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

September 2016

David Neumark
Chancellor's Professor of Economics and
Director, Center for Economics & Public Policy
University of California, Irvine

Ian Burn
Ph.D. candidate in Economics
University of California, Irvine

Patrick Button
Assistant Professor of Economics
Tulane University

Abstract

We design and implement a large-scale field experiment – a resume correspondence study – to address a number of potential limitations of existing field experiments testing for age discrimination, which may bias their results. One limitation that may bias these studies towards finding discrimination is the practice of giving older and younger applicants similar experience in the job to which they are applying, making them "otherwise comparable." The second limitation arises because greater unobserved differences in human capital investment of older applicants may bias existing field experiments against finding age discrimination. We also study ages closer to retirement than in past studies, and use a richer set of job profiles for older workers to test for differences associated with transitions to less demanding jobs ("bridge jobs") at older ages. Based on 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 there is considerably less evidence of age discrimination against men after correcting for the potential biases this study addresses.

* We received generous support from the Alfred P. Sloan Foundation, and helpful comments from seminar participants at Boston College, Georgia State University, IZA, Marquette University, the New School, Ohio State University, the Sloan Foundation, Stanford University, Tulane University, UCI, the University of Tokyo, the University of San Francisco, the University of Wisconsin, Yale Law School, and the NBER Law and Economics Summer 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 McLeod, Matthew Notowidigdo, and anonymous reviewers for very helpful comments. This study was approved by UC Irvine's Institutional Review Board (HS#2013-9942).

1. Introduction

Knowing whether age discrimination deters employment of older workers is critical in light of the challenge posed by population aging to increase the employment of older individuals. The effectiveness of policy reforms addressing this challenge, which focus on labor supply incentives, may be substantially diminished by age discrimination. 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.

The existing research generally concludes that older workers suffer from age discrimination. Most notably, field experiments (audit or correspondence studies) nearly uniformly point to age discrimination in hiring (Bendick et al., 1997; Bendick et al., 1999; Riach and Rich, 2006, 2010; Lahey, 2008).[1] Age discrimination in hiring may be 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 et al., 2009), or return to work after a period of retirement (Maestas, 2010).

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 et al., 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

[1] A summary table is available in the on-line appendix, at (url to be added).

experiment – a resume correspondence study – 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. Based on 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.

2. Past Non-Experimental Research on Age Discrimination

Prior to the passage of the Age Discrimination in Employment Act (ADEA), explicit age restrictions in hiring ads were frequent (U.S. Department of Labor, 1965). Workers in their 50s and early 60s have long had lengthier unemployment durations than many other age groups,[2] although these longer durations could reflect higher reservation wages of unemployed older workers rather than discrimination.[3] There is evidence of negative stereotypes 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 expectation of working past 62 or 65 (Johnson and Neumark, 1997; Adams, 2002).

This evidence based on observational data is generally consistent with age discrimination, but is hardly decisive. Thus, like in research on discrimination along other dimensions, researchers have turned to audit or correspondence 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.

[2] For recent evidence, see http://www.bls.gov/cps/cpsaat31.pdf (viewed September 5, 2016).

[3] Age discrimination leading to longer unemployment durations of older workers could entail considerable lost output and welfare. However, longer unemployment durations could understate such losses. For workers near retirement age – on whom we focus in this study – workers encountering age discrimination may leave the workforce, perhaps in part because they can claim Social Security benefits earlier (consistent with Adams, 2002).

3. Applying Experimental Methods to Studying Age Discrimination in Hiring

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). 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.[4] 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^I, and an unobserved characteristic X^{II} – the source of the uncertainty. In the population, with B a dummy indicator for blacks, expected productivity (P) of blacks and whites differs:

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

In AC studies we create resumes (containing X^I) of identical quality, so that

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

[4] We thank Bentley McLeod for suggesting this framework for thinking about the application of correspondence studies to age discrimination.

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

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

is interpreted as evidence of discrimination against blacks in hiring.[5] If equation (2) holds, we would interpret the evidence as closest to Becker's (1971) employer taste discrimination, with the productivity of blacks undervalued by discriminatory employers. If instead equation (3) holds because

(4) $E(P|B=1,X^I) < E(P|B=0,X^I)$,

there is statistical discrimination, because an employer assumes lower productivity for blacks despite having identical information on black and white applicants. However, both are illegal under U.S. law.[6]

The difference in callback or offer rates conditional on identical resumes does not directly distinguish between taste and statistical discrimination. However, some AC studies try to include 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), to try to rule out statistical discrimination. In that case, 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 as well.

Economists may be interested not only in identifying illegal discrimination, but in understanding its nature. For example, the most natural policy response to taste discrimination à la Becker is to raise the cost to the employer of engaging in discriminatory behavior, effectively restoring equal prices for labor from equally productive groups. In contrast, policy interventions that increase or make more reliable information about workers may reduce statistical discrimination. Although researchers have tried to distinguish between these two models of discrimination in AC studies, the tests require very strong assumptions on what employers know about workers, and when they know it.[7]

---

[5] As in Heckman (1998), we would think of T{.} as a dichotomous outcome based on the probability – given unobservables – that productivity exceeds a threshold for hiring.

[6] 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, viewed September 27, 2015).

[7] For example, some studies add information to a subset of resumes, and interpret a decline in the race gap as evidence of statistical discrimination on the basis of this added information (Kaas and Manger, 2011). But if we add

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^{II} for blacks and whites as σ_B^{II} and σ_W^{II}. In designing an AC study, a researcher chooses the resume quality (the standardized level of productivity X^I, denoted X^{I*}). Suppose that X^{I*} is set low relative to the resumes the employer actually sees. Then if $\sigma_B^{II} > \sigma_W^{II}$, 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, $\sigma_B^{II} < \sigma_W^{II}$, then if X^{I*} is low whites will be favored. Since the researcher does not know whether X^{I*} is low or high, nor whether σ_B^{II} is greater than or less than σ_W^{II}, 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 $T\{E(P|B=1,X^I),B=1\} - T\{E(P|B=0,X^I),B=0\}$, discussed more below. In our view, correspondence studies, especially 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 next.

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

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

There could be animus towards older workers – perhaps less because of "dislike" of older workers

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 (IAT) for implicit discrimination to those who reviewed the applications.

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 couple of 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 level 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 towards finding evidence of discrimination against older workers. In other words, matching on a low level of experience (included in X^l) can lead to

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

which could explain the evidence in past studies.[8]

Statistical Discrimination

It may be particularly difficult to design an AC study of age discrimination that precludes statistical discrimination, which, especially for those more interested in the behavioral source of age discrimination (and not just whether there is illegal discrimination), could limit interest in the findings. First, there may simply be more reasons for employers to statistically discrimination against older workers, although their importance can be overstated. First, some physical capacities that are not conveyed on the resumes can decline with age. However, existing research also points out that some capacities may increase with age.[9]

[8] 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 the female executive secretary applications), or working as a high school teacher (for the 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 10-year job history for all applicants (in part based on conversations with three human resources professionals she cites who said 10-year histories were the "gold standard"). However, the older resumes in either study could convey a negative signal.

