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

Amanda Agan and Sonja Starr[1]

February 26, 2017

ABSTRACT

"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, Larry Katz, Alan Krueger, Steven Levitt, Alex Mas, Ezra Oberfield, Emily Owens, Andrei Shleifer, 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, the Society of Labor Economists Annual Meeting, the IRP Summer Workshop, the American Law and Economics Association Annual Meeting, and the NBER Summer Institute Labor Studies/Crime Session, 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.

[1] Agan: Rutgers University, aagan@econ.rutgers.edu; Starr: University of Michigan, sbstarr@umich.edu.

1. 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 to 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 2017). The "box" referred to in "Ban the Box" (and hereinafter in this paper) 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, and especially for improving hiring of black men (Pinard 2014, Southern Coalition for Social Justice 2013, Clarke 2012, and Community Catalyst 2013), who have recently faced unemployment rates approximately double the national average (Bureau of Labor Statistics 2015).[2] 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 varied whether our applicants had a felony conviction.[3] 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 et al. 2006; Holzer 2007). BTB laws only delay

[2] See, for example, Minnesota Department of Human Rights (2015): "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 (City of New York 2016). Civil rights organizations are also major BTB backers (NAACP 2014, Color of Change 2015).

[3] 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.

(rather than permanently bar) employer access to criminal records; employers may still conduct criminal background checks near the end of the hiring process.[4] 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 to mitigate racial disparities in employment, because black men are more likely to have records (Brame et al. 2014, Shannon et al. 2011).[5]

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 suggested that decision-makers deprived of individualized information might rationally rely on group generalizations instead (Phelps 1972, Arrow 1973, Stoll 2009, Fang and Moro 2011). Observational studies have investigated the application of these theories to employer background checks and to expansion of Internet records databases, producing somewhat mixed results.[6] Other researchers have found evidence of other forms of statistical discrimination in employment.[7]

Alternatively, employers might rely not on accurate information, but on exaggerated assumptions or stereotypes about group differences (for example, Bordalo et al. (2016) provide a theory of the decision process producing stereotyping). While we consider the differences between

[4] In New Jersey and New York City, the two jurisdictions on which this study focuses, employers may conduct background checks anytime after 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 indicated that 97% of them performed some type of background check (National Retail Federation 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).

[5] 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 non-black population. This disparate impact is why EEOC interprets race discrimination law to constrain employers' treatment of criminal records (EEOC 2012).

[6] 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 et al. (2006) and Stoll (2009) find that surveyed employers who report that they use criminal records checks are more likely to hire African-Americans.

[7] Employers appear less willing to hire racial minorities in the absence of drug testing and credit checks (Wozniak 2015, Clifford and Shoag 2016, and Bartik and Nelson 2016). 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.

rational statistical discrimination and stereotyping in the Discussion, it bears emphasis that either would amount to unlawful racial discrimination. Title VII of the Civil Rights Act of 1964, which prohibits race and sex discrimination in employment, does not permit otherwise-illegal treatment be based on group generalizations, even if they are empirically supported.[8] But restrictions on hiring discrimination are famously difficult to enforce, so the fact that racial discrimination would be an unlawful response to BTB does not mean it is unlikely.

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. Finally, we also tested 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.

We report several key findings. The first supports BTB's core premise: when employers ask about them, felony convictions are 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%). 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, do not significantly affect callback rates, and this does 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

[8] 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.

applicants *without* criminal records see a substantial drop in callback rates after BTB, which their white counterparts do not see. Meanwhile, white applicants *with* criminal records see a substantial increase in callbacks, which their black counterparts do 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. As we explain in the Discussion, these generalizations appear exaggerated relative to real-world racial differences in conviction rates. 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.

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 working papers 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.[9] 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 varying characteristics of interest about a job candidate while holding other characteristics constant. It has been used to test

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

employment discrimination based on race, gender, length of unemployment spell, age, commute time, and type of postsecondary education (Neumark 1996; Riach and Rich 2002; Bertrand and Mullainathan 2004; Lahey 2008; Oreopoulos 2011; Neumark 2012; Kroft et al. 2013; Deming et al. 2016; Farber et al. 2015; Neumark et al. 2015; Phillips 2016), as well as the effect of criminal records (Pager 2003; Pager et al. 2009; Uggen et al. 2014; Baert and Verhofstadt 2014; Decker et al. 2014). 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.

2. 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.[10]

2.1 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, online U.S. job site and snagajob.com, the largest site focused on hourly employment.[11] Second, we also directly

[10] In NJ, pre-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.

[11] 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 multi-site aggregators like Indeed (Del Castillo 2016).

searched the employment websites of chain businesses meeting certain criteria.[12] We searched for jobs requiring little work experience, no post-secondary 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 resumes (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 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.[13] As a result, the sample composition by chain, jurisdiction, and specific store is not identical across periods. We address this concern below.

2.2 Applicant Profiles

Our fictitious applicants are all male and approximately 21 to 22 years old.[14] The RAs filled out applications based on profiles that we created using the Resume 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

---

[12] In New Jersey, we applied to businesses 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 could be 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.

[13] 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.

[14] 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.

numbers, information on high school or GED programs, felony conviction status and information about the criminal charge, formatted resume, and answers to many other routine application questions, such as job availability and pay sought (minimum wage).[15]

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:

(1) Has felony criminal conviction or not

a. (Conditional on conviction): convicted of property crime or drug crime

(2) Has 1-year employment gap versus a 0- to 2-month gap (referred to as "no gap" below) between the two past jobs

(3) GED or High School Diploma

Race is indicated via applicant names, as discussed in Section 2.3. The felonies we gave our applicants were nonviolent and fairly minor—either property crimes (e.g., shoplifting, 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.[16]

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.[17] 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 name were designed to have similar connotations, but randomly varied among similar options so as to disguise applicants' characteristics. Most applicant profiles (59%) were sent to only one business, but we sometimes

[15] 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.

[16] 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 1-3 months (Reaves 2013). Incarceration rates are presumably lower yet among first-time offenders with minor felonies, like our fictional applicants.

[17] 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.

used the same pairs to apply to multiple nearby locations of the same chain, as real-world applicants might do.[18] For more details on applicant profiles and application procedures, see Appendix A1.

2.3. 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; Oreopoulous 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 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, 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 below the white SES median, as

---

[18] Our criteria for 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).

[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).

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.

3. 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 Appendices A5 and A6, we replicate several of the tables and figures for New Jersey and New York separately.

3.1 Summary Statistics

Initial summary statistics are presented in Table 1a, 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 (noncompliers). 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

---

[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 footnote 14), 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, non-box 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.

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 non-box employers have no box on their application in any jurisdiction, indicating 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% vs. 10.9%), and lower in New York City than in New Jersey (9.4% vs. 14.7%; see 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 pre-period to 2.8 percentage points in the post-period; 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 1a does not break down callback rates by criminal record status because criminal record is unobserved by most employers (those without the box). Table 1b 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). 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%), though regressions not shown here show that the race-crime interaction is not statistically significant.

3.2. Regression Estimates of Main Effects of Applicant Characteristics

Table 2 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

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

[24] All the results shown in Table 2 are for both periods combined (unlike Table 1b and Figure 1, which were for the pre-period only), but the regression results look similar if only the pre-period observations are used.

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 ($p<0.01$), 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 & Starr (2017), 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 paper, which assesses the effect of the box on racial discrimination.

In Appendix A4 we show that the main effects of race and crime are robust to several alternative specification 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.

4. The Criminal-Record Box and Racial Discrimination

In this Section, we show differences-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 4.1, 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 4.2, we assess the temporal change after

[25] These robustness checks parallel those discussed below concerning the analyses of BTB's effects (Table 5).

[26] In a separate paper, 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.

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 4.3, we employ a triple-differences analysis that combines both these sources of variation.

4.1 Cross-Sectional Differences-in-Differences Estimates

In Table 3, Column 1, we compare race gaps between employers that do and do not have the criminal record "box." We employ a simple differences-in-differences specification for the probability that applicant i to store j receives a callback:

$$Callback_{ij} = \alpha + \beta_1 Box_j + \beta_2 White_i + \beta_3 Box_j \; x \; White_i + \boldsymbol{\Gamma X_i} + \epsilon_{ij} \tag{1}$$

Box_j indicates whether store j has the box, $White_i$ indicates applicant race, $Box_j \; x \; White_i$ is the interaction of those variables, and $\boldsymbol{X_i}$ 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 $Box_j \; x \; White_i$ coefficient in Column 1 indicates that the black-white gap is 2.8 percentage points larger among non-box employers (p<0.05) in the pre-period. 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 ten times as large, 3.1 percentage points compared to a base rate for black applicants of 9.4 percent; in proportional terms, white applicants received 33% more callbacks than black applicants did.

Figure 1 provides a visual representation of this cross-sectional comparison that further breaks down box employers' 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 non-box 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 2, Col. 1). A possible implication is that a substantial share of observed racial discrimination is 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 these patterns do not relate to the box. In 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 (versus others, which are mainly restaurants) *are* noticeably more likely to have the box; however, additional regression analyses (not shown here) indicate that this difference is not what drives the $Box_j x\ White_i$ interaction.[29] Overall callback rates were also nearly identical at the two groups of employers. Still, unobservable differences between the two groups are clearly possible, so the cross-sectional analysis is only suggestive; more rigorous causal identification is left to the analyses below.

4.2 Temporal Differences-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 3, we present differences-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 above. 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 seasons. Accordingly, the vector of

[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 Appendix A3.

[29] Specifically, we add a $Retail_j$ indicator, interacted with $White_i$, to the Col. 1 regression. The $Box_j x\ White_i$ coefficient only increases in magnitude, to 3.4 percentage points ($p<0.05$). Similar analyses also show that the small differences in the other observable employer characteristics in Appendix A3 do not explain the $Box_j x\ White_i$ effect.

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 (4794 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 comply with BTB. The $Box_j \; x \; White_i$ coefficient is -3.6 percentage points in the balanced sample (Col. 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 $Box_j \; x \; White_i$ 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 0.01 to 0.03.

Figure 2 represents these patterns visually (for the balanced sample of box-removers).[31] Like Figure 1, it further decomposes 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 1. 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 differences-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

[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 main term of interest, *Box * White*, retains its interpretation.
[31] The figure looks very similar if done only in the full sample.

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.

