well as interactions between firm decision variables and firm characteristics over which the firm decision process might be expected to vary. The most powerful approach to addressing this problem in pooled samples is to include information on firm specific expectations of performance, see Ross and Yinger (2002) for a detailed discussion.

Finally, housing, labor, and automobile markets are examples of markets that involve differentiate products. Differentiated product markets are described by competition between close substitutions that differ over an attribute space. In such markets, localized shortages and surpluses within attribute space can lead to large price variations within the market, and the specific choices made by consumers can have a large effect on price. As a result, it is very difficult to distinguish between price differences that arise due to discrimination either through price or exclusion from segments of the market and price differences that arise from across group differences in preferences. In fact, this problem is so severe in the housing market that analyses of racial differences in housing price are never discussed as providing evidence of price discrimination, but rather as evidence that minorities are excluded from some segments of the market.

The existing literature does not offer any easy solutions to this problem. The identification of hedonic models is quite difficult even in unconstrained markets. Even the best econometric models for these problems are not designed to cope with the existence of loosely defined and unmeasured barriers that limit an individual's choices within a market. One reasonable approach if the data is available is to identify racial differences based on comparisons within submarkets defined as sets of products that are very near each other in attribute space. For example in housing, spatial fixed effects can be used to test whether whites and minorities pay the same price for housing in the same location. If not, these differences may be attributable to a negotiation disadvantage for minorities as a result of having less options available within the overall market.

Comparisons within submarkets, however, may seriously understate the overall effect of discrimination because discrimination may cause gender, racial or ethnic stratification across these submarkets. The crowding of minorities into specific submarkets is likely to increase prices or decrease wages in those markets, and a comparison within submarkets will not capture those price differences. A complete solution to this problem requires the explicit consideration of group differences in preferences, but the existing literature provides very little if any guidance on the robust estimation of models that include both heterogeneity in preference and discrimination. Manski's (2000) advise to researchers who are attempting to study social interactions may be especially relevant here. He suggests that subjective problems most likely require subjective data and argues that rather than trying to infer preferences from outcomes researchers should attempt to elicit them directly.

Tests for discrimination based on the analysis of performance are subject to all of the concerns raised above concerning traditional regression analysis. Performance analyses based on either transaction data or firm level data with careful controls for the firm's pool of potential customers or employees are likely to provide the most compelling evidence. Even under those circumstances, performance analyses based on samples that are pooled across firms and products will yield biased effects unless appropriate fixed effects are included in the specification.

Performance tests also suffer from additional problems. Performance tests for discrimination in the screening stage obtain their power from selection bias, but selection bias only exists when variables are omitted from the specification that influence screening, omitted variable bias. As a result, we believe that the performance result should not be used to study discrimination in screening. Performance analyses of price discrimination often cannot be directly compared to direct analyses of price discrimination because a consistent price model only requires variables that are observed by the firm, but in a performance, model group differences may be influenced by unobserved factors that are never observed by the firm. The key situation where performance analysis is on an equal footing with direct regression of outcomes is when firms have a continuing relationship with customers or employees and are able to adjust prices as performance is observed.

The paired testing approach offers a very robust solution to the problems of firm and product heterogeneity discussed above. Testing protocols usually are designed to assure that observationally equivalent testers (except for group status) visit the same firm and request information concerning the same product. The debate concerning testing has mostly focused on omitted variable bias and the matching of testers, but just as with outcome data differences across testers can be handled in a regression context. In our opinion, the most important feature of testing is the matching by firm and initial request, a match that is nearly impossible to accomplish in regression analysis of outcomes data.

The paired testing approach does suffer from a number of limitations. Paired testing ignores disparate impact discrimination and can only test discrimination in early stages of the market. As a result, testing must be seen as a complement to regression or performance analyses based on administrative or other publicly available data sets. In addition, a sample of tests is drawn based on specific strategy for gaining entrance to the market, and the strategy may be influenced as much or more by implementation constraints than by the exact context in which discrimination is being measured. As a result, the resulting sample of tests may not describe the discrimination experienced by the average minority in the population, but rather often describes the experience of a "qualified" minority who approaches a randomly selected firm or pursues a randomly selected advertisement.

6. References

Altonji, Joseph G. 1995. "The Effects of High School Curriculum on Education and Labor Market Outcomes." *Journal of Human Resources* 30(3): 409-438.

Altonji, Joseph G. 1993. "The Demand for and Return to Education When Education Outcomes Are Uncertain." *Journal of Labor Economics* 11(1): 48-83.