[9] Belbase et al. (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

Second, employers might expect older workers to have health problems, which could raise absenteeism or pose accommodation costs. While absenteeism costs could matter, health insurance costs are less likely to. 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.[10] As discussed below, the design of our experiment provides some evidence on whether perceived or expected health declines can explain lower callbacks to older applicants, and suggests that they do not.

Third, employers might expect that older workers (our highest age range is 64-66) would be near retirement, and hence be less likely to want to invest in them. This source of statistical discrimination should, however, be relatively unimportant. We study low-skill jobs where training costs and hence turnover 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, in 2015:Q1 data from the Quarterly Workforce Indicators, the separation rate (relative to beginning-of-quarter employment) was 9.9% for workers aged 55-64, and 18.7% for workers aged 25-34 (our youngest age range is 29-31).[11]

Fourth, an older applicant 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 he or she has been at that job level for longer – i.e., has a slower "speed of success" (Tinkham, 2010). On a priori grounds, it is not clear that an employer should be more interested in younger applicants with 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

originality) tend not to decline, into the mid-60s and even beyond. For earlier evidence indicating absence of decline in various work-related capacities, see Posner (1995), Jablonski et al. (1990), and Hellerstein et al. (1999).

[10] See https://www.eeoc.gov/eeoc/publications/age.cfm (viewed August 11, 2016). Section 623 of the ADEA prohibits age discrimination in hiring based on benefits.

[11] See http://qwiexplorer.ces.census.gov/#x=0&g=0 (viewed August 11, 2016).

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, based on the 2015 Current Population Survey Annual Social and Economic Supplement,[12] 21.3% of young women (aged 25-34) had a high school diploma, and 27.9% had a Bachelor's degree, as their highest degree. For older women (aged 60-69), the first percentage was higher (32.6%) and the second lower (16.8%). For men, the differences are qualitatively similar, but less pronounced. The percentages of younger men with at most a high school degree were 28.3%, vs. 24.2% with a Bachelor's degree; the corresponding percentages for older men were 29.0% and 20.1%. Because there is likely more positive selection into higher levels of education for older cohorts, this generates a bias against finding age discrimination. 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.

Our intent in this discussion is not to rule out statistical discrimination as a cause of lower callback rates for older workers. Rather, we highlight potential sources of this kind of discrimination, but also

[12] This data were downloaded from IPUMS-CPS (Flood et al., 2015).

present arguments and evidence why the lower callback rates for older workers we find should by no means be automatically attributed to statistical discrimination.

The Heckman Critique Applied to Older vs. 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^{II} for old and young workers σ_S^{II} and σ_Y^{II}, there is a good reason to suspect that $\sigma_S^{II} > \sigma_Y^{II}$. 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). This variation in investment, which is unlikely to be fully conveyed on the resumes, could lead to σ_S^{II} exceeding σ_Y^{II}. If the study design used relatively low-quality applicants,[13] then the high variance group for older applicants generates a bias 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 we use to try to get at some of the issues 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 race 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. Yet the problems are not unique to the application of these methods to age discrimination, although some are clearly worse and potentially more fundamental in studying age discrimination. 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.

[13] As one example in a different context, Bertrand and Mullainathan (2004, p. 995) claim that they tried to avoid over-qualified applicants who employers might not bother trying to hire.

4. The Experimental Design

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^I. The latent variable model is

(7) $T_i = \alpha + \beta M_i + \gamma S_i + X^I_{iI}\delta + \varepsilon_i$.[14]

The residual ε includes the unobservable worker characteristics, X^{II}. 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, and estimate the model for the separate occupations and then pooled, and separately for men and women.

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 also help address other potential challenges to studying age discrimination in a correspondence study.

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 equal experience to younger applicants; rather, they focused on actual older job applicants who *do* have much more experience.[15] And evidence of age differences in callbacks using experience commensurate with age, rather than equal experience, is

[14] The preceding discussion referred only to younger and older applicants, but the inclusion of middle-aged applicants as well follows naturally.

[15] 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 (viewed March 5, 2013); http://economix.blogs.nytimes.com/2011/05/06/older-workers-without-jobs-face-longest-time-out-of-work/ (viewed March 5, 2013); and http://www.nytimes.com/2009/04/13/us/13age.html?pagewanted=all (viewed March 5, 2013).

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 (as "reasonable factor other than age," or RFOA).[16] Establishing an RFOA would almost surely be easier in older applicant with unusually low experience to younger applicants.

We therefore redefine X^I 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

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

However, we also explore the sensitivity of the results to low versus commensurate experience for older applicants. 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 or more relevant to assessing whether there is age discrimination in hiring.

Correcting for Biases from Differences in the Variance of Unobservables

We also implement the solution proposed in Neumark (2012) to address the Heckman critique. The details are provided in that paper; 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. To focus on essentials, consider estimating the model only for the young and old groups of

[16] See http://www.eeoc.gov/laws/statutes/adea.cfm, *McDonnell Douglas v. Green* (411 U.S. at 792-793, 1973), and http://www1.eeoc.gov//laws/regulations/adea_rfoa_qa_final_rule.cfm?renderforprint=1 (viewed August 4, 2014).

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 $\sigma_S^{II}/\sigma_Y^{II}$ from the ratios of probit coefficients for younger and older applicants.[17] 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, solving 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^I in the latent variable model in equation (7) – restricted to be equal. With data on multiple productivity-related characteristics in X^I, there is an overidentifying restriction that the younger/older ratios of coefficients on any element of this vector are equal (to the same $\sigma_S^{II}/\sigma_Y^{II}$). The method therefore also requires that some applicant characteristics in X^I affect the callback probability (since if all the effects are zero we cannot learn about $\sigma_S^{II}/\sigma_Y^{II}$ 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.

Resume Creation[18]

The core of a correspondence study is the bank of resumes created for the artificial job applicants, since these resumes constitute the study data. 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 over-arching strategy was to use empirical evidence whenever possible in

[17] Think about the standard statistical discrimination framework where an employer puts less weight on an observable signal of productivity the more reliable it is; in the limiting case of an infinite variance of the unobservable, for example, the employer puts zero weight on the observable signal.

[18] The on-line 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 on-line appendix especially useful.

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, which we then scraped for a variety of types of information that we use in our resume design decisions.[19] In addition, we used public-use data to inform many of our decisions.

