



Measuring trends in discrimination with field experiment data

Stephen L. Ross^{a,1}

The passage of major civil rights legislation was accompanied by a large improvement in the economic circumstances of African-Americans (1), and the public consensus against discrimination has grown over time (2). However, the significant postwar progress of African-Americans appears to have slowed or even stagnated over the last several decades (3, 4). The slowing progress of African-Americans in the United States raises questions concerning the persistence of racial discrimination in the United States. A recent study in PNAS (5) examines whether discrimination in hiring has declined over time by conducting a meta-analysis of testing studies or field experiments of discrimination. The study does not find a change in the level of hiring discrimination against African-Americans since 1989, but finds some evidence of declines in discrimination against Hispanics.

To my knowledge, that paper (5) represents the first attempt to measure changes in discrimination by combining data from many different testing efforts or field experiments. Previous attempts to measure changes in discrimination using testing data have relied on replication of earlier testing efforts. For example, the 2000 Housing Discrimination Study (6) measures changes in housing discrimination relative to the 1989 study. However, to accurately measure change, the researchers were required to minimize the number of differences between the studies by restricting the choice of sites or metropolitan areas, limiting changes to the survey instruments, and reanalyzing the 1989 data to match proposed analyses for 2000. By the 2010 study (7), the housing market and the US metropolitan environment had changed so much that these restrictions were not viable and attempts to directly measure changes were abandoned. Further, comparing only two points in time can confound cyclical or temporary changes in the market or economic environment with changes in discrimination. For example, the 2000 study was conducted in the middle of a prolonged housing boom, while the 2010 study followed the housing bust and the associated foreclosure crisis. Given these difficulties, it is very attractive to view the population of existing testing studies for a market as observations essentially randomly

drawn from the US economy over space and time, so that trends in the findings of those studies can be viewed as evidence of trends in discrimination.

Internal Validity

In addition, the recent meta-analysis (5) advocates for its use of testing data, in part, by appealing to the high internal validity of such studies. In testing studies, attributes like education, experience, income, or wealth are assigned randomly to individual tests or pairs of majority and minority testers. This process assures that race or protected class is orthogonal to the attributes observed by the agents in the market being studied. This feature leads to a key advantage of using testing studies for meta-analyses, as opposed to using regression style estimates: The analyst can use the actual counts from each study without worrying about control variables. As a result, the meta-analysis compares simple means or frequencies across samples, yielding, in essence, a weighed mean for the pooled sample. In comparison, meta-analyses of regression studies often compare estimates from a variety of estimation approaches, such as linear regression, regression discontinuity analyses, or difference-in-differences estimates. These models tend to vary in the identity of the marginal individual on whom the effect size estimates are based, and so can potentially confound changes over time with changes in the identification strategy.

Admittedly, methodological differences between audit studies can also influence estimates of discrimination. Historically, a key threat to the individual validity of such studies has been the use of human testers whose personal influence on treatment may correlate with the race or protected status of the tester (8). [Ross (9) argues that tester training can, in principle, mitigate the impact of testers on measures of discrimination. Further, Turner et al. (10) find, at most, a modest relationship between the actual attributes of testers and the treatment of home buyers, and their analysis suggests that these effects, at worst, biased their estimates of discrimination toward zero.] A hiring study in Chicago (11) addresses this concern by conducting tests with mail-in résumés, and, more recently, many studies have

^aDepartment of Economics, University of Connecticut, Storrs, CT 06269-3533

Author contributions: S.L.R. wrote the paper.

The author declares no conflict of interest.

See companion article on page 10870.

¹Email: stephen.l.ross@uconn.edu.

exploited the internet for testing purposes (e.g., ref. 12), again avoiding the need to employ human testers. However, this shift away from in-person tests over time toward résumé or internet tests may create trends in measures of discrimination. While the study by Quillian et al. (5) rules out this possibility in their application, future metastudies should be aware of this concern. More generally, metastudies that attempt to measure changes in phenomena over time should, as a rule, collect data on the details of the studies being combined and test for whether any structural attributes of the studies correlate with the years in which the studies were conducted. [The study by Quillian et al. (SI appendix in ref. 5) contains an example of such an analysis.]

Two other common design features of testing studies are (i) whether or not the study uses pairs of testers or résumés being sent to the same agent and (ii) the method for sampling the tested agents. By pairing testers or résumés, the analyst conditions on the unobservables of the agent being tested, substantially increasing the efficiency of the estimations (9). These efficiency gains have historically been very important for in-person testing, given the high cost of hiring, training, and deploying testers, but with low-cost résumé or internet testing, the use of pairs has become less important. In the study by Quillian et al. (5), the authors must ignore these efficiency gains for paired studies that do not report the rate of equal treatment of tester pairs; as a result, studies involving paired testers may be undercounted in metastudies. [Ondrich et al. (13) find a strong correlation between the treatment of testers who are sent to the same real-estate agent.] Second, a testing study must have some process for sampling individuals or firms in the market being tested. Historically, testing studies have sampled advertisements, for example, of job openings, home sales, or rental housing availability; however, as noted above, newspaper advertisements in many markets are being replaced by internet advertisements or social media outreach. To the extent that research design mirrors reality, the composition of testing studies over sampling strategies may tend to follow changes in the market, and so a metastudy of many testing efforts may be expected to capture both the effect of changing attitudes and the effect of changing search strategies on discrimination.

