American Economic Review: Papers & Proceedings 2016, 106(5): 314–318
http://dx.doi.org/10.1257/aer.p20161010

Determinants of Callbacks to Job Applications: An Audit Study[†]

By Henry S. Farber, Dan Silverman, and Till von Wachter*

We present results of a correspondence audit study of job applications by the unemployed to investigate how unemployment duration, age, and holding a low-level "interim" job affect the likelihood that experienced college-educated females seeking an administrative support job receive a callback from a potential employer.[1] The basic procedure is to send carefully constructed fictitious job applications that vary in key dimensions to posted job openings and to measure callback rates.

Our results are clear. First, there is no relationship between callback rates and the duration of unemployment. Second, workers age 50 and older are significantly less likely than are workers in their 30s and 40s to receive a callback. Third, taking an interim job that is at a lower skill level than the job for which they are applying significantly reduces the likelihood of receiving a callback. Finally, employers who have higher callback rates respond less to observable differences across workers in determining whom to call back.

We also contrast our findings with those of recent papers adopting a similar audit study approach to estimating the effect on callbacks of unemployment duration and other resume characteristics for less experienced workers. The effect of unemployment duration on callbacks varies substantially between studies, and cannot be conclusively regarded as negative. Some of the variation in results could be explained by well-defined resume characteristics, such as time period of application, education, gender, or age. However, based on the existing data, sample sizes, and study protocols it is difficult to assess the sources of the substantial heterogeneity in findings. We conclude that while the audit study approach is a useful tool for understanding the determinants of callbacks, more research with comparable design protocols is needed to assess the external validity of particular findings.

I. Evaluation of Existing Literature

There has been a flurry of resume-based, audit studies of factors determining callback rates for job applications. These were motivated in part by the concern that long unemployment spells in the Great Recession and its aftermath have contributed to a lasting rise in unemployment by depressing job finding rates of particular types of workers. An advantage of audit studies is that they can yield estimates of the causal effect of unemployment duration and other resume characteristics on callbacks in the presence of negative selection.

Another appeal of audit studies of resume characteristics is that the empirical approach can be tied to theoretical predictions from a learning model. The key premise of the model is that employers use observable information available in the worker's job application—such as worker demographics, education, work history, and unemployment experience—to form an expectation about the worker's quality.

The fundamental advantages of the audit method lead the recent studies to share many features and face similar constraints. The authors of each study sent fictitious resumes with random variation in key dimensions (e.g., unemployment duration or age) to actual job postings and measured callback rates. Despite broad similarities, these studies differ substantially along many dimensions. While some are apparent—such as focus on different demographic groups, different periods, or different definitions of the treatment—others are more subtle (e.g., contextual differences due to resume design choices).

* Farber: Princeton University (e-mail: farber@princeton.edu); Silverman: Arizona State University (e-mail: dsilver3@asu.edu); von Wachter: University of California, Los Angeles (e-mail: tvwachter@econ.ucla.edu). The authors thank the Sloan Foundation for financial support and participants in numerous seminars for helpful comments.

[†] Go to http://dx.doi.org/10.1257/aer.p20161010 to visit the article page for additional materials and author disclosure statement(s).

[1] Full details of our study are available in Farber, Silverman, and von Wachter (2015).

Table 1—Audit Studies of Effect of Unemployment History on Callback, Selected Characteristics

Study	Locus	Age and sex	Education	Occupation	Months unemp. dur. (UD)	UD effect on callbacks
Oberholzer-Gee (2008)	Zurich, 1999	25–29 F	—	Administrative assistant	6–30	$+\ UD < 12$, $-\ UD > 8$
Kroft, Lange, and Notowidigdo (2013)	US, 100 MSAs, 2011–2012	19–40 M&F	HS, AS, BA	Administrative, Clerical, Customer service, Sales	1–36	− effect tight LM. Weaker − in slack LM.
Ghayad (2013)	US, 25 MSAs, 2012	25–30 M	BA	Professional, Sales, Administrative	1–12	− effect
Eriksson and Rooth (2014)	Sweden, 2 cities	20–30 M&F	Various skill levels	Various (13)	0–9	− effect low skill, $UD \geq 9$. 0 other.
Nunley et al. (2014)	US, 2013	25–26 M&F	BA	Marketing, Sales, Assistant manager	3–12	0 effect
Farber, Silverman, and von Wachter (2015)	US, 8 MSAs, 2012–2014	35–58 F	BA	Administrative, Clerical	1–12	0 effect

Note: See paper for details.