Basic Parameters

Past studies have tended to use workers near age 30 as the young group, and workers near age 50 as the older group.[20] We include similar age ranges (29-31 and 49-51). But the focus of our study is 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% provide information on high school attendance, and of these 68% (56% 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 those doing the recruiting. 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 where it is less unusual for older workers to apply increases the realism of the resumes, although it potentially excludes jobs with the strongest age

[19] 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.
[20] 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.

discrimination. We put less weight on the second issue because the real effects of age likely preclude older workers from applying to certain jobs (in contrast, say, to what we would do in a study of race discrimination).[21] 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 (five 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 will get male and female applicants.) Table 1 shows the 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% of all recent hires; the corresponding figures for security and retail occupations are 1.00% and 2.09%. For female applicants, recent hires in administrative occupations are 11.57% of all hires of 62-70 year-olds, and 3.77% in retail.[22] Thus, the jobs that we target (especially administrative jobs) capture appreciable shares of new hiring of older workers (and of course higher shares of hiring into low-skill jobs that could plausibly be candidates for our study).[23] Moreover, all of the jobs we target are in the upper tier (and most are in the top decile) of

[21] For example, NFL teams do not hire 60 year-old quarterbacks, no matter how good their past performance.
[22] These calculations differ from Table 1, which reports the percentages of hires in the occupation in specific age ranges.
[23] As additional evidence, Rutledge et al. (2016) compute the ratio of older (50-64) to prime age (30-49) hires in detailed occupations. Retail sales, and security (guards, watchmen, doorkeepers, and protectives services) are in the

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 five years) across all CPS 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%) of all jobs were in the same job as the current job for which the person was seeking work.[24] 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 non-negligible representation of older resumes.[25]

Because low-skill workers have low geographic mobility (Molloy et al., 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 (Lahey, 2008; Bendick et al., 1999), 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

top 10, based on 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.

[24] These are likely lower bounds, because we likely cannot classify all job titles as falling within covered jobs.

[25] 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 towards 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 on-line appendix, CPS 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.

To construct job histories for the resumes, we pool job titles and descriptions from the actual resumes to create a set of entries, with only minor changes to make phrasing, grammar, etc., consistent. We then randomly combined these job entries to create job histories for each of the types of jobs in the study, using a combination of subjective judgement as to what annual job ending probability generated job histories most like those on the downloaded resumes, and using secondary data from the Job Openings and Labor Turnover Survey (JOLTS) 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 based on whether applicants are currently unemployed (with 50% 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.[26]

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 towards 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 skills levels, and then bridging to lower-skill jobs – for the oldest applicants reflect the two alternative patterns of bridging, and for the middle-aged applicants only the

[26] 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 one month, so that unemployment durations did not lengthen during the time the experiment was in the field.

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/non-bridge resumes is M_L, M_{HB}, and M_{HNB}, with B and NB denoting bridge and non-bridge. 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 non-bridge resumes always keep applicants at low skill levels, while the bridge resumes have rising skill levels until the bridging occurs.[27]

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 was 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 large public or private institutions known to be open in a particular period as employers. Among possible employer names valid for each jobs, 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; there was, in particular, no indication that older job applicants limited reported work experience to 10 years. This is reflected in Figure 2, which plots average experience by age – overall in the top panel, and by job in the bottom panel. Both panels indicate that, on average, reported experience on the resumes rises approximately linearly with age.[28]

Skills

To address the Heckman critique, we designate half the resume triplets sent out to be high-skilled

[27] The on-line appendix gives more information on the construction of the bridge resumes and what they look like.
[28] While resumes for older workers did not always feature a complete job history indicating near-continuous work, there was no consistent way that older workers explained gaps when they existed.

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 (B.A. 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 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, did not signal a race other than white, and were not likely to send an unusual signal (positive or negative) about the applicant. 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.

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 send 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, are chosen randomly for each city and day of the month.[29] Other features, including resume templates, were randomly and uniquely assigned to each resume in a triplet to ensure that the applicants were distinguished from each other, and that any other resume characteristics were distributed randomly across the three applicants in each triplet.[30]

Applying for Jobs

We identify jobs to apply for using a common job-posting website. Research assistants read the posts multiple days per week over approximately five 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 fields, and the ads could not require in-person applications, inquiries by phone, or applying on an external website. The ads could not require additional documents we had not prepared (e.g., a salary history, etc.), or skills that our resumes did not have. Other exclusion criteria and quality-control for the selection of ads are described in the on-line 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.

Sample Size

[29] 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 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.

[30] 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 on-line appendix provides resume prototypes that display all the dimensions of variation.

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 x 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 quarter for which they were hired in which the target sample size was reached, which resulted in applications to just over 40,000 jobs.[31]

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 voicemails arriving as attachments that included the phone number of the firm calling and the phone number on the resume. The job-search website we used generates a unique match between email sent in applying for a job and email responses, although sometimes employers responded by email directly to the fictitious applicant. Phone responses are more difficult to match to applications. We purchased 360 on-line phone numbers, enough to assign any incoming call to a unique resume type defined by all of the characteristics by which resumes are distinguished,[32] and automated voicemail message instructed firms to include their name and their number in their message. Members of the research team listened to each voicemail to record the

[31] 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 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 heteroscedastic probit estimation so as not to reject the overidentifying restrictions – and hence could not be further pre-specified. As it turned out, we did not reject overidentifying tests, so we report initial analyses with no specification searching.

[32] 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.

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 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 age of the applicant.

Table 3 reports the matching of responses by voicemail and email to job ids 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.

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% of the total coded as positive); the negative responses were treated the same as no callbacks.

5. Results

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% for middle-aged workers, and about 35% for older workers. For the three-way and two-way tests of independence we strongly reject independence of applicant age and callback rates.[33]

For administrative jobs (Panel B), for which we found by far the most eligible ads (about 61% of the total), the callback rate is 14.4% for young applicants aged 29-31. It is about 29% lower for 49-51

[33] This test treats the observations as independent. In the regression (probit) analyses that follow, the standard errors are clustered appropriately.

year-old applicants (10.3%), and about 47% lower for 64-66 year-old applicants (7.6%). 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% lower – 14.7%, versus 20.9% 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 versus 28.7%), although only marginally significant (p-value = 0.11). 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.[34]

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%). Both are lower than the callback rate for younger applicants (24.3%), with p-values of 0.09 and 0.12. For janitor jobs (Panel F), the callback rate was slightly higher for middle-aged than younger workers. But the callback rate was significantly lower for older applicants (25.9%), 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).

[34] 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.

Second, there is a question of what evidence on callbacks tells us about hiring. For example, if employers believe there is age discrimination, then they may expect older applicants to be more likely to respond positively to a callback. Nondiscriminatory employers might then direct more callbacks to older workers, which would generate a bias against finding evidence of age discrimination (although employers might do the opposite if they have a target share of older worker hires). However, there is evidence that differences in callback rates accurately reflect hiring discrimination. The Bendick et al. (1999) audit study that 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.[35] To be clear, though, we are assuming that our results for callbacks would carry over to job offers.

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 employed/unemployment, and skills. We then add controls for an extensive set of resume features listed in the table notes. The combined specifications also include controls for female occupation. 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 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.[36,37]

[35] An employer is more likely to discriminate at the pre-interview (callback) stage than at the interview 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 who to call back for an interview.

[36] The model is estimated using standard errors clustered at the resume level, since it possible that there are random features of resumes that affect the outcomes. There may also be random influences at the level of the job ad, but we cannot match all responses perfectly to job ads. If we cluster at the job ad level instead of the resume level (for the observations for which we can match to job ads), standard errors are a bit smaller for the comparable sample.