4.3. Triple-Differences Estimates of BTB's Effects

Here, we further analyze the causal effect of BTB on racial discrimination via a differences-in-differences-in-differences analysis, which exploits both sources of variation (cross-sectional and temporal) discussed above. This analysis compares the change 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 4, 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 applicant i to store j in time period t receives a callback:

[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.

$$Callback_{ijt} = \alpha + \beta_1 White_i + \beta_2 Post_t + \beta_3 BoxRemover_j + \beta_4 White_i \ x \ Post_t$$
$$+ \beta_5 White_i \ x \ BoxRemover_j + \beta_6 Post_t \ x \ BoxRemover_j \quad (2)$$
$$+ \beta_7 BoxRemover_j x \ White_i \ x \ Post_t + \boldsymbol{\Gamma X_i} + \epsilon_{ijt}$$

$Callback_{ijt}$ 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. $\boldsymbol{X_i}$ 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 post-periods, we substitute chain fixed effects, interacted with $White_i$ and $Post_t$, in place of the $BoxRemover_j$, $White_i \ x \ BoxRemover_j$, and $Post_t \ x \ BoxRemover_j$ terms in the equation above.[35]

The triple-differences estimate ($BoxRemover_j x \ White_i \ x \ Post_t$) 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 are included. This implies, for example, 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 differences-in-differences analysis but with a third difference (box remover vs. other employers) added. If one compares the triple-differences estimates in Table 4 to the corresponding $Box_j \ x \ White_i$ estimates in Table 3 (Columns 2-4, respectively), in each case, the triple-differences effect estimate is slightly larger in magnitude (with signs reversed, because we

[35] This produces the following estimating equation: $Callback_{ijkt} = \alpha + \beta_1 White_i + \beta_2 Post_t + \sum_{i=1}^{N} \beta_{3i} Chain_k + \beta_4 White_i \ x \ Post_t + White_i \ x \ \sum_{i=1}^{N} \beta_{5i} Chain_k + Post_t \ x \ \sum_{i=1}^{N} \beta_{6i} Chain_k + \beta_7 BoxRemover_j \ x \ White_i \ x \ Post_t + \epsilon_{ijkt}$ 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, *BoxRemover x White x 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 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 3, 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 treated 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 seven.

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 in order to provide equivalent power to estimate effects of a given size; hence, the corresponding Table 3 estimates are more precise even though they are estimated in smaller subsamples. Even so, the triple-differences estimates in Table 4 are significant at the 0.05 threshold or very close to it, with p-values ranging from 0.029 to 0.052. 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 pre-period and those that did not. Although our data are not long enough to compare pre-period 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 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 post-period 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.

4.4 Alternative Specifications and Samples

Table 5, Panels A and B show robustness checks for the balanced-sample estimates of BTB's effects on racial discrimination. Panel 5A shows alternative estimates for the $Box_j \ x \ White_i$ coefficient from the temporal differences-in-differences analysis; the corresponding main-specification result comes from Table 3, Column 2. Panel 5B shows alternative estimates for the $BoxRemover_j x \ White_i \ x \ Post_t$ coefficient from the triple-differences analysis; the corresponding main-specification result comes from Table 4, 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 2.6pp to 5.0pp) 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 0.01 for the double-differences and 0.05 for the triple-differences), with exceptions discussed below.

[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.

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 5A sample and 6.3% in the Panel 5B 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 5B, 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 "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

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

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

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 noncompliance 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.

Nonetheless, for readers who are more interested in BTB's effect on all pre-period box employers in the aggregate, in Column 7, we take an alternate approach, treating the noncompliers 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 * White* variables with *Pre* and *Pre*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 "pre-period box" variable, and change the interaction terms accordingly. These changes reduce the magnitude of the point estimates (to 2.7pp in Panel A and 2.6pp 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 5B, 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 (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

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

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 x White*. Column 8 shows that the *Box Remover x Post x 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 Appendix A3 do not explain the triple-differences effect either.[40]

Finally, in Appendix A7, we recreate the Table 4 analyses substituting the GED and employment gap variables, respectively, for $White$ and its interactions in Equation (2). 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. We do not see strong evidence of these effects, however. For the employment gap, there may be suggestive evidence. The point estimates are nontrivial (around 2.6 percentage points; Table A7.2) and cut in the expected direction. However, these estimates are statistically insignificant. In the GED analysis, the point estimates are also negative but smaller (and again insignificant); they are virtually zero in the full-sample analyses (Table A7.1, Col. 3). So we cannot characterize this as even suggestive evidence that employers are using GED as a proxy for criminal records.

5. Discussion

5.1. 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 et al. 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,

[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 0.05 for both analyses; 0.019 for the temporal differences-in-differences analysis and 0.027 for the triple-differences analysis. If one clusters on the individual store, they are 0.033 and 0.056, respectively.

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 non-box 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 2 shows that 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 (-0.6pp) and blacks with records see a small gain (+0.7pp), but blacks without records see very large losses (-5.1pp) and whites with records see very large gains (+4.5pp).[42]

5.2 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

[41] For the post-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 & 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 2 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 5.4 below for a back-of-the-envelope calculation.

remaining identification threats would have to come from unobserved differences that (1) affect box-remover versus other businesses differently (2) in ways that differ by race (among otherwise-identical applicants) and (3) 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 post-periods were seasonally nearly opposite.

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.[43]

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

[43] A variant of this concern is that BTB might reduce felon applications 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 3 and 4). Any changes to non-box 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 tradeoff between applying to both business types. And given that these employers lack the box, many would likely not *notice* changes in the percentage of their applicants with records, especially subtle changes.

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.

with records even after they "get their foot in the door." 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.

5.3 Mechanisms: Statistical Discrimination vs. 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 policymakers 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, for example, Eberhardt 2004; 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, i.e., 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 (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 2009, 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.[44]

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 Appendix A8, in which the probability of an interview is linear in the perceived probability that an applicant has a criminal record:

$$\begin{aligned} &\mathrm{P}\left(call_{ij} \middle| black_i, no\ box_j\right) \\ &\quad = g_b\,\mathrm{P}\left(call_{ij} \middle|\, black_i, crime_i\right) + (1 - g_b)\mathrm{P}\left(call_{ij} | black_i, no\ crime_i\right) \end{aligned} \tag{3}$$

That is, when there is no box, the probability 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

[44] 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 & 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, for example, Schott Foundation 2015), presumably the race gap in felony convictions must be substantially smaller if assessed within educational categories.

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 2 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 1.6pp secular increase that we observed at employers whose 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 71pp 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 & 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 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. For example, educational status (and specifically, GED versus high school diploma) is a *much* stronger predictor of criminal records than race is. 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). Given this point, 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. In addition, 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.[45]

---

[45] It is also possible that employers are assuming that black applicants have particularly *serious* criminal records qualitatively. While the conviction rate data does not directly get at this possibility, it seems irrational to apply this

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 theory of how stereotypes are formed. Alternate theories are possible, and multiple mechanisms could simultaneously contribute.[46] In any event, our data do support some form of stereotyping explanation, rather than the interpretation that employers are engaging in empirically informed statistical discrimination.

5.4 Likely Market Effects and Policy Implications

BTB may open doors to some applicants with records, but this gain comes at the expense of another group that faces serious employment challenges: black men. 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 & 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 2, adjusting for the 1.6pp secular rise in callbacks observed at control-group companies, as described in Section 5.1 above. Applying these changes to the real-world distribution of records from Bucknor & 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]

---

assumption to a large share of applicants given that felony convictions of *any* sort are relatively infrequent by age 22. Employers need not err on the side of caution to avoid any chance of a serious criminal record; BTB does not prevent them from eventually declining to hire after background checks.

[46] For example, perhaps stereotypes are grounded in longstanding 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.

[47] A similar exercise with the NLSY97 estimates leads to an estimated increase in the race gap of 4.2pp. One complicating factor is that only about half of real-world applicants have racially distinctive names, perhaps reducing the

Policymakers 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. 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). Another possibility is to alter employers' underlying incentives to avoid hiring people with records, perhaps by expanding tax incentives or reducing negligent-hiring liability.[49]

Racial disparity is not the only policy consideration surrounding BTB, and policymakers 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.

---

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 & 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 & 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] A non-representative survey by the Society for Human Resource Management (2012) found that a primary reason companies perform background checks is to reduce negligent-hiring liability.

References

Aigner, D.J. and Glen G. Cain. (1977). “Statistical Theories of Discrimination in Labor Markets”, *Industrial and Labor Relations Review* 30.

Arrow, K. (1973). The Theory of Discrimination. In *Discrimination in Labor Markets* Ashenfelter, O. and A. Rees (Eds) Princeton University Press, 3-33.

Autor, D.H. and D. Scarborough. (2008). “Does Job Testing Harm Minority Workers? Evidence from Retail Establishments,” *The Quarterly Journal of Economics* 123(1): 219-277.

Baert, Stijn and Elsy Verhofstadt. (2014). “Labour Market Discrimination against Former Juvenile Delinquents: evidence from a field experiment”, *Applied Economics* 47(11)

Bartik, A. and S. Nelson. (2016). “Credit Reports as Resumes: The Incidence of Pre-Employment Credit Screening”, MIT Department of Economics Graduate Student Research Paper 16-01

Bertrand, M. and S. Mullainathan. (2004). “Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination”, *American Economic Review* 94(4): 991-1013

Bordalo, Pedro, et al. (2016). “Stereotypes.” *Quarterly Journal of Economics* (forthcoming). Working paper: http://scholar.harvard.edu/files/shleifer/files/stereotypes_bcgs_feb25_paper.pdf.

Brame, R., S.D. Bushway, R. Paternoster and M. G. Turner. (2014). “Demographic Patterns of Cumulative Arrest Prevalence by Ages 18 and 23,” *Crime & Delinquency* 60(3): 471-486.

Bushway, S. (2004). “Labor Market Effects of Permitting Employer Access to Criminal History Records,” *Journal of Contemporary Criminal Justice*. Special Issue on Economics and Crime 20: 276-291.

Center for Health Statistics, Birth/EBC Confidential Data Files, New Jersey Department of Health, Trenton, NJ

City of New York (Jan 2016), “Young Men’s Initiative: Justice” http://www.nyc.gov/html/ymi/html/justice/justice.shtml#ban

Clarke, H. (December 20, 2012). “Protecting the Rights of Convicted Criminals: Ban the Box Act of 2012” *Washington Post.*

Clifford, R. and Shoag, D. (2016). “No More Credit Score: Employer Credit Check Bans and Signal Substitution” Unpublished Manuscript

Color of Change (November 2, 2015). “Civil Rights Group Responds to the ‘Ban the Box’ Executive Order” http://colorofchange.org/press/releases/2015/11/2/civil-rights-group-responds-ban-box-executive-orde/

Community Catalyst. (December 2, 2013). *Banning the Box in Minnesota—and across the United States*, http://www.communitycatalyst.org/blog/banning-the-box-in-minnesota-and-across-the-united-states#.UuG1__Yo46U.

Decker, S., E. Hedberg, and C. Spohn. (2015) “Criminal stigma, race, and ethnicity: The consequences of improsionment for employment” *Journal of Criminal Justice* 43: 108-121

Deming, D., N. Yuchtman, A. Abulafi, C. Goldin and L. Katz (2016). “The Value of Postsecondary Credentials in the Labor Market: An Experimental Study,” *American Economic Review* 106(3): 778-806.

