OK Boomer? Burden of Proof Law and Age Discrimination

Authors:

Scott Barkowski, PhD
Assistant Professor
Department of Economics
University at Buffalo
scottjb@buffalo.edu

Joanne Song McLaughlin, PhD
Associate Professor
Department of Economics
University at Buffalo
jsmclaug@buffalo.edu

December 2024

PRELIMINARY; DO NOT CIRCULATE

ABSTRACT

In this study, we study whether anti-age discrimination legislation can protect older workers from discrimination in the American labor market. Via extensive review of court cases in every state, we document changes over the last fifteen years in state-level age discrimination laws, and then estimate their effects on employment outcomes for older workers. Using a synthetic difference-in-differences approach and data from the Current Population Survey, we find that women ages 50 to 75 are about one percentage point less likely to be retired in response to the law changes. We also find an increase of similar magnitude in employment, though at the lower, ten-percent level of statistical significance. These effects appear to be driven most strongly by those ages 71 to 75, the oldest women in our sample. For men, we find employment among those 60 to 75 also increases by about one percentage point, but again this estimate is also only significant at the ten percent level. Men ages 66 to 70 appear to be driving this finding. Our results are consistent with age discrimination still affecting some workers in the modern economy, and that policy can provide protection and extend lives, particularly those of women.

KEYWORDS: Retirement; employment; labor markets; age discrimination; Age Discrimination in Employment Act (ADEA); law and economics; aging; age discrimination legislation; intersectional discrimination

Introduction

The Age Discrimination in Employment Act of 1967 (ADEA) has provided federal age discrimination protection in American labor markets for almost 60 years, yet surveys still show that a substantial share of older workers believe there is age discrimination in the labor market. For example, a late-2023 survey of individuals 50 and older who were in the labor force found that nearly two-thirds of respondents believe that workers face age discrimination today, and one-fifth claim to have been a victim of it since turning 40 (Choi-Allum 2024). This raises questions on how extensive age discrimination is in the modern labor market and whether a policy like the ADEA can protect older workers from it.

Answering such questions is difficult to do credibly. Comparing younger and older workers does not necessarily answer the question because age is not the only difference between these groups. The best research on these topics in the USA tends to take one of two approaches: estimate changes in employment outcomes using variation in state-level policy (e.g., Neumark and Stock 1999; Adams 2004) or perform a correspondence study in which fictitious resumes are sent to employers and compare call-back rates (e.g., Lahey 2008; Riach and Rich 2010; Neumark, Burn, and Button 2019). The strength of policy-based research is that it can make use of government-collected data that measures outcomes like employment, retirement, and unemployment. Furthermore, such data sources can often reflect the entire economy, not just portions of it. Correspondence studies' major strength is randomized assignment, which provides unmatched internal validity. Both approaches have important weaknesses, however. Correspondence studies are limited in what outcomes they can measure. They can measure call-back rates accurately, but it is unclear if reduced call-back rates translate to actual changes in employment outcomes. For example, if only the best workers are being called back, it is possible that those called back convert interviews to offers at high rates. If conversion rates are high enough, it is possible there might not be large differences in hiring by age despite the call-back rate difference. Additionally, as they are focused on employers, correspondence studies cannot say anything about the perspective of workers. Policy-based research can address many of the shortcomings of correspondence studies measuring outcomes like employment and retirement directly. Moreover, it can use surveys of individuals to bring in the perspective of workers directly. Policy-based research's major weakness is that after the ADEA was instituted, there was a period of over four decades where there were no major policy changes. This meant that there were few opportunities to use policy changes to try to answer these questions. The literature that does exist is based on policy changes from before the ADEA became law, meaning we have no literature using policy changes to

complement work from correspondence studies that reflects the more recent state of the labor market.

The extended period without age discrimination policy changes ended in 2009 when a surprise Supreme Court ruling materially raised the threshold for proving age discrimination in lawsuits under federal law. The Supreme Court ruled in *Gross v. FBL Financial Services, Inc.* ("*Gross*") that plaintiffs alleging age discrimination under the ADEA could not use a "mixed-motives" framework in their claims. This meant that age discrimination could only be found if age discrimination was the causal reason for an adverse employment event, and not if discrimination was only part of the reason (a "motivating factor") for the adverse event.[1] Once federal law changed, the new threshold became the prevailing precedent for age discrimination cases. However, state-level law is not always the same as federal law, and cases began to filter through courts evaluating age discrimination claims made under state laws. Over the 2009 to 2020 period, courts have ruled that thirteen states have laws that allow the original, lower proof threshold in age discrimination cases. That is, they allow for mixed motive claims. These changes in state-level law represent the most recent variation in age discrimination policy in the USA.

