

BAN THE BOX, CRIMINAL RECORDS, AND RACIAL DISCRIMINATION: A FIELD EXPERIMENT*

AMANDA AGAN AND SONJA STARR

“Ban the Box” (BTB) policies restrict employers from asking about applicants’ criminal histories on job applications and are often presented as a means of reducing unemployment among black men, who disproportionately have criminal records. However, withholding information about criminal records could risk encouraging racial discrimination: employers may make assumptions about criminality based on the applicant’s race. To investigate BTB’s effects, we sent approximately 15,000 online job applications on behalf of fictitious young, male applicants to employers in New Jersey and New York City before and after the adoption of BTB policies. These applications varied whether the applicant had a distinctly black or distinctly white name and the felony conviction status of the applicant. We confirm that criminal records are a major barrier to employment: employers that asked about criminal records were 63% more likely to call applicants with no record. However, our results support the concern that BTB policies encourage racial discrimination: the black-white gap in callbacks grew dramatically at companies that removed the box after the policy went into effect. Before BTB, white applicants to employers with the box received 7% more callbacks than similar black applicants, but BTB increased this gap to 43%. We believe that the best interpretation of these results is that employers are relying on exaggerated impressions of real-world racial differences in felony conviction rates. *JEL Codes:* J15, J78, K31, K42.

*The authors gratefully acknowledge generous funding from the Princeton University Industrial Relations Section, the University of Michigan Empirical Legal Studies Center, and the University of Michigan Office of Research, without which this study could not have taken place. We thank Will Dobbie, Henry Farber, Alan Krueger, Steven Levitt, Alex Mas, Ezra Oberfield, Emily Owens, Alex Tabarrok, David Weisbach, Crystal Yang, and seminar participants at Princeton University, Rutgers University, the University of Chicago, the University of Michigan, Yale University, UCLA, the University of Pennsylvania, the University of Toronto, the University of Virginia, the University of Notre Dame, Northwestern University, Harvard, MIT, Harvard Law School, the Society of Labor Economists Annual Meeting, the IRP Summer Workshop, the American Law and Economics Association Annual Meeting, the NBER Summer Institute Labor Studies/Crime Session, and especially Larry Katz, Andrei Shleifer, and our anonymous referees for helpful comments. Finally, we thank every member of our large team of research assistants for their hard work and care, especially head RAs Louisa Eberle, Reid Murdoch, Emma Ward, and Drew Pappas, and our ArcGIS experts Linfeng Li and Grady Bridges. This experiment was initially registered with the AEA RCT Registry on April 16, 2015.

© The Author(s) 2017. Published by Oxford University Press on behalf of the President and Fellows of Harvard College. All rights reserved. For Permissions, please email: journals.permissions@oup.com

The Quarterly Journal of Economics (2018), 191–235. doi:10.1093/qje/qjx028.
Advance Access publication on August 2, 2017.

I. INTRODUCTION

Tens of millions of Americans—disproportionately including black men—have criminal records, and face resulting barriers to employment access. In an effort to help overcome those barriers, and thereby reduce racial disparities in employment, more than 150 jurisdictions and 25 states have recently passed “Ban the Box” (BTB) laws and policies ([Rodriguez and Avery 2016](#)). The “box” referred to in BTB (and hereinafter in this article) is the question on a job application form asking whether the applicant has been convicted of a crime, which is often accompanied by yes and no checkboxes. BTB prohibits employers from asking such questions on initial job applications or in interviews. Most BTB laws apply to public employers only, but nine states and several cities have now extended these restrictions to private employers.

BTB seeks to increase employment of people with criminal records. It is often further presented as an important tool for reducing race gaps in employment, especially for improving hiring of black men ([Clarke 2012](#); [Community Catalyst 2013](#); [Pinard 2014](#); and [Southern Coalition for Social Justice 2017](#)), who have recently faced unemployment rates approximately double the national average ([U.S. Bureau of Labor Statistics 2016](#)).¹ But there is a plausible countervailing concern: absent individualized information, employers might instead rely on race-based assumptions about who is likely to have criminal records. If so, BTB could harm black men, in particular those with no records, who lose the ability to convey that fact on job applications.

We investigate BTB’s effects via a field experiment. We submitted nearly 15,000 fictitious online job applications on behalf of young males to entry-level positions, before and after the effective dates of private-sector BTB laws in New Jersey (March 1, 2015) and New York City (October 27, 2015). We sent these applications in pairs matched on race (black and white) and also randomly

1. See, for example, [Minnesota Department of Human Rights \(2016\)](#): “The Ban the Box law can mitigate disparate impact based on race and national origin in the job applicant pool.” New York City’s public-sector BTB law was passed in 2011 as part of an initiative to improve employment of young black and Latino men ([Astone, Katz, and Gelatt 2014](#)). Civil rights organizations are also major BTB backers ([Color of Change 2015](#); [NAACP 2017](#)).

varied whether the applicant had a felony conviction.² This design allows us to test, among other things, how employer reaction to race changes after BTB's adoption.

The basic premise of BTB laws—that criminal records are a major barrier to employment—finds support in prior research (Pager 2003; Holzer, Raphael, and Stoll 2006, 2007). BTB laws only delay (rather than permanently bar) employer access to criminal records; employers may still conduct criminal background checks near the end of the hiring process.³ But the theory is that after meeting applicants in person, employers are less likely to treat records as being categorically disqualifying: “Rejection is harder once a personal relationship has been formed” (Love 2011). Thus, BTB seeks to help candidates with records to get their feet in the door. If BTB does increase hiring of people with records, that should help mitigate racial disparities in employment, because black men are more likely to have records (Shannon et al. 2011; Brame et al. 2014).⁴

On the other hand, BTB could also inadvertently encourage employers who lack criminal record information to rely on race as a proxy. Theories of statistical discrimination have long

2. We use “criminal record” and “felony conviction” (the type of record we varied experimentally) interchangeably here. Job application questions about records are overwhelmingly limited to convictions (not arrests), and usually to felonies.

3. In New Jersey and New York City, the two jurisdictions on which this study focuses, employers may conduct background checks any time after the first interview (New Jersey) or after a conditional job offer (New York). Nationally, a 2012 survey found that 69% of all employers conducted background checks (Society for Human Resources Management 2012), while a survey of 96 major retail chains in 2011 found that 97% of them performed background checks (Allen 2011). Some BTB laws also substantively restrict the role that criminal records can play in employers' ultimate decisions, but New Jersey's and New York's do not. New Jersey's law affects only the “initial employment application process” (N.J. P.L. 2014, Ch. 32). New York requires employers to consider whether a conviction is job-relevant, but this is a longstanding restriction that was unchanged by BTB (N.Y. Correction Law Sec. 752). All U.S. employers have also long been subject to similar restrictions (requiring nuanced assessments of criminal records) at the federal level, pursuant to the Equal Employment Opportunity Commission's interpretation of the Civil Rights Act of 1964 (EEOC 2012).

4. Brame et al. (2014) find that by age 23, 49% of black men have experienced an arrest, versus 38% of white men; Shannon et al. (2011) estimate that 25% of the U.S. black population has a felony conviction, compared with only 6% of the nonblack population. This disparate impact is why EEOC interprets race discrimination law to constrain employers' treatment of criminal records (EEOC 2012).

suggested that decision makers deprived of individualized information might rationally rely on group generalizations instead (Phelps 1972; Arrow 1973; Aigner and Cain 1977; Stoll 2009; Fang and Moro 2011). Alternatively, employers might rely not on accurate information but on exaggerated assumptions or stereotypes about group differences (e.g., Bordalo et al. 2016 provide a theory of the decision process producing stereotyping). Observational studies have examined whether racial disparities in employment are affected by use of employer background checks and expansion of Internet records databases, producing somewhat mixed results.⁵ Other researchers have found evidence of increased reliance on race in reaction to other policies, such as drug testing or credit checks (Wozniak 2015; Bartik and Nelson 2016; Clifford and Shoag 2016).⁶

Our experimental design allows us to explore several related questions. First, we investigate whether employer callback rates vary by race and by felony conviction status. Second, we estimate how taking criminal history questions off the job application pursuant to BTB changes the race gap. This analysis exploits both the variation over time introduced by BTB and the fact that many employers, even before BTB, chose not to ask such questions on applications. We estimate BTB's effects on racial discrimination at affected employers, and we conduct triple-differences analyses that further difference out changes over the same period among similar employers whose applications were unchanged by BTB. In addition, we also test the effects of two other applicant characteristics that could potentially signal a criminal record: GED (versus an ordinary high-school diploma), and a one-year gap in employment history. Finally, we consider whether our results are consistent with rational statistical discrimination or whether employers are relying on exaggerated assumptions of criminality by race.

We report several key findings. The first supports BTB's core premise: when employers ask about them, felony convictions are

5. Bushway (2004) finds that Internet-based criminal records databases are associated with reduced race gaps in employment; in contrast, Finlay (2014) finds that while young black men without records benefit, these databases' net effect on young black men appears to be negative. Holzer, Raphael, and Stoll (2006) and Stoll (2009) find that surveyed employers who report that they use criminal records checks are more likely to hire African Americans.

6. Autor and Scarborough (2008) find that race gaps in retail hiring were unchanged by adoption of a test on which black applicants scored lower, suggesting that employers statistically discriminated before they used the test.

a major employment barrier. Applicants without convictions were 63% more likely to be called back than those with convictions (5.2 percentage points over a baseline of 8.2%). Second, however, BTB does appear to increase racial discrimination. The black-white gap in callbacks at BTB-affected employers grew by nearly 4 percentage points—a large expansion relative to our overall callback rate (11.7%). In our main specifications, this represented a sixfold increase in racial disparity: before BTB, white applicants received 7% more callbacks than similar black applicants, but after BTB this gap grew to 43% (or 45%, when trends at unaffected employers are further differenced out). The GED and employment gap variables, in contrast, did not significantly affect callback rates, and this did not change significantly after BTB.

The post-BTB increase in racial inequality in callback rates appears to come from a combination of losses to black applicants and gains to white applicants. In particular, black applicants without criminal records saw a substantial drop in callback rates after BTB, which their white counterparts did not see. Meanwhile, white applicants with criminal records saw a substantial increase in callbacks, which their black counterparts did not see. This pattern suggests that when employers lack individualized information, they tend to generalize that black applicants, but not white applicants, are likely to have records. Moreover, this phenomenon may contribute substantially to overall racial discrimination patterns. While we find an overall effect of race that is roughly in line with prior auditing studies, that effect is nearly absent at companies with the box before BTB, suggesting that a large share of observed racial discrimination may be driven by criminal record-based assumptions.

Disparate treatment based on race violates employment discrimination laws, whether it is based on accurate statistical generalizations or on inaccurate stereotypes.⁷ However, the difference between the two mechanisms is important to understanding discrimination processes, which may in turn inform policy responses. We therefore seek to disentangle the two by applying our results to

7. Title VII of the Civil Rights Act of 1964, which prohibits race and sex discrimination in employment, does not permit otherwise illegal treatment based on group generalizations, even if they are empirically supported. For example, in *City of Los Angeles Dep't of Water and Power v. Manhart*, 435 U.S. 702 (1978), the Supreme Court held that an employer could not rely, in designing pension benefits, on the actuarial prediction that women live longer.

a simple model of employer decision making. This exercise implies that employers are relying on greatly exaggerated priors about racial differences in felony conviction rates, relative to the racial differences observed in a range of real-world studies. In addition, employers (even after BTB) do not give significant weight to applicant GED status even though real-world data indicate that having a GED instead of a regular high school diploma is a much stronger predictor of criminal records than race is. Together, these findings strongly imply that employers are not engaging in “rational” statistical discrimination, but rather are relying on stereotypes (see, e.g., [Bordalo et al. 2016](#)).

This study makes several distinct contributions to the literature. First, this is the first experimental study of BTB’s effects, and indeed there is little empirical research of any sort on BTB. Two recent studies investigate BTB’s effects on racial disparities using observational data from employment surveys. [Doleac and Hansen \(2016\)](#) find that in the Current Population Survey there is a decrease in employment for young, low-skill black and Hispanic men after BTB goes into effect. [Shoag and Veuger \(2016\)](#) use American Community Survey data and find in contrast that employment of black men increased after BTB, but do not break this down by age and/or education. These papers, unlike ours, focus primarily on public-sector BTB laws. We hope that our experimental method will shed important light on BTB’s effects and inform ongoing legislative debates about BTB throughout the country.

Second, we use field experimental methods to contribute to the literature on statistical discrimination and stereotyping in employment, which has not generally used such methods.⁸ Although our study is not a pure experiment (a key variable, whether the application asks about records, is not manipulated), our ability to perfectly observe and randomize all of our fictional applicants’ characteristics allows us to avoid many of the most likely threats to causal inference that affect purely observational research.