Doleac, J. and B. Hansen (2016) "Does 'ban the box' help or hurt low-skilled workers? Statistical discrimination and employment outcomes when criminal histories are hidden" Unpublished manuscript. http://jenniferdoleac.com/wp-content/uploads/2015/03/Doleac_Hansen_BanTheBox.pdf

Equal Employment Opportunity Commission (EEOC). (2012). EEOC Enforcement Guidance 915.002. Consideration of Arrest and Conviction Records under Title VII of the Civil Rights Act of 1964.

Eberhardt, J.L. et al. 2004. "Seeing Black: Race, Crime, and Visual Processing," *Journal of Personality and Social Psychology* 87:876.

Emsellem, M. and B. Avery (2016). "Racial Profiling in Hiring: A Critique of New 'Ban-the-Box' Studies." National Employment Law Project Policy Brief.

Fang, H. and A. Moro. (2011). "Theories of Statistical Discrimination and Affirmative Action: A Survey" in Handbooks in Economics: Social Economics eds J. Benhabib, M. Jackson, and A. Bisin

Farber, H., D. Silverman, and T. von Wachter. (2015, Sept 17) "Factors Determining Callbacks to Job Applications by the Unemployed: An Audit Study" Unpublished Manuscript. Accessed at: http://www.irs.princeton.edu/sites/irs/files/event/uploads/audit_hf09.pdf

Finlay, K. (2009). "Effect of Employer Access to Criminal History Data on the Labor Market Outcomes of Ex-Offenders and Non-Offenders", *in Studies of Labor Market Intermediation*, 89 (David H. Autor, ed.).

Finlay, K. (2014). "Stigma in the Labor Market", Unpublished Manuscript

Freeman, R. (2008). " Incarceration, Criminal Background Checks, and Employment in a Low(er) Crime Society," *Criminology & Public Policy* 7: 405-412

Fryer, R.G. Jr and S.D. Levitt (2004). "The Causes and Consequences of Distinctly Black Names", *The Quarterly Journal of Economics* 119(3): 767-805

Heckman, J.J. (1998) "Detecting Discrimination", *Journal of Economic Perspectives* 12(2): 101-116

Heckman, J.J. and P.A. LaFontaine. (2010). "The American High School Graduation Rate: Trends and Levels" *Review of Economics and Statistics* 92(2): 244-262.

Heckman, J., and P. Siegelman (1993). "The Urban Institute Audit Studies: Their Methods and Findings," in ed. M. Fix and R. Struyk, *Clear and Convincing Evidence: Measurement of Discrimination in America,* 187-258.

Holzer, H.J. (2007). "Collateral Costs: The Effects of Incarceration on the Employment and Earning of Young Workers" IZA Discussion Paper No. 3118

Holzer, H.J., S. Raphael and M.A. Stoll (2006). "Perceived Criminality, Criminal Background Checks, and the Racial Hiring Practices of Employer*s," Journal of Law and Economics* 49:451.

Jarosch, G. and L. Pilossoph (2016). "Statistical Discrimination and Duration Dependence in the Job Finding Rate" Unpublished Manuscript

Kroft, K., F. Lange and M. Notowidigdo (2013). "Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment," *Quarterly Journal of Economics* 128(3): 1123-1167

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

Lahey, J. and R. Beasley (2009). "Computerizing Audit Studies," *Journal of Economic Behavior and Organization* 70(3): 508-514

List, J. (2004). "The Nature and Extent of Discrimination in the Marketplace: Evidence from the Field," *The Quarterly Journal of Economics* 119(1): 48-89

Love, M. (2011). "Paying Their Debt to Society: Forgiveness, Redemption, and the Uniform Collateral Consequences of Conviction Act," *Howard Law Journal* 54(3): 753-793

Minnesota Department of Human Rights (2015), "Ban The Box: Overview for Private Employers" http://mn.gov/mdhr/employers/banbox_overview_privemp.html Last Accessed Jan 19, 2016.

NAACP (2014, Jan). *Our Accomplishments,* http://www.naacp.org/pages/2106.

Neumark, D. (1996), "Sex Discrimination in Restaurant Hiring: An Audit Study", *The Quarterly Journal of Economics* 111(3): 915-941.

Neumark, D. (2012) "Detecting Discrimination in Audit and Correspondence Studies", *The Journal of Human Resources* 47(4): 1128-1157

Neumark, D., I. Burn, and P. Button (2015) "Is it Harder for Older Workers to Find Jobs? New and Improved Evidence from a Field Experiment" NBER WP 21669

Nosek, B.A. et al. (2007) "Pervasiveness and Correlates of Implicit Attitudes and Stereotypes," European Review of Social Psychology 2007:1.

Oreopoulos, P. (2011). "Why Do Skilled Immigrants Struggle in the Labor Market? A Field Experiment with Thirteen Thousand Resumes," *American Economic Journal: Economic Policy* 3(4): 148-71.

Pager, D. (March 2003). "The Mark of a Criminal Record," *American Journal of Sociology* 108(5): 937-975.

Pager, D., B. Western, & B. Bonikowski. (2009). "Discrimination in a Low-Wage Labor Market," *American Sociological Review*, 74:777-799.

Phelps, Edmund S. (1972). "The Statistical Theory of Racism and Sexism", *American Economic Review*. 62:659.

Phillips, D. C. (April 2016). "Do Low-Wage Employers Discriminate Against Applicants with Long Commutes? Evidence from a Correspondence Experiment" Unpublished Manuscript

Pinard, M. (January 7, 2014). "Ban the Box in Baltimore," *Baltimore Sun.*

Reaves, B. (December 2013). "Felony Defendants in Large Urban Counties, 2009 – Statistical Tables" US Department of Justice *Bureau of Justice Statistics* Report NCJ 243777

Riach, P.A. and J. Rich (2002). "Field Experiments in Discrimination in the Market Place", *The Economic Journal* 112(483): F480-F518

Rodriguez, M and B. Avery. (February 2017). "Ban The Box: U.S. Cities, Counties, and States Adopt Fair-Chance Policies to Advance Employment Opportunities for People with Past Convictions". National Employment Law Project Guide Accessed June 9, 2016: http://www.nelp.org/publication/ban-the-box-fair-chance-hiring-state-and-local-guide/

Shannon, S., C. Uggen, M. Thompson, J. Schnittker, and M. Massoglia (2011). "Growth in the U.S. Ex-Felon and Ex-Prisoner Population, 1948-2010" Unpublished Manuscript

Shoag, D. and S. Veuger (2016) "No Woman No Crime: Ban the Box, Employment, and Upskilling" Unpubslihed Manuscript

Society for Human Resource Management. (July 19 2012). *Background Checking – the Use of Criminal Background Checks in Hiring Decisions.*

Southern Coalition for Social Justice. (2013). *Ban the Box Community Initiative Guide*, http://www.southerncoalition.org/program-areas/criminal-justice/ban-the-box-community-initiative-guide/.

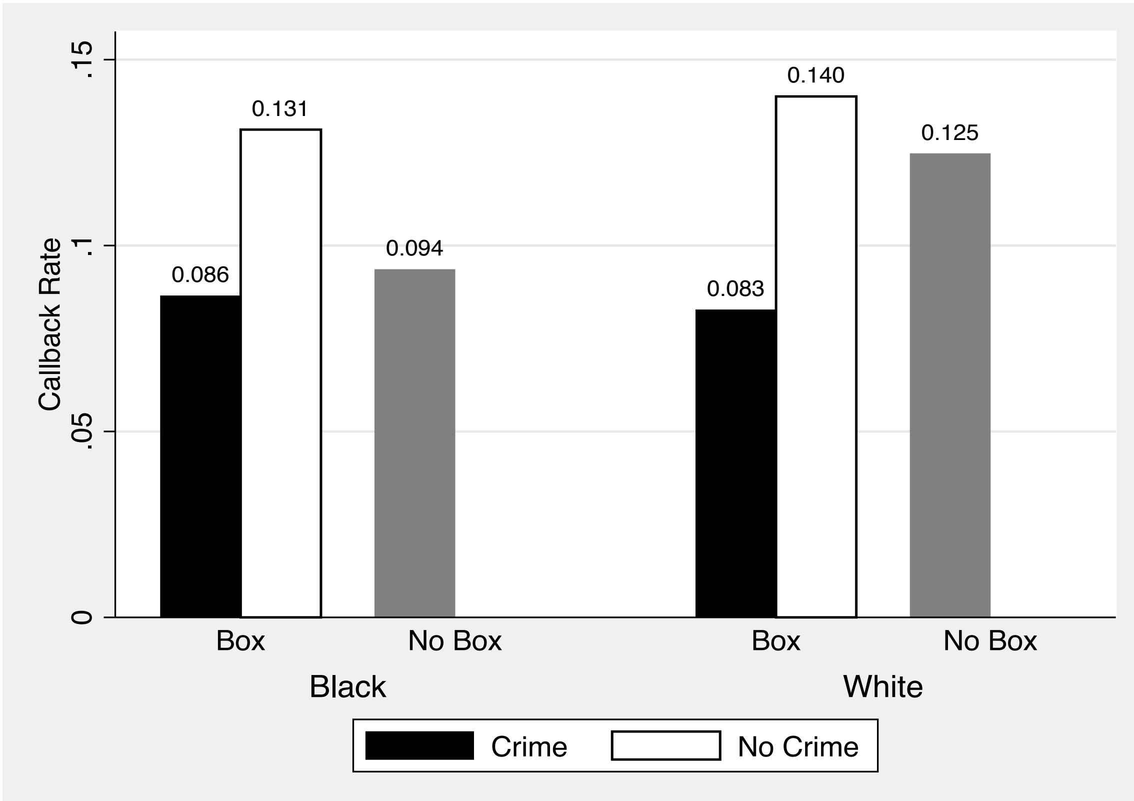
Starr, S. (2015). "Do Ban the Box Laws Reduce Employment Barriers for Black Men?" Unpublished Manuscript.

Stoll, Michael A. (2009). *Ex-Offenders, Criminal Background Checks, and Racial Consequences in the Labor Market*, 1 Univ. of Chicago Legal Forum 381 (2009).

Uggen, C., M. Vuolo, S. Lageson, E. Ruhland, and H. Whitham (2014). "The Edge of Stigma: An Experimental Audit of the Effects of Low-Level Criminal Records on Employment" *Criminology* 52(4): 627-654.

