Università di Pisa
Dipartimento di Informatica

Technical Report: TR-11-16

A multidisciplinary survey on discrimination analysis

Andrea Romei Salvatore Ruggieri

November 7, 2011 (updated April 3, 2013)

Address: Largo B. Pontecorvo 3, 56127 Pisa, Italy. Tel: +39 050 2212700 Fax: +39 050 2212726

A multidisciplinary survey on discrimination analysis

Andrea Romei Salvatore Ruggieri

November 7, 2011 (updated April 3, 2013)

The Knowledge Engineering Review, Vol. 00:0, 1–54. © 2013, Cambridge University Press
DOI: 10.1017/S000000000000000 Printed in the United Kingdom

A multidisciplinary survey on discrimination analysis

ANDREA ROMEI and SALVATORE RUGGIERI

Dipartimento di Informatica, Università di Pisa, Largo B. Pontecorvo 3, 56127 Pisa, Italy
E-mail: {romei, ruggieri}@di.unipi.it

Abstract

The collection and analysis of observational and experimental data represent the main tools for assessing the presence, the extent, the nature, and the trend of discrimination phenomena. Data analysis techniques have been proposed in the last fifty years in the economic, legal, statistical, and, recently, in the data mining literature. This is not surprising, since discrimination analysis is a multi-disciplinary problem, involving sociological causes, legal argumentations, economic models, statistical techniques, computational issues. The objective of this survey is to provide a guidance and a glue for researchers and anti-discrimination data analysts on concepts, problems, application areas, datasets, methods and approaches from a multidisciplinary perspective. We organize the approaches according to their method of data collection as observational, quasi-experimental and experimental studies. A fourth line of recently blooming research on knowledge discovery based methods is also covered. Observational methods are further categorized on the basis of their application context: labor economics, social profiling, consumer markets, and others.

1 Introduction

Discrimination refers to an unjustified difference in treatment on the basis of any physical or cultural trait, such as sex, ethnic origin, religion or political opinions. The problems of assessing the presence, extent, nature, and trends of discrimination and of preventing discrimination in (possibly automated) decision making are thus of primary importance. In the last fifty years, such problems have been investigated from a social, legal, economic, and, recently, from a computer science perspective. The issues of data collection and data analysis are persistent, bringing together all these perspectives. We present an annotated multi-disciplinary bibliography, with a *focus on the methods for data collection and for data analysis*. According to the statistical principles of design, we classify approaches as observational, quasi-experimental, and experimental studies, on the basis of whether data is collected by merely observing existing situations, or by controlling the factors that might affect discriminatory decisions. Since observational studies represent the vast majority, we further categorize observational approaches in terms of context, the main ones being labor economics, social profiling, and consumer markets. We consider a fourth class of recent studies, based on knowledge discovery, which poses new challenges on the discovery and prevention of discrimination.

Discrimination data analysis is very topical. The ease of data storage and retention, the ever increasing computing power, and the development of intelligent data analysis and mining techniques make it possible to apply "in-the-large" and to improve classical statistical and econometric techniques. The reference literature, however, is spread over many disciplines: social sciences, psychology, economics, finance, health research, housing and urban development, statistics, biometrics, econometrics, data mining. Our primary objective is to survey from a data collection and data analysis perspective the main problems, techniques, and approaches adopted in the literature, giving priority to recent works that have not yet been surveyed.

This paper is structured as follows. After introducing the relevant concepts from social and legal perspectives in Section 2, we overview the basic statistical tools adopted for discrimination data analysis

in Section 3, including a discussion of technical issues arising in legal contexts. We then survey four mainstream groups of studies: observational (Sections 4-7), quasi-experimental (Section 8), experimental (Section 9), and knowledge discovery (Section 10). Approaches in the first group are further split according to the application context: labor economics (Section 4), social profiling (Section 5), credit and consumer markets (Section 6), and other contexts (Section 7). Sections 4-9 are organized into three parts: first, we present the general approach of the line of research, with particular emphasis on how data is collected; then, we present some technical details of the underlying data analysis techniques; finally, we provide an annotated bibliography of the literature. Sections 7 and 10 include only the annotated bibliography, since they cover miscellaneous contexts and/or techniques.

2 Background

Discrimination refers to *an unjustified distinction of individuals based on their membership, or perceived membership, in a certain group or category*. Justified distinctions are exceptions explicitly admitted by law, such as imposing a minimum age for voting in elections, or that are proven (sometimes in court) as being objective and legitimate, such as requiring a man for a male character in a film. Some groups, traditionally subject to discrimination, are explicitly listed as "protected groups" by national and international human rights laws. These and other faceted components of the broad definition of discrimination are introduced in the next sub-sections.

2.1 Grounds of discrimination, protected groups, multiple discrimination

Human rights laws prohibit discrimination on the grounds of sex, gender, sexual orientation[1], race, ethnicity[2], skin color, social origin, genetic features, language, religion or belief, political or other personal opinion, membership of a national minority, property, birth, parentage, disability, illness, marital status, or age. Legislations and research studies have evolved with a different focus on vulnerable groups, sometimes restricting themselves to specific settings, including credit and insurance, sale, rental, and financing of housing, personnel selection and wages, access to public accommodation, and education. For instance, discrimination against blacks is dealt with to a large extent by studies from the United States (U.S.), whilst discrimination against Roma people has been mainly considered by European Union (E.U.) studies. In this survey, we do not restrict our attention to a specific group. We assume that, in principle, discrimination can take place based on any physical or cultural trait, and in any setting of everyday life. Hereafter *protected group* will be used to denote a group or category of people that are the subject of discrimination analysis, irrespectively of whether there is any national or international law which explicitly protects the group from discrimination in the specific settings covered by such legislation.

Although the aforementioned grounds for discrimination are typically considered separately, the interaction of multiple forms of discrimination has been receiving increasing attention (European Commission, 2007; ENAR, 2007). An elderly disabled woman for example, could be discriminated against for being above a certain age, because she is a woman, because she is disabled, or any combination of these. *Multiple* discrimination comes into play when a person is discriminated against on the basis of different characteristics *at different times*: each type of discrimination works independently, according to distinct experiences, and multiple discrimination refers to their cumulative impact. When different grounds operate *at the same time*, then this is known as *compound* or *intersectional* discrimination. Compound discrimination occurs when each ground *adds* to discrimination on other grounds, for example

[1] *Sex* refers to a person's biological status, categorized as male, female, or intersex; *gender* refers to the attitudes and behaviors that a culture associates with a person's sex, categorized as masculine, feminine and transgender; *sexual orientation* refers to the sex of those to whom one is sexually and romantically attracted, categorized as homosexual, heterosexual, and bisexual. See American Psychological Association (2011) for a psychological discussion of the differences between the terms, and (Case, 1995) for a discussion with reference to the U.S. anti-discrimination law.

[2] *Race* is a social construct to categorize people into groups. The term is controversial, and with little consensus on its actual meaning. (Blank et al., 2004, Chapter 2) summarize biological and social concepts of race, and discuss U.S. categorizations of races used for data collection, e.g., in census data. *Ethnicity* refers to self-identifying groups based on beliefs concerning shared culture, ancestry and history.

migrant women experiencing both under-employment (such as migrants compared to local residents) and lower pay (such as female workers compared to male workers). Intersectional discrimination occurs when concurrent acts of discrimination result in a specific and distinct form of discrimination, for example as reported by Makkonen (2002), as in the case of Afro-American women stereotypes which when taken in isolation cover neither women nor Afro-Americans.

2.2 Direct discrimination, indirect discrimination, objective justification, tokenism

From a legal perspective, discrimination arises only through the application of different rules or practices to comparable situations or of the same rule or practice to different situations (Tobler, 2008). When such rules or practices explicitly treat one person less favorably on forbidden grounds than another is, has been or would be treated in a comparable situation, this is known as *direct discrimination*, sometimes called *systematic* discrimination or *disparate treatment*. On the other hand, *indirect discrimination*, sometimes called *disparate impact*, is an apparently neutral provision, criterion or practice which results in an unfair treatment of a protected group.

Direct discrimination is intentional and "directed" towards individuals, typically on the basis of their visible traits. One example of an act of direct discrimination would be the refusal to serve Roma people in a restaurant . Direct discrimination is difficult to prove, since the complainant has to demonstrate the intent to discriminate. In the previous example, a sign stating "we do not serve Roma" or some testimony would need to be provided by the complainant.

Indirect discrimination occurs via apparently neutral practices that take into account personal attributes correlated with indicators of race, gender, and other discriminatory grounds. Laws on indirect discrimination are concerned with avoiding the circumvention of the prohibition to discriminate, and to enforce such a prohibition substantively, even in the case of unintentionality. A landmark U.S. law case (*Griggs vs Duke Power Co.*, 1971) concerned the requirement of a high-school diploma for an unskilled job for which no literacy was needed, which practically excluded most black applicants. A presumption of indirect discrimination on a group is typically based on observing that the *effects* of some rules or practices have put the group in an adverse position. The burden of proof then shifts onto the defendant, who has to establish the absence of discrimination by providing an objective and reasonable justification.

Objective justifications are accounted for by anti-discrimination laws. In direct discrimination, a *genuine occupational requirement*, also called a *business necessity* in (U.S. Federal Legislation, 2012, (f)), occurs when the ability to perform a job effectively is substantially reduced for people with certain characteristics. One example is advertising for an actor to interpret a male character in a film is a genuine occupational requirement that overcomes direct discrimination against women. In indirect discrimination, an objective justification of the effects of disparate impact is admitted if "objectively justified by a legitimate aim and the means of achieving that aim are appropriate and necessary" Ellis (2005). For example, selecting applicants on the basis of their truck-driving licence, which is a legitimate aim (for a transport company seeking a truck driver), overcomes indirect discrimination against women, who are less likely to fulfill such a requirement than men. The cause of this indirect discrimination is an unavoidable and justified business necessity.

Examples of practices that are not objective justifications are *tokenism* and *reverse tokenism*. Tokenism (see Kanter, 1977) consists of granting a benefit to a few members of a minority group, called *tokens*, to create the false appearance of inclusiveness and prevent accusations of discrimination. For instance, hiring a black person in a mostly white workplace, or a woman in a traditionally male occupation. Reverse tokenism consists of denying a benefit to a few qualified members of the majority group, again called tokens, to create the false appearance of equality and prevent charges of discrimination. This is achieved by exhibiting these cases, the tokens, who are as qualified as members of minority groups to receive a certain benefit but who have been denied the benefit. For instance, by denying a loan to a qualified domestic worker, a bank sacrifices a good candidate for the sake of exhibiting a qualified comparator against immigrant workers with denied loans.

2.3 *Causes of discrimination*

While it is out of the scope of this survey to clearly identify the social and psychological factors that cause discrimination, we discuss three possible causes with reference to the amount and the type of information that social groups know about each other: prejudice (no information on individuals, no information on their group), statistical thinking (no information on individuals, average performances of people in their group), and unintentionality (information on individuals, no information on effects of decisions). The first two causes form the basis of the economic models of discrimination that will be presented later on.

A common source of discrimination is *prejudice*, which literally means to judge another person or group *a priori*. Prejudice leads to discrimination when it concerns unfairly or unreasonably-formed negative attitudes against a protected group. These negative attitudes are produced by social distance, negative stereotypes, and negative feelings, all of which reinforce one another (Brown, 2010). Social distance means scarce familiarity with the members of the protected group, either voluntary or as the effect of segregation (e.g., by residence location or job type), and the strong influence of the norms of one's own group. Negative feelings are affective components of the belief, at the root, for instance, of racism and sexism. Negative stereotypes are popular beliefs that some negative characteristics are common to all or most of the members of a group. The vicious cycle of discrimination (Newman, 2008) starts from a situation where prejudice causes a protected group to be socially disadvantaged. This is interpreted as evidence that the group is inferior, which, in turn, creates renewed prejudice by increasing social distance, by reinforcing negative stereotypes, and by legitimating negative feelings. Psychologists have investigated situations of anxiety or concerns, called *stereotype threats* (Steele and Aronson, 1995), where individuals have the potential to confirm a negative stereotype of their social group, which then result in the reduced performances of individuals. We refer to Whitley and Kite (2009) for a psychological discussion of prejudice[3].

A second source of discrimination is *statistical thinking*, also called *rational racism* (Harford, 2008). This occurs when the lack of knowledge regarding the skills, productivity, or the impact of an individual is compensated for by a prior knowledge of the average performances of the group or category the individual belongs to. This is the case, for instance, when employers refer directly or indirectly to the average performance of the applicant's racial group as a decision element. Another form of statistical thinking occurs when an employer foresees a negative impact on his/her business in hiring an applicant due to the prejudice of the customers against any worker of the applicant's group.

Finally, *unintentional discrimination* occurs not because of malevolent decisions, but due to a lack of awareness on the effects of the decision. This is the case of indifference, incorrect (execution of) procedures or practices, lack of planning and analysis of decision outcomes. A form of unconscious or implicit discrimination has also been considered in the cognitive, economic, and legal literature (Bertrand et al., 2005; Greenwald and Krieger, 2006; Kang and Lane, 2010). Systematic implicit preferences in favor of socially privileged groups have been observed: young over old, white over black, light skinned over dark skinned, other people over Arab-Muslim, abled over disabled, thin over obese, and straight over gay. Indirect and unintentional discrimination pose considerable problems for data analysts to carefully take into account the effects of decisions due to their unforeseen discriminatory effects.

2.4 *Affirmative actions, reverse discrimination, favoritism*

Affirmative actions, sometimes called *positive actions*, are a range of policies to overcome and to compensate by providing opportunities for those traditionally discriminated against (Sowell, 2005; Holzer and Neumark, 2004; ENAR, 2008). They range from the mere encouragement of under-represented groups to preferential treatment or quotas in favor of these groups. U.S. federal contractors for example, are required to identify and set goals for hiring under-utilized minorities and women (Holzer and Neumark, 2006). Universities have voluntarily implemented admission policies that give preferential treatment to women and minority candidates (Lerner and Nagai, 2000). *Reverse discrimination*, sometimes identified with affirmative actions, is the disadvantage that the non-members of protected groups suffer as a result of

[3]See also http://www.understandingprejudice.org for a collection of prejudice-related resources.

affirmative actions. Affirmative action policies "shall in no case entail as a consequence the maintenance of unequal or separate rights for different racial groups after the objectives for which they were taken have been achieved" (United Nations Legislation, 2012, (d)). It is therefore important to assess and to monitor the application of affirmative actions.

Favoritism (sometimes called *nepotism*) refers to occasions when individuals appear to have been treated better than others for reasons not related to individual merit, business necessity or affirmative actions. Kim (2007) for example, has studied political decisions, such as the distribution of income across regions or groups, which can be discretionary and favor the group or district which a politician belongs to. The difference between affirmative action and favoritism lies then in the group which is favored: in affirmative action, the group is historically disadvantaged and the practice is recommended or required by law. In favoritism, the group is favored for reasons, such as family relationship, that are not supported by explicit rules or legislation. Not surprisingly, favoritism can also refer to members of protected groups, such as in the case of university admissions of black students due to their sporting competitiveness. Discrimination and favoritism are dual concepts: if a protected group is discriminated against in a certain context, then the remaining people in the same context are favored. Strictly speaking, however, we reserve the term favoritism for the unfair (positive) treatment of members of a specific group, and not as the implicit consequence of the discrimination of other groups. Yamagishi et al. (1999) have reviewed the social theories of in-group favoritism.

2.5 Formal equality, substantive equality, group under-representation, prima-facie evidence

Provisions on equality or non-discrimination[4] are firmly embedded within the key human rights treaties of the United Nations Legislation (2012). Anti-discrimination laws, however, have evolved differently in common law countries compared to civil law countries. U.S. Federal Legislation (2012); U.K. Legislation (2012) and Australian Legislation (2012) follow the common law characteristic of "the absence of systematization, or a desire thereof" (Schiek et al., 2007, Introductory Chapter), with the result that laws have been developed on a ground-by-ground basis and with reference to specific contexts, possibly with different rulings from one case to another. European Union Legislation (2012) and E.U. member states follow a principled approach, resulting in laws covering a (long) list of grounds for discrimination. For a more thorough legal discussion and comparison of national and international laws, see the following books: on international group rights by Lerner (2003); Schiek et al. (2007); on E.U. laws by Ellis (2005); E.U. Agency for Fundamental Rights (2011); and on U.S. laws by Bamforth et al. (2008).

The legal debate contrasts *formal equality* with *substantive equality* (Barnard and Hepple, 2000). Formal equality, or equality of treatments, is based on the principle that "things that are alike should be treated alike". Hence individuals should be treated on the basis of their own merits, not on the basis of irrelevant physical or personal characteristics. This is the widespread understanding of non-discrimination today. In substantive equality, or equality of outcomes, individuals are treated differently on the basis of their differences, and in particular of the disadvantages associated with protected groups, with the goal of reaching fair outcomes in decisions. Affirmative actions and the combat of indirect discrimination are based on a distributive principle of justice which in itself is aimed at substantive equality.

Several independent authorities (equality enforcement bodies, regulation boards, consumer advisory councils, commissions) provide advice, monitor, and report on discrimination compliances. For instance, the European Commission[5] publishes an annual report on the progress in implementing the Equal Treatment Directives by the member states (Chopin and Do, 2010); and in the U.S., the Attorney General reports to the Congress regarding the annual referrals to the Equal Credit Opportunity Act. In many cases, anti-discrimination authorities have the power to conduct investigations and inquiries, up to quasi-judicial functions. They carry out research, promote actions in favor of substantive equality, and make

[4]While semantically equivalent, the term "non-discrimination law" recalls a set of negative obligations, while "equality law" recalls, in addition, a set of positive obligations to reach the ideal of equal treatment. See Bell (2002).
[5]See also the European Network of Legal Experts (http://www.non-discrimination.net), the E.U. Agency for Fundamental Rights (http://fra.europa.eu), and the Migration Policy Group (http://www.migpolgroup.com).

recommendations to governments. A comparison of methods adopted by law enforcement officials and by academics (mainly, economists) is reported by Ross and Yinger (2006).

The problems of data collection and assessment, and of discrimination measurement are recognized as fundamental issues to be addressed by both E.U. (Makkonen, 2006, 2007) and U.S. (Blank et al., 2004) authorities. The recourse to data analysis and statistical evidence is also a typical scenario in discrimination litigation, with manuals instructing judges to become acquainted with the characteristics and reasonings of science (see e.g., Kaye and Freedman, 2011). One general legal principle is to consider *group under-representation* in obtaining a benefit, or alternatively *group over-representation* in being denied a benefit, as a quantitative measure of (indirect) discrimination against a protected group. It is commonly agreed, however, that statistical conclusions establish *a prima facie*[6] evidence of discrimination, which may be rebutted by the respondent using further arguments, e.g., an objective justification.

3 Statistical tools

3.1 Statistical data analysis

Statistical tools and methods have been widely adopted, both in legal and in economic analyses, in performing data collection and sampling, statistical inference, correlation and regression. The sources of data are categorized on the basis of the level of control the analyst has: in an experiment, the analyst has control over all independent variables; in quasi-experiments, only a few variables are controlled for; and in observational studies, the analyst has no control over the variables. Observational studies are much more frequent than experiments, with data collected from historical records or from surveys. This gives rise to issues regarding the reliability of the data collection, the completeness of the relevant variables, the absence of confounding variables, and the non-existence of plausible alternative explanations (Fienberg et al., 1995).

As observed by Wingate and Thornton (2000), the frequentist approach to statistics has been mainly adopted as *prima facie* evidence in law cases, whilst the Bayesian approach has seldom been applied. A common tool for statistical analysis is provided by a 2×2, or 4-fold, contingency table, as shown in Fig. 1, or, more generally, by a $k \times 2$ table, where k is the number of groups (identified by race, education level, or other characteristics), and 2 is the number of possible outcomes (in the figure, benefit denied or granted). Different outcomes between groups are measured in terms of the proportion of people in each group with a specific outcome. Figure 1 considers the proportions of benefits denied for the protected group (p_1), the unprotected group (p_2) and the overall sample (p). The differences and rates of these proportions can be seen as the formal counterpart of the legal principle of group under-representation. They are known in statistics as the *risk difference* (RD $= p_1 - p_2$), also known as *absolute risk reduction*; *risk ratio* or *relative risk* (RR $= p_1/p_2$); *relative chance* (RC $= (1 - p_1)/(1 - p_2)$), also known as *selection rate*; *odds ratio* (OR = RR/RC $= p_1(1 - p_2)/(p_2(1 - p_1))$). The versions of RD, RR, and RC when the protected group is compared to the average proportion p, rather than to the proportion of the unprotected group are also reported: *extended difference* (ED $= p_1 - p$); *extended ratio* or *extended lift* (ER $= p_1/p$); and *extended chance* (EC $= (1 - p_1)/(1 - p)$). Level curves of RR, OR, and RD over the 2-D risk plane are shown in Figure 2. A level curve represents the points of the 2-D risk plane with a fixed value of the reference measure. When $p_1 \approx 0$, then OR $\approx$ RR. Apart from this well-known relation, the relative order imposed by the various measures over the points of the 2-D risk plane can be dramatically different (Pedreschi et al., 2012). The names of the measures come mainly from bio-statistics and epidemiological comparative studies between two dichotomous groups. RC comes from the literature on employment discrimination. Recent work on data mining for discrimination discovery has introduced the ED, ER, and EC measures (Pedreschi et al., 2009; Ruggieri et al., 2010b). Finally, the name 2-D risk plane is due to Li and Yang (2007).

Statistical tests and confidence intervals for the difference, ratio, and odds of proportions have been proposed over the last 50 years. Let us denote by π_1 and π_2 the true proportions of p_1 and p_2. Difference,

[6] *Prima facie* is a Latin term meaning "at first look," or "on its face," and refers to evidence which, unless rebutted, would be sufficient to prove a particular proposition or fact.

group	benefit denied	benefit granted	
protected	a	b	n_1
unprotected	c	d	n_2
	m_1	m_2	n

$$p_1 = a/n_1 \quad p_2 = c/n_2 \quad p = m_1/n$$

$$RD = p_1 - p_2 \quad RR = \frac{p_1}{p_2} \quad RC = \frac{1-p_1}{1-p_2} \quad OR = \frac{RR}{RC} = \frac{a/b}{c/d} \quad ED = p_1 - p \quad ER = \frac{p_1}{p} \quad EC = \frac{1-p_1}{1-p}$$

Figure 1 4-fold contingency table and discrimination measures

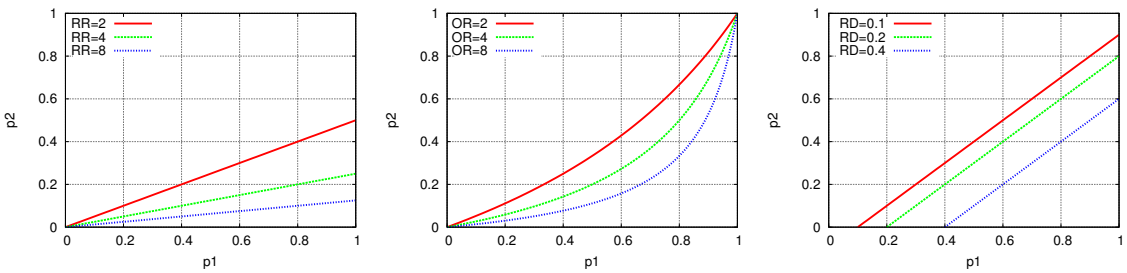

Figure 2 Level curves of RR, OR, and RD over the 2-D risk plane.

ratio and odds of π_1 and π_2 follow discrete distribution probabilities. However, when the numbers in the contingency table are large, the distributions can be asymptotically approximated by a normal or a log-normal distribution. Based on this, Wald confidence intervals can be calculated. As an example, for risk difference, called $\hat{p} = p_1 - p_2$, the confidence interval at the significance level of $100(1-\alpha)\%$ for $\pi_1 - \pi_2$ is $[\hat{p} - d, \hat{p} + d]$ where:

$$d = Z_{1-\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

See Agresti (2002); Farrington and Manning (1990); Fleiss et al. (2003); Leung and Kupper (1981); Newcombe (1998); Tian et al. (2008) for details, other confidence intervals, and for comparisons of estimation methods. In order to improve the approximation of a discrete distribution by the normal or log-normal distribution, several "corrections for continuity" have been proposed, such as Yates's correction and the Mid-p method (Agresti, 2002), or the simple but effective plus-4 method (Agresti and Brian, 2000), which consists of adding $Z_\alpha^2/4$ cases to each cell in the contingency table. When the numbers in a contingency table are very low, the approximation to the normal distribution becomes imprecise. Exact methods have been proposed in the statistical literature, where "exact" means that the actual discrete distribution of the statistical parameter is adopted in computing the confidence intervals. The original work on the subject traces back to Fisher's exact method for a single proportion. The issues here are twofold and contrasting: intervals should ideally have a width that is as strict as possible, yet precise calculations of discrete distributions are computationally expensive.

In order to evaluate whether the difference between groups is due to a third control variable, e.g., high/low education level, a $2 \times 2 \times 2$ table is required, with numbers for each combination of group, education level, and outcome. A Mantel-Haenszel test compares the denial rates for protected and unprotected groups within strata defined by education level. See Agresti (2002); Fleiss et al. (2003) for the Mantel-Haenszel procedure, and Gastwirth (1984); Finkelstein and Levin (2001) for a discussion of its use in legal cases. As observed by Quillian (2006), "the statistical analysis relies upon measuring racial discrimination as a residual: discrimination is the remaining racial [in general, group] difference after statistically accounting for all other race-related [group-related] influences on the outcome", such as the education level in the previous example. However, since it is difficult to know whether all important

influences (called *factors* in statistical terms) have been taken into account[7], "statistical decomposition is better viewed as a method to assess how much of a racial gap can be accounted for by measured factors rather than as a method to measure discrimination per se". The sensitivity of statistical methods regarding the uncertainty due to the omitted variable bias has been studied: for measures over contingency tables (Gastwirth, 1992b); for matched pairs or audits, which will be introduced in Section 8 (Gastwirth et al., 1998); and, for the Mantel-Haenszel test (Yu and Gastwirth, 2008).

For general $k \times 2$ contingency tables, the Pearson χ^2-test is adopted to test the independence of the outcome for the k groups. When the outcome is an ordinal variable, such as a rank, the difference in treatment between a protected and an unprotected group can be evaluated by Wilcoxon's signed-ranks test. For continuous outcomes, such as wages or loan rates, the significance of the average difference of outcome between groups can be evaluated by Student's t-tests. See classical handbooks on non-parametric statistical procedures for details (e.g. Sheskin, 2004; Siegel and Castellan, 1988).