Finally, we make a methodological contribution to the literature on auditing, which has for decades been a central tool for empirical research on discrimination in employment, housing, lending, and other areas. In the employment context, auditing involves

8. See [List \(2004\)](#) for an experimental approach to statistical discrimination in another context, sports card trading.

varying characteristics of interest about a job candidate while holding other characteristics constant. It has been used to test employment discrimination based on race, gender, length of unemployment spell, age, commute time, and type of postsecondary education (Neumark et al. 1996; Riach and Rich 2002; Bertrand and Mullainathan 2004; Lahey 2008; Oreopoulos 2011; Neumark 2012; Kroft, Lange, and Notowidigdo 2013; Farber, Silverman, and Von Wachter 2015; Neumark, Burn, and Button 2015; Deming et al. 2016), as well as the effect of criminal records (Pager 2003; Pager, Bonikowski, and Western 2009; Baert and Verhofs-tadt 2015; Uggen et al. 2014; Decker et al. 2015). But auditing has previously only been used to obtain a one-time snapshot of discrimination patterns—to our knowledge, ours is the first study to use it to assess the effects of a policy. Because researchers cannot randomize the application of the policy itself, using auditing to assess policies requires combining the field-experimental approach with additional methods of causal inference—in this case, difference-in-differences analysis. We believe that combining auditing with quasi-experimental analysis of policy changes enriches the study of discrimination.

II. EXPERIMENTAL DESIGN

We submitted online job applications on behalf of fictitious job applicants to low-skill, entry-level job openings both before and after BTB went into effect in New Jersey and New York City. New Jersey's version of BTB, the Opportunity to Compete Act, became effective March 1, 2015. We submitted applications in New Jersey between January 31 and February 28, 2015 (the pre-BTB period) and between May 4 and June 12, 2015 (the post-BTB period). New York City's BTB law went into effect on October 27, 2015. We submitted applications in New York City between June 10 and August 30, 2015 (the pre-BTB period) and between November 30, 2015, and March 31, 2016 (the post-BTB period). Our main outcome of interest—the “callback”—is whether an employer left a voicemail or email requesting that the applicant contact them or requesting an interview. These phone calls and emails were tracked for eight weeks from the application date.⁹

9. In NJ, pre-BTB period data collection finished after BTB went into effect, but the applications were submitted before it went into effect, so the applications to employers with the box did contain criminal record information.

II.A. Choosing Employers and Job Postings

Our subjects were private, for-profit employers. We relied on two main sources to locate openings. First, we searched two major online job boards: indeed.com, the largest U.S. job website, and snagajob.com, the largest site focused on hourly employment.¹⁰ Second, we also directly searched the employment websites of chain businesses meeting certain criteria.¹¹ We searched for jobs requiring little work experience, no postsecondary education, and no specialized skills: predominantly crew-member restaurant and retail jobs. We focus on these sectors, and specifically on chains, because they typically require online job applications, rather than just evaluating résumés (which do not have a “box” that can be banned). These sectors also are likely to attract applicants with criminal records, who disproportionately have limited work experience and education. The applications were filled out with the help of a large team of University of Michigan student research assistants (RAs). While submitting job applications, the RAs filled out a spreadsheet that indicated details about the employer, position applied to, and the questions asked on the application.

In the post-BTB period, most applications were sent to employers that we had already applied to in the pre-BTB period; these were supplemented with some additional stores from the same chains. Stores thus received up to four applications total, one pair in each period. It was sometimes impossible to send a

10. Prior auditing studies have often identified jobs based on newspaper classified ads; today, these have largely been displaced by online sites, or are included in multisite aggregators like Indeed (Del Castillo 2016).

11. In New Jersey, we applied to chains with at least 30 locations and 300 employees in the state. In New York, we applied to these same chains, plus other chains with at least 20 locations in the city. We excluded employers that did not use online job applications, although the vast majority of chains meeting those size criteria (or advertising on Snagajob or Indeed) do use them. We excluded a few chains due to extremely arduous online application processes, and a few that targeted an overwhelmingly female clientele. Finally, some employers required full SSNs on job applications. For ethical reasons, we avoided using potentially real SSNs, instead using invalid SSNs beginning with 9xx or 666. Some employers had systems that automatically detected these invalid SSNs, and we excluded those businesses. If setting up such a system were correlated with special interest in criminal records, then excluding this pool could reduce our estimates of the effects of a criminal record. However, among employers we did apply to, there was no correlation between whether employers asked for an SSN at all and whether they asked about criminal records.

complete set of four applications, usually because the store was hiring in one period but not the other. In addition, a few RA assignments were not completed before BTB's effective date.¹² As a result, the sample composition by chain, jurisdiction, and specific store is not identical across periods. We address this concern below.

II.B. Applicant Profiles

Our fictitious applicants are all male and approximately 21 to 22 years old.¹³ The RAs filled out applications based on profiles that we created using the *Résumé Randomizer* program created by [Lahey and Beasley \(2009\)](#). Each applicant profile included a name, phone number, address, employment history consisting of two prior jobs, unique email address, two references with phone numbers, information on high school or GED programs, felony conviction status and information about the criminal charge, formatted résumé, and answers to many other routine application questions, such as job availability and pay sought (minimum wage).¹⁴

The profiles were created in pairs of one black and one white applicant, which were assigned to the same store in the same time period. There was a time lag within pairs, with order randomized. In addition to race, other treatment dimensions that we randomized were:

- (i) Has felony criminal conviction or not
 - a. (Conditional on conviction): convicted of property crime or drug crime
- (ii) Has one-year employment gap versus a zero- to two-month gap (referred to as “no gap” below) between the two past jobs
- (iii) GED or high school diploma

12. This occurred mainly in the New Jersey pre-BTB period, which had to be completed relatively quickly. In New Jersey, we filled in these gaps in the post-BTB period whenever possible. In New York City, our pre-BTB wave was quite comprehensive, so we limited the post-BTB wave to stores that we had sent at least one application to pre-BTB.

13. Employers rarely ask about age or high school graduation year due to age discrimination laws but could potentially infer it via length of work history.

14. It was not possible for the profiles to anticipate every question asked, so we relied on the RAs' judgment but provided detailed training about what employers are generally looking for.

Race is indicated via applicant names, as discussed in [Section II.C](#). The felonies we gave our applicants were nonviolent and fairly minor—either property crimes (e.g., shoplifting or receiving stolen property) or drug crimes (e.g., controlled substances possession). Like race, the employment gap and GED variables potentially could be seen by employers as proxies for criminal history.¹⁵

We chose 40 geographically distributed cities/towns in New Jersey and 44 neighborhoods throughout New York City's boroughs to serve as "centers" where the applicants' addresses would be located; each center then served as a base for application to nearby employers.¹⁶ All applicant addresses were in racially diverse, lower- to middle-class neighborhoods. Other applicant characteristics such as work history, address within center, and high school were designed to have similar connotations, but randomized among similar options so they did not appear identical. Most applicant profiles (59% of the sample) were sent to only one store, but we sometimes used the same pairs to apply to multiple nearby locations of the same chain, as real-world applicants might do.¹⁷

15. As of 2005, 13.6% of GEDs were issued in state and federal prisons ([Heckman and LaFontaine 2010](#)). The relationship between GED, race, and criminal records is further addressed in the discussion. The one-year employment gap is meant to signal potential time spent incarcerated. Absence of a gap does not necessarily imply no conviction, however, because offenders are often not incarcerated. Among all individuals charged with felonies in state courts, 62% are not detained before trial; 27% of those convicted receive no incarceration, and of those sentenced to incarceration, approximately 24% received jail sentences of one to three months ([Reaves 2013](#)). Incarceration rates are presumably lower yet among first-time offenders with minor felonies, like our fictional applicants.

16. In New Jersey, we assigned each municipality in the state to its nearest center, minimizing distance. In New York City, because distances are much smaller generally, we prioritized distributing the locations of each chain across centers, and minimized distance within equal-distribution constraints.

17. Our criteria for this grouping differed between New Jersey and New York City. In New Jersey, we were concerned that the same hiring managers might cover multiple locations of chains and might become suspicious upon noticing groups of similar applicants coming within a short time from the same nearby town. Accordingly, we used the same applicant profiles for all locations that were assigned to a given center, as though just one applicant was applying. In New York, our concerns were different: the centers are not towns and likely appear less distinctive to managers, and we had more available time before BTB's effective date, so we were able to space out the timing of our applications (generally by a month or more). Thus, in New York we chose to increase power by sending each application to only one location, except for the largest five chains (in which we sent each applications to up to two or three stores).

For more details on applicant profiles and application procedures, see [Online Appendix A1](#).¹⁸

II.C. Indicating Applicant Race

Race is the central characteristic of interest in our study, and we signal race by the name of the applicant ([Bertrand and Mullainathan 2004](#); [Oreopoulos 2011](#)). To identify racially distinctive names, we used birth certificate data for babies born between 1989 and 1997 from the New Jersey Department of Health (NJDOH), which encompasses the cohort that would include our applicants ([Center for Health Statistics n.d.](#)). We then chose first and last names that were racially distinctive (meeting threshold requirements for the percentage of babies given that name who were black or non-Hispanic white) and common (meeting threshold frequency requirements).¹⁹ Applicants were assigned random first and last names from the appropriate lists, which are provided in [Online Appendix A2](#). The combination of distinctive first and last names should produce a very strong racial signal: according to the birth certificate data, 97% of persons with first and last names on our “black” list are black, and 92% of persons with first and last names on our “white” list are white.

One concern is that racially distinctive names could also signal socioeconomic status (SES), which employers may believe to be correlated with productivity ([Fryer and Levitt 2004](#)). However, our applications provided a great deal of concrete SES information to employers, including work histories, education, current neighborhood, and high school location. With all this information available, employers likely would not need to rely on names to draw SES inferences. To further mitigate this concern, we chose common names (avoiding socioeconomic connotations associated with unusual names) and also limited our white name list to those

18. All appendix material can be found in the [Online Appendix](#).

19. Because blacks are a much smaller fraction of the population, these thresholds varied by race: the minimum percentages were 80% for white first names, 85% for white last names, and 70% for black first and last names, while the minimum frequencies were 450 for white first names, 150 for white last names, 150 for black first names, and 100 for black last names. We eliminated a few first names that were not distinctively male or that had strong associations with a particular religion, to avoid confounding race's effects with other variables. A heavily overlapping list would have been chosen had we followed the approach of [Bertrand and Mullainathan \(2004\)](#) or [Fryer and Levitt \(2004\)](#).

below the white SES median, as measured by maternal education, the best available indicator.²⁰ Finally, racially distinctive names are very common and do not point to an individual being a high- or low-SES outlier within their race.²¹ Thus, even if employers do make assumptions about SES based on such names, similar assumptions would affect a large fraction of real-world job applicants.

III. SUMMARY STATISTICS AND MAIN EFFECTS OF APPLICANT CHARACTERISTICS ON EMPLOYER CALLBACKS

We submitted 15,220 applications, of which 14,637 are in our analysis sample.²² This includes 6,401 applications in New Jersey and 8,236 in New York City. The applications were sent to 4,291 establishments (“stores”) in 293 chains. We begin with summary statistics and then analyze the main effects of applicant characteristics on employer callbacks. The summary statistics and results presented in the tables and figures below combine both jurisdictions; in [Online Appendix A5](#) and [A6](#), we replicate several of the tables and figures for New Jersey and New York separately.

20. It was not possible to create lists that were equivalent on SES; virtually every distinctively white name averages higher than virtually every distinctively black name, due to socioeconomic stratification by race. Although some SES gap remains, it is very similar to the overall SES gap between black and white citizens—that is, choosing distinctive names did not amplify the gap (even if employers were to rely on names to signify SES).

21. In our birth certificate sample, 47% of black children have a racially distinct first name and 36% have a racially distinct last name (as we define distinctiveness, see note 20), while 35% of white children have a racially distinct first name and 65% have a racially distinct last name.

22. The remaining 580 observations (3.8% of those we sent) were dropped for several reasons. First, when an entire chain was applied to only in one period, our key treatment variable (*Box Remover*) could not be coded. Second, some stores had inconsistencies within one or both rounds as to whether the box was present, generally either because of precompliance before BTB’s effective date (occurring between the two applications) or because of RA mistakes (missing disclaimers saying not to answer the criminal record question). In these cases we discarded the observation that was an outlier from the overall chain norm (including the RA-mistake observations, or in the precompliance cases, the later, nonbox observation). Third, we also dropped some businesses (about 1% of the sample) that appeared, mysteriously but presumably due to an administrative mistake, to add the box after BTB and therefore could not be coded as 0 or 1 on the *Box Remover* variable. We add these back in in a robustness check below, with the coding of –1.