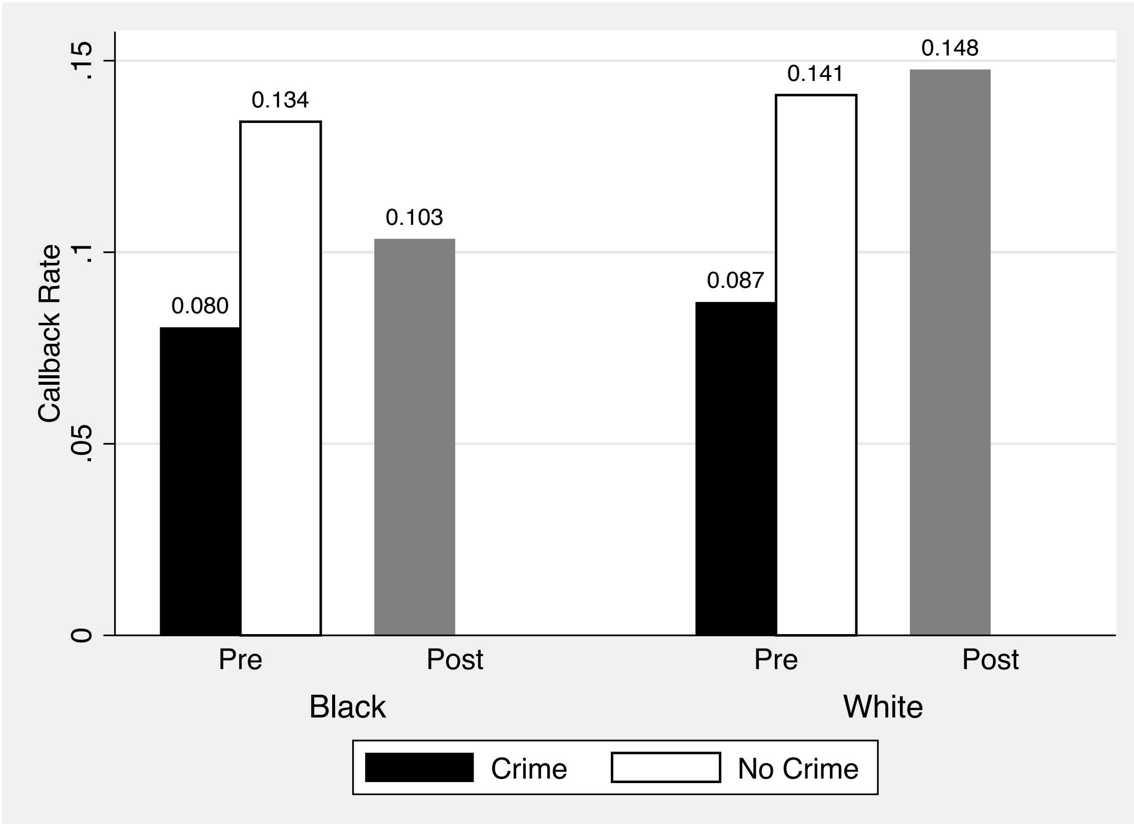
Wozniak, A. (2015, July). "Discrimination and the Effects of Drug Testing on Black Employment". *The Review of Economics and Statistics* 97(3): 548-566

Figure 1: Callback Rates by Race, Crime, and Box: Pre-Period Applications Only



Notes: This figure compares callback rates *within* the pre-period before Ban the Box went into effect, comparing applications with the criminal-records box and those without the box.

Figure 2: Callback Rates by Race, Criminal Record, and Period: Balanced Box Removers Only



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

Table 1a: Means of Applicant and Application Characteristics and Callback Rates by Period

	Pre-Period	Post-Period	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
N	7245	7392	14640

Notes: Callback implies application received a personalized positive response from the employer (either via phone or e-mail). Interview request means the positive response specifically mentioned an interview. Application has box means that the application asked about criminal records. Employment (emp) gap is a 11-13 month employment gap in work history, no emp gap is a 0-2 month gap.

Table 1b: Callback Rates by Crime Status for Stores with the Box in the Pre-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
N	1319	1336	703	633	2655

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

Table 2: Effects of Applicant Characteristics on Callback Rates

	(1)	(2)	(3)
White	0.024***	-0.001	-0.001
	(0.006)	(0.009)	(0.009)
Conviction	-0.014**	-0.052***	
	(0.005)	(0.012)	
GED	-0.004	0.010	0.010
	(0.005)	(0.014)	(0.013)
Employment Gap	0.002	0.011	0.011
	(0.005)	(0.010)	(0.010)
Pre-Period	-0.015		
	(0.010)		
Drug Conviction			-0.050***
			(0.013)
Prop. Conviction			-0.054***
			(0.014)
N	14637	2918	2918
Sample	All	Box	Box
Chain FE	Yes	Yes	Yes
Center FE	Yes	Yes	Yes

Notes: * p<0.1 ** p<0.05 *** p<0.01. 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 (vs black), Conviction is an indicator for whether the applicant has a felony conviction, GED is an indicator for having a GED (vs a regular high-school diploma), and Employment Gap is an indicator for whether the applicant has a 11-13 month gap in work history between the previous two jobs (vs. a 0-2 month gap). "Drug Conviction" and "Prop. Conviction" break the Conviction variable into a categorical variable based on crime type (drug vs. property crime); no conviction is the base category. The Box sample is employers with the box on their application.

Table 3: Effects of the Box on Racial Discrimination: Differences-in-Differences

	(1)	(2)	(3)	(4)	(5)
Box x White	-0.030**	-0.036**	-0.033**	-0.027**	0.002
[*White x Pre, Col 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, Col 5]*	(0.024)	(0.015)	(0.013)	(0.139)	(0.017)
N	7245	3712	4794	4794	7476
Controls	Yes	Yes	Yes	Yes	Yes
Center FE	Yes	No	Yes	Yes	No
Chain FE	No	No	No	Yes	No
Post x Chain FE	No	No	No	Yes	No
White x 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: * p<0.1 ** p<0.05 *** p<0.01. Standard errors in parentheses clustered on chain. 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 4 applications, one white/black pair in each period. Fixed effects can include geographic center, chain, post x chain, and white x 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 (vs regular high-school diploma) and whether they had an employment gap. Box Variation indicates the source of variation in the Box variable – cross-sectional means the variation comes from a comparison of box and no-box stores in the pre-period; 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.

Table 4: Effects of Ban the Box on Racial Discrimination: Triple Differences

	(1)	(2)	(3)
Box Remover x Post x White	0.039*	0.040**	0.035*
	(0.020)	(0.018)	(0.018)
Post x White	-0.002	-0.006	-0.006
	(0.014)	(0.012)	(0.013)
Box Remover x Post	-0.019	-0.011	
	(0.023)	(0.019)	
Box Remover x White	-0.017	-0.021	
	(0.015)	(0.014)	
Box Remover	0.016	0.009	
	(0.028)	(0.024)	
White	0.024**	0.028**	0.098
	(0.012)	(0.011)	(0.129)
Post	0.016	0.012	0.339**
	(0.017)	(0.015)	(0.139)
N	11188	14637	14637
Controls	Yes	Yes	Yes
Center FE	No	Yes	Yes
Chain FE	No	No	Yes
Post x Chain FE	No	No	Yes
White x Chain FE	No	No	Yes
Sample	Balanced	Full	Full

Notes: * p<0.1 ** p<0.05 *** p<0.01. 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-period and then removed it due to BTB. The balanced sample indicates the sample where we sent exactly 4 applications, one white/black pair in each period. Controls are whether the applicant had a GED (vs regular high-school diploma) and whether they had an employment gap. Fixed effects can include, chain, post x chain, white x chain, or center, and are included as indicated.

Table 5: Effects of Ban the Box on Racial Discrimination: Robustness Checks, Balanced Sample

5A. Temporal Differences-in-Differences

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	N/A (Main)	Inter-view	Add Rev. Compl.	Drop RA Errors	NJ Only	NYC Only	Intent-to-Treat	Retail Int.
Box x White	-0.036**	-0.041**	-0.034**	-0.035**	-0.050*	-0.029	-0.027**	
[Pre x White, Col. 7]	(0.014)	(0.016)	(0.013)	(0.013)	(0.028)	(0.019)	(0.013)	
N	3712	3712	3848	3686	1400	2312	4112	

5B: Triple Differences

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
BoxRemover x Post x White	0.039*	0.039**	0.034*	0.038*	0.048	0.034	0.026	0.041*
[Pre-PeriodBox x Post x White, Col. 7]	(0.020)	(0.019)	(0.018)	(0.020)	(0.040)	(0.021)	(0.020)	(0.023)
N	11188	11188	11324	11160	4376	6812	11188	11188

Notes: * p<0.1 ** p<0.05 *** p<0.01. Standard errors clustered on the chain are shown in parentheses. Panel 6A shows robustness checks corresponding to the temporal differences-in-differences specification from Table 3, Col. 2; Panel 6B shows robustness checks corresponding to the triple-differences specification from Table 5, Col. 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 6A regressions are further confined to "box removers" (employers that had the box before BTB and removed it afterward), whereas the Panel 6B regressions include non-box-removers, which provide the third difference.

In each panel, Col. 1 shows the coefficient of interest (*Box x White* or *BoxRemover x Post x White*, respectively) from the respective main specification. The remaining columns show modifications of the sample or specification, labeled as "Change" at the top of the column. Col. 2 uses the interview as the dependent variable (all others use "callback). Col. 3 adds observations ("reverse compliers") that had the box in the *post*-BTB period but *not* in the pre-period; in the triple-differences analysis, these are coded -1 on the *BoxRemover* variable. Col. 4 drops instances where RA erred and answered a box question they were not required to answer or did not answer one they should have. Cols. 5 and 6 are restricted to the New Jersey and New York City subsets of the sample, respectively. Col. 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 6A *Pre* (a pre-BTB period indicator) is substituted for *Box* and the coefficient shown is *Pre x White*; and in Panel 6B *PrePeriodBox* (an indicator for whether the store had the box before BTB) is substituted for *BoxRemover*, and the coefficient shown is *PrePeriodBox x Post x White*. In Col. 8 (Panel 6B only), a *Retail* industry indicator (interacted with Post, White, and Post x White) is added to the regression.

Appendix

A1. Applicant Profile Details

Applicant profiles consist of all information that our RAs might need in order to fill out a given job application. In addition to our characteristics of interest, this included many details such as previous job titles and descriptions, home addresses, names of high schools, references, and e-mail addresses. We wanted to keep these additional characteristics as similar as possible while still introducing slight (random) variation so as not to arouse employer suspicion.

(1) *Work history*: All job applicants have about 3.5 years of work experience: about 2 years as crew members at fast-food chains or convenience stores and about 1.5 years in manual labor jobs such as home improvement, landscaping, or moving. Specific fast-food and convenience-store employers were randomized from lists of chains to which we were not applying. The manual labor jobs were randomly assigned to be in landscaping, paving, moving, home improvement, or lawn care, with fictitious company names randomized from generic-sounding options. Applicants were similarly assigned generic job descriptions implying entry-level, unskilled crew-member positions in the fictitious companies.