In economic research, the analytical tool adopted shifts from contingency tables to regression analysis. Let Y be a continuous dependent variable, such as salary, and $X_1, \ldots, X_k$ be continuous or binary independent variables which are potentially useful (and legally plausible) in explaining Y, such as the employee's education, skills, experience, productivity, job grade, responsibilities, and working time. In addition, let D be a continuous or binary independent variable concerning the grounds of possible discrimination, such as the employee's age range or female gender indicator. Regression analysis consists of extracting a linear model describing Y in terms of $X_1, \ldots, X_k$ and D from data, in our case from employees' data. Technically, it consists of calculating parameters $\alpha, \beta_0, \ldots, \beta_k$ that best fit a set of equations:

$$Y_i = \alpha D_i + \beta_0 + \beta_1 X_{i,1} + \cdots + \beta_k X_{i,k} + \epsilon_i$$

where $i = 1 \ldots n$ is a set of cases, in our example, employees. The ϵ variable models residuals, i.e., the difference between the actual Y_i and the value resulting from the right-hand side (except ϵ_i). The least square minimization method is a common way of estimating the coefficients $\alpha, \beta_1, \ldots, \beta_k$. This is known as the OLS (ordinary least squares) regression method. Models that overcome the linear assumption by constraining the range of the dependent variable are known as Tobit regressions (Amemiya, 1984). Measures of fit include the mean squared error (MSE) and the squared multiple correlation (R^2). Under the assumption that the residuals are identically and independently distributed normal variables with mean zero, α and the β_i's can be interpreted as estimators of the linear model over the entire population. As a consequence, an $\alpha \neq 0$ at some confidence level, can be interpreted as the weight of the discrimination grounds D in affecting the dependent variable. Critical issues in regression analysis include the choice of the independent variables, which could be legally debated in terms of relevance and completeness; the presence of erroneous, missing, and extreme values in data; and the multi-collinearity of variables, namely the overlap of information, especially among D and other variables, e.g., such as in race and education. See the paper by McCabe (1992) and the book by Finkelstein and Levin (2001) for details. Greene (2008) is a classic reference on econometric analysis.

Linear regression can be adopted in the context of binary decisions by considering as the dependent variable a (continuous) function $f(P(Y = granted))$ of the probability $P(Y = granted)$ that the binary variable Y assumes a specific value, say, benefit granted. When $f()$ is the logit function $logit(P) = log(P/(1-P))$, the value α can then be interpreted as the variation coefficient of the logarithm of the odds of the event $Y = granted$ due to the presence of the discrimination grounds, all other predictors being constant. This is known as the logit regression model. When f is the CDF of the unit normal distribution, we end up with the probit regression model. Classical books on generalized linear models are by McCullagh and Nelder (1989); Dobson and Barnett (2008); Hardin and Hilbe (2007).

Finally, approaches based on causal reasoning (Pearl, 2009a,b) have an immediate application in discrimination analysis. The direct casual effect of sex, race or other discrimination grounds, on the dependent variable is the sensitivity of the dependent variable to changes in the discrimination grounds while all other variables are held fixed. Stated in legal terms, "the central question in any employment-discrimination case is whether the employer would have taken the same action had the employee been of

[7] A recurring problem known as *the omitted-variable bias*.

a different race (age, sex, religion, national origin etc.) and everything else had been the same" (*Carson v. Bethlehem Steel Corp.*, 1996).

3.2 Legally-grounded data analysis

As observed by Gastwirth (1992a), the objectives of science and the law often diverge, with rigorous scientific methods conflicting with the adversarial nature of the legal system. Statistical evidence in discrimination cases, mainly in employment, is considered as particularly persuasive. The book by Finkelstein and Levin (2001) reports several law cases to illustrate the statistical methods used to interpret disparity: in pass/fail proficiency tests for Afro-Americans and Hispanic groups; in jury member selection for Mexican-American groups; in hiring black teachers or black lawyers; in promoting black employees of a bank/police department; in granting wage additives to racial groups of employees; in the employment termination of elderly employees; in the time taken to promote female employees; in the salaries between male/female or black/white employees; and in mortgage lending between blacks and whites. The legal principle of *group under-representation* readily fits in the statistical analysis of contingency tables as in Figure 1. The main issues in legal trials consist of determining the relevant population, also known as the *pool of comparators*, the discrimination measure that formalizes group under-representation, and the threshold that constitutes *prima facie* evidence of discrimination.

Relevant population. This means deciding where n is drawn from in Figure 1. There are several possibilities, and the actual choice depends on the specific legal case: the total population of a state, such as in a jury selection discrimination case; the set of eligible people for a position, such as the labor force in an employment discrimination case; and the set of applicants for a benefit, such as in a loan/mortgage discrimination case. The choice of the relevant population is decisive in an indirect discrimination claim, since it influences the measurement of the group's under-representation. Shoben (1992) discusses the various choices in the context of employment discrimination. Interestingly, a well-known defence strategy in discrimination allegations is related to Simpson's paradox of statistical analysis. Bickel et al. (1975) describe a real case of possible discrimination against women in university admissions. The numbers at university level showed a disproportion between male and female candidates. However, when the numbers were split at a departmental level, the disproportion was no longer observed. Since the examination boards worked at a departmental level, then the correct numbers to be considered are the latter, whilst the numbers at a university level contain confounding factors (the boards that took decisions).

Discrimination measures. Various measures are adopted worldwide. U.K. Legislation (2012)(a) mentions risk difference, the E.U. Court of Justice has focused more on the risk ratio (Schiek et al., 2007, Section 3.5), and U.S. laws and courts mainly refer to the selection rate. Notice that the risk ratio is the ratio of the proportions of *benefit denial* between the protected and unprotected groups, while the selection rate is the ratio of the proportions of *benefit granting*. The odds ratio has been widely considered in legal research studies. This is only an approximation of a multi-faceted reality, where the choice of the reference measure and of the methods to evaluate its statistical significance might be part of the debate in a trial (Piette and White, 1999).

Thresholds of group under-representation. While the European Court of Justice has emphasized that a substantial figure of group under-representation needs to be achieved when considering statistical evidence, no strict threshold has been laid down the E.U. On the contrary, two landmark cases led to formal thresholds in the U.S. laws:

- the *four-fifths rule*, from the Equal Employment Opportunity Commission (1978), states that a job selection rate for the protected group of less than 4/5 (i.e., 80%) of the selection rate for the unprotected group represents a *prima facie* evidence of an adverse impact (i.e., indirect discrimination) against the protected group. In terms of the measures in Figure 1, this means verifying that RC < 0.8 at some confidence level (typically, 95%).
- the *Castaneda rule* (named after the *Castadena vs. Partida* law case, 1977, concerning jury selection) states that the number of people of the protected group selected from a relevant population cannot be smaller than 3 standard deviations the number expected in a random selection (Sugrue and Fairley,

1983). In terms of the contingency table in Figure 1, assuming a random binomial distribution with m_2 independent experiments each with the probability of selecting a member of the protected group of $p_0 = n_1/n$ (the fraction of the group over the total population), the actual number b of selected people departs from the expected number $m_2 p_0$ of $(m_2 p_0 - b)/\sqrt{m_2 p_0 (1 - p_0)}$ standard deviations.

A review of statistical evidence as a proof in employment discrimination litigations is reported by Wingate and Thornton (2000). The paper by Kaye and Aickin (1992) contains a collection of works on the subject. A continuously updated book on statistical methods and case laws in the U.S. is due to Paetzold et al. (1994). Discrimination case laws in the E.U. are collected by the European Commission (2010).

4 Observational studies I: labor economic perspective

From now onwards, we detail the mainstream lines of research on discrimination data analysis by distinguishing between studies on the basis of their main approach: observational, quasi-experimental, and experimental. Observational studies collect data by merely observing existing situations, for example from census and survey data. Since the majority of approaches are observational, we further categorize them on the basis of their specific context. We start by considering labor economic models of discrimination and their empirical validation.

4.1 Approach and data collection

In the labor market, different treatments between groups of workers can be measured in terms of their wages (*wage differentials*), of their degree of participation in the labor force (*employment differentials*), or of their degree of segregation in specific occupations or industries (*segregation differentials*). Public surveys routinely collect data on demographic characteristics and attitudes of residents (e.g., in the U.S., the General Social Survey), on the distribution of labor forces in the labor market (e.g., Current Population Survey), on the labor status of a population sample at multiple points in time (e.g., National Longitudinal Surveys), and so on. Race and gender differentials have been persistent over time. In the U.S. for example, the 2010 unemployment rate of black men aged 25-54 was 16.3%, compared to 8.5% of white men [8]. Empirical research techniques have applied statistical inference to collected data both with the purpose of *testing the consequences predicted by a theoretical economic model*, and *to assess the contribution of different types of discrimination* to the different treatments in the labor market overall. Two major theoretical models of economic discrimination have been considered in the economic literature, namely *taste-based* and *statistical* discrimination. See Altonji and Blank (1999) for a comprehensive mathematical introduction to both types.

Taste-based discrimination, originally proposed by the Nobel prize winner Becker (1971), has no rational or economic basis, but only a personal prejudice against protected groups. The marginal cost of a protected worker, say a female, is her wage plus an additional psychological *disutility* factor, expressed as a *discrimination coefficient*. On the other hand, the marginal cost of a male worker with equal productivity consists of his wage alone. The gap between male and female workers increases according to the employer's prejudices, until the labor market tends to a *wages equilibrium* and to segregation of protected workers in less discriminating firms or for specific occupations. Lower profits for discriminatory firms implies that discrimination occurs mainly in low competitive markets. This is known as the *static implication* of Becker's model. On the basis of the identity of the discriminator, the model distinguishes between *employer discrimination* (taste in hiring), *customer discrimination* (taste in buying), and *co-employee discrimination* (taste in co-operating). Non-discriminating employers earn higher profits by hiring members of the protected group, and, in the long run and in a competitive market, discriminatory firms will be driven out of the market – and, a fortiori, discrimination will disappear. This is known as the *dynamic implication* of the Becker model. Both implications represent consequences predicted by the theoretical model. Empirical studies are supposed to confirm or to refute such implications. Extensions of taste-based discrimination, called *search models* (Altonji and Blank, 1999; Lang and Lehmann, 2011),

[8] Aggregated data taken from the U.S. Current Population Survey, available at http://www.bls.gov/cps.

take into account the costs for workers of searching for jobs by interacting with prejudiced and non-prejudiced firms, and, for consumers, the costs of searching for sellers that belongs to their racial group. Finally, Goldberg (1982) models favoritism by assuming that employers receive *positive* utility from employing members of the unprotected group rather than the cost of employing members of the protected group.

Statistical discrimination originated from Arrow (1971) and Phelps (1972), and was systematized by Aigner and Cain (1977); Coate and Loury (1993); Moro and Norman (2004). It assumes statistical thinking by employers, but a scarcity of information on worker productivity at the time of hiring. This market imperfection gives employers an incentive to easily use observable characteristics, such as sex and race, as a proxy for the expected productivity of workers. Employers act rationally to maximize profits by setting wages on the basis of their prior knowledge on the average productivity of the group the worker belongs to. As a result, workers of the protected group who are as equally productive as workers of the unprotected group experience wage differentials. Unlike taste-based discrimination, an increase in information on the productivity of a worker results in lower wage differentials. Some extensions of the statistical discrimination model deal with what happens as the employer's information on workers' productivity changes, e.g., over the course of the time or because of a more comprehensive selection procedure. These dynamic extensions, in contrast to a static model, are known as *employer learning* models. In addition, the differential observability or learnability of a worker's productivity between groups has been taken into account by *screening discrimination* models. Another strand of statistical discrimination models studies how negative rational stereotypes of employers influence hirings and wages on the one hand, and workers' investments, e.g., in education, on the other. Lang and Lehmann (2011) call this class *rational stereotyping* models.

A specific labor context with an unusually good opportunity to study discrimination, is the one of *professional sports*, such as baseball, basketball, football, and soccer. The problem of estimating the productivity of workers, which is the main challenge in both theories of discrimination, is substantially solved here by extensive, publicly available (from online sport almanacs), measures of the performances of players and coaches. For baseball players of the major league for example, the number of hits, doubles, triples, home runs, stolen bases, and other similar metrics are routinely collected. In addition, race, nationality, and salaries of players are also easily obtainable. Research has covered discrimination in hiring, in retaining (over seasons), in segregating (to specific game roles), and in the salaries of players, as well as customer discrimination. Customer discrimination in the context of professional sports is also known as *fan discrimination*. It consists in determining whether the TV audience, the game attendance, or the price of sport cards depend significantly on some grounds of discrimination, such as the race of players or the racial composition of teams. Specific problems in such approaches include determining confounding factors such as the scarcity of a player card, the absence of personal contact between the fans and the players, and the problem of the "common player" price, which is the minimum value a card can take, irrespective of the performance of the player.

Finally, note that the economic models of discrimination assume a unidirectional causal effect of worker productivity on wages. They do not question whether a lower average productivity of a minority group is itself a consequence of previous forms of discrimination, such as a lower education level due to the lower income of minority group families. In other words, they do not take into account the vicious cycle of discrimination described in Section 2.3.

4.2 Data analysis methods

In the field of labor markets, the basic question of data analysis is to test whether wages, once controlled for the appropriate variables of workers' productivity, are affected by sex, race, or other grounds of discrimination. An intuitive inferential approach (see e.g., Cain (1987) and (Blank et al., 2004, Chapter 7)) is to derive a regression model of the wage of employees that separates the discrimination component of race, sex, or other characteristics of the protected group from the productivity components of the worker. Since wage typically has a skewed distribution, the dependent variable is often set as its logarithm. In the

following linear model for example:

$$log(W) = \alpha R + \beta_1 P_1 + \cdots + \beta_n P_n + \epsilon \tag{1}$$

W is the wage or salary, R is an indicator variable of race, sex, or nationality (in general, an indicator of being in the protected group), ϵ is the residual term, and the P_i's are personal characteristics and performance measures of workers. A t-test of the statistical significance of the regression coefficients, and in particular of α, is adopted to test whether membership of the protected group is a significant factor. Standard OLS regression is replaced, in some studies, by a quantile regression estimation [9]. This is the case, for instance, of professional sports, where the log salary distribution tends to have greater kurtosis values than standard occupations. Finally, in order to measure the effects of equality-enforcing laws, a *difference-in-differences* approach, introduced by Ashenfelter and Card (1985), is typically taken. The approach compares the regression coefficients of the wages of the protected and unprotected groups before and after some event, e.g., a new law. See Bertrand et al. (2004) for a critical presentation.

In the labor market literature, wage differentials, rather than absolute wage values, have mainly been considered in the analyses. The classic inferential approach is the Blinder-Oaxaca decomposition (Blinder, 1973; Oaxaca, 1973) of the mean difference in log wages of different groups based on regression models in a counterfactual manner. Oaxaca starts defining a Becker discrimination index as follows:

$$D = \frac{(W_u/W_p) - (M_u/M_p)}{(M_u/M_p)}$$

where W_u/W_p is the ratio of the *observed* unprotected-protected group average wage, and M_u/M_p is the ratio of their marginal products[10], which by assumption is the average wage ratio in the absence of discrimination. In terms of logarithms, the above formula becomes:

$$log\, W_u - log\, W_p = log\, M_u - log\, M_p + log\, (D+1) \tag{2}$$

The difference between the marginal products $log\, M_u - log\, M_p$ is the component of the wage differentials that can be explained by the different levels of productivity of the unprotected and protected groups, whilst $log\, (D+1)$ is the component due to discrimination. Using a linear regression model of wages for a group $g \in \{u, p\}$ (u denotes unprotected groups, and p protected groups):

$$log\, W_g = \alpha_0^g + \sum_i \alpha_i^g X_i^g + \epsilon_g$$

where X_i's are productivity-determining characteristics, and the expected value of ϵ_g is 0, then (2) can be re-written as:

$$log\, W_u - log\, W_p = \sum_i \alpha_i^p (X_i^u - X_i^p) + \sum_i X_i^u (\alpha_i^u - \alpha_i^p) + (\alpha_0^u - \alpha_0^p)$$

The first summation term is an estimate of the differences in productivity, namely $log\, M_u - log\, M_p$, while the second summation term is an estimate of the discrimination component $log\, (D+1)$. Hence, discrimination exists if $\alpha_i^u - \alpha_i^p \neq 0$ for some i, namely if the regression coefficients of the protected and unprotected groups differ significantly, in a statistical sense (e.g., using the Chow's test, see Greene (2008)). See Cotton (1988); Neumark (1988); Oaxaca and Ransom (1994) for refinements of the Blinder-Oaxaca decomposition, Jann (2008) for details of its implementation in a statistical package, Bauer and Sinning (2008); Yun (2004) for extensions to non-linear models, and Bauer and Sinning (2010); Fairlie (2005) for extensions to logit, probit and tobit models.

The regression model approach is subject to three main potential biases. First, the *omitted variable bias* arises when relevant control variables correlated with the group membership are not taken into consideration in the regression model, e.g., because there is no data available. Second, the *included*

[9] See e.g., (Koenker, 2005).

[10] The marginal product is the extra output produced by using one more worker.

variable bias (Killingsworth, 1993) is when some control variable is itself affected by so called, *pre-market* discrimination. As an example, an indicator variable of a previously granted premium to a worker could be biased towards discrimination if the premium is regularly granted only to workers of the unprotected group. This is an instance of the vicious cycle of discrimination described in Section 2.3. Third, the *diverting variables bias* (Yinger, 1998) occurs when the regression model includes variables that are not legitimate controls but that are correlated with the protected group membership. These variables may pick up some of the effect of the group membership variable, leading to the conclusion that there is no discrimination when in fact there is. One example of this is the Armed Forces Qualifying Test (AFQT) score, available from the U.S. National Longitudinal Survey of Youth, which is a test of cognitive skills, taken by male adolescents. When used as a control variable, it greatly reduces the racial wage gap (Neal and Johnson, 1996). However, there is little consensus on what AFQT measures, with possibilities ranging from intelligence to quality of schooling, and thus as to whether it is a legitimate control.

The alternative decomposition of Juhn et al. (1991) (JMP decomposition) has gained greater popularity, since it provides a way to show the effects of omitted variables. Yun (2009) compares it with the Blinder-Oaxaca decomposition. The survey by Altonji and Blank (1999) reports on the use of Blinder-Oaxaca, JMP, and CLem (by Card and Lemieux (1994, 1996)) decompositions to wage changes between groups over time. A survey of decomposition methods in economics can be found in Fortin et al. (2010).

The approaches discussed so far are based on the concept of *within-qualifications discrimination*, i.e., wage discrimination exists if men and women (or, in general, unprotected and protected persons) with similar qualifications receive different wages. A challenging approach is based on the concept of *within-wages discrimination*, i.e., wage discrimination exists if men and women with similar wages have different qualifications. The estimation method involves a reverse regression of qualifications as a function of wages. We refer to Kapsalis (1982, 2010) for details on the approach, and Schaefer and Visser (2004); Racine and Rilstone (1995); Whiteside and Narayanan (1989) for criticisms and a comparison with the direct regression method.

Let us now discuss in greater detail a study that tests the presence of taste-based discrimination, while a paper regarding statistical discrimination will be discussed in Section 7.2. Charles and Guryan (2008) collected data on racial feeling from the General Social Survey (GSS). This includes 26 ordinal scale questions such as "in general, how warm or cool do you feel towards blacks" and "do you think white students and black students should go to the same schools or separate schools?". Data from the GSS covered a period of 30 years and 9 geographic areas, but the set of questions was not constant over the years. The disparate prejudice responses are combined into a prejudice index as follows. First, each response is normalized to its standard score[11]. Let $d^k_{i,t}$ denote the normalized response of individual i in year t to the question k. The prejudice index for i in year t is defined by aggregating over k:

$$D_{i,t} = \sum_k \frac{d^k_{i,t}}{K_t} \tag{3}$$

where K_t is the number of questions asked in year t. Starting from $D_{i,t}$, aggregated measures of prejudice are defined for geographic areas at year t. Charles and Guryan collected also data on wages from the Current Population Survey (CPS), which is a monthly report in the U.S. with hourly and weekly wages, employment and unemployment rates, and other labor market data split by residence, education, experience, race and other characteristics of the labor force population. The residual black-white wage gap is estimated, for each state, by means of an OLS regression:

$$log(w_i) = \alpha_X X_i + \alpha_R R_i + \alpha_T T_i + \alpha_S S_i + \epsilon_i$$

where w_i is the wage of individual i, X_i is a vector of an individual's characteristics (education, experience, etc.), R is a black-race indicator, and T_i (resp., S_i) is a vector of race-specific year (resp., state)

[11] The standard score of a value x for a response is $(x - \mu)/\sigma$, where μ is the mean and σ is the standard deviation of the response over the sample. To account for questions not appearing in all years, authors deviate slightly by taking μ as the mean for the year 1977, which contains almost all questions, and σ is the standard deviation for the first year the question appeared in the GSS data.

variables. The α_R parameter then measures the contribution of being black to the (log of) a worker's wage. To measure how much of this contribution is taste-based, i.e., is due to racial animosity, Charles and Guryan performed a second level regression. The dependent variable is now α_R, and the dependent variables are various measures of prejudice, obtained from (3) by aggregating at the level of states. This second regression explains the black-white wage gap in a state in terms of the prejudice indexes of that particular state. The overall results suggest that racial prejudice accounts for as much as one-fourth of the gap in wages between blacks and whites. As predicted by Becker's model, regions that are more prejudiced exhibit a larger racial wage gap.

4.3 Annotated bibliography

Here, and in later subsections, we provide an annotated bibliography of survey papers and recent advancements on the topic of the section in question.

Surveys and comparisons. The landmark survey by Altonji and Blank (1999) on race and gender discrimination in labor markets covers both the theoretical foundations of economic theories of discrimination and empirical approaches to show direct evidence. In addition, it accounts for pre-market human capital differences, such as education and family background; experience, seniority, training and mobility; the gender wage gap; trends in race and gender differentials; as well as the impact of anti-discrimination policies. Earlier surveys on labor market discrimination go back to Marshall (1974) and to Cain (1987), with the latter also covering empirical studies. A recent and comprehensive review of theories and empirics of racial discrimination in the labor market can be found in Lang and Lehmann (2011). A survey by Fang and Moro (2010) reviews theoretical models of statistical discrimination and affirmative actions. The interdisciplinary economic-legal paper by Donohue (2007) provides an overview of the connections between economic models and empirical findings on the one hand, and the U.S. anti-discrimination laws on the other.

Charles and Guryan (2011) discuss several challenges that "frustrate empirical research" on racial labor market discrimination, including the difficulty of comparing individuals of different races who may systematically differ in terms of some unobserved or unobservable determinants, and the problems with the indirect testing of economic models (i.e., with testing their predictions), rather than directly testing the mechanism assumed by the model (i.e., prejudice for taste-based discrimination, and information limitation for statistical discrimination). Jones (2008) reviews the literature on employment and earnings discrimination for workers with disabilities. The impact of specific anti-discrimination policies is also evaluated. Kunze (2008) reviews data analysis methods from the literature on empirical evidence of gender wage gaps, including OLS regression, Blinder-Oaxaca and JMP decompositions, quantile regression. Weichselbaumer and Winter-Ebmer (2005) conducted a meta regression analysis of the works on gender wage differentials, where each point of data is not an individual worker but a research study. The regression analysis is intended to explain the wage discrimination reported in a study as a function of methodological features adopted by the study. The data was obtained by querying the Economic Literature Index[12] for any reference to "(wage* or salar* or earning*) and (discrimination or differen*) and (sex or gender)", and then manually screening 263 empirical papers. Their results highlight how the selection of the dataset under analysis is critical. For instance, lower wage gaps are found for samples of only new entrants to the labor market, for samples in the public sector, and for samples of high-prestige jobs.

Finally, extensive surveys on racial and ethnic discrimination analysis in professional sports can be found in Kahn (1991b, 2000, 2009). In addition to highlighting the uniqueness of the field and the opportunities for labor market research, Kahn's surveys cover discrimination in hiring, retention, segregation, salary, and consumer discrimination. Frick (2007) reviews papers on remuneration, transfer fees, career duration, contract length, and discrimination in the major European soccer leagues.

[12]http://www.aeaweb.org/econlit

Static and dynamic implications of taste-based discrimination. The direct testing of the implications of the taste-based discrimination model is difficult due to the lack of data on employers, co-employees, or customer preferences. An influential paper by Hellerstein et al. (2002) tests the static implication by relating, through an OLS regression model, firm profitability to the proportion of female workers (here used as a proxy of employer preferences) both in low competition and high competition markets. They also test the dynamic implication by relating firm growth to the proportion of female workers. A recent application of the same approach, called the *market test* methodology, is presented by Sano (2009), where OSL regression is compared with a mixed effect model. Frijters et al. (2003) sample more than 24,000 observations from the U.K. Workplace Employee Relations Survey, which consists of: face-to-face interviews with top management; self-completion questionnaires by employees; and worker-representative interviews. Using job satisfaction as the dependent variable, an ordered probit model is used to evaluate the influence of ethnic minority density in the workplace. The Chinese job advertising market offers a unique opportunity to collect explicit employer preferences on the gender, age, height and beauty of applicants. In contrast to western countries, in fact, Chinese law allows such preferences to be advertised. Kuhn and Shen (2010) analyse job ads through an employer search model that uses an OLS regression of employer preferences. They found that advertising on discriminatory grounds decreases when skill levels rise.

The dynamic implication of Becker's model in the labor market has been investigated mainly indirectly by analyzing the relationship between the level of competition in the product market and employer discrimination. If wage and employment differentials decrease due to an increase in competition, this is an indirect proof of the existence of discrimination. Data from the banking industry have long been exploited. Black and Strahan (2001) and Levine et al. (2008) consider the effects of banking deregulation on (reducing) gender and racial wage gaps, respectively, from data from the Integrated Public Use Microdata Series[13] (IPUMS), and from the U.S. CPS survey. The former study uses a fixed-effect method, the latter OSL and quantile regressions. The effects of globalization on the gender wage gap have been discussed by Oostendorp (2009) and Neumayer and de Soysa (2007). Oostendorp uses OLS regression on data from the International Labor Organization October Inquiry[14] (ILO), a survey of wages around the world, which started in 1924 and now covers 159 occupations in 49 industrial groups from more than 200 countries worldwide. Neumayer and de Soysa adopt a negative binomial regression on data from the Cingranelli-Richards Human Rights Dataset[15], containing standards-based quantitative information on the government's respect of 15 internationally recognized human rights for 195 countries, collected annually since 1981. Along the same lines, Black and Brainerd (2004) use CPS data to study the effect of globalization on gender discrimination in the manufacturing industry. Weichselbaumer and Winter-Ebmer (2007) consider the effects of the adoption of equality laws on wage differentials. Finally, a study on the decrease of gender discriminating employers over time is conducted by Flabbi (2010). He uses a search model estimated on data from the CPS survey over the last 20 years. Not all the studies mentioned above confirm the prediction of Becker model's dynamic implication. For instance, Sano (2009) shows that, in a competitive market, companies with a larger proportion of female employees do not necessarily perform better. He concludes that the Japanese market is not competitive enough to eliminate taste-based discrimination. Oostendorp (2009), in contrast with Neumayer and de Soysa (2007) and Black and Brainerd (2004), concludes that globalization may even increase the gender gap.

Tsao and Pearlman (2010) study firm and consumer discrimination against young black physicians using data from a survey of more than 6,000 physicians. Their difference-in-differences approach shows no strong evidence of a gender wage gap among physicians. Zhang and Dong (2008) compare gender wages with gender productivity both in the public and private sectors of industry in China. They use data at a firm level, so that the wage regression is based on the average wages in a firm. Results are contrasting: in the public sector women receive a wage premium since their productivity is lower; in the

[13]http://www.ipums.org
[14]http://laborsta.ilo.org
[15]http://ciri.binghamton.edu

private sector, women are more productive than males, but suffer from lower wages.

Statistical discrimination and employer learning. Statistical discrimination is dependent on the information available to an employer. A standard approach is thus to observe a shift in discriminatory decisions as soon as a more complete information set is available. The seminal paper by Altonji and Pierret (2001) firstly proposed an employer learning model. They experimented with the approach on the National Longitudinal Survey of Youth (NLSY) dataset, using the AFQT test as a proxy for worker's productivity. More recent works along this line include: Arcidiacono et al. (2010); Lang and Manove (2011), who correlate, through a median regression, the level of education across races with a worker's ability, measured by the AFQT test; Fadlon (2011), using OLS regression in extending the theoretical model to account for employer-employee racial matches; Cheung (2010), analysing data from the Australian NLSY partitioned by the hiring channel[16]. Finally, Wang (2010) considers the height of immigrant workers as grounds for discrimination, and investigates whether it is used as a signal of productivity by employers.