In this paper we use this new state-by-year variation to estimate the effect of age discrimination legislation on employment outcomes for older workers. Using a synthetic difference-in-differences (SDD) approach and data from the Current Population Survey (CPS), we find that women ages 50 to 75 are about one percentage point less likely to be retired in response to the state-level law changes. We also find an increase of similar magnitude in employment, but only at a ten-percent level of statistical significance. These changes appear to be driven most strongly by the oldest women in our sample, those ages 71 to 75. For men, we find employment among the 60 to 75 group also increases by about one percentage point, but, as for women, it is only significant at the ten-percent level. This appears to be driven by men ages 66 to 70. These results show that the laws do appear to make an impact on labor market outcomes for some older workers.

We argue that this paper makes three important contributions to the literature on age discrimination. First, we provide evidence consistent with age discrimination in the labor market, and with policy having some success addressing it. This evidence comes from outcomes that cannot be studied in correspondence studies, including employment, retirement, and being out of the labor force. Second, we document the full history of state-level changes in age discrimination law with respect to the use of mixed motives

[1] This change applied only to age discrimination law, not discrimination on other characteristics, like race and sex, which maintained the lower, mixed-motives proof threshold.

since the 2009 *Gross* ruling through 2020. This adds a source of policy variation in age discrimination law to the literature that is far more recent than was previously available. Since these law changes occurred via case law, documenting them required a thorough review of written opinions from court rulings across all states. Such reviews are time and resource intensive, which underscores the importance of this contribution, and explains (at least in part) why case law reviews are uncommon in economic research. Lastly, we provide evidence from our real-world application that SDD outperforms the more popular methods of difference-in-differences and synthetic control. In our experience, applied researchers seem to have an apprehension towards "synthetic" methods, but our evidence suggests (and our view as researchers is) that applied researchers should be making more use of SDD in applications where difference-in-differences and synthetic control are typically applied.

Legal Background and Significance

At the federal level, Title VII of the Civil Rights Act of 1964 (Title VII) protects workers against employment discrimination based on race, religion, national origin, color, and sex.[2] In contrast, age discrimination is prohibited by a separate law, the Age Discrimination in Employment Act of 1967. For decades, the two laws were widely viewed in the legal community as having the same threshold to prove discrimination, and courts had applied interpretations of Title VII to the ADEA.[3] In 2009, however, the Supreme Court highlighted language differences between the two laws, arguing that they implied different thresholds of proof. The court ruled that, under federal law, plaintiffs in age discrimination claims must prove that age was *the determining factor* ("but-for" cause) in adverse employment decisions rather than *a motivating* factor ("mixed-motive").[4] This "but-for" causation standard is stricter than a "mixed-motive" standard, implying that discrimination was the causal factor, so the ruling effectively made it more difficult for workers to prove age discrimination in the workplace.

Before *Gross*, courts had applied a mixed-motive framework to age discrimination cases based on *Price Waterhouse v. Hopkins* (1989)[5], a Title VII case. Under the Price Waterhouse mixed-motive framework, once the plaintiff proved age played a "motivating part" in the

[2] Sexual orientation and transgender status are covered under Title VII through the Supreme Court ruling in *Bostock v. Clayton County*, 590 U.S. 644 (2020).
[3] *Gross v. FBL Financial Services, Inc.,* 557 U.S. 167, 181 (2009) (Stevens, J., dissenting)
[4] *Gross v. FBL Financial Services, Inc.,* 557 U.S. 167 (2009)
[5] *Price Waterhouse v. Hopkins*, 490 U.S. 228 (1989)

adverse employment decision, the burden of proof shifted to the employer (the defendant). For the employer to prevail, it had to show that it would have made the same adverse employment decision had age of the plaintiff not been considered. In 1991, partly in response to the Price Waterhouse's decision, Congress amended Title VII and explicitly allowed the mixed-motive framework in Title VII cases.[6] This made mixed motives a clear plaintiff strategy in discrimination cases covered by Title VII, such as race and sex. In a fateful step, though, Congress did *not* amend the ADEA as it had Title VII. For the next (nearly) two decades, though, courts continued to allow the mixed-motive framework in ADEA discrimination claims.

When the *Gross* case came before the Supreme Court, the key issue at hand was whether direct evidence was required to allow for judges to instruct juries that they could use the mixed-motive framework in deciding a case. Instead of deciding on this issue, though, the Court ruled that a mixed-motive framework is not allowed *at all* under the ADEA, and the plaintiffs must show that the alleged discriminatory decision would not have been taken "but for" the plaintiff's age. This decision fundamentally changed the burden of proof structure of all age discrimination claims. The rationale for the decision was that since Congress decided not to amend the ADEA in 1991 when it amended Title VII, it must not have intended for mixed motives to be allowed under the ADEA. This received tremendous attention and sparked uproar in the media and among policy, academic, and legal practitioners. Many claims were made that it would be too hard to prove age discrimination, making the ADEA ineffective at protecting older workers.[7] This concern led to policymakers proposing bills to amend the ADEA multiple times, but so far they have failed to pass them.[8]

State Case Laws

[6] The Supreme Court clarified further in 2003 by unanimously deciding that the 1991 amendment does not require *direct* evidence to obtain a mixed-motive instruction (Dessert Palace, Inc. v. Costa).

