Do Experiments Alter Econometrics?
A Historical Perspective Based on the Measurement of Market Discrimination

Cléo Chassonnery-Zaïgouche*

Abstract

Since the 1980s, experimentations have altered the use of econometrics as a dominant method of measurement in economics. The major differences between econometrics and experimentation concern the level and the type of control, which depends ultimately on the nature of the data. How has this renewal of methodology solved the problem of endogeneity, displayed by econometric control? The paper provides a historical answer to this question by focusing on a case study: the debate between indirect and direct methods of measuring market discrimination – *i.e.* the Oaxaca-Blinder wage decompositions vs. field and lab experiments.

**JEL:** B23  
**Keywords:** Econometrics (History of), Discrimination, Wage Decomposition, Testing, Experiments

---

*Centre d’économie de la Sorbonne, University Paris 1 Panthéon-Sorbonne, email: cleo.chassonnery@univ-paris1.fr*
Introduction

Econometrics uses the principle of controlled variations to produce measurements. Since the 1980s, controlled experiments have altered the use of econometrics as a dominant method of measurement. The major differences between econometrics and experimentation concern the level and the type of control, which depends ultimately on the nature of the data. Hence a central question arises: How has this renewal of methodology solved the problem of endogeneity, displayed by econometric control? The paper provides a historical answer to this question by focusing on a case study: the measurement of market discrimination.

“What Has Economics to Say about Racial Discrimination” asked Arrow in 1998? Originally focused on wage inequalities, discrimination has been a long-running concern in economics since the end of the 19th century (Webb, 1891; Fawcett, 1892, 1918; Edgeworth, 1922). Since the 1950s, Nobel Prize economists such as Akerlof, Arrow, Becker, Heckman, Phelps and Stiglitz have analyzed discrimination in an economic framework – as unequal treatment based on criteria such as gender and “race”. Discrimination is supposed to occur when equal productivity is not rewarded by equal pay, or, more generally, when individuals base their decisions on non-economic criteria concerning personal characteristics linked to group affiliation. Initiated by Gary S. Becker’s taste-based model (1955; 1971), the neoclassical framework was completed in the early 1970s by the statistical theories of discrimination (Arrow, 1971; Phelps, 1972). On the one hand, Becker’s model was an illustration of the Chicago movement of the 1950s which lead to the reconfiguration of labor economics along the line of what Heckman called the “analytical labor economics” (Heckman, 2003). The main goal was to find out micro-basis for labor market relations. On the other hand, besides “the human investment revolution in economic thought” (Bowman, 1966), Arrow’s work was representative of the “cognitive revolution” occurring in economics stemming from the rise of the economics of information. Statistical discrimination was further developed along with signal theory and job search models1 (Spence, 1973; Stiglitz, 1973; Lundberg, 1983). Following these theoretical developments, empirical analysis expanded the scope of the economic analysis of discrimination from gender and racial issues to other groups, from labor market to other kind of markets (credit, housing), as well as to non-market

1For a theoretical literature review see Cain (1986), Havet and Sofer (2002), Rodgers (2009) and Gazier (2010).
areas (education, health). Textbooks and historiography focus mainly on the history of theories (Prasch, 2004; Fleury, 2009). This paper provides a first insight on what would be a history of the quantitative analysis of discrimination.

The paper focuses on the debate between indirect and direct methods of measuring discrimination — *i.e.* the Oaxaca-Blinder wage decomposition methods *vs.* field and lab experiments. The first objective of both methods is to quantify discrimination and build evidence of its (non)existence as a phenomenon. Their second objective is to isolate the causal effect of discrimination over economic outcomes. For the wage decomposition method, the measured causal effect is the effect of human capital variables on wage — using Mincer’s wage logarithm — where discrimination is the residual effect. The estimation usually controls for occupation, education, age, etc. Beyond the classical problem of omitted variables, the main criticism addressed to this method is that control variables may be the result of past or pre-market discrimination — *e.g.* discrimination in promotion and in hiring or even auto-selection. The method does not permit to isolate one effect of market discrimination. The experiments movement rely on this criticism to propose new measurements of discrimination.

The first section presents the theoretical foundations and the historical background of the Oaxaca-Blinder decomposition methods and the three major innovations, which occurred since the 1970s: the change of the reference structure, the correction of selection bias, and the extension of the measure beyond the means value. The second section displays the main criticisms made to this methods concerning the impossibility of isolating a causal effect of discrimination. Finally, the last section presents the measure built through experimentation and its major consequence: the fragmentation of identity.

