

Is It Harder for Older Workers to Find Jobs? New and Improved Evidence from a Field Experiment

David Neumark

University of California, Irvine, National Bureau of Economic Research, and IZA Institute of Labor Economics

Ian Burn

Stockholm University

Patrick Button

Tulane University, RAND Corporation, and IZA Institute of Labor Economics

We design and implement a large-scale resume correspondence study to address limitations of existing field experiments testing for age discrimination that may bias their results. One limitation that may bias results is giving older and younger applicants similar experience to make them “otherwise comparable.” A second limitation is that greater unobserved differences in human capital investment of older applicants may bias the results against finding age discrimination. On the basis of over 40,000 job applications, we find robust evidence of age discrimination in hiring against older women, especially those near retirement age, but considerably less evidence of age discrimination against men.

We received generous support from the Alfred P. Sloan Foundation and helpful comments from seminar participants at Boston College, Georgia State University, IZA Institute of Labor Economics, Marquette University, the New School, Ohio State University, the Sloan Foundation, Stanford University, Tulane University, University of California–Irvine, University of Tokyo, University of San Francisco, University of Wisconsin–Madison, Yale Law School, and the National Bureau of Economic Research Law and Economics Summer

Electronically published February 26, 2019

[*Journal of Political Economy*, 2019, vol. 127, no. 2]

© 2019 by The University of Chicago. All rights reserved. 0022-3808/2019/12702-0011\$10.00

I. Introduction

Longer unemployment durations of older workers have long been viewed as potentially reflecting hiring discrimination against older workers. This inference was surely easier to establish when there were explicit maximum age criteria in hiring ads (US Department of Labor 1965), but the persistence of lengthier unemployment durations for older workers, and ongoing Equal Employment Opportunity Commission (EEOC) enforcement activity, at least suggest that older workers are disadvantaged in the job search process.¹

A number of studies have used audit or correspondence (AC) study methods to test for age discrimination in hiring. These past studies nearly uniformly point to age discrimination in hiring (Bendick, Jackson, and Romero 1997; Bendick, Brown, and Wall 1999; Riach and Rich 2006, 2010; Lahey 2008).² However, the existing experimental evidence is potentially flawed in ways that could bias estimates of age discrimination in either direction. One issue is the practice of giving older and younger applicants similar labor market experience, consistent with the standard approach in these studies. However, the absence of relevant experience commensurate with an older applicant's age may be a negative signal, and on real-world resumes, older applicants report experience commensurate with their age. Second, Heckman and Siegelman (1993) and Heckman (1998) have demonstrated that if the groups studied have different variances of unobservables, experimental estimates of discrimination can be biased in either direction (formally, it is unidentified). This problem may be especially salient with respect to age, as the human capital model predicts greater dispersion in unobserved investments among older workers (Mincer 1974; Heckman, Lochner, and Todd 2006). Thus, it is hard to know what to make of the existing experimental evidence of age discrimination.

To provide more compelling evidence on age discrimination, we conducted a large-scale field experiment—a resume correspondence study—

Institute. The views expressed are our own and not those of the foundation. We thank Melody Dehghan, Dominique Dubria, Chenxu Guo, Stephanie Harrington, Kelsey Heider, Matthew Jie, Irene Labadlabad, Benson Lao, Karl Jonas Lundstedt, Catherine Liu, Jason Ralston, Eileen Raney, Nida Ratawessnant, Samantha Spallone, Bua Vanitsthian, Helen Yu, and especially Nanneh Chehras for outstanding research assistance and Scott Adams, Marc Bendick, James Heckman, Richard Johnson, Joanna Lahey, Bentley MacLeod, Doug Miller, Matthew Notowidigdo, and anonymous reviewers for very helpful comments. This study was approved by University of California–Irvine's Institutional Review Board (HS 2013-9942). Data are provided as supplementary material online.

¹ For recent data on unemployment rates, see <http://www.bls.gov/cps/cpsaat31.pdf>. The EEOC reports receipt of 20,857 charges of discrimination under the Age Discrimination in Employment Act (ADEA) in 2016, 2,162 of which are related to hiring. These figures exclude claims filed with state agencies; see https://www1.eeoc.gov/eeoc/statistics/enforcement/statutes_by_issue.cfm?renderforprint=1 and <https://www.eeoc.gov/eeoc/statistics/enforcement/adea.cfm>.

² A summary table is available in the online appendix.

explicitly designed to address the potential limitations and sources of bias in the previous experiments. We also study ages closer to retirement than in past studies and use a richer set of job profiles for older workers to address a number of additional questions and provide some help in distinguishing potential mechanisms of discrimination. On the basis of evidence from over 40,000 job applications, we find robust evidence of age discrimination in hiring against older women, especially those near retirement age. But we find that the evidence for men is less robust and that evidence of age discrimination against them may at least in part reflect the biases this study was designed to assess.

Knowing whether age discrimination deters employment of older workers is critical for at least two economic reasons. First, the aging of the population in the United States (and elsewhere), coupled with lower employment of older individuals, implies a rising dependency ratio and fiscal challenges for Social Security. Increasing work at older ages can help meet the fiscal challenge by increasing payroll tax receipts (US Government Accounting Office 1999). If age discrimination is an important demand-side barrier to extending work lives, then the policy response to population aging may need to include addressing this barrier, in addition to strengthening work incentives for older individuals who might otherwise retire. Age discrimination in hiring may be particularly critical to whether older workers can work substantially longer, because many seniors transition to part-time or shorter-term “partial retirement” or “bridge jobs” at the end of their careers (e.g., Johnson, Kawachi, and Lewis 2009) or return to work after a period of retirement (Maestas 2010). Second, there are economic costs of trying to root out hiring behavior defined as illegal by antidiscrimination laws. For those costs that we can reasonably quantify, we estimate that the costs of potential and actual hiring cases under the ADEA are about \$3.29 billion per year, which is about \$5,300 per firm covered by the ADEA, or about \$35 per covered employee, the vast majority of which are compliance costs.³ Our study is well designed to detect the behavior that these laws are designed to reduce or eliminate and hence

³ The detailed components of this cost estimate are discussed in the online appendix but include, for the purposes of this calculation, monetary damages of EEOC cases (\$4.6 million, or \$7.42 per covered firm per year), the greater of litigation costs or settlement costs paid by the firms (litigation is greater, so \$9.67 million to \$44.48 million per year, or \$15.62 to \$71.76 per covered employer per year), EEOC administrative costs (\$5.0 million, or \$8.06 per covered firm per year), and compliance costs (\$3.24 billion, or \$5,226 per covered firm per year). Using the upper estimate for the litigation costs, this leads to total costs of \$3.29 billion, or \$5,315 per covered firm and \$34.64 per covered employee. This does not include costs such as time spent by executives and management as part of these cases, damages for cases handled at the state level rather than by the EEOC, administrative costs for state agencies that enforce state laws, and productivity losses or gains from induced changes in employers' hiring behavior.

is informative about whether these costs are being borne to address an extant problem.

At the same time, this evidence does not necessarily dictate the appropriate policy response to reduce discrimination, in large part because it is difficult to distinguish between taste-based and statistical discrimination. The most natural policy response to taste discrimination à la Becker (1971) is to raise the cost to employers of engaging in discriminatory behavior, effectively restoring equal prices for labor from equally productive groups. The appropriate policy response to statistical discrimination is more complicated. If statistical discrimination is based on correct stereotypes (i.e., the group averages employers use are correct), then it may not introduce any inefficiency. Policy interventions that increase information about older applicants will help those applicants that defy the stereotypes in positive directions, and vice versa, so that the main rationale for policy intervention may be equity.⁴ If statistical discrimination is based on incorrect negative stereotypes about older workers, then such policy interventions can increase hiring of older workers as a group. While AC methods can establish evidence of discrimination, discerning between taste and statistical discrimination is much more challenging, although we argue that many features of our study design, and the findings, make it relatively more likely that taste discrimination underlies our findings, when we find evidence of age discrimination. Of course, policy makers may have a goal of increasing employment of older workers regardless of the nature of the discrimination, even if such behavior might be hard to rationalize on efficiency grounds.⁵

II. Past Nonexperimental Research on Age Discrimination

As noted above, prior to the passage of the ADEA, explicit age restrictions in hiring ads were frequent (US Department of Labor 1965). In addition, workers in their 50s and early 60s have long had lengthier unemployment durations than many other age groups.⁶ There is evidence of negative ste-

⁴ Affirmative action policies may act in this fashion, reducing reliance on cheap screens like race (Holzer and Neumark 2000; Miller 2017). If productivity depends on the quality of the job match, then statistical discrimination based on correct stereotypes can be inefficient and an information intervention can enhance efficiency.

⁵ Consider, e.g., the example of the Americans with Disabilities Act, which not only bars discrimination against individuals with disabilities but also requires employers to pay reasonable costs of accommodation.

⁶ Age discrimination leading to longer unemployment durations of older workers does not necessarily entail lost output and welfare if it simply reallocates time unemployed from younger to older workers without increasing the total time all workers spend unemployed. However, such reallocation seems likely to generate such losses if older workers respond to long unemployment durations by choosing to retire, in part because they can claim Social Security benefits earlier (consistent with evidence in Adams [2002]).

reotypes regarding older workers in hypothetical scenarios tying attitudes toward older workers to adverse labor market outcomes for them (Gordon and Arvey 2004; Kite et al. 2005). Finally, workers report experiencing age discrimination on the job, and these workers subsequently exhibit more separations, lower employment, slower wage growth, and reduced expectations of working past 62 or 65 (Johnson and Neumark 1997; Adams 2002).

However, these results from observational data are hardly decisive evidence of age discrimination. For example, longer unemployment durations could reflect higher reservation wages of unemployed older workers or narrower search rather than discrimination. Self-reports of age discrimination may reflect other negative outcomes workers experience, followed by leaving the firm, experiencing fewer promotions or raises, and so forth. That is, as is often a concern with observational evidence on discrimination, unobservables may underlie the evidence. Thus, as in research on discrimination along other dimensions, researchers have turned to AC studies to provide more compelling evidence on age discrimination. However, there are numerous challenges to applying such methods to age discrimination, which this paper tries to overcome.

III. Applying Experimental Methods to Studying Age Discrimination in Hiring

A. The General Method

Experimental audit or correspondence (AC) studies of hiring are generally viewed as the most reliable means of inferring labor market discrimination (e.g., Fix and Struyk 1993). While observational studies try to control for productivity differences between groups, AC studies create artificial job applicants in which there are intended to be no average differences by group, so that differences in outcomes likely reflect discrimination. Audit studies use actual applicants coached to act alike and capture job offers, whereas correspondence studies create fake applicants (on paper or electronically) and capture “callbacks” for job interviews. Correspondence studies can collect far larger samples of job applications and outcomes, especially using the internet; because of the time costs of interviews, even large-scale, expensive audit studies typically have sample sizes only in the hundreds. Correspondence studies also avoid “experimenter effects” that can influence the behavior of the actual applicants used in audit studies (Heckman and Siegelman 1993). Correspondence studies have the disadvantage of not capturing actual job offers but just callbacks; however, evidence discussed below indicates that callbacks capture most of the relevant discrimination. For these reasons, our experiment is a correspondence study.

As noted above, there are challenges in applying AC methods to age discrimination. We first outline a framework for thinking about AC studies in a less problematic setting—with reference to race—and then turn to the application to age discrimination.⁷ The underlying idea is that employers try to assess whether an applicant's productivity exceeds a given threshold with sufficiently high probability, and if it does, the applicant is offered a callback or a job. Think of this productivity as a fixed characteristic that depends on observed characteristics X^1 and an unobserved characteristic X^u that is the source of the uncertainty. In the population, with B a dummy indicator for blacks, expected productivity (P) of blacks and whites differs. Supposing the former is lower,

$$E(P|B = 1) < E(P|B = 0). \quad (1)$$

In AC studies we create resumes (containing X^1) intended to be of identical quality, so that

$$E(P|B = 1, X^1) = E(P|B = 0, X^1). \quad (2)$$

A difference in selection for callbacks or job offers (denoted T), such that

$$T\{E(P|B = 1, X^1), B = 1\} < T\{E(P|B = 0, X^1), B = 0\}, \quad (3)$$

is interpreted as evidence of discrimination against blacks in hiring.⁸ If equation (2) in fact holds, we would interpret the evidence as closest to Becker's (1971) employer taste discrimination, with the productivity of blacks undervalued by discriminatory employers. But an alternative is that instead of equation (2),

$$E(P|B = 1, X^1) < E(P|B = 0, X^1). \quad (4)$$

This corresponds to statistical discrimination, because an employer assumes lower productivity for blacks despite having identical information on black and white applicants, which can still generate the difference in outcomes in equation (3). However, both types of discrimination are illegal under US law.⁹

Thus, the difference in callback or offer rates conditional on identical resumes does not directly distinguish between taste and statistical discrimination. Some AC studies try to rule out statistical discrimination

⁷ We thank Bentley MacLeod for suggesting this framework for thinking about the application of correspondence studies to age discrimination.

⁸ As in Heckman (1998), we would think of $T\{\cdot\}$ as a dichotomous outcome based on the probability—given unobservables—that productivity exceeds a threshold for hiring.

⁹ EEOC regulations state, "An employer may not base hiring decisions on stereotypes and assumptions about a person's race, color, religion, sex (including pregnancy), national origin, age (40 or older), disability or genetic information" (<http://www1.eeoc.gov/laws/practices/index.cfm?renderforprint=1>).

by including information on which employers might statistically discriminate, such as including detailed residential information that may hold socioeconomic status and criminality constant (e.g., Bertrand and Mullainathan 2004). In that sense, we can think of equation (2) as the identifying assumption for identifying taste-/animus-based discrimination from an AC study. And when this assumption does not hold, the estimate of discrimination includes statistical discrimination.