[37] 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

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.

A Richer Characterization of Resume Types

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

$$(9) \qquad T_i = \alpha + \beta_L M_{L,i} + \beta_{HB} M_{HB,i} + \beta_{HNB} M_{HNB,i} + \gamma_L S_{L,i} + \gamma_{HB}{}^E S_{HB}{}^E{}_{,i} + \gamma_{HB}{}^L S_{HB}{}^L{}_{,i} + \gamma_{HNB} S_{HNB,i} + X_{il}\delta + \varepsilon_i.$$ [38]

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 experience that is commensurate with age. And we estimate whether there are differences depending on whether applicants are "bridging" to a lower-skilled job, or – for the older applicants – whether they have already done so. For both middle-aged and older applicants, there is one low-experience resume, captured by M_L and S_L. In contrast, there are two different commensurate-experience resumes for middle-aged applicants, with or without bridging (M_{HB} and M_{HNB}, respectively), and three for older applicants, with early bridging, late bridging, or without bridging ($S_{HB}{}^E$, $S_{HB}{}^L$, and S_{HNB}, respectively). We report tests for two hypotheses: (i) no difference in callback rates between the otherwise similar high- and low-experience resumes ($\beta_{HNB} = \beta_L$ and $\gamma_{HNB} = \gamma_L$); and (ii) no difference between the non-bridge and the bridge resumes ($\beta_{HNB} = \beta_{HB}$, $\gamma_{HNB} = \gamma_{HB}{}^E$, and $\gamma_{HNB} = \gamma_{HB}{}^L$).

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 past studies. Second, the results for the low-experience resumes (the estimates of $\beta_L$and

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.)

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

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

γ_L) provide closer comparisons with past work that gave older and younger applicants the same experience (most notably, Lahey, 2008).[39] 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.

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 to 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 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 the same analyses broken down 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 non-bridge resumes.[40,41]

One 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

[39] In addition, our retail and administrative jobs overlap with Lahey (2008) and Riach and Rich (2010).
[40] 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.
[41] One issue discussed in the on-line 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, for middle-aged applicants the statistical evidence of equality of effects for the resumes showing low versus commensurate experience was stronger, with a p-value of 0.06. However, the results for older applicants still did not indicate any difference between low- and high-experience resumes (p-value = 0.85).

applicants reporting low experience – in this case the callback rate is 9.4 percentage points lower than for young applicants, significant at the one-percent level. And the test statistics (Panel A) rejects equality of effects across the two resume types at the five-percent level. Thus, for this occupation, there is arguably a bias against finding age discrimination from using resumes that do not report a "full" job history.

The results for the bridge resumes also provide indirect evidence allaying some of the potential challenges to interpreting lower callback rates for older applicants as reflecting age discrimination. One question was 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). Since employers should know this from past experience, if declining health is an issue, older applicants with "bridge resumes" should experience lower callback rates than other older applicants. We do not find such evidence. A second 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 non-bridge resumes (for resumes with commensurate experience); again, we find no such evidence. Finally, employers might assume that older women showing low experience had many career interruptions to care for children.[42] We are skeptical 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.

Addressing the Heckman Critique

For the analysis of potential biases introduced by differences in the variance of unobservable, we focus on the sharpest results with arguably the greatest policy relevance – the differences in outcomes

[42] Like the actual resumes, our resumes do not show time out for child care. In the resumes we examined for the types of jobs for which we applied, scraping and checking the resume content revealed that is was very unusual for women's resumes to provide explicit information on staying at home to raise children. Well below 1% had any such reference, and most of these had a reference to pairs of words like "child" and "provider" that were more likely to indicate a paid job than staying at home.

between young and old (near retirement age) applicants, without regard to the variation in resume types.[43] As a preliminary step, Table 7 reports models for callbacks that add the interactions between the skill indicators and the dummy variable for older applicants. The skill variables are intended to shift callback probabilities, so that differences in their coefficients for young versus old applicants provide information about the relative variances of the unobservable for young and old 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 lowering the absolute value of the effect of skill for older applicants.[44]

The estimates 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 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).[45] 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.[46]

[43] Implementation of this method in other contexts (e.g., Neumark, 2012; Neumark and Rich, 2016) shows that, unsurprisingly, standard errors of the discrimination estimates are quite a bit larger, so that the bias-corrected estimated for differences between middle-aged and younger applicants, and between resume types, would be likely be uninformative. (This was confirmed in analyses not reported here implementing the bias correction for middle-aged versus younger applicants.)

[44] The standard 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.

[45] The statistical significance of these estimates is not critical. What is critical is that the skill variables have non-zero 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 0.05 for administrative, security, and janitor jobs, but in the 0.5-0.7 range for sales jobs.

[46] Note that the positive interaction of computer skills and "Old" is consistent with statistical discrimination against older applicants, 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 on-line 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.

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.

Table 8 turns to the heteroscedastic probit estimates. The first row of Panel B reports the overall effect from the heteroscedastic probit estimates, which are similar to the probit estimates (Panel A).[47] 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. The p-value is high across all the columns, indicating that we do not reject the overidentifying restrictions; this could be, in part, attributable to the imprecise estimates of the skill variables for some of the jobs, as reported in Table 7.

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 is the potential source of bias in the probit estimates. For all jobs combined, in column (1), the unobservables correction makes little difference. The estimated ratio of standard deviations is a bit greater 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 two rows of the table, however, there is not always strong statistical evidence against the homoscedastic model with equal standard deviations, although we reject the restriction for female sales applicants, and the p-value for janitor jobs is 0.11.[48]

The last two rows of the table decompose the heteroscedastic probit estimates. The "level" effect (labelled "Old-level" in the table) is the unbiased estimate, and the "variance" effect captures the difference in the variance for older applicants. For all jobs combined, the resulting estimate of

[47] The marginal effects are calculated differently from those for the probit estimates in Table 5; see the table notes.
[48] Regardless, imposing equal variances of the unobservable can still lead to biased estimates of discrimination.

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.[49] 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).[50] For female sales applicants, the evidence of discrimination 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 re-

[49] To see this, define the following notation: Φ = the standard normal distribution function; c = the hiring threshold; β_I = the probit coefficient on the observable characteristics on the resume (only one is used here, for simplicity); X^{I*} = the level at which the observable characteristic is set in the experiment; γ = 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:

$$\Pr[T(P(X^{I*},X_S^{II})|S = 1) = 1] = \Phi[(\beta_I X^{I*} + \gamma - c)/\sigma_S^{II}]$$

$$\Pr[T(P(X^{I*},X_Y^{II}) |S = 0) = 1] = \Phi[(\beta_I X^{I*} - c)/\sigma_Y^{II}].$$

If X^{I*} is standardized at a low level, then $\beta_I X^{I*} < c$. In this case, a larger variance for older workers, $\sigma_S^{II} > \sigma_Y^{II}$, can generate $\Phi[(\beta_I^{I*} + \gamma - c)/\sigma_S^{II}] > \Phi[(\beta_I X^{I*} - c)/\sigma_Y^{II}]$ even when $\gamma = 0$ – a bias towards spurious evidence of discrimination *in favor* of older workers.