1 The Oaxaca-Blinder Methods: Decomposing Wage Differentials Into Legitimate/Lawful Inequalities and Outlaw Discrimination

1.1 The Theoretical Fundations of the Measure of Discrimination

The first statistical models come from the expected case method used to decompose the differences in the mortality rates (Kitagawa, 1955) also used in sociology to decompose the wage rates (Turner, 1952; Duncan, 1961). The
methodology was developed in economics in the wake of the operationalization of human capital theory, developed at Chicago and Columbia University. The first quantification of discrimination relies on the specific formalisation of a discrimination coefficient.

**Empirical Interpretations of Becker’s Discrimination Coefficient.**

Becker’s analysis produces a measure of discrimination by the formalisation of a discrimination coefficient. In his dissertation Becker (1955) defines wage discrimination as “the proportionate difference between the currently observed male-female wage ratio and the wage ratio that would prevail in the absence of discrimination” (Berndt, 1991, 182). Formally, the discrimination coefficient corresponds to:

\[
D = \frac{W_H/W_L - MP_H/MP_L}{MP_H/MP_L}
\]  

\[
\ln \frac{W_H}{W_L} = (\ln MP_H - \ln MP_L) + \ln (D + 1)
\]

where \( W_H/W_L \) is the ratio of observed wage rates of \( H \) (for high income group – men or non-discriminated group) and \( L \) (for low income group – women or racialized group), \( MP_H/MP_L \) is the ratio of marginal productivity of \( H \) and \( L \), and (1b) the logarithmic expression in terms of wage differentials. Oaxaca interprets this coefficient, using an equivalence between wage and productivity (Oaxaca, 1973b, 125)

\[
D = \frac{W_H/W_L - (W_H/W_L)^0}{(W_H/W_L)^0}
\]

\[
\ln (D + 1) = \ln W_H - \ln W_L - (\ln W_H - \ln W_L)^0
\]

where \((W_H/W_L)^0\) is the non-discriminatory wage ratio. Hence, the model is based on the following relation:

\[
\left( \frac{W_H^0}{W_L^0} \right) = \frac{MP_H}{MP_L}
\]

In the absence of discrimination, wages equal the marginal productivity of labor. The compensation formula of individuals is supposed to be iden-

\(^2\)For the quantitative aspect, Becker uses extensively Morton Zeman’s empirical analysis of difference of incomes (Zeman, 1955). Zeman did his Ph.D the same year as Becker, with the same supervisor, H. Gregg Lewis. For a history of the importance of Lewis in the making of Chicago Labor Economics see Rees (1976) and Kaufman (2010).

\(^3\)Henry L. Moore, the favourite student of John Bates Clark, provide empirical foundations to the marginal productivity theory (Moore, 1911), that Richard Lester later opposed (?). This controversy is at the core of the opposition between institutionalists and “analytical” labor economists (Stigler, 1947).
tical for every individual. Consequently, the first attempt of quantifying discrimination was elaborated using the famous Mincer wage equation.

**The Mincer Wage Equation.** Human capital theory was developed from a theoretical perspective (Mincer, 1958; Schultz, 1960; Becker, 1960). It was operationalized by means of the Mincer Equation, which links wage to education and experience:

\[
\ln y = \log y_0 + rS + \beta_1 X + \beta_2 X^2 
\]

where \(y\) is observed earnings – \(y_0\) the income of an individual with no education and no experience – decomposed into years of schooling \(S\), a quadratic function representing the years of potential experience on the labor market \(\beta_1 X + \beta_2 X^2\), and \(r\) the rate of return of human capital investments. The right term corresponds to the expected wage rate, according to the human capital theory.

The “mincering” of economics (Rosen, 1992) occurred within different contextual changes happening both inside and outside academic economics: the post-World War II context (Ehrenberg and Smith, 2008, 279), and the Johnson’s “War Against Poverty”, the debate over endogeneous growth (Cot, 1988, 188-198) and the international debates over education policies (Teixeira, 2008). Mincer and Becker worked together in the “legendary” Columbia workshop on labor economics (Heckman, 2003, 245) to built what was first called the “Morningside Heights Model” (Polachek, 1995) – i.e. the human capital theoretical and empirical framework, which was the basis of the “golden age” of analytical labor economics (Grossbard, 2006); a “revolution” in the field (Becker, 2006, 26). These developments constitute the basis of the measurement of discrimination. In order to measure discrimination, the expected wages obtained with a Mincer equation are then compared to a counterfactual situation:

What would have happened to a non-white individual [or a female] if he or she had been white [or a male]? (Blank et al., 2004, 5).

The first type of estimation uses a dummy to explain differences in wage by human capital variables and a residual effect being attributed to discrimination.