Economists may be interested not only in identifying illegal discrimination but in understanding its nature, for both scientific and policy reasons. Researchers have tried to distinguish between these two models of discrimination in AC studies, but the tests require very strong assumptions on what employers know about workers, and when they know it.¹⁰

A more fundamental critique is that even in the best-case scenario when equation (2) holds and there is no statistical discrimination, a difference in the variance of the unobservable determinants of productivity not included in the resumes can generate bias in either direction, rendering discrimination unidentified in AC studies (Heckman and Siegelman 1993; Heckman 1998)—the “Heckman critique.” Denote the standard deviations of the unobservable X^H for blacks and whites as σ_B^H and σ_W^H . In designing an AC study, a researcher chooses the resume quality (the standardized level of productivity X^L , denoted X^L). Suppose that X^L is set low relative to the resumes the employer actually sees. Then if $\sigma_B^H > \sigma_W^H$, blacks are more likely to exceed a given productivity threshold; intuitively, if the resumes are, on average, low-quality, then the low-variance group (whites, in this case) is very unlikely to have high productivity, and vice versa. In contrast, if $\sigma_B^H < \sigma_W^H$ and X^L is low, whites will be favored. Since the researcher does not know whether X^L is low or high nor whether σ_B^H is greater or less than σ_W^H , it is not possible even to sign the bias.

Neumark (2012) proposed a solution to this problem that separately identifies the relative variances of the unobservables and the difference

$$T\{E(P|B = 1, X^L), B = 1\} - T\{E(P|B = 0, X^L), B = 0\},$$

discussed more below. In our view, correspondence studies with this correction for bias from different variances of the unobservables provide the most compelling evidence of labor market discrimination available. However, particular challenges arise in studies of age discrimination, to which we turn in the next subsection.

¹⁰ For example, some studies add information to a subset of resumes and interpret a decline in the race gap as evidence of statistical discrimination based on this added information (Kaas and Manger 2012). But if we add information that is not the basis for employers’ statistical discrimination, then a null finding of no change in offer or callback rates is uninformative. Rooth (2010) pioneered a different approach to learn about the nature of discrimination in an AC study, administering the Implicit Association Test for implicit discrimination to those who reviewed the applications.

It is also useful to think about what the evidence from AC methods can tell us about the fundamental elements of the job search process. In job search models, a searching worker's probability of finding work in a given period is a positive function of the vacancy rate in the market in which he or she is searching and a negative function of the unemployment rate, which can be summarized in the job offer arrival rate.¹¹ Job search models also predict that the probability of a match is a negative function of the job searcher's reservation wage and a negative function of the length of time the person has been unemployed (assuming there is negative duration dependence, as suggested in recent work by Kroft, Lange, and Notowidigdo [2013]).

What would evidence of lower job offer or callback rates to older workers imply about the likelihood of older workers finding a match? Presumably, such evidence speaks most directly to differences in the arrival rate of job offers for older workers. The evidence does not directly capture reservation wages, although it is possible that either the lower job offer arrival rate anticipated *ex ante* or the longer spells of unemployment that result will lower reservation wages of older job applicants, which could counter the lower number of job offers they anticipate or actually receive. This latter channel highlights the fact that job applicants may respond to a lower rate of job offers that AC studies detect in ways that offset what would otherwise be a lower rate of matching to jobs. But this could come at a cost in terms of lower wages. The persistent longer unemployment durations of older workers (Neumark and Button 2014), while potentially attributable to many factors, are at least consistent with reservation wages not declining enough to reduce unemployment durations to those of younger workers.¹²

B. Challenges in Audit and Correspondence Studies of Age Discrimination

The challenges in applying AC methods to age discrimination can be couched in terms of the issues discussed in the previous section. Letting S be a dummy indicator for old ("senior"), equation (2), which we suggested could be interpreted as the key identifying assumption, becomes

$$E(P|S = 1, X^1) = E(P|S = 0, X^1). \quad (5)$$

¹¹ When vacancies are higher, a job searcher is more likely to match to a vacancy; and when the unemployment rate is higher, a job searcher is less likely to match to a vacancy because there is more competition for that vacancy.

¹² For example, older workers may search among a narrower subset of jobs with fewer physical demands or still maintain higher reservation wages owing to the same rising marginal utility of leisure that we think underlies retirement behavior.

There could be animus toward older workers—perhaps less because of “dislike” of older workers than because of negative attributes associated with them—in which case the interpretation of differential treatment based on age in equation (5) parallels other applications of AC methods. However, the assumption that equation (5) holds is more problematic with respect to age for a few reasons.

Matching on Experience

It is not clear how to match older and younger workers on resumes to make it most likely that equation (5) holds. Clearly, a young applicant cannot have the experience of a long-employed older worker. Blind application of the AC study “paradigm” would hence dictate giving older and younger applicants low levels of experience commensurate with the young applicants’ ages. However, this can make the older applicants in the study appear less qualified than the older applicants employers usually see, creating a bias toward finding evidence of discrimination against older workers. In other words, matching on a low level of experience (included in X^1) can lead to

$$E(P|S = 1, X^1) < E(P|S = 0, X^1), \quad (6)$$

which could explain the evidence in past studies.¹³

Statistical Discrimination

There may be reasons for employers to statistically discriminate against older workers. In some cases, there is evidence that may help assess the importance of statistical versus taste discrimination.

First, some physical capacities that are not conveyed on the resumes can decline with age. Existing research also points out that some capacities may increase with age, although we do not know whether the particular capacities important to employers in our study (which may also be legally permissible bona fide occupational qualifications) tend to decline

¹³ Researchers are aware of this problem. Bendick et al. (1997) had both older and younger applicants report 10 years of similar experience on their resumes. However, they had the resumes for older applicants indicate that they had been out of the labor force raising children (for female executive secretary applications) or working as a high school teacher (for male or mixed applications). Lahey (2008) studies women, for whom she argues that time out of the labor force is less likely to be a negative signal. She then includes only a 10-year job history for all applicants (in part based on conversations with three human resources professionals she cites who said that 10-year histories were the “gold standard”). However, the older resumes in either study could convey a negative signal.

or not.¹⁴ Second, and related, employers might expect older workers to have health problems, which could raise absenteeism or pose accommodation costs (Neumark, Song, and Button 2017). While absenteeism costs could matter, health insurance costs may matter less; existing legislation and regulations recognize the potential for higher health insurance costs for older workers and permit, in limited circumstances, reduced health benefits based on age.¹⁵

Third, employers might expect that older workers (our oldest group is 64–66) would be near retirement and hence firms may be less likely to want to invest in them. This source of statistical discrimination should, however, be relatively unimportant. We study low-skill jobs in which training and turnover costs are likely to be minimal. Also, younger workers are more likely to leave an employer for other jobs, and the reason for turnover is irrelevant to the employer. For example, 2015 (quarter 1) data from the Quarterly Workforce Indicators show a lower separation rate for workers aged 55–64 (9.9 percent) than for workers aged 25–34 (18.7 percent).¹⁶ Lower separation rates for older workers overall may represent behavior of high-tenure workers rather than older new hires, and the behavior of old versus young new hires is most relevant. However, other evidence indicates that older new hires have similar or lower separation rates compared to younger new hires within the first year of new employment (Choi and Fernández-Blanco 2017).¹⁷

Fourth, older applicants with experience commensurate to their age applying for the same job as a younger applicant might be viewed as less qualified or having less potential, because they have been at that job level for longer, that is, have a slower “speed of success” (Tinkham 2010); this can be interpreted as older workers searching for new jobs being more adversely selected than younger workers. On a priori grounds, it is not clear that an employer should be more interested in younger applicants with

¹⁴ Belbase, Sanzenbacher, and Gillis (2015) conclude that some abilities tend to decline with age (e.g., explosive strength, manual dexterity, memorization, and spatial orientation), while others (e.g., static strength, stamina, oral comprehension, and originality) tend not to decline, into the mid-60s and even beyond. For earlier evidence indicating the absence of decline in various work-related capacities, see Jablonski, Kunze, and Rosenblum (1990), Posner (1995), and Hellerstein, Neumark, and Troske (1999).

¹⁵ See <https://www.eeoc.gov/eeoc/publications/age.cfm>. Section 623 of the ADEA prohibits age discrimination in hiring based on benefits.

¹⁶ These rates are relative to beginning-of-quarter employment. Our youngest age range is 29–31. For details on the data, see <http://qwexplorer.ces.census.gov/#x=0&g=0>.

¹⁷ Specifically, the authors compute transitions from employment to nonemployment within 1 year of taking a new job, using data from the Survey of Income and Program Participation. They report two computations pertaining to age. First, the average age of “stayers” is 33.8 vs. 28.9 for nonstayers. Second, and better isolating differences at older ages, the authors provided us with calculations showing the overall percentage of new employees moving to nonemployment within the year. This percentage is 36.4 for both ages 25–34 and ages 55–64.

more potential, given that they have high separation rates. Moreover, our young applicants have been in the low-responsibility jobs we study for about 10 years, and older workers often take less demanding jobs on the path to retirement (bridge jobs); so looking for a low-skilled job at an older age would not necessarily convey a particularly negative signal. Finally, we have evidence from our study design, discussed below, to help rule this out.

Fifth, employers may make assumptions about skill differences across cohorts—perhaps most important that older applicants have fewer computer skills. Some of the skill variation we build into the resumes (explained below) allows us to assess whether the differences in callback rates we observe could be due to assumptions about differences in computer skills; we find that this does not account for our evidence.

Sixth, we would expect that older cohorts of women spent more time at home than younger cohorts. Given that, for older cohorts of women, these labor market interruptions would have been many years in the past, it is unlikely they could account for current differences in callback rates by age. Alternatively, employers might discriminate against younger women, expecting them to drop out to care for children. However, this would create a bias against finding discrimination against older women, which would only strengthen our findings. Finally, evidence from differences in outcomes across resume types we use, discussed below, further helps rule out a role for caring for children.

Finally, because educational attainment is higher in younger cohorts, expected ability differences between younger and older cohorts could explain why seemingly identical resumes may not satisfy equation (5). For example, on the basis of the 2015 Current Population Survey Annual Social and Economic Supplement, 21.3 percent of young women (aged 25–34) had a high school diploma, and 27.9 percent had a bachelor's degree, as their highest degree.¹⁸ For older women (aged 60–69), the first percentage was higher (32.6 percent) and the second lower (16.8 percent). For men, the differences are qualitatively similar but less pronounced. The percentages of younger men with at most a high school diploma were 28.3 percent versus 24.2 percent with a bachelor's degree; the corresponding percentages for older men were 29.0 percent and 20.1 percent. More positive selection into higher levels of education for older cohorts could generate a bias against finding discrimination against older, equally educated applicants. However, given that the education difference is considerably stronger for women than for men, yet our evidence of age discrimination is stronger for women, we do not think these cross-cohort education differences drive our results.

¹⁸ These data were downloaded from the Integrated Public Use Microdata Series—Current Population Survey (Flood et al. 2015).

In some cases, the available evidence suggests that statistical discrimination along particular dimensions may not be very likely. However, we cannot rule out statistical discrimination as a cause of lower callback rates for older workers. Moreover, in some ways it may be easier to reject evidence of statistical discrimination because it might have a refutable implication, whereas taste discrimination tends to be a residual explanation. Nonetheless, at a minimum we believe there is evidence to indicate that lower callback rates for older workers should not be automatically attributed to statistical discrimination.¹⁹

In light of this discussion, it is useful to point out both the positive and normative implications of our evidence. As described above with respect to job search, our study yields positive evidence on the role of age in job search and hence job finding behavior—as reflected in the question the title of our paper poses. The potential normative evidence pertains to the implications of the evidence for policy to counter age discrimination, which can be motivated—if there is evidence of age discrimination—by both fairness concerns and the imperative to increase employment of older workers to counter population aging. Our study can establish evidence of age discrimination, and depending on how strongly one views the arguments against interpreting such evidence as reflecting statistical discrimination as opposed to taste discrimination, it can also provide guidance as to appropriate policy responses.

The Heckman Critique Applied to Older versus Younger Workers

The problem of bias from different variances of the unobservable may be particularly salient in an age discrimination study. Denoting the standard deviations of X^H for old and young workers σ_S^H and σ_Y^H , there is a good reason to suspect that $\sigma_S^H > \sigma_Y^H$. Specifically, the human capital model (Mincer 1974) predicts that differential investments in human capital accumulate with age; recent evidence based on wage dispersion is presented in Heckman et al. (2006). And variation in investment is unlikely to be fully conveyed on the resumes. If the study design used relatively low-quality applicants, then the higher variance for older applicants generates a bias

¹⁹ A potential reason to be more skeptical of taste discrimination as a source of differences in callbacks is Becker's (1971) argument that competitive markets may eliminate employers with taste discrimination from the market. However, the claim that competition necessarily eliminates discrimination is often overstated. Even Becker clarified conditions under which employer discrimination could persist, and subsequent theoretical work further undermined the claim that competition necessarily has to undermine employer discrimination (Goldberg 1982; Black 1995).

against finding age discrimination in hiring (the opposite direction from the bias from using low experience for older applicants).

Overall Assessment of Challenges in AC Studies of Age Discrimination in Hiring

The considerations discussed in this section pose the following central question: If we send out resumes for older and younger job applicants and observe a difference in callbacks (including the innovations just discussed), are we confident that a difference in callbacks provides evidence of discrimination that is as convincing as what we would get from a study of, say, racial discrimination? Our answer to this question is “confident, but not quite as confident.” There are serious challenges to using AC methods to study age discrimination. The problems are not unique, however, to the application of these methods to age discrimination, although some could be more severe in this context. The approaches we take in this paper to adapting AC methods to study age discrimination are meant to mitigate these problems, and we think they do so substantially, with the end result being evidence on age discrimination that is compelling—even if, as always in empirical economics, we cannot definitively rule out all other explanations of the evidence.

IV. The Experimental Design

A. Basic Framework

The core analysis uses probit models for callbacks (T) as a function of dummy variables for age (M for middle-aged and S for older/senior) and observables (from the resumes) X^1 . The latent variable model is²⁰

$$T_i = \alpha + \beta M_i + \gamma S_i + X_i^1 \delta + \varepsilon_i. \quad (7)$$

The residual ε includes the unobservable worker characteristics, X^u . In this basic model, the null hypothesis of no discrimination implies that $\beta = 0$ (for middle-aged workers) and $\gamma = 0$ (for older workers). We collect data for multiple occupations and for male and female applicants.

The simple framework is modified in two ways to address the central challenges in applying AC methods to age discrimination: the treatment of experience and different variances of the unobservables. We next explain these modifications and then discuss other features of the study design, some of which help address other potential challenges to studying age discrimination in a correspondence study.

²⁰ The preceding discussion referred only to younger and older applicants, but the inclusion of middle-aged applicants follows naturally.