The estimates in Table 7 can be informative about $\sigma_S^{II}/\sigma_Y^{II}$. A standard probit model identifies coefficients only relative to the standard deviation of the unobservable, so we normalize the standard deviation of the unobservable to one. We can impose this normalization for young applicants ($\sigma_Y^{II} = 1$), which redefines the standard deviation of the unobservable for older applicants to be relative to the standard deviation for younger applicants, or $\sigma_{S/Y}^{II}$. If we assume that β_I is equal for younger and older applicants, then the ratio of the probit coefficients identifies $\sigma_{S/Y}^{II}$.

[50] The evidence of age discrimination for administrative jobs was stronger when the spam responses were dropped, as expected. In estimates corresponding to column (2) of Table 8, when we dropped these responses the "Old-level (marginal)" estimate was −0.081, significant at the one-percent level (versus −0.054, significant only at the 10-percent level, in the table). The "Old-variance (marginal)" estimate remained small and statistically insignificant.

estimated using only the high-experience resumes, the estimated level effect fell by half and was not statistically significant (p-value = 0.38).[51] 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).

The heteroscedastic 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, absent 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 (columns (1), (2), and (4), or three of the five cases where 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 predictive power, so for these occupations additional evidence with skill variables that more strongly shift callbacks would be useful.

6. Conclusions

Past audit and correspondence (AC) studies of age discrimination almost uniformly point to

[51] 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.

discrimination against older workers in hiring. However, there are important challenges in applying AC methods to age discrimination, as a result of which the differences in callback rates observed may not reflect discrimination. In particular, past studies may have been biased *towards* finding age discrimination by giving older applicants limited experience in the job to which they were applying – making their resumes more strictly comparable to those of younger applicants, but also possibly conveying a negative signal. And the past studies fail to address the "Heckman critique," quite possibly generating bias *against* finding age discrimination.[52] Given two potential sources of bias in past research, which may be in offsetting directions, it is hard to know what to make of the past evidence claiming age discrimination in hiring.

To address these issues, we conducted a new correspondence study of age discrimination, adding features to address the two sources of bias just described, 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. In addition, AC studies cannot definitively distinguish among different mechanisms of discrimination – most importantly, taste versus statistical discrimination. However, we believe our data help rule out a number of the most plausible statistical discrimination explanations for the age differences we find.

We have a number of central findings. First, we find the stronger evidence against older applicants near retirement ages (64-66) than for 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. Second, for the most part we find that resumes for older applicants with

[52] The experimental literature on discrimination more generally has failed to address this critique, until recently.

experience comparable to younger applicants does not bias the evidence towards finding age discrimination. However, for one of the three jobs (janitors) to which we send male applicants – and the one job that otherwise provides the strongest evidence of discrimination against older men – our evidence does suggest that the evidence of age discrimination is spuriously generated by using low-experience resumes. Third, 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.

We find much stronger and more robust evidence of age discrimination against older women than against older men.[53] This sex difference is potentially important from the perspective of population aging; women tend to outlive men, and with their lifetime earnings converging to those of men, women's long-term financial security will increasingly depend on their own work history.

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 et al., 1986).[54] If older women suffer from discrimination because of both age and sex, anti-discrimination 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, 2013).

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.

---

[53] Another recent U.S. correspondence study by Farber et al. (2015) provides corroborating evidence of age discrimination against women. The study focuses more on the effect of unemployment duration than on 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).

[54] This is consistent with evidence in Kuhn and Shen (2013) and Hellester et al. (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.

References

Adams, Scott J. 2002. "Passed Over for Promotion Because of Age: An Empirical Analysis of the Consequences." *Journal of Labor Research*, Vol. 23, pp. 447-61.

Becker, Gary S. 1971. *The Economics of Discrimination, Second Edition*. Chicago: University of Chicago Press.

Belbase, Anek, Goeffrey T. Sanzenbacher, and Christopher M. Gillis. 2015. "Does Age-Related Decline in Ability Correspond with Retirement Age?" Center for Retirement Research Working Paper No. 2015-24.

Bendick, Marc, Jr., Lauren E. Brown, and Kennington Wall. 1999. "No Foot in the Door: An Experimental Study of Employment Discrimination Against Older Workers." *Journal of Aging & Social Policy*, Vol. 10, pp. 5-23.

Bendick, Marc, Jr., Charles W. Jackson, and J. Horacio Romero. 1997. "Employment Discrimination Against Older Workers: An Experimental Study of Hiring Practices." *Journal of Aging & Social Policy*, Vol. 8, pp. 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." *American Economic Review*, Vol. 94, pp. 991-1013.

Deutsch, Francine M., Carla M. Zalenski, and Mary E. Clark. 1986. "Is There a Double Standard of Aging?" *Journal of Applied Social Psychology*, Vol. 16, pp. 771-85.

Farber, Henry S., Dan Silverman, and Till von Wachter. 2015. "Factors Determining Callbacks to Job Applications by the Unemployed." NBER Working Paper No. 21689.

Fix, Michael, and Raymond Struyk. 1993. Clear and Convincing Evidence: Measurement of Discrimination in America. Washington, DC: The Urban Institute 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: University of Minnesota.

Gordon, Randall A., and Richard D. Arvey. 2004. "Age Bias in Laboratory and Field Settings: A Meta-Analytic Investigation." *Journal of Applied Social Psychology*, Vol. 34, pp. 468-92.

Heckman, James J. 1998. "Detecting Discrimination." *Journal of Economic Perspectives*, Vol. 12, pp. 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 Hanushek and Welch, eds., Handbook of the Economics of Education, Volume 1, Chapter 7, pp. 307-458.

Heckman, James, and Peter Siegelman. 1993. "The Urban Institute Audit Studies: Their Methods and Findings." In Fix and Struyk, eds., Clear and Convincing Evidence: Measurement of Discrimination in America. Washington, D.C.: The Urban Institute Press, pp. 187-258.

Hellerstein, Judith, David Neumark, and Kenneth Troske. 1999. "Wages, Productivity, and Worker Characteristics." *Journal of Labor Economics* 17(3): 409-46.

Hellester, Miguel Delgado, Peter Kuhn, and Kailing Shen. 2014. "Employers' Age and Gender Preferences: Direct Evidence from Four Job Boards." Unpublished manuscript, University of California, Santa Barbara.

Jablonski, M., K. Kunze, and L. Rosenblum. 1990. "Productivity, Age, and Labor Composition Changes in the U.S. Work Force." In I. Bluestone, R. Montgomery, and J.D. Owen, eds. The Aging of the American Work Force. Detroit, MI: Wayne State University Press.

Jackson, Linda A. 1992. Physical Appearance and Gender: Sociological and Sociocultural Perspectives. Albany, NY: State University of New York Press.