Altonji, Joseph G., and Rebecca M. Blank. 1999. "Race and Gender in the Labor Market." In O.C. Ashenfelter and D. Card, editors, *Handbook of Labor Economics,* 3C. North-Holland, pp. 3144-3259.

Altonji, James G., and Charles R. Pierret. 2001. "Employer Learning and Statistical Discrimination." *The Quarterly Journal of Economics* CXVI(1) (February): 313-348.

Altonji, Joseph G., and James R. Spletzer. 1991. "Worker Characteristics, Job Characteristics and the Receipt of On-the-Job Training." *Industrial and Labor Relations Review* 45(1): 58-79.

Arrow, Kenneth. 1973. "The Theory of Discrimination." In O. Ashenfelter and A. Rees, editors, *Discrimination in Labor Markets.* Princeton, NJ: Princeton University Press, pp. 3-42.

Avery, Robert B., Patricia E. Beeson, and Mark S. Sniderman. 1996. "Accounting for Racial Differences in Housing Credit Markets." In J. Goering and R.Wienk, *Mortgage Lending, Racial Discrimination, and Federal Policy.* Washington, DC, Urban Institute Press, pp. 75-142.

Avery, Robert B., Patricia E. Beeson, and Paul D. Calem. 1997. "Using HMDA Data as a Regulatory Screen for Fair Lending Compliance." *Journal of Financial Services Research,* 11(2) (April): pp. 9-42.

Ayres, Ian. 2001a. "Expert Report." Case No. 3-98-0223, *Carson v. Nissan,* U.S. District Court for the Middle District of Tennessee, May 25.

Ayres, Ian. 2001b. *Pervasive Prejudice? Unconventional Evidence of Race and Gender Discrimination.* Chicago: University of Chicago Press.

Ayres, Ian, and J. Waldfogel. 2001. "A Market Test for Race Discrimination in Bail Setting." In *Pervasive Prejudice? Unconventional Evidence of Race and Gender Discrimination.* Chicago: University of Chicago Press. University of Chicago Press.

Ayres, Ian. 2000. An Analysis of Racial Disparities in the Practices of the Tulsa Police Department. Report commissioned by plaintiff.

Ayres, Ian. 1995. "Further Evidence of Discrimination in New Car Negotiations and Estimates of Its Cause." *Michigan Law Review* 94(1): 109-147.

Ayres, Ian. 1991. "Fair Driving: Gender and Race Discrimination in Retail Car Negotiations." *Harvard Law Review* 104(4): 817-873.

Ayres, Ian, and Peter Siegelman. 1995. "Race and Gender Discrimination in Bargaining for a New Car." *American Economic Review,* 85(3) (June): pp. 304-321.

Bayer, Patrick, Robert McMillan, and Kim Rueben. 2002. "The Causes and Consequences of Residential Segregation: An Equilibrium analysis of Racial Sorting." Unpublished manuscript.

Becker, Gary S. 1971. *The Economics of Discrimination.* Second edition. Chicago: University of Chicago Press. (August).

Becker, Gary S. 1993a. "The Evidence Against Banks Doesn't Prove Bias." *Business Week.* April 19, p. 18.

Becker, Gary S. 1993b. "Nobel Lecture: The Economic Way of Looking at Behavior." *Journal of Political Economy* 101(3): 385-409.

Berkovec, James A., Glenn B. Canner, Stuart A. Gabriel, and Timothy H. Hannan. 1994. "Race, Redlining, and Residential Mortgage Loan Performance." *Journal of Real Estate Finance and Economics,* 9(3) (November): pp. 263-294.

Berkovec, James A., Glenn B. Canner, Stuart A. Gabriel, and Timothy H. Hannan. 1998. "Discrimination, Competition, and Loan Performance in FHA Mortgage Lending." *Review of Economics and Statistics,* 80(2) (May): pp. 241-250.

Berry, S., J. Levinsohn, and A. Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63: 841-890.

Black, Harold A., Robert L. Schweitzer, and Lewis Mandell. 1978. "Discrimination in Mortgage Lending." *American Economic Review* 68(2) (May): 186-191.

Blackburn, McKinley, and Todd Vermilyea. 2001. "Racial Discrimination in Home Purchase Mortgage Lending Among Large National Banks." Unpublished manuscript.