B. *Using Experience Commensurate with Age*

In our view, for both policy and legal reasons, the right comparison for measuring age discrimination is between younger applicants and older applicants who have experience commensurate with their age. The policy debate has focused on whether typical older workers who lose their jobs have difficulty getting hired because of their age. For example, discussions of age discrimination and long unemployment durations faced by older workers during the Great Recession did not consider hypothetical older job applicants who have not worked much and hence have experience equal to that of younger applicants; rather, they focused on actual older job applicants who do have much more experience.²¹ And evidence of age differences in callbacks using experience commensurate with age, rather than equal experience, is more consonant with legal standards for age discrimination. The ADEA makes it unlawful for employers to fail or refuse to hire an individual because of that person's age, with no mention of comparisons using equal experience. In addition, a *prima facie* case for age discrimination requires evidence that the plaintiff was qualified for the job but was not hired and the defendant did not hire the plaintiff yet continued to seek applicants with the plaintiff's qualifications, in which case the burden of proof shifts to the employer to provide a nondiscriminatory explanation (a "reasonable factor other than age," or RFOA).²² Establishing an RFOA would almost surely be easier for an older applicant with unusually low experience.

Therefore, we redefine the resume controls in equation (6) to allow experience to differ with age. Denoting by E_S and E_Y experience "commensurate" with the age of older and younger applicants, respectively, but matching on all other resume characteristics, the identifying assumption becomes

$$E(P|S = 1, X^1, E_S) = E(P|S = 0, X^1, E_Y). \quad (8)$$

Of course, one could argue that giving older workers experience commensurate with age leads to an understatement of discrimination, if in fact this is not representative of older applicants to the jobs we study (i.e., if we created unusually experienced older applicants). However, we present evidence below that such resumes are in fact more representative. Moreover, we explore the sensitivity of the results to using low versus commensurate experience for older applicants.

²¹ See, e.g., http://www.nytimes.com/2013/02/03/business/americans-closest-to-retirement-were-hardest-hit-by-recession.html?pagewanted=all&_r=0, <http://economix.blogs.nytimes.com/2011/05/06/older-workers-without-jobs-face-longest-time-out-of-work/>, and <http://www.nytimes.com/2009/04/13/us/13age.html?pagewanted=all>.

²² See <http://www.eoc.gov/laws/statutes/adea.cfm>, *McDonnell Douglas v. Green* (411 U.S. at 792–793, 1973), and http://www1.eoc.gov/laws/regulations/adea_rfoa_qa_final_rule.cfm?renderforprint=1.

We expand equation (7) to include comparisons between young applicants with typical low experience (Y_L), middle-aged or older applicants with low experience (M_L and S_L), and middle-aged or older applicants with experience commensurate with their age (M_H and S_H). If low-experience resumes send a negative signal, we expect less evidence of discrimination in comparing outcomes between young applicants and middle-aged or older applicants with commensurate experience—comparisons we believe are more relevant to assessing whether there is age discrimination in hiring.

C. Correcting for Biases from Differences in the Variance of Unobservables

We also implement the solution proposed in Neumark (2012) to address the Heckman critique. This method is based on a structural model resulting from the assumption, noted above, that the callback decision is determined by a threshold model, as employers try to assess whether an applicant's productivity likely exceeds a given threshold (as in the original critique). The solution imposes an additional identifying assumption to identify the structural parameter measuring discrimination in hiring (e.g., γ in eq. [7]), distinguishing between what is typically viewed as discrimination (stemming from taste or statistical discrimination) and different treatment stemming from differences in variances of the unobservables.

The details are provided in Neumark (2012); here we discuss the ideas underlying the approach, some potential issues, and implementation. To see the intuition behind the solution, recall that in a probit model, all that is identified is the ratio of the coefficient in the latent variable model to the standard deviation of the unobservable. Consider estimating the model only for the young and old groups of applicants. If we are willing to assume that δ in equation (7) is the same for younger and older applicants, then we can identify the ratio $\sigma_S^{\text{II}}/\sigma_Y^{\text{II}}$ from the ratios of probit coefficients for younger and older applicants.²³ Thus, information from a correspondence study on how variation in observable qualifications is related to callback outcomes can be informative about the relative variance of the unobservables, and this, in turn, solves the problem of identifying the effect of discrimination that the Heckman critique highlights.

The parameters are estimated using a heteroskedastic probit model with variance differing between younger and older applicants, but at least one element of δ —the coefficients on X^1 in the latent variable model like

²³ Think about the standard statistical discrimination framework in which an employer puts less weight on an observable signal of productivity the less reliable it is; in the limiting case of an infinite variance of the unobservable, e.g., the employer puts zero weight on the observable signal.

equation (7)—restricted to be equal. With data on multiple productivity-related characteristics in X^1 , there is an overidentifying restriction that the older/younger ratios of coefficients on any element of this vector are equal (to the same $\sigma_s^{11}/\sigma_Y^{11}$). The method therefore also requires that some applicant characteristics in X^1 affect the callback probability (since if all the effects are zero, we cannot learn about $\sigma_s^{11}/\sigma_Y^{11}$ from these coefficient estimates). AC studies typically do not try to include variables that shift the callback probability, but instead create one “type” of applicant for which there is only random variation in characteristics that are not intended to affect outcomes. However, we build this information into the study design, through assignment of random elements of a skill vector to some resumes. Note that the additional variables we add to the resume that are intended to shift the callback probability are by no means intended to fully capture the unobservables that—if their variances differ across age groups—can create bias. Regardless of what a resume says, even if it went beyond a normal resume, it would not convey reliable information on many characteristics employers might care about—such as those often characterized as “noncognitive skills” (e.g., Heckman and Rubinstein 2001). The characteristics we add are those that might be expected to shift the probability of a callback and hence provide information to identify the heteroskedastic probit model in the face of a difference in the variance of unobservables. And, indeed, as explained below, we do not add resume characteristics that are unusual for resumes; rather, we build systematic variation in conventional resume characteristics into the resumes we send out.

D. Resume Creation

The core of a correspondence study is the bank of resumes created for the artificial job applicants.²⁴ Three goals drove our resume creation. First, we wanted the resumes to be as realistic as possible, so our results are externally valid for actual job applicants. Second, we wanted valid comparisons of older and younger applicants—in part along the lines already discussed. In pursuing these goals, our overarching strategy was to use empirical evidence whenever possible in making decisions about creating the resumes, to minimize decisions that might limit the external or “comparison” validity of the results. In many cases, this empirical evidence came from a large sample of publicly available resumes we downloaded from a popular national job-hunting website. This website has massive numbers of resumes—from thousands to hundreds of thousands in large cities in the jobs we targeted. We downloaded a sample of over 25,000 resumes,

²⁴ The online appendix provides much more detail on the creation of resumes than we provide in this section. Readers interested in many of the “nuts and bolts” of the experimental design may find the online appendix especially useful.

which we then scraped for a variety of types of information that we use in our resume design decisions.²⁵ In addition, we used other data to inform many of our decisions.

Basic Parameters

Past studies have tended to use workers near age 30 as the younger group and workers near age 50 as the older group.²⁶ We include similar age ranges (29–31 and 49–51). But we focus on an older age range (64–66), which is of particular interest in light of policy efforts to induce those near retirement age to work longer. We convey age, on the resumes, via high school graduation year. This is common; in our sample of scraped resumes, 81 percent provide information on high school attendance, and of these 68 percent (56 percent of the total) include high school graduation year.

Given these age ranges, we chose common names (by sex, for first names) for the corresponding cohorts based on data from the Social Security Administration. To focus on age, we chose first and last names that were most likely to signal that the applicant was Caucasian.

AC studies almost always target a subset of jobs to which the resumes are tailored, rather than trying to write generic resumes and applying to a wide variety of jobs. They also generally target fairly low-skill jobs to make it unlikely that candidates or their work histories are known to recruiters. Among these types of jobs, we selected a subset in which there were some low-tenure older workers as well as low-tenure younger workers. Using jobs in which it is less unusual for older workers to apply increases the realism of the resumes, although it potentially excludes jobs with the strongest age discrimination. We put less weight on the second issue because the real effects of age likely preclude older workers from applying for certain jobs.²⁷ If anything, this might generate some bias in our study against finding age discrimination.

We used Current Population Survey (CPS) Tenure Supplement data to identify jobs that are common and have a relatively high representation of older workers with low tenure (5 years or less): retail salespersons, cashiers, janitors and building cleaners, and security guards, for men; and retail salespersons, cashiers, secretaries and administrative assistants, office clerks, receptionists and information clerks, and file clerks, for women. (So only sales jobs got male and female applicants.) Table 1 shows the per-

²⁵ In each occupation and city used in our study, we searched for resumes in three experience groups (3–5 years, 6–10 years, or 10+ years), extracting the greater of all resumes listed or 1,000 resumes, for a total of 25,460 resumes.

²⁶ The younger age range is chosen to capture applicants who are relatively young, but with enough experience to convey an informative job history to employers that can be compared with job histories of older applicants.

²⁷ For example, National Football League teams do not hire 60-year-old quarterbacks, no matter how good their past performance.

TABLE 1
 PERCENTAGES OF RECENT HIRES (<5 Years of Tenure) IN AGE GROUP RELATIVE
 TO ALL HIRES IN OCCUPATION, 2008 AND 2012 CPS TENURE SUPPLEMENTS

OCCUPATION	AGE-SPECIFIC RECENT HIRES/ ALL RECENT HIRES IN OCCUPATION	
	Ages 62–70	Ages 28–32
A. Males		
Average across all occupations	10.79	9.11
Janitors and building cleaners	11.91	2.64
Retail salespersons	11.31	7.55
Security guards and gaming surveillance officers	16.32	8.57
Cashiers	12.62	11.33
B. Females		
Average across all occupations	10.98	7.48
Secretaries and administrative assistants	13.18	3.39
Office clerks, general	10.70	4.34
Retail salespersons	12.35	4.65
Receptionists and information clerks	14.55	6.83
Cashiers	15.60	4.59
File clerks	16.00	5.86

percentages of “recent hires” in these occupations in older (62–70) and younger (28–32) age ranges, relative to all workers in these occupations. We combined these occupations into four groupings of jobs that best capture these occupations on the job search website we use: retail sales (retail salespersons and cashiers), administrative assistant (secretaries and administrative assistants, receptionists and information clerks, office clerks, and file clerks), janitors (janitors and building cleaners), and security guards (security guards and gaming surveillance officers).

Although our study was not meant to provide representative evidence on all older job seekers, the jobs we target are fairly important for hiring of older workers. From CPS data, for the jobs to which we send male applicants, among 62–70-year-olds, recent hires in janitor jobs are 2.16 percent of all recent hires; the corresponding figures for security and retail occupations are 1.00 percent and 2.09 percent. For female applicants, recent hires in administrative occupations are 11.57 percent of all hires of 62–70-year-olds and 3.77 percent in retail.²⁸ Thus, the jobs that we target capture appreciable shares of new hiring of older workers.²⁹ Moreover, all of the jobs we target are in the upper tier (and most are in the top decile)

²⁸ These calculations differ from those in table 1, which reports the percentages of hires in the occupation in specific age ranges.

²⁹ As additional evidence, Rutledge, Sass, and Ramos-Mercado (2019) compute the ratio of older (50–64) to prime-age (30–49) hires in detailed occupations. Retail sales and security (guards, watchmen, doorkeepers, and protective services) are in the top 10, from 1996–2012 CPS data. They also report that the jobs into which older workers tend to be hired are much narrower for less educated workers.

of jobs in terms of the proportions of older people hired. Looking at the distribution of the share of 62–70-year-olds hired recently (tenure less than 5 years) across all occupations, the percentiles for males in the occupations we use are 98.4 for janitors, 96.6 for retail salespersons, 93.3 for security guards and gaming surveillance officers, and 83.9 for cashiers. The percentiles for females are 100 for secretaries and administrative assistants, 96.8 for cashiers, 96.4 for receptionists and information clerks, 95.2 for retail salespersons, 93.6 for office clerks, general, and 85.6 for file clerks.

Examination of our scraped resumes justified the decision to tailor resumes to these specific jobs. We examined the persistence of careers within the occupations we study, using phrases that appear to cover the same jobs (e.g., for administrative assistant: administrative, receptionist, office manager, file manager, file clerk, or secretary). Across resumes, for each type of job included in the study, about one-third (between 29 and 32 percent) of all jobs were in the same job as the current job for which the person was seeking work.³⁰ We also examined the resume database for older applicants in these jobs. Figure 1 displays the age distribution of resumes in each of the four jobs we study and shows a nonnegligible representation of older resumes.³¹

Because low-skill workers have low geographic mobility (Molloy, Smith, and Wozniak 2011), we also target the resumes to jobs in specific cities, with the job and education history on each resume matching the city from which the job ad to which we apply originates. Whereas some studies use only one or two cities (Bendick et al. 1999; Lahey 2008), we chose a broader geographic scope to increase external validity. We also made sure our cities varied on other dimensions that might affect hiring of older workers including variation in state age discrimination laws (see Neumark and Song 2013) and in age composition of the population. Table 2 lists the cities in our study, classified by these characteristics.

Job Histories

To construct job histories for the resumes, we pooled job titles and descriptions from the actual resumes to create a set of entries, with only minor changes to make phrasing, grammar, and so forth consistent. We then ran-

³⁰ These are probably lower bounds, because we likely cannot classify all job titles as falling within covered jobs.

³¹ Because we were more likely to cut off the number of resumes extracted at 1,000 for lower-experience cells, there is likely a bias toward older resumes in these histograms. Offsetting this, older workers looking for jobs may be less likely to post resumes on such a website than younger workers. However, as documented in the online appendix, CPS supplement data on job search methods do not explicitly identify posting resumes on a website, but they otherwise indicate little difference between job search methods by age.