In addition, resource constraints limit the number and type of resumes sent out, the number of jobs to which they are sent, and the number of dimensions of random variation. These differences limit the comparability of studies and, consequently, their usefulness in finding a common answer to the core questions.

Table 1 summarizes a number of audit studies that, in addition to our own, examine directly the effect of unemployment duration on the callback rate. The last column of the table summarizes the main findings. A key message of Table 1 is that the results vary substantially. While some studies find negative effects of unemployment duration on callback rates, others find no effects, or effects only for certain groups or markets. The first study of this kind, Oberholzer-Gee (2008), finds positive effects of unemployment duration on callbacks for spells as long as 12 months, and negative effects for very long unemployment spells. Kroft, Lange, and Notowidigdo (2013), in an influential paper, find negative effects of unemployment duration on callbacks for younger workers that is stronger in cities with low unemployment rates. Ghayad (2013) finds a strong negative effect on callback rates for younger workers. Eriksson and Rooth (2014) find no effects of past or contemporary unemployment on callbacks, except for applicants to low skilled jobs who have been unemployed for nine months. Neither Nunley et al. (2014) nor our study, Farber, Silverman, and von Wachter (2015), find any effects of unemployment duration on callback rates.

The remaining columns of the table give an overview of the broad range of design choices the studies have made. While it is difficult to determine precisely why these studies do not agree in results, it is likely due to the fact that each of them focuses on different subgroups of workers (by age, skill level, occupation, sex, etc). As an illustration of the difficulty in understanding the source of the differing results, we summarize the efforts described in our companion paper (Farber, Silverman, and von Wachter 2015) to reconcile the findings of our audit study (no effect of unemployment duration on the callback rate) with those of Kroft, Lange, and Notowidigdo (2013) (a negative effect of unemployment duration on the callback rate). While we are able to rule out differences in the outcome measure, occupation, skill, and

geographic location as factors that can account for the difference in findings, the studies differ in other dimensions. For example, it is possible that applicant age is an important factor in accounting for the difference in findings (ours being the only study of which we are aware that examines the experience of older applicants). However, we cannot draw any firm conclusions because the samples do not overlap sufficiently in age for a full analysis.

In addition to these immediately apparent and easily quantifiable design choices, audit studies may also differ in ways that are harder to measure. A key difficulty is that little is known about how employers read and respond to the information on resumes. Hence, it is difficult to assess, for example, how different construction of resumes affects employer responses. Employers may respond to contextual cues or interactions between resume characteristics that researchers do not anticipate. Such aspects would neither be controlled for nor documented in the research design, and could contribute to differences in the findings. A broader range of resume attributes can be coded and hence included in the analysis. However, a better understanding of the processing of resumes by employers is likely to be necessary. In the meantime, use of common resume templates and comparable definitions of treatments could yield improvements in the comparability between studies.

Recent audit studies have also examined the effect of resume characteristics other than unemployment duration, in particular age and job quality. So far, there is more agreement among the audit studies of these characteristics. Bendick, Jackson, and Romero (1997); Lahey (2008); Farber, Silverman, and von Wachter (2015); and Neumark, Burn, and Button (2015) all yield the fairly consistent finding that callback rates are lower for older workers. In another example, Farber, Silverman, and von Wachter (2015) and Nunley et al. (2014) both find that non-standard employment histories can have an adverse effect on callback rates.[2]

[2] The literature on such resume characteristics is small, and these findings are hard to compare. Farber, Silverman, and von Wachter (2015) study the effect of low-quality interim jobs for experienced workers, whereas Nunley et al. (2014) study the effect of ill-matched internships for recent college graduates.

II. Design of Our Audit Study

An audit study involves sending fake resumes to actual job postings and measuring callback rates. The main results are estimates of differences in callback rates associated with randomly assigned differences in resume characteristics, such as age, unemployment, or employment history.

To facilitate the tailoring of resumes and reduce idiosyncratic variation in callback rates by job type, we limited applications to white-collar office jobs such as administrative or executive assistants, receptionists, secretaries, and the like. Because these jobs are disproportionately held by women, and gender differences were not our focus, all applicants had female names. Each applicant had a four-year bachelor's degree from comparable institutions. Our fictitious applicants had substantial work histories with three to six white-collar office jobs, depending on age. Prior to the current spell, they had no spells of unemployment longer than a month in the previous five years. Age was not indicated directly in the resumes but could be inferred from year of college completion and work experience. The context of our audit study was nationwide in that we submitted job applications to openings in eight selected cities across the United States.[3]