The speed of employers in learning about the productivity of workers was originally modeled by Lange (2007) and estimated using the NLSY dataset. Hu and Taber (2010) study asymmetric learning models, where employers learn about the productivity of minority workers slower than the productivity of majority workers. They use data from the biennial Displaced Workers Surveys to relate wage losses when moving to a new job due to plant closings or to layoffs. Using again data from the NLSY, Pinkston (2006) confirmed the hypothesis of screening discrimination that employers are less able to evaluate the productivity of workers from one group to another. Variables that are difficult to observe, such as the AFQT score, have a lower influence in an OLS regression for blacks compared to white workers. Holzer et al. (2006) aim at studying the effects of employer-initiated criminal background checks on the hiring of African-Americans. They found that employers who use criminal background checks are more likely to hire black applicants than employers who do not. This confirms that less information (no criminal background check) leads to resorting to race as a proxy of productivity. The impact of race-specific speech patterns, collected from recorded interviews, on the employer's understanding of speaker characteristics has also been evaluated by Grogger (2011) through OSL and median regressions. He found that speakers identified as black earned about 17% less than whites, after controlling for personal and ability characteristics. A worker search model on survey data is adopted by Fryer et al. (2011) to test racial discrimination on search strategies and wage offers collected from unemployed workers. They basically confirm the results reported by Lang and Manove and Arcidiacono et al.. The impact of adopting tests in job hiring, as opposed to informal interviews, is investigated by Autor and Scarborough (2008), and evaluated on a dataset of tests submitted to retail stores. Job tests have the equity-efficiency trade-off objective of raising the productivity of hired workers, whilst not reducing opportunities for protected groups.

Salary discrimination in professional sports. There is an extensive literature on salary discrimination in the context of professional sports. We cover a few recent papers adopting non-standard tools, for older references see Kahn (1991b, 2000, 2009). Palmer and King (2006) compute separate OLS regression models for baseball players with low, medium, and high salaries, in order to reveal discrimination at different level of competition. They found evidence of discrimination in the lower salary groups. Berri and Simmons (2009) argue that quantile regression deals better with the skewness and kurtosis of salary distribution. They found that black football players suffer from an entry barrier. Holmes (2011) contrasts these two types of regression on basketball players' data, and Frick and Deutscher (2009) compare OLS with random effects and median regression models on German professional football data. Cornaglia and Feldman (2011) contrast OLS with fixed effects regression models[17] in analysing the longstanding empirical observation that married baseball players earn significantly more (10% to 40%) than their

[16]Bonoli and Hinrichs (2010) argue that the hiring channel, i.e., the way job applicants contact employers (e.g., through a friend or a relative), which is mostly considered an indicator of motivation, actually plays a stronger role than age or gender as a signal of productivity.
[17]See e.g., Christensen (2002).

single counterparts (everything else being equal). This is known as the "marriage premium", and it may be the effect of employer discrimination: married men are seen as more "stable". The marriage premium is of more general interest, since Neumark (1988) showed that it accounts for about one-third of estimated gender-based wage differentials in the U.S. A fixed effects model was also employed by Yang and Lin (2012) to test wage discrimination against foreign players in the NBA. Goddard and Wilson (2009) adopt a multinomial probit model[18] to characterize transition probabilities of English football players, and use a GHK[19] simulator to estimate regression coefficients. Black players were found to have higher retention probabilities, suggesting a form of discrimination in carrier progression.

Consumer/fan discrimination. Unlike employer and co-employee discrimination, consumer discrimination cannot be eliminated by market competition, as discussed by Kahn (1991a). Evidence of a correlation between the predominant race of customers and the race of the marginal hired worker has been shown by Holzer and Ihlanfeldt (1998). The following include some recent empirical papers. Combes et al. (2011) analyze census data of Afro-immigrant employers in France to estimate the reduction in employment rates in jobs where workers are in contact with customers. The two-step regression model shows that African workers have a low possibility of having a contact job in the areas where the proportion of non-African residents is high. Coyne et al. (2010) test an economic model of the role of entrepreneurs in overcoming consumer discrimination using data on black players in the Major League Baseball. The number of black players employed by a team in a year is set as the dependent variable in a negative binomial regression. Results show that black players have a negative effect on attendance, but a positive effect on the performance of the team. Leonard et al. (2010) use data from 800 retail stores and 70,000 employees matched to census data on the demographics of each store's community, measuring whether lower sales occur in stores whose workforce less closely resembles the potential customers. They found no strong evidence of such a correlation. Ayres et al. (2005) collected survey data from 1066 taxicab drivers on tips received by their customers. A probit regression model shows lower tips for minority drivers and a higher stiffing (i.e., zero tip) rate. In a more recent work, Parrett (2011) uses OLS regression on survey data to compare tips received by male and female servers in restaurants. Discrimination emerges only from customers who rarely frequent the restaurant.

There is an extensive bibliography covering fan discrimination. Let us first consider the approach studying sports card data. A widely cited paper by Nardinelli and Simon (1990) tests whether race directly affects the value of a player in the market for baseball cards. They solve the common player problem by a regression model of $log\ P_p + log\ P_c$, where P_c is the price of the common player and P_p is the component of the card's price related to player performance and race. Hispanics playing as hitters and blacks playing as pitchers have significantly lower card prices, all the other factors being equal. McGarrity et al. (1999) criticize the data collection methodology and the lack of accounting for heteroscedasticity in the regression model. They propose a refined approach based on a tobit regression from a more refined dataset in which card supply is constant. Contrarily to Nardinelli and Simon, they find only little evidence of racial discrimination. A more recent analysis of baseball sports cards data, including a summary of related work, can be found in Primm et al. (2011). To reduce the effect of common cards, they include only the rookie cards of players who had received at least one vote in the baseball's Hall of Fame (HOF) by the Baseball Writers Association of America. They found that for non-white players, only HOF status influences the price of the card. A recent paper on NBA sports cards can be found in Broyles and Keen (2010), who also overview the studies on NBA data. They found no evidence of racial discrimination.

The following are some of the approaches using TV audiences or game attendance and not covered in Kahn's surveys. Aldrich et al. (2005) adopt an OLS regression of the probability of watching a football game based on the race of the players. The paper contrasts TV audience data with GSS survey data on racial attitudes. They found strong evidence of taste for diversity in viewers of all races. Foley and Smith (2007) and Hersch (2010) use data from baseball game attendance. Only the former paper reports evidence of the racial discrimination of fans for the data from the early 1990s. Finally, Wilson and Ying

[18] See e.g., Greene (2008).
[19] Acronym of Geweke-Hajivassilliou-Keane, see Keane (1994).

(2003) test consumer and co-worker nationalistic preferences by measuring the effect of the composition of a team's nationality on fan attendance in the EU's five largest football leagues. OSL regression shows that non-domestic players have a positive influence on club performance and fan attendance. Thus, their under-representation in teams originates from club management.

Coleman et al. (2008) test fan discrimination against Afro-American players using data on votes for the Most Valuable Player competition in the NBA. Hanssen and Andersen (1999) and Jewell et al. (2002) test the same implication by using data from fans voting for baseball's annual All-Star Game and the Hall of Fame in Major League Baseball, respectively. Their results support the view that discriminatory attitudes in the voting procedure have diminished over time, and may even have reversed.

5 Observational studies II: racial profiling

This section and the next cover discrimination as a profiling problem, both in the sense of screening individuals for surveillance or scrutiny (racial profiling) and in the sense of scoring the creditworthiness of loan applicants (credit scoring).

5.1 Approach and data collection

Profiles consist of patterns, rules, or any other form of knowledge used to screen people when searching for those with a certain behavior. They occur in many contexts, from criminal investigations to marketing, from genetic screening to website personalization, from fraud prevention to location-based services. Profiling is the process of extracting profiles, either by manually eliciting them from domain experts or by automatically inferring them from historical data using increasingly sophisticated machine learning and data mining techniques. Profiling also concerns the application of profiles to screen individuals, for example when identifying security risks. The cross-disciplinary collection edited by Hildebrandt and Gutwirth (2008) includes contributions on profiling groups vs profiling individuals, on the technical and social process of profiling, on E.U. legislation, on (data mining) algorithms for profiling, on the relations between profiling and personalization and privacy, and on profiling in specific contexts (e.g. location-based services, web navigation, employment).

Profiling becomes illegal as soon as it results in direct or indirect discrimination against protected groups. In this section, we concentrate on *racial profiling*, defined as "the practice of subjecting citizens to increased surveillance or scrutiny based on racial or ethnic factors rather than reasonable suspicion" (Chan, 2011). Among the possible contexts of racial profiling (e.g. by police, airport security, or income tax fraud investigators), vehicle stops have attracted the vast majority of research. Numerous data collection efforts have been initiated by law enforcement agencies, often as a result of litigation or of legislation, for the purpose of understanding the vehicle stop practices of its officers. Attributes collected concern the stop (time, date, location, reason, duration), driver (race, gender, age), vehicle (make, model), officer (age, gender, race, education, experience), and the outcome of the stop (e.g., warning, citation, arrest, search, seizure of contraband). The objective of data analysis is to identify racial patterns of disparity. Tillyer et al. (2010) classify existing data analyses on the basis of whether they deal with the initial decision to stop or with the outcome of a stop.

In *initial stop* studies, the actual rate of stops by driver's race is compared with benchmark data providing the expected rate of stops assuming no police bias exists. The difficulty of the approach rests in identifying accurate benchmarks of the expected driver population at risk of being stopped. *Post-stop outcome* focuses on identifying racial disparities in a specific outcome of a stop by taking the set of stops as the reference population. An example of post-stop outcome analysis includes checking whether the search for drugs among stopped vehicles is biased against the driver's race.

5.2 Data analysis methods

In the initial stop studies, comparisons between the observed data and the benchmark data simply entail calculating the ratio of the observed number (or proportion) of stops of a racial group over the expected number (or proportion) of people of that racial group in the benchmark data. The statistical significance

of the calculated ratio can be assessed by classical tests of proportions, such as the binomial test or the Z-test. The main problem is then to determine the relevant population, i.e., the benchmark data. Engel and Calnon (2004), and (Blank et al., 2004, Chapter 9) outline the strengths and limitations of six primary data sources and their use in the design of benchmark data: census data, observations of roadway usage, official accident data, assessments of traffic violating behavior, citizen surveys, and internal departmental comparisons. For example, census data was used in early studies only at an aggregate level by considering the race percentages of the total population as the standard of comparison. More refined approaches create baselines of race percentages of the driving population detailed by age and gender even at a city block level. Engel and Calnon (2004) also report a sample study, taken by the Pennsylvania State Police Department, in data collection and benchmark design. Grogger and Ridgeway (2006) depart from the explicit definition of benchmark data. They report that the capability of a police officer of observing the race of a driver before stopping him/her decreases from daylight to darkness. Racial profiling is then tested by a logistic regression in which the dependent variable is the probability that a black driver is stopped given, among the independent variables, a darkness/daylight indicator.

Turning to post-stop outcomes studies, Tillyer et al. (2010) divide the methods from the literature into five classes. The first is multivariate analysis, which typically boils down to a regression model. The weakness of the assumption of the independence of residuals can be mitigated by grouping variables regarding stops and drivers, officers, and other community characteristics into a hierarchical structure. A form of multi-level modeling ensures that the regression coefficients are not biased across the levels of the hierarchical structure (Raudenbush and Bryk, 2002). The omitted-variables bias problem, however, is still recognized as the main limitation of this class of methods. The second class, called the *outcome test* or the *hit-rate test*, consists of studying the relative proportions of an outcome across the drivers' race assuming a condition of equilibrium. Later in this subsection, we detail the approach of Knowles et al. (2001). The main problem with such approaches is to test whether or not the data collected refer to a situation of equilibrium (Engel, 2008). The third class, called *propensity score*[20] methods, consists of matching observations that are the same for all the independent variables except race. Intuitively, the propensity score $w(\mathbf{x})$ is a weight of the distribution $f()$ of the unprotected group (comparison group, in statistical terminology) so that it becomes identical to the distribution of the protected group (treatment group): $f(\mathbf{x}|protected) = w(\mathbf{x})f(\mathbf{x}|unprotected)$. Solving for $w(\mathbf{x})$, and applying the Bayes theorem, we obtain:

$$w(\mathbf{x}) = K\frac{f(protected|\mathbf{x})}{1 - f(unprotected|\mathbf{x})}$$

for a constant K. After weighting the observations of the unprotected group, the only features for which the protected and the unprotected group should differ are race and, possibly, the stop outcome. The degree of representation of stop outcomes over races is then assessed by the discrimination measures in Figure 1, calculated over the weighted observations. Propensity score analysis is also closely linked to causal analysis (Rubin, 2005), since it aims at answering the question "what kind of outcomes would we have observed had the stops involving protected drivers involved unprotected drivers instead?". This class is also subject to the omitted-variable bias problem. The fourth class of methods consists of mapping stops to geography and then performing GIS (Geographic Information Systems) analyses, such as cluster analysis, spatial regression, comparisons with spatial layers regarding community crimes, economic and social indicators. The fifth class of methods deals with trend analysis, as a means to visualize changes in stop outcome rates over time. Temporal analysis techniques, such as ARIMA, can be also adopted.

Let us now detail a specific approach of the outcome test class. With the aim of maximizing an expected payoff, Knowles et al. (2001) propose an *equilibrium model* in which both agents and motorists behave strategically. The paper reports an empirical analysis on data gathered by the Maryland State Police, consisting of 1590 observations including race, sex, year, make, model and date of the vehicle stopped; time, location of stop; and whether illegal drugs (specifying the quantity) were found. We use r to denote race, and c to denote other characteristics that are potentially used by officers in the decision to search cars. Let $P(G|r, c)$ be the probability that a motorist of race r and characteristics c carries drugs. The

[20]The statistical basis of propensity score weighting can be found in Rosenbaum and Rubin (1983).

expected payoff of police in searching him/her is:

$$P(G|c, r) - t_r \tag{4}$$

where t_r is the marginal cost of searching motorists of race r. Similarly, let $\gamma(c, r)$ be the probability that a motorist is searched. The expected payoff of a motorist for carrying drugs is modeled as:

$$\gamma(c, r)(-j(c, r)) + (1 - \gamma(c, r))v(c, r) \tag{5}$$

where $j(c, r)$ is the expected cost of being convicted, and $v(c, r)$ is the expected value for carrying drugs. If (4) and (5) are non-zero, there is an incentive for police (not) to stop motorists of a specific race and, respectively, for motorists of a race (not) to carry drugs. In a condition of equilibrium, both (4) and (5) are zero: police and motorists are willing to randomize their choices.

Racial prejudice or, in the terms of Becker's theory, taste-based discrimination, occurs if there is a preference to search motorists of a specific race. This is modeled by the null hypothesis that t_r is constant over r. Since the t_r's are unknown, this cannot be checked directly. The alternative, suggested by equating (4) to zero, of checking that $P(G|c, r) = P(G|c)$ *for every* possible c would require data to cover every possible context c. Knowles et al. (2001) show that, at equilibrium, if t_r is constant over r, then $D(r)$ is constant as well, where $D(r)$ is the frequency of being found guilty by race r, also called the *hit-rate*. Testing that $D(r)$ is constant can be done by a standard χ^2-test on a contingency table by summarizing the number of guilty and non-guilty searched motorists over race. This test does not need data on any specific context c, and, as a consequence, it does not suffer from the omitted-variable bias problem. If the null hypothesis that t_r is constant cannot be rejected, differences in the probability of search over races can be explained as a form of statistical discrimination. Empirical results on the Maryland State Police dataset show no racial prejudice in general. When restricted to searches for large quantities of drugs, however, white motorists are searched disproportionately compared to the probability of being found guilty.

5.3 Annotated bibliography

Surveys and comparisons. One of the first surveys on racial profiling is by Engel et al. (2002). A more recent paper by Tillyer et al. (2010) reviews initial stop and post-stop outcome approaches. Engel (2008) reviews the outcome test approaches and critiques their main assumptions on the behavior of police and citizens from legal and social perspectives. The adequacy of statistical analysis of racial profiling in addressing legal issues is also discussed by Tillyer et al. (2008). The paper reports an interesting debate on the relative contributions and influences of data analysis and laws. Whitney (2008) overviews statistical data collection and analysis methods in traffic stops and searches from a legal viewpoint, arguing that strong statistical associations could be used in practice by U.S. courts as *prima-facie* evidence of discrimination. Baker and Phillipson (2011) compare the U.S. and U.K. policies in racial and religious profiling, claiming that the E.U. principle of (under-)proportionality is more effective in preventing such forms of discrimination.

Initial stop approaches. Some alternative means for collecting benchmark data have been investigated. Alpert et al. (2004) propose as a reference benchmark the not-at-fault driver demographic data from two-vehicle crashes occurred within the area of vehicle stop points. Jobard and Lévy (2011) gathered stop data by furtively following French police officers and taking photos of the motorists via mobile phones. The benchmark population is obtained by manually recording the characteristics of the drivers entering the place of stop observations. The odds ratio measure is considered in the analysis of the contingency table of stopped vehicles against gender, age, and physical appearance. Results show that police tend to over-stop young men dressed in the style typical of youth culture. Among these, two thirds belong to racial minorities, as estimated from the benchmark data. Quintanar (2009) sets the benchmark population of speeders by collecting the tickets issued by automated cameras. The proportions of women and blacks who were issued tickets by police officers were found to be significantly higher than the benchmark. Using a dataset of stops by the New York Police Department (NYPD), Gelman et al. (2007)

found that, after controlling for population distribution and crime distribution for violent crimes and weapons offenses, blacks and Hispanics are subject to "stop-and-frisk" about twice as often as whites. To handle variations in stop rates in cities where the percentage of blacks is high, they use three different hierarchical Poisson regressions in which the number of stops is modeled according to the percentage of the black population. Ridgeway and MacDonald (2009) propose an "internal benchmark" method consisting of comparing the distribution of stops of an officer with the distribution of all the other officers. In a dataset of 500,000 pedestrian "stop-and-frisks" made by 3,000 officers, the approach detected 15 officers who significantly stopped more blacks and Hispanics than those predicted by the benchmark.

Post-stop outcome approaches. Regression models have been widely adopted in post-stop approaches. See the cited surveys for extensive references. Recent approaches include Anbarci and Lee (2008), who analysed a dataset of more than 16,000 speeding tickets to test for in-group discrimination in granting a discount on the fine to motorists of the same race as the police officer. Using data on the Washington State Patrol, Pickerill et al. (2009) broaden the research from a binary dependent variable (search/no search) to a ternary variable by distinguishing the search into discretionary and mandatory cases. Blalock et al. (2007) consider gender discrimination in issuing traffic tickets, by analysing data of traffic stops from five locations. The first and the third cited papers report conclusions in contrast with conventional wisdom. Anbarci and Lee revealed no same-race preferences; rather, African-American officers are more likely to give a "discount" on tickets in favor of white motorists. (Blalock et al., 2007) found some evidence of gender bias only if location is included in the regression model. In addition, men are more likely to receive tickets only in two out of five locations.

The influential paper by Knowles et al. (2001) has been extended and criticized by several works, as reported in the survey by Tillyer et al. (2010). A well known limitation of the outcome test approach is the *infra-marginality* problem, namely the fact that the *marginal* cost t_r in (4) cannot be directly observed, but only approximated via the *average* cost. Marginal costs can be observed only if all the variables that officers use in taking stop decisions are known, i.e., it is reduced to the problem of the omitted-variable bias. Partial improvements of the work of (Knowles et al., 2001) in this line can be found in Anwar and Fang (2006); Antonovics and Knight (2009). Both papers conduct probit regression analyses on real datasets, which also contain information on the race of the officer. Their basic idea consists of using the latter information to distinguish between taste-based and statistical discrimination: in the presence of taste-based discrimination, the search rate is expected to increase when the race of the officer and the race of the motorist differ. Gardner (2009) critiques the equilibrium conditions of the approach of Knowles et al. (2001) in the case of racially integrated communities, which can be explained as follows. Suppose that blacks and whites obtain illegal drugs from the same distribution networks. Any increase in stopping black motorists who purchase drugs via such a network will deter whites from using the same network. Therefore, the marginal cost t_r cannot be constant over race r, even if there is no taste-based discrimination by the police.

The reference work for the propensity score approach can be found in Ridgeway (2006), who adopts risk difference as a discrimination measure, called the *race effect*, and estimates propensity scores by logistic regression on a dataset of the Oakland Police Department. Results show that black drivers suffer racial disparities both in terms of search rate and duration of the stop.

Related approaches. Besides vehicle stops, racial profiling has been studied, with similar analytical techniques, in other contexts. Adopting a logistic regression model, Gabbidon et al. (2008) consider consumer profiling, e.g., in curtailing shoplifting. They reason on survey data designed to determine the experience encountered by black students as victims of racial profiling. They found racial profiling in several types of business establishments. Alesina and Ferrara (2011) propose an outcome test model for checking racial bias in the context of capital sentencing. They found that the probability of judicial error for minority defendants who killed white victims is greater (about 15%) than the error for minority defendants who killed minority victims. Persico and Todd (2005) extend the theoretical model of (Knowles et al., 2001) to the context of passenger screening for airport security. Leopold and Meints

(2008) review the profiling of employees in automated fraud detection, including the use of intrusion detection systems, skill management tools, and fraud prevention systems. Finally, Schreurs et al. (2008) focus on the applicability of non-discrimination laws to group profiling in the private sector, e.g., in selecting customers for services, offers, or discounts.

6 Observational studies III: credit and consumer markets

6.1 Approach and data collection

Discrimination in the lending process may occur at several steps, from advertising, to pre-application enquiries, to loan approval/denial, and loan administration (Turner and Skidmore, 1999). Loan approval or denial has received the most interest, e.g., in the context of house lending, credit granting, and mortgage lending. Among the various credit markets, mortgage lending has the most extensive literature. The main challenge, is, in any case, the difficulty of estimating the risk of granting a loan to an applicant on the basis of his/her financial capacity and personal characteristics.

In the U.S., the *Home Mortgage Disclosure Act* (HMDA) requires a certain group of lenders to gather census data on their mortgage applications and make it publicly available. Since 1990, the HMDA data has been integrated with the race and income of mortgage loan applicants. One of the first relevant contributions to the collection of data was provided by researchers at the Federal Reserve Bank of Boston in the *Boston Fed Study* (Munnell et al., 1996). They supplemented the original census HMDA data for Boston with additional information on the credit history of more than 3,000 individual applicants, including data from more than one hundred financial institutes. This work remains crucial as no comparable dataset has since been made available. There are a number of criticisms of the Boston Fed study (see e.g., Longhofer and Peters (1999),(Ross and Yinger, 2002, Chapter 5), (Turner and Skidmore, 1999, Chapter 3)) concerning data errors, misclassification problems, endogenous explanatory variables and the omitted variables bias (e.g., loan amounts and indicators of cosigner are missing).

Yezer (2010) discusses three conventional approaches of testing discrimination in loan approval/denial decisions: mortgage rejection, mortgage pricing, and mortgage defaults. In *mortgage rejection*, the disproportionate rate of rejection between racial groups of applicants is considered *prima facie* evidence of discrimination. Empirical studies include the analysis of HMDA data at the *bank level* (i.e., a model for each bank under analysis) or *market level* (i.e., a single model aggregating variables for several banks). An experimental comparison of the two levels of analysis was carried out by Blackburn and Vermilyea (2006). Other sources of data range from micro-lending data to on-line data derived from a peer-to-peer lending site. *Mortgage pricing* concentrates on a dataset of approved loans, by considering whether a protected group is systematically charged with the highest interest rates. Mortgage pricing studies cover gender and racial discrimination in consumer credit, such as credit cards and education loans, in private firm credit, in subprime home loans, and in household credit. *Mortgage default* studies assume the probability of mortgage defaults as a measure of discrimination. If different default rates are observed for equally creditworthy groups that differ on discrimination grounds, this is considered *prima facie* evidence of discrimination.

The spatial clustering of neighborhoods into racially homogeneous areas very often emerges in most cities characterized by ethnic diversity. Schelling's segregation model (Schelling, 1971) shows that there is a natural tendency to spatial segregation, as a collective phenomenon, even if each individual is relatively tolerant – in his famous abstract simulation model, Schelling assumed that a person changes residence only if less than 30% of the neighbors are of his/her own race. This explains why the practice known as *redlining* (see Hillier (2003) and (Turner and Skidmore, 1999, Chapter 4)), which consists of denying or limiting credit on the basis of residence – drawing a red line on the border of an urban neighborhood – is often an indirect way to discriminate on the basis of race. The case of less credit grants is often called *process-based* redlining, while the case of a smaller flow of credit funds is known as *outcome-based* redlining. The percentage of individuals of a protected group living in a neighborhood is often used as a measure of the *level of segregation* of the neighborhood. Empirical works combine HMDA data with census data and credit history, in order to correlate the level of segregation of a neighborhood with its credit denial rate – all other conditions being equal.

6.2 Data analysis methods

Multivariate regression models are the basic tool for data analysis in credit and consumer discrimination. As expected, the probability of rejection/approval is typically taken as the dependent variable for mortgage rejection approaches, the interest rate of the loan for mortgage pricing approaches, and the probability of loan default for mortgage default approaches. An advanced approach regarding credit rejection taking into account the impact of credit history is presented later in this subsection. A discussion of the limitations of data on mortgage defaults, including unobserved variables and sample-selection bias, is reported in Ross and Yinger (1999). Basically, since observed defaults occur for approved loans only, the data on default approaches is a biased sample of loan applications, namely those the loans were granted to. If protected groups are held at a higher standard for loan approval, the sample selection bias is greater for protected group members, hence their propensity to default should be lower. See Ross (1996, 1997) for more details plus a simulation approach to evaluate the impact of sample-selection bias. In terms of the statistical model used to test for the direct and indirect discrimination, Rachlis and Yezer (1993) summarize several criticisms regarding the statistical significance of single-equation regression models.

In the analysis of redlining, the availability of spatial attributes, in addition to an applicant's personal attributes and loan characteristics, means a wide range of specialized tools can be used, including spatial lag and spatial error regression models (LeSage and Pace, 2009), hot spot clustering (Bailey and Gatrell, 1996), and interactive and exploratory spatial analysis (Ward and Gleditsch, 2008). Spatial attributes include aggregate values of census tracts, such average age, average income, number of households, and distribution/segregation of population by race, gender, and so on. Measures of group segregation have been proposed in a large body of literature (James and Tauber, 1985; Reardon and Firebaugh, 2002) to model group dissimilarity, isolation, and concentration, also taking into account the structure of social networks and social interactions (Echenique and Fryer, 2007). A well-known social model of racial segregation motivated by preferences for neighborhood racial composition was originally proposed by Schelling (1971), and tested by Clark (1991). Rogers and McKane (2011) present a unified mathematical framework of Schelling's model and its many variants.