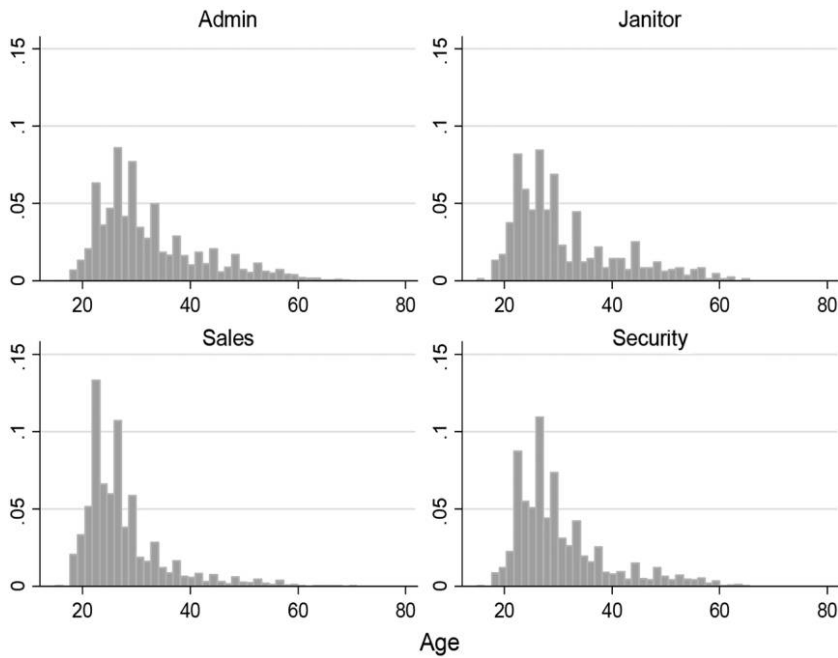


FIG. 1.—Histograms of resumes by age, resume website. Based on the sample of resumes extracted from a resume-posting website, as described in the text. The percentages of observations for ages 62–70 for each job are administrative, 0.73 percent; janitor, 0.49 percent; sales, 0.13 percent; and security, 0.34 percent. Color version available as an online enhancement.

domly combined these job entries to create job histories for each of the types of jobs in the study, using a combination of subjective judgment as to what annual job ending probability generated job histories most like those on the downloaded resumes, and secondary data from the Job Open-

TABLE 2
CITIES IN STUDY, BY PERCENTAGE OF POPULATION AGED 62+,
AND AGE DISCRIMINATION LAWS

	Stronger Laws (Larger Damages)	Weaker Laws (Smaller Damages)
Much older cities	Sarasota (34.7%, 15)	
Older cities	Miami (19.6%, 15)	Pittsburgh (21.7%, 4)
Mixed cities	New York (16.9%, 4), Boston (17.4%, 6), Chicago (15.2%, 15)	Charlotte (15.1%, 15), Phoenix (16.3%, 15), Birmingham (17.6%, 20)
Younger cities	Houston (12.1%, 15), Los Angeles (14.3%, 5)	Salt Lake City (11.6%, 15)

NOTE.—The first number in parentheses is the percentage of the population aged 62 and over, based on 2012 American Community Survey 5-year files. The second number in parentheses is the firm-size cutoff for applicability of the state age discrimination law; the ADEA cutoff is 20. Larger damages refers to more damages being available under state law compared to the ADEA. Nationally, 16.3 percent of the population is aged 62+.

ings and Labor Turnover Survey to mimic the actual monthly pattern of job changes.

We used these to create three job histories for each city and type of job. Each history goes back to 1970 with an essentially continuous work history, aside from job turnover and short spells of unemployment as explained below. To create the job histories of younger applicants, as well as older applicants reporting low experience, we simply truncate the job histories at the appropriate year. For the younger applicants and the middle-aged and older applicants with experience commensurate with age, the job history begins just after the school-leaving age. Thus, the parts of the job histories that overlap all resumes regardless of age or experience look similar across all resumes. We randomly distinguish resumes on the basis of whether applicants are currently unemployed (with 50 percent probability), with all applicants within each triplet of resumes sent to an employer (described below) as either employed (recent job end date listed as “present”) or unemployed.³²

We modified some resumes to learn about potential differences in age discrimination for workers moving into lower-skilled “bridge” jobs. This bridging was reflected on many of the resumes we examined for administrative, sales, and security jobs (but not janitors), which sometimes showed a progression from lower-level to higher-level jobs, and sometimes also (for the oldest workers) a clear downshift toward low-level jobs 8–10 years prior to the end of the job history. Thus, for the administrative, sales, and security resumes, we modified some of the resumes for the middle-aged and oldest applicants to first show rising skill levels and then bridging to lower-skill jobs: for the oldest applicants reflecting the two alternative patterns of bridging and for the middle-aged applicants only the concurrent downshift.

We already defined low- and high-experience resumes for middle-aged and older applicants (M_L , M_H , S_L , and S_H). For middle-aged workers, the notation to further classify bridge/nonbridge resumes is M_L , M_{HB} , and M_{HNB} , with B and NB denoting bridge and nonbridge. For older workers, the resumes are denoted S_L , S_{HB}^E , S_{HB}^L , and S_{HNB} ; the E and L superscripts indicate whether the transition to the bridge job occurs early (years before the current application) or late (contemporaneously with the current application). The low-experience and nonbridge resumes always keep applicants at low skill levels, while the bridge resumes have rising skill levels until the bridging occurs.³³

³² When applicants are unemployed, the resumes indicate that their last job ended in the month prior to the job application. During the course of the field experiment, every month we moved the ending date of the most recent job forward 1 month, so that unemployment durations did not lengthen during the time the experiment was in the field.

³³ The online appendix gives more information on the construction of the bridge resumes and what they look like.

We added employer names and addresses manually to each job on our final job histories to match the cities in which we were applying for jobs. We ensured that the job title and description were realistic for the employer. In addition, we used employers that were active at the time and in the region listed, relying mainly on the actual resumes, supplemented by additional research on chains. In some cases, we added as employers large public or private institutions known to be open in a particular period. The employer names were assigned randomly.

With regard to one of the central issues regarding the job histories, we calculated experience on the downloaded resumes, based on the number of years worked. It is clear that a large share of resumes of older applicants list job experience that is commensurate with their age, including jobs going all the way back to the 1970s and even the 1960s for those who were old enough, and that experience commensurate with age is more representative of the resumes we studied.³⁴ This is reflected in figure 2, which plots average experience by age: overall in panel A and by job in panel B. Both panels indicate that, on average, reported experience on the resumes rises approximately linearly with age.³⁵

Skills

To address the Heckman critique, we designate half the resume triplets sent out to be high-skilled and half to be low-skilled. We choose both general and occupation-specific skills for the jobs for which we apply, based on the downloaded resumes. For each type of high-skill resume, there are seven possible skills, five of which are chosen randomly (so that they are not perfectly collinear within a job). The five general skills that apply to all jobs are a college degree (bachelor of arts for sales, administrative assistant, and security guards, and associate of arts for janitors); fluency in Spanish as a second language; an “employee of the month” award on the most recent job; one of three volunteer activities (food bank, homeless shelter, or animal shelter); and an absence of typographical errors. Two skills are specific to each occupation: for administrative/secretarial jobs, typing 45, 50, or 55 words per minute and facility with relevant computer software (a randomly chosen mix of Quickbooks, Microsoft Office, and inventory management software); for retail/cashier jobs, Microsoft Office and programs used to monitor inventory (VendPOS, AmberPOS, and

³⁴ There was, in particular, no indication that older job applicants limited reported work experience to 10 years.

³⁵ While resumes for older workers did not always feature a complete job history indicating near-continuous work, there was no consistent way in which older workers explained gaps when they existed.

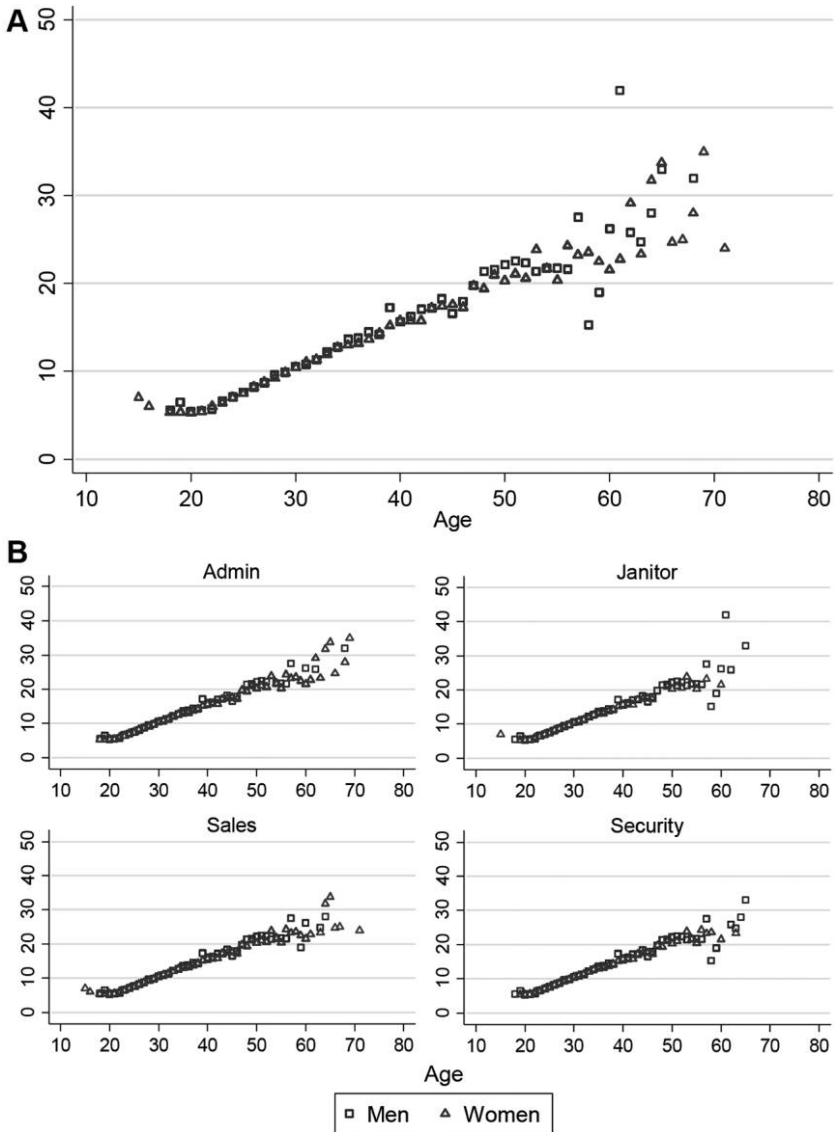


FIG. 2.—Age and experience in the resume sample. *A*, Overall averages by age. *B*, By job. Based on the sample of resumes from a resume-posting website, as described in the text. In the individual-level data, the correlation between age and computed experience is .77. If there are multiple jobs held at the same time, experience is not double-counted. Color version available as an online enhancement.

Lightspeed), and the ability to learn new programs; for security jobs, a state license and CPR training; and for janitor jobs, a certificate in using particular machines and certification in janitorial and cleaning sciences.

Additional Resume Elements

There are a number of additional resume elements that we added. Residential addresses were selected to be realistic for both older and younger applicants and the jobs to which we were applying and to avoid signaling a race other than white or other positive or negative information about the applicants. The addresses were randomly assigned with respect to age, so there is no association between socioeconomic status of the neighborhood and age of applicant.

We randomly assign one of three high schools, and colleges and universities for the high-skilled resumes, for each city, to each applicant in our triplet. We use local schools, colleges, and universities that were in operation since 1960 so that there is no possibility that an applicant attended a school that was not operational at the time. We avoided top-tier/flagship universities whenever possible.

E. Resume Triplets

After creating the final resumes, we combined them into triplets that go out in response to each job for which we apply. The resumes in a triplet are differentiated by age and, for the middle-aged and older applicants, whether they show low or commensurate experience and by the different types of bridge resumes. For age, we sent a triplet consisting of a young applicant and either (1) two older applicants, (2) two middle-aged applicants, or (3) one older applicant and one middle-aged applicant, chosen randomly with probability one-third each. For the middle-aged or older applicants, we also randomly assigned resume type (by experience and bridging): in cases 1 and 2 we sample without replacement two resumes from either the middle-aged or the older resumes, and in case 3 we sample randomly one middle-aged and one older resume. The triplets are also differentiated by sex, chosen randomly for each city and day of the month.³⁶ Other features, including resume templates, were randomly and uniquely assigned to each resume in a triplet to ensure that the applicants were dis-

³⁶ There are two exceptions. First, for sales jobs in one city (New York), a coding error in the triplet randomization generated an excessive share of resume triplets with two older applicants (with early bridge and late bridge job histories). Second, janitors do not get bridge resumes, so we always assigned a middle-age and old resume to each triplet, randomly sampling from the high- or low-experience resumes. Age is always assigned randomly with respect to other resume characteristics.

tintished from each other and that any other resume characteristics were distributed randomly across the three applicants in each triplet.³⁷

F. Applying for Jobs

We identified jobs to apply for using a common job-posting website. Research assistants read the posts multiple days per week over approximately 5 months of data collection, using a detailed protocol to select jobs for the study. Jobs had to be entry level (e.g., not managers or supervisors) in the correct occupations, and the ads could not require in-person applications, inquiries by phone, or application on an external website. The ads could not require additional documents we had not prepared (e.g., a salary history) or skills that our resumes did not have. Other exclusion criteria and quality control for the selection of ads are described in the online appendix. Once a job to apply for was identified, research assistants applied for the jobs using the randomly assigned triplet. Within each triplet, the order of applications was randomized with respect to age, with the resumes generally sent over three consecutive days. We sent triplets of applications in response to 13,371 unique job ads.

G. Sample Size

In an experiment, it is important not to continue to collect data until the estimated differences become statistically significant. The plan in our original proposal was to have three types of resumes—young, old low-experience, and old high-experience—with a target sample size of 11,520 observations calibrated to detect as significant estimated callback differentials similar to those in past studies. Commissioned reviewers of our design protocol suggested expanding to the eight different resumes used in the study, adding the three middle-aged resumes, and splitting the older, high-experience resumes into three groups based on bridging behavior. With eight groups instead of three, this implies a desired sample size of 30,720 ($11,520 \times 8/3$). However, we also decided *ex ante* to keep our research assistants for the job application process working through the end of the last academic quarter for which they were hired in which the target sample size was reached, which resulted in applications to just over 40,000 jobs.³⁸

³⁷ These characteristics included first and last names, school names, addresses, phone numbers, email address formats and domains, cover letter style, and the language describing jobs and skills. The online appendix provides resume prototypes that display all the dimensions of variation.