[7] E.g., see https://www.latimes.com/archives/la-xpm-2009-jun-19-na-court-age-bias19-story.html (viewed November 27, 2024); https://www.nytimes.com/2013/08/27/booming/for-laid-off-older-workers-age-bias-is-pervasive.html?pagewanted=2&hp (viewed November 27, 2024); https://www.theindianalawyer.com/articles/44581-as-age-discrimination-law-turns-50-bias-becomes-subtler-harder-to-prove (viewed November 27, 2024).

[8] U.S. Congress, S. 2189: Protecting Older Workers Against Discrimination Act, 112th Cong., 2012, https://www.congress.gov/bill/112th-congress/senate-bill/2189; U.S. Congress, *H.R. 2062: Protecting Older Workers Against Discrimination Act of 2021*, 117th Cong., 2021, https://www.congress.gov/bill/117th-congress/house-bill/2062; U.S. Congress, S. 1030: Protecting Older Workers Against Discrimination Act, 118th Cong., 2023, https://www.congress.gov/bill/118th-congress/senate-bill/1030.

Beyond its impact on federal law, the decision in *Gross* also influenced how state courts interpreted their state laws, a key aspect of our identification strategy. We thoroughly reviewed case law of each state to identify cases that established the mixed-motive framework as a proof standard. If state courts ruled that a mixed-motive framework is allowed, then in these states, age discrimination lawsuits become easier to win (i.e., the state's laws become "stronger").[9] For states with age discrimination laws, state courts' rationale broadly depends on two questions: 1) what is the structure of state anti-discrimination laws (i.e., is age discrimination governed by a separate law?)?, and (2) what is the role of federal law in interpreting state laws? The answers to these two questions are not mutually exclusive and do not decisively establish whether a mixed-motive framework is allowed, but they guide courts in interpreting their state laws.

For example, in Alaska, both age discrimination and other discrimination are prohibited under the same statute. Given this, the court declined to follow *Gross* because, unlike federal law, there is no difference between the age discrimination law and other discrimination laws.

> *But while we look to federal discrimination law jurisprudence generally, [AK statute] is intended to be more broadly interpreted than federal law to further the goal of eradication of discrimination. The Supreme Court's holding in Gross substantially relied on the differences between the ADEA and Title VII—differences that do not exist in Alaska's anti-discrimination law. […] We decline to follow Gross and we expressly hold that a plaintiff can bring a mixed-motive age discrimination claim under [AK statute]."*

However, in Arizona, even though all protected groups are covered under one statute, the court emphasized that the interpretation of state law follows the counterpart of the federal law, therefore, the court ruled that the Arizona age discrimination law follows *Gross*.

> *Cases analyzing the Age Discrimination in Employment Act of 1967 […] are relevant in interpreting the comparable [Arizona Civil Rights Act] provisions on age discrimination.*

Compiling state age discrimination case laws requires careful reading of court opinions and judicious decisions on which case to use. Our objective is to identify state court

[9] Generally, in these cases, if the employer shows that it would have made the same decision had age not been considered, the remedies are limited to declaratory relief, injunctive relief, attorney's fees, and direct costs. Although employees are not awarded damages or issued a corrective reinstatement, we still view this as stronger state laws because these rulings are made to deter and prevent discrimination.

decisions that will influence the firms' behavior possibly through channels such as alerting firms' attorneys about the notable changes and sending legal updates to their clients. The specific taxonomy of which cases to use is guided by the framework discussed in (Morriss 1995).

As we described above, the state case law change we aim to document is the allowance of mixed-motive in age discrimination claims. To categorize these changes, first, we rely on Neumark and Song (2013) to classify whether each state has state age discrimination law, determine the structure of its anti-discrimination laws, and analyze the definition of unlawful employment practice. Second, we identify age discrimination cases that are brought under each state law, focusing on claims addressed after Gross. Third, we analyze court opinions and identify the first major higher-court decisions or federal courts' decisions that signal the adoption of a mixed-motive framework in state age discrimination claims. We also count federal courts' decisions on state laws because allowing a mixed-motive framework concerns the behavior of firms.[10] For states that have a single law that governs all types of discrimination claims, we also analyze their rulings on other types of discrimination and their rationale. For states that have one law that governs multiple categories and has allowed a mixed-motive framework, if we do not find clear evidence that the federal laws are relevant in interpreting the comparable provision, we classify the state as allowing a mixed-motive framework in age discrimination claims. However, in cases like these, we cross-check with subsequent related court rulings and secondary sources. [11] We also count lower court decisions if they are not reversed or no further decisions are made. If we find cases that explicitly rule that mixed-motive is not allowed, obviously, we classify this state as not allowing a mixed-motive framework. If a state does not have any state laws or has not had any cases, we also consider this state as not allowing a mixed-motive framework.[12] Since case selection determines the timing of treatment, we dedicated considerable time and effort to this step, cross-checking with other resources. All the details on how we code are explained in the Appendix.