---

\[
\ln W_i = \beta_0 + X_i \beta - \delta F_i + \varepsilon_i, \text{ with } i = 1, \ldots, N_m + N_f \tag{5a}
\]
\[
\ln W_m = \beta_0 + X_m \hat{\beta} \tag{5b}
\]
\[
\ln W_f = \beta_0 + X_f \hat{\beta} \tag{5c}
\]
\[
\ln W_m - \ln W_f = (X_m - X_f) \hat{\beta} + \delta \tag{5d}
\]

where \( F \) equals 1 if the individual is a woman or a member of a discriminated group. The differences between the wage rates of the two groups is decomposed in differences of endowment \( X \) that have the same rate of return \( \hat{\beta} \) and a residual \( \delta \) that represents discrimination.

**Differences in Coefficients.** The Oaxaca-Blinder (O-B) model departs from that measurement of discrimination, precisely because:

"\[t\]here is no such thing as a single rate of return to education but rather a different rate of return for each age group. By contrast, Mincer pointed out in *Schooling, Experience and Earnings* that the experience-earnings profiles were relatively parallel for different education groups" (Lemieux, 2006, 128).

The O-B model estimates different coefficients that represent differential returns to human capital variables. Discrimination is now defined as the difference of coefficients. The differential rates of return concerning age groups is extending along gender and racial lines.

### 1.2 The Oaxaca-Blinder Seminal Model

**New Data Sets Driven by Social and Political Demand.** The Oaxaca model was first published in *Discrimination in Labor Market* edited by Ashenfelter and Rees (1973), as an outcome of a Princeton conference on discrimination\(^5\). At the same time that Becker’s work was awakening a renewed interest (second edition of *The Economics of Discrimination*), the opening of new data sets entailed a big push in development of this type of measures. Data sets are one of the results of the "War Against Poverty" that endorses an institutional struggle against racial inequity during the Kennedy and Johnson Administrations. This set of social reform culminate with the

\(^5\)Conference on Discrimination in Labor Markets", 7\(^{th}\) and 8\(^{th}\) October 1971, Princeton University, supported by the Industrial Relation Section and the Woodrow Wilson School of Public and International Affairs.
Civil Right Act of 1964 (Brinkley, 1991, 472), which creates the Equal Employment Opportunity Commission. This commission played a great role in adding the Survey of Economic Opportunity\(^6\) to the Census data and supporting the Panel Study of Income Dynamics of the University of Michigan\(^7\). Oaxaca uses the SEO while Blinder uses the PSID data set (Oaxaca, 1973a; Blinder, 1973). The model of wage decomposition acts as a “mediator” (Morgan and Morrison, 1999) between data and theory.

**A Classical Estimate of Difference in Coefficient Using Multiple Regression.** The Oaxaca-Blinder model estimates two separate equations of the two groups. The difference between the two rates obtained by these estimates is then decomposed in portion attributable to differing endowments \(X\) for an individual \(i\) (\(E\)), a portion attributable to differing coefficients (\(C\)) and an unexplained portion (\(U\)):

\[
\ln{W}_H - \ln{W}_L = \beta_H^0 + \sum_i \beta_H^i X_i^H - (\beta_L^0 + \sum_i \beta_L^i X_i^L) = E + C + U \tag{6a}
\]

\[
E = \sum_i \beta_H^i (X_i^H - X_i^L) \tag{6b}
\]

\[
C = \sum_i X_i^L (\beta_H^i - \beta_L^i) \tag{6c}
\]

\[
U = \beta_H^0 - \beta_L^0 \tag{6d}
\]

For Blinder, discrimination corresponds to \(C + U\), for Oaxaca, it is measured by \(C\).

**Innovations in O-B Decomposition Methods.** The model has been modified since the mid-1970s. Even today three major questions are at the center of these developments\(^8\): How to build the non-discriminatory structure? How to suppress selection biases? Is the average wage a good measurement?

Choosing a non-discriminatory structure implies choosing the coefficient used

---


\(^7\)This data based was produced by the Survey Research Centre of the University of Michigan and is considered as the longest running longitudinal household survey in the world, beginning in 1968.

\(^8\)For a survey of these questions see Cain (1986), (Oaxaca, 2007) and (Beblo et al., 2003).
to estimate the difference in observable characteristics. The coefficient from the supposedly non-discriminated group was generally used before a reflexion on favoritism was introduced:

“The simplest decomposition procedure is to adopt one of the estimated wage structures as the nondiscriminatory norm. Often researchers select the wage structure for the group of workers believed to be dominant in the labor market (at least relative to the comparison group). [...] In the more general approach the nondiscriminatory wage structure is estimated from a pooled sample of the two demographic groups. This approach allows the discrimination component to be further disaggregated into overpayment (favoritism) and underpayment (pure discrimination)” (Neuman and Oaxaca, 2005).