Johnson, Richard W. 2014. "Later Life Job Changes before and after the Great Recession." Draft final report to AARP.

Johnson, Richard W., Janette Kawachi, and Eric K. Lewis. 2009. "Older Workers on the Move:

Recareering in Later Life." Washington, DC: AARP Public Policy Institute.
Johnson, Richard W., and David Neumark. 1997. "Age Discrimination, Job Separations, and Employment Status of Older Workers." *Journal of Human Resources* 32(4): 779-811.
Kaas, Leo, and Christian Manger. 2011. "Ethnic Discrimination in Germany's Labour Market: A Field Experiment." *German Economic Review*, Vol. 13, pp. 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." *Journal of Social Issues* 61(2): 241-66.
Kuhn, Peter, and Kailing Shen. 2013. "Gender Discrimination in Job Ads: Evidence from China." *Quarterly Journal of Economics*, Vol. 128, pp. 287-336
Lahey, Joanna. 2008. "Age, Women, and Hiring: An Experimental Study." *Journal of Human Resources*, Vol. 43, pp. 30-56.
Maestas, Nicole. 2010. "Back to Work: Expectations and Realizations of Work after Retirement." *Journal of Human Resources*, Vol. 45, pp. 718-48.
Mincer, Jacob. 1974. Schooling, Experience, and Earnings. New York: Columbia University Press.
Molloy, Raven, Christopher L. Smith, and Abigail Wozniak. 2011. "Internal Migration in the United States." *Journal of Economic Perspectives*, Vol. 25, pp. 173-96.
Neumark, David. 2012. "Detecting Evidence of Discrimination in Audit and Correspondence Studies." *Journal of Human Resources*, Vol. 47, pp. 1128-57.
Neumark, David, and Judith Rich. 2016. "Do Field Experiments on Labor and Housing Markets Overstate Discrimination? A Re-Examination of the Evidence." NBER Working Paper No. 22278.
Neumark, David, and Joanne Song. 2013. "Do Stronger Age Discrimination Laws Make Social Security Reforms More Effective?" *Journal of Public Economics*, Vol. 108, pp. 1-16.
Posner, Richard A. 1995. *Aging and Old Age*. Chicago: University of Chicago Press.
Riach, Peter A., and Judith Rich. 2010. "An Experimental Investigation of Age Discrimination in the English Labor Market." *Annals of Economics and Statistics*, No. 99/100, pp. 169-85.
Riach, Peter A., and Judith Rich. 2006. "An Experimental Investigation of Age Discrimination in the French Labour Market." IZA Discussion Paper No. 2522.
Rooth, Dan-Olof. 2010. "Automatic Associations and Discrimination in Hiring: Real World Evidence." *Labour Economics*, 17(3), pp. 523-34.
Rutledge, Matthew S., Steven A. Sass, and Jorge D. Ramos-Mercado. 2016. "How Does Occupational Access for Older Workers Differ by Education." Center for Retirement Research at Boston College.
Song, Joanne. 2013. "Falling between the Cracks: Discrimination Laws and Older Women." Unpublished paper. Irvine, CA: University of California, Irvine.
Tinkham, Thomas. 2010. "The Uses and Misuses of Statistical Proof in Age Discrimination Claims." William Mitchell College of Law, Legal Studies Research Paper Series 2010-20.
U.S. Department of Labor. 1965. *The Older American Worker*. Washington, DC: U.S. Government Printing Office.

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

	Age-specific recent hires/all recent hires in occupation	
Occupation	**Age 62 to 70**	**Age 28 to 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%

Table 2: Cities in Study, by Percent 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)

Notes: The first number in parentheses is the percent of population aged 62 and over, based on 2012 ACS five-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. Nationally, 16.3% of the population is aged 62+.

Table 3: Level of Matching of Callbacks

	Matched positive responses			
	Job id match	**Resume match only**	**No responses**	**Total**
Voicemail	2,495	765	N.A.	3,260
Email	2,822	97	N.A.	2,919
All	5,317	862	34,044	40,223

Notes: 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.

Table 4: Callback Rates by Age

		Young (29-31)	**Middle (49-51)**	**Old (64-66)**
A. Combined (N=40,223)				
Callback (%)	No	81.31	84.60	87.86
	Yes	18.69	15.40	12.16
Tests of independence (p-value)	Young vs. middle vs. old (0.00)	Young vs. middle (0.00)	Young vs. old (0.00)	Middle vs. old (0.00)
B. Administration (N=24,350, female)				
Callback (%)	No	85.59	89.70	92.42
	Yes	14.41	10.30	7.58
Tests of independence (p-value)	Young vs. middle vs. old (0.00)	Young vs. middle (0.00)	Young vs. old (0.00)	Middle vs. old (0.00)
C. Sales (N=5,348, males)				
Callback (%)	No	79.11	78.89	85.30
	Yes	20.89	21.09	14.70
Tests of independence (p-value)	Young vs. middle vs. old (0.00)	Young vs. middle (0.90)	Young vs. old (0.00)	Middle vs. old (0.00)
D. Sales (N=4,707, females)				
Callback (%)	No	71.32	74.13	81.57
	Yes	28.68	25.87	18.43
Tests of independence (p-value)	Young vs. middle vs. old (0.00)	Young vs. middle (0.11)	Young vs. old (0.00)	Middle vs. old (0.00)
E. Security (N=4,138, male)				
Callback (%)	No	75.72	78.45	78.26
	Yes	24.28	21.55	21.74
Tests of independence (p-value)	Young vs. middle vs. old (0.16)	Young vs. middle (0.09)	Young vs. old (0.12)	Middle vs. old (0.93)
F. Janitors (N=1,680, male)				
Callback (%)	No	67.92	66.55	74.11
	Yes	32.08	33.45	25.89
Tests of independence (p-value)	Young vs. middle vs. old (0.01)	Young vs. middle (0.66)	Young vs. old (0.03)	Middle vs. old (0.01)

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

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)	-0.033***	-0.033***	-0.035***	-0.035***	-0.015	-0.017	-0.051***	-0.052***	-0.024	-0.023	0.015	0.025
	(0.005)	(0.005)	(0.005)	(0.005)	(0.013)	(0.014)	(0.015)	(0.018)	(0.016)	(0.018)	(0.027)	(0.029)
Old (64-66)	-0.062***	-0.062***	-0.063***	-0.063***	-0.044***	-0.047***	-0.095***	-0.095***	-0.027	-0.029	-0.061**	-0.055*
	(0.004)	(0.005)	(0.005)	(0.005)	(0.012)	(0.014)	(0.014)	(0.017)	(0.017)	(0.017)	(0.027)	(0.029)
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
High-skill indicator	X	X										
Resume features		X		X		X		X		X		X
Female	X	X										
Occupation	X	X										
Callback rate for young (29-31)	18.69		14.41		20.89		28.68		24.28		32.08	
N	40,223		24,350		5,348		4,707		4,138		1,680	
Clusters	3,694		1,052		544		513		893		694	

Notes: 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 based on clustering at the resume level. Significantly different from zero at 1-percent level (***), 5-percent level (**) or 10-percent level (*). Resume features include: template; email script; email format; script subject, opening, body, and signature; and file name format.