All applicants are unemployed at the time of the job application, having ended their most recent job 2 or 3 months before the application is submitted. Descriptions of previous job duties and reasons for leaving jobs varied slightly. Applicants with employment gaps have 11 to 13 months of unemployment between the two past jobs; those without employment gaps have only 0- to 2-month gaps.

(2) *Address and center city:* To ensure that applicants lived near the jobs they were applying to, we chose 40 geographically distributed cities or towns in New Jersey and 44 in New York City to serve as centers where the applicants' addresses would be located; each center then served as the base for applications to jobs located nearby. We first narrowed the entire list of New Jersey cities and towns as well as community districts in New York City to those that were at least 6% black, were at least 20% white, and had median annual incomes less than $100,000. We then used an optimization tool in the ArcGIS software package to select centers that would minimize distance to jobs; in New Jersey this was based on the distribution of postings then found (in January 2015) on snagajob.com, and in New York

City it was based on the locations of employers that we located in the BusinessUSA database. In New Jersey, we assigned every municipality in the state to its nearest center, excluding only a few small towns that were more than 20 miles from any center. In New York City, we minimized distances subject to a constraint of equal distribution of chains across centers—for example, all chains with 44 or fewer locations were distributed such that no more than one location was assigned to each center, while a chain with 45 to 88 locations would be distributed with one to two locations per center, and so forth.

Within each center, eight qualifying addresses were located within census blocks that were at least 10% black and 20% white and that had a median annual income less than $100,000. All addresses came from different streets, and we used Google Street View to ensure that the choices were appropriate residential or mixed-use blocks and that they did not notably differ from one another. Addresses were then slightly changed so as not to represent real addresses, and they were then randomly assigned to applicants.

(3) *High school or GED program:* For diploma earners, high schools for the New Jersey applicants were chosen to be at least 30 miles away from the center where the applicant would apply from, to reduce the chance that the high school would be very familiar to the employer. High schools for New York City applicants were divided equally between New Jersey and upstate New York schools, since similar geographic separation could not be achieved within the city. The high schools used were all at least 10% black, are at least 20% white, have at least 25,000 people, and do not have median incomes more than $100,000. In addition, the high schools do not have median test scores above the 90th or below the 10th percentile in the state. Applicants with GEDs were randomly assigned descriptions and names of New Jersey or New York GED training programs.

(4) *References:* Two fictitious references with phone numbers were created, representing the applicant's supervisors for each of two previous jobs. To complement and strengthen the racial signal provided by our applicant names, the previous supervisor from the manual labor job was given a racially distinctive name suggesting the same race as the applicant. The previous supervisor of the retail or restaurant job was given a race-neutral name. However, no employers ever called the phone numbers that we purchased and provided for the references, suggesting that little attention was likely paid to them.

(5) *Phone number:* Each applicant was assigned a phone number based on center, race, criminal history, and time period. Thus, each center had at least four potential phone numbers during each period; in New York City, because we were sending a larger number of distinct applications per center, we used eight numbers per center/period. The result of this approach is that no store received two applications using the same phone number. That method also helps us identify which application a voicemail belongs to, in case of unclear messages. We purchased these phone numbers from www.callfire.com. The wording and voice on the outgoing voice mail greeting were randomized across several options and designed to sound like a generic voicemail greeting for someone who has not recorded a personalized one.

(6) *E-mail address:* A unique e-mail address was created for each applicant, with the domain always the same but the format randomly varied (incorporating the applicant's name plus varied numbers and punctuation).

(7) *Criminal record:* Applicants with felony convictions were randomly assigned either a property crime or a drug crime. Within those two categories, several potential crimes were chosen—all of them meant to imply similar levels of seriousness. In addition, many applications with the box ask the applicant to "Please explain," and we randomized generic-sounding explanations and expressions of remorse.

We randomly generated each profile using the Resume Randomizer program of Lahey and Beasley (2009). Applicant pairs were always of opposite race, and characteristics were otherwise randomized, with some characteristics forced to vary within pairs to disguise similarity. For example, both members of the pair could have high school diplomas, but never from the same high school or the same town, and so forth. For examples of profiles, which are several pages in length, please e-mail the authors.

A2. Names Used

Table A2.1: White and Black Names Used for Applicants

White Names				Black Names			
First	%White	Last	%White	First	%Black	Last	%Black
SCOTT	88.87	WEBER	94.37	TYREE	97.94	PIERRE	97.78
THOMAS	86.92	ESPOSITO	93.30	TERRELL	96.23	WASHINGTON	90.28
CODY	86.71	SCHMIDT	92.63	DAQUAN	96.04	ALSTON	88.96
RYAN	85.37	BRENNAN	92.45	JAQUAN	95.03	BYRD	85.50
NICHOLAS	84.99	MEYER	92.27	DARNELL	93.43	INGRAM	78.63
DYLAN	84.70	KANE	91.75	JAMAL	91.36	JACKSON	76.32
MATTHEW	83.97	HOFFMAN	91.38	MARQUIS	91.36	BANKS	75.68
JACOB	83.37	RYAN	89.98	JERMAINE	89.45	FIELDS	74.83
KYLE	82.93	WAGNER	89.96	DENZEL	89.27	BRYANT	74.49
TYLER	82.82	HANSEN	89.60	DWAYNE	88.89	WILLIAMS	74.22
SEAN	82.41	SNYDER	88.84	REGINALD	88.41	SIMMONS	72.45
DOUGLAS	81.93	ROMANO	88.84	TYRONE	86.75	CHARLES	72.33
SHANE	81.11	O'NEILL	88.72	MALCOLM	86.06	HAWKINS	70.81
JOHN	80.36	RUSSO	88.67	DARRYL	84.78	ROBINSON	70.70
STEPHEN	80.12	FOX	86.43	TERRANCE	84.12	JENKINS	70.50
		SWEENEY	86.03	MAURICE	82.47	FRANKLIN	70.45
		SULLIVAN	85.08	ISAIAH	74.06	JOSEPH	70.42
				ELIJAH	72.35		

Notes: The %race columns indicate the percentage of male babies born in NJ between 1989 and 1997 with that first or last name that were of that race (i.e. 88.87% of male babies with the first name Scott are White).

A3. Comparing Box and No-Box Stores in the Pre-BTB Period

Table A3.1: Characteristics of Stores With and Without the Box Before BTB

	(1) Box	(2) No Box	(3) Diff
Store CBG %White	58.91	61.12	2.21*
Store CBG %Black	15.37	13.77	-1.60*
Crime Rate	18.48	20.35	1.86***
Avg Num Employees	49.39	43.71	-5.68
Avg Sales Volume	8928.94	7982.05	-946.89
Retail	0.58	0.42	-0.16***
N	1426	2498	3924

Notes: Store CBG %White and Store CBG %Black are the population shares in the store's census block group that are white and black, respectively, from Census data. Crime Rate is the number of index crimes reported to police per 100,000 population in 2015 in the police precinct (for NYC) or the town (for NJ) where the store is located. See Agan & Starr 2017 for details on construction of these variables. Average number of employees and sales volume come from 2015 BusinessUSA data, and represent the average number of employees or average sales volume in establishments of the chain the store belongs to in NJ and NYC in 2015. These are measured at the chain level – every store of the same chain has the same value. Retail is defined using SIC codes provided in the BusinessUSA data.

Table A3.2: Regression of Presence of Box (Before BTB) on Store Characteristics

	(1) Application has Box
Store CBG %White	-0.0004
	(0.0004)
Store CBG %Black	0.0004
	(0.0005)
Crime Rate	-0.0013*
	(0.0005)
Avg Num Employees (10s)	0.0001
	(0.0002)
Avg Sales Volume ($1000s)	-0.0002
	(0.0004)
Retail	0.1383***
	(0.0160)
Constant	0.3422***
	(0.0329)
N	3772

Notes: * p<0.1 ** p<0.05 *** p<0.01. Standard errors in parentheses. See notes to Table A3.1 for definitions of variables.

A4. Robustness Checks for Main Effects of Race and Crime

Table A4.1: Robustness Checks on Main Effect of White

	(1)	(2)	(3)	(4)	(5)
White	0.024***	0.014***	0.024***	0.045***	0.007
	(0.006)	(0.005)	(0.006)	(0.010)	(0.005)
N	14637	14637	14637	6401	8236
Specification	Main	Interview	Ungroup Chain FE	Main	Main
Sample	All	All	All	NJ-All	NYC-All

Notes: Dependent variable is whether the application received a callback. Standard errors clustered on chain in parentheses. Column (1) reproduces the White coefficient from Column 1 of Table 2, and the remaining columns show the White coefficient from different specifications. Column (2) uses interview as the dependent variable rather callback. Column (3) uses ungrouped chain FE, whereas in the main specification the smallest chains were grouped. Columns (4) and (5) separate the sample by jurisdiction (NJ and NYC).

Table A4.2: Robustness Checks on Main Effect of Crime in the Box Sample Only

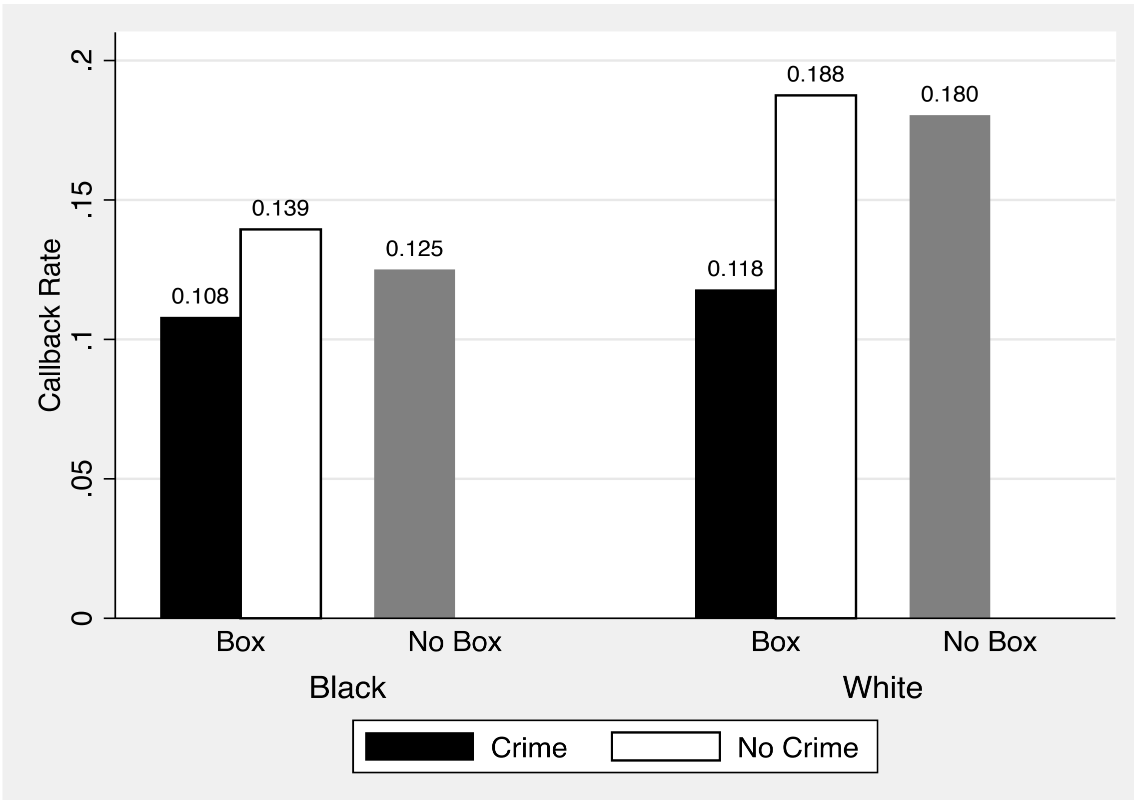
	(1)	(2)	(3)	(4)	(5)
Crime	-0.052***	-0.035***	-0.052***	-0.053**	-0.051***
	(0.012)	(0.006)	(0.012)	(0.022)	(0.016)
N	2918	2918	2918	1156	1762
Specification	Main	Interview	Ungroup Chain FE	Main	Main
Sample	Box	Box	Box	NJ-Box	NYC-Box