[10] (Morriss 1995) suggests that researchers should count federal court decisions when they concern the behavior of the public rather than state courts' behavior.

[11] For example, we classify California as allowing for a mixed-motive following *Harris v. City of Santa Monica,* 56 Cal.4th 203 (Cal. 2013). Although this was pregnancy discrimination, it was a California Supreme Court's decision on mixed-motive under California's Fair Employment and Housing Act and the subsequent age discrimination case cites this case. See *McInteer v. Ashley Distrib. Serv., Ltd.*, 40 F.Supp.3d 1269 (C.D. Cal. 2014)

[12] For example, we code Arkansas not allowing mixed-motive because the Arkansas Civil Rights Act of 1993 does not cover discrimination based on age.

Empirical Strategy

Our empirical strategy relies on the timing of the court decisions establishing mixed motive as a valid plaintiff strategy in an age discrimination case under state-level law. These decisions effectively increased the likelihood of a plaintiff winning an age discrimination case. We, therefore, take individuals living in states that allowed mixed motives as our treatment group and states that have not (through 2020) as our control group. States are considered treated for all years once a court ruled mixed motives were permitted (except as noted below). Embedded in this strategy is the assumption that the federal standard established by Gross that mixed motive was not allowed in age discrimination was the presumed standard by employers and employees in all states before the court clarification that the lower standard applied. Thus, before a state clarified that mixed motive was allowed in a state, we assume that employers, workers, and lawyers considered mixed motive strategies to be against the state's law.

We focus on four primary outcome variables of interest: dummy variables indicating whether someone is retired, employed, unemployed, and not in the labor force. In our primary estimation method, these variables are aggregated to the state level to create a balanced, state-level panel. These outcomes are represented by $y_{s,t}$, where s indicates the state and t the year. We denote the treated states by $mixed_{s,t}$ using the same indices. We estimate two-way fixed-effects models, and their event-study versions, that take the following form:

$$y_{s,t} = mixed_{s,t}\beta + \alpha_s + \gamma_t + u_{s,t}, \quad (1)$$

where α_s and γ_t represent fixed effects for states and years, respectively, and $u_{s,t}$ is the error term. Our parameter of interest is β, represents the average treatment effect on the outcome of a reduction in the state-level legal threshold for proving age discrimination in a lawsuit. That is, it is the effect of making it easier to win an age discrimination case under state law.

As the courts have frequently noted, the goal of the law in allowing mixed motives strategies was to reduce age discrimination in the labor markets. This could only occur if two assumptions hold: 1) age discrimination is, in fact, prevalent in the market, and 2) the law is effective in changing the behavior of market participants. If these assumptions do hold, then a simple model of labor supply and demand can provide helpful theoretical predictions for the sign of the coefficient β. An increase in the cost of discriminating against older individuals (as implied by 2) would cause labor demand for individuals covered by the law (those 40 or older) to increase (shift rightward). Similarly, a reduction in

discrimination would make employment relatively more attractive for workers, implying an increase in labor supply (a rightward shift). Both of these shifts would be consistent with an increase in employment, or a positive value for β when employment is the outcome. A relative increase in labor supply also implies a reduction in retirement and not being in the labor force, and, hence, a negative β value for these outcomes. In contrast, the prediction for the coefficient when unemployment is the outcome is ambiguous, since individuals may leave that state for employment but also may enter it after being retired or out of the labor force, seeking employment opportunities due to the decrease in discrimination. Thus, a positive β estimate for employment and negative ones for retirement and being out of the labor force would be evidence consistent with a reduction in discrimination in the labor market caused by the change in the law. Null estimates would suggest that either age discrimination is not prevalent, or that the law is not effective in changing participant behaviors. The estimate for unemployment will indicate whether flows into employment dominate those from back into the labor force (or vice versa) or whether they approximately balance each other out.[13]

Data

We estimate equation (1) using a sample from the Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC) (Flood et al. 2023). The CPS has a large sample size that covers every state and Washington DC, and its labor market questions have remained consistent over an extended period of time, making it an ideal choice for our study. We include data for the years 2001 through 2020 and exclude individuals serving in the armed forces. Our primary sample of interest consists of individuals aged 50 to 75, being the individuals covered by age discrimination laws and those most likely to experience age discrimination (assuming it exists in the labor market). We calculate weighted sample averages using ASEC sampling weights within year-by-state (plus Washington DC) cells to aggregate our micro-data into a panel that can be used for estimation.