Table 6: Probit Estimates for Callbacks by Age and Resume Type, Marginal Effects, Full Controls

	Combined	Administrative	Sales-Males	Sales-Females	Security	Janitor
	(1)	(2)	(3)	(4)	(5)	(6)
Callback estimates						
Middle, commensurate experience (M_{HNB})	-0.033*** (0.006)	-0.029*** (0.006)	-0.029 (0.017)	-0.057** (0.022)	-0.028 (0.022)	0.015 (0.036)
Middle, commensurate experience, bridge application (M_{HB})	-0.027** (0.006)	-0.027*** (0.006)	0.020 (0.019)	-0.064*** (0.022)	-0.046* (0.022)	…
Middle, experience = young (M_L)	-0.035*** (0.006)	-0.041*** (0.006)	-0.037* (0.018)	-0.027 (0.025)	0.009 (0.028)	0.035 (0.034)
Old, commensurate experience (S_{HNB})	-0.058*** (0.006)	-0.058*** (0.005)	-0.048* (0.023)	-0.080*** (0.021)	-0.050* (0.027)	-0.017 (0.037)
Old, commensurate experience, bridge application, already bridged (S_{HB}^E)	-0.054*** (0.006)	-0.048*** (0.006)	-0.041** (0.017)	-0.100*** (0.022)	-0.026 (0.025)	…
Old, commensurate experience, bridge application (S_{HB}^L)	-0.054*** (0.006)	-0.055*** (0.006)	-0.052*** (0.017)	-0.074*** (0.019)	-0.030 (0.023)	…
Old, experience = young (S_L)	-0.062*** (0.006)	-0.057*** (0.005)	-0.038* (0.020)	-0.099*** (0.022)	-0.003 (0.035)	-0.094*** (0.031)
Tests of restrictions (p-value)						
A. Commensurate experience = low experience						
Middle-aged: $\beta_{HNB} = \beta_L$	0.80	0.12	0.71	0.31	0.23	0.59
Older: $\gamma_{HNB} = \gamma_L$	0.56	0.90	0.73	0.48	0.24	0.05
Joint, middle-aged and older	0.82	0.30	0.88	0.47	0.25	0.12
B. Bridge resumes = non-bridge resumes (all high experience)						
Middle-aged: $\beta_{HNB} = \beta_{HB}$	0.38	0.71	0.02	0.80	0.53	…
Older: $\gamma_{HNB} = \gamma_{HB}^E$	0.57	0.17	0.78	0.46	0.47	…
Older: $\gamma_{HNB} = \gamma_{HB}^L$	0.61	0.69	0.89	0.76	0.51	…
Joint, older ($\gamma_{HNB} = \gamma_{HB}^E$, $\gamma_{HNB} = \gamma_{HB}^L$)	0.83	0.38	0.85	0.54	0.74	…
Joint, middle-aged and older	0.76	0.56	0.10	0.73	0.80	…
Callback rate for young	18.69	14.41	20.89	28.68	24.28	32.08
N	40,223	24,350	5,348	4,707	4,138	1,680
Clusters	3,694	1,052	544	513	893	694

Notes: See notes 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 based on clustering at the resume level. Significantly different from zero at 1-percent level (***), 5-percent level (**) or 10-percent level (*). There are no bridge resumes for janitors.

Table 7: Probit Estimates for Callbacks by Age, Old vs. Young Only, Effects of Skills and Interactions of Old with Skills, Marginal Effects

	Combined	Admin.	Sales-males	Sales-females	Security	Janitor
	(1)	(2)	(3)	(4)	(5)	(6)
Old	-0.071***	-0.090***	-0.062	-0.102	-0.037	0.144
	(0.023)	(0.020)	(0.085)	(0.077)	(0.057)	(0.141)
Common skills						
Spanish	-0.002	0.003	0.007	-0.038	0.081*	-0.021
	(0.011)	(0.010)	(0.025)	(0.037)	(0.045)	(0.047)
Spanish x Old	0.009	0.016	-0.046	0.029	0.038	-0.026
	(0.017)	(0.019)	(0.032)	(0.056)	(0.060)	(0.078)
Grammar	-0.014	-0.019**	-0.017	-0.010	0.025	0.002
	(0.009)	(0.009)	(0.020)	(0.033)	(0.034)	(0.046)
Grammar x Old	0.010	0.031**	0.041	-0.014	-0.019	0.011
	(0.015)	(0.016)	(0.037)	(0.043)	(0.045)	(0.077)
College	0.027***	0.024**	0.008	0.016	0.023	0.125**
	(0.010)	(0.010)	(0.023)	(0.028)	(0.038)	(0.051)
College x Old	-0.016	-0.023*	-0.007	-0.016	0.003	-0.058
	(0.013)	(0.013)	(0.031)	(0.038)	(0.049)	(0.072)
Employee of the month	0.002	0.003	0.033	-0.018	-0.071*	-0.059
	(0.010)	(0.009)	(0.028)	(0.029)	(0.036)	(0.044)
Employee of the month x Old	-0.000	0.002	-0.017	0.042	0.024	0.072
	(0.014)	(0.014)	(0.034)	(0.044)	(0.053)	(0.079)
Volunteer	0.011	0.016*	-0.027	0.022	-0.019	-0.100**
	(0.010)	(0.009)	(0.024)	(0.032)	(0.039)	(0.047)
Volunteer x Old	-0.005	-0.014	0.053	-0.024	-0.034	0.087
	(0.014)	(0.013)	(0.040)	(0.044)	(0.051)	(0.081)
Occupation-specific skills		*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*
Skill 1		-0.012	0.001	0.026	-0.064*	0.132**
		(0.010)	(0.024)	(0.029)	(0.034)	(0.065)
Skill 1 x Old		0.034**	0.034	-0.019	0.111**	-0.140*
		(0.016)	(0.039)	(0.038)	(0.060)	(0.065)
Skill 2		0.021**	0.012	0.008	0.065*	-0.008
		(0.010)	(0.024)	(0.029)	(0.039)	(0.062)
Skill 2 x Old		-0.024*	0.008	-0.039	-0.052	0.007
		(0.012)	(0.036)	(0.037)	(0.044)	(0.086)
N	27,492	16,449	3,570	3,609	2,746	1,118
Number of clusters	2,522	717	359	386	599	462