This type of models (Cotton, 1988; Neumark, 1988; Oaxaca and Ransom, 1994) introduce the favoritism component in a classical O-B decomposition:

\[
\ln W_H - \ln W_L = (X_H - X_L)\hat{\beta}^* + X_H(\hat{\beta}_H - \hat{\beta}^*) + X_L(\hat{\beta}^* - \hat{\beta}_L) \quad (7)
\]

The literature displays different estimations of \( \hat{\beta}^* \), considering the proportion of the two groups in the total population.

The second question concerns the selection bias, which could affect the estimation:

“at the stage of joining the employed labor force and when a specific occupation or an occupational status [e.g. union/nonunion membership, private/public] is chosen” (Neuman and Oaxaca, 2005).

The correction usually adopts a two-step procedure à la Heckman (1976; 1979) using a Inverse of Mill’s Ratio as a regressor (Neuman and Oaxaca, 2001; Neuman, 2004b,a; Neuman and Oaxaca, 2005):

\[
\ln W_H - \ln W_L = (X_H - X_L)\hat{\beta}_H + X_f(\hat{\beta}_H - \hat{\beta}_L) + (\theta_H\hat{\lambda}_H - \theta_L\hat{\lambda}_L) \quad (8)
\]

The bias is then decomposed in a portion link to observables and a portion interpreted as discrimination\(^9\):

\[
\ln W_m - \ln W_f = (X_m - X_f)\hat{\beta}_m + X_f(\hat{\beta}_m - \hat{\beta}_f) + \hat{\sigma}_m\hat{\lambda}_m + \hat{\sigma}_f\hat{\lambda}_f \quad (9a)
\]

where \( \hat{\sigma}_m\hat{\lambda}_m + \hat{\sigma}_f\hat{\lambda}_f = \hat{\sigma}_m\hat{\lambda}_m(\hat{\lambda}_m - \hat{\lambda}_f) + \hat{\sigma}_m\hat{\lambda}_m(\hat{\lambda}_f - \hat{\lambda}_f) \quad (9b)\)

\(^9\)For Oaxaca and Ransom \( \hat{\beta}^* = \Omega\hat{\beta}_H + (I - \Omega)\hat{\beta}_L \) with \( \Omega = (X'X)^{-1}(X_H'X_H) \) (1994; 2007). See also Reimers (1983); Cotton (1988); Neumark (1988). Silber et Weber (1999) use the four type of estimation on israeli data and display slightly different results between methods.

\(^{10}\)This procedure affects the amount of discrimination: “The problem is how to interpret the effects of the selectivity terms along the lines of explained and unexplained (discrimi-
\( \lambda_f^o \) is the mean value of the Inverse Mills Ratio, if women faced the same estimated \( \lambda \) than the men. Two effects of the selection bias are estimated: the effect of gender differences in the probability of being employed – \( \hat{\theta}_m \hat{\sigma}_m (\lambda_m - \lambda_f^o) \) – and an effect of gender differences on the selection parameters – \( \hat{\theta}_m \hat{\sigma}_m \lambda_f^0 - \hat{\theta}_f \hat{\sigma}_f \lambda_f \).

Another development of the O-B decomposition methods extends the analysis to incorporate the dispersion of wages between groups. As an example, Juhn et al. (1993) provide a decomposition for different quantiles:

\[
\ln \bar{W}_H - \ln \bar{W}_L = \hat{\beta}_H (X^H_q - X^L_q) + X^L_q (\hat{\beta}_H - \hat{\beta}_L) + (\varepsilon^H - \varepsilon^L) \tag{10}
\]

Beyond differences of inobservables – \((\varepsilon^H - \varepsilon^L)\) – this method is used with other measurements of dispersion.

The crude description of the evolution of the O-B methodology follows the development of econometrics since the 1970s. There is nothing special about the study of discrimination in the modification of the technical apparatus – the structure of the measure remains the same. The first feature of this general history concerned the abandonment of the search for a true value, illustrated in the corpus by the change in the norm of non-discrimination. There is no non-discriminatory wage structure in absolute but mixed procedures to build a norm of non-discrimination with available descriptive data. The second feature is the correction of the selection bias developed separately and then applied – with a time interval – to the analysis of discrimination. The third element concerns the move from a focus on the average difference between groups to a more precise knowledge of the dispersion within each group.