³⁸ We did not file a pre-analysis plan because the analysis in these studies is standard, entailing testing for differences (by age, in our case) with no controls, and verifying that results are robust to including controls (as expected, given the randomization). We did specify in advance—in a research protocol presented in seminars and for which we commissioned reviews—that we would do the analysis by occupation, pooled across occupations, and separately by sex; the last dated version of this protocol, written before any data were analyzed and while

H. Collecting Responses

Phone numbers and email addresses are included on the resumes, so responses could be received by email or by phone. All responses were forwarded to a central email account, with voice mails arriving as attachments that included the phone number of the firm calling and the phone number on the resume. We had a unique match between the email sent in applying for a job and the email response if the employer sent a direct email reply, which was typical. However, sometimes employers responded in a separate email. Phone responses are more difficult to match to applications. We purchased 360 online phone numbers, enough to assign any incoming call to a unique resume type defined by all of the characteristics by which resumes are distinguished, and used an automated voice mail message to instruct callers to include their name and their number in their message.³⁹ Members of the research team listened to each voice mail to record the response and glean information to match phone responses to specific job ads, which was made much easier since they could be linked to resumes and we knew which resume went to each ad and had other information recorded from the job application process. A similar (simpler) process was used for email responses directly to applicants. In a small number of cases (about 200), we could not match the response to any resume. These cases are dropped because without the resume match we do not know the values for the resume control variables.

Each response was coded as an unambiguous positive response (e.g., “Please call to set up an interview”), an ambiguous response (e.g., “Please return our call, we have a few additional questions”), or an unambiguous negative response (e.g., “Thank you for your interest, but the job has been filled”). To avoid having to classify the ambiguous responses subjectively, they were treated as callbacks (6.6 percent of the total coded as callbacks);⁴⁰ the negative responses were treated the same as no callbacks.

Table 3 reports the matching of responses by voice mail and email to job identifiers or resumes. Even though most responses can be matched to job IDs, we want to make use of all the data. Furthermore, no information beyond that on the resumes is used for the analysis. Thus, we make use of all of these data, and we cluster at the resume level in our statistical estimation.

Nonetheless, there may also be random influences at the level of the job ad, so it is of interest to ask how the standard errors (and hence our

data collection was ongoing, was March 19, 2015. The analysis of bias from different variances of the unobservables is potentially a sequential procedure—involving the specification of skills in the heteroskedastic probit estimation so as not to reject the overidentifying restrictions—and hence could not be further prespecified. As it turned out, we did not reject overidentifying tests, so we report initial analyses with no specification searching.

³⁹ Every resume with the same phone number has a unique first and last name, and all phone responses used a name, so we can always match to the resume.

⁴⁰ Results were very similar if we omitted the small percentage of ambiguous responses.

TABLE 3
LEVEL OF MATCHING OF CALLBACKS

	MATCHED CALLBACK RESPONSES		NO RESPONSES	TOTAL
	Job ID Match	Resume Match Only		
Voice mail	2,495	765	NA	3,260
Email	2,822	97	NA	2,919
All	5,317	862	34,044	40,223

NOTE.—There are 6,179 matched responses to 40,223 resumes that were sent out. Each response received from an employer was matched either to a unique job identifier or to the resume that was sent.

inferences) are affected by clustering at the job ad level as well. This requires multiway clustering (Cameron, Gelbach, and Miller 2011), given that the same resume could be sent to different job ads. The problem with this latter analysis is that we cannot match all responses perfectly to job ads, as table 3 shows. It is undesirable to simply discard the observations that cannot be matched to job ads, because this is not random; all the observations for which we can match to the resume but not the job ad are positive responses (14 percent of our positive responses, as shown in table 3). Moreover, we cannot drop from the sample the other applications that went to the same job ad, for which we received no response, implying that dropping only the positive responses from a triplet generates a bias toward finding no effect of age on callbacks. The potential concern is that by clustering at the level of the resume rather than the job ad, we understate the standard errors. However, we show in the online appendix that, for the subsample for which we can match to job ads, the standard errors (and hence statistical inferences) are not changed by clustering at the resume rather than the job ad level.

V. Results

A. Basic Callback Rates

Table 4 reports raw differences in callback rates for the four occupations combined and for each occupation separately (separating sales by sex). We report statistical tests of whether callback rates are independent of age for all three-way and two-way comparisons. Combining all four occupations, in panel A we find strong overall evidence of age discrimination, with callback rates statistically significantly lower by about 18 percent for middle-aged workers and about 35 percent for older workers. For the three-way and two-way tests of independence we strongly reject independence of applicant age and callback rates.⁴¹

⁴¹ This test treats the observations—which are simply each individual job application—as independent. In the regression (probit) analyses that follow, the standard errors are clustered. However, this has no material impact on the statistical conclusions.

TABLE 4
CALLBACK RATES BY AGE

		Young (29–31)	Middle (49–51)	Old (64–66)
A. Combined (<i>N</i> = 40,223)				
Callback (%)	No	81.31	84.60	87.84
	Yes	18.69	15.40	12.16
Tests of independence	Young vs. middle vs. old	Young vs. middle	Young vs. old	Middle vs. old
(<i>p</i> -value)		(.00)	(.00)	(.00)
B. Administrative (<i>N</i> = 24,350, Females)				
Callback (%)	No	85.59	89.70	92.42
	Yes	14.41	10.30	7.58
Tests of independence	Young vs. middle vs. old	Young vs. middle	Young vs. old	Middle vs. old
(<i>p</i> -value)		(.00)	(.00)	(.00)
C. Sales (<i>N</i> = 5,348, Males)				
Callback (%)	No	79.11	78.81	85.30
	Yes	20.89	21.09	14.70
Tests of independence	Young vs. middle vs. old	Young vs. middle	Young vs. old	Middle vs. old
(<i>p</i> -value)		(.00)	(.00)	(.00)
D. Sales (<i>N</i> = 4,707, Females)				
Callback (%)	No	71.32	74.13	81.57
	Yes	28.68	25.87	18.43
Tests of independence	Young vs. middle vs. old	Young vs. middle	Young vs. old	Middle vs. old
(<i>p</i> -value)		(.00)	(.00)	(.00)
E. Security (<i>N</i> = 4,138, Males)				
Callback (%)	No	75.72	78.45	78.26
	Yes	24.28	21.55	21.74
Tests of independence	Young vs. middle vs. old	Young vs. middle	Young vs. old	Middle vs. old
(<i>p</i> -value)		(.16)	(.09)	(.93)
F. Janitors (<i>N</i> = 1,680, Males)				
Callback (%)	No	67.92	66.55	74.11
	Yes	32.08	33.45	25.89
Tests of independence	Young vs. middle vs. old	Young vs. middle	Young vs. old	Middle vs. old
(<i>p</i> -value)		(.01)	(.03)	(.01)

NOTE.—The *p*-values reported for the tests of independence are from Fisher's exact test (two-sided).

For administrative jobs (panel B), for which we found by far the most eligible ads (about 61 percent of the total), the callback rate is 14.4 percent for young applicants aged 29–31. It is about 29 percent lower for 49–51-year-old applicants (10.3 percent) and about 47 percent lower for 64–66-year-old applicants (7.6 percent). Again, for every comparison we

strongly reject the hypothesis that age of applicant and callback rates are independent.

The next-largest number of applications was in sales. For males (panel C), callback rates for middle-aged versus young applicants were not very different. But the callback rate for older applicants was 30 percent lower: 14.7 percent versus 20.9 percent for young applicants. And the differences between young and old (as well as middle-aged and old) applicants are strongly statistically significant. For female sales applicants (panel D), the callback rate for middle-aged applicants is lower than for younger applicants (25.9 vs. 28.7 percent), although only marginally significant (p -value = .11). The callback differential between old and young applicants is larger (over 10 percentage points). Thus, there is evidence of stronger age discrimination for women than for men in sales.⁴²

There were far fewer ads to apply to for security (around 4,100) and janitor (around 1,700) jobs. For security jobs (panel E), the data indicate roughly equal callback rates for middle-aged and older applicants (around 21.5 percent). Both are lower than the callback rate for younger applicants (24.3 percent), with p -values of .09 and .12. For janitor jobs (panel F), the callback rate was slightly higher for middle-aged than for younger workers. But the callback rate was significantly lower for older applicants (25.9 percent), providing statistically significant evidence of discrimination against the oldest applicants.

What does our evidence imply for hiring opportunities for older workers? There are two key issues. First, our evidence directly pertains only to the occupations for which we have data and can only be suggestive about the full set of jobs to which older workers might apply. Nonetheless, three conclusions seem fair: (1) the distribution of ads to which we applied is to some extent representative of hiring opportunities for older workers, at least in this set of jobs and on the job-listing website we used; (2) the large number of administrative job ads, coupled with the sex composition of new, older hires in this occupation, suggests that our results speak more to hiring of older women than of older men; and (3) most important, at least for the jobs we study, the evidence of age discrimination in hiring is stronger for women, as it is stronger in the female than in the male jobs and for women in the mixed job (sales).

Second, there is a question of what evidence on callbacks tells us about hiring. The literal meaning of the evidence—and how it is usually interpreted (e.g., Bertrand and Mullainathan 2004)—is that a group that experiences a lower callback rate has to apply to more jobs to receive a callback. However, we believe that differences in callbacks are likely to translate quite closely into differences in job offers. A priori, an employer is more

⁴² Note that the callback rates at all ages are higher for women than for men. Similarly, Bertrand and Mullainathan (2004) did not find discrimination against women in retail.

likely to discriminate at the pre-interview (callback) stage than at the interview (job offer) stage. Because company personnel systems often create data records for those interviewed, discrimination in offering jobs to applicants may be much easier to detect than discrimination in deciding whom to call back for an interview. Indeed, there is evidence to support this presumption. The Bendick et al. (1999) audit study of age discrimination captured differences in outcomes at different stages of the application process and found that three-quarters of the overall discriminatory difference in treatment occurred at the pre-interview stage. Studies of ethnic discrimination by the International Labor Organization, discussed in Riach and Rich (2002), provide estimates of differences at the selection for interview stage and the job offer stage and find that around 90 percent of the discrimination that is detected occurs at the selection for interview stage. And Neumark (1996) finds similar evidence in an audit study of sex discrimination that also included a callback stage.

Finally, in line with the earlier discussion of what the evidence can tell us about the likelihood of older workers finding a match, the basic job search model with a constant reservation wage strategy derives the average duration of unemployment as inversely proportional to the job offer rate and directly proportional to the hazard rate for exiting unemployment (Cahuc, Carcillo, and Zylberberg 2014, 264).⁴³ Thus, all else the same, the 35 percent lower callback rate for older versus younger applicants in panel A would imply about 54 percent longer unemployment durations. Of course in reality, older applicants might adjust in a way that would lower their unemployment durations. Nonetheless, this crude estimate corresponds closely to CPS data on unemployment durations (for incomplete spells). In 2014 (when our data were collected), the ratio of average duration for 55–64-year-olds versus 25–34-year-olds was 1.48, and the ratio for 65+ versus 25–34 was 1.59.⁴⁴

B. Multivariate Estimates for Young, Middle-Aged, and Old Applicants

In table 5, we report results of probit estimates for callbacks (showing marginal effects). In each case, we first report results with controls for the city, the order in which applications were submitted, current employment/unemployment, and skills. We then add controls for an extensive set of resume features listed in the table note. The combined specifications also include controls for occupation and sex. In general, the random assignment of group membership to resumes in AC studies implies that the controls should not affect the estimated differences associated with group

⁴³ In both cases, the job offer arrival rate is multiplied by the probability that the job offer exceeds the constant reservation wage.

⁴⁴ See <https://www.bls.gov/cps/aa2014/cpsaat31.pdf>.

TABLE 5
PROBIT ESTIMATES FOR CALLBACKS BY AGE, MARGINAL EFFECTS

	COMBINED		ADMINISTRATIVE		SALES—MALES		SALES—FEMALES		SECURITY		JANITOR	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Callback estimates:												
Middle (49–51)	-.033*** (.005)	-.033*** (.005)	-.035*** (.005)	-.035*** (.005)	-.015 (.013)	-.017 (.014)	-.051*** (.015)	-.052*** (.018)	-.024 (.016)	-.023 (.018)	.015 (.027)	.025 (.029)
Old (64–66)	-.062*** (.004)	-.062*** (.005)	-.063*** (.005)	-.063*** (.005)	-.044*** (.012)	-.047*** (.014)	-.095*** (.014)	-.095*** (.017)	-.027 (.017)	-.029 (.017)	-.061*** (.027)	-.055* (.029)
Controls:												
City, order, unemployed	X	X	X	X	X	X	X	X	X	X	X	X
Skills	X	X	X	X	X	X	X	X	X	X	X	X
High-skill indicator		X										
Resume features		X										
Female		X										
Occupation		X				X		X				X
Callback rate for young (29–31)		18,69		14,41	20,89				24,28		32,08	
Observations		40,223		24,350	5,348			4,707	4,138		1,680	
Clusters		3,694		1,052	544			513	893		694	

NOTE.—Marginal effects are reported, computed as the discrete change in the probability associated with the dummy variable, evaluating other variables at their means. Standard errors are computed on the basis of clustering at the resume level. Resume features include template; email script; email format; script subject, opening, body, and signature; and file name format.

* Significantly different from zero at the 10 percent level.

** Significantly different from zero at the 5 percent level.

*** Significantly different from zero at the 1 percent level.

membership. This does not necessarily apply to AC studies of age discrimination, because of the issue of conditioning on experience (and because, in our study, the older resumes vary in terms of both experience and bridge jobs). Nonetheless, the estimates in table 5 are quite similar to those in table 4. For the larger samples (all jobs combined and administrative jobs) the estimated differentials are nearly identical to those in table 4, and for the smaller samples (the remaining jobs) the estimates are similar.⁴⁵

Table 5 echoes table 4 in pointing to unambiguous evidence of age discrimination for female job applicants, for both the middle-aged and older groups. For males the evidence is less clear; we never find statistically significant evidence of age discrimination for the middle-aged relative to the younger applicants, and the evidence for older applicants is weaker—with smaller estimated differentials in sales and only marginally significant evidence in security.