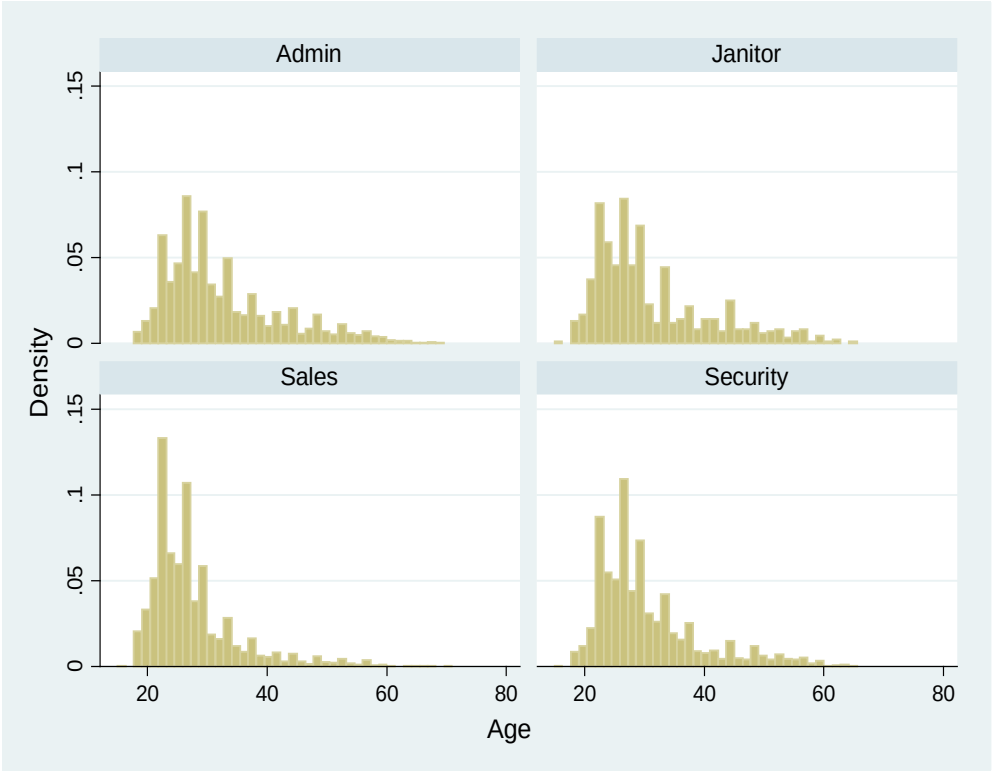
Notes: See notes to Table 5. Standard errors are computed based on clustering at the resume level. Significantly different from zero at 1-percent level (***), 5-percent level (**) or 10-percent level (*). Control variables correspond to the first specification for each occupation in Table 5 (odd-numbered columns). 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 zero). Only the skill variable interactions (and main effects) are reported. In column (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 dummy variable, evaluating other variables at their means. (See text for explanation.)

Table 8: Heteroscedastic Probit Estimates for Callbacks by Age, Old vs. Young Only (Corrects for Potential Biases from Difference in Variance of Unobservables)

	Combined	Administrative	Sales-males	Sales-females	Security	Janitor
	(1)	(2)	(3)	(4)	(5)	(6)
	5 common skills	All skills	All skills	All skills	All skills	All skills
A. Probit estimates						
Old (marginal)	-0.062***	-0.067***	-0.044***	-0.093***	-0.028	-0.062**
	(0.006)	(0.005)	(0.012)	(0.014)	(0.017)	(0.028)
B. Heteroscedastic probit estimates						
Old (marginal)	-0.060***	-0.068***	-0.049***	-0.074***	-0.022	-0.049*
	(0.006)	(0.006)	(0.012)	(0.015)	(0.020)	(0.029)
Overidentification test: ratios of coefficients on skills for old relative to young are equal (p-value, Wald test)	0.99	0.78	0.88	0.91	0.85	1.00
Standard deviation of unobservables, old/young	1.09	0.94	0.84	1.44	1.16	1.66
Test: standard vs. heteroscedastic probit (p-value, log-likelihood test)	0.28	0.56	0.26	0.02	0.22	0.11
Old-level (marginal)	-0.080***	-0.054*	-0.005	-0.161***	-0.058*	-0.153*
	(0.022)	(0.028)	(0.039)	(0.034)	(0.030)	(0.082)
Old-variance (marginal)	0.020	-0.014	-0.043	0.086**	0.036	0.104
	(0.023)	(0.029)	(0.040)	(0.040)	(0.035)	(0.092)
N	27,492	16,449	3,570	3,609	2,746	1,118

Notes: 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. In this table, we use the continuous version of the partial derivative, because this version gives an unambiguous decomposition of the estimates from the heteroscedastic probit model (Neumark, 2012); thus, the probit marginal effects differ slightly from Table 5. Significantly different from zero at 1-percent level (***), 5-percent level (**) or 10-percent level (*). Control variables correspond to the first specification for each occupation in Table 5 (odd-numbered columns), except that skill vector is as noted. Callback rates for young and old applicants are as in Table 4.

Figure 1: Histograms of Resumes by Age, Resume Website



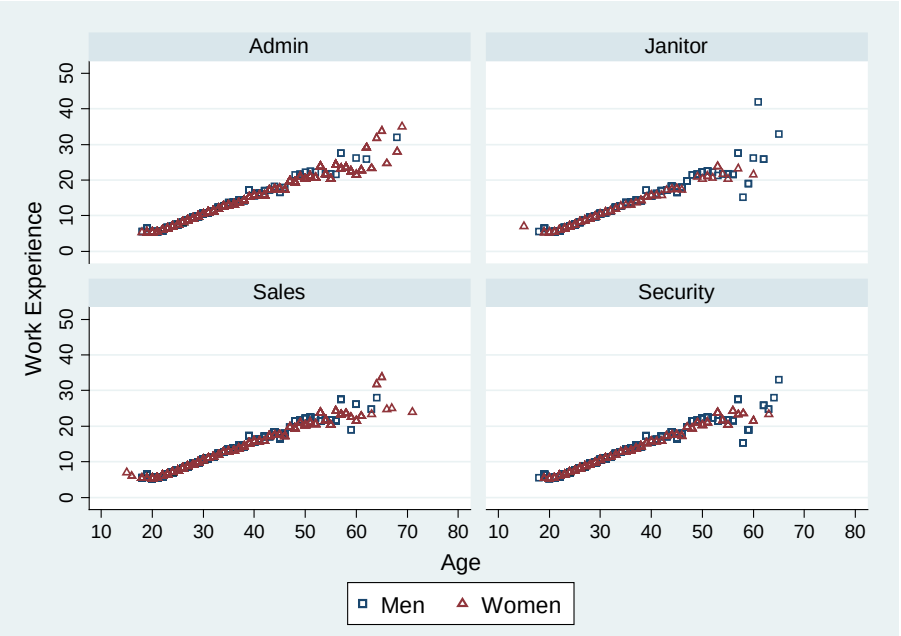
Notes: Based on the sample of resumes extracted from a resume-posting website, as described in text. The percent of observations for ages 62-70 for each job are: administrative – 0.73%; janitor – 0.49%; sales – 0.13%; security – 0.34%.

Figure 2: Age and Experience in Resume Sample

A. Overall averages by age



B. By job



Notes: Based on the sample of resumes from a resume-posting website, as described in text. In the individual-level data, the correlation between age and computed experience is 0.77. If there are multiple jobs held at the same time, experience is not double-counted.