2 A Change in the History of Econometrics?

The Role of the Traditional Criticism of Econometrics in the Rise of Experimentation

Empirical analysis in economics was renewed by the use of experimental methods, challenging classical econometrics. The two methods, applied to the study of discrimination, have two historical backgrounds: one stems from
quantitative sociology and demography for wage differentials (Palmer, 1947; Turner, 1952); the other, from experimental psychology – especially in the spirit of the Clark’s dolls study\textsuperscript{11} (Clark and Clark, 1950). Are these histories related? The present paper argues that a part of the experiments movement in labor economics, in the case of discrimination analysis, could be understood as a challenge to existing measurement techniques, and essentially to multiple regression analysis. The historiographic consequence is to insert the development of experiments within a history of econometrics\textsuperscript{12} (Morgan, 1990, 2) or quantification in economics.

2.1 Traditional Limits of O-B

The traditional O-B decomposition contributed to change the belief that the wage gap between groups was entirely due to differences in personal characteristics (Bergmann, 1973). By drawing a frontier between legitimate inequalities and illegitimate discrimination, the method endorses a specific delimitation of discrimination by focusing on a pure wage discrimination. What is specific to the study of discrimination is a debate on the perimeter of discrimination, along with the question of which control variables to include in the regression.

The analysis was developed along extensive lines, largely dependent on datasets opening in an expanding numbers of countries\textsuperscript{13}. The corpus displays a variety of results according to each type of methodologies and the number of variables integrated:

“[t]hese various studies demonstrate that measures of discrimination by race vary considerably among racial and ethnic groups in the United States. They also appear to suggest that if one includes an expanded set of explanatory variables in regression equations, such as facility in the English language, the effects of discrimination are estimated to be smaller” (Berndt, 1991, 187).

“[I]t is clear that the magnitude of the estimated effects of discrimination crucially depends upon the choice of control variables for the wage regressions” (Oaxaca, 1973a, 699).

The classical arbitration between available tools and variables constitutes the basis of empirical strategies. Classical problems of econometric measurement

\textsuperscript{11}This study have a major impact in the enactment of the end of institutional segregation, and is quoted along with Myrdal’s \textit{American Dilemma} (1944) in the \textit{Brown vs board of Education} decision (U.S. Supreme Court 347 U.S. 483, May 17, 1954, footnote p.11, http://laws.findlaw.com/us/347/483.html [last access on 12 november 2013].

\textsuperscript{12}For a general history see Morgan (1990).

\textsuperscript{13}For a geographical review, see (Beblo et al., 2003).
arise and will become the basis of both the diversification of identification strategies and new methods.

<table>
<thead>
<tr>
<th>Problem</th>
<th>Definition</th>
<th>Consequences</th>
<th>Appropriate method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selectivity</td>
<td>non participants are not randomly distributed over the population</td>
<td>in case of positive selection: over-estimated wages, hence under-estimated pay gap</td>
<td>all except OLS or selection bias correction</td>
</tr>
<tr>
<td>Endogeneity</td>
<td>error term is correlated with at least one of the explanatory variables</td>
<td>biased estimated coefficients</td>
<td>Instrumental variable estimation</td>
</tr>
<tr>
<td>Heterogeneity</td>
<td>unobservable individual characteristics and the wage or explanatory variables</td>
<td>impact of the correlated explanatory variable is underestimated if correlation negative; <em>vice versa</em></td>
<td>FE panel estimators</td>
</tr>
</tbody>
</table>

Tabl. 1: Overview of methodological problem when estimating wage equations, adapted from Beblo *et al.* (2003, 25)

All the authors recognize the limits of the methodology itself and acknowledge the epistemological constraints of an indirect method of measurement. As revealing the difficulty of integrating all unobservable individual-level characteristics that are correlated with productivity, discrimination as a residual may reflect only the incomplete list of variables or the incomplete structure of the model.

More than the existence of an alternative identification procedure, a major question occurs when one might consider differences in these endowment variables as being the result of pre-market or previous market discrimination. O-B models display an endogeneity and heterogeneity problem: some control variables, if not all, may be influenced of other types of discrimination.

2.2 The Measurement Perimeter: Challenging the Definition of Discrimination

"The Statistical decomposition of racial gaps in social outcomes using multivariate regression and related techniques is a valuable tool for understanding the sources of racial differences. However, such decompositions using data sets with limited numbers of explanatory variables, such as the *Current Population Survey* or
the decennial census, do not accurately measure the portion of those differences that is due to current discrimination” (Blank et al. 2004, 8).

This final statement engages not only a methodology and its capacity to measure, but the content of the measurement. Methodological choices reflect conceptions of discrimination. Discrimination may have cumulative effects at different periods of an individual’s life. Even if one focuses on a strict definition of wage discrimination, the controlling procedure is complicated by the existence of different types of discrimination occurring at different time. Discrimination is thus likely to be under-estimated. Three types of discrimination may be included in the control variables: discrimination in hiring, in education and in overall labor market experience.