TABLE I
MEANS OF APPLICANT AND APPLICATION CHARACTERISTICS AND
CALLBACK RATES BY PERIOD

	Pre-BTB	Post-BTB	Combined
Characteristics			
White	0.502	0.497	0.500
Conviction	0.497	0.513	0.505
GED	0.498	0.502	0.500
Employment gap	0.492	0.504	0.498
Application has box	0.366	0.036	0.199
Results			
Callback rate	0.109	0.125	0.117
Interview req.	0.060	0.067	0.063
Callback rate by characteristics			
Black	0.099	0.111	0.105
White	0.120	0.139	0.129
GED	0.106	0.127	0.117
HSD	0.113	0.122	0.118
Emp. gap	0.110	0.126	0.118
No emp. gap	0.109	0.124	0.116
<i>N</i>	7,245	7,392	14,637

Notes. “Callback” means application received a personalized positive response from the employer (either voicemail or e-mail). Interview request means the positive response message specifically mentioned an interview. “Application has box” means that the application asked about criminal records. “Employment gap” or “Emp. gap” means an 11–13-month employment gap in work history, while “No emp. gap” means a 0–2-month gap.

III.A. Summary

Initial summary statistics are presented in [Table I](#), by period and overall. As expected, approximately 50% of our applications had each of our randomized characteristics, but our other characteristic of interest—whether the application had the criminal-record “box”—could not be randomized. Among pre-BTB applications, 36.2% had the box; after BTB, 3.6% still had it (non-compliers). Thus, 33% of the sample (4,793 observations, 1,383 employers, and 71 chains) consisted of “box removers”: employers that had the box before BTB, but not after. The rate at which employers had the box before BTB may seem surprisingly low, given earlier studies finding rates as high as 80% (see [Uggen et al. 2014](#), reporting results from 2007–2008). A plausible explanation for this difference is the recent success of the BTB movement, which has affected employers (especially national chains) even in jurisdictions without a BTB law. Most of the nonbox employers have no box on their application in any jurisdiction, indicating

TABLE II
CALLBACK RATES BY CRIME STATUS FOR STORES WITH THE BOX IN THE PRE-BTB PERIOD

	No crime	Crime	Property	Drug	Combined
Callback rate	0.136	0.085	0.084	0.085	0.110
Callback black	0.131	0.086	0.091	0.081	0.109
Callback white	0.140	0.083	0.077	0.089	0.111
<i>N</i>	1,319	1,336	703	633	2,655

Notes. Sample restricted to pre-BTB period applications where the application asked about criminal records. Callback implies application received a personalized positive response from the employer.

that very few reflect early compliance with the New Jersey or New York BTB laws specifically.

Overall, 11.7% (1,715) of our applications received callbacks.²³ This rate was higher in the post-BTB period (12.5% versus 10.9%), and lower in New York City than in New Jersey (9.4% versus 14.7%; see [Online Appendix](#) Tables A5.1a and A6.1a). Among the callbacks, about 55% specifically mentioned an interview (though many others were likely seeking interviews even if the message did not so specify). The race gap in callback rates grew from 2.1 percentage points in the preperiod to 2.8 percentage points in the postperiod; these averages do not differentiate box-remover employers from others and mask large changes occurring at box-remover employers, as shown below. Callback rates hardly differed by GED/diploma status or employment gap.

[Table I](#) does not break down callback rates by criminal record status because criminal record is unobserved by most employers (those without the box). [Table II](#) thus shows separate summary statistics limited to pre-BTB applications to employers with the box. Callback rates were 60% higher for applicants without criminal records (5.1 percentage points, over a base rate of 8.5%). Applicants with drug and property crime convictions had similar callback rates—perhaps surprisingly, as one might have expected employers to be particularly concerned about potential employee theft. When employers asked about records, we saw essentially no race gap in callback rates (11.1% for whites, 10.9% for blacks).

23. This rate is similar to other recent audit studies: [Kroft, Lange, and Notowidigdo \(2013\)](#) had a positive response rate of 11.6%; [Farber, Silverman, and Von Wachter \(2015\)](#) had a 10.4% callback rate; [Deming et al. \(2016\)](#) had an 8.2% callback rate.

TABLE III
EFFECTS OF APPLICANT CHARACTERISTICS ON CALLBACK RATES

	(1)	(2)	(3)
White	0.024*** (0.006)	-0.001 (0.009)	-0.001 (0.009)
Conviction	-0.014** (0.005)	-0.052*** (0.012)	
GED	-0.004 (0.005)	0.010 (0.014)	0.010 (0.013)
Employment gap	0.002 (0.005)	0.011 (0.010)	0.011 (0.010)
Pre-BTB period	-0.015 (0.010)		
Drug conviction			-0.050*** (0.013)
Prop. conviction			-0.054*** (0.014)
N	14,637	2,918	2,918
Sample	All	Box	Box
Chain FE	Yes	Yes	Yes
Center FE	Yes	Yes	Yes

Notes. Dependent variable is whether the application received a callback. Standard errors clustered on chain in parentheses. Chain and geographic center fixed effects are included in all regressions. White is an indicator for race (versus black), Conviction is an indicator for whether the applicant has a felony conviction, GED is an indicator for having a GED (versus a regular high-school diploma), and Employment gap is an indicator for whether the applicant has an 11–13-month gap in work history between the previous two jobs (versus a 0–2-month gap). “Drug conviction” and “Prop. conviction” break the Conviction variable into a categorical variable based on crime type (drug versus property crime); no conviction is the base category. The box sample is employers with the box on their application. *10%, **5%, and ***1% significance level.

The advantage for applicants without records was slightly larger for white applicants (5.7 percentage points, or 69% above the base rate of 8.3%) than for black applicants (4.5 percentage points, or 52% above the base rate of 8.6%), although regressions not shown here show that the race-record interaction is not statistically significant.

III.B. Regression Estimates of Main Effects of Applicant Characteristics

Table III provides regression estimates of the main effects of race, record, GED/diploma status, and employment gap on callback rates; the regressions also include fixed effects for the chain (with the smallest chains grouped by business category) and the

geographic center.²⁴ These estimates parallel the summary statistics, which is not surprising given that applicant characteristics were distributed randomly. Column (1) shows results for the full sample. White applicants were about 2.4 percentage points (23%) more likely to receive a callback. In contrast, callback rates did not vary based on the GED and employment gap variables.

To assess the effect of having a criminal record when employers observe it, we show analyses in columns (2) and (3) that are limited to employers with the box. Column (2) shows a 5.2-percentage-point criminal record effect ($p < .01$), which translates into a 63% higher callback rate for applicants without records compared to the 8.2% baseline for applicants with records. Column (3) shows that this effect was similar for property and drug crimes. In [Agan and Starr \(2017a\)](#), we report additional statistics detailing the criminal record effect, including variations by applicant race, industry, local demographics, and crime rates. Meanwhile, the main effect of race is economically and statistically insignificant in the box sample, a point further examined in the remainder of this article, which assesses the effect of the box on racial discrimination.

In [Online Appendix A4](#) we show that the main effects of race and crime are robust to several alternative specifications and samples.²⁵ The biggest differences are geographic. The “white” effect is far larger in New Jersey (4.5 percentage points, or 37% more callbacks for whites) than in New York City (0.7 percentage points, or 8% more callbacks for whites).²⁶ The criminal record effect, in contrast, is larger in New York City, at least in proportional terms.²⁷ At box employers, applicants without records receive 45% more callbacks than those with records in New Jersey; in New York City, applicants without records receive 78% more callbacks.

24. All the results shown in [Table III](#) are for both periods combined (unlike [Table II](#) and [Figure II](#), which were for the pre-BTB period only), but the regression results look similar if only the pre-BTB observations are used.

25. These robustness checks parallel those discussed below concerning the analyses of BTB’s effects ([Table VI](#)).

26. In a separate article, we further explore the geographic variations in the race effect. We find that businesses in whiter neighborhoods much more strongly favor white applicants, suggesting that New York City’s greater racial diversity could partially (but not fully) explain its smaller race gap. For a preliminary version, see [Agan and Starr \(2016\)](#).

27. New Jersey’s callback rate was higher, so similar percentage-point gaps translate into different proportional effects.

IV. THE CRIMINAL RECORD BOX AND RACIAL DISCRIMINATION

In this section, we show difference-in-differences analyses that shed light on the effect of criminal record information on racial discrimination in callbacks. We exploit two different sources of variation in whether employers have the box. First, in [Section IV.A](#), we briefly examine the cross-sectional variation between employers that (before BTB) chose to ask about criminal records (“box employers”) and those that did not. Second, in [Section IV.B](#), we assess the temporal change after BTB for employers that had the box before BTB and then removed it (and show that no similar change existed for employers unaffected by BTB). Third, in [Section IV.C](#), we employ a triple-differences analysis that combines both these sources of variation.

IV.A. Cross-Sectional Difference-in-Differences Estimates

In [Table III](#), column (1), we compare race gaps between employers that do and do not have the criminal record “box.” We employ a simple difference-in-differences specification for the probability that applicant i to store j receives a callback:

$$\begin{aligned} \text{Callback}_{ij} = & \alpha + \beta_1 \text{Box}_j + \beta_2 \text{White}_i \\ (1) \quad & + \beta_3 \text{Box}_j \times \text{White}_i + \Gamma \mathbf{X}_i + \epsilon_{ij}. \end{aligned}$$

Box_j indicates whether store j has the box, White_i indicates applicant race, $\text{Box}_j \times \text{White}_i$ is the interaction of those variables, and \mathbf{X}_i is a vector of control variables: GED, employment gap, and geographic center.

In column (1), we present results from [equation \(1\)](#) in the pre-BTB sample only. This sample includes 7,245 observations in 3,874 stores and 293 chains. The $\text{Box}_j \times \text{White}_i$ coefficient in column (1) indicates that the black-white gap is 2.8 percentage points larger among nonbox employers ($p < .05$) in the preperiod. Among box employers, the race gap is just 0.3 percentage points (in proportional terms, white applicants received 2% more callbacks than black applicants did). Among employers without the box, the gap was about 10 times as large, 3.1 percentage points compared to a base rate for black applicants of 9.4%; in proportional terms, white applicants received 33% more callbacks than black applicants did.

[Figure I](#) provides a visual representation of this cross-sectional comparison that further breaks down box employers’

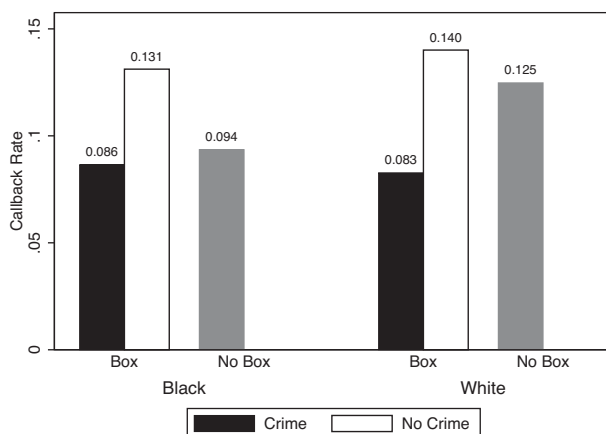


FIGURE I

Callback Rates by Race, Crime, and Box: Preperiod Applications Only

This figure compares callback rates within the preperiod before Ban the Box went into effect, comparing applications with the criminal record question box and those without the box.

callback rates by applicants' conviction status. This figure shows that at box employers, while conviction status itself dramatically affects callback rates, race makes little difference. Among applicants with records, black callback rates are slightly higher (8.6% versus 8.3%), while among applicants without records, white callback rates are higher (14.0% versus 13.1%); neither difference is significant. Again, only at nonbox employers does a significant race gap emerge (12.5% versus 9.4% for white and black applicants, respectively).

This analysis suggests that employers who lack individualized information might be relying on race-based assumptions about criminal record status. Conversely, the absence of a race effect among box employers is intriguing, providing a sharp contrast from other auditing studies, which do not differentiate by "box" status and which generally do find race gaps (as we did in the full sample; see [Table III](#), column (1)). A possible implication is that a substantial share of the racial discrimination observed in other studies might be driven by employers who lack criminal record information and make negative assumptions about black applicants' criminality.