C. A Richer Characterization of Resume Types

We next turn to models incorporating all of the resume types, expanding equation (7) to be⁴⁶

$$T_i = \alpha + \beta_L M_{L,i} + \beta_{HB} M_{HB,i} + \beta_{HNB} M_{HNB,i} + \gamma_L S_{L,i} + \gamma_{HB}^E S_{HB}^E{}_{,i} + \gamma_{HB}^L S_{HB}^L{}_{,i} + \gamma_{HNB} S_{HNB,i} + X_i \delta + \varepsilon_i. \quad (9)$$

We use this model to compare the findings for older versus younger applicants depending on whether the resumes for the older applicants show the same low experience as the younger resumes or instead experience that is commensurate with age. To do this, we test the equality of age differences using low-experience and high-experience/nonbridge resumes, reporting tests for the hypotheses $\beta_{HNB} = \beta_L$ and $\gamma_{HNB} = \gamma_L$; the comparison to high-experience/nonbridge resumes is apt because these are the “conventional” high-experience resumes. Then we test whether the different kinds of bridging to lower-skill jobs matter. Specifically, we test the hypotheses $\beta_{HNB} = \beta_{HB}$, $\gamma_{HNB} = \gamma_{HB}^E$, and $\gamma_{HNB} = \gamma_{HB}^L$, corresponding to one test for middle-aged applicants and two for older applicants.

Equation (9) also touches base closely with past AC studies of age discrimination. First, it gives evidence on callback rates for applicants near 50 versus applicants near 30, similar to the age ranges considered in the

⁴⁵ Perhaps not surprisingly given the large sample and differences in parameter estimates, we strongly reject the pooling restrictions implied by combining the results for all occupations. (For this test, because the skill indicators vary by occupation, we simply use the high-skill indicator for the models for each occupation, and we estimate separate models by sex for both sales and the combined occupations, to avoid non-nested models.)

⁴⁶ The exception is for janitors, for which we do not construct bridge resumes, and hence estimate

$$T_i = \alpha + \beta_L M_{L,i} + \beta_{HNB} M_{HNB,i} + \gamma_L S_{L,i} + \gamma_{HNB} S_{HNB,i} + X_i \delta + \varepsilon_i.$$

past studies. Second, the results for the low-experience resumes (the estimates of β_L and γ_L) provide closer comparisons with past work that gave older and younger applicants the same experience (most notably, Lahey [2008]).⁴⁷ Thus, one can view the full model in equation (9) as giving the study a treatment arm that mimics past studies and treatment arms that provide evidence from using commensurate experience and from using bridge resumes.

Our estimates indicate that, with one exception, the results are insensitive to the two types of differences in the career paths of older applicants indicated on the resumes: equal experience versus experience commensurate with age, and bridge versus nonbridge resumes. For all jobs combined, all three estimates for middle-aged applicants indicate lower callback rates than for young applicants, with a small range of estimates (2.7–3.5 percentage points lower). All four estimates for older applicants are also strongly significantly different from zero, indicating lower callback rates than for young applicants regardless of resume type, with a similarly small range of estimates. Consistent with the point estimates, panel A of table 6 shows no significant differences between estimated callback rates for resumes showing low experience versus experience commensurate with age, for middle-aged or older applicants, suggesting that low experience on resumes for these applicants does not lead to spurious evidence of age discrimination. Panel B reveals no significant differences in the estimated effects of resumes showing applicants bridging to a lower-skilled job (M_{HB} or O_{HB}^L) or already having done so (O_{HB}^E).

The remaining columns of table 6 report results by occupation. Looking at applicants for administrative jobs, sales jobs (male or female), and security jobs, the conclusions from the key statistical tests are similar. In almost every case we do not reject hypotheses—whether for middle-aged and older applicants separately or considered together—that the estimated effects are equal regardless of experience or for the different bridge or nonbridge resumes.⁴⁸

⁴⁷ In addition, our retail and administrative jobs overlap with those in Lahey (2008) and Riach and Rich (2010).

⁴⁸ The one exception (in 20 tests), for the restriction $M_{HNB} = M_{HB}$ for male applicants in sales, is not in the direction of lower callbacks for bridge resumes. One issue discussed in the online appendix is the treatment of spam responses to our job applications. We retained these observations because we could not be sure we identified all spam responses, and from the point of view of a job applicant, a spam response is an unproductive response to a job application. On the other hand, we would expect the retention of these responses to lead to understating age discrimination, because the spam ads generate null responses in a manner that should be unrelated to age. Nearly all the spam ads were for administrative assistant jobs, so this issue has no bearing on results for other occupations. For the analysis discussed here, when we dropped ads we identified as spam, for middle-aged applicants the statistical evidence of equality of effects for the resumes showing low vs. commensurate experience was stronger, with a p -value of .06. However, the results for older applicants still did not indicate any difference between low- and high-experience resumes (p -value = .85).

The exception is for janitors. In table 6, we find no evidence of discrimination against older janitor applicants showing high experience but strong evidence of discrimination against older janitor applicants reporting low experience; in this case the callback rate is 9.4 percentage points lower than for young applicants, significant at the 1 percent level. And the test statistic (panel A) rejects equality of effects across the two resume types at the 5 percent level for older applicants. Thus, for this occupation, there is arguably a bias against finding age discrimination from using resumes that do not report a “full” job history.⁴⁹

The results for the bridge resumes also provide indirect evidence suggesting that the lower callback rates for older applicants do not reflect statistical discrimination. One question was whether older applicants showing longer experience at low-level jobs conveyed a negative signal (lower “speed of success”). In this case, the bridge resumes are useful because they distinguish older workers who have, instead, risen to higher job levels over much of their job history. Thus, the speed-of-success hypothesis would predict higher callback rates for older applicants with bridge resumes than with nonbridge resumes (for resumes with commensurate experience); we find no such evidence.⁵⁰ Finally, employers might assume that older women showing low experience had many career interruptions to care for children.⁵¹ We are skeptical that this would matter for low-skill jobs and resumes showing 10 or more years of recent experience anyway, and the absence of different effects, for women, of resumes with low versus commensurate experience provides consistent evidence.

D. Addressing the Heckman Critique

For the analysis of potential biases introduced by differences in the variances of unobservables, we focus on the sharpest results with arguably

⁴⁹ Although the other point estimates in table 6 sometimes point to lower callbacks for the low-experience middle-aged or older applicants, this evidence is not consistent or statistically significant. This suggests that the effect of age outweighs the effects of these experience differences, possibly because in the kinds of low-skill jobs we study there is not a lot of human capital accumulation beyond 10 or so years.

⁵⁰ One might also argue that the absence of differences between bridge and nonbridge resumes speaks to the question of whether perceived health differences drive lower callback rates for older applicants. We know that workers sometimes move to bridge jobs/partial retirement because of declining health (Johnson et al. 2009; Johnson 2014). If the kinds of job changes associated with bridge resumes are associated with health declines and declining health is an issue, we might expect older applicants with bridge resumes to experience lower callback rates than other older applicants, which we do not find. Of course the bridge resumes cannot speak to expectations about future health declines for older applicants.

⁵¹ Our resumes do not show time out for child care, which matches actual resumes. In the resumes we examined for the types of jobs for which we applied, scraping and checking the resume content revealed that it was very unusual for women’s resumes to provide explicit information on staying at home to raise children. Well below 1 percent had any such reference, and most of these had a reference to pairs of words such as “child” and “provider” that were more likely to indicate a paid job than staying at home.

TABLE 6
 PROBIT ESTIMATES FOR CALLBACKS BY AGE AND RESUME TYPE, MARGINAL EFFECTS, FULL CONTROLS

	(1)	(2)	(3)	(4)	(5)	(6)
	Combined	Administrative	Sales—Males	Sales—Females	Security	Janitor
Callback estimates:						
Middle, commensurate experience (M_{HNB})	-.033*** (.006)	-.029*** (.006)	-.029 (.017)	-.057** (.022)	-.028 (.022)	.015 (.036)
Middle, commensurate experience, bridge application (M_{HB})	-.027** (.006)	-.027*** (.006)	.020 (.019)	-.064*** (.022)	-.046* (.022)	...
Middle, experience = young (M_L)	-.035*** (.006)	-.041*** (.006)	-.037* (.018)	-.027 (.025)	.009 (.028)	.035 (.034)
Old, commensurate experience (S_{HNB})	-.058*** (.006)	-.058*** (.005)	-.048* (.023)	-.080*** (.021)	-.050* (.027)	-.017 (.037)
Old, commensurate experience, bridge application, already bridged (S_{HB}^L)	-.054*** (.006)	-.048*** (.006)	-.041** (.017)	-.100*** (.022)	-.026 (.025)	...
Old, commensurate experience, bridge application (S_{HB}^L)	-.054*** (.006)	-.055*** (.006)	-.052*** (.017)	-.074*** (.019)	-.030 (.023)	...
Old, experience = young (S_L)	-.062*** (.006)	-.057*** (.005)	-.038* (.020)	-.099*** (.022)	-.003 (.035)	-.094*** (.031)

A. Commensurate Experience = Low Experience

.80	.12	.71	.31	.23	.59
.56	.90	.73	.48	.24	.05
.82	.30	.88	.47	.25	.12

B. Bridge Resumes = Nonbridge Resumes (All High Experience)

.38	.71	.02	.80	.53	...
.57	.17	.78	.46	.47	...
.61	.69	.89	.76	.51	...
.83	.38	.85	.54	.74	...
.77	.56	.10	.73	.80	...
18.69	14.41	20.89	28.68	24.28	32.08
40,223	24,350	5,348	4,707	4,138	1,680
3,694	1,052	544	513	893	694

Tests of restrictions (p -value):

Middle-aged: $\beta_{HNB} = \beta_L$

Older: $\gamma_{HNB} = \gamma_L$

Joint, middle-aged and older

Middle-aged: $\beta_{HNB} = \beta_{HB}$

Older: $\gamma_{HNB} = \gamma_{HB}^E$

Older: $\gamma_{HNB} = \gamma_{HB}^L$

Joint, older ($\gamma_{HNB} = \gamma_{HB}^E, \gamma_{HNB} = \gamma_{HB}^L$)

Joint, middle-aged and older

Callback rate for young

Observations

Clusters

NOTE.—See the note to table 5. Control variables correspond to the second specification for each occupation (and combined occupations) in table 5 (even-numbered columns). Standard errors are computed on the basis of clustering at the resume level. There are no bridge resumes for janitors.

* Significantly different from zero at the 10 percent level.

** Significantly different from zero at the 5 percent level.

*** Significantly different from zero at the 1 percent level.

the greatest policy relevance: the differences in outcomes between young and old (near retirement age) applicants, without regard to the variation in resume types.⁵²

As a preliminary step, table 7 reports models for callbacks that add interactions between the skill indicators and the dummy variable for older applicants. Recall that correcting for biases from differences in the variance of unobservables relies on differences (if they exist) between the probit coefficient estimates on the skill-related variables for younger and older applicants. For example, if the unobserved variance is larger for older workers, then for any skill, the interaction should be the opposite sign of the main skill effect and lower the absolute value of the effect of skill for older applicants.⁵³ The interactive model reported in table 7 is not needed to implement the bias correction. That is done using the heteroskedastic probit model. However, the estimates of the interactive model provide information on the differences in the coefficients on these skill-related variables that provide the basis for identifying the heteroskedastic probit model. In addition, the overidentification test concerns the ratios of these coefficients for each skill-related variable included in the model and is computed on the basis of the interactive model.⁵⁴

The estimates in table 7 sometimes give a fairly clear indication of what to expect from the heteroskedastic probit estimates regarding the relative variances of the unobservables. In column 1, for all jobs combined, college significantly predicts hiring, and the interaction points to a smaller

⁵² Implementation of this method in other contexts (e.g., Neumark 2012; Neumark and Rich 2019) shows that, unsurprisingly, standard errors of the discrimination estimates are quite a bit larger, so that the bias-corrected estimates for differences between middle-aged and younger applicants, and between resume types, would likely be uninformative. (This was confirmed in analyses not reported here implementing the bias correction for middle-aged vs. younger applicants.)

⁵³ The correct computation of marginal effects for interactions accounts for changes in each variable in the interactions. Our interest, though, is in the signs and magnitudes of the underlying probit coefficients on the “old” and the “old-skill” interactions, of which the marginal effects reported here are approximately rescaled versions.

⁵⁴ To be clear, returning to eq. (7) and focusing only on the older vs. younger applicants, we estimate a probit model corresponding to the latent variable model

$$T_i = \alpha + \gamma S_i + X^1_{i1} \delta + X^1_{i2} \cdot S_i \lambda + \varepsilon_i,$$

i.e., adding in a full set of interactions between the dummy variable for older applicants and the skill-related variables in the model (X^1). The overidentification restriction tested is

$$\delta_1/(\delta_1 + \lambda_1) = \delta_2/(\delta_2 + \lambda_2) = \dots = \delta_K/(\delta_K + \lambda_K)$$

for the K element of X^1 , where subscripts on the parameters indicate the corresponding elements of δ and λ . In principle, it would be ideal to estimate a model with all of the age-skill interactions as a heteroskedastic probit model. However, this led to convergence problems for some samples, which is not surprising given that it can be hard to distinguish how age shifts the variance of the error term from a rich set of interactions between age and the resume elements in the linear index function. For the samples for which the model did converge, the estimates were very similar to those in table 7 (as were the results of the overidentification tests reported in table 8, with p -values within .01).

effect for older workers, consistent with a larger variance for them. The same is true for janitors, for both college and technical skills and volunteering (although volunteering has an unexpected negative main effect).⁵⁵ Similarly, for administrative jobs, three of the main skill effects have statistically significant positive effects—college, volunteer, and words per minute—the interactions are negative, and the summed effects for older workers are smaller than the main effects (in absolute value). However, for these jobs the effect of computer skills is larger for older applicants.⁵⁶

For sales workers the evidence is less clear. The skill variables have weaker effects, and for both male and female applicants the combined main effects and interactions are often of opposite sign but do not consistently point to smaller absolute effects for older applicants. For security workers, Spanish strongly predicts hiring, although the interaction suggests a larger effect for older applicants, consistent with a lower variance of the unobservable for them. For other skills the estimates point to diminished effects for older applicants, consistent with a larger variance of the unobservable for older applicants.

Finally, table 8 turns to the heteroskedastic probit estimates. The first row of panel B reports the overall effect from the heteroskedastic probit estimates, which are similar to the probit estimates (panel A).⁵⁷ Next, we report the p -value from the overidentification test that the ratios of the skill coefficients between younger and older workers are equal across all of the skills (based on the models estimated in table 7). The p -value is high across all the columns, indicating that we do not reject the overidentifying restrictions.

The third row of panel B reports the estimated ratio of the standard deviation of the unobservable for old relative to young applicants. A ratio different from one can cause bias in the estimate of discrimination. For all jobs combined, in column 1, the unobservables correction makes little difference. The estimated ratio of standard deviations is a bit greater

⁵⁵ The statistical significance of these estimates is not critical. What is critical is that the skill variables have nonzero effects on callback rates, and how strong these are will influence how informative the heteroskedastic probit estimates are. Regardless, in models without skill-age interactions, the p -value for the joint test of the skill variables was below .05 for administrative, security, and janitor jobs but in the .5–.7 range for sales jobs.