Occupational segregation can be interpreted as a result of discrimination in hiring. More generally the distribution of individual across occupations is interpreted as a choice in human capital investment, but alternative characterizations interpret it as a result of discriminatory practices. As human capital variables are treated as revealed preferences wage decomposition could not assess for this type of discrimination. Inspired by segmentation theories (Doeringer and Piore, 1971) and the “crowding hypothesis” (Bergmann, 1971), this criticism questions the implicit representation of an isolated labor market.

Wage discrimination does not disentangle market discrimination – demand side – from pre-market discrimination – supply-side (Butler, 1982). Pre-market discrimination might possibly have a great impact on market discrimination and greatly influence control variables:

“Many individual-level characteristics that we would like to include as controls may be influenced by discrimination. For example, because blacks tend to have lower levels of experience than whites, it would seem natural to include experience in the wage regressions. However, if blacks have lower levels of experience because of past discrimination, then including experience will lead the coefficient on race to understate the true extent of discrimination” (Antonovics et al. 2005, 921).

---

14 In his analysis of racial discrimination, Myrdal develops a “cumulative causation principle” that he borrows from Wicksell’s conception of causality (Myrdal, 1944, 1065-1070).

15 Some decomposition methods include a type a occupational discrimination. Notably, Brown et al. (1980) add to a standard O-B decomposition an effect of across occupation gap and an effect of employment shares across occupations. For a review of literature using this method (four articles), see Beblo et al. (2003, 36-39).
The same remark applies to education. Furthermore, feedback effects may influence human capital variables, contributing to alter the demarcation between what is exogeneous and what is endogeneous. If a type of self-fulfilling prophecies do exist, O-B wage decomposition does not correctly measure discrimination (Prasch, 2004, 174):

“If the employer’s belief that race is associated with low productivity is false, a feedback mechanism may nonetheless set up a self-fulfilling prophesy that will confirm that belief” (Reich, 1981, 104-105)

In more general terms, differences in productivity may be the result of discrimination. In that case, dependent variables become the actual independent (or explanatory) variables:

“[I]n other words, women’s current choices may well reflect adaptation to expected future discrimination by gender. According to this line of reasoning, discrimination by gender not only manifests itself in differences in regression coefficients, [...] but is also evident in distinctive career choices and in other variables that should be treated as endogeneous, not exogenous, in regression equations” (Berndt, 1991, 190).

The objective of the econometric measure of discrimination is to quantify discrimination and enlighten causal relations. In the case of O-B decomposition, the causal relation is the effect of human capital variables on wage, and their non effect – discrimination – when applied to certain groups. Two positions are taken on the possible evolution of the measure. The first clearly expects more accurate measures of productivity in the future through an improvement of the statistical data:

“If it were possible to control for virtually all sources of variation in wages, one could pretty well eliminate labor market discrimination as a significant factor in determining wage differentials by sex (or race)” (Oaxaca, 1973a, 699).

In this perspective, employers are rational, individuals do differ in terms of productivity and there is a correlation between being a member of a group

\[^{16}\text{For example, one might argue that the reason women tend to choose certain careers and develop distinctive personality traits is not because of male/female differences in inherent endowments or preferences, but rather because females expect to encounter future labor market discrimination, perhaps in the form of occupational discrimination” (Berndt, 1991, 189-190).}\]
and the under-investment in human capital. A second position expects the persistence of an unexplained residual event if we perfectly measure productivity. The implicit question besides those speculations on the future of measurement lies in the responsibility of the act of discriminating. Who is responsible for discrimination? Members of minority groups that underinvest in human capital or employers, that have tastes against them or bad information on them? By producing data with new tools, experimentations carry out a change of viewpoint on both the possibility of measure and the question of responsibility of discrimination.

3 “Caught in the Act”: Building Evidence of Discrimination

Experiments on discrimination became important in the 1980s as a tool of evaluation and as a matter of academic discussions (Rich and Judge, 2010). Even if field and lab experiments were developed outside of economic department at first, the rhetoric they developed was a challenge to existing methodologies – in our case study, a critique towards wage differentials. This element allows to insert their contribution within the history of discrimination measurement along with regression analysis. Historically applied to housing market\(^{17}\) the EEOC and others public agencies now recognize results from experiments as conclusive or valid evidence.

3.1 Challenging Wage Decomposition by a New Direct Method of Measurement?

Beyond the Gun Smoke For experimentators,

“while wage regressions often provide interesting descriptions of the data, they do not provide solid evidence of discrimination” (Antonovics et al., 2005, 921).