However, the two groups of employers could differ in other ways, such that the patterns we observe do not relate to the

box. In [Online Appendix A3](#), we show that the two groups do not vary substantially across most observable characteristics: index crime rates of store neighborhood, black and white percentages of the neighborhood population, average number of employees, and average sales volume.²⁸ Retail employers are noticeably more likely to have the box than other employers (which are mainly restaurants); however, additional regression analyses (not shown here) indicate that this difference is not what drives the $Box_j \times White_i$ interaction.²⁹ Overall callback rates were also nearly identical at the two groups of employers. Still, unobservable differences between the two groups are possible, so the cross-sectional analysis is only suggestive; more rigorous causal identification is left to the analyses below.

IV.B. Temporal Difference-in-Differences Estimates of BTB's Effects

Because BTB introduced exogenous variation in whether employers have the box, we need not rely only on comparing different groups of employers. In columns (2), (3), and (4) of [Table IV](#), we present difference-in-differences analyses exploiting this temporal variation, limited to “box remover” employers: those that had the box before BTB and removed it afterward. Like the column (1) analysis, these regressions implement [equation \(1\)](#). However, the source of variation in the Box variable is now time (and the intervening policy change), rather than cross-sectional variation among businesses. Thus, Box_j becomes equivalent to an indicator for the pre-BTB period.

In column (2), the analysis is limited to box-remover stores to which we were able to send a complete set of four applications (the “box-remover balanced sample”): 3,712 observations in 928 stores and 62 chains. This limitation means that the sets of employers being compared are identical before and after BTB, so there is no cross-sectional variation being inadvertently introduced due, for example, to different openings being available in different

28. Because the industry difference may be correlated with other characteristics, we also run a simple regression of whether a store has the box on all these characteristics, showing that the only characteristic with a substantial and significant effect is the retail indicator. See [Online Appendix A3](#).

29. Specifically, we add a $Retail_j$ indicator, interacted with $White_i$, to the column (1) regression. The $Box_j \times White_i$ coefficient only increases in magnitude, to 3.4 percentage points ($p < .05$). Similar analyses also show that the small differences in the other observable employer characteristics in [Online Appendix A3](#) do not explain the $Box_j \times White_i$ effect.

TABLE IV
EFFECTS OF THE BOX ON RACIAL DISCRIMINATION: DIFFERENCE-IN-DIFFERENCES

	(1)	(2)	(3)	(4)	(5)
Box \times white	-0.030**	-0.036**	-0.033**	-0.027**	0.002
(White \times pre, column (5))	(0.015)	(0.014)	(0.014)	(0.013)	(0.014)
White	0.032***	0.044***	0.040***	0.123	0.022**
	(0.012)	(0.013)	(0.012)	(0.132)	(0.009)
Box	0.015	0.003	-0.002	-0.345**	-0.016
(Pre, column (5))	(0.024)	(0.015)	(0.013)	(0.139)	(0.017)
<i>N</i>	7,245	3,712	4,794	4,794	7,476
Controls	Yes	Yes	Yes	Yes	Yes
Center FE	Yes	No	Yes	Yes	No
Chain FE	No	No	No	Yes	No
Post \times chain FE	No	No	No	Yes	No
White \times chain FE	No	No	No	Yes	No
Box variation	Cross-section	Temporal	Temporal	Temporal	None
Sample	Pre-BTB	Box remover -balanced	Box remover -full	Box remover -full	Other empl. balanced

Notes. Standard errors clustered on chain in parentheses. Dependent variable is whether the application received a callback. Box removers are stores that had the box in the pre-BTB period and removed it after BTB. “Box removers-balanced” consists of box remover stores to which we sent exactly four applications, one white/black pair in each period. Fixed effects can include geographic center, chain, post \times chain, and white \times chain, and are included as indicated; note that because of the inclusion of interacted fixed effects in column (4), the white and box coefficients are not meaningful. Controls are whether the applicant had a GED (versus regular high-school diploma) and whether he had an employment gap. Box variation indicates the source of variation in the box variable: “Cross-section” means the variation comes from a comparison of box and nonbox stores in the preperiod; “Temporal” means the variation is pre- and post-BTB, triggered by the implementation of the BTB policy. In the last column, which is shown as a comparison point, there is no box variation; the pattern over the same time period is shown for companies that did not change their job applications. *10%, **5%, and ***1% significance level.

seasons. Accordingly, the vector of controls in column (2) includes GED status and employment gap but not the geographic center, which is already perfectly balanced. In column (3), we show nearly the same analysis but in the full sample of box removers, adding back the center fixed effects. This sample is larger (4,794 observations in 1,383 stores and 71 chains), at the cost of some imbalance in the employers represented across time periods. In column (4), also carried out in the full box-remover sample, we add chain fixed effects and interact them with Box_j and $White_i$, which accounts for the pre/post imbalances in chains but not individual stores.³⁰

Each of these analyses shows that racial discrimination increased substantially when these companies removed the box to

30. In this regression, the main effects of Box and $White$ do not have a meaningful interpretation because those variables’ effects are spread across the interacted fixed effects. The principal term of interest, $Box \times White$, retains its interpretation.

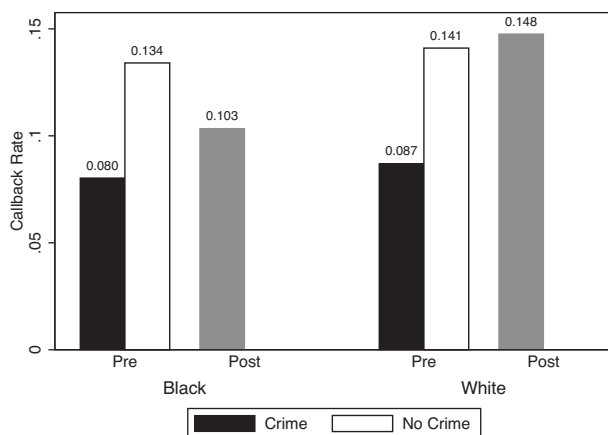


FIGURE II

Callback Rates by Race, Criminal Record, and Period: Balanced Box Removers Only

This figure compares callback rates before and after Ban the Box went into effect, among companies that had the criminal record question box before BTB and removed it afterward, in the balanced sample only (i.e., stores to which we sent complete application pairs in both the pre-BTB and post-BTB periods).

comply with BTB. The $Box_j \times White_i$ coefficient is -3.6 percentage points in the balanced sample (column (2))—that is, when these employers had the box, the race gap was 3.6 percentage points smaller than after it was removed. In the pre-BTB period, the white callback rate was about 0.8 percentage points higher than the black baseline of 10.7% ; in proportional terms, whites received 7% more callbacks. In the post-BTB period, after these companies dropped the box, this race gap ballooned. The white callback rate was 4.4 percentage points higher than the black baseline of 10.4% ; in proportional terms, whites now received 43% more callbacks. The $Box_j \times White_i$ coefficient is slightly smaller in the full sample (3.3 percentage points and 2.8 percentage points without and with the interacted chain fixed effects, respectively), but the overall pattern is very similar: a multifold enlargement of the black/white gap after BTB. The effects are all statistically significant, with p -values ranging from $.01$ to $.03$.

Figure II represents these patterns visually (for the balanced sample of box-removers).³¹ Like Figure I, it further decomposes

31. The figure looks very similar if the full box-remover sample is used instead.

the callback rate when employers have the box (here labeled as “Pre”) based on criminal record status. The pattern is similar to that observed in Figure I. Before BTB, callback rates are only slightly higher for white applicants (whether with or without records), but they become substantially higher (15% versus 10.3% for black applicants) after BTB.

The temporal difference-in-differences analysis is more causally rigorous than the cross-sectional comparison: it compares results across the same employers, just a few months apart, facing identical pools of fictional applicants. It seems unlikely that non-box-related differences would explain such a sharp increase in the race gap, and this intuition is further supported by the striking similarity to the cross-sectional results. Still, the estimates could potentially be confounded by trends unrelated to BTB. For example, if adoption of BTB reflects a motivation to address racial disparities in employment, that motivation could in theory affect disparity trends in other ways.³² An initial check of this possibility is to run a similar difference-in-differences analysis in the sample of employers whose applications were unchanged after BTB (predominantly employers that never had the box). These results are in column (5), which reflects the same specification as column (2), carried out in the balanced sample.³³ In sharp contrast to the box-remover employers, among other employers there was essentially no change in the black-white gap between the pre-BTB and post-BTB periods (indeed, the sign is flipped, though this difference is insignificant).³⁴ To further address these concerns, we turn to the triple-differences analysis.

IV.C. Triple-Differences Estimates of BTB's Effects

Here, we further analyze the causal effect of BTB on racial discrimination via a difference-in-difference-in-differences analysis, which exploits both sources of variation (cross-sectional and temporal) discussed above. This analysis compares the change

32. Seasonal variation is also possible, although this possibility is mitigated by the fact that the timing of the NYC and NJ studies was nearly seasonally opposite.

33. This analysis substitutes *pre* for *box* in equation (1), but it is still directly parallel to columns (2) through (4), which could have those variables labeled either way (within the box-remover sample, *pre* is equivalent to *box*).

34. Column (5) does show an overall higher callback rate in the post-BTB period, which is also seen in the box-remover stores, and which is presumably unrelated to BTB.

TABLE V
EFFECTS OF BAN THE BOX ON RACIAL DISCRIMINATION: TRIPLE DIFFERENCES

	(1)	(2)	(3)
Box Remover × post × white	0.039* (0.020)	0.040** (0.018)	0.035* (0.018)
Post × white	−0.002 (0.014)	−0.006 (0.012)	−0.006 (0.013)
Box Remover × post	−0.019 (0.023)	−0.011 (0.019)	
Box Remover × white	−0.017 (0.015)	−0.021 (0.014)	
Box Remover	0.016 (0.028)	0.009 (0.024)	
White	0.024** (0.012)	0.028** (0.011)	0.098 (0.129)
Post	0.016 (0.017)	0.012 (0.015)	0.339** (0.139)
N	11,188	14,637	14,637
Controls	Yes	Yes	Yes
Center FE	No	Yes	Yes
Chain FE	No	No	Yes
Post × chain FE	No	No	Yes
White × chain FE	No	No	Yes
Sample	Balanced	Full	Full

Notes. Standard errors in parentheses clustered on chain. Dependent variable is whether the application received a callback. A “box remover” store is one that had the box in the pre-BTB period and then removed it due to BTB. The balanced sample consists of stores to which we sent exactly four applications, one white/black pair in each period. Controls are whether the applicant had a GED (vs. regular high-school diploma) and whether he had an employment gap. Fixed effects for the chain, post × chain, white × chain, and geographic center are included as indicated. *10%, **5%, and ***1% significance level.

in racial discrimination after BTB at box-remover employers to changes over the same period at other employers. This analysis will “difference out” the effect of any non-BTB temporal differences so long as they affect both sets of employers similarly. Similarly, the analysis effectively controls for unobserved cross-sectional differences between the two groups of employers so long as they are time-invariant over the period in question. The change in the race gap that remains after this differencing-out is interpreted as the causal effect of a chain’s compliance with BTB.