[13] These predictions would be less certain in a model that included time. If workers expect that a discrimination reduction will persist, they might exit employment and enter unemployment in search of a better employer match. Similarly, better future opportunities would make additional schooling more attractive, potentially leading to exit from current employment and entrance into being in school and out of the labor force. Neither of these seem likely to reduce employment, however, since the core of this argument for both is that employment as a state is more attractive than before, and since, in the case of education, the individuals covered by the law are older and a group that is unlikely to return to school in great numbers.

In 2014 the CPS ASEC underwent an experimental redesign that implemented new income questions on a 3/8ths portion of the full sample (IPUMS CPS, n.d.). The remainder 5/8ths of the sample received the old questions. For our purposes, the redesign of the income questions is not relevant, but the weights released by the Census Bureau for the 2014 survey—which *are* used in our analysis—were calculated differently from other years. To allow for national-level statistics to be calculated using both portions, the Census Bureau's ASEC sampling weight is nationally representative for both portions separately, so the weights sum across all observations to twice the national population level. Since we use the sampling weights to collapse our data to a state-by-year panel for our analysis, we only use the larger, 5/8 portion of the 2014 ASEC in our sample, which, when weighted, represents its own nationally representative sample.

Table 1 presents descriptive statistics (calculated using sampling weights) for individuals in treated states and those in states that are never treated (our control group). We also include corresponding estimates for both men and women, separately. All statistics are for the year 2009, just before the Gross decision and the first state decisions to allow mixed motives. As can be seen in the table, there are important differences across the groups for several characteristics, which underscores the importance using an empirical strategy that corrects for state-level differences between the groups.

Estimation

The primary method we rely on to estimate equation (1) is synthetic difference-in-differences (SDD) (Arkhangelsky et al. 2021), which we implement using Stata software and the "sdid" and "sdid_event" add-on packages (StataCorp 2024; Clarke et al. 2023; Ciccia 2024). We calculate standard errors that are clustered at the state-level for all estimates (Bertrand, Duflo, and Mullainathan 2004).

The SDD method builds on the more widely known methods of difference-in-differences (DD) (and its two-way fixed-effects generalization) and synthetic control (SCM) CITE ORIGINAL SCM PAPER. Like DD, SDD accounts for state-level (more generally, unit-level) unobserved factors via inclusion of state-level fixed effects. Like SCM, SDD weights each unit in the estimation process, thereby creating a synthetic control group where the differences between the treatment and control group pre-period outcome values are minimized. Unlike SCM, however, SDD adds a time-period weight to the process that also minimizes differences between pre- and post-period control group outcome values. That is, SDD includes both unit and time weights in its estimation procedure.

Specifically, the SDD method chooses the parameters β, α_s, and γ_t for equation (1) by solving the following least squares problem implied by equation (1):

$$\Sigma_{s=1}^{S}\Sigma_{t=1}^{T}\left(y_{s,t} - mixed_{s,t}\beta - \alpha_s - \gamma_t\right)^2 \widehat{\omega_s}\widehat{\lambda_t}, \quad (2)$$

where S and T are the total number of states and time periods (51 and 20 in our application), respectively. The weights $\widehat{\omega_s}$ and $\widehat{\lambda_t}$ in equation (2) are chosen to reduce the impact of control units that are dissimilar from the treatment group during the pre-period, and pre-period years that are dissimilar from the post-period years (based on a comparison of control group values). (Arkhangelsky et al. 2021) provides the specific details on the calculation of the weights. Since the SDD method is designed to create a strong control group based on pre-period outcome values, we do not include controls besides the state and year fixed effects when we use the SDD method. In essence, we view the weights in SDD as serving a similar role to what controls variables contribute in a DD approach.

Equation (2) shows that SDD is a generalization of DD and SCM.[14] Conceptually, SDD's setup has two important advantages over these other methods. First, as noted by (Arkhangelsky et al. 2021), the DD approach uses its unit and time fixed effects to address any sources of bias, while the SCM uses its time fixed effects and unit weights to address potential bias. SDD, in contrast, uses both time and unit fixed effects *and* both time and unit weights. Thus, SDD has more tools to address sources of bias than these other methods. Second, in relation to the SCM specifically, SDD adds a regularization term in the unit-weight minimization problem, which penalizes putting more weight on fewer states, and only requires the trends to be parallel instead of being right on top of each other. In practice, this means that the SDD method is less likely to put all the weight on only a few control states, thereby forcing more states to contribute to the synthetic control. Intuitively, this suggests that SDD would be relatively robust to shocks that affect only a few control states. The SCM, on the other hand, would be more heavily affected if heavily weighted states happened to experience a random, post-period shock.