Notes: Dependent variable is whether the application received a callback. Standard errors clustered on chain in parentheses. Column (1) reproduces the Crime coefficient from Column 2 of Table 2, and the remaining columns show the Crime coefficient from different specifications. Column (2) uses interview as the dependent variable rather callback. Column (3) uses ungrouped chain FE rather than grouped. Columns (4) and (5) separate the sample in the NJ sample and the NYC sample.

A5. Analysis Tables for NJ Only

This appendix recreates Figures 1 and 2 and Tables 1-4 for NJ Applications Only.

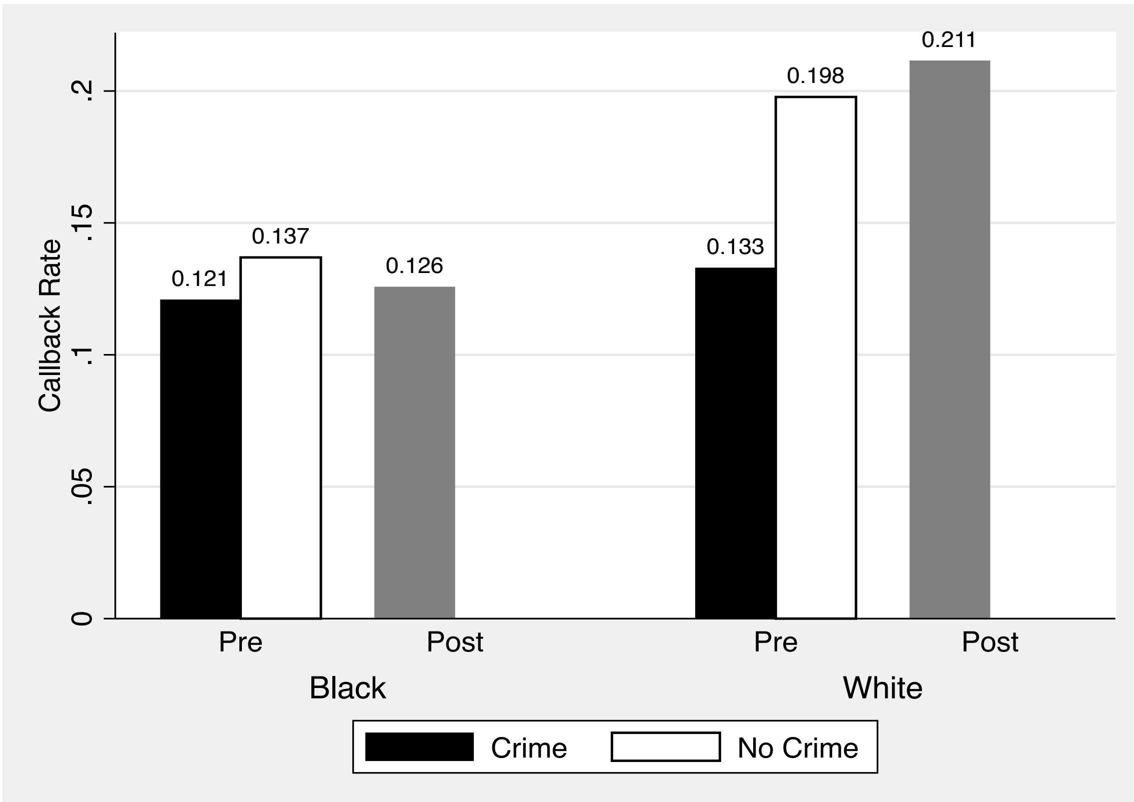
Figure A5.1: Callback Rates by Race, Crime, and Box: Pre-Period NJ Applications Only



Notes: Limited to only NJ applications. This figure compares callback rates *within* the pre-period before Ban the Box went into effect, comparing applications with the box (application which ask about criminal records) and those without (applications that do not ask about criminal records).

Figure A5.2: Callback Rates by Race, Criminal Record, and Period:

New Jersey Balanced Sample, Box Removers Only



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

Table A5.1a: Means of Applicant and Application Characteristics and Callback Rates by Period, NJ Only

	Pre-Period	Post-Period	Combined
Characteristics:			
White	0.507	0.495	0.500
Crime	0.498	0.504	0.501
GED	0.506	0.513	0.510
Employment Gap	0.503	0.504	0.504
Application has Box	0.362	0.034	0.181
Results:			
Callback Rate	0.147	0.146	0.147
Interview Req	0.081	0.076	0.078
Callback Rate by Characteristics:			
Black	0.125	0.124	0.124
White	0.170	0.170	0.170
GED	0.139	0.143	0.142
HSD	0.156	0.150	0.152
Emp Gap	0.145	0.149	0.147
No Emp Gap	0.150	0.144	0.146
Observations	2864	3537	6401

Notes: Sample limited to NJ applications. Callback implies application received a personalized positive response from the employer (either via phone or e-mail). Interview request means the positive response specifically mentioned an interview. Application has box means that the application asked about criminal records. Employment (emp) gap is a 11-13 month employment gap in work history, no emp gap is a 0-2 month gap.

Table A5.1b: Callback Rates by Crime Status for Stores with the Box in the Pre-Period, NJ Only

	No Crime	Crime	Property	Drug	Combined
Callback Rate	0.164	0.113	0.102	0.127	0.138
Callback Black	0.139	0.108	0.087	0.139	0.124
Callback White	0.188	0.118	0.118	0.118	0.151
Observations	507	530	293	237	1037

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

Table A5.2: Effects of Applicant Characteristics on Callback Rates NJ ONLY

	(1)	(2)	(3)
White	0.045***	0.026	0.025
	(0.010)	(0.021)	(0.021)
Crime	-0.015**	-0.053**	
	(0.007)	(0.022)	
GED	-0.016**	-0.003	-0.002
	(0.008)	(0.028)	(0.028)
Employment Gap	0.001	-0.006	-0.006
	(0.007)	(0.012)	(0.012)
Pre-Period	-0.003		
	(0.014)		
Drug Crime			-0.042
			(0.030)
Property Crime			-0.063**
			(0.025)
N	6401	1156	1156
Sample	All	Box	Box
Chain FE	Yes	Yes	Yes
Center FE	Yes	Yes	Yes

Notes: * $p<0.1$ ** $p<0.05$ *** $p<0.01$. This table recreates Table 2 for NJ applications only. 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 (vs black), Conviction is an indicator for whether the applicant has a felony conviction, GED is an indicator for having a GED (vs a regular high-school diploma), and Employment Gap is an indicator for whether the applicant has a 11-13 month gap in work history between the previous two jobs (vs. a 0-2 month gap). "Drug Conviction" and "Prop. Conviction" break the Conviction variable into a categorical variable based on crime type (drug vs. property crime); no conviction is the base category. The Box sample is employers with the box on their application.

Table A5.3: Effects of the Box on Racial Discrimination: Differences-in-Differences NJ ONLY

	(1)	(2)	(3)	(4)	(5)
Box x White	-0.031	-0.050*	-0.037	-0.032	-0.002
[*White x Pre, Col 5]*	(0.032)	(0.028)	(0.031)	(0.031)	(0.030)
White	0.057***	0.086***	0.072***	0.143*	0.041**
	(0.022)	(0.013)	(0.017)	(0.074)	(0.018)
Box	-0.002	0.003	-0.009	-1.111***	0.009
[*Pre, Col 5]*	(0.035)	(0.025)	(0.019)	(0.087)	(0.029)
N	2864	1400	2054	2054	2976
Controls	Yes	Yes	Yes	Yes	Yes
Center FE	Yes	No	Yes	Yes	No
Chain FE	No	No	No	Yes	No
Post x Chain FE	No	No	No	Yes	No
White x 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: * p<0.1 ** p<0.05 *** p<0.01. This table recreates Table 3 for NJ Applications Only. Standard errors in parentheses clustered on chain. 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 4 applications, one white/black pair in each period. Fixed effects can include geographic center, chain, post x chain, and white x 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 (vs regular high-school diploma) and whether they had an employment gap. Box Variation indicates the source of variation in the Box variable – cross-sectional means the variation comes from a comparison of box and no-box stores in the pre-period; 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.