We sent our fictitious resumes in matched pairs or quadruples to openings posted on a large online job board. The experiment proceeded in four rounds, and our procedures evolved with experience. Our richest variation is in round four, where we submitted four applications per posting with within-posting variation in unemployment duration, age, and whether the applicant held an interim job. We focus our discussion here only on the 6,072 applications submitted to 1,118 job openings in round four (fielded April–August 2014). Unemployment duration indicated on each resume took on one of five values (0, 4, 12, 24, or 52 weeks) with equal

[3] Because we wanted to allow for differences in treatment effects by local unemployment rates, four of the cities we chose had relatively low unemployment rates (average 6.1 percent) in 2012 (Dallas, Texas; Omaha, Nebraska; Pittsburgh, Pennsylvania; and Portland, Maine) and four had relatively high unemployment rates (average 9.2 percent) in 2012 (Charlotte, North Carolina; Chicago, Illinois; Sacramento, California; and Tampa, Florida).

TABLE 2—LOGIT, RANDOM EFFECTS LOGIT, AND CONDITIONAL LOGIT ESTIMATES: ODDS RATIOS

Variable	All logit (1)	All RE logit (2)	All FE logit (3)	1 callback logit (4)	2 callbacks logit (5)	3 callbacks logit (6)
Unemp duration	1.001	1.004	1.005	1.001	1.010	1.000
	(0.002)	(0.003)	(0.004)	(0.005)	(0.006)	(0.012)
Age 55–58	0.688	0.533	0.534	0.380	0.365	1.019
	(0.056)	(0.064)	(0.064)	(0.087)	(0.132)	(0.455)
Interim job	0.842	0.736	0.725	0.485	0.998	1.428
	(0.074)	(0.095)	(0.099)	(0.099)	(0.217)	(0.565)
Number apps	6,072	6,072	1,092	600	340	152

Notes: Columns 1 and 2 include an indicator for low local unemployment rate. $\hat{\rho} = 0.704$ (s.e. 0024) for the RE logit model. Numbers in parentheses are standard errors. Standard errors in columns 1 and 4–6 are clustered at the job posting level.

probability. Age took on four possible values (37, 42, 57, or 58 years) with equal probability, and we define the two highest values of age to indicate older workers. Interim jobs were held by about one-half of the applicants. There was random variation within job posting in all three dimensions.

III. Results of our Audit Study

The callback rate in round four of our study was 9.1 percent. There was no systematic variation in callback rates by the duration of unemployment spell (p-value of test for independence $= 0.623$). Older workers had a significantly lower callback rate than younger workers (7.6 percent versus 10.7 percent with p-value of test for independence < 0.0005). Finally, workers who held a low-level interim job had a significantly lower callback rate than workers who did not hold such a job (8.4 percent versus 9.9 percent with p-value of test for independence $= 0.044$).

In order to explore alternative statistical models that allow for correlation across applications to the same job posting, we present a series of logit analyses, the results of which are contained in the first three columns of Table 2. Our first model (column 1) uses both within- and between-posting variation in applicant characteristics and does not account for within-posting correlation. The random effects model (column 2) accounts for the fact that job postings are randomly drawn from the underlying population and may differ in their mean callback rate. This model is appropriate (yields consistent estimates) when the baseline variation across job postings in their callback rates is uncorrelated with the observed applicant characteristics of interest. Finally (column 3), we present estimates of Chamberlain's fixed effect logit model which is robust to violation of the assumption that posting-specific variation in callbacks is independent of applicant characteristics.[4]

We report our estimates of the logit models in Table 2 in terms of odds ratios, which, for small probabilities, are approximately the ratio of probabilities of callback given treatment versus no treatment. We obtain three main findings: (i) There is no detectable effect of unemployment duration on callback rates.[5] In none of our models can we reject this null hypothesis. (ii) There is a precisely estimated negative effect (an estimated odds ratio less than one) of age on the callback rate. (iii) There is a substantial negative effect of holding an interim job on the callback rate. As expected, the results are confirmed in the random effects and fixed effects logit.[6]

[4] Estimation of the fixed-effect logit model relies only on the 1,092 applications to the 273 job postings for which there was within-job-posting variation in callback (1–3 callbacks).

[5] This result is robust to a less constrained specification for the unemployment duration variable (four indicator variables for the five distinct values of unemployment duration).

[6] This is expected given the fact that we assigned key characteristics on resumes sent to specific job postings randomly. The random-effect estimates should be consistent and efficient relative to the fixed-effect estimates, and this is verified by a Hausman test (p-value $= 0.85$).

A prediction from a simple learning model is that employers will be less selective on worker characteristics when they have a greater need for new hires, as measured here by the posting-specific callback rates. We investigate this directly using separate logit analyses of the probability of callback for postings with 1, 2, and 3 callbacks. The results of this analysis are contained in columns 4–6 of Table 2. Note that this analysis is in the spirit of the Chamberlain fixed-effect logit model, which conditions on the number of within-posting callbacks.