Blau, Francine D., and Lawrence M. Kahn. 1997. "Swimming Upstream: Trends in the Gender Wage Differential in the 1980's." *Journal of Labor Economics* 15(1, part 1): 1-42

Blau, Francine D. 1998. "Trends in the Well-Being of American Women, 1970-1*995*." *Journal of Economic Literature* 36(1): 112-65.

Blau, Francine D., Patricia Simpson, and Deborah Anderson. 1998. "Continuing Progress? Trends in Occupational Segregation in the United States in the 1970's and 1980's." Working Paper # 6716. Cambridge, MA: National Bureau of Economic Research.

Blau, Francine D., and Andrea H. Beller. 1988. "Trends in Earnings Differentials by Gender (1971-81)." *Industrial and Labor Relations Review* 41(4): 513-529.

Boggs, Roderic V. O., Joseph M. Sellers, and Marc Bendick Jr. 1993. "Use of Testing in Civil Rights Enforcement." In Michael Fix and Raymond J. Struyk, editors, *Clear and Convincing Evidence.* Washington, DC: Urban Institute Press.

Borooah, V.K. (2001) "Racial bias in police stops and searchers: An economic analysis." *European Journal of Political Economy*, 17, 17-37.

Bound, John, and Harry J. Holzer. 1993. "Industrial Shifts, Skill Levels and the Labor Market for White and Black Males." *The Review of Economics and Statistics* 75(3): 387-396.

Bound, John, and Richard B. Freeman. 1992. "What Went Wrong? The Erosion of Relative earnings and Employment Among Young Black Men in the 1980's." *Quarterly Journal of Economics* 107(1): 201-232.

Bratsberg, Bernt, and Dek Terrell. 1998. "Experience, Tenure, and Wage Growth of Young Black and White Men." *Journal of Human Resources* 33(3): 658-683.

Brown, Charles, and Mary Corcoran. 1997. "Sex Based Differences in School Content and the Male/Female Wage Gap." *Journal of Labor Economics* 15(3, part1): 431-465.

Brown, Lynn E. and Geoffrey M. Tootell. 1995. Mortgage lending in Boston: A response to the critics. *New England Economic Review* Sept/Oct: 53-78.

Buist, Henry, Peter Linneman, and Isaac F. Megbolugbe. 1999. "Residential Lending Discrimination and Lender Compensation Policies," *Journal of the American Real Estate and Urban Economics Association,* 27(4) (Winter): pp. 695-717.

Cameron, Stephen V., and James J. Heckman. 1993. "The Nonequivalence of High School Equivalents." *Journal of Labor Economics* 11(1): 1-47.

Card, David, and Thomas Lemieux. 1996. "Wage Dispersion, Returns to Skill and Black-White Wage

Differentials." *The American Economic Review* 74: 319-361.

Card, David, and Alan B. Krueger. 1992. "School Quality and Black-White Relative Earnings: A Direct Assessment." *Quarterly Journal of Economics* 107: 151-200.

Carr, James H. and Isaac F. Megbolugbe. 1993. The Federal Reserve Bank of Boston study of mortgage lending discrimination revisited. *Journal of Housing Research* 4: 277-313.

Chambers, Daniel N. 1992. "The Racial Housing Price Differential and Racially Transitional Neighborhoods." *Journal of Urban Economics* 32: 214-232.

Charles, Kerwin Kofi, and Erik Hurst. 2000. "The Transition to Home Ownership and the Black-White Wealth Gap." *Review of Economics and Statistics* 8(4): 281-297.

Chay, Kenneth Y., and David S. Lee. 2000. "Changes in Relative Wages in the 1980s: Returns to Observed and Unobserved Skills and Black-White Wage Differentials." *Journal of Econometrics* 99(1): 1-38.

Couch, Kenneth and Mary Daly. 2002. Black-white wage inequality in the 1990s : A decade of progress. *Economic Inquiry* 40 : 31-41.

Courant, Paul N. 1978. "Racial Prejudice in a Search Model of the Urban Housing Market." *Journal of Urban Economics* 5(3) (July): 329-345.

Courchane, Marsha, David Nebhut, and David Nickerson. 2000. "Lessons Learned: Statistical Techniques and Fair Lending." *Journal of Housing Research* 11(2): 277-295.

Courchane, Marsha, and David Nickerson. 1997. "Discrimination Resulting from Overage Practices." *Journal of Financial Services Research* 11(1-2) (April): 133-151.

Crawford, Gordon W. and Eric Rosenblatt. 1999. "Differences in the Cost of Mortgage Credit: Implications for Discrimination." *Journal of Real Estate Finance and Economics* 19(2) (September): 147-159.