Table A5.4: Effects of Ban the Box on Racial Discrimination, Triple Differences NJ ONLY

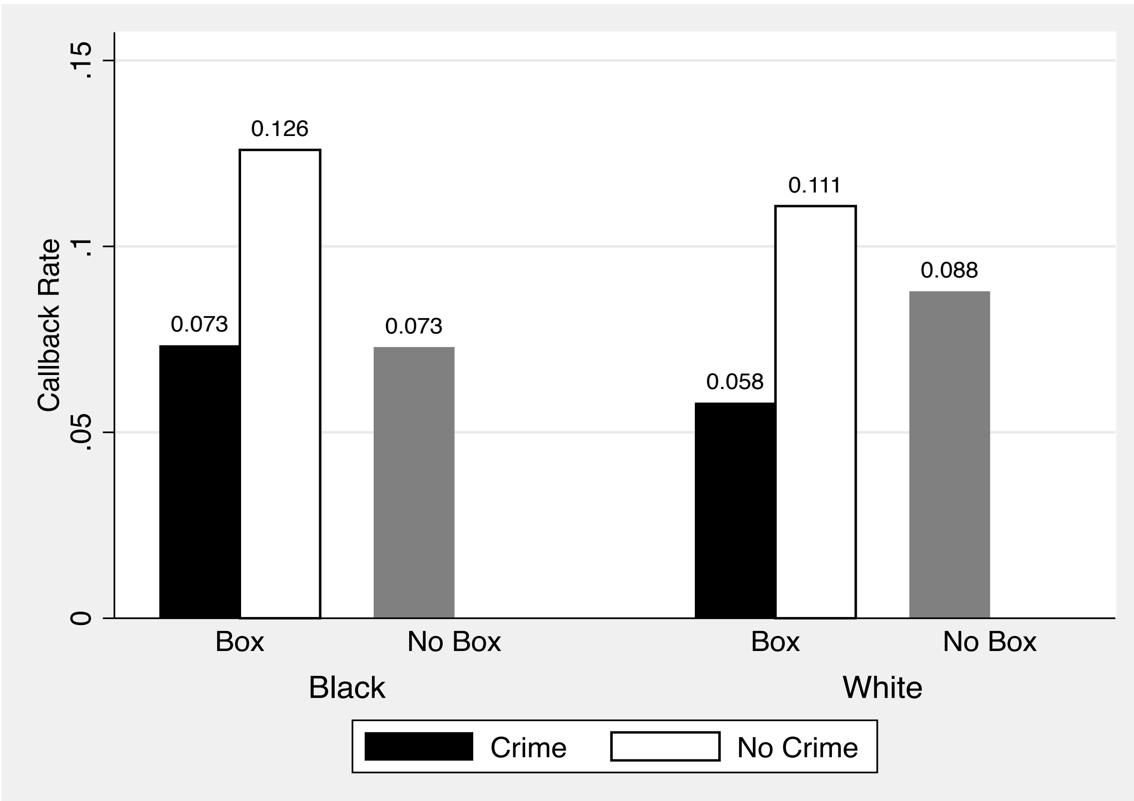
	(1)	(2)	(3)
Box Remover x Post x White	0.048	0.056	0.044
	(0.040)	(0.038)	(0.038)
Post x White	0.003	-0.017	-0.010
	(0.030)	(0.024)	(0.023)
Box Remover x Post	0.005	0.009	
	(0.037)	(0.027)	
Box Remover x White	-0.002	-0.018	
	(0.032)	(0.031)	
Box Remover	0.002	-0.004	
	(0.040)	(0.033)	
White	0.039*	0.052**	0.009
	(0.021)	(0.021)	(0.034)
Post	-0.009	-0.001	1.009***
	(0.029)	(0.022)	(0.034)
Observations	4376	6401	6401
R^2	0.006	0.031	0.215
Controls	Yes	Yes	Yes
Center FE	No	Yes	Yes
Chain FE	No	No	Yes
Post x Chain FE	No	No	Yes
White x Chain FE	No	No	Yes
Sample	Balanced	Full	Full

Notes: * p<0.1 ** p<0.05 *** p<0.01. This table recreates Table 4 for NJ applications only. 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-period and then removed it due to BTB. The balanced sample indicates the sample where we sent exactly 4 applications, one white/black pair in each period. Controls are whether the applicant had a GED (vs regular high-school diploma) and whether they had an employment gap. Fixed effects can include, chain, post x chain, white x chain, or center, and are included as indicated

A6. Analysis Tables for NYC Only

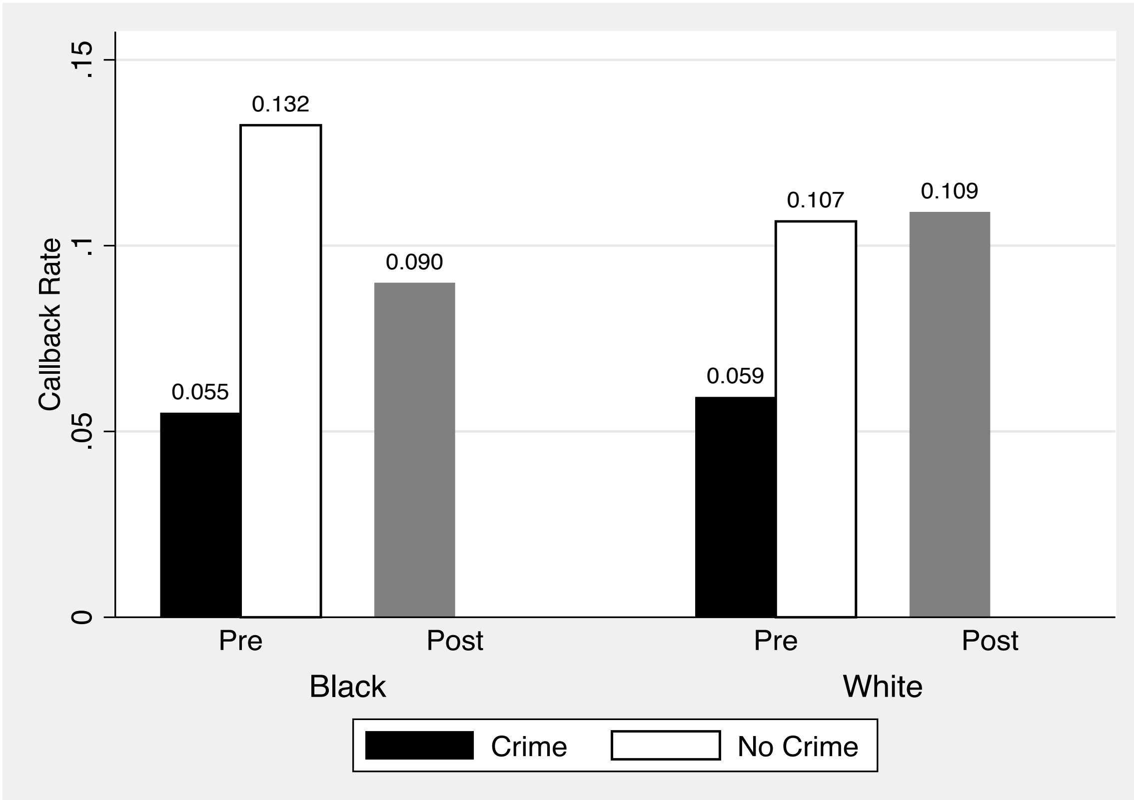
This appendix recreates Figures 1 and 2 and Tables 1-4 for NYC Applications Only.

Figure A6.1: Callback Rates by Race, Crime, and Box: Pre-Period NYC Applications Only



Notes: Limited to only NYC applications. This figure compares callback rates *within* the pre-period before Ban the Box went into effect, comparing applications with the box (application which ask about criminal records) and those without (applications that do not ask about criminal records).

Figure A6.2: Callback Rates by Race, Criminal Record, and Period:
New York City Balanced Sample, Box Removers Only



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

Table A6.1a: Means of Applicant and Application Characteristics and Callback Rates by Period, NYC Only

	Pre-Period	Post-Period	Combined
Characteristics:			
White	0.500	0.499	0.499
Crime	0.496	0.521	0.508
GED	0.492	0.493	0.492
Employment Gap	0.486	0.504	0.494
Application has Box	0.369	0.037	0.214
Results:			
Callback Rate	0.085	0.105	0.094
Interview Req	0.046	0.059	0.052
Callback Rate by Characteristics:			
Black	0.083	0.099	0.090
White	0.087	0.110	0.098
GED	0.083	0.112	0.097
HSD	0.086	0.098	0.092
Emp Gap	0.086	0.104	0.095
No Emp Gap	0.084	0.105	0.094
Observations	4381	3855	8236

Notes: Sample limited to NYC applications. Callback implies application received a personalized positive response from the employer (either via phone or e-mail). Interview request means the positive response specifically mentioned an interview. Application has box means that the application asked about criminal records. Employment (emp) gap is a 11-13 month employment gap in work history, no emp gap is a 0-2 month gap.

Table A6.1b: Callback Rates by Crime Status for Stores with the Box in the Pre-Period, NYC Only

	No Crime	Crime	Property	Drug	Combined
Callback Rate	0.118	0.066	0.071	0.061	0.092
Callback Black	0.126	0.073	0.093	0.052	0.099
Callback White	0.111	0.058	0.046	0.069	0.085
Observations	812	806	410	396	1618

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

Table A6.2: Effects of Applicant Characteristics on Callback Rates: NYC Only

	(1)	(2)	(3)
White	0.007	-0.018**	-0.018**
	(0.005)	(0.009)	(0.009)
Crime	-0.014*	-0.051***	
	(0.008)	(0.016)	
GED	0.004	0.014	0.014
	(0.006)	(0.010)	(0.010)
Employment Gap	0.001	0.021*	0.021*
	(0.006)	(0.012)	(0.012)
Pre-Period	-0.024		
	(0.017)		
Drug Crime			-0.058***
			(0.017)
Property Crime			-0.045***
			(0.016)
Observations	8236	1762	1762
Sample	All	Box	Box
Chain FE	Yes	Yes	Yes
Center FE	Yes	Yes	Yes

Notes: * p<0.1 ** p<0.05 *** p<0.01. This table recreates Table 2 for NJ applications only. 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 (vs black), Conviction is an indicator for whether the applicant has a felony conviction, GED is an indicator for having a GED (vs a regular high-school diploma), and Employment Gap is an indicator for whether the applicant has a 11-13 month gap in work history between the previous two jobs (vs. a 0-2 month gap). "Drug Conviction" and "Prop. Conviction" break the Conviction variable into a categorical variable based on crime type (drug vs. property crime); no conviction is the base category. The Box sample is employers with the box on their application.

Table A5.3: Effects of the Box on Racial Discrimination: Differences-in-Differences NYC ONLY

	(1)	(2)	(3)	(4)	(5)
Box x White	-0.029**	-0.029	-0.026	-0.022	0.005
[*White x Pre, Col 5]*	(0.013)	(0.019)	(0.017)	(0.017)	(0.010)
White	0.015*	0.018	0.014	0.245***	0.010
	(0.009)	(0.018)	(0.013)	(0.048)	(0.008)
Box	0.026	0.003	0.001	-0.207***	-0.032
[*Pre, Col 5]*	(0.028)	(0.029)	(0.027)	(0.009)	(0.023)
N	4381	2312	2740	2740	4500
Controls	Yes	Yes	Yes	Yes	Yes
Center FE	Yes	No	Yes	Yes	No
Chain FE	No	No	No	Yes	No
Post x Chain FE	No	No	No	Yes	No
White x 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: * p<0.1 ** p<0.05 *** p<0.01. This table recreates Table 3 for NYC Applications Only. Standard errors in parentheses clustered on chain. 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 4 applications, one white/black pair in each period. Fixed effects can include geographic center, chain, post x chain, and white x 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 (vs regular high-school diploma) and whether they had an employment gap. Box Variation indicates the source of variation in the Box variable – cross-sectional means the variation comes from a comparison of box and no-box stores in the pre-period; 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.