In Table V, columns (1) and (2), which are carried out in the balanced sample and the full sample, respectively, we apply the following triple differences estimating equation for the probability that applicant i to store j in time period t receives a

callback:

$$\begin{aligned}
 \text{Callback}_{ijt} &= \alpha + \beta_1 \text{White}_i + \beta_2 \text{Post}_t + \beta_3 \text{BoxRemover}_j \\
 &+ \beta_4 \text{White}_i \times \text{Post}_t + \beta_5 \text{White}_i \times \text{BoxRemover}_j \\
 &+ \beta_6 \text{Post}_t \times \text{BoxRemover}_j \\
 (2) \quad &+ \beta_7 \text{BoxRemover}_j \times \text{White}_i \times \text{Post}_t + \Gamma \mathbf{X}_i + \epsilon_{ijt}.
 \end{aligned}$$

Callback_{ijt} indicates whether the applicant received a callback, and Post_t indicates whether the application was sent post-BTB. Box Remover_j is an indicator (coded at the store level) for whether the store had the box before BTB and removed it after BTB. It is coded as 0 if the store never had the box, and also in the rarer case of stores that had the box and failed to remove it after BTB. \mathbf{X}_i is a vector of control variables including GED, employment gap, and (in the full-sample analysis only) geographic center fixed effects. The main effect of interest is the triple-differences coefficient, β_7 , which tells us how the employer callback gap for whites versus blacks changes differentially after BTB for box-remover versus other stores. In column (3), to account for imbalances across chains in the pre- and postperiods, we substitute chain fixed effects, interacted with White_i and Post_t , in place of the BoxRemover_j , $\text{White}_i \times \text{BoxRemover}_j$, and $\text{Post}_t \times \text{BoxRemover}_j$ terms in the equation above.³⁵

The triple-differences estimate ($\text{BoxRemover}_j \times \text{White}_i \times \text{Post}_t$) in every column is large: 3.9 percentage points in the balanced sample, and 4.0 or 3.5 percentage points in the full sample depending on whether the interacted chain fixed effects

35. This produces the following estimating equation: $\text{Callback}_{ijkt} = \alpha + \beta_1 \text{White}_i + \beta_2 \text{Post}_t + \sum_{i=1}^N \beta_{3i} \text{Chain}_k + \beta_4 \text{White}_i \times \text{Post}_t + \text{White}_i \times \sum_{i=1}^N \beta_{5i} \text{Chain}_k + \text{Post}_t \times \sum_{i=1}^N \beta_{6i} \text{Chain}_k + \beta_7 \text{BoxRemover}_j \times \text{White}_i \times \text{Post}_t + \epsilon_{ijkt}$ where now k indexes chains to which store j belongs, and Chain_k are dummy variables for each of the chains in our sample. Note that this analysis does not provide meaningful estimates for the effects of White or Post because they are diffused across the interacted fixed effects. However, $\text{BoxRemover} \times \text{White} \times \text{Post}$ retains the same interpretation as in equation (2) above. The smallest chains (fewer than three locations or 12 total observations) are combined into industry-category groups; these chains represent about 9% of the sample. Use of the original coding does not affect the coefficient. In addition, in the unusual cases where “box remover” status varies between stores within a chain (usually between New Jersey and New York City), we assign separate Chain fixed effects to box-remover and non-box-remover subsets of such chains. The result is that the Chain fixed effects perfectly parallel the Box Remover variable.

are included. This implies that in the balanced sample the white callback-rate advantage over identical black applicants grew by 3.9 percentage points after BTB. These samples and specifications parallel the three analyses presented in the temporal difference-in-differences results above, but with a third difference (box remover versus other employers) added. If one compares the triple-differences estimates in Table V to the corresponding $Box_j \times White_i$ estimates in Table IV (columns (2)–(4), respectively), in each case, the triple-differences effect estimate is slightly larger in magnitude (with signs reversed, because we are now evaluating the effect of removing the box). This pattern suggests that unrelated temporal trends are not what drove the large post-BTB expansion in racial discrimination at companies that removed the box. As we saw in Table IV, column (5), no such changes were observed at companies whose job applications were unaffected by BTB—indeed, if anything, the underlying trends cut slightly in the opposite direction.

Instead, these analyses provide evidence that BTB increases racial discrimination in employer callbacks. Prior to the adoption of BTB, racial disparities are somewhat larger among the stores that do not have the box. After BTB, that difference flips. The growth in the “white” effect after BTB is quite dramatic. In column (1) (the balanced-sample analysis, which we consider our main triple-differences specification), the estimated race gap at box-remover stores goes from 0.7 percentage points before BTB to 4.7 points after. Comparing this to the baseline callback rate, whites receive 6.7% more callbacks than similar black candidates when employers are able to observe criminal records, but 45.2% more callbacks than similar black candidates when employers cannot observe records. In other words, the race gap grows by a factor of almost 7.

Despite its advantages in terms of causal identification, the triple-differences approach comes at a cost in statistical power. Three-way interactions demand much larger samples than analyses of main effects or two-way interactions do to provide equivalent power to estimate effects of a given size; hence, the corresponding Table IV estimates are more precise even though they are estimated in smaller subsamples. Even so, the triple-differences estimates in Table V are significant at the .05 threshold or very close to it, with p -values ranging from .029 to .052. Moreover, they confirm the patterns observed with more

precision in the double-differences analyses—and provide additional confidence in interpreting these estimates as causal.

Interpretation of the triple-differences estimates as causal does rely on the assumption that, absent BTB, trends in employer callback differences by race would have been the same for employers that had the box in the preperiod and those that did not. Although our data are not long enough to compare preperiod trends, we believe the assumption is very plausible. For a vast majority of employers in our sample (even those that are franchised), the job applications are standardized nationally at the chain level, with built-in variations accommodating local differences in BTB laws.³⁶ Thus, the decision to include or not include the box on the application is made at the chain level, whereas callback decisions are made at the individual store level by store managers, or in some chains by local managers who supervise a small subset of locations. In that sense, whether a store has the box should be exogenous to the decision makers we are studying. Thus, we consider it appropriate to consider “box remover” status as a “treatment” variable; from the perspective of the local decision makers, it is something that is imposed on them, rather than a choice.

Moreover, there is no qualitative reason to believe that box-remover chains differ from other chains in any way that would affect hiring trends in a racially disparate way (see [Online Appendix A3](#) for characteristics of box-remover and other employers). To pose a threat to identification in the triple-differences analysis, hiring differences across those two groups of employers would have to be racially disparate in a way that also differs over the short time between our pre- and postperiod applications (about four months on average). Note that not having the box does not generally reflect lack of interest in criminal records; chains with and without the box routinely do back-end background checks, and their applications usually warn applicants of this fact.

IV.D. Alternative Specifications and Samples

[Table VI](#), Panels A and B show robustness checks for the balanced-sample estimates of BTB’s effects on racial discrimination. Panel A shows alternative estimates for the $Box_j \times White_i$ coefficient from the temporal difference-in-differences analysis; the corresponding main specification result

36. For such chains, applications that normally have the box will usually either omit it if the store being applied to is in a BTB jurisdiction, or instruct the applicant not to answer the question if applying in certain jurisdictions.

TABLE VI
EFFECTS OF BAN THE BOX ON RACIAL DISCRIMINATION: ROBUSTNESS CHECKS, BALANCED SAMPLE

	Main (1)	Interview (2)	Add rev. compl. (3)	Drop RA errors (4)	NJ only (5)	NYC only (6)	Intent-to-treat (7)	Retail int. (8)
Panel A: Temporal difference-in-differences								
Box \times white	-0.036** (0.014)	-0.041** (0.016)	-0.034** (0.013)	-0.035** (0.013)	-0.050* (0.028)	-0.029 (0.019)	-0.027** (0.013)	
(<i>Pre \times white, column (7)</i>)								
<i>N</i>	3,712	3,712	3,848	3,686	1,400	2,312	4,112	
Panel B: Triple differences								
Box Remover \times post \times white	0.039* (0.020)	0.039** (0.019)	0.034* (0.018)	0.038* (0.020)	0.048 (0.040)	0.034 (0.021)	0.026 (0.020)	0.041* (0.023)
(<i>Preperiod box \times post \times white, column (7)</i>)								
<i>N</i>	11,188	11,188	11,324	11,160	4,376	6,812	11,188	11,188

Notes. Standard errors clustered on the chain in parentheses. Panel A shows robustness checks corresponding to the temporal difference-in-differences specification from Table IV, column (2); Panel B shows robustness checks corresponding to the triple-differences specification from Table V, column (1). All analyses shown are carried out in the balanced sample, consisting of stores to which we sent two applications each in the pre-BTB and post-BTB periods. Except as described below, the Panel A regressions are further confined to “box removers” (employers that had the box before BTB and removed it afterward), whereas the Panel B regressions include non-box-removers, which provide the third difference. In each panel, column (1) shows the coefficient of interest (*Box \times white* or *BoxRemover \times Post \times White*, respectively) from the respective main specification. The remaining columns show modifications of the sample or specification. Column (2) uses the interview as the dependent variable (all others use “callback”). Column (3) adds observations (“reverse compliers”) that had the box in the post-BTB period but not in the preperiod; in the triple-differences analysis, these are coded -1 on the *Box Remover* variable. Column (4) drops instances where RAs erred and answered a box question they were not required to answer or did not answer one they should have. Columns (5) and (6) are restricted to the New Jersey and New York City subsets of the sample, respectively. Column (7) shows an “intent to treat”-style analysis that groups companies that had the box but failed to omit it after BTB (noncompliers) with the box removers rather than with non-box-removers. In both panels, this requires changing the variable of interest and its interactions: in Panel A *Pre* (a pre-BTB period indicator) is substituted for *Box* and the coefficient shown is *Pre \times White*; and in Panel B *PrePeriodBox* (an indicator for whether the store had the box before BTB) is substituted for *Box Remover*, and the coefficient shown is *Pre period box \times Post \times White*. In column (8) (Panel B only), a *Retail* industry indicator (interacted with *Post*, *White*, and *Post \times White*) is added to the regression. * 10%, ** 5%, and *** 1% significance level.

comes from Table IV, column (2). Panel B shows alternative estimates for the $\text{BoxRemover}_j \times \text{White}_i \times \text{Post}_t$ coefficient from the triple-differences analysis; the corresponding main specification result comes from Table V, column (1). The main specification coefficient is shown in column (1) of each panel for comparison purposes. The two sets of results shown complement one another, one more precise and the other addressing additional causal inference challenges. In every specification and sample, point estimates indicate an economically large (ranging from 2.6 percentage points to 5.0 percentage points) post-BTB increase in racial discrimination. The effect sizes and precision are similar to the main specifications for most variants (with p -values close to .01 for the double-differences and .05 for the triple-differences), with exceptions discussed below.

Column (2) in both panels replaces the callback outcome variable with another variable called “interview”: whether an employer’s message specifically mentioned an interview. In percentage-point terms, the estimates are similar, but because baseline “interview” rates were much lower (7.5% in the Panel A sample and 6.3% in the Panel B sample) than the corresponding callback rates, BTB’s apparent effect on “interview” disparities was considerably larger in proportional terms. However, we believe the interview variable is not really a good measure of whether an interview is sought (whether the employer happened to say that word generally seemed arbitrary), and that “callback” is thus the better measure.

Columns (3) and (4) in both panels alter choices that we made about how to deal with small groups of “problem” observations. Column (3) adds a group we had excluded: “reverse complier” stores that had no box before BTB but (apparently due to administrative mistakes) added it after BTB. *Box Remover* cannot be coded as 0 or 1 for these observations, but in Panel B, column (3), we code it as -1 , reflecting the reversal of the usual treatment direction.³⁷ The effect sizes in both panels are smaller, but only slightly. In column (4), we exclude a small number of observations (about 0.3% of the full balanced sample and 0.7% of the box-remover sample) in which an RA mistakenly answered a

37. The relationship between treatment and the passage of time is inverted for these observations, making these specifications diverge from a standard difference-in-differences analysis. This is the primary reason we excluded them from both sets of the main analyses.

“box” question that she was not required to answer, or vice versa.³⁸ Excluding them leaves both estimates virtually unchanged.

Columns (5) and (6) divide the sample between New Jersey and New York City, respectively. The large reduction in sample size renders these analyses underpowered for the purpose of estimating triple differences (or even two-way interactions confined to the box-remover subset of each sample), and thus these estimates are quite imprecise and should not be given much interpretive weight. In any event, although the New Jersey point estimates are larger in percentage-point terms, they are very similar in relative terms, once one accounts for New Jersey’s substantially higher callback rate (14.7% versus 9.0% in New York City, in the balanced samples).

In our main analyses, we treated the companies that retained the box after BTB (noncompliers) as part of the non-box-remover control group. We consider this to be the most appropriate categorization, because their applications did not change and hiring managers continued to be informed of criminal records, so one should expect no BTB-driven changes in racial discrimination. Moreover, the failure to comply clearly appears to be the result of choices (or, almost certainly, administrative mistakes) made at chains’ national headquarters. Because this is effectively exogenous from the perspective of the local managers whose decisions are being studied, we do not particularly worry about non-compliance introducing treatment selection bias. If one’s interest is in BTB’s net market effects on box employers, noncompliance might reasonably be viewed as an offsetting component of those effects—albeit in most cases a temporary component, lasting until employers discover the mistake.³⁹ But we are more interested conceptually in understanding how access to criminal record information affects hiring managers’ use of race-based assumptions about records and their resulting willingness to racially discriminate. This access did not change for noncompliant chains, so we consider them untreated observations.

38. The main sample kept RA-error cases if the same error was made consistently within the store; we coded *Box Remover* according to how the RA interpreted the application, since that tracked the information about criminal records that the RA provided to the employer.

39. We have been able to recheck most noncomplier applications, confirming that as of February 2017, most have complied with BTB.