Cross, Harry, G. Kenny, J. Mell, and W. Zimmerman. 1990. *Employer Hiring Practices: Differential Treatment of Hispanic and Anglo Job Seekers.* Washington, DC: Urban Institute Press.

Cutler, David M., Edward L. Glaeser, and Jacob L. Vigdor. 1999. "The Rise and Decline of the American Ghetto." *Journal of Political Economy* 107(3): 455-506.

D'Amico, Ronald and Nan Maxwell. 1994. The impact of post-school joblessness on male black-white wage differentials. *Industrial Relations* 33: 184-205.

Day T. and Stan J. Liebowitz. 1996. Mortgages, Minorities, and HMDA. Paper presented at the Federal Reserve Bank of Chicago. April 1996.

Darity, William A. Jr., and Patrick L. Mason. 1998. "Evidence on Discrimination in Employment: Codes of Color, Codes of Gender." *Journal of Economic Perspectives* 12(2) (Spring): 63-90.

Darity, William, Jr, David K. Guilkey, and William Winfrey. 1996. "Explaining Differences in

Economic Performance among Racial and Ethnic Groups in the USA: The Data Examined." *American Journal of Economics and Sociology* 55(4): 411-425.

Deng, Yongheng, Stephen L. Ross, and Susan M. Wachter. In Press. "Racial Differences in Homeownership: The Effect of Residential Location." *Regional Science and Urban Economics*.

Dharmapala, Dhammika and Stephen L. Ross. 2003. Racial bias in motor vehicle searches: Additional theory and evidence. Unpublished manuscript.

Dietrich, Jason. 2001. "Missing Race Data in HMDA and the Implications for the Monitoring of Fair Lending Compliance." Economic and Policy Analysis Working Paper 2001-1. Washington, D.C.: Office of the Comptroller of the Currency:

Donohue, John J., and James Heckman. 1991. "Continuous Versus Episodic Change: The Impact pf Civil Rights Policy on the Economic Status of Blacks." *Journal of Economic Literature* 29: 1603-1643.

Donohue, John J., and Steven D. Levitt. 1998. "The Impact of Race on Policing, Arrest Patterns and Crime." Working Paper No. 6784. Cambridge, MA: National Bureau of Economic Research.

Duca, John V., and Stuart S. Rosenthal. 1994. "Borrowing Constraints and Access to Owner-Occupied Housing." *Regional Science and Urban Economics* 24(3) (June): 301-322.

Epple, Dennis. 1987. "Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products." *Journal of Political Economy* 95: 828-859.

Equal Employment Opportunity Commission. 1978. *Uniform Guidelines on Employee Selection Procedures*. Code of Federal Regulations, Title 41, Volume 1, Part 60-3. Washington, D.C.: U.S. Government Printing Office.

Filer, Randall K. 1993. "The Usefulness of Predicted Values for Prior Work Experience in Analyzing Labor Market Outcomes for Women." *Journal of Human Resources* 28(3): 519-537.

Follain, James R., and Stephen Malpezzi. 1981. "Another Look at Racial Differences in Housing Prices." *Urban Studies* 18: 195-203.

Foster, Angela Williams, Faith Mitchell, and Stephen E. Fienberg. 2002. *Measuring Housing Discrimination in a National Study: Report of a Workshop*. Washington D.C.: National Academy Press.

Galster, George C. 1996. "Comparing Loan Performance between Races as a Test for Discrimination." *Cityscape: A Journal of Policy Development and Research* 2(1): 33-39.

Glennon, Dennis and Mitchell Stengel. 1994. An evaluation of the Federal Reserve Bank of Boston's study of racial discrimination in mortgage lending. Economic and Policy Analysis Working Paper No. 94-2, Office of the Comptroller of the Currency, Washington, D.C.

Godfrey, Erin, Stephen L. Ross, and Margery A. Turner. 2003. Tester heterogeneity and paired testing in the mortgage market. Unpublished manuscript.

Goldberg, Pinilopi Koujianoi. 1996. "Dealer Price Discrimination in New Car Purchases: Evidence from the Consumer Expenditure Survey." *Journal of Political Economy* 1043: 622-654.

Goldberger, Arthur S. 1984. "Reverse Regression and Salary Discrimination." *Journal of Human Resources* 19(3): 293-318.