Let us now concentrate on a specific approach in the class of credit rejection. Han (2011) stratifies racial disparities by the length of the borrower's credit history. The objective is to study whether creditors get private information about borrowers (their capacity to return the loan) from observable characteristics and the length of the creditor's history. He follows an approach inspired by Altonji and Pierret (2001) in which lenders play the role of employers, borrowers the role of workers, and the credit history takes the role of the workers' productivity. Han uses the data collected in Munnell et al. (1996), with a modified regression model that accounts for the length t of credit history, defined as follows: $t = 0$ if the applicant has neither a consumer history, nor a mortgage history; $t = 1$ if the applicant only has a consumer history; $t = 2$ if the applicant has both a consumer and a mortgage history. A probit regression model is adopted, with $Y \geq 0$ denoting loan application approval:

$$Y = \sum_{i=0}^{2} (b_{xi}X + b_{gi}G + \gamma_i)\mathbf{1}_{t=i} + \alpha H_c \mathbf{1}_{t=1} + (\beta_1 H_c + \beta_2 H_m)\mathbf{1}_{t=2} + \epsilon$$

where:

- X is a vector of the characteristics of the loan (e.g., whether it is fixed-rate, the loan maturity in years, the vacancy rate, etc.), of the applicant (e.g., age, marital status, number of the applicant's dependents), and of the neighborhood;
- G is the racial indicator: 1 for black applicants and 0 for white applicants;
- γ_i is a factor accounting for the length i of the applicant's credit history;
- $\mathbf{1}_{t=i}$ for $t = 0, 1, 2$ are indicators of the non-zero length t of the applicant's credit history;
- H_c (resp. H_m) is an indicator of payment issues in the credit history (resp., mortgage history), such as the 60 days delinquency in consumer credit (resp., late payments in mortgage history).

Note that all the coefficients vary according to the length of the credit history. To test for the presence of taste-based discrimination, it is necessary to check whether a significant racial disparity coefficient

b_{gi} is non-zero for all credit history lengths $i = 0, 1, 2$. The disutility of lending to blacks, in other words, persists irrespectively of the information on their credit history. On the other hand, the presence of statistical discrimination can be checked by observing coefficients b_{gi} whose absolute value declines as i increases. Intuitively, for loan applicants with no credit history, a lender uses the observable characteristic G, the race, as a proxy for estimating his/her private information, thus supporting the assumption that the lack of data favors statistical discrimination. As the length of credit history increases, the lender relies only on information on X, H_c and H_m. Experimental analyses confirm that the importance of race in explaining loan approval decisions decreases with the length of credit history.

6.3 Annotated bibliography

Surveys and comparisons. A critical survey of theoretical models and empirical studies since 1975 on racial disparities in mortgage lending markets in the context of the fair housing legislation can be found in LaCour-Little (1999). The book by Ross and Yinger (2002) is a comprehensive discussion of the literature on mortgage discrimination in the U.S. Yezer (2010) categorizes the statistical models of mortgage discrimination into mortgage rejection, pricing, and default. Problems with such models are discussed, including sources of bias, lack of theoretical support, and unsatisfactory causation analysis. The collection edited by Turner and Skidmore (1999) reviews mortgage lending research, with chapters on issues and challenges, on pre-market discrimination, on criticisms of the Boston Fed Study, and on redlining. Dymski (2006) summarizes the main lines of research in credit and housing markets, covering the legal context, theoretical models, empirical studies on redlining, mortgage and housing discrimination, small business credit market discrimination, and the subprime "predatory lending" market. Yinger (1998) is a survey of the evidence of discrimination in consumer markets, including house rentals and sales, car sales, and product prices at fast-food restaurants. Finally, Hillier (2003) and (Turner and Skidmore, 1999, Chapter 4) review the main methodologies and types of data that have been investigated for redlining.

Credit rejection. This strand of the literature considers market-level models, typically by academic studies, and lender-specific models, typically by bank regulators such as the Office of the Comptroller of the Currency (OCC) in the U.S. Blackburn and Vermilyea (2006) compare data from eight banks and argue that market-level models produce larger estimated racial effects. The authors also provide a brief summary of the literature. Recent approaches using lender-specific models include Dietrich and Johannsson (2005), where logistic regression is adopted to test for age and gender discrimination on HMDA data, and Dietrich (2009), where HMDA data from 22 banks is analyzed through an OLS regression to unveil multiple discrimination on the grounds of ethnicity, race and gender. Clarke et al. (2009) compare the robustness of logit regression models with probit, gompit and complementary log-log models on data from five different banks. Evidence of discrimination is affected by both the regression method and the data sampling approaches adopted in the analyses. Goenner (2010) observes that different conclusions have been reached by researchers when extracting models from different subsets of variables from the Boston Fed Study dataset. He adopts a Bayesian Model Averaging approach[21] in order to average the results over a collection of models. From this, no significant evidence of racial discrimination can be concluded in the Boston Fed dataset. We also highlight the interesting study by Pope and Sydnor (2011) on race discrimination in *peer-to-peer lending*, which is an online market where individual borrowers and lenders directly engage in credit transactions with no intermediary. Data is from a specialized website (prosper.com) where borrowers present themselves through pictures and lenders browse loan listings. More than 110,000 listings are considered in the study, using OLS and logistic regression. Listings with a picture of a black person were 34% less likely to get funding than those with a picture of a white person.

Indirect discrimination of credit scoring in mortgage and consumer credit has been assessed by Avery et al. (2010), who had access to a unique dataset comprising 312 credit features of 301,536 individuals drawn from the records of a U.S. national credit reporting agency. Their regression model did not report evidence of disparate impact by race, ethnicity, or gender, and only a limited disparate impact by age,

[21] See Hoeting et al. (1999) for an introduction.

with older individuals having lower credit scores than young people. The impact of credit-based car insurance scores on protected groups of consumers was assessed by the Federal Trade Commission (2007), including the use of credit scores as a proxy for race and ethnicity.

Credit pricing. Typical factors affecting the price of a loan, i.e., its annual percentage rate (APR), include: risk elements (credit risk and prepayment risk), expenses (overheads and services), loan characteristics (type and channel of delivery), and negotiation factors. See Avery et al. (2005) for a more detailed list and discussion. Higher priced loans may hide discrimination if, in addition to the aforementioned factors, race or gender affect the price of loans. The in-depth study by Courchane (2007) found that only a small percentage of APR racial differentials cannot be explained by typical factors. Edelberg (2007) highlights a lower unexplained racial disparity in data prior to 1995, while Weller (2008) argues that taste-based and statistical discrimination persist. Such contrasting conclusions are frequent in the literature: they are mostly due to the use of different datasets and/or different explanatory variables and methods. Cheng et al. (2011) argue that women pay higher rates because they tend to choose recommended lenders, while men tend to search for the lowest rate. Gender discrimination in loan amounts was also analyzed by Agier and Szafarz (2013) on a dataset of 34,000 loan applications to a single microfinance institution. Evidence was found of a "glass ceiling" against female entrepreneurs in terms of loan size.

The subprime crisis in the late 2000s prompted specific studies on mortgage redlining in the "predatory lending" market, even though the problem was already recognized in the early 2000s (e.g. Calem et al., 2004). The issue here is whether minorities are less creditworthy on average than whites, which would then justify a higher rate of subprime in minority populated neighborhoods. Bocian et al. (2008); Reid and Laderman (2009); Squires et al. (2009); Wyly et al. (2008) and Vicki et al. (2009) relate race or racial segregation with high lending costs. The first two papers combine HMDA data with a proprietary database of subprime loans. The last three papers use HMDA data and census data to estimate neighborhood racial composition via dissimilarity and isolation indexes. All the studies found strong evidence of racial and ethnic segregation in the subprime market. Ding et al. (2008) adopt multivariate regression on HDMA data of a large U.S. city, suggesting that the role of funding by government agencies, such as the Federal Housing Administration should be expanded. Rugh and Massey (2010) analyse data in order to isolate subprime lending as the causal mechanism through which segregation influenced foreclosure. They adopt a two-stage OLS regression model on data obtained from RealtyTrac, the U.S. largest provider of foreclosure listings. Finally, an interdisciplinary economic-social science study of the emergence of the subprime crisis can be found in Dymski et al. (2011).

Credit default. Credit default consists of estimating the probability of the default of an applicant on the basis of his/her financial capacity, of the risk of the local market, the loan terms, and the applicant's personal characteristics – including race, gender, and other discrimination grounds. If discrimination exists, protected groups should have lower default rates than non-protected ones. Berkovec et al. (1994) adopt a linear regression model on a dataset of mortgages insured by the U.S. Federal Housing Administration. A higher likelihood of default for black householders was found, which then justifies higher credit denial or interest rates. A refined study, accounting for the market concentration of lenders, is reported by Berkovec et al. (1998). The recent paper by Brown and Simpson (2010) reviews the role of default rates in mortgage discrimination research, and conducts a simulation analysis to compare the performance of reverse regression (used by LaCour-Little (1996)) versus the commonly-adopted logit regression models. The latter methodology outperforms reverse regression. Finally, Chan et al. (2011) relate the 90-day mortgage default rate available in HDMA data to neighborhood characteristics, hence highlighting possible sources of redlining.

Credit discrimination at firm-level. Blanchflower et al. (2003) compare the credit rejection of black-owned and white-owned small businesses. The paper deals with data from the U.S. National Surveys of Small Business Finances, finding that the probability of having the credit denied for a black owner is twice that of a white owner. The same source of data is also used in Cavalluzzo et al. (2002); Blanchard

et al. (2008); Asiedu et al. (2012), in analysing gender, race and ethnic discrimination against small business owners. Muravyev et al. (2009) consider gender using cross-national data from the Business Environment and Enterprise Performance Survey, an establishment level survey which since 1999 has been conducted by the European Bank for Reconstruction and Development and the World Bank. Many of the aforementioned studies cover both loan-denial decisions (using probit regression) and differences in the interest rate charged (using OLS regression). Data include control variables on credit history and the financial health of the firm. Blanchard et al. (2008) also discuss an approach to differentiate between taste-based and statistical discrimination by identifying variables that might be associated with lender prejudice. Albareto and Mistrulli (2011); Alesina et al. (2008) adopt probit regression to investigate whether micro firms (i.e., sole proprietorship) run by migrants or women pay more for credit than firms run by native or male entrepreneurs. They had access to unique datasets from the Central Credit Register run by the Bank of Italy. Fraser (2009) considers the small business credit market, using probit regression on data from the U.K. Survey of SME Finances. With the exception of (Blanchard et al., 2008), the research confirms the presence of gender discrimination in firms, in terms of higher interest rates (in Italy, (Alesina et al., 2008)), of restrictions in funding start-up companies (in Europe, (Muravyev et al., 2009)), and for small firms in terms of a market concentration (in U.S., (Cavalluzzo et al., 2002)).

Redlining. "The Color of Money" is a famous series of newspaper articles[22] highlighting the issues of redlining. Updates and extensions of the original study are reported by Holloway and Wyly (2001).

Outcome-based redlining considers the supply of mortgage credit to a neighborhood as the dependent variable, normalized by the ratio of loan requests over the number of residential properties in the neighborhood. An early survey of outcome-based redlining can be found in Schill and Wachter (1993). Phillips-Patrick and Rossi (1996) highlight that using regression models on data aggregated at the level of neighborhoods may lead to inconclusive and misleading results.

Process-based redlining concentrates on the probability of credit denial in minority neighborhoods compared to majority neighborhoods. Approaches using the Boston Fed Study data include Hunter and Walker (1996), Tootell (1996) and, more recently, Ross and Tootell (2004). The second and the third papers highlight indirect forms of redlining, such as the practice of forcing applicants of a minority neighborhood to buy private mortgage insurance. The sociological implications of insurance redlining are discussed by Squires (2003). One particular line of research has tested the theory, originally proposed by Lang and Nakamura (1993), which assumes the number of recent sales in a neighborhood as an "information externality" able to affect loan decision outcomes. Blackburn and Vermilyea (2007) was the first empirical paper in these lines. A logistic regression on HDMA and census data shows that, after controlling for information externality, the percentage of racial minorities in a neighborhood positively affects the decision outcome in an area. Blank et al. (2005) report three types of analyses: a dissimilarity index of a census tract, comparing the expected approval rate against the observed one; a three-way cross-tabulation approach; and a logistic regression. Structured interviews with bank officers were used to complement the study. All methodologies support the evidence of discrimination against blacks and Hispanics in housing and mortgage lending markets. Turning to other forms of credit, Campbell et al. (2008) use a logit regression on observations from a single bank on unsecured consumer loans, arguing that, due to the lack of collateral, they are more prone to discriminatory decisions compared to mortgages. Cohen-Cole (2009) and Brevoort (2011) link location-based information on race with individual credit files, using a proprietary dataset of credit histories from a major credit bureau. The former paper finds redlining practises against minority neighborhoods. The latter paper, however, shows that such results are largely due to omitted variables and data sampling bias.

Interesting case studies involving the use of descriptive statistics, GIS mapping, multivariate regression or generalized linear mixed modeling are reported by Hernandez (2009); Zonta (2012); Kuebler (2012); Silverman (2005, 2008). All of them integrate HMDA data with census data. While the first four papers found a significant correlation of negative decisions with race and/or ethnicity, the latter shows that disparities in lending outcome are affected by education, neighborhood socioeconomic distress,

[22] http://powerreporting.com/color

residential stability, and the age of neighborhood housing stock, but not by race. The case study by Aalbers (2007) found a positive correlation between redlining and low income, unemployment, and ethnicity. Aalbers adopts discriminant analysis for correlation, using Wilks lambda test of significance, and hierarchical and k-means cluster analysis to detect outlier neighborhoods. Finally, a probit regression is adopted by Ezeala-Harrison et al. (2008) on a dataset of micro-data of house loans. They found that minorities have great difficulties in accessing credit for housing in white neighborhoods, whilst this is not the case in black neighborhoods.

Discrimination in consumer markets. The practice of a retailer, wholesaler, or manufacturer selling the same product, with the same marginal cost, at different prices based on a buyer's willingness to pay is known as *price discrimination* (Armstrong, 2006). Price discrimination is not considered part of the economics of discrimination: while it is rare nowadays for black consumers to pay a higher price for the same quantity of items, this certainly occurred in the past in Jim Crow law states in the U.S. Nevertheless, discrimination in specific consumer markets has been considered in the past. Differential treatment in pricing in the car sale market was investigated using regression analysis of U.S. nationwide data (Goldberg, 1996). Legal issues of profiling customers and consumers for automated individual recommendations/promotions in marketing campaigns, possibly resulting in differential treatment, are discussed by Körffer and Meints (2008). The disparate treatment of the U.S. Harmonized Tariff Schedules on imported goods for women is discussed in Bertero (2011). Finally, Bayer et al. (2012) analyse a large dataset of proprietary transactions, with a rich set of features on prices, on buyer's and seller's characteristics, and on houses. Their regression model shows that black and Hispanic buyers pay a premium of about 3% compared to white buyers.

7 Observational studies IV: other contexts

This section surveys a few other contexts where observational techniques of discrimination analysis have been applied. We cover gender bias in academia, discrimination in healthcare, and in social interactions.

7.1 Gender bias in academic grants

The problem of gender differences in research is of concern to all major funding institutions. The E.U. regularly publishes a report on the status of gender funding in the member states (European Commission, 2009), and it promotes gender equality in scientific research[23]. E.U. legislation includes an explicit resolution on women and science (Council of the E.U., 1999), which notably preceded the resolutions on racial and employment equality. The National Science Foundation (NSF) in the U.S. supports the development of systemic approaches to increase the representation and advancement of women in academic science, technology, engineering, and mathematics through the ADVANCE program[24]. Broader overviews of studies and findings on gender in (scientific and technological) research have been conducted by the European Commission (2012) and by UNESCO (2007).

Forms of gender discrimination may explain women's under-representation in academia, both past and present. The surveys by Bentley and Adamson (2003) and Ceci and Williams (2011) cover multidisciplinary literature on gender differences in manuscript reviewing, grant funding, university admission, and hiring and promoting in research.

The influential paper by Wennerås and Wold (1997) reports a study on post-doctoral fellowship applications to the Swedish Medical Research Council (MRC) in 1995. A total of 62 applications were submitted by men and 52 by women: 16 men were funded (25.8%) versus 4 women (7.7%). Applicant's sex and scientific competence are considered as independent variables in a linear regression model estimating the score assigned by the reviewers. Scientific competence as a control factor is measured in terms of the number of published journal articles, their citation count, and total impact of those journals. The regression shows that "a female applicant had to be 2.5 times more productive than the average male

[23] http://www.yellowwindow.be/genderinresearch
[24] http://www.portal.advance.vt.edu

applicant to receive the same score". However, subsequent studies by several funding societies in Europe and North America fail to show evidence of sex bias in approval rates (Ceci and Williams, 2011). In fact, Sandström and Hällsten (2008) analysed data on applications to the MRC in 2004 and found a *reversed* gender bias, namely a small but significant effect in favor of funding womens grants compared to men with the same scientific competence score.

RAND (2005) investigates grant applications in the period 2001-2003 to the NSF, the National Institutes of Health (NIH), and the Department of Agriculture. No evidence of gender bias was generally found after controlling for age, academic degree, institution, grant type, institute, and application year. There were two exceptions, partly explainable by the lack of further control variables. First, women received only 63% of the amount of funding awarded to men by the NIH. Secondly, women who applied in 2001 were less likely than men to submit applications in the next two years. Wilson (2004) explain the lower amount of funding granted to women by their marginalization within the scientific community, by their segregation to lower rank positions, and by their smaller social networks – all of these factors affecting their chances of funding possibilities.

Ley and Hamilton (2008) highlight that, whilst there is now generally gender equality between students and instructors, there is still a striking drop in the roles of assistant professors and professors – i.e., a *glass ceiling* in science. The authors analysed more than 100 thousand applications between 1996 and 2007 to NIH grant programs to determine whether gender differences occur at some stage of a researcher's career, which may explain the observed attrition. While they found a decrease in female applicants for grants throughout a researcher's career, there is substantial equity of the rates of funded applications between males and females at all stages of the process.

Marsh et al. (2008) summarize the major findings of an eight year research program on the analysis of peer reviews in grant applications to the Australian Research Council. Their dataset includes 2,331 proposals rated by 6,233 external assessors, out of a total of 10,023 reviews. They consider issues such as: reliability of reviews, in the sense of an agreement of reviewers across individual proposals and across disciplines; trustworthiness of reviewers nominated by applicants; bias of national reviewers, who give more favorable evaluations than international ones; the positive influence of academic rank, in the sense that professors are more likely to be funded due to their experience and successful research track records; and the positive influence of the prestige of the university affiliation and of the applicant's age. They also consider the influence of an applicant's gender, finding that 15.2% of funded applications were led by females, which was exactly the same percentage as female applicants. Once again, although women are under-represented in the applications pool, they are equally represented in the funded pool.

Regarding the analytical methodology, research on peer review studies has carried out statistical analyses mainly by means of variants of regression, correlation, Z-tests of proportions, and more rarely by analysis of variance and discriminant function analysis. Multi-stage peer reviews have been also analysed with latent Markov models (Bornmann et al., 2008). Multi-level regression[25], adopted by Jayasinghe et al. (2003) and Mutz et al. (2012), includes a subject level random variable modeling variation in a cluster of data. For instance, Jayasinghe et al. (2003) adopt multi-level regression to take into account correlations in the cluster of ratings of a reviewer, and in the cluster of ratings in the same field of study. Another interesting methodology is meta-analysis, the objective of which is to detect which conditions (e.g., application type, research area, country, etc.) affect the ratio of success between women and men. Bornmann et al. (2007) collected data from 66 peer reviewed procedures, and, using generalized regression models, concluded that there were overall robust gender differences, with men having greater odds of approval than women by about 7%, after controlling for discipline, funding agencies, and year. The subsequent study Marsh (2009), using multi-level regression, revised the conclusion for research grant applications, but it remained the same for fellowship applications.

[25] See Goldstein (2011).

7.2 Discrimination in healthcare

The labor market models of statistical and screening discrimination have been studied in the context of healthcare. "Expert" patients are shown to receive different treatment, for example obstetricians, midwives and doctors have a higher rate of Caesarean sections than other patients. Grytten et al. (2011) analyse the Norway Medical Birth Registry data, consisting of about 2.5 million records, using logit regression. Their results support a form of screening discrimination, where expert patients exhibit better communication with doctors. Chandra and Staiger (2010) and Anwar and Fang (2011) differentiate taste-based and statistical discrimination in a dataset of heart attack treatments and of emergency ward operations, respectively. Mcguire et al. (2008) test racial and ethnicity discrimination in the treatment of depression using survey data. Finally, Green et al. (2007) use implicit association tests to measure the implicit bias of physicians in treating thrombolysis in black vs white patients. Blacks patients are unconsciously perceived as less cooperative. Implicit association tests[26] assess unconscious preferences by measuring the time it takes subjects to match representatives of social groups to particular attributes (e.g., good, bad, cooperative, and stubborn). A recent review on racial and ethnic discrimination in healthcare is reported by Shavers et al. (2012).

Let us now turn to discuss in more detail an interesting paper by Balsa et al. (2005) on statistical discrimination models in healthcare. They carried out an empirical study of the disparities originating in clinical encounters, where statistical discrimination is interpreted as the "rational application of probabilistic decision rules when uncertainty surrounds the doctor's estimate of a patient's health status". Prior experience may induce a doctor to take into account race or ethnicity as a factor in clinical decisions, for example because a given symptom is more predictive of a certain disease in a racial group compared to other groups. Data was collected from a medical observational study covering 11,664 patients on depression, hypertension, and diabetes. Complementary data from national surveys was used to estimate the prevalence of diseases and risk factors nationwide, which represents doctors' prior knowledge. The authors first consider a "traditional disparities" probit model:

$$Pr(diagnosis) = \Phi(\delta_1 S + \delta_2 R + \beta X + \eta Y) \tag{6}$$

where the probability of a certain diagnosis is estimated from indicators (0/1 values) of patient symptoms (S), minority race (R), other characteristics of the patient (X), and the doctor's characteristics (Y). A negative and significant δ_2 is usually claimed as evidence of discrimination. However, nothing can be concluded on the reasons for such a disparate treatment, for example whether it is due to racial prejudice or to prior knowledge. The model is then augmented as follows:

$$Pr(diagnosis) = \Phi(\delta_1 S + \delta_2 R + \delta_3 S\theta + \delta_4 \theta + \delta_5 RS + \beta X + \eta Y) \tag{7}$$

to take into account the prior knowledge θ (prevalence of the disease among similar individuals), and the joint effects of race and symptoms (modeled by the product of indicators RS) and of symptoms and prior knowledge ($S\theta$). A Bayesian information criterion is adopted to test which model best fits the data. For instance, (7) performs better than (6) for hypertension diagnoses – hence the contribution of prior knowledge is significant. A decrease in absolute value of the regression coefficient δ_2 of the race indicator R from (6) to (7) measures the extent to which prior knowledge explains the disparate treatment, hence, the extent of statistical discrimination. For hypertension, the decrease is 75%. Finally, the coefficient δ_5 enables miscommunication problems (i.e., screening discrimination) to be tested between doctors and minority patients. In fact, a negative and significant δ_5 means that the way doctors perceive a signal of disease depends on the racial group of the patient.

However, the possibly discriminatory nature of the prior knowledge θ is not discussed in the paper. If past diagnoses were biased due to existing "background" discrimination, then θ is itself affected by discrimination. Thus, (7) performs better than (6) as a consequence of the included-variable bias. Rather than being an explanatory variable, prior knowledge has the effect of perpetuating the vicious cycle of discrimination described in Section 2.3.

[26] https://implicit.harvard.edu/implicit

7.3 Social interactions

In the labor market, social interactions refer to forms of externalities not regulated by the price mechanism, in which actions of a reference group affect an individual's preferences (see Scheinkman (2008) for a theorical and historical perspective, and Ioannides (2008) for an empirical perspective). The *social multiplier* is a ratio measure of the effects on the average action of individuals with or without the influence of their peers. In the context of discrimination in labor economics, Montgomery (1991, 1994) firstly modeled job searches through off-line social networks, assuming ties between workers and employers that satisfy a binomial probability distribution. In equilibrium, wages are determined not by workers' actual skills but by the number of social ties they hold. As a consequence, if white workers have a better network, they have an advantage over equally productive black workers. A richer network structure is assumed in the work by Calvo-Armengol and Jackson (2004), yet the underlying economic model adopted is simpler. van der Leij et al. (2009) analyse data from the U.S. Social Capital Benchmark Survey 2000 and from the Estonian Labor Force Survey, claiming that social network segregation, rather than discrimination, explains the racial wage gap. Inter-ethnic communication is estimated by aggregate telephone call data between geographical regions. Ioannides and Loury (2004) survey the economic literature on job information networks, social interactions, and wage inequalities. More recent models of job inequality explained by the structure of social networks can be found in Calvo-Armengol and Jackson (2007) and Kim (2009).

Beyond the labor market research, social interactions can provide a fertile ground for the emerging area of social network analysis. We believe that extending concepts and methods to the analysis of discrimination and segregation in large on-line social networks is a promising direction for future research. An example of the potential of such approaches is the work of Stephens-Davidowitz (2012), who used the number of Google search queries with racially charged terms as a proxy to estimate racial animosity towards Barack Obama in the 2008 presidential elections.

8 Quasi-experimental studies

This and the following section are concerned with quasi-experimental and experimental approaches to data collection and analysis, as a means to overcoming the limitations of observational data in controlling the factors which may affect a decision.

8.1 Approach and data collection

Auditing, also known as field experiments, follows a quasi-experimental approach to investigate the presence of discrimination by controlling some of the factors that may influence decision outcomes. It consists of using pairs of *testers*, also called *auditors*, who have been matched to be similar on all characteristics that may influence the outcome except race, gender, or other grounds of possible discrimination. The tester pairs are then sent into one or more situations in which discrimination is suspected, for example renting an apartment or applying for a job, and the decision outcome is recorded. The difference in outcomes between the paired groups then provides a measure of discrimination. This is a quasi-experimental approach since some factors are not controlled by the researcher, for example the environment of a job selection procedure or the presence of other applicants.

Situation-testing occurs as a special case of auditing when the testers come into contact with the decision maker. This is the case, for instance, of job interviews involving human testers, who are selected and trained in advance to act in a similar way to each other. The strong point in favor of situation-testing is that testers can record the cause of discrimination, such as prejudice or stereotypes, hence allowing for a causal analysis of the discrimination cases. The main limitation of situation-testing is the cost of data collection. A sufficiently large number of audits needed to produce statistically significant results is expensive. In addition, situation-testing cannot be applied at all in some contexts, for example in wage increases, or in the disparate application of contractual terms (e.g., in house lending, see Roscigno et al. (2009)). Moreover audit studies are typically not available in a time series, which is a pre-requisite for analysing changes over time.

	unprotected group		
protected group	benefit denied	benefit granted	
benefit denied	a	b	
benefit granted	c	d	
			n

$$GR = b/n \quad NR = (b - c)/n$$

Figure 3 4-fold contingency table and discrimination measures for audit studies.

In *correspondence-testing*, the data scarcity problem is mitigated by designing paired ad-hoc fake resumes or application forms to be sent to advertised vacancies. In a landmark study, Bertrand and Mullainathan (2004) prepared a collection of 5,000 resumes from standard templates, which they randomly assigned to either a typical white American name (e.g. Emily or Greg) or an African-American sounding name (e.g. Lakisha and Jamal). Resumes were sent to newspaper help-wanted ads in Chicago and Boston, and the invitation to interview was taken as the decision outcome of the experiment. Resumes with white-sounding names had a 10.08% chance of receiving a callback, while equivalent resumes with African American sounding names had a 6.70% chance of being called back. In addition, higher quality resumes elicited 30% more callbacks for white-sounding names and only 9% more for African-American sounding name. The absence of personal contact between the decision maker and the testers represents the main criticism against the correspondence-testing approach. Larger opportunities for data collection are offered by Internet job advertisement services, known as *e-recruiting*,which enable automated searches for job position adverts in large databases, and automated submission and notification of applications. In addition, contexts other than employment can be covered, such as discrimination in product advertising in internet marketing.