Nonetheless, for readers who are more interested in BTB's effect on all preperiod box employers in the aggregate, in column (7), we take an alternate approach, treating the non-compliers as equivalent to box removers—effectively, an intent-to-treat analysis. Thus, in Panel A, we add them to the sample and replace the *Box* and *Box* \times *White* variables with *Pre* and *Pre* \times *White* (evaluating the effect of the passage of time on businesses that initially had the box, regardless of whether they actually removed the box). In Panel B, we replace the “box-remover” treatment variable with a “preperiod box” variable, and change the interaction terms accordingly. These changes reduce the magnitude of the point estimates (to 2.7 percentage points in Panel A and 2.6 percentage points in Panel B); they remain economically quite large, and statistically significant in Panel A but lose statistical significance in the triple differences specification in Panel B. This change is not surprising: recoding 400 noncomplier observations as though they were box removers naturally attenuates the estimates of the effects of box removal toward zero.

In column (8) of Panel B, we address a concern implicating only the triple-differences analysis. While the “box-remover” employers were qualitatively similar across numerous dimensions to the control-group employers ([Online Appendix A3](#)), they did include a larger share of retail and lower share of restaurant employers. One might worry that these industries experienced different time trends (affecting black and white applicants differently), and that this might explain our results. Although we know of no specific such trend to worry about, to test the general theory, we added an indicator for *Retail*, interacted with *White*, *Post*, and *Post* \times *White*. Column (8) shows that the *Box Remover* \times *Post* \times *White* coefficient is if anything slightly larger when these terms are added, albeit less precise. Similar analyses (not shown here) demonstrate that the other employer characteristics described in [Online Appendix A3](#) do not explain the triple-differences effect either.⁴⁰

40. We also tested alternative clusterings of standard errors. The clustering shown in all tables is on the chain, because whole chains are likely susceptible to serially correlated shocks. The chain also encompasses the smaller units according to which the applications we sent were grouped (the store, or sometimes a small group of stores). If one clusters on the geographic center instead, the *p*-values for our main specifications are easily below .05 for both analyses, .019 for the temporal difference-in-differences analysis, and .027 for the triple-differences analysis. If one clusters on the individual store, they are .033 and .056, respectively.

TABLE VII
EFFECTS OF BTB ON CALLBACKS OF GED HOLDERS VERSUS HS GRADS: TRIPLE
DIFFERENCES

	(1)	(2)	(3)
Box Remover \times post \times GED	-0.012 (0.025)	-0.010 (0.019)	-0.002 (0.019)
Post \times GED	0.021* (0.013)	0.009 (0.010)	-0.000 (0.009)
Box Remover \times post	0.006 (0.030)	0.014 (0.024)	
Box Remover \times GED	0.021 (0.022)	0.021 (0.015)	
Box Remover	-0.003 (0.029)	-0.012 (0.026)	
GED	-0.019* (0.011)	-0.012 (0.008)	0.401*** (0.129)
Post	0.004 (0.016)	0.005 (0.015)	0.473*** (0.174)
<i>N</i>	11,188	14,637	14,637
Controls	Yes	Yes	Yes
Center FE	No	Yes	Yes
Chain FE	No	No	Yes
Post \times chain FE	No	No	Yes
White \times chain FE	No	No	Yes
Sample	Balanced	Full	Full

Notes. This table recreates Table V, substituting GED for White. Standard errors in parentheses clustered on chain. Dependent variable is whether the application received a callback. The balanced sample consists of stores to which we sent exactly four applications, one white/black pair in each period. Fixed effects can include chain, post \times chain, white \times chain, or center, and are included as indicated. *10%, **5%, and ***1% significance level.

We also varied whether our applicants had GEDs or high school diplomas and whether they had a one-year employment gap or not. These characteristics are also correlated with criminal records in the real world, so one might expect the weight that employers place on them to also increase after BTB. In fact, as we show later, having a GED instead of a regular diploma is a much stronger predictor of whether one has a criminal record than race is (and moreover, in contrast to race, GED status is a lawful basis for employer decision making). Thus, one might expect that after BTB, rational employers who seek to avoid interviewing people with records would rely heavily on diploma type as a proxy. Accordingly, in Table VII we recreate the analysis in Table V substituting in the GED variable for *White*. The point estimates are negative, but small and insignificant; they are

virtually zero in the full-sample analyses (Table VII, column (3)). So, we cannot characterize this as even suggestive evidence that employers are using GED as a proxy for criminal records after BTB. In Online Appendix A7 we perform a similar analysis with the employment gap variable. Though the point estimates are non-trivial (around 2.6 percentage points; Table A7.2) and cut in the expected direction, these estimates are statistically insignificant.

V. DISCUSSION

V.A. *Who Gains and Loses from BTB?*

Our results produce mixed implications. On the one hand, they confirm BTB's basic premise: having a record poses an obstacle to employment. When employers had the box, applicants without records received 62% more callbacks than identical applicants with records did, even though those records entailed just one conviction for a minor, nonviolent felony, more than two years prior. This finding is consistent with prior auditing studies (Pager 2003; Pager, Bonikowski, and Western 2009), but it is useful to confirm it in a newer, much larger sample and a setting (online job applications) that is central to the modern job market. The practical effect of the criminal record penalty could be mitigated by the fact that most employers had no box even before BTB. But absent BTB, employers may ask about records at interviews and check records at any time, so employers' disfavoring of applicants with records may matter even without the box itself.

However, our findings also show that BTB increases racial discrimination. At "box remover" companies, the black-white gap increased sixfold: white applicants received 7% more callbacks than similar black applicants before BTB, and 43% more after BTB. Differencing out trends among nonbox employers only strengthens this conclusion, increasing the estimated growth in the gap slightly. Black applicants saw their callback rates fall by two percentage points after BTB, while white applicants saw theirs rise by two percentage points.

More specifically, as one would expect, BTB's main negative impact appears to fall on black applicants without records, while it is mainly white applicants with records whose callback rates go up. Figure II shows that after BTB, at box-remover employers in the balanced sample, callback rates for black applicants with records increased from 8.0% to 10.3%; for white applicants

with records they increased from 8.7% to 14.8%. Callback rates for black applicants without records decreased from 13.4% to 10.3%; for white applicants without records they actually increased from 14.1% to 14.8%.⁴¹ However, overall callback rates increased in our whole sample over this time period, making this picture look overly rosy across the board. If we subtract out the 1.6 percentage point increase that occurred at companies whose applications were unaffected by BTB, white applicants without records see a small decrease from BTB (−0.6 percentage point) and blacks with records see a small gain (+0.7 percentage point), but blacks without records see very large losses (−5.1 percentage points) and whites with records see very large gains (+4.5 percentage points).⁴²

V.B. Mechanisms: Statistical Discrimination versus Stereotyping

Our results imply that after BTB, employers use race to proxy for convictions, increasing racial discrimination. Are these employer assumptions empirically accurate, or are they relying on stereotypes about black criminality? Even accurate statistical discrimination is illegal, and conflicts with the policy objective of reducing employment disparities—but policy makers and scholars might still be interested in disentangling these mechanisms of discrimination. To make progress on this task requires outside data and some assumptions about the employer decision process.

It would not be surprising if employers made assumptions about black applicants' likely criminal records, even if those assumptions are not well founded. Lab experiments have consistently found that most Americans subconsciously associate blackness and criminality (see, e.g., [Eberhardt et al. 2004](#);

41. For the post-BTB period, these calculations apply the same averages to those with and without records because, in our experiment, these applicants became indistinguishable once employers no longer asked about records.

42. Because the overall rise in callbacks appears unrelated to BTB and should be filtered out, it is misleading to suggest, as [Emsellem and Avery \(2016\)](#) do (relying on our full-sample results), that our study shows that BTB increased overall black callback rates. Note also that the [Figure II](#) numbers relied on here slightly understate the increase in the race gap that we found in the triple-differences regression analysis. In addition, because the primary negative effect of BTB is on black applicants without convictions, and in the real world most black men do not have felony convictions, the real-world effects on black applicants may well be worse than in our study, where half of our black applicants stood to benefit. See [Section V.B](#) for a back-of-the-envelope calculation.

Nosek et al. 2007). Bordalo et al. (2016) offer a general theoretical model for how generalizations based on a “kernel of truth” (such as somewhat higher black conviction rates) may become greatly exaggerated in the eyes of decision makers. They posit that decision makers save cognitive resources by relying on heuristics—representative types—that they then use to predict characteristics of interest. In our context, their theory implies that when hiring managers formulate expectations about black candidates, they overweight traits disproportionately represented in black populations, that is, felony convictions—even if the actual difference is small.

In assessing the accuracy of employer priors, a threshold challenge is to determine the real-world distribution of felony convictions in our relevant population. There are no comprehensive national data on the number of people with felony convictions overall or by subgroups (Bucknor and Barber 2016), but none of the data sources that are available vindicate the employer behavior we observe. For example, the National Longitudinal Survey of Youth 1997 (NLSY97) contains self-reported information about convictions at particular ages, and about other characteristics like race, education, and work history. Based on our calculations, these data show that by our applicants’ age (22), 15.7% of non-Hispanic white males and 18.5% of black males have received adult criminal convictions. This gap, quite small to begin with (and much smaller than analogous race gaps in arrest and incarceration rates), disappears entirely once differences in educational attainment are accounted for. Because education is observable to employers and randomized in our study, if the NLSY97 data are representative, well-informed employers should not have assumed that our black applicants had higher conviction rates than our white applicants at all.

However, some researchers have critiqued the accuracy of self-report studies like the NLYS97, including as a tool for estimating race gaps (see Piquero and Brame 2008, for a review of criticisms). Moreover, the NLYS97 does not distinguish felonies from misdemeanors. Other data sources have drawn on correctional data to estimate felony conviction rates by race. Shannon et al. (2011) estimate that among adults of all ages and genders, 6% of non-African-Americans and 25% of African Americans have felony convictions. Bucknor and Barber (2016) estimate that among men of all ages, approximately 10% of white men and 42% of black men have felony convictions. Although these gaps

are much larger than those observed in the NLSY97, they are not broken down by age or other characteristics, and it is likely they would not be nearly as stark if limited to the subset paralleling our applicants: young men with only high-school-level education.⁴³

Keeping in mind this wide range of estimates, we turn to the question of what our data can tell us about employer priors about black and white criminality. To do so, we employ a very simple model of employer decision making, outlined in detail in [Online Appendix A8](#), in which the probability of an interview is linear in the perceived probability that an applicant has a criminal record:

$$(3) \quad \begin{aligned} P(\text{call}_{ij} | \text{black}_i, \text{no box}_j) &= g_b P(\text{call}_{ij} | \text{black}_i, \text{crime}_i) \\ &+ (1 - g_b) P(\text{call}_{ij} | \text{black}_i, \text{no crime}_i). \end{aligned}$$

That is, when there is no box, the probability that a black applicant i gets a callback from store j is a weighted average of the probability that the store would call back a black applicant with a criminal record and a black applicant without one (if that information were known). g_b is then the object of interest—the employer's estimate of the probability that a black applicant has a criminal record. An analogous equation could be written for white applicants to estimate g_w .

We use the callback rates reported in [Figure II](#) for the pre- and post-BTB periods among box-remover employers to back out g_b and g_w . Because the post-BTB period had a higher callback rate across the board, we first subtract out the 1.6 percentage points secular increase that we observed at employers whose

43. Because convictions can be accrued throughout life, the estimates for both whites and blacks (and thus, presumably, the percentage point gap between the two) would be much lower if the comparison focused on 22-year-olds. [Reaves \(2013\)](#) finds that the average age of a felony defendant is 32. Also, neither study estimates race breakdowns conditional on educational status or other socioeconomic characteristics, but [Bucknor and Barber \(2016\)](#) confirm that education and felony conviction rates are very strongly correlated: they estimate that about 64% of men without any high-school-level diploma have felony convictions (nearly 4 times the rate of male high-school graduates and 17 times the rate of men with any college). Because black men are highly overrepresented among dropouts and attend college at lower rates (see, e.g., [Schott Foundation 2015](#)), presumably the race gap in felony convictions must be substantially smaller if assessed within educational categories.

applications were unaffected by BTB. After this adjustment, inserting the observed callback rates into [equation \(3\)](#) implies that (on average) employers' priors for the probability that black and white applicants had felony convictions are 87% and 16%, respectively (a 71 percentage point difference). These priors appear sharply exaggerated, even relative to the largest estimates in the empirical literature. Even the black-white gap among males of all ages estimated by [Bucknor and Barber \(2016\)](#) is only about 32 percentage points. At the other extreme, the NLSY97 data suggest that rational employers should assume a 3 percentage point gap even if they ignore socioeconomic observables, or no gap if they are informed of the rates within educational categories. In [Online Appendix A8](#), we explain what assumptions would justify [equation \(3\)](#). These are necessarily strong and simplified, although we also show that alternate sets of assumptions can lead us to quite similar estimated employer priors.