Goldsmith, Arthur H., Jonathan Veum, and William Darity Jr. 1997. "The Impact of Psychological and Human Capital on Wages." *Economic Inquiry* 35(4) (October): 815-829.

Gottschalk, Peter. 1997. "Inequality, Income Growth, and Mobility: The Basic Facts." *Journal of Economic Perspectives* 11(2): 21-40.

Graddy, Kathryn. 1997. Do fast-food chain prices discriminate on the race and income characteristics of the area. Review of Economics and Statistics 15: 391-401.

Grogger, Jeff, and Eric Eide. 1995. "Changes in College Skills and the Rise in the College Wage Premium." *Journal of Human Resources* 30(2): 280-310.

Gyourko, Joseph, Peter Linneman, and Susan Wachter. 1999. "Analyzing the Relationships among Race, Wealth, and Home Ownership in America." *Journal of Housing Economics* 8(2) (June): 63-89.

Han, Song. 2002. Learning and statistical discrimination in lending. Unpublished Manuscript.

Han, Song. 2003. Discrimination in lending: Theory and Evidence. Unpublished Manuscript.

Heckman, James J. 1998. "Detecting Discrimination." *Journal of Economic Perspectives* 12: 101-116.

Heckman, James J., and Peter Siegelman. 1993. "The Urban Institute Studies: Their Methods and Findings." In M. Fix and R. Struyk, editors, *Clear and Convincing Evidence: Testing for Discrimination in America.* Washington, DC: The Urban Institute Press.

Hellerstein, Judith K., and David Neumark. 1999. "Sex, Wages, and Productivity: An Empirical Analysis of Israeli Firm-Level Data." *International Economic Review* 40(1): 95-123.

Hellerstein, Judith K., David Neumark, and Kenneth R. Troske. 1999. "Wages, Productivity and Worker Characteristics: Evidence From Plant-Level Production Functions and Wage Equations." *Journal of Labor Economics* 17(3): 409-446.

Hellerstein, Judith K., David Neumark, and Kenneth R. Troske. 2002. Market forces and sex discrimination. *Journal of Human Resources* 37: 353-380.

Holzer, Harry J., and Jens Ludwig. 2002. "Measuring Discrimination in Education: Are Methodologies From Labor and Housing Markets Useful?" Paper presented at the National Research Council Workshop on Discrimination in Education, July 1, 2002.

Holzer, Harry J. and Keith R. Ihlanfeldt. 1998. "Customer Discrimination and Employment Outcomes for Minority Workers." *Quarterly Journal of Economics* 113(3): 835-867.

Horne, David. 1997. Mortgage lending, race, and model specification. *Journal of Financial Services Research* 11: 43-68.

Ihlanfeldt, Keith R., and Benjamin P. Scafidi. 2002. "Black Self-Segregation as a Cause of Neighborhood Racial Segregation: Evidence From the Multi-City Study of Urban Inequality." *Journal of Urban Economics* 51: 366-390.

James, F., and S.W. DelCastillo. 1991. "Measuring Job Discrimination by Private Employers Against Young Black and Hispanic Males Seeking Entry Level Work in the Denver Metropolitan Area." Unpublished Manuscript, Denver, CO: University of Colorado.

Johnson, William R., and Derek Neal. 1998. "Basic Skills and the Black-White Earnings Gaps." In Jenkins, Christopher and M. Peterson, editors, *The Black-White Test Score Gap.* Washington, DC: Brookings.

Johnson, William R., and Derek Neal. 1996. The roel of premarket factors in Black-White wage differentials. *Journal of Political Economy* 104: 869-895.

Johnson, James H., Jr., Elisa Jayne Bienenstock, and Jennifer A.Stoloff. 1995. "An Empirical Test of the Cultural Capital Hypothesis." *Review of Black Political Economy* 23(4): 7-27.

Juhn, Chinhui , and Kevin M. Murphy. 1997. "Wage Inequality and Family Labor Supply." *Journal of Labor Economics* 15(1, part 1): 72-97.

Juhn, Chinhui. 1992. "Decline of Male Labor Market Participation: The Role of Declining Market Opportunities." *Quarterly Journal of Economics* 107(1): 79-121.

Juhn, Chinhui, Kevin Murphy, and Brooks Pierce. 1991. "Accounting for the Slowdown in Black-White Wage Convergence." In Marvin Kosters, editor, *Workers and Their Wages.* Washington, DC: AEI Press.

Kahn, Lawrence M., and Peter D. Sherer. 1988. "Racial Differences in Professional Basketball Players' Compensation." *Journal of Labor Economics* 6(1): 40-61.