⁵⁶ Note that the positive interaction of computer skills and “old” is consistent with statistical discrimination against older applicants, with employers assuming they have lower computer skills when computer skills are not listed. However, the age gap is far larger than this estimate (0.09 vs. 0.034), so this by no means accounts for the age difference. For sales occupations—where computer skills are also one of the skills we sometimes add—there is no clear age-skill pattern. In the online appendix we present additional evidence and show that accounting for this difference in the effects of computer skills in the correction for bias from different variances of the unobservable does not alter the qualitative conclusions regarding age discrimination for older applicants either with or without computer skills.

⁵⁷ The marginal effects are calculated differently from those for the probit estimates in table 5; see the table note.

TABLE 7
 PROBIT ESTIMATES FOR CALLBACKS BY AGE, OLD VERSUS YOUNG ONLY, EFFECTS OF SKILLS AND INTERACTIONS OF OLD WITH SKILLS, MARGINAL EFFECTS

	Combined (1)	Administrative (2)	Sales—Males (3)	Sales—Females (4)	Security (5)	Janitor (6)
Old	-.071*** (.023)	-.090*** (.020)	-.062 (.085)	-.102 (.077)	-.037 (.057)	.144 (.141)
Common skills:						
Spanish	-.002 (.011)	.003 (.010)	.007 (.025)	-.038 (.037)	.081* (.045)	-.021 (.047)
Spanish × old	.009 (.017)	.016 (.019)	-.046 (.032)	.029 (.056)	.038 (.060)	-.026 (.078)
Grammar	-.014 (.009)	-.019** (.009)	-.017 (.020)	-.010 (.033)	.025 (.034)	.002 (.046)
Grammar × old	.010 (.015)	.031** (.016)	.041 (.037)	-.014 (.043)	-.019 (.045)	.011 (.077)
College	.027*** (.010)	.024** (.010)	.008 (.023)	.016 (.028)	.023 (.038)	.125** (.051)
College × old	-.016 (.013)	-.023* (.013)	-.007 (.031)	-.016 (.038)	.003 (.049)	-.058 (.072)
Employee of the month	.002 (.010)	.003 (.009)	.033 (.028)	-.018 (.029)	-.071* (.036)	-.059 (.044)
Employee of the month × old	-.000 (.014)	.002 (.014)	-.017 (.034)	.042 (.044)	.024 (.053)	.072 (.079)
Volunteer	.011 (.010)	.016** (.009)	-.027 (.024)	.022 (.032)	-.019 (.039)	-.100** (.047)
Volunteer × old	-.005 (.014)	-.014 (.013)	.053 (.040)	-.024 (.044)	-.034 (.051)	.087 (.081)

	Skill 1: Computer; Skill 2: Words per Minute	Skill 1: Computer; Skill 2: Customer Service	Skill 1: Computer; Skill 2: Customer Service	Skill 1: CPR; Skill 2: License	Skill 1: Technical Skills; Skill 2: Certificate
Occupation-specific skills:					
Skill 1	-.012 (.010)	.001 (.024)	.026 (.029)	-.064* (.034)	.132** (.065)
Skill 1 × old	.034** (.016)	.034 (.039)	-.019 (.038)	.111** (.060)	-.140* (.065)
Skill 2	.021** (.010)	.012 (.024)	.008 (.029)	.065* (.039)	-.008 (.062)
Skill 2 × old	-.024* (.012)	.008 (.036)	-.039 (.037)	-.052 (.044)	.007 (.086)
Observations	27,492	3,570	3,609	2,746	1,118
Clusters	2,522	359	386	599	462

NOTE.—See the note to table 5. Standard errors are computed on the basis of clustering at the resume level. Control variables correspond to the first specification for each occupation in table 5 (odd-numbered columns). In this table we report estimates of the probit model including interactions between resume elements and the indicator for old applicants. Because differences in the standard deviations of the unobservable would generate differences in all coefficients, all controls are interacted with age (so the main effect of “old” is for those with all of the variables interacted with age set to 0). Only the skill variable interactions (and main effects) are reported. In col. 1 we omit the controls for occupation and sex because the variance of the unobservable may vary by occupation or sex; a comparison of the probit estimates in tables 5 and 8 shows that this has no effect on the estimated effects of age (as we would expect, given randomization). Marginal effects are reported, computed as the discrete change in the probability associated with the variable, evaluating other variables at their means.

* Significantly different from zero at the 10 percent level.

** Significantly different from zero at the 5 percent level.

*** Significantly different from zero at the 1 percent level.

TABLE 8
HETEROSKEDASTIC PROBIT ESTIMATES FOR CALLBACKS BY AGE, OLD VERSUS YOUNG ONLY
(Corrects for Potential Biases from Difference in Variances of Unobservables)

	Combined (1)	Administrative (2)	Sales—Males (3)	Sales—Females (4)	Security (5)	Janitor (6)
Skill vector	All skills		All skills	All skills	All skills	All skills
Old (marginal)	-.062*** (.006)	-.067*** (.005)	-.044*** (.012)	-.093*** (.014)	-.028 (.017)	-.062*** (.028)
	A. Probit Estimates					
Old (marginal)	-.060*** (.006)	-.068*** (.006)	-.049*** (.012)	-.074*** (.015)	-.022 (.020)	-.049* (.029)
Overidentification test: ratios of coefficients on skills for old relative to young are equal (<i>p</i> -value, Wald test)	.99	.78	.88	.91	.85	1.00
Standard deviation of unobservables, old/young	1.09	.94	.84	1.44	1.16	1.66
Test: homoskedastic vs. heteroskedastic probit (<i>p</i> -value, Wald test for equal variances)	.39	.63	.28	.03	.31	.23
	B. Heteroskedastic Probit Estimates					

Old-level (marginal)	-.054**	-.005	-.161***	-.058*	-.153*
	(.028)	(.039)	(.034)	(.030)	(.082)
Old-variance (marginal)	-.014	-.043	.086**	.036	.104
	(.023)	(.029)	(.040)	(.035)	(.092)
Observations	27,492	16,449	3,609	2,746	1,118

NOTE.—Marginal effects are reported, computed as the change in the probability associated with the dummy variable, using the continuous approximation, evaluating other variables at their means. Denote the control variables in probit X_k and their coefficients ψ , and the variance of the unobservable $[\exp(Z\theta)]^2$. For a variable X_k that is also in Z , a change in X_k shifts both the variance and the level of the latent variable. Using the continuous version of the partial derivative to compute marginal effects from the heteroskedastic probit model, there is a unique decomposition of the effect of a change in a variable X_k into these two components (Cornelißen 2005). With the variables in Z arranged such that the k th element of Z is X_k , the partial derivative (Cornelißen 2005) is

$$\partial P(\text{callback})/\partial X_k = \phi(X\psi/\exp(Z\theta)) \cdot \{\psi_k/\exp(Z\theta)\} + \phi(X\psi/\exp(Z\theta)) \cdot \{(-X\psi \cdot \theta_k)/\exp(Z\theta)\}.$$

The first part of the sum is the partial derivative with respect to changes in X_k affecting only the level of the latent variable—corresponding to the counterfactual of X_k changing the valuation of the worker without changing the variance of the unobservable. The second part is the partial derivative with respect to changes via the variance of the unobservable. The table reports these two separate effects as well as the overall marginal effect, and standard errors are calculated using the delta method. (See Neumark [2012] for additional discussion. One can decompose the partial derivative from the heteroskedastic probit model based on the partial derivative for discrete variables calculated from differences in the cumulative normal distribution functions, but then the decomposition is not unique.) Because this table uses the continuous version of the partial derivative, the probit marginal effects differ slightly from those in table 5. The overidentification test is based on the estimates in table 7. Control variables correspond to the first specification for each occupation in table 5 (odd-numbered columns), except that the skill vector is as noted. Callback rates for young and old applicants are as in table 4.

* Significantly different from zero at the 10 percent level.

** Significantly different from zero at the 5 percent level.

*** Significantly different from zero at the 1 percent level.

than one (1.09), as table 7 foreshadowed. Similarly, we find evidence of a higher variance for older applicants for security and janitor jobs. For the other jobs, the evidence from table 7 was less clear, and the evidence in table 8 is indeed mixed, with evidence of a larger variance of the unobservable for sales jobs (females), but not for administrative jobs or sales jobs for males. As reported in the following row of the table, however, there is not always strong statistical evidence against the homoskedastic model with equal standard deviations, although we reject the restriction for female sales applicants.⁵⁸

The last two rows of the table decompose the heteroskedastic probit estimates. The “level” effect (labeled “old-level” in the table) is the unbiased estimate, and the “variance” effect captures the difference in the variance for older applicants.⁵⁹ For all jobs combined, the resulting estimate of discrimination (-0.080) is slightly larger in absolute value, which, together with the slightly larger variance of the unobservable for older workers, is consistent with the resumes being, on average, lower-quality than what employers observe, in which case the higher variance for older workers generates a bias against finding age discrimination.⁶⁰ We also find either similar or stronger evidence of discrimination for the two jobs to which females apply. For administrative applicants the level effect (-0.054) is close to the effect estimated from the probit model (-0.067) and, while less precise owing to the more demanding estimation, is still significant at the 10 percent level (consistent with the relative standard deviations being close to one).⁶¹ For female sales applicants, the evidence of discrimination

⁵⁸ Regardless, imposing equal variances of the unobservable can still lead to biased estimates of discrimination.

⁵⁹ This is explained further in the table note.

⁶⁰ To see this, define the following notation: Φ is the standard normal distribution function; c is the hiring threshold; β_1 is the probit coefficient on the observable characteristics on the resume (only one is used here, for simplicity); X^{1*} is the level at which the observable characteristic is set in the experiment; γ is the discrimination coefficient; and σ_s^{II} and σ_y^{II} are the standard deviations of the unobservable for senior (older) and younger workers. Then the hiring probabilities for older and younger applicants are, respectively,

$$\begin{aligned}\Pr[T(P(X^{1*}, X_s^{II})|S = 1) = 1] &= \Phi[(\beta_1 X^{1*} + \gamma - c)/\sigma_s^{II}], \\ \Pr[T(P(X^{1*}, X_y^{II})|S = 0) = 1] &= \Phi[(\beta_1 X^{1*} - c)/\sigma_y^{II}].\end{aligned}$$

If X^{1*} is standardized at a low level, then $\beta_1 X^{1*} < c$. In this case, a larger variance for older workers, $\sigma_s^{II} > \sigma_y^{II}$, can generate

$$\Phi[(\beta_1 X^{1*} + \gamma - c)/\sigma_s^{II}] > \Phi[(\beta_1 X^{1*} - c)/\sigma_y^{II}]$$

even when $\gamma = 0$ —a bias toward spurious evidence of discrimination in favor of older workers.

⁶¹ The evidence of age discrimination for administrative jobs was stronger when the spam responses were dropped, as expected. In estimates corresponding to col. 2 of table 8, when we dropped these responses the old-level (marginal) estimate was -0.081 , significant at the 1 percent level (vs. -0.054 , significant only at the 10 percent level, in the table). The old-variance (marginal) estimate remained small and statistically insignificant.

strengthens; the estimated level effect is -0.161 versus the probit estimate of -0.093 .

For male job applicants the findings are more mixed but overall still do not provide clear evidence of age discrimination. For sales jobs, the estimate of discrimination falls to near zero (-0.005) from an estimate of -0.044 . For security jobs, the evidence of discrimination strengthens, with the estimate rising from -0.028 to -0.058 (significant at the 10 percent level). For janitor jobs, the bias-corrected estimate of discrimination is much larger, -0.153 versus -0.062 , but recall from table 6 that low-experience resumes generate spurious evidence of age discrimination. Indeed, when the unobservables analysis was reestimated using only the high-experience resumes, the estimated level effect fell by half and was not statistically significant (p -value = .38).⁶² Discounting this occupation, then, the unobservables correction leaves relatively little evidence of age discrimination for men (only for security workers, and then significant only at the 10 percent level).

Thus, the heteroskedastic probit estimation that addresses the Heckman critique reinforces the evidence of age discrimination for women. The evidence for administrative applicants (all of whom are female) is reinforced, and the evidence of discrimination for female sales workers becomes considerably stronger. For men, the analysis generally weakens the evidence of age discrimination, except for security jobs.

What is gained by using the more complicated and less precise estimator that addresses the Heckman critique? First, without this correction it is not clear what we identify, and there is a good a priori reason to expect unequal variances by age. Second, in our view the less conclusive evidence for men is not an argument against the approach, especially when coupled with evidence that, for women, it delivers informative estimates; rather, it shows us that the evidence for men is in fact unclear. At the same time, there sometimes is not strong statistical evidence against the restriction to equal variances of the unobservables. On this point, we would plead "common practice." Labor economists generally implement less precise procedures intended to eliminate bias even when the data do not reject the hypothesis of no bias; put differently, we do not commonly weight variance heavily (if at all) in using a mean-squared error criterion for choosing estimators. Finally, the bias-corrected estimates are in some cases sufficiently precise that the larger uncorrected estimates of discrimination would remain significant with the standard errors resulting from the heteroskedastic probit estimation (cols. 1, 2, and 4, or three of the five cases in which the probit effects are significant). Nonetheless, for the sales applications we saw that the skill variables used to correct the bias did not have strong

⁶² In addition, the ratio of standard deviations of the unobservables fell from 1.66 to 1.33, consistent with the low-experience resumes providing less information to employers.

predictive power, so for these occupations additional evidence with skill variables that more strongly shift callbacks would be useful.

VI. Conclusions

We conducted a new correspondence study of age discrimination, adding features to address two potential sources of bias in past studies, as well as other potential challenges to interpreting differences in callback rates as evidence of age discrimination. Our correspondence study is by far the largest that has been attempted, with about 40,000 job applications, and strives to maximize the credibility of its findings by grounding the many design elements that make up such studies in empirical evidence on job applicants and the job application process. Even with our innovations, there may still be challenges in using AC methods to study age discrimination. We believe we have presented an objective discussion of all of these challenges and how our study helps address them, but readers will of course have to make their own assessment of our efforts on which to base their interpretation of the evidence.