8.2 Data analysis methods

The surveys by Bendick (2007); Riach and Rich (2002); Yinger (1998) assume the *net rate of discrimination* as a measure of discrimination in audit studies. Consider Figure 3. Each cell in the 4-fold contingency table reports the number of paired audits satisfying the cell coordinates. n is the total number of audits. Two measures are defined on the contingency table: the *gross rate of discrimination* GR, defined as the share of audits in which the auditor of the protected group is treated less favorably than his or her teammate; and the *net rate of discrimination* NR, defined as the gross rate minus the share of audits in which the auditor of the protected group is treated more favorably than his or her teammate.

Another example of data analysis is the internet-based auditing approach of Nunley et al. (2011). In the context of eBay auctions, they test the presence of in-group racial discrimination, which occurs when individuals prefer to interact with people of their own racial group. Their experiment defines two groups of fictitious users, say whites and blacks, and a set of low-price items. The race of bidders, which is unknown, is inferred by splitting the set of products into two groups for which the relative racial compositions of the bidders is likely to differ. For each product of both groups, two different seller accounts are opened choosing a typical "white" name (e.g. Jake or Dustin) and a typical "black" name (e.g. Tyrone or Jamal). Each seller uses a similar picture and description of the product. The experiment consisted of 288 auctions over a period of 11 months. Auctions lasted five days and began with an initial price of one cent. To prevent buyers recognizing that the two accounts were from the same source, an auction for a given pair began two days prior to the other, and this delay was alternated between black and white sellers. The price received by a product from a prospect buyer is modeled by an OLS regression:

$$P = \beta_0 + \beta_1 R + \beta_2 E + \beta_3 Y + \beta_4 W + \epsilon \tag{8}$$

where P is the price received, R is an indicator denoting whether the seller's and buyer's race match, E indicates whether the auction was the first or the last in its pair, Y is a boolean denoting whether the seller has a good rating (a yellow star, namely 10 positive feedbacks from past buyers), W captures other

determinants of the price dependent on the selling period and the product type, and ϵ is the residual term. The OLS regression model is estimated for sub-samples of low/high competition products, for different levels of seller feedback, and for predominantly white and black products. The level of competition for a product is measured by the number of possible auctions for similar products on eBay in the same time period. A positive value for β_1 in (8) is an indication of in-group discrimination, and it turns out to be higher for low competitive markets and for low seller ratings, which is in line with the economic models of discrimination. For instance, the mean price paid for a product is \$6.37 if a seller's race and buyer's race match, and \$5.93 otherwise.

8.3 Annotated bibliography

Surveys and comparisons. Bendick (2007) reviews more than 30 situation-testing studies in employment discrimination in the U.S. The net rate of discrimination in these studies ranges from 2% to 38% when comparing Afro-American to European-American testers, from 12% to 25% when comparing Hispanic to Anglo named auditors[27], from 8% to 40% when comparing females to males, and from 17% to 41% when comparing older and younger auditors. Rorive (2009) covers the use of situation-testing in the E.U. member states. Another summary of recent audit studies in employment discrimination can be found in Pager (2007), also including an interesting appendix on the design and implementation of a generic situation-testing study. Riach and Rich (2002) review and compare the statistical significance of field experiments on racial, sex, and disability discrimination in employment, and on discrimination in housing sales and rentals. The collection edited by Fix and Tuner (1999) covers the use of auditing to measure racial discrimination in housing, employment, commercial transactions, and business. Criticisms of the conclusions drawn from audit methods are discussed by Heckman and Siegelman (1993) and Heckman (1998), while Riach and Rich (2004) comment on the ethical implications of such methods. Quillian (2006) discusses how the measurement of discrimination through audit methods should incorporate psychological theories of prejudice.

Situation-testing. Pager and Quillian (2005) show that blind-methods, i.e., experiments with information (partly) hidden from the participants, are essential in gathering trustworthy data. They found that employers who indicated a greater likelihood of hiring ex-offenders in a (non-blind) survey were no more likely than other employers to hire an ex-offender in a (blind) audit study. In order to overcome problems related to the selection and training (e.g., motivation, fairness, unbiased behavior) of auditors, Moreno et al. (2004) select auditors from real job-seekers retrieved from a job intermediation agency database. Massey and Lundy (2001) correlate speech patterns with racial discrimination in housing by conducting audits through telephone calls with speakers exhibiting a "black sounding" or a "white sounding" accent. They found that black women face higher fees and a higher probability of being refused a rental.

Bendick et al. (2010) analyse the racial discrimination of restaurant employers from two perspectives. From an observational perspective, regression analysis over census data reveals no adverse impact over race, gender, and nationality on earnings after controlling for workers education and experience. However, applied to a specific sector, namely fine dine establishments, a situation-testing perspective reveals a high wage gap for these groups. Pager et al. (2009) study low-wage labor markets by analysing situation-testing data through a hierarchical logistic regression accounting for a random effect for each employer by capturing correlations between observations of the same employer[28]. Their results show that black applicants are half as likely as equally qualified white applicants to receive a callback or job offer.

Paired tests were extensively conducted in the U.S. House Discrimination Survey, the results of which are reviewed by Turner et al. (2002); Turner and Ross (2005). Differential treatment in pricing car sales has been investigated using audit studies since the 1990s (Ayres and Siegelman, 1995). Siegelman (1999) summarizes approaches and legal cases using auditing in public accommodation and in consumer markets discrimination such as in car rentals, shopping, taxicab stops, and home repairs. In the context

[27]Bendick (2007) reports a single "outlier" study, whose methodology and rigor have been seriously questioned, which found a net rate of -10%, denoting a better treatment of the Hispanic testers compared to their counterparts.
[28]The statistical basis of the approach can be found in Ghosh et al. (2000).

of consumer services, Myers et al. (2010) deal with the unequal treatment of female customers in coffee shops in terms of time-to-be-served, rather than the cost of the service. The differential treatment is found to increase with the proportion of male waiters.

Correspondence-testing. Experiments in the line with Bertrand and Mullainathan (2004) have been recently conducted by Banerjee et al. (2009), who sent fictitious resumes in response to job openings to test for discrimination against caste and religious groups in India; and by Kaas and Manger (2012), who sent applications to student internship advertisements to test for discrimination against immigrant students. (Banerjee et al., 2009) found no significant differences in the call-back rates, whilst (Kaas and Manger, 2012) found that German-sounding names compared to Turkish-sounding names raise the average probability of a callback by about 14%. Neumark (2010) proposes a statistical test to evaluate whether the data collected in an audit study shows biased racial discrimination due to the fact that applicants from different groups exhibit different variances of their characteristics, yet having the same average values. They use the same data as Bertrand and Mullainathan (2004). Other grounds of discrimination covered by correspondence-testing in job applications include sex (Riach and Rich, 2006; Booth and Leigh, 2010), obesity and physical appearance (Rooth, 2009; Bóo et al., 2012), sexual orientation (Drydakis, 2009), ethnicity (McGinnity et al., 2009; Carlsson and Eriksson, 2012), or a combination of these (Patacchini et al., 2012). For grounds not related to the name of an applicant (e.g., obesity or sexual orientation), the authors add a manipulated portrait photo to the fake CV or an explicit reference to well-known homosexual communities. Most of the cited works adopt statistical regression for the data analysis. Results reveal preferences for women in several types of jobs, such as trainee chartered accountants and secretaries. However, (Booth and Leigh, 2010) observe that this occurs only for jobs where the percentage of females is already greater than 80% – thus highlighting a form of occupational segregation. Other findings include a negative bias in Europe against gay men ((Drydakis, 2009)), yet not against gay women ((Patacchini et al., 2012)), and a positive bias, known as the "beauty premium", in favor of attractive females in low-skilled jobs ((Patacchini et al., 2012)).

In addition to recording the outcome of job applications, Carlsson and Rooth (2007) also conduct interviews with employers who invited at least one applicant, in order to gather information on the gender and ethnicity of the recruiter and on the characteristics of the employer's company, with the intended objective of extracting a meaningful classification of discriminatory as opposed to non-discriminatory companies. They found a greater negative bias in small firms and for low-skilled occupations. Arai et al. (2008) extend the same experiment to a two-stage correspondence test. In the first stage, employers receive CVs of an equally observable quality between the two groups of applicants. In the second stage, CVs of the protected group have an advantage of two years experience. The aim is to study the degree of reduction in discrimination. Results show that differences in call-backs disappear at the second stage for females, whilst remaining for males belonging to an ethic minority. Linear regression models are adopted in estimating call-back probabilities. A similar approach is conducted by Zussman (2012) in analyzing 384 fictitious ads for used vehicles, distinguished by typical Arab and Jewish names. In the follow-up experiment, potential buyers who responded to the ads are contacted to participate in a survey measuring their racial attitudes. Finally, a variant of correspondence-testing has been adopted in order to assess gender discrimination in schools. Hinnerich et al. (2011) compare grades assigned to a written test by teachers who knew (non-blind test) or did not know (blind test) student names. The study found no differences in grading between male and female students in both non-blind and blind test evaluations.

Internet-based testing. Correspondence-testing can be applied in-the-large by resorting to data crawled from the Internet. One example is Booth et al. (2010), who tackle the critical issue of writing CVs starting from templates and real CVs crawled from the web. They submitted more than 4,000 fictional applications for entry-level jobs to an on-line e-recruiting web site. The applicant's name was set as an indicator of ethnicity, sex, or both. The authors report a ratio of callback rates of Anglo-Saxon sounding names over minority (e.g., indigenous, Italian, Chinese) names ranging from 1.12 to 1.64. For male applicants, the ratio ranges from 1.21 to 1.76; for females, from to 1.03 to 1.82. Another recent example includes Ahmed

et al. (2013), who code homosexuality by stating that an applicant lives with a partner of the same sex. The synthetic generation of resumes is discussed by Lahey and Beasley (2009), who designed a parametric tool to mitigate the bias due to the limited number of (template) CVs that can be generated manually. The legal implications of possible discrimination in e-recruiting, compared to classical means of recruiting, are discussed by Hogler et al. (1998). Eriksson and Lagerstrøm (2012) on the other hand, use an Internet database of job applications where job seekers can choose whether or not to reveal sex and ethnicity. From a sample of 18,000 real job seekers in Sweden, the callback rate is calculated for the full sample or for blind observations only. A Poisson model is adopted to estimate the number of contacts. They found no evidence of ethnic discrimination and only a little evidence of gender discrimination.

The impact of a seller's race in online auctions is investigated by Doleac and Stein (2010). They consider iPods ads posted through local on-line classified advertisements featuring a photograph of the product held by a black hand, a white hand, or a tattooed hand. Discrimination on the grounds of race and of a tattoo was found. Similar to (Nunley et al., 2011), they use the number of competing sellers as a measure of market competition. The disparities disappear in competitive markets. The quality of reactions from potential buyers (categorized through good/bad e-mails) is also analysed. Results do not vary w.r.t. the quality of the e-mails. Regarding the analytical methods, they adopt several regression models, including negative binomial regression, probit and OLS.

Discrimination in on-line rental housing is investigated by Ahmed and Hammarstedt (2008). Fictitious e-mails with ethnic and gender sounding names are sent in response to house rental ads. To distinguish between taste-based and statistical discrimination, Bosch et al. (2010) prepared two types of e-mails. The first contains no information about the candidate other than his/her name. The second email contains a detailed description of his/her current high-quality employment status. The same implications are studied by Hanson and Hawley (2011), Baldini and Federici (2011) and Friedman et al. (2010), with the additional purpose of determining racial discrimination across neighborhoods. All these works show a negative bias against minorities. The introduction of positive information in the e-mails (e.g., high social class, employment information, education level) reduces or eliminates such bias.

9 Experimental Studies

9.1 Approach and data collection

Field experiments construct control groups by matching similar individuals, for example applicants for a job in situation-testing, and then observing the outcome of a quasi-experiment in a *natural* environment, e.g., in a job selection procedure. Empirical data from field experiments reflect a variety of environmental factors: disentangling these factors may be difficult if not impossible. Controlled experiments are conducted in an *artificial* environment, such as a laboratory, under tightly controlled conditions, including the selection of treatment and control groups and strict rules regarding their behavior and actions. On the one hand, the impact of a specific factor can be evaluated by systematically (e.g., randomly) varying it. On the other hand, confounding variables and other extraneous stimuli can be minimized. In other words, whilst observational and quasi-experimental studies can only conclude correlations, for example, whether grounds for discrimination co-occur with a negative decision, controlled experiments lead to causation conclusions, i.e., whether a negative decision is due to the presence of grounds for discrimination. Controlled experiments are very useful to test the predictions of theoretical models or to pre-test the impact of some ruling or laws before their application. In addition, controlled experiments are repeatable, by definition, and less expensive than field experiments. The main criticism against controlled experiments is that they suffer from a lack of realism, also called the *external validity* problem, due to the selection of individuals from a specific population (e.g., volunteer students) and to the influence on their behavior of the artificial experimental environment. See Charness and Kuhn (2011); Levitt and List (2007) for an in-depth discussion of these issues.

There are two main classes of controlled experiments: laboratory and natural experiments.

Levitt and List (2007) review five classes of games played by individuals or groups, which are adopted in the economic literature on *laboratory experiments* to measure social preferences, including fairness, trust, and conditional reciprocity. Fershtman and Gneezy (2001) for instance, adopt a trust game to test for

ethnic discrimination. The trust game assumes two players. "Player A" is given a fixed amount of money and he/she is asked to transfer a certain amount to "player B". The transferred amount is triplicated. Then, "player B" can choose to transfer any part of the received amount back to "player A". Players A and B are randomly paired from Israeli male students of different ethnicities. The lower average amount of money transferred to players of a specific ethnicity, compared to others, is considered evidence of discrimination against this ethnicity. Fershtman and Gneezy found that the average transfer to players of European or American origin was about three times the amount transferred to players of Asian or African origin.

Natural experiments occur in real life (controlled) situations. The experimenter only observes the behavior of participants, who typically are not aware of the experiment. Television game shows are a typical example, where discriminatory choices of participants can be studied in a controlled environment. In addition to the criticism of external validity, natural experiments also have the problem that not all factors can be controlled for, for example how participants are selected in a TV game show. A detailed example is discussed in the next subsection.

9.2 Data analysis methods

Controlled experiments are aimed at measuring *reactions* to a person's characteristics while manipulating only the person's attributes that are grounds for possible discrimination, such as race, sex, and age. Negative reactions that constitute an expression of various degrees of discrimination include: verbal hostility (e.g., negative personal comments), disparate written ratings, non-verbal indicators of hostility (e.g., seating distance, tone of voice, facial expressions), choice of whether to associate with the person (e.g., in a work team, in competitive games), and deviation from a baseline (e.g., from the optimal strategy in a game). Most of these reactions are typically not detectable in observational data. Measures of disproportionality in reactions include those in Figure 1. The main critical issue in experimental design is the construction of the control and treatment groups, which, in our context, means the construction of the protected and unprotected group members. In laboratory experiments, a randomization selection process is believed to result in a balance over omitted variables, thus overcoming the omitted-variable bias of non-experimental approaches. However, a causation conclusion of the findings of an experiment can be only drawn for the, possibly biased, sample of population considered in the experiment, not for the general population. This is known as the *internal validity* problem of controlled experiments.

We will now focus on a natural experiment based on the *Weakest Link* television game show. In the Weakest Link, players compete with one another in a series of rounds. In each round, contestants take turns answering trivia questions, with the goal of building a "chain" of answers. The longer the chain, the higher the jackpot increase. At the conclusion of each round, players secretly vote for a contestant to be eliminated, i.e., the weakest link of the chain. The one who receives the highest number of votes is eliminated from the next round. The last two contestants compete for the whole accumulated jackpot on the basis of the number of correct answers in the last round. Levitt (2004) collected data from 161 shows out of a total of 3,273 person-votes. The recorded attributes include race, age, education, and region of residence of the players, and their performances in each round, including the number of questions, percentage of correct answers, number of votes received. The aim of the work is to test for the presence of taste-based and statistical discrimination. If there is no discrimination, the following optimal strategy of contestants should be observed: in the early rounds, votes should be uniformly distributed; in the middle rounds, the weakest players should be voted for, irrespective of their race, since they do not contribute to increasing the jackpot; and in the final rounds, the strongest players should be voted for, irrespective of their race, since the other players would prefer not to compete with them in the last round. On the other hand, taste-based discrimination implies persistent punishment in the number of votes against the protected group in all rounds. Statistical discrimination, instead, implies higher votes for players of the protected group in the early rounds, since the race of the other players is the only perceived characteristic of their performance; however, in the middle and final rounds, the behavior should switch to the optimal strategy of no-discrimination. Levitt adopts OLS regression models to estimate the number of votes received by a player in a round with independent variables including the contestant's attributes and performances in previous rounds. Three separate models are computed, for the early, middle, and

final rounds of the game. The author claims that a negative binomial regression, typically adopted for estimating count variables, yields identical results. The results of the regression models are consistent with statistical discrimination against Hispanics and with taste-based discrimination against the elderly.

9.3 Annotated bibliography

Surveys and comparisons. A review of laboratory studies from the early 1980s is reported by Crosby et al. (1980). Harrison and List (2004) propose a taxonomy of experiments ranging from conventional lab experiments, to field experiments in a natural environment without the awareness of participants. Charness and Kuhn (2011) survey the contribution of lab experiments to labor economics, including studies on discrimination. Methodological issues are also discussed, such as when and why to adopt lab experiments, and how they compare to field experiments. Levitt and List (2007) warn about the limits of generalizing results obtained from lab experiments, particularly from experiments measuring social preferences through games. (Blank et al., 2004, Chapter 6) present an overview of experimental and audit methods with a discussion of their strengths and weaknesses.

Laboratory experiments. In labor market research, Feltovich and Papageorgiou (2004) aim to investigate how quickly employers learn about the ability of a group of workers and, consequently, update their prior opinions. They design a lab experiment in which subjects (modeling employers) draw four cards (modeling new employees) from two buckets (modeling white and black groups) for a total of nine rounds, with the first six rounds having a bias of card values in favor of the "white" bucket. Page's test for ordered alternatives[29] is adopted for the statistical validation of the results, which confirm that players tend to the optimal strategy of choosing from the "white" bucket during early rounds, and randomly for late rounds. Dickinson and Oaxaca (2009) consider "second moment" statistical discrimination by risk-adverse employers offering women lower wages because of their perceived higher *variance* in productivity (vs lower *average* productivity in "first moment" discrimination). They design a job auction game to simulate a labor market. At each round of the game, employers decide the wage distribution among workers, and the workers decide the productivity distribution among employers. Evidence of employer statistical discrimination is found by comparing the wage distribution among the various rounds. Lab experiments were revealed to be essential in complex contexts, e.g., due to the scarcity of reliable data or to the complexity of psycho-social interactions. Examples include the study of discrimination against previously convicted job applicants (Falk et al., 2008), and discrimination in the co-employment of temporary and permanent workers (Güth et al., 2010). Both papers use a gift-exchange game, where employer players offer a wage to worker players, who in turn decide their working input. (Güth et al., 2010) classify workers into temporary or permanent, and wages into publicly known or not. They found discrimination against permanent workers when wages are not public. (Falk et al., 2008) add to the CVs of workers information on previous conviction. Ex-offenders suffer from a negative bias in offered wages. Finally, List (2004) argues that laboratory experiments can complement audit studies in establishing the taste-based or statistical nature of observed discrimination. He integrates audits with games in the sports card market.

A few study focus on favoritism towards attractive people, known as the *beauty premium* (Hamermesh and Biddle, 1994), through a collaborative game. Andreoni and Petrie (2008) run different experiments in which the employer player is or is not given information on the skills of worker players. Differential wages are offered to good-looking employee players, the beauty premium, when this information is not given. Mobius and Rosenblat (2006) also run experiments in which the employer player has different interactions with worker players, ranging from only looking at the CV with photo, to having phone interviews, to face-to-face interviews. The authors claim that the beauty stereotype explains only about 30% of the differential treatments. The impact of stereotypical beliefs regarding looks and accents is also tested by Rödin and Özcan (2011) using a betting game. In each round of the game, employer players decide workers' wages for a pool of candidates. In the first round, the only information available to employers is a facial portrait

[29] See Siegel and Castellan (1988).

of each worker. In the second round a voice message is added, and finally, in the last round the educational level is also revealed. Results show that: in the first round, candidates that are perceived as non-Swedish are given lower wages; and in the second round, candidates speaking Swedish with a foreign accent are given lower wages, hence the speech-based stereotype dominates over the look-based stereotype.

In the context of group formation, Falk and Zehnder (2007) study discrimination in team trusting by a variant of the investment game where pairs of players are selected from residents in the same or in different districts of the same city. They found evidence of in-group favoritism, where groups are made up of people from the same district. Another variant of the investment game is used by Slonim and Guillen (2010) in dealing with gender preferences in group member selection. They show discriminatory choices of men against men due to taste-based discrimination, and of women against women due to expectations or stereotypes. Castillo and Petrie (2010) assess racial differences in the *rankings* of preferences, claiming that the results of their analyses are the same along several regression models (random-effects OLS, random-effects Tobit, OLS regression with standard errors clustered on the individual doing the ranking, rank-ordered logit, and fixed-effect regressions). They use a group formation game in which, at each round, participants decide between a private and a public investment. Only the latter yields a return to all the members of the group. In the first two sessions of the game (each consisting of ten rounds), players are chosen randomly to make up the group. In the third session, players can choose their own group. Players use race to select individuals from their own group in the first rounds of the session, but not in the final rounds. This is claimed as evidence of statistical discrimination.

Natural experiments. Discrimination analysis of data gathered from the Weakest Link game show is also reported by Antonovics et al. (2005). In addition to taste-based and statistical discrimination, they test strategic discrimination defined as the use of gender and race as "focal points for collusive behavior". However, they found no evidence of strategic discrimination. In terms of analysis, they adopt conditional logit regression models. Other studies on the same game show include: Antonovics et al. (2009), testing whether the gender of the opponent affects performance; and Bagues and Villadoniga (2008), testing a statistical discrimination model of the "similar-to-me" effects on evaluators. The former paper shows that the results obtained by replicating the Weakest Link in a lab experiment differ from those of the TV show. The latter paper finds in-group preferences in votes with respect both to gender and age. Lee (2009) analyses data on the number of viewers of the *American Idol* TV show, collected through a Nielsen People Meter system, showing that the racial composition of viewers determines the race of the person eliminated, while the racial composition of the contestants affects viewing patterns.

The classic game known as the "prisoner's dilemma" has inspired several TV game shows where members of a team compete for a prize either by cooperating with each other or by 'defecting' from the group. List (2006) analyses data from *Friend or Foe?* using probit regression to detect statistical discrimination in the selection of team members. Men discriminate against women by cooperating less in mixed-gender groups compared to an all-men group. This is not the case for women. Dilks et al. (2010) aim at extending economic theories of discrimination by incorporating insights from structural social psychology. They test implications of the extension by conditional logit regression on data from the *Survivor* reality show.

The "beauty premium" has also been studied using natural experiments. Belot et al. (2012) disentangle the sources of favoritism to attractive people by analysing data from a TV game show based on the prisoner's dilemma. They use a conditional logit regression model, which shows that less attractive contestants are more likely to be expelled by the game in the earlier rounds. No evidence of taste-based discrimination is found by Darai and Grätz (2012) in the analysis of data gathered from *Golden Balls*, a game show based on the prisoner's dilemma. Other types of data sources have also been considered. A recent work by Cipriani and Zago (2011) studies favoritism towards attractive students in taking exams at university. The productivity of a student is measured by the sum of his/her grades. A regression model is calculated describing student's productivity in terms of his/her characteristics and his/her physical attractiveness. They contrast differences in productivity between written exams, where the beauty premium should be absent, and oral exams. The related work by Breda and Ly (2012) uses a

natural experiment to study gender differences between oral and blind written tests. The study shows that professors tend to discriminate in favor of females in typical male-concentrated disciplines, and in favor of males in female-concentrated ones. The effectiveness of blind decisions in reducing gender discrimination has been evaluated in an influential study regarding auditions for orchestras by Goldin and Rouse (2000). By using probit regression, they found that, when a black screen is used to hide the identity of an applicant, the probability of success for women increases.

10 Knowledge discovery studies

Discrimination data analysis has recently been tackled from a knowledge discovery, or data mining, perspective which is orthogonal to the method of data collection (observational, quasi-experimental, or experimental). The perspective addresses discrimination along two lines: discrimination discovery and discrimination prevention.

10.1 Discrimination discovery from data

Discrimination discovery from data consists in the actual discovery of discriminatory situations and practices hidden in a large amount of historical decision records. Unlike hypothesis testing, where a statistical analysis is adopted to confirm a predetermined hypothesis of discrimination, the aim of discrimination discovery is to *unveil* contexts of possible discrimination. The legal principle of under-representation with reference to legally-grounded measures of the degree of discrimination suffered by protected groups has inspired existing approaches for discrimination discovery based on pattern mining. Starting with a dataset of historical decision records, Pedreschi et al. (2008); Ruggieri et al. (2010b) propose the extraction of classification rules of the form $\mathbf{A}, \mathbf{B} \rightarrow \mathbf{C}$, called *potentially discriminatory* (PD) rules, to unveil contexts $\mathbf{B}$ of the dataset where the protected group $\mathbf{A}$ suffered from under-representation w.r.t the positive decision $\mathbf{C}$ or from over-representation w.r.t the negative decision $\mathbf{C}$. $\mathbf{A}$ is a non-empty itemset, whose elements belong to a fixed set of protected groups. $\mathbf{C}$ is a class item denoting the negative decision, e.g., credit denial, application rejection, job firing, and so on. Finally, $\mathbf{B}$ is an itemset denoting a context of possible discrimination. The degree of over-representation is measured by the ER measure (called *extended lift* by the authors) from Figure 1, where the contingency table refers to only those records satisfying $\mathbf{B}$. For example:

$$\text{RACE=BLACK, PURPOSE=NEW_CAR} \rightarrow \text{CREDIT=NO}$$

is a PD rule about denying credit (the decision $\mathbf{C}$) to blacks (the protected group $\mathbf{A}$) among those applying for credit in order to buy a new car (the context $\mathbf{B}$). PD rules are ranked according to their measure value. The approach has been implemented on top of an Oracle database (Ruggieri et al., 2010c) by relying on tools for frequent itemset mining. A GUI for visual exploratory analysis has been developed by Gao and Berendt (2011). Extensions of the approach by Pedreschi et al. (2009); Ruggieri et al. (2010a) deal with: (1) any measure built from a contingency table as in Figure 1; (2) the statistical significance of measures of discrimination of an extracted rule; (3) the discovery of affirmative actions and favoritism; and (4) the filtering of rules that can be supported by a genuine occupational requirement. The rankings imposed by the discrimination measures in Figure 1 are investigated by Pedreschi et al. (2012): the choice of the reference measure critically affects the rankings of PD rules, with the RC (mainly adopted in the U.S.) and the RR (mainly adopted in the E.U.) measures exhibiting the largest differences. The approach has also been extended by Luong (2011) to unveil and represent discrimination in a "rich semantic" form for semi-structured business data based on support from ontologies.