As has been noted in a significant recent literature, treatment effect heterogeneity can pose problems for two-way fixed-effect designs, and staggard adoption DD designs like ours are a canonical case where heterogeneity is presumed to exist (Baker, Larcker, and Wang 2022; de Chaisemartin and D'Haultfœuille 2023; Roth et al. 2023). SDD handles staggard adoption designs like ours in a manner similar to other approaches suggested in the literature, such as by (Callaway and Sant'Anna 2021). That is, for each cohort of states adopting mixed motives in a given year, the process calculates a separate SDD, using only

[14] The process that determines the SDD weights is also a generalization of that used in the SCM.

the states changing their law that year as the treatment group and those states that are never treated as the control group. The estimates for each separate adoption cohort are then aggregated via weighted average, where the weight is calculated as the adoption cohort's number of post-period years in effect as a share of all post-period years in effect across all adoption cohorts (Arkhangelsky et al. 2021; Clarke et al. 2023). Thus, our approach fully accounts for potential heterogeneity across adoption cohorts.

With respect to inference, (Arkhangelsky et al. 2021) show that the SDD estimator is normally distributed asymptotically. They study three methods to estimate the asymptotic variance, including bootstrap, jackknife, and placebo inference approaches. We estimate our standard errors using the placebo method (with 500 repetitions) for all regressions since simulations (Arkhangelsky et al. 2021) carried out suggest it performed well compared to the other methods. Additionally, the placebo method can also be used for the SCM approach. As we note below, we also compare our estimates to ones implemented using the SCM. The placebo method allows us to use the same variance estimation method for both approaches, thereby increasing the comparability of our results across the two methods.

The placebo method works by randomly assigning treatment to states in the sample, then calculating the estimator using the randomly assigned groups numerous times to obtain a distribution of estimates. The variance is then estimated from the distribution of placebo estimates. An important aspect of this process is that only the original control states—the actual states that never allowed mixed motives—are used in the placebo process. This implies an assumption that the control units have the same variance structure as the complete distribution of states. Thus, by choosing this method we implicitly assume a form of homoskedasticity across units, an important potential weakness in our application. That said, this process does allow for serial correlation within states (i.e., clustering) since the placebo treatment assignments are made at the state-level (and the full panel of years for a state all receive the same assignment). So, to the extent the serial correlation for the full distribution can be estimated by the never treated units only, this will give correct inference.

Synthetic difference-in-differences outperforms other methods in our context

To augment the conceptual arguments we make above in support of our use of SDD, we next provide empirical evidence that SDD outperforms DD and SCM in our particular application. This evidence comes from comparison of estimates from the three methods

for four placebo outcomes that we expect the law changes would not affect. These outcomes include: 1) being White race, 2) being Black race (where multi-race categories are not considered black or white), 3) age, and 4) sex. We interpret statistically significant results for these outcomes as evidence that a method is having difficulty addressing sources of bias in our application. As with our main results, we do not include any controls other than the state and year fixed effects when we estimate via SDD, and for similar reasons we also do not include additional controls aside from the year fixed effects for SCM. For DD we include the state and year fixed effects as well as sets of fixed effects for all values of age, education, race, Hispanic ethnicity, and sex (where sex and race controls are excluded when those variables are used as outcomes).

Table 2 presents results from estimating β via each method on our main sample of individuals ages 50 to 75, and for men and women separately. The top panel shows that when SDD is the estimation method, there are no statistically significant estimates. In contrast, for both DD and SCM we obtain estimates that are generally larger in magnitude, and in the case of the age outcome, statistically significant (at 5 percent). We supplement Table 2 with event study plots for our full sample in Online Appendix Figure 1. These show that, in addition to the issues with age, we obtain statistically significant estimates for some post periods via the DD and SCM approaches. Specifically, for DD white race and sex exhibit these issues, and for SCM, sex does. Meanwhile, the event study plots for SDD show no significant (at 5 percent) estimates for any post-period across all outcomes. Overall, we find that DD and SCM face some problems addressing unobserved factors as well as SDD, and so we use SDD as our primary estimation method for our analysis.

In their article developing SDD, (Arkhangelsky et al. 2021) perform two Monte Carlo studies based on simulated data to compare their estimator to DD and SCM (and other related methods), showing that it outperforms them in most scenarios.[15] Our assessment here also shows SDD outperforming these other methods, but provides an important comparison based on actual CPS data that is not simulated and real state-level policies. We, therefore, view one of the contributions of this study to be providing a real-world policy evaluation example where SDD appears to be more appropriate than DD and SCM.

Results

[15] The data generating process for one of these studies is based on CPS data. They estimate a model and the variance of an error distribution using their CPS data, and then use this model and distribution to generate outcome values.