We have a number of central findings. First, we find much stronger and more robust evidence of age discrimination against older women than against older men.⁶³ Second, we find stronger evidence of discrimination against older applicants near retirement ages (64–66) than against middle-aged workers (49–51), who have been the focus of past research. This new evidence on retirement-age workers is relevant to policy efforts to encourage older people to work longer. Third, for the most part we find that using resumes for older applicants with experience comparable to that of younger applicants (as in past studies) does not bias the evidence toward finding age discrimination. However, for one of the three jobs (janitors) to which we send male applicants—the one job that otherwise provides the strongest evidence of discrimination against older men—our evidence does suggest that using low-experience resumes generates spurious evidence of age discrimination. Fourth, we find that the evidence of age discrimination for women is robust to correcting for the bias identified by the Heckman critique, while the evidence of age discrimination for men is not robust. AC studies cannot definitively distinguish among different mechanisms of discrimination—most importantly, taste versus statistical discrimination. However, we believe our analysis and results make it less likely that some of the most plausible sources of statistical discrimination against older workers explain our findings.

⁶³ Another recent US correspondence study by Farber, Silverman, and von Wachter (2017) provides corroborating evidence of age discrimination against women. The study focuses more on the effect of unemployment duration than of age discrimination but finds evidence of lower callback rates for women aged 55–58 (compared to 35–37 and 40–42) who apply to administrative support jobs (one of the jobs in this study).

Why might older women be more likely to experience age discrimination than older men? Evidence suggests that physical appearance matters more for women (Jackson 1992) and that age detracts more from physical appearance for women than for men (e.g., Deutsch, Zalenski, and Clark 1986).⁶⁴ If older women suffer from discrimination because of both age and sex, antidiscrimination laws may be less effective than thought; because Title VII of the Civil Rights Act, which prohibits sex discrimination, is separate from the ADEA, “intersectional” claims of age discrimination against older women are difficult to bring before the courts (Song 2017).

We do not know whether these factors explain our evidence. But the stronger and more robust evidence of age discrimination against older women than against older men suggests that researchers should do more to see if this finding, itself, is robust, to understand the sources of these differences, and potentially to point out how policy efforts to extend working lives might productively focus on reducing discriminatory barriers to older women’s employment.

We believe that our experimental design and analysis substantially improve on the prior research. At the same time, we want to be clear that there are some potential limitations in our study that future research could potentially address; some are specific to studying age, and some are more general. First, it is difficult to distinguish between statistical and taste discrimination, yet the distinction is important for both understanding behavior and designing policy responses. One could imagine a follow-on experimental study that tried to focus specifically on this question, perhaps by explicitly signaling health differences or by eliciting information on selection decisions from employers. Second, the method used to correct for bias from differences in the variance of the unobservable (the Heckman critique) hinges on at least one coefficient on the skill-related resume characteristics being equal across age groups. While there is an overidentifying test, the identifying restriction is untestable, and there are reasons to expect the effects of at least some of these characteristics to differ by age.⁶⁵ Other potential solutions to the Heckman critique that do not rely on this assumption would hence be valuable. Third, given that older and younger workers differ on experience, the standard paradigm of making applicants identical on all characteristics except age (in this case) is likely inappropriate, but at the same time it is not crystal clear what

⁶⁴ This is consistent with evidence in Kuhn and Shen (2013) and Hellester, Kuhn, and Shen (2014) from job descriptions posted on internet job boards in China and Mexico on which employers often express preferences for workers based on age and sex. These papers find a “twist” in relative preference away from women with age, with greater preference for women in job descriptions seeking young workers and for men in job descriptions seeking older workers.

⁶⁵ For example, there are cohort differences (which to employers are the same as age differences) in the returns to schooling (Heckman et al. 2006; Lemieux 2006).

the alternative should be. We have designed older applicants' resumes with matched versus commensurate experience, and other experience patterns, and find that the results are generally (but not always) robust. Still, the interpretation of an age difference in callback rates is not necessarily identical to the interpretation of similar differences in audit or correspondence studies of other groups, and there may be alternative ways to address the age-experience issue. Finally, both the original Heckman critique and the solution used in this paper are based on an assumption that the callback/hiring process is based on a threshold model. It could be useful to think about how to interpret data from AC studies in the context of other models of hiring decisions, which could also, perhaps, lead to different experimental designs to learn more about the nature of discrimination.

References

- Adams, Scott J. 2002. "Passed Over for Promotion Because of Age: An Empirical Analysis of the Consequences." *J. Labor Res.* 23 (Summer): 447–61.
- Becker, Gary S. 1971. *The Economics of Discrimination*. 2nd ed. Chicago: Univ. Chicago Press.
- Belbase, Anek, Goeffrey T. Sanzenbacher, and Christopher M. Gillis. 2015. "Does Age-Related Decline in Ability Correspond with Retirement Age?" Working Paper no. 2015-24 (September), Center Retirement Res., Boston Coll.
- Bendick, Marc, Jr., Lauren E. Brown, and Kennington Wall. 1999. "No Foot in the Door: An Experimental Study of Employment Discrimination Against Older Workers." *J. Aging and Soc. Policy* 10 (4): 5–23.
- Bendick, Marc, Jr., Charles W. Jackson, and J. Horacio Romero. 1997. "Employment Discrimination Against Older Workers: An Experimental Study of Hiring Practices." *J. Aging and Soc. Policy* 8 (4): 25–46.
- Bertrand, Marianne, and Sendhil Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *A.E.R.* 94 (September): 991–1013.
- Black, Daniel A. 1995. "Discrimination in an Equilibrium Search Model." *J. Labor Econ.* 13 (April): 309–34.
- Cahuc, Pierre, Stéphane Carcillo, and André Zylberberg. 2014. *Labor Economics*. 2nd ed. Cambridge, MA: MIT Press.
- Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller. 2011. "Robust Inference with Multiway Clustering." *J. Bus. and Econ. Statis.* 29 (2): 238–49.
- Choi, Sekyu, and Javier Fernández-Blanco. 2017. "A Note on US Worker Turnover." *Oxford Bull. Econ. and Statis.* 79 (2): 276–89.
- Cornelißen, Thomas. 2005. "Standard Errors of Marginal Effects in the Heteroskedastic Probit Model." Paper no. 320. Diskussionspapiere des Fachbereichs Wirtschaftswissenschaften, Universität Hannover.
- Deutsch, Francine M., Carla M. Zalenski, and Mary E. Clark. 1986. "Is There a Double Standard of Aging?" *J. Appl. Soc. Psychology* 16 (December): 771–85.
- Farber, Henry S., Dan Silverman, and Till von Wachter. 2017. "Factors Determining Callbacks to Job Applications by the Unemployed: An Audit Study." *RSF: Russell Sage Found. J. Soc. Sci.* 3 (3): 168–201.
- Fix, Michael, and Raymond Struyk, eds. 1993. *Clear and Convincing Evidence: Measurement of Discrimination in America*. Washington, DC: Urban Inst. Press.

- Flood, Sarah, Miriam King, Steven Ruggles, and J. Robert Warren. 2015. *Integrated Public Use Microdata Series, Current Population Survey: Version 4.0*. Machine-readable database. Minneapolis: Univ. Minnesota.
- Goldberg, Matthew S. 1982. "Discrimination, Nepotism, and Long-Run Wage Differentials." *Q.J.E.* 97 (May): 307–19.
- Gordon, Randall A., and Richard D. Arvey. 2004. "Age Bias in Laboratory and Field Settings: A Meta-Analytic Investigation." *J. Appl. Soc. Psychology* 34 (March): 468–92.
- Heckman, James J. 1998. "Detecting Discrimination." *J. Econ. Perspectives* 12 (Spring): 101–16.
- Heckman, James J., Lance L. Lochner, and Petra E. Todd. 2006. "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond." In *Handbook of the Economics of Education*, vol. 1, edited by Eric Hanushek and Finis Welch, chap. 7. Amsterdam: Elsevier.
- Heckman, James J., and Yona Rubinstein. 2001. "The Importance of Noncognitive Skills: Lessons from the GED Testing Program." *A.E.R.* 91 (May): 145–49.
- Heckman, James J., and Peter Siegelman. 1993. "The Urban Institute Audit Studies: Their Methods and Findings." In *Clear and Convincing Evidence: Measurement of Discrimination in America*, edited by Michael Fix and Raymond Struyk, 187–258. Washington, DC: Urban Inst. Press.
- Hellerstein, Judith, David Neumark, and Kenneth Troske. 1999. "Wages, Productivity, and Worker Characteristics." *J. Labor Econ.* 17 (July): 409–46.
- Hellester, Miguel Delgado, Peter Kuhn, and Kailing Shen. 2014. "Employers' Age and Gender Preferences: Direct Evidence from Four Job Boards." Manuscript, Dept. Econ., Univ. California, Santa Barbara.
- Holzer, Harry J., and David Neumark. 2000. "What Does Affirmative Action Do?" *Indus. and Labor Relations Rev.* 53 (January): 240–71.
- Jablonski, M., K. Kunze, and L. Rosenblum. 1990. "Productivity, Age, and Labor Composition Changes in the U.S. Work Force." In *The Aging of the American Work Force*, edited by I. Bluestone, R. Montgomery, and J. D. Owen, 304–38. Detroit: Wayne State Univ. Press.
- Jackson, Linda A. 1992. *Physical Appearance and Gender: Sociological and Sociocultural Perspectives*. Albany: State Univ. New York Press.
- Johnson, Richard W. 2014. "Later Life Job Changes Before and After the Great Recession." Report, AARP, Washington, DC.
- Johnson, Richard W., Janette Kawachi, and Eric K. Lewis. 2009. "Older Workers on the Move: Recareering in Later Life." Public Policy Inst., AARP, Washington, DC.
- Johnson, Richard W., and David Neumark. 1997. "Age Discrimination, Job Separations, and Employment Status of Older Workers." *J. Human Resources* 32 (Autumn): 779–811.
- Kaas, Leo, and Christian Manger. 2012. "Ethnic Discrimination in Germany's Labour Market: A Field Experiment." *German Econ. Rev.* 13 (February): 1–20.
- Kite, Mary E., Gary D. Stockdale, Bernard E. Whitley Jr., and Blair T. Johnson. 2005. "Attitudes Toward Younger and Older Adults: An Updated Meta-Analytic Review." *J. Soc. Issues* 61 (June): 241–66.
- Kroft, Kory, Fabian Lange, and Matthew Notowidigdo. 2013. "Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment." *Q.J.E.* 128 (August): 1123–67.
- Kuhn, Peter, and Kailing Shen. 2013. "Gender Discrimination in Job Ads: Evidence from China." *Q.J.E.* 128 (February): 287–336.
- Lahey, Joanna. 2008. "Age, Women, and Hiring: An Experimental Study." *J. Human Resources* 43 (Winter): 30–56.

- Lemieux, Thomas. 2006. "The 'Mincer Equation' Thirty Years after *Schooling, Experience, and Earnings*." In *Jacob Mincer: A Pioneer of Modern Labor Economics*, edited by S. Grossbard, 127–45. New York: Springer.
- Maestas, Nicole. 2010. "Back to Work: Expectations and Realizations of Work after Retirement." *J. Human Resources* 45 (Summer): 718–48.
- Miller, Conrad. 2017. "The Persistent Effect of Temporary Affirmative Action." *American Econ. J.: Appl. Econ.* 9 (3): 152–90.
- Mincer, Jacob. 1974. *Schooling, Experience, and Earnings*. New York: Columbia Univ. Press.
- Molloy, Raven, Christopher L. Smith, and Abigail Wozniak. 2011. "Internal Migration in the United States." *J. Econ. Perspectives* 25 (Summer): 173–96.
- Neumark, David. 1996. "Sex Discrimination in Restaurant Hiring: An Audit Study." *Q.J.E.* 111 (August): 915–41.
- . 2012. "Detecting Evidence of Discrimination in Audit and Correspondence Studies." *J. Human Resources* 47 (Fall): 1128–57.
- Neumark, David, and Patrick Button. 2014. "Did Age Discrimination Protections Help Older Workers Weather the Great Recession?" *J. Policy Analysis and Management* 33 (Summer): 566–601.
- Neumark, David, and Judith Rich. 2019. "Do Field Experiments on Labor and Housing Markets Overstate Discrimination? A Re-examination of the Evidence." *ILR Rev.* 72 (1): 223–52.
- Neumark, David, and Joanne Song. 2013. "Do Stronger Age Discrimination Laws Make Social Security Reforms More Effective?" *J. Public Econ.* 108 (December): 1–16.
- Neumark, David, Joanne Song, and Patrick Button. 2017. "Does Protecting Older Workers from Discrimination Make It Harder to Get Hired? Evidence from Disability Discrimination Laws." *Res. on Aging* 39 (1): 29–63.
- Posner, Richard A. 1995. *Aging and Old Age*. Chicago: Univ. Chicago Press.
- Riach, Peter A., and Judith Rich. 2002. "Field Experiments of Discrimination in the Market Place." *Econ. J.* 112 (November): F480–F518.
- . 2006. "An Experimental Investigation of Age Discrimination in the French Labour Market." IZA Discussion Paper no. 2522, Inst. Study Labor, Bonn.
- . 2010. "An Experimental Investigation of Age Discrimination in the English Labor Market." *Ann. Econ. and Statis.* 99/100 (July/December): 169–85.
- Rooth, Dan-Olof. 2010. "Automatic Associations and Discrimination in Hiring: Real World Evidence." *Labour Econ.* 17 (June): 523–34.
- Rutledge, Matthew S., Steven A. Sass, and Jorge D. Ramos-Mercado. 2017. "How Does Occupational Access for Older Workers Differ by Education?" *J. Labor Res.* 38:283–305.
- Song, Joanne. 2017. "Falling between the Cracks: Discrimination Laws and Older Women." Manuscript, Dept. Econ., SUNY–Buffalo.
- Tinkham, Thomas. 2010. "The Uses and Misuses of Statistical Proof in Age Discrimination Claims." Legal Studies Research Paper Series, no. 2010-20, William Mitchell Coll. Law, St. Paul, MN.
- US Department of Labor. 1965. *The Older American Worker*. Washington, DC: US Government Printing Office.
- US Government Accounting Office. 1999. *Social Security Reform: Implications of Raising the Retirement Age*. Letter Report, August 7. <https://www.gpo.gov/fdsys/pkg/GAOREPORTS-HEHS-99-112/html/GAOREPORTS-HEHS-99-112.htm>.