In no case is the callback rate significantly related to unemployment duration. The effect of age is present only for applications to jobs with one or two callbacks. The effect of reporting the holding of an interim job is present only for applications to jobs with one callback. This pattern suggests that employers who are eager to hire—and hence have a higher callback rate for their job posting—are less choosy, i.e., resume characteristics appear to matter less in determining callback. This supports the view that a strong labor market can play an important role in reducing the disadvantage of particular types of applicants (e.g., older applicants) in searching for jobs.

The results of our and other audit studies of the determinants of callbacks thus reveal the ability of this method to estimate important causal effects of worker characteristics that would be exceedingly difficult to estimate from observational data. At the same time, the variability of results across superficially similar studies indicates that studies with similar design protocols are needed to assess comparability and external validity of specific findings.

REFERENCES

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* 8 (4): 25–46.

Eriksson, Stefan, and Dan-Olof Rooth. 2014. "Do Employers Use Unemployment as a Sorting Criterion when Hiring? Evidence from a Field Experiment." *American Economic Review* 104 (3): 1014–39.

Farber, Henry S., Dan Silverman, and Till von Wachter. 2015. "Factors Determining Callbacks to Job Applications by the Unemployed: An Audit Study." National Bureau of Economic Research Working Paper 21689.

Ghayad, Rand. 2013. "The Jobless Trap." Unpublished.

Kroft, Kory, Fabian Lange, and Matthew J. Notowidigdo. 2013. "Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment." *Quarterly Journal of Economics* 128 (3): 1123–67.

Lahey, Joanna N. 2008. "Age, Women, and Hiring: An Experimental Study." *Journal of Human Resources* 43 (1): 30–56.

Neumark, David, Ian Burn, and Patrick Button. 2015. "Is it Harder for Older Workers to Find Jobs? New and Improved Evidence from a Field Experiment." National Bureau of Economic Research Working Paper 21669.

Nunley, John M., Adam Pugh, Nicholas Romero, and Richard Alan Seals, Jr. 2014. "Unemployment, Underemployment, and Employment Opportunities: Results from a Correspondence Audit of the Labor Market for College Graduates." Unpublished.

Oberholzer-Gee, Felix. 2008. "Nonemployment Stigma as Rational Herding: A Field Experiment." *Journal of Economic Behavior and Organization* 65 (1): 30–40.

This article has been cited by:

1. Stephen R. G. Jones, W. Craig Riddell. 2019. Unemployment, Marginal Attachment, and Labor Force Participation in Canada and the United States. *Journal of Labor Economics* **37**:S2, S399-S441. [Crossref]
2. Gregor Jarosch, Laura Pilossoph. 2019. Statistical Discrimination and Duration Dependence in the Job Finding Rate. *The Review of Economic Studies* **86**:4, 1631-1665. [Crossref]
3. Henry S. Farber, Chris M. Herbst, Dan Silverman, Till von Wachter. 2019. Whom Do Employers Want? The Role of Recent Employment and Unemployment Status and Age. *Journal of Labor Economics* **37**:2, 323-349. [Crossref]
4. Stijn Baert, Dieter Verhaest. 2019. Unemployment or Overeducation: Which is a Worse Signal to Employers?. *De Economist* **167**:1, 1-21. [Crossref]
5. Alexander Torgovitsky. 2019. Nonparametric Inference on State Dependence in Unemployment. *SSRN Electronic Journal* . [Crossref]
6. Eva Van Belle, Valentina Di Stasio, Ralf Caers, Marijke De Couck, Stijn Baert. 2018. Why Are Employers Put Off by Long Spells of Unemployment?. *European Sociological Review* **34**:6, 694-710. [Crossref]
7. . Bibliography 285-295. [Crossref]
8. Stijn Baert. Hiring Discrimination: An Overview of (Almost) All Correspondence Experiments Since 2005 63-77. [Crossref]
9. Francisco Perez-Arce, Maria Prados, Tarra Kohli. 2018. The Decline in the U.S. Labor Force Participation Rate. *SSRN Electronic Journal* . [Crossref]
10. Henry S. Farber. 2017. Employment, Hours, and Earnings Consequences of Job Loss: US Evidence from the Displaced Workers Survey. *Journal of Labor Economics* **35**:S1, S235-S272. [Crossref]
11. Alexander Torgovitsky. 2015. Partial Identification of State Dependence. *SSRN Electronic Journal* . [Crossref]