[UNDER CONSTRUCTION]

Conclusion

[UNDER CONSTRUCTION]

References

Adams, Scott J. 2004. "Age Discrimination Legislation and the Employment of Older Workers." *Labour Economics* 11 (2): 219–41. https://doi.org/10.1016/j.labeco.2003.06.001.

Arkhangelsky, Dmitry, Susan Athey, David A. Hirshberg, Guido W. Imbens, and Stefan Wager. 2021. "Synthetic Difference-in-Differences." *American Economic Review* 111 (12): 4088–4118. https://doi.org/10.1257/aer.20190159.

Baker, Andrew C., David F. Larcker, and Charles C.Y. Wang. 2022. "How Much Should We Trust Staggered Difference-in-Differences Estimates?" *Journal of Financial Economics* 144 (2): 370–95. https://doi.org/10.1016/j.jfineco.2022.01.004.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *The Quarterly Journal of Economics* 119 (1): 249–75.

Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics*, Themed Issue: Treatment Effect 1, 225 (2): 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001.

Choi-Allum, Lona. 2024. "Age Discrimination Among Adults Age 50-Plus in the Labor Force." Washington, DC: AARP Research. https://doi.org/10.26419/res.00545.005.

Ciccia, Diego. 2024. "A Short Note on Event-Study Synthetic Difference-in-Differences Estimators." arXiv. https://doi.org/10.48550/arXiv.2407.09565.

Clarke, Damian, Daniel Pailañir, Susan Athey, and Guido Imbens. 2023. "Synthetic Difference In Differences Estimation." arXiv. https://doi.org/10.48550/arXiv.2301.11859.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2023. "Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey." *The Econometrics Journal* 26 (3): C1–30. https://doi.org/10.1093/ectj/utac017.

Flood, Sarah, Miriam King, Renae Rodgers, Steven Ruggles, J. Robert Warren, Daniel Warren, Annie Chen, et al. 2023. "IPUMS, Current Population Survey: Version 11.0." Minneapolis, MN: IPUMS. https://doi.org/10.18128/D030.V11.0.

IPUMS CPS. n.d. "2014 3/8 ASEC File Information." Accessed November 25, 2024. https://cps.ipums.org/cps/three_eighths.shtml.

Lahey, Joanna N. 2008. "Age, Women, and Hiring An Experimental Study." *Journal of Human Resources* 43 (1): 30–56. https://doi.org/10.3368/jhr.43.1.30.

Morriss, Andrew P. 1995. "Developing a Framework for Empirical Research on the Common Law: General Principles and Case Studies of the Decline of Employment-at-Will." *Case Western Reserve Law Review* 45 (4): 999–1148.

Neumark, David, Ian Burn, and Patrick Button. 2019. "Is It Harder for Older Workers to Find Jobs? New and Improved Evidence from a Field Experiment." *Journal of Political Economy* 127 (2): 922–70. https://doi.org/10.1086/701029.

Neumark, David, and Wendy A. Stock. 1999. "Age Discrimination Laws and Labor Market Efficiency." *Journal of Political Economy* 107 (5): 1081–1125. https://doi.org/10.1086/250091.

Riach, Peter A., and Judith Rich. 2010. "An Experimental Investigation of Age Discrimination in the English Labor Market." *Annals of Economics and Statistics*, no. 99/100, 169–85. https://doi.org/10.2307/41219164.

Roth, Jonathan, Pedro H. C. Sant'Anna, Alyssa Bilinski, and John Poe. 2023. "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature." *Journal of Econometrics* 235 (2): 2218–44. https://doi.org/10.1016/j.jeconom.2023.03.008.

StataCorp. 2024. "Stata Statistical Software: Release 18.5." College Station, TX: StataCorp LP.

Tables and Figures

Table 1 Descriptive Statistics [HERE]

Table 2: Comparison of estimates produced by different estimation methods for placebo outcomes

Outcome	Men	Women	All
Method: SDD			
A. Black	0.0024	-0.0012	-0.0003
	(0.0040)	(0.0036)	(0.0035)
B. White	-0.0047	-0.0013	-0.0035
	(0.0054)	(0.0048)	(0.0045)
C. Age	-0.0309	-0.0503	-0.0550
	(0.1122)	(0.1069)	(0.1018)
D. Sex			0.0004
			(0.0021)
Method: DD			
A. Black	0.0032	0.0003	0.0004
	(0.0045)	(0.0048)	(0.0036)
B. White	-0.0090	-0.0027	-0.0035
	(0.0059)	(0.0054)	(0.0051)
C. Age	-0.1958***	-0.0434	-0.1154*
	(0.0688)	(0.0872)	(0.0631)
D. Sex			0.0058
			(0.0040)
Method: SCM			
A. Black	0.0056	0.0053	0.0041
	(0.0055)	(0.0053)	(0.0050)
B. White	-0.0100	-0.0041	-0.0069
	(0.0084)	(0.0075)	(0.0093)
C. Age	-0.2441***	-0.1655**	-0.2212***
	(0.0929)	(0.0803)	(0.0744)
D. Sex			-0.0002
			(0.0018)

Note: statistically significant estimates for two-tailed tests at the one, five, and ten-percent levels are indicated by ***, **, and *, respectively.