External Validity

While testing studies are thought to have high internal validity, questions have been raised concerning the external validity of such tests. Many researchers have argued that testing studies understate discrimination because they cannot go far enough into the process and, for example, miss the final hiring or rental decision (14). Alternatively, market discrimination is defined by Heckman (8) as the impact of discrimination on market outcomes, drawing a distinction from adverse treatment, as captured by testing studies. Testing

studies exploit some type of portal to expose testers or résumés to a random sample of opportunities in the market, but if minorities can avoid firms that discriminate, the results of such tests may substantially overstate the levels of discrimination experienced in the marketplace. For example, while testing evidence suggests a negative effect of “black” names on interview decisions (11), others (15) find no effect of “black” names on adult outcomes. Similarly, housing prices have been used to detect massive reductions in the exclusion of African-Americans from predominantly white neighborhoods in the postwar period, by demonstrating that the great migration generated a very large price premium in African-American neighborhoods, which then fell and reversed as the racial barriers to white neighborhoods fell or disappeared (16, 17). We do not know if testing studies would have captured these massive declines. Finally,

The analysis in the study by Quillian et al. provides convincing evidence that hiring discrimination against African-Americans has been relatively stable over the last few decades.

the narrative or anecdotal evidence from the 1989 Housing Discrimination Study was striking, with African-American testers often being explicitly excluded from housing or actively avoided by housing providers once their race was observed. Overall, the verbal treatment of African-American and Hispanic homeowners during the 2000 study was polite and professional, representing a dramatic change in the social environment encountered during the housing search process that was not necessarily captured by the more modest changes in the measures of adverse treatment in the study by Ross and Turner (6).

In summary, the use of meta-analyses to measure changes in discrimination appears to offer great potential, especially when applied to testing or field experiments. The analysis in the study by Quillian et al. (5) provides convincing evidence that hiring discrimination against African-Americans has been relatively stable over the last few decades. The empirical evidence is important for its direct policy relevance, and also because it might help explain the persistent low levels of employment among African-Americans (18) and the stagnation of declines in the black-white wage gap (19). I look forward to seeing future efforts to use meta-analysis to measure trends in discrimination in other markets. However, I also caution researchers that they cannot simply combine all of the studies that they find and hope to obtain convincing evidence of trends in discrimination. If metastudies are to be used to test for changes in discrimination, they will require significant efforts to document and exploit the institutional details of all studies included in the analysis.

- 1 Donahue J-J, III, Heckman J (1991) Continuous versus episodic change: The effect of federal civil rights policy on the economic status of blacks. *J Econ Lit* 29:1603–1643.
- 2 Bobo L-D, Charles C-Z, Krysan M, Simmons A-D (2012) The real record on racial attitudes. *Social Trends in American Life: Findings from the General Social Survey Since 1972*, ed Marsden P-V (Princeton Univ Press, Princeton), pp 38–83.
- 3 Chinhui J, Murphy K-M, Pierce B (1991) Accounting for the slowdown in black-white wage convergence. *Workers and Their Wages: Changing Patterns in the United States*, eds Chinhui J, Murphy K-M, Pierce B (American Enterprise Institute Press, Washington, DC), pp 107–143.
- 4 Neal D (2006) Why has black-white skill convergence stopped? *Handbook of the Economics of Education*, eds Hanushek EA, Welch F (Elsevier, Amsterdam), Vol 1, pp 511–576.
- 5 Quillian L, Pager D, Hexel O, Midtbøen AH (2017) Meta-analysis of field experiments shows no change in racial discrimination in hiring over time. *Proc Natl Acad Sci USA* 114:10870–10875.
- 6 Ross S-L, Turner M-A (2005) Housing discrimination in metropolitan America: Explaining changes between 1989 and 2000. *Soc Probl* 52:152–180.
- 7 Turner M-A, et al. (2013) *Housing Discrimination Against Racial and Ethnic Minorities 2012* (US Department of Housing and Urban Development/HUD User, Washington, DC).
- 8 Heckman J (1998) Detecting discrimination. *J Econ Perspect* 12:101–116.
- 9 Ross S-L (2002) Paired testing and the 2000 Housing Discrimination Study. *Measuring Housing Discrimination in a National Study: Report of a Workshop*, eds Feinberg S, Foster AW (National Resource Council/National Academy, Washington, DC), pp 49–66.

- 10 Turner M-A, Ross S-L, Galster G, Yinger J (2002) *Discrimination in Metropolitan Housing Markets: National Results from Phase 1 of HDS2000* (Department of Housing and Urban Development/HUD User, Washington, DC).
- 11 Bertrand M, Mullainathan S (2004) Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *Am Econ Rev* 94:991–1013.
- 12 Hanson A, Hawley Z (2011) Do landlords discriminate in the rental housing market? Evidence from an internet field experiment in US cities. *J Urban Econ* 70:99–114.
- 13 Ondrich J, Ross S-L, Yinger J (2000) How common is housing discrimination: Improving on traditional measures. *J Urban Econ* 47:470–500.
- 14 Pitingolo R, Ross S-L (2015) Housing discrimination among available housing units in 2012: Do paired-testing studies understate housing discrimination? *Cityscape* 17:61–85.
- 15 Fryer R, Levitt S (2004) The causes and consequences of distinctively black names. *Q J Econ* 119:767–805.
- 16 Cutler D-M, Glaeser E-L, Vigdor J-L (1999) The rise and decline of the American ghetto. *J Polit Econ* 107:455–506.
- 17 Ross S-L (2010) Understanding racial segregation: What is known about the effect of housing discrimination. *Neighborhood and Life Changes: How Place Matters in Modern American?* eds Newburger H, Birch E-L, Wachter S-M (Univ of Pennsylvania Press, Philadelphia), pp 288–301.
- 18 Fairlie R-W, Sundstrom W-A (1999) The emergence, persistence, and recent widening of the racial unemployment gap. *Ind Labor Relat Rev* 52:252–270.
- 19 Chinhui J (2003) Labor market dropouts and trends in the wages of black and white men. *Ind Labor Relat Rev* 56:643–662.