Moreover, even without relying on the estimates from this modeling exercise, more basic reasons suggest employers are getting it wrong. First, educational status (specifically, GED versus high school diploma) is a much stronger predictor of criminal records than race is. For example, among white and black men in the NLSY97, the chance of a conviction by age 22 was 35.8% for GED-holders and 12.1% for HSD-holders (and there is no race gap conditional on education). However, [Table VII](#) shows that employers do not change their reaction to a GED after BTB goes into effect—that is, they do not seem to use the GED as a proxy for records. No plausible model of well-informed employer decision making could explain why BTB greatly increases the effect of race and not that of a GED, suggesting that employers are instead relying on stereotypes or implicit racial bias. Second, consider the mere fact that (after adjusting for the secular rise in callbacks) the post-BTB black callback rate substantially declined compared to the rates when half of our black applicants had observable felony convictions. This suggests that post-BTB employers are assuming that considerably more than half of our black applicants have felony convictions—but by any plausible real-world measure, the appropriate assumption would be less than half, probably far less for young men with diplomas and several years of work experience.⁴⁴

44. It is also possible that employers are assuming that black applicants have particularly serious criminal records qualitatively. While the conviction rate

In short, the pattern observed here is most consistent with a stereotyping model (such as that in [Bordalo et al. 2016](#)), in which small real-world differences are greatly exaggerated. Our data provide no means of testing how employers come to their seemingly incorrect assumptions, so we cannot offer a direct empirical test of [Bordalo et al.'s \(2016\)](#) theory of how stereotypes are formed. Alternate theories are possible, and multiple mechanisms could simultaneously contribute.⁴⁵ In any event, our data do support some form of stereotyping or bias explanation, rather than the interpretation that employers are engaging in empirically informed statistical discrimination.

V.C. Identification Challenges and Limitations

Our research design provides a strong basis for interpreting our estimates as causal. Because our black and white applicants to all employers in both periods have the same characteristics, and because our results hold when changes at businesses unaffected by BTB are filtered out, any remaining identification threats would have to come from unobserved differences that (i) affect box-remover versus other businesses differently, (ii) in ways that differ by race (among otherwise identical applicants), and (iii) differ across time periods as well. Such a difference is of course possible, but there is no obvious candidate for what it might be. This is especially so because the gap between the pre- and post-BTB periods is short, because the groups of businesses are qualitatively similar, and because we see approximately the same effect in New Jersey and New York City although their pre- and postperiods were seasonally nearly opposite.

data does not directly get at this possibility, it seems irrational to apply this assumption to a large share of applicants given that felony convictions of any sort are relatively infrequent by age 22. Nor do employers need to err on the side of caution at the callback stage to avoid any chance of a serious criminal record; BTB does not prevent them from eventually declining to hire after background checks.

45. For example, perhaps stereotypes are grounded in long-standing cultural biases with no empirical foundation. Or perhaps employers confuse the distribution of criminal convictions (the relevant distribution for our purposes, since convictions are what job applications ask about) with larger gaps in other outcomes like police stops. Or perhaps employers ignore the fact that race gaps are likely smaller after conditioning on other observable characteristics.

A potential concern is that BTB might encourage real-world applicants with records to apply to box-remover companies, affecting the competition our fictional applicants face. But such a change should affect all our applicants; there is no reason to expect it to cause employers to treat black applicants more adversely than identical whites, and even if there were, that would merely provide another mechanism by which BTB increases racial discrimination. Moreover, we think BTB probably did not substantially affect applicant pools, especially within the short period covered by our study. Many applicants likely do not know which employers have the box before they actually see it (usually on one of the final screens of the application): we ourselves could find no resources listing employers with and without the box. Applicants would also have to know about BTB and its effective date and be so discouraged by the box that they avoid applying, yet not discouraged by the fact that even post-BTB, employers conduct background checks.⁴⁶

Our study has important limitations. Our applicants were only black and white men; dividing the sample into additional groups would have created serious statistical power concerns. We also mainly focused on chain employers in the retail and restaurant industries. These are important sectors for employment of people with records, but whether our results apply to other sectors or to smaller employers remains an open question for future research.

Perhaps the most significant limitation is that we were unable to study effects of BTB on ultimate hiring patterns, only callbacks, so we do not know whether firms avoid hiring applicants with records even after they “get their foot in the door.”

46. A variant of this concern is that BTB might discourage people with records from applying to employers that never had the box. But this theory is even less likely to explain our results, which are driven almost entirely by changes among box-remover employers, not other employers (see [Tables V and VI](#)). Changes to nonbox employers' applicant pools would likely be even more subtle than changes to box employers' pools, as their applications do not change, and for most applicants there is likely no trade-off between applying to both business types. And given that these employers lack the box, many would likely not notice subtle changes in the percentage of their applicants with records. Another concern is that BTB could encourage even employers that never had the box to racially discriminate in callbacks because they know that (per BTB) they won't be able to screen out candidates with records at the interview. We do not observe such a change, however—and if anything, this would downward bias our triple-differences estimate.

Still, BTB is meant precisely to impact the initial stage of the hiring process (the stage at which most job applicants are filtered out), and our study speaks to those impacts. Moreover, a potential worst-of-both-worlds scenario is that BTB could have the negative consequence of excluding black applicants from callbacks, even if it does *not* have its intended positive consequence of increasing hiring of people with records. Note that, in line with our results, [Doleac and Hansen \(2016\)](#), using the Current Population Survey, find reduced employment of young black men in jurisdictions that adopt BTB, and (for private employers) an increase in employment of young white men.

VI. CONCLUSION

BTB policies are well intentioned, and they do appear to help some applicants with records to gain access to job interviews. However, our field experiment shows that this gain comes at the expense of another group that faces serious employment challenges: black men. Our findings support the hypothesis that employers substantially increase discrimination on the basis of race after BTB goes into effect.

One caveat is that our estimates (like those of auditing studies generally; see [Heckman and Siegelman 1993](#); [Heckman 1998](#)) do not directly speak to changes in actual markets. Real-world applicants are not divided 50/50 between identical black male and white male candidates (and no other groups). And if BTB helps black men with records while hurting black men without records, the net effect on black male employment would depend on the real-world sizes of these groups.

That said, back-of-the-envelope calculations point to an enlargement of the black-white employment gap. To render these calculations conservative, we apply the conviction-rate figures from [Bucknor and Barber \(2016\)](#), with the largest race gap we have found in the literature. Suppose all black and white men were subject to changes paralleling the pattern in [Figure II](#), adjusting for the 1.6 percentage point secular rise in callbacks observed at control-group companies, as described in [Section V.A](#) above. Applying these changes to the real-world distribution of records from [Bucknor and Barber \(2016\)](#) implies that black callback rates would fall by 2.7 points, while white callback rates would rise by 0.1 points—a net rise of 2.8 percentage points in the black-white gap. To put this in perspective, this

is one-quarter of our overall callback rate, and would be enough to quintuple the underlying pre-BTB black-white gap observed in our sample.⁴⁷

Policy makers might also consider whether other interventions could offset BTB's adverse effects on black candidates. If laws against racial discrimination in hiring were effectively enforced, BTB would not have this unintended consequence.⁴⁸ This, to be sure, is easier said than done, as hiring discrimination laws are notoriously difficult to enforce. The intuition behind BTB perhaps suggests one potential innovation: employers could blind themselves to names (and other potentially racially identifying information unrelated to job qualifications).

Alternatively, policy makers could seek alternate strategies to increase hiring of people with records. One possibility is expanding successful job training and placement programs in correctional or prisoner reentry settings, though these programs have had mixed results (Visher, Winterfield, and Coggeshall 2005; Seiter and Kadela 2009; Jacobs 2012).⁴⁹ In addition, tax credits can incentivize hiring of disadvantaged groups (Katz 1998). Although a federal Work Opportunity Tax Credit (WOTC) already extends to persons hired within one year of conviction or release from prison, take-up has been limited, perhaps implying that the credit should be increased, better publicized, and/or made administratively less burdensome (Holzer, Raphael, and Stoll 2007). Other approaches

47. A similar exercise with the NLSY97 estimates leads to an estimated increase in the race gap of 4.2 percentage points. One complicating factor is that only about half of real-world applicants have racially distinctive names, perhaps reducing the relative impact of the racial-discrimination effect. However, this point may be offset by the fact that real-world applicants (unlike our fictional ones) often have other racial signals on their job applications, such as their neighborhood or high school. Moreover, even if we cut the expected losses to black and white applicants without records in half, the exercise above using Bucknor and Barber's numbers would still lead to a growth in the black-white gap of 1.4 percentage points. Of course, a full analysis would also have to consider the fact that white and black men have different distributions of other characteristics as well, and that they are not the only two groups competing for jobs.

48. Thus, we do not disagree with Emsellem and Avery (2016) that the "root of the problem" is employers' reliance on race-based assumptions about criminality; however, unless some other strategy for changing that employer behavior can be found, BTB is likely to have the unintended consequences we identify.

49. Some job training programs for disadvantaged persons more generally have been more successful (see, e.g., Roder and Elliot 2014, discussing the Year Up program) and perhaps could be adapted to target people with records.

could seek to mitigate employers' fear of negligent-hiring liability, which surveys show is a common reason that employers investigate applicants' records (Society for Human Resource Management 2012).⁵⁰ For example, some jurisdictions offer certificates of "employability" (or rehabilitation); these may increase employer confidence that applicants are not a substantial risk, and if litigation does arise, employers can point to the certificates to show that they took adequate care in hiring. A recent audit study found that applicants with convictions and employability certificates were called back at nearly the same rate as applicants with clean records (Leasure and Anderson 2016).

Racial disparity is not the only policy consideration surrounding BTB, and policy makers could seek to prioritize opportunities for people with records in spite of BTB's unintended racial consequences or to mitigate those consequences in other ways. But to the extent that advocates hope that BTB itself will reduce racial disparity in employment, that hope appears misguided.

RUTGERS UNIVERSITY
UNIVERSITY OF MICHIGAN

SUPPLEMENTARY MATERIAL

An Online Appendix for this article can be found at *The Quarterly Journal of Economics* online. Data and code replicating the tables and figures in this article can be found in Agan and Starr (2017b), in the Harvard Dataverse, doi:10.7910/DVN/VPHMNT.

REFERENCES

- Agan, Amanda Y., and Sonja B. Starr, "Ban the Box, Criminal Records, and Statistical Discrimination: A Field Experiment," University of Michigan Law and Econ Research Paper No. 16-012 (2016), <https://ssrn.com/abstract=2795795>.
—, "The Effect of Criminal Records on Access to Employment," *American Economic Review*, 107 (2017a), 560–64.

50. Negligent hiring (i.e., hiring an unfit employee) is a legal theory by which employers can be held responsible for torts committed by their employees, and such suits can be costly, with settlements averaging \$1.6 million (Connerly, Avery, and Bernardy 2001). Human resources guides consistently advise employers that liability is more likely if they fail to check applicants' backgrounds and to reject applicants with criminal records that suggest a risk (Connerly, Avery, and Bernardy 2001; Levashina and Campion 2009).