Table 3, Part 1: Estimates of equation 1 for retirement and employment

Age Group	Men	Women	All
Outcome: Retirement			
50 to 75	-0.0050	-0.0109**	-0.0066
	(0.0057)	(0.0049)	(0.0043)
50 to 60	-0.0018	-0.0024	-0.0009
	(0.0039)	(0.0037)	(0.0033)
60 to 75	-0.0056	-0.0096	-0.0046
	(0.0080)	(0.0089)	(0.0070)
62 to 65	0.0091	-0.0169	-0.0046
	(0.0141)	(0.0165)	(0.0122)
66 to 70	-0.0138	-0.0116	-0.0110
	(0.0162)	(0.0156)	(0.0119)
71 to 75	0.0042	-0.0162	-0.0069
	(0.0160)	(0.0123)	(0.0112)
Outcome: Employment			
50 to 75	0.0098*	0.0095*	0.0085*
	(0.0056)	(0.0058)	(0.0050)
50 to 60	0.0105	0.0005	0.0052
	(0.0072)	(0.0084)	(0.0071)
60 to 75	0.0123*	0.0027	0.0051
	(0.0066)	(0.0081)	(0.0061)
62 to 65	0.0018	0.0094	0.0045
	(0.0134)	(0.0146)	(0.0114)
66 to 70	0.0205	0.0185	0.0179*
	(0.0133)	(0.0139)	(0.0092)
71 to 75	-0.0052	0.0200**	0.0119
	(0.0170)	(0.0099)	(0.0100)

Table 3, Part 2: Estimates of equation 1 for not in the labor force and unemployment

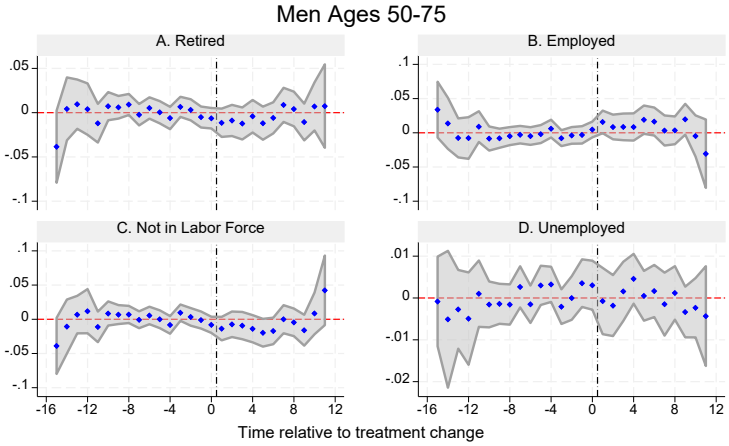
Age Group	Men	Women	All
Outcome: Not in the Labor Force			
50 to 75	-0.0092	-0.0090	-0.0075
	(0.0059)	(0.0058)	(0.0051)
50 to 60	-0.0072	-0.0003	-0.0040
	(0.0063)	(0.0079)	(0.0059)
60 to 75	-0.0117*	-0.0015	-0.0037
	(0.0070)	(0.0084)	(0.0064)
62 to 65	0.0008	-0.0075	-0.0024
	(0.0129)	(0.0157)	(0.0115)
66 to 70	-0.0207	-0.0145	-0.0160*
	(0.0130)	(0.0139)	(0.0095)
71 to 75	0.0012	-0.0239**	-0.0147
	(0.0166)	(0.0099)	(0.0099)
Outcome: Unemployment			
50 to 75	0.0001	-0.0000	0.0005
	(0.0021)	(0.0015)	(0.0013)
50 to 60	-0.0017	0.0005	-0.0002
	(0.0029)	(0.0027)	(0.0020)
60 to 75	-0.0009	-0.0002	-0.0008
	(0.0021)	(0.0012)	(0.0012)
62 to 65	-0.0003	-0.0021	-0.0011
	(0.0034)	(0.0028)	(0.0022)
66 to 70	0.0033	-0.0046**	-0.0020
	(0.0031)	(0.0022)	(0.0020)
71 to 75	0.0041	0.0011	0.0022
	(0.0028)	(0.0019)	(0.0016)

Figure 1: Event study estimates for equation 1

1. Full Sample



2. Men



3. Women



Online Appendix

Online Appendix Figure 1: Event study estimates for placebo outcomes by estimation method