There are two major drawbacks of the approach. First, there is no control, within a context $\mathbf{B}$, of the characteristics of the protected group $\mathbf{A}$ as opposed to others in such a context. Since the discrimination measures are defined as aggregated values, they mix decisions for people that may be very different in terms of the skills required to obtain a benefit. This results in an overly large number of rules that need to be further screened to filter out explainable discrimination. Second, the result of the knowledge discovery process is a (possibly large) set of classification rules, which provide *local* and *overlapping*

niches of possible discrimination: a *global* description is lacking of who is and is not discriminated against. Luong et al. (2011) exploit the idea of situation-testing. For each member of the protected group with a negative decision outcome, testers with similar characteristics are searched for in a dataset of historical decision records. If there are significantly different decision outcomes between the testers of the protected group and the testers of the unprotected group, the negative decision can be ascribed to a bias against the protected group, thus labeling the individual as discriminated against. Similarity is modeled via a distance function. Testers are searched for among the k-nearest neighbors, and the difference is measured by one of the functions in Figure 1 calculated over the two sets of testers. After this kind of labeling, a global description of those labeled as discriminated against can be extracted as a standard classification task. A real case study in the context of the evaluation of scientific projects for funding is presented by Romei et al. (2012).

The approaches described so far assume that the dataset under analysis contains attributes that denote protected groups. This may be not the case when such attributes are not available, or not even collectable at a micro-data level, as in the case of the loan applicant's race. Ruggieri et al. (2010a,b) adopt a form of rule inference to cope with the indirect discovery of (either direct or indirect) discrimination. Intuitively, a PD rule $\mathbf{A}, \mathbf{B} \rightarrow \mathbf{C}$, which is not directly computable, can be deduced from classification rules without protected items, e.g., of the form $\mathbf{D}, \mathbf{B} \rightarrow \mathbf{C}$, where $\mathbf{D}$ is an itemset of properties strongly correlated in the context $\mathbf{B}$ with some protected group $\mathbf{A}$ (e.g., low education level with minority race in some city neighborhood). The correlation information is called background knowledge, and is itself coded as an association rule.

10.2 Discrimination prevention in data mining

Data mining and machine learning models extracted from historical data may discover traditional prejudices that are endemic in reality (i.e., taste-based discrimination), or patterns of lower performances, skills or capacities of protected-by-law groups (i.e., statistical discrimination). Mining algorithms may then assign the status of general rules to such practices, with the result that these rules may be deeply hidden within a data mining classifier. Sensitive application fields include personnel selection[30] and human resources management, gender identification, fraud detection in insurance claims, intrusion and crime detection, and credit scoring. Discrimination prevention in data mining and machine learning consists of extracting models (typically, classifiers) that trade off accuracy with non-discrimination.

Discrimination prevention has been recognized as an issue in a tutorial by (Clifton, 2003, Slide 19) where the danger of building classifiers capable of racial discrimination in home loans has been put forward. For example, mortgage redlining can be easily recognized as a common pattern in loan data. Similar warnings were raised in the context of automated credit scoring systems, for example by the Consumer Federation of America & National Credit Reporting Association (2002). The naïve approach of deleting attributes that denote protected groups from the original dataset does not prevent a classifier from indirectly learning discriminatory decisions, since other attributes (sometimes called *redundant encodings*) that are strongly correlated with them could be used as proxies by the mining algorithm. This has been repeatedly observed in several contexts: credit scoring (Fortowsky and LaCour-Little, 2001), predictive statistics (Pope and Sydnor, 2007), and data mining (Ruggieri et al., 2010b; Calders and Žliobaitė, 2013). The earliest approaches to developing classifiers that do not incur in indirect discrimination were proposed in the area of credit scoring. Fortowsky and LaCour-Little (2001) originally considered logistic regression models, proposing a re-weighting of regression coefficients. The incorporation of anti-discrimination law requirements in credit scoring systems is currently challenging – for a recent approach, see Lieli and White (2010). As for data mining approaches, we foresee four non mutually-exclusive strategies regarding discrimination prevention.

The first strategy consists of a controlled distortion of the training set (a pre-processing approach). Kamiran and Calders (2012) compare sanitization techniques such as changing class labels based on

[30] Where automated systems are designed to be "capable of reproducing the judgement of the recruitment consultant" (Kessler et al., 2009).

prediction confidence, instance re-weighting, and sampling. Žliobaitye et al. (2011) prevent excessive sanitization by taking into account legitimate explanatory variables that are correlated with grounds of discrimination, i.e., genuine occupational requirements. The approach of Luong et al. (2011) extends to discrimination prevention by changing the class label of individuals that are labeled as discriminated against. Using additional background knowledge, Hajian and Domingo-Ferrer (2012) perturb the training set so as to reduce the degree of indirect discrimination, which is measured in terms of the number of rules that could be inferred using the approach of Ruggieri et al. (2010a,b). The advantage of pre-processing approaches is that they are independent of the learning model and algorithm at hand.

The second strategy is to modify the classification learning algorithm (an in-processing approach), by integrating it with anti-discrimination criteria. Calders and Verwer (2010) consider three approaches to deal with naïve Bayes models, two of which consist in modifications to the learning algorithm: training a separate model for each protected group; and, adding a latent variable to model the class value in the absence of discrimination. Kamiran et al. (2010) modify the entropy-based splitting criterion in decision tree induction to account for attributes denoting protected groups. Kamishima et al. (2012) measure the indirect causal effect of variables modeling grounds of discrimination on the independent variable in a classification model by their mutual information. Then they apply a regularization (i.e., a change in the objective minimization function) to probabilistic discriminative models, such as logistic regression.

The third strategy is to post-process the classification model once it has been extracted. Pedreschi et al. (2009) alter the confidence of classification rules inferred by the CPAR algorithm. Calders and Verwer (2010) act on the probabilities of a naïve Bayes model. Kamiran et al. (2010) re-label the class predicted at the leaves of a decision tree induced by C4.5.

Finally, the fourth strategy assumes no change in the construction of a classifier. At the time of application, instead, predictions are corrected to keep proportionality of decisions among protected and unprotected groups. Kamiran et al. (2012) propose correcting predictions of probabilistic classifiers that are close to the decision boundary, given that (statistical) discrimination may occur when there is no clear feature supporting a positive or a negative decision.

There is a summary of the contributions of research groups in discrimination discovery and prevention in a recent book edited by Custers et al. (2013). The book covers the design of data-driven technologies, such as data mining and profiling, and this design is privacy and discrimination aware. Legal and ethical issues of discrimination, privacy, and profiling are also dealt with. The chapter by Gellert et al. (2013) reports a comparative analysis of data protection and anti-discrimination legislations.

A technical approach that tackles both privacy in disclosing data mining models and discrimination in applying such models is discussed by Hajian et al. (2012). The work considers classification rule models and measures privacy by k-anonymity and discrimination by the number of PD rules. Dwork et al. (2012) propose a model of fairness of classifiers and relate it to differential privacy in databases. The model imposes that the predictions over two similar cases are also similar. The similarity of cases is formalized by a distance measure between tuples. The similarity of predictions is formalized by the distance between the distributions of probability assigned to class values.

11 Conclusions

We have provided an annotated bibliography of the main references and recent approaches on discrimination data analysis. Our focus has been on analytical methods, whereas pure psychological or sociological studies have not been covered. Nevertheless, the relevant literature is abundant and spread over many disciplines. We cited 422 references, 50% of which have appeared in the last five years (2008-2012), and 73% of which have appeared in the last decade (2003-2012). This demonstrates a never-ending interest in the topic, which has recently been stimulated by the advent of the *knowledge society*. A society where decisions – small or big, by business or policy makers – can be taken on the basis of models, patterns, profiles, and rules of human behavior distilled from the ubiquitous digital traces generated as a side effect of our (on-line and off-line) lives.

Acknowledgements

We are grateful to Franco Turini and to the anonymous reviewers for several helpful comments.

References

M. Aalbers. What types of neighbourhoods are redlined? *Journal of Housing and the Built Environment*, 22(2): 177–198, 2007.

I. Agier and A. Szafarz. Microfinance and gender: Is there a glass ceiling on loan size? *World Development*, 42(0): 165 – 181, 2013.

A. Agresti. *Categorical Data Analysis*. Wiley Series in Probability and Statistics. Wiley-Interscience, 2 edition, 2002.

A. Agresti and C. Brian. Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *The American Statistician*, 54(4):280–288, 2000.

A. Ahmed, L. Andersson, and M. Hammarstedt. Are gay men and lesbians discriminated against in the hiring process? *Southern Economic Journal*, 79(3):565–585, 2013.

A. M. Ahmed and M. Hammarstedt. Discrimination in the rental housing market: A field experiment on the internet. *Journal of Urban Economics*, 64(2):362–372, 2008.

D. J. Aigner and G. G. Cain. Statistical theories of discrimination in labor markets. *Industrial and Labor Relations Review*, 30:175–187, 1977.

G. Albareto and P. Mistrulli. Bridging the gap between migrants and the banking system. Economic working paper 794, Bank of Italy, Economic Research Department, 2011. http://www.bancaditalia.it.

E. M. Aldrich, P. S. Arcidiacono, and J. L. Vigdor. Do people value racial diversity? Evidence from Nielsen ratings. *Journal of Economic Analysis & Policy*, 5(1):Art. 4, 2005.

A. F. Alesina and E. L. Ferrara. A test of racial bias in capital sentencing. Working Paper 16981, National Bureau of Economic Research, 2011. http://www.nber.org.

A. F. Alesina, F. Lotti, and P. E. Mistrulli. Do women pay more for credit? Evidence from Italy. Working Papers 14202, National Bureau of Economic Research, 2008. http://www.nber.org.

G. P. Alpert, M. R. Smith, and R. G. Dunham. Toward a better benchmark: Assessing the utility of not-at-fault traffic crash data in racial profiling research. *Justice Research and Policy*, 6(1):43–70, 2004.

J. G. Altonji and R. M. Blank. Race and gender in the labor market. In O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics*, volume 3, Part C, pages 3143–3259. Elsevier, 1999.

J. G. Altonji and C. R. Pierret. Employer learning and statistical discrimination. *The Quarterly Journal of Economics*, 116(1):313–350, 2001.

T. Amemiya. Tobit models: A survey. *Journal of Econometrics*, 24(1-2):3–61, 1984.

American Psychological Association. Guidelines for psychological practice with lesbian, gay, and bisexual clients, 2011. http://www.apa.org.

N. Anbarci and J. Lee. Speed discounting and racial disparities: Evidence from speeding tickets in Boston. Discussion Paper 3903, Institute for the Study of Labor (IZA), 2008. http://ftp.iza.org.

J. Andreoni and R. Petrie. Beauty, gender and stereotypes: Evidence from laboratory experiments. *Journal of Economic Psychology*, 29(1):73–93, 2008.

K. L. Antonovics and B. G. Knight. A new look at racial profiling: Evidence from the Boston Police Department. Working Paper 10634, National Bureau of Economic Research, 2009. http://www.nber.org.

K. L. Antonovics, P. Arcidiacono, and R. Walsh. Games and discrimination: Lessons from the Weakest Link. *Journal of Human Resources*, 40(4):918–947, 2005.

K. L. Antonovics, P. Arcidiacono, and R. Walsh. The effects of gender interactions in the lab and in the field. *Review of Economics and Statistics*, 91(1):152–162, 2009.

S. Anwar and H. Fang. An alternative test of racial prejudice in motor vehicle searches: Theory and evidence. *American Economic Review*, 96(1):127–151, 2006.

S. Anwar and H. Fang. Testing for the role of prejudice in emergency departments using bounceback rates. Working Paper 16888, National Bureau of Economic Research, 2011. http://www.nber.org.

M. Arai, M. Moa Bursell, and L. Nekby. Between meritocracy and ethnic discrimination: The gender difference. Research Papers in Economics 2008:4, Institute for the Study of Labor (IZA), 2008. http://ftp.iza.org.

P. Arcidiacono, P. Bayer, and A. Hizmo. Beyond signaling and human capital: Education and the revelation of ability. *American Economic Journal: Applied Economics*, 2(4):76–104, 2010.

M. Armstrong. Recent developments in the economics of price discrimination. In R. Blundell, W. K. Newey, and T. Persson, editors, *Proc. of World Congress on Advances in Economics and Econometrics Theory and Applications*, volume 2, pages 1–46. Cambridge University Press, 2006.

K. J. Arrow. The theory of discrimination. In O. Ashenfelter and A. Rees, editors, *Discrimination in Labor Markets*, pages 3–33. Princeton University Press, 1971.

O. Ashenfelter and D. Card. Using the longitudinal structure of earnings to estimate the effect of training programs. *The Review of Economics and Statistics*, 67(4):648–660, 1985.

E. Asiedu, J. A. Freeman, and A. Nti-Addae. Access to credit by small businesses: How relevant are race, ethnicity, and gender? *American Economic Review: Papers & Proceedings*, 102(3):532–537, 2012.

Australian Legislation. (a) Age Discrimination Act, 2004; (b) Australian Human Rights Commission Act, 1986; (c) Disability Discrimination Act, 1992; (d) Racial Discrimination Act, 1975; (e) Sex Discrimination Act, 1984; (f) Victoria Equal Opportunity Act, 1995, (g) Queensland Anti-Discrimination Act, 1991, 2012. http://humanrights.gov.au.

D. H. Autor and D. Scarborough. Does job testing harm minority workers? Evidence from retail establishments. *Quarterly Journal of Economics*, 123(1):219–277, 2008.

R. B. Avery, G. B. Canner, and R. E. Cook. New information reported under HMDA and its application in fair lending enforcement. *Federal Reserve Bulletin*, 91(3):344–394, 2005.

R. B. Avery, K. Brevoort, and G. B. Canner. Does credit scoring produce a disparate impact? Finance and economics discussion series, Division of Research & Statistics and Monetary Affairs, Federal Reserve Board, 2010. http://works.bepress.com/kbrevoort/9.

I. Ayres and P. Siegelman. Race and gender discrimination in bargaining for a new car. *The American Economic Review*, 85(3):304–321, 1995.

I. Ayres, F. E. Vars, and N. Zakariya. To insure prejudice: Racial disparities in taxicab tipping. *Yale Law Journal*, 114 (7):1613–1674, 2005.

M. F. Bagues and M. J. P. Villadoniga. Why do i like people like me? Working Paper 2008-06, Universidad Carlos III, Departamento de Economía de la Empresa, 2008. http://papers.ssrn.com.

T. Bailey and T. Gatrell, editors. *Interactive Spatial Data Analysis*. Prentice Hall, 1996.

A. Baker and G. Phillipson. Policing, profiling and discrimination law: US and European approaches compared. *Journal of Global Ethics*, 7(1):105–124, 2011.

M. Baldini and M. Federici. Ethnic discrimination in the Italian rental housing market. *The Journal of Housing Economics*, 20(1):1–14, 2011.

A. I. Balsa, T. G. Mcguire, and L. S. Meredith. Testing for statistical discrimination in health care. *Health Services Research*, 40(1):227–252, 2005.

N. Bamforth, M. Malik, and C. O'Cinneide. *Discrimination Law: Theory & Context, Text and Materials*. Sweet & Maxwell, 1 edition, 2008.

A. Banerjee, M. Bertrand, S. Datta, and S. Mullainathan. Labor market discrimination in Delhi: Evidence from a field experiment. *Journal of Comparative Economics*, 37(1):14–27, 2009.

C. Barnard and B. Hepple. Substantive equality. *The Cambridge Law Journal*, 59(3):562–585, 2000.

T. Bauer and M. Sinning. An extension of the Blinder-Oaxaca decomposition to nonlinear models. *AStA Advances in Statistical Analysis*, 92:197–206, 2008.

T. K. Bauer and M. Sinning. Blinder–Oaxaca decomposition for tobit models. *Applied Economics*, 42(12):1569–1575, 2010.

P. Bayer, M. D. Casey, F. Ferreira, and R. McMillan. Price discrimination in the housing market. Working Paper 18069, National Bureau of Economic Research, 2012. http://www.nber.org.

G. S. Becker. *The Economics of Discrimination (Economic Research Studies)*. University of Chicago Press, Chicago, USA, 2 edition, 1971.

M. R. Bell. *Anti-Discrimination Law and the European Union*. Oxford University Press, 2002.

M. Belot, V. Bhaskar, and J. Van de Ven. Beauty and the sources of discrimination. *Journal of Human Resources*, 47 (3):851–872, 2012.

M. Bendick. Situation testing for employment discrimination in the United States of America. *Horizons Stratégiques*, 3(5):17–39, 2007.

M. Bendick, R. E. Rodriguez, and S. Jayaraman. Employment discrimination in upscale restaurants: Evidence from matched pair testing. *Social Science Journal*, 47(4):802–818, 2010.

J. T. Bentley and R. Adamson. Gender differences in the careers of academic scientists and engineers: A literature review. Special report, National Science Foundation, 2003. http://www.nsf.gov.

J. A. Berkovec, G. B. Canner, S. A. Gabriel, and T. H. Hannan. Race, redlining, and residential mortgage loan performance. *Journal of Real Estate Finance and Economics*, 9(3):263–294, 1994.

J. A. Berkovec, G. B. Canner, S. A. Gabriel, and T. H. Hannan. Discrimination, competition, and loan performance in FHA mortgage lending. *The Review of Economics and Statistics*, 80(2):241–250, 1998.

D. J. Berri and R. Simmons. Race and the evaluation of signal callers in the National Football League. *Journal of Sports Econonomics*, 10(1):23–43, 2009.

A. P. Bertero. Gender discrimination in the United States tariff schedule: Does unequal tariff treatment of mens and womens products constitute an equal protection violation?, 2011. Working Paper, http://ssrn.com.

M. Bertrand and S. Mullainathan. Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review*, 94(4):991–1013, 2004.

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

M. Bertrand, D. Chugh, and S. Mullainathan. Implicit discrimination. *American Economic Review*, 95(2):94–98, 2005.

P. J. Bickel, E. A. Hammel, and J. W. O'Connell. Sex bias in graduate admissions: Data from Berkeley. *Science*, 187 (4175):398–404, 1975.

S. E. Black and E. Brainerd. Importing equality? The impact of globalization on gender discrimination. *Industrial and Labor Relations Review*, 57(4):540–559, 2004.

S. E. Black and P. E. Strahan. The division of spoils: Rent-sharing and discrimination in a regulated industry. *American Economic Review*, 91(4):814–831, 2001.

M. L. Blackburn and T. Vermilyea. A comparison of unexplained racial disparities in bank-level and market-level models of mortgage lending. *Journal of Financial Services Research*, 29(2):125–147, 2006.

M. L. Blackburn and T. Vermilyea. The role of information externalities and scale economies in home mortgage lending decisions. *Journal of Urban Economics*, 61(1):71–85, 2007.

G. Blalock, J. L. DeVaro, S. Leventhal, and D. H. Simon. Gender bias in power relationships: Evidence from police traffic stops. Working paper, Cornell University, School of Industrial and Labor Relations, 2007. http://ssrn.com.

L. Blanchard, B. Zhao, and J. Yinger. Do lenders discriminate against minority and woman entrepreneurs? *Journal of Urban Economics*, 63(2):467–497, 2008.

D. G. Blanchflower, P. B. Levine, and D. J. Zimmerman. Discrimination in the small-business credit market. *The Review of Economics and Statistics*, 85(4):930–943, 2003.

E. Blank, P. Venkatachalam, L. McNeil, and R. Green. Racial discrimination in mortgage lending in Washington, D.C.: A mixed methods approach. *The Review of Black Political Economy*, 33(2):9–30, 2005.

R. M. Blank, M. Dabady, and C. F. Citro, editors. *Measuring Racial Discrimination - Panel on Methods for Assessing Discrimination*. National Academies Press, 2004.

A. Blinder. Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources*, 8(Fall): 436–455, 1973.

D. G. Bocian, K. S. Ernst, and W. Li. Race, ethnicity and subprime home loan pricing. *Journal of Economics and Business*, 60(1-2):110–124, 2008.

G. Bonoli and K. Hinrichs. Statistical discrimination and employers' recruitment practices for low skilled workers. Working Paper 10-2010, Reconciling Work and Welfare in Europe, 2010. http://ssrn.com.

F. L. Bóo, M. A. Rossi, and S. Urzúa. The labor market return to an attractive face: Evidence from a field experiment. Discussion Paper 6356, Institute for the Study of Labor (IZA), 2012. http://ftp.iza.org.

A. L. Booth and A. Leigh. Do employers discriminate by gender? A field experiment in female-dominated occupations. *Economics Letters*, 107(2):236–238, 2010.

A. L. Booth, A. Leigh, and E. Varganova. Does racial and ethnic discrimination vary across minority groups? Evidence from a field experiment. Discussion paper 4947, Institute for the Study of Labor (IZA), 2010. http://ftp.iza.org.

L. Bornmann, R. Mutz, and H.-D. Daniel. Gender differences in grant peer review: A meta-analysis. *Journal of Informetrics*, 1(3):226–238, 2007.

L. Bornmann, R. Mutz, and H.-D. Daniel. Latent markov modeling applied to grant peer review. *Journal of Informetrics*, 2(3):217–228, 2008.

M. Bosch, M. A. Carnero, and L. Farr. Information and discrimination in the rental housing market: Evidence from a field experiment. *Regional Science and Urban Economics*, 40(1):11–19, 2010.

T. Breda and S. T. Ly. Do professors really perpetuate the gender gap in science? evidence from a natural experiment in a French higher education institution. Discussion Papers CEEDP0138, Centre for the Economics of Education (CEE), 2012. http://cee.lse.ac.uk.

K. P. Brevoort. Credit card redlining revisited. *Review of Economics and Statistics*, 93(2):714–724, 2011.

C. L. Brown and W. G. Simpson. An analysis of alternative methodologies and interpretations of mortgage discrimination research using simulated data. *Academy of Banking Studies Journal*, 9(2):65–75, 2010.

R. Brown. *Prejudice: Its Social Psychology*. Wiley-Blackwell, 2 edition, 2010.

P. Broyles and B. Keen. Consumer discrimination in the NBA: An examination of the effect of race on the value of basketball trading cards. *The Social Science Journal*, 47(1):162–171, 2010.

G. G. Cain. The economic analysis of labor market discrimination: A survey. In O. Ashenfelter and R. Layard, editors, *Handbook of Labor Economics*, volume 1, pages 693–781. Elsevier, 1987.

T. Calders and I. Žliobaitė. Why unbiased computational processes can lead to discriminative decision procedures. In B. H. M. Custers, T. Calders, B. W. Schermer, and T. Z. Zarsky, editors, *Discrimination and Privacy in the Information Society*, volume 3 of *Studies in Applied Philosophy, Epistemology and Rational Ethics*, pages 43–57. Springer, 2013.

T. Calders and S. Verwer. Three naive bayes approaches for discrimination-free classification. *Data Mining & Knowledge Discovery*, 21(2):277–292, 2010.

P. S. Calem, K. Gillen, and S. Wachter. The neighborhood distribution of subprime mortgage lending. *Journal of Real Estate Finance and Economics*, 29(4):393–410, 2004.

A. Calvo-Armengol and M. O. Jackson. The effects of social networks on employment and inequality. *The American Economic Review*, 94(3):426–454, 2004.

A. Calvo-Armengol and M. O. Jackson. Networks in labor markets: Wage and employment dynamics and inequality. *Journal of Economic Theory*, 132(1):27–46, 2007.

R. Campbell, B. Roberts, and K. Rogers. An evaluation of lender redlining in allocation of unsecured consumer credit. *Urban Studies*, 45(5):1243–1254, 2008.

D. Card and T. Lemieux. Changing wage structure and black-white differentials among men and women: A longitudinal analysis. *American Economic Review*, 84:29–33, 1994.

D. Card and T. Lemieux. Wage dispersion, returns to skill, and black-white wage differentials. *Journal of Econometrics*, 74:319–361, 1996.

M. Carlsson and S. Eriksson. Do reported attitudes towards immigrants predict ethnic discrimination? Working Paper 2012:6, Department of Economics, Uppsala University, Sweden, 2012. http://swopec.hhs.se.

M. Carlsson and D.-O. Rooth. Evidence of ethnic discrimination in the Swedish labor market using experimental data. *Labour Economics*, 14(4):716–729, 2007.

M. A. C. Case. Disaggregating gender from sex and sexual orientation: The effeminate man in the law and feminist jurisprudence. *The Yale Law Journal*, 105(1):1–105, 1995.

M. Castillo and R. Petrie. Discrimination in the lab: Does information trump appearance? *Games and Economic Behavior*, 68(1):50–59, 2010.

K. Cavalluzzo, L. Cavalluzzo, and J. Wolken. Competition, small business financing, and discrimination: Evidence from a new survey. *Journal of Business*, 75(4):641–680, 2002.

S. J. Ceci and W. M. Williams. Understanding current causes of women's underrepresentation in science. *Proc. of the National Academy of Sciences*, 108(8):3157–3162, 2011.

J. Chan. Racial profiling and police subculture. *Canadian Journal of Criminology and Criminal Justice*, 53(1):75–78, 2011.

S. Chan, M. Gedal, V. Been, and A. F. Haughwout. The role of neighborhood characteristics in mortgage default risk: Evidence from New York City. Working Paper 4551, NYU Wagner School and Furman Center for Real Estate & Urban Policy, 2011. http://www.ssrn.com.

A. Chandra and D. O. Staiger. Identifying provider prejudice in healthcare. Working Paper 16382, National Bureau of Economic Research, 2010.

K. K. Charles and J. Guryan. Prejudice and wages: An empirical assessment of Becker's 'The economics of discrimination'. *Journal of Political Economy*, 116(5):773–809, 2008.

K. K. Charles and J. Guryan. Studying discrimination: Fundamental challenges and recent progress. *Annual Review of Economics*, 3:479–511, 2011.

G. Charness and P. Kuhn. Lab labor: What can labor economists learn from the lab? In O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics*, volume 4, pages 229–330. Elsevier, 2011.

P. Cheng, Z. Lin, and Y. Liu. Do women pay more for mortgages? *The Journal of Real Estate Finance and Economics*, 43(4):423–440, 2011.

S. Cheung. A test of employer learning in the labour market for young Australians. *Applied Economics Letters*, 17 (1):93–98, 2010.

I. Chopin and T. U. Do. *Developing Anti-Discrimination Law in Europe*. European Network of Legal Experts in Anti-Discrimination, 2010. http://ec.europa.eu.

R. Christensen. *Plane Answers to Complex Questions: The Theory of Linear Models*. Springer, New York, 3 edition, 2002.

G. P. Cipriani and A. Zago. Productivity or discrimination? beauty and the exams. *Oxford Bulletin of Economics and Statistics*, 73(3):428–447, 2011.

W. A. V. Clark. Residential preferences and neighborhood racial segregation: A test of the Schelling segregation model. *Demography*, 28(1):1–19, 1991.

J. Clarke, N. Roy, and M. Courchane. On the robustness of racial discrimination findings in mortgage lending studies. *Applied Economics*, 41(18):2279–2297, 2009.

C. Clifton. Privacy preserving data mining: How do we mine data when we aren't allowed to see it? In *Proc. of the ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD 2003), Tutorial, Washington, DC (USA)*, 2003. http://www.cs.purdue.edu.

S. Coate and G. C. Loury. Will affirmative-action policies eliminate negative stereotypes? *American Economic Review*, 83(5):1220–1240, 1993.

E. Cohen-Cole. Credit card redlining. Working Paper QAU08-1, Federal Reserve Bank of Boston, 2009. http://www.bos.frb.org.

B. J. Coleman, J. M. DuMond, and A. K. Lynch. An examination of NBA MVP voting behavior: Does race matter? *Journal of Sports Economics*, 9(6):606–627, 2008.

P.-P. Combes, B. Decreuse, M. Laouenan, and A. Trannoy. Customer discrimination and employment outcomes: Theory and evidence from the French labor market, 2011. http://hal.archives-ouvertes.fr.