The economic definition of discrimination became more and more important in legal court (Ashenfelter and Oaxaca, 1987). The social use of such models is ensured by the measurement they allow, as a monitoring tool and as rhetoric argument in policy making. Wage differentials are widely used in evaluation procedures especially to compare levels of discrimination in different countries and/or at different times, to measure the effect of market

\(^{17}\)The Urban Institute of Washington D.C. was crucial in the demand and financement of audit studies.
structure (competition and monopoly) on the amount of discrimination and to evaluate anti- or affirmative action policies. In 2001, the Committee on National Statistics set up a panel on methods for assessing discrimination. The committee validate field and lab experiments as the more reliable tool to measure discrimination – compared to statistical analysis of observable data and survey data. This position is based on the strong internal validity of such methodology. The objective of experiments is to build evidence of discrimination rather than making aggregated quantification. Oaxaca describe the measurement of discrimination by wage differential as the measure of the “gunsmoke”: what is measured is the effect of discriminatory behaviors. The criticism of wage differential models will essentially reflect the indirect aspect of the method, which assimilates discrimination to gunsmoke but says nothing about who pulls the trigger and how and why the shot was fired off. The objective of experimentation is to “catch in the act” the discriminatory agent – as express in the title of Yinger (1986). The principle of such experiments aims at providing a framework to observe behaviors in a “controlled” environment. Laboratory, natural and field experiments offer different levels and types of “control” on a real or lab context, but rely on the same principle. As discrimination is an unobservable element, experiments intend to make it visible.

Diverse Levels of Control: Lab, Field and Natural Experiments. Audit testing consists in matching couples of applicants who differ only in terms of personal characteristics or group affiliation, and comparing the outcomes (treatment in a broad sense: interview, job offer or refusal, interview length etc.) between the representatives of the two demographic groups. Discrimination is defined as “a systematically less favorable treatment for the minority auditors” (Yinger, 1986, 881). The testing could be base on a written test (letter and/or CV), a job interview or a housing visit (Yinger, 1986; Turner et al., 1991; Ayres et al., 1995; Neumark et al., 1996; Yinger, 1998; Bertrand and Mullainathan, 2003). Lab experiments essentially use games to test a proxy for generosity, mistrust, or, at any rate, positive or negative perception of individuals toward a particular a group (Fershtman and Gneezy, 2001; Oaxaca and Dickinson, 2006; Castillo and Petrie, 2010). If mistrust against individuals from one particular group can systematically be

---

18Supported by the Ford Foundation and the Andrew W. Mellon Foundation, Agriculture and Education Federal Departments. Results were published in Blank et al. (2004), and as a book: America Becoming: Racial Trends and Their Consequences (National Research Council, 2001).

19Testing has been conducted for over 30 years across 10 countries, for a review see Riach and Rich (2002) and Bendick (2007).
found, this behavior is associated with discrimination. Exchanges between individuals take place using money (virtually or not) as a measuring rod for behavior toward a particular group. For example, in a dictator game between individuals from two different backgrounds, a player A is said to discriminate if the player from group A systematically gives less money to a player from group B. Group A and group B are built as homogeneous demographic groups (women, individuals with a religious representative name, with the same age etc.). A natural experiment is a real event that unintentionally reproduces the conditions of a controlled experiment. The only example we found on discrimination deals with a TV show (“The Weakest Link”) and does not display a significant amount of discrimination against a particular minority or against women \(^{20}\), but a large amount of “strategic discrimination” against men by women participants (Antonovics et al. 2005).

**The Rhetoric of a Direct Method.** The “profession’s enthusiasm” for the wage differentials as a measure of discrimination led progressively to “an increasing vocal minority [...] expressing concern about the specification problems associated with the regression procedure” (Riach and Rich, 1991, 143). The experimental method highlights the search for “catching the discriminator in the act” the discriminators. Concerning the criticism of wage differentials, objections mainly concern the confusion between current discrimination and pre-market discrimination, “that is, differences in worker’s access to quality schooling, cultural enrichment, health care, residential location, and other factors which equip job seekers with prerequisites for many better-paid, more prestigious jobs” (Bendick, 2007, 18). Audit testing controls these elements by selecting (or creating) comparable pre-market experience during testers’ training or in a fictional CV. In the lab, these elements are not necessary for the study of decision, making the results less representative of real markets but more striking on “pure” discrimination. More generally,

“discrimination can, after all, take many forms besides sex-based wage differentials, and seems unlikely to be completely captured by so crude a measure as a log-earnings regression” (Kuhn, 1990, 297)

This method is based on a rhetorical discourse that moves from considerations on the (economic) characteristics of the victims to the characteristics of the discriminators. In that sense, experiments call for a focus on the re-

\(^{20}\)For another study on the same TV show, see Levitt (2003).
sponsability of the act of discriminating while the use of wage differentials represents the advent of a more “supply-side” analysis.