Table A6.4: Effects of BTB on Racial Discrimination, Triple Differences: NYC ONLY

	(1)	(2)	(3)
Box Remover x Post x White	0.034	0.027	0.027
	(0.021)	(0.020)	(0.020)
Post x White	-0.004	-0.002	-0.005
	(0.010)	(0.011)	(0.011)
Box Remover x Post	-0.035	-0.027	
	(0.036)	(0.034)	
Box Remover x White	-0.025*	-0.023*	
	(0.014)	(0.012)	
Box Remover	0.027	0.017	
	(0.029)	(0.028)	
White	0.014	0.012	0.233***
	(0.010)	(0.008)	(0.052)
Post	0.032	0.026	0.193***
	(0.023)	(0.022)	(0.018)
N	6812	8236	8236
Controls	Yes	Yes	Yes
Center FE	No	Yes	Yes
Chain FE	No	No	Yes
Post x Chain FE	No	No	Yes
White x Chain FE	No	No	Yes
Sample	Balanced	Full	Full

Notes: * p<0.1 ** p<0.05 *** p<0.01. This table recreates Table 4 using NYC applications only. 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-period and then removed it due to BTB. The balanced sample indicates the sample where we sent exactly 4 applications, one white/black pair in each period. Controls are whether the applicant had a GED (vs regular high-school diploma) and whether they had an employment gap. Fixed effects can include, chain, post x chain, white x chain, or center, and are included as indicated

A7. Triple Differences with GED and Emp Gap

Table A7.1: Effects of Ban the Box on GED vs High School Diploma, Triple Differences

	(1)	(2)	(3)
Post x Treated x GED	-0.012	-0.010	-0.002
	(0.025)	(0.019)	(0.019)
Post x GED	0.021*	0.009	-0.000
	(0.013)	(0.010)	(0.009)
Post x Treated	0.006	0.014	
	(0.030)	(0.024)	
Treated x GED	0.021	0.021	
	(0.022)	(0.015)	
BoxRemover	-0.003	-0.012	
	(0.029)	(0.026)	
Post	0.004	0.005	0.473***
	(0.016)	(0.015)	(0.174)
N	11188	14637	14637
Controls	Yes	Yes	Yes
Cente rFE	No	Yes	Yes
Chain FE	No	No	Yes
Post x Chain FE	No	No	Yes
White x Chain FE	No	No	Yes
Sample	Balanced	Full	Full

Notes: This table recreates Table 4, substituting GED for White. Standard errors in parenthesis clustered on chain. Dependent variable is whether the application received a callback. The balanced sample indicates the sample of observations where we sent exactly 4 applications, one white/black pair in each period. Fixed effects can include, chain, post x chain, white x chain, or center, and are included as indicated.

Table A7.2: Effects of Ban the Box on Emp Gap vs No Emp Gap, Triple Differences

	(1)	(2)	(3)
Post x Treated x Emp Gap	-0.026	-0.026	-0.022
	(0.024)	(0.020)	(0.021)
Post x Emp Gap	0.014	0.011	0.008
	(0.015)	(0.014)	(0.013)
Post x Treated	0.013	0.022	
	(0.024)	(0.020)	
Treated x Emp Gap	0.013	0.019	
	(0.015)	(0.014)	
BoxRemover	0.002	-0.011	
	(0.030)	(0.027)	
Post	0.008	0.004	0.631^{***}
	(0.019)	(0.017)	(0.151)
N	11188	14637	14637
Controls	Yes	Yes	Yes
Center FE	No	Yes	Yes
Chain FE	No	No	Yes
Post x Chain FE	No	No	Yes
White x Chain FE	No	No	Yes
Sample	Balanced	Full	Full

Notes: This table recreates Table 4, substituting Emp Gap for White. Standard errors in parenthesis clustered on chain. Dependent variable is whether the application received a callback. The balanced sample indicates the sample of observations where we sent exactly 4 applications, one white/black pair in each period. Fixed effects can include, chain, post x chain, white x chain, or center, and are included as indicated.

Appendix A8: Backing out Employer Priors

A8.1 Deriving Equation 3.

In Section 5.3, we presented Equation (3), a linear-in-prior-probabilities equation that helps us to back out employer priors about criminality by race. Here, we discuss further how this simple model is derived, and what assumptions it depends on.

Let $u(x_i, b_i, \epsilon_{ij})$ be the latent utility that store j would derive from interviewing applicant i. x_i is a vector of (non-crime) characteristics of applicant i (which include race); $b_i \in \{c, nc, ni\}$ is the information about the applicant's criminal history, which can be crime, no crime, or no information; ϵ_{ij} is a match-specific shock. The store j interviews applicant i if:

$$u_{ij}(x_i, b_i, \epsilon_{ij}) > u_j^*$$

That is, they interview the applicant if their expected utility with that applicant is above a store-specific threshold.

Let $g_j(x_i)$ be the belief of store j that applicant i has committed a crime, given observable characteristics x_i. Then, for the case where store j has no information about criminal history, the store's beliefs must satisfy:

$$u_{ij}(x_i, ni, \epsilon_{ij}) = g_j(c|x_i)u_{ij}(x_i, c, \epsilon_{ij}) + \left(1 - g_j(c|x_i)\right)u_{ij}(x_i, nc, \epsilon_{ij})$$

Without loss of generality, we can define the match-specific shock ϵ_{ij} so that:

$$u_{ij}(x_i, b_i, \epsilon_{ij}) = \bar{u}(x_i, b_i) + \epsilon_{ij}$$

From here, to derive Equation (3), one must also assume that in the relevant range, ϵ_{ij} is distributed **uniformly**, with parameters A and B, that is**:**

$$\Pr(\epsilon_{ij} < \epsilon) = A_j + B_j\epsilon$$

Then, when employers have information about criminal convictions (i.e., the box):

$$\Pr(callback_{ij}|x_i, c) = \Pr\left(\epsilon_{ij} > u_j^* - \bar{u}(x_i, c)\right) = A_j + B_j\left(u_j^* - \bar{u}(x_i, c)\right)$$

$$\Pr(callback_{ij}|x_i, nc) = \Pr\left(\epsilon_{ij} > u_j^* - \bar{u}(x_i, nc)\right) = A_j + B_j(u_j^* - \bar{u}(x_i, nc))$$

When employers lack information about criminal convictions:

$$\begin{aligned}\Pr(callback_{ij}|x_i, ni) &= \Pr\left(\epsilon_{ij} > u_j^* - \bar{u}(x_i, nc)\right)\\ &= \Pr(\epsilon_{ij} > u_j^* - g(x_i)(\bar{u}(x_i, c) - (1 - g(x_i))\bar{u}(x_i, nc))\\ &= A_j + B_j[u_j^* - g(x_i)(\bar{u}(x_i, c) - \left(1 - g(x_i)\right)\bar{u}(x_i, nc)]\\ &= g(x_i)\left[A_j + B_j\left[\left(u_j^* - \bar{u}(x_i, c)\right)\right]\right] + (1 - g(x_i)\left[A_j + B_j\left[\left(u_j^* - \bar{u}(x_i, nc)\right)\right]\right] \Rightarrow\end{aligned}$$

$$\Pr(callback_{ij}|x_i, ni) = g(x_i)\Pr(callback_{ij}|x_i, c) + \left(1 - g(x_i)\right)\Pr(callback_{ij}|\, x_i, nc)$$

Additionally, if race and crime are additively separable from other characteristics in the store's latent utility, then because the randomization implies that race and crime are independent of other characteristics, $g(x_i)$ can be rewritten as a function of race only, $g(r_i)$.

This is Equation (3) in Section 5.3. We can then use the adjusted results from Figure 2 to back out $g(r_i)$. As described in Section 5.3, this equation implies that (on average) employers' priors for the probability that black and white applicants had felony convictions are 87% and 16%, respectively.

A8.2 Alternative distributions

What if instead the match-specific shocks were drawn from a different distribution? We will not get the linear equation above. However, if we make some additional (strong) assumptions that all stores have the same cutoff u^* and same beliefs $g(r_i)$, that the distribution of match-specific shocks is the same for blacks and whites, and that race is the only observable that employer's use to infer then we can make some progress.

Given these assumptions, our model has the following parameters: $u^*, \bar{u}(w,nc), \bar{u}(b,nc), \bar{u}(w,c), \bar{u}(b,c), g(w), g(b)$, and the parameters of the chosen distribution for the match specific shocks.

Given the parameters of the distribution and normalizing $\bar{u}(w,nc) = 0$ so that the utility of candidates by race and crime status are all *relative* to white candidates with no felony, then we will have 6 parameters in our model: $u^*, \bar{u}(b,nc), \bar{u}(w,c), \bar{u}(b,c), g(w), g(b)$. And from our results, we have 6 moments (using the results from Figure 2):

Table A8.1: Available moments from the data (from Figure 2)

$\Pr(callback_{ij} \mid w,c)$	0.087
$\Pr(callback_{ij} \mid w,nc)$	0.141
$\Pr(callback_{ij} \mid b,c)$	0.080
$\Pr(callback_{ij} \mid b,nc)$	0.134
$\Pr(callback_{ij} \mid w,ni)$	0.135
$\Pr(callback_{ij} \mid b,ni)$	0.09

Notes: These numbers come from Figure 2, which shows raw callback rates by period and race, and by criminal record status in the pre-BTB period, for box-remover stores in the balanced sample. Because the post-BTB period had a higher callback rate across the board, we subtract out the 1.6pp secular increase that we observed at employers whose applications were unaffected by BTB from the post-period callback rates that we use to estimate the no information moments in the table.

Thus, we can use these 6 moments to back-out the 6 parameters of the model, including the desired priors $g(w)$ *and* $g(b)$, under several different distributions for the match-specific shocks.

If the match-specific shocks are distributed as a normal, with the above assumptions, we back out very similar priors to the uniform case: the prior probability that a black applicant has a criminal record is 85% and that the prior probability a white applicant has a criminal record is 14%. This remains true for the log normal distribution, and the beta distribution (under a large range of choices for alpha and beta). [50]

[50] Because expected utilities are preserved under linear transformations, the choice of mean and variance for the normal distributions will not change the priors, we can normalize them to whatever we choose. Changing the beta distribution

Thus, under the simplified and arguably strong assumptions laid out above, the distribution of match-specific shocks that we choose matters little in terms of getting a difference in priors that is quite large, on the order of 68-72 percentage points.[51]

The assumption that all stores have the same cut-off/utility serves a similar purpose as the uniform distribution assumption does above; it means that we do not have to be able to observe the dispersion in the utilities/cut-offs across stores. Once we relax this assumption, or the assumption that the match-specific shock distributions have the same parameters for blacks and whites, then solving this problem becomes more complicated. It may be possible, in theory, to rationalize any set of priors, and this exercise thus cannot be seen as conclusive. However, under various sets of strong but common modeling assumptions, we get priors on the probabilities that white and black candidates have criminal convictions that are very far apart.

parameters does not result in a linear transformation, however we calculated priors under a very wide range of alpha and betas (.5/.5; 100/100; 1/100; 100/1) and calculated very similar priors

[51] Full equations and simulations for distributions other than uniform available on request.