Consumer Federation of America & National Credit Reporting Association. Credit score accuracy and implications for consumers, 2002. http://www.consumerfed.org.

F. Cornaglia and N. E. Feldman. Productivity, wages, and marriage: The case of Major League Baseball. Discussion Paper 5695, Institute for the Study of Labor (IZA), 2011. http://ftp.iza.org.

J. Cotton. On the decomposition of wage differentials. *The Review of Economics and Statistics*, 70(2):236–243, 1988.

Council of the E.U. Resolution 1999/C 201/01 on Women and Science, 1999. http://eur-lex.europa.eu.

M. J. Courchane. The pricing of home mortgage loans to minority borrowers: How much of the APR differential can we explain? *Journal of Real Estate Research*, 29(4):399–440, 2007.

C. J. Coyne, J. Isaacs, and J. Schwartz. Entrepreneurship and the taste for discrimination. *Journal of Evolutionary Economics*, 20(4):609–627, 2010.

F. Crosby, S. Bromley, and L. Saxe. Recent unobtrusive studies of black and white discrimination and prejudice: A literature review. *Psychological Bulletin*, 87(3):546–563, 1980.

B. H. M. Custers, T. Calders, B. W. Schermer, and T. Z. Zarsky, editors. *Discrimination and Privacy in the Information Society*, volume 3 of *Studies in Applied Philosophy, Epistemology and Rational Ethics*. Springer, 2013.

D. Darai and S. Grätz. Facing a dilemma: Cooperative behavior and beauty. Working paper 82, University of Zurich, Department of Economics, 2012. http://www.merlin.uzh.ch.

D. L. Dickinson and R. L. Oaxaca. Statistical discrimination in labor markets: An experimental analysis. *Southern Economic Journal*, 76(1):16–31, 2009.

J. Dietrich. Does multiple jeopardy exist in mortgage markets? Economics Working Paper 2009-3, Office of the Comptroller of the Currency, 2009. http://www.occ.treas.gov.

J. Dietrich and H. Johannsson. Searching for age and gender discrimination in mortgage lending. Economics Working Paper 2005-2, Office of the Comptroller of the Currency, 2005. http://www.occ.treas.gov.

L. M. Dilks, S. R. Thye, and P. A. Taylor. Socializing economic theories of discrimination: Lessons from Survivor. *Social Science Research*, 39(6):1164–1180, 2010.

L. Ding, J. Ratcliffe, M. Stegman, and R. Quercia. Neighborhood patterns of high-cost lending: The case of Atlanta. *Journal of Affordable Housing*, 17(3):194–211, 2008.

A. J. Dobson and A. G. Barnett. *Introduction to Generalized Linear Models*. Chapman and Hall/CRC, 3 edition, 2008.

J. L. Doleac and L. C. D. Stein. The visible hand: Race and online market outcomes. Discussion Paper 09-015, Stanford Institute for Economic Policy Research, 2010. http://www.stanford.edu.

J. J. Donohue. The law and economics of antidiscrimination law. In A. M. Polinsky and S. Shavell, editors, *Handbook of Law and Economics*, pages 1387–1472. Elsevier, 2007.

N. Drydakis. Sexual orientation discrimination in the labour market. *Labour Economics*, 16(4):364–372, 2009.

C. Dwork, M. Hardt, T. Pitassi, O. Reingold, and R. S. Zemel. Fairness through awareness. In S. Goldwasser, editor, *Proc. of the 3rd Int. Conf. on Innovations in Theoretical Computer Science (ITCS 2012)*, pages 214–226. ACM, 2012.

G. A. Dymski. Discrimination in the credit and housing markets: Findings and challenges. In W. M. Rodgers, editor, *Handbook on the Economics of Discrimination*, pages 215–259. Edward Elgar Publishing, 3 edition, 2006.

G. A. Dymski, J. Hernandez, and L. Mohanty. Race, power, and the subprime/foreclosure crisis: A mesoanalysis. Economics Working Paper 669, Levy Economics Institute of Bard College, 2011. http://ssrn.com.

F. Echenique and R. G. Fryer. A measure of segregation based on social interactions. *The Quarterly Journal of Economics*, 122(2):441–485, 2007.

W. Edelberg. Racial dispersion in consumer credit interest rates. Finance and Economics Discussion Series 2007–2028, Board of Governors of the Federal Reserve System (U.S.), 2007.

E. Ellis. *EU Anti-Discrimination Law*. Oxford University Press, 2005.

ENAR. European network against racism, Fact sheet 33: Multiple discrimination, 2007. http://www.enar-eu.org.

ENAR. European network against racism, Fact sheet 35: Positive actions, 2008. http://www.enar-eu.org.

R. S. Engel. A critique of the "outcome test" in racial profiling research. *Justice Quarterly*, 25(1):1–36, 2008.

R. S. Engel and J. M. Calnon. Comparing benchmark methodologies for police-citizen contacts: Traffic stop data collection for the Pennsylvania State Police. *Police Quarterly*, 7(1):97–125, 2004.

R. S. Engel, J. M. Calnon, and T. J. Bernard. Theory and racial profiling: Shortcomings and future directions in research. *Justice Quarterly*, 19(2):1–36, 2002.

Equal Employment Opportunity Commission. Uniform guidelines on employee selection procedure, 1978. 43 FR 38295, http://www.gpo.gov.

S. Eriksson and J. Lagerstrøm. Detecting discrimination in the hiring process: Evidence from an Internet-based search channel. *Empirical Economics*, 43(2):537–563, 2012.

E.U. Agency for Fundamental Rights. *Handbook on European non-discrimination law*. European Court of Human Rights, 2011. http://fra.europa.eu.

European Commission. Tackling multiple discrimination : Practices, policies and laws, 2007. Directorate General for Employment, Social Affairs and Equal Opportunities, Unit G.4, http://ec.europa.eu.

European Commission. The gender challenge in research funding: Assessing the European national scenes, 2009. Directorate General for Research, Science, Economy and Society, Unit L.4, http://ec.europa.eu.

European Commission. Compilation of case law on the equality of treatment between women and men and on non-discrimination in the european union, 2010. Directorate General for Employment, Social Affairs and Equal Opportunities, Unit G.2, http://ec.europa.eu.

European Commission. Meta-analysis of gender and science research, 2012. Directorate General for Research and Innovation, Sector B6.2, http://www.genderandscience.org.

European Union Legislation. (a) European Convention on Human Rights, 1950; (b) Racial Equality Directive, 2000; (c) Employment Equality Directive, 2000; (d) Gender Goods and Services Directive, 2004; (e) Gender Employment Directive, 2006; (f) Equal Treatment Directive (proposal), 2008, 2012. http://eur-lex.europa.eu.

F. Ezeala-Harrison, G. Glover, and J. Shaw-Jackson. Housing loan patterns toward minority borrowers in Mississippi: Analysis of some micro data evidence of redlining. *The Review of Black Political Economy*, 35(1):43–54, 2008.

Y. Fadlon. Statistical discrimination and the implications of employer-employee racial matches, 2011. Draft manuscript, https://my.vanderbilt.edu/yarivfadlon.

R. W. Fairlie. An extension of the Blinder–Oaxaca decomposition technique to logit and probit models. *Journal of Economic and Social Measurement*, 30:305–316, 2005.

A. Falk and C. Zehnder. Discrimination and in-group favoritism in a citywide trust experiment. Working Paper iewwp318, Institute for Empirical Research in Economics, University of Zurich, 2007. http://www.iew.uzh.ch.

A. Falk, G. Walkowitz, and W. Wirth. Do ex-offenders face discrimination in the labor market? Because of expected inferior reciprocity?, 2008. Unpublished manuscript, http://www.uni-graz.at.

H. Fang and A. Moro. Theories of statistical discrimination and affirmative action: A survey. In J. Benhabib, M. O. Jackson, and A. Bisin, editors, *Handbook of Social Economics*, volume 1A, pages 134–200. Elsevier, 2010.

C. P. Farrington and G. Manning. Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. *Statistics in Medicine*, 9(12):1447–1454, 1990.

Federal Trade Commission. Credit-based insurance scores: Impacts on consumers of automobile insurance, 2007. http://ftc.gov.

N. Feltovich and C. Papageorgiou. An experimental study of statistical discrimination by employers. *Southern Economic Journal*, 70(4):837–849, 2004.

C. Fershtman and U. Gneezy. Discrimination in a segmented society: An experimental approach. *The Quarterly Journal of Economics*, 116(1):351–377, 2001.

S. E. Fienberg, S. H. Krislov, and M. L. Straf. Understanding and evaluating statistical evidence in litigation. *Jurimetrics Journal of Law, Science, and Technology*, 36(1):1–32, 1995.

M. O. Finkelstein and B. Levin, editors. *Statistics for Lawyers*. Springer-Verlag, 2 edition, 2001.

M. Fix and M. Tuner, editors. *A National Report Card on Discrimination in America: The Role of Testing*. The Urban Institute, 1999.

L. Flabbi. Prejudice and gender differentials in the US labor market in the last twenty years. *Journal of Econometrics*, 156(1):190–200, 2010.

J. L. Fleiss, B. Levin, and M. C. Paik. *Statistical Methods for Rates and Proportions*. Wiley, 3 edition, 2003.

M. Foley and F. H. Smith. Consumer discrimination in professional sports: New evidence from Major League Baseball. *Applied Economics Letters*, 14(13):951–955, 2007.

N. Fortin, T. Lemieux, and S. Firpo. Decomposition methods in economics. Working Paper 16045, National Bureau of Economic Research, 2010. http://www.nber.org.

E. Fortowsky and M. LaCour-Little. Credit scoring and disparate impact. Working paper, Wells Fargo Home Mortagage, 2001. http://fic.wharton.upenn.edu/fic.

S. Fraser. Is there ethnic discrimination in the UK market for small business credit? *International Small Business Journal*, 27(5):583–607, 2009.

B. Frick. The football players' labor market: Empirical evidence from the major European leagues. *Scottish Journal of Political Economy*, 54(3):422–446, 2007.

B. Frick and C. Deutscher. Salary determination in the German "Bundesliga": A panel study. IASE Conference Paper 0811, International Association of Sports Economists, 2009.

S. Friedman, G. D. Squires, and C. Galvan. Cybersegregation in Boston and Dallas: Is Neil a more desirable tenant than Tyrone or Jorge?, 2010. Presented at the Population Association of America 2010 Annual Meeting, http://paa2010.princeton.edu.

P. Frijters, M. A. Shields, N. Theodoropoulos, and S. Wheatley Price. Testing for employee discrimination using matched employer-employee data: Theory and evidence. Discussion paper 807, Institute for the Study of Labor (IZA), 2003. http://ftp.iza.org.

R. G. Fryer, D. Pager, and J. L. Spenkuch. Racial disparities in job finding and offered wages. Working Paper 17462, National Bureau of Economic Research, 2011. http://www.nber.org.

S. L. Gabbidon, R. Craig, N. Okafo, L. N. Marzette, and S. A. Peterson. The consumer racial profiling experiences of Black students at historically Black colleges and universities: An exploratory study. *Journal of Criminal Justice*, 36(4):354–361, 2008.

B. Gao and B. Berendt. Visual data mining for higher-level patterns: Discrimination-aware data mining and beyond. In *Proc. of Belgian Dutch Conf. on Machine Learning (Benelearn 2011), The Hague (The Netherlands)*, pages 45–52, 2011.

J. Gardner. Deterrence externalities and racial bias in law enforcement. Working Paper 88, Carnegie Mellon University, 2009. http://www.heinz.cmu.edu.

J. L. Gastwirth. Statistical methods for analyzing claims of employment discrimination. *Industrial and Labor Relations Review*, 38(1):75–86, 1984.

J. L. Gastwirth. Statistical reasoning in the legal setting. *The American Statistician*, 46(1):55–69, 1992a.

J. L. Gastwirth. Methods for assessing the sensitivity of statistical comparisons used in Title VII cases to omitted variables. *Jurimetrics Journal*, 33:19–34, 1992b.

J. L. Gastwirth, A. M. Krieger, and P. R. Rosenbaum. Dual and simultaneous sensitivity analysis for matched pairs. *Biometrika*, 85(4):907–920, 1998.

R. Gellert, K. D. Vries, P. D. Hert, and S. Gutwirth. A comparative analysis of anti-discrimination and data protection legislations. In B. H. M. Custers, T. Calders, B. W. Schermer, and T. Z. Zarsky, editors, *Discrimination and Privacy*

in the Information Society, volume 3 of *Studies in Applied Philosophy, Epistemology and Rational Ethics*, pages 61–89. Springer, 2013.

A. Gelman, J. Fagan, and A. Kiss. An analysis of the New York City Police Departments "stop-and-frisk" policy in the context of claims of racial bias. *Journal of the American Statistical Association*, 102(479):813–823, 2007.

M. Ghosh, M.-H. Chen, A. Ghosh, and A. Agresti. Hierarchical Bayesian analysis of binary matched data. *Statistica Sinica*, 10:647–657, 2000.

J. Goddard and J. O. S. Wilson. Racial discrimination in English professional football: Evidence from an empirical analysis of players' career progression. *Cambridge Journal of Economics*, 33(2):295–316, 2009.

C. Goenner. Discrimination and mortgage lending in Boston: The effects of model uncertainty. *Journal of Real Estate Finance and Economics*, 40(3):260–285, 2010.

M. S. Goldberg. Discrimination, nepotism, and long-run wage differentials. *The Quarterly Journal of Economics*, 97 (2):307–19, 1982.

P. K. Goldberg. Dealer price discrimination in new car purchases: Evidence from the consumer expenditure survey. *Journal of Political Economy*, 104(3):622–654, 1996.

C. Goldin and C. Rouse. Orchestrating impartiality: The impact of "blind" auditions on female musicians. *American Economic Review*, 90(4):715–741, 2000.

H. Goldstein. *Multilevel Statistical Models*. Wiley, 4 edition, 2011.

A. Green, D. Carney, D. Pallin, L. Ngo, K. Raymond, L. Iezzoni, and M. Banaji. Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of General Internal Medicine*, 22: 1231–1238, 2007.

W. H. Greene. *Econometric Analysis*. Prentice-Hall, 7 edition, 2008.

A. G. Greenwald and L. H. Krieger. Implicit bias: Scientific foundations. *California Law Review*, 94(4):945–967, 2006.

J. T. Grogger. Speech patterns and racial wage inequality. *Journal of Human Resources*, 46(1):1–25, 2011.

J. T. Grogger and G. Ridgeway. Testing for racial profiling in traffic stops from behind a veil of darkness. *Journal of the American Statistical Association*, 101(475):878–887, 2006.

J. Grytten, I. Skau, and R. Sørensen. Do expert patients get better treatment than others? Agency discrimination and statistical discrimination in obstetrics. *Journal of Health Economics*, 30(1):163–180, 2011.

W. Güth, M. G. Kocher, and V. Popova. Co-employment of permanently and temporarily employed agents. Jena Economic Research Paper 2010–016, Friedrich-Schiller-University Jena, Max-Planck-Institute of Economics, 2010. http://www.econ.mpg.de.

S. Hajian and J. Domingo-Ferrer. A methodology for direct and indirect discrimination prevention in data mining. *IEEE Transactions on Knowledge and Data Engineering*, page to appear, 2012.

S. Hajian, A. Monreale, D. Pedreschi, J. Domingo-Ferrer, and F. Giannotti. Injecting discrimination and privacy awareness into pattern discovery. In J. Vreeken, C. Ling, M. J. Zaki, A. Siebes, J. X. Yu, B. Goethals, G. I. Webb, and X. Wu, editors, *Proc. of the IEEE ICDM 2012 Int. Workshop on Discrimination and Privacy-Aware Data Mining (DPADM)*, pages 360–367. IEEE Computer Society, 2012.

D. Hamermesh and J. Biddle. Beauty and the labor market. *The American Economic Review*, 84(5):1174–1194, 1994.

S. Han. Creditor learning and discrimination in lending. *Journal of Financial Services Research*, 40(1-2):1–27, 2011.

A. Hanson and Z. Hawley. Do landlords discriminate in the rental housing market? Evidence from an Internet field experiment in U.S. cities. *Urban Economics*, 70(2-3):99–114, 2011.

F. A. Hanssen and T. Andersen. Has discrimination lessened over time? a test using baseball's all-star vote. *Economic Inquiry*, 37(2):326–52, 1999.

J. Hardin and J. Hilbe. *Generalized Linear Models and Extensions*. Stata Press, 2 edition, 2007.

T. Harford. *The Logic of Life*. The Random House Publishing Group, 2008.

G. W. Harrison and J. A. List. Field experiments. *American Economic Literature*, 42(4):1009–1055, 2004.

J. Heckman. Detecting discrimination. *The Journal of Economic Perspectives*, 12(2):101–116, 1998.

J. Heckman and P. Siegelman. The Urban Institute audit studies: Their methods and findings. In M. Fix and R. Struyk, editors, *Clear and convincing evidence: Measures of discrimination in America*, pages 187–248. The Urban Institute Press, 1993.

J. K. Hellerstein, D. Neumark, and K. R. Troske. Market forces and sex discrimination. *Journal of Human Resources*, 37(2):353–380, 2002.

J. Hernandez. Redlining revisited: Mortgage lending patterns in sacramento 1930-2004. *International Journal of Urban and Regional Research*, 33(2):291–313, 2009.

P. L. Hersch. Customer discrimination against black Major League Baseball pitchers reconsidered. *Applied Economics Letters*, 17(2):205–208, 2010.

M. Hildebrandt and S. Gutwirth, editors. *Profiling the European Citizen: Cross-Disciplinary Perspectives*. Springer, 2008.

A. E. Hillier. Spatial analysis of historical redlining: A methodological explanation. *Journal of Housing Research*, 14(1):137–168, 2003.

B. T. Hinnerich, E. Höglin, and M. Johannesson. Are boys discriminated in Swedish high schools? *Economics of Education Review*, 30(4):682–690, 2011.

J. Hoeting, D. Madigan, A. Raftery, and C. Volinsky. Bayesian model averaging: A tutorial. *Statistical Science*, 14 (4):382–401, 1999.

R. L. Hogler, C. Henle, and C. Bemus. Internet recruiting and employment discrimination: A legal perspective. *Human Resource Management Review*, 8(2):149–164, 1998.

S. R. Holloway and E. K. Wyly. "The color of money" expanded: Geographically contingent mortgage lending in atlanta. *Journal of Housing Research*, 12(1):55–90, 2001.

P. Holmes. New evidence of salary discrimination in Major League Baseball. *Labour Economics*, 18(3):320–331, 2011.

H. J. Holzer and K. R. Ihlanfeldt. Customer discrimination and employment outcomes for minority workers. *The Quarterly Journal of Economics*, 113(3):835–867, 1998.

H. J. Holzer and D. Neumark. *The Economics of Affirmative Action*. Cheltenham: Edward Elgar, 2004.

H. J. Holzer and D. Neumark. Affirmative action: What do we know? *Journal of Policy Analysis and Management*, 25(2):463–490, 2006.

H. J. Holzer, S. Raphael, and M. A. Stoll. Perceived criminality, criminal background checks, and the racial hiring practices of employers. *Journal of Law & Economics*, 49(2):451–480, 2006.

L. Hu and C. Taber. Displacement, asymmetric information and heterogeneous human capital. *Journal of Labor Economics*, 29(1):113–152, 2010.

W. C. Hunter and M. B. Walker. The cultural affinity hypothesis and mortgage lending decisions. *The Journal of Real Estate Finance and Economics*, 13(1):57–70, 1996.

Y. M. Ioannides. Social interactions (empirics). In S. N. Durlauf and L. E. Blume, editors, *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, 2008.

Y. M. Ioannides and L. D. Loury. Job information networks, neighborhood effects, and inequality. *Journal of Economic Literature*, 42(4):1056–1093, 2004.

D. R. James and K. E. Tauber. Measures of segregation. *Sociological Methodology*, 13:1–32, 1985.

B. Jann. The Blinder–Oaxaca decomposition for linear regression models. *Stata Journal*, 8(4):453–479, 2008.

U. W. Jayasinghe, H. W. Marsh, and N. W. Bond. A multilevel cross-classified modeling approach to peer-review of grant proposals. *Journal of the Royal Statistical Society*, 166(3):279–300, 2003.

R. T. Jewell, R. W. Brown, and S. E. Miles. Measuring discrimination in Major League Baseball: Evidence from the baseball hall of fame. *Applied Economics*, 34(2):167–77, 2002.

F. Jobard and R. Lévy. Racial profiling: The Parisian police experience. *Canadian Journal of Criminology and Criminal Justice*, 53(1):87–93, 2011.

M. K. Jones. Disability and the labour market: A review of the empirical evidence. *Journal of Economic Studies*, 35 (5):405–424, 2008.

C. Juhn, K. M. Murphy, and B. Pierce. Accounting for the slowdown in black-white wage convergence. In M. H. Kosters, editor, *Workers and Their Wages: Changing Patterns in the United States*, pages 107–143. AEI Press, 1991.

L. Kaas and C. Manger. Ethnic discrimination in Germany's labour market: A field experiment. *German Economic Review*, 13(1):1–20, 2012.

L. M. Kahn. Customer discrimination and affirmative action. *Economic Inquiry*, 29(3):555–71, 1991a.

L. M. Kahn. Discrimination in professional sports: A survey of the literature. *Industrial and Labor Relations Review*, 44(3):395–418, 1991b.

L. M. Kahn. The sports business as a labor market laboratory. *Journal of Economic Perspectives*, 14(3):75–94, 2000.

L. M. Kahn. The economics of discrimination: Evidence from basketball. Discussion Paper 3987, Institute for the Study of Labor (IZA), 2009. http://ftp.iza.org.

F. Kamiran and T. Calders. Data preprocessing techniques for classification without discrimination. *Knowledge and Information Systems*, 33:1–33, 2012.

F. Kamiran, T. Calders, and M. Pechenizkiy. Discrimination aware decision tree learning. In G. I. Webb, B. Liu, C. Zhang, D. Gunopulos, and X. Wu, editors, *Proc. of the IEEE Int. Conf. on Data Mining (ICDM 2010)*, pages 869–874. IEEE Computer Society, 2010.

F. Kamiran, A. Karim, and X. Zhang. Decision theory for discrimination-aware classification. In M. J. Zaki, A. Siebes, J. X. Yu, B. Goethals, G. I. Webb, and X. Wu, editors, *Proc. of the IEEE Int. Conf. on Data Mining (ICDM 2012)*, pages 924–929, 2012.

T. Kamishima, S. Akaho, and J. Sakuma. Fairness-aware classifier with prejudice remover regularizer. In *Proc. of the Eur. Conf. on Machine Learning and on Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD 2012)*, volume 7524 of *LNCS*, pages 35–50. Springer, 2012.

J. Kang and K. Lane. Seeing through colorblindness: Implicit bias and the law. *UCLA Law Review*, 58(2):465–520, 2010.

R. M. Kanter. Some effects of proportions on group life: Skewed sex ratios and responses to token women. *American Journal of Sociology*, 82(5):965–990, 1977.

C. Kapsalis. A new measure of wage discrimination. *Economics Letters*, 9(3):287–293, 1982.

C. Kapsalis. Wage discrimination: The case for reverse regression. Technical Report 27331, University Library of Munich, Germany, 2010. http://mpra.ub.uni-muenchen.de.

D. H. Kaye and M. Aickin, editors. *Statistical Methods in Discrimination Litigation*. Marcel Dekker, Inc., 1992.

D. H. Kaye and D. A. Freedman. Reference guide on statistics. In *Reference Manual on Scientific Evidence*, pages 83–178. Federal Judicial Center, 3 edition, 2011.

M. P. Keane. A computationally practical simulation estimator for panel data. *Econometrica*, 62(1):95–116, 1994.

R. Kessler, N. Béchet, J. M. T. Moreno, M. Roche, and M. El-Bèze. Job offer management: How improve the ranking of candidates. In J. Rauch, Z. W. Ras, P. Berka, and T. Elomaa, editors, *Proc. of the Int. Symposium on Foundations of Intelligent Systems (ISMIS 2009)*, volume 5722 of *LNCS*, pages 431–441. Springer, 2009.

M. R. Killingsworth. Analyzing employment discrimination: From the seminar room to the courtroom. *American Economic Review*, 83(2):67–72, 1993.

K.-H. Kim. Favoritism and reverse discrimination. *European Economic Review*, 51(1):101–123, 2007.

Y. C. Kim. Lifetime network externality and the dynamics of group inequality. Technical Report 18767, University Library of Munich, Germany, 2009. http://mpra.ub.uni-muenchen.de.

J. Knowles, N. Persico, and P. Todd. Racial bias in motor vehicle searches: Theory and evidence. *Journal of Political Economy*, 109(1):203–229, 2001.

R. Koenker. *Quantile Regression*. Cambridge University Press, Cambridge, 2005.

B. Körffer and M. Meints. Profiling of customers and consumers – customer loyalty programs and scoring practices. In M. Hildebrandt and S. Gutwirth, editors, *Profiling the European Citizen: Cross-Disciplinary Perspectives*, chapter 11, pages 201–215. Springer, 2008.

M. Kuebler. Lending in the modern era: Does racial composition of neighborhoods matter when individuals seek home financing? A pilot study in New England. *City & Community*, 11(1):31–50, 2012.

P. J. Kuhn and K. Shen. Gender discrimination in job ads: Theory and evidence. Discussion Paper 5195, Institute for the Study of Labor (IZA), 2010. http://ftp.iza.org.

A. Kunze. Gender wage gap studies: Consistency and decomposition. *Empirical Economics*, 35(1):63–76, 2008.

M. LaCour-Little. Application of reverse regression to Boston Federal Reserve data refutes claims of discrimination. *Journal of Real Estate Research*, 11(1):1–12, 1996.

M. LaCour-Little. Discrimination in mortgage lending: A critical review of the literature. *Journal of Real Estate Literature*, 7(1):15–49, 1999.

J. N. Lahey and R. A. Beasley. Computerizing audit studies. *Journal of Economic Behavior & Organization*, 70(3): 508–514, 2009.

K. Lang and J.-Y. K. Lehmann. Racial discrimination in the labor market: Theory and empirics. Working Paper 17450, National Bureau of Economic Research, 2011. http://www.nber.org.

K. Lang and M. Manove. Education and labor market discrimination. *American Economic Review*, 101(4):1467–1496, 2011.

W. W. Lang and L. I. Nakamura. A model of redlining. *Journal of Urban Economics*, 33(2):223–234, 1993.

F. Lange. The speed of employer learning. *Journal of Labor Economics*, 25(1):1–35, 2007.

J. Lee. American Idol: Evidence on same-race preferences. *The Berkeley Electronic Journal of Economic Analysis & Policy*, 9(1):Article 28, 2009.

J. S. Leonard, D. I. Levine, and L. Giuliano. Customer discrimination. *The Review of Economics and Statistics*, 92 (3):670–678, 2010.

N. Leopold and M. Meints. Profiling in employment situations (fraud). In M. Hildebrandt and S. Gutwirth, editors, *Profiling the European Citizen: Cross-Disciplinary Perspectives*, pages 236–257. Springer, 2008.

N. Lerner. *Group Rights and Discrimination in International Law*. Martinus Nijhoff Publishers, 2 edition, 2003.

R. Lerner and A. K. Nagai. Reverse discrimination by the numbers. *Journal Academic Questions*, 13(3):71–84, 2000.

J. P. LeSage and R. K. Pace, editors. *Introduction to spatial econometrics*. Chapman and Hall/CRC, 2009.

H. M. Leung and L. L. Kupper. Comparisons of confidence intervals for attributable risk. *Biometrics*, 37(2):293–302, 1981.