Kiel, Katherine A., and Jeffrey E. Zabel. 1996. "House Price Differentials in U.S. Cities: Household and Neighborhood Racial Effects." *Journal of Housing Economics* 5: 143-165.

King, A. Thomas, and Peter Mieszkowski. 1973. "Racial Discrimination, Segregation, and the Price of Housing." *Journal of Political Economy* 81: 590-606.

King, A. Thomas. 1980. "Discrimination in Mortgage Lending: A Study of Three Cities." Washington, DC: Federal Home Loan Bank Board.

Knowles, John, Nicola Persico, and Petra Todd. 2001. "Racial Bias in Motor Vehicle Searches: Theory and Evidence." *Journal of Political Economy* 109(1) (February): 203-229.

Lawton, Rachel. 1996. "Preapplication Mortgage Lending Testing Program: Lender Testing by a Local

Agency." In John Goering and Ron Wienk, editor, *Mortgage Lending, Racial Discrimination, and Federal Policy.* Washington, D.C.: Urban Institute Press: pp. 611-622.

Leibowitz, Arleen , and Jacob Alex Klerman. 1995. "Explaining Changes in Married Mothers' Employment over Time." *Demography* 32(3): 365-378.

Leonard, Jonathan. 1989. Gender and affirmative action. Journal of Economic Perspectives 3: 61-75.

Light, Audrey, and Manuelita Ureta. 1995. "Early Career Work Experience and Gender Wage Differentials." *Journal of Labor Economics* 13(1): 121-154.

Lundberg, Shelly J. 1991. "The Enforcement of Equal Opportunity Laws under Imperfect Information: Affirmative Action and Alternatives." *The Quarterly Journal of Economics,* 106(1) (February): pp. 309-326.

Lundberg, Shelly J., and Richard Startz. 2000. "Inequality and Race: Models and Policy." In Kenneth Arrow, Samuel Bowles and Steven Durlauf, editors, *Meritocracy and Inequality.* Princeton, NJ: Princeton University Press,

Maddala, G. S., and Robert P. Trost. 1982. "On Measuring Discrimination in Loan Markets." *Housing Finance Review* 1: 245-268.

Manski, Charles F. 2000. "Economic Analysis of Social Interactions." *Journal of Economic Perspectives* 14: 115-136.

Maxwell, Nan. 1994. "The Effect on Black-White Wage Differences of Difference in the Quantity and Quality of Education." *Industrial and Labor Relations Review* 47 (January): 249-264.

Munnell, Alicia H., Geoffrey M. B. Tootell, Lynn E. Browne, and James McEneaney. 1996. "Mortgage Lending in Boston: Interpreting HMDA Data." *American Economic Review* 86(1) (March): 25-53.

Murphey, S.A. 2001. "Audit Studies and the Assessment of Discrimination." In A.W. Foster, F. Mitchell and S.E. Fienberg, editors, *Measuring Housing Discrimination in a National Study: A Report of a Workshop.* Washington, DC: National Academy Press,

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

O'Neill, June. 1990. "The Role of Human Capital in Earnings Differences Between Black and White Men." *The Journal of Economic Perspectives* 4(4) (Fall): 25-45.

Ondrich, Jan, Stephen L. Ross, and John Yinger. 2002a. "Geography of Housing Discrimination." *Journal of Housing Research* 12(2): 217-238

Ondrich, Jan, Stephen Ross, and John Yinger. In Press. Now You See It, Now You Don't: Why Do Real Estate Agents Withhold Houses from Black Customers? *Review of Economics and Statistics*.

Ondrich, Jan, Stephen L. Ross, and John Yinger. 2000. "How Common is Housing Discrimination? Improving on Traditional Measures." *Journal of Urban Economics* 47(2): 470-500.

Ondrich, Jan, Alex Stricker, and John Yinger. 1999. "Do Landlords Discriminate? The Incidence and Causes of Racial Discrimination in Rental Housing Markets." *Journal of Housing Economics* 8(3) (September): 185-204.

Ondrich, Jan, Alex Stricker, and John Yinger. 1998. "Do Real Estate Brokers Choose to Discriminate? Evidence from the 1989 Housing Discrimination Study." *Southern Economic Journal* 64(April): 880-901.

Page, Marianne 1995. "Racial and Ethnic Discrimination in Urban Housing Markets: Evidence From a Recent Audit Study." *Journal of Urban Economics* 38: 183-206.