- , “Replication Data for: ‘Ban the Box, Criminal Records and Statistical Discrimination: A Field Experiment,’” Harvard Dataverse (2017b), doi:10.7910/DVN/VPHMNT.
- Aigner, Dennis J., and Glen G. Cain, “Statistical Theories of Discrimination in Labor Markets,” *Industrial and Labor Relationship Review*, 30 (1977), 175–187.
- Allen, Kathy Grannis, “NRF Survey Finds Nearly All Retailers Rely on Background Checks to Keep Consumers, Companies Safe,” *National Retail Federation*, October 4, 2011, <https://nrf.com/media/press-releases/nrf-survey-finds-nearly-all-retailers-rely-background-checks-keep-consumers>.
- Arrow, Kenneth Joseph, “The Theory of Discrimination,” in *Discrimination in Labor Markets*, Orley Ashenfelter and Albert Rees, eds. (Princeton, NJ: Princeton University Press, 1973).
- Astone, Nan, Michael Katz, and Juila Gelatt, “Innovations in NYC Health & Human Services Policy: Young Men’s Initiative,” Urban Institute, February 2014.
- Baert, Stijn, and Elsy Verhofstadt, “Labour Market Discrimination Against Former Juvenile Delinquents: Evidence from a Field Experiment,” *Applied Economics*, 47 (2015), 1061–1072.
- Bartik, Alexander Wickman, and Scott Nelson, “Credit Reports as Resumes: The Incidence of Pre-Employment Credit Screening,” MIT Department of Economics Graduate Student Research Paper 16-01 (2016).
- Bertrand, Marianne, and Sendhil Mullainathan, “Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination,” *American Economic Review*, 94 (2004), 991–1013.
- Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer, “Stereotypes,” *Quarterly Journal of Economics*, 131 (2016), 1753–1794.
- Brame, Robert, Shawn D. Bushway, Ray Paternoster, and Michael G. Turner, “Demographic Patterns of Cumulative Arrest Prevalence by Ages 18 and 23,” *Crime and Delinquency*, 60 (2014), 471–486.
- Bucknor, Cherrie, and Alan Barber, “The Price We Pay: Economic Costs of Barriers to Employment for Former Prisoners and People Convicted of Felonies,” CEPR Working Paper No. 2016-07 2016.
- Bushway, Shawn D., “Labor Market Effects of Permitting Employer Access to Criminal History Records,” *Journal of Contemporary Criminal Justice*, 20 (2004), 276–291.
- Center for Health Statistics. Birth/EBC Confidential Data Files, New Jersey Department of Health, Trenton, NJ.
- Clarke, H., “Protecting the Rights of Convicted Criminals: Ban the Box Act of 2012,” *Washington Post*, December 20, 2012.
- Clifford, Robert, and Daniel Shoag, “No More Credit Score’: Employer Credit Check Bans and Signal Substitution,” Unpublished manuscript, 2016.
- Color of Change, “Civil Rights Group Responds to the ‘Ban the Box’ Executive Order,” last modified November 2, 2015, <http://colorofchange.org/press/releases/2015/11/2/civil-rights-group-responds-ban-box-executive-order/>.
- Community Catalyst, “Banning the Box in Minnesota—and across the United States,” last modified December 2, 2013, <http://www.communitycatalyst.org/blog/banning-the-box-in-minnesota-and-across-the-united-states#.YuG1.Yo46U>.
- Connerly, Mary L., Richard D. Avery, and Charles J. Bernardy, “Criminal Background Checks for Prospective and Current Employees: Current Practices among Municipal Agencies,” *Public Personnel Management*, 30 (2001), 173–183.
- Decker, Scott H., Natalie Ortiz, Cassia Spohn, and Eric Hedberg, “Criminal Stigma, Race, and Ethnicity: The Consequences of Imprisonment for Employment,” *Journal of Criminal Justice*, 43 (2015), 108–121.

- Del Castillo, Christine, "Does Anyone Advertise Jobs in Newspapers Anymore," Workable (blog), May 19, 2016, <https://resources.workable.com/blog/newspaper-job-ads>.
- Deming, David J., Noam Yuchtman, Amira Abulafi, Claudia Goldin, and Lawrence F. Katz, "The Value of Postsecondary Credentials in the Labor Market: An Experimental Study," *American Economic Review*, 106 (2016), 778–806.
- Doleac, Jennifer L., and Benjamin Hansen, "Does 'Ban the Box' Help or Hurt Low-Skilled Workers? Statistical Discrimination and Employment Outcomes When Criminal Histories Are Hidden," NBER Working Paper 22469, 2016.
- Eberhardt, Jennifer L., Phillip Atiba Goff, Valerie J. Purdie, and Paul G. Davies, "Seeing Black: Race, Crime, and Visual Processing," *Journal of Personality and Social Psychology*, 87 (2004), 876.
- Emsellem, Maurice, and Beth Avery, "Racial Profiling in Hiring: A Critique of New 'Ban the Box' Studies," National Employment Law Project, 2016.
- Fang, Hanming, and Andrea Moro, "Theories of Statistical Discrimination and Affirmative Action: A Survey," in *Handbooks of Social Economics Vol 1A*, Jess Benhabib, Matthew O. Jackson, and Alberto Bisin, eds. (Amsterdam: North-Holland, 2011), 133–200.
- Farber, Henry S., Dan Silverman, and Till Von Wachter, "Factors Determining Callbacks to Job Applications by the Unemployed: An Audit Study," NBER Working Paper 21689, 2015.
- Finlay, Keith, "Stigma in the Labor Market," Unpublished manuscript, 2014.
- Fryer, Roland G., and Steven D. Levitt, "The Causes and Consequences of Distinctively Black Names," *Quarterly Journal of Economics*, 119 (2004), 767–805.
- Heckman, James J., "Detecting Discrimination," *Journal of Economic Perspectives*, 12 (1998), 101–116.
- Heckman, James J., and Paul A. LaFontaine, "The American High School Graduation Rate: Trends and Levels," *Review of Economics and Statistics*, 92 (2010), 244–262.
- Heckman, J., and P. Siegelman, "The Urban Institute Audit Studies: Their Methods and Findings," in *Clear and Convincing Evidence: Measurement of Discrimination in America*, Michael Fix and Raymond Struyk, eds. (Washington, DC: Urban Institute Press, 1993).
- Holzer, Harry J., Steven Raphael, and Michael A. Stoll, "Perceived Criminality, Criminal Background Checks, and the Racial Hiring Practices of Employers," *Journal of Law and Economics*, 49 (2006), 451–480.
- , "The Effect of an Applicant's Criminal History on Employer Hiring Decisions and Screening Practices: Evidence from Los Angeles," in *Barriers to Reentry? The Labor Market for Released Prisons in Post-Industrial America*, Shawn Bushway, Michael Stoll, and David Weiman, eds. (New York: Russel Sage Foundation, 2007).
- Jacobs, Erin, "Returning to Work after Prison—Final Results from the Transitional Jobs Reentry Demonstration," May 10, 2012, <http://dx.doi.org/10.2139/ssrn.2056045>.
- Katz, Lawrence F. "Wage Subsidies for the Disadvantaged," in *Generating Jobs*, Richard B. Freeman and Peter Gottschalk, eds. (New York: Russel Sage Foundation, 1998).
- Kroft, Kory, Fabian Lange, and Matthew J. Notowidigdo, "Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment," *Quarterly Journal of Economics*, 128 (2013), 1123–1167.
- Lahey, Joanna N., "Age, Women, and Hiring an Experimental Study," *Journal of Human Resources*, 43 (2008), 30–56.
- Lahey, Joanna N., and Ryan A. Beasley, "Computerizing Audit Studies," *Journal of Economic Behavior and Organization*, 70 (2009), 508–514.
- Leasure, Peter, and Tia Stevens Andersen, "The Effectiveness of Certificates of Relief as Collateral Consequence Relief Mechanisms: An Experimental Study," *Yale Law & Policy Review*, *Inter Alia*, 35 (2016).
- Levashina, Julia, and Michael A. Campion, "Expected Practices in Background Checking: Review of the Human Resources Management

- Literature," *Employee Responsibilities and Rights Journal*, 21 (2009), 231–249.
- List, John A., "The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field," *Quarterly Journal of Economics*, 119 (2004), 49–89.
- Love, Margaret Colgate, "LAWYERING: Paying Their Debt to Society: Forgiveness, Redemption, and the Uniform Collateral Consequences of Conviction Act," *Howard Law Journal*, 54 (2011), 753–795.
- Minnesota Department of Human Rights, "Ban The Box: Overview for Private Employers," accessed January 19, 2016, <https://mn.gov/mdhr/employers/criminal-background/>.
- NAACP, "Ban the Box," accessed May 3, 2017, <http://www.naacp.org/campaigns/ban-the-box/>.
- Neumark, David, "Detecting Discrimination in Audit and Correspondence Studies," *Journal of Human Resources*, 47 (2012), 1128–1157.
- Neumark, David, Roy J. Bank, and Kyle D. Van Nort, "Sex Discrimination in Restaurant Hiring: An Audit Study," *Quarterly Journal of Economics*, 111 (1996), 915–941.
- Neumark, David, Ian Burn, and Patrick Button, "Is It Harder for Older Workers to Find Jobs? New and Improved Evidence from a Field Experiment," NBER Working Paper 21669, 2015.
- Nosek, Brian A., Frederick L. Smyth, Jeffrey J. Hansen, Thierry Devos, Nicole M. Lindner, Kate A. Ranganath, Colin Tucker Smith, Kristina R. Olson, Dolly Chugh, Anthony G. Greenwald, and Mahzarin R. Banaji, "Pervasiveness and Correlates of Implicit Attitudes and Stereotypes," *European Review of Social Psychology*, 18 (2007), 36–88.
- Oreopoulos, Philip, "Why Do Skilled Immigrants Struggle in the Labor Market? A Field Experiment with Thirteen Thousand Resumes," *American Economic Journal: Economic Policy*, 3 (2011), 148–171.
- Pager, Devah, "The Mark of a Criminal Record," *American Journal of Sociology*, 108 (2003), 937–975.
- Pager, Devah, Bart Bonikowski, and Bruce Western, "Discrimination in a Low-Wage Labor Market: A Field Experiment," *American Sociological Review*, 74 (2009), 777–799.
- Phelps, Edmund S., "The Statistical Theory of Racism and Sexism," *American Economic Review*, 62 (1972), 659–661.
- Pinard, M., "Ban the Box in Baltimore," *Baltimore Sun*, January 7, 2014.
- Piquero, Alex R., and Robert W. Brame, "Assessing the Race–Crime and Ethnicity–Crime Relationship in a Sample of Serious Adolescent Delinquents," *Crime and Delinquency*, 54 (2008), 390–422.
- Reaves, Brian A., "Felony Defendants in Large Urban Counties, 2009-Statistical Tables," US Department of Justice, 2013.
- Riach, Peter A., and Judith Rich, "Field Experiments of Discrimination in the Market Place," *Economic Journal*, 112 (2002), F480–F513.
- Roder, Anne, and Mark Elliott, "Sustained Gains: Year Up's Continued Impact on Young Adults' Earnings," Economic Mobility Corporation Report (May 2014), <https://economicmobilitycorp.org/uploads/sustained-gains-economic-mobility-corp.pdf>.
- Rodriguez, Michelle N., and Beth Avery, "Ban the Box: US Cities, Counties, and States Adopt Fair Hiring Policies," National Employment Law Project (2016), accessed June 9, 2016, <http://www.nelp.org/publication/ban-the-box-fair-chance-hiring-state-and-local-guide/>.
- Scarborough, David, "Does Job Testing Harm Minority Workers? Evidence from Retail Establishments," *Quarterly Journal of Economics*, 123 (2008), 219–277.
- Schott Foundation, "Black Lives Matter: Schott 50-State Report on Public Education and Black Males," 2015, <http://blackboysreport.org/national-summary/>.
- Seiter, Richard P., and Karen R. Kadela, "Prisoner Reentry: What Works, What Does Not, and What Is Promising," *Crime and Delinquency*, 49 (2009), 360–388.

- Shannon, Sarah, Christopher Uggen, Melissa Thompson, Jason Schnittker, and Michael Massoglia, "Growth in the U.S. Ex-Felon and Ex-Prisoner Population, 1948–2010," Unpublished manuscript, 2011.
- Shoag, Daniel, and Stan Veuger, "No Woman No Crime: Ban the Box, Employment, and Upskilling," Unpublished manuscript, 2016.
- Society for Human Resource Management, "Background Checking—The Use of Criminal Background Checks in Hiring Decisions," last modified July 19, 2012, <https://www.shrm.org/hr-today/trends-and-forecasting/research-and-surveys/pages/criminalbackgroundcheck.aspx>.
- Southern Coalition for Social Justice, "Ban the Box Community Initiative Guide," accessed May 3, 2017, <http://www.southerncoalition.org/program-areas/criminal-justice/ban-the-box-community-initiative-guide/>.
- Stoll, Michael A., "Ex-Offenders, Criminal Background Checks, and Racial Consequences in the Labor Market," *University of Chicago Legal Forum* (2009), Article 11.
- Uggen, Christopher, Mike Vuolo, Sarah Lageson, Ebony Ruhland, and Hilary K. Whitham, "The Edge of Stigma: An Experimental Audit of the Effects of Low-level Criminal Records on Employment," *Criminology*, 52 (2014), 627–654.
- U.S. Bureau of Labor Statistics, "Labor Force Characteristics by Race and Ethnicity, 2015," U.S. Department of Labor, September 2016, <https://www.bls.gov/opub/reports/race-and-ethnicity/2015/home.htm>.
- U.S. Equal Employment Opportunity Commission (EEOC), "Consideration of Arrest and Conviction Records in Employment Decisions under Title VII of the Civil Rights Act of 1964," EEOC Enforcement Guidance, April 25, 2012, https://www.eeoc.gov/laws/guidance/arrest_conviction.cfm.
- Visher, Christy A., Laura Winterfield, and Mark B. Coggeshall, "Ex-Offender Employment Programs and Recidivism: A Meta-Analysis," *Journal of Experimental Criminology*, 1 (2005), 295–316.
- Wozniak, Abigail, "Discrimination and the Effects of Drug Testing on Black Employment," *Review of Economics and Statistics*, 97 (2015), 548–566.