R. Levine, A. Levkov, and Y. Rubinstein. Racial discrimination and competition. Working Paper 14273, National Bureau of Economic Research, 2008. http://www.nber.org.

S. D. Levitt. Testing theories of discrimination: Evidence from Weakest Link. *Journal of Law & Economics*, 47(2): 431–452, 2004.

S. D. Levitt and J. A. List. What do laboratory experiments measuring social preferences reveal about the real world? *Journal of Economic Perspectives*, 21(2):153–174, 2007.

T. J. Ley and B. H. Hamilton. The gender gap in NIH grant applications. *Science*, 322(5907):1472–1474, 2008.

J. Li and Q. Yang. Strong compound-risk factors: Efficient discovery through emerging patterns and contrast sets. *IEEE Trans. on Information Technology in Biomedicine*, 11(5):544–552, 2007.

R. P. Lieli and H. White. The construction of empirical credit scoring rules based on maximization principles. *Journal of Econometrics*, 157(1):110–119, 2010.

J. A. List. The nature and extent of discrimination in the marketplace: Evidence from the field. *The Quarterly Journal of Economics*, 119(1):49–89, 2004.

J. A. List. Friend or foe? A natural experiment of the prisoner's dilemma. *Review of Economics and Statistics*, 88(3): 463–471, 2006.

S. D. Longhofer and S. R. Peters. Why is mortgage discrimination illegal? A fresh look at the mortgage discrimination debate. *Regulation, Cato Institute*, 22(4):28–36, 1999.

B. T. Luong. *Generalized discrimination discovery on semi-structured data supported by ontology*. PhD thesis, IMT Institute for Advanced Studies. Lucca, Italy, 2011.

B. T. Luong, S. Ruggieri, and F. Turini. k-NN as an implementation of situation testing for discrimination discovery and prevention. In C. Apté, J. Ghosh, and P. Smyth, editors, *Proc. of the ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD 2011)*, pages 502–510. ACM, 2011.

T. Makkonen. Compound and intersectional discrimination: Bringing the experiences of the most marginalized to the fore, 2002. Unpublished manuscript, Institute for Human Rights, Abo Alademi University.

T. Makkonen. *Measuring Discrimination: Data Collection and the EU Equality Law*. European Network of Legal Experts in Anti-Discrimination, 2006. http://www.migpolgroup.com.

T. Makkonen. *European handbook on equality data*. European Commission, 2007. Directorate-General for Employment, Social Affairs and Equal Opportunities, Unit G.4, http://bookshop.europa.eu.

H. W. Marsh. Gender effects in the peer reviews of grant proposals: A comprehensive meta-analysis comparing traditional and multilevel approaches. *Review of Educational Research*, 79(3):1290–1326, 2009.

H. W. Marsh, U. W. Jayasinghe, and N. W. Bond. Improving the peer-review process for grant applications: Reliability, validity, bias, and generalizability. *American Psychologist*, 63(3):160–168, 2008.

R. Marshall. The economics of racial discrimination: A survey. *Journal of Economic Literature*, 12(3):849–871, 1974.

D. S. Massey and G. Lundy. Use of black english and racial discrimination in housing markets. *Urban Affairs Review*, 36(4):452–469, 2001.

G. P. McCabe. Regression analysis in discrimination cases. In D. H. Kaye and M. Aickin, editors, *Statistical Methods in Discrimination Litigation*, pages 69–84. Marcel Dekker, Inc., 1992.

P. McCullagh and J. A. Nelder. *Generalized Linear Models*. Chapman and Hall, 2 edition, 1989.

J. McGarrity, H. Palmer, and M. Poitras. Consumer racial discrimination: A reassessment of the market for baseball cards. *Journal of Labor Research*, 20(2):247–258, 1999.

F. McGinnity, J. Nelson, P. Lunn, and E. Quinn. Discrimination in recruitment - evidence from a field experiment. Report, The Equality Authority and The Economic and Social Research Institute, 2009. http://www.esri.ie.

T. G. Mcguire, J. Z. Ayanian, D. E. Ford, R. E. M. Henke, K. M. Rost, and A. M. Zaslavsky. Testing for statistical discrimination by race/ethnicity in panel data for depression treatment in primary care. *Health Research and Educational Trust*, 43(2):531–551, 2008.

M. Mobius and T. Rosenblat. Why beauty matters. *American Economic Review*, 96:222–235, 2006.

J. D. Montgomery. Social networks and labor-market outcomes: Toward an economic analysis. *American Economic Review*, 81(5):1407–1418, 1991.

J. D. Montgomery. Weak ties, employment, and inequality: An equilibrium analysis. *American Journal of Sociology*, 99(5):1212–1236, 1994.

M. Moreno, H. Ñopo, J. Saavedra, and M. Torero. Gender and racial discrimination in hiring: A pseudo audit study for three selected occupations in metropolitan Lima. Working Paper 321, Econometric Society, 2004. http://econpapers.repec.org.

A. Moro and P. Norman. A general equilibrium model of statistical discrimination. *J. Econ. Theory*, 114(1):1–30, 2004.

A. H. Munnell, G. M. B. Tootell, L. E. Browne, and J. McEneaney. Mortgage lending in Boston: Interpreting HMDA data. *American Economic Review*, 86(1):25–53, 1996.

A. Muravyev, O. Talavera, and D. Schfer. Entrepreneurs' gender and financial constraints: Evidence from international data. *Journal of Comparative Economics*, 37(2):270–286, 2009.

R. Mutz, L. Bornmann, and H.-D. Daniel. Does gender matter in grant peer review? An empirical investigation using the example of the Austrian Science Fund. *Journal of Psychology*, 220:121–129, 2012.

C. K. Myers, M. Bellows, H. Fakhoury, D. Hale, A. Hall, and K. Ofman. Ladies first? A field study of discrimination in Coffee Shops. *Applied Economics*, 42(14):1761–1769, 2010.

C. Nardinelli and C. Simon. Customer racial discrimination in the market for memorabilia: The case of baseball. *The Quarterly Journal of Economics*, 105(3):575–95, 1990.

D. A. Neal and W. R. Johnson. The role of premarket factors in black-white wage differences. *Journal of Political Economy*, 104(5):869–895, 1996.

D. Neumark. Employers discriminatory behavior and the estimation of wage discrimination. *Journal of Human Resources*, 23(3):279–295, 1988.

D. Neumark. Detecting discrimination in audit and correspondence studies. Working Paper 16448, National Bureau of Economic Research, 2010. http://www.nber.org.

E. Neumayer and I. de Soysa. Globalisation, women's economic rights and forced labour. *The World Economy*, 30 (10):1510–1535, 2007.

R. G. Newcombe. Interval estimation for the difference between independent proportions: comparison of eleven methods. *Statistics in Medicine*, 17(8):873–89, 1998.

D. M. Newman. *Sociology: Exploring the Architecture of Everyday Life*. Pine Forge Press, 7 edition, 2008.

J. M. Nunley, M. F. Owens, and R. S. Howard. The effects of competition and information on racial discrimination: Evidence from a field experiment. *Journal of Economic Behavior & Organization*, 80(3):670–679, 2011.

R. L. Oaxaca. Male-female wage differential in urban labour markets. *International Economic Review*, 14(3):693–705, 1973.

R. L. Oaxaca and M. R. Ransom. On discrimination and the decomposition of wage differentials. *Journal of Econometrics*, 61(1):5–21, 1994.

R. Oostendorp. Globalization and the gender wage gap. *World Bank Economic Review*, 23(1):141–161, 2009.

R. L. Paetzold, S. L. Willborn, and D. C. Baldus. *The Statistics of Discrimination: Using Statistical Evidence in Discrimination Cases*. Shepard's/McGraw-Hill, 1994.

D. Pager. The use of field experiments for studies of employment discrimination: Contributions, critiques, and directions for the future. *The ANNALS of the American Academy of Political and Social Science*, 609(1):104–133, 2007.

D. Pager and L. Quillian. Walking the talk? What employers say versus what they do. *American Sociological Review*, 70(3):355–380, 2005.

D. Pager, B. Western, and B. Bonikowski. Discrimination in a low-wage labor market: A field experiment. Discussion Paper 4469, Institute for the Study of Labor (IZA), 2009. http://ftp.iza.org.

M. Palmer and R. King. Has salary discrimination really disappeared from Major League Baseball? *Eastern Economic Journal*, 32(2):285–297, 2006.

M. Parrett. Customer discrimination in restaurants: Dining frequency matters. *Journal of Labor Research*, 32(2): 87–112, 2011.

E. Patacchini, G. Ragusa, and Y. Zenou. Unexplored dimensions of discrimination in Europe: Religion, homosexuality and physical appearance. The 14th European Conf. of the Fondazione R. DeBenedetti (fRDB), 2012. http://www.frdb.org.

J. Pearl. Causal inference in statistics: An overview. *Statistics Surveys*, 3:96–146, 2009a.

J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York, USA, 2 edition, 2009b.

D. Pedreschi, S. Ruggieri, and F. Turini. Discrimination-aware data mining. In Y. Li, B. Liu, and S. Sarawagi, editors, *Proc. of the ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD 2008)*, pages 560–568. ACM, 2008.

D. Pedreschi, S. Ruggieri, and F. Turini. Measuring discrimination in socially-sensitive decision records. In *Proc. of the SIAM Int. Conf. on Data Mining (SDM 2009)*, pages 581–592. SIAM, 2009.

D. Pedreschi, S. Ruggieri, and F. Turini. A study of top-k measures for discrimination discovery. In S. Ossowski and P. Lecca, editors, *Proc. of ACM Int. Symposium on Applied Computing (SAC 2012)*, pages 126–131. ACM, 2012.

N. Persico and P. Todd. Passenger profiling, imperfect screening, and airport security. *American Economic Review*, 95(2):127–131, 2005.

E. S. Phelps. The statistical theory of racism and sexism. *American Economic Review*, 62(4):659–661, 1972.

F. J. Phillips-Patrick and C. V. Rossi. Statistical evidence of mortgage redlining? a cautionary tale. *Journal of Real Estate Research*, 11(1):13–24, 1996.

J. M. Pickerill, C. Mosher, and T. Pratt. Search and seizure, racial profiling, and traffic stops: A disparate impact framework. *Law & Policy*, 31(1):1–30, 2009.

M. J. Piette and P. F. White. Approaches for dealing with small sample sizes in employment discrimination litigation. *Journal of Forensic Economics*, 12(1):43–56, 1999.

J. C. Pinkston. A test of screening discrimination with employer learning. *Industrial and Labor Relations Review*, 59 (2):267–284, 2006.

D. G. Pope and J. R. Sydnor. Implicit statistical discrimination in predictive models. Working Paper 2007-09-11, Risk Management and Decision Processes Center, The Wharton School of the University of Pennsylvania, 2007. http://opim.wharton.upenn.edu.

D. G. Pope and J. R. Sydnor. Whats in a picture? Evidence of discrimination from Prosper.com. *Journal of Human Resources*, 46(1):53–92, 2011.

E. Primm, N. L. Piquero, A. R. Piquero, and R. M. Regoli. Investigating customer racial discrimination in the secondary baseball card market. *Sociological Inquiry*, 81(1):110–132, 2011.

L. Quillian. New approaches to understanding racial prejudice and discrimination. *Annual Review of Sociology*, 32 (1):299–328, 2006.

S. M. Quintanar. Man vs. machine: An investigation of speeding ticket disparities based on gender and race. Departmental Working Paper 2009-16, Department of Economics, Louisiana State University, 2009. http://bus.lsu.edu.

M. B. Rachlis and A. M. J. Yezer. Serious flaws in statistical tests for discrimination in mortgage markets. *Journal of Housing Research*, 2(4):315–336, 1993.

J. Racine and P. Rilstone. The reverse regression problem: Statistical paradox or artefact of misspecification? *The Canadian Journal of Economics*, 28(3):502–531, 1995.

RAND. Is there gender bias in federal grant programs?, 2005. RAND Infrastructure, Safety, and Environment Brief RB-9147-NSF, http://rand.org.

S. W. Raudenbush and A. S. Bryk. *Hierarchical Linear Models*. Sage Publications, 2 edition, 2002.

S. F. Reardon and G. Firebaugh. Measures of multigroup segregation. *Sociological Methodology*, 32(1):33–67, 2002.

C. Reid and E. Laderman. The untold costs of subprime lending: Examining the links among higher-priced lending, foreclosures and race in California. Community Development Investment Center Working Paper 2009-09, Federal

Reserve Bank of San Francisco, 2009.
P. A. Riach and J. Rich. Field experiments of discrimination in the market place. *The Economic Journal*, 112(483): 480–518, 2002.
P. A. Riach and J. Rich. Deceptive field experiments of discrimination: Are they ethical? *Kyklos*, 57(3):457–470, 2004.
P. A. Riach and J. Rich. An experimental investigation of sexual discrimination in hiring in the English labor market. *The B.E. Journal of Economic Analysis & Policy*, 6(2):Art. 1, 2006.
G. Ridgeway. Assessing the effect of race bias in post-traffic stop outcomes using propensity scores. *Journal of Quantitative Criminology*, 22(1):1–29, 2006.
G. Ridgeway and J. M. MacDonald. Doubly robust internal benchmarking and false discovery rates for detecting racial bias in police stops. *Journal of the American Statistical Association*, 104(486):661–668, 2009.
M. Rödin and G. Özcan. Is it how you look or speak that matters? An experimental study exploring the mechanisms of ethnic discrimination. Research Papers in Economics 2011:12, Stockholm University, Department of Economics, 2011. http://www2.ne.su.se.
T. Rogers and A. J. McKane. A unified framework for Schelling's model of segregation. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(7):P07006, 2011.
A. Romei, S. Ruggieri, and F. Turini. Discovering gender discrimination in project funding. In *Proc. of the IEEE ICDM 2012 Int. Workshop on Discrimination and Privacy-Aware Data Mining (DPADM)*, pages 394–401. IEEE Computer Society, 2012.
D.-O. Rooth. Obesity, attractiveness, and differential treatment in hiring: A field experiment. *Journal of Human Resources*, 44(3):710–735, 2009.
I. Rorive. Proving Discrimination Cases - the Role of Situation Testing, 2009. Centre For Equal Rights & Migration Policy Group, http://www.migpolgroup.com.
V. Roscigno, D. L. Karafin, and G. Tester. The complexities and processes of racial housing discrimination. *Social Problems*, 56(1):49–69, 2009.
P. R. Rosenbaum and D. B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.
S. L. Ross. Mortgage lending discrimination and racial differences in loan default. *Journal of Housing Research*, 7 (1):117–126, 1996.
S. L. Ross. Mortgage lending discrimination and racial differences in loan default: A simulation approach. *Journal of Housing Research*, 8(2):277–297, 1997.
S. L. Ross and G. M. B. Tootell. Redlining, the community reinvestment act, and private mortgage insurance. *Journal of Urban Economics*, 55(2):278–297, 2004.
S. L. Ross and J. Yinger. The default approach to studying mortgage discrimination: A rebuttal. In M. A. Turner and F. Skidmore, editors, *Mortgage lending discrimination: A review of existing evidence*, pages 107–127. The Urban Institute, 1999.
S. L. Ross and J. Yinger. *The Color of Credit: Mortgage Discrimination, Research Methodology, and Fair-Lending Enforcement*. MIT Press Books. The MIT Press, 2002.
S. L. Ross and J. Yinger. Uncovering discrimination: A comparison of the methods used by scholars and civil rights enforcement officials. *American Law and Economics Review*, 8(3):562–614, 2006.
D. B. Rubin. Causal effects using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100(469):322–331, 2005.
S. Ruggieri, D. Pedreschi, and F. Turini. Integrating induction and deduction for finding evidence of discrimination. *Artificial Intelligence and Law*, 18(1):1–43, 2010a.
S. Ruggieri, D. Pedreschi, and F. Turini. Data mining for discrimination discovery. *ACM Trans. on Knowledge Discovery from Data*, 4(2):Article 9, 2010b.
S. Ruggieri, D. Pedreschi, and F. Turini. DCUBE: Discrimination discovery in databases. In A. K. Elmagarmid and D. Agrawal, editors, *Proc. of the ACM SIGMOD Int. Conf. on Management of Data (SIGMOD 2010)*, pages 1127–1130. ACM, 2010c.
J. S. Rugh and D. S. Massey. Racial segregation and the American foreclosure crisis. *American Sociological Review*, 75(5):629–651, 2010.
U. Sandström and M. Hällsten. Persistent nepotism in peer-review. *Scientometrics*, 74(2):175–189, 2008.
S. Sano. Testing the taste-based discrimination hypothesis: Evidence from data on japanese listed firms. *Japan Labor Review*, 6(1):36–50, 2009.
K. Schaefer and M. L. Visser. Reverse regression and orthogonal regression in employment discrimination analysisn. *Journal of Forensic Economics*, 16(3):283–298, 2004.
J. A. Scheinkman. Social interactions (theory). In S. N. Durlauf and L. E. Blume, editors, *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, 2008.
T. C. Schelling. Dynamic models of segregation. *Journal of Mathematical Sociology*, 1(May):143–186, 1971.
D. Schiek, L. Waddington, and M. Bell, editors. *Cases, Materials and Text on National, Supranational and International Non-Discrimination Law*. Hart Publishing, 2007.
M. H. Schill and S. M. Wachter. A tale of two cities: Racial and ethnic geographic disparities in home mortgage lending in Boston and Philadelphia. *Journal of Housing Research*, 4(2):245–275, 1993.

W. Schreurs, M. Hildebrandt, E. Kindt, and M. Vanfleteren. The role of data protection law and non-discrimination law in group profiling in the private sector. In M. Hildebrandt and S. Gutwirth, editors, *Profiling the European Citizen: Cross-Disciplinary Perspectives*, pages 258–287. Springer, 2008.

V. L. Shavers, P. Fagan, D. Jones, W. M. P. Klein, J. Boyington, C. Moten, and E. Rorie. The state of research on racial/ethnic discrimination in the receipt of health care. *Am J Public Health*, 102(5):953–966, 2012.

D. J. Sheskin, editor. *Handbook of Parametric and Non-Parametric Statistical Procedure*. Chapman & Hall/CRC, 3 edition, 2004.

E. Shoben. Defining the relevant population in employment discrimination cases. In D. H. Kaye and M. Aickin, editors, *Statistical Methods in Discrimination Litigation*, pages 55–68. Marcel Dekker, Inc., 1992.

S. Siegel and N. J. Castellan. *Nonparametric statistics for the behavioral sciences*. McGraw-Hill, 2 edition, 1988.

P. Siegelman. Racial discrimination in 'everyday' commercial transactions: What do we know, what do we need to know, and how can we find out? In M. Fix and M. Tuner, editors, *A National Report Card on Discrimination in America: The Role of Testing*. The Urban Institute, 1999.

R. M. Silverman. Redlining in a majority black city?: Mortgage lending and the racial composition of Detroit neighborhoods. *The Western Journal of Black Studies*, 29(1):531–541, 2005.

R. M. Silverman. Mortgage lending disparities in metropolitan Buffalo: Implications for community reinvestment policy. *Journal of Regional Analysis & Police*, 38(1):36–44, 2008.

R. Slonim and P. Guillen. Gender selection discrimination: Evidence from a trust game. *Journal of Economic Behavior & Organization*, 76(2):385–405, 2010.

T. Sowell, editor. *Affirmative Action Around the World: An Empirical Analysis*. Yale University Press, 2005.

G. D. Squires. Racial profiling, insurance style: Insurance redlining and the uneven development of metropolitan areas. *Journal of Urban Affairs*, 25(4):391–410, 2003.

G. D. Squires, D. S. Hyra, and R. N. Renner. Segregation and the subprime lending crisis. Briefing Paper 244, Economic Policy Institute, 2009. http://www.epi.org.

C. M. Steele and J. Aronson. Stereotype threat and the intellectual test performance of african americans. *Journal of Personality and Social Psychology*, 69(5):797–811, 1995.

S. I. Stephens-Davidowitz. The effects of racial animus on a black presidential candidate: Using Google search data to find what surveys miss. Working paper, Harvard University, Department of Economics, 2012. http://ssrn.com.

T. J. Sugrue and W. B. Fairley. A case of unexamined assumptions: The use and misuse of the statistical analysis of Castaneda/Hazelwood in discrimination litigation. *Boston College Law Review*, 24(4):925–960, 1983.

M. Tian, M. L. Tang, H. K. T. Ng, and P. S. Chan. Confidence intervals for the risk ratio under inverse sampling. *Statistics in Medicine*, 27(17):3301–3324, 2008.

R. Tillyer, R. S. Engel, and J. Wooldredge. The intersection of racial profiling research and the law. *Journal of Criminal Justice*, 36(2):138–153, 2008.

R. Tillyer, R. S. Engel, and J. C. Cherkauskas. Best practices in vehicle stop data collection and analysis. *Journal of Police Strategies and Management*, 33(1):69–92, 2010.

C. Tobler. *Limits and potential of the concept of indirect discrimination*. European Network of Legal Experts in Anti-Discrimination, 2008. http://www.migpolgroup.com.

G. M. B. Tootell. Redlining in Boston: Do mortgage lenders discriminate against neighborhoods? *The Quarterly Journal of Economics*, 111(4):1049–1079, 1996.

T.-Y. Tsao and A. Pearlman. Decomposition of the black-white wage differential in the physician market. Economics Working Paper wp588, The Levy Economics Institute, 2010. http://www.levyinstitute.org.

M. A. Turner and S. L. Ross. How racial discrimination affects the search for housing. In X. de Souza Briggs, editor, *The Geography of Opportunity*, pages 81–100. Brookings Institution Press, 2005.

M. A. Turner and F. Skidmore, editors. *Mortgage lending discrimination: A review of existing evidence*. The Urban Institute, 1999. http://www.urban.org.

M. A. Turner, S. L. Ross, G. C. Galster, and J. Yinger. *Discrimination in Metropolitan Housing Markets: National Results from Phase I of HDS2000*. Urban Institute, Dep. Hous. Urban Dev, 2002. http://www.urban.org.

U.K. Legislation. (a) Sex Discrimination Act, 1975, (b) Race Relation Act, 1976, (c) Equality Act, 2010, 2012. http://www.statutelaw.gov.uk.

UNESCO. *Science, Technology and Gender: An International Report*. UNESCO Publishing, 4 edition, 2007.

United Nations Legislation. (a) Universal Declaration of Human Rights, 1948, (b) International Covenant for Civil and Political Rights, 1966, (c) International Covenant on Ecomomic, Social and Cultural Rights, 1966, (d) Convention on the Elimination of All forms of Racial Discrimination, 1966, (e) Convention on the Elimination of All forms of Discrimination Against Women, 1979, 2012. http://www.ohchr.org.

U.S. Federal Legislation. (a) Equal Credit Opportunity Act, 1974; (b) Fair Housing Act, 1968; (c) Employment Act, 1967; (d) Equal Pay Act, 1963; (e) Pregnancy Discrimination Act, 1978; (f) Civil Right Act, 1964, 2012. http://www.eeoc.gov.

I. Žliobaitye, F. Kamiran, and T. Calders. Handling conditional discrimination. In *Proc. of the IEEE Int. Conf. on Data Mining (ICDM 2011)*, pages 992–1001. IEEE Computer Society, 2011.

M. van der Leij, M. Rolfe, and O. Toomet. Social networks and the economic performance of minorities. Research paper, EQUALSOC Network of Excellence, 2009. http://equalsoc.org.

B. Vicki, I. Ellen, and J. Madar. The high cost of segregation: Exploring racial disparities in high-cost lending. *Fordham Urban Law Journal*, 36(3):361–393, 2009.

S.-Y. Wang. Statistical discrimination, productivity and the height of immigrants. Working Paper 3344, eSocialSciences, 2010. http://www.esocialsciences.org.

M. D. Ward and K. S. Gleditsch, editors. *Spatial Regression Models*. Quantitative Applications in the Social Sciences. Sage Publications, Inc, 2008.

D. Weichselbaumer and R. Winter-Ebmer. A meta-analysis of the international gender wage gap. *Journal of Economic Surveys*, 19(3):479–511, 2005.

D. Weichselbaumer and R. Winter-Ebmer. The effects of competition and equal treatment laws on gender wage differentials. *Economic Policy*, 22:235–287, 2007.

C. E. Weller. Credit access, the costs of credit and credit market discrimination. Working Paper 171, Political Economy Research Institute, University of Massachusetts at Amherst, 2008. http://www.peri.umass.edu.

C. Wennerås and A. Wold. Nepotism and sexism in peer-review. *Nature*, 387(5):341–343, 1997.

M. M. Whiteside and A. Narayanan. Reverse regression, collinearity, and employment discrimination. *Journal of Business & Economic Statistics*, 7(3):403–406, 1989.

B. E. Whitley and M. E. Kite. *The Psychology of Prejudice and Discrimination*. Wadsworth Publishing, 2 edition, 2009.

M. Whitney. The statistical evidence of racial profiling in traffic stops and searches: Rethinking the use of statistics to prove discriminatory intent. *Boston College Law Review*, 49(1):263–299, 2008.

D. P. Wilson and Y.-H. Ying. Nationality preferences for labour in the international football industry. *Applied Economics*, 35(14):1551–1559, 2003.

R. Wilson. Where the elite teach, it's still a man's world. *The Chronicle of Higher Education*, 51(15), 2004.

P. H. Wingate and G. C. Thornton. Statistics and employment discrimination law: An interdisciplinary review. In J. Martocchio, H. Liao, and A. Joshi, editors, *Research in Personnel and Human Resources Management*, volume 19, pages 295–337. Emerald Group Publishing Ltd, 2000.

E. K. Wyly, M. Moos, H. Foxcroft, and E. Kabahizi. Subprime mortgage segmentation in the American urban system. *Journal of Economic and Social Geography*, 99(1):3–23, 2008.

T. Yamagishi, N. Jin, and T. Kiyonari. Bounded generalized reciprocity: Ingroup boasting and ingroup favoritism. In E. J. Lawler and M. W. Macy, editors, *Advances in Group Processes*, volume 16, pages 161–197. Jai Press Inc., 1999.

C.-H. Yang and H.-Y. Lin. Is there salary discrimination by nationality in the NBA? Foreign talent or foreign market. *Journal of Sports Economics*, 13(1):53–75, 2012.

A. M. Yezer. A review of statistical problems in the measurement of mortgage market discrimination and credit risk. Report, Research Institute for Housing America, 2010. http://www.housingamerica.org.

J. Yinger. Evidence on discrimination in consumer markets. *The Journal of Economic Perspectives*, 12(2):23–40, 1998.

B. Yu and J. L. Gastwirth. A method of assessing the sensitivity of the Cochran-Mantel-Haenszel test to an unobserved confounder. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 366(1874):2377–2388, 2008.

M.-S. Yun. Decomposing differences in the first moment. *Economics Letters*, 82(2):273–278, 2004.

M.-S. Yun. Wage differentials, discrimination and inequality: A cautionary note on the Juhn, Murphy and Pierce decomposition method. *Scottish Journal of Political Economy*, 56(1):114–122, 2009.

L. Zhang and X.-Y. Dong. Male-female wage discrimination in Chinese industry - investigation using firm-level data. *Economics of Transition*, 16(1):85–112, 2008.

M. M. Zonta. Applying for home mortgages in immigrant communities: the case of Asian applicants in Los Angeles. *Environment and Planning*, 44(1):89–110, 2012.

A. Zussman. Ethnic discrimination and identity manipulation. Working paper, Department of Economics, The Hebrew University of Jerusalem, 2012.