Paglin, Morton, and Anthony M. Rufolo. 1990. "Heterogenous Human Capital, Occupational Choice and Male-Female Earnings Differences." *Journal of Labor Economics* 8(1, part 1): 123-144.

Peterson, Richard L. 1981. "An Investigation of Sex Discrimination in Commercial Banks' Direct Consumer Lending." *The Bell Journal of Economics,* 12(2) (Autumn): pp. 547-561.

Phelps, Edmund S. 1972. "The Statistical Theory of Racism and Sexism." *American Economic Review,* 62(4) (September): pp. 659-661.

Racine, Jeff, and Paul Rilstone. 1995. "The Reverse Regression Problem: Statistical Paradox or Artifact of Misspecification?" *Canadian Journal of Economics* 28(3): 502-531.

Rodgers III, William, and William Spriggs. 1996. "What Does AFQT Really Measure: Race, Wages, Schooling and the AFQT Score." *The Review of Black Political Economy* 24(4) (Spring): 13-46.

Rodgers III, William, William Spriggs, and Elizabeth Waaler. 1997. "The Role of Premarked Factors in Black-White Differences." Unpublished manuscript. College of William and Mary.

Ross, Stephen, L. In Press. "Segregation and Racial Preferences: An Analysis of Choice Based on Satisfaction and Outcome Measures." *Annales d'Economie et Statistique*.

Ross, Stephen L. 2002. "Paired Testing and the 2000 Housing Discrimination Study." In S. E. Feinberg, F. Mitchell and A.W. Foster, Report editors, *Measuring Housing Discrimination in a National Study.* Washington, D.C.: National Academy Press.

Ross, Stephen L. 2000. "Mortgage Lending, Sample Selection, and Default." *Real Estate Economics* 28: 581-621.

Ross, Stephen L. 1997. "Mortgage Lending Discrimination and Racial Differences in Loan Default: A Simulation Approach." *Journal of Housing Research* 8(2): 277-297.

Ross, Stephen L. 1996. "Mortgage Lending Discrimination and Racial Differences in Loan Default." *Journal of Housing Research* 7(1): 117-126.

Ross, Stephen L., and John Yinger. 2002. *The Color of Credit: Mortgage Lending Discrimination, Research Methodology, and Fair-Lending Enforcement.* Cambridge, MA: MIT Press.

Ross, Stephen L., and John Yinger. 2002. Detecting discrimination. Presented at the National Resource Council Panel on Methods for Assessing Discrimination on July 2, 2002

Ross, Stephen L., and John Yinger. 1999a. "Does discrimination in mortgage lending exist? The Boston Fed study and its critics." In Margery Austin Turner and Felicity Skidmore, editors, *Mortgage Lending Discrimination: A Review of Existing Evidence.* Washington, DC: The Urban Institute, pp. 43-83.

Ross, Stephen L., and John Yinger. 1999b. "The Default Approach to Studying Mortgage Discrimination: A Rebuttal." In Margery Austin Turner and Felicity Skidmore, editors, *Mortgage Lending Discrimination: A Review of Existing Evidence.* Washington, DC: The Urban Institute, pp. 107-127.

Roychoudhury, C., and A.C. Goodman. 1996. "Evidence of Racial Discrimination in Different Dimensions of Housing Search." *Real Estate Economics* 24: 161-178.

Roychoudhury, C., and A.C. Goodman. 1992. "An Ordered Probit Model for Estimating Racial Discrimination Through Fair Housing Audits." *Journal of Housing Economics* 2: 358-373.

Schafer, Robert. 1979. "Racial Discrimination in the Boston Housing Market." *Journal of Urban Economics* 6: 1176-1196.

Schafer, Robert, and Helen F. Ladd. 1981. *Discrimination in Mortgage Lending.* Cambridge, MA: MIT Press.

Scheessele, Randall M. 1998. "HMDA Coverage of the Mortgage Market." Housing Finance Working Paper #HF-007. Washington, DC: U.S. Department of Housing and Urban Development, Office of Policy Development and Research.

Schnare, Ann B. 1976. "Racial and Ethnic Price Differentials in an Urban Housing Market." *Urban Studies* 13: 107-120.

Smith, Robin, and Michelle DeLair. 1999. "New Evidence From Lender Testing: Discrimination at the Pre-Application Stage." In Margery Austin Turner and Felicity Skidmore, editor, *Mortgage Lending Discrimination: A Review of Existing Evidence.* The Urban Institute, pp. 23-41.