3.2 Experiments as a Perfect Substitute for Wage Differentials?

The nature of the experimental data allows to test for potential discrimination where wage differentials measure effective discrimination by using administrative statistics measures. The question (here) is to isolate a causality and to control for other sources of discrimination remainings, even when producing experimental data.

The Limits of Control. In the audit studies, the experimenter controls four steps in the experiments: 1) the recruitment of paired testers, 2) their training and the matching process that permits making equivalent “application”, 3) the experiment management itself and 4) the replication of the experiments. The main weakness of this approach is the lack of generalization – i.e. a weak external validity. Results depend on samples and locations of the experiments and on the objectivity of testers:

“Auditors are sometimes instructed on the “problem of discrimination in American society” prior to sampling firms, so they may have been coached to find what the audit agencies wanted to find. [...] Audit evidence does not translate into actual employment experiences and wages obtained by actors who purposively search markets” (Heckman, 1998, 104)

This last problem has been tentatively addressed by moving such experiments into the lab and by selecting participants randomly. Within the lab, the simplified framework of the game works as a representation of a real market. Several assumptions are made to control for other criteria different than race or gender. However, these assumptions do not really ensure the isolation of discriminatory behavior. The impossibility of generalization and the doubt on the representativeness of games as a proxy for real markets limits the usefulness of such experiments.

Heckman’s Criticism. The experimental procedure, especially the audit testing, give rise to renewed criticisms on the part of econometricians, notably by Heckman (Heckman, 1998). Heckman first enacts the prevalence of decomposition methods over experimental results:
“[M]ost of the disparity in earnings between blacks and whites in the labor market of the 1990s is due to differences in skills they bring to the market, and not to discrimination within the labor market. [...] Labor market discrimination is no longer a first-order quantitative problem in American society” (Heckman, 1998, 101)

Heckman especially insists on the existence of pre-market differences and unobservables quoting a study that measures the effect of IQ tests on wages (Neal and Johnson, 1996). The effect of the results of the Armed Forces Qualifications Test (AFQT) on wages still displays a “black-white wage gap”, a decade after the test was taken. He gives the precision that this element is not a measure enforcing genetic determination but reflecting differences in schooling and family environments.

The major methodological criticism is based on the existence of unobservable characteristics that are not distributed similarly within each group:

“[T]his evidence [of disparate treatments] has to be taken with a grain of salt, because it is based on the implicit assumption that the distribution of unobserved productivity is the same in both race groups” (Heckman, 1998, 105).

“A related crucial assumption behind the audit method is that it typically assumes that the outcome being studied is a linear function of the relevant variables of the outcome can be transformed to be so; that is, more skill leads in a linear way to a greater chance of employment or higher wages. (Heckman, 1998, 109).

He recalls that the relation between job and relevant characteristics is usually non-linear ultimately “depending on the distribution of unobserved characteristics for each race group” (Heckman, 1998, 111). The two methods are not substitutes and can be used as complement (Darity and Mason, 1998; Adida et al., 2010).

**Building a Non-Onthological Definition of Identity : From a Dummy to a Cliché.** One of the specific consequence of the development of experiments is the reconfiguration of the representation of the economic agent through the modeling of identity, not in terms of census category but in a more decomposed form (language, skin color, signaling ... ). Discrimination has a formal or universal definition and a variety of forms (discrimination in hiring process, income disparity) and “contents” (gender, race, age, etc.). Hence, the morphology of the concept leads mechanically to more empirical and practical works, not only in order to test theories, but also in order to
produce knowledge and evidence in different kinds of local situations. Discrimination, both as a concept and as a phenomenon, therefore requires different empirical investigations. Thus, the methodological framework changed to fit discrimination characteristics. The main consequence of this, regarding the reflection upon discrimination, concerns the contribution of experiments, leading to a non-ontological definition of identity: as a signal or as a cliché, i.e., a visible characteristic that does not depend on a true identity or identification.

Conclusion

The controversy shows a focus on an evidence-based research framework turned toward a policy-purpose, which ultimately rely on the reliability of the measurement. The debate over the type and the extent of discrimination translates into questions about controlling procedures. Wage differentials constitute supply-side analysis of discrimination while experiments try to provide direct measurements of discriminatory behavior from the demand side of the labor market (1995). Historically, the research for differences is prior to the measurement of discriminating behaviors.

How economists explain the world is important in order to understand how they conceive and propose political solutions. Becker’s “taste for discrimination” theory does not imply an authoritarian change of preferences but an economic sanction by the market mechanism. Statistical theories involve political action to eliminate differences due to education or educational policies to improve market signals of individuals from racial or sexual minority groups. Experiments widen the range of political uses of economics’ results without resolving methodological tensions about the isolation of causal effects.

References


25