Smith, Shanna L. 1993. "The Fair-Housing Movement's Alternative Standard for Measuring Housing Discrimination: Comment." In Michael Fix and Raymond J. Struyk, editors, *Clear and Convincing Evidence.* Washington, DC: Urban Institute Press.

Smith, Shanna L. and Cathy Cloud. 1996. "The Role of Private, Nonprofit Fair Housing Enforcement Organizations in Lending Testing." In John Goering and Ron Wienk, editors, *Mortgage Lending, Racial Discrimination, and Federal Policy.* Washington, DC: Urban Institute Press, pp. 589-610.

Stengel, Mitchell , and Dennis Glennon. 1999. "Evaluating Statistical Models of Mortgage Lending Discrimination: A Bank-Specific Analysis," *Real Estate Economics,* 27(2) (Summer): pp. 299-334.

Szymanski, Stefan. 2000. "A Market Test for Discrimination in the English Professional Soccer Leagues." *Journal of Political Economy* 108(3): 590-603.

Telles, Edward, and Edward Murguia. 1990. "Phenotypic Discrimination and Income Differences Among Mexican Americans." *Social Science Quarterly* 71(4): 682-696.

Turner, Margery A., Stephen L. Ross, George Galster, John Yinger. 2002. *Discrimination in Metropolitan Housing Markets: National Results from Phase I of the HDS 2000.* U.S. Department of Housing and Urban Development. Washington D.C.

Turner, Margery A. and Stephen L Ross. 2003. Housing discrimination in American: Findings from the latest national paired testing study

Turner, Margery A., Erin Godfrey, Stephen L. Ross, and Robin Smith. 2003. Other things being equal: A paired testing study of mortgage lending discrimination. Unpublished manuscript.

Turner, Margery A., Michael Fix, and Raymond J. Struyk. 1991. *Opportunities Denied, Opportunities Diminished: Racial Discrimination in Hiring.* Washington, DC: Urban Institute.

Van Order, Robert, and Peter Zorn. 1995. "Testing for Discrimination: Combining Default and Rejection Data." In Anthony M. Yezer, editor, *Fair Lending Analysis: A Compendium of Essays on the Use of Statistics.* Washington, DC: American Bankers Association, pp. 105-112.

Veum, Jonathan R. 1996. "Gender and Race Differences in Company Training." *Industrial Relations* 35(1): 32-44.

Wellington, Alison J. 1993. "Changes in the Male/Female Wage Gap (1976-85)." *Journal of Human Resources* 28(2): 383-411.

White, Katherine. 1997. "Simultaneity Issues in the Relationship of Income and Intelligence." Undergraduate Senior Honors Thesis. Department of Economics, University of North Carolina at Chapel Hill.

White, Paul F., and Michael J. Piette. 1998. "The Use of "Reverse Regression" in Employment Discrimination Analysis." *Journal of Forensic Economics* 11(2): 127-138.

Wienk, R.E., C.E. Reid, J.C. Simonson, and F.J. Eggers. 1979. *Measuring Discrimination in American Housing Markets: The Housing Market Practices Survey.* U.S. Department of Housing and Urban Development, Washington, DC.

Wissoker, Douglas A, Wendy Zimmermann, and George Galster. 1998. *Testing for Discrimination in Home Insurance.* Washington, DC: Urban Institute.

Wood, Robert G., Mary E. Corcoran, and Paul N. Courant. 1993. "Pay Differences among the Highly Paid: The Male-Female Earnings Gap in Lawyers' Salaries." *Journal of Labor Economics* 11(3): 417-441.

Yinger, John. 1998. "Housing Discrimination is Still Worth Worrying About." *Housing Policy Debate* 9(4): 893-927.

Yinger, John. 1997. "Cash in Your Face: The Cost of Racial and Ethnic Discrimination in Housing." *Journal of Urban Economics* 42(3) (November): 339-365.

Yinger, John. 1995. *Closed Doors, Opportunities Lost: The Continuing Costs of Housing Discrimination.* New York: Russell Sage Foundation.

Yinger, John. 1993. "Access Denied, Access Constrained: Results and Implications of the 1989 Housing Discrimination Study." In Michael Fix and Raymond J. Struyk, editors, *Clear and Convincing Evidence.* Washington, DC: Urban Institute Press.

Yinger, John. 1978. "The Black-White Price Differential in Housing: some Future Evidence." *Land Economics* 54: 187-206.